RAIRO. RECHERCHE OPÉRATIONNELLE

J. L. CHANDON
J. LEMAIRE
J. POUGET

Construction de l'ultramétrique la plus proche d'une dissimilarité au sens des moindres carrés

RAIRO. Recherche opérationnelle, tome 14, nº 2 (1980), p. 157-170

<http://www.numdam.org/item?id=RO_1980__14_2_157_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 14, n° 2, mai 1980, p. 157 à 170)

CONSTRUCTION DE L'ULTRAMÉTRIQUE LA PLUS PROCHE D'UNE DISSIMILARITÉ AU SENS DES MOINDRES CARRÉS (*)

par J. L. CHANDON ([1]), J. LEMAIRE et J. POUGET ([2])

Résumé. — *Soit un ensemble X d'objets à classer sur la base d'une matrice de dissimilarités entre paires d'objets. On se propose de réaliser une classification hiérarchique des objets de X, telle que la transformation des dissimilarités initiales d en dissimilarités ultramétriques δ induise une distorsion minimale au sens des moindres carrés.*

Une propriété générale d'un optimum global est démontrée. Elle sert de fondement à l'élaboration d'un algorithme hiérarchique ascendant garantissant l'obtention d'un optimum global. Ce premier algorithme, relativement lent, est complété par un algorithme hiérarchique divisif rapide qui permet de classer un grand nombre d'objets et qui conduit généralement à un optimum local de meilleure qualité que ceux obtenus par des algorithmes hiérarchiques classiques.

Abstract. — *X is a set of objects to be classified on the basis of the pairwise dissimilarity matrix D. This paper present a hierarchical classification algorithm that minimize the sum of squared differences between the original dissimilarities d and the final ultrametric dissimilarities δ resulting from the classification.*

A general property of the global optimum is demonstrated. It is then converted into a branch and bound algorithm that insure that the true global optimum will be reached. However, this algorithm is rather slow. It is supplemented by a divisive algorithm that can handle large sets of objects. This second algorithm may lead to more local optimum but it often improve upon the results of the classical hierarchical algorithms.

1. PRÉSENTATION DU PROBLÈME

X désigne un ensemble fini d'objets. On notera $\mathbb{P}$ l'ensemble des paires d'objets de X, c'est-à-dire l'ensemble des parties à deux éléments de X. Pour simplifier, une telle partie $\{x, y\}$ sera notée xy.

Étant donné une dissimilarité d sur X, c'est-à-dire une application de $\mathbb{P}$ dans $\mathbb{R}$ [$d(xy)$ sera noté d_{xy}] on se propose de réaliser une classification hiérarchique indicée des objets de X. S. C. Johnson [13] et J. P. Benzecri [3] ont démontré conjointement qu'il suffisait pour cela de transformer d en une dissimilarité ultramétrique δ (en abrégé DU) vérifiant l'inégalité ultramétrique

$$\forall xy, \quad yz \in \mathbb{P} : \quad \delta_{xz} \leqq \max(\delta_{xy}, \delta_{yz}). \tag{1}$$

(*) Reçu novembre 1978.
([1]) I.A.E., Aix-en-Provence.
([2]) LASSY (Laboratoire de Signaux et Systèmes), Équipe de Recherche associée au C.N.R.S., Université de Nice, Nice.

R.A.I.R.O. Recherche opérationnelle/Operations Research, 0399-0842/1980/157 /$ 5.00

L'écart entre deux dissimilarités d et δ peut être mesuré par la norme euclidienne ([3]) :

$$\|d-\delta\|=[\sum_{xy}(d_{xy}-\delta_{xy})^2]^{1/2}. \tag{2}$$

Cet article présente une méthode permettant de résoudre le problème :

$$\min_{\delta DU}\|d-\delta\|. \tag{3}$$

R. R. Sokal et F. J. Rholf [17] puis R. R. Sokal et P. H. A. Sneath [18] furent les premiers à poser ce problème sous une forme équivalente (en termes de corrélation cophénétique).

Devant la complexité combinatoire et la nature non convexe du problème, plusieurs algorithmes de recherche d'un optimum local ont été proposés. Une première famille d'algorithmes repose sur des opérations locales consistant à échanger ou à supprimer certains sommets de la représentation arborescente des dissimilarités ultramétriques : J. A. Hartigan [11], D. G. Caroll et J. J. Chang [5]. Une deuxième famille d'algorithme utilise une méthode de gradient : D. J. Caroll et S. Pruzansky [6].

Signalons également les études empiriques de Sokal et Rholf [17], Boyce [4], Farris [9] et Jambu [12] qui montrent que parmi la famille des algorithmes ascendants hiérarchiques de G. N. Lance et W. T. Williams [15], l'algorithme de la moyenne est celui qui conduit le plus fréquemment à la plus petite valeur du critère $\|d-\delta\|$.

On trouvera dans la thèse de Davy de Virville [7] et dans l'article de D. Florens et E. O. Lochard [10] une étude des propriétés mathématiques d'un optimum *global*. Notre article explore cette voie de recherche en utilisant très largement le concept de préordonnance ultramétrique binaire introduit par I. C. Lerman (par exemple [16]) et la technique de regression isotone.

Cette nouvelle formulation du problème nous a permis de découvrir une propriété originale d'un optimum global. Celle-ci induit un algorithme hiérarchique ascendant généralisant l'algorithme de la moyenne et garantissant l'obtention d'un optimum global. Cet algorithme permet également de mieux comprendre l'efficacité de l'algorithme de la moyenne constatée empiriquement pour minimiser ce critère.

([3]) D. Defays [8] étudie un problème analogue avec la norme

$$\|d-\delta\|=\sum_{xy}|d_{xy}-\delta_{xy}|.$$

Cependant, cet algorithme étant relativement lent (il devient impraticable dès que le nombre d'objets dépasse 15), nous avons été conduit à définir sur la base de la propriété mentionnée ci-dessus un algorithme hiérarchique divisif rapide qui permet de classer un grand nombre d'objets et qui conduit généralement à un optimum local de meilleure qualité que ceux obtenus par les algorithmes hiérarchiques classiques.

2. FORMALISATION DU PROBLÈME (3) EN TERMES DE PRÉORDONNANCE ULTRAMÉTRIQUE

Une préordonnance ultramétrique sur X (en abrégé PU) est un préordre total sur $\mathbb{P}$ noté $\Gamma : \Gamma_1 < \Gamma_2 < \ldots < \Gamma_K$ et vérifiant :

$$\forall k : \quad xy, yz \in \bigcup_{j=1}^{k} \Gamma_j \quad \Rightarrow \quad xz \in \bigcup_{j=1}^{k} \Gamma_j. \tag{4}$$

La définition d'une DU δ équivaut à celle d'une PU Γ et d'une suite numérique $\delta_1 < \delta_2 < \ldots < \delta_K$ en posant :

$$\forall k, \quad \forall xy \in \Gamma_k : \quad \delta_{xy} = \delta_k.$$

Le cône convexe des DU construites de cette manière à partir d'une PU Γ sera noté Δ_Γ. Ainsi, le problème (3) peut se formuler

$$\min_{\Gamma PU} \min_{\delta \in \Delta_\Gamma} \| d - \delta \| . \tag{5}$$

L'opportunité d'une technique de régression isotone apparaît clairement en considérant la fermeture $\overline{\Delta}_\Gamma$ de Δ_Γ ; cette fermeture contient toutes les DU δ définies comme précédemment mais à partir d'une suite numérique non décroissante : $\delta_1 \leqq \delta_2 \leqq \ldots \leqq \delta_K$. Non seulement le problème

$$\min_{\Gamma PU} \min_{\delta \in \overline{\Delta}_\Gamma} \| d - \delta \| , \tag{6}$$

équivaut au problème (3) car on obtient toujours une PU en regroupant des classes consécutives d'une PU, mais en plus, cette dernière formulation conduit tout naturellement au problème suivant de régression isotone lorsque Γ est fixée

$$\min_{\delta \in \overline{\Delta}_\Gamma} \| d - \delta \| . \tag{7}$$

L'algorithme J. B. Kruskal [14] — *cf.* aussi D. Florens et E. O. Lochard [10] — fournit aisément la solution $\hat{d}_\Gamma$ du problème (7).

On l'obtient en regroupant des classes consécutives de Γ. Notant $\Sigma_1 < \Sigma_2 < \ldots < \Sigma_L$ la PU obtenue on a

$$\forall l, \quad \forall xy \in \Sigma_l : \quad \hat{d}_\Gamma(xy) = \bar{d}_{\Sigma_l}$$

(dans toute la suite, si Q est une partie de $\mathbb{P}$, $\bar{d}_Q$ désignera la moyenne sur Q des d_{xy}).

Il reste donc à résoudre le problème

$$\min_{\Gamma \mathrm{PU}} \| d - \hat{d}_\Gamma \| . \tag{8}$$

Nous allons établir une propriété possédée par ses solutions. Le principal résultat est le théorème 1 qui permet de restreindre considérablement le nombre de PU à examiner pour obtenir les solutions du problème (3) *via* les solutions « binaires » du problème (8). Auparavant, il convient de préciser quelques concepts classiques.

3. SUR LES PRÉORDONNANCES ULTRAMÉTRIQUES

Le caractère récurrent de l'algorithme proposé au paragraphe 5 nécessite le concept de *section commençante de PU* sur X (en abrégé SCPU). On désigne ainsi une famille de parties de $\mathbb{P}$ notée : $\Gamma_1 < \Gamma_2 < \ldots < \Gamma_L$ et pouvant être complétée pour former une

$$\mathrm{PU} : \quad \Gamma_1 < \Gamma_2 < \ldots < \Gamma_L < \Gamma_{L+1} < \ldots < \Gamma_K .$$

En particulier, toute PU est une SCPU et il en est de même pour la famille vide.

PROPOSITION 1 : *Soit $\Gamma_1 < \Gamma_2 < \ldots < \Gamma_L$ une SCPU la relation définie par :*

$$x \sim y \qquad si \quad x = y \quad ou \quad xy \in \Gamma_1 \cup \ldots \cup \Gamma_L$$

est une relation d'équivalence sur X.

Démonstration : Seule la transitivité n'est pas évidente. Supposons $x \sim y$ et $y \sim z$. Si deux quelconques des trois éléments x, y et z sont égaux, il est clair que $x \sim z$. Sinon xy et yz appartient à $\bigcup_{j=1}^{L} \Gamma_j$ et il en est de même pour xz d'après la propriété (4).

PROPOSITION 2 : *Soient $\Gamma_1 < \Gamma_2 < \ldots < \Gamma_L < \Gamma_{L+1}$ une SCPU et Π la partition de X induite par la relation d'équivalence associée à la SCPU $\Gamma_1 < \Gamma_2 < \ldots < \Gamma_L$* (proposition 1). *La relation définie par :*

$$C \approx D \qquad si \quad C = D \quad ou \quad \exists\, xy \in \Gamma_{L+1} : \quad x \in C \quad \text{et} \quad y \in D$$

est une relation d'équivalence sur Π. *De plus :*

$$\Gamma_{L+1} = \bigcup_{\substack{C,\, D \text{ classes distinctes de } \Pi \\ C \approx D}} \langle C, D \rangle,$$

où

$$\langle C, D \rangle = \{ xy;\ x \in C \text{ et } y \in D \}.$$

Démonstration : Là encore, seule la transitivité n'est pas évidente. Supposons $C \approx D$ et $D \approx E$. Si deux quelconques des trois classes C, D *et* E sont égales, il est clair que $C \approx E$. Sinon, soient $x \in C$, y, $u \in D$ et $v \in E$ tels que xy et uv appartiennent à Γ_{L+1}. y et u étant dans la même classe de Π, on a

$$y = u \qquad \text{ou} \qquad yu \in \Gamma_1 \cup \ldots \cup \Gamma_L.$$

Ainsi xy et yv ou xy, yu et uv appartiennent toutes à $\Gamma_1 \cup \ldots \cup \Gamma_L \cup \Gamma_{L+1}$. La propriété (4) d'ultramétricité implique alors :

$$xv \in \Gamma_1 \cup \ldots \cup \Gamma_{L+1}.$$

En fait, $xv \in \Gamma_{L+1}$ car x et v ne sont pas dans une même classe de Π. On en conclue que $C \approx E$. Si maintenant $xy \in \Gamma_{L+1}$, il existe deux classes distinctes C et D de Π telles que $x \in C$ et $y \in D$. L'existence de la paire xy assure que $C \approx D$. De plus $xy \in \langle C, D \rangle$. Réciproquement si C et D sont deux classes distinctes de Π telles que $xy \in \langle C, D \rangle$ et $C \approx D$, il existe une paire $uv \in \Gamma_{L+1}$ telle que $u \in C$ et $v \in D$. x, u d'une part et y, v d'autre part, étant dans la même classe de Π, en supprimant éventuellement certaines paires dans les cas d'égalité, on en déduit que xu, uv, vy appartiennent toutes à $\Gamma_1 \cup \ldots \cup \Gamma_{L+1}$. Comme précédemment, l'appartenance de xy à Γ_{L+1} en découle.

Cette proposition précise la construction des SCPU. Inversement :

PROPOSITION 3 : *Soient* $\Gamma_1 < \Gamma_2 < \ldots < \Gamma_L$ *une SCPU,* Π *la partition sur* X *induite par la relation d'équivalence associée (proposition* 1) *et* $\approxeq$ *une relation d'équivalence sur* Π.

En posant

$$\Gamma_{L+1} = \bigcup_{\substack{C,\, D \text{ classes distinctes de } \Pi \\ C \approxeq D}} \langle C, D \rangle.$$

On obtient une SCPU :

$$\Gamma_1 < \ldots < \Gamma_L < \Gamma_{L+1}.$$

Avec les notations de la proposition 2, $\approx$ *et* $\approxeq$ *coïncident et la partition* Π' *induite par la relation d'équivalence associée à cette nouvelle SCPU est obtenue à partir de* Π *en réunissant les classes* $\approx$*-équivalentes.*

Démonstration : On utilise le même type de raisonnement que précédemment.

En particulier *les sections commençantes de préordonnances ultramétriques binaires sur* X (en abrégé SCPUB) sont les SCPU $\Gamma_l < \ldots < \Gamma_L$ telles que pour tout $l < L$, Γ_{l+1} soit égal à $\langle C, D \rangle$ où C et D sont deux classes distinctes de la partition induite par la relation d'équivalence associée à la SCPU $\Gamma_1 < \ldots < \Gamma_l$.

Comme toute SCPU peut être obtenue à partir d'une SCPUB en réunissant des classes consécutives de cette dernière (cela résulte immédiatement de la proposition 2), on obtient une nouvelle formulation du problème (3) :

$$\min_{\Gamma \text{PUB}} \| d - \hat{d}_\Gamma \|. \tag{9}$$

L'énumération de toutes les PUB est vite fastidieuse : si $n = \text{card}(X)$, il y en a $n!(n-1)!/2^{n-1}$ (M. Barbut [1] par exemple). Une option simplificatrice consiste à construire une PUB Γ, par récurrence, sans retour en arrière. C'est le cas en particulier pour l'algorithme de la moyenne. A partir d'une SCPUB : $\Gamma_1 < \ldots < \Gamma_l$, la $l+1$-ième itération construit : $\Gamma_{l+1} = \langle C, D \rangle$, où C, D sont deux classes distinctes de la partition Π à distance moyenne D_{CD} minimale :

$$D_{CD} = \frac{1}{\text{card}(C)\,\text{card}(D)} \sum_{\substack{x \in C \\ y \in D}} d_{xy}.$$

Observons que $\bar{d}_{\Gamma_{l+1}} = D_{CD}$. Il est bien connu que la PUB ainsi construite ne comporte pas d'inversions : $\bar{d}_{\Gamma_1} \leqq \bar{d}_{\Gamma_2} \leqq \ldots \leqq \bar{d}_{\Gamma_K}$ (par exemple *cf.* J. P. Benzécri [2]).

Dans ce cas, le problème de régression isotone évoqué au paragraphe 2 admet comme solution triviale

$$\forall k, \quad \forall xy \in \Gamma_k : \quad \hat{d}_\Gamma(xy) = \bar{d}_{\Gamma_k}$$

et $\| d - \hat{d}_\Gamma \|^2$ est égal à l'inertie intraclasse $\sum_k I(\Gamma_k)$.

Nous allons voir que cette propriété est vraie également pour toute solution du problème (9).

4. PROPRIÉTÉ D'UNE SOLUTION DU PROBLÈME (9)

THÉORÈME 1 : *Si* $\Gamma_1 < \ldots < \Gamma_K$ *est une solution du problème* (9) *alors :*

$$\bar{d}_{\Gamma_1} \leqq \bar{d}_{\Gamma_2} \leqq \ldots \leqq \bar{d}_{\Gamma_K}.$$

On dira que Γ *est sans inversion.*

Démonstration : Notons $\Sigma_1 < \ldots < \Sigma_L$ la PU associée à $\hat{d}_\Gamma$.

Les classes de Σ étant des réunions de classes consécutives de Γ, il existe une partition de $\{1, \ldots, K\}$ en L classes contiguës $K_1, \ldots, K_L$ telles que :

$$\forall l : \quad \Sigma_l = \bigcup_{k \in K_l} \Gamma_k.$$

D'autre part,

$$\hat{d}_\Gamma(xy) = \bar{d}_{\Sigma_l} \qquad \text{si} \quad xy \in \Sigma_l \qquad \text{et} \qquad \bar{d}_{\Sigma_1} < \ldots < \bar{d}_{\Sigma_L}.$$

[*cf.* problème (7)].

Nous allons construire une PUB $\Theta_1 < \ldots < \Theta_M$ et une partition de $\{1, \ldots, M\}$ en L classes contigues $M_1, \ldots, M_L$ telles que

$$\left.\begin{array}{r} \forall l : \quad \Sigma_l = \bigcup_{m \in M_l} \Theta_m \\ \forall l : \quad (\bar{d}_{\Theta_m})_{m \in M_l} \text{ non décroissante,} \end{array}\right\} \qquad (10)$$

cette PUB nous permettra de montrer que $\bar{d}_{\Gamma_k} = \bar{d}_{\Sigma_l}$ pour tout l et tout $k \in K_l$.

Cette construction est réalisée par récurrence. Supposons déjà construite une SCPUB $\Theta_1 < \ldots < \Theta_N$ et une partition de $\{1, \ldots, N\}$ en $L_1 < L$ classes contiguë $M_1, \ldots, M_{L_1}$ telles que

$$\forall l \leqq L_1 : \quad \Sigma_l = \bigcup_{m \in M_l} \Theta_m,$$

$$\forall l \leqq L_1 : \quad (\bar{d}_{\Theta_m})_{m \in M_l} \text{ non décroissante.}$$

Soient Π la partition de X associée à cette SCPUB (proposition 1) (ou, ce qui revient au même, à la SCPU $\Sigma_1 < \ldots < \Sigma_{L_1}$) et $\approx$ la relation d'équivalence associée à la classe Σ_{L_1+1} (proposition 2). Sur Π définissons la dissimilarité :

$$D^*_{CD} = D_{CD} \quad \text{si } C \approx D, \qquad = \infty \quad \text{sinon } (^4)$$

(D_{CD} est défini à la fin du paragraphe 3). Avec cette dissimilarité, l'algorithme de la moyenne construit sur Π une PUB

$$\Phi_1 < \ldots < \Phi_P < \Phi_{P+1} < \ldots < \Phi_Q$$

sans inversion; avec des notations évidentes on a :

$$\bar{D}^*_{\Phi_1} < \ldots < \bar{D}^*_{\Phi_P} < \bar{D}^*_{\Phi_{P+1}} = \ldots = \bar{D}^*_{\Phi_Q} = \infty.$$

([4]) En fait, une très grande valeur positive.

Prolongeons alors la SCUB $\Theta_1 < \ldots < \Theta_N$ en posant

$$\Theta_{N+p} = \bigcup_{CD \in \Phi_p} \langle C, D \rangle \qquad \text{pour} \quad p=1, \ldots, P$$

et définissons

$$M_{L_1+1} = \{ N+1, \ldots, N+P \}$$

(Θ est la PUB compatible avec Σ la plus conforme avec l'algorithme de la moyenne).

Grâce à la proposition 2, on prouve aisément que

$$\Sigma_{L_1+1} = \bigcup_{m \in M_{L_1+1}} \Theta_m.$$

La suite $(\bar{d}_{\Theta_m})_{m \in M_{L_1+1}}$ est non décroissante puisque $\bar{d}_{\Theta_{N+p}} = \bar{D}^*_{\Phi_p}$. Enfin, il est facile de vérifier que

$$\Theta_1 < \ldots < \Theta_N < \Theta_{N+1} < \ldots < \Theta_{N+P}$$

est encore une SCPUB. Ceci garantit bien l'existence d'une PUB $\Theta_1 < \ldots < \Theta_M$ possédant les propriétés (10).

En fait, une telle PUB est sans inversion. En effet, comme $(\Sigma, \hat{d}_\Gamma)$ est solution du problème (5) et $\hat{d}_\Gamma \in \bar{\Delta}_\Theta$, $(\Theta, \hat{d}_\Gamma)$ est solution du problème (6). On en déduit que $\hat{d}_\Theta = \hat{d}_\Gamma$. Par conséquent $\hat{d}_\Theta$ croît strictement d'une classe Θ_m à une classe Θ_{m+1} lorsque $m \in M_l$ et $m+1 \in M_{l+1}$ car on passe de Σ_l à Σ_{l+1}, et $\bar{d}_{\Sigma_l} < \bar{d}_{\Sigma_{l+1}}$. On en déduit que nécessairement : $\bar{d}_{\Theta_m} < \bar{d}_{\Theta_{m+1}}$ (propriété de la régression isotone). Grâce à la deuxième propriété de (10), il en résulte que $\bar{d}_{\Theta_1} \leqq \ldots \leqq \bar{d}_{\Theta_M}$.

On peut dire alors que

$$\hat{d}_\Theta (x, y) = \bar{d}_{\Theta_m} \qquad \text{si} \quad xy \in \Theta_m,$$

car il s'agit bien de la meilleure approximation dans le cadre d'une régression isotone, et comme $\hat{d}_\Theta = \hat{d}_\Gamma$, on peut en conclure que, sur chaque classe M_l, la suite $(\bar{d}_{\Theta_m})_{m \in M_l}$ est constante. Compte tenu du mode de construction de Θ utilisant l'algorithme de la moyenne, ceci n'est possible que si, pour tout entier $L_1 < L$, la dissimilarité moyenne D_{CD} est la même pour toute paire CD de classes de la partition Π induite par la relation d'équivalence associée à la SCPU $\Sigma_1 < \ldots < \Sigma_{L_1}$ et vérifiant $C \approx D$ (cette valeur commune étant d'ailleurs $\bar{d}_{\Sigma_{L_1+1}}$). Pour conclure, il suffit de remarquer que si $k \in K_{L_1+1}$ la classe Γ_k est une réunion d'ensembles $\langle C, D \rangle$, C et D étant deux classes distinctes de Π telles que $C \approx D$. En effet, observons que si $xy \in \Gamma_k$, comme $xy \in \Sigma_{L_1+1}$, il existe deux classes distinctes de Π : C et D telles que $C \approx D$ (proposition 2). Dès lors, $\langle C, D \rangle \subset \Gamma_k$ car, si $u \in C$ et $v \in D$, ux, xy, et yv sont toutes les trois dans $\Gamma_1 \cup \ldots \cup \Gamma_k$ et il en est de même pour uv d'après la propriété (4). En fait $uv \in \Gamma_k$ sinon appliquant cette même propriété à xu, uv et vy, on en déduirait que $xy \in \Gamma_k$. Conséquence,

$$\bar{d}_{\Gamma_k} = D_{CD} = \bar{d}_{\Sigma_{L_1+1}}$$

pour tout $k \in K_{L_1+1}$ et la suite des $\bar{d}_{\Gamma_k}$ est donc bien non décroissante.

COROLLAIRE 1 : *Si* $\Gamma_1 < \ldots < \Gamma_K$ *est une solution du problème* (9) :

$$\forall k : \quad \bar{d}_{\Gamma_k} \leqq \bar{d}_{\Gamma_{k+1}} \leqq \bar{d}_{\Gamma_{k+1} \cup \ldots \cup \Gamma_K}.$$

Démonstration : La première inégalité résulte immédiatement du théorème 1. La seconde en résulte aussi car

$$\bar{d}_{\Gamma_k \cup \ldots \cup \Gamma_K} = \frac{\operatorname{card}(\Gamma_k)\, \bar{d}_{\Gamma_k} + \ldots + \operatorname{card}(\Gamma_K)\, \bar{d}_{\Gamma_K}}{\operatorname{card}(\Gamma_k) + \ldots + \operatorname{card}(\Gamma_K)} \geqq \bar{d}_{\Gamma_{K_k}}.$$

COROLLAIRE 2 : *Le problème* (9) *équivaut au problème* (11)

$$\min_{\substack{\Gamma_1 < \ldots < \Gamma_K \\ PUB \text{ sans inversion}}} \sum_{k=1}^{K} I(\Gamma_k).$$

(*critère de l'inertie intraclasse*).

Démonstration :

$$\| d - \hat{d}_\Gamma \|^2 = \sum_{k=1}^{K} \sum_{xy \in \Gamma_k} (d_{xy} - \bar{d}_{\Gamma_k})^2 = \text{inertie intraclasse}$$

puisque le problème de régression isotone est trivial.

Le caractère récurrent de la construction ascendante d'une PUB suggère un algorithme de branchement dont le corollaire 1 va limiter considérablement le nombre de branches à explorer.

5. ALGORITHME DE BRANCHEMENT RÉSOLVANT LE PROBLÈME (11)

L'algorithme du type BRANCH and BOUND décrit ci-dessous construit une arborescence où chaque chemin issu de la racine correspond à une SCPUB. Tout sommet S est caractérisé par une matrice symétrique. Les lignes et les colonnes de cette matrice correspondent aux classes de la partition Π associée à la SCPUB $\Gamma_1 < \ldots < \Gamma_k$ définie par le chemin finissant en S (proposition 1). Les valeurs de cette matrice sont les distances moyennes D_{CD} entre classes de Π. On partitionne un tel sommet en réunissant deux classes C, D de Π ce qui revient à prolonger la SCPUB par $\Gamma_{k+1} = \langle C, D \rangle$. Comme $\bar{d}_{\Gamma_{k+1}} = D_{CD}$, il est inutile d'envisager les branches pour lesquelles D_{CD} n'est pas dans l'intervalle $[\bar{d}_{\Gamma_k}, \bar{d}_{\mathbb{P}-\Gamma_1 \cup \ldots \cup \Gamma_k}]$ (corollaire 1).

Par ailleurs, on peut aussi négliger l'exploration des sommets pour lesquels l'inertie intraclasse partielle $\sum_{j=1}^{k} I(\Gamma_j)$ dépasse la meilleure valeur de l'inertie intraclasse déjà obtenue. Les passages d'une matrice à la suivante s'effectuent rapidement en utilisant les formules récurrentes de l'algorithme de la moyenne.

Signalons enfin que l'exploration de l'arborescence peut être réalisée en « préordre » ce qui nécessite seulement le stockage des informations relatives à un chemin de longueur maximale au plus.

Exemple : (cf. *fig.*) : 19 sommets de l'arborescence sont créés au lieu des 251 possibles.

REMARQUES : 1° Cet algorithme généralise l'algorithme de la moyenne. A chaque itération, au lieu de se limiter au regroupement des classes C et D de Π qui minimisent D_{CD}, on envisage tous les regroupements pour lesquels D_{CD} tombe dans l'intervalle d'admissibilité. Cependant la similitude des calculs peut expliquer le bon comportement de l'algorithme de la moyenne.

2° Sur les exemples, il apparaît clairement que le critère minimisé – l'inertie intraclasse – dépend essentiellement des dernières classes formées. Cette remarque motive l'heuristique du paragraphe 6 et justifie son remarquable comportement.

3° Évidemment, cet algorithme de branchement devient vite impraticable dès que le nombre d'objets augmente. Néanmoins, on peut toujours l'appliquer dans ce sens, à partir d'une SCPUB, sur la partition Π qu'elle induit lorsque card (Π) est faible (< 15). Une bonne SCPUB peut être obtenue par exemple en utilisant l'algorithme de la moyenne ou en utilisant l'algorithme du paragraphe 6.

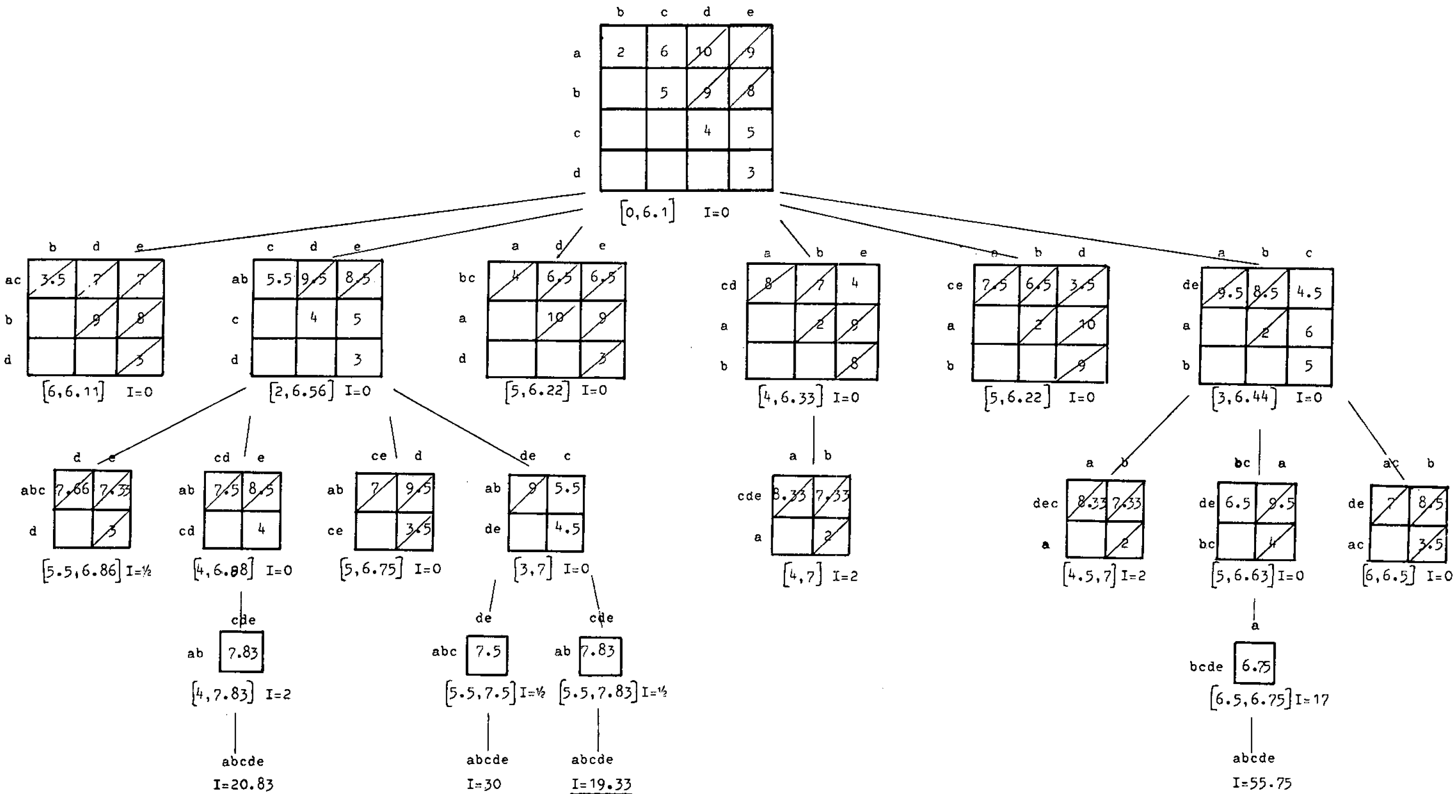
6. ALGORITHME DIVISIF D'APPROXIMATION D'UNE SOLUTION DU PROBLÈME (11)

Cet algorithme construit une PUB $\Gamma_1 < \ldots < \Gamma_K$, en commençant par Γ_K et en terminant par Γ_1. La construction de chaque classe Γ_k est obtenue en partitionnant un sous-ensemble X_k de X en deux classes X_k^1 et X_k^2, puis en posant

$$\Gamma_k = \langle X_k^1, X_k^2 \rangle.$$

Ces deux classes sont construites progressivement en minimisant dans chaque étape, l'accroissement de l'inertie intraclasse résultant du classement d'un nouvel élément de X_k.

Exemple d'utilisation de l'algorithme hiérarchique ascendant de branchement pour déterminer la préordonnance ultramétrique optimale au sens du critère des moindres carrés.



OPTIMUM: ab < de < cd=ce < ac=ad=ae=bc=bd=be (algorithme de la moyenne)
2 3 4.5 7.83

Sous chaque matrice, on indique l'intervalle $[a, b]$ signalé au paragraphe 5. Les cases barrées correspondent à des classes C et D telles que D_{CD} n'appartienne pas à $[a, b]$. I est l'inertie intraclasse de la SCPUB associée.

Dans ce qui suit, lorsque C et D sont deux parties disjointes de X :

$$I(\langle C, D\rangle)=\sum_{xy\in\langle C,D\rangle}(d_{xy}-\overline{d}_{\langle C, D\rangle})^2.$$

Initialisation :

1. $k\leftarrow n-1$ [$n=\operatorname{card}(X)$];
2. $X_k\leftarrow X$.

Itération k :

1. $xy\leftarrow$ une paire de X_k, de dissimilarité maximale;
2. $X_k^1\leftarrow\{x\}$ et $X_k^2\leftarrow\{y\}$;
3. pour tout objet $z\in X_k-X_k^1\cup X_k^2$:

$$\left\{\begin{array}{l} 3.1.\ I_z^1\leftarrow I(\langle X_k^1\cup\{z\}, X_k^2\rangle),\\ 3.2.\ I_z^2\leftarrow I(\langle X_k^2\cup\{z\}, X_k^1\rangle),\\ 3.3.\ I_z\leftarrow\min(I_z^1, I_z^2);\end{array}\right.$$

4. z^* étant un objet minimisant I_z :
si $I_{z^*}=I_{z^*}^1$, faire $X_k^1\leftarrow X_k^1\cup\{z^*\}$,
sinon faire $X_k^2\leftarrow X_k^2\cup\{z^*\}$;
5. si $X_k\neq X_k^1\cup X_k^2$, aller en 3,
sinon faire $\Gamma_k\leftarrow\langle X_k^1, X_k^2\rangle$.

Test d'arrêt et incrémentation :

1. si $k=1$, stop;
2. sinon faire :

$$\left\{\begin{array}{l} 2.1.\ k\leftarrow k-1,\\ 2.2.\ X_k\leftarrow X_l^i \text{ non encore partitionné, et contenant une paire de dissimi-}\\ \text{larité maximale } (i=1, 2 \text{ et } l<k),\\ 2.3.\ \text{aller à } \textit{itération } k.\end{array}\right.$$

Grâce à la proposition 3, $\Gamma_1<\ldots<\Gamma_{n-1}$ est une PUB.

Exemple : Les données sont celles de la figure.

$k=4$ $X_4=\{a, b, c, d, e\}$,
$X_4^1=\{a\}$ et $X_4^2=\{d\}$.

$X_4^1=\{a, b\}$ et $X_4^2=\{d\}$,

$X_4^1=\{a, b\}$ et $X_4^2=\{d, e\}$,
$X_4^1=\{a, b\}$ et $X_4^2=\{c, d, e\}$.

$k=3$ $X_3=\{c, d, e\}$,
$X_3^1=\{c\}$ et $X_3^2=\{e\}$,
$X_3^1=\{c\}$ et $X_3^2=\{d, e\}$.

	z	I_z^1	I_z^2	
	b	0,5	32	z^*
	c	18	8	
←	e	24,5	0,5	
	c	20,67	14,75	
←	e	28,67	2	z^*
←	c	29,5	18,83	z^*
←	d	2	0,5	z^*

$$
\begin{array}{ll}
k=2 & X_2=\{d, e\}, \\
 & \underline{X_2^1=\{d\} \quad \text{et} \quad X_2^2=\{e\}.} \\
k=1 & X_1=\{a, b\}, \\
 & \underline{X_1^1=\{a\} \quad \text{et} \quad X_1^2=\{b\}.}
\end{array}
$$

On détermine ainsi la PUB : $ab < de < cd=ce < ac=ad=ae=bc=bd=be$ précédemment obtenue et optimale.

REMARQUES : 1. Cet algorithme ne conduit pas toujours à une PUB solution du problème (11). En particulier, la PUB qu'il détermine peut présenter des inversions.

2. Néanmoins, comme il limite au mieux l'accroissement de l'inertie intraclasse dans la formation progressive des classes, et comme il construit d'abord les classes finales qui contribuent le plus à cette inertie, son efficacité est grande. Ainsi, sur des exemples classiques, le critère des moindres carrés a toujours été amélioré. Pour plus de détails concernant ces exemples, on pourra consulter le rapport de J. L. Chandon (1978) de l'I.A.E. d'Aix-en-Provence (n° W.P. 123).

Programmes : des programmes FORTRAN des deux algorithmes peuvent être obtenus auprès de J. L. Chandon.

BIBLIOGRAPHIE

1. M. BARBUT, *Partitions d'un ensemble fini : leur treillis (cosimplex) et leur représentation géométrique*, Mathématiques et Sciences humaines, n° 22, 1968, p. 5-22.
2. J. P. BENZÉCRI, *Problèmes et méthodes de la taxinomie*, Publications de l'I.S.U.P., Université de Paris-VI, 1965.
3. J. P. BENZÉCRI *et al.*, *La taxinomie*, tome 1 de : *l'analyse des données, Dunod, Paris*, 1973.
4. A. J. BOYCE, *Mapping Diversity: A Comparative Study of Some Numerical Methods*, in *Numerical Taxonomy*, COLE, éd., Academic Press, 1969.
5. J. D. CARROLL et J. J. CHANG, *A Method for Fitting a Class of Hierarchical tree Structure Models to Dissimilarities Data and Its Application to Some « Body Parts » Data of Miller's, Proceedings*, 81st Annual Convention A.P.A., 1973.
6. J. D. CARROLL et S. PRUZANSKY, *Fitting of Hierarchical Tree Structure (HTS) Models, Mixtures of HTS, and Hybrid Models, via Mathematical Programming and Alternating Least Squares, in: Theory, Methods and Applications of Multidimensional Scaling and Related Techniques*, U.S. Japan Seminar, University of California, 1975.
7. M. DAVY DE VIRVILLE, *L'Analyse des proximités*, Thèse de 3e cycle, Université de Paris-VI, 1972.
8. D. DEFAYS, *Recherche des ultramétriques à distance minimum d'une similarité donnée*, Bulletin de la Société royale des Sciences de Liège, vol. 44, nos 5-6, 1975, p. 330-343.

9. J. S. Farris, *On the Cophenetic Correlation Coefficient*, Systematic Zoology, vol. 18, 1969, p. 279-285.
10. D. Florens et E. O. Lochard, *Taxonomie automatique : la sous-dominante dans son cadre général et « la plus proche du sens des moindres carrés »*, Cahier du Bureau Universitaire de Recherche opérationnelle, n° 25, 1976, p. 39-60.
11. J. A. Hartigan, *Representation of Similarity Matrices by Trees*, J. Amer. Stat. Assoc., vol. 62, 1967, p. 1140-1158.
12. M. Jambu, *Techniques de classification automatique appliquées à des données de Sciences humaines*, Thèse de Doctorat de 3e cycle. Paris, 1972.
13. S. C. Johnson, *Hierarchical Clustering Schemes*, Psychometrica, vol. 32, 1967, p. 241-254.
14. J. B. Kruskal, *Multidimensional Scaling by Optimizing Goodness of fit to a Non Metric Hypothesis*, Psychometrica, vol. 29, 1964, p. 1-7.
15. G. N. Lance et W. T. Williams, *A General Theory of Classification Sorting Strategies: 1. Hierarchical Systems; 2, Clustering Systems*, Computer Journal, vol. 9-10, 1967, p. 373-380.
16. I. C. Lerman, *Les bases de la classification automatique*, Gauthier-Villars, Paris, 1970.
17. R. R. Sokal et F. J. Rholf, *The Comparison of Dendograms by Objective Methods*, Taxon, vol. 11, 1962, p. 33-40.
18. R. R. Sokal et P. H. A. Sneath, *Principles of Numerical Taxonomy*, Freeman, 1963.