Oblique rotation in correspondence analysis: a step forward in the search of the simplest interpretation

Urbano Lorenzo-Seva*[*]
Michel van de Velden**
Henk A. L. Kiers***

June, 2007

Econometric Institute Report EI 2007-25

Abstract

Correspondence analysis (CA) is a popular method that can be used to analyze relationships between categorical variables. It is closely related to several popular multivariate analysis methods such as canonical correlation analysis and principal component analysis. Like principal component analysis, CA solutions can be rotated orthogonally as well as obliquely to simple structure, without affecting the total amount of explained inertia. However, some specific aspects of CA prevent standard rotation procedures from being applied in a straightforward fashion. In particular, the role played by weights assigned to points and dimensions, and the duality of CA solutions, are unique to CA. For orthogonal simple structure rotation, recently procedures have been proposed. In this paper, we construct oblique rotation methods for CA that takes into account these specific difficulties. We illustrate the benefits of our oblique rotation procedure by means of two illustrative examples.

Keywords: Correspondence analysis, simple structure, oblique rotation.

* Rovira i Virgili University, Tarragona, Spain
** Erasmus University, Rotterdam, The Netherlands
*** University of Groningen, Groningen, The Netherlands

1. Introduction

In the context of exploratory factor analysis (EFA), it is usually accepted that the best possible interpretation is the simplest one. This implies that maximizing factor simplicity helps to optimize factor interpretation. Kaiser (1974) defined factor simplicity as the simplest possible solution in which each variable is generated by a single component, or factor. In order to achieve such factor simplicity, many rotation criteria have been proposed in the literature. It is important to note that rotation does not change a complex factor solution into a simple factor solution: it helps to determine the position of axes that will allow the simplest possible interpretation of data. However, if a simple interpretation does not exist, then rotating the axes will be of no help at all.

Orthogonal and oblique rotation involve the same problem of minimizing a complexity (or maximizing a simplicity) criterion. Only the imposed constraints differ. Since fewer constraints are imposed in oblique rotation, it is generally possible to obtain a lower value of the complexity function and thus make the factor solution simpler than in orthogonal rotation (Browne, 2001). From this point of view, Browne suggests that oblique rotation is probably more appropriate in most practical situations.

Correspondence analysis (CA) is a popular method that can be used to analyze relationships between categorical variables. The method is closely related to several popular multivariate analysis methods such as canonical correlation analysis and principal component analysis (e.g. Greenacre, 1984, Tenenhaus & Young, 1985). The relationship with principal component analysis and factor analysis has led some authors to refer to the method as principal component analysis of categorical data. Despite this close relationship with a method in which rotation is quite common, rotation in CA has not received much attention: Adachi (2004) considers rotation in multiple correspondence analysis (MCA), and Van de Velden and Kiers (2003, 2005), and Greenacre (2006a) explicitly consider rotation in CA.

One reason for the scarcity of work on rotation in CA may be the emphasis that is usually put on graphical representations in CA. Graphical representations are feasible if a low-dimensional solution (usually two) provides an adequate approximation of the data. Rotation, on the other hand, is particularly useful when solutions are of higher

dimensionality. For many applications of CA, the two-dimensional solution is deemed sufficient, so a rotation procedure is not needed. If a CA solution, however, requires more than two dimensions to account for sufficient variation, the numerical output must be studied in the same way as it is in factor analysis. In these cases, rotation is an important tool for facilitating interpretation of the data.

Another reason for the under-representation of rotation in CA is that some specific aspects of CA prevent straightforward application of standard rotation procedures. In particular, the role played by weights assigned to points and dimensions in CA, and the asymmetry of the solutions and the corresponding scaling options are unique to CA. The latter issue, which is caused by the duality of the CA solution (i.e. CA simultaneously yields a solution for rows and columns of a data table), is of particular importance when considering oblique rotations. In this paper, we build on the work by van de Velden and Kiers (2005) and Greenacre (2006a) to construct oblique rotation methods for CA that take into account the specific difficulties of CA.

Greenacre and Blasius (1994, page viii) point out the importance of the data and of the features in the data *revealing themselves* rather than limiting one's vision of the data to restrictively and subjectively defined statistical models. Benzécri, one of the inventors of correspondence analysis, expressed it as follows: "The model must follow the data, and not the other way around". Nowadays, this philosophical principle can still be observed in researchers who use correspondence analysis. For example, Greenacre (2006b, page 61) points out that in spite of the theoretical difficulties involved in justifying the full-space chi-square geometry in the multiple correspondence analysis context, it is regularly applied because it successfully recovers interesting patterns of associations between variables. Oblique rotation in CA can be seen as one step further that helps to interpret complex datasets, and to better understand the association between the different categories of the variables. We do not suggest that rotated solutions can replace the original CA solutions: oblique rotation is merely a source of additional information that can be considered. If it helps the data to reveal themselves, then it should be welcome.

We shall first review some aspects of oblique rotation in the context of EFA: oblique quartimin rotation, an algorithm to implement it, and the usual interpretation of

the oblique solution. Then we shall consider the rotation in the context of CA, and propose an algorithm to rotate symmetrical coordinates. Finally, the usefulness of oblique rotation in correspondence analysis is illustrated with two numerical examples.

2. Rotation in the EFA context

We shall now overview three aspects of oblique rotation in the context of EFA: (a) the oblique Quartimin criterion is defined; (b) the Gradient Projection algorithm is briefly described; and (c) the interpretation of the rotated solution is discussed.

2.1. Oblique Quartimin rotation

The Quartimin rotation criterion is a member of Carroll's oblimin family, and with $\mathbf{\Lambda}$ denoting the matrix of interest, it can be expressed as

$$Q(\mathbf{\Lambda}) = \mathrm{tr}\mathbf{\Lambda}^{2\prime}(\mathbf{\Lambda}^2\mathbf{N}), \tag{1}$$

where $\mathbf{\Lambda}^2 = \mathbf{\Lambda}\cdot\mathbf{\Lambda}$ is the element-wise product of $\mathbf{\Lambda}$ with itself, $\mathbf{N}$ is a square matrix with zeros on the diagonal and ones elsewhere, and tr$\mathbf{H}$ refers to the trace of matrix $\mathbf{H}$. In the context of EFA in which it was proposed (Jennrich & Sampson, 1966; see also Clarkson & Jennrich, 1988), $\mathbf{\Lambda}$ is an obliquely rotated pattern matrix from an initial orthogonal pattern matrix $\mathbf{A}$,

$$\mathbf{\Lambda} = \mathbf{A}\left(\mathbf{U}'\right)^{-1}, \tag{2}$$

where $\mathbf{U}$ is an arbitrary nonsingular matrix with columns of unit length. The aim is to minimize the function

$$f\left(\mathbf{U}\right) = Q\left(\mathbf{A}\left(\mathbf{U}'\right)^{-1}\right), \tag{3}$$

over all nonsingular matrices $\mathbf{U}$ whose columns have unit length. In the context of factor analysis, Kaiser (1958) pointed out that a systematic bias is observed in rotated loading matrices which can be attributed to the divergent weights that are implicitly attached to the variables by their communalities. To resolve this, he proposed to minimize

$$f\left(\mathbf{T}\right) = Q\left(\mathrm{B}\left(\mathbf{U}'\right)^{-1}\right), \tag{4}$$

where $\mathbf{B}$ is the corresponding row-wise normalized matrix obtained from $\mathbf{A}$ (i.e. $\mathbf{B} = \mathrm{diag}\left(\mathbf{A}\mathbf{A}'\right)^{-1/2}\mathbf{A}$, and diag($\mathbf{H}$) refers to a diagonal matrix that contains the diagonal

elements of matrix **H**). Although Kaiser proposed the normalization for orthogonal rotations of **A**, this normalization is now also advised in oblique rotations (Browne, 2001). Once the optimal rotation matrix **U** has been obtained, the rotated pattern matrix is obtained by expression (2), and the inter-factor correlation matrix is obtained as

$$\mathbf{\Phi} = \mathbf{U}'\mathbf{U}\,. \tag{5}$$

Note that **U** is constrained so that diag(**U**'**U**) = **I** (i.e. equal column sums of squares in **U**). In EFA, this constraint is imposed because it ensures that, if component scores are standardized and uncorrelated before rotation, after rotation they are still standardized.

2.2. An algorithm for oblique rotation

Jennrich (2002) proposed a simple general algorithm for oblique rotation, which can be easily adapted to minimize several rotation criteria. The gradient projection (GP) algorithm has two steps: (a) computation of the gradient of the rotation criterion, and (b) projection onto a manifold of matrices with unit length columns. As the projection step is very simple, implementation of the algorithm mainly involves computing the corresponding gradient. Jennrich (2002) shows that the gradient of any oblique rotation at **U** is given by

$$\mathbf{G} = -\left(\mathbf{\Lambda}'\mathbf{G}_{\mathrm{q}}\mathbf{U}^{-1}\right)', \tag{6}$$

where $\mathbf{G}_q$ is the particular gradient of the rotation criterion that is implemented. For example, when the quartimin rotation criterion in expression (1) is implemented, the gradient $\mathbf{G}_q$ of Q at $\mathbf{\Lambda}$ is

$$\mathbf{G}_q = \mathbf{\Lambda} \cdot \left(\mathbf{\Lambda}^2\mathbf{N}\right). \tag{7}$$

2.3. Interpretation of the rotated solution

In the context of EFA, matrix $\mathbf{\Lambda}$ is known as a pattern matrix, and the following product is known as the reproduced correlation matrix,

$$\mathbf{\Lambda\Phi\Lambda}' = \mathbf{R}^*. \tag{8}$$

In addition, the product $\widetilde{\mathbf{\Lambda}} = \mathbf{\Lambda\Phi} = \mathbf{AU}$ is known as the structure matrix, and can also be used to compute the reproduced correlation matrix,

$$\mathbf{\Lambda}\widetilde{\mathbf{\Lambda}}' = \mathbf{R}^*. \tag{9}$$

As has already been pointed out, the aim of the rotation is to determine the position of the axes that allows the simplest possible interpretation of data. Intuitively, it seems that a graphical representation of both data and rotated axes should help to interpret the data. When axis rotation was initially proposed it was thought that the graphical representation would be essential for the correct interpretation of the rotated solution (see for example Thurstone, 1947). Nowadays, the graphical representation is only used for pedagogical purposes. The reason for this is that the large number of variables and rotated components (or factors) in the datasets mean that the graph is not interpretable at all. So nowadays, the solution is usually interpreted by the numerical outcome, not the graphical outcome, and graphical representations of a rotated factor analysis are rarely found in scientific journals. The question is: which matrices should be interpreted? A popular procedure is to study the pattern matrix $\mathbf{\Lambda}$, and then the inter-factor correlation matrix $\mathbf{\Phi}$. Alternatively, the structure matrix $\widetilde{\mathbf{\Lambda}}$ could be studied. Although $\mathbf{\Lambda}$ and $\widetilde{\mathbf{\Lambda}}$ may be very different if dimensions are strongly correlated, they are closely related ($\widetilde{\mathbf{\Lambda}} = \mathbf{\Lambda\Phi}$) so that conclusions should be the same, no matter which matrix ($\mathbf{\Lambda}$ or $\widetilde{\mathbf{\Lambda}}$) is studied. Note that studying the structure is like studying the pattern and the inter-factor correlation matrix at once. However, studying first the pattern, and then the inter-factor correlation matrix is a much more parsimonious and simpler procedure: it is like solving a complex problem using progressive steps. A final consideration is that, if $\mathbf{\Lambda}$ has a simple structure, $\widetilde{\mathbf{\Lambda}}$ must also be simple, but in a different metric.

Another aspect to consider is which values in the pattern matrix $\mathbf{\Lambda}$ must be interpreted. If a variable has a loading value close to zero it means that the variable is not related to the component (or factor). However, a salient loading (i.e. a high loading) indicates that the variable is related to the component (or factor). Arbitrary values of 0.3 or 0.4 are frequently used in practice as a threshold between salient and not salient loadings. Some rotation methods use the mean of the squared loadings in the component (or factor) as a threshold: only loadings whose squared value is larger than this mean are considered as salient loadings (see, for example, Lorenzo-Seva, 1999). Once the salient loadings have been determined, the interpretation of each component (or factor) is based on the characteristics of variables with a salient loading on the corresponding component (or factor).

3. Rotation in the CA context

As a large number of matrices is involved in CA, we shall first define them to avoid confusions in the text that follows. Then we shall introduce four aspects: (a) oblique rotation in the context of CA; (b) an algorithm for simultaneous rotation of symmetrical coordinates to maximize the quartimin criterion; (c) the interpretation of the rotated solution; and, finally, (d) some further considerations.

3.1. Correspondence analysis

Let $\mathbf{F}$ be an $n \times p$ contingency table divided by the total number of observations, $\mathbf{1}_i$ an $i \times 1$ vector of ones, $\mathbf{r} = \mathbf{F}\mathbf{1}_p$, $\mathbf{c} = \mathbf{F}'\mathbf{1}_n$, $\mathbf{D}_r = \text{diag}(\mathbf{r})$, and $\mathbf{D}_c = \text{diag}(\mathbf{c})$. By assigning weights to the rows and the columns of a matrix of deviations from independence, we obtain the following matrix of standardized residuals,

$$\widetilde{\mathbf{F}} = \mathbf{D}_r^{-1/2}\left(\mathbf{F} - \mathbf{r}\mathbf{c}'\right)\mathbf{D}_c^{-1/2}. \tag{10}$$

The aim in correspondence analysis is to find k dimensional coordinate matrices $\mathbf{X}$ and $\mathbf{Y}$, for row and column points respectively, in such a way that the loss function

$$\phi(\mathbf{X},\mathbf{Y}) = \left\|\widetilde{\mathbf{F}} - \mathbf{D}_r^{1/2}\mathbf{X}\mathbf{Y}'\mathbf{D}_c^{1/2}\right\| \tag{11}$$

is minimized, where $||\mathbf{H}||^2$ denotes the sum of squared elements of $\mathbf{H}$. Let

$$\widetilde{\mathbf{F}} = \mathbf{U}\boldsymbol{\Gamma}\mathbf{V}', \tag{12}$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix with singular values on the diagonal, in weakly descending order, and $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$. As Van de Velden and Kiers (2005) pointed out, $\phi(\mathbf{X}, \mathbf{Y})$ is minimized by

$$\mathbf{X} = \mathbf{D}_r^{-1/2}\mathbf{U}_k\boldsymbol{\Gamma}_k^{\alpha} \tag{13}$$

and

$$\mathbf{Y} = \mathbf{D}_c^{-1/2}\mathbf{V}_k\boldsymbol{\Gamma}_k^{1-\alpha}, \tag{14}$$

where $\mathbf{U}_k$ and $\mathbf{V}_k$ are respectively the $n \times k$ and $p \times k$ matrices of singular vectors corresponding to the k largest singular values gathered in the $k \times k$ diagonal matrix $\boldsymbol{\Gamma}_k$. Three choices are usually considered for α:

(a) $\alpha = 0$: The column coordinates $\mathbf{Y}$ are referred to as *principal coordinates* and the row coordinates $\mathbf{X}$ as *standard coordinates*.

(b) $\alpha = 1$: The column coordinates **Y** are referred to as *standard coordinates* and the row coordinates **X** as *principal coordinates*.

(c) $\alpha = .5$: Both column and row coordinates are referred to as *symmetrical coordinates*.

An important feature of **X** and **Y** is that they satisfy the biplot requirement proposed by Gabriel (1971): the matrix product **XY'** must optimally approximate $\widetilde{\mathbf{F}}$. The biplot requirement is satisfied because, for any choice of α, the product $\mathbf{D}_r^{\frac{1}{2}}\mathbf{XY'}\mathbf{D}_c^{\frac{1}{2}}$ is always the same.

The distinction between principal, standard and symmetrical coordinates is in fact fundamental in CA:

(a) *Principal coordinates* are the coordinates that are described in the analysis. If $\alpha = 0$, these coordinates are related to column variables, and if $\alpha = 1$, to row variables.

(b) *Standard coordinates* are considered to be the describing variable. If $\alpha = 0$, these coordinates are related to row variables; and, if $\alpha = 1$, to column variables. An important feature of standard coordinates is that

$$\mathbf{X'D}_r\mathbf{X} = \mathbf{I} \tag{15}$$

for $\alpha = 0$, and

$$\mathbf{Y'D}_c\mathbf{Y} = \mathbf{I} \tag{16}$$

for $\alpha = 1$. This means that (row or column) standard coordinates are uncorrelated and standardized.

(c) When *symmetrical coordinates* are chosen, both column and row coordinates are to be described. In this situation,

$$\mathbf{X'D}_r\mathbf{X} = \mathbf{Y'D}_c\mathbf{Y} = \mathbf{\Gamma}_k . \tag{17}$$

This means that row and column symmetrical coordinates are uncorrelated, and show the same spread.

3.2. Oblique rotation in the context of CA

Row and column coordinates, **X** and **Y**, are usually inspected in order to explain the meaning of the k dimensions. As in the context of EFA, it is accepted that the best

possible interpretation is the simplest one. Note that expressions (13) and (14) do not ensure any simplicity in the coordinates. However, rotation can be used to maximize the simplicity in rotated row coordinates $\widetilde{\mathbf{X}}$, and rotated column coordinates $\widetilde{\mathbf{Y}}$. If simplicity is really maximized in $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$, then the interpretation of the dimensions should be optimized.

Van de Velden (2000) and Van de Velden and Kiers (2005) proposed an orthogonal rotation procedure for matrices **X** and **Y**. As has been pointed out above, **X** and **Y** satisfy the biplot requirement. In the orthogonal rotation proposed by Van de Velden and Kiers (2005), the rotated matrices also satisfy the requirement. If **M** is an orthogonal rotation matrix (**M'M**=**MM'**=**I**), then $\mathbf{D}_r^{1/2}\mathbf{XMM'Y'}\mathbf{D}_c^{1/2} = \mathbf{D}_r^{1/2}\mathbf{XY'}\mathbf{D}_c^{1/2}$. Note that both matrices **X** and **Y** are post-multiplied by **M**, so that the same rotation matrix is used to rotate both matrices. In the oblique rotation, matrices **X** and **Y** cannot be rotated by the same rotation matrix, if the biplot requirement is to be satisfied.

As already explained, the value of α changes the meaning of coordinates **X** and **Y**, making them principal, standard or symmetrical coordinates. In addition, simplicity can be maximized in $\widetilde{\mathbf{X}}$, or in $\widetilde{\mathbf{Y}}$, or in both simultaneously. We propose two different procedures: one for $\alpha = 0$ or $\alpha = 1$, and another for $\alpha = .5$.

In the first procedure, simplicity should be maximized in the column (or row) principal coordinates. More concretely, when $\alpha = 0$, we need an oblique rotation matrix **U** that maximizes simplicity in $\widetilde{\mathbf{Y}}$, constrained to diag(**U**'**U**)= **I**. Once **U** is available, it can be used to rotate **X** and **Y** in the following way: $\widetilde{\mathbf{X}} = \mathbf{XU}$ and $\widetilde{\mathbf{Y}} = \mathbf{Y}(\mathbf{U}')^{-1}$, such that the biplot requirement is satisfied: $\mathbf{D}_r^{1/2}\widetilde{\mathbf{X}}\widetilde{\mathbf{Y}}'\mathbf{D}_c^{1/2} = \mathbf{D}_r^{1/2}\mathbf{XUU}^{-1}\mathbf{Y'D}_c^{1/2} = \mathbf{D}_r^{1/2}\mathbf{XY'D}_c^{1/2}$. In addition, the constraint diag(**U**'**U**) = **I** ensures that the rotated standard coordinates $\widetilde{\mathbf{X}}$ are standardized. The rotation matrix **U** can be obtained to maximize simplicity in $\widetilde{\mathbf{Y}}$ in terms of the quartimin criterion, and the algorithm explained in Section 2.2 in the EFA context can be used to obtain it. When $\alpha = 1$, we need an oblique rotation matrix **U** that maximizes simplicity in $\widetilde{\mathbf{X}}$, also constrained to diag(**U**'**U** = **I**). Once **U** is available, it can be used to rotate **X** and **Y** in the following way: $\widetilde{\mathbf{X}} = \mathbf{X}(\mathbf{U}')^{-1}$ and $\widetilde{\mathbf{Y}} = \mathbf{XU}$. Again, the biplot requirement is satisfied: $\mathbf{D}_r^{1/2}\widetilde{\mathbf{X}}\widetilde{\mathbf{Y}}'\mathbf{D}_c^{1/2} = \mathbf{D}_r^{1/2}\mathbf{X}(\mathbf{U}')^{-1}\mathbf{U'Y'D}_c^{1/2} = \mathbf{D}_r^{1/2}\mathbf{XY'D}_c^{1/2}$, and

the rotated standard coordinates $\tilde{\mathbf{Y}}$ are standardized. As before, the rotation matrix **U** can be obtained to maximize simplicity in $\tilde{\mathbf{X}}$ in terms of the quartimin criterion, and the algorithm explained in Section 2.2 in the EFA context can be used to obtain it. The inter-dimension correlation matrix is obtained as $\mathbf{\Phi} = \mathbf{U}'\mathbf{U}$.

In the second procedure (i.e. α = .5), the simplicity should be simultaneously maximized in the column and row symmetrical coordinates. In principle, this can be done by finding the oblique rotation matrix **T** that jointly simplifies

$$\tilde{\mathbf{X}}_1 = \mathbf{X}(\mathbf{T}')^{-1} \text{ and } \tilde{\mathbf{Y}}_1 = \mathbf{Y}\mathbf{T}, \tag{19}$$

or

$$\tilde{\mathbf{X}}_2 = \mathbf{X}\mathbf{T} \text{ and } \tilde{\mathbf{Y}}_2 = \mathbf{Y}(\mathbf{T}')^{-1}. \tag{20}$$

In addition, in order to preserve the common spread of symmetrical coordinates defined in expression (17), the oblique rotation matrix **T** will be constrained to $\text{diag}(\mathbf{T}'\mathbf{T}) = \mathbf{\Gamma}_k$. As both approaches approximate $\tilde{\mathbf{F}}$, to choose one of them would be arbitrary. In addition, if **T** were to be constrained to **T**'**T**=**TT**'=**I**, the rotated coordinates $\tilde{\mathbf{X}}_1$ would be equal to $\tilde{\mathbf{X}}_2$. The same applies to the pair $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Y}}_2$. Note that $\tilde{\mathbf{X}}_1$ is known in the EFA context as a pattern matrix, and $\tilde{\mathbf{X}}_2$ is its corresponding structure matrix. Because they are a pattern and the corresponding structure matrix, the inter-dimensions variance-covariance matrix $\mathbf{S} = \mathbf{T}'\mathbf{T}$ relates them,

$$\tilde{\mathbf{X}}_2 = \mathbf{X}\mathbf{T} = \mathbf{X}(\mathbf{T}')^{-1}\mathbf{T}'\mathbf{T} = \tilde{\mathbf{X}}_1\mathbf{S}. \tag{21}$$

This relationship also holds between $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Y}}_2$,

$$\tilde{\mathbf{Y}}_1 = \mathbf{Y}\mathbf{T} = \mathbf{Y}(\mathbf{T}')^{-1}\mathbf{T}'\mathbf{T} = \tilde{\mathbf{Y}}_2\mathbf{S}. \tag{22}$$

In addition, if $\tilde{\mathbf{X}}_1$ is a pattern matrix and $\tilde{\mathbf{Y}}_1$ is a structure matrix,

$$\tilde{\mathbf{X}}_1\tilde{\mathbf{Y}}_1' = \tilde{\mathbf{X}}_1\mathbf{S}\tilde{\mathbf{Y}}_2' = \tilde{\mathbf{X}}_1\mathbf{S}^{1/2}\mathbf{S}^{1/2}\tilde{\mathbf{Y}}_2'. \tag{23}$$

Also

$$\tilde{\mathbf{X}}_2\tilde{\mathbf{Y}}_2' = \tilde{\mathbf{X}}_1\mathbf{S}\tilde{\mathbf{Y}}_2' = \tilde{\mathbf{X}}_1\mathbf{S}^{1/2}\mathbf{S}^{1/2}\tilde{\mathbf{Y}}_2'. \tag{24}$$

An overall approach can be found by defining $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ as

$$\tilde{\mathbf{X}} = \mathbf{X}(\mathbf{T}')^{-1}(\mathbf{T}'\mathbf{T})^{1/2} = \tilde{\mathbf{X}}_1\mathbf{S}^{1/2}, \tag{25}$$

and

$$\tilde{\mathbf{Y}} = \mathbf{Y}(\mathbf{T}')^{-1}(\mathbf{T}'\mathbf{T})^{1/2} = \tilde{\mathbf{Y}}_2\mathbf{S}^{1/2}. \quad (26)$$

Note that now $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ are neither pattern matrices, nor structure matrices. Instead, they can be defined as pseudo-structure matrices. Note that the biplot requirement is again satisfied,

$$\mathbf{D}_r^{1/2}\tilde{\mathbf{X}}\tilde{\mathbf{Y}}'\mathbf{D}_c^{1/2} = \mathbf{D}_r^{1/2}\mathbf{X}(\mathbf{T}')^{-1}(\mathbf{T}'\mathbf{T})(\mathbf{T})^{-1}\mathbf{Y}'\mathbf{D}_c^{1/2} = \mathbf{D}_r^{1/2}\mathbf{X}\mathbf{Y}'\mathbf{D}_c^{1/2}. \quad (27)$$

Finally,

$$\mathrm{tr}\left(\tilde{\mathbf{X}}'\mathbf{D}_r\tilde{\mathbf{X}}\right) = \mathrm{tr}\left(\tilde{\mathbf{Y}}'\mathbf{D}_c\tilde{\mathbf{Y}}\right) = \mathrm{tr}\left(\mathbf{X}'\mathbf{D}_r\mathbf{X}\right) = \mathrm{tr}\left(\mathbf{Y}'\mathbf{D}_c\mathbf{Y}\right) = \mathrm{tr}\left(\boldsymbol{\Gamma}_k\right). \quad (28)$$

Note that, in order to preserve the common spread of symmetrical coordinates defined in expression (17), the oblique rotation matrix $\mathbf{T}$ was constrained to $\mathrm{diag}(\mathbf{T}'\mathbf{T}) = \boldsymbol{\Gamma}_k$. This yields common spread for $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{X}}_2$ (which could be defined as structure matrices), but it does not for $\tilde{\mathbf{Y}}_2$ or $\tilde{\mathbf{X}}_1$. This situation is similar in EFA context, where the spread is preserved only for the structure matrix that is used to obtain the reproduced correlation matrix (not in the pattern matrix). In CA, where $\tilde{\mathbf{X}}\tilde{\mathbf{Y}}' = \tilde{\mathbf{X}}_1\tilde{\mathbf{Y}}_1' = \tilde{\mathbf{X}}_2\tilde{\mathbf{Y}}_2'$, the spread is also preserved in a structure matrix.

In EFA context, simplicity is usually maximized in a pattern matrix. Here, the matrices that are equivalent to a pattern matrix are $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{Y}}_2$. Therefore, a rotation matrix $\mathbf{T}$ should be chosen to simultaneously maximize any simplicity criterion in $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{Y}}_2$. No algorithm is available for this situation, so in the next Section we propose an algorithm to obtain a rotation matrix $\mathbf{T}$ that maximizes the quartimin criterion simultaneously in both matrices.

3.3. An algorithm for simultaneous rotation of symmetrical coordinates to maximize quartimin criterion

Let $\mathbf{X}$ and $\mathbf{Y}$ be symmetrical coordinate matrices of order $n \times k$ and $p \times k$, respectively. Finally, let $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{Y}}_2$ be defined by expressions (19) and (20), respectively. The aim is to find a rotation matrix $\mathbf{T}$ of order $k \times k$ that rotates $\mathbf{X}$ and $\mathbf{Y}$ so that simplicity is maximized in both $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{Y}}_2$, constrained to $\mathrm{diag}(\mathbf{T}'\mathbf{T}) = \boldsymbol{\Gamma}_k$. The simultaneous quartimin criterion can be defined as

$$Q\left(\widetilde{\mathbf{X}}_1, \widetilde{\mathbf{Y}}_2\right) = \frac{1}{2}\,\mathrm{tr}\widetilde{\mathbf{X}}_1^{2\prime}\left(\widetilde{\mathbf{X}}_1^{2}\mathbf{N}\right) + \frac{1}{2}\,\mathrm{tr}\widetilde{\mathbf{Y}}_2^{2\prime}\left(\widetilde{\mathbf{Y}}_2^{2}\mathbf{N}\right). \tag{29}$$

The two terms in (29) are the quartimin simplicity criteria for $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Y}}_2$, respectively, as defined in expression (1). The following function must be minimized

$$f(\mathbf{T}) = Q\left(\mathbf{X}(\mathbf{T}')^{-1}, \mathbf{Y}(\mathbf{T}')^{-1}\right) \tag{30}$$

over all nonsingular matrices $\mathbf{T}$ satisfying $\mathrm{diag}(\mathbf{T}'\mathbf{T}) = \mathbf{\Gamma}_k$. To minimize expression (30), we use the GP algorithm presented in Section 2.2. For this particular rotation, the partial gradients $\mathbf{G}_{qx}$ of Q at $\widetilde{\mathbf{X}}_1$ and $\mathbf{G}_{qy}$ of Q at $\widetilde{\mathbf{Y}}_2$ can be obtained as

$$\mathbf{G}_{qx} = \widetilde{\mathbf{X}}_1 \cdot \left(\widetilde{\mathbf{X}}_1^{2}\mathbf{N}\right), \text{ and } \mathbf{G}_{qy} = \widetilde{\mathbf{Y}}_2 \cdot \left(\widetilde{\mathbf{Y}}_2^{2}\mathbf{N}\right), \tag{31}$$

respectively. The gradient $\mathbf{G}$ of Q at $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Y}}_2$ then becomes

$$\mathbf{G} = -\frac{1}{2}\left(\widetilde{\mathbf{X}}_1'\mathbf{G}_{qx}\mathbf{T}^{-1}\right)' - \frac{1}{2}\left(\widetilde{\mathbf{Y}}_2'\mathbf{G}_{qy}\mathbf{T}^{-1}\right)'. \tag{32}$$

Once $\mathbf{T}$ has been obtained, both $\mathbf{X}$ and $\mathbf{Y}$ must be post multiplied as proposed in expressions (19) and (20) in order to obtain $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Y}}_2$, respectively. The inter-dimension covariance matrix is obtained as $\mathbf{S} = \mathbf{T}'\mathbf{T}$, whereas the inter-dimension correlation matrix $\mathbf{\Phi}$ is obtained by standardizing the variance-covariance matrix,

$$\mathbf{\Phi} = \mathrm{diag}(\mathbf{S})^{-1/2}\mathbf{S}\,\mathrm{diag}(\mathbf{S})^{-1/2} = \mathrm{diag}(\mathbf{T}'\mathbf{T})^{-1/2}\mathbf{T}'\mathbf{T}\,\mathrm{diag}(\mathbf{T}'\mathbf{T})^{-1/2}. \tag{33}$$

Finally, the pseudo-structure matrices $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ can be obtained by expressions (25) and (26), respectively.

3.4. Interpretation of the rotated solution

When the number of columns and rows in $\mathbf{F}$ is high, and more than two dimensions are needed to explain a substantial amount of variance, then the typical graphical plot will be difficult to interpret. In such complex situations, researchers tend to study the numerical outcome, as in the context of EFA. The question is: which matrices should be studied to best understand the data? Before answering this question, we need to consider further the product of the rotated coordinates.

We have proposed two different rotation procedures to be applied in two different situations: one to rotate principal and standard coordinates, and the other to rotate symmetrical coordinates. In the first situation, $\mathbf{X}$ or $\mathbf{Y}$ contains principal coordinates. If $\mathbf{X}$

is the matrix that contains the principal coordinates (i.e. α = 0), then the product of rotated coordinates is,

$$\tilde{\mathbf{X}}\tilde{\mathbf{Y}}' = \mathbf{X}\mathbf{U}\mathbf{U}^{-1}\mathbf{Y}' = \mathbf{X}(\mathbf{U}')^{-1}\mathbf{U}'\mathbf{U}\mathbf{U}^{-1}\mathbf{Y}' = \mathbf{X}(\mathbf{U}')^{-1}\mathbf{\Phi}\mathbf{U}^{-1}\mathbf{Y}' = \tilde{\mathbf{X}}_p\mathbf{\Phi}\tilde{\mathbf{Y}}_p' = \tilde{\mathbf{X}}_s\tilde{\mathbf{Y}}_p', \quad (34)$$

where $\tilde{\mathbf{X}}_p$, $\tilde{\mathbf{Y}}_p$, and $\tilde{\mathbf{X}}_s$ are defined as $\mathbf{X}(\mathbf{U}')^{-1}$, $\mathbf{Y}(\mathbf{U}')^{-1}$, and $\mathbf{X}(\mathbf{U}')^{-1}\mathbf{\Phi}$, respectively. If $\mathbf{Y}$ is the matrix that contains the principal coordinates (i.e. α = 1), the product of rotated coordinates is,

$$\tilde{\mathbf{X}}\tilde{\mathbf{Y}}' = \mathbf{X}(\mathbf{U}')^{-1}\mathbf{U}'\mathbf{Y}' = \mathbf{X}(\mathbf{U}')^{-1}\mathbf{U}'\mathbf{U}\mathbf{U}^{-1}\mathbf{Y}' = \mathbf{X}(\mathbf{U}')^{-1}\mathbf{\Phi}\mathbf{U}^{-1}\mathbf{Y}' = \tilde{\mathbf{X}}_p\mathbf{\Phi}\tilde{\mathbf{Y}}_p' = \tilde{\mathbf{X}}_p\tilde{\mathbf{Y}}_s' \quad (35)$$

where $\tilde{\mathbf{Y}}_s$ is defined as $\mathbf{Y}(\mathbf{U}')^{-1}\mathbf{\Phi}$. Expressions (34) and (35) show that the procedures proposed when α = 0 or α = 1 are equivalent. The difference between them is that simplicity is maximized in $\tilde{\mathbf{X}}_p$ or $\tilde{\mathbf{Y}}_p$ (depending on which matrix contains the principal coordinates). In the context of EFA, $\tilde{\mathbf{X}}_p$ and $\tilde{\mathbf{Y}}_p$ are known as pattern matrices, whereas $\tilde{\mathbf{X}}_s$ and $\tilde{\mathbf{Y}}_s$ are known as structure matrices. As in the context of EFA, if a pattern is simple, its corresponding structure matrix must also be simple to some extent: the elements in the pattern matrix with the largest values must also be the elements with the largest values in the structure matrix. Thus, the simplicity is, to some extent, preserved.

The product of rotated coordinates suggests two sets of matrices that could be interpreted in this situation: (a) $\tilde{\mathbf{Y}}_p$ and $\tilde{\mathbf{X}}_s$ if α = 0, or $\tilde{\mathbf{X}}_p$ and $\tilde{\mathbf{Y}}_s$ if α = 1; or, alternatively, (b) $\tilde{\mathbf{X}}_p$, $\tilde{\mathbf{Y}}_p$, and $\mathbf{\Phi}$ for α = 0 or α = 1. Option (a) means that one pattern matrix and one structure matrix would be studied. As we have already observed in the context of EFA, studying a structure matrix is like studying a pattern and the inter-dimensions correlation matrix at once. So option (b) could be more convenient: again, it is like solving a problem using progressive steps. Matrices $\tilde{\mathbf{X}}_p$ and $\tilde{\mathbf{Y}}_p$ help to name dimensions, and to decide which (row and column) variables are best related to each dimension. Matrix $\mathbf{\Phi}$ shows how strongly the dimensions are related with one another.

In the second situation, $\mathbf{X}$ and $\mathbf{Y}$ contain symmetrical coordinates (i.e. α = .5). In this situation, the product of rotated coordinates is,

$$\tilde{\tilde{\mathbf{X}}}\tilde{\tilde{\mathbf{Y}}}' = \mathbf{X}(\mathbf{T}')^{-1}(\mathbf{T}'\mathbf{T})^{1/2}(\mathbf{T}'\mathbf{T})^{1/2}\mathbf{T}^{-1}\mathbf{Y}' = \mathbf{X}(\mathbf{T}')^{-1}\mathbf{S}\mathbf{T}^{-1}\mathbf{Y}' = \tilde{\tilde{\mathbf{X}}}_1\mathbf{S}\tilde{\tilde{\mathbf{Y}}}_2' = \tilde{\tilde{\mathbf{X}}}_{ps}\tilde{\tilde{\mathbf{Y}}}_{ps}', \quad (36)$$

where $\widetilde{\mathbf{X}}_1$, $\widetilde{\mathbf{Y}}_2$, $\widetilde{\mathbf{X}}_{ps}$, and $\widetilde{\mathbf{Y}}_{ps}$ are defined as $\mathbf{X}(\mathbf{T}')^{-1}$, $\mathbf{Y}(\mathbf{T}')^{-1}$, $\mathbf{X}(\mathbf{T}')^{-1}\mathbf{S}^{1/2}$, and $\mathbf{Y}(\mathbf{T}')^{-1}\mathbf{S}^{1/2}$, respectively. Note that in this situation, simplicity is maximized in the pattern matrices $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Y}}_2$ simultaneously. $\widetilde{\mathbf{X}}_{ps}$ and $\widetilde{\mathbf{Y}}_{ps}$ are not pattern matrices nor structure matrices. We shall refer to them as pseudo-structure matrices.

Again, the product of rotated coordinates suggests two sets of matrices that could be interpreted in this situation: (a) $\widetilde{\mathbf{X}}_{ps}$ and $\widetilde{\mathbf{Y}}_{ps}$; or, alternatively, (b) $\widetilde{\mathbf{X}}_1$, $\widetilde{\mathbf{Y}}_2$, and $\mathbf{S}$. Option (a) would be as complex as studying two structure matrices (note that we call them pseudo-structure matrices). Option (b) may again be more convenient because it involves interpreting matrices that are simpler. The matrices $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Y}}_2$ would help to name the dimensions, and to decide which (row and column) variables are best related to each dimension. The matrix $\mathbf{S}$ will inform about how strongly the dimensions are related to one another. However, it is usually easier to interpret standardized covariance matrices, so matrix $\mathbf{\Phi}$ defined in expression (33) may be preferred to $\mathbf{S}$.

Another aspect to consider is which coordinate values in the rotated patterns must be interpreted in order to name the dimensions. If a variable shows a coordinate value close to zero on a dimension, it means that the variable does not behave substantively different from the average variable in this dimension. However, a salient coordinate (i.e. a coordinate with a high absolute value) on a dimension indicates that the corresponding variable shows a particular behavior on that dimension. If a variable has a salient coordinate on a dimension, then the variable is related to the dimension. We need a threshold between salient and non-salient coordinates. As in EFA, we propose to compute the mean of the squared coordinates for each dimension in each rotated pattern. Then, each squared coordinate is compared to its corresponding mean: only coordinates whose squared value is larger than the mean are considered as salient coordinates. Labels can then be assigned to each dimension according to the characteristics of the variables with salient coordinates in the dimension. Once labels have been assigned to dimensions, the interpretation of all the (column and row) variables depends on the values of the (salient and non-salient) coordinates in the dimensions. Note that if a variable systematically has close-to-zero coordinates in the dimensions, it may behave no differently from the average variable in any dimension (which in fact can be interesting information).

3.5. Further considerations

In the context of EFA, loading matrices are frequently weighted before rotations are computed. After rotation, the original distances of points from the origin are reestablished, so the interpretation is not affected by the weights applied. The most popular weighting scheme is the row-wise normalization of the loading matrix already explained in Section 2.1 (Kaiser, 1958).

In the context of CA, other weighting schemes may also be applicable. Let $\mathbf{W}_x$ and $\mathbf{W}_y$ be diagonal matrices, with weights on the diagonal and zeros elsewhere. $\mathbf{W}_x$ and $\mathbf{W}_y$ are weighting matrices related to the coordinate matrices **X** and **Y**, respectively. The aim is to weight the rows of **X** and **Y** during the rotation, so that the products $\mathbf{W}_x\mathbf{X}$ and $\mathbf{W}_y\mathbf{Y}$ are rotated (instead of **X** and **Y**). Three options for $\mathbf{W}_x$ and $\mathbf{W}_y$ can be considered:

(a) Each weighting matrix is defined as an identity matrix. With this option, no weight is actually applied. This is the scheme used by Van de Velden and Kiers (2005).

(b) Due to the specific weighting in correspondence analysis, infrequently observed points are sometimes positioned relatively far away from the origin. In this situation, these particular points may play an important role in determining the rotation angle. Greenacre (2006a) suggests that the coordinates are rescaled using $\mathbf{W}_x = \text{diag}(\mathbf{r})$ and $\mathbf{W}_y = \text{diag}(\mathbf{c})$. This weighting procedure would place infrequent points close to the origin, while others would remain far away.

(c) As explained in Section 2.1, in the context of EFA it is customary to avoid different distances of points from the origin, because this tends to produce systematic bias in the rotated matrices. Kaiser (1958) proposed a row-wise normalization of the matrix to be rotated, which has become usual practice nowadays. In CA, this procedure would imply that coordinates are rescaled using $\mathbf{W}_x = \text{diag}\left(\mathbf{X}'\mathbf{X}\right)^{-1/2}$ and $\mathbf{W}_y = \text{diag}\left(\mathbf{Y}'\mathbf{Y}\right)^{-1/2}$. With this scheme, all the rows have the same influence on the final position of the axes.

The original distances of points from the origin must be reestablished after rotation. We should point out that none of these weighting schemes is considered to be generally better than any other. So the characteristics of each particular dataset must be considered to choose a weighting.

After rotation, an increase in simplicity can be assessed. This can be done by computing simplicity indices such as Bentler's (1977). As Bentler's index does not depend on the scale of the columns of the matrix evaluated, it can be used in the context of CA. Simplicity should be assessed in the rotated coordinate matrix (or matrices) where simplicity was aimed to be maximized. Bentler's simplicity index ranges between 0 and 1, where a value of 0 indicates maximum complexity, and a value of 1 indicates maximum simplicity.

4. Illustrative examples

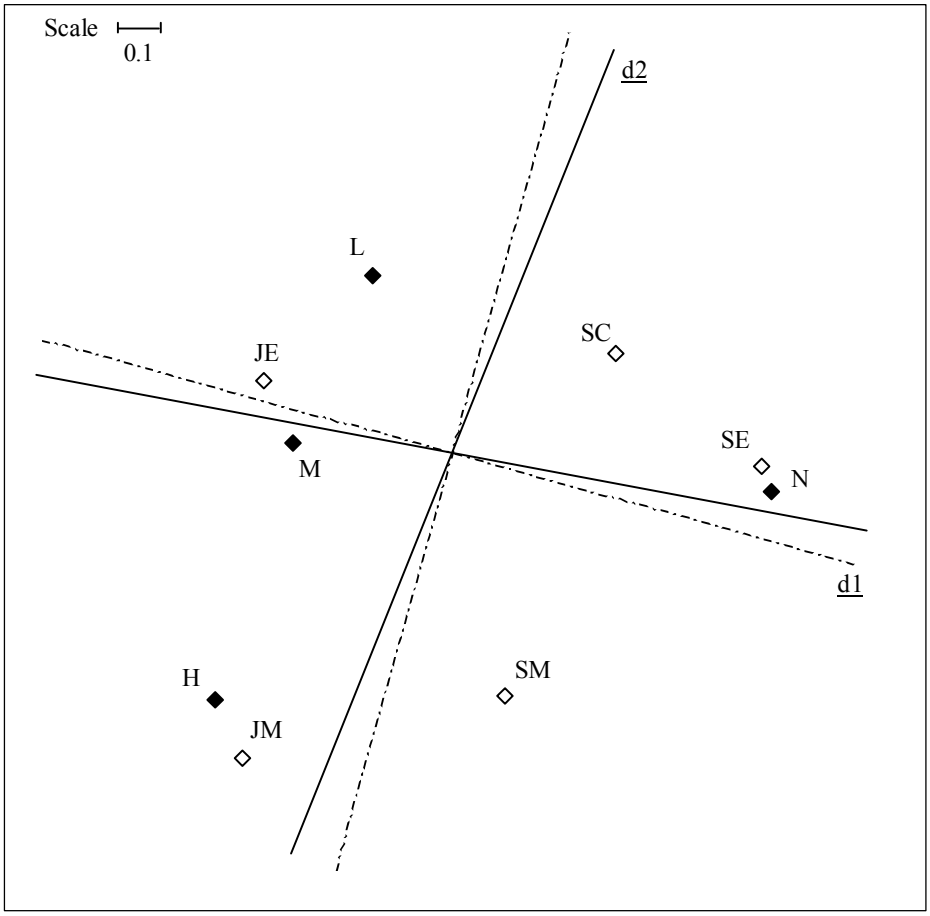
We present two illustrative examples. The first example, illustrates the procedure using a small data set that permits a graphical depiction of the results. The second example concerns a much larger data set, in which a graphical presentation would be complex to interpret. The interpretation of the numerical CA results is therefore the only choice for this dataset and, as we will see, rotation simplifies this interpretation considerably.

4.1. First illustrative example: smoking habits and types of workers

This example was presented by Greenacre (1984, p. 55), and consists of artificial data on the smoking habits of different types of workers in a company. Smoking habits were None (N), Light (L), Medium (M), and Heavy (H); while the types of workers were Senior Managers (SM), Junior Managers (JM), Senior Employees (SE), Junior Employees (JE), and Secretaries (SC). The corresponding contingency table of order 4×5 was analyzed using CA. The first eigenvalue was .07 (87.8% explained inertia), and the second was .01 (11.8% explained inertia). Greenacre (1984) interpreted two dimensions, so we retained the same number of dimensions. As we aim to describe both variables (smoking habits and employees) and their relationship, we computed symmetrical coordinates (i.e. α=.5).

Figure 1

Graphical presentation of rotated axes. Dotted and solid line axes correspond to pseudo-structure axes, and pattern axes, respectively.



Note: Row variables (different employees in the company) represented by hollow rhombi are SM = Senior Managers; JM = Junior Managers; SE = Senior Employees; JE=Junior Employees; and SC = Secretaries. Column variables (smoking habits) represented by solid rhombi are N = None; L = Light; M = Medium; and H = Heavy. Finally, d1 and d2 stand for Dimension 1, and Dimension 2, respectively. See text for further details.

Table 1

Symmetrical coordinates for first illustrative example. Salient coordinates printed in bold face (see text for further details).

Variable	Unrotated symmetrical coordinates		Pseudo-structure coordinates		Pattern coordinates	
	d1	d2	d1	d2	d1	d2
Senior Managers (SM)	0.13	-0.61	0.29	-0.55	0.74	**-1.89**
Junior Managers (JM)	-0.50	-0.77	-0.27	-0.88	-0.24	**-2.76**
Senior Employees (SE)	0.73	-0.03	0.71	0.17	**1.33**	0.32
Junior Employees (JE)	-0.45	0.18	-0.48	0.05	**-0.95**	0.33
Secretaries (SC)	0.38	0.25	0.30	0.35	0.48	1.03
None (N)	0.75	-0.10	0.75	0.11	**1.43**	0.13
Light (L)	-0.19	0.45	-0.30	0.38	-0.71	1.32
Medium (M)	-0.38	0.02	-0.37	-0.08	-0.69	-0.14
Heavy (H)	-0.56	-0.63	-0.37	-0.75	-0.48	**-2.34**

Table 1 shows the unrotated coordinates, and Figure 1 shows the corresponding graphical representation of points in the two dimensional space. The first dimension (horizontal axis) differentiates habit N (Non Smoking) from the other smoking habits, and also orders habits between H and N. This dimension also separates Juniors and Seniors, with JE and SE values at opposite ends. The second dimension (vertical axis) also differentiates between smoking habits, with H and L values at opposite ends. This dimension separates Managers and Secretaries, with values JM and SC at opposite ends. The unrotated solution has one drawback: both dimensions are related to the differentiation of smoking habits, so they somehow appear redundant. The inherent complexity of the substantive interpretation was obvious in the coordinates related to category H: these were substantial (that is to say, extreme) in both dimensions (values of -0.56 and -0.63).

We obliquely rotated the symmetrical coordinates following the procedure explained in Section 3.3. As weighting scheme explained in Section 3.5, we chose approach (b), the method proposed by Greenacre (2006a), so that infrequent categories have little influence on the final position of the axes.

Table 2

Bentler's simplicity index for different coordinate matrices

	Unrotated symmetrical coordinates	Pseudo-structure coordinates	Pattern coordinates
Employees coordinates	.859	.951	.973
Smoking habits coordinates	.782	.921	.968

We computed Bentler's simplicity index in order to assess whether the rotated solution was simpler than the unrotated solution. Table 2 shows the corresponding values of the index for each matrix. As expected, values of the simplicity index reported that the pattern matrices were the simplest matrices and, therefore the most easily interpreted. As explained in Section 3.4, we interpret the solution by studying the rotated pattern matrices and the standardized inter-dimensions covariance matrix.

Table 1 also shows the rotated coordinates related to the pseudo-structure matrices (i.e. matrices $\widetilde{\mathbf{X}}_{ps}$ and $\widetilde{\mathbf{Y}}_{ps}$ for row and columns variables, respectively) and the pattern matrices (i.e. matrices $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Y}}_2$ for row and column variables, respectively). We determined the salient coordinates in the rotated patterns using the procedure explained in Section 3.4. These coordinates are printed in bold face.

We first consider smoking habits. On dimension d1 only one coordinate (related to category N) was salient. This dimension seemed to differentiate habit N from the other smoking habits. In addition, the coordinate value for category H in d1 was the closest to the origin of the axes. Only one coordinate (related to category H) was salient on dimension d2, too. Now the coordinate value for category N in d2 was the closest to the origin. We could call d1 *non-smoking habits*, and d2 *strong smoking habits*. Let us now consider the types of workers in the company. On dimension d1, two coordinates (related to SE and JE) were salient, each with a different sign. Because of this configuration of different signs, it could be said that this dimension was bipolar. On dimension d2, two coordinates (related to categories SM and JM) were also salient, both with the same sign. Now we could call d1 *Employees*, and d2 *Managers*. Finally, we can put all this together to interpret the smoking habits of the workers in the company. Dimension d1 was a

bipolar dimension, and showed that Junior Employees had a tendency to non-smoking habits, whereas Senior Employees had a tendency towards smoking habits. Dimension d2 was a unipolar dimension, and showed that Managers had a tendency towards strong smoking habits, especially Junior Managers.

Table 3

Inter-dimensions variance/covariance matrices for first illustrative example

Dimensions	Raw variance/covariance matrix		Standardized variance/covariance matrix	
	d1	d2	d1	d2
d1	0.27	--	1	--
d2	0.04	0.10	.25	1

Table 3 shows the inter-dimensions variance/covariance matrix (i.e. matrix **S**) and the corresponding standardized matrix $\mathbf{\Phi}$. The correlation between dimensions was .20. If d1 and d2 are related to non-smoking habits and strong smoking habits, respectively, one would expect these two dimensions to be orthogonal (i.e. a correlation of zero). However, it must be noted that the dimensions are also related to the types of workers in the company, and that each type of worker was not so simple as to be related to only one smoking behavior. Secretaries was the most complex group: they were between low and non-smoking habits. Senior managers were also a complex group: they were close to strong smoking habits but some of them had no smoking habits.

This illustrative example is simple enough to allow a comprehensive graphical representation of rotated axes. Figure 1 shows the rotated solution, where dotted lines correspond to pseudo-structure axes, and solid lines correspond to pattern axes. Please note that points are presented unrotated, whereas axes are rotated. It can be observed that the rotation drew the axes closer to the points, which helped to simplify the interpretation of the coordinates.

4.2. Second illustrative example: Spanish tourist destinations from different countries

This example consists of data collected from the Spanish Hotel Occupancy Survey during the year 2005, which was carried out by the Spanish National Statistics Institute. The data is available on the Institute's web site (www.ine.es). The data used in this analysis corresponds to the number of travelers checked in by Spanish provinces, and the country of residence of the traveler.

The 50 provinces and 20 residence countries led to a contingency table of 1,000 cells, and a total frequency of 25,092,282 registered visits. We analyzed the contingency table using CA to interpret the possible underlying dimensions. The first five eigenvalues were: 0.2926 (63.8% explained inertia), 0.0687 (14.98% explained inertia), 0.0325 (7.08% explained inertia), 0.0266 (5.79% explained inertia), and 0.0119 (2.59% explained inertia). To determine the dimensionality of a solution, Greenacre and Blasius (1994, p. 29) suggested to discard those dimensions for which the explained inertia is below average (100 divided by the minimum number of rows and columns, minus one). With this contingency table, the average explained variance was 4.76%, so four dimensions were retained.

Again, we are interested in the relationship between row and column variables (Spanish province and Country of residence, respectively), so we computed symmetrical coordinates (i.e. α=.5). The corresponding graphical representation of points in the four-dimensional space was ignored, as this would mean having to interpret 6 plots simultaneously. The two matrices of unrotated symmetrical coordinates were of orders 50×4 and 20×4, and turned out to be quite complex: Bentler's simplicity index (1977) reported a value of .03 and .02 for row and column coordinates, respectively (note that a value of zero would mean maximum complexity). For example, Portugal showed coordinate values of -2.09, -3.82, -1.79, and 0.11; the United States showed coordinate values of -1.63, 0.05, 2.07, and -2.06; and Pontevedra showed coordinate values of -1.43, -2.26, -1.03, and 0.19. So interpreting the unrotated solution seemed a difficult task.

Table 4

Row pattern coordinates for the second illustrative example. Salient coordinates printed in bold face. (See text for further details)

Spanish provinces	d1	d2	d3	d4
Alacant	**-2.06**	0.46	0.17	0.88
Tarragona	**-1.70**	0.43	-0.46	-1.37
Murcia	**-1.04**	-0.05	0.14	0.23
Málaga	**-0.93**	-0.02	-0.31	0.63
Almería	**-0.89**	0.72	-0.03	0.30
Lugo	**0.67**	-0.02	1.41	-0.62
León	**0.70**	-0.25	1.04	-0.74
Cádiz	**0.94**	0.41	0.15	-0.30
Madrid	0.26	**-2.20**	0.55	0.89
Seville	0.25	**-1.57**	0.51	0.03
Segovia	0.05	**-1.47**	0.45	-0.24
Toledo	0.30	**-1.33**	0.92	-0.58
Granada	0.28	**-1.27**	-0.07	-0.36
Barcelona	-0.15	**-1.21**	-0.43	0.09
Salamanca	0.02	**-0.99**	2.47	-0.48
València	-0.09	**-0.89**	-0.04	-0.18
Santa Cruz de Tenerife	-0.44	**0.79**	-0.36	0.68
Las Palmas	**0.96**	**1.00**	-0.48	0.63
Balearic Islands	**0.68**	**1.23**	-0.44	0.53
Cáceres	0.02	-0.05	**2.92**	-0.06
La Coruña	0.40	-0.51	**3.66**	0.54
Zamora	-0.06	0.19	**3.87**	-0.46
Huelva	-0.15	**0.98**	**4.09**	1.31
Guadalajara	0.12	0.21	**4.20**	-0.69
Badajoz	0.07	0.25	**6.73**	1.38
Orense	0.03	**0.75**	**7.75**	0.44
Pontevedra	0.07	**0.79**	**9.49**	1.22
Huesca	-0.15	0.15	-0.16	**-4.17**
Lleida	0.02	0.10	0.62	**-4.10**
Girona	-0.15	0.10	-0.88	**-3.15**
Castelló	0.18	-0.17	-0.32	**-2.82**
Palencia	-0.03	0.58	2.44	**-2.28**
Guipúzcoa	0.11	-0.72	0.13	**-2.19**
Valladolid	-0.11	0.10	2.46	**-2.16**
Navarre	0.55	-0.37	0.20	**-1.92**
Burgos	0.09	-0.20	1.26	**-1.78**

Teruel	-0.38	-0.12	0.04	**-1.72**
Jaén	-0.15	-0.22	-0.14	**-1.72**
Albacete	-0.16	-0.07	0.80	**-1.61**
Soria	-0.22	0.20	1.85	**-1.57**
Ciudad Real	-0.18	-0.53	0.47	**-1.53**
Ávila	0.29	-0.63	2.24	0.34
Asturias	-0.08	-0.21	1.69	-0.29
Zaragoza	0.34	-0.44	1.69	-1.10
Álava	-0.16	-0.02	1.46	-0.80
Cuenca	-0.26	-0.69	0.82	0.14
La Rioja	0.19	-0.10	0.26	-1.38
Córdoba	0.49	-0.67	-0.27	-1.35
Vizcaya	0.05	-0.63	0.19	-0.77
Cantabria	-0.32	-0.07	0.59	-0.65

Table 5

Column pattern coordinates for the second illustrative example.

Salient coordinates printed in bold face (see text for further details).

Country of residence	d1	d2	d3	d4
Russia	**-1.51**	-0.09	-0.55	**-1.37**
United Kingdom	**-0.86**	0.49	-0.25	0.68
Belgium	**-0.78**	-0.02	-0.25	-0.90
Luxembourg	**0.75**	0.53	-0.26	0.40
Germany	**1.10**	1.20	-0.37	0.32
Greece	0.41	**-3.52**	-0.02	**1.59**
The United States	0.34	**-3.21**	0.16	**1.15**
Portugal	0.08	-0.25	**5.51**	0.76
France	-0.23	-0.44	-0.18	**-2.72**
Finland	-0.37	-0.80	-0.31	**1.04**
Austria	0.58	-0.35	0.03	0.48
Italy	0.32	-1.13	-0.16	-0.15
Poland	0.03	-1.05	-0.07	-0.33
The Netherlands	-0.42	-0.73	-0.37	-0.10
Switzerland	0.20	-0.54	-0.27	-0.42
Czech Republic	0.30	-0.31	-0.79	-0.94
Norway	0.25	-0.36	-0.15	0.74
Sweden	0.44	0.00	-0.43	0.30
Denmark	0.39	-0.34	-0.29	0.64
Ireland	-0.28	-0.40	-0.14	0.53

We obliquely rotated the symmetrical coordinates following the procedure explained in Section 3.3 with the aim of obtaining a simple interpretation. We chose approach (c) described in Section 3.5 as the weighting scheme, so that all the variables have the same influence on the final positions of the axes. Tables 4 and 5 show the rotated coordinates related to the pattern matrices for rows and columns (i.e. matrices $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Y}}_2$), respectively. We determined the salient coordinates in the rotated patterns using the procedure explained in Section 3.4. These coordinates are printed in bold face. Now, Bentler's simplicity index (1977) reported a value of .943 and .803 for row and column rotated coordinates, respectively. Visual inspection of Tables 4 and 5 shows that each dimension was related to a relatively low number of salient coordinates, while only five Spanish provinces and three Countries of residence had salient coordinates in more than one dimension. To help to understand the rotated solution, a map of Iberia (obtained from www.pupilvision.com) is presented in Figure 2 with the names of provinces included. We studied the salient coordinates and arrived at the following interpretation of dimensions:

Figure 2

Map of Iberia with the names of provinces included.

(a) Dimension d1: the configuration of coordinates related to Spanish provinces defined the dimension as bipolar. The typical Mediterranean beaches that represent a cheap tourist option were in the negative pole. The provinces in the positive pole were mainly islands or provinces with Atlantic beaches. The configuration of coordinates related to Country of residence was also bipolar. People from Russia, United Kingdom, and Belgium seemed to prefer the Mediterranean beaches, whereas people from Germany and Luxembourg seemed to prefer the Atlantic beaches.

(b) Dimension d2: the configuration of coordinates related to Spanish provinces also defined this dimension as bipolar. In the negative pole there were provinces with historic cities. The provinces in the positive pole were islands, or provinces in the north or south of Portugal. The configuration of coordinates related to Country of residence was unipolar, and showed that people from Greece and the United States seemed to prefer to visit the provinces with historic cities.

(c) Dimension d3: all the salient coordinates in this dimension were positive. All the provinces that border with Portugal had a salient positive coordinate in this dimension. The only exception was Salamanca, even though its coordinate was also large in this dimension. Although its coordinate was not salient, Girona was the province with the largest negative coordinate: this is the most eastern province of the country, the furthest from Portugal, and close to the border with France. The configuration of coordinates related to Country of residence showed that only Portugal had a salient coordinate in this dimension. This meant that people from Portugal visited the Spanish provinces that are close to their border.

(d) Dimension d4: all the salient coordinates in this dimension were negative. The largest coordinates were for provinces in the North-East of the country, on the border with France. Although its coordinate was not salient, Badajoz, a province in the South-West of the country and bordering Portugal, was the province with the largest positive coordinate. The configuration of coordinates related to Country of residence showed that France and, to a lesser extent, Russia had a salient coordinate in this dimension. This meant that mainly French people visited provinces close to the eastern border with France.

Nine (out of 50) provinces showed no salient coordinates. However, the configuration of coordinates allows an interpretation that is coherent with the characteristics of the provinces. Of the 10 residence countries with no salient coordinates, only three showed systematically low and similar coordinates across dimensions (Sweden, Denmark, and Ireland).

Table 6

Standardized inter-dimensions variance/covariance matrix for second illustrative example

	d1	d2	d3	d4
d1	1.00			
d2	-.07	1.00		
d3	.02	-.06	1.00	
d4	.02	.15	-.22	1.00

Table 6 shows the standardized inter-dimensions covariance matrix. Two inter-dimension correlations were substantive. The largest correlation (-.22) was between dimensions d3 and d4. Note that d3 sorts Spanish provinces from West to East in the country, whereas d4 sorts Spanish provinces from South-West to North-East in the country. So the correlation between these two dimensions can be easily understood. The negative values of the correlation means that people from a particular country who visited provinces in the West did not visit the provinces in the East. The second correlation (.15) was between dimensions d2 and d4. It can be interpreted as follows: French people visit the provinces closest to their border (dimension d2), but also provinces with historic cities.

As the correlation values between dimensions were not high, we also computed orthogonal varimax rotation (Van de Velden, 2000; Van de Velden & Kiers, 2005) in order to compare the oblique and the orthogonal solutions. Bentler's simplicity index (1977) applied to the orthogonally rotated solution reported a value of .796 and .733 for row and column rotated coordinates, respectively. As these indexes were lower than the

ones obtained for the oblique solution (.943 and .803, respectively), we concluded that, even if the correlation between dimensions were not high, the oblique rotation achieved a solution considerably simpler than the orthogonal one. So the oblique rotation seems justified for his dataset.

Finally we analyzed the contingency table for only the provinces and residence countries that showed salient coordinates in at least one dimension in the previous analysis. The 41 provinces and 10 residence countries led to a contingency table of 410 cells, and a total frequency of 19,040,788 registered visits. The variance explained by the eigenvalues suggested that four dimensions should be retained. So we ran the analysis using the same procedure as in the previous analysis. To assess the similarity between both analyses, we computed the congruence index (Tucker, 1951) between columns in the pattern coordinates (for Spanish provinces and Country of residence). Lorenzo-Seva and Ten Berge (2006) suggested that a value in the range [.85, .94] corresponds to a fair similarity, while a value higher than .95 implies that the two columns compared can be considered equal. The values of the index ranged from .93 to .99 for seven of the comparisons. Only one value was outside this range (.85). The lowest congruence was obtained for Spanish provinces in dimension d2. In fact, the negative pole of the dimension considered the same provinces in almost the same order (only Granada and Toledo permuted their positions). However, in the positive pole the islands disappeared and were substituted by Tarragona and Palencia. None of the provinces in this pole had well-known historic cities. We concluded that eliminating the provinces and countries that were closest to the origin of dimensions (i.e. variables with non salient coordinates in any dimension) did not substantially change the meaning of the dimensions.

5. Discussion

In this paper we considered oblique rotation as an additional tool for interpreting solutions in CA. If the dimensionality of the CA solution is such that a graphical representation of the solution is no longer possible, an interpretation of the numerical CA coordinates is needed. Similar as in principal component analysis and factor analysis, rotation can in such cases be a useful tool. The rotational freedom was earlier studied and

exploited by van de Velden and Kiers (2003, 2005), however, in these papers attention was restricted to orthogonal (varimax) rotation.

In factor analysis, oblique rotation is often favored over orthogonal rotation (see, for example, Browne, 2001), especially in case of the interpretation of results with more than two factors. It therefore seems logical to likewise use oblique rotation in CA. The introduction of oblique rotation in CA, however, is not straightforward. For example, unlike the situation in the orthogonal setting, oblique rotation yields different ($\mathbf{T}$ and $\left(\mathbf{T}'\right)^{-1}$) rotation matrices for the two sets of points. Moreover, the different scaling options for the sets of coordinates, that is, principal, standard and symmetrical coordinates for row and columns, require different approaches. Depending on the choice of the CA scaling, different oblique rotation procedures must be considered. With respect to this it is important to note that we only consider CA solutions that satisfy, before and after rotation, the so-called biplot property. That is, the inner products of the row and column coordinates approximate the elements of the matrix of standardized residuals. Using the Quartimin criterion, we derived, by using the gradient method, two oblique rotation procedures. The first approach yields simple structure for one set of principal coordinates. The second approach simultaneously maximizes simplicity in the two sets of symmetrical coordinates.

To interpret the oblique rotation results, we suggested methods that are similar to those that are common practice in factor analysis. For the asymmetric solutions (i.e., one set of coordinates is in principal and the other in standard coordinates) we propose to interpret the rotated pattern matrices $\mathbf{X}_p$ and $\mathbf{Y}_p$ together with the interdimensions correlation matrix $\mathbf{\Phi}$. The pattern matrices can be used to give labels to dimensions and to assess which variables are best related to the dimensions. The interdimensions correlation matrix $\mathbf{\Phi}$ indicates how closely dimensions are related to each other. For the simultaneous rotation of the symmetrical coordinates, we propose a similar scheme.

Greenacre (2006a) pointed out that the role of the weights, which was ignored in van de Velden and Kiers (2005), may be of considerably importance in rotation. To resolve this problem, he proposed the elimination of the effect of the weights by rescaling the coordinates before rotation. In this paper we consider this option as well as an alternative that considers rotation of the row-wise normalized coordinates. This latter

approach, which is common in EFA, ensures that all rows have the same influence on the final positions of the axes.

We illustrated our oblique rotation method by applying it to one small example that permitted a graphical representation, and a larger example that required a four dimensional solution. This last example, based on real data, clearly demonstrates the practical value of oblique rotation. The complex original solution is rotated to a simple rotated solution whose interpretation is relatively straightforward. Moreover, comparing the obliquely rotated solution with an orthogonally rotated solution, shows that abandoning the orthogonality constraint leads to a considerably larger simplicity of the solution.

Although we think that oblique rotation should certainly be considered when trying to interpret high-dimensional CA solutions, we do not claim that oblique rotation always provides a useful or better solution than the original solution. There may be cases in which rotation does not simplify the solution, or that lead to simple solutions that are still not easy to interpret. Also, there may be specific reasons that prohibit the use of rotation. For example, if the original distribution of inertia over the dimensions is instrumental, rotation cannot be applied. However, as was illustrated by the examples in this paper, we do believe that rotation can provide a useful additional tool for interpreting a CA solution.

References

Adachi, K. (2004). Oblique promax rotation applied to solutions in multiple correspondence analysis. Behaviormetrika, 31, 1-12.

Bentler, P. M. (1977). Factor simplicity index transformations. Psychometrika, 42, 277-295.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research, 36, 111-150.

Clarkson, D. B., & Jennrich, R. I. (1988). Quartic rotation criteria and algorithms. Psychometrika, 53, 251-259.

Gabriel, K. R. (1971). The biplot graphical display of matrices with applications in Principal Component Analysis. Biometrika, 58, 453-467.

Greenacre, M. J. (1984). Theory and Applications of Correspondence Analysis. London: Academic Press.

Greenacre, M. J. (2006a). Tying up the loose ends in simple correspondence analysis. http://www.econ.upf.es/docs/papers/downloads/940.pdf.

Greenacre, M. J. (2006b). From simple to multiple correspondence analysis. In M. J. Greenacre and J. Blasius, (Eds.), Multiple correspondence analysis and related methods (41-76). Boca-Raton, FL: Chapman-Hall.

Greenacre, M. J., & Blasius, J. (1994). Correspondence Analysis in the Social Sciences: Recent developments and Applications. London: Academic Press.

Jennrich, R. I. (2002). A simple general method for oblique rotation. Psychometrika, 67, 7-20.

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings, Psychometrika, 31, 313-323.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, 187-200.

Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39, 31-36.

Lorenzo-Seva, U. (1999). Promin. A method for oblique factor rotation. Multivariate Behavioral Research, 34, 347-365.

Lorenzo-Seva, U., & Ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology, 2, 57-64.

Tenenhaus, M., & Young, F. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis ans other methods for quantifying categorical multivariate data. Psychometrika, 50, 91-119.

Thurstone, L. L. (1947). Multiple factor analysis. Chicago, IL: University of Chicago Press.

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. Personnel Research Section Report No.984, Washington D.C.: Department of the Army.

Van de Velden, M. & Kiers, H. A. L. (2003). An application of rotation in correspondence analysis. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, and J.J. Meulman, (Eds.), New Developments in Psychometrics (471-478). Tokyo: Springer Verlag.

Van de Velden, M. & Kiers, H. A. L. (2005). Rotation in correspondence analysis. Journal of Classification, 22, 251–271.

Van de Velden, M. (2000). Topics in Correspondence Analysis. Ph.D. Thesis, Tinbergen Institute Research Series, 238: University of Amsterdam.