Statistical Methods & Applications (2004) 13: 139–150
DOI: 10.1007/s10260-004-0089-z

SMA

A general algorithm to fit constrained DEDICOM models

Roberto Rocci

Dip. SEFeMEQ, University of Rome "Tor Vergata", Italy (e-mail: roberto.rocci@uniroma2.it)

Abstract. The DEDICOM model is a model to analyze square tables describing asymmetric relationships among n entities. Its importance in the asymmetric multidimensional scaling literature is due to the fact that several authors showed a large class of models to be simply a constrained version of DEDICOM. A typical example is the Generalized GIPSCAL proposed by Kiers & Takane. In this paper we present a new algorithm capable to fit, in the least squares sense, any DEDICOM constrained model.

Key words: Asymmetric multidimensional scaling, DEDICOM model, penalized least squares

1. Introduction

Multidimensional scaling is a set of techniques to analyze a data matrix $\mathbf{X} = [x_{ij}]$ whose rows and columns correspond to the same set of n objects (e.g. correlation matrices, proximity data). When the intensity of the relation depends on the order in which the objects are considered, the data matrix $\mathbf{X}$ is asymmetric. Typical examples are: similarity ratings, preferences, flows (e.g. import-export, brand switching), contingency tables (e.g. occupational mobility, word associations), etc.. Standard multidimensional scaling (Gower, 1966), designed for symmetric data, can not be applied in this case and suitable techniques are required for analyzing asymmetry not solely due to random noise. Several models have been proposed (for a review, see Zielman and Heiser, 1996) to deal with asymmetric data. Among them, DEDICOM (Harshman et al., 1982) is one of the most important. This model can be formulated as

$$\mathbf{X} = \mathbf{A}\mathbf{R}\mathbf{A}' + \mathbf{E}, \tag{1}$$

where $\mathbf{A}$ is a $n \times p$ ($n > p$) matrix of loadings (weights) for the n objects on p dimensions, $\mathbf{R}$ is a $p \times p$ matrix representing relations among the p dimensions and $\mathbf{E}$ is a residual term. The aim of the model is to explain the asymmetric relations among the n objects by the relations among a small set of dimensions which can be regarded as "aspects" of the objects. The relevance of each aspect on a particular object is indicated by the entries of $\mathbf{A}$. The parameters are usually estimated by minimizing the sum of squared residuals. Several Alternating Least Squares (ALS) algorithms have been proposed to solve this minimization problem. They can be divided into two classes. The first contains algorithms, which we call left-and-right, where the left and the right hand side matrix $\mathbf{A}$ in $\mathbf{ARA}'$ are treated independently. Mostly, the left and right hand matrices are equal upon convergence and thus the proper model has been fitted. The second class contains algorithms that do not distinguish the two left and right matrices but impose the constraint $\mathbf{A}'\mathbf{A} = \mathbf{I}$. This does not cause any loss of fit because every non-singular linear transformation of $\mathbf{A}$ can be undone by applying the inverse transformation on $\mathbf{R}$ (Kiers, 1989; Kiers et al., 1990).

The importance of DEDICOM in multidimensional scaling literature is due to the fact that a large class of models to analyze asymmetric proximity data are simply constrained versions of it. In fact, it can be shown that the Generalized GIPSCAL (Kiers & Takane, 1994), Generalized EG (Rocci & Bove, 2002), EG (Escoufier & Grorud, 1980) and the Gower decomposition (Gower, 1977) are particular cases of the DEDICOM model. They simply correspond to a DEDICOM model where the relation matrix $\mathbf{R}$ has a particular form. Several other models could be obtained by imposing some constraints on the matrix $\mathbf{R}$. For example, in some applications it could be useful to set some elements of $\mathbf{R}$ equal to zero to simplify the interpretation of the results.

This class of models gives rise to the need for a unique algorithm capable to fit the DEDICOM model and its constrained versions. In this way the researcher can easily implement a model selection procedure to find the model suitable for the data. At the moment such algorithm does not exist. Our aim is to fill this gap by extending the generalized Takane's algorithm (Kiers et al., 1990), which has been proposed to fit the DEDICOM model and appeared to be very efficient in most practical cases.

The material of this paper is organized as follows. First, in Sect. 2 we examine some constrained versions of DEDICOM. The new algorithm is introduced in Sect. 3, while in Sect. we present the results of a simulation study to test the performances of the algorithm. We conclude with a general discussion in Sect. 5.

2. Constrained DEDICOM models

One of the most important DEDICOM constrained model (cDEDICOM for short) is Generalized GIPSCAL (GG). This model generalizes, but also simplifies, the GIPSCAL model proposed by Chino (1990) and it allows a graphical representation of the results which is absent in the DEDICOM model. It is based on the decomposition of the elementary datum x_{ij} into two parts: symmetry (i.e. $s_{ij} = 1/2(x_{ij}+x_{ji})$)

and skew-symmetry (i.e. $k_{ij} = {}^1\!/_2(x_{ij} - x_{ji})$). The aim is to find a simultaneous graphical representation of the two parts. The model can be formulated as

$$x_{ij} = \sum_{m=1}^{p/2} [b(a_{i(2m-1)}a_{j(2m-1)} + a_{i(2m)}a_{j(2m)}) \\ +d_m(a_{i(2m-1)}a_{j(2m)} - a_{i(2m)}a_{j(2m-1)})] + e_{ij}, \qquad (2)$$

where b can be only 1 or 0 and $d_m \geq 0(m = 1, 2, \ldots, p/2)$. When $b = 1$ and $p = 2$, it allows us to represent object i as a point lying in a plane (bimension) having coordinates (a_{i1}, a_{i2}). The obtained configuration of n points represents simultaneously symmetry and skew-symmetry. In fact, let $\hat{x}_{ij}$ be the fitted data pertaining to the pair (i, j), we have

$$\hat{s}_{ij} = \frac{1}{2}(\hat{x}_{ij} + \hat{x}_{ji}) = a_{i1}a_{j1} + a_{i2}a_{j2}, \qquad (3)$$

that is, for any pair of points the scalar product describes the symmetric component of the data. Furthermore,

$$\hat{k}_{ij} = \frac{1}{2}(\hat{x}_{ij} - \hat{x}_{ji}) = d_1(a_{i1}a_{j2} - a_{i2}a_{j1}) \qquad (4)$$

that is, twice the area of the triangle having the two points and the origin as vertices, multiplied by d_1, describes the absolute value of the skew-symmetric component, whose algebraic sign is associated with the orientation of the plane (positive counter-clockwise, negative clockwise).

When $b = 1$ and $p > 2$, the symmetric component of the data is represented by

$$\hat{s}_{ij} = \sum_{m=1}^{p/2} (a_{i(2m-1)}a_{j(2m-1)} + a_{i(2m)}a_{j(2m)}) \qquad (5)$$

which corresponds to the sum of the scalar products between a pair of points on each bimension, while the skew-symmetric component is represented by

$$\hat{k}_{ij} = \sum_{m=1}^{p/2} d_m(a_{i(2m-1)}a_{j(2m)} - a_{i(2m)}a_{j(2m-1)}) \qquad (6)$$

that is, twice the sum of the triangle areas taking into account the algebraic sign and the weight d_m of each bimension. It follows that to get the intensity of the relation between two objects we have to consider all the diagrams simultaneously (one for each bimension) summing algebraically scalar products and weighted areas.

The GG model is a constrained version of DEDICOM. This can be easily seen, by rewriting (2) in matrix notation as

$$\mathbf{X} = \mathbf{A}(b\mathbf{I} + \mathbf{KD})\mathbf{A}' + \mathbf{E}, \qquad (7)$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{D} = \text{diag}(d_1, d_1, , d_2, d_2, \ldots, d_{p/2}, d_{p/2})$ and $\mathbf{K}$ is a block-diagonal matrix having on its main diagonal the matrix

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{8}$$

repeated $p/2$ times. It follows that GG is a DEDICOM model where $\mathbf{R}$ is constrained to have a particular form. Kiers & Takane (1994) have shown that this constraint could be non-active, since every matrix $\mathbf{R}$ of even order having a positive semi-definite symmetric part (i.e. $1/2(\mathbf{R} + \mathbf{R}'))$ can be always rewritten in the form $\mathbf{R} = \mathbf{U}(\mathbf{I} + \mathbf{KD})\mathbf{U}'$.

Several models can be considered as constrained versions of GG and then cDEDICOM models: the Generalized Escoufier & Grorud model (Rocci &Bove, 2002), which can be obtained by setting $b = 1$ and $d_1 = d_2 = \ldots = d_{p/2}$, the Escoufier & Grorud model (Escoufier & Grorud, 1980), where $b = 1$ and $d_1 = d_2 = \ldots = d_{p/2} = 1$, and the Gower decomposition (Gower, 1977), where $b = 0$ and only the skew-symmetric part of the data is analyzed. Of course, the last model is used only when the data matrix is skew-symmetric or when the symmetric and the skew-symmetric parts of the data are analyzed separately.

In the next section we will introduce a general algorithm capable to fit, in a least squares sense, the models considered above and other cDEDICOM models.

3. A general algorithm

In this section we consider the problem of fitting a cDEDICOM model. The least squares loss function to be minimized can be written as

$$f(\mathbf{A}, \mathbf{R}) = 2 \left\| \mathbf{X} - \mathbf{ARA}' \right\|^2, \tag{9}$$

where $\|\cdot\|$ indicates the Frobenius matrix norm and the multiplicative factor 2 is inserted only for convenience. An ALS algorithm can be formulated by alternatively updating $\mathbf{A}$ and $\mathbf{R}$. In the next subsections we will show how to perform those updates.

3.1. Update of $\mathbf{A}$

The update of the loading matrix can be done by adopting a left-and-right approach, i.e. (9) is modified as

$$f_\lambda(\mathbf{A}, \mathbf{B}, \mathbf{R}) = \left\| \mathbf{X} - \mathbf{ARB}' \right\|^2 + \left\| \mathbf{X} - \mathbf{BRA}' \right\|^2 + \lambda \left\| \mathbf{A} - \mathbf{B} \right\|^2 \tag{10}$$

which coincides with (9) when $\mathbf{A} = \mathbf{B}$, and the current value of the loading matrix $\mathbf{A}$ is updated with the minimizer $\mathbf{B}$ of (10) calculated for given values of $\mathbf{A}$ and λ.

The underlying idea is the following. Since the two matrices $\mathbf{ARB}'$ and $\mathbf{BRA}'$ approximate the same matrix $\mathbf{X}$, when we minimize (10), then $\mathbf{A}$ tends to be equal to $\mathbf{B}$. To force this "natural" tendency, we also add a penalty term which penalizes those solutions where $\mathbf{A}$ is different from $\mathbf{B}$. The parameter λ calibrates the weight of the penalty and it has to be chosen by the user. In the following, we first discuss the choice of λ, then the minimization of (10) with respect to $\mathbf{B}$. We start from the following result.

Result 1. Let $\mathbf{B}$ be such that

$$f_\lambda(\mathbf{A}, \mathbf{B}, \mathbf{R}) < f_\lambda(\mathbf{A}, \mathbf{A}, \mathbf{R}), \tag{11}$$

then $f(\mathbf{B}, \mathbf{R}) < f(\mathbf{A}, \mathbf{R})$ if $\lambda \geq \lambda^*$, where

$$\lambda^* = \frac{\text{vec}(\mathbf{A}'\mathbf{A} - \mathbf{B}'\mathbf{B})'\mathbf{S}\ \text{vec}(\mathbf{A}'\mathbf{A} - \mathbf{B}'\mathbf{B}) - 2\text{vec}(\mathbf{A} - \mathbf{B})'\mathbf{T}\ \text{vec}(\mathbf{A} - \mathbf{B})}{\text{vec}(\mathbf{A} - \mathbf{B})'\ \text{vec}(\mathbf{A} - \mathbf{B})} \tag{12}$$

being $\mathbf{S}$ and $\mathbf{T}$ the symmetric parts of $\mathbf{R} \otimes \mathbf{R}$ and $\mathbf{R} \otimes \mathbf{X}$, respectively.

Proof. The statement is proven if we are able to show that

$$f_\lambda(\mathbf{A}, \mathbf{A}, \mathbf{R}) + f_\lambda(\mathbf{B}, \mathbf{B}, \mathbf{R}) - 2f_\lambda(\mathbf{A}, \mathbf{B}, \mathbf{R}) \leq 0. \tag{13}$$

In fact, by combining (13) with (11), we have

$$f_\lambda(\mathbf{A}, \mathbf{A}, \mathbf{R}) + f_\lambda(\mathbf{B}, \mathbf{B}, \mathbf{R}) \leq 2f_\lambda(\mathbf{A}, \mathbf{B}, \mathbf{R}) < 2f_\lambda(\mathbf{A}, \mathbf{A}, \mathbf{R}) \tag{14}$$

which implies

$$f_\lambda(\mathbf{B}, \mathbf{B}, \mathbf{R}) < f_\lambda(\mathbf{A}, \mathbf{A}, \mathbf{R}), \tag{15}$$

i.e., the statement. To show that (13) is true when $\lambda \geq \lambda^*$, we note that

$$\begin{aligned} f_\lambda(\mathbf{A}, \mathbf{B}, \mathbf{R}) &= 2\left\|\mathbf{X}\right\|^2 + \left\|\mathbf{ARB}'\right\|^2 + \left\|\mathbf{BRA}'\right\|^2 - 2\text{tr}(\mathbf{X}'\mathbf{ARB}') \\ &\quad -2\text{tr}(\mathbf{X}'\mathbf{BRA}') + \lambda\left\|\mathbf{A} - \mathbf{B}\right\|^2 \\ &= 2\left\|\mathbf{X}\right\|^2 + \text{vec}(\mathbf{A}'\mathbf{A})'(\mathbf{R} \otimes \mathbf{R} + \mathbf{R}' \otimes \mathbf{R}')\ \text{vec}(\mathbf{B}'\mathbf{B}) + \\ &\quad -2\text{vec}(\mathbf{A})'(\mathbf{R} \otimes \mathbf{X} + \mathbf{R}' \otimes \mathbf{X}')\ \text{vec}(\mathbf{B}) + \lambda\left\|\mathbf{A} - \mathbf{B}\right\|^2 \\ &= 2\left\|\mathbf{X}\right\|^2 + 2\tilde{\mathbf{a}}'\mathbf{S}\tilde{\mathbf{b}} - 4\mathbf{a}'\mathbf{T}\mathbf{b} + \lambda(\mathbf{a} - \mathbf{b})'(\mathbf{a} - \mathbf{b}), \end{aligned} \tag{16}$$

where $\tilde{\mathbf{a}} = \text{vec}(\mathbf{A}'\mathbf{A})$, $\tilde{\mathbf{b}} = \text{vec}(\mathbf{B}'\mathbf{B}), \mathbf{a} = \text{vec}(\mathbf{A}), \mathbf{b} = \text{vec}(\mathbf{B}), \mathbf{S} = 0.5(\mathbf{R} \otimes \mathbf{R} + \mathbf{R}' \otimes \mathbf{R}')$ and $\mathbf{T} = 0.5(\mathbf{R} \otimes \mathbf{X} + \mathbf{R}' \otimes \mathbf{X}')$. It follows that (13) can be rewritten as

$$\begin{aligned} &f_\lambda(\mathbf{A}, \mathbf{A}, \mathbf{R}) + f_\lambda(\mathbf{B}, \mathbf{B}, \mathbf{R}) - 2f_\lambda(\mathbf{A}, \mathbf{B}, \mathbf{R}) = \\ &\quad = 2\tilde{\mathbf{a}}'\mathbf{S}\tilde{\mathbf{a}} - 4\mathbf{a}'\mathbf{T}\mathbf{a} + 2\tilde{\mathbf{b}}'\mathbf{S}\tilde{\mathbf{b}} - 4\mathbf{b}'\mathbf{T}\mathbf{b} - 4\tilde{\mathbf{a}}'\mathbf{S}\tilde{\mathbf{b}} + 8\mathbf{a}'\mathbf{T}\mathbf{b} - 2\lambda(\mathbf{a} - \mathbf{b})'(\mathbf{a} - \mathbf{b}) \\ &\quad = 2(\tilde{\mathbf{a}} - \tilde{\mathbf{b}})'\mathbf{S}(\tilde{\mathbf{a}} - \tilde{\mathbf{b}}) - 4(\mathbf{a} - \mathbf{b})'\mathbf{T}(\mathbf{a} - \mathbf{b}) - 2\lambda(a - \mathbf{b})'(\mathbf{a} - \mathbf{b}), \end{aligned} \tag{17}$$

which is negative or zero when λ satisfies inequality (12). □

The above result states that the choice of λ greater the particular threshold λ^*, assures that the minimization of (10) with respect to $\mathbf{B}$ gives us an update for the loading matrix which decreases the loss function (9). However, the threshold λ^* depends on both the current value $\mathbf{A}$ and the new value $\mathbf{B}$. This implies that we should know in advance the value of the new update before computing it. This problem can be solved by finding an upper bound for the threshold that does not depend on $\mathbf{B}$. This upper bound can be found when the loading matrix is constrained to be columnwise orthonormal, as stated by the following corollary.

Corollary 1. Let $\mathbf{A}$ and $\mathbf{B}$ be columnwise orthonormal matrices such that $f_\lambda(\mathbf{A},\mathbf{B},\mathbf{R}) < f_\lambda(\mathbf{A},\mathbf{A},\mathbf{R})$, then $f(\mathbf{B},\mathbf{R}) < f(\mathbf{A},\mathbf{R})$ if

$$\lambda \geq \alpha, \tag{18}$$

where α is the largest eigenvalue of $-2\mathbf{T} = -(\mathbf{R}\otimes\mathbf{X} + \mathbf{R}'\otimes\mathbf{X}')$.

Proof. The proof consists in showing that $\alpha \geq \lambda^*$. Firstly, we note that in this case $\mathbf{A}'\mathbf{A} - \mathbf{B}'\mathbf{B} = \mathbf{0}$ in (12). Secondly, we notice that the ratio

$$\frac{-2\mathrm{vec}(\mathbf{A}-\mathbf{B})'\mathbf{T}\ \mathrm{vec}(\mathbf{A}-\mathbf{B})}{\mathrm{vec}(\mathbf{A}-\mathbf{B})'\ \mathrm{vec}(\mathbf{A}-\mathbf{B})} \tag{19}$$

is the Rayleigh quotient of $-2\mathbf{T}$. The statement follows because it is well known that the Rayleigh quotient of any symmetric matrix assumes as maximum value the largest eigenvalue of the matrix itself. □

The above corollary is equivalent to result 2 by Kiers et al. (1990), where the unconstrained DEDICOM model has been considered. When the loading matrix is unconstrained, we can use alternatively the following corollary.

Corollary 2. Let $\mathbf{B}$ satisfying $f_\lambda(\mathbf{A},\mathbf{B},\mathbf{R}) < f_\lambda(\mathbf{A},\mathbf{A},\mathbf{R})$, then $f(\mathbf{B},\mathbf{R}) < f(\mathbf{A},\mathbf{R})$ if

$$\lambda \geq \mu\left(\|\mathbf{A}-\mathbf{B}\| + 2\|\mathbf{A}\|\right)^2 + \alpha, \tag{20}$$

where μ and α are the largest eigenvalues of $-2\mathbf{T} = -(\mathbf{R}\otimes\mathbf{X} + \mathbf{R}'\otimes\mathbf{X}')$ and $\mathbf{S} = 0.5(\mathbf{R}\otimes\mathbf{R} + \mathbf{R}'\otimes\mathbf{R}')$, respectively.

Proof. The proof consists in showing that $\mu\left(\|\mathbf{A}-\mathbf{B}\| + 2\|\mathbf{A}\|\right)^2 + \alpha \geq \lambda^*$. First, we note that

$$\begin{aligned}
&\mathrm{vec}(\mathbf{A}'\mathbf{A}-\mathbf{B}'\mathbf{B})'\mathbf{S}\ \mathrm{vec}(\mathbf{A}'\mathbf{A}-\mathbf{B}'\mathbf{B}) = \\
&\leq \mu\mathrm{vec}(\mathbf{A}'\mathbf{A}-\mathbf{B}'\mathbf{B})'\ \mathrm{vec}(\mathbf{A}'\mathbf{A}-\mathbf{B}'\mathbf{B}) && \text{(21.a)}\\
&\leq \mu\left\|\mathbf{C}'\mathbf{C}-\mathbf{A}'\mathbf{C}-\mathbf{C}'\mathbf{A}\right\|^2 && \text{(21.b)}\\
&\leq \mu\left(\|\mathbf{C}'\mathbf{C}\| + \|\mathbf{A}'\mathbf{C}\| + \|\mathbf{C}'\mathbf{A}\|\right)^2 && \text{(21.c)}\\
&\leq \mu\left(\|\mathbf{C}\|^2 + 2\|\mathbf{A}\|\,\|\mathbf{C}\|\right)^2, && \text{(21.d)}
\end{aligned}$$

where (21.a) derives from the upper bound of the Rayleigh quotient, (21.b) is obtained by setting $\mathbf{C} = \mathbf{A}-\mathbf{B}$, (21.c) derives from the triangular inequality and (21.d) follows from the fact that $\|\mathbf{A}'\mathbf{C}\| = \|\mathbf{C}'\mathbf{A}\| \leq \|\mathbf{A}\|\,\|\mathbf{C}\|$. Then, by recalling that (see proof of corollary 1)

$$\frac{-2\mathrm{vec}(\mathbf{A}-\mathbf{B})'\mathbf{T}\ \mathrm{vec}(\mathbf{A}-\mathbf{B})}{\mathrm{vec}(\mathbf{A}-\mathbf{B})'\ \mathrm{vec}(\mathbf{A}-\mathbf{B})} \leq \alpha, \tag{22}$$

it follows

$$\mu\left(\|\mathbf{A}-\mathbf{B}\|+2\|\mathbf{A}\|\right)^{2}+\alpha=\frac{\mu\left(\|\mathbf{A}-\mathbf{B}\|^{2}+2\|\mathbf{A}\|\,\|\mathbf{A}-\mathbf{B}\|\right)^{2}}{\|\mathbf{A}-\mathbf{B}\|^{2}}+\alpha\geq\lambda^{*}. \tag{23}$$

□

Corollary 2 gives us a threshold that can not be computed without the prior knowledge of the new update. However, we know that the norm of the difference between the new and the old value of the loading matrix is a decreasing function of λ, if $\mathbf{B}$ is computed by minimizing (10). This implies that if we compute (20) by using the norm $\|\mathbf{A}-\mathbf{B}\|$, where $\mathbf{B}$ have been computed by minimizing (10) with $\lambda=0$, then we obtain a value for λ greater than the threshold. The choice of λ will be also discussed in the next section.

Then, the update of the weight matrix can be done by minimizing (10) for a given value of λ.

When the weight matrix is constrained to be columnwise orthonormal, i.e. $\mathbf{A}'\mathbf{A}=\mathbf{B}'\mathbf{B}=\mathbf{I}$, the minimum of (10) with respect to $\mathbf{B}$ can be computed by noting that

$$\begin{aligned} f_{\lambda}(\mathbf{A},\mathbf{B},\mathbf{R}) &= 2\|\mathbf{X}\|^{2}+\|\mathbf{BRA}'\|^{2}+\|\mathbf{ARB}'\|^{2}+\lambda(\|\mathbf{A}\|^{2}+\|\mathbf{B}\|^{2}) \\ &\quad -\mathrm{tr}(\mathbf{B}'\mathbf{X}'\mathbf{AR}+\mathbf{B}'\mathbf{XAR}'+\lambda\mathbf{B}'\mathbf{A}) \qquad (24)\\ &= 2\|\mathbf{X}\|^{2}+2\|\mathbf{R}\|^{2}+2\lambda p-\mathrm{tr}(\mathbf{B}'\mathbf{X}'\mathbf{AR}+\mathbf{B}'\mathbf{XAR}'+\lambda\mathbf{B}'\mathbf{A}) \\ &= 2\|\mathbf{X}\|^{2}+2\|\mathbf{R}\|^{2}+2\lambda p-\mathrm{tr}(\mathbf{B}'\mathbf{W}). \end{aligned}$$

The implication is that (10) attains a minimum when $\mathrm{tr}(\mathbf{B}'\mathbf{W})$ is a maximum, i.e. when

$$\mathbf{B}=\mathbf{PQ}', \tag{25}$$

where $\mathbf{PLQ}'$ is the singular value decomposition of $\mathbf{W}$ (see Cliff, 1966).

Finally, it can be easily shown that (10) attains its unconstrained minimum for

$$\mathbf{B}=(\mathbf{X}'\mathbf{AR}+\mathbf{XAR}'+\lambda\mathbf{A})(\mathbf{R}'\mathbf{A}'\mathbf{AR}+\mathbf{RA}'\mathbf{AR}'+\lambda\mathbf{I})^{-1}. \tag{26}$$

In the next subsection we will show how to update $\mathbf{R}$ to fit a cDEDICOM model.

3.2. Update of $\mathbf{R}$

DEDICOM model

The update is

$$\mathbf{R}=(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{XA}(\mathbf{A}'\mathbf{A})^{-1}. \tag{27}$$

GG model

The matrix $\mathbf{R}$ is constrained to be of the form $\mathbf{I}+\mathbf{KD}$. The update can be easily found by writing (9) as

$$\begin{aligned} f(\mathbf{A},\mathbf{R}) &= 2\left\|\mathbf{X}-\mathbf{ARA}'\right\|^2 \\ &= 2\left\|\mathbf{X}^s-\mathbf{AA}'\right\|^2 + 2\left\|\mathbf{X}^k-\mathbf{AKDA}'\right\|^2 \\ &= 2\left\|\mathbf{X}^s-\mathbf{AA}'\right\|^2 + 2\left\|\mathrm{vec}(\mathbf{X}^k)-(\mathbf{A}\times\mathbf{AK})\mathbf{d}\right\|^2, \end{aligned} \tag{28}$$

where $\mathbf{X}^s$ and $\mathbf{X}^k$ are the symmetric and skew-symmetric parts of $\mathbf{X}$, $\times$ indicates the Khatri-Rao (column-wise Kronecker) product of matrices and $\mathbf{d}$ the vector containing the diagonal elements of $\mathbf{D}$. It implies that the update of $\mathbf{R}$ can be obtained by setting

$$[d_1, d_2, ..., d_{p/2}] = \mathrm{vec}(\mathbf{X}^k)'\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1} \tag{29}$$

where $\mathbf{C}=(\mathbf{A}\times\mathbf{AK})(\mathbf{I}_p\otimes\begin{bmatrix}1\\1\end{bmatrix})$.

A different update can be found by using the Kiers & Takane theorem (1994). In fact they show that a DEDICOM solution is a GG solution if and only if the symmetric part of the matrix $\mathbf{R}$ is positive semi-definite. It follows that we can fit the GG model exactly as DEDICOM, i.e. we can require $\mathbf{A}$ columnwise orthonormal, by imposing the constraint that $\mathbf{R}^s$, i.e. the symmetric part of $\mathbf{R}$, must be positive semi-definite. To impose this constraint we can follow this strategy. First we rewrite (9) as

$$\begin{aligned} f(\mathbf{A},\mathbf{R}) &= 2\left\|\mathbf{X}-\mathbf{ARA}'\right\|^2 \\ &= 2\left\|\mathbf{X}^s-\mathbf{AR}^s\mathbf{A}'\right\|^2 + 2\left\|\mathbf{X}^k-\mathbf{AR}^k\mathbf{A}'\right\|^2 \\ &= 2\left\|\mathbf{A}'\mathbf{X}^s\mathbf{A}-\mathbf{R}^s\right\|^2 + 2\left\|\mathbf{A}'\mathbf{X}^k\mathbf{A}-\mathbf{R}^k\right\|^2 + 2\|\mathbf{X}\|^2 - 2\|\mathbf{A}'\mathbf{XA}\|^2. \end{aligned} \tag{30}$$

Note that (30) includes two independent minimization problems. The solution of the first one is obtained by taking $\mathbf{R}^s=\mathbf{PLP}'$ where $\mathbf{P}$ are the eigenvectors of $\mathbf{A}'\mathbf{X}^s\mathbf{A}$ associated to the positive eigenvalues and $\mathbf{L}$ is the diagonal matrix of the corresponding eigenvalues. The solution of the second minimization problem is simply $\mathbf{R}^k=\mathbf{A}'\mathbf{X}^k\mathbf{A}$.

GEG model

In this case the matrix $\mathbf{R}$ is constrained to be of the form $\mathbf{I}+d\mathbf{K}$. It follows that the update of $\mathbf{R}$ simply resolves in the update of d. We can write

$$\begin{aligned} f(\mathbf{A},\mathbf{R}) &= 2\left\|\mathbf{X}-\mathbf{ARA}'\right\|^2 \\ &= 2\left\|\mathbf{X}^s+\mathbf{X}^k-\mathbf{AA}'-d\mathbf{AKA}'\right\|^2 \\ &= 2\left\|\mathbf{X}^s-\mathbf{AA}'\right\|^2 + 2\left\|\mathbf{X}^k-d\mathbf{AKA}'\right\|^2, \end{aligned} \tag{31}$$

which attains its minimum with respect to d when

$$d=\frac{\mathrm{tr}(\mathbf{X}^k\mathbf{AKA}')}{\mathrm{tr}(\mathbf{AKA}'\mathbf{AKA}')}. \tag{32}$$

Other models

To fit other cDEDICOM models it is important to note that

$$f(\mathbf{A}, \mathbf{R}) = 2\left\|\mathbf{X} - \mathbf{A}\mathbf{R}\mathbf{A}'\right\|^2 = 2\left\|\mathrm{vec}(\mathbf{X}) - (\mathbf{A} \otimes \mathbf{A})\mathrm{vec}(\mathbf{R})\right\|^2, \quad (33)$$

and consequentely when we are fitting a cDEDICOM model, the update of $\mathbf{R}$ can be always treated as a constrained linear regression.

4. Simulation study

In the previous section, we have shown that the same model can be fitted by following different strategies. In particular, DEDICOM and GG models can be fitted by either requiring or not the weight matrix to be orthonormal. In this section we will perform a simulation study to compare the two options. The same study will be also used to test two different strategies in the choice of λ. In general, the results in Sect. 3.1 indicate that there exists a threshold, say λ_0, such that if $\lambda \geq \lambda_0$ then the value of $\mathbf{B}$, minimizer of (10), decreases (9). We do not know the exact value of λ_0, but we can compute only some upper bounds of it as shown in the previous sections. Then the first strategy is to use a value of λ equal to the proper upper bound (formula (18) or (20)). However, such upper bounds could be too far from the true λ_0 implying only a small decrease for the loss (9), since obviously for $\lambda \to +\infty$ then $f(\mathbf{B}, \mathbf{R}) \to f(\mathbf{A}, \mathbf{R})$. It follows that the speed of convergence of the algorithm may be improved by initially setting $\lambda = 0$ and then increasing this value till function (9) stops decreasing. In the sequel, this strategy will be called "forward selection" of λ and will be compared in our study with the one consisting in setting λ equal to the proper upper bound.

The design of the simulation is the following. We generated data matrices by a DEDICOM model according to the following four factors, each with two levels: a) low (10% of total variability) or high (60% of total variability) error level; b) $n = 20$ or $n = 40$ objects; c) $p = 2$ or $p = 4$ components. The DEDICOM model was fitted on each data matrix by using four different implementations of the algorithm: *i)* without orthonormality constraints and λ set to the upper bound (20) (no orth $\lambda = \mathrm{up}$); *ii)* without orthonormality constraints and forward selection of λ (no orth λ forw); *iii)* with orthonormality constraints and λ set to the upper bound (18) (orth λ =up); *iv)* with orthonormality constraints and forward selection of λ (orth λ forw). We recorded the Sum of Squared Residuals (SSR), the CPU time in seconds that has been used to run the algorithm till convergence and the number of iterations (one iteration corresponds to one update of $\mathbf{A}$ and $\mathbf{R}$). For each combination of factor levels (8 in total) we generated 250 data matrices, we ran the four implementations of the algorithm for fitting a DEDICOM model, computed the average of SSR, CPU time and number of iterations. The simulation has been implemented in MATLAB and the computations have been done on a PC equipped with a 2.0 GHz Pentium 4. The results are reported in Table 1.

From Table 1, we can see that the best strategy for the choice of λ, in terms of fit, number of iterations and CPU time, is the forward selection. Among these two algorithms the version with orthonormality constraints seems to be more precise

Table 1. Average Sum of Squared Residuals (SSR), CPU time in seconds (CPU secs) and number of iterations (#Iterations) for 4 different algorithms to fit the DEDICOM model applied on 2000 simulated data sets with different error level, number of objects and components

Error	#comp	#objects		no orth $\lambda = \text{up}$	no orth $\lambda = \text{forw}$	orth $\lambda = \text{up}$	orth λ forw
Low	2	20	SSR%	10.99	9.854	10.616	9.851
			CPU secs	0.603	0.011	0.079	0.016
			#Iterations	320.256	11.224	53.700	13.024
		40	SSR%	11.324	10.138	10.511	9.808
			CPU secs	1.451	0.017	0.181	0.019
			#Iterations	315.352	10.856	40.624	10.180
	4	20	SSR%	8.202	8.015	8.158	8.006
			CPU secs	4.919	0.009	0.521	0.015
			#Iterations	1085.512	12.964	126.076	14.796
		40	SSR%	9.166	9.069	9.168	9.083
			CPU secs	16.737	0.015	2.167	0.027
			#Iterations	886.388	12.484	118.116	14.532
High	2	20	SSR%	49.473	48.627	49.181	47.964
			CPU secs	0.703	0.150	0.127	0.099
			#Iterations	304.488	25.108	71.524	48.520
		40	SSR%	50.620	48.693	50.219	48.416
			CPU secs	2.582	0.030	0.474	0.101
			#Iterations	426.628	26.068	87.092	38.256
	4	20	SSR%	33.052	33.032	32.972	32.991
			CPU secs	5.082	0.023	0.669	0.030
			#Iterations	842.756	25.700	123.128	26.008
		40	SSR%	39.757	39.581	39.647	39.503
			CPU secs	23.381	0.028	3.346	0.036
			#Iterations	872.640	22.704	126.304	22.372

while the other one seems to be faster. However, the differences are, in our opinion, not significant.

The Generalized GIPSCAL (GG) model was fitted to the same datasets by using the same four different implementations of the algorithm as before. The results are reported in Table 2.

In this case we obtained results similar to the DEDICOM case. The best strategy for λ is the forward selection with orthonormality constraints. The difference is that now the algorithm with orthonormality constraints and forward selection for λ is even the fastest.

From the simulation study we conclude that the forward selection for λ gives the best results and, whenever it is possible, the algorithm that uses orthonormality constraints on the loading matrix should be preferred.

Table 2. Average Sum of Squared Residuals (SSR), CPU time in seconds (CPU secs) and number of iterations (#Iterations) for 4 different algorithms to fit the Generalized GIPSCAL model applied on 2000 simulated data sets with different error level, number of objects and components

Error	#comp	#objects		no orth λ =up	no orth λ forw	orth λ = up	orth λ forw
Low	2	20	SSR%	44.671	43.534	43.379	43.389
			CPU secs	1.446	0.325	0.173	0.015
			#Iterations	757.164	376.096	120.208	10.716
		40	SSR%	44.515	44.297	43.504	43.530
			CPU secs	3.191	0.655	0.432	0.024
			#Iterations	709.384	330.520	108.148	11.648
	4	20	SSR%	35.310	35.441	35.278	35.204
			CPU secs	15.408	0.688	1.141	0.029
			#Iterations	3482.100	446.780	275.880	17.636
		40	SSR%	36.375	37.124	36.549	36.299
			CPU secs	63.538	0.630	5.262	0.047
			#Iterations	3551.468	283.036	292.700	17.492
High	2	20	SSR%	65.022	65.499	65.827	65.293
			CPU secs	1.011	0.356	0.210	0.129
			#Iterations	437.536	307.688	117.108	59.520
		40	SSR%	67.061	69.117	68.306	67.919
			CPU secs	2.816	0.526	0.849	0.190
			#Iterations	453.040	290.900	156.288	64.644
	4	20	SSR%	53.779	53.880	53.724	53.672
			CPU secs	19.951	0.605	1.221	0.155
			#Iterations	3432.924	327.348	221.400	52.540
		40	SSR%	57.544	57.769	57.501	57.478
			CPU secs	91.908	0.611	5.922	0.202
			#Iterations	3528.576	250.500	222.600	51.256

5. Discussion

In this paper we present a new algorithm to fit cDEDICOM models, which can be considered a generalization of the algorithm proposed by Kiers et al. (1990) for fitting the DEDICOM model. The algorithm follows an alternating least squares scheme, is very easy to implement and, as shown in the simulation study, it is also very fast. In the simulation study we tested two different strategies of selection for λ. However, other strategies can be implemented. For example, we can first minimize (10) with $\lambda = 0$, compute the loss (9) and repeat the minimization if the loss does not decrease by setting λ to the proper upper bound (18) or (20). In a simulation study, not reported here for sake of brevity, this strategy, suggested by Kiers et al. (1990), has been tested by giving results only slightly worse than the forward strategy.

The algorithm here proposed has been tested for the DEDICOM and Generalized GIPSCAL models but it can be also used to fit the Generalized Escoufier

& Grorud, Escoufier & Grorud and Gower decomposition models. However, we recall that for the last two models it can be also used a routine able to compute the singular value decomposition of a real matrix (see Escoufier & Grorud, 1980, and Gower, 1977). Finally, we note that the algorithm can be easily extended to fit every cDEDICOM model, i.e. a model which can be considered as a constrained DEDICOM model by assuming a particular form for $\mathbf{R}$.

References

Chino N (1990) A generalized inner product model for the analysis of asymmetry. *Behaviormetrika* 27: 25–46

Cliff N (1966) Orthogonal rotation to congruence. *Psychometrika* 31: 33–42

Escoufier Y, Grorud A (1980) Analyse Factorielle des matrices carrees non symetriques. In: Diday E et al. (ed) *Data Analysis and Informatics*, vol 1. North Holland, Amsterdam, pp 263–276

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53: 325–338

Gower JC (1977) The analysis of asymmetry and orthogonality. In: Barra JR et al. (eds) *Recent developments in Statistics*. North-Holland, Amsterdam, pp 109–123

Harshman RA, Green PE, Wind Y, Lundy ME (1982) A Model for the Analysis of Asymmetric Data in Marketing Research. *Marketing Science* 1: 205–242

Kiers HAL, TAKANE Y (1994) A generalization of GIPSCAL for the Analysis of Nonsymmetric Data. *Journal of Classification* 11: 79–99

Kiers HAL (1989) An alternating least squares algorithm for fitting the two- and three-way DEDICOM model and the IDIOSCLA model. *Psychometrika* 54: 515–521

Kiers HAL, Ten Berge JMF, Takane Y, De Leeuw J (1990) A generalization of Takane's algorithm for DEDICOM. *Psychometrika* 55: 151–158

Rocci R, Bove G (2002) Rotation Techniques in Asymmetric Multidimensional Scaling. *Journal of Computational & Graphical Statistics* 11: 405–419

Zielman B, Heiser WJ (1996) Models for asymmetric proximities. *British Journal of Mathematical and Statistical Psychology* 49: 127–147