

REVUE DE STATISTIQUE APPLIQUÉE

E. DIDAY

Introduction à l'analyse factorielle typologique

Revue de statistique appliquée, tome 22, n° 4 (1974), p. 29-38

http://www.numdam.org/item?id=RSA_1974__22_4_29_0

© Société française de statistique, 1974, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

INTRODUCTION A L'ANALYSE FACTORIELLE TYPOLOGIQUE

E. DIDAY

Ingénieur de Recherche à l'Institut de Recherche
en Informatique et Automatique (I.R.I.A.)

Il s'agit d'une méthode qui permet de généraliser l'analyse factorielle classique par la recherche simultanée de k variétés d'inertie minimum, au lieu d'une seule. On utilise l'algorithme des nuées dynamiques avec des variétés pour noyau. Avec les hypothèses introduites, il est démontré que l'algorithme converge vers une solution qui tend à minimiser les inerties locales de sous-nuages par rapport à des variétés locales. Les résultats théoriques sont illustrés par un programme et trois exemples simples d'où il ressort notamment que :

1) les plans factoriels (ou axes) locaux expliquent une part d'inertie plus importante que les plans factoriels (ou axes) classiques du nuage ;

2) pour la reconnaissance de formes les solutions obtenues sont proches de l'optimum global dans le cas de figures planes simples (caractères de l'alphabet).

PLAN

- 1/ Formalisation du problème
- 2/ L algorithme et étude de sa convergence
- 3/ Exemples d'utilisations
 - 3.1/ Le programme
 - 3.2/ Reconnaissance du "squelette" de la lettre A.

Conclusion

INTRODUCTION

Il s'agit de détecter des axes factoriels, ou des plans factoriels *locaux* qui tiennent compte des étendues locales du nuage E. Contrairement à l'analyse factorielle discriminante, l'analyse factorielle typologique ne nécessite pas de connaissances a priori sur ces agglomérats. *Son but est d'obtenir parmi toutes les*

(1) Article remis le 11/12/73. Révisé le 25/3/74.

partitions de E , celle dont la somme des moments d'inertie de chaque classe par rapport aux axes factoriels qu'elle engendre soit minimum.

1. FORMALISATION DU PROBLEME

Nous supposons $E \subset \mathbb{R}^p$ et muni d'une mesure $m : E \rightarrow \mathbb{R}^+$. Nous noterons x le vecteur ayant pour origine l'origine des axes de \mathbb{R}^p et pour extrémité l'élément x de E . Le barycentre de $A \subset E$ sera noté $G(A)$. La mesure de ressemblance de deux éléments de E sera définie par une distance d telle que $d(x,y) = Q(x-y, x-y)$ où Q est une forme quadratique. Soit q l'application linéaire $E \rightarrow E$ associée à Q par la relation $Q(x,y) = q(x) \cdot y$.

Posons $L = P(E) \times \mathcal{V}$ où \mathcal{V} est l'espace des variétés affines de \mathbb{R}^p et $P(E)$ est l'ensemble des parties de E . Nous supposons d'abord pour simplifier que ces variétés sont du premier ordre. (la généralisation est immédiate).

L'algorithme des nuées dynamiques (c.f. [2] et [3]) va être utilisé avec les applications D , R , f et g suivantes :

$$- D : E \times L \rightarrow \mathbb{R}^+$$

est définie par

$$D(x,\lambda) = m(x) [\vec{U}_\Delta \cdot (X - \vec{G}(A))]^2$$

où $\lambda = (A, \Delta)$ et \vec{U}_Δ est un vecteur unitaire du plan (Δ, x) et perpendiculaire à Δ (on a donc $q(\Delta) \cdot v_\Delta = 0$) ; on note Δ (resp. Δ_i) un vecteur unitaire de Δ (resp. Δ_i). On voit que $D(x, \lambda)$ exprime le moment d'inertie du vecteur $x - G(A)$ par rapport à une droite parallèle à Δ et passant par le barycentre de A .

- P_k est l'ensemble des partitions de E contenant au maximum k classes.

- T est l'ensemble des entiers compris entre 1 et k

- $L_k = L = (L_1, \dots, L_k) / L_i = (A, \Delta) \in P(E) \times \mathcal{V}$ autrement dit $L_k = (P(E) \times \mathcal{V})^k$

- $R : L_k \times T \times P_k \rightarrow \mathbb{R}^+$ où

$$R(L, i, P) = \sum_{x \in P_i} D(x, L_i) = \sum_{x \in P_i} m(x) u_{\Delta_i} \cdot ((x - G(A_i))^2)$$

si $L_i = (A_i, \Delta_i)$ et $P = (P_1, \dots, P_i, \dots, P_k)$

$R(L, i, P)$ exprime l'inertie de la classe P_i par rapport à l'axe Δ_i .

- f est l'application $L_k \rightarrow P_k$ telle que $f(L) = \Pi$ si $L = (L_1, \dots, L_k)$.

avec

$$L_i = (P_i, \Delta_i), \Pi = (\Pi_1, \dots, \Pi_k)$$

et

$$\Pi_i = \{x \in E / D(x, L_i) \leq D(x, L_j)\}$$

(en cas d'égalité on affecte x à la classe de plus petit indice).

– g est l'application $\mathbf{P}_k \rightarrow \mathbf{L}_k$ telle que $g(\mathbf{P}) = \mathbf{L}$ avec

$\mathbf{L} = (\mathbf{L}_1, \dots, \mathbf{L}_k)$ et $\mathbf{L}_i = (\mathbf{P}_i, \Delta_i)$ telle que Δ_i minimise $R((\mathbf{P}_i, \cdot), i, \mathbf{P})$ parmi toutes les variétés affines d'ordre 1 de \mathbf{R}^p . Autrement dit Δ_i maximise

$$J(\Delta_i) = \sum_{x \in \mathbf{P}_i} m(x) [q(\Delta_i) \cdot (x - G(\mathbf{P}_i))]^2 ;$$

il en résulte la proposition suivante :

Proposition 1 :

Δ_i est le vecteur propre de plus grande valeur propre de $\mu \circ q$, où μ est la forme quadratique associée à la matrice d'inertie *M de \mathbf{P}_i .

En effet : maximiser $J(\Delta_i)$ revient à maximiser

$$\sum_{x \in \mathbf{P}_i} m(x) \bar{\Delta}_i' \tilde{Q}z \cdot \bar{\Delta}_i \tilde{Q}z$$

où $\bar{\Delta}_i$ et z sont les vecteurs colonnes respectivement associés à $\vec{\Delta}_i$ et $\vec{z} = \vec{x} - \vec{G}(\mathbf{P}_i)$ dans \mathbf{R}^p ; On note \tilde{Q} la matrice associée à la forme quadratique Q . D'où

$$J(\Delta_i) = \sum_{x \in \mathbf{P}_i} m(x) \bar{\Delta}_i' \tilde{Q}z \cdot z' Q \Delta_i,$$

ce qui implique

$$J(\Delta_i) = \bar{\Delta}_i' \tilde{Q}M\tilde{Q} \bar{\Delta}_i = q(\vec{\Delta}_i) \mu \circ q(\vec{\Delta}_i).$$

Afin de simplifier les notations posons $v = \Delta_i$. Le problème se ramène à maximiser $q(v) \cdot \mu \circ q(v)$ sachant que $q(v) \cdot v = 1$. Ou encore à

Maximiser $Q(v, \mu \circ q(v))$

sous la contrainte $Q(v, v) = 1$

Des résultats classiques d'analyse fonctionnelle (c.f. [4] et [5]) permettent alors d'affirmer :

1) $\tilde{Q}^{-1} \tilde{Q} M \tilde{Q}' v^* = J(v^*) v^*$.

2) $\exists v^* : \text{Max } J(v) = J(v^*)$.

D'après 1), $J(v^*)$ est une valeur propre de $M Q'$ d'après 2) $J(v^*)$ est la plus grande des valeurs propres.

-- Le critère W que l'on cherche à optimiser est l'application $\mathbf{V} = \mathbf{L}_k \times \mathbf{P}_k \rightarrow \mathbf{R}^+$ telle que

$$W(v) = \sum_{i=1}^k R(\mathbf{L}_i, i, \mathbf{P}_i) = \sum_{i=1}^k \sum_{x \in \mathbf{P}_i} m(x) [\vec{U} \Delta_i (\vec{x} - \vec{G}(\mathbf{A}_i))]^2$$

$$\mathbf{L} = (\mathbf{L}_1, \dots, \mathbf{L}_k) \text{ avec } \mathbf{L}_i = (\mathbf{A}_i, \Delta_i).$$

* On a $M(j, 1) = \sum_{x \in \mathbf{P}_i} (m(x) (x^j - x_G^j) (x^1 - x_G^1))$, où x^j est la coordonnée j de \vec{x} .

Si $V = (L, P)$ avec $L = (L_1, \dots, L_k)$ où $L_i = (P_i, \Delta_i)$ on voit que minimiser W revient à chercher parmi tous les couples (L, P) celui qui minimise le moment d'inertie de chaque classe P_i de la partition P par rapport à l'axe Δ_i .

Les suites v_n et u_n que nous allons maintenant définir permettent une bonne approche de ce problème.

II – L'ALGORITHME ET ETUDE DE SA CONVERGENCE

Soit l'application $h : V_k \rightarrow V_k$ telle que

$$v = (L, P) \in V_k \Rightarrow h(v) = (g(f(L)), f(L)).$$

La suite v_n est définie à partir de

$$v_0 = (L^{(0)}, P^{(0)}) \text{ où } L^{(0)} = g(P^{(0)})$$

par

$$v_{n+1} = h(v_n) \quad \forall n \geq 0.$$

La suite u_n est définie par $u_n = W(v_n)$.

Il n'est pas possible d'appliquer ici, directement les résultats obtenus pour l'algorithme de base, car les applications D et R ne sont pas définies sur les mêmes espaces.

L'application $S : L_k \times L_k \rightarrow R^+$ est définie ici par

$$S(L, M) = \sum_{i=1}^k R(L, i, f(M)) = \sum_{i=1}^k \sum_{x \in \pi_i} D(x, L_i)$$

où

$$f(M) = \Pi = (\Pi_1, \dots, \Pi_k) \text{ et } L_i = (A_i, \Delta_i) \in P(E) \times \mathfrak{A}$$

Rappelons (c.f. [2]) que R est dite carrée si

$$S(L, M) \leq S(M, M) \Rightarrow S(L, L) \leq S(L, M)$$

Proposition 2

L'application R est carrée.

Démonstration

$$S(L, L) = \sum_{i=1}^k R(L_i, i, f(L))$$

où

$$f(L) = P = (P_1, \dots, P_k) \in P_k$$

et

$$L = (L_1, \dots, L_k) \in L_k.$$

On a donc :

$$S(L, M) = \sum_{i=1}^k \sum_{x \in \pi_i} D(x, L_i) \quad (1)$$

$$S(L, L) = \sum_{i=1}^k \sum_{x \in P_i} D(x, L_i) \quad (2)$$

Le même élément x peut apparaître dans (1) sous la forme $D(x, L_i)$ et dans (2) sous la forme $D(x, L_j)$. Or par construction même de P_i , on a nécessairement $D(y, A_i) \leq D(y, A_j)$. Comme chaque élément x apparaît une fois et une seule dans (1) et (2) puisque P et Q sont des partitions de E , on peut faire le même raisonnement $\forall x \in E$, d'où $S(L, L) \leq S(L, M)$. L'application P est donc bien carrée.

Proposition 3

La suite u_n converge en décroissant.

Démonstration

Soit

$$v_n = (L^{(n)}, P^{(n)}) \text{ alors } W(v_n) = \sum_{i=1}^k R(L^{(n)}, i, P^{(n)}).$$

On a

$$v_{n+1} = h(v_n) = (g(f(L^{(n)})), f(L^{(n)})) = (L^{(n+1)}, P^{(n+1)}) \Rightarrow$$

$$L^{(n+1)} = g(P^{(n+1)}) = (L_1^{(n+1)}, L_2^{(n+1)}, \dots, L_k^{(n+1)})$$

où

$$L_i^{(n+1)} = (P_i^{(n+1)}, \Delta_i^{(n+1)}).$$

D'autre part : $S(L^{(n)}, L^{(n)}) < S(L^{(n)}, L^{(n-1)})$ d'après la proposition 2 d'où

$$\sum_{i=1}^k R(L^{(n)}, i, P^{(n+1)}) \leq \sum_{i=1}^k R(L^{(n)}, i, P^{(n)})$$

car R est carrée.

On a de plus

$$u_{n+1} = \sum_{i=1}^k R(L^{(n+1)}, i, P^{(n+1)}) \sum_{i=1}^k R(L^{(n)}, i, P^{(n+1)})$$

en effet :

$$R(L^{(n+1)}, i, P^{(n+1)}) = \sum_{x \in P_i} (n+1)^{m(x)} [\vec{U}_{\Delta_i^{(n+1)}}(\vec{x} - \vec{G}(P_i^{(n+1)}))]^2$$

$$R(L^{(n)}, i, P^{(n+1)}) = \sum_{x \in P_i} (n+1)^{m(x)} [\vec{U}_{\Delta_i^{(n)}}(\vec{x} - \vec{G}(P_i^{(n)}))]^2$$

d'où l'inégalité, puisque $\Delta_i^{(n+1)}$ est l'axe d'inertie de $P_i^{(n+1)}$ ce qui n'est pas généralement le cas de $\Delta_i^{(n)}$

Remarque :

On peut facilement généraliser l'algorithme et la proposition au cas de variétés affines d'ordre plus grand que 1. (Par exemple pour les variétés affines d'ordre plus grand que 1. (Par exemple pour les variétés affines d'ordre 2, les Δ_i

seront des plans d'inertie et on sera conduit à calculer les deux vecteurs propres de plus grande valeur propre associés à A_i).

III – EXEMPLES D'UTILISATION (réalisés par Mme OK-SAKUN)

3.1 – Le programme *

En entrée

- a) Le tableau des données.
- b) Le nombre maximum de classes désirées : k .
- c) La dimension des variétés affines recherchées : $p = 1, 2, 3$ (droites, plan, . . .)

La partition de départ est tirée au hasard.

En sortie

- a) La partition obtenue à la convergence.
- b) Pour chaque classe
 - les p axes de la variété cherchée
 - l'inertie expliquée par ces p axes.
- c) Une mesure de la qualité de la partition obtenue (critère)
- d) Une visualisation graphique des résultats.
 - pour donner une idée des performances du programme expérimental.

– étant donné une population de 394 individus caractérisés par 16 paramètres, le nombre de classes demandé étant 4, le programme a nécessité pour la recherche de plans factoriels locaux et l'impression graphique de 5 plans factoriels (plan global et 4 plans locaux) un temps de passage total de 6 minutes sur 10070 CII.

– L'analyse factorielle globale de la même population avec impression de 3 graphiques seulement nécessite sur 10070 CII 3.60 minutes.

3.2 – Reconnaissance du squelette de la lettre A

L'algorithme a été appliqué de la façon suivante :

– E est un nuage de 119 points de \mathbf{R}^2 dont la configuration dans le plan est celle de la lettre A (fig. 1). E est muni d'un système de masses uniforme : $\forall x \in E, \mu(x) = 1/\text{card}(E)$.

– La métrique dont on munit \mathbf{R}^2 est la métrique euclidienne usuelle.

– \mathcal{V} est l'ensemble des droites du plan

– k est le nombre de classes demandées (fixé a priori). L'algorithme a été appliquée de la manière suivante :

1) avec $k = 3$ (k optimal étant donné la configuration de l'ensemble de 119 points)

* Ce programme peut être obtenu sur simple demande à l'I.R.I.A. durant l'année suivant la date de parution de l'article.

		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	
	24 25 27 29 30 32 35 36 38 40 41 43 46 47 49	26 28 31 33 34 37 39 42 44 45 48 50 51 52 53	
	54 55 62 63 64 65 69 70 71 72 73 74 76 77 78 80 81 83 85 86 89 91 92 94	56 57 58 59 60 61 66 67 68 75 79 82 84 87 88 90 93 94	
96 97 99 102 103 105 106 108 111 112 114 116 117		98 100 101 104 107 109 110 113 115 118 119	

Figure 1

2) avec $k = 6$

3) avec $k = 2$

— la visualisation des résultats consiste

a) en une représentation dans le 1er plan factoriel de l'analyse globale (analyse en composantes principales du nuage de 119 points) de l'ensemble des points désignés chacun par le numéro de la classe à laquelle ils appartiennent.

b) en une projection sur ce plan du centre de gravité de chaque classe P_i et de l'axe Δ_i qui lui est associé à la convergence de l'analyse factorielle typologique.

Dans le 1er cas (nombre de classes fixé à 3, on a obtenu 3 axes reproduisant exactement la configuration de la lettre A (fig. 2) ; mais il faut préciser que ce résultat n'est pas obtenu quel que soit le tirage aléatoire de départ et que l'existence de "noeuds" rend la convergence très lente (de l'ordre de 10–12 itérations et même jusqu'à 20 pour un des tirages).

Toutefois, le choix que l'on peut faire parmi les résultats des différents tirages. d'après la valeur du critère à optimiser, correspond toujours à la partition dont les axes reconstituent la lettre A.

Le critère permet de juger de la qualité relative des différents résultats.

D'autre part quelques modifications de l'algorithme de base ont permis *dans ce cas particulier* d'obtenir les "bons" axes quel que soit le tirage aléatoire de départ (pour 8 tirages consécutifs) et avec un nombre réduit d'itérations (en 6 itérations (3 fois), en 7 itérations (2 fois), en 9 et en 11 itérations (1 fois)), et des deux premiers axes de chacune des classes obtenues à la convergence de l'algorithme.

Dans le second cas (nombre de classes fixé à 6). On a obtenu suivant les tirages 5 ou 6 axes (le nombre de classes peut en effet diminuer). Parmi les axes obtenus pour 6 tirages, on retrouve toujours 2 axes de A, le 3ème axe (la barre) n'apparaissant que dans 3 tirages, les autres variant selon les tirages.

Dans le troisième cas (nombre de classes fixé à 2). On retrouve les axes 1 et 2 du A, légèrement déplacés vers l'intérieur, les points et la "barre" étant également répartis entre les deux classes.

L'inertie totale de chacune des 2 classes est expliquée à plus de 98 % par son premier axe, alors que l'axe 1 de l'analyse factorielle globale explique seulement 66 % de l'inertie du nuage.

CONCLUSION

Sur le plan de la recherche les perspectives offertes par l'analyse factorielle typologique sont multiples : il s'agit d'étendre les résultats obtenus au cas de métriques locales (par exemple avec la distance du χ^2 on doit obtenir des analyses factorielles des correspondances locales à la convergence de l'algorithme améliorée). On peut définir la notion d'axes "forts" et "faibles" et en déduire des modèles linéaires locaux ou des liaisons linéaires particulières pour les classes ayant une très forte inertie axiale.

Analyse factorielle typologique, données : 119 pts à
2 coordonnées
(Lettre A)

Algorithme appliqué en demandant 3 classes

Visualisation des résultats

on a projeté sur le 1er plan factoriel de l'analyse
globale (A.C.P.)

- Les individus désignés par leur numéro de classe
- Les centres de gravité des classes
- Les deux premiers axes de chaque classe.

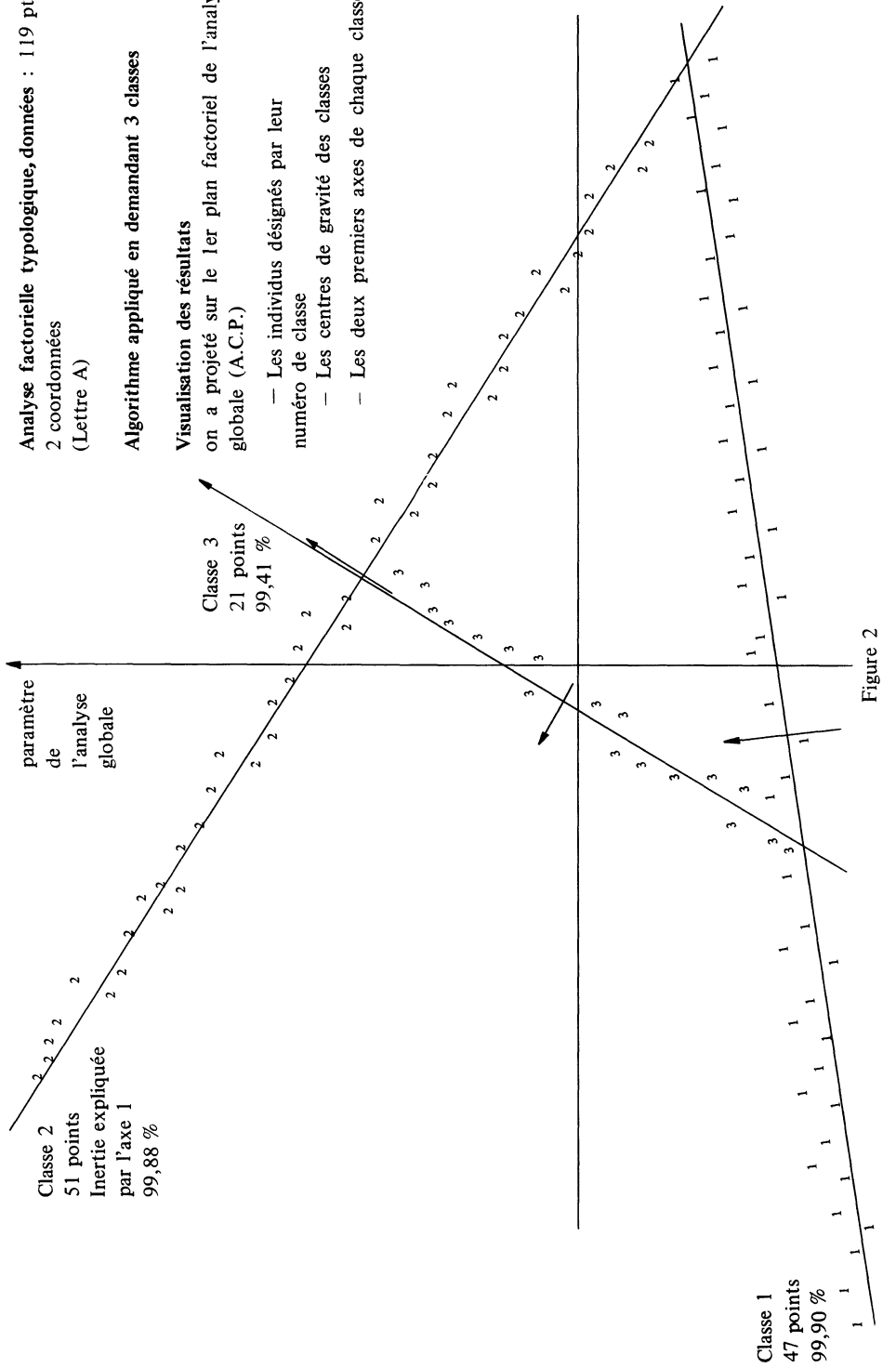


Figure 2

Sur le plan pratique les résultats déjà obtenus devraient permettre de détecter des tendances locales et donner ainsi la possibilité d'étudier leur évolution (dans le cas de données dépendant du temps par exemple). Ils devraient permettre également l'obtention de variétés avec forte inertie locale (ce qui peut fournir une visualisation par une sorte de "grossissement à la loupe", intéressante dans le cas de grands tableaux de données).

BIBLIOGRAPHIE

- [1] BENZECRI J.P. – "Analyse factorielle des correspondances" Publications du Lab. de Stat. Math. de l'Université de Paris VI.
- [2] DIDAY E. (1973) – "The dynamic cluster Method in non-hierarchical clustering" vol. 2 – n° 1.
- [3] DIDAY E. (1973) – "Optimisation en classification automatique et reconnaissance des formes" – RIRO.
- [4] DHUIME G. – "Analyse factorielle" – *cahier IRIA* n° 5,
- [5] RIESZ et NAGY (1968) – "Leçons d'analyse fonctionnelle" Gauthiers-Villars – Paris.
- [6] SANDOR G., DIDAY E., LECHEVALLIER Y., BARRE J. (1972) – "Une étude des corrélations entre les modifications des protéines sériques en pathologie humaine" – *C.R. Acad. Sc. Paris*, t 274, pp 331–334.