Computational Statistics & Data Analysis 18 (1994) 97–119
North-Holland

The ACT (STATIS method)

Christine Lavit, Yves Escoufier
Unité de Biométrie, ENSA.M-INRA-UM II, Montpellier, France

Robert Sabatier
Faculté de Pharmacie, Département Pharmaceutique de Physique Industrielle, Statistique et Informatique, UM I, Montpellier, France

Pierre Traissac
ORSTOM, Montpellier, France

Abstract: ACT (STATIS method) is a data analysis technique which computes Euclidean distances between configurations of the same observations obtained in K different circumstances, and thus handles three-way data as a set of K matrices. In this article, the recent developments of the ACT technique are fully described – concepts and theorems related to Euclidean scaling being discussed in the Appendix – and the software manipulation is illustrated on real data.

Keywords: Euclidean vector spaces; Spectral decomposition; RV coefficient

1. Introduction

The ACT (STATIS method) is an exploratory technique of multivariate data analysis based on linear algebra and especially Euclidean vector spaces (ACT stands for Analyse Conjointe de Tableaux, STATIS stands for Structuration des Tableaux A Trois Indices de la Statistique). It has been devised for multiway data situations on the basic idea of computing Euclidean distances between configurations of points (Escoufier, 1973).

At the time or writing, ACT (STATIS method) and ACT (dual STATIS method) can be obtained from CISIA as executable codes running on a PC under DOS. This package encloses also the Fortran 77 code which runs on various mainframes. A more general flexible software will be implemented at the end of 1993 in SPAD distributed by CISIA too. Besides, writing and running the ACT procedure in your own environment is quite simple since only usual

Correspondence to: C. Lavit, Unité de Biométrie, ENSA.M-INRA-UM II, Place Viala, 34060 Montpellier, France.

0167-9473/94/$07.00
SSDI 0167-9473(92)00003-A

routines about matrix computations and eigensystems are required.

For the time being, the software ACT (STATIS method) handles input data of this sort:

– A set of K matrices $X_1, \dots, X_K$, each X_k of dimension $I \times J_k$ is a data matrix of J_k quantitative variables measured on the same I observations.

– A set of I weights $m_1, \dots, m_I$.

– Parameters to specify directives of computation and output.

This program centers obligatorily each variable of each X_k according to the weights $m_1, \dots, m_I$. Denoting by W_k the scalar products between observations at stage k and D the diagonal matrix

$$\begin{pmatrix} m_1 & & \\ & \ddots & \\ & & m_I \end{pmatrix}$$

rescaled to obey $\Sigma m_i = 1$, we obtain the following results:

– Euclidean distances between W_k and $W_{k'}$ or, in other words, between configurations of observations at stages k and k'. These distances are derived from the scalar product

$$\operatorname{Tr}(W_k D W_{k'} D) \quad \text{or} \quad \frac{\operatorname{Tr}(W_k D W_{k'} D)}{\sqrt{\operatorname{Tr}\left[(W_k D)^2\right]}\sqrt{\operatorname{Tr}\left[(W_{k'} D)^2\right]}}.$$

– Trajectories which reflect the contribution of each observation to the Euclidean distance between the W_k's.

– A compromise matrix of dummy scalar products between observations, computed as a weighted sum of the W_k's.

The software ACT (dual STATIS method) deals with the following input data:

– A set of K matrices $X_1, \dots, X_K$ corresponding to the same J quantitative variables measured on K different groups of $I_1, \dots, I_K$ observations.

– K sets $D_1, \dots, D_K$ of weights and, as before, parameters to specify directives of computation and output.

If $V_k = X_k' D_k X_k$ denotes the covariance matrix at stage k, three main results are then carried out:

– Euclidean distances between the covariance matrices V_k and $V_{k'}$ derived from the scalar product $\operatorname{Tr}(V_k V_{k'})$.

– Trajectories which reflect the contribution of each variable to the Euclidean distance between the V_k's.

– A compromise covariance matrix computed as a weighted sum of the V_k's.

The next version of the software will let the user work on centred scalar products (as it was the case for W_k), non-centred scalar products (as it was the case for V_k) or scalar products centred on one on the I observations, and compute scalar products of the form $\operatorname{Tr}(W_k S W_{k'} S)$, S being any positive definite symmetric matrix.

In the following chapter, we recall well-known underlying mathematics and geometric properties of scalar products derived from usual data matrices. Chapter 3 describes the up-to-date developments of the ACT technique. The software manipulation is illustrated on real data in chapter 4. The Appendix reviews concepts and theorems related to Euclidean scaling methods from an algebraic viewpoint.

To conclude this chapter, it should be noted that, in a recent work, Franc pointed out connections between different multiway techniques by rewriting them in a multilinear algebra context (Franc, 1992). A n-way data matrix is considered as an element of the tensor product of n vector spaces. And solutions provided by the different methods (TUCKER, PARAFAC/ CANDECOMP, CANDELINK among others) are expressed as sums of tensor products of vectors which obey specific constraints.

2. A few remarks on scalar products between observations

2.1. Scalar products derived from data matrices

Let $X_{I\times J}$ be a data matrix which consists of J variables measured on I observations.

Example 1. We describe the set of observations as:
- Vectors $x_1, x_2, \ldots, x_I$ of a real vector space of dimension J, on which we define the scalar product: $(x_i \mid x_{i'}) = \Sigma_j X_{ij} X_{i'j}$. The scalar product $(x_i \mid x_{i'})$ is the element $W_{ii'}$ of $W = XX'$.
- Points of a metric space since $d^2(i, i') = \Sigma_j (X_{ij} - X_{i'j})^2$.
- A J-dimensional configuration of points $M_1, M_2, \ldots, M_I$. The coordinates of M_i are simply the elements $X_{i1}, \ldots, X_{iJ}$ of X. Thus we have

$$\overrightarrow{\mathrm{OM}_i}.\overrightarrow{\mathrm{OM}_{i'}} = W_{ii'} \quad \text{and} \quad \left\| \overrightarrow{\mathrm{M}_i\mathrm{M}_{i'}} \right\| = \mathrm{d}(i, i').$$

Note that $d^2(i, i') = \| x_i - x_{i'} \|^2$ where $\| \cdot \|$ is the norm induced by the scalar product $(.\mid.)$. Then

$$\mathrm{d}^2(i, i') = W_{ii} + W_{i'i'} - 2W_{ii'}$$

which implies that

$$W_{ii'} = \tfrac{1}{2}\left[\mathrm{d}^2(0, i) + \mathrm{d}^2(0, i') - \mathrm{d}^2(i, i')\right].$$

Example 2. Suppose now that the columns of $X_{I\times J}$ are centred according to the weights $m_1, \ldots, m_I$ with the constraint $\Sigma_i m_i = 1$. That is to say $\Sigma_i m_i X_{ij} = 0$ for $j = 1, \ldots, J$. Therefore we describe the set of variables as:
- Vectors $x_1, x_2, \ldots, x_J$ of a real vector space of dimension I. If we want the

scalar product $(x_j \mid x_{j'})$ to be the covariance between variables j and j', it is necessary that:

$$(x_j \mid x_{j'}) = \sum_i m_i X_{ij} X_{ij'}.$$

The scalar product $(x_j \mid x_{j'})$ is the element jj' of $X'DX$ where D denotes the diagonal matrix whose elements are $m_1, \ldots, m_I$.

– A different approach is to consider

$$\begin{pmatrix} X_{1j} \\ \vdots \\ X_{Ij} \end{pmatrix}$$

as the matrix of x_j in an orthogonal but unstandardized basis $(e_1, \ldots, e_I)$ assuming $\| e_i \|^2 = m_i$ for $i = 1, \ldots, I$. Then $x_j = \Sigma_i X_{ij} e_i$ implies $(x_j \mid x_{j'}) = \Sigma_i X_{ij} X_{ij'} (e_i \mid e_i) = \Sigma_i m_i X_{ij} X_{ij'}$. Since the basis $(e_1, \ldots, e_I)$ is unstandardized, we do not have any trivial configuration of the variables using the original coordinates $X_{1j}, \ldots, X_{Ij}$.

Generalization Considering that the columns of $X_{I \times J}$ are the rows of the transpose $X'_{J \times I}$, the two former examples are particular cases of the following general formulation. Let $X_{I \times J}$ be a data matrix. We describe the set of the rows as vectors $x_1, \ldots, x_I$ of a real vector space of dimension J. We define the scalar product $(x_i \mid x_{i'}) = \Sigma_j \Sigma_{j'} Q_{jj'} X_{ij} X_{i'j'}$ as the element $W_{ii'}$ of $W = XQX'$, Q being a positive definite symmetric matrix. Then

$$\begin{pmatrix} X_{i1} \\ \vdots \\ X_{iJ} \end{pmatrix}$$

is the matrix of x_i in any basis $(e_1, \ldots, e_J)$ obeying $(e_j \mid e_{j'}) = Q_{jj'}$. The rows of $X_{I \times J}$ can be plotted as points $M_1 \ldots M_I$ generated by the original coordinates in a set of axes neither orthogonal nor standardized. Actually this representation is unsatisfactory for our perception used to classic Euclidean geometry in which axes are assumed to be rectangular and to have unit length. It is the reason why methods based on the spectral decomposition of W calculate coordinates of the points $M_1, \ldots, M_I$ in a new suitable set of axes.

2.2. *Configuration of points*

Let us consider the matrix $W_{I \times I}$ of scalar products between I elements, forgetting the way it has been calculated. The problem is to plot the I elements in a set of orthonormal axes. We briefly note a few results, more or less known, which are detailed in the Appendix.

W is a positive semi-definite symmetrix matrix. From the spectral decomposition of W:

$$W = P_{I\times r}\Lambda_{r\times r}P'_{r\times I} \quad \text{with} \quad P'P = I_{r\times r} \quad \text{(Identity matrix)},$$

r denoting the rank of W, we obtain an Euclidean configuration $M_1, \ldots, M_I$ in a r-dimensional space. The r coordinates of M_i are the elements of the ith row of $P\Lambda^{1/2}$. Consequently, the scalar products $W_{ii'}$ and the distances $\mathrm{d}(i, i')$ can be calculated in a classical Euclidean context since

$$\overrightarrow{\mathrm{OM}_i}.\overrightarrow{\mathrm{OM}_{i'}} = W_{ii'} \quad \text{and} \quad \left\|\overrightarrow{\mathrm{M}_i\mathrm{M}_{i'}}\right\| = \mathrm{d}(i, i').$$

The spectral decomposition of W is a particular singular value decomposition as it is defined in the Appendix. Let S be any $I \times I$ positive definite symmetric matrix. It will be shown that W can be written as

$$W = P_{I\times r}\Sigma_{r\times r}P'_{r\times I} \quad \textit{with} \quad P'SP = I_{r\times r} \quad \text{(Identity matrix)}.$$

The columns of P are the eigenvectors of the self adjoint matrix WS, and Σ is the diagonal matrix whose elements are the corresponding eigenvalues. If we denote by r the rank of W (or WS), we obtain another set of r coordinates of M_i by taking the elements of the ith row of $P\Sigma^{1/2}$. Similarly the scalar products $W_{ii'}$ and the distances $\mathrm{d}(i, i')$ can be calculated in an usual Euclidean context since

$$\overrightarrow{\mathrm{OM}_i}.\overrightarrow{\mathrm{OM}_{i'}} = W_{ii'} \quad \text{and} \quad \left\|\overrightarrow{\mathrm{M}_i\mathrm{M}_{i'}}\right\| = \mathrm{d}(i, i').$$

Remark. If the scalar products have been derived from a data matrix $X_{I\times J}$, and hence from J coordinates, the rank of $W_{I\times I}$ is less than the minimum of I and J.

2.3. Centred scalar products

Let us consider again a matrix of I observations on J variables and suppose now that the mean of the I observations is significant. Then we can deal with the centred matrix X obeying $\Sigma_i X_{ij} = 0$, or more generally $\Sigma_i m_i X_{ij} = 0$ if the observations are weighted, in order to shift the mean to the origin. This derivation is equivalent to the transformation of the original W in W^G defined by

$$W^G_{ii'} = W_{ii'} - W_{i\cdot} - W_{i'\cdot} + W_{\cdot\cdot}$$

where $W_{i\cdot} = \Sigma_{i'} m_{i'} W_{ii'} / \Sigma_i m_i$ and $W_{\cdot\cdot} = \Sigma_i \Sigma_{i'} m_i m_{i'} W_{ii'} / (\Sigma_i m_i)^2$.

The singular value decomposition of W^G, or $W^G S$, S being any positive definite symmetric matrix, provides a configuration $M_1, \ldots, M_I$ in a r-dimensional space in which the origin O is the centroid G of $M_1, \ldots, M_I$ weighted by

$m_1, \ldots, m_I$. The rank r of W^G (or $W^G S$) is less than the minimum of $I-1$ and J. Note that the case

$$S = D = \begin{pmatrix} m_1 & & \\ & \ddots & \\ & & m_I \end{pmatrix}$$

with the constraint $\Sigma_i m_i = 1$ gives the principal components of the PCA of the centred data matrix $X_{I \times J}$ when the eigenvalues of $W^G D$ are ordered in the usual way. This choise leads to the agreeable interpretation of eigenvalues as variances of principal components.

2.4. Scalar product-like derived from dissimilarity matrices

If the data are available in the form of an $I \times I$ matrix Δ of dissimilarities between pairs of observations, some results are still valid:

– Choosing the observation i_0 as origin, we can calculate the scalar product-like matrix $W^{M_{i_0}}$ of dimension $I-1$, defined by

$$W_{ii'}^{M_{i_0}} = \tfrac{1}{2}\left[\Delta_{i_0 i}^2 + \Delta_{i_0 i'}^2 - \Delta_{ii'}^2\right].$$

And reciprocally, given W we deduce

$$\Delta_{ii'}^2 = W_{ii} + W_{i'i'} - 2W_{ii'}.$$

– On the other hand, if we suppose that the mean of the I observations is significant, we can deal with the so-called Torgerson symmetric matrix W^G defined by

$$W_{ii'}^G = \tfrac{1}{2}\left[\Delta_{i\cdot}^2 + \Delta_{i'\cdot}^2 - \Delta_{ii'}^2 - \Delta_{\cdot\cdot}^2\right]$$

where $\Delta_{i\cdot} = \Sigma_{i'} m_{i'} \Delta_{ii'} / \Sigma_i m_i$ and $\Delta_{\cdot\cdot} = \Sigma_i \Sigma_{i'} m_i m_{i'} \Delta_{ii'} / (\Sigma_i m_i)^2$, $m_1, \ldots, m_I$ being the weights of the observations.

– It is well known that $W^{M_{i_0}}$ (or W^G) is positive semi-definite if the dissimilarities are Euclidean distances and reciprocally. In this particular case, the singular value decomposition of W or WS provides a configuration $M_1, \ldots, M_I$ of the observations in a r-dimensional space whether W is centred or not. In both cases, the rank r of W (or WS) is less than $I-1$.

3. Strategy for ACT: Scalar products between configurations

The central idea of the ACT technique is to compare configurations of the same observations obtained in different circumstances. Thus we need to introduce a measure of similarity between two configurations. This is equivalent to define a distance between the corresponding scalar product matrices. We can use the classic Euclidean norm

$$\| W_1 - W_2 \|^2 = \sum_i \sum_{i'} \left[(W_1 - W_2)_{ii'}\right]^2 = \mathrm{Tr}\left[(W_1 - W_2)^2\right],$$

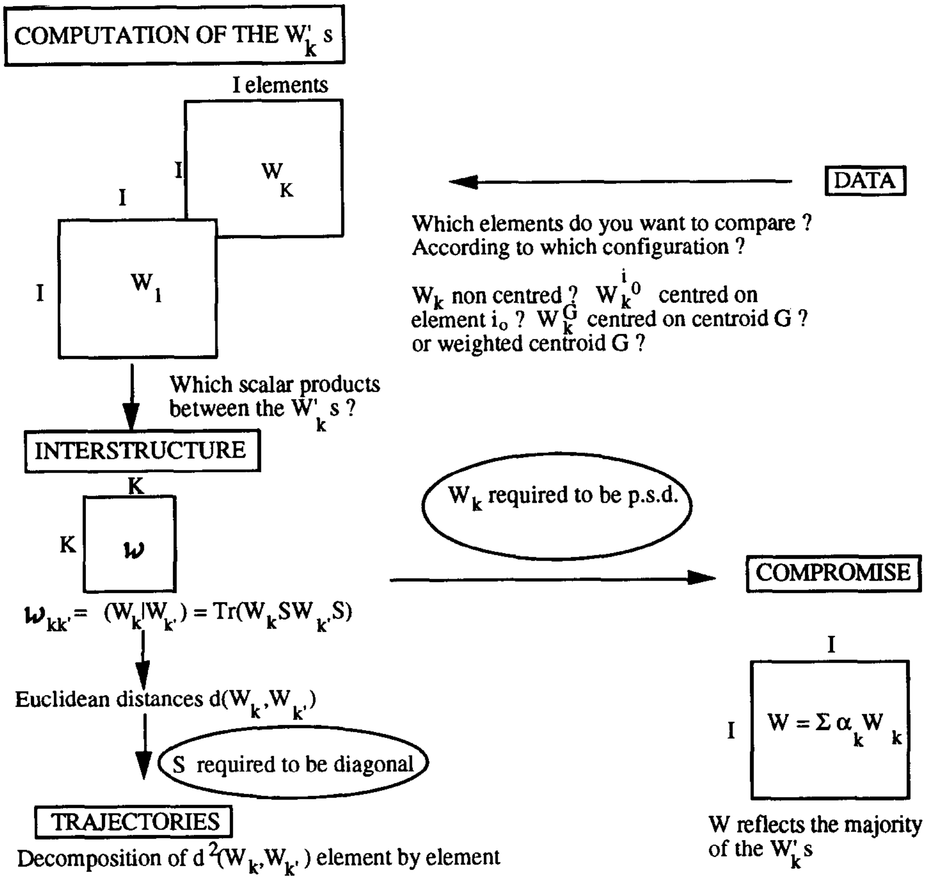

Figure 1. The ACT procedure

or a weighted version

$$\| W_1 - W_2 \|^2 = \sum_i \sum_{i'} S_{ii} S_{i'i'} \left[(W_1 - W_2)_{ii'} \right]^2,$$

or, more generally

$$\| W_1 - W_2 \|^2 = \mathrm{Tr}\left[(W_1 S - W_2 S)^2 \right]$$

for any positive definite symmetric matrix S. Although distances between symmetric matrices have been studied for a long time, it is more convenient to argue their properties as a particular case of distances between linear mappings. Definitions and algebraic derivations are discussed in the Appendix.

According to the data and the objectives of the analysis, we have first to decide which elements have to be compared. Then, it is necessary to specify from which origin we compute the scalar products W_k's (whereas this question does not arise when we compute distances). We will illustrate these decisive choices and their consequences on a few examples in the last section.

3.1. Interstructure

First step: which distance between the W_k's? Properties of these distances are discussed in the Appendix. We will simply list the different choices.
- $d^2(W_k, W_{k'}) = \Sigma_i \Sigma_{i'} [(W_k - W_{k'})_{ii'}]^2$ deduced from the scalar product $(W_k \mid W_{k'}) = \mathrm{Tr}(W_k W_{k'}) = \mathrm{Tr}(W_k S W_{k'} S)$ with S equal to the identity matrix.
- $d^2(W_k, W_{k'}) = \Sigma_i \Sigma_{i'} S_{ii} S_{i'i'} [(W_k - W_{k'})_{ii'}]^2$ deduced from $(W_k \mid W_{k'}) = \mathrm{Tr}(W_k S W_{k'} S)$ with S diagonal. Note that for W_k^G centred on a weighted centroid G with weights $m_1, \ldots, m_I$, the matrix S can be any diagonal matrix and not necessarily

$$\begin{pmatrix} m_1 & & \\ & \ddots & \\ & & m_I \end{pmatrix}.$$

- $d^2(W_k, W_{k'}) = \mathrm{Tr}[(W_k S - W_{k'} S)^2]$ deduced from the scalar product $(W_k \mid W_{k'}) = \mathrm{Tr}(W_k S W_{k'} S)$ where S is any positive definite symmetric matrix.

Second step: do we compare the W_k's or the normed W_k's? A large distance $d(W_k, W_{k'})$ points out a strong difference between W_k and $W_{k'}$. Difference in shape or difference in size? To eliminate the second effect, we can compare the normed scalar products

$$\frac{W_k}{\| W_k \|} = \frac{W_k}{\sqrt{\mathrm{Tr}(W_k S)^2}}$$

and thus calculate

$$\left(\frac{W_k}{\| W_k \|} \,\middle|\, \frac{W_{k'}}{\| W_{k'} \|} \right) = \frac{\mathrm{Tr}(W_k S W_{k'} S)}{\sqrt{\mathrm{Tr}(W_k S)^2} \sqrt{\mathrm{Tr}(W_{k'} S)^2}}.$$

Graphical representation of the interstructure. Let $W_{K \times K}$ be the interstructure matrix whose elements are the scalar products $(W_k \mid W_{k'})$. To plot the K stages in a two or three-dimensional space, say a h-dimensional space, we use the least squares approximation W_h of W, equal to the h first elements of the spectral decomposition $\sigma_1 p_1 p_1' + \cdots + \sigma_r p_r p_r' = P \Sigma P'$ of W, with $P'P = I$. Stage k is plotted as point M_k whose coordinates are the h first elements of the kth row of $P \Sigma^{1/2}$. The points $M_1, \ldots, M_K$ satisfy $\overrightarrow{\mathrm{OM}_k} \cdot \overrightarrow{\mathrm{OM}_{k'}} = (W_h)_{kk'}$. And the loss function is

$$\| W - W_h \|^2 = \sum_k \sum_{k'} [(W - W_h)_{kk'}]^2 = \sum_{l=h+1}^{r} \sigma_l^2.$$

On this graph, scalar products are not easily readable, except for the norm $\| W_k \|$ approximated by the length of vector $\overrightarrow{\mathrm{OM}_k}$, and for the scalar product between normed W_k and normed $W_{k'}$ approximated by the cosine of $(\overrightarrow{\mathrm{OM}_k}, \overrightarrow{\mathrm{OM}_{k'}})$.

Distances $d(W_k, W_{k'})$ can also be readable on a graph using the first elements of the spectral decomposition of the interstructure matrix W^G, centred on the centroid G of the K stages, in a similar way. On these two graphs, the projected distance induced by least squares approximation is systematically lower than $d(W_k, W_{k'})$. Besides, it is possible to distort Euclidean distances $d(W_k, W_{k'})$ into ultrametric distances and build a dendrogram, or use another technique of visual information.

3.2. Compromise

In this section, the W_k's are required to be positive semi-definite whereas this restrictive assumption was not necessary in the interstructure derivations.

Property 1. *Let $\mathbf{W}_1 = \sigma_1 p_1 p_1'$ be the first element of the spectral decomposition of the interstructure matrix $\mathbf{W}$. Components of p_1 can be chosen positive.*

To establish this particular case of the Frobenius theorem, see Property 8 in the Appendix which implies that all the elements of $\mathbf{W}$ are positive. Consequently, in the configuration of the K stages, cosines are positive and angles $(\overrightarrow{\mathrm{OM}_k},\overrightarrow{\mathrm{OM}_{k'}})$ acute. Thus the points M_k are situated inside a convex cone.

Definition 1. The $I \times I$ compromise matrix W is defined as a weighted sum $\sum_k \alpha_k W_k$. The coefficient α_k is the coordinate of stage k in the one-dimensional plot deduced from the first element $\mathbf{W}_1$ of the spectral decomposition of $\mathbf{W}$.

Property 2. *As the W_k's are positive semi-definite and the α_k's are positive, the compromise matrix W is positive semi-definite. Thus W will be considered as a scalar product matrix which induces a compromise configuration of the I elements.*

Property 3. *The compromise matrix W is the linear combination of the W_k's the most related to each W_k. In other words, W maximizes*

$$\sum_l \Big(\sum_k \alpha_k W_k \mid W_l \Big)^2 / \sum_k \alpha_k^2.$$

Recalling that $\mathbf{W}_{kl}$ equals $(W_k \mid W_l)$, we develop the numerator into

$$\sum_k \sum_{k'} \alpha_k \alpha_{k'} \sum_l (W_k \mid W_l) = \sum_k \sum_{k'} \alpha_k \alpha_{k'} (\mathbf{W}^2)_{k'k}.$$

In the spectral decomposition of $\mathbf{W}$, p_1 is the eigenvector of $\mathbf{W}\mathbf{W}^* = \mathbf{W}^2$ associated to the largest eigenvalue (see Property 7 of the Appendix). Thus the quotient can be written as the Rayleigh quotient

$$\frac{x'\mathbf{W}\mathbf{W}^* x}{\| x \|^2} \quad \text{with} \quad x = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{pmatrix},$$

which is maximum for $x = p_1$.

Constraints on the α_k's. If the W_k's are centred on a centroid G, or a particular element i_0, the compromise matrix W will obviously be centred in the same way. Similarly, if the origin is equidistant from the I elements at each stage k, this property holds for W. However when the W_k's are correlation matrices, it is necessary to rescale the α_k's with the constraint $\Sigma_k \alpha_k = 1$ in order to obtain a compromise of the same nature. If the W_k's are normed, the compromise will be rescaled to be normed too.

Interpreting the compromise. If W_k and $W_{k'}$ correspond to similar configurations in shape and size, the angle $(\overrightarrow{\mathrm{OM}}_k, \overrightarrow{\mathrm{OM}}_{k'})$ is small and the lenghts $\| \overrightarrow{\mathrm{OM}}_k \|, \| \overrightarrow{\mathrm{OM}}_{k'} \|$ are nearly the same. This case leads to identical values for α_k and $\alpha_{k'}$. On the other hand, a large difference between the two configurations induces either a large angle $(\overrightarrow{\mathrm{OM}}_k, \overrightarrow{\mathrm{OM}}_{k'})$ or unequal lenghts for $\overrightarrow{\mathrm{OM}}_k$ and $\overrightarrow{\mathrm{OM}}_{k'}$, and different values for α_k and $\alpha_{k'}$ in both cases. Consequently W gives relatively less weight to outliers, and leads to a compromise configuration which reflects the inter-element distances as they are seen by the majority.

Graphical representation of the compromise. To plot the I elements according to the compromise in a h-dimensional space, we use the least squares approximation W_h of W, equal to the h first elements of the singular value decomposition $\sigma_1 p_1 p_1' + \cdots + \sigma_r p_r p_r' = P \Sigma P'$ of W, with $P'SP = I$. Element i is plotted as point M_i whose coordinates are the h first elements of the ith row of $P\Sigma^{1/2}$.

The points $M_1, \ldots, M_I$ satisfy $\overrightarrow{\mathrm{OM}}_i \cdot \overrightarrow{\mathrm{OM}}_{i'} = (W_h)_{ii'}$ and the loss function is

$$\| W - W_h \|^2 = \mathrm{Tr}\left[(WS - W_h S)^2 \right] = \sum_{l=h+1}^{r} \sigma_l^2 .$$

Note that it is logical, but not necessary, to take the same matrix S as the one used in the computation of the distances $\mathrm{d}(W_k, W_{k'})$. Think, for example, to the case where S is not diagonal.

3.3. Trajectories

In this section, S is required to be diagonal. Then $\mathrm{d}^2(W_k, W_{k'})$ can be written element by element, as the sum $\Sigma_i S_{ii} (\Sigma_{i'} S_{i'i'} [(W_k - W_{k'})_{ii'}]^2)$ and split into contributions of the different elements. This decomposition leads to an $I \times [\frac{1}{2}K(K-1)]$ matrix from which we can detect which elements are strongly perturbed from one configuration to another.

However knowing the direction of the perturbation needs further investigations. We consider the W_k's and the compromise W as matrices of linear mappings between an Euclidean vector space F and its dual F^* (see Appendix, Section 6).

$$w_k^i = \begin{pmatrix} (W_k)_{i1} \\ \vdots \\ (W_k)_{iI} \end{pmatrix}$$

is a vector of F^* which appears in the decomposition of $d^2(W_k, W_{k'}) = \Sigma_i S_{ii} \| w_k^i - w_{k'}^i \|^2$. It is now of interest to know the direction of the different vectors $w_k^i - w_{k'}^i$. Eigenvectors $p_1, \ldots, p_r$ of WS which occur in the singular value decomposition $\sigma_1 p_1 p_1' + \cdots + \sigma_r p_r p_r'$ of W, can be completed to form a s-orthonormal basis of F^*. The idea is to express w_k^i in this new basis.

Graphical representation of the trajectories. Unfortunately, if we want to plot the I elements as they are seen by the different stages in a unique two-dimensional space, it is obvious that none of the subspaces spanned by $p_1, \ldots, p_r$ corresponds to the best choice. Nevertheless, we decide to restrict w_k^i to its projection on the two-dimensional space spanned by p_1 and p_2.

In this space, we know that a least squares approximation of the compromise configuration is provided by the points $M_1, \ldots, M_I$ whose coordinates are the components of $\sigma_1^{1/2}\ p_1 = \sigma_1^{-1/2} WSp_1$ (namely the projections on p_1 rescaled by $\sigma_1^{-1/2}$) and $\sigma_2^{1/2} p_2 = \sigma_2^{-1/2} WSp_2$. In order to draw the trajectory of element i around its compromise position M_i, we rescale the projections of the w_k^i's for $k = 1, \ldots, K$, in the same way. More precisely, we plot the I elements, as they are seen by stage k, by the points $M_1^k, \ldots, M_I^k$ whose coordinates are the I components of $\sigma_1^{-1/2} W_k Sp_1$ on the first axis and $\sigma_2^{-1/2} W_k Sp_2$ on the second axis. In spite of this obvious lack of optimality, numerous examples show that, in practice, the vector $\overrightarrow{\mathrm{M}_i^k \mathrm{M}_i^{k'}}$ gives a good idea of the importance and the direction of the change of position of element i between the stages k and k'.

3.4. Some applications of the ACT technique

Example 1. The first set of data, fully discussed in Chapter 4 consists of K matrices $X_1, \ldots, X_K$ of I rows and J columns, corresponding to the judgment of K students on I of their professors according to J criteria. In this particular case, we decide to work on the matrices $W_k^G = X_k X_k'$ after having centred each column of X_k. In other words we compare the configurations of the I professors as they are seen by each student, after having standardized severe and generous students to the same level of notation. Distances $d(W_k, W_{k'})$ are interpreted as disagreements between judgments. The compromise matrix W provides a configuration of the I professors reflecting the majority opinion. And the trajectories point out who are the professors on which students are not in agreement.

On the other hand we can decide to study the correlation matrices $V_k = (1/I) X_k' X_k$ (the columns of X_k being standardized to have mean 0 and unit variance). The objectives are then different. Forgetting the configurations of professors, we are now interested in associations or oppositions between criteria. Are they different from one student to another? Which criteria make the difference? Answers are given by the interstructure and the trajectories, whereas the compromise correlation matrix summarizes the V_k's.

Example 2. Let us consider a set of five sociological data matrices X_{54}, X_{62}, X_{68}, X_{75}, X_{82} describing the evolution of the working population of I communes around Montpellier in the last forty years. It happens that the definition of occupation groups changed from 1975 to 1982. Thus, the population has been classified into nine occupation groups for the four first census returns and only eight groups for the last one. Nevertheless the five configurations of the I communes can be compared through the matrices W_k^G (where the centroid G, weighted or not, represents a dummy average commune) or through $W_k^{i_0}$ (if we take Montpellier, indexed by i_0, as a reference).

Example 3. The third set of data consists of J variables characterizing different stages of the sleep, collected on I_1 narcoleptics and I_2 persons in good health, after 16 hours, 20 hours and 24 hours of wakefulness. Comparing the six correlation matrices by means of the interstructure and the trajectories can be a first approach to understand how this disease disturbs the distribution of the different stages of the sleep, whereas the compromise matrix is not of interest here.

Example 4. The last set of data concerns I sites described by three specialists: a botanist, a pedologist and a biologist. The botanist describes the floristic composition of the sites by an $I \times J_1$ presence–absence data matrix. The pedologist provides an $I \times J_2$ data matrix of chemical characteristics of the soil, and the biologist gives an $I \times J_3$ data matrix concerning the abundance of various earthworms.

If two sites are claimed similar by the biologist, are they found similar by the pedologist or the botanist? To deal with such a situation, we have to discuss with each specialist which dissimilarity or distance is appropriate, then derive scalar product-like matrices W_1, W_2 and W_3 centred on a reference site i_0, and compare them by means of the interstructure for which the positivity of the W_k's is not required. Further information can be extracted from the decomposition of the squared distance $\mathrm{d}^2(W_k^{i_0}, W_{k'}^{i_0})$ between specialists into contributions of the different sites.

Suppose now that the data are available in the form of K dissimilarity matrices, each matrix corresponding to a dissimilarity coefficient δ_k whose square root is an Euclidean distance as Rogers and Tanimoto, Russel and Rao or Ochiaï coefficients (Fichet and Le Calvé, 1984). In this case, instead of performing only the first step of ACT on the scalar product-like matrices straightforward derived from the δ_k's, we can operate the whole procedure on the positive semi-definite W_k's derived from the square roots $\sqrt{\delta_k}$ as follows:

$$(W_k^{i_0})_{ii'} = \tfrac{1}{2}[\delta_k(i_0, i) + \delta_k(i_0, i') - \delta_k(i, i')].$$

The compromise matrix W induces Euclidean distances between sites which satisfy $\mathrm{d}^2(i, i') = \sum_k \alpha_k \delta_k(i, i')$. And, all things considered, as $\sqrt{\delta}$ is a monotonic function of δ, performing ACT on $\sqrt{\delta}$ instead of δ turns out to be a useful insight into the structure of the original dissimilarity matrices.

4. Input and output of the 01.90 version of ACT (STATIS method)

Input. $K = 8$ students judge $I = 11$ professors by means of $J = 7$ criteria: competence, lucid explanation, pedagogy, cheerfulness, is the professor dynamic? accessible? does the student find the subject interesting? The data are given in Table 1.

"Practice" having missing values, is considered as a supplementary row and ignored when we compute the W_k's and the compromise. The other active rows have identical weights $m_i = 1/10$. Practice has weight 0. The preprocessing of the data consists of centering each column of each X_k according to those weights. Thus we compare the 10×10 matrices W_k^G centred on the centroid G of the ten active rows, W_4^G being derived from only six criteria.

Interstructure output. The norms of the W_k's are derived from the scalar product $(W_k \mid W_{k'}) = \operatorname{Tr}(W_k S W_{k'} S)$ where S is diagonal and equal to

$$\begin{pmatrix} 1/10 & & \\ & \ddots & \\ & & 1/10 \end{pmatrix}.$$

These norms are equal to 119, 151, 87, 78, 101,117, 137 and 102. Since they are different, we decide to compute the Euclidean distances between the normed W_k's deduced from the RV-coefficients (Definition 9 of the Appendix). We note that the opinion of judges no. 2 and no. 8 fairly differs from the others.

Compromise output. The compromise matrix W is defined as the linear combination $\Sigma_k \alpha_k W_k$ of normed W_k's with α_k equal to the coordinate of judge k on the first axis of Figure 2. W gives relatively less weight to judges nr. 2 and nr. 8, and leads to a configuration of the professors which reflects principally the opinion of the six judges left.

α_k	0.18	0.08	0.17	0.17	0.22	0.20	0.21	0.11
$\mathrm{d}^2(W, W_k)$	0.46	1.32	0.60	0.61	0.19	0.33	0.27	1.06

Since the W_k's are centred on the centroid G of the active rows, the compromise W is centred in the same way. In addition, as the data are available in the form of observations $\times$ variables matrices, we have $J_1 + \ldots + J_8$ variables at our disposal. We can then compute their correlations with the coordinates of the compromise points in order to explain the position of the different professors.

Trajectories output. In the interstructure output, we noted that the distances between judge no. 2 and the others were quite important. It is interesting to

Table 1

computer science	judge nr. 1	judge nr. 2	judge nr. 3	judge nr. 4
architecture	13 18 8 1 13 3 20	18 18 10 2 14 12 18	15 18 10 4 10 0 18	14 17 7 12 2 10
theory	12 10 5 3 9 5 7	6 4 8 2 4 12 16	12 5 4 8 10 8 10	6 5 7 5 10 7
languages	4 10 6 6 6 10 2	2 2 2 4 2 20 10	10 12 5 6 5 3 10	3 2 3 2 4 3
practice	17 16 17 17 17 18 18	18 20 18 12 14 18 18	14 13 15 12 12 10 13	14 12 14 10 15 13
economics	15 12 13 14 13 18 9	14 16 14 12 10 18 10	13 12 12 12 12 11 11	7 6 7 4 16 5
accounting	16 18 15 11 13 15 15	18 18 18 12 16 16 18	17 15 15 14 12 14 14	16 16 14 12 12 13
management	12 12 9 1 11 7 9	18 20 18 10 12 18 14	19 16 18 13 15 12 16	16 14 10 7 3 12
information system design	10 18 0 2 12 2 17	16 18 18 16 14 18 16	16 2 2 2 2 2 12	5 5 5 2 6 9
statistics	17 12 16 19 20 19 13	18 16 18 12 12 18 14	15 6 14 12 15 10 5	10 11 10 5 14 12
operations research	11 8 5 1 8 1 15	18 16 14 0 14 6 16	15 13 8 2 2 2 10	13 13 7 10 3 14
English	17 18 15 10 13 15 15	10 4 2 0 4 4 8	15 10 10 0 12 5 4	17 13 10 13 10 9

computer science	judge nr. 5	judge nr. 6	judge nr. 7	judge nr. 8
architecture	14 15 0 0 7 0 15	19 18 5 3 15 1 18	15 15 0 0 11 5 13	18 14 11 17 15 15 16
theory	12 8 8 8 15 16 9	11 14 13 8 17 15 17	12 15 12 11 14 15 13	18 18 17 12 17 18 15
languages	5 2 0 0 2 8 0	8 4 1 2 1 1 1	10 0 5 5 1 5 1	17 18 18 10 12 14 15
practice	16 15 16 14 10 14 16	14 12 13 15 17 18 17	20 15 14 13 16 14 18	
economics	12 10 15 15 10 15 7	9 12 15 17 16 14 15	15 12 15 16 15 18 14	16 15 15 12 17 18 17
accounting	16 15 15 12 14 13 14	19 17 18 15 17 3 15	16 13 15 11 14 14 14	16 16 12 5 10 14 16
management	8 14 9 2 15 8 3	15 13 15 12 17 5 15	12 15 12 10 12 6 18	14 10 11 9 14 10 16
information system design	12 7 5 7 5 10 14	12 5 3 2 10 15 13	15 10 0 15 12 15 10	19 18 19 17 17 14 12
statistics	18 10 18 14 15 15 8	12 15 15 10 18 17 15	20 20 20 20 20 20 20	1 2 4 2 2 9 2
operations research	10 8 5 2 11 8 10	14 10 10 5 12 10 17	10 11 13 4 2 5 12	15 11 12 7 14 14 16
English	14 13 15 5 10 10 12	18 15 17 15 17 12 17	17 17 15 12 14 14 15	13 15 10 11 12 10 17

Table 2
RV-coefficients

1							
0.17	1						
0.45	0.28	1					
0.43	0.30	0.42	1				
0.69	0.22	0.53	0.54	1			
0.48	0.16	0.54	0.63	0.76	1		
0.63	0.18	0.49	0.41	0.87	0.73	1	
0.42	0.11	0.31	0.23	0.29	0.19	0.40	1

Table 3
$d^2(k, k') = 2(1 - RV)$

0							
1.65	0						
1.10	1.44	0					
1.15	1.39	1.16	0				
0.62	1.56	0.94	0.92	0			
1.04	1.68	0.92	0.74	0.48	0		
0.75	1.64	1.03	1.17	0.26	0.54	0	
1.17	1.78	1.39	1.55	1.42	1.62	1.20	0

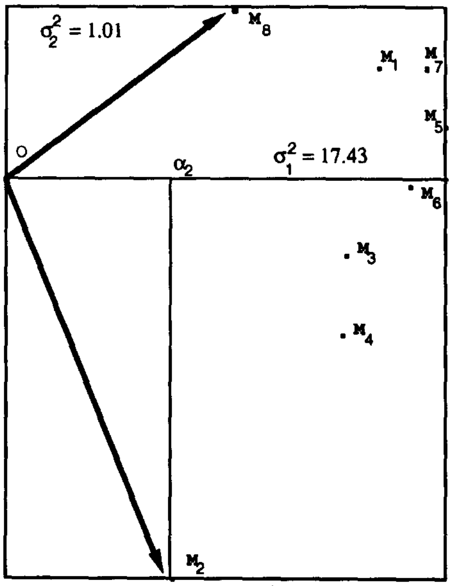

Fig. 2. Graphical representation of the interstucture. Loss function: $\| W - W_h \|^2 = \sigma_3^2 + \cdots + \sigma_8^2 = 1.77$

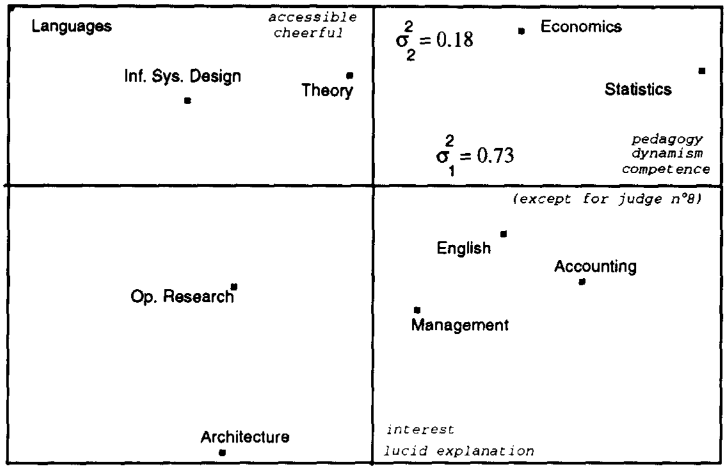

Fig. 3. Graphical representation of the compromise. Loss function: $\| W - W_h \|^2 = \mathrm{Tr}((W - W_h)S)^2 = (\frac{1}{10})^2 \mathrm{Tr}(W - W_h)^2 = \sigma_3^2 + \cdots + \sigma_9^2 = 0.09$

split those distances into contributions $\frac{1}{10} \| w_2^i - w_k^i \|^2$ of the different elements i to detect who are the professors on which the opinion of judge no. 2 differs from the others. Table 4 shows that the difference comes essentially from languages, English, statistics and information system design.

For sake of readibility, only three trajectories are drawn on Figure 4: languages, architecture and the supplementary element practice. The star corresponds to the compromise point. Recall that the vector between judges no. 2 and no. 6 is the projection of $w_6^i - w_2^i$ on the two-dimensional space spanned by eigenvectors p_1 and p_2 of the compromise. (For practice, we compute the centred scalar products between this supplementary row and the active ones and

Table 4
decomposition of the squared distance between judge no. 2 and judge no. k in parts $\frac{1}{10} \| w_2^i - w_k^i \|^2 / \mathrm{d}^2(2, k)$ explained by professor i

squared distance $\mathrm{d}^2(2, k)$ between judge no. 2 and	no. 1	no. 3	no. 4	no. 5	no. 6	no. 7	no. 8
	1.65	1.44	1.39	1.56	1.68	1.64	1.78
contributions (percentage)							
architecture	2.51	3.01	3.36	9.47	2.41	5.70	2.56
theory	7.97	7.56	9.51	9.61	10.48	9.50	8.06
languages	10.41	14.53	16.94	18.62	27.26	22.84	12.48
economics	6.24	2.58	7.07	4.70	3.36	2.84	3.64
accounting	8.41	9.38	11.64	8.08	8.69	5.47	5.53
management	6.94	11.41	7.09	7.20	4.62	4.22	3.64
inf. sys. design	12.85	27.80	15.87	8.93	14.35	6.71	12.47
statistics	16.66	6.65	3.37	10.78	4.93	16.89	37.07
operations research	6.82	6.68	1.14	3.99	2.57	6.41	1.55
English	21.19	10.39	24.01	18.64	21.31	19.41	12.99

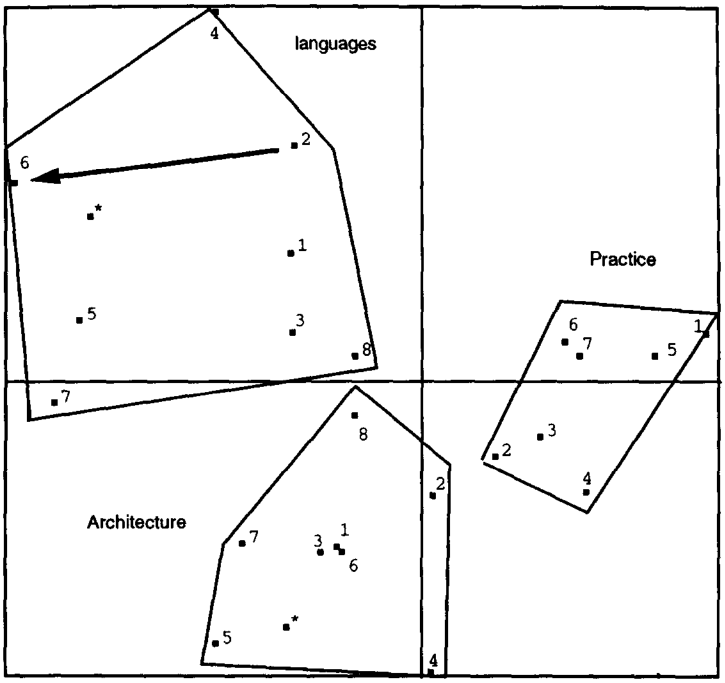

Fig. 4. Graphical representation of the trajectories.

project this vector in the same way). Keeping this distortion in mind, we note that all the vectors $w_k^i - w_{k'}^i$, have significant norms for languages. It means that languages contributes for a large part to any distances $d^2(k, k')$ between judges (and not only between judge no. 2 and the others), and between judges and the compromise. It seems on the contrary, that disagreements between judgments are less crucial for architecture and practice.

Appendix: Euclidean distances between scalar product matrices

We review definitions and properties of linear mappings associated with Euclidean scaling methods. See Rao (1973, 1980), Robert and Escoufier (1976), Sabatier (1987) and Lavit (1988) for proofs and complements.

1. Euclidean vector space

Definition 1. An Euclidean vector space (E, s_J) is the association of a J-dimensional real vector space E and a scalar (or inner) product s_J. $(x \mid y)_{s_J}$ denotes the scalar product of x and y.

Definition 2 (Dual of an euclidean vector space). The dual E^* of an euclidean vector space E is the set $L(E, \mathbb{R})$ of linear mappings between E and the set of real numbers. Note that elements of E^* correspond simply to vectors of E written as row matrices.

This supplementary mathematical concept will appear to be an appropriate tool later.

Property 1. *It is convenient to consider the scalar product on E as an one to one linear correspondence $\mathscr{S}$ between E and E^* (which implies that $\mathscr{S}^{-1}$ does exist). $\forall x \in E$, $\mathscr{S}(x) \in E^*$ is defined as the row vector which, applied to y, gives*

$$(\mathscr{S}(x))(y) = (x \mid y)_s.$$

Definition 3. s^* is defined as follows:

$$\forall u \quad \forall v \in E^* \quad (u \mid v)_{s^*} = v(\mathscr{S}^{-1}(u))$$

is a scalar product on E^*, called dual scalar product of s.

2. Adjoint of a linear mapping

Notation. E and F being two vector spaces, $L(E, F)$ denotes the vector space of linear mappings between E and F.

Definition 4 (Transpose of a linear mapping). Let $\mathscr{A}$ be any element of $L(E, F)$. The transpose of $\mathscr{A}$ is the element $\mathscr{A}'$ of $L(F^*, E^*)$ defined as:

$$\forall x \in E \quad \forall v \in F^* \quad (\mathscr{A}'(v))(x) = v(\mathscr{A}(x)).$$

Definition 5 (Adjoint of a linear mapping). Let $\mathscr{A}$ be any linear mapping between (E, s_J) and (F, s_I). The adjoint of $\mathscr{A}$ is the element $\mathscr{A}^*$ of $L(F, E)$ defined as:

$$\forall x \in E \quad \forall y \in F \quad (x \mid \mathscr{A}^*(y))_{s_J} = (\mathscr{A}(x) \mid y)_{s_I}$$

or $\mathscr{A}^* = \mathscr{S}_J^{-1}\mathscr{A}'\mathscr{S}_I$.

3. Self adjoint operator

Definition 6. An element of $L(E, E)$ is called operator of E.

Definition 7. An operator of (E, s) is self adjoint if $\mathscr{A} = \mathscr{A}^*$, or $\mathscr{S}\mathscr{A} = \mathscr{A}'\mathscr{S}$.

Property 2 (Spectral decomposition of a self adjoint operator $\mathscr{A}$ of (E, s)). *x is an eigenvector of $\mathscr{A}$ if there exists a real number σ such that $\mathscr{A}(x) = \sigma x$. It can be shown that there exists a s-orthonormal basis of eigenvectors. In other words, the matrix A can be written as $A = P\Sigma P'$ with $P'SP = I$. Σ is diagonal and its elements are the eigenvalues of A, the columns of P are the corresponding eigenvectors, S is the matrix of s and I is the identity matrix.*

4. Singular value decomposition

Property 3. *Let $\mathscr{A}$ be any linear mapping between (E, s_J) and (F, s_I). Then $\mathscr{A}^*\mathscr{A}$ and $\mathscr{A}\mathscr{A}^*$ satisfy the following statements:*
– *$\mathscr{A}^*\mathscr{A}$ is a self adjoint operator of E, and it is positive, that is to say:*

$$\forall x \in E \quad (\mathscr{A}^*\mathscr{A}(x) \mid x)_{s_J} \geq 0.$$

$\mathscr{A}\mathscr{A}^*$ is a positive self adjoint operator of F, as well.
– $\operatorname{Ker}(\mathscr{A}^*\mathscr{A}) = \operatorname{Ker}(\mathscr{A})$.
– *$\mathscr{A}^*\mathscr{A}$ and $\mathscr{A}\mathscr{A}^*$ have the same eigenvalues. If x is an eigenvector of $\mathscr{A}^*\mathscr{A}$ obeying $\| x \|_{s_J} = 1$, and σ the corresponding eigenvalue, then $y = \sigma^{-1/2} \mathscr{A}(x)$ is the corresponding eigenvector of $\mathscr{A}\mathscr{A}^*$ obeying $\| y \|_{s_I} = 1$.*
– *$\mathscr{A}$, $\mathscr{A}^*$, $\mathscr{A}\mathscr{A}^*$ and $\mathscr{A}^*\mathscr{A}$ have identical rank.*

Property 4 (Singular value decomposition of $\mathscr{A}$). *Let r be the rank of $\mathscr{A}$. The operator $\mathscr{A}^*\mathscr{A}$ being self adjoint and positive, have positive eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$. The square roots $\sigma_1, \ldots, \sigma_r$ are called singular values of $\mathscr{A}$. Let $p_1, \ldots, p_r$ be a set or s_I-orthonormal eigenvectors of $\mathscr{A}\mathscr{A}^*$, and $q_1, \ldots, q_r$ a set of s_J-orthonormal eigenvectors of $\mathscr{A}^*\mathscr{A}$, corresponding to the eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$. Then, the matrix A_{IxJ} can be written as the sum*

$$A = \sigma_1 p_1 q_1' S_J + \cdots + \sigma_r p_r q_r' S_J,$$

or as the matrix product

$$A = P \Sigma Q' S_J \quad \text{with} \quad P' S_I P = I \quad \text{and} \quad Q' S_J Q = I.$$

Σ_{rxr} is a diagonal matrix whose elements are the singular values $\sigma_1, \ldots, \sigma_r$ and the columns of P_{Ixr} (respectively Q_{Jxr}) are the eigenvectors $p_1, \ldots, p_r$ (respectively $q_1, \ldots, q_r$) of $\mathscr{A}\mathscr{A}^*$ (respectively $\mathscr{A}^*\mathscr{A}$).

5. Euclidean distances between linear mappings

Definition 8 (Scalar product on $L(E, F)$. Let $\mathscr{A}$ and $\mathscr{B}$ two linear mappings between (E, s_J) and (F, s_I). Then

$$(\mathscr{A} \mid \mathscr{B}) = \operatorname{Tr}(\mathscr{A}\mathscr{B}^*) = \operatorname{Tr}(\mathscr{A}\mathscr{S}_J^{-1}\mathscr{B}'\mathscr{S}_I)$$

defines a scalar product on L(E, F). The induced norm is $d^2(\mathscr{A}, \mathscr{B}) = \| \mathscr{A} - \mathscr{B} \|^2 = \operatorname{Tr}((\mathscr{A} - \mathscr{B})\mathscr{S}_J^{-1}(\mathscr{A} - \mathscr{B})'\mathscr{S}_I)$.

Property 5 (Approximation of a linear mapping). *Let $\mathscr{A}$ be any linear mapping of rank r between E and F. Let us consider its singular value decomposition with singular values $\sigma_1, \ldots, \sigma_r$ in decreasing order. Then*

$$\min_{\substack{\mathscr{B} \in L(E,F) \\ \text{rank of } \mathscr{B} \leq h}} \| \mathscr{A} - \mathscr{B} \|^2 = \| \mathscr{A} - \mathscr{A}_h \|^2 = \sum_{l=h+1}^{r} \sigma_l^2,$$

where $\mathscr{A}_h$ is the sum of the h first elements of the singular value decomposition of $\mathscr{A}$.

6. Embedding W into an Euclidean vector space

Let us consider again a matrix W_{IxI} of scalar products (or scalar product-like) between I elements.

(a) W can be viewed as an element of $L(F, F^*)$. Let F be an I-dimensional vector space, and $(f_1, \ldots, f_I)$ a basis of F, each vector f_i being associated to one of the I elements. The symmetric matrix W can be regarded as the matrix of a bilinear mapping w between $F \times F$ and the set of real numbers $\mathbb{R}$, defined on the basis vectors by $w(f_i, f_{i'}) = W_{ii'}$.

Now, as it has been done for a scalar product in Property 1, we associate to w the following element $\mathscr{W}$ of $L(F, F^*)$. For any vector x of F, $\mathscr{W}(x)$ is the element of F^*, which, applied to any vector y of F, gives

$$(\mathscr{W}(x))(y) = w(x, y).$$

(b) Euclidean structure of F. To calculate euclidean distances between the w's as we did in Section 5 between linear mappings, we need to enrich the structure of F with a scalar product denoted $\mathscr{S}^{-1}$ instead of the straightforward designation $\mathscr{S}$ for a simple question of notation. Thus, the dual scalar product on F^* corresponds to $\mathscr{S}$ and the adjoint mapping of $\mathscr{W}$ is simply $\mathscr{S}\mathscr{W}\mathscr{S}$ (and not $\mathscr{S}^{-1}\mathscr{W}\mathscr{S}^{-1}$).

(c) Euclidean structure of $L(F, F^*)$. Suppose that W_1 and W_2 are two matrices of scalar products (or scalar product-like) between the same I elements. We associate to W_1 and W_2 the corresponding linear mappings $\mathscr{W}_1$ and $\mathscr{W}_2$ between (F, s^{-1}) and (F^*, s). Then the scalar product

$$(\mathscr{W}_1 \mid \mathscr{W}_2) = \mathrm{Tr}(\mathscr{W}_1 \mathscr{W}_2^*) = \mathrm{Tr}(W_1 S W_2 S)$$

induces the Euclidean distance

$$\mathrm{d}^2(W_1, W_2) = \mathrm{Tr}[(W_1 - W_2)S(W_1 - W_2)S].$$

This general formulation includes two usual distances:
- If S is the identity matrix I, then $\mathrm{d}^2\ (W_1, W_2) = \Sigma_i \Sigma_{i'} [(W_1 - W_2)_{ii'}]^2$. The basis $(f_1, \ldots, f_I)$ is required to be s-orthonormal.
- If S is diagonal, then $\mathrm{d}^2(W_1, W_2) = \Sigma_i \Sigma_{i'} S_{ii} S_{i'i'} [(W_1 - W_2)_{ii'}]^2$. The basis $(f_1, \ldots, f_I)$ is required to be s-orthogonal but not necessarily standardized. In other words, weights of the elements are taken into account in the calculation of $\mathrm{d}(W_1, W_2)$.

The more general case where S is not diagonal can be interpreted as some exogenous constraint of contiguity between elements, which should be taken into consideration in the calculation of $\mathrm{d}(W_1, W_2)$.

7. Special case of positive semi-definite W's.

Property 6. *If $\mathscr{W}$ is p.s.d., then $\mathscr{W}\mathscr{S}$ is a positive self adjoint operator of F^*.*

Property 7 (Singular value decomposition of $\mathscr{W}$). *If $\mathscr{W}$ is p.s.d., the singular value decomposition of $\mathscr{W}$ is deduced from the spectral decomposition of the self-adjoint operator $\mathscr{W}\mathscr{S}$. Namely $W=\sigma_1 p_1 p_1' + \cdots + \sigma_r p_r p_r' = P\Sigma P'$ with $P'SP=I$, where the singular values $\sigma_1,\ldots,\sigma_r$ are the eigenvalues of $\mathscr{W}\mathscr{S}$, and $p_1,\ldots,p_r$ are eigenvectors of $\mathscr{W}\mathscr{S}$.*

It can be shown that eigenvectors $p_1,\ldots,p_r$ of $\mathscr{W}\mathscr{W}^*$ are eigenvectors of $\mathscr{W}\mathscr{S}$ and $q_1=\mathscr{S}(p_1),\ldots,q_r=\mathscr{S}(p_r)$ are eigenvectors of $\mathscr{W}^*\mathscr{W}$. Then the singular value decomposition of $W=P\Sigma Q'S^{-1}$ turns out to $P\Sigma P'$.

If S is the identity matrix, the singular value decomposition of W is simply designated as spectral decomposition of $W: P\Sigma P'$ with $P'P=I$.

Property 8. *If $\mathscr{W}$ is p.s.d., $(\mathscr{W}_1 \mid \mathscr{W}_2)=\mathrm{Tr}(\mathscr{W}_1\mathscr{W}_2^*)$ is a positive real number.*

Property 9. *Suppose that $\mathscr{W}_1$ p.s.d. has rank r and $\mathscr{W}_2$ p.s.d. has rank h less than r. If $\sigma_1,\ldots,\sigma_r$ denote the singular values of $\mathscr{W}_1$ in decreasing order, then the scalar product between $\mathscr{W}_1/\|\mathscr{W}_1\|$ and $\mathscr{W}_2/\|\mathscr{W}_2\|$ verify*

$$0 \le \frac{(\mathscr{W}_1 \mid \mathscr{W}_2)}{\|\mathscr{W}_1\|.\|\mathscr{W}_2\|} \le 1 - \frac{\sum_{l=h+1}^{r} \sigma_l^2}{2\sum_{l=1}^{r}\sigma_l^2}.$$

To establish this inequality, apply Property 5 to $\mathscr{A}=\mathscr{W}_1/\|\mathscr{W}_1\|$ and $\mathscr{B}=\mathscr{W}_2/\|\mathscr{W}_2\|$ recalling that $(\mathscr{A}\mid\mathscr{B})=\frac{1}{2}(\|\mathscr{A}\|^2+\|\mathscr{B}\|^2-\|\mathscr{A}-\mathscr{B}\|^2)$. This property must be kept in mind while interpreting the ACT's results on Example 4 of Chapter 3, as the number of flower species might be five times greater than the number of variables measured by the pedologist.

8. Special case of matrices $W=XQX'$.

Let X_1 be an $I\times J_1$ matrix of I observations on J_1 variables, and X_2 an $I\times J_2$ matrix of the same observations on J_2 variables. The J_1 columns of X_1 are centred according to weights $m_1,\ldots,m_I$, with the constraint $\Sigma m_i=1$, and J_2 columns of X_2 are likewise centred according to the same weights.

Thus $W_1=X_1Q_1X_1'$ and $W_2=X_2Q_2X_2'$ are two matrices, obtained in two different circumstances, of centred scalar products between the I elements. If, in addition, we choose S equal to the diagonal matrix

$$\begin{pmatrix} m_1 & & \\ & \ddots & \\ & & m_I \end{pmatrix}$$

usually denoted D, the scalar product $\mathrm{Tr}(\mathscr{W}_1\mathscr{W}_2^*)=\mathrm{Tr}(W_1DW_2D)$ have the following interesting statistical interpretation.

Property 10. *$\mathrm{Tr}(W_1 D W_2 D)$ is the sum of the squared covariances between each variable of X_1 and each variable of X_2. If each column of X_1 and X_2 is standardized to have unit variance, $\mathrm{Tr}(W_1 D W_2 D)$ is the sum of the squared correlations.*

Definition 9. The scalar product between normed W's

$$\frac{\mathrm{Tr}(W_1 D W_2 D)}{\sqrt{\mathrm{Tr}(W_1 D)^2}\sqrt{\mathrm{Tr}(W_2 D)^2}}$$

is known as Rv-coefficient [Robert and Escoufier, 1976].

Remark. Whether S is equal to D with $m_1 = \cdots = m_I$ or equal to the identity matrix, the scalar product

$$\frac{\mathrm{Tr}(W_1 S W_2 S)}{\sqrt{\mathrm{Tr}(W_1 S)^2}\sqrt{\mathrm{Tr}(W_2 S)^2}}$$

has the same value.

References

Carlier, A., Ch. Lavit, M. Pages, M.O. Pernin and J.Ch. Turlot, A comparative review of methods which handle a set of indexed data tables, Coppi and Bolasco, (ed.), in: Multiway Data Analysis, (North-Holland, Amsterdam, 1989) 85–101.

Escoufier, Y., Le traitement des variables vectorielles, *Biometrics*, 29 (1973) 751–760.

Escoufier, Y., Operators related to a data matrix, in: Barra et al. (ed.), *Recent Developments in Statistics*, (North-Holland, Amsterdam, 1977) 125–131.

Fichet, B. and G. Le Calve, Structure géométrique des principaux indices de dissimilarité sur signes de présence–absence, *Statistique et Analyse de Données*, **9** (3) (1984) 11–44.

Franc, A., Etude algébrique des multi-tableaux. Apports de l'algèbre tensorielle, doctoral thesis, University of Montpellier II, 1992.

Lavit, Ch., *Analyse Conjointe de Tableaux Quantitatifs*, (Masson, Paris, 1988).

Rao, C.R., *Linear Statistical Inference and its Applications*, (Wiley, New York, 1973).

Rao, C.R., Matrix approximations and reduction of dimensionality in multivariate statistical analysis, in: Krishnaiah, (ed.), *Multivariate Analysis*, (North-Holland, Amsterdam, 1980) 3–22.

Robert, P. and Y. Escoufier, A unifying tool for linear multivariate statistical methods: the RV coefficient, *Applied Statistics C*, **25** (3) (1976) 257–265.

Sabatier, R., Méthodes factorielles en analyse des données, doctoral thesis, University of Montpellier II, 1987.

Availability of the software

Distributor: CISIA, 1, Avenue Herbillon, 94160 Saint Mandé France. Tel.: (1) 43 74 20 20. Fax: (1) 43 74 17 29

Author: Christine Lavit, Unité de Biométrie, ENSA.M.-INRA-UM II, Place Viala - 34060 Montpellier, France.

Release for IBM-compatible micro-computers: User's guide + executable code for ACT (STATIS method) and ACT (dual STATIS method) + source code written in portable FORTRAN 77 on floppy disks. Minimal hardware required to run the executable code: 512 K RAM.

Other computers: The source code is available for implementing the software on other kind of computers, work stations under UNIX and mainframes as IBM, VAX, UNIVAC...

Cost: 1500 FF.