Structural Space Scaling

Kojiro Shojima
(The National Center for University Entrance Examinations)

1. Introduction

Multidimensional scaling (MDS) is a statistical model for plotting items (variables, nodes, units, or subjects) in a multidimensional space and visualizing the inter-item relationships in this space. In MDS, all the items are located in one and only space, and their relationships are considered in that space. However, if this space consists of some subspaces, to show the relationships between the subspaces is effective with regard to understanding the analyzed data. The purpose of this study is to propose structural space scaling (SSS), which is a technique in the context of MDS to describe the relationships among subspaces.

2. Path Diagram

Path diagrams are used to show analysis results in some statistical models such as graphical models, structural equation models, and Bayesian networks. These are very useful when visually expressing the inter-variable relationships. SSS also uses path diagrams to express the inter-subspace structure in the original multidimensional space.

Fig. 1 illustrates an original multidimensional space, that contains eight items: $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_8$. Of these eight, items 1–4 exist in subspace S_1 and items 5–8 exist in subspace S_2. Furthermore, S_1 has dimensions f_{11} and f_{12}, and S_2 has dimensions f_{21} and f_{22}.

If there is a causal relationship such that S_2 is created by S_1, the relationship is expressed as Fig. 2. The four ellipses represent the dimensions making the subspaces, and the rectangles stand for the eight items. The strength of the linear relationship between dimensions f_{11} and f_{12} that make S_1 is expressed as r_1 in the path diagram. In addition, as items 1 and 2 exist in dimension f_{11}, the arrows from f_{11} are pointing toward the two items in the diagram, where $a_1 = |\overrightarrow{O_1X_1}|$ and $a_2 = |\overrightarrow{O_1X_2}|$.

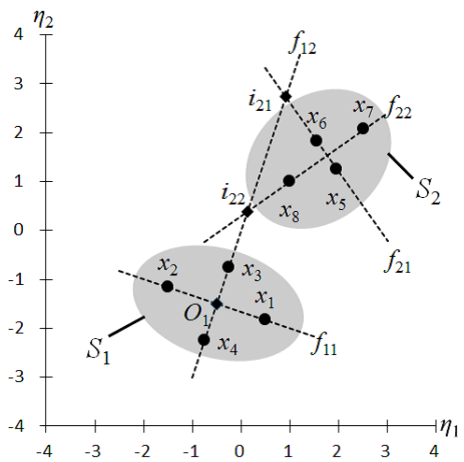

Figure 1: Subspaces in Map

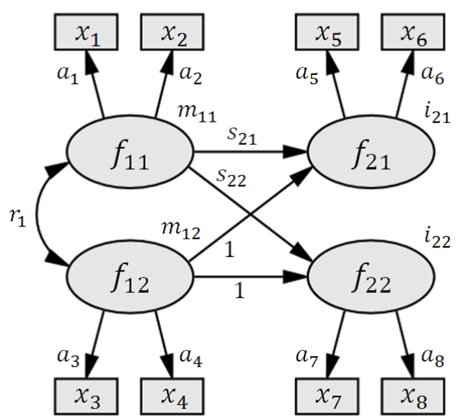

Figure 2: Path Diagram

Furthermore, in the diagram, dimension f_{21} receives two arrows from f_{11} and f_{12}, and the numbers over the two paths are s_{21} and 1, respectively. This means that the slope of line f_{21} is s_{21} from the viewpoint of S_1. In other words, f_{21} is a vector in which f_{11} changes s_{21} when f_{12} runs one unit. Finally, i_{21} denotes the intercept where f_{21} intersects f_{12} when $f_{11} = 0$.

3. Estimation

The matrix representation of the structure shown by Fig. 2 is expressed as follows:

$$\begin{bmatrix} f_{11} \\ f_{12} \\ f_{21} \\ f_{22} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} \\ \\ i_{21} \\ i_{22} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{bmatrix} + \begin{bmatrix} & & & & \\ & & & & \\ s_{21} & 1 & & & \\ s_{22} & 1 & & & \\ a_1 & & & & \\ a_2 & & & & \boldsymbol{O}_{12\times 8} \\ & a_3 & & & \\ & a_4 & & & \\ & & a_5 & & \\ & & a_6 & & \\ & & & a_7 & \\ & & & a_8 & \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{21} \\ f_{22} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} + \begin{bmatrix} f_{11} \\ f_{12} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{bmatrix}. \tag{1}$$

The above equation is simply written as

$$\boldsymbol{t} = \boldsymbol{\mu} + \boldsymbol{\Lambda t} + \boldsymbol{u}, \tag{2}$$

and can be transformed into

$$\boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{\Lambda})^{-1}(\boldsymbol{\mu} + \boldsymbol{u}). \tag{3}$$

Using selection matrix $\boldsymbol{G}$ for taking $\boldsymbol{x}$ out of $\boldsymbol{t}$, we can obtain

$$\boldsymbol{x} = \boldsymbol{Gt} = \boldsymbol{G}(\boldsymbol{I} - \boldsymbol{\Lambda})^{-1}(\boldsymbol{\mu} + \boldsymbol{u}). \tag{4}$$

Then, the Euclidean distance matrix between the items is given by

$$\boldsymbol{\Delta}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \mathrm{diag}(\boldsymbol{xx}')\boldsymbol{1}' - 2\boldsymbol{xx}' + \boldsymbol{1}\mathrm{diag}(\boldsymbol{xx}')'. \tag{5}$$

Therefore, the least squares function to be minimized is obtained as

$$S(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \mathrm{tr}(\boldsymbol{D} - \boldsymbol{\Delta})'(\boldsymbol{D} - \boldsymbol{\Delta}), \tag{6}$$

where $\boldsymbol{D}$ is the data matrix.

4. Discussion

SSS is useful in understanding the inter-subspace relationships if the original space can be divided into some subspaces that contain some causal structures. In addition, path diagrams are good at expressing high-dimensional structures. In MDS, the number of dimensions in the (original) space is usually two or three due to difficulties in visual expression. However, if the data have an essentially high-dimensional structure, the silhouette of the structure projected onto the two- or three-dimensional space does not represent what the structure is really like. However, if such a structure is drawn as a path diagram, we can easily visualize it even if it has four or more dimensions.