

STATISTIQUE ET ANALYSE DES DONNÉES

GILDAS BROSSIER

Classification hiérarchique à partir de matrices carrées non symétriques

Statistique et analyse des données, tome 7, n° 2 (1982), p. 22-40

http://www.numdam.org/item?id=SAD_1982__7_2_22_0

© Association pour la statistique et ses utilisations, 1982, 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/>

CLASSIFICATION HIERARCHIQUE
A PARTIR DE MATRICES CARREES NON SYMETRIQUES
Gildas BROSSIER

U.E.R. Sciences et Techniques
Université de Haute Bretagne - 35043 RENNES Cédex

Résumé : *L'ensemble à classer est caractérisé par une matrice carrée non symétrique, telle qu'une matrice d'échange, de migration, de préférence, de confusion, ...etc.*

Cette matrice se décompose de façon canonique en sa partie symétrique et antisymétrique. Après avoir fait une classification hiérarchique sur la partie symétrique, on y intègre une représentation de l'antisymétrie.

On justifie cette représentation par des propriétés d'optimalités et on comparera avec les méthodes factorielles d'analyse des matrices non symétriques à l'aide d'un exemple réel.

Abstract : *The data set is characterized by an asymmetric square matrix as for exchange, migration, preference or confusion data.*

This matrix is decomposed into its canonical symmetric and skew-symmetric elements. After computing a hierarchical clustering on the symmetric matrix we introduce a graphical representation of the skew-symmetric matrix on the dendogram.

We justify this graphical representation by some properties of optimality and we compare, on a real data set, with the factorial analysis methods for asymmetric matrices.

Mots clés : *Classification hiérarchique*
Matrices non symétriques
Ordres compatibles avec une hiérarchie.

I - INTRODUCTION

On s'intéresse ici aux matrices carrées non symétriques où les lignes et les colonnes représentent le même ensemble d'objets. Ce type de matrices est assez fréquent et recouvre toute une classe de problèmes. Il suffit de penser aux matrices de préférences, d'échanges, de migration, de confusion, etc.. (voir pour un exemple le § IX).

Si on veut tenir compte du caractère non symétrique de ces matrices on ne peut retenir aucune méthode usuelle d'analyse des données, car elles supposent toutes la symétrie de la relation entre les éléments. C'est pourquoi, deux méthodes spécifiques ont déjà été développées dans le cadre de l'analyse factorielle ([3 | , [5]). Nous comparerons nos résultats avec ceux de ces méthodes sur un exemple réel.

Dans le cadre des méthodes de classification on ne peut retenir l'approche de Cunningham ([4]) car les arbres bidirectionnels qu'il construit ne peuvent plus être calculés au-delà d'une dizaine d'éléments et la représentation est difficile à lire. C'est pourquoi l'approche classificatoire de ce type de problème reste à faire.

Pour de telles données, il peut paraître paradoxal de penser à une classification, par essence même, symétrique. Il est donc nécessaire de se poser la question de savoir quels peuvent être les objectifs d'un utilisateur qui désire une classification sur une matrice non symétrique. La méthode devra répondre à ces objectifs.

Pour illustrer notre propos, supposons que nous soyions en présence d'une matrice d'échange entre pays.

- A l'évidence, dans un premier temps, il faut former les classes des pays qui se ressemblent, c'est à dire ici, les pays entre lesquels les échanges sont importants; c'est la classification usuelle sur la partie symétrique de la matrice.

- Dans un deuxième temps, il faut mettre en évidence les excédents et les déficits dans les échanges entre les pays; c'est l'analyse de la partie antisymétrique de la matrice qu'il faut intégrer dans la représentation finale.

- Enfin, dans un troisième temps, il faut évaluer le sens général dans lequel s'effectuent les échanges; c'est l'ordre induit par la partie antisymétrique de la matrice.

On va chercher à répondre à ces différents objectifs pour finalement construire une représentation des données qui met en évidence ces différents aspects.

II - DECOMPOSITION DE LA MATRICE ET RAPPELS

Nous rappelons ici deux propriétés élémentaires que nous utilisons.

Soit D une matrice $n \times n$, non symétrique, et D^t sa transposée.

Propriété 1

D se décompose de façon unique en la somme d'une matrice symétrique S et d'une matrice anti-symétrique, A

$$D = A + S \text{ avec } A = \frac{1}{2} (D - D^t) \text{ et } S = \frac{1}{2} (D + D^t)$$

Propriété 2

Soient \mathcal{A} l'ensemble des matrices $n \times n$ antisymétriques et \mathcal{S} celui des matrices $n \times n$ symétriques, alors \mathcal{A} est orthogonal à \mathcal{S}

En effet :

$$\begin{aligned} \forall A \in \mathcal{A} \\ \forall S \in \mathcal{S} \quad \langle A, S \rangle = \text{tr } A \cdot S^t = \sum_{i,j} A_{ij} \cdot S_{ij} = 0 \end{aligned}$$

Ceci nous permet de dire que toute approximation de D , au sens de la norme L^2 , par une matrice symétrique (resp. antisymétrique), revient au même que l'approximation de la seule partie symétrique (resp. antisymétrique) de D .

On est donc conduit à approximer séparément la partie symétrique et la partie antisymétrique.

Rappelons d'autre part, voir par exemple [3], que toute matrice antisymétrique peut se décomposer de façon canonique en une somme de matrices antisymétriques de rang, 2. C'est à dire :

$$A = \sum_{i=1}^{[n/2]} \lambda_i (u_i v_i^t - v_i u_i^t)$$

où les vecteurs u_i (resp. v_i) sont orthogonaux entre eux et les λ_i sont des réels positifs.

Alors on voit que le cas le plus simple de matrice antisymétrique est celui où elle s'écrit sous la forme $A = (e \mathbf{1}^t - \mathbf{1} e^t)$

où $\mathbf{1}$ est le vecteur dont toutes les composantes sont égales à 1. Alors les éléments de A s'écrivent $A_{ij} = e_i - e_j$

Nous dirons dans ce cas que la matrice A est engendrée par le vecteur e ou qu'elle est de la forme $e_i - e_j$

III - REPRESENTATION GRAPHIQUE D'UNE MATRICE NON SYMETRIQUE

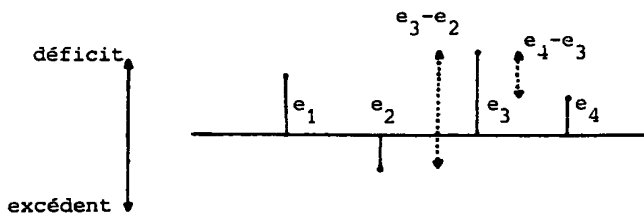
Soit $D = A + S$ la matrice non symétrique à analyser. Nous allons évidemment approximer la partie symétrique S , par une ultramétrique U et lui associer un arbre hiérarchique la représentant (§ IV).

Pour représenter la partie antisymétrique A , nous allons l'approximer par une matrice antisymétrique E de la forme $e_i - e_j$, c'est à dire engendrée par un seul vecteur, e , que nous calculerons (§ V).

Si le terme S_{ij} de la partie symétrique représente le niveau moyen des échanges entre les éléments i et j , le terme antisymétrique A_{ij} (resp. A_{ji}) représente l'excédent ou le déficit dans les échanges de i vers j (resp. de j vers i).

Donc en approximant la matrice A par une matrice E , engendrée par le vecteur e , nous associons à chaque élément i , un excédent (ou déficit) e_i . Ainsi nous approximos le terme A_{ij} par la différence des excédents, $e_i - e_j$.

Le vecteur e peut se représenter alors par le diagramme suivant :



La différence $e_i - e_j$ apparaît ici clairement comme étant le déséquilibre dans les échanges entre i et j .

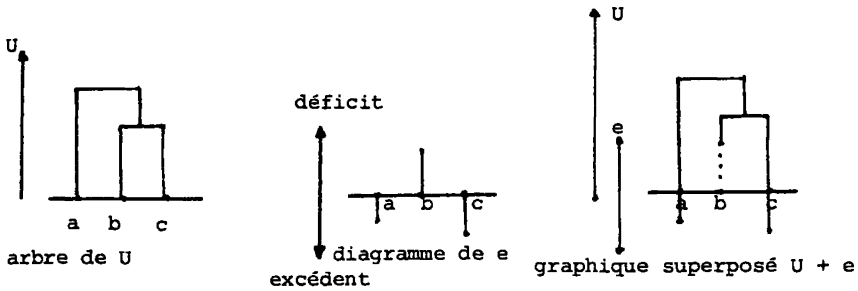
On va maintenant ajouter à l'ultramétrique U_{ij} , la quantité $e_i - e_j$ et nous allons superposer les deux représentations pour reconstruire de façon approchée D_{ij}

$$D_{ij} = S_{ij} + A_{ij} \approx U_{ij} + (e_i - e_j)$$

La superposition des deux graphiques se fait en allongeant ou en

racourcissant la longueur des branches de l'arbre à chaque noeud terminal i , d'une longueur égale à e_i .

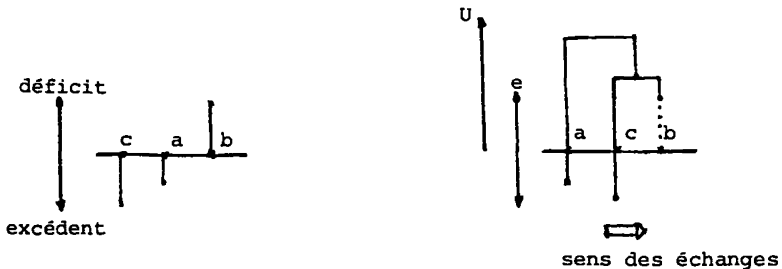
schémas explicatifs :



De cette façon on a une représentation simultanée de la partie symétrique et de la partie antisymétrique, mettant en évidence à la fois le niveau moyen des échanges, ainsi que les déséquilibres entre ceux-ci.

Pour rendre encore plus lisible cette représentation et visualiser l'ordre général dans lequel s'effectuent ces échanges, on va ordonner les noeuds terminaux de la hiérarchie selon un ordre aussi proche que possible de l'ordre induit par le vecteur e : de l'élément le plus excédentaire vers l'élément le plus déficitaire.

Schémas



Il est bien évident que nous ne pourrions pas en général prendre l'ordre exact induit par e car cela conduirait à faire des croisements dans les branches de l'arbre (voir [1]).

Ainsi sur notre exemple l'ordre cab conduit à un graphique avec croisement. L'ordre est dit non compatible avec la hiérarchie.

Nous serons donc amenés à rechercher un ordre compatible avec la hiérarchie, c'est à dire tel que l'on puisse représenter la hiérarchie sans croisement, qui soit le plus proche possible de l'ordre induit par e (§ VI).

IV - APPROXIMATION DE LA PARTIE SYMETRIQUE

On retombe ici dans le problème classique de l'approximation d'une matrice de similarité S par une ultramétrie. On sait qu'il n'existe pas de solution algorithmique efficace à la recherche de l'ultramétrie à distance minimale au sens des moindres carrés de S, voir [2] .

On est donc amené à utiliser un des algorithmes classiques de classification ascendante hiérarchique au gré des besoins de l'utilisateur. Cependant, faute d'un autre choix, on peut utiliser l'algorithme de la moyenne (average linkage) sachant que :

- le résultat est interprété comme un niveau moyen d'échange,
- cet algorithme constitue en général une bonne approximation, au sens des moindres carrés, de la matrice initiale,
- on utilisera au § VI la notion de moyenne sur une classe.

V - APPROXIMATION DE LA PARTIE ANTISYMETRIQUE

Afin de pouvoir la représenter, il faut approcher la matrice A, antisymétrique, par une matrice E, antisymétrique de la forme $e_i - e_j$. On démontre la proposition suivante :

Proposition 1

Soit A une matrice carrée $n \times n$, antisymétrique

Alors la matrice E, carrée, $n \times n$, antisymétrique, de la forme

$E_{ij} = e_i - e_j$ telle que : $\|A - E\|$ minimal

est définie par $e_i = A_{i.}$ ou $A_{i.} = \sum_j A_{ij}$

Démonstration

On note par e, le vecteur d'éléments e_i , $i = 1, \dots, n$ engendrant E. Le vecteur e n'est défini qu'à une constante additive, m, près.

En effet si e^* est défini par $e_i^* = e_i + m$, alors la matrice E^* engendrée par e^* est égale à E. On peut donc supposer le vecteur e centré. ce qui fait que $\|E\|^2 = 2 n \|e\|^2$

$$\begin{aligned} \text{Donc } \|A - E\|^2 &= \|A\|^2 + \|E\|^2 - 2 \operatorname{tr} AE^t \\ &= \|A\|^2 + 2n \|e\|^2 - 4 \sum_i e_i A_{i.} \text{ en notant } A_{i.} = \sum_j A_{ij} \end{aligned}$$

Le minimum de $\|A - E\|^2$ est atteint en choisissant e proportionnel au vecteur $(A_{i.})$, $i = 1, n$, la constante de proportionnalité étant égale à $\frac{1}{n}$

$$\text{d'où } e_i = \frac{A_{i.}}{n} \text{ la valeur de } \|A - E\|^2 \text{ étant égale à } \|A\|^2 - \frac{2}{n} \|A_{i.}\|^2 \quad \square$$

Donc, d'une part on a la meilleure approximation de A , d'autre part on peut effectivement interpréter les e_i , comme les excédents moyens de i .
On peut remarquer qu'il peut être plus simple lors de la lecture de la représentation finale de la matrice D de ne pas considérer les distances $U_{ij} + (e_i - e_j)$ mais $U_{ij} + e_i$, c'est à dire de lire la représentation en tenant compte de l'excédent du point de départ, sans tenir compte de celui du point d'arrivée.

Cette lecture des résultats, si elle est moins bonne n'est pas pour autant erronée grâce à la proposition 2.

Proposition 2

Soit A une matrice carrée $n \times n$, alors la matrice F , carrée, $n \times n$, de la forme $F_{ij} = f_i$ telle que :

$$\|A - F\| \text{ minimal, est définie par } f_i = \frac{A_{i.}}{n}.$$

$$\text{La valeur du minimum étant égale à } \|A\|^2 - \frac{1}{n} \|A_{i.}\|^2$$

Démonstration

Elle est analogue à celle de la propriété 3 \square

Autrement dit pour une matrice A donnée, la représentation est valide pour les deux modèles puisque $\forall i, f_i = e_i$.

VI - APPROXIMATION DE L'ORDRE SUR LA HIERARCHIE

L'ordre général des échanges, c'est à dire des éléments les plus excédentaires vers les plus déficitaires, est celui induit par le vecteur e. Le problème est donc d'ordonner les noeuds terminaux de l'arbre hiérarchique selon un ordre aussi proche que possible de l'ordre O défini par :

Si O_i est le rang de l'élément i, $O_i < O_j$ si $e_i > e_j$. On suppose-
ra que O est un ordre total et donc que $e_i \neq e_j$ si $i \neq j$

Rappelons les deux propriétés suivantes, voir [1]

Propriété 3

Un ordre O est compatible avec une hiérarchie U, si on peut représenter celle-ci par un arbre plan sans croisement dont les noeuds terminaux sont ordonnés selon l'ordre O, et réciproquement.

Propriété 4

Une ultramétrique U sur un ensemble E de cardinal égal à n, admet au moins 2^{n-2} ordres compatibles.

Si on note par \mathcal{O}_U , l'ensemble des ordres sur E, compatibles avec U, si O est l'ordre sur E induit par le vecteur e, et représenté par son vecteur rang $(O_i, i = 1, \dots, n)$ alors il s'agit de trouver :

$R \in \mathcal{O}_U$, $R = (R_i, i = 1, \dots, n)$ tel que $\|O - R\|$ soit minimal, ou ce qui revient au même, $\rho(O, R)$ maximal. ρ étant le ρ de Spearman et R_i étant le vecteur rang associé à l'ordre R.

On va construire R selon l'algorithme suivant :

Algorithme 1

On va ordonner la hiérarchie U par un algorithme descendant qui va orienter chacun des noeuds de l'arbre de la façon suivante :

Soit un noeud qui se divise en deux branches A et B. On choisira l'ordre AB si $\bar{O}_A < \bar{O}_B$ Avec :

$$\bar{O}_A = \frac{1}{n_A} \sum_{i \in A} O_i, \quad n_A = \text{cardinal de la classe A.}$$

Cette procédure engendre un ordre R sur les noeuds terminaux de U, et par construction, R est compatible avec U.

Théorème :

Soit O un ordre, sur \mathcal{E} , donné

Soit \mathcal{O}_u l'ensemble des ordres sur \mathcal{E} compatibles avec U

Soit R l'ordre sur \mathcal{E} engendré par l'algorithme 1

alors :

$$\min_{s \in \mathcal{O}_u} \|O - s\| = \|O - R\|$$

Démonstration

Nous allons d'abord démontrer le lemme suivant :

lemme : Soit $s \in \mathcal{O}_u$, soit s^* l'ordre qui se déduit de S en permutant un noeud de la hiérarchie (le noeud AB dans S, devient le noeud BA dans s^*) sans changer les ordres restreints à A et à B.

alors $\bar{O}_A < \bar{O}_B \iff \|O - s\| < \|O - s^*\|$

$$K = \|O - s\|^2 - \|O - s^*\|^2 = \sum_i (O_i - S_i)^2 - \sum_i (O_i - S_i^*)^2$$

$$(1/2)K = -\sum_{i \in A \cup B} O_i S_i + \sum_i O_i S_i^* \text{ . pour } i \notin A \cup B \text{ } S_i = S_i^* \text{ donc}$$

$$(1/2)K = -\sum_{i \in A \cup B} O_i S_i + \sum_{i \in A \cup B} O_i S_i^* = \sum_{i \in A} O_i (S_i^* - S_i) + \sum_{i \in B} O_i (S_i^* - S_i)$$

S_i et S_i^* étant compatibles avec U, les 2 classes A et B sont contigües

Donc ayant supposé la classe A avant la classe B dans S (et l'inverse dans s^*), on a :

$$\text{pour } i \in A \quad S_i^* = S_i + n_B$$

$$i \in B \quad S_i^* = S_i - n_A \text{ , } n_A \text{ et } n_B \text{ étant les cardinaux de A et}$$

de B.

$$(1/2)K = n_B \sum_{i \in A} O_i - n_A \sum_{i \in B} O_i$$

$$\text{donc } \|O - s\| < \|O - s^*\| \iff \bar{O}_A < \bar{O}_B \quad \square$$

donc l'algorithme 1 qui oriente chaque noeud selon le critère A avant B si

$\bar{O}_A < \bar{O}_B$ génère l'ordre R optimal. En effet :

Supposons qu'il existe un ordre R' différent de R et tel que $\|O - R'\| < \|O - R\|$ comme R est généré par l'algorithme 1, et que R' est différent de R , cela veut dire qu'il existe au moins un noeud CD , tel que pour R' , $\bar{O}_C > \bar{O}_D$ et C avant D (sinon R' serait égal à R).

En appliquant le lemme au noeud CD , on trouve un ordre R'^* déduit de R' en permutant le noeud CD , meilleur que R' . Donc R' ne peut pas être à la fois optimal et différent de R .

VIII - MISE EN OEUVRE DE LA METHODE

Si on a à analyser une matrice non symétrique il faut donc successivement :

- décomposer en partie symétrique et antisymétrique,
- faire une classification sur la partie symétrique en appelant un programme existant,
- calculer le vecteur e , qui est la marge de la matrice antisymétrique,
- ordonner la hiérarchie selon l'algorithme 1, ce qui nécessite le calcul de $2(n - 2)$ moyennes et $(n - 2)$ tests,
- modifier la longueur des branches en ajoutant e_i ,

On voit donc qu'il suffit d'ajouter quatre petits programmes autour du programme de classification dont dispose l'utilisateur.

IX - EXEMPLE D' APPLICATION ET COMPARAISONS

Nous allons analyser le tableau des échanges intracommunautaires entre les 9 pays de la Communauté Européenne (la Belgique et le Luxembourg ayant été regroupés en un seul pays). Ce tableau est extrait des statistiques de bases de la Communauté, EUROSTAT, 1981 et concerne les échanges pour l'année 1980.

| | Allemagne | France | Italie | Pays-Bas | Belgique | G.B. | Irl. | Dan. | Grèce |
|------------|-----------|--------|--------|----------|----------|------|------|------|-------|
| Allemagne | | 15 708 | 11923 | 12 297 | 10 152 | 9209 | 486 | 2645 | 1064 |
| France | 14 542 | | 9971 | 3 707 | 7 459 | 6290 | 396 | 564 | 472 |
| Italie | 10 707 | 9 121 | | 1 746 | 1 853 | 3834 | 157 | 385 | 625 |
| Pays-Bas | 16 225 | 5 262 | 2993 | | 8 454 | 4826 | 285 | 1091 | 295 |
| Belgique | 9 962 | 8 117 | 2553 | 6 449 | | 4039 | 161 | 503 | 150 |
| Grande-Br. | 9 162 | 5 246 | 3182 | 4 522 | 4 169 | | 4417 | 1688 | 349 |
| Irlande | 599 | 544 | 190 | 285 | 193 | 2893 | | 41 | 28 |
| Danemark | 2 327 | 621 | 631 | 480 | 239 | 1809 | 52 | | 46 |
| Grèce | 1 086 | 331 | 350 | 216 | 78 | 206 | 6 | 27 | |

Sur ce tableau nous avons effectué 3 analyses : celle proposée par Constantine et Gower [3] , celle proposée par Escoufier et Gorud [5] , et celle que nous venons de présenter.

D'autre part, il est d'usage en économie internationale d'effectuer ce que l'on appelle une analyse de structure, en pondérant le volume des échanges par le Produit Intérieur Brut (P.I.B.) des pays concernés. Si D_{ij} est le volume des échanges du pays i vers le pays j , et p_i le P.I.B. du pays i , on considérera le tableau D^* défini par :

$$D_{ij}^* = D_{ij} / (p_i \cdot p_j)$$

Sur ce tableau D^* nous avons également effectué les 3 analyses.

| | | | |
|-----------|-----|-------------|-----|
| Allemagne | 557 | Royaume-Uni | 292 |
| France | 417 | Irlande | 10 |
| Italie | 236 | Danemark | 48 |
| Pays-Bas | 108 | Grèce | 28 |
| Belgique | 82 | | |

P I B en 1979

1 - Rappel de la méthode proposée par Constantine et Gower

Après avoir décomposé D en $A + S$, A la partie antisymétrique et S la partie symétrique, ils représentent séparément A et S .

S est représenté de façon classique dans le plan de ses deux premiers vecteurs propres et A est décomposé selon :

$$A = \sum_{i=1}^{n/2} \lambda_i (U_i V_i' - V_i U_i')$$

En ne retenant que les premiers vecteurs U_1 et V_1 ils obtiennent une représentation plane de l'antisymétrie. Si O est l'origine, et x, y deux points du plan, l'antisymétrie est proportionnelle à la surface du triangle oxy et son signe est donné par l'orientation du triangle.

2 - Rappel de la méthode proposée par Escoufier et Grorud

D étant décomposée en A et S, ils proposent une représentation conjointe de A et de S en analysant la matrice complexe hermitienne $H = A + iS$
H s'écrit :

$$H = \sum_{j=1}^n \lambda_j (U_j + iV_j)$$

En ne retenant que les deux premiers vecteurs U_1 et V_1 on a une représentation plane et simultanée de la symétrie et de l'antisymétrie. L'interprétation de l'antisymétrie étant analogue à celle de la méthode précédente.

3 - Analyse de la matrice D, du volume des échanges

- Méthode Constantine et Gower

La partie symétrique des échanges est représentée sur la figure 1a. L'axe 1 est trivialement celui du volume des échanges à partir ou à destination de chacun des pays. On retrouve ainsi dans l'ordre les pays dont les commerces intra-communautaires sont les plus importantes : l'Allemagne, puis la France, les Pays-Bas ...etc. L'axe 2 est un axe géographique Nord-Sud traduisant la proximité entre les pays qui favorise les échanges.

La partie antisymétrique est représentée sur la figure 1.b. Elle met en évidence le rôle particulier des Pays-Bas dont les échanges sont excédentaires par rapport à tous les autres pays européens. A un degré moindre, on remarque l'excédent de l'Allemagne sur la France, de la France sur l'Italie, ...etc.

L'Irlande, le Danemark et la Grèce ont un volume d'échange trop faible pour être bien interprétés.

- Méthode Escoufier et Grorud (figure 2)

L'axe horizontal est ici aussi celui du volume des échanges pour chacun des pays. On y retrouve les mêmes valeurs que dans l'analyse précédente. L'antisymétrie est surtout le fait encore des Pays-Bas et l'on remarque également la Belgique par rapport à la France et l'Italie. Ici encore le Danemark, l'Irlande et la Grèce ne jouent aucun rôle.

- Classification (figure 3)

Elle met en évidence, d'abord, la classe d'échange France-Allemagne qui est le fait essentiel, en volume, pour les échanges européens et n'apparaissait pas dans les analyses précédentes. A ce noyau central se rajoutent l'Italie et la Grande-Bretagne formant ainsi le groupe des échanges importants. A côté se forme la classe naturelle des Pays-Bas et de la Belgique avec un volume d'échange important. Les autres pays s'agrégeant à des niveaux très faibles correspondant au faible flux commercial.

L'analyse de l'antisymétrie met en évidence le rôle particulier des Pays-Bas que l'on avait également remarqué dans les précédentes analyses, ainsi que les excédents Allemagne-France et France-Italie.

4 - Analyse de la matrice D^* de structure

Le volume des échanges est maintenant ramené au poids économique de chaque pays. On va donc analyser non plus le volume brut des échanges mais leurs structures.

- Méthode Constantine et Gower (figure 4)

La partie symétrique est représentée sur la figure 4.a. L'axe 1 représente toujours le poids des échanges par pays, mais l'axe 2 sépare nettement deux groupes de pays. Le groupe Pays-Bas - Belgique, et le groupe Angleterre - Irlande mettant en évidence l'importance des échanges entre ces pays. On visualise aussi ce qui semble être un troisième groupe formé par les autres pays. Or, l'Allemagne et la France échangent relativement à leur PIB, beaucoup plus avec les Pays-Bas et la Belgique qu'entre eux. Il ne s'agit donc pas d'un troisième groupe.

La partie antisymétrique est représentée sur la figure 4.b. Cette représentation met en évidence uniquement la très forte dissymétrie des échanges entre la Grande-Bretagne et l'Irlande. Le reste des autres pays n'intervenant pas à ce niveau. En traçant le plan des vecteurs 3 et 4, on aurait ainsi la même opposition entre les Pays-Bas et la Belgique, et il faut aller chercher les vecteurs 5 et 6 pour voir représenter correctement les autres pays.

- Méthode Escoufier et Grorud (figure 5)

L'axe horizontal représente, comme dans le cas symétrique de l'analyse précédente, le poids des échanges entre les pays. Avec l'axe vertical on voit la dissymétrie des échanges Grande-Bretagne - Irlande et Pays-Bas - Belgique (à la condition de savoir associer correctement les deux couples).

Les autres pays comme dans l'analyse précédente sont écrasés par ces deux paires de pays et semblent former un groupe à part, ce qui n'est pas le cas.

- Classification

On met là aussi en évidence les deux couples Pays-Bas - Belgique et Grande-Bretagne - Irlande, les classes sont formées et l'on voit très nettement le sens des échanges.

A la différence des autres analyses les autres pays ne forment pas une classe à part, mais se rattachent au groupe Pays-Bas-Belgique avec lequel ils ont effectivement leurs échanges les plus importants, bien que le niveau soit faible (2000) comparé à celui des deux classes (12.000 et 9.000)

X - CONCLUSION

Reprenant les objectifs assignés à une classification d'une matrice non symétrique, on se rend compte, à travers l'exemple, que la différence essentielle avec les autres méthodes est lié au fait que l'outil de base est une classification hiérarchique. Et que c'est à la représentation de cette classification que l'on va intégrer la dissymétrie et l'ordre qu'elle induit.

Donc, comme dans le cas symétrique le choix entre classification hiérarchique et analyse factorielle dépend des objectifs de l'utilisateur, les deux méthodes restant complémentaires.

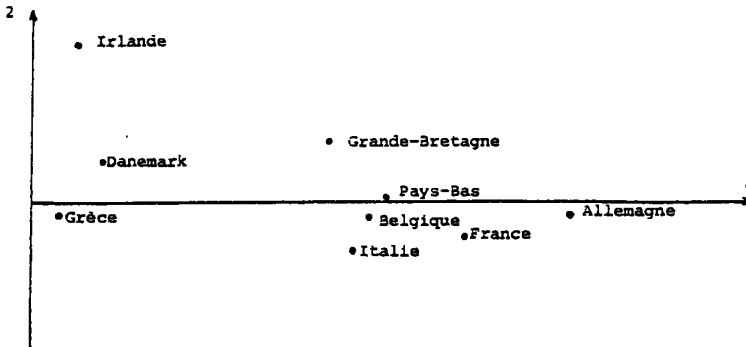


Figure 1a
Partie symétrique

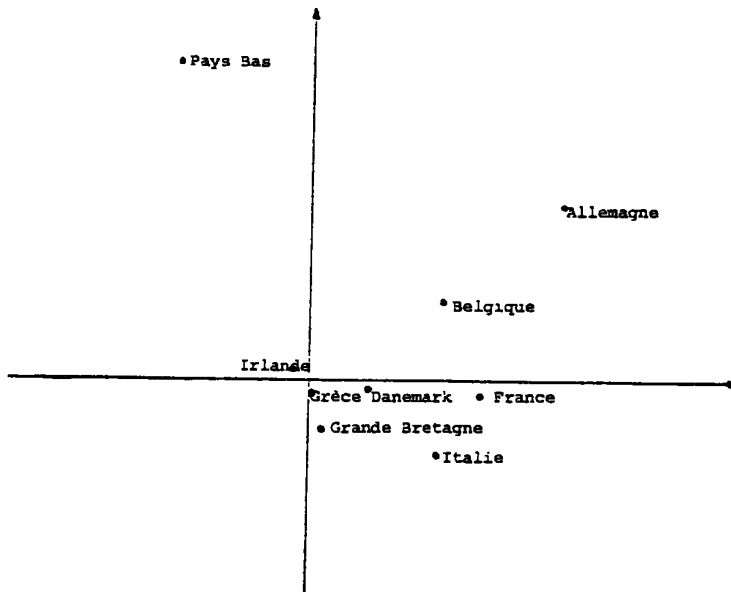


Figure 1 b
Partie anti-symétrique
matrice des volumes d'échanges
méthode de Constantine et Gower

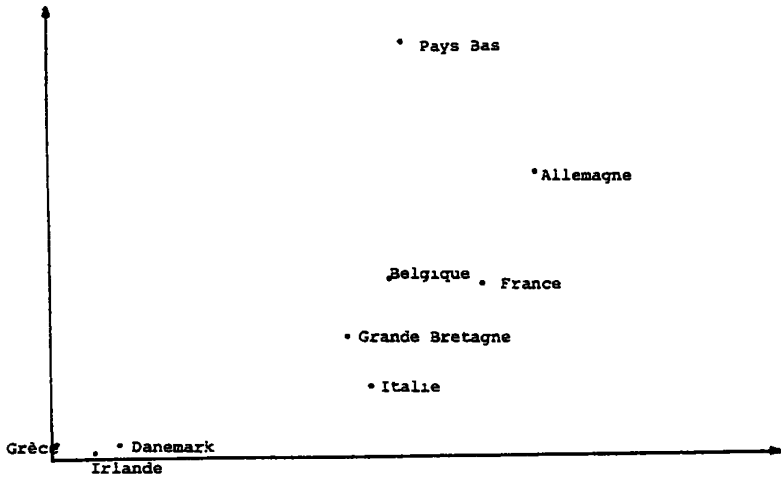


Figure 2

méthode Escoufier et Growud
matrice de volume des échanges

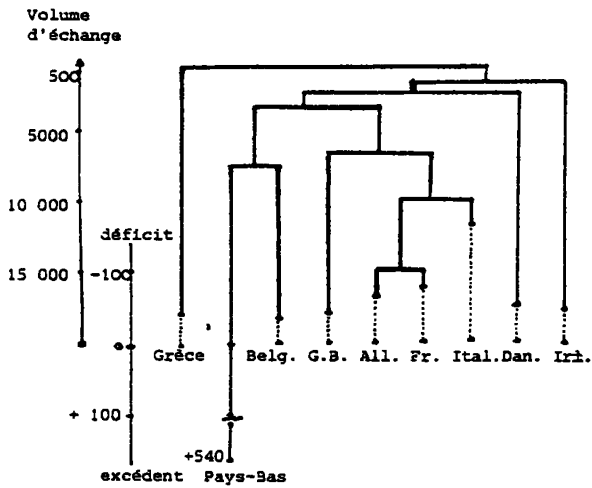


Figure 3

Classification

matrice de volume des échanges

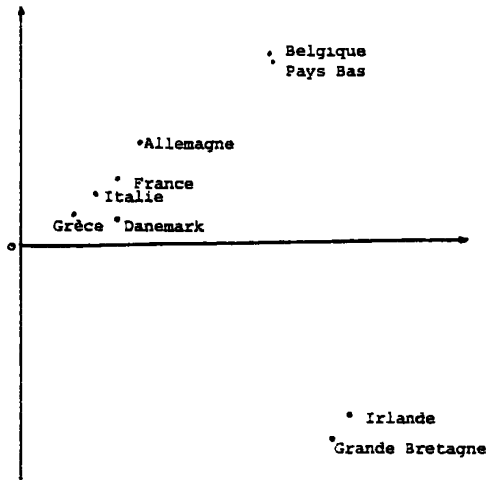


Figure 4.a
Partie symétrique

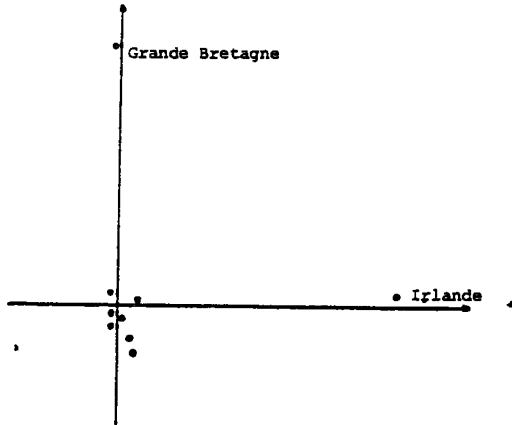


Figure 4.a
Partie anti-symétrique

méthode de Constantine et Gower
matrice de structure des échanges

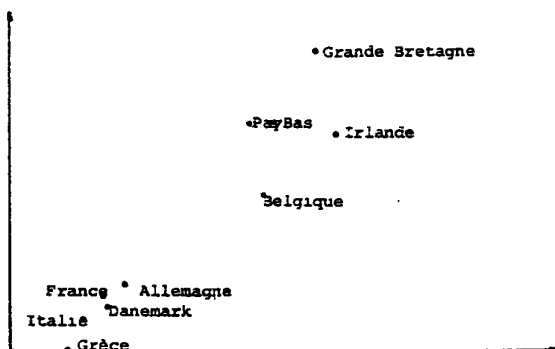


Figure 5

matrice de structure d'échanges

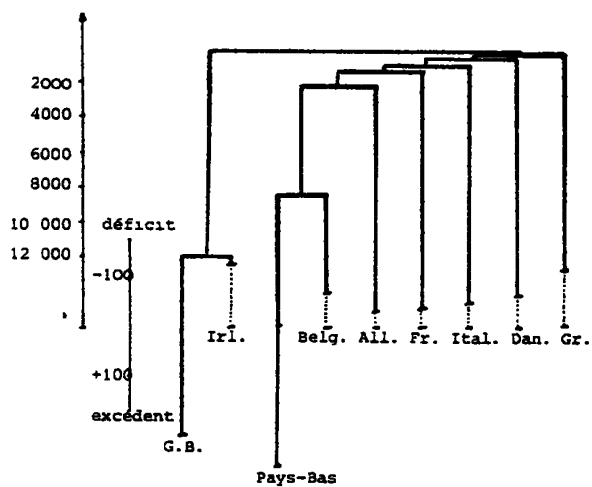


Figure 6

Classification

matrice de structure des échanges

XI - BIBLIOGRAPHIE

- (1) BROSSIER G. 1980 - Représentation ordonnée des classifications hiérarchiques - statistique et analyse des données 2 - 1980 - pp. 31-44.
- (2) CHANDON J.L. et 1980 - Analyse typologique - Masson ed.Paris
PINSON S.
- (3) CONSTANTINE A.G. et 1978 - Graphical representation of asymmetric
GOWER J.C. matrices - Applied statistics - 27 n°3 pp.297-304
- (4) CUNNINGHAM J.P. 1978 - Free trees and bidirectional trees as
representation of psychological distances -
Journal of mathematical psychology. 17, pp 165-188
- (5) ESCOUFIER Y. et 1979 - Analyse factorielle des matrices carrées
GRORUD A. non symétriques - 2èmes Journées Internationales -
Analyse des données et informatique - IRIA - sept.79
Versailles.
- (6) GOWER J.C. 1977 - The analysis of asymetry and orthogonality,
in "Recent developments in statistics", Barra ed.
North-Holland.