



HAL
open science

Nouvelles représentations graphiques en classification automatique .Donnees chaotiques et donnees coherentes

Edwin Diday

► **To cite this version:**

Edwin Diday. Nouvelles représentations graphiques en classification automatique .Donnees chaotiques et donnees coherentes. RR-0569, INRIA. 1986. inria-00075985

HAL Id: inria-00075985

<https://inria.hal.science/inria-00075985>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

IRIA

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France

Tél. (1) 39 63 55 11

Rapports de Recherche

N° 569

NOUVELLES REPRÉSENTATIONS GRAPHIQUES EN CLASSIFICATION AUTOMATIQUE

DONNÉES CHAOTIQUES ET DONNÉES COHÉRENTES

Edwin DIDAY

Septembre 1986

NOUVELLES REPRESENTATIONS GRAPHIQUES EN CLASSIFICATION AUTOMATIQUE.**NEW GRAPHICAL REPRESENTATIONS IN CLUSTERING****E. DIDAY.****RESUME.**

Partant d'un indice de dissimilarité quelconque, nous proposons d'abord une nouvelle forme de représentation complémentaire aux ultramétriques et appelés ultramines; cette représentation peut produire des informations intéressantes : par exemple, être plus fidèle aux données initiales et faire ressortir les contrastes beaucoup mieux que les ultramétriques. On présente ensuite la notion de "courbe compatible" qui permet une représentation hiérarchique curviligne alors qu'elle était seulement rectiligne jusqu'à présent; cette nouvelle représentation facilite l'expression de classifications sous contrainte d'une autre ultramétrie d'une ultramine ou d'un indice de dissimilarité quelconque.

On introduit enfin la notion de données "chaotiques" ou "cohérentes" et des critères permettant de savoir si les données seront plus fidèles à une ultramétrie (resp. un indice pyramidal) ou à une ultramine (resp. un indice anti-pyramidal).

ABSTRACT.

Having chosen a dissimilarity index, hierarchical clustering is one of the most popular techniques used to represent clusters ; in fact, the fit between the graphical representation of hierarchies (i.e. dendrograms) and the data is very good only when the two following conditions are satisfied.

- 1) The dissimilarity index satisfies the "ultrametric inequality".
- 2) There exists an order which is "compatible" with the dissimilarity index". This paper is devoted to new kinds of graphical representation when those conditions are not satisfied. When the first condition is not satisfied at all, we introduce a complementary notion of ultrametric called "ultramine". When the second condition is not satisfied at all, a new dissimilarity index called "anti-pyramidal" which is an extension of ultramine is considered. We introduce the notion of "compatible curve" which is useful to represent hierarchies under constraints. Graphical representations of those notions are given. We give a measure which indicate how much the data are ultrametric or ultramine and order compatible or not for the chosen dissimilarity index. Each one of the new graphical representations sheds new and interesting light on the data.

1. INTRODUCTION

Avec l'amélioration rapide des outils graphiques, les aspects graphiques prennent une importance croissante en traitement des données ; en classification automatique, les hiérarchies par la commodité de leur interprétation visuelle constituent depuis longtemps une forme de représentation très populaires.

De nombreux ouvrages ont été consacrés aux hiérarchies : le lecteur pourra se reporter par exemple à Benzecri et col. (1980), Diday et al (1982), Jambu (1978), Lerman (1981) et Sokal et Sneath (1963).

Les ultramétriques ont été depuis longtemps l'outil privilégié pour approximer les données initiales par une hiérarchie puisque l'on a montré qu'il existe une bijection entre ultramétriques et hiérarchies indicées. En regardant les choses de plus près, on s'aperçoit cependant que les ultramétriques ne sont fidèles qu'à une classe particulière de données ; en effet, on peut montrer que la matrice des distances associées à une ultramétrique peut toujours être ordonnée de façon à être de type Robinson (lignes et colonnes croissantes à partir de la diagonale principale, voir Hubert (1974)) ; il en résulte que la matrice des distances données par l'utilisateur sera d'autant plus fidèle à une ultramétrique qu'elle satisfait un tel ordre. Si les données initiales sont très éloignées de cette condition, les ultramétriques sont inaptes à les représenter ; on s'est intéressé à cette problématique ; une nouvelle classe de mesures de ressemblances graphiquement représentable appelées "ultramines" a été définie ; elle est apte à représenter des données "non ultramétriques", de type "antirobinson" (lignes et colonnes décroissantes à partir de la diagonale principale).

Dans la figure 1 (tétraèdre écrasé) on représente une configuration de 4 points abcd représentables par une ultramine et non une ultramétrique la figure 2 présente un cas inverse.

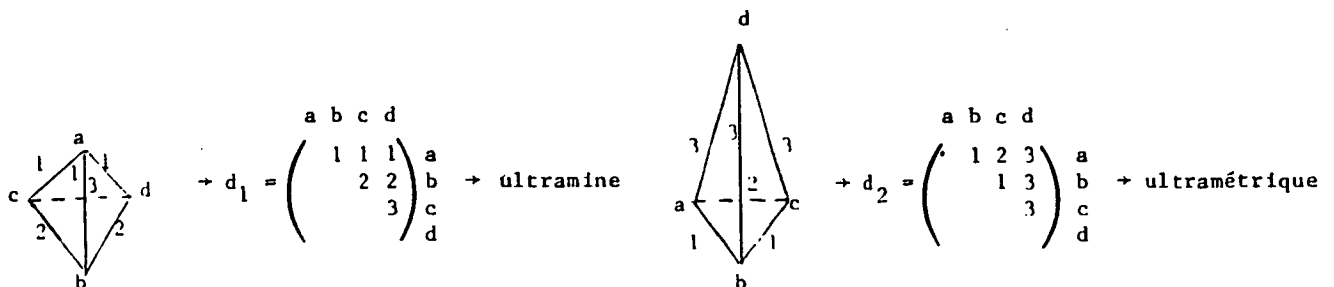


Figure 1

Figure 2

Les propriétés théoriques des ultramines ont été largement étudiées par Batbedat (1984) dans le cas où elles sont réduites à des indices de "dissimilarités" infra-isocèles ; la plupart d'entre elles se démontrent facilement de façon "duale" aux propriétés classiques des ultramétriques. B. Monjardet m'a signalé que les ultramines ont été appelées "relation d'équivalence floue" si on restreint leurs valeurs dans l'intervalle $[0,1]$ (dans le cas où l'ultramine ne peut prendre que les valeurs 0 ou 1, on montre facilement qu'elle définit une relation d'équivalence).

Dans toute la suite on considère un ensemble $\Omega = \{w_1, \dots, w_n\}$ d'individus muni d'un indice de dissimilarité ; autrement dit, d'une application $d : \Omega \times \Omega \rightarrow \mathbb{R}^+$, symétrique et telle $d(w,w) = 0$. Si $d(w,w) \geq d(w,w') \quad \forall w, w' \in \Omega$, d est un indice de similarité. On appelle mesure (ou indice) de ressemblance tout indice qui appartient à l'ensemble réuni des indices de similarité et de dissimilarité. On définit d'abord les ultramines ; il résulte de cette définition que ce sont des indices de similarité dont l'une des propriétés caractéristiques est que tous les triangles sont isocèles avec la base plus grande que les côtés. Cela revient à dire que chaque côté de n'importe quel triangle est plus grand* que le plus petit (le Min) des deux autres d'où l'appellation "d'ultramine". Afin de faciliter l'étude des propriétés ultérieures on restreint les ultramines δ à l'ensemble $M(X)$ des ultramines qui prennent une même valeur $\delta(w,w)$ pour chaque élément de Ω . On en déduit que les ultramétriques et les ultramines restreintes à $M(X)$ sont en bijection. On montre ensuite que l'enveloppe inférieure des ultramines supérieures à l'indice de dissimilarité initiale est une ultramine ce qui n'est pas le cas des ultramétriques ; il en résulte un encadrement de l'indice initial par la sous-dominante ultramétrique et la surdominante ultramine. Cet indice initial peut être aussi bien un indice de similarité qu'un indice de dissimilarité ; il peut paraître étonnant de vouloir approximer un indice de dissimilarité par une ultramine qui est un indice de similarité ; en fait, cette approximation est intéressante car ce sont surtout les termes différents qui jouent un rôle important dans la représentation et son interprétation ; d'autre part, il est toujours possible de transformer artificiellement un indice de dissimilarité en une similarité en conservant les valeurs de la dissimilarité pour les couples d'éléments distincts et en remplaçant chaque terme de la diagonale de la matrice de dissimilarité par le plus grand terme de la ligne et de la colonne correspondante. Pour construire la surdominante ultramine d'une mesure de ressemblance on montre

* au sens large.

qu'il suffit de prendre l'inverse de la sous-dominante de l'inverse de cette mesure. On montre ensuite que la surdominante ultramétrique d'un indice de dissimilarité est plus fidèle aux oppositions alors que la sous-dominante ultramétrique est plus fidèle aux ressemblances.

L'étude du lien entre les ultramétriques et les ordres sur Ω montre que pour toute ultramétrique u il existe un ordre O tel que la matrice des distances $M(u, O)$ de terme générique $u(w_i, w_j)$ et dont les lignes et colonnes respectent l'ordre O est anti-robinson. On en déduit que l'on peut plonger l'ensemble des ultramétriques dans un ensemble plus vaste qui est celui des indices antipyramidaux. Ces résultats permettent de donner une représentation graphique des ultramétriques et des indices antipyramidaux sous forme de hiérarchies ou de pyramides anti-indicées. En profitant de la propriété d'ordre des ultramétriques (existence d'un ordre compatible) et des ultramétriques on étudie aussi la possibilité de représentation sous forme curviligne. On énonce d'abord le problème sous sa forme analytique puis on propose une euristique permettant de représenter les éléments sur une "courbe compatible" respectant l'ordre et les distances. Comme pour les hiérarchies et les pyramides on est amené à recourir à un anti-indicage pour représenter une ultramétrique ou un indice anti-pyramidale sur une courbe compatible. On remarque que la représentation visuelle habituelle d'une hiérarchie est rectiligne puisqu'elle s'appuie sur une droite. En remplaçant la droite, qui est une courbe compatible particulière, par une courbe compatible quelconque exprimant une contrainte externe, on obtient une "hiérarchie curviligne" d'interprétation plus riche. On introduit ensuite la notion de données chaotiques ou cohérentes pour un indice de dissimilarité s ; on montre que s'il existe 4 individus dont les dissimilarités sont toutes différentes, il ne peut exister simultanément un ordre anticompatible et un ordre compatible avec s . S'il existe un ordre anticompatible avec s on dira que les données sont chaotiques (c'est donc le cas quand s est une ultramétrique ou un indice antipyramidale) ; s'il existe un ordre compatible avec s on dira que les données sont cohérentes (c'est le cas quand s est une ultramétrique ou un indice pyramidale) ; on peut obtenir des chaînes ou des agrégats fidèles aux données initiales dans le cas de données cohérentes ; par contre, dans le cas de données chaotiques ce sont les classes d'éléments opposés qui sont "fidèles" aux données initiales. On donne enfin des indicateurs permettant de mesurer le degré de chaotisation ou de cohérence des données pour un indice de dissimilarité.

2. DEFINITION ET PROPRIETES D'UNE ULTRAMINE

2.1. Définition

Une ultramine est une application u de $\Omega \times \Omega \rightarrow \mathbb{R}^+$ qui satisfait aux deux conditions suivantes :

$$i) \forall w_i, w_j \in \Omega \quad u(w_i, w_j) = u(w_j, w_i) \text{ (symétrie).}$$

$$ii) \forall w_i, w_j, w_k \in \Omega \quad u(w_i, w_j) \geq \text{Min} \{u(w_i, w_k), u(w_k, w_j)\} \text{ (inégalité ultramine).}$$

Comme première conséquence de cette définition, on voit qu'une ultramine n'est pas une métrique puisque l'on peut avoir des triangles dont un côté est plus grand que la somme des deux autres. Une seconde conséquence importante concerne les couples d'éléments identiques, on a en effet la relation : $\forall w_i, w_j \in \Omega$, $u(w_i, w_i) \geq u(w_i, w_j)$ puisque d'après l'inégalité ultramine on a :

$$\forall w_i, w_j \in \Omega, u(w_i, w_i) \geq \text{Min} \{u(w_i, w_j), u(w_j, w_i)\} = u(w_i, w_j).$$

Il en résulte qu'une ultramine est un indice de similarité.

2.2. Une restriction simplicatrice de l'ensemble des ultramines : l'ensemble $M(X)$

On ne perd pas beaucoup en généralité en identifiant les ultramines qui prennent les mêmes valeurs sur tous les couples d'éléments différents ; autrement dit, en ne considérant que l'ensemble M des ultramines satisfaisant à la condition :

$$\forall w_i, w_j \in \Omega \text{ tel que } w_i \neq w_j \text{ on a : } u(w_i, w_j) = u'(w_i, w_j) \text{ implique } u = u'.$$

On peut dire aussi que l'ensemble M est l'espace quotient des ultramines par la relation d'équivalence $u \sim u' \Leftrightarrow \{\forall (w_i, w_j) \text{ avec } w_i \neq w_j, u(w_i, w_j) = u'(w_i, w_j)\}$.

On note $M(X)$ l'ensemble inclus dans M , des ultramines u dont la valeur pour les couples d'éléments identiques est définie par $X : (u(w_1, w_1), \dots, u(w_n, w_n))$.

Autrement dit, $\forall u, u' \in M(X)$ on a pour tout i $u(w_i, w_i) = u'(w_i, w_i)$ on pose $X_i = u(w_i, w_i)$ et donc $X = (X_1, \dots, X_n)$.

La restriction des ultramines à un espace $M(X)$ permettra de prouver qu'il existe une bijection entre les ultramétries et les ultramines (voir proposition 1) ; l'égalité des $u(w_i, w_i)$ pour i fixé quand u varie sera aussi nécessaire pour prouver l'existence et donc l'unicité de la surdominante ultramine (voir la proposition 2).

Proposition 1

. Une condition nécessaire et suffisante pour qu'un indice de similarité soit une ultramine est que tous les triangles de Ω soient isocèles avec la base plus grande que les côtés.

. Il existe une bijection entre un ensemble $M(X)$ d'ultramines et les ultramétries.

. Toute ultramine satisfait à l'inégalité :

$$(2) \quad u(w_i, w_k) \geq \text{Min} \{u(w_i, w_{i+1}), u(w_{i+1}, w_{i+2}), \dots, u(w_{k-1}, w_k)\}$$

Démonstration

Tous les triangles de Ω munis d'une ultramine sont isocèles avec la base plus grande que les côtés ; en effet, étant donné un triangle quelconque (i, j, k) , considérons le plus petit côté, d'après l'inégalité ultramine il est plus grand que les deux autres ; réciproquement, si tous les triangles sont isocèles avec la base plus grande que les côtés on voit que l'inégalité ultramine est satisfaite ; dans le cas où deux sommets du triangle sont confondus l'inégalité est encore vraie puisqu'il s'agit d'un indice de similarité.

. Il existe une bijection entre l'ensemble des ultramétries U et l'ensemble des ultramines $M(X)$. En effet, soit f l'application $U \rightarrow M(X)$ telle que $f(u) = \delta$ avec $\delta(w_i, w_j) = (u(w_i, w_j))^{-1}$ si $w_i \neq w_j$ et $\delta(w_i, w_i) = X_i$. $f(u) = \delta$ est bien une ultramine puisque tout triangle isocèle avec la base plus petite que les côtés se transforme en un triangle isocèle avec la base plus grande que les côtés ; elle est surjective puisque tout triangle isocèle avec la base plus grande que les côtés (même s'il ne vérifie pas l'inégalité triangulaire !) peut être obtenu à partir d'un triangle

isocèle avec la base plus petite que les côtés ; l'application est injective puisque l'on s'est restreint aux ultramines qui sont identiques dès qu'elles prennent les mêmes valeurs sur les couples d'éléments différents.

Enfin l'inégalité (2) se démontre facilement par récurrence en partant de l'inégalité ultramine.

3. CALCUL DES ULTRAMINES

3.1. Surdominante ultramine

3.1.1. Définition et propriétés

Etant donné un ensemble d'ultramines $M(X)$ et une mesure de ressemblance d , la surdominante ultramine u_d de d dans $M(X)$ est la borne inférieure des ultramines de $M(X)$ supérieures à d .

Autrement dit $u_d = \inf \{u / u \geq d, u \in M(X)\}$, (on dit que $u \geq d$ si $\forall w_i, w_j \in \Omega$ $u(w_i, w_j) \geq d(w_i, w_j)$).

Remarquons que cette définition n'a de sens que si $\forall i=1, \dots, n$ on a $d(w_i, w_i) \leq X_i$ dans toute la suite d étant donné $X = (X_1, \dots, X_n)$ sera choisi de cette façon ; dans la pratique on calculera d'abord la surdominante ultramine sur les couples d'éléments différents, puis on ajustera au mieux les termes diagonaux de la matrice de ressemblance associée à la surdominante ultramine de façon à s'écarter le moins possible des $d(w_i, w_i)$ tout en restant dans $M(X)$; plus précisément si $d(w_i, w_i)$ est supérieur à tous les $d(w_i, w_j)$ pour w_i fixé on prendra $u_d(w_i, w_i) = d(w_i, w_i)$ dans le cas contraire on prendra $u_d(w_i, w_i) = \max_j d(w_i, w_j)$.

Il résulte facilement de cette définition que la surdominante ultramine (comme la surdominante ultramétrique) existe et est unique ; cependant, contrairement à la surdominante ultramétrique qui n'est pas nécessairement une ultramétrique, la surdominante ultramine est une ultramine comme le prouve le résultat suivant.

Proposition 2

La surdominante ultramine u_d d'une mesure de ressemblance d est une ultramine.

Démonstration

L'ensemble des $u(w_i, w_j)$ pour i et j fixé et u variant dans $M(X)$ étant minoré par d admet une borne inférieure ; soit $u_d(w_i, w_j)$ cette borne ; il existe donc une suite $\delta_n(w_i, w_j)$ avec $\delta_n \in M(X)$ qui converge vers $\delta_n(w_i, w_j)$. Par passage à la limite de cette suite, on voit facilement que $u_d(w_i, w_j) = u_d(w_j, w_i)$ pour tout couple de Ω et que $u_d(w_i, w_i) = X_i$ pour $i=1, \dots, n$. Pour que u_d soit dans $M(X)$ il reste à vérifier l'inégalité ultramétrique. Cette inégalité est satisfaite car on a successivement les relations suivantes :

$$\begin{aligned} u_d(w_i, w_j) &= \text{Min} \{ u(w_i, w_j) / u \geq d, u \in M(X) \} \\ &\geq \text{Min} \{ \text{Min} (u(w_i, w_k), u(w_k, w_j)) / u \geq d, u \in M(X) \} \\ &\geq \text{Min} \{ \text{Min} (u(w_i, w_k) / u \geq d, u \in M(X)), \text{Min} \{ u(w_k, w_j) / u \geq d, u \in M(X) \} \} \\ &\geq \text{Min} \{ u_d(w_i, w_k), u_d(w_k, w_j) \} \quad \square \end{aligned}$$

D'autres propriétés de la surdominance analogues ou complémentaires de celles obtenues dans le cas des ultramétriques (voir Diday et al (1982)) peuvent être utiles :

. Si D est un indice qui mesure la "distance" entre deux mesures de ressemblance telle que :

$$D(d, \delta) = \left[\sum_{i \in \Omega} \sum_{j \in \Omega} |d(i, j) - \delta(i, j)|^\alpha \right]^{\frac{1}{\alpha}} \text{ avec } \alpha > 0, \text{ il résulte}$$

facilement de la proposition précédente que :

$$\inf \{ D(d, \delta) / \delta \in M(X), \delta \geq d \} = D(d, u_d).$$

. D'autre part, on sait que si δ_d est la sous-dominante de l'ensemble des ultramétriques U on a :

$$\inf \{ D(d, \delta) / \delta \in U, \delta < d \} = D(d, \delta_d).$$

On en déduit un encadrement d'une mesure de ressemblance d par deux solutions optimales u_d et δ_d sur les sous-espaces qui les concernent :

$$\delta_d \leq d \leq u_d$$

La borne inférieure des ultramétriques supérieures n'existe généralement pas, on peut néanmoins déduire d'une hiérarchie du saut maximum une ultramétrique supérieure à d qui l'approxime d parfois mieux que δ_u comme le montrera l'exemple 2 ci-dessous. D'autre part comme il existe une bijection d'après la proposition 2 entre les ultramétriques et les ultramines, leur nombre est le même et donc on peut supposer que de façon générale on approche la mesure de ressemblance initiale avec la surdomine par dessus d'aussi près qu'avec la sous dominante par dessous (sauf bien sûr pour les couples d'éléments identiques, l'avantage allant par exemple à la sous-dominante s'ils sont à distance nulle).

3.1.2. Construction de la surdominante ultramétrique

On peut construire la surdomine en rendant successivement tous les triangles isocèles avec la base plus grande que les côtés (on égalise les 2 plus petits côtés au plus grand) et on recommence le procédé tant que tous les triangles ne satisfont pas cette propriété.

Un autre algorithme plus efficace consiste à utiliser la sous-dominante de "l'inverse" de l'indice de ressemblance à représenter comme nous allons maintenant le montrer.

On note D l'ensemble des indices de dissimilarité sur Ω et $D^{-1}(X)$ l'ensemble des inverses d^{-1} des éléments $d \in D$ dont les valeurs pour les couples d'éléments identiques sont définis par $X = (X_1, \dots, X_n)$ avec $X_i = d^{-1}(w_i, w_i)$.

Exemple 1

Soit $d \in D$ un indice de dissimilarité et $\Omega = (w_1, w_2, w_3)$ tel que la matrice de dissimilarité associée à d soit

$$d = \begin{array}{ccc} & w_1 & w_2 & w_3 \\ \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 4 \\ 2 & 4 & 0 \end{pmatrix} & w_1 \\ & w_2 \\ & w_3 \end{array}$$

si $d^{-1} \in D^{-1}(X)$ avec $X = (1, 2, 5)$ on a alors :

$$d^{-1} = \begin{pmatrix} 1 & 1 & 0,50 \\ 1 & 2 & 0,25 \\ 0,50 & 0,25 & 5 \end{pmatrix}$$

Soit $U^{-1}(X) \subset D^{-1}(X)$ l'ensemble des ultramétries inverses sur Ω dont les valeurs pour les éléments identiques sont définies par X . On a le résultat suivant :

Proposition 3

L'ensemble des ultramétries inverses $U^{-1}(X)$ est identique à l'ensemble des ultramines $M(X)$.

Démonstration

Selon la proposition 1 on sait qu'il existe une bijection entre l'ensemble des ultramétries U et $M(X)$, comme il existe une bijection entre $U^{-1}(X)$ et U il en résulte que $U^{-1}(X)$ et $M(X)$ sont en bijection. Chaque élément de $U^{-1}(X)$, étant un triangle isocèle avec la base plus grande que les côtés, est un élément de $M(X)$; de plus dans $U^{-1}(X)$ comme dans $U(X)$ la ressemblance des termes identiques est définie par X donc la bijection est en fait une identité. \square

De cette proposition, on déduit le résultat suivant qui permet de construire la surdominante ultramine à l'aide d'une sous-dominante ultramétrique.

Proposition 4

La surdominante ultramine u_d d'une mesure de ressemblance d pour un ensemble d'ultramines $M(X)$ est l'inverse de la sous-dominante de $d^{-1} \in D^{-1}(X)$.

Démonstration

Par définition la sous-dominante de d^{-1} est

$$\delta_{d^{-1}} = \sup \{u/u \leq d^{-1}, u \in U\}, \text{ or comme il existe une bijection entre } U \text{ et}$$

$U^{-1}(X)$, on peut écrire :

$$\delta_{d^{-1}}^{-1} = \inf \{u^{-1}/u^{-1} \geq d, u^{-1} \in U^{-1}(X)\}.$$

D'après la proposition 4 on sait que $U^{-1}(X)$ et $M(X)$ sont en bijection d'où :

$\delta_{d^{-1}}^{-1} = \inf \{u/u \geq d, u \in M(X)\}$ ce qui prouve bien que $\delta_{d^{-1}}^{-1} = u_d$ la surdominante ultramétrique de d pour $M(X)$. □

Une autre façon de calculer la surdominante de d est donnée par la proposition suivante ; soit $C(i,j)$ une chaîne $w_i w_{i+1}, w_{i+1} w_{i+2}, \dots, w_{j-1} w_j$ qui relie w_i à w_j dans Ω et soit $f(C(i,j)) = \min\{d(w_\ell, w_{\ell+1}) / \ell=i, \dots, j-1\}$.

Autrement dit, on considère que la "longueur" d'une chaîne est la "longueur" de sa plus courte arête.

On a alors le résultat suivant :

Proposition 5

Si u est une mesure de ressemblance définie sur Ω telle que $u(w_i, w_j) = \max \{f(C) / C \in C(i,j)\}$ où $C(i,j)$ est l'ensemble des chaînes qui relient w_i à w_j dans Ω et pour $i=1, \dots, n$ $u(w_i, w_i) = X_i$, alors $u = u_d$ la surdominante ultramétrique de d dans $M(X)$.

Démonstration

Montrons d'abord que u est une ultramétrie : pour cela il suffit de remarquer que $f(C_{ij}^*) \geq f(C_{i\ell}^* \cup C_{\ell j}^*)$ où C_{ij}^* est parmi les chaînes qui relient i à j celle dont la plus courte arête est la plus grande. Comme $f(C_{i\ell}^* \cup C_{\ell j}^*) = \min \{f(C_{i\ell}^*), f(C_{\ell j}^*)\}$ on a $f(C_{ij}^*) \geq \min \{f(C_{i\ell}^*), f(C_{\ell j}^*)\}$ d'où $u(w_i, w_j) \geq \min \{u(w_i, w_\ell), u(w_\ell, w_j)\}$. Donc u est une ultramétrie.

Il reste à vérifier que u est la surdominante. On remarque d'abord que $u(w_i, w_j) \geq d(w_i, w_j) \forall w_i, w_j \in \Omega$ car la chaîne qui relie directement w_i à w_j a pour "longueur" $d(w_i, w_j)$. Enfin, d'après la proposition 2, toute ultramétrie satisfait à l'inégalité (2) et en particulier la surdominante ultramétrique d'où :

$$u_d(w_i, w_j) \geq \min \{u_d(w_i, w_{i+1}), \dots, u_d(w_{j-1}, w_j)\} \geq u(w_i, w_j)$$

car la chaîne $w_i w_{i+1}, \dots, w_{j-1} w_j$ n'est pas forcément celle dont la plus courte arête est la plus grande, on en déduit que $u_d \geq u \geq d$ et donc que $u = u_d$. □

On peut donc dire que $u(w_i, w_j)$ est la plus grande arête parmi les plus courtes arêtes des chaînes qui relient w_i à w_j .

Il en résulte que la surdominante ultramétrique comme la sous-dominante ultramétrique est sujette à un "effet de chaîne". Pour la surdominante ultramétrique il s'exprime de la façon suivante : s'il existe une chaîne qui relie deux éléments de Ω dont la plus courte arête est "grande", les deux points sont considérés comme éloignés au sens de la surdominante ultramétrique même s'ils sont proches au sens de la mesure de ressemblance initiale (voir figure 3 où d est la distance euclidienne du plan).

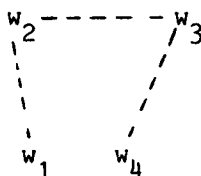


Figure 3

$$u_d(w_1, w_4) = d(w_2, w_3) \gg d(w_1, w_4)$$

$$\delta_d(w_2, w_3) = d(w_1, w_4) \ll d(w_2, w_3)$$

En règle générale la surdominante ultramétrique déforme les petites distances surtout s'il existe des éléments très éloignés mais respecte les distances des éléments éloignés ; ceci n'est pas le cas de la sousdominante ultramétrique qui elle, déforme les distances des éléments éloignés dès qu'il existe une chaîne d'éléments successivement proches qui les relie, par contre elle reste fidèle aux éléments proches. On peut résumer ces propriétés en disant que "la surdominante ultramétrique représente bien les oppositions alors que la sousdominante ultramétrique représente bien les ressemblances".

Exemple 2

Soit $\Omega = \{i, j, k, l\}$ et d défini par la matrice de dissimilarité

$$M(d, \theta) = \begin{matrix} & \begin{matrix} i & j & k & l \end{matrix} \\ \begin{pmatrix} 0 & 1 & 4 & 3 \\ & 0 & 2 & 5 \\ & & 0 & 6 \\ & & & 0 \end{pmatrix} & \begin{matrix} i \\ j \\ k \\ l \end{matrix} \end{matrix}$$

où $\mathbf{0}$ est l'ordre des colonnes (i.e. des lignes).

On a donc

$$M(d^{-1}, \mathbf{0}) = \begin{pmatrix} X_1 & 1 & 1/4 & 1/3 \\ & X_2 & 1/2 & 1/5 \\ & & X_3 & 1/6 \\ & & & X_4 \end{pmatrix} \begin{matrix} i \\ j \\ k \\ l \end{matrix}$$

La sous-dominante de d^{-1} dans $M(X)$ se déduit facilement de la hiérarchie du saut minimum construite à partir de d^{-1} , on obtient :

$$M(\delta_d^{-1}, \mathbf{0}) = \begin{pmatrix} X_1 & 1/4 & 1/4 & 1/4 \\ & X_2 & 1/5 & 1/5 \\ & & X_3 & 1/6 \\ & & & X_4 \end{pmatrix} \begin{matrix} i \\ j \\ k \\ l \end{matrix}$$

D'après la proposition 5, la surdominante ultramétrique est

$$M(u_d, \mathbf{0}) = \begin{pmatrix} X_1 & 4 & 4 & 4 \\ & X_2 & 5 & 5 \\ & & X_3 & 6 \\ & & & X_4 \end{pmatrix} \begin{matrix} i \\ j \\ k \\ l \end{matrix}$$

On peut maintenant choisir les X_i d'après la remarque faite en 3.1.1 : $X_1 = 4$, $X_2 = 5$, $X_3 = X_4 = 5$. On aurait obtenu le même résultat en utilisant la "technique des chaînes" définie par la proposition 6. La sousdominante ultramétrique de d étant

$$M(\delta_d, \mathbf{0}) = \begin{pmatrix} 0 & 1 & 2 & 3 \\ & 0 & 2 & 3 \\ & & 0 & 3 \\ & & & 0 \end{pmatrix} \begin{matrix} i \\ j \\ k \\ l \end{matrix}$$

On vérifie bien que la surdominante ultramétrique respecte bien les grandes distances : $d(w_k, w_l)$, $d(w_l, w_j)$, $d(w_i, w_k)$ alors que la sousdominante respecte surtout les petites : $d(w_i, w_j)$, $d(w_j, w_k)$, $d(w_i, w_l)$.

3.2. Calcul d'autres ultramines

En utilisant la proposition 4 qui indique que l'inverse d'une ultramétrie δ dans $D^{-1}(X)$ est une ultramine de $M(X)$, on peut calculer une ultramétrie à l'aide de l'un des indices d'agrégation usuels puis inverser cette ultramétrie de façon à obtenir une ultramine de $M(X)$.

En s'inspirant de la technique qui permet de calculer la surdominante ultramine grâce à la proposition 4, une autre stratégie consiste d'abord à inverser la mesure de ressemblance initiale puis à calculer une ultramétrie à l'aide de l'un des indices usuels à l'aide de cette mesure inverse, on inverse enfin cette ultramétrie pour obtenir l'ultramine désirée.

On peut alors se poser la question de savoir si les deux stratégies sont équivalentes. L'exemple suivant montre qu'elles ne le sont pas.

Exemple

$$\Omega = \{i, j, k\}, M(d, 0) = \begin{pmatrix} 0 & 1 & 2 \\ & 0 & 1 \\ & & 0 \end{pmatrix} \begin{matrix} i \\ j \\ k \end{matrix}$$

si l'on décide d'utiliser l'indice d'agrégation du saut maximum on obtient deux solutions (voir figure 4) :

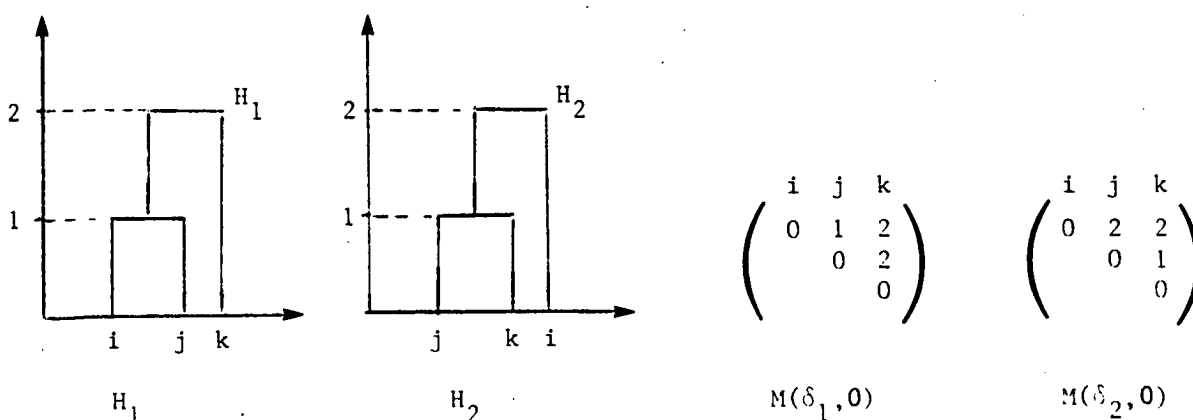


Figure 4

Les deux hiérarchies du saut maximum et les ultramétriques qu'elles induisent.

Il correspond aux deux hiérarchies H_1 et H_2 , deux ultramétries δ_1 et δ_2 (qui prouvent d'ailleurs que la surdominante ultramétrique n'est pas une ultramétrie puisqu'on voit facilement que c'est d). Les inverses de δ_1 et δ_2 dans $M(X)$ sont définies par :

$$M(\delta_1^{-1}, 0) = \begin{pmatrix} X_1 & 1 & 1/2 \\ & X_2 & 1 \\ & & X_3 \end{pmatrix} \begin{matrix} i \\ j \\ k \end{matrix} \quad M(\delta_2^{-1}, 0) = \begin{pmatrix} X_1 & 1/2 & 1/2 \\ & X_2 & 1 \\ & & X_3 \end{pmatrix} \begin{matrix} i \\ j \\ k \end{matrix}$$

δ_1^{-1} et δ_2^{-1} sont deux ultramines de $M(X)$.

En utilisant la seconde stratégie on a :

$$M(d^{-1}, 0) = \begin{pmatrix} X_1 & 1 & 1/2 \\ & X_2 & 1 \\ & & X_3 \end{pmatrix} \begin{matrix} i \\ j \\ k \end{matrix}$$

avec l'indice du saut maximum sur d^{-1} on obtient une ultramétrie δ :

$$M(\delta, 0) = \begin{pmatrix} X_1 & 1 & 1/2 \\ & X_2 & 1 \\ & & X_3 \end{pmatrix} \begin{matrix} i \\ j \\ k \end{matrix}$$

d'où l'ultramétrie δ^{-1} telle que :

$$M(\delta^{-1}, 0) = \begin{pmatrix} X_1 & 1 & 2 \\ & X_2 & 1 \\ & & X_3 \end{pmatrix} \begin{matrix} i \\ j \\ k \end{matrix}$$

qui est différente de δ_1^{-1} et δ_2^{-1} .

Les deux stratégies sont donc différentes et l'on peut se poser la question de savoir laquelle est la "meilleure". La seconde stratégie est certainement la meilleure au sens où elle donne une ultramétrie dont les valeurs prises sont plus proches de la mesure de ressemblance initiale d ; en effet, une ultramétrie est surtout fidèle aux faibles distances et déforme généralement les grandes ; en prenant l'inverse d'une ultramétrie calculée avec un indice d'agrégation classique on obtient donc une ultramétrie qui n'est ni fidèle aux petites distances (puisque l'on prend l'inverse de ces distances) ni aux grandes. Par contre la seconde stratégie donne une ultramétrie qui reste fidèle aux grandes distances.

Exemple 4

Reprenons l'exemple 2 où c'est l'indice d'agrégation du saut minimum qui avait été choisi, la sous-dominante δ_d est définie par

$$M(\delta_d, \mathbf{0}) = \begin{pmatrix} 0 & 1 & 2 & 3 \\ & 0 & 2 & 3 \\ & & 0 & 3 \\ & & & 0 \end{pmatrix}$$

donc l'écart avec d par la première stratégie est :

$$|M(d, \mathbf{0}) - M(\delta_d^{-1}, \mathbf{0})| = \frac{1}{2} \sum_{i,j} |d(i,j) - \delta_d^{-1}(i,j)| = 19$$

Alors que la deuxième stratégie donne l'ultramétrie u_d et

$$|M(d, \mathbf{0}) - M(u_d, \mathbf{0})| = 7 < 19.$$

Avec l'exemple 3 où l'indice d'agrégation choisi est celui du saut maximum la deuxième stratégie donne un écart nul, ce qui n'est pas le cas de la première.

4. ORDRE ASSOCIE A UNE ULTRAMETRIE4.1. Ordres anticompatibles et matrices antirobinson

Nous allons montrer qu'à toute ultramétrie, on peut associer un ordre anticompatible et une matrice antirobinson qui se définissent de la façon suivante.

Définition d'un ordre anticompatible

Etant donné une mesure de ressemblance d , un ordre \mathbf{O} sur Ω est dit anticompatible si $w_i < w_j < w_k$ au sens de \mathbf{O} implique $d(w_i, w_k) \leq \min(d(w_i, w_j), d(w_j, w_k))$ (1).

Définition d'une matrice Robinson et antirobinson

Une matrice est dite Robinson (resp. antirobinson) si elle est symétrique et si ses lignes et colonnes croissent (resp. décroissent) à partir de la diagonale principale comprise.

De ces définitions on déduit immédiatement la proposition suivante :

Proposition 6

Les deux propriétés suivantes sont équivalentes :

- 1) d et $\mathbf{0}$ sont anticompatibles et $\forall w_i, w_j \in \Omega \quad d(w_i, w_j) \leq d(w_i, w_i)$
- 2) $M(d, \mathbf{0})$ est antirobinson.

Démonstration

On voit facilement que s'il existe un ordre anticompatible $\mathbf{0}$ pour une mesure de ressemblance d , la matrice de similarité associée $M(d, \mathbf{0})$ est anticompatible ; réciproquement, si $M(d, \mathbf{0})$ est antirobinson alors l'ordre des lignes et colonnes est anticompatible ; en effet, il suffit de remarquer que le terme $d(w_i, w_j)$ est supérieur à tous les termes qui la précèdent sur la ligne $d(w_i, w_\ell)$ avec $i \leq \ell \leq j$ qui le suivent sur la même colonne ($d(w_\ell, w_j)$ avec $i \leq \ell \leq j$) et qui suivent cette ligne et précèdent cette colonne ($d(w_\ell, w_k)$ avec $i \leq \ell \leq k < j$). \square

4.2. Existence d'un ordre anticompatible pour une matrice ultramine.

Nous allons montrer qu'à toute ultramine u on peut associer un ordre $\mathbf{0}$ tel que $M(u, \mathbf{0})$ soit antirobinson.

Proposition 7

Si u est une ultramine on peut ordonner les lignes et colonnes selon un ordre $\mathbf{0}$ tel que la matrice $M(u, \mathbf{0})$ soit antirobinson.

Démonstration

Soit $\mathbf{0}$ un ordre sur Ω obtenu en construisant une suite w_1, \dots, w_n de la façon suivante : w_1 est choisi au hasard, w_2 est l'élément le plus éloigné de w_1 (au sens de u), w_{i+1} est l'élément le plus éloigné de w_i parmi les éléments restants ; autrement dit on a : $u(w_\ell, w_{\ell+1}) = \text{Max} \{u(w_\ell, w_j) / j > \ell\}$.

Il en résulte que si $w_i \leq w_j$ selon l'ordre $\mathbf{0}$ ainsi construit on a : $u(w_i, w_j) \leq u(w_i, w_{i+1})$ or $u(w_{i+1}, w_j) \geq \text{Min} \{u(w_i, w_j), u(w_i, w_{i+1})\}$ puisque u est une ultramine d'où $u(w_{i+1}, w_j) \geq u(w_i, w_j)$; si i est l'indice des lignes et k celui des colonnes de $M(u, \mathbf{0})$, on voit que cette matrice est croissante en colonnes.

On a donc $u(w_i, w_{j+1}) \leq u(w_j, w_{j+1})$ (puisque $w_i \leq w_j$ selon $\mathbf{0}$ et donc dans $M(u, \mathbf{0})$ la ligne i précède la ligne j) ; il en résulte que $u(w_i, w_j) \geq u(w_i, w_{j+1})$ puisque $u(w_i, w_j) \geq u(w_i, w_{j+1}) \geq \text{Min} \{u(w_i, w_{j+1}), u(w_j, w_{j+1})\}$ et donc que $M(u, \mathbf{0})$ est également décroissante en ligne. C'est donc une matrice antirobinson. \square

Des propositions 6 et 7 on déduit immédiatement les résultats suivants.

Proposition 8

1) A toute ultramétrie u on peut associer un ordre $\mathbf{0}$ tel que u et $\mathbf{0}$ soient anticompatibles.

2) Si la suite w_1, \dots, w_n respecte l'ordre $\mathbf{0}$ la chaîne $w_1 w_2, w_2 w_3, \dots, w_{n-1} w_n$ est celle dont la somme des longueurs des arêtes $u(w_i, w_{i+1})$ est la plus grande.

Exemple 5

Considérons l'ultramétrie définie sur w_i, w_j, w_k, w_l dans le figure 5

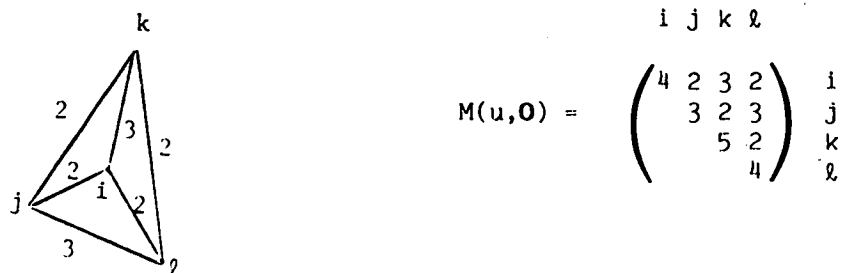


Figure 5

En construisant l'ordre $\mathbf{0}'$ défini dans la démonstration de la proposition 8 on obtient

$$M(u, \mathbf{0}') = \begin{matrix} & \begin{matrix} l & j & i & k \end{matrix} \\ \begin{matrix} l \\ j \\ i \\ k \end{matrix} & \begin{pmatrix} 4 & 3 & 2 & 2 \\ & 3 & 2 & 2 \\ & & 4 & 3 \\ & & & 5 \end{pmatrix} \end{matrix}$$

On voit que $M(u, O')$ est une matrice antirobinson que l'ordre $O' = \{w_l w_j w_i w_k\}$ est anticompatible avec u et que la chaîne $w_l w_j, w_j w_i, w_i w_k$ de longueur $3+2+3 = 8$ est la plus grande. Remarquons enfin qu'aucun ordre ne peut rendre la matrice $M(u, O)$ robinson.

5. EXTENSION DES ULTRAMINES AUX INDICES ANTIPYRAMIDAUX

5.1. Les indices antipyraxidaux

On sait que les ultramétriques peuvent être plongées dans l'ensemble des indices pyramidaux (voir Diday 1984), on a un résultat analogue pour les ultramines ; en conservant la propriété des ultramines concernant l'existence d'un ordre anticompatible et en supprimant la contrainte due à l'inégalité ultramine, on plonge les ultramines dans un ensemble plus vaste : celui des indices antipyraxidaux. Plus précisément ces indices sont définis de la façon suivante :

Définition

Un indice antipyraxidal est une mesure de ressemblance pour laquelle il existe un ordre anticompatible.

Comme on sait d'après la proposition 8 qu'à toute ultramine on peut associer un ordre anticompatible, on a le résultat suivant :

Proposition 9

L'ensemble des ultramines est inclus dans l'ensemble des indices antipyraxidaux.

5.2. Tableau résumé des principales propriétés

On sait d'après la proposition 1 que l'ensemble des ultramines de $M(X)$ et des ultramétriques $\delta \in U$ sont en bijection. L'ensemble des matrices ultramines et des matrices ultramétriques sont en bijection puisque ces ensembles sont respectivement en bijection avec $M(X)$ et U . D'après la proposition 7 nous savons que si $M(u, O)$ est une matrice ultramine (donc u une ultramine), il existe un ordre O tel que U et O soient anticompatibles. Cette dernière propriété est d'une part équivalente au fait que u soit un indice antipyraxidale par définition et d'autre part équivalente au fait que $M(u, O)$ soit antirobinson d'après la proposition 6. Toutes les conditions

portant sur les ultramétriques ont été démontrées par exemple dans Diday (1982) ; enfin en définissant l'ensemble $P(X)$ des indices antipyramidaux (où X est le vecteur exprimant la valeur de l'indice pour les couples d'éléments identiques) analogue à celui qui a été défini pour les ultramines et en remarquant que si l'on inverse les valeurs prises par un indice pyramidale δ on obtient un indice antipyramidale U de $P(X)$ et réciproquement en associant au vecteur nul des $\delta(w,w)$ le vecteur X des $u(w_1,w)$ on définit une bijection de $P(X)$ dans l'ensemble des indices pyramidaux. Comme il existe une bijection entre l'ensemble des indices antipyramidaux et les matrices antirobinson d'après la proposition 6 entre les indices pyramidaux et les matrices robinson d'après Diday (1982), il en résulte l'existence d'une bijection entre les matrices robinson et les matrices antirobinson (que l'on aurait pu d'ailleurs démontrer directement en remarquant qu'en inversant les termes d'une matrice antirobinson en dehors de ceux de la diagonale principale, on obtient une matrice robinson). Tous ces résultats sont résumés dans le tableau 1.

6. REPRESENTATIONS GRAPHIQUES DES ULTRAMINES ET DES INDICES ANTIPYRAMIDAUX

6.1. Représentation par hiérarchies et pyramides anti-indicées

La représentation visuelle de toute ultramine de $M(X)$ peut se faire par une hiérarchie : en effet, on sait qu'il existe une bijection entre les hiérarchies et les ultramétriques voir Diday et al (1982) et ces dernières et $M(X)$; pratiquement, ayant une ultramine, on calcule son "inverse" pour avoir l'ultramétrie correspondante que l'on représente par une hiérarchie ; pour que cette hiérarchie soit plus fidèle à l'ultramine on exprime les hauteurs des paliers par les valeurs de l'ultramine ; de cette façon les hauteurs des paliers n'auront pas l'interprétation habituelle : plus ils sont petits, plus les individus qui leur sont associés devront être considérés comme éloignés ; cela revient à anti-indicer la hiérarchie et donc à définir un indice f de $P(\Omega)$ (l'ensemble des parties de Ω) dans \mathbb{R}^+ satisfaisant à la propriété suivante : $h, h' \in P(\Omega) : h \supset h' \Rightarrow f(h') < f(h)$; f est appelé anti-indice. Avec cet indiçage les longueurs des branches associées aux singletons peuvent être ajustées pour tenir compte des valeurs des termes $X_i = u(w_1, w_i)$. On peut représenter un indice antipyrimal par une pyramide puisqu'il existe une bijection entre les pyramides et les indices pyramidaux (voir Diday 1984) et ces derniers et les indices antipyrimal d'un ensemble $P(X)$. Cependant, pour que les valeurs de l'indice antipyrimal correspondent aux hauteurs des paliers il faut anti-indicer la pyramide comme cela a été fait pour représenter une ultramine par une hiérarchie anti-indicée ; l'application f d'anti-indiçage se définit dans les deux cas de façon identique.

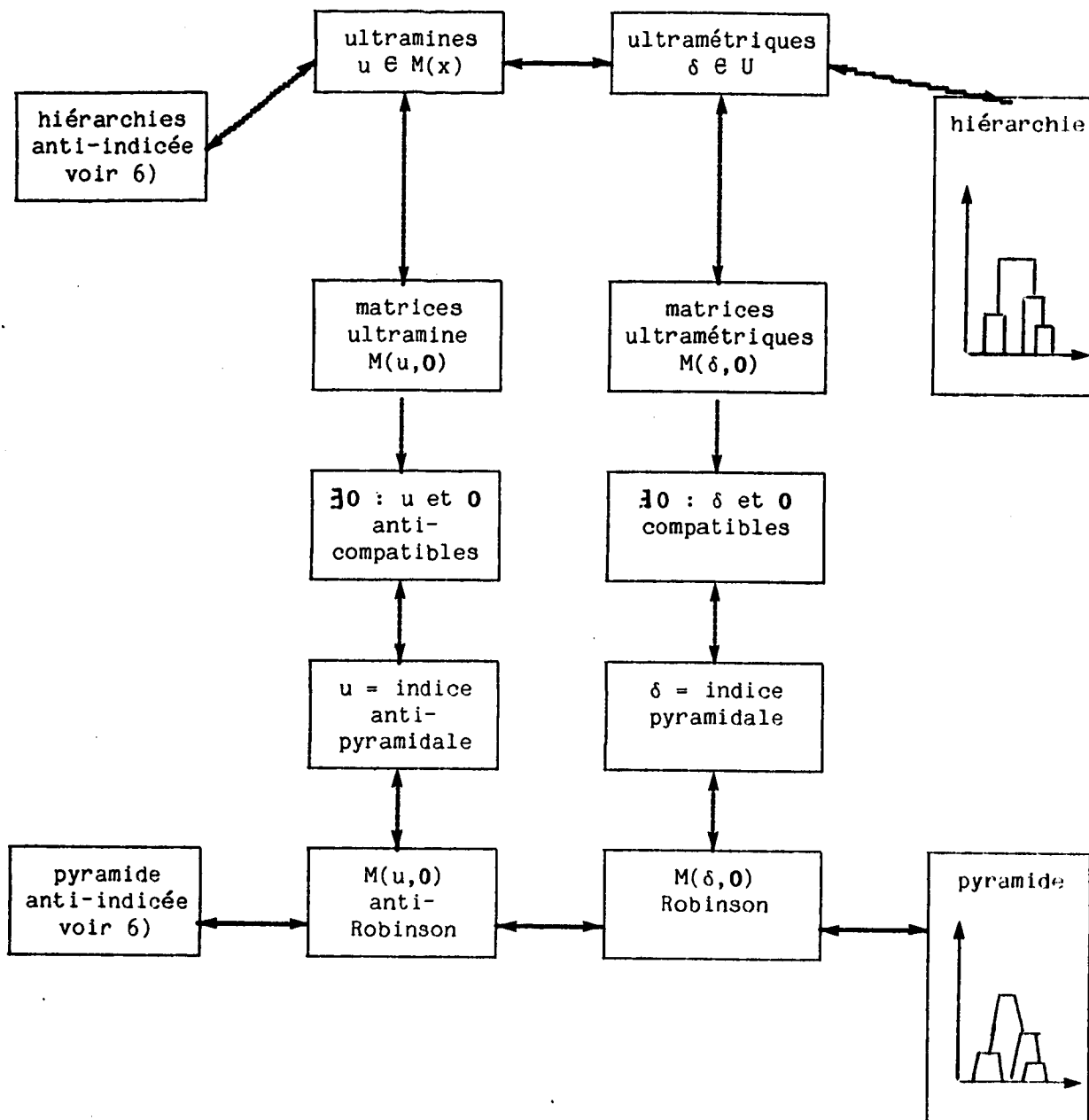


Tableau 1

Résumé des différents liens

le signe \longleftrightarrow (resp. \rightarrow) indique l'existence d'une bijection (resp. implication)

Exemple 6

On considère un ensemble $\Omega = \{a,b,c,d\}$ une ultramétrie u et un indice antipyrâmidal dont les matrices $M(u,0)$, $M(s,0)$ et leur représentation antihierarchique et antipyrâmidal sont données figure 6.

$$M(u,0) = \begin{matrix} & a & b & c & d \\ \begin{matrix} a \\ b \\ c \\ d \end{matrix} & \begin{pmatrix} 5 & 3 & 2 & 1 \\ & 4 & 2 & 1 \\ & & 5 & 1 \\ & & & 2 \end{pmatrix} & \end{matrix}$$

$$M(s,0) = \begin{matrix} & a & b & c & d \\ \begin{matrix} a \\ b \\ c \\ d \end{matrix} & \begin{pmatrix} 5 & 3 & 2 & 1 \\ & 4 & 3 & 1 \\ & & 3 & 2 \\ & & & 2 \end{pmatrix} & \end{matrix}$$

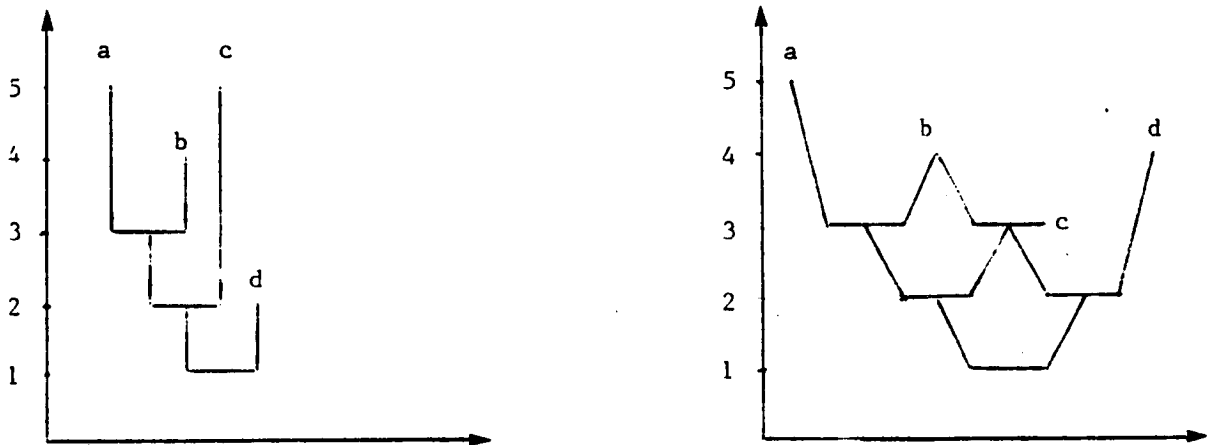


Figure 6

6.2. Représentation curviligne6.2.1. Le problème général

On sait qu'une ultramétrie ou un indice antipyrâmidal ayant été calculé on peut leur associer un ordre tel que la matrice des distances associée à cet ordre soit antirobinson. Le problème que l'on se pose ici est celui de la recherche d'une courbe sur laquelle on puisse disposer les individus selon cet ordre de façon à respecter au mieux les distances définies par l'ultramétrie ou l'ultramétrique.

6.2.2. Expression analytique du problème

Le problème de la représentation curviligne des ultramétriques et des indices pyramidaux se pose de façon tout à fait analogue et nous en avons donné l'expression analytique dans E. DIDAY (1984) (p. 65). Dans le cas des ultramines et des indices antipyramidaux il s'énonce sous une forme analogue :

.trouver une courbe C du plan représentée par une application $f \Omega \rightarrow C$: $f(w_i) = z_i = (x_i, y_i) \in C$ telle que si l'on représente les w_i sur cette courbe en respectant un ordre O on minimise par exemple, le critère suivant :

$$V(s,d) = \sum_{i,j} (s(w_i, w_j) - D(w_i, w_j))^2$$

où D est la distance euclidienne du plan et où s est l'indice associé à l'ultramine ou à l'indice antipyramidale considéré (on peut normaliser s et D en les divisant par leur plus grande valeur).

Si f est une fonction à m paramètres on doit résoudre un problème d'optimisation à $\frac{n(n-1)}{2}$ équations à n-1 contraintes (pour respecter O) et à 2n+m inconnus (car il faut trouver les 2n points (x_i, y_i) et les m coefficients qui définissent f).

L'exemple suivant montre que ce problème n'a pas toujours de solution

Exemple 7

Considérons l'indice antipyramidale s : $M(s,0) =$

| | a | b | c | d |
|---|---|---|-----|-----|
| a | 0 | 6 | 5,5 | 3 |
| b | | 0 | 7 | 4,5 |
| c | | | 0 | 7,2 |
| d | | | | 0 |

On peut la représenter sous forme curviligne comme indiqué dans la figure 7 ; à l'aide des médiatrices des segments ab, ac, bc, et dc et en utilisant le fait qu'un cinquième point e doit être à une distance supérieure à la longueur dc, on voit qu'un tel pont ne peut être positionné dans le plan. (e doit être par exemple au-dessus de la médiatrice de bc).

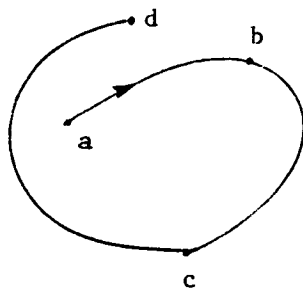


Figure 7

6.2.3. Représentation à l'aide de courbes compatibles

Définition

Une courbe est dite "compatible" avec un indice de dissimilarité d si et seulement si tout triplet x, y, z , ordonné sur cette courbe est tel que $d(x, z) \geq \text{Max} \{d(x, y), d(y, z)\}$. S'il s'agit d'une courbe du plan et si d est la distance euclidienne on dira simplement que "la courbe est compatible".

Il résulte de cette définition que tout ensemble fini de points ordonné selon un ordre O sur une courbe compatible avec d est tel que d et O soient compatibles. Ces courbes existent puisque les droites constituent un exemple simple de courbes compatibles.

On donne figure 8 des exemples de courbes compatibles et non compatibles.

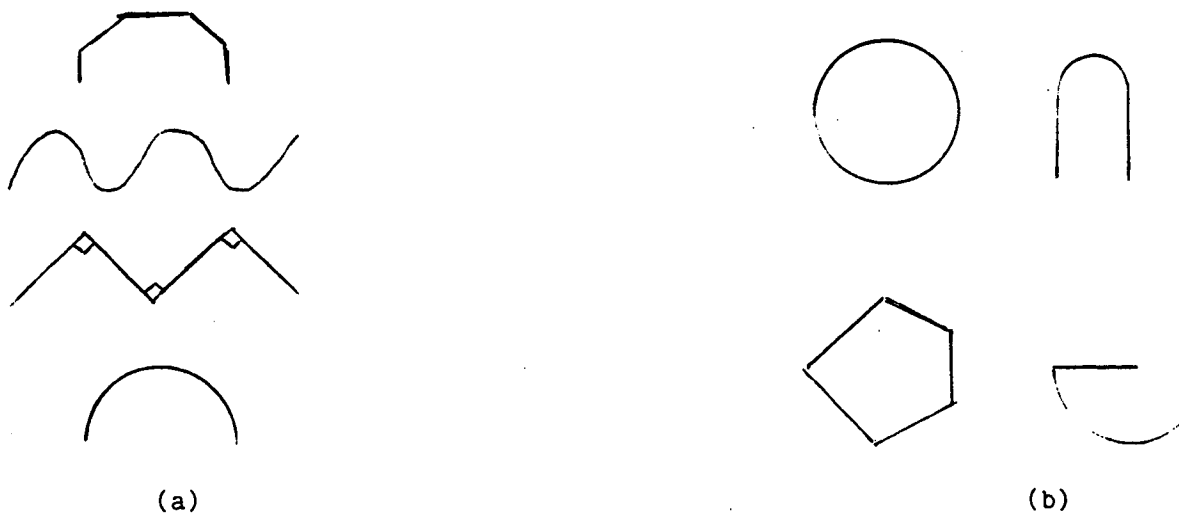


Figure 8

(a) courbes compatibles, (b) courbes non compatibles avec la distance euclidienne

On peut se poser la question de savoir s'il existe des "courbes anticompatibles" ; autrement dit, des courbes sur lesquelles trois points ordonnés w_1, w_2, w_3 , satisfont à l'inégalité ultramine : $d(w_1, w_3) \geq \text{Min}(d(w_1, w_2), d(w_2, w_3))$. Une telle courbe ne peut être visualisée car si elle contient un segment de droite ou si on peut l'approximer par un tel segment aussi petit soit-il elle ne satisfait plus à l'inégalité ultramine sur cette portion et elle n'est donc pas anticompatible.

Pour représenter à l'aide d'une courbe une matrice anticompatible provenant d'une ultramine ou d'un indice antipyramide, on peut procéder à l'aide d'un anti-indiçage comme on l'a fait pour les hiérarchies et les pyramides.

On transforme d'abord la matrice antirobinson en inversant chacun des ses termes et en remplaçant les termes diagonaux par 0, on obtient ainsi une matrice robinson qu'il s'agit de représenter sur une courbe compatible.

Exemple 8

Etant donnée la matrice antirobinson suivante :

$$M(s, 0) = \begin{array}{cccc} & a & b & c & d \\ \begin{pmatrix} 30 & 25 & 16 & 14 \\ & 50 & 40 & 25 \\ & & 60 & 58 \\ & & & 70 \end{pmatrix} & a \\ & & & & b \\ & & & & c \\ & & & & d \end{array}$$

il s'agit d'en donner une représentation curviligne sur un demi-cercle. Pour cela, on inverse d'abord chaque valeur de $M(s, 0)$ (sauf les termes de la diagonale principale, qu'on annule) on obtient ainsi la matrice robinson suivante :

$$M(s^{-1}, 0) = \begin{array}{cccc} & a & b & c & d \\ \begin{pmatrix} 0 & 4 & 6 & 7 \\ & 0 & 2,5 & 4 \\ & & 0 & 1,7 \\ & & & 0 \end{pmatrix} & a \\ & & & & b \\ & & & & c \\ & & & & d \end{array}$$

Cette matrice est représentée sur le demi-cercle de la figure 9 en positionnant "bien" les points a, b, c, d de façon à respecter au mieux les distances s' . Finalement on anti-indiçage les segments ab, bc... et on représente les valeurs $s(a, a), s(b, b)$...

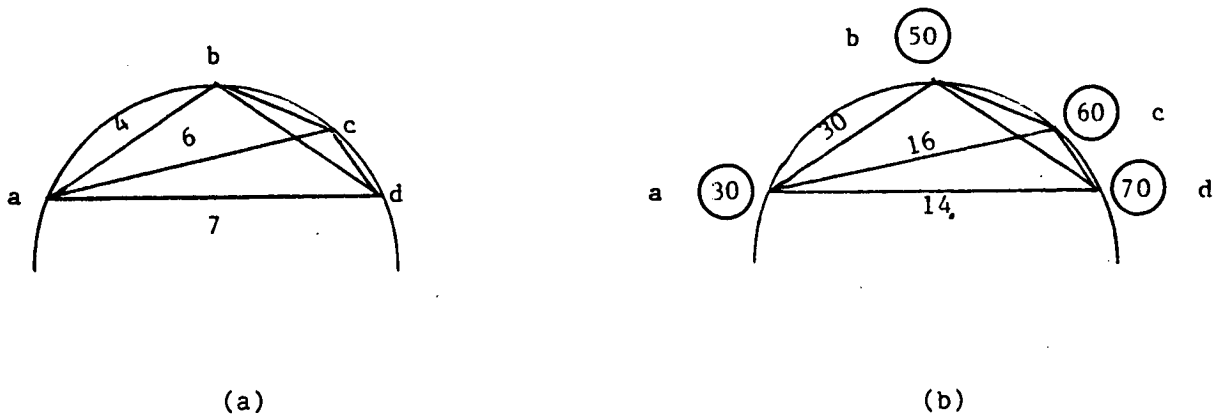


Figure 9

(a) représentation curviligne de $M(s', 0)$, (b) représentation curviligne de $M(s, 0)$

6.2.4. Un algorithme général pour la représentation à l'aide de courbes compatibles

Il s'agit de proposer une euristique permettant de trouver simultanément une courbe et des points sur cette courbe de façon que leur distance euclidienne d respecte au mieux celle d'une ultramétrie ou d'un indice antipyramide (obtenus par "inversion" d'une ultramétrie ou d'un indice antipyramide ou plus généralement d'un indice de dissimilarité quelconque). Cet euristique peut s'énoncer sous la forme suivante :

1. On dispose les points w_1, \dots, w_n dans l'ordre et à égale distance sur une droite.
2. Chaque segment $w_i w_{i+1}$ est découpé en intervalles égaux dont les sommets forment un réseau.
3. On décale chaque point pris successivement vers la gauche ou vers la droite sur le réseau tant que l'ordre initial est respecté et que le critère
$$W = \sum_{i,j} |d(w_i, w_j) - s(w_i, w_j)|$$
 (où d est la distance euclidienne du plan et s l'indice que l'on veut représenter) s'améliore.
4. On revient en 2 tant que le critère continue à s'améliorer (en recommençant 3 s'il le faut plusieurs fois sur l'ensemble des points).
5. On "incurve" la droite aux endroits à forte densité et on revient en 1 tant que le critère s'améliore.

Remarque : au lieu de partir d'une droite, on peut aussi partir d'une courbe présentant des sommets associées à chaque noyau d'une classification faite au préalable sur le tableau de dissimilarité défini par s . Ces noyaux doivent respecter l'ordre O sur la courbe. On représente ensuite tous les points à intervalles égaux en respectant O . A, partir de cette position on peut appliquer l'algorithme énoncé précédemment (voir la figure 10).

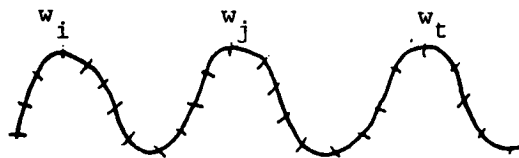


Figure 10

Représentation curviligne d'un indice de dissimilarité s ;

w_i, w_j, w_k , sont les noyaux d'une partition obtenue sur tableau de distance.

6.2.5. Hierarchies curvilignes

La représentation hiérarchique habituelle est une représentation simultanément hiérarchique (le dendrogramme) et rectiligne (la droite sur laquelle s'appuie le dendrogramme) ; cette représentation est de lecture simple car elle exprime deux formes de compatibilité.

a) la compatibilité entre l'ordre O des individus (défini par les racines du dendrogramme) et l'ultramétrie induite par le hiérarchie puisqu'il n'y a pas de croisements (voir Diday (1982)).

b) la compatibilité entre O et la distance euclidienne puisque la droite est une courbe compatible.

Comme la droite ne constitue qu'un cas particulier de courbe compatible, la représentation visuelle classique des hiérarchies à l'aide d'un dendrogramme appuyé sur une droite peut être étendue en remplaçant la droite par toute courbe du plan compatible avec la distance euclidienne.

L'intérêt d'une telle représentation est de permettre de visualiser la hiérarchie sous une forme "curviligne" à l'aide d'un dendrogramme et d'une courbe compatible exprimant une contrainte externe (un autre indice de dissimilarité, une ultramétrique etc...) Par exemple, si les individus sont des communes la hiérarchie initiale peut exprimer des classes définies à partir de paramètres économiques et l'indice de dissimilarité de la contrainte : la position géographique.

Pratiquement, il s'agit de trouver une courbe compatible et la position sur cette courbe des individus de façon à ce qu'ils respectent un ordre compatible avec la hiérarchie et que leur distance euclidienne satisfasse au mieux à la contrainte.

Dans la figure 11b on représente la hiérarchie 11a sous contrainte de la courbe compatible formée d'un demi-cercle sur lequel les individus sont positionnés de façon à satisfaire à l'indice anti-pyramidal défini dans l'exemple du § 6.2.3.

Dans la figure 11d on visualise la hiérarchie 11c sous une autre contrainte à l'aide d'une autre courbe compatible.

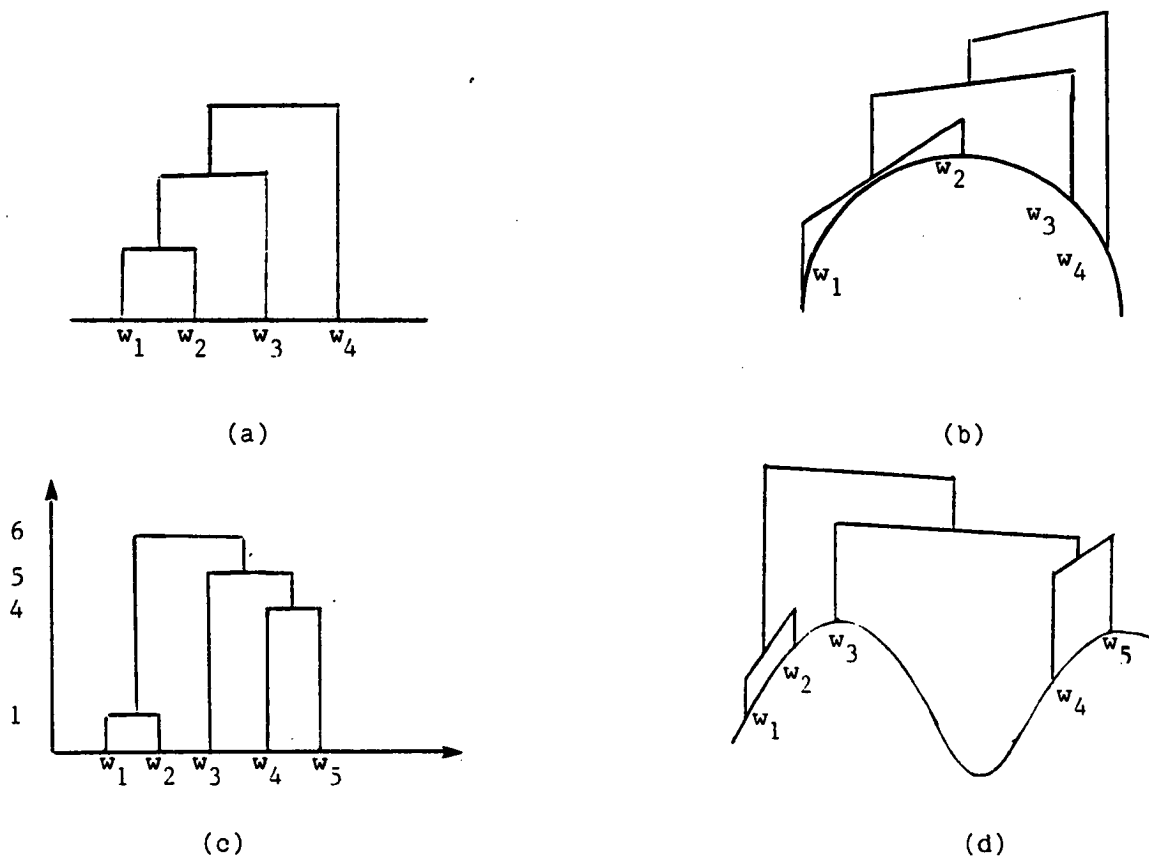


Figure 11

La hiérarchie de la figure (a) (resp. (c)) est représentée sous contrainte en (b) (resp. (d))

Il y a quatre façons de représenter une hiérarchie sur une courbe compatible suivant que la hiérarchie représente une ultramétrie ou une ultramine et que l'ordre des individus sur la courbe est compatible ou anti-compatible avec la distance euclidienne.

7. DONNEES CHAOTIQUES ET DONNEES COHERENTES

7.1. Définition et propriétés

Etant donné un indice de dissimilarité s , il se prête bien à une représentation par une ultramine ou par un indice antipyramidale quand il existe un ordre O sur Ω tel que s et O soient anticompatibles. Un tel ordre est "chaotique" puisqu'il exprime le fait que deux éléments de Ω sont d'autant plus "proches" selon cet ordre qu'ils sont "éloignés" selon s ; nous avons vu aussi qu'il n'existe pas de courbe permettant de représenter un tel ordre. Nous dirons que des données sont "chaotiques" pour un indice de dissimilarité s s'il existe un ordre O sur Ω tel que s et O soient anticompatibles et s'il n'existe pas d'ordre O' tel que s et O' soient compatibles ; on dira qu'elles sont "cohérentes" dans le cas contraire, c'est à dire s'il existe un ordre compatible avec s et pas d'ordre anticompatible. Le tableau 2 permet de caractériser ces deux types de données.

| | Indicage | Représentation graphique | Classes caractéristiques |
|--------------------|------------------------------------|---|--------------------------|
| Données chaotiques | Ultramines Indice antipyramidal | Hiérarchie } Pyramide } anti-indicée Fidélité aux grandes distances | Classes d'oppositions |
| Données cohérentes | Ultramétriques Indice pyramidal | Hiérarchies } Pyramides } indicées Fidélité aux petites distances | Chaînes Agrégats |

Tableau 2

Des données peuvent être "amorphes" c'est-à-dire en même temps "chaotiques" et "cohérentes" c'est par exemple le cas quand tous les couples ont la même dissimilarité. La proposition suivante montre qu'il existe une vaste classe de données qui ne peuvent être à la fois chaotiques et cohérentes.

Proposition 10

Si pour quatre éléments de Ω toutes les valeurs prises par l'indice de dissimilarité sont distinctes alors les données ne peuvent être amorphes.

Démonstration

Il suffit de démontrer que si toutes les valeurs prises par un indice de dissimilarité s sont distinctes pour 4 éléments ordonnés selon O alors si $M(s, O)$ est robinson (resp. antirobinson) alors il n'existe par O' tel que $M(s, O')$ soit antirobinson (resp. robinson).

Supposons que $\Omega = \{a, b, c, d\}$, que O respecte l'ordre $a b c d$ et que $M(s, O)$ soit robinson. Si toutes les valeurs prises par s sont distinctes, il existe un couple unique par exemple (a, b) de plus petite dissimilarité ; s'il existait un ordre antirobinson O' ses extrêmes seraient forcément a et b . Supposons que O' commence par a et finisse par b (la démonstration est bien sûr identique dans le cas inverse) ; il reste deux ordres possibles pour O' soit $acdb$ soit $adcb$; dans les deux cas $M(s, O')$ n'est pas antirobinson car dans le premier cas on a : $s(a, c) < s(a, d)$ et dans le second $s(d, c) < s(d, b)$ puisque O est compatible avec s . Donc O' n'existe pas. Si Ω contient plus de quatre éléments, a fortiori il n'existe pas d'ordre antirobinson puisque l'on pourrait alors extraire quatre éléments de Ω ordonnés selon O qui formeraient un ordre antirobinson si on les ordonnait selon O' ce qui est contradictoire avec ce qui vient d'être démontré.

Si $M(s, O)$ est antirobinson on démontrerait de façon tout à fait analogue qu'il n'existe pas d'ordre O' qui soit robinson avec s . \square

Remarques : dans le cas où Ω est formé de 3 éléments $a b c$ ordonnés selon un ordre robinson, on voit qu'il peut exister, même si les valeurs de l'indice sont distinctes, un ordre antirobinson sur l'exemple de la figure 12. Sur la figure 13 on montre qu'il peut exister un ordre robinson et un ordre antirobinson sur les mêmes données avec $\text{card } \Omega > 3$, si les valeurs prises par s ne sont pas toutes différentes.

$$\begin{array}{ccc}
 a & b & c \\
 \left(\begin{array}{ccc} 0 & 1 & 3 \\ & 0 & 2 \\ & & 0 \end{array} \right) & \begin{array}{l} a \\ b \\ c \end{array} & \\
 b & c & a \\
 \left(\begin{array}{ccc} 0 & 2 & 1 \\ & 0 & 3 \\ & & 0 \end{array} \right) & \begin{array}{l} b \\ c \\ a \end{array} &
 \end{array}$$

Figure 12

abc est un ordre Robinson
bca est antirobinson

$$\begin{array}{cccc}
 a & b & c & d \\
 \left(\begin{array}{cccc} 0 & 3 & 3 & 1 \\ & 0 & 3 & 3 \\ & & 0 & 5 \\ & & & 0 \end{array} \right) & \begin{array}{l} a \\ b \\ c \\ d \end{array} & \\
 c & b & a & d \\
 \left(\begin{array}{cccc} 0 & 3 & 3 & 5 \\ & 0 & 3 & 3 \\ & & 0 & 1 \\ & & & 0 \end{array} \right) & \begin{array}{l} c \\ b \\ a \\ d \end{array} &
 \end{array}$$

Figure 13

abcd est un ordre Robinson
cbad est antirobinson

7.2. Indicateur de chaos et de cohérence des données pour un indice de dissimilarité s

7.2.1. A l'aide des ultramines et de ultramétriques

Les données* munies d'un indice de dissimilarité s sont dites ultramines (resp. ultramétriques) si tous les triangles sont isocèles avec la base plus grande (resp. petite) que les côtés. Dans ce cas on sait que s est une ultramine (resp. ultramétrique) et qu'il existe un ordre O tel que M(s,0) soit antirobinson (resp. robinson). Les données sont donc plus ou moins chaotiques ou cohérentes suivant qu'elles sont plus ou moins ultramines ou ultramétriques. On peut imaginer de nombreux indicateurs utilisant la notion de triangles isocèles ; nous proposons le suivant : soit T l'ensemble des triangles de Ω muni d'un indice de dissimilarité s ; soient a_t, b_t, c_t les longueurs au sens de s par ordre de grandeur décroissante du triangle $t \in T$; l'écart entre les deux plus grands côtés est $a_t - b_t$, celui entre les deux plus petits est $b_t - c_t$ et donc l'écart entre ces deux écarts est $a_t - 2b_t + c_t = E(t)$; si $E(t) > 0$ le triangle est plutôt isocèle avec la base plus petite que les côtés ; si $E(t) < 0$ c'est le cas contraire qui se produit ; si $E(t) = 0$ le triangle est équilatéral. Si réciproquement les données sont ultramines, (resp. ultramétriques) on a $E(t) \leq 0$ (resp. $E(t) \geq 0$) pour tout $t \in T$.

D'où l'indicateur suivant :

$$W_1(\Omega, s) = \frac{6}{n(n-1)(n-2)-6N} \sum_{t \in T} \text{sign } E(t)$$

où N est le nombre de triangles équilatéraux.

* En fait il vaudrait mieux parler du couple (données, indice).

On voit que l'on a $W_1(\Omega, s) = 1$ si les données sont ultramines ; $W_1(\Omega, s) = 1$ si elle sont ultramétriques enfin $W_1(\Omega, s) = 0$ si elle sont "amorphes", c'est à dire si toutes les distances au sens de s sont égales (autrement dit tous les triangles sont équilatéraux).

7.2.2. Indicateur de données chaotiques ou cohérentes

On peut imaginer beaucoup d'indicateurs permettant de savoir si les données sont plus ou moins chaotiques ou cohérentes sans passer par les notions d'ultramines ou d'ultramétriques. Nous proposons un indicateur basé sur la notion d'ordre partiel compatible ou anticompatible ; pour les ordres partiels de plus grande taille (i.e. contenant le plus d'éléments possible) on peut utiliser un algorithme du type indiqué dans la démonstration de la proposition. On part d'un élément quelconque de Ω et l'on cherche l'élément le plus proche au sens de s en le plaçant à gauche ou à droite et en recommençant ainsi de suite avec les extrémités de la suite obtenues tant que l'ordre partiel ainsi construit reste compatible (ou anticompatible) avec s .

De cette façon on calcule tous les ordres partiels de plus grande taille en recommençant avec le même procédé tant qu'il reste des éléments n'appartenant à aucun des ordres déjà obtenus.

Notons O_1, O_2, \dots, O_k (resp O'_1, \dots, O'_k) les plus grands ordres partiels compatibles (resp. anticompatibles) avec s , rangés par ordre de taille décroissante. On peut alors proposer l'indicateur suivant :

$$W_2(\Omega, s) = \frac{1}{n-3} \sum_{i=1}^{\text{Min}(k, k')} (\text{card } \Omega_i - \text{card } \Omega'_i)$$

où Ω_i est l'ensemble des éléments de Ω concernés par l'ordre O_i .

D'après la proposition 10 si $\text{card } \Omega > 3$ et s'il existe un ordre robinson (resp antirobinson) alors $0 \leq W_2(\Omega, s) \leq 1$ (resp. $-1 \leq W_2(\Omega, s) < 0$). Autrement dit plus W_2 est proche de -1 plus les données sont chaotiques plus W_2 est proche de $+1$ plus elles sont cohérentes.

Le cas extrême de données complètement chaotiques et pas du tout cohérentes est obtenu quand $k'=1$ et quand le plus grand ordre cohérent contient trois éléments ; on a donc

$$W_2(\Omega, s) = \frac{1}{n-3} (\text{card } \Omega'_1 - \text{card } \Omega_1) = \frac{1}{n-3} (3-n) = -1.$$

CONCLUSION

Partant d'un indice de dissimilarité quelconque, nous avons étudié une forme de représentation complémentaire aux hiérarchies qui peut produire des informations intéressantes : être parfois plus fidèle aux données initiales, faire ressortir les contrastes beaucoup mieux que par les ultramétriques ce qui peut permettre par exemple de mieux interpréter une analyse factorielle, ou en traitement d'image parallèlement aux zones homogènes détectées par un partitionnement ou une ultramétrique faire apparaître les zones de contraste. Les ultramines et les indices anti-pyramidaux peuvent aussi être utilisés pour faire apparaître l'écart entre une hiérarchie, une pyramide, une analyse factorielle ou une représentation quelconque et les données initiales : la représentation ultramine par exemple, calculée sur l'écart, permet de détecter les plus grandes erreurs, là où l'ultramétrique fait ressortir surtout les plus petites. L'intérêt d'une représentation par ultramine ou par indice anti-pyramidal peut se mesurer a priori par les critères que nous avons donnés ; pratiquement, partant d'un indice de dissimilarité, on calcule son indicateur de chaotisation W_1 ou W_2 , s'il est grand on remplace chaque élément diagonal de la matrice de dissimilarité par la plus grande valeur de la ligne et de la colonne correspondante puis on utilise les représentations ultramines, antipyramidales ou curvilignes. Réciproquement, si l'indice de départ est un indice de similarité qui satisfait à un ordre robinson on remplace les termes de la diagonale principale de l'indice de similarité par 0 puis on utilise les techniques de représentation hiérarchiques ou pyramidales usuelles.

De nombreuses directions de recherche restent ouvertes, d'abord étendre l'approche à d'autres formes de distances que les ultramétriques, par exemple, les quadrangulaires ou les distances à centre et chercher les "anti-représentations" correspondantes. En ce qui concerne la représentation graphique des ultramines et des indices anti-pyramidaux, trouver la solution du problème dont la forme analytique a été énoncée en 6.2.2. ; étudier et préciser l'algorithme général de représentation des courbes compatibles tel qu'il est énoncé en 6.2.4. ; étudier de façon précise les indicateurs de chaotisation qui ont été donnés pour en déduire des tests non paramétrique, en proposer d'autres etc...

BIBLIOGRAPHIE

- [1] BATBAT A. (1984) - "La surdominante et la chaîne dominante ramifiée". Cahier de DEA, Université des Sciences et Techniques du Languedoc, UER de Mathématiques, Place Eugène Bataillon, Montpellier, France.
- [2] DIDAY E. (1982) - "Croisements ordres et ultramétriques : application à la recherche de consensus en classification automatique", Rapport de recherche n° 144 - INRIA, Rocquencourt 78150.
- [3] DIDAY E., LEMAIRE J., POUGET J., TESTU F. (1982) - "Elements d'analyse des données" DUNOD.
- [4] DIDAY E., (1982) - "Croisements ordres et ultramétriques", Math. Sciences Humaines, 21ème année, n°83, 31-54.
- [5] DIDAY E., (1984) - "Une représentation des classes empiétantes : les pyramides" - Rapport INRIA n° 291, Rocquencourt 78150 FRANCE.
- [6] HUBERT (1974) - "Some applications of graph theory to clustering". Psychometrika 39, 283-309.
- [7] JAMBU M. (1978) - "Classification automatique pour l'analyse des données" DUNOD.
- [8] LERMAN I.C. (1981) - "Classification et analyse ordinaire des données" DUNOD.
- [9] SNEATH PHA, SOKAL RR (1973) - "Numerical Taxonomy" Freeman.

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100