

# Two-stage estimation of ATRISCAL coordinates

Kojiro Shojima

(The National Center for University Entrance Examinations)

## 1. Stress Function

This study proposes an estimation framework of asymmetric triangulation scaling (ATRISCAL; Shojima, 2009). ATRISCAL is an asymmetric multidimensional scaling, and the object of the analysis is an asymmetric correct response rate matrix obtained from binary (correct/incorrect) test data.

From Shojima (2010a), the stress function of ATRISCAL is given as

$$F^*(\mathbf{X}) = \frac{F(\mathbf{X})}{T(\mathbf{X})} = \frac{\sum_{c,r}^{n+1} \lambda_{c|r} (p_{c|r} - \pi_{c|r})^2}{\sum_{c,r}^{n+1} \delta_{c|r} \lambda_{c|r} (\pi_{c|r} - \bar{p})^2}, \quad (1)$$

where  $\mathbf{X} = \{x_{rm}\}$  ( $(n+1) \times 3$ ) is a matrix consisting of three-dimensional coordinates of  $n+1$  items, although the coordinates to be estimated are of  $n$  items because the coordinate of the  $n+1$ -th item is fixed.

In addition,  $\pi_{c|r}$  in the stress function is expressed as follows:

$$\pi_{c|r} = \frac{|\overrightarrow{OX_{rc}}|}{|\overrightarrow{OX_r}|} = \frac{\sqrt{|\overrightarrow{OX_r}|^2 |\overrightarrow{OX_c}|^2 - \overrightarrow{OX_r} \cdot \overrightarrow{OX_c}}}{|\overrightarrow{OX_r}| |\overrightarrow{X_r X_c}|} = \frac{\sqrt{|\mathbf{x}_r|^2 |\mathbf{x}_c|^2 - (\mathbf{x}'_r \mathbf{x}_c)^2}}{|\mathbf{x}_r| |\mathbf{x}_c - \mathbf{x}_r|}. \quad (2)$$

Shojima (2010a) proposed a method for minimizing Equation (1) by steepest descent method (SDM). However, it is difficult to prepare an adequate initial value of  $\mathbf{X}$  for starting the SDM. The stress function of ATRISCAL is very complicated; therefore, the estimate of  $\mathbf{X}$  easily falls into a local minima. That is, it is primarily important to determine the initial value of  $\mathbf{X}$  to start exploring the coordinate estimates using the SDM.

Therefore, in this study, a two-stage estimation of the ATRISCAL coordinates is proposed. In the first stage, a simple genetic algorithm (SGA) is used to determine the initial values for the SDM. In the second stage, the final estimate of  $\mathbf{X}$  is obtained by the SDM using the initial values obtained in the first stage.

## 2. First Stage by Simple Genetic Algorithm

### 2.1 Chromosome

The genetic algorithm is a technique used for combinatorial optimization. Let  $\mathbf{X}$  be a chromosome and the  $r$ -th row vector in  $\mathbf{X}$ ,  $\mathbf{x}_r$ , be the  $r$ -th locus. Let us also assume that the number of generations is  $T$  and the population is  $Q$ . In addition, let  $\mathbf{X}^{(t,q)}$  be the chromosome of individual  $q$  ( $= 1, \dots, Q$ ) in generation  $t$  ( $= 1, \dots, T$ ). Of the  $r$ -th ( $= 1, \dots, n$ ) locus of individual  $q$ ,  $\mathbf{x}_r^{(1,q)}$ , the  $x$ - and  $y$ -coordinates are generated from  $U[-1, 1]$ , and the  $z$ -coordinate is generated from  $U[0, 1]$ . However, some item coordinates are fixed to satisfy the condition shown by Shojima (2010a) to avoid spacial indeterminacy.

### 2.2 Selection

Selection is carried out for the  $Q$  individuals in generation  $t$  with respect to the size of the stress function. Let  $p_s$  be the survival rate, and the number of survivals by selection is  $S = \text{Round}(p_s Q)$ . That is, by calculating the stress function values of the  $Q$  individuals, we adopt the individuals with the  $S$  smallest

values of the stress function as the parents for the next  $t + 1$ -th generation, and we discard the remaining  $Q - S$  individuals.

### 2.3 Crossover

During crossover, the  $Q$  individuals in the next  $t+1$ -th generation are produced by crossing the  $S$  survivals in generation  $t$ . However, as an elite-preserving strategy, the individuals with the  $E$  ( $< Q$ ) smallest values of the stress function are used as the individuals in generation  $t + 1$ , without changing their chromosomes. The remaining  $Q - E$  individuals are produced by randomly crossing the chromosomes of the  $S$  survivals in generation  $t$ . That is, the chromosome of individual  $q$  ( $> E$ ) in generation  $t + 1$  is created by

$$\mathbf{X}^{(t+1,q)} = \mathbf{z} \odot \mathbf{X}^{(t,F)} + (\mathbf{1} - \mathbf{z}) \odot \mathbf{X}^{(t,M)} \quad (q = 1, 2, \dots, Q), \quad (3)$$

where  $\mathbf{X}^{(t,F)}$  and  $\mathbf{X}^{(t,M)}$  are the father and the mother of child  $\mathbf{X}^{(t+1,q)}$ , and the parents are randomly chosen from the  $S$  survivals in generation  $t$ . In addition,  $\mathbf{1}$  ( $n \times 1$ ) is a vector in which all elements are 1, and  $\mathbf{z} = \{U|U \sim \text{Bernoulli}(0.5)\}$  ( $n \times 1$ ) is a vector in which the  $r$ -th element is a dichotomous random variable used for choosing either the father's or the mother's element in the  $r$ -th locus.

The process from 2.1 to 2.3 is repeatedly applied through  $T$  generations. Finally, the  $S$  survivals in the last  $T$ -th generation,  $\mathbf{X}^{(T,1)}, \dots, \mathbf{X}^{(T,S)}$ , are adopted as the estimates at the first stage.<sup>1</sup>

### 3. Second Stage by SDM

In the second stage, the final estimates of the coordinates of the  $n$  items are obtained. Here, the  $S$  survivals obtained from the first stage,  $\mathbf{X}^{(T,1)}, \dots, \mathbf{X}^{(T,S)}$ , are used as the initial values for the SDM. That is,  $S$  estimates,  $\hat{\mathbf{X}}^{(1)}, \dots, \hat{\mathbf{X}}^{(S)}$ , can be obtained by applying the SDM with respect to the  $S$  initial values. The first derivatives of the stress function required for the SDM are proposed by Shojima (2010a). Finally, the estimate of the ATRISCAL coordinates,  $\hat{\mathbf{X}}$ , is determined as the estimate with the smallest value of the stress function among  $F^*(\hat{\mathbf{X}}^{(1)}), \dots, F^*(\hat{\mathbf{X}}^{(S)})$ . That is, the estimate is given by

$$\hat{\mathbf{X}} = \arg \min_{s \in S} F^*(\hat{\mathbf{X}}^{(s)}). \quad (4)$$

#### References

- Nakayama, A. & Yokoyama, S. (2010) Multidimensional scaling *The 13th Spring Seminar of The Behaviorometric Society of Japan*, (pp.40-57).
- Okada, A., & Imaizumi, T. (1987) Nonmetric multidimensional scaling of asymmetric proximities. *Behaviorometrika*, **21**, 81-96.
- Shojima, K. (2009) Asymmetric triangulation scaling: A multidimensional scaling for visualizing inter-item dependency structure. *Proceedings of The 7th annual meeting of the Japan Association for Research on Testing*, 88-91.
- Shojima, K. (2010a) ATRISCAL coordinate estimation by steepest descent method. *Proceedings of The 8th annual meeting of the Japan Association for Research on Testing*, in press.
- Shojima, K. (2010b) *Exametrika 4.3*. ([www.rd.dnc.ac.jp/~shojima/exmk/index.htm](http://www.rd.dnc.ac.jp/~shojima/exmk/index.htm)).

(shojima@rd.dnc.ac.jp)

<sup>1</sup>In the software for ATRISCAL, Exametrika (Shojima, 2010b), the parameters in the SGA at the first stage are set to as  $(T, Q, p_s, E) = (20, 1000, 0.1, 1)$ .