MATHÉMATIQUES ET SCIENCES HUMAINES

TAOUFIK BENKARAACHE

L'ultramétrique inférieure maximum d'une dissimilarité à valeurs dans un inf-demi-treillis

Mathématiques et sciences humaines, tome 143 (1998), p. 27-40

<http://www.numdam.org/item?id=MSH_1998__143__27_0>

Math. Inf. Sci. hum., (36e année, n°143, *1998*, pp.27-40)

L'ULTRAMÉTRIQUE INFÉRIEURE MAXIMUM D'UNE DISSIMILARITÉ À VALEURS DANS UN INF-DEMI-TREILLIS

Taoufik BENKARAACHE[1]

RÉSUMÉ — *Les dissimilarités sont habituellement à valeurs dans l'ensemble des réels positifs* $\mathbb{R}^+$. *Cet ensemble est riche en propriétés dont, par exemple, celles liées à l'ordre total de* $\mathbb{R}$. *Plusieurs auteurs ont montré que certains résultats fondamentaux relatifs aux dissimilarités restent valables quand on remplace* $\mathbb{R}^+$ *par un ensemble ordonné* L *plus général.*

Dans ce papier nous proposons deux méthodes d'approximation d'une dissimilarité d *par l'ultramétrique inférieure maximum (la sous-dominante ultramétrique) quand* d *est à valeurs dans un inf-demi-treillis fini.*

ABSTRACT — The subdominant ultrametric of a dissimilarity taking its values in a meet semilattice. *Dissimilarity functions are usually defined to take values in the set* $\mathbb{R}^+$ *of nonnegative reals. This set has various properties as, for instance, those related to the intrinsic linear order of* $\mathbb{R}$. *But many authors have shown that a number of basic results relative to dissimilarity functions are preserved when* $\mathbb{R}^+$ *is replaced by any partially ordered set. In this paper we propose two methods for the fitting of a given dissimilarity approximation by the subdominant ultrametric when this dissimilarity takes values in a meet semi-lattice.*

1. INTRODUCTION

Dans le cadre de la généralisation des structures utilisées en classification, et connaissant l'importance des considérations ordinales dans cette discipline, plusieurs auteurs se sont consacrés à l'étude des relations mutuelles qui lient les ensembles ordonnés et la classification mathématique (Janowitz 1978, Barthélemy et al. 1984, Leclerc 1994, ...), ainsi qu'au développement de la méthodologie qui vise à obtenir des modèles de classification à partir de dissimilarités en mettant l'accent sur l'aspect ordinal (Johnson 1967, Jardine et Sibson 1971, Benzécri 1973, Critchley et Van Cutsem 1994, etc.). L'une de ces généralisations réside dans le fait que les dissimilarités considérées sont à valeurs dans un ensemble L non nécessairement totalement ordonné : en effet, Janowitz (1978), puis Barthélemy et al. (1984) avaient déjà posé le problème de la recherche et de l'étude des méthodes de classification en des termes tout-à-fait ordinaux, dans le cadre de théorie de la résiduation. Les L-dissimilarités et les L-ultramétriques peuvent se ramener aux fonctions résiduelles et résiduées. Le problème de l'ajustement d'une L-dissimilarité par une L-ultramétrique peut être aussi formulé en termes d'applications résiduelles. Notre contexte étant un peu limité, on renvoie pour la théorie de la résiduation sur l'ouvrage de Blyth et Janowitz (1972) ou, plus récemment, l'article de Leclerc (1994).

Cette nouvelle direction qu'offrent les dissimilarités à valeurs dans un ensemble ordonné donne la possibilité d'étudier plusieurs cas particuliers intéressants tels que les dissimilarités vectorielles ou encore la comparaison des ultramétriques (classifications hiérarchiques) usuelles

[1] Faculté d'économie et de droit, Mohammédia, Maroc.

qui sont à valeurs dans un ensemble totalement ordonné (LTO) (Benkaraache et Van Cutsem 1993, 1994). Il est aussi intéressant de savoir si les résultats classiques sur l'approximation ultramétrique d'une dissimilarité à valeurs dans un ensemble LTO restent valables et comment ils peuvent être généralisés. Ce dernier point est justement le but du présent travail. Il s'agit de construire l'ultramétrique inférieure maximum d'une dissimilarité à valeurs dans un ensemble ordonné dont l'existence et l'unicité ont été établies dans Critchley et Van Cutsem 1994. Ainsi nous montrons que deux méthodes usuelles (quand L est LTO) s'étendent à des ensembles non totalement ordonnés : ces deux méthodes sont la méthode du lien simple (voir Sibson 1972) et la méthode algébrique introduite dans Van Cutsem (1983).

Dans la deuxième section nous rappelons quelques définitions sur les ensembles ordonnés et les L-dissimilarités ainsi que sur les L-ultramétriques et leur caractérisation par les partitions qu'elles induisent aux différents niveaux de L. Les troisième et quatrième sections sont consacrées à l'étude de l'ultramétrique inférieure maximum. Après avoir rappelé le théorème sur son existence et son unicité, nous présentons une généralisation de la méthode du lien simple au cas où L est un inf-demi-treillis complet, puis nous étendons la méthode algébrique de Van Cutsem 1983 au cas où L est un treillis distributif.

2. L-DISSIMILARITÉS ET L-ULTRAMÉTRIQUES À VALEURS DANS UN ENSEMBLE ORDONNÉ

2.1. Notations et définitions

Rappelons d'abord les notations et les définitions sur les ensembles ordonnés que nous utiliserons dans la suite. Pour plus de détail sur les ensembles ordonnés on peut consulter par exemple Davey et Priestley (1990).

Soit S un ensemble fini de cardinal n et L un ensemble partiellement ordonné. La relation d'ordre partiel sur L (réflexive, transitive et antisymétrique) est notée $\leq$. Rappelons qu'un majorant (resp. minorant) d'une partie de L est un élément de L qui est supérieur (resp. inférieur) ou égal à tous les éléments de cette partie. Le supremum (resp. infimum) d'une partie L (quand il existe) est le plus petit des majorants (resp. plus grand des minorants) communs à cette partie.

Un treillis est un ensemble ordonné où toute paire d'éléments admet un infimum (inf-demi-treillis) et un supremum (sup-demi-treillis). On utilisera les notations suivantes (quand elles existent) : $\inf(x, y) = x \wedge y$ et $\sup(x, y) = x \vee y$.

Nous supposons, dans toute la suite, que L vérifie au moins les deux conditions suivantes :

(LMIN) : L admet un élément minimum noté 0.

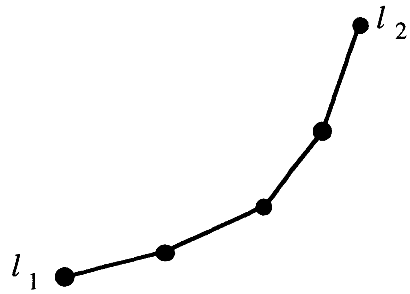
($\bar{\mathrm{L}}$) : tout sous ensemble non vide de L qui admet un majorant possède un supremum.

On montre qu'avec ces deux conditions, toute partie de L admet un infimum. L'ensemble ordonné L est donc un inf-demi-treillis.

Nous supposons en plus, dans toute la suite, que L est fini.

Comme tout inf-demi-treillis fini est complet, l'ensemble ordonné L est donc complet.

Nous représentons un tel ensemble ordonné par son diagramme : deux éléments de L comparables pour l'ordre $\leq$ y sont liés par un chemin d'éléments de L qui est totalement ordonné par $\leq$, l'élément inférieur se trouvant plus bas :

$l_1 \leq l_2 \qquad \equiv$ 

Quand le chemin se réduit à deux éléments $\{l_1 , l_2\}$, on dit que l'élément l_2 couvre l'élément l_1 .

Enfin, la notation $l_1 \not\leq l_2$ signifie l_1 n'est pas inférieur ni égal à l_2 pour l'ordre $\leq$.

Soit S un ensemble fini. Une dissimilarité d sur S et à valeurs dans L (ou une L-dissimilarité) est une application de $S \times S$ dans L qui vérifie $d(a, a) = 0$ pour tout $a \in S$.

Nous ne considérons que les L-dissimilarités symétriques, i.e., les L-dissimilarités d vérifiant $d(a, b) = d(b, a)$ pour tous $a, b \in S$.

Une L-dissimilarité d est dite propre si $d(a, b) = 0$ implique $a = b$.

L'ensemble des L-dissimilarités sur S est muni de l'ordre usuel défini comme suit : si d_1 et d_2 sont deux L-dissimilarités sur S, $(d_1 \leq d_2)$ si $d_1(a, b) \leq d_2(a, b)$ pour tous $a, b \in S$. La dominance est dite stricte s'il existe un couple d'éléments de S pour lequel l'inégalité précédente est stricte.

2.2. L-ultramétrique à valeurs dans un ensemble ordonné

DÉFINITION 1. Soit S un ensemble fini et L un ensemble ordonné. Une L-dissimilarité u sur S est une L-ultramétrique sur S si :

$$\forall\, x \in L, \forall\, (a, b, c) \in S \times S \times S,$$

$$[u(a, b) \leq x \text{ et } u(b, c) \leq x] \quad \Rightarrow \quad u(a, c) \leq x .$$

On dit aussi que le triplet (a, b, c) est u-ultramétrique.

La définition précédente est équivalente à ce que tout majorant de deux des trois éléments $u(a, b), u(b, c)$ et $u(a, c)$ est un majorant du troisième.

Dans le cas où L est un treillis, on a équivalence entre la définition 1 et chacune des propriétés suivantes (Benkaraache 1993) :

(i) Pour tous éléments $a, b, c \in S$:

$$u(a, c) \vee u(c, b) = u(a, b) \vee u(b, c) = u(a, b) \vee u(a, c) = u(a, c) \vee u(b, c) \vee u(a, b).$$

(ii) Pour tous éléments $a, b, c \in S : u(a, b) \leq u(a, c) \vee u(c, b)$.

Dans le cas où L est un produit fini d'ensembles LTO, on montre facilement qu'une L-dissimilarité est une L-ultramétrique si et seulement si toutes ses composantes sont des ultramétriques. Une dissimilarité u à valeurs dans un ensemble LTO est une ultramétrique si elle vérifie l'inégalité ultramétrique suivante :

$$u(a, c) \leq \max (u(a, b), u(b, c) \text{ pour tous } a, b, c \in S.$$

Critchley et Van Cutsem (1994) ont obtenu plusieurs caractérisations d'une L-ultramétrique dont une à l'aide des partitions induites par l'ultramétrique pour chaque valeur de L :

PROPOSITION 1. Soit S un ensemble quelconque, L un ensemble ordonné vérifiant la condition (LMIN), et u une L-dissimilarité sur S. Les deux propositions suivantes sont équivalentes :

(1) u est une L-ultramétrique.

(2) Pour tout $x \in L$, la relation binaire P_x définie par $a\, P_x\, b$ si et seulement si $u\,(a, b) \leq x$ est une relation d'équivalence.

Une L-ultramétrique définit donc, pour chaque élément x de L, une partition dont les classes sont les classes d'équivalence de la relation P_x. Cependant, l'ensemble de partitions P_x ne définit une L-ultramétrique que si pour tout couple d'éléments $a, b \in S$, l'ensemble $\{x \in L \,/\, a\, P_x\, b\}$ a un minimum. La L-ultramétrique induite dans ce cas est définie par $u\,(a, b) = \min\,\{x \in L \,/\, a\, P_x\, b\}$.

Dans ce qui suit, nous assimilerons toute relation d'équivalence P_x à la partition en classes d'équivalence qu'elle induit sur S.

Pour un élément a de S, la classe de P_x qui contient a est la boule (au sens de u) de centre a et de rayon x : $B_u\,(a, x) = \{b \in S \,/\, u\,(a, b) \leq x\}$.

3. ULTRAMÉTRIQUE INFÉRIEURE MAXIMUM D'UNE DISSIMILARITÉ À VALEURS DANS UN INF-DEMI-TREILLIS COMPLET

Dans le cas où L n'est pas totalement ordonné, le théorème suivant (Critchley et Van Cutsem 1994) assure l'existence et l'unicité de l'ultramétrique inférieure maximum :

THÉORÈME 1. Soit L un inf-demi-treillis complet et d une L-dissimilarité sur S. On pose $F_d = \{u$ L-ultramétrique sur $S \,/\, u \leq d\}$. Alors, F_d admet un élément maximum noté $\hat{u}_d$. De plus, $\hat{u}_d = d \Leftrightarrow d \in F_d \Leftrightarrow d$ est L-ultramétrique.

L'ultramétrique inférieure maximum, ou la sous-dominante de d, est l'élément maximum $\hat{u}_d$ de l'ensemble F_d. Elle vérifie (par définition de la maximalité) : si v est une L-dissimilarité de F_d telle que $\hat{u}_d \leq v \leq d$, alors $v = \hat{u}_d$.

Dans le cas classique où L est totalement ordonné (par exemple où $L = \mathbb{R}^+$), on sait déterminer l'unique ultramétrique inférieure maximum d'une dissimilarité d. Il existe plusieurs méthodes et algorithmes qui permettent sa construction : à l'aide de la méthode de réduction des triplets (Roux 1968), en déterminant l'arbre de longueur minimum (Gower et Ross 1969), par l'algorithme du lien simple (single linkage) (Sibson 1972) qui a reçu divers noms, par une méthode algébrique (Van Cutsem 1983) qui s'inscrit dans toute une classe de constructions de dissimilarités et qui sera généralisée plus loin. Toutes ces méthodes ne sont pas équivalentes du point de vue de la complexité algorithmique. La plus efficace est celle de l'arbre minimum, en $O(n^2)$. Le lien simple est en $O(n^3)$. La méthode de Roux (comme celle de Van Cutsem) consiste en une suite d'itérations, chacune en $O(n^3)$. La méthode de Van Cutsem est essentiellement théorique et a l'intérêt d'avoir un "domaine de généralisation" plus étendu.

3.1. Cas où L est un produit d'ensembles totalement ordonnés.

Dans le cas particulier où L est un produit d'ensembles LTO, on a le résultat suivant :

THÉORÈME 2. (Critchley, Communication personnelle)
On suppose que $L = \Pi\{L_i, i \in I\}$ où L_i est LTO et I un ensemble fini. Soit d une L-dissimilarité. Alors, l'ultramétrique inférieure maximum de d est le produit des ultramétriques inférieures maximum des composantes de d :

Si $d = (d_i, \ i \in I)$, alors $\hat{u}_d = (\hat{u}_{d_i}, \ i \in I)$.

Ce résultat découle du fait qu'une L-dissimilarité est une L-ultramétrique si et seulement si toutes ses composantes sont des L_i-ultramétriques. On peut donc, dans ce cas, utiliser les méthodes habituelles de calcul de la sous dominante sur chaque composante L_i de L.

3.2. Construction de l'ultramétrique inférieure maximum dans le cas où L est un inf-demi-treillis fini : généralisation de la méthode du lien simple.

Soit d une L-dissimilarité sur S, a un élément de S et x un élément de L. On note $B_d(a, x)$ la boule (au sens de d) de centre a et de rayon x, et $G(S, x)$, le graphe sur S au seuil x, i.e. le graphe où $\{\alpha, \beta\}$ est une arête si $d(\alpha, \beta) \leq x$.

Introduisons alors $\overline{B_d}(a, x)$, l'ensemble de tous les points de S reliés par un chemin (ou une chaîne) du graphe seuil $G(S, x)$.

PROPOSITION 2. Soit d une L-dissimilarité sur S et x un élément de L. Alors :

1) La relation binaire sur S définie par $a \, R_x \, b \Leftrightarrow b \in \overline{B_d}(a, x)$ est une relation d'équivalence.
2) Si $x \leq x'$ alors R_x est plus fine que $R_{x'}$.

Preuve.

1) Soient a, b et c des éléments quelconques de S. R_x est réflexive car $a \in \overline{B_d}(a, x)$. Elle est symétrique. En effet, si $b \in \overline{B_d}(a, x)$, il existe une chaîne C_{ab} reliant a et b et dont toutes les arêtes sont de longueurs inférieures à x (au sens de d). On a donc aussi $a \in \overline{B_d}(b, x)$ (en considérant la chaîne C_{ba} obtenue en "inversant" C_{ab}), c'est-à-dire $b \, R_x \, a$. R_x est enfin transitive : Si $a \, R_x \, b$ et $b \, R_x \, c$, il suffit de considérer la chaîne C_{ac} reliant a et c dans l'union $C_{ab} \cup C_{bc}$ (il en existe toujours une).

2) Résulte du fait que si $x \leq x'$: $\overline{B_d}(a, x) \subset \overline{B_d}(a, x)$ pour tout a de S. □

REMARQUE 1.

i. L'ensemble $L_{ab} = \{x \in L \,/\, a \, R_x \, b\}$ est non vide car il contient l'élément $d(a, b)$.

ii. L'application $x \mapsto R_x$ est le dendrogramme (ou stratification) défini dans Critchley et Van Cutsem (1994) et aussi (bien avant) par Janowitz (1978) (quand L est un treillis) sous forme d'une application résiduelle. Pour $L = \mathbb{R}^+$, on retrouve les dendrogrammes au sens de Jardine et Sibson (1971).

DÉFINITION 2. Avec les mêmes notations que ci-dessus (S est un ensemble fini et L est un inf-demi-treillis complet), on définit une application α sur l'ensemble $D(S, L)$ des L-dissimilarités sur S à valeurs dans $D(S, L)$ par : $d \mapsto \alpha(d)$ où $\alpha(d)$ est définie par : $\alpha(d)(a, b) = \wedge \{x \in L \,/\, a\, R_x\, b\} = \wedge L_{ab}$ pour tous $a, b \in S$.

On définit de même les itérées de d par α : si $n \geq 2$,

$$\alpha^n(d) = \alpha^{n-1}(\alpha(d)).$$

REMARQUE 2. L'ensemble L_{ab} de la définition 2 s'écrit aussi :

$L_{ab} = \{x \in L \,/\, b \in \overline{B_d}(a, x)\}$

$= \{x \in L$ / il existe une chaîne C_{ab} reliant a et b dans S telle que, pour tout $r \in C_{ab}$ (r est une arête de la chaîne C_{ab}), $d(r) \leq x\}$.

Les deux lemmes suivants serviront dans la suite :

LEMME 1. Avec les mêmes notations que ci-dessus, on a :

$$\alpha(d) \leq d.$$

Preuve. Évidente à partir de la remarque 2. □

LEMME 2. Soit d une L-dissimilarité sur S. La suite $(\alpha^n(d))_{n \geq 1}$ est stationnaire :

$$(\exists\, n \geq 1) \text{ tel que : } \alpha(d) > \alpha^2(d) > \ldots > \alpha^n(d) = \alpha^{n+1}(d) = \ldots .$$

On notera dans la suite u_d la limite de cette suite.

Preuve. Comme les ensembles L et S sont finis l'ensemble $D(S, L)$ est aussi fini. Donc toute suite décroissante sur $D(S, L)$ devient nécessairement stationnaire à partir d'un certain rang. □

On a alors une caractérisation des L-ultramétriques qui est donnée par :

THÉORÈME 3. Soit L un inf-demi-treillis et S un ensemble fini. Soit d une L-dissimilarité sur S. Alors les deux propositions suivantes sont équivalentes :

i) d est une L-ultramétrique.

ii) $d = \alpha(d)$.

Preuve.

<u>(i) ⇒ (ii)</u> : d'après le lemme 1., il suffit de montrer que $d \leq \alpha(d)$. Soit (a, b) un couple d'éléments de S, $L_{ab} = \{x \in L \,/\, a\, R_x\, b\}$. Si $x \in L_{ab}$, alors il existe une chaîne entre a et b, notée $C = (a_o = a, a_1, \ldots, a_k = b)$ telle que $d(a_{i-1}, a_i) \leq x$, $i = 1, \ldots, k$. Il s'ensuit que $d(a, b) \leq x$ puisque d est une L-ultramétrique. Donc $d(a, b) \leq \wedge L_{ab}$, c'est-à-dire $d(a, b) \leq \alpha(d)(a, b)$.

<u>(ii) ⇒ (i)</u> : soient a, b et c trois éléments quelconques de S. On a :

$d(a, b) = \wedge \{x \in L \,/\, a\, R_x\, b\}$ et $d(b, c) = \wedge \{x \in L \,/\, b\, R_x\, c\}$.

Soit alors x_0 un élément de L qui vérifie $d(a, b) \leq x_0$ et $d(b, c) \leq x_0$. On a alors $a\, R_{x_0}\, b$ et $b\, R_{x_0}\, c$, donc aussi $a\, R_{x_0}\, c$ (R_x est transitive), c'est-à-dire $x_0 \in L_{ac}$. Et comme $d(a, c) = \wedge L_{ac}$ (par hypothèse), on a $d(a, c) \leq x_0$.

On a donc montré que d est une L-ultramétrique. □

PROPOSITION 3. Soient d_1 et d_2 deux L-dissimilarités sur S. Avec les mêmes notations que ci-dessus, on a $(d_1 \leq d_2) \Rightarrow (\alpha(d_1) \leq \alpha(d_2))$.

Preuve. On note par $R_{i,x}$ $(i = 1, 2)$ la relation d'équivalence associée à d_i $(i = 1, 2)$ (voir proposition 2) au niveau x et soit (a, b) un couple d'éléments de S. Par hypothèse, on a : $d_1(a, b) \leq d_2(a, b)$. Donc : $a R_{2,x} b$ implique que a $R_{1,x} b$.
Si x_0 est un élément de l'ensemble $\{x \in L \,/\, a R_{2,1} b\}$, x_0 est aussi un élément de l'ensemble $\{x \in L \,/\, a R_{1,x} b\}$, c'est-à-dire $\{x \in L \,/\, a R_{2,x} b\} \subseteq \{x \in L \,/\, a R_{1,x} b\}$ et $\wedge \{x \in L \,/\, a R_{2,x} b\} \geq \wedge \{x \in L \,/\, a R_{1,x} b\}$. D'où $\alpha(d_1)(a, b) \leq \alpha(d_2)(a, b)$. □

On conclut enfin par le résultat suivant :

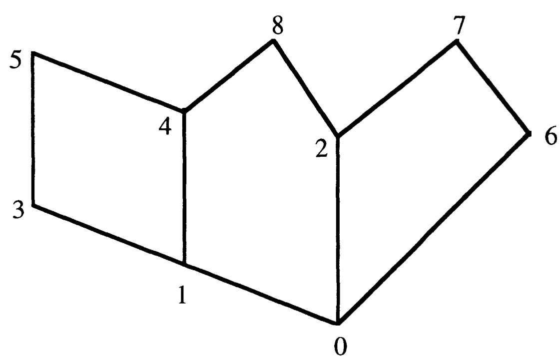
COROLLAIRE 1. Soit d une L-dissimilarité et v une L-ultramétrique telle que $v \leq d$. Alors $v \leq u_d$ de sorte que u_d est l'ultramétrique inférieure maximum de d.

Preuve. Nous savons, en vertu du théorème 3, que $\alpha^n(v) = v$ pour tout $n \geq 1$. En appliquant la proposition 3, nous obtenons $\alpha^n(v) = v \leq \alpha^n(d)$, pour tout $n \geq 1$. Comme u_d est la limite de la suite $\alpha^n(d)$ (voir lemme 2), il vient que $v \leq u_d$.

D'où le résultat. □

3.3. Exemple 1.

Soit $S = \{a, b, c\}$ et L l'inf-demi-treillis complet défini par le graphe suivant :



Soit D la L-dissimilarité sur S suivante :

$$D = \begin{array}{c|ccc} & a & b & c \\ a & \mathbf{0} & 4 & 5 \\ b & 4 & \mathbf{0} & 7 \\ c & 5 & 7 & \mathbf{0} \end{array}$$

Calcul de $\alpha(D)$: (Π_0 désigne la partition en singletons)

x	R_x
0	Π_0
1	Π_0
2	Π_0
3	Π_0
4	$(a\,b)(c)$
5	S
6	Π_0
7	$(a)(b\,c)$
8	$(a\,b)(c)$

donc :

(i, j)	L_{ij}	$\alpha(D)(i, j)$
(a, b)	4, 5, 8	4
(a, c)	5	5
(b, c)	7, 5	0

d'où,

$$\alpha(D) = \begin{array}{cccc} & a & b & c \\ a & \mathbf{0} & 4 & 5 \\ b & 4 & \mathbf{0} & 0 \\ c & 5 & 0 & \mathbf{0} \end{array}$$

$\alpha(D)$ n'est pas une L-ultramétrique, il faut donc calculer $\alpha^2(D)$:

— Calcul de $\alpha^2(D)$:

x	R_x
0	$(a)(b\,c)$
1	$(a)(b\,c)$
2	$(a)(b\,c)$
3	$(a)(b\,c)$
4	S
5	S
6	$(a)(b\,c)$
7	$(a)(b\,c)$
8	S

donc :

(i, j)	L_{ij}	$\alpha^2(D)(i, j)$
(a, b)	4, 5, 8	4
(a, c)	4, 5, 8	4
(b, c)	L	0

d'où,

$$\alpha^2(D) = \begin{array}{cccc} & a & b & c \\ a & \mathbf{0} & 4 & 4 \\ b & 4 & \mathbf{0} & 0 \\ c & 4 & 0 & \mathbf{0} \end{array}$$

En calculant $\alpha^3(D)$, on trouve : $\alpha^3(D) = \alpha^2(D)$. Donc α^2(D) est l'ultramétrique inférieure maximum de D.

4. CAS OÙ LA DISSIMILARITÉ EST À VALEURS DANS UN TREILLIS DISTRIBUTIF

Nous allons présenter un algorithme qui permet d'obtenir en moins de n-1 étapes (n étant le cardinal de S), l'ultramétrique inférieure maximum d'une dissimilarité à valeurs dans un treillis distributif. Notre approche est une généralisation de la méthode algébrique proposée par Van Cutsem (1983) quand la dissimilarité est à valeurs dans un ensemble totalement ordonné. Nous commençons ce paragraphe par montrer que l'hypothèse de la distributivité de L est nécessaire pour introduire l'opération $*$ à un ordre supérieur. Nous démontrons au passage quelques propriétés simples qui ne nécessitent pas la distributivité, en particulier une caractérisation de la L-ultramétricité. Nous montrons ensuite (l'hypothèse de distributivité devient alors nécessaire) que la limite de la suite des itérées de la L-dissimilarité d pour l'opération $*$ converge vers l'ultramétrique inférieure maximum de d.

Soit L un treillis. Ses opérations inf et sup seront notées $\wedge$ et $\vee$ respectivement. Le treillis L est dit distributif lorsque :

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z),$$

ou de manière équivalente,

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$$

pour tous $x, y, z \in L$.

4.1. Définition et propriétés de l'opération $*$ entre L-dissimilarités

DÉFINITION 3. Soit L un treillis distributif et $\wedge$ et $\vee$ les deux opérations qui lui sont associées. Si f et g sont deux L-dissimilarités sur S, alors l'opération algébrique $*$ entre f et g est définie par :

$$\forall (a,b) \in S^2, \qquad (f * g)\,(a, b) = \wedge\{f\,(a, c) \vee g\,(c, b) \,/\, c \in S\ \}.$$

On remarque que $(f * g)(a, b)$ est définie à l'aide des chemins C_{ab} à deux arêtes. La distributivité de L est nécessaire pour définir l'opération $*$ à un ordre $k > 2$:

$$f^{*k}\,(a, b) = (f^{*(k-1)} * f)\,(a, b).$$

En effet,

PROPOSITION 4. Si le treillis L est distributif, l'opération $*$ définie ci-dessus est associative.

Preuve. Soit (f, g, h) un triplet de L-dissimilarités et (a, b) un couple d'éléments de S.

On a alors,

$$
\begin{aligned}
((f * g) * h)(a, b) &= \wedge\{(f * g)(a, c) \vee h(c, b) \,/\, c \in S\}. \\
&= \wedge\{\wedge\{f(a, d) \vee g(d, c) \,/\, d \in S\} \vee h(c, b) \,/\, c \in S\}. \\
&= \wedge\{\wedge\{[f(a, d) \vee g(d, c)] \vee h(c, b) \,/\, d \in S\} / c \in S\} \quad \text{(distributivité).} \\
&= \wedge\{\wedge\{f(a, d) \vee [g(d, c) \vee h(c, b)] \,/\, c \in S\} \,/\, d \in S\} \quad (\vee \text{ est associative).} \\
&= \wedge\{f(a, d) \vee (\wedge\{[g(d, c) \vee h(c, b)] \,/\, c \in S\}) \,/\, d \in S\} \quad \text{(distributivité).} \\
&= \wedge\{f(a, d) \vee (g * h)(d, b) \,/\, d \in S\}. \\
&= (f * (g * h))(a, b).
\end{aligned}
$$

□

Énonçons tout d'abord quelques propriétés de l'opération $*$ qui ne nécessitent pas la distributivité de L :

PROPOSITION 5. Soit S un ensemble fini et L un inf-demi-treillis. Soient f et g deux L-dissimilarités sur S. Alors :

(i) $f * g \leq f$ et $f * g \leq g$

(ii) $f * g \leq f \wedge g$

(iii) Si $f \geq g$, alors $f * h \geq g * h$ pour toute L-dissimilarité h sur S.

Preuve.
(i) $f * g(a, b) \leq f(a, b) \vee g(b, b) = f(a, b)$ et de même pour la deuxième inégalité.
(ii) résulte de (i).
(iii) Soit $(a, b) \in S \times S$; on a $(f * h)(a, b) = \{f(a, c) \vee h(c, b) \,/\, c \in S\}$.
Or, pour tout $c \in S$, on a :

$$f(a, c) \vee h(c, b) \geq g(a, c) \vee h(c, b) \geq (g * h)(a, b).$$

donc $\wedge\{f(a, c) \vee h(c, b) \,/\, c \in S\} \geq (g * h)(a, b).$ □

On a alors une caractérisation de la L-ultramétricité par le théorème suivant (où la distributivité de L n'est toujours pas nécessaire puisqu'on n'utilise que la définition de l'opération $*$ à l'ordre 1) :

THÉORÈME 4. Si L est un treillis et f une L-dissimilarité sur S, alors les deux propositions suivantes sont équivalentes :

1) f est une L-ultramétrique

2) $f * f = f$.

Preuve.
1) $\Rightarrow$ 2). Supposons f L-ultramétrique et considérons $a, b \in S$. Alors pour tout $c \in S$: $f(a,c) \vee f(c,b) \geq f(a,c)$ et $f(a,c) \vee f(c,b) \geq f(c,b)$ impliquent $f(a,c) \vee f(c,b) \geq f(a,b)$ (de par la L-ultramétricité de f). Donc $(f*f)(a,b) = \wedge\{f(a,c) \vee f(c,b) / c \in S\} \geq f(a,b)$. Et puisque dans le membre de gauche de l'inégalité il existe un terme égal à $f(a,b)$ (pour $c = a$ ou b), l'égalité est atteinte : $(f*f)(a,b) = f(a,b)$.

2) $\Rightarrow$1). Supposons maintenant que $f*f=f$. Soient a, b et c des éléments de S et $x \in L$ tels que $f(a,c) \leq x$ et $f(c,b) \leq x$. Alors :

$$f(a,b) = f*f(a,b) = \wedge\{f(a,e) \vee f(e,b) / e \in S\} \leq f(a,c) \vee f(c,b) \leq x. \qquad \square$$

4.2. Calcul de l'ultramétrique inférieure maximum quand L est un treillis distributif

Supposons dans la suite que L est un treillis distributif et montrons le théorème suivant :

THÉORÈME 5. Soit S un ensemble fini, L un treillis distributif et f une L-dissimilarité sur S. La suite des L-dissimilarités $\{(f^{*k}) / k \geq 1\}$ est décroissante et il existe un entier h_0 tel que $f^{*h} = f^{*h_0}$ pour tout $h \geq h_0$. De plus, f^{*h_0} est une L-ultramétrique.

Preuve. La décroissance est une conséquence de la proposition (5 - ii) en faisant $f = g$.

Comme l'ensemble des valeurs prises par les L-dissimilarités (f^{*k}) sont bornées par celles de f et L vérifie (LMIN), la suite $\{(f^{*k}),\ k \geq 1\}$ étant décroissante et à valeurs dans un ensemble fini est nécessairement stationnaire. Il existe donc $k_0 \geq 1$ tel que

$$f^{*(k_0+1)} \in \left\{f, f^{*2}, \ldots, f^{*k_0}\right\}.$$

Il existe alors $h_0 \in [1, \ldots, k_0]$ tel que : $f^{*(k_0+1)} = f^{*h_0}$. Comme (f^{*k}) est décroissante, on a donc $f^{*h_0} \geq f^{*h_0+1} \geq \ldots \geq f^{*k_0+1} = f^{*h_0}$. D'où $f^{*h_0} = f^{*h_0+1} = \ldots = f^{*k_0+1}$.

Donc $(\forall h > h_0)\quad f^{*h} = f^{*h_0}\ (= \hat{f})$.

$\hat{f}$ est une ultramétrique puisque $\hat{f} * \hat{f} = f^{*h_0} * f^{*h_0} = f^{*2h_0} = \hat{f}$. $\square$

Le terme h_0 de la démonstration précédente n'est autre que le degré d'une dissimilarité défini dans Van Cutsem (1983) dans le cas où L est LTO.

THÉORÈME 6. Sous les mêmes notations que ci-dessus, l'ultramétrique inférieure maximum de f est égale à $\hat{f}$.

Preuve. Par définition de la sous-dominante $\hat{u}_f$ de f, on a : $f \geq \hat{u}_f \geq \hat{f}$. (1)

On compose par $\hat{f}$ les trois membres de (1), on obtient :

$$\hat{f} = f * \hat{f} \geq \hat{u}_f * \hat{f} \geq \hat{f} * \hat{f} = \hat{f} \text{ (en utilisant le théorème 4)}$$

donc $\hat{u}_f * \hat{f} = \hat{f}$. (2)

D'autre part, on a aussi : $f \geq \hat{u}_f \Rightarrow f * \hat{u}_f \geq \hat{u}_f * \hat{u}_f = \hat{u}_f$

donc $f^{*k} * \hat{u}_f \geq \hat{u}_f$ pour tout $k \geq 1$. (3)

En combinant les relations (2) et (3), on obtient $\hat{f} \geq \hat{u}_f$. Or d'après (1), on a aussi $\hat{u}_f \geq \hat{f}$, d'où l'égalité. □

Il reste à montrer maintenant que la limite est atteinte en moins de n-1 étapes.

LEMME 3. Avec les mêmes notations que ci-dessus, on a :

$$f^{*k}(a, b) = \min \{ \mid C_k \mid / \ C_k \in C_K(a, b) \}$$

où $\quad |C_k| = \vee\{f(a, a_1), ..., f(a_{k-1}, b)\}$ quand $C_k = (a, a_1, ..., a_{k-1}, b)$.

et $\quad C_k(a, b)$ désigne l'ensemble des chaînes reliant a et b avec un nombre d'arêtes au plus égal à k.

Preuve. Utilisons un raisonnement par récurrence. Le résultat est vrai pour $k = 2$:

$f^{*2}(a, b) = \wedge\{f(a, c) \vee f(c, b) / c \in S\} = \wedge\{ \mid C \mid / C \in C_2(a, b)\}$.

Supposons que le résultat est vrai à l'ordre k - 1. A l'ordre k on a :

$$\begin{aligned} f^{*k}(a, b) &= (f^{*k-1} * f)(a, b) \\ &= \wedge\{f^{*k-1}(a, c) \vee f(c, b), c \in S\} \\ &= \wedge\{[\wedge\{ \mid C_{k-1} \mid / C_{k-1} \in C_{k-1}(a, c)\}] \vee f(c, b) / c \in S\} \\ &= \wedge\{ \mid C'_k(c) \mid / c \in S\} \end{aligned}$$

où $C'_k(c) = C_{k-1} + (c, b)$, (chaîne reliant a à c avec au plus k-1 arêtes, ensuite reliée à b à l'aide de l'arête (c, b)). Donc $C'_k(c)$ est une chaîne de $C_k(a, b)$.

Par conséquent, $f^{*k}(a, b) = \wedge\{ \mid C_k \mid / C_k \in C_k(a, b)\}$. □

THÉORÈME 7. Soit f une L-dissimilarité. Si le cardinal de S est égal à n, alors $f^{*n} = f^{*n-1}$.

Preuve. Soit a et b deux éléments distincts de S. Il suffit de vérifier que $C_{n-1}(a, b) = C_n(a, b)$ pour a, b donnés dans S, et d'en tirer l'égalité souhaitée.

Clairement $C_{n-1}(a, b) \subset C_n(a, b)$. Supposons qu'un élément de $C_n(a, b)$ ait n arêtes. Alors cet élément compterait n+1 sommets, ce qui est impossible. D'où $C_n(a, b) \subset C_{n-1}(a, b)$.

Par conséquent $f^{*(n-1)} = f^{*n}$. □

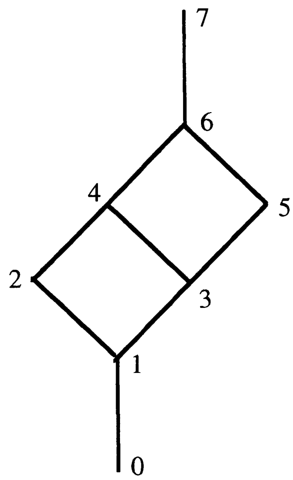
4.4. Exemple 2

$S = \{a, b, c, d, e\}$.

Soit la L-dissimilarité D sur S suivante :

$$D = \begin{array}{c|ccccc} & a & b & c & d & e \\ \hline a & \mathbf{0} & 3 & 2 & 6 & 5 \\ b & 3 & \mathbf{0} & 3 & 1 & 4 \\ c & 2 & 3 & \mathbf{0} & 4 & 5 \\ d & 6 & 1 & 4 & \mathbf{0} & 7 \\ e & 5 & 4 & 5 & 7 & \mathbf{0} \end{array}$$

où les valeurs prises par D sont dans le treillis distributif suivant :



$$D^{*2} = \begin{array}{c|ccccc} & a & b & c & d & e \\ \hline a & \mathbf{0} & 3 & 1 & 3 & 3 \\ b & 3 & \mathbf{0} & 3 & 1 & 3 \\ c & 1 & 3 & \mathbf{0} & 3 & 3 \\ d & 3 & 1 & 3 & \mathbf{0} & 4 \\ e & 3 & 3 & 3 & 4 & \mathbf{0} \end{array}$$

$$D^{*3} = \begin{array}{c|ccccc} & a & b & c & d & e \\ \hline a & \mathbf{0} & 3 & 1 & 3 & 3 \\ b & 3 & \mathbf{0} & 3 & 1 & 3 \\ c & 1 & 3 & \mathbf{0} & 3 & 3 \\ d & 3 & 1 & 3 & \mathbf{0} & 3 \\ e & 3 & 3 & 3 & 3 & \mathbf{0} \end{array}$$

et $D^{*4} = D^{*3}$

Donc l'ultramétrique inférieure maximum de D est égale à D^{*3}.

BIBLIOGRAPHIE

BARTHÉLEMY, J.P., LECLERC, B., MONJARDET, B., "Ensembles ordonnés et taxonomie mathématique", in M. Pouzet and D. Richard (eds.), *Orders : Description and roles, Annals of Discrete Mathematics* , 23, 1984, 523 - 548.

BENKARAACHE, T., "Problèmes de validité en classification et quelques généralisations aux ultramétriques à valeurs dans un ensemble ordonné", Thèse de doctorat, Université Joseph Fourier, Grenoble, 1993.

BENKARAACHE, T., VAN CUTSEM, B., "Complexité d'une hiérarchie et comparaison de hiérarchies : une revue et quelques nouvelles idées ", *Communication aux Journées de Statistiques de Vannes*, 1993.

BENKARAACHE, T., VAN CUTSEM, B., "Comparison of hierarchical classifications", *Rapport technique du laboratoire LMC-IMAG,* RT 100, Grenoble, 1994.

BENZECRI, J.P., *L'analyse des données.* I - *La taxinomie*, Paris, Dunod, 1973.

CRITCHLEY, F., VAN CUTSEM, B., "An order - theoretic unification and generalization of certain fundamental bijections", in Bernard Van Cutsem (ed.), *Classification and Dissimilarity Analysis*, *Lecture Notes in Statistics*, vol. 93, New York, Springer Verlag, 1994, 87-148.

DAVEY, B.A., PRIESTLEY, H. A., *Introduction to lattice and order*, Cambridge (U. K.), Cambridge University Press, 1990.

GOWER, J.C., ROSS, J.S. , "Minimum Spaning trees and Single Linkage Cluster Analysis", *Applied Stat.* 18, 1969, 54 - 64.

JANOWITZ, M.F., "An order theoretic model for cluster analysis", *SIAM J. Appl. Math.* 34, 1978, 55-72.

JARDINE, N., SIBSON, R., *Mathematical taxonomy*, London, Wiley, 1971.

JOHNSON, S.C., "Hierarchical clustering schemes", *Psychometrika* 32, 1967, 241-254.

LECLERC, B., "The residuation model for the ordinal construction of dissimilarities and other valued objects", in Bernard Van Cutsem (ed.), *Classification and Dissimilarity Analysis*, *Lecture notes in Statistics* 93, New York, Springer Verlag, 1994, 149-172.

ROUX, M., "Un algorithme pour construire une hiérarchie particulière", Thèse de 3ème cycle, Université de Paris VI, 1968.

SIBSON, R., "Slink : an optimally efficient algorithm for the single link cluster method", *Computer Journal* 16, n°1, 1972, 30-34.

VAN CUTSEM, B., "Ultramétriques, distances, φ-distances maximales dominées par une dissimilarité", *Statistique et Analyse des Données* 8, 1983, 42-63.