Ann. Inst. Statist. Math.
31 (1979), Part A, 115-124

OPTIMAL SCALING FOR ARBITRARILY ORDERED CATEGORIES

YUTAKA TANAKA

(Received May 30, 1978; revised Jan. 20, 1979)

Abstract

The methods of optimal scaling are usually formulated as the maximization problem of a ratio of quadratic forms, and the optimal scores are obtained by solving an eigenequation. However, there sometimes exist order relations among categories. For such cases, Bradley, Katti and Coons [2] proposed an algorithm to maximize the criterion under complete order restrictions. Nishisato and Arri [7] discussed the case of partial order and proposed an algorithm using separable programming. Their method is, however, limited to a special type of partial order. Avoiding this limitation, the present paper gives a generalized formulation applicable to arbitrary order restrictions and proposes an efficient algorithm using Wolfe's reduced gradient method. Numerical examples are provided to show the validity, the rapidness of convergence and the stability of the procedure.

1. Introduction

The methods of optimal scaling proposed by Fisher [4], Hayashi [5], Tanaka [9], [10], etc. are formulated as a problem of maximization of a ratio of quadratic forms, i.e.

$$Q=\boldsymbol{t}'A\boldsymbol{t}/\boldsymbol{t}'B\boldsymbol{t}\to\max \tag{1.1}$$

and the optimal scores are obtained by solving the eigenequation of A relative to B, where $\boldsymbol{t}=[t_1, t_2, \cdots, t_r]'$ denotes a score vector and the matrices A and B are defined according to the criterion adopted for optimal scaling ([9], [10]). However, we sometimes meet the situations where there are some reasons to believe that order relations should exist among categories. In such cases the optimal scores obtained by the ordinary methods do not always satisfy the same relations, and then we have to try the analysis several times after pooling categories with disordered scores until we obtain the scores satisfying the order relations required. In order to obtain an order-preserving solution in applying

Fisher's method, Bradley, Katti and Coons [2] proposed an algorithm to maximize the criterion Q under the order restrictions $t_1 \geqq t_2 \geqq \cdots \geqq t_r$. Tanaka [8] showed that the similar algorithm can be applied to Hayashi's second method of quantification with ordered categorical outside variable. Nishisato and Arri [7] considered the case of partial order and proposed an algorithm using separable programming. Their method is, however, limited to a special type of partial order, where the categories are expressed by a connected graph without any circuit. Avoiding this limitation, the present paper gives a generalized formulation applicable to arbitrary order restrictions and proposes an efficient algorithm using Wolfe's reduced gradient method [11].

2. Formulation of the problem—order restrictions and the corresponding transformations of variables

Assume an arbitrary set of order restrictions on the scores $t_1, t_2, \cdots, t_r$, and apply a transformation $z_{jj'} = t_j - t_{j'}$ corresponding to a restriction $t_j \geqq t_{j'}$. Using matrix notations the transformation is expressed by

$$z(c\times 1) = T(c\times r)t(r\times 1) \tag{2.1}$$

where

$$T = [n_{ij}]$$

$$n_{ij} = \begin{cases} 1 & \text{for some } j\,, \\ -1 & \text{for some } j' \neq j, \text{ for each } i\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

and where $\boldsymbol{t} = [t_1, t_2, \cdots, t_r]'$ denotes a score vector, $\boldsymbol{z} = [z_1, \cdots, z_c]'$ a transformed variable vector, and T a transformation matrix.

In order to investigate the properties of the transformation matrix, we shall define "connectedness" between categories as follows.

DEFINITION. Two categories t_j and $t_{j'}$ are "connected" with each other and expressed as $t_j \sim t_{j'}$, when

i) there exists a chain of inequalities, which starts from t_j and ends at $t_{j'}$, or

ii) there exists a connection between the jth and the j'th columns on T for some $i_0, i_1, \cdots, i_k$th rows, that is, there exists at least a sequence $\{n_{i_0 j}, n_{i_0 j_1}, n_{i_1 j_1}, \cdots, n_{i_k j'}\}$, where

$$n_{i_0 j} \neq 0,\ n_{i_0 j_1} \neq 0, \cdots, n_{i_k j'} \neq 0\ . \tag{2.2}$$

Also we define that a category is connected with itself, i.e. $t_j \sim t_j$.

The connectedness defined above is equivalent to that in graph theory when the categories $t_1, t_2, \cdots, t_r$ and the order relations among them are expressed as vertices and edges, respectively ([1], [3]).

Directly from the above definition it can be shown that the connected relation is an equivalent relation and makes a classification into equivalent classes of categories. Thus, for arbitrary order restrictions the set of categories $\{t_1, t_2, \cdots, t_r\}$ are classified into mutually exclusive and exhaustive subsets $\mathcal{T}_1, \mathcal{T}_2, \cdots, \mathcal{T}_m$ according to connected relations. Corresponding to this classification the transformation matrix T can be expressed as a block diagonal form such that

$$T(c\times k)=\text{block diag}\,\{T_1, T_2, \cdots, T_m\}\ , \tag{2.3}$$

where the submatrix $T_k = T_k(c_k \times r_k)$ transforms the subset of r_k categories $\mathcal{T}_k = \{t_{(k)1}, t_{(k)2}, \cdots, t_{(k)r_k}\}$ into the subset of c_k transformed variables $\mathcal{Z}_k = \{z_{(k)1}, z_{(k)2}, \cdots, z_{(k)c_k}\}$.

Lemma 1. *The rank of the submatrix $T_k = T_k(c_k \times r_k)$ is $r_k - 1$, and accordingly the rank of the transformation matrix $T(c\times r)$ is $r-m$.*

The proof is easily derived from the fact that the rank of T_k is equal to or greater than $r_k - 1$ because of the connected relations of the categories in the subset and the fact that the number of linearly independent contrasts of r_k variables is at most $r_k - 1$.

If we add a row, say,

$$z_{(0)k} = t_{(k)1}\ , \tag{2.4}$$

to a transformation $\boldsymbol{z}_{(k)} = T_k \boldsymbol{t}_{(k)}$, where $\boldsymbol{z}_{(k)} = [z_{(k)1}, \cdots, z_{(k)c_k}]'$, $\boldsymbol{t}_{(k)} = [t_{(k)1}, \cdots, t_{(k)r_k}]'$, we obtain a transformation submatrix with rank r_k.

Lemma 2. *We can construct a transformation matrix $T[(c+m)\times r]$ of rank r, i.e.*

$$T[(c+m)\times r] = \begin{array}{c|c|c|c|c|} & r_1 & r_2 & \cdots & r_m \\ \hline c_1 & T_1 & 0 & \cdots & 0 \\ \hline c_2 & 0 & T_2 & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline c_m & 0 & 0 & \cdots & T_m \\ \hline m & \multicolumn{4}{c|}{T_0} \\ \hline \end{array} \tag{2.5}$$

where

$$T_0=[n_{ij}^0] ,$$

$$n_{ij}^0=\begin{cases} 1 & \text{for } t_j \in \mathcal{T}_0 , \\ 0 & \text{otherwise} , \end{cases}$$

and $\mathcal{T}_0$ *is a set of elements* $\{t_{(k)1},\ k=1, 2, \cdots, m\}$ *chosen in* (2.4).

Using the matrix (2.5) the transformation is expressed as

$$\mathbf{z}[(c+m)\times 1]=T[(c+m)\times r]\mathbf{t}(r\times 1) , \tag{2.6}$$

where

$$\mathbf{z}[(c+m)\times 1]=[\mathbf{z}'_{(1)}, \mathbf{z}'_{(2)}, \cdots, \mathbf{z}'_{(m)}, \mathbf{z}'_{(0)}]' ,$$

$$\mathbf{t}(r\times 1)=[\mathbf{t}'_{(1)}, \mathbf{t}'_{(2)}, \cdots, \mathbf{t}'_{(m)}]' ,$$

$$\mathbf{z}_{(0)}=[z_{(0)1}, z_{(0)2}, \cdots, z_{(0)m}]' .$$

From the nature of optimal scaling criteria, the location of scores may be arbitrarily determined. Therefore, without loss of generality, we may specify a score $t_{(m)1}$ zero. Then the criterion is expressed as

$$Q=\tilde{\mathbf{t}}'\tilde{A}\tilde{\mathbf{t}}/\tilde{\mathbf{t}}'\tilde{B}\tilde{\mathbf{t}} , \tag{2.7}$$

where $\tilde{A}$, $\tilde{B}$ and $\tilde{\mathbf{t}}$ indicate the matrices made from A, B and $\mathbf{t}$ by eliminating a row and/or column corresponding to the specified score. The transformation becomes

$$\tilde{\mathbf{z}}[(c+m-1)\times 1]=\tilde{T}[(c+m-1)\times(r-1)]\tilde{\mathbf{t}}[(r-1)\times 1] , \tag{2.8}$$

where the rank of $\tilde{T}$ is $r-1$ and the meaning of the symbol ($\tilde{\ }$) is the same as in the case of $\tilde{A}$, $\tilde{B}$ and $\tilde{\mathbf{t}}$.

Because $\tilde{T}$ is of full rank, it has a left inverse matrix $(\tilde{T}'\tilde{T})^{-1}\tilde{T}'$, and (2.8) has an inverse transformation such that

$$\tilde{\mathbf{t}}=(\tilde{T}'\tilde{T})^{-1}\tilde{T}'\tilde{\mathbf{z}} . \tag{2.9}$$

Hence we finally have the following theorem.

THEOREM. *The problem of optimal scaling under arbitrary order restrictions can be always transformed to*

$$Q(\tilde{\mathbf{z}})=\tilde{\mathbf{z}}'C\tilde{\mathbf{z}}/\tilde{\mathbf{z}}'D\tilde{\mathbf{z}}\rightarrow \max \tag{2.10}$$

subject to

$$\text{(i)}\quad \mathbf{z}_{(k)}=[z_{(k)1}, z_{(k)2}, \cdots, z_{(k)c_k}]'\geqq 0 , \qquad k=1, 2, \cdots, m , \tag{2.11}$$

(2.12) (ii) $\boldsymbol{a}'_{(kj)}\boldsymbol{z}_{(k)}=0\,, \qquad j=1, 2, \cdots, c_k-r_k+1,\ k=1, 2, \cdots, m\,,$

where

$$C=\tilde{T}(\tilde{T}'\tilde{T})^{-1}\tilde{A}(\tilde{T}'\tilde{T})^{-1}\tilde{T}'\,, \qquad D=\tilde{T}(\tilde{T}'\tilde{T})^{-1}\tilde{B}(\tilde{T}'\tilde{T})^{-1}\tilde{T}'$$

and where $\boldsymbol{a}_{(kj)}$ *is a coefficient vector with* 0'*s and* ± 1'*s for* $\boldsymbol{z}_{(k)}$ *in the equality restrictions corresponding to linear dependencies among* $\{z_{(k)1}, z_{(k)2}, \cdots, z_{(k)c_k}\}$, $k=1, 2, \cdots, m$. *The number of equality restrictions is given as the difference between the number of transformed variables* $\{z_{(k)j}\}$ *and the number of linearly independent variables.*

When we represent the given order restrictions in a graph, we can easily obtain the equality restrictions corresponding to circuits in the graph. The cases discussed by Bradley, Katti and Coons [2] and Nishisato and Arri [7] are special cases with mutually connected categories and no equality restriction.

3. Application of a nonlinear programming technique to the optimization problem

3.1. *Wolfe's reduced gradient method* ([11])

As shown in the previous section the problem of optimal scaling under generalized order restrictions reduces finally to (2.10)–(2.12), i.e. the problem of maximizing a nonlinear objective function under nonnegativity and linear equality restrictions.

Wolfe's reduced gradient method was proposed just for such type of nonlinear programming problem ([11]) and is known as an efficient algorithm comparing with other competing methods ([6]). From such a viewpoint we shall use it to solve our optimal scalig problem numerically.

3.2. *Numerical examples*

Table 1, which shows the data for a five-treatment experiment with a five-point scoring scale, is taken from the study of Bradley, Katti and Coons [2] (p. 366, example 3). Now consider two kinds of order restrictions and apply our generalized method.

Table 1. A numerical example (Bradley, Katti and Coons [2])

		Response categories*					Total
		1 (t_1)	2 (t_2)	3 (t_3)	4 (t_4)	5 (t_5)	
Treat-ments	1	9	5	9	13	4	40
	2	7	3	10	20	4	44
	3	14	13	6	7	0	40
	4	11	15	3	5	8	42
	5	0	2	10	30	2	44
Total		41	38	38	75	18	210

* Notes: Category 1=Excellent; Category 2=Good;
Category 3=Fair; Category 4=Poor;
Category 5=Terrible.

(i) The case of order restrictions $t_1 \geqq \{t_2, t_3\} \geqq t_4 \geqq t_5$

The order restrictions are expressed by a graph in Fig. 1, which shows that all categories construct an equivalent class because they are connected with each other and that there exists only one circuit t_1—t_2—t_4—t_3—t_1.

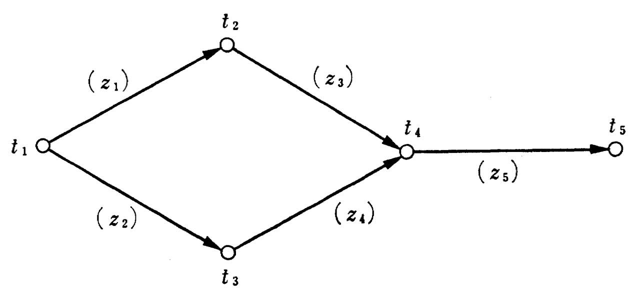

Fig 1. Expression of the order restrictions $t_1 \geqq \{t_2, t_3\} \geqq t_4 \geqq t_5$ by a (directed) graph

According to the order restrictions, the transformation matrix $T[(c+m) \times r]$ in (2.5) is expressed as follows.

$$(3.1) \qquad T=\left[\begin{array}{c|ccccc|} 1 & -1 & & & \\ 1 & & -1 & & \\ & 1 & & -1 & \\ & & 1 & -1 & \\ & & & 1 & -1 \\ \hline 1 & & & & \end{array}\right].$$

$$\Vert$$

$$\tilde{T}$$

Then the problem becomes

$$Q=\tilde{z}'C\tilde{z}/\tilde{z}'D\tilde{z}\rightarrow \max\ , \tag{3.2}$$

subject to

$$\text{(i)}\quad \tilde{z}=[z_1, z_2, \cdots, z_5]'\geqq 0\ , \tag{3.3}$$

$$\text{(ii)}\quad z_1-z_2+z_3-z_4=0\ , \tag{3.4}$$

where

$$C=\begin{bmatrix} 0.310250 & 0.252340 & 0.336210 & 0.394120 & 0.162770 \\ 0.252340 & 2.099860 & 2.888190 & 1.040670 & -0.280090 \\ 0.336210 & 2.888190 & 3.988550 & 1.436570 & -0.300860 \\ 0.394120 & 1.040670 & 1.436570 & 0.790020 & 0.141990 \\ 0.162770 & -0.280090 & -0.300860 & 0.141990 & 0.835500 \end{bmatrix},$$

$$D=\begin{bmatrix} 12.998770 & 3.498790 & -0.210730 & 9.289268 & 1.757150 \\ 3.498790 & 12.998779 & 9.289268 & -0.210730 & 1.757150 \\ -0.210730 & 9.289268 & 17.703537 & 8.203548 & 5.014279 \\ 9.289268 & -0.210730 & 8.203548 & 17.703537 & 5.014279 \\ 1.757150 & 1.757150 & 5.014279 & 5.014279 & 16.457138 \end{bmatrix}.$$

Application of the reduced gradient method to the optimization problem (3.2)-(3.4) yields the result shown in Table 2. Normalizing so as to satisfy $t_1=1.0$ and $t_5=0.0$, the optimal scores are given as

$$\boldsymbol{t}=[1.0000\quad 1.0000\quad 0.1435\quad 0.0000\quad 0.0000]'\ .$$

Table 2. Solution obtained for $t_1\geqq\{t_2, t_3\}\geqq t_4\geqq t_5$

Cycle	z_1	z_2	z_3	z_4	z_5	$Q(z)$
0	*1.00000	1.00000	1.00000	1.00000	1.00000	0.1224780
1	*0.0^4757	1.10058	1.70579	0.60522	0.58841	0.1978295
2	0.0^6542	*1.10058	1.70580	0.60521	0.58839	0.1978315
3	0.08964	*1.85925	2.03080	0.26120	0.0^5930	0.2423639
4	0.0^5140	*1.75611	2.03277	0.27666	0.0^5930	0.2435769
5	0.0^5140	*1.74103	2.03097	0.28994	0.0^5930	0.2435847
6	0.0^5140	*1.74063	2.03092	0.29029	0.0^5930	0.2435847
7	0.0^5140	*1.73886	2.03070	0.29184	0.0^5930	0.2435848
8	0.0^5140	*1.73990	2.03084	0.29094	0.0^5930	0.2435847
9	0.0^5140	1.73944	2.03078	0.29134	0.0^5930	0.2435848

Notes: (1) The iterative procedure stops when the norm of gradient vector for non-basic variables is smaller than 10^{-5}.

(2) The value with symbol (*) indicates that it is selected as a basic variable in each cycle.

Although the set of order restrictions $t_1 \geqq \{t_2, t_3\} \geqq t_4 \geqq t_5$ does not contain $t_2 \geqq t_3$, the solution obtained satisfies also the condition $t_2 \geqq t_3$. Therefore, the optimal scores obtained should coincide with those under the complete order restrictions $t_1 \geqq t_2 \geqq t_3 \geqq t_4 \geqq t_5$. In fact, the method of Bradley et al. gives the optimal scores

$$\boldsymbol{t} = [1.0000 \quad 1.0000 \quad 0.1434 \quad 0.0000 \quad 0.0000]'$$

under the complete order restrictions.

(ii) The case of order restrictions $t_1 \geqq \{t_2, t_3, t_4, t_5\}$

This set of order restrictions, which was investigated by Nishisato and Arri [7], is expressed by a graph in Fig. 2. As this graph does not contain any circuit, there exists no equality restriction. In this case the reduced gradient method is equivalent to the ordinary gradient method and yields the result $\boldsymbol{t} = [1.0000\ 1.0000\ -1.4260\ -2.1681\ 0.0000]'$, which coincides well with the solution $\boldsymbol{t} = [1.000\ 1.000\ -1.426\ -2.168\ 0.000]'$ by Nishisato and Arri [7].

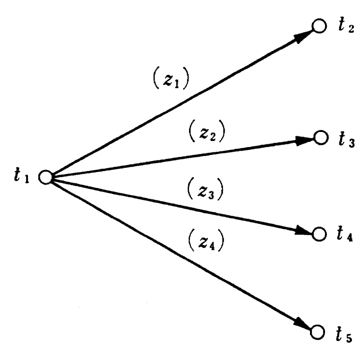

Fig. 2. Expression of the order restrictions $t_1 \geqq \{t_2, t_3, t_4, t_5\}$ by a (directed) graph

4. Discussion

Methods of optimal scaling under order restrictions have been investigated by several authors such as Bradley, Katti and Coons [2], Nishisato and Arri [7] among others. Bradley et al. [2] studied the case of complete order $t_1 \geqq t_2 \geqq \cdots \geqq t_r$. Nishisato and Arri [7] extended this to the case of partial order. Their extension is, however, limited to the case that all categories are connected with each other and can be expressed by a connected graph containing no circuit. In preceding sections we gave a generalized formulation for arbitrary partial order. One of our motivations behind this generalization is to scale optimally a movement of grade or scoring from pre- to post-treatment.

These problems of optimal scaling under order restrictions are formulated as nonlinear programming problems or optimization problems under constraints. But generally a nonlinear programming problem is not easy to solve numerically even if the formulation is made. The method by Bradley et al. [2] is a kind of gradient method and the iterative procedure converges rapidly to optimal solution according to our experiences. It is not applicable, however, to the case of partial order in their original form. Nishisato and Arri [7] applied separable programming based on polygonal approximations after transforming all functions in the constraints and the objective function to separable forms. In order to obtain a solution accurately, a large number of mesh points should be chosen around the initial values, which are obtained sufficiently near the optimal values. They obtained initial values by solving eigenvalue problems iteratively after pooling disordered categories. In this procedure to obtain initial approximations, however, there exists a possibility to pool excessively, so that the initial approximations are not always sufficiently near the optimal values. In such cases it is required to choose mesh points and initial values carefully and to solve a large scale linear programming problem.

Because of these difficulties of separable programming we applied the Wolfe's reduced gradient method [11]. It is known as a very efficient algorithm to solve a special type of nonlinear programming problem with linear equality constraints for nonnegative variables [6]. Thus it may be said that the method is just suitable for our problem (2.10)–(2.12). In fact the procedure converges rapidly to the optimal values, starting from arbitrarily chosen initial values $z_1=z_2=\cdots=1.0$. It may be much better to use the conjugate gradient algorithm than to use the ordinary steepest ascent algorithm in the reduced gradient method, especially when the number of unknown parameters is large.

Acknowledgement

The author would like to express his hearty thanks to Professor C. Asano, Research Institute of Fundamental Information Science, Kyushu University, for his advice and encouragement. He is also grateful to the referee for his valuable comments, which were useful to improve the paper.

TAKEDA CHEMICAL INDUSTRIES, LTD.*

* Now at Okayama University.

REFERENCES

[1] Biggs, N. (1974). *Algebraic Graph Theory*, Camridge Univ. Press.

[2] Bradley, R. A., Katti, S. K. and Coons, I. J. (1962). Optimal scaling for ordered categories, *Psychometrika*, **27**, 355-374.

[3] Busacker, R. G. and Saaty, T. L. (1965). *Finite Graphs and Networks: An Introduction and Applications*, McGraw-Hill.

[4] Fisher, R. A. (1958). *Statistical Methods for Research Workers*, 13th ed. Oliver & Boyd.

[5] Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view, *Ann. Inst. Statist. Math.*, **3**, 69-98.

[6] Jacobi, S. L. S., Kowalik, J. S. and Pizzo, J. T. (1972). *Iterative Methods for Nonlinear Optimization Problems*, Prentice-Hall.

[7] Nishisato, S. and Arri, P. S. (1975). Nonlinear programming approach to optimal scaling of partially ordered categories, *Psychometrika*, **40**, 525-548.

[8] Tanaka, Y. (1972). On quantification theory with ordered categorical outside variable, *Standardization and Quality Control*, **25**, 31-39 (in Japanese).

[9] Tanaka, Y. (1978). Some generalized methods of optimal scaling and their asymptotic theories: The case of multiple responses-multiple factors, *Ann. Inst. Statist. Math.*, **30**, A, 329-348.

[10] Tanaka, Y. (1979). Some generalized methods of optimal scaling and their asymptotic theories: The case of single response-multiple factors, to appear.

[11] Wolfe, P. (1963). Methods of Nonlinear Programming, *Recent Advances in Mathematical Programming* R. L. Graves and P. Wolfe (eds.), McGraw-Hill.