Journal of Classification 15:245-263 (1998)

Joint Orthomax Rotation of the Core and Component Matrices Resulting from Three-mode Principal Components Analysis

Henk A.L. Kiers

University of Groningen

Abstract: The analysis of a three-way data set using three-mode principal components analysis yields component matrices for all three modes of the data, and a three-way array called the core, which relates the components for the different modes to each other. To exploit rotational freedom in the model, one may rotate the core array (over all three modes) to an optimally simple form, for instance by three-mode orthomax rotation. However, such a rotation of the core may inadvertently detract from the simplicity of the component matrices. One remedy is to rotate the core only over those modes in which no simple solution for the component matrices is desired or available, but this approach may in turn reduce the simplicity of the core to an unacceptable extent. In the present paper, a general approach is developed, in which a criterion is optimized that not only takes into account the simplicity of the core, but also, to any desired degree, the simplicity of the component matrices. This method (in contrast to methods for either core or component matrix rotation) can be used to find solutions in which the core and the component matrices are *all* reasonably simple.

Keywords: Three-way data; Simple structure rotation; Varimax.

The author is obliged to Marieke Timmerman and Jos ten Berge for useful comments on an earlier version of this paper, and to Pieter Kroonenberg for providing a computerized version of the Osgood and Luria data.

Author's address: Henk A.L. Kiers, Heymans Institute, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands; email: h.a.l.kiers@ppsw.rug.nl

1. Introduction

Tucker's (1966) three-mode factor analysis, or, more appropriately, three-mode principal components analysis (3MPCA), is a method for analyzing three-way data by using three sets of components (one for each mode) and a so-called core array relating the three sets of components to each other. For example, three-way data may consist of scores of individuals (mode A), on a set of variables (mode B), at a number of different occasions (mode C). Let the data array be denoted by **X**, and its elements by x_{ijk}, where i refers to individual i, j refers to variable j, and k refers to occasion k, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$. Then the 3MPCA model can be described as

$$\hat{x}_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip}\, b_{jq}\, c_{kr}\, g_{pqr}\ , \qquad (1)$$

where $\hat{x}_{ijk}$ is the model estimate for x_{ijk}, **A**, **B**, and **C** (with elements a_{ip}, b_{jq}, and c_{kr}, respectively) are component matrices of orders $I \times P$, $J \times Q$, and $K \times R$, respectively, **G** is a $P \times Q \times R$ three-way array known as the core, with elements g_{pqr}, $p = 1, \ldots, P$, $q = 1, \ldots, Q$, and $r = 1, \ldots, R$. The elements of the core indicate the joint impact of the components from the different modes. As proposed by Kroonenberg and de Leeuw (1980), the 3MPCA model is fitted to the data by minimizing the sum of squared residuals, $\Sigma_i \Sigma_j \Sigma_k (x_{ijk} - \hat{x}_{ijk})^2$, over **A**, **B**, **C**, and **G**.

As already noted by Tucker (1966), the 3MPCA model is not fully determined. Specifically, he showed that postmultiplying **A**, **B**, and **C** by nonsingular matrices can always be compensated for by applying the inverses of these matrices to the core array. That is, it can be verified that $\tilde{\mathbf{A}} = \mathbf{AS}^{-1}$, $\tilde{\mathbf{B}} = \mathbf{BT}^{-1}$, $\tilde{\mathbf{C}} = \mathbf{CU}^{-1}$, and $\tilde{\mathbf{G}}$ with elements

$$\tilde{g}_{uvw} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} s_{up} t_{vq} u_{wr} g_{pqr}\ , \qquad (2)$$

$u = 1, \ldots, P$, $v = 1, \ldots, Q$, and $w = 1, \ldots, R$, give the same model estimates (in $\hat{\mathbf{X}}$) as **A**, **B**, **C**, and **G** do.

To exploit the above indeterminacy in 3MPCA, Kiers (1997) suggested orthogonally rotating the core to simple structure (i.e., with most elements either large or small in absolute sense), so as to enhance interpretability of the core. Specifically, he proposed to maximize a class of criteria that measure simplicity of the core as the sum of the "orthomax" (see Jennrich 1970) values of the core elements in respectively all frontal, lateral, and horizontal core planes. The maximization can be performed over rotation matrices for all three modes, or a subset thereof.

Kiers' (1997) three-mode orthomax approach can be used successfully for simplifying the core, but may at the same time reduce the interpretability of the component matrices. Often interpretability is enhanced by rotation to simple structure, although sometimes component matrices are more easily interpreted when they contain other patterns, such as polynomial trends (when data entries refer to successive time points). Here, we deal only with situations where interpretability is enhanced by simple stucture. To avoid problems in interpretation of component matrices, Kiers proposed to first rotate a subset of the component matrices to simple structure, and next rotate the core only in the directions for which no simple associated component matrix is desired. However, this two-step approach need not always work. Specifically, when simple component matrices are desired for all three modes, fixing them at a simple solution would leave no opportunity to rotate the core to some degree of simplicity as well. Similarly, even when only one or two component matrices are fixed, it may be impossible to simplify the core as much as desired. This two-step approach emphasizes simplicity of the subset of component matrices, and simplicity of the core is deemed of secondary importance. A more balanced approach would attach equal importance to simplicity of the core and to simplicity of the component matrices, or, involve any other weighting of importances. *It is the purpose of the present paper to develop a general procedure for joint orthomax rotation of the core and of the component matrices, allowing for any combination of weights attached to simplicity of the core versus simplicity of the component matrices.* The present approach subsumes the three-mode orthomax core rotation approach as a special case; the present approach, however, has a considerably wider scope in that any position can be taken with respect to the tradeoff between simplicity of the core and of the component matrices, whereas with the two-step approach one can only choose between the extreme positions. In particular, one can search for a solution where the core is considerably simpler than with the two-step approach, and where the component matrices are still sufficiently simple for easy interpretation.

The next section proposes a general criterion for joint simplicity of the core and of the component matrices and explains how different sets of weights can be used to obtain different tradeoffs between simplicity of the core and simplicity of the component matrices, an extreme case of which is that used in the two-step approach. Then, it describes how the joint orthomax criterion can be optimized over **S**, **T**, and **U**. The performance of the algorithm is assessed in a small Monte Carlo study. Finally, the method is illustrated by means of an analysis of empirical data.

2. A Criterion for Joint Orthomax Simplicity of the Core and of the Component Matrices

The present section supplements Kiers' (1997) orthomax criterion for the core with terms assessing orthomax simplicity of the component matrices. The orthomax criterion (Crawford and Ferguson 1970; Jennrich 1970), which has varimax (Kaiser 1958) and quartimax (Carroll 1953; Saunders 1953; Ferguson 1954; Neuhaus and Wrigley 1954) as special cases, is, for an $m \times r$ loading matrix Λ, given by m^{-1} times the function

$$f_{OR}(\Lambda, \gamma) = \sum_{l=1}^{r} \left[\sum_{i=1}^{m} \lambda_{il}^4 - \frac{\gamma}{m} \left[\sum_{i=1}^{m} \lambda_{il}^2 \right]^2 \right], \tag{3}$$

where λ_{il} denotes the element (i,l) of Λ, and γ is the parameter monitoring the choice of the orthomax criterion (e.g., $\gamma = 0$ yields the quartimax criterion, and $\gamma = 1$ yields the varimax criterion). The three-mode orthomax criterion proposed by Kiers is given by

$$f(\mathbf{S},\mathbf{T},\mathbf{U}) = \sum_{l=1}^{3} w_l f_{OR}(\tilde{\mathbf{G}}^l, \gamma_l), \tag{4}$$

where $\tilde{\mathbf{G}}^1$, $\tilde{\mathbf{G}}^2$, and $\tilde{\mathbf{G}}^3$ denote the matrices whose columns consist respectively of the vectorized horizontal, lateral and frontal slabs of the core $\tilde{\mathbf{G}}$, which is the original core array rotated by the orthonormal matrices **S**, **T**, and **U**, according to (2); w_1, w_2, and w_3 are fixed prespecified weights; and γ_1, γ_2, and γ_3 are prespecified values of the γ parameter in the orthomax function. The first term measures simplicity of the (horizontal) slabs associated with the individuals, the second term measures simplicity of the (lateral) slabs associated with the variables, and the third term measures simplicity of the (frontal) slabs associated with the occasions.

To supplement this criterion with terms measuring simplicity of the A-, B-, and C-mode component matrices, we propose here to add to (4) the orthomax criteria applied to $\tilde{\mathbf{A}} = \mathbf{AS}'$, $\tilde{\mathbf{B}} = \mathbf{BT}'$, and $\tilde{\mathbf{C}} = \mathbf{CU}'$, multiplied by an importance weight. We thus obtain the joint orthomax criterion for the core and the component matrices:

$$f_{cc}(\mathbf{S},\mathbf{T},\mathbf{U}) = \sum_{l=1}^{3} w_l f_{OR}(\tilde{\mathbf{G}}^l, \gamma_l) + w_A f_{OR}(\tilde{\mathbf{A}}, \gamma_A)$$

$$+ w_B f_{OR}(\tilde{\mathbf{B}}, \gamma_B) + w_C f_{OR}(\tilde{\mathbf{C}}, \gamma_C), \tag{5}$$

where w_A, w_B, and w_C denote the weights attached to the orthomax criteria applied respectively to $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$, with respective γ-parameters γ_A, γ_B, and

γ_C. Joint orthomax rotation of the core and the component matrices consists of maximizing f_{cc} for a priori chosen weights w_1, w_2, w_3, w_A, w_B, and w_C, and a priori chosen orthomax parameters γ_1, γ_2, γ_3, γ_A, γ_B, and γ_C, with $\tilde{\mathbf{G}}$ as defined in (2), and $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$ as defined above, over orthonormal matrices **S**, **T**, and **U**. Here, we assume that the matrices **A**, **B** and **C** are columnwise orthonormal, and hence so will be $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$. However, the joint orthomax procedure to be presented here may just as well be applied to matrices **A**, **B**, and **C** that are normalized in a different manner (see also Discussion), but this would require a reconsideration of the choices of weights to be discussed in the next subsection.

2.1 Choosing the Weights in the Joint Orthomax Criterion

The weights and the orthomax parameters in the joint orthomax criterion can be specified at will. However, in practice, some guidelines for these choices may be helpful. First the weights for the terms corresponding to the core (w_1, w_2, and w_3) could be chosen as described and motivated by Kiers (1997). Note that he has shown that the choices of the (nonzero) weights and of the gamma parameters are related, and therefore these weights can be fixed in a conceptually useful way, without loss of generality. He advocated using the "natural" weights $w_1 = (QR)^{-1}$, $w_2 = (PR)^{-1}$, and $w_3 = (PQ)^{-1}$, or taking $w_l = 0$ if simplicity in direction l (l = 1,2,3) is to be ignored. Next, the choices of the weights for the terms corresponding to simplicity of **A**, **B**, and **C** could be chosen such that the associated criterion values are of the same order of magnitude, both in comparison to each other, and in comparison to the terms associated with simplicity of the core. As a starting point for finding such "standard" weights, we consider two benchmark situations: In the first, all component matrices and the core are absolutely *not simple* (NS), having all elements equal, in absolute sense; in the second situation, all component matrices and the core are highly *simple* in a *systematic* way (SS), with, per column or plane, *half* the elements equal, in absolute sense, and nonzero, and the other half of the elements equal to 0. In these benchmark situations the amount of simplicity in the core and in each of the component matrices can be judged equal, and hence we choose the "standard" weights such that the ensuing joint orthomax criterion reflects this equality.

To derive weights that satisfy the above requirement, we first consider the NS situation. In the NS situation, the columnwise orthonormality of **A** ensures that **A** has all *squared* elements equal to I^{-1}, hence, the orthomax value, for any choice of γ, is, according to (3),

$$f_{OR}(\mathbf{A},\gamma) = \sum_{l=1}^{P}\sum_{i=1}^{I}(I^{-1})^2 - \gamma I^{-1}\sum_{l=1}^{P}\left[\sum_{i=1}^{I} I^{-1}\right]^2$$

$$= PII^{-2} - \gamma I^{-1}P = I^{-1}P(1-\gamma)\,.$$

Similarly, $f_{OR}(\mathbf{B},\gamma) = J^{-1}Q(1-\gamma)$ and $f_{OR}(\mathbf{C},\gamma) = K^{-1}R(1-\gamma)$. Furthermore, let g denote each squared element in the core; hence $g = (PQR)^{-1}\|\mathbf{G}\|^2$. Then $f_{OR}(\mathbf{G}^1,\gamma) = (PQRg^2 - \gamma(QR)^{-1}P(QRg)^2 = (PQRg^2 - \gamma PQRg^2) = PQRg^2(1-\gamma) = (PQR)^{-1}\|\mathbf{G}\|^4(1-\gamma)$; similarly, $f_{OR}(\mathbf{G}^2,\gamma) = f_{OR}(\mathbf{G}^3,\gamma) = (PQR)^{-1}\|\mathbf{G}\|^4(1-\gamma)$. Hence, the total contribution of the core is $(w_1 + w_2 + w_3)\ (PQR)^{-1}\|\mathbf{G}\|^4(1-\gamma)$. It is now clear that, by taking $w_A = IP^{-1}$, $w_B = JQ^{-1}$, and $w_C = KR^{-1}$, the contributions of the last three orthomax terms in (5) equal $(1-\gamma)$, and hence are equal to each other; the total contribution of the core is normalized to $(1-\gamma)$ by dividing the earlier defined weights by $(w_1 + w_2 + w_3)\ (PQR)^{-1}\|\mathbf{G}\|^4$. By this procedure of normalizing the earlier chosen "natural" weights for the three core terms, we keep the relative contributions of the three core terms the same as with the natural weights advocated by Kiers (1997). It can be shown analogously that, using the above weights, the contributions of the component matrices, and the total contribution of the core all equal $(2-\gamma)$ in the SS situation. Because, in the two benchmark situations, these weights lead to equal influences of the separate parts of the 3MPCA solution, we consider these weights as "standard" weights:

$$w_1 = v_1(v_1 + v_2 + v_3)^{-1}(PQR)\|\mathbf{G}\|^{-4}, \text{ where } v_1 = (QR)^{-1} \text{ or } v_1 = 0;$$

$$w_2 = v_2(v_1 + v_2 + v_3)^{-1}(PQR)\|\mathbf{G}\|^{-4}, \text{ where } v_2 = (PR)^{-1} \text{ or } v_2 = 0;$$

$$w_3 = v_3(v_1 + v_2 + v_3)^{-1}(PQR)\|\mathbf{G}\|^{-4}, \text{ where } v_3 = (PQ)^{-1} \text{ or } v_3 = 0;$$

$$w_A = IP^{-1};$$

$$w_B = JQ^{-1};$$

$$w_C = KR^{-1}.$$

The standard weights chosen above lead to the "standard joint orthomax criterion." It can be verified that the criterion has two additional useful properties:

1. The standard joint orthomax criterion is insensitive to an overall rescaling of the original data, and hence of the array **G**.
2. The standard joint orthomax criterion is insensitive to replacing one of the component matrices by a supermatrix consisting of two or more of the same, below each other. Thus such a change of the

component matrix which does affect the *size* but not the *simplicity* of the matrix does not affect the value of the joint orthomax criterion. This observation demonstrates that the criterion value does not depend intrinsically on the size of the data matrix.

We do not claim that the standard weights chosen above are the only or even the best ones to equalize the influence of the different terms. However, they do fulfil the practical requirement of having a standard (by convention) combination of the orthomax criteria equalizing the influence of different terms, and which is not sensitive to unimportant changes (like scale) in the data. Therefore, these standard weights can be used as a reference point. When, in practice, one chooses weights different from the standard weights (e.g., to attach more importance to some criteria than to others), this difference can most easily be referred to by expressing the weights relative to the standard weights.

As an extreme choice of weights, one might set certain weights to 0, and hence attach no importance whatsoever to the simplicity of certain matrices. In fact, in this way the two-step approach (where, first, certain component matrices are chosen optimally by means of orthomax rotation and then the core is simplified by rotation over the other rotation matrices) emerges as the special case in which certain weights are chosen equal to 0, certain weights are taken infinite, and the other weights are taken as the standard weights. For example, when $\tilde{\mathbf{B}}$ is obtained by varimax rotation (and this rotation is compensated for in the core), and afterwards orthomax rotation is applied to the core, using only **S** and **U**, this approach is equivalent to maximization of f_{cc} with w_1, w_2, and w_3 taken "standard," $w_A = w_C = 0$ and w_B infinite.

In addition to the weights, one has to choose the orthomax parameters. These can be chosen differently for each term in the criterion f_{cc}, depending on the actual orthomax criterion one wishes to employ for assessing simplicity in each of the six different respects. For instance, one might choose $\gamma_1 = \gamma_2 = \gamma_3 = 0$ to ensure that quartimax is used for assessing simplicity of the core, and at the same time use $\gamma_A = 1$, $\gamma_B = 1$, and $\gamma_C = 1/2$ so as to use varimax for rotation of **A** and **B** and a blend of varimax and quartimax for assessing simplicity of **C**. For conceptual ease, one will tend to use the same orthomax criterion for all component matrices, which will often be the varimax criterion.

3. An Algorithm for Joint Orthomax Rotation of the Core and of the Component Matrices

To maximize (5) over **S**, **T**, and **U** we propose to use an algorithm which is a modification of the three-way orthomax algorithm proposed by Kiers (1997). As in that paper, we propose to increase the function value, after choosing initial values for the rotation matrices **S**, **T**, and **U**, by iteratively updating them until the function value stabilizes. We only describe the update for **S**, because **T** and **U** can be updated fully analogously. To update **S**, we have to maximize

$$f_{cc}(\mathbf{S},*) = \sum_{l=1}^{3} w_l f_{OR}(\tilde{\mathbf{G}}^{l},\gamma_l) + w_A f_{OR}(\mathbf{AS}',\gamma_A)\,,$$

in which the first term can be combined into one weighted orthomax function (see Kiers 1997) to yield

$$\begin{aligned} f_{cc}(\mathbf{S},*) &= (w_1+w_2+w_3) f_{OR}(\mathbf{H}_f'\mathbf{S}',\gamma_1 w_1 (w_1+w_2+w_3)^{-1}) \\ &\quad + w_A f_{OR}(\mathbf{AS}',\gamma_A) \\ &= f_{OR}((w_1+w_2+w_3)^{1/4}\mathbf{H}_f'\mathbf{S}',\gamma_1 w_1 (w_1+w_2+w_3)^{-1}) \\ &\quad + f_{OR}(w_A^{1/4}\mathbf{AS}',\gamma_A)\,, \end{aligned} \tag{7}$$

where $\mathbf{H}_f$ denotes the supermatrix containing the frontal slabs of the core rotated by the current matrices **T** and **U**. Clearly, (7) consists of the sum of two orthomax functions, and hence cannot be optimized by the standard orthomax procedure, as is done in the core rotation approach. To maximize (7) over **S**, we use an alternative iterative planar rotation procedure, which finds the optimal **S** by iteratively updating all pairs of columns of $\mathbf{S}'$, where each pairwise updating finds the optimal rotation of the current $\mathbf{S}'$ in a plane.

To find the optimal rotation angle, we can use Jennrich's (1970) general procedure for maximization of a "symmetric fourth degree rotation criterion," where "symmetric" refers to the fact that the criterion, as a function of the rotation angle θ, is invariant under an interchange of the columns or a reflection of the columns of $\mathbf{S}'$. The present criterion clearly is symmetric (as is the original orthomax criterion to which Jennrich applied his procedure). Because of this symmetry, the criterion can be written using θ as $F(\theta) = c + a\cos(4\theta) + b\sin(4\theta)$ (see Jennrich 1970), where c denotes a constant, and a and b denote two coefficients which Jennrich suggests computing via evaluation of F at three values of θ. A somewhat more efficient procedure is to use the explicit expressions for a and b as derived by Clarkson and Jennrich (1988, equations (13) and (15)) for the orthomax criterion. In the present case, where we have a sum of two orthomax criteria, we can obtain a

and b by simply summing the corresponding coefficients found for the two orthomax criteria at hand. In fact, this procedure is essentially a generalization of the planar rotation approach employed by Hakstian's (1976) procedure for joint *varimax* rotation of two matrices.

3.1 Testing the Algorithm by a Small Monte Carlo Study

We programmed the algorithm in MATLAB (Mathworks Inc. 1992); the associated m-files are available from the author upon request. After some preliminary runs, we tested the algorithm more systematically with a small Monte Carlo study. The primary aim was to see whether the algorithm is prone to hitting local optima, and whether it is computationally feasible (i.e., not prohibitively slow).

We constructed 80 sets consisting of component matrices **A**, **B**, and **C**, and a core array **G**. The first twenty of these were constructed as random rotations of the simple matrices

$$\mathbf{A}_0 = \frac{1}{2}\begin{bmatrix}\mathbf{I}_3\\ \mathbf{I}_3\\ \mathbf{I}_3\\ \mathbf{I}_3\end{bmatrix},\quad \mathbf{B}_0 = \left(\frac{1}{3}\right)^{1/2}\begin{bmatrix}\mathbf{I}_3\\ \mathbf{I}_3\\ \mathbf{I}_3\end{bmatrix},\quad \mathbf{C}_0 = \left(\frac{1}{2}\right)^{1/2}\begin{bmatrix}\mathbf{I}_3\\ \mathbf{I}_3\end{bmatrix},$$

and associated counter-rotations of the simple core array $\mathbf{G}_0$ with frontal planes

$$\begin{bmatrix}1&0&0\\0&0&0\\0&1&0\end{bmatrix},\quad \begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix},\quad \text{and}\quad \begin{bmatrix}0&1&0\\0&0&0\\0&0&1\end{bmatrix},$$

respectively (where the core was chosen in the same way as in Kiers 1997). For these rotated versions of the simple matrices and the core, we expected that the maximum of the joint orthomax function is attained for $\mathbf{A}_0$, $\mathbf{B}_0$, $\mathbf{C}_0$, and $\mathbf{G}_0$. The next sixty sets were also based on randomly rotated versions of $\mathbf{A}_0$, $\mathbf{B}_0$, $\mathbf{C}_0$, and $\mathbf{G}_0$, but in addition to these rotations, various amounts of random noise were added to the three matrices and to the core. The rationale behind this construction of arrays was that, at least when little noise is added, the search for a simple solution for all three matrices as well as the core will be meaningful. Specifically, when noise was added, we constructed **A** as

$$\mathbf{A} = GS(\mathbf{A}_0\mathbf{S}_r + \alpha\mathbf{N}),$$

where $GS()$ denotes the Gram-Schmidt orthonormalization of the matrix in parentheses, $\mathbf{S}_r$ denotes a random rotation matrix, α the noise level (varying as $(.2)^{1/2}$, $(.4)^{1/2}$, and $(.6)^{1/2}$, for groups of twenty data sets each), and **N** a matrix with random elements drawn from a normal distribution with variance equal to the sum of squares of the elements of $\mathbf{A}_0$ and hence of $\mathbf{A}_0\mathbf{S}_r$. Thus, 100 α^2 can be interpreted as the added percentage of noise. **B** and **C** are constructed analogously from $\mathbf{B}_0$ and $\mathbf{C}_0$; **G** is constructed from $\mathbf{G}_0$ by first applying the inverses of the rotations used for the component matrices to it, next adding noise to the core (in the same way, using the same percentages, as for the component matrices), and finally applying the inverses of the transformations involved in the Gram-Schmidt orthonormalizations (of the component matrices) to the core. This procedure guarantees that, as the noise level tends to zero, the set $\{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{G}\}$ tends toward a rotated version of $\{\mathbf{A}_0,\mathbf{B}_0,\mathbf{C}_0,\mathbf{G}_0\}$.

As has been mentioned, the main interest in this Monte Carlo study was to investigate computational efficiency and proneness to local optima. Each data set was analyzed by the standard joint *varimax* procedure (i.e., all gamma parameters were taken equal to 1), using five randomly started runs of the algorithm. The best of the resulting five solutions was considered the global optimum; the others were considered local optima if they differed from the purported global optimum by more than .01%; we will use this terminology even though it is technically not correct, because the best value need not be the best attainable, and the "local optima" may refer to solutions that are not even locally optimal (e.g., because they may refer to saddle points, or solutions not yet converged). The algorithm was considered to have converged if a full iteration (i.e., a cycle in which all three rotation matrices in turn are optimized, given the others) failed to increase the criterion by more than .0001%.

We first checked how many of the 400 runs led to local optima. There were only three, so it can be concluded that, at least for the standard varimax procedure and the present type of data, the algorithm is not vulnerable to local optima.

We also checked to see if the purported global optima for the first twenty data sets corresponded to the underlying simple solution given by $\{\mathbf{A}_0,\mathbf{B}_0,\mathbf{C}_0,\mathbf{G}_0\}$; the result was positive. For the other 60 data sets, no such absolute check was available, but we verified that, on average, the solutions became less similar to the basic underlying set $\{\mathbf{A}_0,\mathbf{B}_0,\mathbf{C}_0,\mathbf{G}_0\}$ as the noise level increased. A detailed study of similarity to the underlying sets was not carried out, because even solutions with low similarity to the underlying sets might nevertheless be simple.

The second concern in this Monte Carlo study is the efficiency of the algorithm. On a Pentium 100MHz pc, each full analysis (including five runs) required less than 47 seconds, while the mean was only 13.6 seconds. The problem size employed here can be considered fairly representative of that encountered in practice. However, to see what happens for considerably larger sizes, two sets with $I = J = K = 40$ and $P = Q = R = 4$ (which can be considered large) were run, and they took respectively 26 and 12 seconds.

Therefore, it seems safe to conclude that the algorithm is not slow, and is thus feasible in practice.

4. Illustrative Analysis

To illustrate the joint varimax procedure, and the manner in which weights can be chosen satisfactorily, we chose to reanalyze a reasonably sized data set ($15 \times 10 \times 6$), which has been published completely, and previously analyzed with 3MPCA by Kroonenberg (1983, 1985): the data published by Osgood and Luria (1954; as reprinted in Snider and Osgood 1969, pp. 505-517). The data are scores of a single psychiatric patient with respect to 15 concepts (see Table 1), on 10 semantic differential scales (see Table 1), over six administrations, each referring to a particular state this alleged multiple personality patient was in (''Eve White'' in two cases, ''Eve Black'' in two other cases, and ''Jane'' in the last two cases). For more details, see Osgood and Luria (1954). Following Kroonenberg (1983, pp. 228-241; see also Kroonenberg 1985), we analyzed these data by a 3MPCA and preprocessed the data just as he did (by only subtracting the scale mean). For the dimensonalities of the three component matrices, we decided to take $P = 3$, $Q = 3$, and $R = 2$, on the basis of a comparison of results of 3MPCA's with varying dimensionalities. The ensuing solution accounted for 72.2% of the sum of squares of the preprocessed data. The resulting unrotated, or, ''principal axes'' solution (i.e., where **A** contains the leading eigenvectors of $\mathbf{X}_f(\mathbf{C}'\mathbf{C} \otimes \mathbf{B}'\mathbf{B})\mathbf{X}_f'$, with $\mathbf{X}_f$ the supermatrix with frontal planes of **X** next to each other, and **B** and **C** contain the leading eigenvectors of analogously defined matrices) is given in Table 1.

4.1 Choosing the Weights for the Joint Varimax Criterion

It can be seen from Table 1 that, in the principal axes solution, matrix **C** is quite simple, but for **A** and **B**, as well as for the core, a further simplification would help. Therefore, we applied our joint orthomax method, using the standard weights, as well as various modifications of these weights, to put more emphasis on simplicity of certain matrices and less on that of others. In all cases we set all gamma parameters to 1, thus using varimax throughout. The modified weights are expressed relative to the standard weights. Table 2 contains an overview of the varimax values of the core and of component matrices for all combinations we employed. The first line gives the varimax values for the unrotated solution. The second gives these values for the case where the weights for the component matrices are all 0. This case is equivalent to three-mode varimax and hence yields the maximally attainable varimax value for the core (which is underlined for that reason).

TABLE 1

Unrotated (Principal Axes) 3MPCA Solution for the Osgood and Luria data

A (concepts)	1	2	3
my doctor	.19	.54	–.03
peace of mind	.19	.45	–.19
my father	.25	.32	.17
my mother	.23	.19	.30
self-control	.29	.20	–.02
love	.40	–.12	–.02
sex	.26	–.24	–.03
fraud	–.31	.10	.37
hatred	–.26	.13	.43
my spouse	.24	–.17	–.09
me	–.08	.31	.23
my job	.33	–.17	.08
mental sickness	.25	–.18	.43
confusion	.10	–.19	.50
child	.30	–.05	.12

B (scales)	1	2	3
small-large	.34	–.09	.28
dirty-clean	.39	–.13	.24
valuable-worthless	–.42	.01	–.17
tasty-distasteful	–.38	–.24	–.01
weak-strong	.33	.26	–.04
deep-shallow	–.31	.42	–.01
tense-relaxed	.21	.62	–.31
fast-slow	–.04	.22	.77
active-passive	–.30	.45	.30
cold-hot	.25	.18	–.24

C (administrations)	1	2
Eve White 1	.48	.09
Eve White 2	.44	.15
Eve Black 1	–.09	.64
Eve Black 2	–.13	.75
Jane 1	.51	.05
Jane 2	.54	.03

Frontal Core Slabs

	C1			C2		
	B1	B2	B3	B1	B2	B3
A1	43.42	–0.98	7.48	–12.30	–4.39	3.91
A2	9.76	0.70	0.49	32.61	–1.79	–8.69
A3	–1.49	–21.72	1.56	–0.88	–8.97	–4.76

TABLE 2

Varimax Values for the Core and the Component Matrices for Various Joint Varimax Rotations Applied to the 3MPCA Solution for the Osgood and Luria Data

Relative weight for			Varimax value of			
A	B	C	Core	A	B	C
unrotated			4.23	1.06	1.48	1.16
0	0	0	<u>5.12</u>	0.95	1.65	0.84
0.5	0.5	0.5	5.06	0.88	1.92	1.04
1.0	1.0	1.0	4.99	0.85	1.95	1.13
1.5	1.5	1.5	4.92	0.84	1.97	1.18
2.0	2.0	2.0	4.85	0.84	1.99	1.20
2.5	2.5	2.5	4.78	0.85	2.01	1.21
3.0	3.0	3.0	3.84	1.15	2.01	1.24
3.5	3.5	3.5	3.28	1.31	2.03	1.23
4.0	4.0	4.0	2.82	1.41	2.05	1.23
4.5	4.5	4.5	2.45	1.48	2.07	1.24
5.0	5.0	5.0	2.14	1.53	2.08	1.24
5.5	5.5	5.5	1.88	1.56	2.09	1.25
6.0	6.0	6.0	1.65	1.59	2.10	1.25
100.0	100.0	100.0	0.75	1.66	2.12	1.27
∞	∞	∞	0.75	<u>1.66</u>	<u>2.12</u>	<u>1.27</u>
3.0	1.0	1.0	4.78	1.29	1.91	0.48
3.0	1.0	1.5	4.03	1.34	1.87	0.99
3.0	1.0	2.0	3.98	1.27	1.89	1.13
3.0	1.0	2.5	3.99	1.22	1.90	1.19

*) Underlined values indicate maximally attainable values.

The case where the weights for the component matrices are all 100, on the other hand, turns out to be almost equivalent to that where each component matrix is rotated by varimax separately, and the core is counter-rotated (which is indicated by infinite relative weights in Table 2). This case thus gives the other extreme, where the varimax values for the component matrices are maximal (and also underlined). The other cases are between these extremes, and can be compared to these extremes to choose one's favorite description.

It can be seen from Table 2 that, when the relative weights for **A**, **B**, and **C** together increase, the varimax value of the core decreases, and those for **B** and **C** increase, whereas that for **A** first *decreases*, but then again increases, when taking weights of 2.5 or higher. Interestingly, at the point where the values for **A** start increasing, the varimax value of the core drops considerably. Apparently, simplicity of the core and of **A** are largely incompatible, whereas simplicity of the core and of **B** and **C** is quite feasible: When the relative weights are all 2.5, the core and matrices **B** and **C** have varimax values larger than 93% of their maximal varimax values (whereas for **A** the varimax value is only 51% of its maximum).

In an attempt nevertheless to find solutions where both **A** and the core are reasonably simple, a number of choices for the relative weights were considered where the weight for **A** is the largest. When taking the relative weights equal to {3,1,1}, we did get reasonable simplicity of **A**, but then simplicity of **C** was reduced enormously. Therefore, in the next trials, we gradually increased the weight for **C** as well. This situation caused the varimax value for **C** to increase rapidly, whereas that for **A**, **B**, and the core did not change much. In fact, with the set of relative weights {3,1,2} we obtained varimax values for the core and the component matrices that were all higher than 75% of the associated maximal varimax values.

4.2 Interpreting the Results

From the above discussion of the results presented in Table 2, we conclude that the solutions with relative weights sets {2.5,2.5,2.5} and {3,1,2} strike the most interesting compromises of simplicity in the core and the component matrices. The first solution closely approaches maximal simplicity in the core and in **B** and **C**, at the cost of **A**; the second gives relatively high simplicity for the core as well as all three component matrices but is not very close to the maximally attainable simplicity. These solutions are given respectively in the left and right panels of Table 3; in the component matrices, values below −.35 and above .35 are underlined; in the cores, values below − 10 and above +10 are underlined. Matrix **A** in the first solution is not at all simple (with many intermediate values, i.e., between .20 and .35 in absolute sense), whereas the matrix **A** of the second solution is quite simple indeed. Therefore, we will only interpret the second solution.

The first component in the second solution refers to peace of mind, which is apparently related to the patient's father and doctor. The second component contrasts fraud and hatred against love and sex. The third

TABLE 3

Two Solutions from Joint Varimax Rotation of the Core and the Component Matrices of the 3MPCA Solution for the Osgood and Luria Data

	Relative Weights {2.5,2.5,2.5}			Relative Weights {3,1,2}		
A	1	2	3	1	2	3
my doctor	.30	.49	−.06	.52	.23	−.09
peace of mind	.28	.39	−.22	.46	.13	−.23
my father	.31	.27	.16	.40	.09	.16
my mother	.27	.16	.29	.30	.04	.30
self-control	.32	.13	−.03	.34	−.07	.00
love	.37	−.20	−.01	.20	−.36	.07
sex	.21	−.29	−.02	.02	−.35	.05
fraud	−.28	.18	.36	−.15	.37	.29
hatred	−.22	.20	.42	−.09	.37	.35
my spouse	.20	−.22	−.08	.05	−.30	−.02
me	−.01	.33	.21	.17	.32	.16
my job	.28	−.23	.09	.11	−.33	.16
mental sickness	.21	−.21	.44	.05	−.20	.49
confusion	.06	−.18	.51	−.07	−.08	.54
child	.28	−.10	.12	.18	−.21	.17
B	1	2	3	1	2	3
small-large	.38	−.13	.19	.35	−.16	.23
dirty-clean	.43	−.15	.14	.40	−.18	.18
valuable-worthless	−.44	.01	−.11	−.42	.04	−.15
tasty-distasteful	−.36	−.27	−.06	−.37	−.24	−.08
weak-strong	.30	.30	.02	.32	.27	.03
deep-shallow	−.34	.35	.17	−.32	.39	.13
tense-relaxed	.11	.71	−.08	.18	.70	−.10
fast-slow	.05	−.07	.80	−.03	−.03	.80
active-passive	−.30	.27	.47	−.31	.32	.43
cold-hot	.20	.29	−.19	.24	.26	−.18
C	1	2		1	2	
Eve White 1	.48	.07		.48	−.07	
Eve White 2	.45	.13		.47	.00	
Eve Black 1	−.07	.64		.13	.63	
Eve Black 2	−.10	.75		.13	.75	
Jane 1	.51	.03		.50	−.12	
Jane 2	.54	.01		.52	−.15	

Frontal Core Slabs

	C1			C2			C1			C2		
	B1	B2	B3	B1	B2	B3	B1	B2	B3	B1	B2	B3
A1	44.7	2.4	2.2	−6.2	−5.9	0.5	40.3	−1.1	2.8	1.4	−2.2	−6.2
A2	1.5	0.5	−1.6	32.6	6.9	−12.6	−12.4	−1.8	−8.1	36.5	5.0	−7.5
A3	0.2	−21.3	−6.5	−2.3	−6.2	−6.7	2.0	−22.5	−5.7	−9.1	−0.8	−3.5

component mainly refers to confusion and mental sickness. These dimensions hence globally refer to Comforters, Bad versus Good, and Instabilizors. The components for the semantic differential scales can be labeled respectively as Potency versus Evaluation, Relaxedness, and Inactivity, thus deviating considerably from the (all too) standard semantic differential distinctions of Evaluation, Potency, and Activity: Apparently, a more important dimension in these data is the Relaxedness (versus tension) dimension. The components for the third mode clearly describe Eve White and Jane (component 1) and Eve Black (component 2). Finally, in the core, the highest element (40.3) refers to the strong contribution of Comforters to Potency versus Evaluation when the patient ''is'' either Eve White or Jane. The second highest element (36.5) refers to the strong contribution of Bad concepts to Potency for Eve Black (in line with the perverseness of Eve Black). The third highest element (-22.5) indicates the negative contribution of Instabilizors to Relaxedness for Eve White and Jane. The other contributions can be similarly interpreted.

The present illustrative analysis served to demonstrate the flexibility of the joint orthomax method. We showed that by varying the relative weights for the component matrices, one can focus on different compromises of simplicity of the different parts of the solution and thus search for the ideal compromise. Kiers' (1997) core rotation procedure does not allow such compromises, because that method requires deciding whether or not to rotate a component matrix so that the *core* is optimally simple, which will lead to insufficient simplicity of either the component matrix or the core matrix, and there is no intermediate approach. The present analysis demonstrates that such intermediate solutions may well give the most interesting compromises.

5. Discussion

The present paper offers a procedure for rotating the core and the component matrices to simplicity, without forcing the user to choose for optimal simplicity of some matrices at the cost of a considerable loss of simplicity of the others. Using the method proposed here, we can search for solutions for which all parts of the solution (or a prespecified subset of matrices) are relatively simple. In the illustrative analysis presented in this paper, it is demonstrated that the new procedure can indeed find solutions that make all matrices of interest sufficiently simple, in the sense that they lead to interpretable solutions.

We have seen in the illustrative analysis that the flexibility of the present method allows one to change relative weights for **A**, **B**, and **C** to levels such that the simplicity of these matrices as well as of the core is satisfactory. In fact, in this way, one may aim to find solutions in which the varimax values of the core and of the component matrices are all higher than a certain percentage of the maximally attainable varimax value. Specifically, one might wish to maximize the smallest of the four thus-defined percentages. An alternative would be to search for solutions with a maximal average of thus-defined percentages. Unfortunately, we do not have a procedure for finding relative weights that actually optimize such criteria, although we do think that the trial-and-error procedure used in our illustrative analysis does a good job. Moreover, it should be noted that the method is meant for seeking a useful and easily interpretable description among infinitely many equivalent solutions, all of which are equally "correct." Therefore, the fact that subjective decisions (for instance, concerning the choice of weights) partly determine the outcome of the procedure is not harmful, as long as they do lead to clear and useful descriptions of the data.

A drawback of the flexibility of the joint varimax approach is that, in practice, it may not be easy to decide on an appropriate compromise of simplicity of the core and of the respective component matrices. Moreover, a comparison of solutions based on different sets of weights would be more complete if it would involve the full rotated component matrices, rather than just their varimax values (or percentages of the maximal varimax values). However, since this approach would require comparison of many solutions each consisting of three component matrices and a core, the strategy is virtually unfeasible in practice, and may, adversely, lead one to settle for only one analysis based on standard weights. We propose to handle the flexibility by comparing solutions according to simplicity indices for all constituents of the solution, but further research may lead to better procedures to deal with the flexibility, which is not just an aspect of the present criterion, but is inherent in the wish to simplify the interpretation of all parts of a 3MPCA solution *jointly*.

In the analyses in the present paper, we consistently set all gamma parameters equal to 1, thus combining only varimax criteria. The reason is that the varimax criterion is known to work well for rotation of component matrices, whereas, for instance, the quartimax criterion ($\gamma = 0$) is notorious for its tendency to yield a "general factor" (one column with relatively high elements throughout), even when the more simple situation of different components being related to different, mostly exclusive, subsets of variables can be attained. For rotation of the core, Kiers (1997) reported that the varimax criteria tended to work well too, in general, but that for certain situations it may be advantageous to take γ larger than 1 (as demonstrated by a case where taking $\gamma = 3$ yielded considerably better results). Therefore, we recommend starting one's analysis by using $\gamma = 1$ for all terms, and to consider alternative choices for γ (e.g., > 1, in particular for the core), only when this approach does not lead to acceptable solutions.

An obvious extension of the present procedure is the one to solutions of N-way PCA. Because the core rotation algorithm can easily be extended to the N-way case, by simply cycling over N rather than three updatings of rotation matrices, so can the *joint* orthomax procedure.

The 3MPCA solution is usually taken such that the component matrices are columnwise orthonormal. However, this restriction is made merely for identification purposes, and any nonsingular transformation of these matrices (when supplemented with the inverse transformations of the core) yields the same model fit. Hence, to find interpretable solutions, one need not restrict oneself to the orthogonal rotations. In fact, an oblique rotation procedure for rotating the core to simplicity has been proposed by Kiers (1998). However, combining this criterion with criteria for simplicity of the (counter-rotated) component matrices seems all but straightforward. An alternative procedure for obtaining oblique rotations of the component matrices could be based on rotation of rescaled versions of the component matrices (compare Harris and Kaiser 1964). For instance, first normalizing the columns of **A**, **B**, and **C** to sums of squares equal to the associated eigenvalues, and administering that solution to the present joint orthomax procedure would reduce to searching for oblique rotations that simplify the component matrices and the core jointly. Of course, other rescalings could be considered for this purpose as well.

An alternative to *rotating* the core and components to simplicity, is to constrain the core and components to be simple, in that certain elements are required to be zero (see, Kiers 1992; Kiers and Smilde 1998). This approach, of course, requires insight in which elements should be constrained to zero. The rotation procedure proposed here could be used to choose which elements to set to zero, and, upon comparison of fit values of different models, one can check if the chosen constraints are sensible.

References

CARROLL, J. B. (1953), "An Analytic Solution for Approximating Simple Structure in Factor Analysis," *Psychometrika, 18*, 23-38.

CLARKSON, D. B., and JENNRICH, R. I. (1988), "Quartic Rotation Criteria and Algorithms," *Psychometrika, 53*, 251-259.

CRAWFORD, C. B., and FERGUSON, G. A. (1970), "A General Rotation Criterion and Its Use in Orthogonal Rotation," *Psychometrika, 35*, 321-332.

FERGUSON, G. A. (1954), "The Concept of Parsimony in Factor Analysis," *Psychometrika, 19*, 281-290.

HAKSTIAN, A. R. (1976), "Two-matrix Orthogonal Rotation Procedures," *Psychometrika, 41*, 267-272.

HARRIS, C. W., and KAISER, H. F. (1964), "Oblique Factor Analytic Solutions by Orthogonal Transformations," *Psychometrika, 29*, 347-362.

JENNRICH, R. I. (1970), ''Orthogonal Rotation Algorithms,'' *Psychometrika, 35*, 229-235.

KAISER, H. F. (1958), ''The Varimax Criterion for Analytic Rotation in Factor Analysis,'' *Psychometrika, 23*, 187-200.

KIERS, H. A. L. (1992) ''TUCKALS Core Rotations and Constrained TUCKALS Modelling,'' *Statistica Applicata, 4*, 659-667.

KIERS, H. A. L. (1998), ''Three-way SIMPLIMAX for Oblique Rotation of the Three-mode Factor Analysis Core to Simple Structure,'' *Computational Statistics and Data Analysis, xx*, xxx-xxx.

KIERS, H. A. L. (1997), ''Three-mode Orthomax Rotation,'' *Psychometrika, 62*, 579-598.

KIERS, H. A. L., and SMILDE, A. K. (1998), ''Constrained Three-mode Factor Analysis as a Tool for Parameter Estimation with Second-order Instrumental Data,'' *Journal of Chemometrics, 12*, 125-147.

KROONENBERG, P. M. (1983), *Three-Mode Principal Component Analysis: Theory and Applications*, Leiden: DSWO Press.

KROONENBERG, P. M. (1985), ''Three-Mode Principal Components Analysis of Semantic Differential Data: The Case of a Triple Personality,'' *Applied Psychological Measurement, 9*, 83-94.

KROONENBERG, P. M., and DE LEEUW, J. (1980), ''Principal Component Analysis of Three-mode Data by Means of Alternating Least Squares Algorithms,'' *Psychometrika, 45*, 69-97.

MATHWORKS INC. (1992), *MATLAB*, Natick, MA: Mathworks Inc.

NEUHAUS, J. O., and WRIGLEY, C. (1954), ''The Quartimax Method: An Analytic Approach to Orthogonal Simple Structure,'' *British Journal of Mathematical and Statistical Psychology, 7*, 81-91.

OSGOOD, C. E., and LURIA, Z. (1954), ''A Blind Analysis of a Case of Multiple Personality,'' *Journal of Abnormal and Social Psychology, 49*, 579-591.

SAUNDERS, D. R. (1953), ''An Analytic Method for Rotation to Orthogonal Simple Structure,'' Research Bulletin, RB 53-10. Princeton: Educational Testing Service.

SNIDER, J. G., and OSGOOD, C. E. (1969), *Semantic Differential Technique: A Source Book*, Chicago: Aldine.

TUCKER, L. R (1966), ''Some Mathematical Notes on Three-mode Factor Analysis,'' *Psychometrika, 31*, 279-311.