

ISRAËL-CÉSAR LERMAN

**Convergence optimale de l'algorithme
de « réallocation-recentrage » dans le cas
continu le plus simple**

RAIRO. Recherche opérationnelle, tome 20, n° 1 (1986), p. 19-50

http://www.numdam.org/item?id=RO_1986__20_1_19_0

© AFCET, 1986, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

CONVERGENCE OPTIMALE DE L'ALGORITHME DE « RÉALLOCATION-RECENTRAGE » DANS LE CAS CONTINU LE PLUS SIMPLE (*)

par Israël-César LERMAN ⁽¹⁾

Résumé. — *Après avoir rappelé les différents types de problèmes de convergence liés à cet algorithme, encore appelé des moyennes mobiles » ou de sa généralisation : « algorithme des nuées dynamiques », nous précisons la position originale de notre étude. Il s'agit de dégager les conditions de convergence vers la partition globalement optimale — du point de vue du critère de l'inertie expliquée — dans la situation topologique la plus simple de classifiabilité : une suite de deux intervalles finis uniformément chargés d'un même axe. Ces conditions — auxquelles la méthode des pôles d'attraction apporte sa contribution — ont un caractère très spécifique.*

Mots clés : Classification non hiérarchique; optimum global.

Abstract. — *They are different types of convergence problems concerning the "k-means algorithm" or its generalization, the "dynamic cluster algorithm". After recalling them, we show the specificity of our study where we consider a priori the easiest topological situation of classifiability in a continuous case. Then, we determine general conditions on the initial state of the algorithm, for which the issue is necessarily the global optimal partition. We show that these conditions are satisfied by the "poles of attraction" clustering method.*

Keywords: Non-hierarchical classification; global optimum.

PLAN

1. *Position du problème*
 1. 1. Rappel de l'Algorithme.
 1. 2. Description générale dans le cas continu de l'algorithme.
 1. 3. Apport de notre analyse.
 1. 4. Un autre problème de convergence.
 1. 5. Idée du travail proposé.
 2. *Théorème de convergence optimale*
 2. 1. Considérations et notations préliminaires.
 2. 2. Résultat principal.
 3. *Étude de contre-exemples.*
 3. 1. Étude du premier contre-exemple.
 3. 2. Étude du second contre-exemple.
 4. *Contribution de la méthode des pôles d'attraction.*
 4. 1. Détermination du premier pôle.
 4. 2. Détermination du deuxième pôle.
 5. *Conclusion*
- Bibliographie.*

(*) Reçu en novembre 1984.

(¹) I.R.I.S.A., Campus Universitaire de Beaulieu, avenue du Général-Leclerc, 35042 Rennes Cedex, France.

1. POSITION DU PROBLÈME

1.1. Rappel de l'Algorithme

Il s'agit de l'un des algorithmes les plus classiques de la classification automatique non hiérarchique. Il est connu sous le nom de l'algorithme des « k -means » dans la littérature anglophone où il a été introduit et auquel on a donné de nombreuses variantes [Thorndike (1953) D. J. Hall, G. H. Ball (1965) McQueen (1967)]. Nous avons préféré — de façon plus suggestive — l'appeler : algorithme de « réallocation-recentrage »; en effet, ce dernier titre reflète les deux articulations fondamentales de cet algorithme dont nous allons rappeler la version la plus classique qui est celle qui nous intéressera ici.

On considère la description d'un ensemble fini E d'objets au moyen d'un ensemble fini de variables et on admet pour cette description, une représentation de E au moyen d'un nuage de points $\mathcal{N}(E)$ dans un espace euclidien qu'on peut — pour fixer les idées — supposer être l'espace géométrique $\mathbb{R}^p$ muni de sa métrique euclidienne ordinaire. En identifiant E à sa représentation dans $\mathbb{R}^p$, nous noterons :

$$\mathcal{N}(E) = \{ (x, \mu_x) / x \in E \},$$

où le « poids » μ_x (nombre réel positif) peut définir l'« importance » *a priori* de l'objet x pour la classification à établir sur E . Dans la plupart des applications et même des études théoriques, on peut — si ce n'est pas déjà le cas naturellement se ramener à la situation où $\mu_x = 1$ pour tout x appartenant à E (fini).

L'entier k étant fixé, le but de cet algorithme est la construction d'une partition de E en k classes au plus, optimisant au mieux un critère de cohésion des classes formées. Le nombre de classes obtenues résulte en fait de l'aléa du déroulement de l'algorithme, il reste dans la pratique très près de k .

Le critère que nous choisissons de considérer compte tenu de la représentation de E et de la nature même de l'algorithme, est celui de l'inertie. Plus précisément, il s'agit de minimiser, pour $h \leq k$,

$$\sum_{1 \leq j \leq h} \sum_{x \in E_j} \mu_x \|x - g_j\|^2, \quad (1)$$

où $P = \{ E_j / 1 \leq j \leq h \}$ est la partition en h classes et où $\{ g_j / 1 \leq j \leq h \}$ est le système des centres de gravité des classes E_j (g_j : centre de gravité de $E_j / 1 \leq j \leq h$).

Pour le démarrage de l'algorithme, nous supposons la donnée *a priori* d'un ensemble quelconque de k points $\{ o_1^{(0)}, o_2^{(0)}, \dots, o_k^{(0)} \}$ qui définit le système

initial de « centres d'attraction ». Le premier pas de l'algorithme consiste à définir un « zonage » de E en attachant chacun des sommets du nuage au centre d'attraction le plus proche. En cas d'ex aequo dans la proximité, l'usage est d'attribuer le sommet au centre d'indice le plus petit. On définit ainsi une partition de E en $k(1)$ classes :

$$\{E_1^{(1)}, E_2^{(1)}, \dots, E_{k(1)}^{(1)}\} \quad \text{où } k(1) \leq k;$$

en effet, certains centres, ne correspondant pas à des sommets du nuage, peuvent demeurer seuls.

De façon plus précise, l'hyperplan médiateur d'un même couple de sommets $(o_j^{(0)}, o_h^{(0)})$ tel que $j < h$, découpe le nuage $\mathcal{N}(E)$ en deux morceaux :

$$\mathcal{E}_{jh} = \{x/x \in E, d(x, o_j^{(0)}) \leq d(x, o_h^{(0)})\}$$

et :

$$\mathcal{E}_{hj} = \{x/x \in E, d(x, o_j^{(0)}) > d(x, o_h^{(0)})\}, \quad (2)$$

de sorte que la « zone » attribuée à $o_j^{(0)}$, se trouve définie par l'intersection :

$$\mathcal{Z}(o_j^{(0)}) = \left(\bigcap_{h>j} \mathcal{E}_{jh} \right) \cap \left(\bigcap_{l<j} \mathcal{E}_{jl} \right). \quad (3)$$

$\mathcal{Z}(o_j^{(0)})$ — s'il n'est pas vide — définit l'une des classes $E_j^{(1)}$, $1 \leq j \leq k(1)$.

A chacune des classes $E_j^{(1)}$, on associe son *centre de gravité* formant ainsi une nouvelle suite de centres d'attraction : $(o_1^{(1)}, o_2^{(1)}, \dots, o_{k(1)}^{(1)})$. L'ensemble des centres ainsi obtenus est de cardinal inférieur ou égal à $k(1)$; en effet, il peut se faire qu'un même centre de gravité corresponde à plus d'une seule classe.

D'où une nouvelle décomposition de E en classes dont les centres de gravité définiront les nouveaux centres d'attraction et ainsi de suite... Nous allons voir que dans le cadre défini ci-dessus, le processus — dont chaque pas diminue le critère (1) — converge nécessairement et *on espère* que l'ensemble des derniers centres viendra occuper les régions à forte « densité » du nuage $\mathcal{N}(E)$. Un des résultats de notre analyse consiste précisément à montrer que cette espérance n'est pas toujours fondée.

Sous l'appellation « Algorithme des nuées dynamiques (A.N.D.) », E. Diday [Diday (1972), (1980)] a introduit en France ce type d'algorithme en lui apportant une certaine formalisation et en cherchant à généraliser au maximum sa portée. La généralisation porte sur la notion de représentation d'une classe qui ne se fait plus nécessairement au moyen d'un centre de gravité, mais à partir de la définition d'un « noyau » de la classe; lequel peut par exemple — dans la situation décrite ci-dessus — correspondre à un sous-

ensemble de faible cardinal de la classe. Ce dernier, définissant un « squelette » de la classe, « épouserait mieux sa forme » que ne le ferait un centre de gravité. Le deuxième aspect de l'extension de l'algorithme porte sur la nature de l'espace de représentation des données et d'ailleurs, la notion de « noyau » dépend de cet espace et du problème de reconnaissance posé. Le troisième aspect de l'extension concerne la nature de la structure de type classificatoire visée dans l'algorithme (e. g. une hiérarchie de partie [Ralambondrainy et Diday (1984)]). Enfin, un problème important de métriques et critères se pose, pour alimenter de façon cohérente ces développements de l'algorithme.

Du point de vue de l'analyse des données réelles, un aspect inconfortable de cette méthode concerne d'une part, le choix *a priori* et donc non sans arbitraire de la structure du système initial de noyaux et d'autre part, le caractère aléatoire de ce choix. Dans la situation d'un nuage de points dans un espace euclidien et pour une représentation d'une même classe par un de ses éléments — version souvent utilisée dans l'AND. — nous proposons à travers la méthode des pôles d'attraction » [Lerman, Leredde (1977), Leredde (1979), Lerman (1981), chap. 8], un choix objectif à partir d'indices statistiques, d'un système initial de noyaux.

La construction de ces indices — qui suppose établie la table des proximités ou distances entre éléments de l'ensemble E à classifier — est basée sur l'étude de la distribution des proximités des différents sommets de E à un même point x de E , lorsque x varie dans E . Pour un nombre de classes fixé, ces indices déterminent des pôles — éléments de E — d'entraînement des classes.

Il s'est avéré expérimentalement que — pour un choix cohérent des coefficients de détermination des pôles et du critère d'affectation des sommets de $\mathcal{N}(E)$ aux pôles — on obtenait une quasi-stabilité de la partition formée autour des centres de gravité des classes entraînées par les pôles. Ce résultat expérimental se trouvera conforté par celui théorique que nous obtiendrons dans le « cas le plus pur » où la détermination d'un système de noyaux formé des pôles d'attraction conduira à la solution optimale.

Pour la description de l'algorithme de « reallocation-recentrage », on introduit deux fonctions. La première que nous notons π , pour rappeler que son objet est de former une partition, est de « reallocation » de l'ensemble des sommets du nuage $\mathcal{N}(E)$ autour des centres d'attraction formant le système des noyaux. Dans notre cadre, $\mathcal{L}_k$ désignant l'ensemble des parties de E de cardinal inférieur ou égal à k et $\mathcal{P}_k$ l'ensemble des partitions de E en au plus k classes, π est une application de $\mathcal{L}_k$ dans $\mathcal{P}_k$ dont le sens a été précisé ci-dessus [cf. (3)].

La deuxième fonction que nous notons v , pour rappeler que son objet est de former un système de noyaux, est de « recentrage » des classes. Il s'agit d'une application de $\mathcal{P}_k$ dans $\mathcal{L}_k$, associant à chaque classe son centre d'attraction.

Si (L, P) est un élément de $\mathcal{L}_k \times \mathcal{P}_k$, on désigne par $\omega(L, P)$ la valeur du critère (1) pour (L, P) ; plus précisément,

$$\omega(L, P) = \sum_{1 \leq j \leq h} \min \{ \mathcal{M}_2(E_j, o) / o \in L \}, \quad (4)$$

où nous avons noté $P = \{ E_j / 1 \leq j \leq h \}$ ($h \leq k$) et où :

$$\mathcal{M}_2(E_j, o) = \sum_{x \in E_j} \mu_x \|x - o\|^2 \quad (5)$$

est le moment d'inertie de E_j par rapport à o .

Considérons la partition $\pi(L)$ que nous pouvons noter $\pi(L) = \{ G(o) / o \in L \}$ où $G(o)$ est la classe attribuée au centre d'attraction o . On a :

$$\omega[L, \pi(L)] \leq \omega(L, P). \quad (6)$$

En effet, si x — sommet quelconque de E — est un point de $E_j \cap G(o)$ et si nous désignons par o_j le point de L qui rend minimum (5), la contribution de x à $\omega(L, P)$ [cf. (4)] est $\mu_x d^2(x, o_j)$. Celle à $\omega[L, \pi(L)]$ est $\mu_x d^2(x, o)$. Or, de par l'affectation de x dans la définition de $\pi(L)$, on a :

$$\mu_x d^2(x, o) \leq \mu_x d^2(x, o_j), \quad (7)$$

d'où le résultat.

D'autre part, de par la propriété caractéristique d'un centre de gravité, on a de façon immédiate :

$$\omega[v(P), P] \leq \omega(L, P). \quad (8)$$

Ainsi, chacune des deux articulations d'un même pas de l'algorithme — réallocation définie par π et recentrage défini par v — fait diminuer la quantité critère qui est l'inertie perdue par la classification.

Bien qu'il s'agisse de déterminer une « bonne » partition de E , minimisant au mieux l'inertie perdue, il faut regarder l'algorithme de « réallocation-recentrage » comme procédant à la recherche d'un « bon » système de noyaux-centres d'attraction. $L^{(0)}$ désignant le système initial de noyaux, lequel pouvant être choisi de façon quelconque, on a :

$$\left. \begin{aligned} L^{(1)} &= v \circ \pi[L^{(0)}], \\ L^{(2)} &= v \circ \pi[L^{(1)}], \dots, L^{(r)} = v \circ \pi[L^{(r-1)}], \dots \end{aligned} \right\} \quad (9)$$

Chaque étape faisant diminuer la quantité critère ω qui reste positive, le processus converge nécessairement et s'arrête après le t -ième pas si :

$$v \circ \pi [L^{(t)}] = L^{(t)}, \quad (10)$$

système de noyaux qu'il n'est plus possible d'améliorer, point fixe de l'application $v \circ \pi$. Le couple $[L^{(t)}, \pi(L^{(t)})]$ est le résultat de l'algorithme et on peut écrire :

$$L^{(t)} = (v \circ \pi)^t (L^{(0)}), \quad (11)$$

t étant défini comme le plus petit entier pour lequel :

$$(v \circ \pi)^{t+1} (L^{(0)}) = (v \circ \pi)^t (L^{(0)}). \quad (12)$$

Dans la pratique et en tenant compte de la précision calcul de l'ordinateur, on arrête le processus dès que la différence entre le premier et le second membre de (12) devient inférieure à un certain seuil.

Le résultat précédent sur la décroissance du critère ω [cf. (6) et (8) cidessus] peut directement être étendu dans le cas continu où μ définit une mesure positive finie et différentiable sur $\mathbb{R}^p$. Dans ce contexte, une quantité telle que (5) s'écrit :

$$\mathcal{M}_2(E_j, o) = \int_{E_j} d^2(x, o) d\mu(x). \quad (13)$$

Toutefois, dans ce cadre continu, on n'est plus certain de la finitude de t .

1.2. Description générale dans le cas continu de l'algorithme

Nous voulons ici mentionner la nature des résultats obtenus dans [Diebolt, Simon et Miranker (1977)]. Ces auteurs proposent une description dans un cadre topologique général de l'algorithme dans sa version « nuées dynamiques » où une classe se trouve représentée par un sous-ensemble de points. L'espace de représentation des unités de données est assimilé à un espace métrique compact (X, d) , sur lequel se trouve introduite une mesure de probabilité μ que supporte une σ -algèbre borélienne $\mathcal{S}$ dans X . On se donne par ailleurs une fonction de dissemblance δ sur X , qui est supposée une application continue de $X \times X$ dans $\mathbb{R}^+$. On peut pratiquement supposer que $\delta = \varphi \circ d$ où $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est une fonction continue et croissante, avec

$\varphi(0)=0$. On suppose de plus que :

$$\text{diam}_\delta X = \sup \{ \delta(x, y) / (x, y) \in X \times X \} = 1.$$

De la sorte l'élément $d^2(x, o) d\mu(x)$ composant de (13) devient :

$$\hat{A}(x) = \int_A \delta(x, y) d\mu(y), \quad (14)$$

où le centre d'attraction o devient le noyau $A \in \mathcal{S}$ et où d^2 devient δ .

En fait, on travaille au niveau de l'ensemble S des classes d'équivalence de $\mathcal{S}$, où la relation d'équivalence est définie par :

$$A \sim B \Leftrightarrow \mu(A \triangle B) = 0.$$

On travaille — sans perte de généralité — dans le cas de la recherche d'une partition en deux classes, mais on suppose satisfaite toute hypothèse évitant l'ambiguïté dans le résultat de chacune des deux articulations fondamentales de l'algorithme : recentrage et reallocation.

C'est ainsi qu'on suppose que l'application de S dans l'ensemble $\mathcal{C}(X)$ des fonctions continues sur $X : A \rightarrow \hat{A}$, est *injective*.

La métrique qu'on introduit naturellement sur S est celle Δ de la différence symétrique (pour A et B appartenant à S , $A \triangle B = \mu[(A - B) \cup (B - A)]$) et (S, Δ) est un espace métrique complet.

Δ est étendue à S^2 au moyen de l'expression :
pour :

$$\mathbf{A} = (A_1, A_2) \quad \text{et} \quad \mathbf{B} = (B_1, B_2)$$

appartenant à S^2 ,

$$\mathbf{A} \triangle \mathbf{B} = \max[\Delta(A_1, B_1), \Delta(A_2, B_2)]. \quad (15)$$

Les métriques qui peuvent correspondre de façon cohérente à Δ , sur l'espace des fonctions $\hat{S} = \{ \hat{A} / A \in S \}$, sont D et D_2 , où pour $\hat{A}$ et $\hat{B}$ appartenant à $\hat{S}$,

$$D(\hat{A}, \hat{B}) = \| \hat{A} - \hat{B} \| \quad \text{et} \quad D_2(\hat{A}, \hat{B}) = \| \hat{A} - \hat{B} \|_2, \quad (16)$$

où $\| \cdot \|$ désigne la norme de la convergence uniforme dans $\mathcal{C}(X)$ et où $\| \cdot \|_2$ désigne la norme $L^2(X, \mu)$.

Dans ces conditions, sans être compacts, $(\hat{S}, D)$ et $(\hat{S}, D_2)$ sont respectivement, relativement compacts dans :

$$[\mathcal{C}(X), \| \cdot \|] \quad \text{et} \quad [L^2(X, \mu), \| \cdot \|_2].$$

$\mathcal{P}$ désignant l'ensemble des partitions en deux classes de X et $\hat{K}$, l'ensemble $\hat{S}^2$ auquel on ôte la diagonale, on reprend le critère d'adéquation ω [cf. (4)] entre un élément $\mathbf{P}=(X_1, X_2)$ de $\mathcal{P}$ et un élément $\hat{\mathbf{A}}=(\hat{A}_1, \hat{A}_2)$ de $\hat{K}$:

$$\omega(\hat{\mathbf{A}}, \mathbf{P}) = \omega_0(\hat{A}_1, X_1) + \omega_0(\hat{A}_2, X_2), \quad (17)$$

où :

$$\omega_0(\hat{A}_i, X_i) = \iint_{A_i \times X_i} \delta(x, y) d\mu(x) d\mu(y) = \int_{X_i} \hat{A}_i d\mu = \int_{A_i} \hat{X}_i d\mu. \quad (18)$$

On considère les extensions naturelles des métriques précédentes sur $\hat{K} \times \mathcal{P}$. Insistons plus particulièrement sur :

$$e = D^2 \times \Delta^2 \quad \text{et} \quad e_2 = D_2^2 \times \Delta^2. \quad (19)$$

On établit alors des propriétés d'absolue continuité de ω sur $\hat{K} \times \mathcal{P}$ muni de e ou de e_2 . On établit d'autre part — dans des conditions adéquates — des propriétés de continuité pour chacune des applications π (réallocation) et ν (recentrage). Enfin, relativement à l'application $\tau = \nu \circ \pi$ de :

$$(\{\hat{\mathbf{A}} \in \hat{K} / \mu[(\pi \mathbf{A})_1] \mu[(\pi \mathbf{A})_2] \neq 0\}, D_2, D \text{ ou } \Delta),$$

dans $(\hat{K}, D_2, D \text{ ou } \Delta)$, on établit sa continuité.

On démontre alors dans le cadre le plus général les propriétés de décroissance correspondantes à (6) et à (8) :

$$\omega(\hat{\mathbf{A}}, \mathbf{P}) \geq \omega[\hat{\mathbf{A}}, \pi(\hat{\mathbf{A}})] \quad (20)$$

et :

$$\omega(\hat{\mathbf{A}}, \mathbf{P}) \geq \omega[\nu(\mathbf{P}), \mathbf{P}]. \quad (21)$$

Le résultat acquis le plus important de cette étude concerne la description topologique des différentes issues possibles dans le déroulement de l'algorithme dont un même pas se trouve défini par la fonction τ .

La première issue est celle où au bout d'un nombre fini d'étapes, on tombe sur une partition dégénérée réduite à une seule classe; en d'autres termes, il existe un plus petit entier $m \geq 1$, tel que :

$$\pi(\hat{\mathbf{A}}^{(m)}) = \mathbf{P}^{(m+1)} = \{X\},$$

où nous avons noté : $\hat{\mathbf{A}}^{(m)} = \tau^{(m)}[\hat{\mathbf{A}}^{(0)}]$ avec le système initial de noyaux $\hat{\mathbf{A}}^{(0)}$.

La deuxième issue est celle où au bout d'un nombre fini d'étapes on obtient un système stable de noyaux : il existe un plus petit entier $m \geq 1$, tel que :

$$\hat{\mathbf{A}}^{(m+1)} = \tau[\hat{\mathbf{A}}^{(m)}] = \hat{\mathbf{B}}.$$

On a alors trivialement la convergence — au sens de D , D_2 ou Δ — de $\hat{\mathbf{A}}^{(n)}$ vers $\hat{\mathbf{B}}$ et celle — au sens de Δ — de $\mathbf{P}^{(n)}$ vers $\mathbf{Q} = \pi(\hat{\mathbf{B}})$.

La troisième issue, qui demande l'analyse la plus élaborée, est celle où pour tout n ,

$$\mathbf{A}^{(n+1)} \neq \hat{\mathbf{A}}^{(n)} \quad \text{et} \quad \pi[\hat{\mathbf{A}}^{(n+1)}]$$

non dégénérée. Dans ces conditions, la suite $\{\hat{\mathbf{A}}^{(n)}/n \geq 1\}$ possède au moins un point d'accumulation $\mathbf{B}$ dans la fermeture $\bar{\hat{K}}$ de $\hat{K}$ (relativement à D ou à D_2), pour lequel des propriétés de stabilité sont établies concernant la partition et le noyau images de $\mathbf{B}$ par π et τ . La sous-suite $\hat{\mathbf{A}}^{(n_k)}$ convergente vers $\mathbf{B}$ au sens de D (resp. D_2), donne lieu — via π — à une suite de partitions $\mathbf{P}^{(n_k)}$ qui converge — au sens de Δ — vers $\pi(\mathbf{B})$. Si le point d'accumulation $\mathbf{B}$ tombe dans $\hat{K}$, alors :

$$\omega(\bar{\mathbf{B}}) = \inf \{ \omega(\hat{\mathbf{A}}^{(n)})/n \geq 0 \}.$$

Enfin, si $\lim_{n \rightarrow \infty} D \text{ (resp. } D_2) (\hat{\mathbf{A}}^{(n)}, \mathbf{B}) = 0 \nsubseteq -\mathbf{B}$ ne donnant lieu ni à ambiguïté d'affectation, ni à partition dégénérée — alors $\mathbf{B} \in \hat{K}$, $\Delta\text{-}\lim \hat{\mathbf{A}}^{(n)} = \mathbf{B}$ et on a les propriétés de stabilité qu'on vient de mentionner.

L'étude se termine par une conjecture reliant la convergence au sens de Δ à celle au sens de D (resp. D_2) de la sous-suite $\hat{\mathbf{A}}^{(n_k)}$ vers le point d'accumulation $\bar{\mathbf{B}}$, supposé appartenir à $\hat{K}$.

Pour nous résumer, l'étude de J. Diebolt, J. C. Simon et W. L. Miranker fournit une description dans un cadre topologique par trop général de l'espace de représentation des données et des différents éléments nourrissant l'algorithme de « réallocation-recentrage » dans sa version « nuées-dynamiques ». Ainsi se trouvent précisée de façon cohérente la nature de la fonction critère ω et des métriques par rapport auxquelles la convergence de l'algorithme a lieu. On y trouve d'autre part une description topologique des différents types d'issues possibles dans le déroulement de l'algorithme et certaines informations concernant la nature de la convergence.

1.3. Apport de notre analyse

Toutefois, l'étude précédente ne nous informe guère sur la nature de la solution obtenue par l'algorithme et son caractère d'optimalité, compte tenu de μ et du système initial de noyaux $\mathbf{A}^{(0)}$. Notre propre contribution se situe délibérément dans cette dernière optique. D'autre part, notre démarche sera en quelque sorte opposée. En effet, partisans du degré de généralité nécessaire, nous considérerons un type pur de situation mathématique aussi simple que possible et nous étudierons le comportement de l'algorithme et la nature de la solution finale, compte tenu des caractéristiques de la situation et du choix du système initial de noyaux.

La situation mathématique la plus simple de la parfaite classifiabilité est définie par — comme il est indiqué dans la figure ci-dessous — une distribution continue de masses sur un axe, chargeant de façon uniforme et avec une densité à 1, une suite non connexe formée de deux intervalles de longueurs respectives l_1 et l_2 et qu'on désignera par S_1 et S_2 ; alors que T — de longueur t

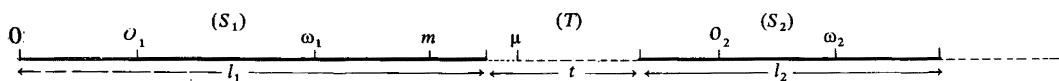


Figure 1

indiquera l'intervalle séparant S_1 de S_2 . Précisons qu'en dehors de S_1 et S_2 , la densité de la distribution est nulle. Dans cette figure nous avons porté quelques éléments qui seront utilisés aux paragraphes suivants : (o_1, o_2) [resp. $(\omega_1, \omega_2)] \in S_1 \times S_2$ et m (resp. μ) qui est le milieu du segment $[o_1, o_2]$ (resp. $[\omega_1, \omega_2]$).

On ne peut admettre — comme il arrive parfois — de voir poser une axiomatique d'une classification, sans que cette parfaite partition en deux classes S_1 et S_2 , satisfasse une telle axiomatique. Il s'agit là en effet d'une question de *cohérence logique*. D'autre part, un algorithme de classification ascendante hiérarchique (réunion des paires également proches de classes et fermeture transitive) utilisant un critère compatible avec la distance euclidienne ordinaire, conduit au bout d'une seule étape à la découverte de la classification en deux classes S_1 et S_2 .

On peut croire que l'algorithme de « reallocation-recentrage » dans sa version originelle — que nous avons rappelée au début du paragraphe 1.1 — conduira « aisément » à la solution optimale. En fait, il n'en est rien et ce sont des conditions bien précises sur la comparaison des amplitudes l_1 , t , l_2 et la disposition initiale du système de centres d'attraction qui assureront la convergence vers la solution optimale. Le but majeur de ce papier — construit autour des théorèmes du paragraphe II — est de montrer ce fait en précisant les conditions auxquelles la méthode des pôles d'attraction apporte sa contribution.

Dans son ouvrage et en quelques lignes J. P. Benzecri considère bien la situation que nous étudions ici pour mentionner l'aspect nécessaire de la condition établie dans le cadre de la propriété 2 du paragraphe II ci-dessous [Benzecri (1973) TIB n° 9, § 2.2.2].

Signalons pour terminer que différents chercheurs ont pu s'intéresser à l'analyse expérimentale de ce type d'algorithme dans le cas d'une distribution unidimensionnelle [e. g. [Lechevalier (1974)]].

1.4. Un autre problème de convergence

Pour éviter toute confusion, nous voulons mentionner ici un problème de convergence d'une toute autre nature, résolu par D. Pollard [Pollard (1981)].

L'ensemble E des unités de données est regardé comme la réalisation d'un échantillon aléatoire provenant d'une distribution de probabilité P' sur $\mathbb{R}^p$, de sorte que le nuage $\mathcal{N}(E)$ (cf. § 1.1) est équipondéré ($\mu_x = 1$, pour tout $x \in E$). On associe à $E = \{x_i / 1 \leq i \leq n\}$, une distribution de probabilité empirique P'_n sur $\mathbb{R}^p$, plaçant en chaque sommet x_i la masse $1/n$. Ainsi, relativement à un système L de k centres d'attraction, l'expression du critère ω peut se mettre sous la forme :

$$\omega(L, P'_n) = \int \min \{ \|x - o\|^2 / o \in L \} P'_n(dx) \quad (22)$$

et — comme nous l'avons souligné ci-dessus — le problème de la classification est la détermination du système optimal L_n formé de k centres d'attraction qui rend minimal (22).

Pour chaque L fixé, la loi forte des grands nombres permet de montrer que — pour n tendant vers l'infini — on a la convergence presque sûre de la valeur du critère :

$$\omega(L, P'_n) \rightarrow \omega(L, P') = \int \min \{ \|x - o\|^2 / o \in L \} P'(dx), \text{ (p. s.)} \quad (23)$$

Dans ces conditions et dans la mesure où le système optimal de centres d'attraction $\bar{L}$ pour P' — formé d'au plus k sommets — est unique, on peut espérer montrer la convergence dans la « proximité » entre L_n et $\bar{L}$.

Plus précisément et pour k fixé, le résultat principal du travail que nous venons de citer, est d'établir la convergence presque sûre de $L_n(k)$ [système optimal (pour P'_n) vers $\bar{L}(k)$]. Cette convergence presque sûre est une convergence point par point qui suppose un étiquetage $o_{n1}, o_{n2}, \dots, o_{nk}$ des points dans $L_n(k)$ et un étiquetage $o_1, o_2, \dots, o_k$ des points dans $\bar{L}$, tels que $o_{nj} \rightarrow o_j$ (p. s.), pour tout $j = 1, 2, \dots, k$.

Pour réaliser cette convergence p. s. d'un ensemble vers un autre, on utilise la distance de Hausdorff qui, rappelons le, se trouve définie comme suit : A et B étant deux parties quelconques de $\mathbb{R}^p$,

$$H(A, B) = \sup[\rho(A, \{B\}), \rho(B, \{A\})], \quad (24)$$

où :

$$\rho(A, \{B\}) = \sup \{ d(a, B) / a \in A \},$$

représente la distance maximale entre un point de A et l'ensemble B , globalement pris;

$$d(a, B) = \inf \{ d(a, b) / b \in B \}.$$

où d est la distance euclidienne ordinaire.

De sorte que si A contient k sommets, B au plus k sommets et $H(A, B) < \delta$, où δ est choisi strictement inférieur à la moitié de la plus petite distance séparant deux points de A , alors B doit contenir exactement k sommets, chacun correspondant à un sommet de A et s'y trouvant dans une boule ouverte de centre de ce sommet et de rayon δ .

Ainsi, le résultat de D. Pollard est d'une nature fondamentale et classique en statistique mathématique. Il étudie en effet la convergence de la solution optimale — supposée donnée — de l'algorithme dans le cas fini, vers la solution optimale de ce même algorithme — toujours supposer donnée — dans le cas continu infini, lorsque le cas fini converge vers le cas infini.

J. Lemaire [Lemaire (1983)] a défini un cadre très général topologique et probabiliste permettant de situer le problème de D. Pollard par rapport à d'autres de même nature et déjà étudiés dans le cadre d'autres critères tels que ceux du « maximum de vraisemblance » et de la « vraisemblance classifiante ».

1.5. Idée du travail proposé

Le point de départ de l'algorithme de « reallocation-recentrage » pour la détermination d'une partition en deux classes de l'ensemble E des unités de données, est le choix d'un couple $(o_1^{(0)}, o_2^{(0)})$ de centres d'attraction, appartenant à $E \times E$ (cf. § 1.1). En supposant que E se trouve représenté par la situation mathématiquement pure définie au paragraphe 1.3 ci-dessus ($E = S_1 \cup S_2$), on se pose la question de savoir s'il n'y a pas de conditions générales — incluant le choix du système initial de centres d'attraction — qui assurent la convergence optimale de l'algorithme où au bout d'un nombre fini ou infini d'étapes, le couple de centres d'attraction viendra occuper le couple de centres de gravité de S_1 et de S_2 qui sont respectivement les milieux de S_1 et de S_2 .

En cas de convergence optimale, à partir d'un certain rang r , on a $(o_1^{(r)}, o_2^{(r)}) \in S_1 \times S_2$. On peut supposer que $(o_1^{(0)}, o_2^{(0)}) \in S_1 \times S_2$, dès lors qu'une méthode nous permet d'assurer cette appartenance pour le système initial de centres d'attraction. C'est précisément le rôle de la méthode des pôles d'attraction.

Toutefois, il importe que l'image par $\tau = \nu \circ \pi$ (cf. § 1.3) d'un couple de centres d'attraction du rectangle $S_1 \times S_2$, tombe dans $S_1 \times S_2$. Nous commencerons par préciser une condition sur la comparaison des amplitudes l_1 et l_2 de S_1 et de S_2 , qui assure ce fait.

En munissant $S_1 \times S_2$ de la distance : somme des valeurs absolues des différences des coordonnées de même nom i ($i=1, 2$), le rectangle $S_1 \times S_2$ est un espace métrique complet.

Or — dans la mesure où elle est unique — la solution du problème d'optimisation (g_1, g_2) apparaît comme solution de l'équation $\tau(g_1, g_2) = (g_1, g_2)$. De sorte que l'algorithme de « réallocation-recentrage » ne correspond à rien d'autre qu'à une matérialisation de la « méthode des approximations successives » qui est une des méthodes les plus puissantes pour démontrer l'existence de solutions d'une équation et parfois, pour calculer effectivement cette solution. Toutefois, cette méthode ne converge vers la solution de l'équation que sous certaines conditions particulières que nous allons rappeler.

Disons d'abord qu'une application ϕ d'un espace métrique F dans un espace métrique G est contractante de rapport ρ où $0 \leq \rho < 1$, si pour tout couple de points (y, y') de F , on a :

$$d_G[\phi(y), \phi(y')] \leq \rho d_F(y, y'), \quad (25)$$

où d_F (resp. d_G) est la distance dont se trouve muni F (resp. G).

Dans ces conditions, on a le théorème bien connu (Choquet (1964)) :

THÉORÈME : *Soit F un espace métrique complet et soit ϕ une application contractante de F dans lui-même. Pour tout $y^{(0)} \in F$, la suite des transformés successifs $y^{(n)} = \phi^{(n)}(y^{(0)})$ du point $y^{(0)}$, converge vers un point g qui est solution de l'équation $y = \phi(y)$. Cette solution g est la seule solution de l'équation.*

Nous montrerons que si $t < \min(l_1, l_2)$, l'application τ restreinte à $S_1 \times S_2$ est contractante. Dans ces conditions, pourvu que $\tau(g_1, g_2) = (g_1, g_2)$ [g_1 (resp. g_2) est le milieu de S_1 (resp. S_2)], quel que soit le point source $(o_1^{(0)}, o_2^{(0)})$ de $S_1 \times S_2$ et pourvu que τ soit une application de $S_1 \times S_2$ dans $S_1 \times S_2$, la suite des itérés du point source converge nécessairement vers (g_1, g_2) .

Au paragraphe 3, deux contre-exemples permettront de se rendre compte du caractère quasi nécessaire des conditions posées pour la convergence optimale.

Au paragraphe 4 nous montrerons comment la méthode des pôles d'attraction assure l'appartenance de $(o_1^{(0)}, o_2^{(0)})$ à $S_1 \times S_2$.

2. THÉORÈME DE CONVERGENCE OPTIMALE

2.1. Considérations et notations préliminaires (cf. fig. 1, § 1.3)

Soient (o_1, o_2) et (ω_1, ω_2) deux systèmes initiaux quelconques de noyaux-singletons tels que :

$$(o_1, o_2) \in S_1 \times S_2 \quad \text{et} \quad (\omega_1, \omega_2) \in S_1 \times S_2. \quad (1)$$

En considérant comme origine de l'axe supportant les deux intervalles, l'origine du premier intervalle S_1 , on désignera par x_1 (resp. x_2) l'abscisse du point o_1 (resp. o_2) et par ξ_1 (resp. ξ_2) l'abscisse du point ω_1 (resp. ω_2).

Rappelons ici que pour un système de noyaux (n_1, n_2) — où on suppose n_1 à gauche de n_2 — $\pi(n_1, n_2)$ définit une partition en deux classes que délimite la médiatrice du segment $n_1 n_2$. La classe C_i attribuée à n_i se trouve définie par :

$$G_i = \{p/p \in S_1 \cup S_2 \text{ et } d(p, n_i) < d(p, n_{(3-i)})\}. \quad (3)$$

$v(C_1, C_2)$ associe à (C_1, C_2) , $(g(C_1), g(C_2))$ où $g(C_i)$ ($i=1, 2$) est le centre de gravité de la classe C_i .

On comprend dans ces conditions que l'analyse va devoir s'organiser par rapport aux positions respectives des milieux des segments $o_1 o_2$ et $\omega_1 \omega_2$. On désignera par m (resp. μ) l'abscisse du milieu de $o_1 o_2$ (resp. $\omega_1 \omega_2$) : $m = (x_1 + x_2)/2$ [resp. $\mu = (\xi_1 + \xi_2)/2$].

On a pour m les trois cas suivants (numérotés et encerclés d'un rond) :

① : $m < l_1$; ② : $l_1 \leq m < l_1 + t$ et ③ : $l_1 + t \leq m < l_1 + t + l_2$.

Respectivement, on a pour μ les trois cas suivants (numérotés et encadrés d'un carré) :

Ⓐ : $\mu < l_1$; Ⓑ : $l_1 \leq \mu < l_1 + t$ et Ⓒ : $l_1 + t \leq \mu < l_1 + t + l_2$.

Neuf situations seront donc considérées résultant du croisement des trois cas définis pour m et des trois cas correspondants définis pour μ .

Désignons par $\mathcal{Z}_i = \mathcal{Z}(o_i)$ la zone attribuée à o_i et par $\zeta_i = \zeta(\omega_i)$ celle attribuée à ω_i :

$$\mathcal{Z}_i = \{p/p \in S_1 \cup S_2 \quad \text{et} \quad d(p, o_i) < d(p, o_{(3-i)})\} \quad (3')$$

et :

$$\zeta_i = \{p/p \in S_1 \cup S_2 \text{ et } d(p, \omega_i) < d(p, \omega_{(3-i)})\}.$$

L'influence de l'attribution du point tel que $d(p, o_1) = d(p, o_2)$ [resp. $d(p, \omega_1) = d(p, \omega_2)$] est nulle, de sorte qu'en désignant par y l'abscisse

d'un point courant p de l'axe, on définira :

$$\left. \begin{aligned} \mathcal{Z}_1 &= \{p/o < y \leq m\}, & \mathcal{Z}_2 &= \{p/m < y \leq l_1 + t + l_2\}, \\ \zeta_1 &= \{p/o < y \leq \mu\} & \text{et} & \quad \zeta_2 = \{p/\mu < y \leq l_1 + t + l_2\} \end{aligned} \right\}. \quad (4)$$

Nous allons calculer, dans chacun des cas ①, ② et ③, le nouveau système de noyaux que nous noterons $(x'_1, x'_2) = [g(\mathcal{Z}_1), g(\mathcal{Z}_2)]$.

Cas ① : $x_1 + x_2 \leq 2l_1$

$$\begin{aligned} x'_1 &= g(\mathcal{Z}_1) = \frac{m}{2} = \frac{1}{4}(x_1 + x_2). \\ x'_2 &= g(\mathcal{Z}_2) = \frac{\lambda_1}{\lambda_1 + l_2} \times \frac{(l_1 + m)}{2} + \frac{l_2}{\lambda_1 + l_2} \cdot \frac{2l_1 + l_2 + 2t}{2}, \end{aligned}$$

où nous avons noté λ_1 la différence positive $(l_1 - m)$.

$$x'_2 = \frac{1}{2(\lambda_1 + l_2)} [\lambda_1(l_1 + m) + l_2(2l_1 + l_2 + 2t)].$$

Cas ② : $2l_1 < (x_1 + x_2) \leq 2(l_1 + t)$ (pour ce cas on peut consulter la propriété ci-dessous) :

$$x'_1 = g(\mathcal{Z}_1) = \frac{l_1}{2} \quad \text{et} \quad x'_2 = \frac{2l_1 + l_2 + 2t}{2}.$$

Cas ③ : $2(l_1 + t) < (x_1 + x_2) \leq 2(l_1 + t + l_2)$:

$$x'_1 = g(\mathcal{Z}_1) = \frac{l_1}{(l_1 + \lambda_2)} \times \frac{l_1}{2} + \frac{\lambda_2}{(l_1 + \lambda_2)} \times \frac{(l_1 + t + m)}{2},$$

où nous avons noté λ_2 pour $m - (l_1 + t)$.

$$\begin{aligned} x'_1 &= \frac{1}{2(\lambda_2 + l_1)} [l_1^2 + \lambda_2(l_1 + t + m)], \\ x'_2 &= g(\mathcal{Z}_2) = \frac{1}{2} [m + (l_1 + t + l_2)]. \end{aligned}$$

On considérera de même le nouveau système de noyaux $(\xi'_1, \xi'_2) = [g(\zeta_1), g(\zeta_2)]$ dans chacun des cas ①, ② et ③ :

Cas ① : $\xi_1 + \xi_2 \leq 2l_1$.

En posant $v_1 = (l_1 - \mu) > 0$, on a :

$$\xi'_1 = g(\zeta_1) = \frac{\mu}{2} = \frac{1}{4}(\xi_1 + \xi_2),$$

$$\xi'_2 = g(\zeta_2) = \frac{1}{2(v_1 + l_2)} [v_1(l_1 + \mu) + l_2(2l_1 + l_2 + 2t)].$$

Cas ② : $2l_1 < (\xi_1 + \xi_2) \leq 2(l_1 + t)$:

$$\xi'_1 = g(\zeta_1) = \frac{l_1}{2} \quad \text{et} \quad \xi'_2 = g(\zeta_2) = \frac{2l_1 + l_2 + 2t}{2}.$$

Cas ③ : $2(l_1 + t) < (\xi_1 + \xi_2) \leq 2(l_1 + t + l_2)$.

En posant $v_2 = \mu - (l_1 + t)$, on a :

$$\xi'_1 = g(\zeta_1) = \frac{1}{2(v_2 + l_1)} [l_1^2 + v_2(l_1 + t + \mu)],$$

$$\xi'_2 = g(\zeta_2) = \frac{1}{2} [\mu + (l_1 + t + l_2)].$$

Nous allons à présent établir une condition suffisante pour laquelle l'image par $\tau = v \circ \pi$ d'un élément [représentant un système (o_1, o_2) de centres d'attraction] de $S_1 \times S_2$, tombe dans $S_1 \times S_2$. Sans aucunement restreindre la généralité, on peut supposer — comme la figure 1 le suggère — que $l_1 \geq l_2$.

PROPRIÉTÉ 1. — Si $l_1 \leq \sqrt{3}l_2$, l'image par $\tau = v \circ \pi$ de $((o_1, o_2) \in S_1 \times S_2$ est un point (o'_1, o'_2) qui appartient à $S_1 \times S_2$.

Avec les notations précédentes, nous allons considérer chacun des cas ①, ② et ③.

Cas ① :

Il est clair que l'abscisse x'_1 de o'_1 est strictement inférieure à l_1 . L'essentiel est de montrer que l'abscisse x'_2 de o'_2 est supérieure à $(l_1 + t)$ (i. e. $o'_2 \in S_2$). Rappelons que :

$$x'_2 = \frac{\lambda_1}{(\lambda_1 + l_2)} \left(\frac{l_1}{2} + \frac{m}{2} \right) + \frac{l_2}{(\lambda_1 + l_2)} \cdot \frac{l_2}{2} + \frac{l_2}{(\lambda_1 + l_2)} \cdot (l_1 + t). \quad (5)$$

En ajoutant et en retranchant $\lambda_1(l_1 + t)/(\lambda_1 + l_2)$, on obtient :

$$x'_2 = (l_1 + t) + \frac{\lambda_1}{(\lambda_1 + l_2)} \left(\frac{l_1}{2} + \frac{m}{2} - l_1 - t \right) + \frac{l_2}{(\lambda_1 + l_2)} \cdot \frac{l_2}{2}$$

$$= (l_1 + t) + \frac{1}{(\lambda_1 + l_2)} \left[-\lambda_1 \left(\frac{\lambda_1}{2} + t \right) + \frac{l_2^2}{2} \right]. \quad (6)$$

Dans l'expression de x'_2 , le complément à $(l_1 + t)$ est positif si et seulement si :

$$l_2^2 \geq \lambda_1 (\lambda_1 + 2t) = (\lambda_1 + t)^2 - t^2, \quad (7)$$

soit :

$$(\lambda_1 + t)^2 \leq l_2^2 + t^2. \quad (8)$$

Or, puisque $o_2 \in S_2$, on a :

$$m \geq \frac{x_2}{2} \geq \frac{l_1 + t}{2}, \quad (9)$$

par conséquent,

$$\lambda_1 + t = l_1 - m + t \leq \frac{l_1 + t}{2}, \quad (10)$$

et :

$$(\lambda_1 + t)^2 \leq \frac{1}{4}(l_1 + t)^2$$

Mais si $l_1 \leq \sqrt{3} l_2$, on a :

$$\frac{1}{4}(l_1 + t)^2 \leq l_2^2 + t^2. \quad (11)$$

En effet cette dernière inégalité revient à cette :

$$3t^2 - 2l_1 t + (4l_2^2 - l_1^2) \geq 0, \quad (11')$$

qui se trouve assurée dès que ce trinôme du second degré en t a son discriminant Δ' négatif. Or :

$$\Delta' = l_1^2 - 3(4l_2^2 - l_1^2) = 4(l_1^2 - 3l_2^2). \quad (12)$$

C.Q.F.D.

Cas ② :

Ce cas est immédiat. En effet, (o'_1, o'_2) n'est autre que (g_1, g_2) où g_1 (resp. g_2) est le milieu de S_1 (resp. S_2).

Cas ③ :

Ce cas est symétrique de ①, les rôles de S_1 et de S_2 se trouvant intervertis. La même démonstration que dans le cas ① est valable; il suffit en effet de renverser le sens de l'axe et de prendre comme origine l'extrémité de S_2 . Compte tenu de ci-dessus, $(o'_1, o'_2) \in S_1 \times S_2$ si $l_2 \leq \sqrt{3} l_1$. Or nous avons supposé que $l_2 \leq l_1$, donc *a fortiori* $l_2 \leq \sqrt{3} l_1$.

PROPRIÉTÉ 2 : La condition nécessaire et suffisante pour que $\tau(g_1, g_2) = (g_1, g_2)$ est que t soit supérieure à $(l_1 - l_2)/2$, où on suppose $l_1 \geq l_2$.

$\tau(g_1, g_2) = (g_1, g_2)$ si et seulement si le milieu du segment (g_1, g_2) tombe dans T . L'abscisse de ce milieu est :

$$\frac{1}{2} \left(\frac{l_1}{2} + l_1 + t + \frac{l_2}{2} \right).$$

Il s'agit d'écrire que cette abscisse est comprise entre l_1 et $(l_1 + t)$:

$$l_1 \leq \frac{1}{2} \left(\frac{l_1}{2} + l_1 + t + \frac{l_2}{2} \right) \leq l_1 + t,$$

ce qui donne :

$$t \geq \frac{1}{2} (l_1 - l_2). \quad (13)$$

2.2. Résultat principal

Nous munissons $S_1 \times S_2$ de la distance : somme des valeurs absolues des différences des coordonnées de même nom. Ainsi, la distance entre (o_1, o_2) et (ω_1, ω_2) est définie par :

$$d[(o_1, o_2), (\omega_1, \omega_2)] = |x_1 - \xi_1| + |x_2 - \xi_2|. \quad (14)$$

THÉORÈME 1 : Si le système initial de noyaux-singletons (i. e. centres d'attraction) est formé d'un couple (o_1, o_2) de sommets appartenant respectivement aux deux intervalles S_1 et S_2 de longueurs l_1 et l_2 et si la longueur t de l'intervalle T séparant S_1 et S_2 est strictement inférieure à $\min(l_1, l_2)$, alors l'image par $\tau = v \circ \pi$ de (o_1, o_2) est contractée. En d'autres termes, il existe ρ ($0 \leq \rho < 1$) pour lequel :

$$d[(o'_1, o'_2), (\omega'_1, \omega'_2)] \leq \rho d[(o_1, o_2), (\omega_1, \omega_2)]. \quad (15)$$

Nous allons à présent comparer — pour les neuf situations définies ci-dessus : $(\textcircled{i}, \textcircled{j})$, $1 \leq i, j \leq 3$ —

$$d[(o'_1, o'_2), (\omega'_1, \omega'_2)] = |x'_1 - \xi'_1| + |x'_2 - \xi'_2|$$

à $d[(o_1, o_2), (\omega_1, \omega_2)]$ [cf. (14) ci-dessus].

Cas (①, ①) :

$$\begin{aligned} |x'_1 - \xi'_1| + |x'_2 - \xi'_2| &= \frac{1}{4} |(x_1 - \xi_1) + (x_2 - \xi_2)| \\ &+ \frac{1}{2} \left| \frac{(l_1 - m)(l_1 + m) + l_2(2l_1 + l_2 + 2t)}{(l_1 + l_2 - m)} \right. \\ &\quad \left. - \frac{(l_1 - \mu)(l_1 + \mu) + l_2(2l_1 + l_2 + 2t)}{(l_1 + l_2 - \mu)} \right|. \quad (16) \end{aligned}$$

Après avoir posé $L = (l_1 + l_2)$ développons le contenu de la deuxième expression sous valeur absolue qui peut alors se mettre sous la forme :

$$\begin{aligned} &\frac{-L(m^2 - \mu^2) + l_1^2(m - \mu) + m\mu(m - \mu) + l_2(L + l_1 + 2t)(m - \mu)}{(L - m)(L - \mu)} \\ &= \frac{(m - \mu)[-L(m + \mu) + m\mu + l_1^2 + l_2(L + l_1 + 2t)]}{(L - m)(L - \mu)}. \end{aligned}$$

En introduisant dans le dernier crochet du numérateur $(L - m)(L - \mu)$ et en développant le reste par rapport à l_1 et l_2 , on obtient :

$$(m - \mu) \left[1 + \frac{2l_2 t}{(L - m)(L - \mu)} \right]. \quad (17)$$

Or on est dans une situation caractérisée par $m \leq l_1$ et $\mu \leq l_1$. D'où :

$$(L - m)(L - \mu) \geq l_2^2.$$

Finalement, après mise en forme, on a :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| \leq \frac{1}{2} \left(1 + \frac{t}{l_2} \right) (|x_1 - \xi_1| + |x_2 - \xi_2|). \quad (18)$$

Cas (②, ②) :

$$|x'_1 - \xi'_1| = \left| \frac{m}{2} - \frac{l_1}{2} \right| = \frac{1}{2}(l_1 - m) = \frac{1}{2}\lambda_1, \quad (19)$$

$$|x'_2 - \xi'_2| = \frac{1}{2} \left| \frac{\lambda_1(l_1 + m) + l_2(2l_1 + l_2 + 2t)}{(\lambda_1 + l_2)} - (2l_1 + l_2 + 2t) \right|. \quad (20)$$

En regroupant (19) et (20), après calcul on obtient :

$$\begin{aligned} |x'_1 - \xi'_1| + |x'_2 - \xi'_2| &= \frac{\lambda_1(\lambda_1 + l_2 + t)}{(\lambda_1 + l_2)} \\ &= \frac{(l_1 - m)(l_1 + l_2 + t - m)}{(l_1 + l_2 - m)} = \left[1 + \frac{t}{(\lambda_1 + l_2)} \right] \lambda_1, \end{aligned} \quad (21)$$

qui ne dépend que de $(x_1 + x_2)$.

Où le cas [2] entraîne que :

$$\begin{aligned} \lambda_1 &= \left(l_1 - \frac{x_1 + x_2}{2} \right) < \frac{1}{2} [(\xi_1 + \xi_2) - (x_1 + x_2)] \\ &\leq \frac{1}{2} (|\xi_1 - x_1| + |\xi_2 - x_2|). \end{aligned} \quad (22)$$

De sorte que :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| < \frac{1}{2} \cdot \left[1 + \frac{t}{(\lambda_1 + l_2)} \right] (|\xi_1 - x_1| + |\xi_2 - x_2|). \quad (23)$$

Cas (①, ③) :

$$\begin{aligned} |x'_1 - \xi'_1| &= \frac{1}{2} \left| \frac{m(l_1 + v_2) - (l_1^2 + v_2(\mu + l_1 + t))}{(l_1 + v_2)} \right| \quad \text{où } v_2 = \mu - (l_1 + t) \\ &= \frac{1}{2} \left| \frac{(l_1 + v_2)(m - \mu) + (l_1 - v_2)t}{(l_1 + v_2)} \right|. \end{aligned} \quad (24)$$

$$\begin{aligned} |x'_2 - \xi'_2| &= \frac{1}{2} \left| \frac{\lambda_1(l_1 + m) + l_2(2l_1 + l_2 + 2t) - (l_2 + \lambda_1)(\mu + l_1 + l_2 + t)}{(l_2 + \lambda_1)} \right|, \\ \text{où } \lambda_1 &= l_1 - m \\ &= \frac{1}{2} \left| \frac{(l_2 + \lambda_1)(m - \mu) + (l_2 - \lambda_1)t}{(l_2 + \lambda_1)} \right|. \end{aligned} \quad (25)$$

Ainsi,

$$\begin{aligned} |x'_1 - \xi'_1| + |x'_2 - \xi'_2| &= \frac{1}{2} \left(\left| (\mu - m) + \frac{(v_2 - l_1)}{(v_2 + l_1)} t \right| \right. \\ &\quad \left. + \left| (\mu - m) + \frac{(\lambda_1 + l_2)}{(\lambda_1 + l_2)} t \right| \right) \end{aligned} \quad (26)$$

$(v_2 - l_1)/(v_2 + l_1) = (\mu - 2l_1 - t)/(\mu - t)$ est une fonction strictement croissante de μ , dont la plus petite (resp. grande) valeur est — pour $(l_1 + t) < \mu \leq (l_1 + l_2 + t)$ (cas $\textcircled{3}$) — -1 (resp. $(l_2 - l_1)/(l_1 + l_2)$). En particulier $|(v_2 - l_1)/(v_2 + l_1)| \leq 1$.

De même $(\lambda_1 - l_2)/(\lambda_1 + l_2) = (l_1 - l_2 - m)/(l_1 + l_2 - m)$ est une fonction strictement décroissante de m , dont la plus petite (resp. grande) valeur est — pour $m < l_1$ (cas $\textcircled{1}$) — -1 [resp. $(l_1 - l_2)/(l_1 + l_2)$]. En particulier $|(\lambda_1 - l_2)/(\lambda_1 + l_2)| \leq 1$.

Comme $m \leq l_1$ et $\mu > (l_1 + t)$ [Cas ($\textcircled{1}$, $\textcircled{3}$)], on a :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| = (\mu - m) + \frac{1}{2} \left[\frac{v_2 - l_1}{v_2 + l_1} + \frac{\lambda_1 - l_2}{\lambda_1 + l_2} \right] t. \quad (27)$$

Considérons maintenant le complément à $(\mu - m)$ dans le second membre de (27). D'après ce qui précède, la valeur la plus petite (la plus négative) de la contribution de ce complément est égale à $-t$ et se trouve atteinte pour $m = l_1$ et $\mu = l_1 + t$, où alors le second membre de (27) devient nul. La valeur la plus grande de cette contribution qui est atteinte pour $m = 0$ et $\mu = (l_1 + l_2 + t)$ est égale à :

$$\frac{1}{2} \left[\frac{(l_2 - l_1)}{(l_1 + l_2)} + \frac{(l_1 - l_2)}{(l_1 + l_2)} \right] = 0. \quad (28)$$

Donc de toute façon :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| \leq (\mu - m) \leq \frac{1}{2} (|\xi_1 - x_1| + |\xi_2 - x_2|). \quad (29)$$

Cas ($\textcircled{2}$, $\textcircled{1}$) :

Ce cas est équivalent à celui ($\textcircled{1}$, $\textcircled{2}$). Pour obtenir le résultat qui lui correspond, il y a lieu de permuter les rôles des deux intervalles S_1 et S_2 . Dans ces conditions, il s'exprime comme suit :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| < \frac{1}{2} \left[1 + \frac{t}{(v_2 + l_1)} \right] (|\xi_1 - x_1| + |\xi_2 - x_2|). \quad (30)$$

Cas ($\textcircled{2}$, $\textcircled{2}$) :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| = 0. \quad (31)$$

Cas (2), (3) :

$$\begin{aligned}
 |x'_1 - \xi'_1| + |x'_2 - \xi'_2| &= \frac{1}{2} \left\{ \left| l_1 - \frac{[l_1^2 + v_2(l_1 + t + \mu)]}{(l_1 + v_2)} \right| \right. \\
 &\quad \left. + \left| 2l_1 + l_2 + 2t - \mu - l_1 - l_2 - t \right| \right\} \\
 &= \frac{1}{2} \left\{ \frac{v_2(t + \mu)}{(l_1 + v_2)} + \mu - (l_1 + t) \right\} \\
 &= \frac{[\mu - (l_1 + t)]\mu}{(\mu - t)}. \quad (32)
 \end{aligned}$$

où la dernière relation est obtenue en tenant compte de $v_2 = \mu - (l_1 + t)$.

Or :

$$\begin{aligned}
 \frac{\mu}{(\mu - t)} \cdot [\mu - (l_1 + t)] &\leq \frac{\mu}{(\mu - t)} \left[\frac{(\xi_1 + \xi_2)}{2} - \frac{(x_1 + x_2)}{2} \right] \quad (\text{voir cas 2}) \\
 &\leq \frac{\mu}{2(\mu - t)} (|\xi_1 - x_1| + |\xi_2 - x_2|). \quad (33)
 \end{aligned}$$

La fonction $\mu/(\mu - t)$ est strictement décroissante par rapport à μ . Sa valeur maximale est atteinte pour $\mu = (l_1 + l_2 + t)$ (voir condition (3)) où elle vaut $[1/2 + t/2(l_1 + l_2)]$. Donc :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| \leq \frac{1}{2} \left[1 + \frac{t}{(l_1 + l_2)} \right] (|\xi_1 - x_1| + |\xi_2 - x_2|). \quad (34)$$

Cas (3), (1) :

Ce cas est analogue à celui (1), (3) précédemment traité et par rapport auquel on a à permuter les rôles des deux intervalles S_1 et S_2 . De sorte qu'on aboutit à la même inégalité que (19) ci-dessus.

Cas (3), (2) :

Ce cas est de même nature que celui (2), (3) qu'on vient d'examiner et où on permutera les rôles de S_1 et de S_2 , m et μ , $\lambda_2 = m - (l_1 + t)$ et v_2 . Par conséquent, on obtient la même inégalité (34).

Cas (3), (3) :

Ce cas est équivalent au cas (1), (1), il suffit en effet d'invertir les rôles des deux intervalles S_1 et S_2 . Dans ces conditions, on obtient [voir (18)] :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| \leq \frac{1}{2} \left(1 + \frac{t}{l_1} \right) (|x_1 - \xi_1| + |x_2 - \xi_2|). \quad (35)$$

On peut constater que dans tous les cas de figure on a :

$$|x'_1 - \xi'_1| + |x'_2 - \xi'_2| \leq \frac{1}{2} \left[1 + \frac{t}{\min(l_1, l_2)} \right] (|x_1 - \xi_1| + |x_2 - \xi_2|). \quad (36)$$

Comme nous avons supposé que t est strictement inférieur à $\min(l_1, l_2)$, on a :

$$\frac{1}{2} \left[1 + \frac{t}{\min(l_1, l_2)} \right] < 1. \quad (37)$$

Par conséquent l'application $\tau = v \circ \pi$ de réallocation (à partir de la plus petite distance) et de recentrage (au moyen du centre de gravité) est *contractante* de rapport défini par le premier membre de (37). Dans le cadre du théorème, alors que l'ensemble source de l'application τ est $S_1 \times S_2$, l'ensemble but est *a priori* $(S_1 \cup T \cup S_2)^2$.

THÉOREME 2 : *Si le système initial de centres d'attraction est formé d'un couple (o_1, o_2) de sommets appartenant respectivement aux intervalles S_1 et S_2 de longueurs l_1 et l_2 , si la longueur t de l'intervalle T séparant S_1 et S_2 est telle que :*

$$\frac{1}{2}(l_1 - l_2) \leq t < \min(l_1, l_2)$$

et si $\max(l_1, l_2) \leq \sqrt{3} \min(l_1, l_2)$, alors l'algorithme de réallocation-recentrage (défini par l'application τ) converge vers la solution optimale formée du couple de centres de gravité (g_1, g_2) . Cette solution est unique.

En vertu de la propriété 1 du paragraphe 2.1, τ est une application de $S_1 \times S_2$ dans $S_1 \times S_2$. Compte tenu du théorème 1 ci-dessus, cette application est contractante, enfin $\tau(g_1, g_2) = (g_1, g_2)$. D'où (cf. rappels du paragraphe 1.5) le résultat.

On remarquera que si $l_1 = l_2 = l$, les conditions du théorème sont satisfaites si $t < l$.

Ayant supposé — sans perte de généralité — que $l_1 \geq l_2$, on remarquera que la condition $l_1 \leq \sqrt{3} l_2$ suffit pour assurer l'existence d'un intervalle possible des valeurs de t ; en effet, cette condition est équivalente à :

$$\frac{1}{2}(l_1 - l_2) \leq \frac{1}{2}(\sqrt{3} - 1) l_2$$

qui est strictement inférieur à l_2 . Ainsi par exemple, si $l_1 = 12$ cm et $l_2 = 8$ cm, on a bien $l_1 \leq \sqrt{3} l_2$ et toute valeur de t telle que 2 cm $\leq t < 8$ cm permet d'assurer la convergence optimale.

3. ÉTUDE DE CONTRE-EXEMPLES

Nous allons étudier ici deux situations qui ne respectent pas les conditions du théorème précédent, où nous allons établir que l'algorithme ne converge pas vers la solution optimale (g_1, g_2) où — rappelons le — g_1 (resp. g_2) est le centre de gravité (ici le milieu) de S_1 (resp. S_2). Dans la première situation $t < 1/2(l_1 - l_2)$ et dans la seconde $t > l_2 = \min(l_1, l_2)$ avec non respect de la condition $l_1 \leq \sqrt{3} l_2$.

Nous allons ci-dessous (§ 3 et 4) simplifier quelque peu les notations en remplaçant l_1 par a et l_2 par b ($a \geq b$).

3.1. Étude du premier contre-exemple

(o_1, o_2) étant un couple de centres d'attraction et (o'_1, o'_2) son image par τ , rappelons que nous désignons par m l'abscisse du milieu du segment $o_1 o_2$ et par m' celle du milieu du segment $o'_1 o'_2$.

PROPRIÉTÉ. — Si $t < 1/2(a - b)$, alors $m < a$ implique $m' < a$.

Nous allons commencer par calculer m' par rapport à a, b et m . On a :

$$x'_1 = \frac{m'}{2}, \quad x'_2 = \left[\frac{1}{2}(a - m)(a + m) + b \left(a + t + \frac{b}{2} \right) \right] (a - m + b). \quad (1)$$

Ainsi :

$$\begin{aligned} m' &= \frac{m}{4} + \frac{(a - m)(a + m) + b(2a + b + 2t)}{4(a - m + b)} \\ &= \frac{1}{4} \left[\frac{m(a + b - m) + (a - m)(a + m) + b(2a + b + 2t)}{(a - m + b)} \right] \\ &= \frac{1}{4} \times \left[\frac{-2m^2 + m(a + b) + (a + b)^2 + 2bt}{(a - m + b)} \right]. \end{aligned}$$

Formons à présent $(m' - m)$:

$$\begin{aligned} m' - m &= \frac{1}{4} \times \left[\frac{-2m^2 + m(a + b) + (a + b)^2 + 2bt - 4m(a + b - m)}{(a - m + b)} \right] \\ &= \frac{2(a + b - m)^2 - (a + b)(a + b - m) + 2bt}{4(a + b - m)} \\ &= \frac{1}{4}(a + b - 2m) + \frac{bt}{2(a + b - m)}. \quad (2) \end{aligned}$$

Posons à présent $\alpha = (a - m)$. On a :

$$m' = (a - \alpha) + \frac{1}{4}(b - a + 2\alpha) + \frac{bt}{2(b + \alpha)} = \frac{3}{4}a + \frac{b}{4} - \frac{\alpha}{2} + \frac{bt}{2(b + \alpha)}. \quad (3)$$

On a :

$$m' < \frac{3}{4}a + \frac{b}{4} - \frac{\alpha}{2} + \frac{t}{2}. \quad (4)$$

Or si $t < 1/2(a - b)$, on a l'inégalité plus forte :

$$m' < \frac{3}{4}a + \frac{b}{4} + \frac{1}{4}(a - b) - \frac{\alpha}{2} = \left(a - \frac{\alpha}{2}\right). \quad (5)$$

COROLLAIRE : Si le système initial (o_1, o_2) de centres d'attraction appartient à $S_1 \times S_2$ et est tel que $m < a$, si de plus $a \leq \sqrt{3}b$ et si enfin t est strictement inférieur à $(1/2)(a - b)$ ($a > b$), alors l'algorithme de « reallocation-recentrage » converge vers une solution unique qui n'est pas optimale.

Pour que la preuve soit claire posons :

$$t = \frac{1}{2}(a - b) - \gamma(a - b), \quad (6)$$

où on suppose $a > b$ et $\gamma > 0$. D'après le calcul précédent [cf. (3)], on a :

$$\alpha' = a - m' = \frac{\alpha}{2} + \frac{1}{4}(a - b) - \frac{bt}{2(b + \alpha)}, \quad (7)$$

$$\alpha' > \frac{\alpha}{2} + \frac{1}{4}(a - b) - \frac{t}{2} > \frac{\alpha}{2} + \frac{\gamma}{2}(a - b). \quad (8)$$

Donc, au bout de la r -ième application de l'algorithme, il restera toujours :

$$\alpha^{(r)} = a - m^{(r)} > \frac{\gamma}{2}(a - b) \quad (9)$$

et donc une distance supérieure à un seuil constant strictement positif entre le nouveau système de centres d'attraction $(o_1^{(r)}, o_2^{(r)})$ et (g_1, g_2) qui définit la solution optimale.

D'autre part, $t < (1/2)(a - b)$ implique *a fortiori* $t < b$ puisque déjà $a \leq \sqrt{3}b$. L'application définie par τ de $S_1 \times S_2$ dans $S_1 \times S_2$ (cf. propriété 1, § 2. 1) est donc *contractante* (cf. théorème 1, § 2. 2). D'où le résultat de convergence non optimale.

On peut remarquer que si t est exactement égal à $(1/2)(a-b)$, la convergence devient optimale. En effet, la relation (7) devient :

$$\alpha' = f(\alpha) = \frac{\alpha}{2} + \frac{1}{4}(a-b) - \frac{b}{(b+\alpha)} \times \frac{1}{4}(a-b). \quad (10)$$

On a bien $f(0) = 0$ et l'application f qu'on vient de définir est contractante :

$$f(\alpha_2) - f(\alpha_1) = \left[\frac{1}{2} + \frac{1}{4} \cdot \frac{(a-b)b}{(b+\alpha_1)(b+\alpha_2)} \right] (\alpha_2 - \alpha_1) < \frac{3}{4} (\alpha_2 - \alpha_1), \quad (11)$$

puisque $a \leq \sqrt{3}b \Rightarrow (a-b) < b$.

La convergence vers la solution optimale où $\alpha=0$ s'obtient ici pour r tendant vers l'infini.

3.2. Étude du second contre-exemple

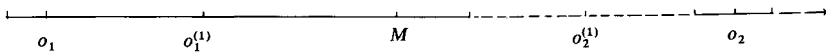
Considérons l'exemple suivant où $l_1 = 12$ cm, $t = 6$ cm et $l_2 = 2$ cm :



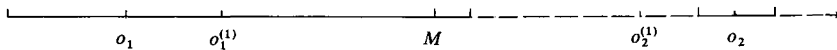
On désignera par (x_1, x_2) le couple des abscisses du système initial (o_1, o_2) de centres d'attraction et par $(x_1^{(r)}, x_2^{(r)})$ le couple des abscisses du r -ième transformé ($o_1^{(r)}, o_2^{(r)}$) du système initial de noyaux. $M^{(r)}$ est le milieu du segment reliant $o_1^{(r)}$ à $o_2^{(r)}$, d'abscisse $m^{(r)}$.

On a un premier point fixe de l'application τ qui est défini par la solution optimale (g_1, g_2) où g_1 (resp. g_2) d'abscisse 6 (resp. 19) est le milieu du premier segment (resp. du second).

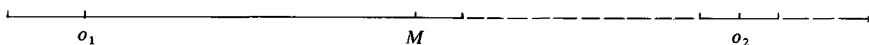
Si $(x_1, x_2) = (1, 19)$, $m = 10$ et $(x_1^{(1)}, x_2^{(1)}) = (5, 15)$ définit un point fixe de l'application τ , mais qui ne correspond pas à la solution optimale :



De même, si $(x_1, x_2) = (3, 19)$, $m = 11$ et $(x_1^{(1)}, x_2^{(1)}) = (5.5, 16.5)$ définit un point fixe de l'application τ , mais qui ne correspond pas à la solution optimale :



Un même point fixe de l'application τ peut correspondre à la limite — non nécessairement optimale — de l'algorithme; ainsi, supposons qu'on démarre avec $(x_1, x_2) = (2, 19)$ où $m = 10.5$:



on a, pour la suite des points $(o_1^{(r)}, o_2^{(r)})$ et en gardant dans les calculs deux chiffres significatifs :

$$\begin{array}{lll}
 (x_1^{(1)}, x_2^{(1)}) = (5.25, 15.68) & \text{et} & m = 10.46, \\
 (x_1^{(2)}, x_2^{(2)}) = (5.23, 15.62) & \text{et} & m = 10.43, \\
 (x_1^{(3)}, x_2^{(3)}) = (5.219, 15.58) & \text{et} & m = 10.40, \\
 (x_1^{(4)}, x_2^{(4)}) = (5.20, 15.53) & \text{et} & m = 10.365, \\
 (x_1^{(5)}, x_2^{(5)}) = (5.18, 15.48) & \text{et} & m = 10.33, \\
 (x_1^{(6)}, x_2^{(6)}) = (5.165, 15.435) & \text{et} & m = 10.30, \\
 (x_1^{(7)}, x_2^{(7)}) = (5.15, 15.40) & \text{et} & m = 10.275, \\
 (x_1^{(8)}, x_2^{(8)}) = (5.14, 15.36) & \text{et} & m = 10.25, \\
 (x_1^{(9)}, x_2^{(9)}) = (5.125, 15.325) & \text{et} & m = 10.225, \\
 (x_1^{(10)}, x_2^{(10)}) = (5.11, 15.30) & \text{et} & m = 10.20, \\
 (x_1^{(11)}, x_2^{(11)}) = (5.10, 15.26) & \text{et} & m = 10.18,
 \end{array}$$

... qui converge lentement vers le point fixe (5, 15) ci-dessus considéré.

4. CONTRIBUTION DE LA MÉTHODE DES PÔLES D'ATTRACTION

Le principe de cette méthode a pour la première fois été exposé dans [I. C. Lerman et H. Leredde (1977)]. Pour un exposé détaillé, on pourra consulter [Leredde (1979)] dont certains aspects sont repris dans [I. C. Lerman (1981), chap. 8]. Cette méthode présente différentes stratégies cohérentes pour un certain type de données, lesquelles restent très voisines dans leurs résultats lorsqu'il s'agit de déterminer seulement deux pôles entraînant une première segmentation en deux classes de l'ensemble des unités de données.

Ici nous allons considérer une stratégie cohérente avec la nature de la donnée formée d'une suite de deux intervalles supportés par un axe (cf. *fig. 1*, § 1. 2). Les notations sont les mêmes que ci-dessus et on continuera à supposer que a (longueur de l'intervalle gauche) et supérieure à b (longueur de l'intervalle droite). Le premier pôle P sera défini comme étant le point de l'axe par rapport auquel le moment d'ordre 2 est maximal. Le second pôle Q sera défini comme étant le point de l'axe pour lequel est maximal le produit de la distance au premier pôle P , par le moment d'ordre 2 à ce point Q .

4.1. Détermination du premier pôle

L'abscisse p du premier pôle d'attraction P est défini comme rendant maximale la quantité :

$$\mathcal{M}_2(P) = \int_0^a (p-x)^2 dx + \int_{a+t}^{a+b+t} (p-x)^2 dx \quad (1)$$

$$= -\frac{1}{3}[(p-a)^3 - p^3 + (p-a-b-t)^3 - (p-a-t)^3]. \quad (2)$$

En développant — après calcul — on obtient :

$$\mathcal{M}_2(P) = (ap^2 - a^2 p) + \{[p-(a+t)]^2 b - [p-(a+t)] b^2\} + \left(\frac{a^3}{3} + \frac{b^3}{3}\right), \quad (3)$$

$$= ap(p-a) + b[p-(a+t)][p-(a+b+t)] + \left(\frac{a^3}{3} + \frac{b^3}{3}\right). \quad (4)$$

Il s'agit d'un trinôme du second degré en p , dont la dérivée par rapport à p peut s'écrire :

$$\begin{aligned} a^2 + b^2 + 2a(p-a) + 2b(p-a-b-t) \\ = 2\{p(a+b) - [a^2 + 2b(a+t) + b^2]\}, \end{aligned} \quad (5)$$

qui s'annule pour :

$$p_0 = \frac{a^2 + 2b(a+t) + b^2}{2(a+b)} \quad (6)$$

Ainsi, $\mathcal{M}_2(P)$ est une fonction de p à valeurs positives [cf. (1)], décroissante pour $p \leq p_0$ et croissante pour $p \geq p_0$. Sa valeur maximale est donc atteinte soit pour $p=0$, soit pour $p=a+b+t$.

Calculons dans ces conditions $\mathcal{M}_2(P)$ pour $p=0$ — qu'on notera $\mathcal{M}_2(0)$ — et $\mathcal{M}_2(P)$ pour $p=a+b+t$ — qu'on notera — $\mathcal{M}_2(a+b+t)$:

$$\begin{aligned} \mathcal{M}_2(0) &= b^2(a+t) + b(a+t)^2 + \left(\frac{a^2}{3} + \frac{b^3}{3}\right) \\ &= b(a+t)(a+b+t) + \left(\frac{a^3}{3} + \frac{b^3}{3}\right). \end{aligned} \quad (7)$$

$$\mathcal{M}_2(a+b+t) = a(a+b+t)(b+t) + \left(\frac{a^3}{3} + \frac{b^3}{3}\right). \quad (8)$$

La différence :

$$\mathcal{M}_2(a+b+t) - \mathcal{M}_2(0) = t(a-b)(a+b+t). \quad (9)$$

Ayant supposé a supérieur à b , le maximum de $\mathcal{M}_2(P)$ est atteint pour $p = a + b + t$ où P est l'extrémité droite de l'intervalle S_2 le plus à droite.

4.2. Détermination du deuxième pôle d'attraction

La règle de choix a été ci-dessus précisée. Le deuxième pôle d'attraction Q d'abscisse q , est déterminé de façon à rendre maximale la quantité positive (qui ne s'annule que si $q = a + b + t$) :

$$\begin{aligned} & (a+b+t)-q \left[\int_0^a (q-x)^2 dx + \int_{a+t}^{a+b+t} (q-x)^2 dx \right] \quad (10) \\ & = [(a+b+t)-q] \left\{ (aq^2 - a^2 q) + [b(q-(a+t))]^2 \right. \\ & \quad \left. - b^2(q-(a+t))] + \left(\frac{a^3}{3} + \frac{b^3}{3} \right) \right\}. \quad (11) \end{aligned}$$

La dérivée par rapport à q de ce polynôme du troisième degré en q , s'écrit :

$$\begin{aligned} & [(a+b+t)-q] \{ (a^2 + b^2) + 2[a(q-a) + b(q-a-b-t)] \} \\ & - \left\{ aq(q-a) + b(q-a-t)(q-a-b-t) \right. \\ & \quad \left. + \left(\frac{a^2}{3} + \frac{b^2}{3} \right) \right\}. \quad (12) \end{aligned}$$

Pour simplifier les écritures, nous allons poser :

$$L = (a+b+t) \quad \text{et} \quad r = (L-q).$$

L'expression (12) devient :

$$\begin{aligned} & r[(a^2 + b^2) + 2a(L-a-r) - 2br] - \left\{ a(L-r)(L-a-r) \right. \\ & \quad \left. - br(b-r) + \left(\frac{a^3}{3} + \frac{b^3}{3} \right) \right\} \quad (13) \end{aligned}$$

et après mise en forme, elle s'écrit :

$$-3(a+b)r^2 + 2(2aL - a^2 + b^2)r - \left[aL(L-a) + \left(\frac{a^3}{3} + \frac{b^3}{3} \right) \right]. \quad (14)$$

Ce polynôme du second degré en r définit une fonction parabolique d'abord croissante puis décroissante. Sa valeur maximale est atteinte pour :

$$r_0 = \frac{1}{3} \left[(a+b) + \frac{2at}{(a+b)} \right]. \quad (15)$$

où la dérivée de (14) s'annule.

Nous allons à présent dégager une condition suffisante pour laquelle la valeur maximale de (14) est négative. Son expression est :

$$-\frac{[(a+b)^2 + 2at]^2}{3(a+b)} + 2\frac{[(a+b)^2 + 2at]^2}{3(a+b)} - \left[a(a+b+t)(b+t) + \left(\frac{a^3}{3} + \frac{b^3}{3} \right) \right]. \quad (16)$$

dont le signe est le même que celui de :

$$[(a+b)^2 + 2at]^2 - (a+b)[3a(a+b+t)(b+t) + a^3 + b^3]. \quad (17)$$

Le calcul détaillé de cette expression donne :

$$a^3t + a^2t^2 - a^2bt - 2ab^2t - 3abt^2, \quad (18)$$

qui est négative dès lors que :

$$a^2 + at < ab + 2b^2 + 3bt. \quad (19)$$

Si $a \leq 2b$, cette inégalité se trouve satisfaite; en effet dans ce cas,

$$a^2 + at \leq 2ab + 2bt \quad (20)$$

et l'inégalité :

$$2ab + 2bt < ab + 2b^2 + 3bt, \quad (21)$$

se réduit à :

$$a < 2b + t. \quad (22)$$

C.Q.F.D.

LEMME : Si $a > b$, le premier pôle d'attraction P est l'extrémité droite de l'intervalle le plus à droite. Si de plus $a \leq 2b$, le second pôle d'attraction Q est l'extrémité gauche de l'intervalle le plus à gauche ($q=0$).

THÉORÈME : Si $b < a \leq \sqrt{3}b$ et si $(1/2)(a-b) \leq t < b$, alors le système initial de noyaux-singletons défini par le couple des deux premiers pôles d'attraction,

conduit par l'algorithme de « reallocation-recentrage » à la solution optimale définie par le couple (g_1, g_2) des centres de gravité des deux intervalles.

Il suffit d'appliquer le théorème 2 du paragraphe 2.2 précédent et de tenir compte du lemme qu'on vient d'établir.

Il faut savoir que le nombre de pôles d'attraction à extraire peut être déterminé automatiquement dans la méthode de par la qualité de la partition obtenue par allocation, au moyen d'un critère global tel que (1) du paragraphe 1.1. Dans le cas que nous avons étudié ici, on doit pouvoir établir que la meilleure partition est autour de deux pôles.

5. CONCLUSION

Ce travail qui se situe parfaitement dans le cadre de « la reconnaissance des formes par algorithmes » [J. C. Simon (1984)] présente différents aspects.

Il met en évidence — par rapport à la situation mathématique la plus pure et la plus simple de classifiabilité — des propriétés non « triviales »; au moins dans le sens où les praticiens de ce type d'algorithmes n'en ont pas véritablement conscience.

D'autre part et de façon liée, nous montrons la vulnérabilité de l'algorithme de reallocation-recentrage par rapport au choix du système initial de noyaux formé de centres d'attraction. C'est une des raisons profondes et implicites qui ont conduit E. Diday [Diday (1972)] à considérer une répétition de l'algorithme pour arriver aux « formes fortes » qui — pour des raisons statistiques non explicites — doivent correspondre à des agrégats de relative forte cohésion. Nous montrons ici le rôle important que peuvent avoir les pôles d'attraction dans la procédure d'initialisation.

Enfin, cette étude s'inscrit dans un nouveau type de recherches où on considère une situation mathématique ou statistique de pure classification et où on étudie le comportement de l'algorithme compte tenu de son état initial. La généralisation de ce type d'études pose des problèmes difficiles où la simulation informatique pourra apporter une intéressante contribution.

BIBLIOGRAPHIE

- J. P. BENZECRI, *L'analyse des Données*, tome I. *La Taxinomie*, Dunod, Paris, 1973.
- G. CHOQUET, *Cours d'Analyse*, tome II. *Topologie*, Masson, Paris, 1964.
- E. DIDAY, *Optimisation en classification automatique et reconnaissance des formes*, Note scientifique n° 6, supplément au bulletin de l'I.R.I.A. n° 12. Paru également en 1973 dans la revue de l'A.F.I.R.O., série verte, 1972.

- E. DIDAY *et al.*, *Optimisation en classification automatique*, Publications de l'I.N.R.I.A., tomes I et II, 1980.
- J. DIEBOLT, J. C. SIMON et W. L. MIRANKER, *The Dynamic Cluster Algorithm with Continuous Data*, Research Report, RC 6743, I.B.M. Research Division, 1977.
- D. J. HALL et G. H. BALL, *Isodata a Novel Method of Data Analysis and Pattern Classification*, Technical (Report, 5 RI project 5533, Stanford Research Institute, Mento Park, California, U.S.A., 1965.
- Y. LECHEVALLIER, *Optimisation de quelques critères en classification automatique*, Thèse de 3^e cycle, Université Paris-VI, 1974.
- J. LEMAIRE, *Propriétés asymptotiques en classification (consistance des solutions de problèmes approchés)*, Rapport interne, L.A.S.S.Y.-E.R.A. n° 855, Université de Nice, 1983.
- H. LEREDDE, *La méthode des pôles d'attraction; la méthode des pôles d'agrégation : deux nouvelles familles d'algorithmes en classification automatique et sériation*, vol. 1 : méthodes et exemples réels; vol. 2 : programmes. Thèse de 3^e cycle, Université Paris-VI, 1979.
- I. C. LERMAN et H. LEREDDE, *La méthode des pôles d'attraction*, Actes du colloque : *Journées Analyse des Données et Informatique*, I.R.I.A., Versailles, septembre 1977.
- I. C. LERMAN, *Classification et analyse ordinale des données*, Dunod, Paris, 1981.
- J. B. MACQUEEN, *Some Methods for Classification and Analysis of Multivariate Observations*, Proc. of the 5th Berkeley Symposium on Mathematical Statistical and Probability, vol. 1, 1967, p. 281-297.
- D. POLLARD, *Strong Consistency of k-means Clustering*, The Annals of Statistics, vol. 9, n° 1, 1981, 135-140.
- J. C. SIMON, *La reconnaissance des formes par algorithmes*, Masson, Paris, 1984.
- R. L. THORNDIKE, *Who belongs in the Family*, Psychometrika, 18, 1953, p. 267-276.