ELSEVIER Computational Statistics & Data Analysis 35 (2001) 417–428

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

A generalization of principal component analysis to K sets of variables

Ph. Casin

Faculté de Droit et d'Economie, Ile du Saulcy, 57005 METZ, Cedex 1, France

Received 1 November 1999; accepted 1 May 2000

Abstract

The aim of this paper is to introduce a new method, generalized principal component analysis (GPCA), which is a generalization of principal component analysis (PCA), to several data tables. GPCA is a method for both finding common dimensions in several sets of variables and giving a description of each set of variables: GPCA takes into account both the correlation structure within sets and relationships between sets. Two sorts of orthogonal basis are provided; the first basis is useful to represent each set of variables (as PCA does) and the second is useful to represent associations between sets of variables (as canonical correlation analysis does). An example using real data (evolution of five characteristics of car markets from 86 to 93 for 8 countries) illustrates the method.

Keywords: Generalized canonical analysis; Principal component analysis; Canonical correlation analysis

1. Introduction

Canonical correlation is a technique for analyzing linear relationships between two sets of variables due to Hotelling (1936). Many generalizations of Hotelling's analysis to three or more sets of variables have been proposed (Horst, 1961; Mc Keon, 1966; Carroll, 1968; Kettenring, 1971; Saporta, 1975; Van de Geer, 1984; Kiers et al., 1994).

But canonical correlation analysis capitalizes exclusively on relationships between sets and ignores the correlation structure within sets. Principal component analysis (PCA) is a technique which describes the correlation structure, but for only one set of variables.

0167-9473/01/$ - see front matter
PII: S 0167-9473(00)00024-4

The aim of this paper is to introduce a generalization of PCA to several data tables, generalized principal component analysis (GPCA), which takes into account both correlation structure within sets and relationships between sets (Casin, 1996).

2. Notation

The rows of K data tables refer to the same set of individuals, but their columns are associated to different sets of variables indexed by k; let X_k $(k = 1,\ldots,K)$ be the $n \times m_k$ matrix of the kth set of m_k variables on the n individuals and let $X = [X_1,\ldots,X_k,\ldots,X_K]$ be an $n \times m$ matrix, where $m = \sum_{k=1}^{K} m_k$.

Each variable $X_{k,j}$ (the jth column of X_k) has zero mean and unit-variance; the rank of X is assumed to be m. These latter assumptions are made mainly in order to keep the notation simple. W_k is the subspace of R^n spanned by the columns of X_k.

$R(x, y)$ and $\mathrm{Cov}(x, y)$ are, respectively, the correlation coefficient and the covariance between two variables x and y. $\mathrm{Var}(x)$ is the variance of the variable x.

3. Generalized principal component analysis

3.1. The aim of the method

The aim of the method is both finding common dimensions in K sets of variables and describing each set of variables. GPCA provides components in different sets of variables, which are mutually uncorrelated within these sets, and which are related to components of the other sets.

3.2. The method

3.2.1. The first step

At the first step, GPCA is the first step of a PCA on the supermatrix X. As a matter of fact, PCA finds an auxiliary variable z^1 such that $\sum_{k=1}^{K}\sum_{j=1}^{m_k} R^2(z^1, X_{k,j})$ is maximized over z^1, under the constraint $\mathrm{Var}(z^1) = 1$. Let z_k^1 be the projection of z^1 onto W_k; then $R^2(z^1, X_{k,j}) = \mathrm{Cov}^2(z^1, X_{k,j}) = \mathrm{Cov}^2(z_k^1, X_{k,j})$ because $z^1 - z_k^1$ is orthogonal to $X_{k,j}$. Consequently,

$$R^2(z^1, X_{k,j}) = \mathrm{Var}(z_k^1) R^2(z_k^1, X_{k,j}),$$

and, because $\mathrm{Var}(z_k^1) = R^2(z^1, z_k^1)$:

$$\sum_{k=1}^{K}\sum_{j=1}^{m_k} R^2(z^1, X_{k,j}) = \sum_{k=1}^{K}\left(R^2(z^1, z_k^1)\sum_{j=1}^{m_k} R^2(z_k^1, X_{k,j})\right).$$

So, PCA can be seen as a method to optimize a compromise between two conflicting objectives:

(1) Maximizing $R^2(z^1, z_k^1)$ over z^1 for $k = 1,\ldots,K$, in other words, maximizing correlations between z^1 and its projections onto spaces W_k. The higher the $R^2(z^1, z_k^1)$

for $k = 1,\ldots,K$ are, the closer the components z_k^1 are to each other; the first objective is finding common variables in the K sets.

(2) Maximizing $\sum_{j=1}^{m_k} R^2(z_k^1, X_{k,j})$ for $k=1,\ldots,K$ over z^1. $\sum_{j=1}^{m_k} R^2(z_k^1, X_{k,j})$ is the variance of the variables in the kth group explained by z_k^1; consequently, the second objective is optimally describing each of the K sets of variables.

3.2.2. The other steps

The search continues beyond the first step.

In order to provide a decomposition of the variance of the variables in the different sets, GPCA computes an orthogonal basis of each subspace W_k. Consequently, at the second step, the constraints are: for $k=1,\ldots,K$ $R(z_k^1, z_k^2)=0$, where z_k^2 is the second component of the kth data set.

In other words, for a particular value of k, z_k^2 is a linear combination of variables $X_{k,j}^2$ (for $j=1,\ldots,m_k$), where $X_{k,j}^2$ is the residual of the regression of $X_{k,j}$ on z_k^1.

Let X_k^2 be the data table whose columns are variables $X_{k,j}^2$ (for $j=1,\ldots,m_k$), W_k^2 be the subspace of W_k spanned by variables $X_{k,j}^2$ (for $j=1,\ldots,m_k$), and X^2 be the supermatrix $[X_1^2,\ldots,X_k^2,\ldots,X_K^2]$.

Let z^2 be the second auxiliary of GPCA; z^2 is the first principal component (PC) issued from a PCA (using a covariance with z^2. matrix) on the supermatrix X^2, and z_k^2 is the variable of W_k^2 which has the highest correlation coefficient.

As a matter of fact, PCA on the supermatrix X^2 finds an auxiliary variable z^2 such that $\sum_{k=1}^{K}\sum_{j=1}^{m_k} \mathrm{Cov}^2(z^2, X_{k,j}^2)$ is maximized over z^2 under the constraint $\mathrm{Var}(z^2)=1$. Then

$$\begin{aligned}
\mathrm{Cov}^2(z^2, X_{k,j}^2) &= \mathrm{Cov}^2(z_k^2, X_{k,j}^2) \quad \text{because } z^2 - z_k^2 \text{ is orthogonal to } X_{k,j}^2 \\
&= \mathrm{Cov}^2(z_k^2, X_{k,j}) \quad \text{because } X_{k,j} - X_{k,j}^2 \text{ is orthogonal to } W_k^2 \text{ and} \\
&\qquad\qquad\qquad\qquad\quad z_k^2 \text{ belongs to } W_k^2 \\
&= \mathrm{Var}(z_k^2) R^2(z_k^2, X_{k,j}) \\
&= R^2(z^2, z_k^2) R^2(z_k^2, X_{k,j})
\end{aligned}$$

and consequently

$$\sum_{k=1}^{K}\sum_{j=1}^{m_k} \mathrm{Cov}^2(z^2, X_{k,j}^2) = \sum_{k=1}^{K} \left(R^2(z^2, z_k^2) \sum_{j=1}^{m_k} R^2(z_k^2, X_{k,j}) \right).$$

So, in the second step, GPCA optimizes a compromise between two conflicting objectives:

(1) Maximizing $R^2(z^2, z_k^2)$ over z^2, for $k = 1,\ldots,K$; the first objective is finding common variables in the K sets, under the constraint that z_k^2 must be orthogonal to the first component of the set, z_k^1: the higher the $R^2(z^2, z_k^2)$ are, the closer the components z_k^2 are to each other.

(2) Maximizing $\sum_{j=1}^{m_k} R^2(z_k^2, X_{k,j})$ for $k=1,\ldots,K$ over z^2; $\sum_{j=1}^{m_k} R^2(z_k^2, X_{k,j})$ is the variance of the variables in the kth group explained by z_k^2; the second objective is optimally describing the variance which has not been explained in the first step,

under the constraint that z_k^2 must be orthogonal to the first component of the set, z_k^1.

Let us now consider the rth stage, in order to provide an orthogonal basis of each subspace $W_k \cdot z_k^r$, the rth component of the k-set of variables, is orthogonal to previous components of this set. In other words, z_k^r is a linear combination of variables $X_{k,j}^r$ (for $j = 1,\ldots,m_k$), where $X_{k,j}^r$ is the residual of the regression of $X_{k,j}$ on $z_k^1,\ldots,z_k^{r-1}$.

Let X_k^r be the data table whose columns are variables $X_{k,j}^r$ (for $j = 1,\ldots,m_k$), W_k^r be the subspace of W_k spanned by variables $X_{k,j}^r$ (for $j = 1,\ldots,m_k$), and X^r be the supermatrix $[X_1^r,\ldots,X_k^r,\ldots,X_K^r]$.

At the rth step, GPCA is the first step of a PCA (using a covariance matrix) of the supermatrix X^r, and finds an auxiliary variable z^r such that $\sum_{k=1}^K \sum_{j=1}^{m_k} \mathrm{Cov}^2(z^r, X_{k,j}^r)$ is maximized over z^r, under the constraint $\mathrm{Var}(z^r) = 1$. z_k^r is the projection of z^r onto W_k^r.

And, at the rth step, GPCA optimizes a compromise between two conflicting objectives, finding common dimensions in the K sets of variables, and describing each set of variables: the relationship between z^r and the kth group of variables is measured by $R^2(z^r, z_k^r) \sum_{j=1}^{m_k} R^2(z_k^r, X_{k,j})$.

4. GPCA's properties

Property 1. *If $K = 1$, GPCA reduces to PCA using a correlation matrix.*

Proof. If $K = 1$, at the rth step, the problem to be solved is maximizing $\sum_{k=1}^K \sum_{j=1}^{m_k} \mathrm{Cov}^2(z^r, X_{k,j}^r)$ under the constraints: for $s = 1,\ldots,r-1$, $R(z^r, z^s) = 0$.

Because of the constraints

$$\sum_{k=1}^{K} \sum_{j=1}^{m_k} \mathrm{Cov}^2(z^r, X_{k,j}^r) = \sum_{k=1}^{K} \sum_{j=1}^{m_k} \mathrm{Cov}^2(z^r, X_{k,j}) = \sum_{k=1}^{K} \sum_{j=1}^{m_k} R^2(z^r, X_{k,j}).$$

This problem is solved by PCA using a correlation matrix. This property justifies the name "generalized principal component analysis".

Property 2. *For $r \neq s$, $R(z^r, z^s) = 0$.*

Proof. Let us suppose that $r > s$, $X_{k,j}^r$ is orthogonal to z_k^s; on the other hand, because z_k^s is the projection of z^s onto the space spanned by the columns of X_k^s, $z^s = z_k^s + e_k^s$, where e_k^s is orthogonal to the columns of X_k^s and consequently e_k^s is orthogonal to $X_{k,j}^r$. Then

$$(X_{k,j}^r)' z^s = (X_{k,j}^r)' z_k^s + (X_{k,j}^r)' e_k^s = 0.$$

z^r is a linear combination of the variables $X_{k,j}^r$, then z^r is orthogonal to z^s.

This property is very interesting; it means that GPCA provides an orthogonal basis constituted by auxiliary variables. This basis is useful to describe correlations of all

original variables between sets, or in other words, is useful to represent an association between sets of variables, as CCA does.

Property 3. *If* $r > s$, *then* $R(z_k^r, z^s) = 0$ *for* $k = 1, \ldots, K$

Proof. From the proof of Property 2, z^s is orthogonal to variables $X_{k,j}^r$; for a particular value of k, since z_k^r is the projection of z^r onto the space spanned by the variables $X_{k,j}^r$, z_k^r is a linear combination of variables $X_{k,j}^r$, for $j = 1, \ldots, m_k$, then z_k^r is orthogonal to z^s.

5. Plots based on canonical variables and auxiliary variables

GPCA provides two different types of graphical representations. These graphical representations are used to represent the components z_k^r which explain a substantial amount of the variance of the original variables $X_{k,j}$ of the kth set of variables.

For these components, one of the two different types of graphical representations is choosen according to values of $R^2(z^r, z_k^r)$. The first of these two graphical representations answers the question regarding how variables from different sets are related and the second one answers the question regarding how variables from the same set are related.

When $R^2(z^r, z_k^r)$ is nearly unity, for some or for all values of k, two plots are useful to represent the association between the sets of variables:

(1) a plot of the original variables (correlations of original variables $X_{k,j}$ with auxiliary variables);
(2) a plot of the individuals (scores of individuals for variables z_k^r and z^r).

These plots are close to those used with canonical analysis and its extensions.

When $R^2(z^r, z_k^r)$ is not nearly unity for a particular value of k, two plots can be drawn to represent relations within the kth data table: GPCA computes an orthonormal basis of the subspace spanned by the columns of X_k. Therefore, a plot of the variables $X_{k,j}$, for $j = 1, \ldots, m_k$ and a plot of the individuals (scores of individuals for variables z_k^r) can be drawn and PCA rules of interpretation can be used (Jolliffe, 1986).

6. An application of GPCA

GPCA allows for different variables in the K sets of variables, but an interesting particular case, especially in economics, deals with the same variables indexed by time; then, the aim of GPCA is to study the stability or the change of the relationships between variables, and to study the stability or the change of the proximities between individuals.

Here, each of $K = 9$ data tables (years from 86 to 94) describes the rate of growth of the $m = 5$ ($m_k = m$, for $k = 1, \ldots, 9$) following variables:

Table 1
Results of GPCA

	86	87	88	89	90	91	92	93	94
CC1	0.68	0.91	0.95	0.53	0.99	0.99	0.97	0.95	0.97
PCAC1	0.91	2.83	2.97	1.52	0.63	1.08	1.67	2.33	1.51
CC2	0.97	0.75	0.92	0.97	0.98	0.94	0.78	0.93	0.91
PCAC2	0.78	0.56	0.90	0.83	2.81	1.91	1.55	1.19	1.27
CC3	0.99	0.37	0.57	0.78	0.55	0.86	0.86	0.93	0.51
PCAC3	1.68	0.90	0.17	1.13	0.91	0.79	0.39	0.83	1.09
CC4	0.46	0.64	0.08	0.97	0.73	0.93	0.97	0.84	0.44
PCAC4	0.45	0.45	0.50	1.23	0.38	0.63	0.62	0.27	0.86
CC5	0.80	0.70	0.10	0.57	0.44	0.73	0.11	0.61	0.34
PCAC5	0.17	0.25	0.46	0.28	0.27	0.59	0.76	0.39	0.27

R1: number of registrations of private cars,
R2: number of registrations of commercial cars,
R3: number of registrations of private cars with diesel engine,
R4: production of private cars,
R5: production of commercial cars, for $n = 8$ countries: Germany, Belgium, France, United Kingdom, Italy, Portugal, The Netherlands, Spain.

The results of GPCA are as follows (see Table 1): $R^2(z^r, z_k^r)$ is denoted by CCr (as canonical criterion at the rth step) and $\sum_{j=1}^{m_k} R^2(z_k^r, X_{k,j})$ is denoted by PCAr (as PCA criterion at the rth step).

In the first step, all data tables have high values for PCAC (except "90") and for CC (except "86" and "89"); in the second step, all data tables, except "86", "87" and "89" have a high value for PCAC and all, except "87" and "92", have a high value for CC. Consequently, plots representing the association between sets of variables are used.

For the first two steps, Fig. 1 describes the relations between all variables (axes are the variables z^1 and z^2):

Variables which are highly correlated with the first two auxiliary variables come from sets which have high values for both CCAC and PCAC : 87 (R3, R4 and R5), 88 (R1, R2 and R3), 91 (R4), 92 (R1), 93 (R2, R3 and R5) and 94 (R2 and R4) for the first axis, and 88 (R4), 90 (R1, R2, R4 and R5), 91(R1 and R2), 92 (R2 and R5) and 93 (R4) for the second axis.

It is worth noting that the variables R2 and R4 are well described for most years by the two first axes.

For the first two steps, Fig. 2 describes trajectories of individuals:

Points for the countries are based on z^1 and z^2, whereas those for countries in years are based on z_k^1 and z_k^2. All points are not based on coordinates with respect to

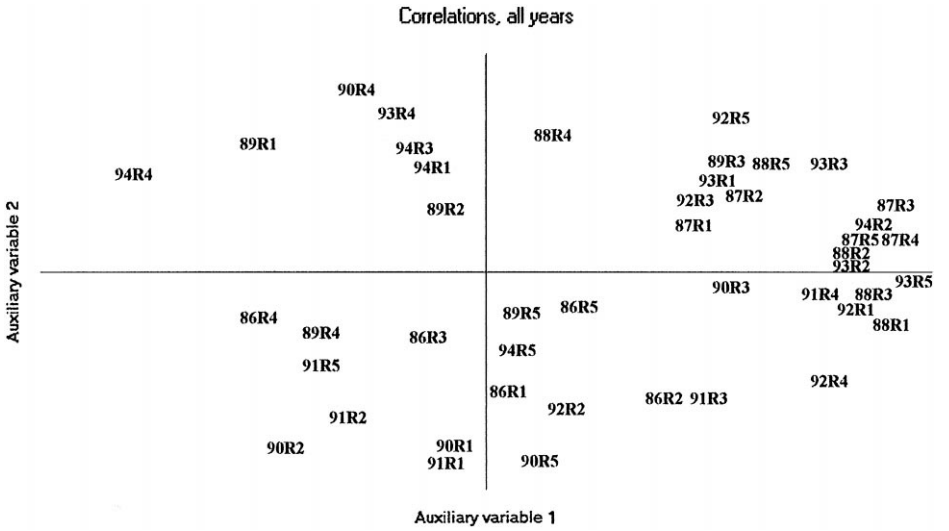

Fig. 1. Plot of the variables, for all years.

the same axes, but z^1 and z_k^1 are very close to each other in most years and similarly for z^2 and z_k^2 and superimposing these points is nice as a way of visualization.

N is the letter used for The Netherlands. N*xx* is the value for this country in the year *xx*: for instance, N92 is the value in 92 for The Netherlands. S is used for Spain, G for Germany, P for Portugal, U for United-Kingdom and F for France.

Now turning briefly to the position of the observations, the point near the right-hand side (Portugal) corresponds to high values for R2 (88, 93 and 94) and R4 (in 87 and 91) and a low value for R4 in 94, whereas the points near the left-hand side (Italy, The Netherlands) correspond to low values for R2 (88, 93 and 94) and R4 (in 87 and 91) and a high value for R4 in 94.

The second axis contrasts UK (low values for R2 in 90, 91 and 92, high values for R4 in 88, 90 and 93) with Germany (high values for R2 in 90, 91 and 92, low values for R4 in 88, 90 and 93).

In the third step, only "86" have a high value for both PCAC and CC; in the fourth step, only "89" has a high value for PCAC whereas, in the fifth step, only "86" has a high value for PCAC. For these steps and for components which have a high value for PCAC, plots to describe relations within sets can be drawn.

Let us now consider results of the first data table (year 86), for example. This set of variables is not very well described by the two first steps: values of PCAC are 0.91 for the first step, 0.78 for the second step; components of the third and the fifth steps have high values (1.68 and 1.17, respectively) for PCAC, whereas the component of the fourth step explains a low amount of variance (PCAC equals 0.45).

Fig. 3 shows the correlations for the third and the fifth steps, and the first data table, (axes are the components z_3^3 and z_3^5).

Rules of interpretation of Fig. 4 are PCA rules: the third component contrasts Germany (high value for R3 in 86) with Spain (low value for R3 in 86), whereas

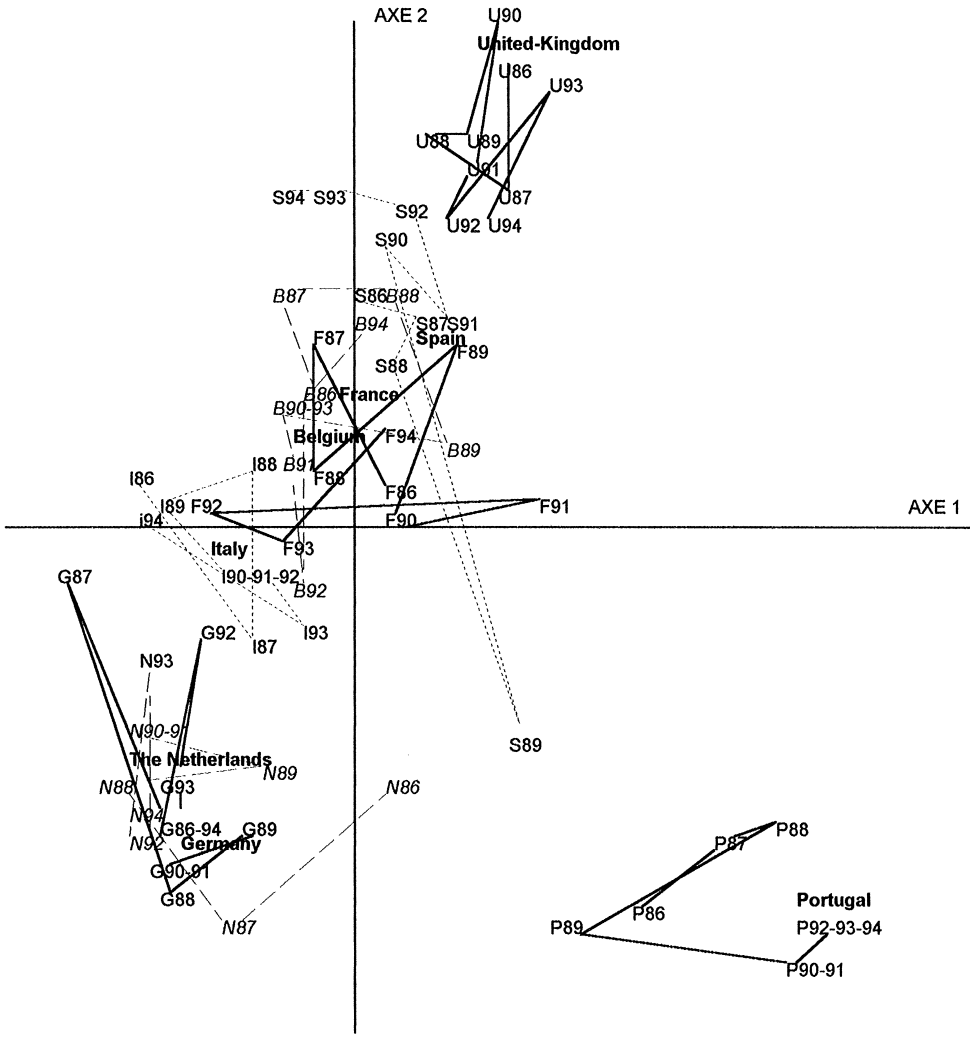

Fig. 2. Representation of individuals, all years.

the fifth component contrasts Spain (high value for R1 in 86) with Italy (low value for R1 in 86).

7. Comparison with some other methods

In this section, GPCA is compared with previous techniques to describe linear relations among K sets of variables, first with Carroll's Generalized canonical analysis, which is the most famous generalization of canonical analysis, and second with techniques based on PCA.

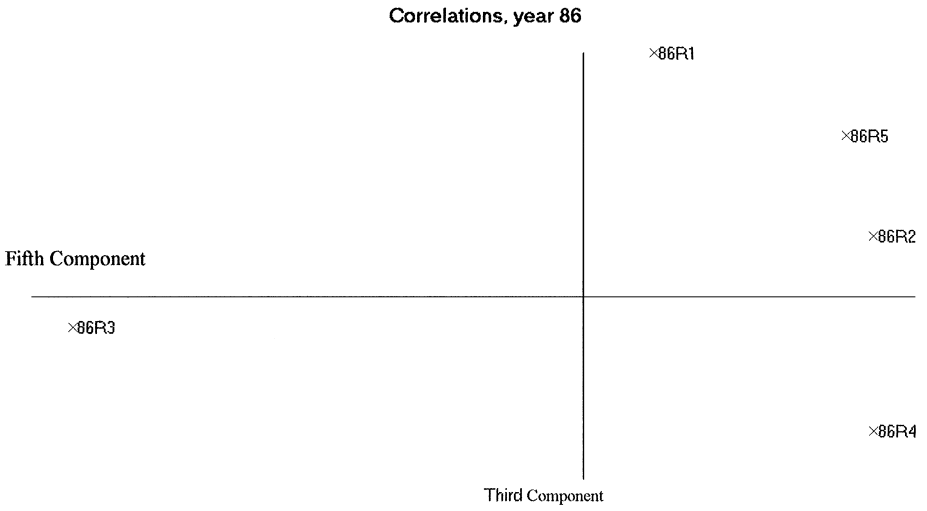

Fig. 3. Correlations, year 86.

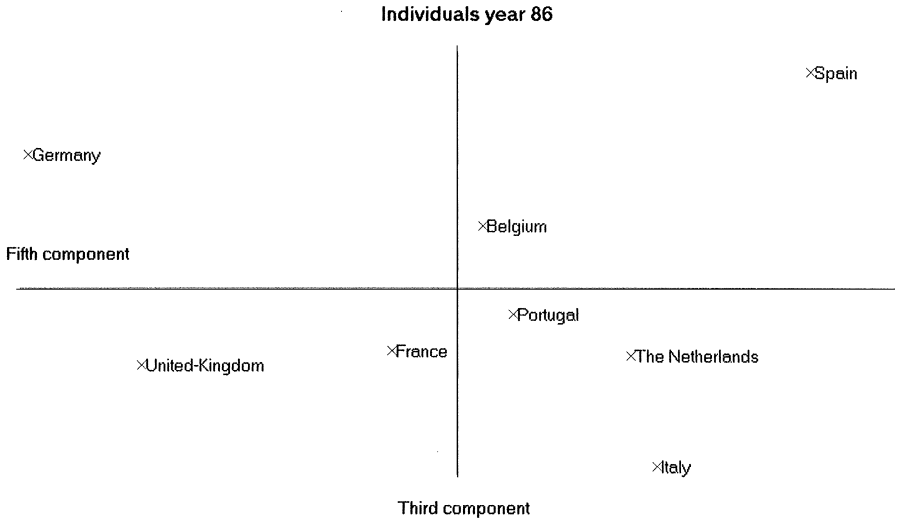

Fig. 4. Representation of individuals, year 86.

7.1. Comparison with Carroll's generalized canonical analysis (GCA)

In the first step, the problem to be solved by Carroll's (1968) GCA is maximizing $\sum_{k=1}^{K} R^2(z^1, z_k^1)$ over z^1 and z_k^1; in the rth step, the problem to be solved is maximizing $\sum_{k=1}^{K} R^2(z^r, z_k^r)$ over z^r and z_k^r, under the constraints: for $s = 1, \ldots, r-1$ $R(z^r, z^s) = 0$.

In GCA, the fact that z_k^r is the projection of z^r onto W_k is a consequence of the optimality of the optimization criterion used, whereas in GPCA the fact that z_k^r is the projection of z^r onto W_k^r is imposed as a constraint. In GCA, the relationship between z^r and the kth group of variables is measured by the square of the multiple

correlation coefficient $R^2(z^r, z_k^r)$ whereas in GPCA this relationship is measured by $R^2(z^r, z_k^r) \sum_{j=1}^{m_k} R^2(z_k^r, X_{k,j})$.

The main difference between GCA and GPCA is that GCA only focuses on relations between sets, and consequently does not take into account relationships between variables of the same group, and does not provide a basis of each subspace W_k.

7.2. Comparison with PCA

Like GPCA, PCA on the supermatrix X focuses on relations between sets as well as relations within sets, but with GPCA far fewer components are required to describe relations within the kth set of variables: for a particular value of k, PCA provides a decomposition of the variance of the variable $X_{j,k}$ into m auxiliary variables, whereas with GPCA, this decomposition needs only m_k components.

7.3. Comparison with multiple factor analysis (MFA)

MFA (Escofier and Pagès, 1994) consists of a PCA of the supermatrix $[\beta_1 X_1, \ldots, \beta_k X_k, \ldots, \beta_K X_K]$, where β_k is the inverse of the root of the first eigenvalue of PCA of the data table X_k: the groups of variables are normalized such that their first principal components have a variance equal to 1.

Like PCA on the supermatrix X, MFA takes into account variance within sets of variables, but does not provide a basis of each subspace W_k; consequently, with GPCA far fewer components are required to describe relations within the kth set of variables, for a given value of k.

7.4. Comparison with PCASUP

Let us suppose that the number of variables is the same in each set, say M; PCASUP (Kiers, 1991) is a PCA on the data table $[\text{Vect}(X_1), \ldots, \text{Vect}(X_K)]$ where $\text{Vect}(X_k)$ is the matrix X_k, strung out row wise into a column vector of order nM.

This PCA does not compute correlations between variables, and consequently does not describe relations between original variables $X_{j,k}$.

7.5. Comparison with MAXBET

MAXBET (Van de Geer, 1984) uses PCA's criterion on the supermatrix X, but with other constraints. Let $\tilde{z}_k^r$ be the rth component of the kth data table, then $\tilde{z}_k^r = X_k t_k^r$, where t_k^r is a column vector with m_k rows, and the constraint is that the columns t_k^r $r = 1, 2, \ldots$, form an orthonormal basis, $k = 1, \ldots, K$. Thus, MAXBET is a generalization of PCA with respect to vectors t_k^r. MAXBET does not provide an uncorrelated basis of each subspace W_k, and consequently does not provide a decomposition of the variance of the variables within each set in terms of uncorrelated components.

7.6. Comparison with the "Analyse de la co-inertie" method (ACOM)

Let the Euclidean metric associated with the subject subspace $\Re^{m_k}$ be defined by the positive definite matrix M_k, and π_k be a positive weight associated with the kth set of variables.

In the first step, Chessel and Hanafi's (1996) ACOM consists of maximizing $\sum_{k=1}^{K} \pi_k((v^1)'X_kM_ku_k^1)^2$ over v^1 and u_k^1.

In the rth step, the problem solved by ACOM is maximizing $\sum_{k=1}^{K} \pi_k((v^r)'X_kM_ku_k^r)^2$ over v^r and u_k^r, under the constraints: for $s=1,\ldots,r-1$ $(v^r)'v^s=0$ and for $s=1,\ldots,r-1$ $(u_k^r)'M_ku_k^s = 0$ $(k = 1,\ldots,K)$.

ACOM is a generalization of PCA in the subject space $\Re^m$ whereas GPCA is a generalization in the variable space; the two generalizations are not equivalent: ACOM computes an orthonormal basis of each subject subspace $\Re^{m_k}$, and does not compute an orthonormal basis of each variable subspace W_k. Consequently, ACOM does not provide a decomposition of the variance of the variables within each set in terms of uncorrelated components.

8. GPCA'software

GPCA's software will be available from CISIA (261, rue de Paris, 93556 Montreuil Cedex, France) in October 2000, as a part of the SPAD-TM package.

9. Conclusion

GPCA generalizes PCA to more than one set of variables; this new method both finds common dimensions in several sets of variables and gives a description of each set of variables. GPCA provides useful plots in order to represent the association between the sets of variables or to represent the specificities of each set of variables.

References

Carroll, J.D., 1968. Generalization of canonical correlation analysis to three or more sets of variables. Proceedings of the 76th Convention of the American Psychology Association, Vol. 3. 227–228.

Casin, Ph., 1996. L'analyse en composantes principales généralisée. Rev. Statist. Appl. 3, 63–81.

Chessel, D., Hanafi, M., 1996. Analyse de la co-inertie de K nuages de points. Rev. Statist. Appl. 19, 35–60.

Escofier, B., Pagès, J., 1994. Multiple factor analysis (AFMULT package). Comput. Statist. Data Anal. 18, 121–140.

Horst, P., 1961. Relation among m sets of variables. Psychometrica 2, 129–149.

Hotelling, H., 1936. Relation between two sets of variables. Biometrika 28, 321–377.

Jolliffe, I.T., 1986. Principal Component Analysis. Springer, New York.

Kiers, 1991. Hierarchical relations among three-way methods. Psychometrika 56 (3), 449–470.

Kiers, H.A.L., Cleroux, R., TEN BERGE, J.M.F., 1994. Generalized canonical analysis based on optimizing matrix correlations and a relation with IDIOSCAL. Comput. Statist. Data Anal. 18, 331–340.

Kettenring, R.J., 1971. Canonical analysis of several sets of variables. Biometrica 3, 433–450.
Mc Keon, J.J., 1966. Canonical analysis: Some relations between canonical correlation, factor analysis, discriminant function analysis, and scaling theory. Psychometric Monograph 13.
Saporta, G., 1975. Liaisons entre plusieurs ensembles de variables et codage de données qualitatives. Thèse de 3ème cycle, Université de Paris VI.
Van de Geer, J., 1984. Linear relations among k sets of variables. Psychometrica 1, 79–94.