

# STATISTIQUE ET ANALYSE DES DONNÉES

ROGER LAFOSSE

## **Métriques et analyse factorielle comparative de deux tableaux ou plus**

*Statistique et analyse des données*, tome 11, n° 3 (1986), p. 51-75

[http://www.numdam.org/item?id=SAD\\_1986\\_\\_11\\_3\\_51\\_0](http://www.numdam.org/item?id=SAD_1986__11_3_51_0)

© Association pour la statistique et ses utilisations, 1986, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**METRIQUES ET ANALYSE FACTORIELLE**  
**COMPARATIVE DE DEUX TABLEAUX OU PLUS**

Roger Lafosse  
Laboratoire Modèles et Logiciels d'Analyse des Données  
Université Paul Sabatier  
31062 TOULOUSE CEDEX

Résumé :

*Est proposée une analyse factorielle de la ressemblance et de la différence entre plusieurs tableaux d'individus, à partir de la mise en oeuvre d'un algorithme, que nous décrivons.*

*Quelques apports, concernant les semi-métriques dans un espace somme directe de sous-espaces vectoriels, permettent de présenter cette analyse sur plusieurs tableaux en prolongement d'analyses factorielles connues.*

Abstract

*We propose an analysis of the likeness and of some differences between several tables of cases, by defining principal components built with an algorithm optimizing a likeness criterion.*

*Some contributions, concerning with metrics divided into blocks, permit us to link this analysis with some known analyses.*

Mots clefs : rotation procuste, analyse canonique, métrique, ajustement.

Indices STMA : 06-900    00-60    07-020

Manuscrit reçu le 29 juillet 1986

Révisé le 27 avril 1987

## INTRODUCTION

S'intéresser à plusieurs tableaux c'est le plus souvent se poser d'abord le problème de leur comparaison . Une de ces études comparatives peut s'exprimer en termes de ressemblance entre nuages de points individus associés à ces tableaux. Une analyse de la différence entre ces nuages peut alors se concevoir, en se référant à ce qui a été reconnu comme le plus ressemblant entre ces tableaux, en termes d'ajustements successifs, dans l'esprit de Lingoes J.C. et Borg I. (1978).

Nous nous intéressons ainsi à la recherche de directions principales associées aux tableaux respectifs, sur lesquelles les nuages respectifs projetés apparaissent les plus ressemblants. Pour ce faire, est proposée une généralisation d'un critère de ressemblance et un algorithme permettant d'optimiser ce critère (la convergence de l'algorithme n'est pas établie). Ceci nous permet de construire une rotation de plusieurs tableaux, vers l'un d'eux pris pour cible, à partir d'une superposition d'axes orthogonaux. Cette approche, qui se veut être un prolongement à plus de deux tableaux, d'une présentation possible de la rotation procuste orthogonale entre deux tableaux (Lafosse, 1985), est différente de celle considérée par Gower (1975) et par Berge (1977) qui optimisent la statistique procuste (Sibson, 1978) à l'aide d'algorithmes itératifs, permettant à chaque pas de la diminuer.

L'aspect factoriel est en effet privilégié dans notre démarche, au détriment peut être d'une très bonne diminution de la statistique. (Cependant, sur le seul exemple traité par nous, les résultats obtenus s'avèrent bons du point de vue de l'optimisation de cette statistique). Notre méthode permet de fournir les expressions matricielles des rotations à appliquer aux tableaux vers l'un d'entre eux pris pour cible, le rapprochement ne se faisant pas vers un tableau compromis. De plus, ces expressions sont décomposables en termes dont les importances relatives sont évaluées.

Une deuxième étape dans l'étude de la différence entre tableaux par rapport à l'un d'eux pris pour modèle, consiste à proposer un ajustement global, en procédant par affinités orthogonales, généralisant ainsi pour plusieurs tableaux une approche précédemment considérée pour deux tableaux (Lafosse,1985). Un indice de liaison

globale permettant d'évaluer l'intensité de l'ajustement au tableau cible est alors proposée. C'est pour nous un ajustement de type procustéen en ce sens que la démarche ne peut que diminuer la valeur de la Statistique Procuste et aussi dans la mesure où nous ne voulons pas limiter le cadre procustéen aux seules rotations : le bandit Procuste a bâti sa (mauvaise) réputation en déformant les voyageurs plus qu'en les couchant sur un lit.

Les semi-métriques, considérées par blocs, sur espaces vectoriels somme directe d'espaces vectoriels, ont été introduites pour la première fois en tant que telles par Schektman (1975) qui définit une famille particulière, nommées métriques à effets relationnels. ( A un moment de l'exposé nous considérons l'une des métriques de cette famille). Nous apportons quelques développements quant à la connaissance de cet outil, sans nous référer dans un premier temps à une analyse de données. Notamment, nous proposons une écriture de ces semi-métriques où apparaissent clairement tous les termes des analyses canoniques qu'elles définissent implicitement, ce qui revient à proposer un mode de construction de semi-métriques dont nous usons par la suite. Ces semi-métriques nous aident à aborder la complexité des outils euclidiens que l'on veut mettre en oeuvre quand il s'agit de préciser les dispositions relatives de plus de deux nuages de points : si on modifie celle de deux nuages par rapport à un troisième, il faut contrôler la disposition relative de ces deux nuages, sinon on risque de bâtir des modifications impossibles à réaliser. De plus, ces semi-métriques nous permettent d'établir un lien entre notre analyse et d'autres analyses connues. Elles participent donc à une justification de nos calculs à plusieurs égards. Elles ne sont pas cependant indispensables pour une présentation des résultats.

L'exposé est rédigé en considérant deux, puis seulement trois tableaux, pour simplifier le propos et l'écriture. Rien cependant, ne nous semble s'opposer à une mise en oeuvre de la méthode et de l'algorithme en prenant un nombre de tableaux plus grand.

Notons que si l'algorithme ne convergerait pas vers la solution cherchée, la solution obtenue reste cependant utilisable pour effectuer les ajustements proposés, ces derniers devenant alors sans doute moins efficaces.

## A - METRIQUES ET ANALYSE DE DEUX TABLEAUX

### 1 - PROPRIETES DES METRIQUES DANS $E_1 \oplus E_2$

#### 1.1 - Notations

Les notations prises sont, pour l'essentiel, celles de Cailliez et Pagès (1976). Sont confondues les notations de vecteurs, applications linéaires, etc., et leurs représentations matricielles dans les bases de référence choisies une fois pour toutes. Soient  $E_1 = R^{m_1}$  et  $E_2 = R^{m_2}$ , deux espaces vectoriels sur  $R$ . On note  $E_1^* = O_1 \oplus E_2$  et  $E_2^* = E_1 \oplus O_2$  les sous-espaces vectoriels de l'espace somme directe  $E_1 \oplus E_2$ ,  $O_1$  étant le vecteur nul de  $E_1$  et  $O_2$  celui de  $E_2$ .

#### 1.2 - Rappels

Ces rappels ont été rassemblés ou établis par Tebboune (1981).  $E_1$  et  $E_2$  sont munis des métriques (forme définies positives)  $M_1$  et  $M_2$ . On considère l'écriture matricielle par blocs des semi-métriques (formes semi définies positives) dans  $E_1 \oplus E_2$  :

$$M = \begin{pmatrix} M_1 & M_{12} \\ M_{21} & M_2 \end{pmatrix}$$

$M_{12} = M_{21}'$  (transposée de  $M_{21}$ ) est le bloc inter définissant la disposition relative des sous-espaces  $E_1^*$  et  $E_2^*$ , une fois choisis les blocs intra  $M_1$  et  $M_2$ .

Cette séparation d'information peut être aussi réalisée à l'aide de la décomposition suivante :

$$M = M_D P, \quad M_D = \begin{pmatrix} M_1 & O \\ O & M_2 \end{pmatrix}, \quad P = \begin{pmatrix} I_1 & P_{12} \\ P_{21} & I_2 \end{pmatrix}$$

$I_1$  et  $I_2$  étant matrices identité de dimensions respectives  $m_1$  et  $m_2$ . On a :

$$P_{12} = M_1^{-1} M_{12}, \quad P_{21} = M_2^{-1} M_{21} \quad \text{et} \quad P_{12} = M_1^{-1} P_{21}' M_2$$

$P_{12}$ , application linéaire de  $(E_2, M_2)$  dans  $(E_1, M_1)$ , est donc opérateur adjoint de  $P_{21}$ .

P opérateur de  $E_1 \oplus E_2$ ,  $M_D$ -symétrique et M-symétrique (la M-symétrie s'étend naturellement au cas où M est une semi-métrique et non une métrique), caractérise la disposition des sous-espaces  $E^*_1$  et  $E^*_2$  dans  $E_1 \oplus E_2$  pour  $M_D$  fixée.

$$P = P_{E^*_1} + P_{E^*_2} \text{ avec } P_{E^*_1} = \begin{pmatrix} I_1 & P_{12} \\ 0 & 0 \end{pmatrix} \text{ et } P_{E^*_2} = \begin{pmatrix} 0 & 0 \\ P_{21} & I_2 \end{pmatrix}$$

$P_{E^*_1}$  est le projecteur M orthogonal de  $E_1 \oplus E_2$  dans  $E^*_1$ .

$P_{E^*_2}$  est le projecteur M orthogonal de  $E_1 \oplus E_2$  dans  $E^*_2$ .

M est définie si et seulement si P est régulière.

### 1.3- Analyse canonique dans $(E_1 \oplus E_2, M)$

L'analyse canonique d'un couple de sous-espaces d'un espace euclidien a été étudiée par Dempster (1968), Dauxois, Pousse (1976), Cazes et coll. (1977), Pages, Cailliez, Escoufier (1979), etc..

Cette notion reprise ici, est étendue en ce sens que des métriques intra différentes sont considérées dans  $E_1$  et  $E_2$  (sans référence à une analyse de données).

Soient les décompositions spectrales suivantes,  $P_{12}$  étant l'adjoint de  $P_{21}$  :

$$P_{12} P_{21} = \sum_{i=1}^{m_1} s_i^2 u_i u_i' M_1 \text{ et } P_{21} P_{12} = \sum_{i=1}^{m_2} s_i^2 v_i v_i' M_2$$

Les vecteurs colonnes  $u_j$  sont  $M_1$  orthonormés, les vecteurs colonnes  $v_j$  sont  $M_2$  orthonormés. Le rang de  $P_{12}$  (ou de  $P_{21}$ ) étant égal à r, seules les r premières valeurs propres sont non nulles.

On a ainsi les éléments permettant d'écrire les décompositions en valeurs singulières (où  $s_j > 0, j = 1 \text{ à } r$ )

$$P_{12} = \sum_{j=1}^r s_j u_j v_j' M_2 \text{ et } P_{21} = \sum_{j=1}^r s_j v_j u_j' M_1$$

avec

$$P_{12} v_j = s_j u_j \qquad P_{21} u_j = s_j v_j \qquad j = 1, \dots, r$$

Et P peut donc s'écrire :

$$P = \begin{pmatrix} I_1 & \sum_{j=1}^r s_j u_j v_j' M_2 \\ \sum_{j=1}^r s_j v_j u_j' M_1 & I_2 \end{pmatrix} \quad (1)$$

avec les métriques identités  $I_1 = \sum_{i=1}^{m_1} u_i u_i' M_1$  et  $I_2 = \sum_{i=1}^{m_2} v_i v_i' M_2$

P, qui est somme de deux projecteurs orthogonaux, a ses valeurs propres comprises entre 0 et 2. Donc  $0 < s_j \leq 1$ . L'écriture (1) permet de vérifier aisément que  $s_j$  est le coefficient canonique associé au couple canonique  $(u_j, v_j)$ ,  $j = 1$  à  $r$ .

Nous confondrons désormais les notations  $u_j \in E_1$  et  $u_j \in E^*_1$ , ou  $v_j \in E_2$  et  $v_j \in E^*_2$ . La matrice colonne  $\begin{pmatrix} u \\ v \end{pmatrix}$  sera noté  $u + v$ .

Les vecteurs  $u_j + v_j$ ,  $j = 1$  à  $r$  sont vecteurs propres de P associés aux valeurs propres  $1 + s_j$ , les vecteurs propres  $u_j - v_j$ ,  $j = 1$  à  $r$ , étant associés aux valeurs  $1 - s_j$ , les autres valeurs propres valant 1.

Nous retrouvons donc des propriétés bien connues, souvent établies dans le cadre d'une analyse de deux tableaux mettant en jeu la somme de deux projecteurs orthogonaux sur deux sous-espaces supplémentaires.

Réciproquement, étant donné un système de vecteurs  $M_1$  orthonormés  $\{u_j\}$ ,  $j = 1$  à  $r$ , et celui des vecteurs  $M_2$  orthonormés  $\{v_j\}$ ,  $j = 1$  à  $r$ , la métrique qui permet de préciser la disposition relative de ces deux systèmes, connaissant l'angle  $\alpha_j$  entre  $u_j$  et  $v_j \forall j$ , est la métrique  $M = M_D P$ , avec P écrite selon (1), en posant  $s_j = \cos \alpha_j$ .

Cas particulier :

Les r angles canoniques sont nuls.

$P_{12} = \sum_{j=1}^r u_j v_j' M_2$  est opérateur partiellement isométrique de  $(E_2, M_2)$  dans  $(E_1, M_1)$ .

Le vecteur  $u_j$  est donc alors confondu avec  $v_j$ ,  $\forall j = 1$  à  $r$  dans sa représentation par M dans  $E_1 \oplus E_2$ .

Remarques

Les écritures précédentes sont uniques aux indéterminations près dues à l'existence éventuelle d'un sous-espace propre de dimension supérieure à 1 dans les équations aux valeurs propres. De plus, tout vecteur  $u_j + v_j$  peut être remplacé par le vecteur  $-u_j -v_j$ .

2 - AJUSTEMENT DE TABLEAUX ET METRIQUES

2.1 - Introduction

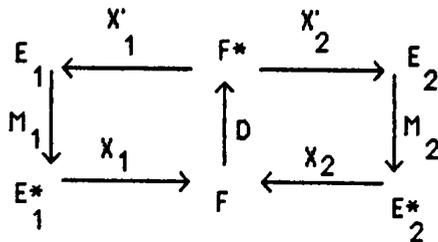
Pour  $k = 1$  et  $2$ ,  
 $(X_k, M_k, D)$  sont les deux triplets soumis à analyse. Les tableaux  $X_k$  sont formés de  $n$  lignes  $x_i^k, i = 1$  à  $n$ , affectées des poids  $p_i$  vérifiant  $\sum_{i=1}^n p_i = 1$ .

$x_i^1$  est apparié à  $x_i^2, \forall i = 1$  à  $n$ .

$E_k$  désigne l'espace vectoriel de dimension  $m_k$  engendré à partir des vecteurs  $x_i^k$ ; les points  $x_i^k, i = 1$  à  $n$ , sont ceux du nuage  $N_k$  dans leurs représentations par  $M_k$ , métrique sur  $E_k$ .

L'espace  $F$ , engendré par les  $m_1 + m_2$  vecteurs colonnes centrés de  $R^n$  formant les tableaux  $X_1$  et  $X_2$ , est muni de la métrique diagonale  $D$  des poids  $p_i$ .

Les vecteurs dans  $F, E_k^*$  et  $E_k$  se correspondent par dualité selon le schéma (Cailliez et Pagès, 1976).



Notons  $V_1 = X'_1 DX_1, V_2 = X'_2 DX_2, V_{12} = X'_1 DX_2$  et  $V_{21} = X'_2 DX_1$ . La

correspondance entre  $(E_1, M_1)$  et  $(E_2, M_2)$  est précisée par l'opérateur  $V_{12} M_2$ , ou l'opérateur adjoint  $V_{21} M_1$ , de rang  $r$ . La décomposition en valeurs singulières des opérateurs

$$V_{12} M_2 = \sum_{j=1}^r \mu_j u_j v_j' M_2 \quad \text{ou} \quad V_{21} M_1 = \sum_{j=1}^r \mu_j v_j u_j' M_1$$

où  $\mu_j^2$  est valeur propre de  $V_{12} M_2 V_{21} M_1$  et de  $V_{21} M_1 V_{12} M_2$ , met en

evidence le fait que cette correspondance entre espaces métriques engendrés, est décrite par celles existant entre les vecteurs  $M_1$  orthonormés  $u_j$  et les vecteurs  $M_2$  orthonormés  $v_j$ ,  $j = 1$  à  $r$  :  $V_{12} M_2 v_j = \mu_j u_j$  et  $V_{21} M_1 u_j = \mu_j v_j$ .

Le plus souvent vouloir associer une semi-métrique  $M$  dans  $E_1 \oplus E_2$  à une analyse de deux tableaux, c'est soit s'intéresser à une disposition relative particulière des deux nuages de points associés, soit, et cela autant par un choix particulier des métriques intra que du bloc inter, pour obtenir des variables ayant des propriétés particulières (notamment celles associables par dualité aux vecteurs canoniques implicitement définis pas  $M$  dans  $E_1 \oplus E_2$ ).

Dans ce contexte de métrique associée à une analyse, seuls des tableaux où les individus sont appareillés sont en général considérés.

Des choix de semi-métriques (les plus intéressants pour nous) sont rappelés ci-dessous en précisant la nature des vecteurs canoniques associés.

## 2.2- Quelques choix de métriques

### Rotation Procuste orthogonale

$$M = \begin{pmatrix} M_1 & M_1 V_{12} M_2 (V_{21} M_1 V_{12} M_2)^{-1/2} \\ M_2 V_{21} M_1 (V_{12} M_2 V_{21} M_1)^{-1/2} & M_2 \end{pmatrix}$$

est la métrique procuste orthogonale écrite par Croquette (1980), étendant la notion de rotation à celle d'isométrie partielle entre espaces métriques. Le résultat de la rotation apparaît clairement indépendant de la cible choisie dans cette écriture, où les rôles joués par les deux sous-espaces sont symétriques.

En l'écrivant sous la forme :

$$M = M_D \begin{pmatrix} I_1 & \sum_{j=1}^r u_j v_j' M_2 \\ \sum_{j=1}^r v_j u_j' M_1 & I_2 \end{pmatrix} = M_D P$$

on met en évidence les deux systèmes d'axes orthogonaux superposés dans la rotation (ceux de la décomposition en valeurs singulières de  $V_{12}M_2$  ou  $V_{21}M_1$ ). Notons que les valeurs propres de  $P$  sont égales à 2 ou 1.

### Ajustement procustéen

On ajuste un des deux nuages à l'autre pris pour cible, en le déformant :  $r$  affinités orthogonales de centre  $O$  et de rapport  $k_j$  sont effectuées dans chacune des  $r$  directions orthogonales superposées par rotation procuste orthogonale (Lafosse, 1985). En complétant le système des  $r$  vecteurs  $v_j$  pour obtenir une base orthonormée de  $E_2$ , on peut écrire,

$$M_2 = M_2 I_2 = M_2 \sum_{j=1}^{m_2} v_j v_j' M_2$$

La métrique traduisant la nouvelle configuration du nuage  $N_2$  s'écrit alors :

$$A_2 = M_2 \sum_{j=1}^{m_2} k_j^2 v_j v_j' M_2$$

avec  $k_j = 1$  pour  $r < j \leq m_2$ .

Le système des vecteurs  $b_j = v_j / k_j$  est  $A_2$  orthonormé et l'opérateur

$A_{12} = \sum_{j=1}^r u_j v_j' A_2 = \sum_{j=1}^r k_j u_j v_j' M_2$ , appliqué au nuage  $N_2$ , permet d'observer dans  $(E_1, M_1)$  le

résultat de l'ajustement au nuage  $N_1$ .

Dans  $E_1 \oplus E_2$ , la métrique exprimant le résultat de cet ajustement est :

$$N = \begin{pmatrix} M_1 & 0 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} I_1 & \sum_{j=1}^r u_j b'_j A_2 \\ \sum_{j=1}^r b_j u'_j M_1 & I_2 \end{pmatrix}$$

Dans le cas où toutes les valeurs  $k_j$  sont inférieures ou égales à 1, il est possible de réaliser simultanément la superposition, par rotations et par contractions

avec l'opérateur  $A_{12} = \sum_{j=1}^r k_j u_j v'_j M_2$ , de sorte que la disposition relative

des deux nuages décrite par  $N$  soit aussi décrite par :

$$M_D P = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} I_1 & P_{12} \\ P_{21} & I_2 \end{pmatrix}, \text{ où } P_{12} = A_{12}.$$

C'est ce qui se passe (cf. [11]) dans le cas particulier où  $M_1$  et  $M_2$  sont métriques de Mahalanobis dans  $E_1$  et  $E_2$  respectivement, avec la métrique à effet relationnel (Schektman, 1978) :

$$M_D P = \begin{pmatrix} V_1^{-1} & V_1^{-1} V_{12} V_2^{-1} \\ V_2^{-1} V_{21} V_1^{-1} & V_2^{-1} \end{pmatrix}$$

Dans  $E_1 \times E_2$ , l'analyse canonique de  $E_1$  et  $E_2$  induite par cette métrique admet pour vecteurs canoniques et pour coefficients canoniques les axes canoniques et les coefficients de corrélation canoniques de l'analyse canonique classique de deux tableaux.

Toutes ces métriques correspondent à un ajustement entre tableaux.

## B - METRIQUES ET ANALYSE DE TROIS TABLEAUX

### 1 - SEMI METRIQUE DANS $E = E_1 \oplus E_2 \oplus E_3$ . GENERALITES

Nous considérons ici trois espaces vectoriels  $E_1, E_2, E_3$  sur  $R$ , de dimensions respectives  $m_1, m_2, m_3$ , avec  $m = m_1 + m_2 + m_3$  ; la plupart des notions introduites pourraient être prolongées à plus de trois espaces sans difficulté. Les notations adoptées viennent à la suite de celles définies pour les semi métriques dans  $E_1 \oplus E_2$ .

Soit l'écriture par blocs d'une semi métrique dans  $E_1 \oplus E_2 \oplus E_3$

$$M = \begin{pmatrix} M_1 & M_{12} & M_{13} \\ M_{21} & M_2 & M_{23} \\ M_{31} & M_{32} & M_3 \end{pmatrix}$$

Les sous-matrices de  $M$ ,

$$\begin{pmatrix} M_1 & M_{12} \\ M_{21} & M_2 \end{pmatrix} \quad \begin{pmatrix} M_1 & M_{13} \\ M_{31} & M_3 \end{pmatrix} \quad \begin{pmatrix} M_2 & M_{23} \\ M_{32} & M_3 \end{pmatrix}$$

sont des semi-métriques définissant les dispositions relatives des sous-espaces considérés deux à deux. (cette condition nécessaire n'est pas suffisante pour conclure que  $M$  est une semi-métrique).

$$\text{Posons } M = M_D P \quad , \quad M_D = \begin{pmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{pmatrix} \quad , \quad P = \begin{pmatrix} I_1 & P_{12} & P_{13} \\ P_{21} & I_2 & P_{23} \\ P_{31} & P_{32} & I_3 \end{pmatrix}$$

$M_D$  est supposée définie positive.

$P_{12}, P_{13}, P_{23}$  sont opérateurs adjoints de  $P_{21}, P_{31}, P_{32}$  respectivement, sachant que pour tout  $i \neq j$ ,  $P_{ij}$  est application de  $(E_j, M_j)$  dans  $(E_i, M_i)$ .

En posant :

$$P_{E_1} = \begin{pmatrix} I_1 & P_{12} & P_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad P_{E_2} = \begin{pmatrix} 0 & 0 & 0 \\ P_{21} & I_2 & P_{23} \\ 0 & 0 & 0 \end{pmatrix} \quad P_{E_3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ P_{31} & P_{32} & I_3 \end{pmatrix}$$

$P = P_{E_1} + P_{E_2} + P_{E_3}$  est la somme de trois projecteurs  $M$  orthogonaux sur  $E_1 \times \{O_2\} \times \{O_3\}$ ,  $\{O_1\} \times E_2 \times \{O_3\}$ ,  $\{O_1\} \times \{O_2\} \times E_3$ , notés  $E_1^*$ ,  $E_2^*$ ,  $E_3^*$ , respectivement.  $P$  a donc ses valeurs propres comprises entre 0 et 3.

Soit la décomposition spectrale usant de la  $M_D$  symétrie de  $P$  (qui est aussi  $M$  symétrique) :

$$P = \sum_{j=1}^m \lambda_j Y_j Y_j' M_D$$

où  $Y_j = x_j + y_j + z_j$  est vecteur propre de  $P$ , de  $M_D$  norme 1, associé à la valeur propre  $\lambda_j$  qui lui correspond ;  $x_j \in E_1^*$ ,  $y_j \in E_2^*$  et  $z_j \in E_3^*$ . On a :

$$P_{E_1^*}(Y_j) = \lambda_j x_j \quad P_{E_2^*}(Y_j) = \lambda_j y_j \quad \text{et} \quad P_{E_3^*}(Y_j) = \lambda_j z_j$$

Soient  $\alpha_j, \beta_j, \gamma_j$  les cosinus des  $M$  angles aigus respectifs des vecteurs  $x_j, y_j, z_j$  avec leur vecteur somme  $Y_j$ . On note la norme d'un vecteur quelconque de  $E_k$  ou  $E_k^*$ ,

$k = 1, 2$  ou  $3$  :  $\| \cdot \| = \| \cdot \|_M = \| \cdot \|_{M_D} = \| \cdot \|_{M_k}$ . On a :  $\alpha_j = \frac{\sqrt{\lambda_j} \|x_j\|}{\|Y_j\|}$ , car  $\sqrt{\lambda_j}$  est la

$M$  norme de  $Y_j$ . Comme  $\|Y_j\|_{M_D} = 1$ ,  $\lambda_j = \alpha_j^2 + \beta_j^2 + \gamma_j^2$ .

Notons  $u_j, v_j, w_j$  les vecteurs  $x_j, y_j, z_j$  respectivement normés à l'unité. On peut écrire :

$$P = \sum_{j=1}^m \begin{pmatrix} \alpha_j & u_j \\ \beta_j & v_j \\ \gamma_j & w_j \end{pmatrix} (\alpha_j u_j' \quad \beta_j v_j' \quad \gamma_j w_j') M_D$$

contrairement à ce qui se passe dans une écriture analogue relative à deux sous-espaces seulement, aucune relation d'orthogonalité n'existe entre les vecteurs  $u_j$ , ou entre les vecteurs  $v_j$ , ou entre les vecteurs  $w_j$ .

Il n'est plus possible ici de changer un angle, pour définir une nouvelle disposition relative des sous-espaces, sans tenir compte des liens existant entre l'ensemble de tous les vecteurs, tels que :

$$\sum_{i=1}^m \alpha_i^2 = m_1 \quad \sum_{i=1}^m \beta_i^2 = m_2 \quad \sum_{i=1}^m \gamma_i^2 = m_3 \quad \sum_{i=1}^m \alpha_i^2 (u'_i M_1 u_k)^2 = 1, \quad \forall k = 1 \text{ à } m$$

Comme  $\lambda_j = \alpha_j^2 + \beta_j^2 + \gamma_j^2$ ,  $\forall j = 1 \text{ à } m$ , l'analyse canonique généralisée des sous-espaces  $E_1^*$ ,  $E_2^*$  et  $E_3^*$  dans  $(E_1 \oplus E_2 \oplus E_3, M)$  a pour solution, au sens de Carroll (1968), les triplets  $(x_j, y_j, z_j)$ , le classement des vecteurs  $x_j + y_j + z_j$  se faisant selon les valeurs propres de P rangées dans l'ordre décroissant.

## 2 - AJUSTEMENT DE TROIS TABLEAUX

### 2.1 - Introduction

Nous venons de voir que la difficulté pour choisir une disposition relative de trois sous-espaces ou de trois nuages de points provenait de l'outil euclidien mis en oeuvre, les choix des angles ne pouvant se traiter de façon indépendante comme pour deux tableaux de variables centrées.

Nous proposons cependant une solution au problème de l'ajustement global de deux tableaux vers un troisième pris pour cible, par rotations ou par rotations et affinités orthogonales. L'aspect global provient de ce que la méthode consiste à déterminer d'abord de façon globale des directions principales à superposer.

Pour alléger l'écriture, l'analyse des trois triplets  $(X_k, M_k, D)$ ,  $k = 1 \text{ à } 3$ , pouvant être remplacée par celle des triplets  $(X_k, M^{1/2}_k, D)$ , nous travaillons souvent avec les métriques identités sur  $E_1, E_2$  et  $E_3$  sans perte de généralité. Mais présentons d'abord un cas particulier où une solution est connue.

### 2.2 - Ajustement global de nuages normalisés

Dans le cas où les métriques intra relatives aux trois espaces engendrés à partir des trois tableaux sont celles respectives de Mahalanobis, nous avons une solution

au problème de l'ajustement global par rotations et affinités orthogonales des trois nuages. En effet, la procédure d'ajustement, vue pour deux nuages avec la métrique à effets relationnels, peut être appliquée simultanément aux trois nuages considérés deux par deux, en aboutissant à une disposition globale compatible entre les trois nuages, comme le montre l'écriture de la métrique à effet relationnel :

$$M = M_D P = \begin{pmatrix} Y_1^{-1} & 0 & 0 \\ 0 & Y_2^{-1} & 0 \\ 0 & 0 & Y_3^{-1} \end{pmatrix} \begin{pmatrix} I_1 & Y_{12} Y_2^{-1} & Y_{13} Y_3^{-1} \\ Y_{21} Y_1^{-1} & I_2 & Y_{23} Y_3^{-1} \\ Y_{31} Y_1^{-1} & Y_{32} Y_2^{-1} & I_3 \end{pmatrix}$$

Les trois nuages étant normalisés par l'emploi des métriques de Mahalanobis [14] (les ellipsoïdes d'inertie sont alors hypersphériques), l'ajustement ainsi réalisé est particulièrement efficace. Aussi, les premières valeurs propres de P sont-elles souvent proches de 3.

Les analyses associées à cette métrique, notamment analyse des correspondances, analyse discriminante, sont largement développées, le plus souvent à partir de l'analyse canonique généralisée au sens de Carroll (1968), Saporta (1975), c'est-à-dire à partir du calcul des premiers vecteurs propres de P. (voir par exemple Casin et Turlot, 1986).

### 2.3 - Critère de ressemblance maximum

Cette notion de ressemblance a été définie (Lafosse, 1985) comme le résultat d'un compromis recherché, en trouvant un sous-espace commun aux espaces  $E_1$  et  $E_2$ , entre une inertie maximum des sous-nuages projetés dans ce sous-espace et une proximité optimisée pour ces sous-nuages entre points appareillés. La démarche utilisée consistait à définir pas à pas des couples  $(u, v)$ ,  $u \in E_1$ ,  $v \in E_2$ , tels que soit atteint

$$\max u' V_{12} v \quad \|u\| = \|v\| = 1$$

critère à rapprocher de l'expression de la statistique Procuste

$$u' V_1 u + v' V_2 v - 2 u' V_{12} v$$

qui représente la distance, entre points appareilles, du nuage issu de  $X_1$ , projete sur l'axe de support  $u$ , et du nuage issu de  $X_2$  projeté sur l'axe de support  $v$ , quand  $u$  et  $v$  sont superposés.

Nous remarquons que relativement à trois vecteurs  $u, v$  et  $w$  superposés issus des trois tableaux  $X_1, X_2$  et  $X_3$ , la statistique procuste entre les nuages respectifs projetés

$$\text{vaut : } 2 (u' V_1 u + v' V_2 v + w' V_3 w - u' V_{12} v - u' V_{13} w - v' V_{23} w)$$

et qu'une extension naturelle a trois tableaux du critère de ressemblance entre trois nuages projetés sur un même axe obtenu par superposition de trois vecteurs  $u \in E_1$ ,  $v \in E_2$  et  $w \in E_3$ , peut être posée égale à :

$$\alpha = u' V_{12} v + v' V_{23} w + w' V_{31} u \quad (1)$$

avec  $\|u\| = \|v\| = \|w\| = 1$ .

#### 2.4 - Algorithme de definition d'un triplet

Soit donc le problème de la recherche d'un triplet  $(u, v, w)$  optimisant  $\alpha$ . Ce problème semble n'avoir pour solution actuellement que celles construites à partir d'algorithmes itératifs, où la convergence vers la solution ne serait pas prouvée (Takeuchi, Yanai, Mukherjee, 1982 ou Yanai 1985). L'algorithme proposé ci-dessous s'est avéré expérimentalement convergent et seule la comparaison du résultat obtenu avec ceux fournis par d'autres approches, nous a persuadé de l'intérêt possible de cette technique, basée sur la remarque suivante :

$$\max_{u, v, w} (u' V_{12} v + v' V_{23} w + w' V_{31} u), \text{ avec } \|u\| = \|v\| = \|w\| = 1$$

$$\Leftrightarrow \max_{x, y, z} (x' \frac{V_{12}}{a} y + y' \frac{V_{23}}{b} z + z' \frac{V_{31}}{c} x)$$

$$\text{avec } a = \|x\| \|y\|, \quad b = \|y\| \|z\|, \quad c = \|z\| \|x\| \quad (2)$$

### Description de l'algorithme

Les tableaux  $X_1, X_2$  et  $X_3$  doivent avoir la même inertie (sinon la divergence peut se produire).

Nous calculons le vecteur propre associé à la plus grande valeur propre  $\mu_1$  de la matrice blocs

$$\begin{pmatrix} 0 & V_{12} & V_{13} \\ V_{21} & 0 & V_{23} \\ V_{31} & V_{32} & 0 \end{pmatrix}$$

D'où le vecteur  $x_1 + y_1 + z_1$  et le calcul des valeurs  $a_1, b_1$  et  $c_1$  selon (2).

Puis nous calculons le vecteur propre associé à la plus grande valeur  $\mu_2$  de la matrice

$$\begin{pmatrix} 0 & V_{12}/a_1 & V_{13}/c_1 \\ V_{21}/a_1 & 0 & V_{23}/b_1 \\ V_{31}/c_1 & V_{32}/b_1 & 0 \end{pmatrix}$$

D'où un nouveau vecteur  $x_2 + y_2 + z_2$  et la mise à jour des valeurs  $a_2, b_2$  et  $c_2$  selon (2).

L'étape  $i+1$  consiste à renouveler la deuxième étape avec le triplet  $(a_i, b_i, c_i)$  obtenu à la  $i^{\text{ème}}$  étape, en calculant le vecteur propre associé à la plus grande valeur propre  $\mu_{i+1}$ .

Posons  $r_i = x_i / \|x_i\|$ ,  $s_i = y_i / \|y_i\|$ ,  $t_i = z_i / \|z_i\|$

et  $\alpha_i = r'_i V_{12} s_i + s'_i V_{23} t_i + t'_i V_{31} r_i$ .

L'arrêt de la procédure est décidé à partir du moment où  $|\alpha_i - \mu_i / 2|$  est suffisamment proche de zéro. Soit  $n$  cette dernière étape.  $\alpha_n$  est la valeur optimale proposée :

$$r'_n V_{12} s_n + r'_n V_{13} t_n = \mu_n x'_n x_n$$

$$s'_n V_{21} r_n + s'_n V_{23} t_n = \mu_n y'_n y_n$$

$$t'_n V_{31} r_n + t'_n V_{32} s_n = \mu_n z'_n z_n$$

et donc  $r'_n V_{12} s_n + s'_n V_{23} t_n + t'_n V_{31} s_n = \mu_n / 2 \equiv \alpha$  car les vecteurs propres  $x_i + y_i + z_i$  sont normés,  $\forall i$ .

## 2.5 - Définition de triplets successifs et rotation

Posons  $u_1 = r_n$ ,  $v_1 = s_n$  et  $w_1 = t_n$ .  $(u_1, v_1, w_1)$  est donc le premier triplet obtenu sur lesquels les trois nuages respectifs associés projetés apparaissent les plus ressemblants selon notre critère. Dans le cas de deux tableaux, les couples successifs  $(u_j, v_j)$ ,  $(u_j$  superposé à  $v_j$ ,  $\forall j = 1$  à  $r$ , réalisant la rotation procuste orthogonale) sont tels que

$$u_j \perp u_i \text{ et } v_j \perp v_i, \quad \forall i = 1 \text{ à } r, \quad \forall j = 1 \text{ à } r, \quad i \neq j$$

Nous imposons donc, pour définir les couples successifs, les contraintes supplémentaires:

$$u_j \perp u_i, \quad v_j \perp v_i \text{ et } w_j \perp w_i, \quad \forall i, \forall j, \quad i \neq j.$$

Une façon de réaliser cet objectif est la suivante : le calcul du deuxième triplet  $(u_2, v_2, w_2)$  se fera en appliquant l'algorithme aux tableaux

$X_1 - X_1 u_1 u'_1$ ,  $X_2 - X_2 v_1 v'_1$  et  $X_3 - X_3 w_1 w'_1$  après avoir ramené ceux-ci à un même moment d'inertie par rapport à O (sinon l'algorithme peut ne pas converger). Le troisième triplet  $(u_3, v_3, w_3)$  se calcule à partir des tableaux :

$$X_1 - X_1 \sum_{j=1}^2 u_j u'_j, \quad X_2 - X_2 \sum_{j=1}^2 v_j v'_j \text{ et } X_3 - X_3 \sum_{j=1}^2 w_j w'_j \quad \text{etc.}$$

jusqu'à l'obtention de  $k$  triplets ainsi définissables et tels que :

$$u'_j V_{12} v_j > 0, \quad v'_j V_{23} w_j > 0, \quad w'_j V_{31} u_j > 0, \quad \forall j = 1 \text{ à } k$$

Dans la recherche d'un  $k+1$ ème triplet, un des trois termes apparaît négatif : les trois systèmes de vecteurs orthonormés superposables  $\{u_j\}$ ,  $\{v_j\}$  et  $\{w_j\}$ ,  $j = 1$  à  $k$  ne peuvent être complétés à l'aide de nouveaux vecteurs qui seraient également superposables :  $k$  apparaît ainsi comme dimension maximale du sous-espace [

commun pouvant être formé dans ces rotations orthogonales.

Soient  $X_1 - X_1 \sum_{j=1}^k u_j v'_j$ ,  $X_2 - X_2 \sum_{j=1}^k v_j v'_j$  et  $X_3 - X_3 \sum_{j=1}^k w_j w'_j$  les trois tableaux obtenus à

l'issu du calcul des  $k$  triplets. Les espaces respectifs engendrés à partir de ces tableaux sont orthogonaux à  $\mathbb{C}$ , mais non forcément orthogonaux deux à deux : les rotations procustes orthogonales considérées sur ces tableaux pris deux par deux, permettent de superposer les deux vecteurs des couples ainsi définis. On obtient donc :

$k_{12}$  nouveaux vecteurs  $u_{12j}$  superposables à  $k_{12}$  vecteurs  $v_{12j}$

$k_{13}$  nouveaux vecteurs  $u_{13j}$  superposables à  $k_{13}$  vecteurs  $w_{13j}$

et  $k_{23}$  nouveaux vecteurs  $v_{23j}$  superposables à  $k_{23}$  vecteurs  $w_{23j}$ .

Finalement sont obtenus des systèmes de vecteurs orthonormés, superposables trois à trois ou deux à deux.

Posons :

$$P_{12} = \sum_{j=1}^k u_j v'_j + \sum_{j=k+1}^{k+k_{12}} u_{12j} v'_{12j}$$

$$P_{13} = \sum_{j=1}^k u_j w'_j + \sum_{j=k+1}^{k+k_{13}} u_{13j} w'_{13j}$$

$$P_{23} = \sum_{j=1}^k v_j w'_j + \sum_{j=k+1}^{k+k_{23}} v_{23j} w'_{23j}$$

La métrique, disposant les trois nuages de sorte que toutes les superpositions soient effectuées, peut s'écrire :

$$M = \begin{pmatrix} I_1 & P_{12} & P_{13} \\ P'_{12} & I_2 & P_{23} \\ P'_{13} & P'_{23} & I_3 \end{pmatrix}$$

Les valeurs propres de cette matrice sont égales à 1, 2 ou 3. Cette semi-métrique peut être donnée comme une solution au problème des rotations procustes orthogonales, où la recherche de composantes principales orthogonales successives, a pris le pas

sur celle de la minimisation de la statistique procuste considérée globalement sur les nuages. Ainsi, les plans définissables dans  $E_1, E_2$  ou  $E_3$ , à l'aide des premiers vecteurs respectifs donnent, une fois la rotation effectuée, une représentation des trois nuages respectifs qui y sont projetés (la même dans  $E_1, E_2$  ou  $E_3$ , s'il n'existe pas de sous-espaces orthogonaux à  $\tilde{L}$ ), constituant un résumé intéressant pour une observation simultanée. Il s'agit d'une observation de la ressemblance permettant donc en particulier, de repérer éventuellement quelques individus porteurs d'une différence importante par rapport à l'ensemble des autres individus.

Cet ajustement par rotation peut servir de support à un ajustement plus poussé, obtenu par déformation de deux nuages vers un troisième pris pour cible, permettant une meilleure juxtaposition entre points appareillés, selon le critère de la statistique procuste. Cette démarche peut par exemple, être alors conçue dans le but d'une comparaison entre trois codages, comme cela est abordé en [12]. C'est un tel ajustement qui est considéré par la suite.

## 2.6 - Ajustement par affinités orthogonales

La procédure consistant à effectuer des affinités orthogonales relativement aux directions superposables, est étendue ici à trois tableaux,  $X_1$  étant le tableau de cible. Calculons les deux coefficients permettant le rapprochement entre les trois nuages projetés sur un même axe par minimisation de la statistique procuste.

Notons  $\{x_{i1}\}, \{x_{i2}\}$  et  $\{x_{i3}\}$ ,  $i = 1$  à  $n$ , les coordonnées des points des trois nuages sur une des directions communes obtenue par superposition de trois vecteurs  $u, v$  et  $w$ .

$$\frac{1}{2} \min_{k_2, k_3} \sum_{i=1}^n p_i (x_{i1} - k_2 x_{i2})^2 + \sum_{i=1}^n p_i (x_{i1} - k_3 x_{i3})^2 + \sum_{i=1}^n p_i (k_2 x_{i2} - k_3 x_{i3})^2$$

$$\Leftrightarrow \min \text{var } \xi + k_2^2 \text{var } \eta + k_3^2 \text{var } \psi - k_2 \text{cov}(\xi, \eta) - k_3 \text{cov}(\xi, \psi) - k_2 k_3 \text{cov}(\eta, \psi)$$

en posant :  $\xi = X_1 u, \eta = X_2 v$  et  $\psi = X_3 w$ .

D'où

$$k_2 = \frac{2 \operatorname{var} \psi \operatorname{cov}(\xi, \eta) + \operatorname{cov}(\eta, \psi) \operatorname{cov}(\xi, \psi)}{4 \operatorname{var} \psi \operatorname{var} \eta - \operatorname{cov}^2(\eta, \psi)}$$

$$k_3 = \frac{2 \operatorname{var} \eta \operatorname{cov}(\xi, \psi) + \operatorname{cov}(\eta, \psi) \operatorname{cov}(\xi, \eta)}{4 \operatorname{var} \psi \operatorname{var} \eta - \operatorname{cov}^2(\eta, \psi)}$$

Soit, en écrivant les coefficients de corrélations linéaires

$$\rho_{12} = \rho(\xi, \eta), \quad \rho_{13} = \rho(\xi, \psi), \quad \rho_{23} = \rho(\eta, \psi)$$

$$k_2 = \frac{\sqrt{\operatorname{var} \xi}}{\sqrt{\operatorname{var} \eta}} \frac{2 \rho_{12} + \rho_{13} \rho_{23}}{4 - \rho_{23}^2} \quad \text{et} \quad k_3 = \frac{\sqrt{\operatorname{var} \xi}}{\sqrt{\operatorname{var} \psi}} \frac{2 \rho_{13} + \rho_{12} \rho_{23}}{4 - \rho_{23}^2}$$

Le minimum vaut :

$$a_j^2 = \operatorname{var} \xi_j \left( 1 - \frac{\rho_{12j}^2 + \rho_{13j}^2 + \rho_{12j} \rho_{13j} \rho_{23j}}{4 - \rho_{23j}^2} \right)$$

relativement à un triplet  $(u_j, v_j, w_j)$  précédemment défini,  $j = 1$  à  $k$ .

Quant aux supplémentaires orthogonaux, considérés deux à deux, s'ils existent, ils correspondent à un seul coefficient (de régression linéaire simple), permettant d'ajuster le nuage  $N_2$  projeté sur  $v_{12j}$  (ou le nuage  $N_3$  projeté sur  $w_{13j}$ ), à  $N_1$  projeté sur le vecteur  $u_{12j}$  (ou  $u_{13j}$ ) correspondant. Ils sont donc du type

$$k_{12} = \frac{\sqrt{\operatorname{var} \xi}}{\sqrt{\operatorname{var} \eta}} \rho(\xi, \eta) \quad \text{ou} \quad k_{13} = \frac{\sqrt{\operatorname{var} \xi}}{\sqrt{\operatorname{var} \psi}} \rho(\xi, \psi)$$

avec des minima correspondants de la forme

$$\operatorname{var} \xi (1 - \rho^2(\xi, \eta)) \quad \text{ou} \quad \operatorname{var} \xi (1 - \rho^2(\xi, \psi))$$

Notons  $k_{2j}, j = 1, \dots, h_2$  et  $k_{3j}, j = 1, \dots, h_3$  tous ces coefficients, en posant  $h_2 = k + k_{12}$  et  $h_3 = k + k_{13}$ . Soient les métriques (on complète les 2 systèmes orthonormés considérés de façon à former deux bases orthonormées de  $E_2$  et de  $E_3$  respectivement)

$$A_2 = \sum_{j=1}^{m_2} k_{2j}^2 v_j v_j' \quad A_3 = \sum_{j=1}^{m_3} k_{3j}^2 w_j w_j'$$

où  $v_j = v_{12j}$  pour  $j = k+1$  à  $h_2$   
 $w_j = w_{13j}$  pour  $j = k+1$  à  $h_3$   
 $k_{2j} = 1$  pour  $j = k_{2+1}$  à  $m_2$  et  $k_{3j} = 1$  pour  $j = h_3 + 1$  à  $m_3$ .

Posons

$$b_j = v_j / k_{2j}, \quad j = 1 \text{ à } m_2$$

$$c_j = w_j / k_{3j}, \quad j = 1 \text{ à } m_3$$

Posons

$$u_j = u_{12j} \text{ pour } j = k+1 \text{ à } h_2$$

$$u^*_j = u_j \text{ pour } j = 1 \text{ à } k \text{ et } u^*_j = u_{13j} \text{ pour } j = k+1 \text{ à } h_3.$$

La métrique décrivant la disposition des trois nuages après ajustement global à  $X_1$  par rotations et affinités orthogonales, s'écrit donc :

$$N = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{pmatrix} \begin{vmatrix} I_1 & \sum_{j=1}^{h_2} u_j b'_j A_2 & \sum_{j=1}^{h_3} u^*_j c'_j A_3 \\ \sum_{j=1}^{h_2} b_j u'_j & I_2 & \sum_{j=1}^k b_j c'_j A_3 \\ \sum_{j=1}^{h_3} c_j u^*_j & \sum_{j=1}^k c_j b'_j A_2 & I_3 \end{vmatrix}$$

**Indice global de mesure de l'intensité de l'ajustement :**

Bien que faisant jouer un rôle particulier au tableau  $X_1$ , cet ajustement est global.

Posons :

$$R_j^2 = \frac{\rho_{12j}^2 + \rho_{13j}^2 + \rho_{12j} \rho_{13j} \rho_{23j}}{4 - \rho_{23j}^2}, \text{ pour } j = 1 \text{ à } k$$

Soient  $R_j = \rho(\xi_j, \eta_j)$ , avec  $\xi_j = X_1 u_{12j}$  et  $\eta_j = X_2 v_{12j}$  pour  $j = k+1$  à  $h_2$ , et soient  $R_j = \rho(\xi_j, \psi_j)$ , avec  $\xi_j = X_1 u_{13j}$  et  $\psi_j = X_3 w_{13j}$  pour  $j = k-h_2+1$  à  $h$  ( $h-h_2=k_{13}$ ,  $h \leq m_1$ ) les deux familles de coefficients de corrélations déterminés dans les ajustements relatifs à deux directions superposées.

En référence à la forme prise par l'indice de Lingoes et Schönemann entre deux

tableaux, ajustés avec une méthode analogue (Lafosse, 1985), nous proposons pour indice de mesure d'intensité de l'ajustement obtenu de  $X_2$  et  $X_3$  à  $X_1$  :

$$I = \frac{\sum_{j=1}^h R_j^2 \text{var } \xi_j}{\text{tr } V_1}$$

### 3 - RESULTATS DE LA ROTATION SUR UN EXEMPLE

Les trois tableaux considérés sont ceux fournis dans l'exemple donné en démonstration du macro genproc de la bibliothèque Nag. Nous avons centré ces tableaux, puis nous les avons ramené à la même inertie 1. Le premier tableaux  $X_1$  est pris pour cible et  $X_2$  et  $X_3$  sont transformés par les rotations  $P_{12}$  et  $P_{13}$  respectivement (cf. chapitre 2.5). Les trois tableaux, qui ont chacun 7 variables et 9 individus, permettent de définir 6 triplets  $(u_j, v_j, w_j)$ ,  $j = 1$  à 6, associés aux 6 mesures de ressemblance successives  $\alpha_j$  :

$$\begin{aligned} \alpha_1 &= 1.6322, & \alpha_2 &= .2121, & \alpha_3 &= .1428, & \alpha_4 &= .0571 \\ \alpha_5 &= .0282, & \alpha_6 &= .0079 \end{aligned}$$

Les mesures associées aux couples éventuels relatifs à des espaces orthogonaux, sont ici négligeables. Les résultats sont les suivants :

a) Initialement

$$\text{tr}(V_1 + V_2 + V_3) = 3$$

et la statistique procuste  $SP(p_i = \frac{1}{9}, \forall i)$

$$SP = \sum_{i=1}^9 p_i (x_{i_1} - x_{i_2})^2 + \sum_{i=1}^9 p_i (x_{i_2} - x_{i_3})^2 + \sum_{i=1}^9 p_i (x_{i_3} - x_{i_1})^2$$

vaut ici :

$$\frac{1}{2} SP = \text{tr}(V_1 + V_2 + V_3) - \text{tr}(V_{12} + V_{13} + V_{23}) = 1.3705$$

b) après rotation :

$$P_{12} = \sum_{j=1}^6 u_j v_j \quad P_{13} = \sum_{j=1}^6 u_j w_j$$

les nouveaux tableaux sont alors  $X_1, X_2 P_{12}$  et  $X_3 P_{13}$ .

Les inerties des tableaux transformés se sont légèrement modifiées (résultat de la dimension négligée et peut être aussi d'une imprécision des calculs effectués) :

inertie totale : 2.9922

$$\frac{1}{2} SP = .9119$$

Cette valeur peut être comparée à la valeur impossible à atteindre, qui aurait été obtenue si les trois rotations procustes effectuées sur les tableaux pris deux par deux pouvaient aboutir à une disposition finale des trois tableaux réalisable.

$$\frac{1}{2} SP = \frac{3 - \text{tr}(V_{12} V_{21})^{1/2} - \text{tr}(V_{13} V_{31})^{1/2} - \text{tr}(V_{23} V_{32})^{1/2}}{2} = .8889$$

Si, avant le calcul de la statistique procuste, on avait multiplié par des scalaires les tableaux  $P_{12} X_2$  et  $P_{13} X_3$  pour leur redonner une inertie égale à 1, on aurait obtenu la valeur .9141, au lieu de la valeur .9119.

On donne ci-dessous les expressions de  $P_{12}$  et  $P_{13}$

0.5086	-0.1931	-0.3565	0.3517	0.1903	0.1622	-0.4694
-0.5069	0.1552	0.2093	0.6119	0.1917	-0.4193	-0.2673
-0.3467	0.0436	-0.3051	-0.5485	0.5971	-0.0237	-0.3556
0.2316	-0.2150	0.7354	-0.0105	0.5596	0.2061	0.0594
0.1836	0.3298	-0.2384	0.3531	0.2842	0.1787	0.0262
-0.1228	0.3852	-0.2043	0.1361	0.3104	0.2877	0.6091
-0.3141	-0.7942	-0.2770	0.2249	0.1245	0.0871	0.3110

0.2207	0.1830	-0.4786	0.3349	-0.2354	0.5922	0.0013
-0.0374	0.5093	0.6738	-0.1112	0.0187	0.3807	0.3332
0.6544	0.3377	-0.1940	-0.3617	0.5217	-0.0799	-0.1018
-0.5887	0.0533	-0.2359	0.0866	0.7208	0.2589	0.0273
0.2324	0.1036	0.0863	0.5738	0.0656	0.1756	-0.0289
0.2662	-0.3352	0.3669	0.5316	0.3735	-0.1828	-0.0188
0.2226	-0.6707	0.0390	-0.3128	0.0936	0.4760	0.3773

## **BIBLIOGRAPHIE**

- [1] **BERGE (1977)**  
Orthogonal Procrustes rotation for two or more matrices.  
Psychometrika vol 42 n° 2.
- [2] **CAILLIEZ, F., PAGES, J.P. (1976)**  
Introduction à l'analyse des données. SMASH, 9, Rue Durban - 75016 PARIS
- [3] **CARROLL J.D. (1968)**  
A generalization of canonical correlation analysis to three or more sets of variables. Proc. 76<sup>th</sup> Conv. amer., Psych. Assoc.
- [4] **CASIN P., TURLOT J.C. (1986)**  
Une présentation de l'analyse canonique généralisée dans l'espace des individus. Revue de Stat. Appliquées, vol. 35, n° 3.
- [5] **CAZES P., BAUMERDER A. BONNEFOUS S., PAGES J.P. (1977)**  
Codage et analyse des tableaux logiques. Introduction à la pratique des variables qualitatives. Cahiers du Buro, n° 27, Université Paris VI.
- [6] **CROQUETTE A. (1980)**  
Quelques résultats synthétiques en analyse des données multidimensionnelle : optimalité et métriques à effets relationnels. Thèse de 3<sup>ème</sup> cycle, UPS Toulouse.
- [7] **DAUXOIS J., POUSSE A. (1976)**  
Les analyses factorielles en calcul de probabilités et en statistiques : essai d'étude synthétique. Thèse UPS TOULOUSE.
- [8] **DEMPSTER A.P. (1968)**  
Elements of continuous multivariate analysis.  
Readings, Massachussets, Addison Wesley.
- [9] **GOWER (1975)**  
Generalized Procrustes analysis. Psychometrika, vol. 40, n° 1
- [10] **LAFOSSE, R. (1985)**  
Une nouvelle analyse procustéenne de deux tableaux, appariement typique et atypique de deux nuages.  
4<sup>th</sup>. Int. Symp. data Analysis and Informatics INRIA
- [11] **LAFOSSE (1985)**  
Analyses procustéennes de deux tableaux. Thèse de 3<sup>ème</sup> cycle, Toulouse.

- [12] LAFOSSE, R., CAELEN, J., ROC, P. (1986)  
Recoding a table in order to increase its similarity with another table, using a procustean technique.  
2<sup>nd</sup> Catalan Int. Symp. on Statistics - BARCELONE, SPAIN
- [13] LINGOES J.C., BORG I. (1978)  
A direct approach to individual differences scaling using increasingly complex transformations.
- [14] LINGOES, J.C., SHONEMANN, P.H. (1974)  
Alternative measures of fit for the Shönemann-Carroll matrix fitting algorithm.  
Psychometrika, Vol. 39, n° 4.
- [15] MAHALANOBIS (1936)  
On the generalized distance in statistics. Proc. Nat. Inst. Sc. India 12.
- [16] PAGES J.P., CAILLIEZ F., ESCOUFIER Y. (1979)  
Analyse factorielle : un peu d'histoire et de géométrie. R.S.A. 27 n° 1.
- [17] SAPORTA G. (1975)  
Liaison entre plusieurs ensembles de variables et codage de données qualitatives. Thèse de 3<sup>ème</sup> cycle, Université Curie, Paris VI.
- [18] SCHEKTMAN, Y. (1978)  
Contribution à la mesure en facteurs dans les sciences expérimentales et à la mise en oeuvre des calculs statistiques.  
Thèse, U.P.S. TOULOUSE.
- [19] SIBSON R. (1978)  
Studies in the robustness of multidimensional scaling : procustes statistics J.R. Stat. Soc. B vol 40 n° 2.
- [20] TAKEUCHI, K. YANAI H., MUKHERJEE B.N. (1982)  
The foundation of multivariate analysis. Wiley Eastern - New Delhi.
- [21] TEBBOUNE F.O. (1981)  
Contribution aux analyses en composantes principales sous contraintes linéaires. Thèse de 3<sup>ème</sup> cycle- UPS Toulouse
- [22] YANAI H. (1985)  
Some generalizations of correspondance analysis in terms of projection operators. 4<sup>th</sup> int. Symp. Data analysis and informatics - INRIA