

ANALYSE FACTORIELLE DES
MATRICES CARREES NON SYMETRIQUES

ESCOUFIER Yves
GRORUD Axel
C.R.I.G
U.S.T.L Montpellier

Summary : The hermitian matrices properties give the basis for a new method in the exploration of asymmetric square matrices in which the rows and the columns are classified by the same set of entities. In the particular case of preference data, the present approach can be interpreted as the result of a complex coding of the preferences given by the individuals. This interpretation leads to the proposition of a method for the joint analysis of several individual preference matrices.

Résumé : S'appuyant sur les propriétés des matrices hermitiennes, on propose une méthode originale pour l'analyse des matrices carrées non-symétriques dans lesquelles les lignes et les colonnes correspondent aux mêmes objets. On montre, dans le cas des matrices de préférence, que cette approche peut être interprétée comme le résultat d'un codage complexe des préférences formulées par chaque individu. Ce résultat conduit à proposer une méthode pour l'analyse conjointe de plusieurs matrices de préférence individuelle.

I. INTRODUCTION

I.1. De nombreuses situations expérimentales conduisent à l'étude des matrices carrées non-symétriques :

Matrice de confusion : On dispose d'une batterie de p signaux (des lettres de l'alphabet morse par exemple). Un expérimentateur choisit un des signaux de façon aléatoire et l'émet. Un sujet doit restituer le signal qu'il a reçu. La matrice de confusion est la matrice $p \times p$ dont l'élément n_{kl} est le nombre de fois où le sujet a restitué le signal l alors que le signal k était émis.

Matrice d'échange entre régions : L'étude porte sur un ensemble de p régions et on s'intéresse aux migrations de population entre les régions. La matrice d'échange est la matrice $p \times p$ dont l'élément n_{kl} est le nombre des personnes qui habitaient à une certaine date dans la région k et qui habitent au moment de l'enquête dans la région l .

Matrice de préférence : p produits sont présentés à n sujets. La matrice $p \times p$ des préférences a pour élément n_{kl} , le nombre de fois où le produit l a été préféré au produit k . Notons que dans cet exemple la définition de la diagonale est ambiguë : on peut choisir 0, n , ou $\frac{n}{2}$ selon qu'on considère que k n'est jamais préféré à l ; que k est toujours préféré à l ; qu'il y a indétermination.

Bilan d'un jeu : p joueurs se sont affrontés 2 à 2 dans un jeu dont l'enjeu est financier. Une matrice $p \times p$ peut être constituée dont l'élément n_{kl} est le gain du joueur k sur le joueur l . Dans une telle matrice $n_{kl} = -n_{lk}$ et $n_{kk} = 0$.

I.2. Pour traiter de telles données, le statisticien est enclin à se ramener à des techniques qu'il connaît bien au risque d'oublier la nature spécifique des données qu'il traite.

La tentation pourra être de considérer de telles matrices comme un cas particulier de matrices de dimension $p \times q$ et d'employer des méthodes développées pour de telles matrices : analyse en composantes principales, analyse des correspondances si tous les éléments sont positifs. On doit s'interroger alors sur le fait que de telles méthodes fournissent des représentations distinctes pour les objets-lignes et les objets-colonnes alors qu'ici ce sont les mêmes objets qui définissent les lignes et les colonnes.

Une autre tentation pourra être de symétriser artificiellement la matrice de façon à pouvoir utiliser les méthodes adaptées à l'étude des matrices symétriques : factorisation ; multidimensionnal scaling. Il faut alors bien garder en mémoire que les résultats obtenus ont trait à la matrice artificiellement symétrisée et non à la matrice de départ.

Le but de notre travail est de proposer une nouvelle approche de l'étude de telles matrices, approche qui essaie de ne pas oublier que les objets-lignes et les objets-colonnes appartiennent au même ensemble.

A la suite des travaux de A. GRORUD (6), la démarche conduit à faire un détour par le corps des nombres complexes, ce qui est rare en analyse des données mais ne présente pas de difficultés majeures. Cette approche conduit très près des propositions faites par A.G. CONSTANTINE et J.C. GOWER (3) pour l'étude des matrices carrées non-symétriques, ce qui est un encouragement à poursuivre dans ce sens. De plus, elle fournit une vision tout à fait nouvelle des données de préférence puisqu'elle conduit à envisager des codages complexes.

II. RAPPELS SUR LES MATRICES HERMITIENNES

Ce paragraphe a pour but de fournir les éléments mathématiques qui constituent la base de la méthode proposée.

II.1. Soit H une matrice $p \times p$, à éléments complexes. On dit que H est hermitienne si pour tout couple (k, l) :

$$H_{kl} = \overline{H_{lk}}, \text{ où } \overline{H_{lk}} \text{ est le complexe conjugué de } H_{lk}.$$

En écrivant que $H = A + iB$, cet énoncé est équivalent à l'énoncé : H est hermitienne si et seulement si A est symétrique ($A = {}^t A$) et B est antisymétrique (${}^t B = -B$).

Comme les matrices symétriques réelles, les matrices hermitiennes possèdent des valeurs propres et des vecteurs propres. Les vecteurs propres sont des vecteurs à composantes complexes mais les valeurs propres ont la propriété d'être réelles. Si toutes les valeurs propres sont positives, on dit que la matrice hermitienne est définie positive.

Soit $w = u + iv$ un vecteur propre de $H = A + iB$ associé à la valeur propre λ . Des calculs simples permettent d'établir les propriétés suivantes :

Propriété 1 : Le vecteur $\begin{pmatrix} u \\ -v \end{pmatrix}$ est vecteur propre de la matrice $E = \begin{pmatrix} A & | & -B \\ \hline -B & | & A \end{pmatrix}$

pour la valeur propre λ .

Propriété 2 : Si $\begin{pmatrix} u \\ v \end{pmatrix}$ est vecteur propre de la matrice E pour la valeur propre λ , le vecteur $\begin{pmatrix} -v \\ u \end{pmatrix}$ l'est aussi pour la même valeur propre.

Ces deux propriétés fournissent un moyen simple pour calculer les valeurs propres et vecteurs propres de H . On calculera en fait ceux de E et on ne retiendra que l'un des deux vecteurs propres associés à la même valeur propre.

II.2. Appelons s^j et w^j les deux vecteurs propres de E , normés à 1, associés à la même valeur propre λ_j ($j = 1, \dots, p$). On sait que pour E qui est réelle et symétrique :

$$E = \sum_{j=1}^p \lambda_j (s^j t_{s^j} + w^j t_{w^j})$$

où t_{s^j} est le transposé de s^j

$$s^j = \begin{pmatrix} u^j \\ v^j \end{pmatrix} \text{ et } w^j = \begin{pmatrix} -v^j \\ u^j \end{pmatrix}$$

On en déduit aisément que :

$$A = \sum_{j=1}^p \lambda_j (u^j t_{u^j} + v^j t_{v^j}) \quad B = \sum_{j=1}^p \lambda_j (v^j t_{u^j} - u^j t_{v^j})$$

Limitons notre intérêt aux vecteurs s^1 et w^1 associés à la plus grande valeur propre λ_1 et considérons les matrices :

$$A^* = \lambda_1 (u^1 t_{u^1} + v^1 t_{v^1})$$

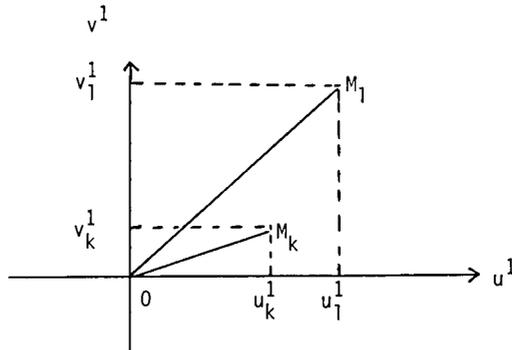
$$B^* = \lambda_1 (v^1 t_{u^1} - u^1 t_{v^1})$$

A^* et B^* sont des approximations de A et B . Lorsque toutes les valeurs propres sont positives (et nous verrons comment nous pourrions nous limiter à ce cas) l'intérêt de A^* et B^* réside dans le fait que :

$$E^* = \begin{pmatrix} A^* & | & -B^* \\ \hline B^* & | & A^* \end{pmatrix}$$

est la meilleure approximation de rang 2 de E ; ce qui revient à dire que $H^* = A^* + iB^*$ est la meilleure approximation de rang 1 de H .

Adoptons alors une représentation qui consiste à associer les deux axes d'un repère orthonormé de R^2 respectivement à u^1 et v^1 .



Le produit scalaire des vecteurs $\sqrt{\lambda_1} \begin{pmatrix} u & 1 \\ -k & 1 \end{pmatrix}$ et $\sqrt{\lambda_1} \begin{pmatrix} u & 1 \\ v & 1 \end{pmatrix}$ est égal à A_{k1}^* et

fournit donc une approximation de A_{k1} .

La norme de $\sqrt{\lambda_1} \begin{pmatrix} u & 1 \\ -k & 1 \end{pmatrix}$ est une approximation de A_{kk} .

On a donc une représentation approchée de A , tout à fait en accord avec les représentations usuellement faites en analyse des données.

Mais on a plus car B_{k1}^* est égal au double de la surface algébrique du triangle de côté $0 M_k$, $0 M_1$, multipliée par λ_1 . On a donc là une information en signe et en intensité qui ne fait pas partie des outils classiques de l'analyse des données. Elle nous permettra de voir que B_{k1}^* est positif ou négatif selon que

$0 M_k$ succède ou précède $0 M_1$ lorsqu'on tourne autour de l'origine dans le sens des aiguilles d'une montre ; la surface du triangle nous renseignera sur la valeur absolue de B_{k1}^* . En particulier la colinéarité de $0 M_k$ et $0 M_1$ indique que B_{k1}^* est nul.

III. APPLICATION A DES MATRICES CARREES NON-SYMETRIQUES

III.1. L'idée simple à laquelle on est conduit après les réflexions précédentes est de substituer à une matrice carrée non-symétrique, la matrice :

$$H = \left(\frac{C + {}^tC}{2} \right) + i \left(\frac{C - {}^tC}{2} \right)$$

$A = \frac{C + {}^tC}{2}$ est symétrique et mesure la part de symétrie présente dans la matrice C.

$B = \frac{C - {}^tC}{2}$ est antisymétrique et mesure la part d'antisymétrie contenue dans la matrice initiale. On a les propriétés suivantes :

Propriété 3 : On ne modifie pas la représentation en ajoutant une constante aux éléments diagonaux de C.
Par construction on a : $\text{diag. } (B) = 0$ et $\text{diag. } (A) = \text{diag. } (C)$.
Le résultat découle directement du fait que E et E + kI ont les mêmes vecteurs propres. Si λ est une valeur propre de E, $\lambda + k$ est la valeur propre correspondante de E + kI.

Cette propriété permet de ne pas se soucier du signe des valeurs propres de E. Le vecteur associé à la plus grande valeur propre est toujours celui qui est retenu.

Propriété 4 : Si C est centrée en ligne et en colonne pour des poids d_i ($i = 1, \dots, p$), la représentation obtenue est centrée pour les mêmes poids.

Soit D la matrice diagonale $p \times p$, d'éléments diagonaux $D_{ii} = d_i$.
Considérer que la matrice initiale est centrée en ligne et en colonne revient à considérer une matrice de la forme :

$$(I - \underline{\underline{11}}'D) C (I - \underline{\underline{D11}}')$$

On a alors pour partie symétrique :

$$(I - \underset{\sim}{1}\underset{\sim}{1}'D) \frac{C + C'}{2} (I - D\underset{\sim}{1}\underset{\sim}{1}')$$

et pour partie antisymétrique :

$$(I - \underset{\sim}{1}\underset{\sim}{1}'D) \frac{C - C'}{2} (I - D\underset{\sim}{1}\underset{\sim}{1}')$$

Le calcul permet de vérifier que le vecteur $\begin{pmatrix} D & 1 \\ \dots & \dots \\ D & 1 \end{pmatrix}$

est vecteur propre de la matrice E pour la valeur propre 0.

Puisque E est symétrique, ses vecteurs propres sont orthogonaux. Il s'ensuit que pour le vecteur s^1 :

$$t_u^1 D \underset{\sim}{1} + t_v^1 D \underset{\sim}{1} = 0$$

et pour le vecteur w^1 :

$$- t_v^1 D \underset{\sim}{1} + t_u^1 D \underset{\sim}{1} = 0$$

On a déduit :

$$t_u^1 D \underset{\sim}{1} = t_v^1 D \underset{\sim}{1} = 0$$

Remarque : Dans leurs travaux A.G. CONSTANTINE et J.C. GOWER propose de représenter séparément A et B. Nous proposons de les représenter conjointement. La représentation séparée a l'avantage d'être optimale pour A et B mais a l'inconvénient de ne pas être conjointe. La représentation conjointe a l'avantage d'être conjointe, optimale pour E ; elle a l'inconvénient de ne pas être optimale pour A et B séparément.

III.2. Exemples :

Pour illustrer notre proposition, nous avons retenu de présenter les résultats obtenus pour les matrices de LE NY et ROTHKOPF qui sont proposées également par J.P. BENZECRI (2).

a) Dans les données de LE NY, on dispose d'une matrice 6 x 6. Les objets étudiés sont des sons de fréquence croissante. n_{k1} est le nombre de fois où le son 1 a été donné en réponse alors que le son k était émis.

Réponses

Stimuli

233	96	54	26	6	0
94	142	118	40	16	6
20	62	122	110	71	31
11	22	76	139	130	38
3	4	15	54	174	166
4	0	3	18	98	293

La figure 1 fournit la représentation obtenue dans l'analyse de cette matrice. On voit que cette représentation suggère une énumération des sons en fonction de leur fréquence. On peut trouver plusieurs explications à cette disposition. Si on remarque que n_{k1} est toujours inférieur à n_{1k} , on peut interpréter la position des points en référence à la dissymétrie de la matrice. On peut dire aussi que les points sont rangés dans un ordre qui est sensiblement fonction de l'utilisation du son comme réponse (somme de la colonne) que cette utilisation soit faite à tord (éléments non-diagonaux) ou à raison (élément diagonal). Ce classement correspond aussi au classement auquel conduit le critère défini par la différence entre le nombre de fois où le stimulus provoque une fausse réponse et le nombre de fois où il est utilisé comme fausse réponse.

La figure 2 fournit la représentation obtenue dans l'analyse de la matrice centrée déduite de la matrice initiale par la transformation rappelée dans la proposition 4. L'allure parabolique de cette représentation rappelle curieusement l'effet Guttman.

b) La matrice des données de ROTHKOPF (8) concerne les lettres et les chiffres de l'alphabet Morse. Elle est donc de dimensions 36 x 36. Nous ne la reproduisons pas ici. Dans cette matrice n_{k1} est la proportion des fois où les sujets ont cru entendre deux signaux identiques alors qu'on a émis k suivi de l.

Lorsqu'on applique la méthode que nous proposons, on obtient bien sûr une représentation des 36 points. Pour simplifier l'interprétation nous n'avons représenté que les 26 lettres dans la figure 3. Les 10 chiffres sont représentés dans la figure 4.

On constate dans la figure 3 que la distance d'une lettre à l'origine est à peu près proportionnelle à la longueur du signal qui lui est associé (le trait et l'espace séparant deux caractères valent 3 points). On note aussi que :

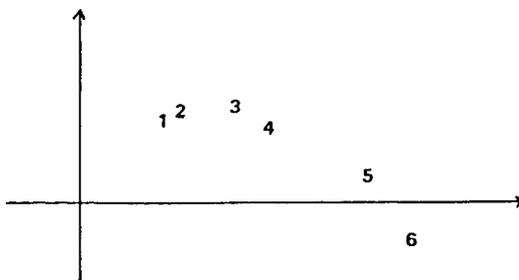


FIGURE 1

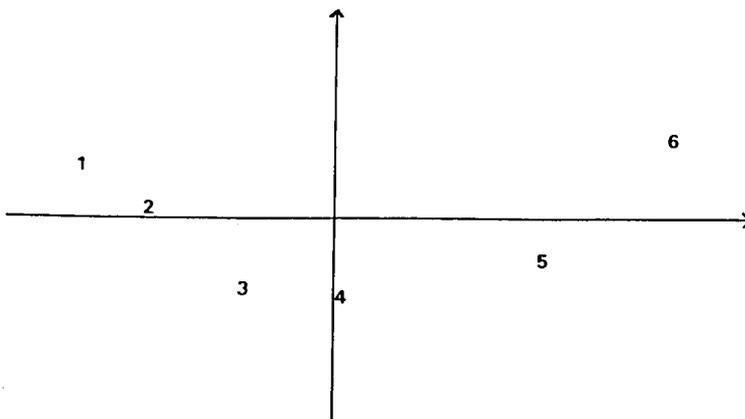


Figure 2

u (.. —) précède d (—..)
 w (.— —) précède g (— —.)
 v (... —) précède b (—...)

On est conduit à interpréter que les lettres associées à des signaux courts sont moins souvent impliquées dans les erreurs que les lettres associées à des signaux longs et que, pour des signaux de même longueur, les lettres commençant par des points le sont moins que les lettres commençant par les traits. La distance à l'origine apparaît ici comme liée à l'intensité des erreurs dans laquelle la lettre est concernée (la partie symétrique).

En se rapportant aux données, on remarque que la lettre u (.. —) entraîne moins souvent une fausse réponse quand elle est émise en premier que quand elle est émise en second. Au contraire, la lettre y (.— —) entraîne plus souvent une fausse réponse quand elle est émise en premier que quand elle est émise en second. On peut donc penser que u et ses voisines sont des lettres que les sujets identifient bien quand elles sont émises en premier. Ils les mémorisent bien et sont assez souvent capables de reconnaître si le signal qui les suit leur est identique ou non. Au contraire, y et ses voisines sont plus rarement identifiées et mémorisées ; les sujets ont de la difficulté à dire si le signal qui les suit leur est identique ou non. On a donc ici une illustration de la représentation des dissymétries : les lettres placées vers le haut de la représentation sont les lettres qui entraînent peu d'erreurs quand elles sont émises en premier ; les lettres placées vers le bas sont les lettres qui provoquent des erreurs quand elles sont émises en premier.

La figure 4 montre que les chiffres ont une organisation qui leur est propre. La superposition des figures 3 et 4 permet de constater que 5 (.....) se confond facilement avec h (....) et zéro (— — — —) avec la lettre 0 (— — —). Sur l'axe horizontal on retrouve les signaux commençant par des points avant les signaux de même longueur commençant par des traits. On peut également dire de 5 (.....), et de zéro (— — — —) ce qui a été dit de u : ce sont des signaux assez bien reconnus quand ils sont émis en premier.

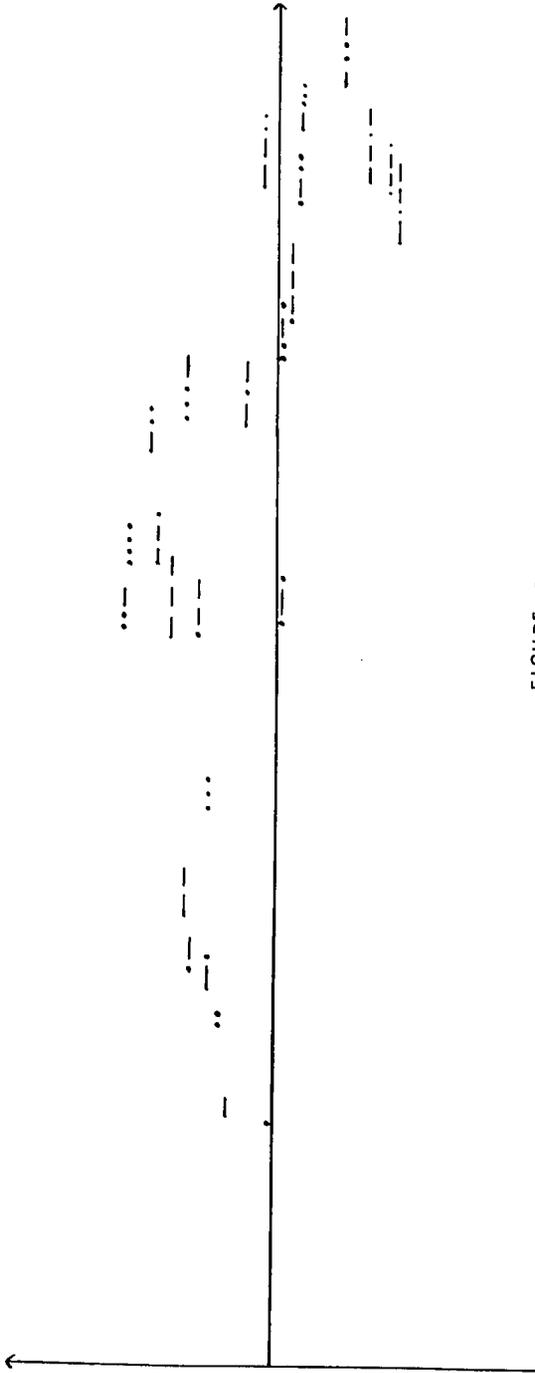
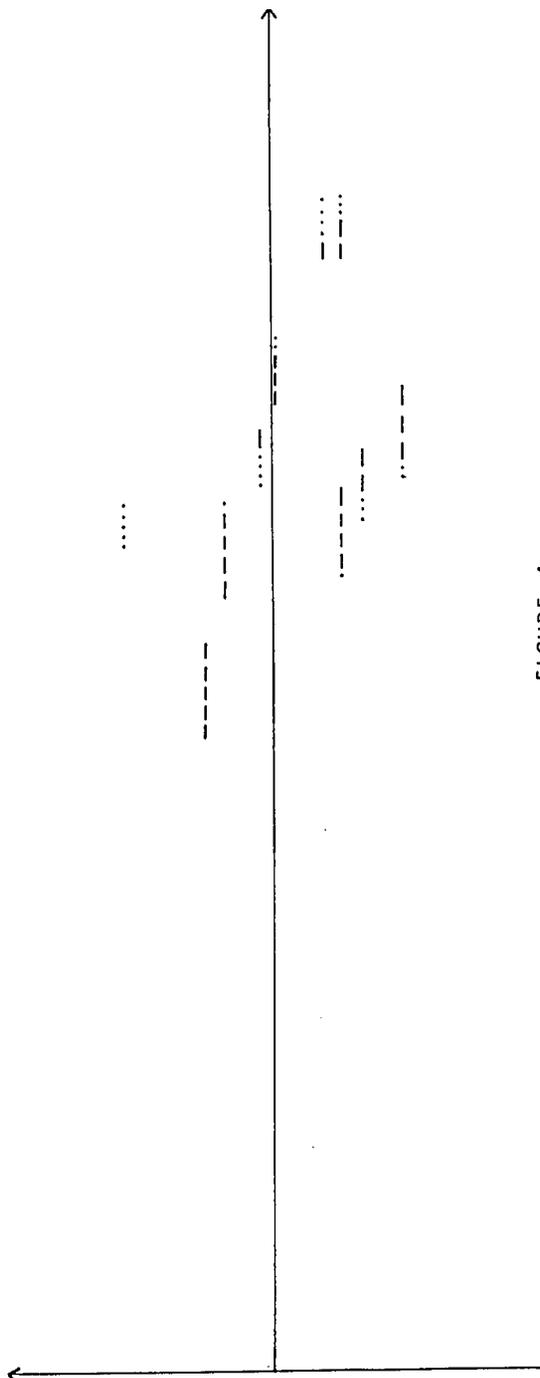


FIGURE 3



IV. LE CAS PARTICULIER DES MATRICES DE PREFERENCE

IV.1. Soit C une matrice de préférence ; l'élément (k,l) de C est n_{kl} le nombre de fois où l'objet k a été préféré à l'objet l. Si n sujets ont été interrogés ; $n_{kl} + n_{lk} = n$. La propriété 3 permet de ne pas se soucier outre mesure de la valeur des éléments diagonaux de C. Pour faire un choix nous les prendrons égaux à $\frac{n}{2}$.

Les éléments des matrices A et B définies précédemment sont pour tout couple (k,l) :

$$A_{kl} = \frac{n}{2}$$

pour tout k : $B_{kk} = 0$

$$\begin{aligned} \text{pour tout couple : } (k,l), k \neq l : B_{kl} &= \frac{n_{kl} - n_{lk}}{2} = \frac{n - 2 n_{lk}}{2} \\ &= - \frac{n - 2 n_{kl}}{2} \end{aligned}$$

Considérons un codage complexe des préférences défini de la façon suivante.

A l'individu $j(j=1 \dots n)$, on associe la matrice S^j , $p \times p$, définie par :

$$\text{pour } k \neq l : S_{kl}^j = \frac{1}{2} (1 + i) \text{ si } k \text{ est préféré à } l$$

$$S_{kl}^j = \frac{1}{2} (1 - i) \text{ si } l \text{ est préféré à } k$$

$$\text{pour tout } k : S_{kk}^j = \frac{1}{2}$$

$$\text{Soit } S = \sum_{j=1}^n S^j$$

$$\text{pour tout } k : S_{kk} = \frac{n}{2}$$

$$\text{pour } k \neq l : S_{kl} = \frac{n_{kl}}{2} (1 + i) + \frac{n - n_{kl}}{2} (1 - i)$$

$$S_{kl} = \frac{n}{2} - i \frac{n - 2 n_{kl}}{2}$$

On a donc $S = A + iB$. La procédure que nous avons proposé dans le paragraphe III correspond donc à un codage complexe implicite des données de préférences. A chaque individu est associé une matrice hermitienne S^j . Le point de vue de l'ensemble des individus se traduit par S la somme des S^j .

Remarque : On peut écrire la matrice S^j sous la forme :

$$2 S^j = \underset{\sim}{1} + i N$$

où $\underset{\sim}{1}$ est une matrice $p \times p$ dont tous les éléments sont égaux à 1 et N une matrice antisymétrique.

En associant à $2 S^j$ la matrice $E^j = \begin{pmatrix} 1 & | & N \\ \hline \underset{\sim}{1} & | & \underset{\sim}{1} \\ N & | & \underset{\sim}{1} \end{pmatrix}$

On constate que pour tout vecteur $w = \begin{pmatrix} u \\ \vdots \\ v \end{pmatrix}$ on a :

$${}^t_w E^j w = \left(\sum_{i=1}^p u_i \right)^2 + \left(\sum_{i=1}^p v_i \right)^2 \geq 0$$

ce qui établit que les matrices E^j et donc S^j sont définies positives.

IV.2. Dans la ligne des travaux de H. L'HERMIER (7), Y. ESCOUFIER et H. L'HERMIER (4), P. CAZES et al. (1), on peut se demander si le point de vue de l'ensemble des individus est bien traduit par S . Ne pourrait-on pas cons-

truire une matrice $S^* = \sum_{j=1}^n \alpha_j S^j$ qui soit un meilleur compromis que S ?

Suivant ces travaux on est conduit à considérer successivement :

i) $\text{Tr} (H_1 \overline{{}^t H_2})$ produit scalaire sur l'ensemble des matrices hermitiennes $p \times p$.

ii) F matrice $n \times n$, d'éléments $F_{ij} = \text{Tr} (S^i \overline{{}^t S^j})$.

Parce que les S^j sont définies positives, les éléments de F sont tous positifs.

(La factorisation de F fournit une représentation des individus. Des individus qui se ressemblent, c'est-à-dire qui ont donné des

des préférences peu différentes sont voisins).

iii) α le premier vecteur propre de F , d'éléments $\alpha_1, \dots, \alpha_n$ (Les α_1 sont tous positifs puisque les éléments de F sont positifs).

On peut alors définir $S^* = \sum_{j=i}^p \alpha_j S^j$ et démontrer que S^* est la combinaison linéaire des S^j qui maximise $\sum_{j=i}^n (\text{Tr}(S^* \overline{t} S^j))^2$ sous la contrainte

$$\sum_{i=1}^n \alpha_i^2 = 1.$$

- (1) P. CAZES, S. BONNEFOUS, A. BAUMERDER, J.P. PAGES (1976).
- Description cohérente des variables qualitatives prises globalement et de leurs modalités.
S.A.D n° 2 p. 48-62.
- (2) J.P. BENZECRI (1973) - L'analyse des données, Tome 1. Dunod (p. 28).
- (3) A.G. CONSTANTINE and J.C. GOWER (1978) - Graphical Representation of Asymmetric Matrices. Appl. Statist. 27. n° 3 p. 297-304.
- (4) Y. ESCOUFIER et H. L'HERMIER (1978) - A propos de la comparaison graphique des matrices de variance. Biom. J., Vol. 20 n° 5 p. 491-497.
- (5) J.C. GOWER (1977) - The analysis of asymmetry and orthogonality. In Recent Developments in Statistics (J. Barra et al. ed.) Amsterdam North-Holland pp. 109-123.
- (6) A. GRORUD (1978) - Analyse factorielle de données complexe. Thèse 3ème cycle U.S.T.L Montpellier.
- (7) H. L'HERMIER (1976) - Structuration des tableaux à trois indices de la Statistique. Thèse 3ème cycle, U.S.T.L Montpellier.
- (8) E.Z. ROTHKOPFF (1957) - A measure of stimulus similarity and errors in some paired-associated learning tasks. J. Exp. Psych., 53, 94 - 101.