ELSEVIER Computational Statistics & Data Analysis 37 (2001) 49–64

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Factorial k-means analysis for two-way data

Maurizio Vichi[a,*], Henk A.L. Kiers[b]

[a]*Dipartimento di Statistica Probabilità e Statistiche Applicate, Università di Roma "La Sapienza", P.le A. Moro 5, I-00185, Italy*

[b]*Heymans Institute (PA), Grote Kruisstraat 2/1, 9712 TS Groningen, Netherlands*

Received 1 March 2000

Abstract

A discrete clustering model together with a continuous factorial one are fitted simultaneously to two-way data, with the aim of identifying the best partition of the objects, described by the best orthogonal linear combinations of the variables (factors) according to the least-squares criterion. This methodology named for its features *factorial k-means analysis* has a very wide range of applications since it fulfills a double objective: data reduction and synthesis, simultaneously in the direction of objects and variables; variable selection in cluster analysis, identifying variables that most contribute to determine the classification of the objects. The least-squares fitting problem proposed here is mathematically formalized as a quadratic constrained minimization problem with mixed variables. An iterative alternating least-squares algorithm based on two main steps is proposed to solve the quadratic constrained problem. Starting from the cluster centroids, the subspace projection is found that leads to the smallest distances between object points and centroids. Updating the centroids, the partition is detected assigning objects to the closest centroids. At each step the algorithm decreases the least-squares criterion, thus converging to an optimal solution. Two data sets are analyzed to show the features of the factorial k-means model. The proposed technique has a fast algorithm that allows researchers to use it also with large data sets.

Keywords: Cluster analysis; Factorial model; Tandem analysis; k-means algorithm

1. Introduction

Often the relationships within a set of objects are explored by fitting discrete classification models as partitions, n-trees, hierarchies, via non-parametric techniques of

* Corresponding author.

E-mail addresses: maurizio.vichi@uniroma.it (M. Vichi), h.a.l.kiers@ppsw.rug.nl (H.A.L. Kiers).

PII: S 0167-9473(00)00064-5

clustering. In other situations, the relationships within a set of variables are explored by using continuous models (e.g., principal component analysis (PCA) or factor analysis (FA)), that detect non-observable dimensions summarizing the common information in the data set.

When the number of variables is large or when it is believed that some of these do not contribute much to identify the clustering structure [1] in the data set, researchers apply the continuous and discrete models sequentially, frequently carrying out a PCA and then applying a clustering algorithm on the object scores on the first few components. However, De Sarbo et al. (1990), and De Soete and Carroll (1994) warn against this approach, called "*tandem analysis*" by Arabie and Hubert (1994), because PCA or FA may identify dimensions that do not necessarily contribute much to perceive the clustering structure in the data and that, on the contrary, may obscure or mask the taxonomic information.

In an attempt to focus on the main information in the data, conveyed in a limited number of dimensions, De Soete and Carroll (1994) proposed an alternative to the k-means procedure such that each cluster is represented by a point (which they call a "centroid") in a low-dimensional space, chosen such that these points are closest to the objects associated with the cluster at hand. In a second step, they project the objects back into the space containing the centroids, and hence they give a low-dimensional representation of objects and clusters. Thus, in their procedure, a subspace of the full data space is found such that the actual data points (in the full space) have the smallest sum of squared distances to the centroids (which lie in a subspace only).

De Soete and Carroll's (1994) method may still fail to find an interesting clustering residing in a subspace of the data, when the data have much variance in directions orthogonal to the one capturing the interesting clustering. This is because the variance in such directions may contribute considerably to the sum of squared distances between the data points and the centroids. However, considering that one is only interested in a subspace representation of the data, it seems more consistent to find this subspace such that the *projected* data points (i.e., projected onto this subspace) have the smallest sum of squared distances to centroids in the same subspace. In the present paper, a procedure for this very purpose is described. The procedure, which seems closely related to a procedure described (in French) by Diday et al. (1979), combines k-means cluster analysis with aspects of factor analysis and PCA (finding the best subspace) and is, therefore, called factorial k-means analysis here.

Various alternative methods combining cluster analysis and the search for a low-dimensional representation have been proposed, but these focus on multidimensional scaling or unfolding analysis (e.g., Heiser, 1993; De Soete and Heiser, 1993), whereas here the more simple idea of component analysis (and the ensuing projection on a subspace) is followed. In various instances (e.g., DeSarbo et al., 1991), a mixture modeling approach is taken for simultaneously performing multidimensional scaling and fuzzy cluster analysis of two-way data. This maximum likelihood

[1] The terms taxonomic information, clustering structure, classification are used interchangeably.

approach is based on normal distribution assumptions, which are avoided in our approach.

In the present paper, first, it is shown that tandem analysis may not be appropriate in recognizing the taxonomic information in the data set. This demonstrates that it is worthwhile to study new methodologies that help investigators to select relevant components to identify the best partition of the objects in the observed data, as is done in factorial k-means analysis.

Section 2 shows an example in which tandem analysis as well as De Soete and Carroll's (1994) method completely fail to correctly identify three well-separated classes masked by randomly generated variables. Sections 3 and 4 provide the notation and the factorial k-means model. In Section 4, also an extension of the model towards using more than one clustering is described briefly. An efficient alternating least-squares (ALS) algorithm for the factorial k-means model is given in Section 5, and it is indicated how to modify this for its extensions. Two data sets are analyzed in Sections 6 and 7. In particular, it is shown that the factorial k-means model correctly identifies three well-separated classes masked by randomly generated variables considered in Section 2. In Section 7, the last short-term macroeconomic data (September 1999) provided by the Organization for the Economic Cooperation and Development (OECD) are investigated using factorial k-means analysis and tandem analysis. Some conclusions follow in Section 8.

2. Masking of the clustering structure, tandem analysis and k means in low-dimensional space

An example of the masking of the clustering structure due to the inclusion of irrelevant variables may help to clarify the problems discussed here. In Fig. 1, the plot of 42 objects according to two variables is shown; this plot is similar to that given in Gordon (1999). Examining the location of the objects, disregarding the labels for the moment, it can be clearly observed that the data comprise a well-defined clustering structure in three separated classes, with centers located at the vertices of an equilateral triangle with sides of length six. However, the 42 objects were also described by four other variables randomly generated by a normal distribution with mean 0 and variance 6. On the 42×6 data matrix, the k-means clustering algorithm (Ball and Hall, 1967; MacQueen, 1967) was applied. The results of the classification are illustrated, in Fig. 1, by the labels (1–3) identifying the class membership of the 42 objects.

It can be noticed that the clustering structure observed in the two-dimensional data set has been masked, even if not completely hidden, by the six-dimensional data set. In particular, 17 of the 42 objects have been incorrectly classified.

Using PCA to extract the most relevant information in the data set, a researcher may think to reduce the masking effect but this does not happen as can be seen from Figs. 2(a), (b), (c) and (d), where PCA is applied on the six-dimensional data set and sequentially k-means is applied on the two-, three-, four-, and five-dimensional components' scores, respectively (when the k-means algorithm is applied, respectively,

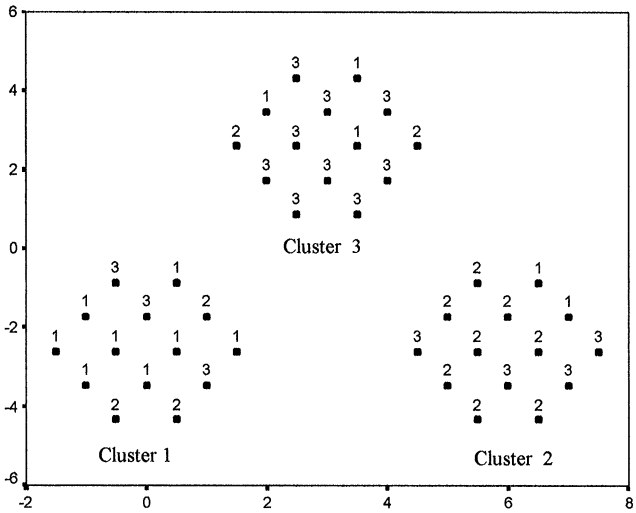

Fig. 1. Classification of 42 objects described by six variables two of which determine the location of the points in the plot (three classes) and the other four variables are randomly generated by normal distribution.

on the scores of 3, 4, and 5 principal components, the location of the objects in the plots (b), (c) and (d) is given by the two original variables that define the three well-separated classes, while the class membership, i.e., the labels in these plots are specified by the results of the respective tandem analyses). The percentages of the total variance explained by the different PCA solutions are shown in Table 1.

In Fig. 2(a) the first two principal components do not represent the original three well-separated classes and objects that should be seen in the same classes are not correctly classified. The situation becomes slightly better when one increases the number of principal components on which the k-means algorithm is applied (Fig. 2(d)), in this case the number of not correctly classified objects is slightly reduced (from 17 to 12).

The tandem analysis as carried out here is the one usually applied. One could, however, consider that using standardized principal components distorts the mutual distances between the objects (see, e.g., Kiers, 1997 in a similar context). To avoid this problem, one should replace the standardized component scores by components scores multiplied by the square roots of the eigenvalues. In this way, the distances between objects are conserved much better, and, in fact, taking all components would reproduce them exactly. When doing this for the present situation, the outcomes using 2, 3, 4 or 5 components hardly differed from the outcomes obtained for standardized components. Therefore, it can be concluded that the poor results of tandem analysis are not due to a bad representation of the data space, but to the fact that the principal components are not related to the best clusterings.

Because we know that, in the present case, the clustering in fact resides in a low-dimensional subspace of the full data space, we also applied DeSoete and Carroll's (1994) method to these data, specifying the centroids to be in a two-

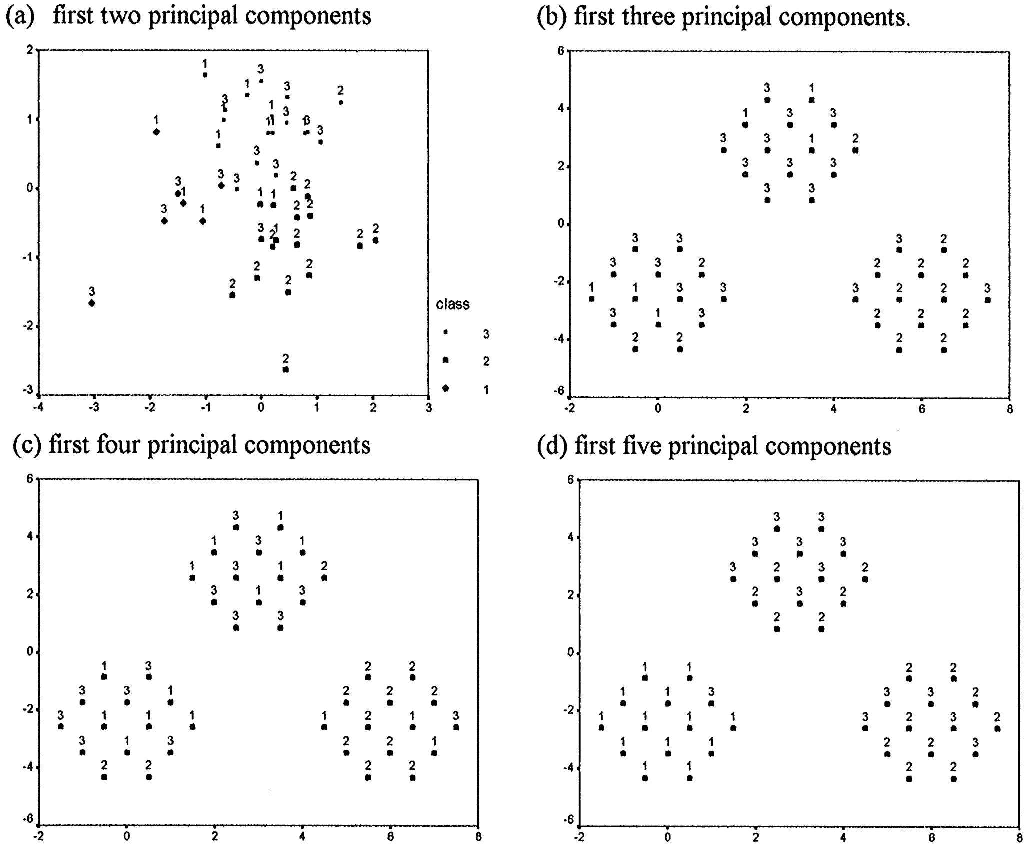

Fig. 2. Tandem analysis. k-means clustering computed on: (a) the first two components' scores; (b) three components' scores; (c) four components' scores; (d) five components' scores. Note. Classifications (b), (c) and (d) are represented on the two variables that defined the three well-separated classes in Fig. 1.

Table 1
Explained total variance and cumulated variance

Components	Eigenvalue	% variance	% cumulated
1	1.444	24.072	24.072
2	1.329	22.152	46.223
3	1.137	18.950	65.173
4	0.787	13.114	78.287
5	0.663	11.049	89.336
6	0.640	10.664	100.000

dimensional subspace. The result is given in Fig. 3. It can be seen that their method also fails to recover the actual clustering in the data.

We will now turn to the factorial k-means procedure, which is better suited for handling such data, but before doing so, we give the notation used in the present paper.

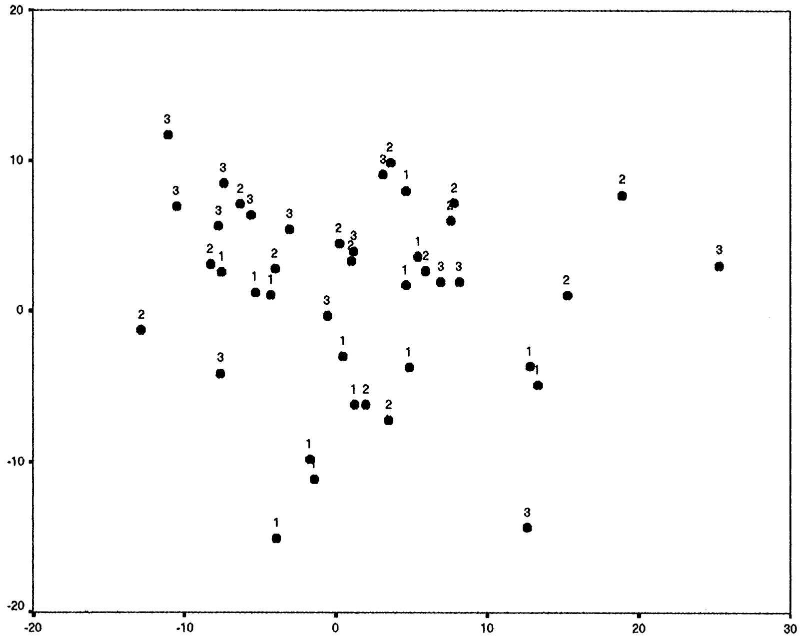

Fig. 3. De Soete and Carroll's k-means clustering in a low-dimensional space.

3. Notation

Let $\boldsymbol{X} = [x_{ip}]$ be a $(n \times k)$ two-way two-mode (objects and variables) data matrix, regarding the k-variate profiles of n objects.

For the convenience of the reader, the terminology used in this paper is listed here:

n, k	number of objects to be partitioned and number of variables
c	number of classes of the partition
m	number of factors that describe (explain) the partition
$I = \{o_1, \ldots, o_n\}$;	set of n objects to be partitioned
$P = \{P_1, \ldots, P_c\}$	partition of I into c classes, where P_j is the jth class of I
$\boldsymbol{U} = [u_{ij}]$	$(n \times c)$ matrix specifying for each o_i the membership to each class P_j, i.e., $u_{ij} = 1$ if o_i belongs to P_j, $u_{ij} = 0$ otherwise
$\boldsymbol{J} = \boldsymbol{I}_n - (1/n)\mathbf{1}_n\mathbf{1}_n'$	$(n \times n)$ centering operator, where $\mathbf{1}_n$ is a n-vector of unit elements
$\boldsymbol{A} = [a_{pl}]$	$(k \times m)$ columnwise orthonormal matrix. The a_{pl} are the coefficients of the linear combinations of the observed variables

$\boldsymbol{Y} = [y_{il} = \sum_{p=1}^{k} a_{pl}x_{ip}]$	$(n \times m)$ factor matrix, where y_{il} is the score of o_i for the lth factor y_l
$\bar{\boldsymbol{Y}} = [(\sum_{i=1}^{n} u_{ij})^{-1} \sum_{i=1}^{n} y_{il}u_{ij}]$	$(c \times m)$ matrix containing centroids $\bar{y}_{jl}$ of P_j based on factors y_l

The model proposed here defines an optimal partition P of I into c classes P_j $(j = 1,\dots,c)$ and, simultaneously, a subset of factors y_l $(l = 1,\dots,m)$, which are linear combinations of the k variables.

4. The factorial *k*-means model

The model is mathematically specified as follows:

$$\boldsymbol{XAA}' = \boldsymbol{U\bar{Y}A}' + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{E}$ is a matrix of error components. Model (1) describes an orthogonal projection onto a subspace spanned by the columns of a columnwise orthonormal matrix $\boldsymbol{A}$ $(k \times m)$. The coordinates of the projections onto the basis are given by the components y_l $(l = 1,\dots,m)$, collected in the matrix $\boldsymbol{Y} = \boldsymbol{XA}$. Within this subspace, hence, with these components, a partition of the objects is sought such that the objects are "closest" to the centroids of the clusters of objects. Model (1) is fitted in the least-squares sense by minimizing

$$\begin{aligned} F(\boldsymbol{A}, \boldsymbol{U}, \bar{\boldsymbol{Y}}) &= ||\boldsymbol{XAA}' - \boldsymbol{U\bar{Y}A}'||^2 \\ &= ||\boldsymbol{XA} - \boldsymbol{U\bar{Y}}||^2 \end{aligned} \tag{2}$$

subject to the constraints that $\boldsymbol{U}$ is binary and has only one unit element per row, and that $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$. The optimal $\bar{\boldsymbol{Y}}$ can be expressed in terms of $\boldsymbol{A}$ and $\boldsymbol{U}$ as $\bar{\boldsymbol{Y}} = (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{XA}$, so it remains to minimize

$$F(\boldsymbol{A}, \boldsymbol{U}) = ||\boldsymbol{XA} - \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{XA}||^2, \tag{3}$$

corresponding to the *within*-classes deviance of the partition P of I described by the factors $\boldsymbol{XA}$. It can be easily observed that function F can be decomposed into the two terms:

$$F(\boldsymbol{A}, \boldsymbol{U}) = tr(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{XA}) - tr(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{XA}), \tag{4}$$

where the first component is the total deviance of $\boldsymbol{XA}$ and the second component is the *between*-classes deviance of P.

To minimize (2), or equivalently, (4), an ALS algorithm will be described in the next section. Here, some further properties of the model will be described, as well as an extension of it will be given. First of all, we note that the method is not sensitive to the addition of a constant to columns of $\boldsymbol{X}$. This follows at once from the fact that columnwise centering of $\boldsymbol{X}$ does not change the function value of (3), as can be seen as follows:

$$
\begin{aligned}
F^*(\boldsymbol{A},\boldsymbol{U}) &= ||\boldsymbol{JXA} - \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{JXA}||^2 \\
&= ||\boldsymbol{XA} - (1/n)\mathbf{1}_n\mathbf{1}_n'\boldsymbol{XA} - \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{XA} \\
&\quad + (1/n)\mathbf{1}_n\mathbf{1}_n'\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{XA}||^2 \\
&= ||\boldsymbol{XA} - \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{XA}||^2,
\end{aligned} \tag{5}
$$

where it is used that $\mathbf{1}_n'\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}' = \mathbf{1}_n'$, because the column space of $\boldsymbol{U}$ contains the vector $\mathbf{1}_n$. This result implies that the method is insensitive to the choice of origin taken for the data. Furthermore, we can therefore, to standardize the procedure, always analyze the columnwise centered version of $\boldsymbol{X}$.

In factorial k-means, one has to choose a number of clusters to use, as well as the number of components to be used. These choices should, however, not be made independent of each other. This is because the rank of $\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{JXA}$ is equal to $\min(c-1, m)$, hence, it seems rather wasteful to take the number of components larger than $(c-1)$, as this corresponds to describing a low-dimensional configuration of centroids in more dimensions than necessary. Conversely, it may sometimes be interesting to take *fewer* components than the number of clusters minus 1. This is particularly useful in cases where the clusters nearly fall in a subspace of dimensionality lower than $(c-1)$, for instance, having four clusters (nearly) in a plane. Therefore, it is recommended to first choose the number of clusters – either on the basis of substantive information, or applying one of the procedures reviewed in the comparative study by Milligan and Cooper (1985), or by gradually increasing the number of clusters – and to take $m = c-1$. If it turns out that $\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{JXA}$ can be represented very well by fewer components (as can be seen, for instance, upon checking whether the last singular values are close to zero), then it is advised to rerun the analysis with fewer components. By thus verifying solutions for different numbers of clusters, one can select the solution that gives the best interpretable results.

A third remark on the factorial k-means model concerns a second indeterminacy of the model. As is readily verified, $||\boldsymbol{XA} - \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{XA}||^2$ does not change by replacing $\boldsymbol{A}$ by $\boldsymbol{AT}$, where $\boldsymbol{T}$ is an orthonormal rotation matrix. This property is a direct consequence of the fact that rotation of the projected configuration of the object points does not alter their distances to the centroids. In cases of two-dimensional solutions, this rotational indeterminacy is of little consequence, because it implies that any rotation of a plot of the configuration is just as good. In cases of three- and higher-dimensional solutions, visualizations are more difficult, and using the rotational indeterminacy to ensure that the configuration is located close to the axes (e.g., by Varimax rotation, Kaiser, 1958) may actually help in interpreting the solution.

To conclude this section, we mention the possibility for extending the model to search for more than one clustering in the same data. This may be relevant in cases where different directions in data space (or different subsets of variables) pertain to different clusterings of the objects. Suppose the first factorial k-means solution is given by the matrices $\boldsymbol{U}_1$ and $\boldsymbol{A}_1$, and a second clustering described mainly by those original variables that did not contribute much to identify $\boldsymbol{U}_1$, is present in the data.

Then such a second partition can be found by minimizing

$$||XA_2 - U_2(U_2'U_2)^{-1}U_2'XA_2||^2, \tag{6}$$

subject to $A_2'A_2 = I$ and $A_2'A_1 = 0$.

Possibly, further solutions can be found by minimizing $||XA_3 - U_3(U_3'U_3)^{-1}U_3'XA_3||^2$, subject to $A_3'A_3 = I$ and $A_3'A_1 = 0$, and $A_3'A_2 = 0$, and so on. Rather than proceeding in this successive manner, one might also find several solutions at once by minimizing

$$G(A_1,\dots,A_S,U_1,\dots,U_S) = \sum_{s=1}^{S} ||XA_s - U_s(U_s'U_s)^{-1}U_s'XA_s||^2 \tag{7}$$

subject to $A_s'A_s = I$ and $A_s'A_t = 0$, $(s \neq t)$. In the next section, it will be indicated how these problems can be solved.

5. An alternating least-squares algorithm

The constrained problem of minimizing (2) can be solved using an ALS algorithm, as follows:

Step 0: First, initial values are chosen for A, U, and $\bar{Y}$. Values for A and U can be chosen randomly, or in a rational way, and in both cases should satisfy the constraints on A and U. Initial values for $\bar{Y}$ are then given at once by $(U'U)^{-1}U'XA$.

Step 1: We minimize $F(A,U,\bar{Y}) = ||XA - U\bar{Y}||^2$ with respect to U, given the current estimate of A and $\bar{Y}$. This problem is solved for the different rows of U independently by taking $u_{ij} = 1$, if $F(A,[u_{ij}]) = \min\{F(A,[u_{iv}])\colon v = 1,\dots,m\}$ and $u_{ij} = 0$ otherwise.

Step 2: We update A and implicitly $\bar{Y}$ given U by minimizing (4) over A. This problem is solved by taking the first m eigenvectors of $X'(U(U'U)^{-1}U' - I_n)X$ (e.g., see Ten Berge, 1993). From the optimal A, we obtain $\bar{Y} = (U'U)^{-1}U'XA$, and thus have updated A and $\bar{Y}$ jointly.

Step 3: Finally, we compute the function value for the present values of A, U, and $\bar{Y}$. When the updates of A, U, and $\bar{Y}$ have decreased the function value F, we update A, U, and $\bar{Y}$ once more according to Steps 1 and 2. Otherwise, the process has converged.

The above algorithm monotonically decreases the loss function and, because this is bounded below, will converge to a solution which is at least a local optimum of (1). Because of the binary constraint on U, the method can be expected to be rather sensitive to local optima. Hence, it is recommended to either use many randomly started runs (and retain only the best solution) to decrease the chance of missing the global optimum, or to use a good rational start, obtained from, for instance, an earlier analysis of the data. In some test runs with data for which a perfect solution (i.e., with $F = 0$) was available, it turned out that using 100 random starts usually suffices. Considering that each run usually requires only few iterations, running 100 analyses can be done quite efficiently.

At the end of Section 5, two methods were described for fitting more than one clustering simultaneously to the data. In particular, it was first proposed to minimize

Table 2
Correlation between the first two dimensions of the factorial k-means analysis and the six variables

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
Dim 1	0.956	−0.287	−0.067	0.289	−0.171	−0.140
Dim 2	−0.308	−0.943	0.158	−0.008	−0.091	0.201

function (6) subject to $A_2'A_2 = I$ and $A_2'A_1 = 0$. This problem is very similar to that of minimizing (2), except for the constraint $A_2'A_1 = 0$. To implement this constraint, we write $A_2 = N_{A_1}B_2$, where N_{A_1} denotes a columnwise orthonormal basis for the null space of A_1'. Now we have the problem of minimizing

$$||XN_{A_1}B_2 - U_2(U_2'U_2)^{-1}U_2'XN_{A_1}B_2||^2, \tag{8}$$

subject to $B_2'N_{A_1}'N_{A_1}B_2 = B_2'B_2 = I$, which reduces to the ordinary factorial k-means problem with X replaced by XN_{A_1}, A by B_2, and U by U_2, and can hence be solved as before.

The problem of minimizing function (7) subject to $A_s'A_s = I$ and $A_s'A_t = 0$, $(s \neq t)$ is somewhat more involved, and will not be described in full detail here. The basic idea is that, again, an ALS algorithm is used, after writing the function as

$$G(A_1,\ldots,A_S,\ U_1,\ldots,U_S,\ \bar{Y}_1,\ldots,\bar{Y}_S) = \sum_{s=1}^{S} ||XA_s - U_s\bar{Y}_s||^2, \tag{9}$$

subject to $A_s'A_s = I$ and $A_s'A_t = 0, (s \neq t)$. Now, G is minimized alternately over U_s, $\bar{Y}_s$ and A_s, keeping the other matrices fixed. Minimizing G over U_s is done as in ordinary factorial k-means, but with A replaced by A_s and $\bar{Y}$ by $\bar{Y}_s$. Minimizing G over $\bar{Y}_s$ is done by taking $\bar{Y}_s = (U_s'U_s)^{-1}U_s'XA_s$. Minimizing G over $A_1,\ldots,A_S$ can now be done over all A_s simultaneously by writing $A = [A_1,\ldots,A_S]$, so that we have to minimize

$$\sum_{s=1}^{S} ||XA_s - U_s\bar{Y}_s||^2 = ||XA - [U_1\bar{Y}_1 \ldots U_S\bar{Y}_S]||^2 \tag{10}$$

subject to $A'A = I$. A solution for this problem can be found by means of the iterative method proposed by Green and Gower (1979) (also see Ten Berge, 1993).

6. Simulated data

The six-dimensional simulated data shown in Fig. 1, were classified using the factorial k-means model, fixing $c = 3$ and $m = 2$. In Table 2, the correlations between each factor defined by model 1 and the six original variables are shown.

Fig. 4 shows the classification of the 42 objects described by six variables and represented on the first two factors of the factorial k-means model. It can be noted that the random noise generated by the normal variables is drastically reduced by the factorial k-means analysis that, moreover, recovers the well-separated clustering structure, and correctly classifies all 42 objects.

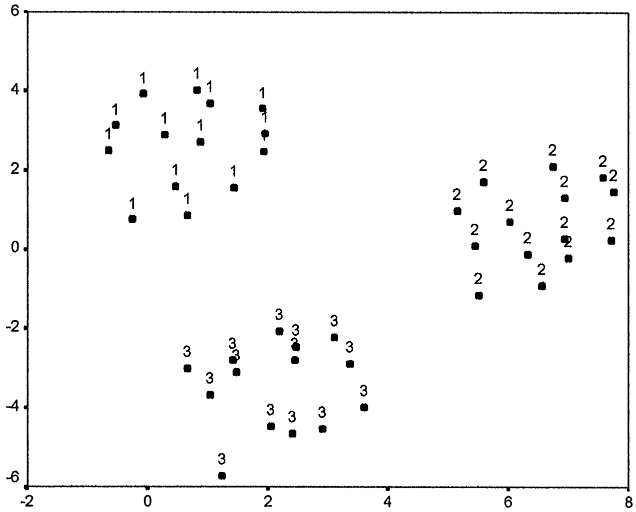

Fig. 4. Classification of 42 objects represented on the first two dimensions of the factorial k-means analysis.

7. Latest short-term macroeconomic scenario

The latest short-term scenario (September 1999) on macroeconomic performance of national economies of 20 countries, members of the OECD, has been considered to test the ability of the factorial k-means analysis in identifying classes of similar economies and help to understand the relationships within the set of observed economic indicators.

The performance of the economies reflects the interaction of six main economic indicators: gross domestic product (GDP), leading indicator (LI), unemployment rate (UR), interest rate (IR), trade balance (TB), net national savings (NNS) (Table 3).

The analysis started first with the classification of the countries. Before applying the k-means algorithm, the variables were standardized following the recommendations by Milligan and Cooper (1988). No significant correlation is observed between the six economic indicators. The results of k-means applied on the six standardized indicators are

First class	Australia, Canada, Finland, France, Spain, Sweden, United States
Second class	Greece, Mexico, Portugal
Third class	Austria, Belgium, Denmark, Germany, Italy, Japan, Netherlands, Norway, Switzerland, United Kingdom

The classification into three groups reflects a striking feature of economic developments in recent quarters, perceived by several economic institutions, that is, the emergence of divergent cyclical conditions across the countries especially within the euro area. Sizeable differences in growth are shown between the first and third

Table 3
Latest short-term indicators and economic performance indicators (percentage change from the previous year, September 1999)[a,b]

Country	Gross domestic product (GDP)	Leading indicator (LI)	Unemploy-ment rate (UR)	Interest rate (IR)	Trade balance (TB)	Net national savings (NNS)	Classmember-ship obtained by the factorial k-means model
Australia (A-lia)	4.8	8.4	8.1	5.32	0.70	4.70	1
Canada (Can)	3.2	2.5	8.4	5.02	1.60	5.20	1
Finland (Fin)	3.9	−1.0	11.8	3.60	8.80	7.70	1
France (Fra)	2.3	0.7	11.7	3.69	3.90	7.30	1
Spain (Spa)	3.6	2.5	19.0	4.83	1.20	9.60	1
Sweden (Swe)	4.1	1.1	8.9	4.20	7.00	4.00	1
United States (USA)	4.1	1.4	4.5	5.59	−1.40	7.00	1
Netherlands (Net)	2.9	1.6	4.2	3.69	7.00	15.80	1
Greece (Gre)	3.2	0.6	10.3	11.70	−8.30	8.00	2
Mexico (Mex)	2.3	5.6	3.2	20.99	0.00	12.70	2
Portugal (Por)	2.8	−7.5	4.9	4.84	−8.70	14.00	2
Austria (A-tria)	1.1	0.6	4.7	3.84	−0.60	9.40	3
Belgium (Bel)	1.4	−0.1	9.6	3.64	4.50	12.40	3
Denmark (Den)	1.0	1.5	5.3	4.08	3.30	5.00	3
Germany (Ger)	0.8	−2.0	9.5	3.74	1.50	7.70	3
Italy (Ita)	0.9	−0.4	12.3	6.08	4.30	8.20	3
Japan (Jap)	0.1	5.4	4.2	0.74	1.20	15.10	3
Norway (Nor)	1.4	0.9	3.3	4.47	7.10	15.10	3
Switzerland (Swi)	1.1	2.1	3.8	1.84	4.40	13.20	3
United Kingdom (UK)	1.2	4.9	6.4	7.70	−0.50	4.80	3

[a]Source: OECD, Paris (1999).
[b]Definitions and notes: GDP – Percentage change from previous year; seasonally adjusted except for Portugal; LI – A composite indicator based on other indicators of economic activity (qualitative opinions on production or employment, housing permits, financial or monetary series, etc.), which signals cyclical movements in industrial production from 6 to 9 months in advance; UR – percentage of civilian labor force; standardized unemployment rate; national definitions for Mexico, Switzerland; seasonally adjusted; IR – 3 months, except for Greece (12 months); TB – (goods and services) % of GDP at current price 1997; NNS – % of GDP at current prices 1997.

class, that represent fast and slow growth patterns. The center of the growth continues to be concentrated in the United States, while growth in the countries such as Portugal and Spain reflect underlying processes or structural convergence, where less advanced economies in euro area close in upon the more developed ones. There is also divergence between core countries of the monetary European Union: Germany and Italy on one side and the rest of euro area (particularly France) on the other. The long-standing structural regional problems in eastern Germany and southern Italy reduce the abilities of these economies to adapt to changing conditions. A growth below potential rates is also observed for the United Kingdom. Mexico and Greece, that are in the second group, are characterized by a high inflation, while Greece and Portugal have a very negative trade balance.

Table 4
Correlation between variables and two factors specified by the factorial k-means analysis[a]

	GDP	IR	LI	UR	NNS	TB
Dim 1	0.175	**−0.711**	**0.297**	0.260	**−0.329**	**0.717**
Dim 2	**0.956**	**0.309**	−0.003	0.245	−0.133	−0.184

[a]Contributions larger or equal than 0.3 in absolute values are highlighted.

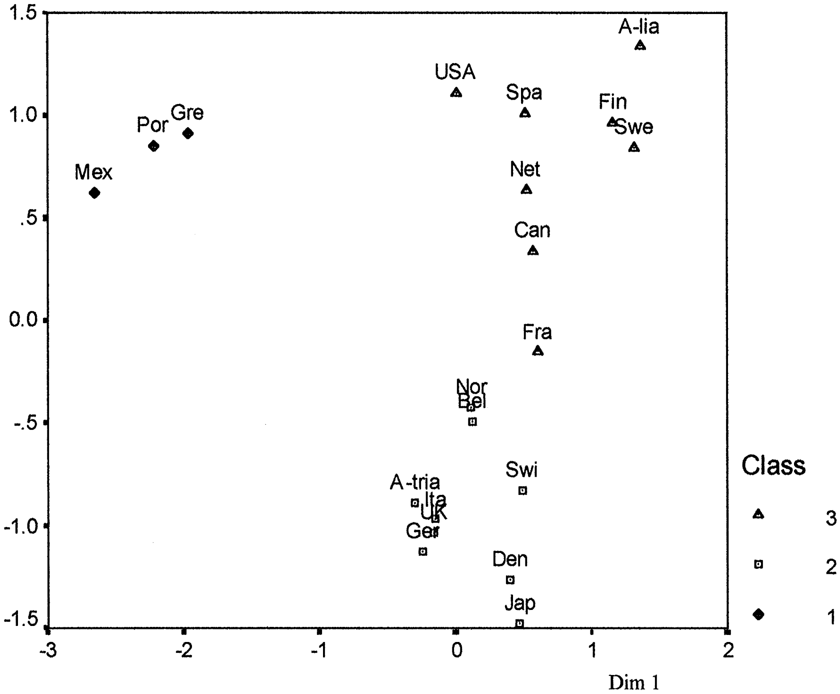

Fig. 5. Classification of the OECD countries given by the factorial k-means analysis, for $c = 3$ and $m = 2$.

The same classification (excluding the Netherlands that moves into class 1) was obtained by the factorial k-means analysis, fixing $c = 3$ and $m = 2$. The plot of the classification on the two dimensions defined by model 1 is shown in Fig. 5.

By examining the correlation between factors defined by model 1 and the six economic indicators, the researcher has the advantage to detect the original variables that most contribute to determine the classification. The first dimension is characterized mainly by interest rate, and the trade balance, but also by net national savings and leading indicator, whereas the second dimension is influenced almost exclusively by gross domestic product and interest rate. The unemployment rate is moderately correlated with both dimensions (Table 4).

A successive *tandem analysis* was carried out by computing the first two principal components and classifying countries on the basis of first two objects scores. The results are shown in Fig. 6. This classification differs for four countries with respect to the classification with all six variables: Italy, UK, Netherlands and Portugal.

Table 5
Correlation between variables and the first two principal components[a]

	GDP	IR	LI	UR	NNS	TB
Dim 1	**0.749**	0.232	0.249	**0.648**	**−0.792**	−0.081
Dim 2	−0.077	**−0.834**	−0.276	**0.442**	−0.112	**0.674**

[a]Contributions larger than 0.3 in absolute values are highlighted.

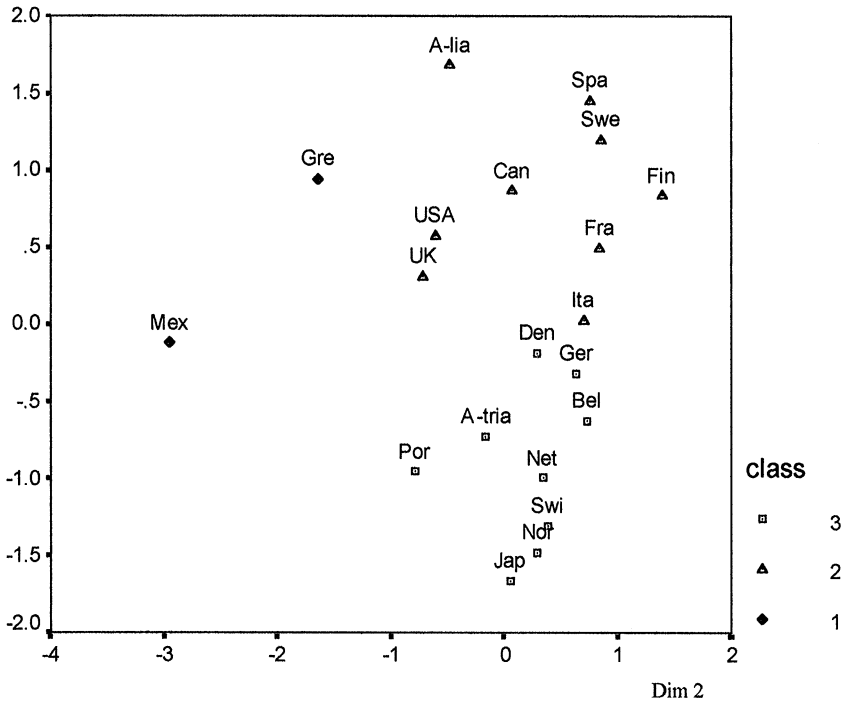

Fig. 6. Tandem analysis. k-means classification computed on the first two principal components.

The first PCA dimension is characterized mainly by net national savings, gross domestic product, whereas the second PCA dimension by interest rate and trade balance. The unemployment rate characterizes both dimensions (Table 5).

Comparing the two graphical representations in Figs. 5 and 6, it can be observed that the factorial k-means analysis, more clearly shows the three classes detected by the classification on the six variables than the tandem analysis. In Fig. 5, the Netherlands is positioned close to Canada and this is explained by their similar growth pattern (GDP: 2.9, 3.2, respectively). By contrast, in Fig. 6, the Netherlands is located close to Switzerland but these have very different growth pattern (GDP: 2.9, 1.1, respectively). Italy, Germany, Austria and United Kingdom, which have similar development (GDP: 0.9, 0.8, 1.1, 1.2, respectively), using factorial k-means analysis, are located (Fig. 5) close to each other, whereas this is not so evident with tandem analysis (Fig. 6).

8. Conclusions

In this paper factorial k-means analysis has been introduced. It is a discrete and continuous model that simultaneously classifies objects and finds a subset of factors that best describe the classification. The methodology allows one to select the most relevant components for the classification. A representation in a reduced number of dimensions can be given to help the interpretation of relationships within the set of variables and the set of objects. The least-squares fit of the factorial k-means model is mathematically formalized as a mixed quadratic constrained problem. An ALS algorithm is proposed to efficiently solve the quadratic constrained problem.

This methodology is a valid alternative to the frequently used tandem analysis. With respect to this last sequential approach of factorial and classification analysis, the factorial k-means analysis identifies main factors and the partition of the objects simultaneously by minimizing a single objective function. By contrast, tandem analysis produces factors and an optimal classification, thus minimizing two different objective functions that, as shown in this paper, may work in contradiction, identifying factors that do not contribute much to perceive the clustering structure, and that conversely, may in part obscure or mask it. With factorial k-means analysis, we do run a risk as well. Especially, in case of data containing dimensions with very low variance, factorial k-means may focus on these dimensions primarily because these will contribute little to the loss function (2). Therefore, after a factorial k-means, it is particularly important to check whether the obtained clustering indeed makes sense in terms of the full data set. To avoid this problem, one could, for instance, first eliminate such trivial dimensions from the data.

Factorial k-means analysis has a fast ALS algorithm that allows its application also with large data sets. The methodology can, therefore, be recommended as an alternative to the widely used tandem analysis.

References

Arabie, P., Hubert, L., 1994. Cluster analysis in marketing research. In: Bagozzi, R.P. (Ed.), Handbook of marketing research. Blackwell, Oxford.

Ball, G.H., Hall, D.J., 1967. A clustering technique for summarizing multivariate data. Behav. Sci. 12, 153–155.

DeSarbo, W.S., Jedidi, K., Cool, K., Schendel, D., 1990. Simultaneous multidimensional unfolding and cluster analysis: an investigation of strategic groups. Marketing Lett. 2, 129–146.

DeSarbo, W.S., Howard, D.J., Jedidi, K., 1991. MULTICLUS: a new method for simultaneous performing multidimensional scaling and clustering. Psychometrika 56, 121–136.

De Soete, G., Carroll, J.D., 1994. k-means clustering in a low-dimensional Euclidean space. In: Diday, E., et al. (Eds.), New Approaches in Classification and Data Analysis. Springer, Heidelberg, pp. 212–219.

De Soete, G., Heiser, W.J., 1993. A latent class unfolding model for analyzing single stimulus preference ratings. Psychometrika 58, 545–565.

Diday, E., et al., 1979. Optimisation en classification automatique. INRIA, Vol. 1. Le Chesnay, France.

Gordon, A.D., 1999. Classification, 2nd Edition. Chapman & Hall, London.

Green, B.F., Gower, J.C., 1979. A problem with congruence. Paper presented at the Annual Meeting of Psychometric Society. Monterey, USA.

Heiser, W.J., 1993. Clustering in low-dimensional space. In: Opitz, O., Lausen, B., Klar, R. (Eds.), Information and Classification. Springer, Berlin, pp. 162–173.

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

Kiers, H.A.L., 1997. Discrimination by means of components that are orthogonal in the data space. J. Chemometrics 11, 533–545.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Le Cam, L.M., Neyman, J. (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Statistics. University of California Press, Berkeley, CA, pp. 281–297.

Milligan, G.W., Cooper, M., 1985. An extimation of procedures for determining the number of clusters in a data set. Psychometrika 50, 159–179.

Milligan, G.W., Cooper, M.C., 1988. A study of standardization of variables in cluster analysis. J. Classification 5, 181–204.

Ten Berge, J.M.F., 1993. Least Squares Optimization. DSWO Press, Leiden.