50376 1987 71



# THESE

Nº d'ordre 734

présentée à

### L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

pour l'obtention du grade de

Docteur ès-Sciences Physiques

par

Abderrahmane TOUZANI



CLASSIFICATION AUTOMATIQUE PAR DETECTION DES CONTOURS DES MODES DES FONCTIONS DE DENSITE DE PROBABILITE MULTIVARIABLES ET ETIQUETAGE PROBABILISTE

Soutenue le 26 juin 1987 devant la commission d'examen :

MM.	P.	VIDAL
	J.G.	POSTAIRE
	B.	DUBUISSON
	J.	KITTLER
	Ρ.	DEVIJVER
	M.	STAROSWIECKI

Président Rapporteur et Directeur de Recherche Rapporteur Examinateur Examinateur

### SOMMAIRE

# SECTION I

# METHODES DE CLASSIFICATION STATISTIQUES ET CONTOURS DES FONCTIONS DE DENSITE

### **CHAPITRE I : LA CLASSIFICATION AUTOMATIQUE**

- I.1. Perception et classification automatique
- I.2. Automatisation du processus de classification
- I.3. L'approche métrique
- I.4. L'approche statistique

I.4.1 Analyse des melanges et classification

I.4.2 Recherche des groupements

I.5 Contours des modes et classification automatique

- Références I

# CHAPITRE II : IDENTIFICATION DES FONCTIONS DE DENSITE NORMALES MULTIVARIABLES PAR ANALYSE DE LEUR CONTOUR

- II.1. Le choix du modéle normal
- II.2. Les fonctions de densité de probabilité normales
- II.3. Calcul du gradient d'une fonction de densité normale

II.3.1. Expression réduite d'une fonction de densité normale

II.3.2. Détermination de la norme du gradient

II.4. Contour d'une fonction de densité multivariable

II.4.1. Analyse des extréma de la norme

II.4.2. Définition et propriétés du contour d'une fonction de densité normale

II.5. Analyse géométrique du contour d'une fonction de densité

**II.6.** Conclusion

- Références II

### SECTION II

# EXTRACTION DES CONTOURS ET CLASSIFICATION AUTOMATIQUE

CHAPITRE III : ESTIMATION ET FILTRAGE DES FONCTIONS DE DENSITE DE PROBABILITE

**III.1.** Introduction

III.2. Estimation non paramétrique

III.2.1. Principe

III.2.2. Comportement asymptotique

III.2.3. Méthode du noyau

III.2.4. Algorithme rapide d'estimation des fonctions de densite de probabilite

III.2.4.(a) Position du problème

III.2.4.(b) Principe

III.3. Filtrage non linéaire

III.3.1. Principe du filtre médian

III.3.2. Analyse statistique

III.3.3. Principe de programmation

III.3.4 Résultats expérimentaux

**III.4.** Conclusion

- Références III

### CHAPITRE IV : DETECTION ET EXTRACTION DES CONTOURS DES FONCTIONS DE DENSITE DE PROBABILITE

**IV.1.** Introduction

IV.2. Detection des contours

IV.2.1. Principe

IV.2.2. Généralisation de l'opérateur de Roberts

IV.2.3. Généralisation de l'opérateur de Prewitt

IV.3. Comportement des opérateurs différentiels multidimensionnels

IV.4. Extraction des contours

IV.4.1. Principe d'extraction

IV.4.2. Exemple

**IV.5.** Conclusion

- Références IV

CHAPITRE V : APPLICATION DES TECHNIQUES DE RECHERCHE DES CONTOURS DES MODES EN CLASSIFICATION AUTOMATIQUE

V.1. Introduction

۰.

V.A. Classification paramétrique

V.A.1. Determination des paramétres d'une distribution normale multidimensionnelle

V.A.1.1. Localisation du contour d'une fonction de densité normale

V.A.1.2. Modélisation du contour

V.A.1.3. Exemples d'application

V.A.2. Identification des mélanges gaussiens

V.A.3. Résultats expérimentaux

V.B. Classification non paramétrique

V.B.1. Recherche des groupements par détection des contours des modes

V.B.2. Résultats expérimentaux

V.2. Conclusion

- Références V

# SECTION III

# APPLICATION DES TECHNIQUES D'ETIQUETAGE PROBABILISTE A LA RECHERCHE DES GROUPEMENTS

# CAPITRE VI : EXTRACTION DES CONTOURS PAR ETIQUETAGE PROBABILISTE ITERATIF

VI.1. Introduction

VI.2. Etiquetage initial

VI.3. Coefficients de compatibilité

VI.4. Etiquetage probabiliste itératif

VI.5. Résultats expérimentaux

VI.6. Conclusion

- Références

### CHAPITRE VII : DETECTION DIRECTE DES MODES PAR ETIQUETAGE PROBABILISTE ITERATIF

VII.1. Introduction

- VII.2. Détermination des probabilités initiales
- VII.3. Coefficients de compatibilité et étiquetage itératif
- VII.4. Résultats expérimentaux

VII.5. Conclusion



- Références

١

### CONCLUSION GENERALE

## Références (conclusion)

# SECTION I

# METHODES DE CLASSIFICATION STATISTIQUES ET CONTOURS DES FONCTIONS DE DENSITE

### CHAPITRE I

# LA CLASSIFICATION STATISTIQUE

### **I.1. PERCEPTION ET CLASSIFICATION AUTOMATIQUE**

L'histoire des civilisations montre que, de tout temps, l'homme a essayé de se faire aider par des machines. Celles-ci lui permettent de travailler plus vite et mieux, tout en limitant les efforts nécessaires. L'automatisation des tâches manuelles a réellement commencé avec l'apparition de la machine à vapeur et atteint aujourd'hui son point culminant avec l'utilisation des robots. L'automatisation des tâches intellectuelles est beaucoup plus récente et son histoire est étroitement liée à celle de l'informatique.

Parmi les facultés intellectuelles de l'homme, la perception, c'est-àdire l'acquisition, l'évaluation et l'interprétation de données sensorielles, est certainement l'une des plus importantes. Tout être vivant évolué est en fait capable de percevoir une grande quantité d'informations sur son environnement et de les interpréter pour agir, réagir et finalement survivre. Malgré la simplicité apparente de cette démarche intellectuelle, il est actuellement impossible de connaître les processus mis en jeu au niveau du système nerveux. Il est cependant possible d'aboutir à des résultats très semblables par des voies tout à fait différentes des voies biologiques : les nombreuses recherches dans le domaine de la reconnaissance des formes n'ont pour principale motivation que la volonté de donner à des machines des capacités de perception aussi semblables que possibles à celles de l'être humain.

La reconnaissance des formes, prise dans son sens le plus large, est un vaste domaine d'activité que l'on ne peut aborder sans préciser quelques concepts mathématiques simples. On supposera d'abord que tout "objet" ou tout "événement" peut être décrit par un nombre fini N de variables  $x_1, x_2, ..., x_n, ..., x_N$  dont les valeurs peuvent être connues par le biais de capteurs ou d'instruments de mesure. L'ensemble de ces variables, ou attributs, constitue une observation de l'objet considéré qui peut être représentée par un vecteur :

 $X = \begin{bmatrix} x_1, x_2, \dots, x_n, \dots, x_N \end{bmatrix}^T$ dans un espace euclidien à N dimensions.

La reconnaissance des formes est constituée d'un ensemble de disciplines fortement interdépendantes portant principalement sur le traitement de l'information et la théorie de la décision. Ces disciplines, très liées à l'automatique sur le plan purement scientifique et à l'informatique industrielle sur le plan pratique, débouchent sur un ensemble de techniques qui permettent un traitement et une interprétation automatiques des observations issues de l'environnement étudié. Parmi ces techniques, nous nous intéressons, dans ce mémoire, à la classification automatique dont l'objet premier et de définir des groupements (ou classes) dans un ensemble d'observations inconnues a priori. Son but principal est donc de condenser les informations multiples observées sur un ensemble d'objets ou d'événements, la description complexe et détaillée de chacun d'eux étant remplacée par son appartenance à une classe bien définie. La classification automatique apparaît donc comme une technique fondamentale pour quiconque désire donner à la machine des facultés de perception. De la prolifération des détails recueillis sous la forme d'observations multidimensionnelles, elle permet de déboucher sur une représentation structurée et conceptualisée de l'environnement étudié, chaque classe correspondant à un type bien défini d'objet ou d'événement.

#### **1.2. AUTOMATISATION DU PROCESSUS DE CLASSIFICATION**

Avant l'apparition des ordinateurs, les techniques permettant de découvrir les propriétés structurales d'un ensemble d'observations faisaient surtout appel aux capacités de perception du système visuel humain. En effet, l'observateur n'a aucun problème pour discerner des groupements de points, pourvu qu'ils soient disposés dans un espace à deux ou trois dimensions. Cependant, les observations recueillies sur les objets que l'on étudie sont généralement plus nombreuses, ce qui rend l'examen visuel impossible. Au début du XXème siècle apparaissent, sous l'impulsion des psychologues /1/ /2/, des techniques qui visent à transformer l'espace de représentation des observations en une image bidimensionnelle. Des procédures classiques telles que l'analyse en composantes principales /3/ ou l'analyse factorielle des correspondances /4/ /5/, visent à projeter les données de dimension élevée sur un espace bi ou tridimensionnel. Il existe également d'autres méthodes qui permettent une réduction substancielle de la dimension de l'espace de représentation, tout en conservant la structure des données /6/ /7/ /8/. Ces méthodes, qui nécessitent finalement un examen visuel pour analyser les données, ne sont pas, au sens strict du terme, des méthodes de classification automatique.

En fait, il faudra attendre l'essor des calculateurs numériques pour qu'apparaissent des méthodes d'exploration multidimensionnelle entièrement automatisées. Les travaux de Mahalanobis /9/, comme ceux de Fisher /10/, qui introduisit le premier la notion de fonction discriminante, restèrent longtemps sans retombées pratiques, faute de moyens de calcul et de mémorisation des données.

Aujourd'hui, on assiste à une pénétration sans cesse croissante des techniques de classification automatique dans les différents secteurs de l'activité scientifique, industrielle et économique, essentiellement favorisée par la diffusion des moyens informatiques.

Cependant, la classification automatique apparaît encore trop souvent comme un arsenal de techniques qui n'ont en commun que leur finalité et qui font appel à autant de notions mathématiques et de concepts scientifiques qu'il existe d'algorithmes. L'utilisateur reste souvent perplexe devant le foisonnement des techniques qui permettent d'identifier les classes en présence dans un ensemble de données multidimensionnelles. Selon le nombre d'observations à classer et la dimension des données, selon la possibilité d'utiliser un modèle paramètrique et compte-tenu des informations dont il dispose a priori, l'analyste doit choisir une stratégie parmi un ensemble de méthodes très diverses et souvent peu comparables, chacune ayant ses points forts et ses faiblesses. Ce manque de cohérence et d'unité apparaît surtout en classification non supervisée, c'est-à-dire lorsqu'il s'agit d'identifier les classes en présence dans un échantillon à partir de la seule information qui peut être extraite des observations à classer. Ce type de situation correspond à des démarches exploratoires, où on ne dispose d'aucune information a priori sur les données à classer, ne serait-ce que sous la forme de quelques prototypes. Dans certaines situations, il arrive même que l'analyste ne connaisse pas le nombre de classes en présence.

Il est toutefois possible de dégager deux grandes approches au problème de la classification non supervisée : l'approche métrique et l'approche statistique.

#### I.3. L'APPROCHE METRIQUE

Le problème de la classification automatique peut être abordé comme la recherche d'une partition de l'ensemble des observations telle que chaque individu de l'échantillon analysé ressemble plus aux individus de sa classe qu'à ceux des autres classes. Les méthodes permettant de résoudre ce problème nécessitent généralement la définition d'une métrique sur l'espace de représentation des données. En effet, la mesure la plus naturelle de la ressemblance entre deux individus est la distance séparant les deux observations correspondantes /11/.

Certaines procédures basées sur cette notion de distance sont essentiellement intuitives /12/, mais la plupart font appel à des critères algèbriques pour mesurer la qualité des groupements d'observations obtenus /13/ /14/. Le problème est alors d'extrémiser le critère. Certains critères, de type quadratique, mesurent la dispersion intragroupe /15/ /16/ /17/ /18/ /19/. D'autres sont calculés à partir des matrices de covariance ou de dispersion /20/ /21/ /22/ /23/.

Les méthodes hiérarchiques opèrent par séparations ou groupements successifs des observations pour aboutir à des partitions significatives /4/ /24/ /25/ /26/. Par ailleurs, presque toutes ces méthodes supposent connu le nombre de classes en présence et les résultats dépendent du nombre de groupements demandés par l'analyste. Lorsqu'on analyse un échantillon inconnu, ce choix est généralement difficile à justifier /32//33/.

Notons enfin que, pour certaines de ces méthodes, les résultats sont conditionnés par l'ordre dans lequel les observations sont prises en considération /14//16/.

En utilisant des méthodes de classification basées sur des notions métriques, les différentes classes mises en évidence sont définies par l'ensemble des observations qui les composent. On ne dispose donc d'aucune information de nature statistique sur la distribution des observations assignées aux différentes classes, ni d'aucune information sur les éventuelles erreurs de classification associées à la méthode utilisée. Il semble pourtant évident que seule une procédure de classification qui assigne toute observation à la classe pour laquelle la probabilité d'erreur est minimale peut être considérée comme optimale, en ce sens qu'aucune autre stratégie ne pourrait donner un taux d'erreur plus petit.

L'ensemble de ces critiques constitue la principale raison de notre attachement à l'approche statistique, présentée au paragraphe suivant.

#### I.4. L'APPROCHE STATISTIQUE

Les inconvénients et les limitations liés aux approches métriques peuvent être évités en faisant appel à des concepts de distance. On ne peut à proprement parler de méthode statistique que lorsqu'il est fait explicitement appel à la notion de fonction de densité de probabilité. Si on connaît la classe à laquelle appartient un objet, l'observation qui lui est associée peut être considérée comme une réalisation particulière d'une variable aléatoire dont les propriétés statistiques sont caractéristiques de la classe considérée. La théorie des graphes permet également de formaliser l'analyse des ressemblances entre les observations et de découvrir des groupements significatifs dans les échantillons analysés /27/ /28/.

Ces méthodes, à l'exception d'une seule fondée sur l'utilisation d'un critère invariant /29/, sont sensibles à des changements d'échelle sur les attributs. La figure I.1. illustre ce phénomène dont on tente de minimiser les effets en normalisant les données avant leur traitement /30//31/.



Dans ces conditions, la distribution des observations pour des objets tirés aléatoirement et indépendamment de l'ensemble des classes constituant l'échantillon étudié est un mélange pondéré des lois de probabilité relatives à ces différentes classes. Le coefficient de pondération de chaque loi est la probabilité a priori de la classe correspondante.

Les méthodes de classification statistiques consistent à analyser la fonction de densité sous-jacente à la distribution des observations disponibles pour extraire toute l'information nécessaire à leur classification. Selon que l'on accepte ou non l'hypothèse paramètrique, c'est-à-dire l'existence d'un modèle analytique pour les fonctions de densité, on débouche sur des méthodes d'analyse de mélanges ou des méthodes de recherche des groupements.

#### I.4.1. Analyse des mélanges et classification

La connaissance du nombre de classes en présence dans un échantillon et, pour chacune d'elles, de sa fonction de densité de probabilité et de sa probabilité a priori permet de connaître la probabilité d'erreur associée au classement d'une observation donnée dans chacune des classes /34/. La règle de classement qui assigne toute observation à la classe pour laquelle la probabilité d'erreur est minimale est désignée sous le nom de procédure de classement optimale /35/.

Dans la pratique, les données nécessaires pour calculer les probabilités d'erreur de classement d'une observation dans les différentes classes ne sont pas, en général, disponibles. Dans certain cas, on ne connaît ni le nombre de classes en présence, ni la fonction de densité de probabilité, ni la probabilité a priori associées à chacune d'elles. On peut toutefois compenser ce manque de connaissances sur le mélange par les informations apportées par les observations à classer elles-mêmes. En supposant que les fonctions de densité des différentes classes appartiennent à un ensemble de fonctions représentables par quelques paramètres, soit en d'autres termes, sous l'hypothèse paramètrique, le problème de l'optimisation du processus de classification se trouve posé en termes d'analyse des mélanges. Il existe de nombreuses techniques d'analyse de mélanges monovariables /36/ /37/ /38/ /39/ 40/ 41/ /42/, mais aucune d'entre elles n'a pu être étendue au cas multivariable.

Daly /43/, puis Hillborn et Lainiotis /44/ ont utilisé des techniques d'apprentissage bayesien qui nécessitent la connaissance a priori du nombre de classes présentes dans les mélanges à identifier et conduisent à des calculs prohibitifs. La solution proposée par Makov et Smith /45/, plus réaliste, est malheureusement limitée à des problèmes à deux classes.

Day /46/ étend les méthodes d'estimation par maximum de vraisemblance au cas multidimensionnel, mais ne traite, sur le plan pratique, que le cas de classes ayant mêmes matrices de covariance.

Wolfe /33/, dans le même esprit, se limite au cas de matrices de covariance égales ou diagonales. La méthode de Shroeder /47/, proche de celle des nuées dynamiques /48/, s'applique à n'importe quel mélange de lois de même type, mais nécessite la connaissance a priori du nombre de composantes du mélange. Kazakos /49/ utilise également la méthode du maximum de vraisemblance, mais ne détermine que les probabilités a priori des différentes classes.

Chien et Fu /50/, puis Young et Coraluppi /51/, ont attaché leurs noms aux techniques d'apprentissage stochastique dont l'utilisation semble pratiquement limitée au cas monovariable. Seuls des problèmes multivariables à deux classes, avec matrices de covariance ou vecteurs moyenne connus a priori, ont pu être abordés par Mizoguchi et Shimura/52/.

La méthode proposée par Cooper et Cooper /53/, puis améliorée par Cooper /54/, ne permet de traiter que le cas de composantes qui ne diffèrent que par leurs vecteurs-moyenne. L'utilisation d'histogrammes, et plus particulièrement de quartiles, proposée par Patrick et Hancock /55/ semble réservée, au niveau pratique, au cas monovariable.

Postaire et Vasseur /56/ /57/ proposent d'analyser la convexité de la fonction de densité sous-jacente pour approcher tous les paramètres nécessaires à la description d'un mélange gaussien totalement inconnu. Bien que limitée aux mélanges de distributions normales, leur approche ne nécessite aucune information a priori sur les données et ne fait intervenir aucune hypothèse restrictive. Elle permet d'envisager une classification optimale pour un échantillon totalement inconnu. Il apparaît cependant que la mise en oeuvre du test de convexité demeure très sensible aux irrégularités de la distribution des observations disponibles et demande beaucoup de soin pour ajuster les paramètres de réglage.

#### I.4.2. Recherche des groupements

Lorsqu'on abandonne l'hypothèse paramètrique, on peut supposer que chaque groupement d'observations correspond à un mode de la fonction de densité sous-jacente /58/ /59/ /60/ /61/ /62/ et le problème de la classification automatique se trouve posé en termes de détection des modes.

La majorité des méthodes proposées jusqu'à présent dans cette optique nécessitent en premier lieu une estimation de la fonction de densité à partir des observations. Les modes sont alors détectés en remontant les pentes de la fonction de densité en faisant appel à des techniques basées sur l'utilisation du gradient de cette fonction /63/ /64/ /65/ /66/. Une variante intéressante de cette approche consiste à estimer directement le gradient à partir des observations /67/. Une autre procédure de recherche de maxima locaux consiste à transformer les observations en une séquence dont l'analyse permet de détecter les modes en présence /68/.

Toutes ces procédures sont connues pour être sensibles aux irrégularités rencontées dans les distributions réelles d'observations ainsi qu'aux variations spatiales non significatives de l'estimateur de la fonction de densité ou de son gradient. En pratique, elles tendent à générer de nombreux modes parasites qui sont généralement difficiles à distinguer des véritables modes de la fonction de densité.

Une approche très différente a été proposée par Vasseur et Postaire /62//69/ qui considèrent les modes comme des régions de l'espace où la fonction de densité est concave plutôt que comme des maxima locaux de cette fonction. Les opérations d'intégration qui sont à la base de l'analyse de la convexité de la fonction de densité sous-jacente permettent théoriquement de filtrer les variations accidentelles de cette fonction pour ne mettre en évidence que ses modes véritables. Cependant, la sensibilité de l'algorithme par rapport à l'ajustement de ses paramètres de fonctionnement est telle qu'il est extrêmement délicat d'obtenir des résultats fiables.

### **I.5. CONTOURS DES MODES ET CLASSIFICATION AUTOMATIQUE**

L'introduction du concept de convexité a permis, pour la première fois, de proposer une approche statistique unique pour l'analyse des mélanges comme pour la détection des modes en classification automatique /69/. L'intérêt d'une telle démarche est de disposer d'une famille d'algorithmes qui peuvent s'adapter à la majorité des situations rencontrées dans la pratique quotidienne de l'analyse des données multidimensionnelles.

Une critique importante qui peut cependant être adressée à l'ensemble de ces méthodes est leur grande sensibilité par rapport à certains paramètres de réglage des algorithmes, notamment le pas de discrétisation de l'espace des données. Pratiquement, seuls des analystes expérimentés peuvent utiliser ces algorithmes avec profit.

Poursuivant des objectifs comparables, nous proposons, dans ce mémoire, une nouvelle approche statistique unifiée aux problèmes de classification. Basée sur la détection et l'extraction des contours des modes des fonctions de densité, elle se veut plus robuste et plus facile à mettre en oeuvre que l'approche par analyse de convexité.

Notre démarche est très voisine de celle pratiquée en analyse des images numériques. Pour segmenter une image, on peut en effet chercher à reconstituer des régions au sein desquelles certaines propriétés de luminance, de texture, demeurent relativement stables /70/ /71/ /72/. Cette technique est comparable à l'analyse de la convexité des fonctions de densité qui permet de reconstituer les modes d'une distribution sous la forme de régions où cette convexité est constante. Lorsque les procédures de reconstitution de régions ne permettent pas de segmenter une image, l'analyste à recours aux techniques de détection et d'extraction des contours, qui sont les frontières entre les différentes régions constituant l'image /73/ /74/ /75/. La nouvelle approche que nous proposons procède de la même démarche, puisqu'il ne s'agit plus de caractériser un mode de la fonction de densité comme une région à convexité constante, mais comme une région délimitée par son contour. Comme en traitement d'images, les deux approches ne sont pas exclusives, mais doivent plutôt être considérées comme complémentaires l'une de l'autre.

Si la notion de contour dans une image est bien connue /76/ /77/ /78/, celle de contour pour une fonction de densité de probabilité se doit d'être précisée. Pour ce faire, il est montré comment une fonction de densité multivariable normale peut être entièrement caractérisée par son contour, défini comme le lieu des points où le module du gradient de la fonction présente des maxima locaux. Ce contour est une surface dont les propriétés géomètriques sont étroitement liées aux paramètres statistiques de la distribution considérée (Chapitre II).

Dans les problèmes pratiques, on s'intéresse aux contours de la fonction de densité sous-jacente à la distribution des observations constituant l'échantillon à analyser. Dans le contexte de la classification automatique, cette fonction n'est pas connue explicitement et doit être estimée à partir des observations disponibles par des techniques non paramétriques. Une généralisation multidimensionnelle du filtre médian est alors introduite et utilisée pour éliminer les effets du bruit tout en préservant les contours de la fonction de densité estimée (Chapitre III).

Deux opérateurs différentiels multidimensionnels sont alors utilisés pour estimer le module du gradient de la fonction de densité estimée-filtrée. Une procédure séquentielle d'extraction permet de reconstituer les contours à partir de l'ensemble des points de l'espace où le module du gradient de la fonction de densité présente des valeurs élevées (Chapitre IV).

La description des fonctions de densité par leurs contours permet alors d'envisager des problèmes de classification optimale par analyse des mélanges sous l'hypothèse paramètrique normale (Chapitre V.A.). Le rejet de tout modèle analytique conduit à des procédures de recherche de groupements où les modes de la fonction de densité, définis par leurs contours, peuvent être de formes quelconques (Chapitre V.B.).

L'approche proposée peut être mise en défaut lorsque les contours des classes en présence sont peu marqués, soit qu'ils soient noyés dans du bruit, soit par suite de chevauchements inter-classes trop importants. On montre alors comment les techniques d'étiquetage probabiliste permettent de lever les ambiguités que l'on peut rencontrer au niveau de l'extraction des contours (Chapitre VI).

Ces mêmes techniques ont également été appliquées avec succès au problème de la détection des modes considérés comme des maxima locaux des fonctions de densité de probabilité (Chapitre VII).

Cette approche met à la disposition de l'analyste des algorithmes robustes qui lui permettent de mettre en évidence les classes en présence dans un échantillon soit sous la forme de régions à fortes valeurs de la fonction de densité sous-jacente, soit sous la forme des contours délimitant ces régions.

#### REFERENCES I

- /1/ SPEARMAN, C.
   "General intelligence objectively determined and measured"
   Amer. Jour. Psychol., Vol. 15, pp 202-292, 1904
- /2/ BURT, C.
   "Experimental tests of general intelligence"
   Brit. Jour. Psychol., Vol. 3, pp 94-177, 1909
- /3/ LEBART, L., MORRINEAU, J.A. et FENELON, J.P. "Traitement des données statistiques" DUNOD, Paris, 1979
- /4/ CAILLEZ, F. et PAGES, J.P. "Introduction à l'analyse des données" Smash, Paris, Chpt. 11 à 13, 1976
- /5/ BENZECRI, J.P. "L'analyse de données" DUNOD, Paris, 1980
- /6/ KRUSKAL, J.B.
   "Multidimensional scaling by optimizing goodness of fit to a non
   metric hypothesis"
   Psychometrika, Vol. 29, pp 1-27, 1964
- /7/ FUKUNAGA, K. et MANTOCK, J.M. "A non parametric two dimensional display for classification" I.E.E.E. Trans. Pattern Anal. Machine Intell., Vol. PAMI-4, N° 4, pp 427-436, 1982
- /8/ KAKUSHO, O. et MIZOGUCHI, R. "A new algorithm for non-linear mapping with application to dimension and cluster analysis" Pattern Recognition, Pergamon Press, Vol. 16, Nº 1, pp 109-117, 1983
- /9/ MAHALANOBIS, P.C.
   "On tests and measures of group divergence"
   Journ. Asiat. Soc. Beng., Vol. 26, pp 541-588, 1930
- /10/ FISHER, R.A.
   "The use of multiple measurements in taxonomic problems"
   Ann. Eugenics, Vol. 7, part. II, pp 179-188, 1936

/11/ BOURTON, M. "Contribution à l'analyse, le traitement et la reconnaissance des formes ponctuées. Application à la classification de couches géologiques" Thèse de Docteur ès-Sciences, Univ. Lille 1, 1975 /12/ BATCHELOR, B.G. et WILKINS, B.R. "Methods for location of clusters of patterns to initialize a learning machine" Elect. Letters, Vol. 5, Nº 20, pp 481-483, 1969 /13/ BALL, G.H. "Data Analysis in the Social Sciences : what about the details" Proc. FJCC, Spartan Books, Washington, pp 533-560, 1965 /14/ FUKUNAGA K. et KOONTZ, W.L.G. "A criterion and an algorithm for grouping data" I.E.E.E. Trans. Computers, Vol. C-19, pp 917-923, 1970 /15/ BALL, G.H. et HALL, D.J. "A clustering technique for summarizing multivariate data" Behavioral Sc., Vol. 12, pp 153-155, 1967 /16/ DIDAY, E. "Une nouvelle méthode en classification automatique et reconnaissance des formes : La méthode des nuées dynamiques" Rev. Stat. Appl., Vol. 19, Nº 2, pp 20-33, 1971 /17 MAC QUEEN, J. "Some methods for classification and analysis of multivariate observation" Proc. 5<sup>th</sup> Berkeley Symp. Math. Stat. & Prob., Lecam & Neyman, Eds. Univ. Calif. Press, Vol. 1, pp 281-297, 1967 /18/ SEBESTYEN, G.S. "Pattern Recognition by an adaptative process of sample set construction" I.E.E.E. Trans. Inf. Theory, Vol. IT-8, pp 82-91, 1962 /19/ THORNDIKE, R.L. "Who belong in the family" Psychometrica, Vol. 18, pp 267-276, 1953 /20/ FRIEDMAN, H.P. et ROBIN, J. "On some invariant criteria for grouping data" J. Amer. Stat. Ass., Vol. 62, pp 1159-1178, 1967

- /21/ VOGEL, M.A. et WONG, A.K.C.
   "P.F.S. Clustering Method"
   I.E.E.E. Trans. Pattern Anal. Machine Intell., Vol. PAMI-1, N° 3,
   pp 237-245, 1979
- /22/ WILKS, S.S.
   "Multidimensional statistical scatter"
   Contrib. to Prob. Stat. I. Olkin, Eds. Stanford, Calif., 1960
- /23/ YAN, S.S. et CHANG, S.C.
   "A direct method for cluster analysis"
   Pattern Recognition, Pergamon Press, Vol. 7, pp 215-224, 1975
- /24/ LANE, G.N. et WILLIAMS, W.T. "A general theory of classificatory sorting strategies : 1 - Hierarchical Systems"
- /25/ LUKASOVA, A.
   "Hierarchical agglomerative clustering procedure"
   Pattern Recognition, Pergamon Press, Vol. 11, pp 365-381, 1979
- /26/ WISHART, D.
   "Estimating the modes of a multivariate sample density"
   Proc. Joint Meeting of the Classification Soc. British Pattern
   Recognition Ass., London, 1978
- /27/ AUGUSTSON, J.G. et MINKER, J. "An analysis of some graph theoretical cluster techniques" J. ACM, Vol. 17, pp 571-588, 1970
- /28/ ZAHN, C.T.
   "Graph theoretical methods for detecting and describing gestalt
   clusters"
   I.E.E.E. Trans. Computers, Vol. C-20, pp 68-86, 1971
- /29/ FRIEDMAN, H.P. et ROBIN, J.
   "On some invariant criteria for grouping data"
   J. Amer. Stat. Ass., Vol. 62, pp 1159-1178, 1967
- /30/ ANDREWS, H.C. "Introduction to mathematical techniques in pattern recognition" Wiley-Interscience, New-York, 1972
- /31/ MEISEL, W.S.
   "Computer-oriented approaches to pattern recognition"
   Academic Press, New-York, 1972
- /32/ DUBES, R. et JAIN, A.K.
   "Validity studies in clustering methodologies"
   Pattern Recognition, Pergamon Press, Vol. 11, pp 235-254, 1979

/33/ WOLFE, J.H. "Pattern Clustering for multivariate mixture analysis" Mult. Behavioral Research, Vol. 5, pp 329-350, 1970 /34/ DUDA, R.O. et HART, P.E. "Pattern Classification and scene analysis" J. Wiley, New-York, Chpt. 2, 1973 /35/ POSTAIRE J.G. "De l'image à la décision. Analyse des images numériques et théorie de la décision" DUNOD, Paris, 1987 /36/ PEARSON, K. "Contribution to the mathematical theory of evolution" Philos. Trans. Royal Soc. of London, Vol. 185, pp 71-110, 1894 /37/ BUCHAMAN-WOLLASTON, H.G. et HODGESSON, W.G. "A new method of treating frequency curves in fishery statistics, with some results" Journ. Cons., Vol. 4, pp 207-225, 1929 /38/ BHATTACHARYA, C.G. "A simple method of resolution of a distribution into gaussian components" Biometrics, 1967 /39/ DOETSCH, G "Zerlegung einer funktion in gaushe fehlerkurven und zeitliche zuruckverfolgung eines temperaturzustandes" Math. Zeitschrift, Vol. 41, pp 283-318, 1936 /40/ RAO, G.R. "Utilization of multiple measurements in problem of biological classification" Jour. R. Statist. Soc. B., Vol. 10, Nº 2, pp 159-193, 1948 /41/ BENZECRI, J.P. "La régression" Publ. Labo. Statist. Math. Univ. Paris VI, 1972 /42/ CAZES, P. "Décomposition d'un histogramme en composantes gaussiennes" Rev. Statist. App., Vol. 24, Nº 1, pp 63-82, 1976 /43/ DALY, R.F. "The adaptative binary-detection problem on the real line" Techn. Report 2003-3, Stand ford Univ. Calif., 1962

- /44/ HILLBORN, G. et LAINIOTIS, D.G. "Optimal unsupervised learning multicategory dependent hypothesis pattern recognition" I.E.E.E. Trans. INf. Theory, Vol. IT-14, pp 468-470, 1968
- /45/ MAKOV, V.E. et SMITH, A.F.M. "A quasi-Bayses unsupervised learning procedure for priors" I.E.E.E. Trans. Inf. Theory, Vol. IT-23, N° 6, pp 761-764, 1977
- /46/ DAY, N.E.
   "Estimating the components of a mixture of normal distributions"
   Biometrika, Vol. 56, pp 463-474, 1969
- /47/ SCHROEDER, A. "Analyse d'un mélange de distributions de probabilité de même type" Rev. Stat. Appl., Vol. 24, N° 1, pp 39-62, 1976
- /48/ DIDAY, E. "Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes" Thèse d'Etat, Univ. Paris VI, 1972
- /49/ KAZAKOS, D.
   "Recursive estimation of prior probabilities using a mixture"
   I.E.E.E. Trans. Inf. Theory, Vol. IT-23, N° 2, pp 203-211, 1977
- /50/ CHIEN, Y.T. et FU, F.S. "On Bayesian learning and stochastic approximation" I.E.E.E. Trans. Syst. Sc. Cyb., Vol. SSC-3, pp 28-38, 1967
- /51/ YOUNG, T.Y. et CORALUPPI, C.
   "Stochastic estimation of a mixture of normal density functions
   using an information criterion"
   I.E.E.E. Trans. Inf. Theory, Vol. IT-16, N° 3, pp 258-263, 1970
- /52/ MIZOGUCHI, R. et SHIMURA, M. "An approach to unsupervised learning Classification" I.E.E.E. Trans. Comput., Vol. C-24, N° 10, pp 979-983, 1975
- /53/ COOPER, D.B. et COOPER, P.W. "Non supervised adaptive signal detection and pattern recognition" Inform. & Control., Vol. 7, pp 416-444, 1964
- /54/ COOPER, P.W.
   "Some topics on non supervised adaptive detection for multivariate
   normal distributions"
   Computer Inf. Sc. II, Acd. Press, New-York, pp 123-146, 1967
- /55/ PATRICK, E. et HANCOCK, J.
   "Nonsupervised sequential classification and recognition of
   patterns"
   I.E.E.E. Trans. Inf. Theory, Vol. IT-12, Nº 4, pp 362-372, 1966

/56/ POSTAIRE J.G. et VASSEUR, C.P.A. "An approximate solution to normal mixture identification with application to unsupervised pattern classification" I.E.E.E. Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, N° 2, pp 163-179, 1981

/57/ POSTAIRE J.G. "Fonctions convexes et optimisation du processus de classification automatique : I. Identification des mélanges gaussiens par estimation de la convexité des fonctions de densité multivariables" RAIRO-Automatique, AFCET, Vol. 16, Nº 4, pp 357-379, 1982

- /58/ GITMAN, I.
   "An algorithm for nonsupervised pattern classification"
   I.E.E.E. Trans. Syst. Man, Cybern., Vol. SCM-3, pp 66-74, 1973
- /59/ GITMAN, I. et LEVINE, M.D.
   "An algorithm for detecting unimodal fuzzy sets and its applica tions as a clustering technique"
   I.E.E.E. Trans. Computers, Vol. C-19, pp 583-593, 1970
- /60/ HARTIGAN, J.A. "Clusters as modes" Proc. 1st-Int. Symp. Data ANal. Inf., Versailles, pp 433-448, 1977
- /61/ MIZOGUCHI, R. et SHIMURA, M. "Non parametric learning without a teacher based on mode estimation" I.E.E.E. Trans. Computers, Vol. C-25, pp 1109-1117, 1976
- /62/ VASSEUR, C.P.A. et POSTAIRE J.G. "A Convexity testing method for cluster analysis" I.E.E.E. Trans. Syst. Man Cyber., Vol. SMC-10, N° 3, pp 145-179, 1980
- /63/ KATZ, J.O. et ROHLF, F.J.
   "Function-point cluster analysis"
   Systematic Zoology, Vol. 22, pp 295-301, 1973
- /64/ JOHNSTON, B., BAILEY, T. et DUBES, R. "A variation on a nonparametric clustering method" I.E.E.E. Trans. Pattern Anal. Machine Intell., Vol. PAMI-1, N° 4, pp 400-408, 1979
- /65/ KOONTZ, W.L.G., NARENDRA, P.M. et FUKUNAGA, K. "A nonparametric valley-seeking technique for cluster analysis" I.E.E.E. Trans. Computers, Vol. C-21, Nº 2, pp 171-178, 1972
- /66/ KOONTZ, W.L.G., NARENDRA, P.M. et FUKUNAGA, K. "A graph Theoriticapproach to nonparametric cluster analysis" I.E.E.E. Trans. Comput., Vol. C-25, N° 9, pp 936-944, 1976

/67/ FUKUNAGA, K. et HOSTETLER, L.D. "The estimation of the gradient of a density function with applications in pattern recognition" I.E.E.E. Trans. Inf. Theory, Vol. IT-21, Nº 1, pp 32-40, 1975 /68/ KITTLER, J. "A locally sensitive method for cluster analysis" Pattern Recognition, Pergamon Press, Vol. 8, pp 23-33, 1976 /69/ POSTAIRE J.G. "Une approche statistique unique pour l'analyse des mélanges et la détection des modes en classification automatique" Rev. Stat. Appl., Vol. 31, Nº 4, pp 17-36, 1983 /70/ PAVLIDIS, T. "Structural pattern recognition" Springer, New-York, 1977 /71/ GONZALEZ, R.C. et WINTZ, R. "Digital image processing" Addison - Wesley, 1977 /72/ ROSENFELD, A. et DAVIS, L.S. "Image segmentation and image models" Proc. I.E.E.E., Vol. 67, pp 764-772, 1979 /73/ FU, K.S. et MUI, J.K. "A Survey on image segmentation" Pattern Recognition, Pergamon Press, Vol. 31, pp 3-16, 1981 /74/ DAVIS, L.S., ROSENFELD, A et WESZKA, J.S. "Region extraction by averaging and thresholding" I.E.E.E. Trans. Syst. Man Cybern., Vol. 5, pp 383-388, 1975 /75/ COLEMAN, G.B. et ANDREWS, H.C. "Image segmentation by clustering" Proc. I.E.E.E., Vol. 67, pp 773-785, 1979 /76/ DAVIS, L.S. "A survey of edge detection techniques" Comput. Graphics, image Processing, Vol. 4, pp 248-278, 1975 /77/ GRIFFITH, A.K. "Mathematical models for automatic line detection" J. Ass. Comput. Mach., Vol. 20, Nº 1, pp 62-80, 1972 /78/ MARR, D. et HILDRETH, E. "Theory of edge detection" MIT, Artificial Intell. Laboratory, AI Memo, Nº 518, 1979.

### CHAPITRE II

# IDENTIFICATION DES FONCTIONS DE DENSITE NORMALES MULTIVARIABLES PAR ANALYSE DE LEUR CONTOUR

### **II.1. LE CHOIX DU MODELE NORMAL**

Le concept de contour d'une fonction de densité de probabilité, sur lequel s'appuieront les techniques de classification automatique paramètrique et non paramètrique présentées en seconde partie de ce mémoire, est introduit sous l'hypothèse paramètrique normale. En effet, l'utilisation d'un modèle analytique s'est avérée nécessaire pour cerner avec rigueur et précision ce nouveau concept. Le choix du modèle Gaussien a été motivé par son très vaste champ d'application et son emploi quasi-universel. On sait qu'en classification paramètrique, l'hypothèse normale est généralement admise lorsqu'on dispose de peu d'information a priori sur la distribution des observations associées aux individus d'une classe et que ces individus sont très nombreux /1//2/.

### **II.2. LES FONCTIONS DE DENSITE DE PROBABILITE NORMALES**

Les fonctions de densité de probabilité normales multivariables sont définies sur  $\mathbb{R}^{N}$  sous la forme :

$$p(X) = \frac{1}{(2\pi) \frac{N}{2} |S|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (X - \overline{X})^{T} S^{-1} (X - \overline{X})\right]$$

où :

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N \end{bmatrix}^{\mathrm{T}}$$

 $\overline{X}$  est le vecteur-moyenne de la distribution normale :

$$\overline{\mathbf{X}} = \mathbf{E}[\mathbf{X}]$$

$$= \left[\mathbf{E}(\mathbf{x}_1), \ \mathbf{E}(\mathbf{x}_2), \ \dots, \ \mathbf{E}(\mathbf{x}_n), \ \dots, \ \mathbf{E}(\mathbf{x}_N)\right]^{\mathrm{T}}$$

$$= \left[\overline{\mathbf{x}}_1, \ \overline{\mathbf{x}}_2, \ \dots, \ \overline{\mathbf{x}}_n, \ \dots, \ \overline{\mathbf{x}}_N\right]^{\mathrm{T}}$$

avec :

$$\overline{\mathbf{x}}_{n} = \mathbf{E}(\mathbf{x}_{n})$$
,  $n = 1, 2, \dots, N$   
=  $\int_{-\infty}^{+\bar{\infty}} \mathbf{x}_{n} \mathbf{p}(\mathbf{x}) d\mathbf{x}$ 

Le symbole E(x) représente l'espérance mathématique de la variable aléatoire scalaire x, |S| est le déterminant de la matrice de covariance S, de la forme :



où:

$$\sigma_{i,j} = E\left[(x_i - \bar{x}_i) (x_j - \bar{x}_j)\right]$$

La matrice de covariance S est toujours réelle symétrique  $(S_{i,j} = S_{j,i})$  et semi définie positive /3/ /4/. Dans tout ce qui suit, nous ne considérons que les cas où cette matrice est définie positive, rejetant ainsi toute situation où les observations analysées appartiendraient à un sous-espace de R<sup>N</sup>.

L'identification des paramètres de la fonction de densité normale p(X)à partir d'un échantillon d'observations, qui se traduit par la détermination du vecteur-moyenne  $\overline{X}$  et des éléments de la matrice de covariance S, a fait l'objet de nombreux travaux (cf. Chapitre I, paragraphe (I.4.1)). Comme nous l'avons signalé la majorité de ces méthodes présente un certain nombre d'inconvénients au niveau de leur mise en oeuvre et/ou des hypothèses souvent restrictives qu'elles supposent. Or, d'un point de vue pratique, il existe de nombreuses situations où on ne possède aucune information a priori sur la distribution des observations si ce n'est l'hypothèse de normalité.

Pour ces raisons, il semble intéressant de proposer une nouvelle approche qui permette de déterminer les paramètres d'une distribution normale et qui soit applicable à l'analyse des mélanges gaussiens.

#### **II.3. CALCUL DU GRADIENT D'UNE FONCTION DE DENSITE NORMALE**

L'évaluation du gradient de la fonction p(X) en tout point de  $\mathbb{R}^N$  nous permettra de définir son contour avec précision.

#### II.3.1. Expression réduite d'une fonction de densité normale

Pour simplifier la présentation des calculs et sans aucune perte de généralité, considérons une fonction de densité normale de vecteur-moyenne nul :

$$p(X) = \frac{1}{(2\pi)^{\frac{N}{2}} |S|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} X^{T} S^{-1} X\right]$$
$$= C. \exp \left[-\frac{1}{2} X^{T} S^{-1} X\right]$$

avec :

$$C = \frac{1}{(2\pi)^{\frac{N}{2}} |S|^{\frac{1}{2}}}$$

L'étude de toute distribution normale de vecteur-moyenne non nul peut être ramenée à ce cas par une simple translation de l'origine de vecteur  $\overline{X}$ .

Considérons la forme quadratique Q de matrice  $S^{-1}$ :

$$Q = X^{T} S^{-1} X$$

Comme la matrice  $S^{-1}$  est réelle et symétrique, elle peut être diagonalisée par une transformation orthogonale /5/. Dans la base des vecteurs propres de la matrice  $S^{-1}$ , la forme quadratique Q s'écrit :

$$Q = Y^T D Y$$

avec : X = V Y

où : 
$$V = \begin{bmatrix} V_1, V_2, \dots, V_N \end{bmatrix}$$

définit le changement de base qui permet de passer d'une représentation dans l'espace de départ  $\mathbb{R}^N$  à une représentation dans un nouvel espace  $\mathbb{R}^{N}$  où l'inverse de la matrice de covariance est une matrice diagnonale de la forme :



Dans la base des vecteurs propres de  $S^{-1}$ , la fonction de densité normale considérée prend la forme réduite :

$$p(Y) = C \exp\left[-\frac{1}{2} Y^{T} D Y\right]$$

qui se prête mieux au calcul de son gradient que la forme initiale.

### II.3.2. Détermination de la norme du gradient

Soit  $\nabla_{\mathbf{y}}$  l'opérateur gradient usuel tel que :

$$\nabla_{\mathbf{Y}}\left[\mathbf{p}(\mathbf{Y})\right] = \begin{bmatrix}\frac{\partial \mathbf{p}(\mathbf{Y})}{\partial \mathbf{y}_{1}}, \frac{\partial \mathbf{p}(\mathbf{Y})}{\partial \mathbf{y}_{2}}, \dots, \frac{\partial \mathbf{p}(\mathbf{Y})}{\partial \mathbf{y}_{n}}, \dots, \frac{\partial \mathbf{p}(\mathbf{Y})}{\partial \mathbf{y}_{N}}\end{bmatrix}^{\mathrm{T}}$$

En utilisant l'expression de p(Y), il vient :

$$\frac{\partial p(Y)}{\partial y_n} = -C.(d_n y_n) \exp\left[-\frac{1}{2} Y^T D Y\right]$$

ce qui permet d'écrire :

$$\nabla_{\mathbf{Y}}\left[\mathbf{p}(\mathbf{Y})\right] = -\mathbf{C}\cdot\left[\mathbf{d}_{1}\mathbf{y}_{1}, \mathbf{d}_{2}\mathbf{y}_{2}, \dots, \mathbf{d}_{n}\mathbf{y}_{n}, \dots, \mathbf{d}_{N}\mathbf{y}_{N}\right]^{\mathrm{T}} \exp\left[-\frac{1}{2}\mathbf{Y}^{\mathrm{T}} \mathbf{D}\mathbf{Y}\right]$$

La norme enclidienne du vecteur gradient est de la forme :

$$\left\| \nabla_{\mathbf{Y}} \left[ \mathbf{p}(\mathbf{Y}) \right] \right\| = C \sqrt{\sum_{n=1}^{N} (d_n y_n)^2} \exp \left[ -\frac{1}{2} \mathbf{Y}^T \mathbf{D} \mathbf{Y} \right]$$

Il s'agit maintenant de localiser les points de l'espace R'N où cette norme présente des valeurs extrêmales.

### **II.4. CONTOUR D'UNE FONCTION DE DENSITE NORMALE MULTIVARIABLE**

### II.4.1. Analyse des extrema de la norme du gradient

Pour localiser les points de l'espace  $\mathbb{R}^{N}$  où la norme du gradient de p(Y) est maximale, on se place dans l'hyperplan ( $\Delta$ ) passant par l'origine et défini par :

$$\mathbf{T}.\mathbf{Y}=\mathbf{0}$$

avec :

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n, \dots, \mathbf{t}_N \end{bmatrix}$$

Dans ces conditions, une condition nécessaire d'extremum de  $\left\| \nabla_{Y} \left[ p(Y) \right] \right\|$  sous la contrainte T.Y = 0 peut être obtenue en utilisant la méthode des multiplieurs de Lagrange /6/, /7/.

On considère la fonction

$$h(Y,\lambda) = \left| \left| \nabla_{Y} \left[ p(Y) \right] \right| \right| - \lambda T. Y$$

qu'il s'agit d'extrémaliser par rapport aux composantes du vecteur Y et à  $\lambda$ . On obtient ainsi les conditions nécessaires :

$$\begin{cases} \nabla_{\mathbf{Y}} \left[ h(\mathbf{Y}, \lambda) \right] = 0 \\ \frac{\partial}{\partial \lambda} \left[ h(\mathbf{Y}, \lambda) \right] = 0 \end{cases}$$

En utilisant l'expression de  $h(Y, \lambda)$  il vient :

$$\begin{cases} \nabla_{\mathbf{Y}} \begin{bmatrix} \| \nabla_{\mathbf{Y}} \left[ \mathbf{p}(\mathbf{Y}) \right] \| \end{bmatrix} - \lambda \nabla_{\mathbf{Y}} \left[ \mathbf{T} \cdot \mathbf{Y} \right] = 0 \\ \mathbf{T} \cdot \mathbf{Y} = 0 \end{cases}$$

On reconnaît l'équation de l'hyperplan ( $\Delta$ ) dans la seconde condition.

En tenant compte de l'expression réduite de p(Y), on calcule les composantes du gradient  $\nabla_{Y} \left\| \nabla_{Y} \left[ p(Y) \right] \right\|$ :

$$\frac{\partial \left\| \nabla_{Y}\left[ p(Y) \right] \right\|}{\partial y_{n}} = C \cdot \frac{d_{n}y_{n}\left[ d_{n} - \left( \sum_{n=1}^{N} \left( d_{n} y_{n} \right)^{2} \right) \right]}{\sqrt{\sum_{n=1}^{N} \left( d_{n} y_{n} \right)^{2}}} \exp \left[ -\frac{1}{2} Y^{T} D Y \right]$$

Une condition d'extrema de  $\left\| \nabla_{\mathbf{Y}} \left[ \mathbf{p}(\mathbf{Y}) \right] \right\|$  s'écrit alors :

$$C \frac{\exp\left[-\frac{1}{2}Y^{T} D Y\right]}{\sqrt{\sum_{n=1}^{N} (d_{n}y_{n})^{2}}} \left[d_{n} - \sum_{n=1}^{N} (d_{n}y_{n})^{2}\right] d_{n}y_{n} = \lambda t_{n} , n = 1, 2, \dots, N$$
$$\sum_{n=1}^{N} (d_{n}y_{n})^{2}$$

En multipliant chacune des N premières égalités par  $y_n$ , n = 1, 2, ..., N, on obtient les n équations suivantes :

$$C \frac{\exp\left[-\frac{1}{2} Y^{T} D Y\right]}{\sqrt{\sum_{n=1}^{N} (d_{n} y_{n})^{2}}} \left[ (d_{n} y_{n})^{2} - d_{n} y_{n} \sum_{n=1}^{N} (d_{n} y_{n})^{2} \right] = \lambda t_{n} y_{n} , n=1,2,...,N$$

Puis, en sommant ces n égalités membre à membre et en utilisant la condition  $\sum_{n=1}^{N} t_n y_n = 0$ , on obtient une nouvelle condition nécessaire d'extréma :

$$C \frac{\exp - \frac{1}{2} Y^{T} D Y}{\sqrt{\sum_{n=1}^{N} (d_{n} y_{n})^{2}}} \left[ \sum_{n=1}^{N} (d_{n} y_{n})^{2} - \sum_{n=1}^{N} (d_{n} y_{n})^{2} \sum_{n=1}^{N} d_{n} y_{n}^{2} \right] = \dots$$
  
$$\dots = \lambda \sum_{n=1}^{N} t_{n} y_{n} = 0$$

La quantité

 $C \frac{\exp \left[-\frac{1}{2} Y^{T} D Y\right]}{\sqrt{\sum_{n=1}^{N} (d_{n} y_{n})^{2}}} \text{ ne pouvant jamais s'annuler,}$ 

cette condition est équivalente à :

$$\sum_{n=1}^{N} (d_n y_n)^2 - \sum_{n=1}^{N} (d_n y_n)^2 \sum_{n=1}^{N} d_n y_n^2 = 0$$

Soit, après simplification :

$$\sum_{n=1}^{N} d_n y_n^2 = 1$$

Cette dernière équation indique que sous la contrainte T.Y = 0, les maxima locaux du module du gradient de p(Y) sont situés sur l'hyperellipsoïde ( $\Gamma$ ) d'équation :



Figure II.1. : Cas bidimensionnel :

 $\bigcirc : \text{Localisation des extrema} \\ de \left\| \nabla_{y} \left[ P(Y) \right] \right\| \text{ dans le sous espace à une} \\ \text{dimension défini par } t_{1}y_{1} + t_{2}y_{2} = 0$ 

Ce résultat étant indépendant des valeurs des coefficients  $t_1, ..., t_n, ..., t_N$  définissant l'hyperplan ( $\Delta$ ), on en déduit que l'hyperellipsoïde ( $\Gamma$ ) représente le lieu géomètrique des maxima locaux du module du gradient dans tout hyperplan passant par l'origine. La figure II.1. illustre cette propriété dans le cas bidimensionnel.

II.4.2. Définition et propriétés du contour d'une fonction de densité normale

La détermination des extrema du module du gradient d'une fonction de densité normale multivariable sous contraintes nous permettent de définir avec rigueur la notion de contour pour de telles fonctions.

En revenant à l'espace de départ  $\mathbb{R}^N$  l'hyperellipsoïde ( $\Gamma$ ) a pour équation

$$(\mathbf{X} - \mathbf{\overline{X}})^{\mathrm{T}} \, \mathbf{S}^{-1} \, (\mathbf{X} - \mathbf{\overline{X}}) = 1$$

Définition :

On appelle contour de la fonction de densité normale

$$p(X) = \frac{1}{(2\pi)^{\frac{N}{2}} |S|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (X - \overline{X})^{T} S^{-1} (X - \overline{X})\right]$$

l'hyperellipsoïde d'équation :

$$(\mathbf{X} - \overline{\mathbf{X}})^{\mathrm{T}} \, \mathbf{S}^{-1} \, (\mathbf{X} - \overline{\mathbf{X}}) = 1$$

#### Propriété 1 :

Le contour est le lieu géomètrique des points de l'espace  $\mathbb{R}^N$  où, dans tout hyperplan passant par la moyenne  $\overline{X}$ , le module du gradient de la fonction p(X) présente des extrema locaux.

Il est intéressant de remarquer que le contour de p(X) ainsi défini coïncide avec la surface délimitant le domaine de concavité de p(X). En effet, il a été montré, en analysant le Hessien de p(X), que cette fonction est concave à l'intérieur du domaine défini par l'hyperellipsoïde ( $\Gamma$ ) /8/.

On peut donc énoncer la seconde propriété :

#### <u>Propriété 2</u> :

L'intérieur du contour de p(X) constitue le domaine de concavité de cette fonction.

# II.5. ANALYSE GEOMETRIQUE DU CONTOUR D'UNE FONCTION DE DENSITE

Dans l'espace  $R'^N$ , le lieu géomètrique des maxima locaux du module du gradient d'une fonction de densité est donné par l'équation :

$$Y^T DY = 1$$

qui est celle d'un hyperellipsoïde centré à l'origine (cf figure II.2). Ses axes principaux ont les mêmes directions que les vecteurs propres de la matrice de
covariance S. De plus, les longueurs  $L_n$ , n = 1, 2, ..., N de ses axes sont liées aux éléments diagonaux  $d_n$ , n = 1, 2, ..., N de la matrice diagonale D par les relations :

$$L_n = \frac{2}{\sqrt{d_n}}$$
,  $n = 1, 2, ..., N$ 





Figure II.2. : Le contour d'une fonction de densité de probabilité normale bidimensionnelle.

La matrice de covariance S peut être ainsi déterminée à partir des caractéristiques géomètriques de l'hyperellipsoïde ( $\Gamma$ ). En effet, la transformation orthogonale qui diagonalise la matrice S<sup>-1</sup> est définie par la matrice V =  $\begin{bmatrix} V_1, V_2, ..., V_N \end{bmatrix}$ dans laquelle les vecteurs V<sub>n</sub>, n = 1, 2, ..., N sont des vecteurs propres normés de S<sup>-1</sup>, c'est-à-dire les vecteurs unitaires des axes principaux de l'hyperellipsoïde.

En tenant compte des relations  $L_n^2 = 4/d_n$ , n = 1, 2, ..., N et de l'équation :

$$Q = Y^T V^T S^{-1} V Y$$
$$= Y^T D Y$$

on obtient l'expression de l'inverse de la matrice de covariance :



Les coordonnées du centre de l'hyperellipsoïde ( $\Box$ ) dans  $\mathbb{R}^N$  indiquent la valeur du vecteur moyenne  $\overline{X}$ .

La connaissance du contour, c'est-à-dire le lieu géomètrique des maxima locaux du module du gradient de la fonction de densité sous-jacente déterminés dans les hyperplans passant par le vecteur moyenne nous permet d'évaluer les paramètres d'une distribution normale. Schématiquement on procède par étapes :

- Premièrement, le centre de l'hyperellipsoïde ( $\cap$ ) indique le vecteurmoyenne  $\overline{X}$  de la distribution.
- Deuxièmement, les axes principaux de cet hyperellipsoïde définissent les vecteurs propres de l'inverse de la matrice de covariance.
- Troisièmement, les longueurs des axes principaux de ([]) donnent les valeurs propres de l'inverse de la matrice de covariance.
- Finalement, la connaissance de ces valeurs propres et de ces vecteurs propres permet de déterminer l'inverse de la matrice de covariance (S<sup>-1</sup>), puis la matrice de covariance elle-même.

#### **II.6. CONCLUSION**

L'identité du contour d'une fonction de densité normale avec la frontière de son domaine de concavité laisse envisager une exploitation fructueuse de ce nouveau concept. En effet, il a été montré comment l'analyse géomètrique des domaines de concavité des mélanges gaussiens permettait de proposer des procédures de classification automatique optimales au sens de la théorie de la décision par minimisation du taux d'erreur /8/.

Tout laisse donc à penser que lorsqu'on sera capable de détecter les contours des fonctions de densité, qui font appel à des concepts de valeurs extremales de leur gradient plutôt qu'à celui d'analyse de leur concavité, on pourra déboucher sur des procédures aussi intéressantes. Avant d'aborder l'analyse des mélanges gaussiens, il convient maintenant de considérer les problèmes liés à l'estimation des fonctions de densité à partir des observations disponibles, estimations sur lesquelles nous appliquerons des opérateurs différentiels destinés à faire apparaître les contours de ces fonctions.

# REFERENCES II

- /1/ DUDA, R.O. et HART, P.E. "Pattern Classification and Scene Analysis" J. Wiley, New-York, Chapitre 2, 1973
- /2/ SIMON, J.C.
   "La Reconnaissance des formes par Algorithmes"
   MASSON, Paris, 1984
- /3/ CRAMER, H. "Mathematical methods of statistics" Princeton, New-York, Chapitre 24, pp 310-316, 1963
- /4/ SEBER, G.A.F.
   "Multivariate Observations"
   J. Wiley, New-York, 1984
- /5/ GANTMACHER, F.R.
   "The theory of matrices"
   Chelsea Publishing C<sup>o</sup>, New-York, 1960
- /6/ MORRISSON, D.F. "Multivariate Statistical Methods" Mc. Graw Hill, 2<sup>nd</sup> Edition, Chapitre 2, pp 70-75, 1967
- /7/ STOER, J. et WITZALL, C. "Convexity and Optimization in finite dimensions I" Springer-Verlag, Berlin-Heidelberg, Chapitre 4, 1970
- /8/ POSTAIRE, J.G. et VASSEUR, C.P.A.
   "An approximate solution to normal mixture identification with application to unsupervised pattern classification" I.E.E.E. Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, N° 2, pp 163-178, 1981.

# SECTION II

# EXTRACTION DES CONTOURS ET CLASSIFICATION AUTOMATIQUE

# CHAPITRE III

# ESTIMATION ET FILTRAGE DES FONCTIONS DE DENSITE DE PROBABILITE

## **III.1. INTRODUCTION**

Dans un problème de classification non supervisé, la première difficulté rencontrée lorsqu'on utilise une approche statistique, est l'estimation de la fonction de densité de probabilité sous-jacente à la distribution des observations à classer.

On aborde dans ce chapitre un rappel sur les estimateurs non paramètriques utilisant la méthode du noyau et les problèmes liés à leur mise en oeuvre sur calculateur numérique.

Une généralisation de l'opération de filtrage non linéaire de type "median" est ensuite présentée pour améliorer la qualité de ces estimateurs. Ce filtre permet d'éliminer les effets du bruit tout en préservant les caractéristiques des contours des modes.

**III.2. - ESTIMATION NON PARAMETRIQUE** 

### III.2.1. Principe

Supposons que l'on dispose d'un échantillon  $\mathcal{H} = \begin{bmatrix} X_1, ..., X_j, ..., X_Q \end{bmatrix}$  de Q observations indépendantes et identiquement distribuées suivant une fonction de densité de probabilité f(X). La probabilité P pour qu'une observation X soit située à l'intérieur d'un domaine R(X<sub>0</sub>) centré au point X<sub>0</sub>, est donnée par :

(3.1.) 
$$P = \int_{R(X_0)} f(X) dX$$

La probabilité  $P_q$  pour que q des Q observations de l'échantillon soient situées dans le domaine  $R(X_0)$  est donnée par la loi binomiale :

$$(3.2.) \qquad P_q = \begin{pmatrix} Q \\ q \end{pmatrix} P^q (1 - P)^{Q-q}$$

L'espérance mathématique de q est alors :

$$(3.3.) E(q) = Q.P$$

ce qui permet d'affirmer que q/Q est un estimateur non biaisé de P.

Si l'on suppose que f(X) ne présente pas de variations importantes à l'intérieur du domaine  $R(X_0)$  la probabilité P peut se mettre sous la forme :

(3.4.) 
$$P = f(X) \cdot V(X_0)$$

où  $V(X_0)$  est le volume du domaine  $R(X_0)$ .

L'estimation de la fonction de densité en  $X_0$  est alors obtenue par combinaison des équations (3.3) et (3.4) :

(3.5.) 
$$\hat{\mathbf{f}}(\mathbf{X}_{o}) = \frac{\frac{\mathbf{q}}{\mathbf{Q}}}{\mathbf{v}(\mathbf{X}_{o})}$$

La question essentielle à résoudre est celle du comportement asymptotique de cet estimateur.

#### III.2.2. Comportement asymptotique

L'utilisation de cet estimateur simple soulève plusieurs problèmes. Si  $V(X_0)$  reste fixe alors que le nombre Q d'observations disponibles augmente indéfiniment, le rapport q/Q converge vers P (dans le sens probabiliste), et l'estimateur  $\hat{f}(X_0)$  tend vers la valeur moyenne de f(X) sur  $R(X_0)$ .

Mais, dans la pratique, le nombre d'observations disponibles n'est jamais illimité. Dans ces conditions, l'ajustement de la taille du domaine  $R(X_0)$ utilisé pour estimer f(X) en  $X_0$  est de la toute première importance. Si le domaine  $R(X_0)$  est trop petit, l'estimateur est nul dans presque tout l'espace, excepté lorsque le point d'estimation  $X_0$  est situé dans le voisinage immédiat d'une observation, ce qui conduit à des valeurs élevées de  $\hat{f}(X_0)$ . L'estimateur, très hératique, est alors inexploitable. A l'inverse, si le domaine  $R(X_0)$  est trop grand, on assiste à un phénomène de lissage et l'estimateur manque de résolution.

Pour assurer la convergence de l'estimateur  $\hat{f}(X_0)$  vers  $f(X_0)$ , il a été montré que le volume  $V(X_0)$  doit être lié au nombre total Q d'observations possibles.

Si pour un échantillon de taille Q, on note  $V_Q(X_0)$  le volume du domaine  $R_Q(X_0)$  utilisé,  $q_Q$  le nombre d'observations situées dans ce domaine et  $\hat{f}_Q(X_0)$  l'estimateur correspondant, on montre /1/ /2/ /3/ que  $\hat{f}_Q(X_0)$  converge en moyenne quadratique vers  $f(X_0)$  si :

1)  $\lim_{Q \to \infty} V_Q(X_0) = 0$ 2)  $\lim_{Q \to \infty} q_Q = \infty$ 3)  $\lim_{Q \to \infty} \frac{q}{Q} = 0$ 

La première condition indique  $V_Q(X_0)$  doit décroître lorsque Q augmente mais, d'après la deuxième condition, suffisamment lentement pour que  $R_Q(X_0)$  contienne de plus en plus d'observations. En pratique, on peut fixer le volume  $V_Q(X_0)$  en fonction du nombre Q d'observations disponibles. On aboutit alors à la méthode du noyau /4//5/. Parmi les possibilités offertes à l'analyste, on peut citer :

1) 
$$V_Q(X_0) = V_0 Q^{-\alpha}$$
 avec  $\alpha \in ]0, 1[$   
2)  $V_Q(X_0) = V_0 / \log Q$ 

Quelque soit la valeur du paramètre  $V_0$ , ces relations assurent la convergence asymptotique de l'estimateur. Nous verrons cependant qu'avec un nombre Q limité d'observations, le choix de  $V_0$  conditionne fortement les résultats de l'estimation.

Une autre possibilité consiste à se fixer le nombre d'observations  $q_Q$ en fonction du nombre total d'observations Q et faire croître le volume  $V_Q(X_0)$ jusqu'à ce qu'il englobe  $q_Q$  observations voisines de  $X_0$ . Il s'agit alors de la méthode de  $q_Q$  plus proches voisins /6/, /7/. Des relations du type :

1) 
$$q_Q = q_0 \sqrt{Q}$$
  
2)  $q_Q = q_0 \log Q$ 

assurent la convergence asymptotique de l'estimateur.

Comme pour le paramètre  $V_0$ , le choix de  $q_0$ , peut être théoriquement quelconque pour assurer la convergence de l'estimateur lorsque la taille Q de l'échantillon augmente indéfiniment. Dans la pratique, avec des échantillons de taille finie, ce choix conditionne la qualité de l'estimation.

La méthode du noyau, comme celle de  $q_Q$  plus proches voisins, présentent des avantages spécifiques qu'il n'y a pas lieu de discuter ici. Le lecteur intéressé pourra se reporter aux références /8/, /9/, /10/, /11/. Compte-tenu de l'usage qui sera fait des fonctions de densité estimées, nous ne nous intéresserons, dans tout ce qui suit, qu'à la méthode du noyau dont la mise en oeuvre est la mieux adaptée aux outils développés dans ce mémoire.

## III.2.3. Méthode du noyau

La méthode du noyau, initialement proposée par Rosenblatt /12/ dans le cas monovariable, puis approfondie par Parzen /13/, a été étendue au cas multidimensionnel par Cacoullos et Murthy /14/, /15/. Dans un espace de dimension N, l'estimateur est de la forme :

$$\hat{f}_Q(X) = \frac{1}{Q} \sum_{q=1}^{Q} h_Q^{-N} \phi(\frac{X - X_q}{h_Q})$$

où  $\phi$  (X) est une fonction réelle positive appelée "noyau" (ou "fenêtre") de "largeur"  $h_Q$ . L'estimateur  $\hat{f}_Q(X)$  ainsi défini converge en moyenne quadratique vers f(X) /13/, /16/, /17/ sous les conditions suivantes :

Il existe différents types de noyau  $\phi(X)$  répondant à ces conditions.

L'un des noyaux les plus utilisés dans la pratique est le noyau cubique :

$$h_Q^{-N} \phi \left( \frac{X}{h_Q} \right) = \begin{cases} \left( \frac{1}{2h_Q} \right)^N \sin \left| \frac{x_n}{h_Q} \right| \leq 1, & n = 1, 2, \dots, N \\ 0 & \text{autrement} \end{cases}$$

où: 
$$X = \begin{bmatrix} x_1, x_2, \dots, x_n, \dots, x_N \end{bmatrix}^T$$

Ce choix est généralement motivé par des considérations de programmation sur calculateur numérique. En effet, le calcul des valeurs absolues est beaucoup plus rapide que le calcul des normes euclidiennes associées au noyau de type exponentiel. Dans ce sens, l'estimateur  $\hat{f}_Q(X)$  est obtenu d'une façon simple, car il est très facile et rapide de compter le nombre d'observations situées dans l'hypercube de côté  $h_Q$  centré en  $X_0$ . La position d'une observation par rapport à cet hypercube peut être déterminée en testant ses composantes une à une. Il en résulte qu'un grand nombre d'observations situées à l'extérieur de cet hypercube peuvent être éliminées avant d'avoir testé toutes leurs composantes.

Cependant, malgré ces avantages apparents, l'utilisation de la technique d'estimation des fonctions de densité employant le noyau cubique reste limitée à des données de faibles dimensions. En effet, le nombre de points nécessaires pour décrire explicitement une fonction multivariable croit exponentiellement avec la dimension de l'espace de représentation. Ainsi, lorsque l'opération relativement simple et rapide d'estimation de la fonction de densité en un point doit être répétée pour tous les points où on désire estimer cette fonction de densité, on aboutit à des temps de calcul prohibitifs dès que s'élève la dimension des données.

Pour pallier cet inconvénient, nous allons utiliser un algorithme rapide d'estimation des fonctions de densité de probabilité qui a été développé par Postaire et Vasseur /18/.

# III.2.4. Algorithme rapide d'estimation des fonctions de densité de probabilité

# III.2.4.(a) Position du problème

L'implantation classique de la procédure d'estimation d'une fonction de densité de probabilité par la méthode du noyau cubique nécessite une discrétisation de l'espace  $\mathbb{R}^N$  de représentation des données. Schématiquement, en normalisant les plages de variation de chacune des coordonnées de  $\mathbb{R}^N$  de telle sorte que les observations soient toutes situées dans un hypercube de côté unité et en divisant ces plages en  $(1/h_Q)$  intervalles adjacent égaux, on obtient une discrétisation sur  $(1/h_Q)^N$  hypercubes de côté  $h_Q$ . La fonction de denstité est alors estimée aux centres de ces hypercubes qui constituent les noeuds d'un réseau hypercubique (cf. figure III.1.).





Cette représentation explicite de la fonction de densité sous jacente en ces points de discrétisation nécessite le dénombrement des observations situées dans chacun des  $(1/h_Q)^N$  hypercubes précédemment définis. En remarquant que le nombre d'hypercubes non vides ne peut dépasser celui des observations disponibles, on constate que ce dénombrement conduit souvent à mettre en évidence des hypercubes vides, surtout pour des données de forte dimension pour lesquelles

$$(1/h_Q)^N > > > Q$$

En fait, pour un échantillon de taille Q, quelle que soit la dimension des données, l'estimation de la fonction de densité est nulle, sauf en un nombre limité de points de discrétisation au plus égal à Q. Dans la pratique, le nombre d'estimations non nulles est généralement très nettement inférieur au nombre de points de discrétisation de la fonction de densité.

L'algorithme rapide d'estimation est basé sur une procédure permettant de déterminer directement les hypercubes non vides, sans avoir à prendre en considération l'ensemble des hypercubes centrés en tous les points de discrétisation où la densité doit être estimée. Pour une taille d'échantillon donnée, le temps de calcul est ainsi limité au temps nécessaire à la détermination d' un nombre d'estimations élémentaires au maximum égal au nombre Q d'observations.

# III.2.4.(b) Principe

Avant tout traitement, il est généralement conseillé de normaliser les plages de variation de chacune des composantes des observations /19/. L'espace R<sup>N</sup> est normalisé de telle sorte que les valeurs extrêmes de chaque composante soient O et 1.

Le coefficient de normalisation de chacune des composantes  $x_{q,1}$ ,  $x_{q,2}$ , ...,  $x_{q,n}$ , ...,  $x_{q,N}$  des observations  $X_q$ , q = 1, 2, ..., Q est défini par

$$a_n = \underset{q}{\operatorname{Max}} x_{q,n} - \underset{q}{\operatorname{Min}} x_{q,n}, \qquad n = 1, 2, \dots, N$$

En faisant subir à l'espace  $\mathbb{R}^{N}$  une translation de vecteur :

$$B = \left[ (-x_1^m), (-x_2^m), \dots, (-x_n^m), \dots, (-x_N^m) \right]^T$$
$$x_n^m = \underset{q}{\text{Min}} x_{q,n} \qquad n = 1, 2, \dots, N$$

désigne la plus petite valeur du nième caractère, suivie d'une transformation diagonale de matrice :



les coordonnées des observations dans le nouvel espace  $R_1^N$  sont :

$$X'_{q} = \begin{bmatrix} x'_{q,1}, \dots, x'_{q,n}, \dots, x'_{q,N} \end{bmatrix}^{T}$$
  $q = 1, 2, \dots, Q$   
 $X'_{q} = A \cdot \begin{vmatrix} x_{q} + B \end{vmatrix}$ 

avec :

où:

Ces nouvelles coordonnées vérifient les relations suivantes :

$$0 \leq x'_{q,n} \leq 1$$
   
 $\forall q, q = 1, 2, ..., Q$   
 $\forall n, n = 1, 2, ..., N$ 

Chaque axe de l'espace normalisé est partitionné en  $(1/h_Q)$  intervalles égaux et adjacents. Cette discrétisation définit un réseau de  $(1/h_Q)^N$  hypercubes adjacents de côté  $h_Q$  qui couvre tout l'espace  $R_1^N$ .

La normalisation des données est suivie d'un changement d'échelle de l'espace  $R_1^N$ , obtenu par la transformation linéaire :

$$Y_{q} = Z \cdot X_{q}$$
$$Z \cdot = \left(\frac{1}{h_{Q}}\right)^{T} \left[ I^{N} \right]$$

de matrice

où  $\begin{bmatrix} I & N \end{bmatrix}$  est la matrice unité d'ordre N et où  $(1/h_Q)$  est un nombre entier. Dans cette transformation linéaire, les composantes de l'espace  $R_1^N$  sont divisées par  $(h_Q)$  de telle sorte qu'un hypercube de côté  $h_Q$  de l'espace  $R_1^N$  devient un hypercube de côté unité dans l'espace transformé, noté  $R_2^N$ .

La procédure rapide permet d'estimer la fonction de densité sousjacente uniquement aux centres des hypercubes non vides du réseau de discrétisation défini sur R<sub>2</sub><sup>N</sup>, dont le nombre ne peut dépasser celui des observations disponibles. Les performances remarquables de cet algorithme résultent d'une technique de programmation qui tire partie de la normalisation préalable de l'espace R<sub>2</sub><sup>N</sup> pour dresser directement la liste des hypercubes non vides à partir du fichier des données. De plus, cette technique permet de connaître simultanément le nombre d'observations situées dans chacun des hypercubes non vides, donc la valeur de l'estimateur en son centre. Sans aucun calcul, on sait d'autre part que la valeur de l'estimateur est nulle aux centres des hypercubes vides qui, surtout dans les espaces de grande dimension, constituent l'écrasante majorité des hypercubes du réseau de discrétisation.

On désigne par  $P_r$ , r = 1, 2, ..., R, R Q, les centres des hypercubes non vides qui sont appelés *points d'échantillonnage* dont les positions dans l'espace  $R_2^N$  sont repérées par des coordonnées à valeurs entières.

Avec un nombre limité d'observations, l'ajustement du paramètre  $h_Q$  constitue un problème délicat. Si le pas de discrétisation est trop petit, l'estimateur présente trop de variabilité pour être exploitable, en particulier dans les régions de

faible densité. A l'opposé, s'il est trop grand, l'estimateur manque de résolution et des modes significatifs de la fonction de densité risquent de passer inaperçus.

Il existe des techniques basées sur l'analyse de la stabilité du nombre de modes détectés qui permettent d'ajuster ce paramètre sans connaissance a priori sur les données à traiter /20/, /21/. Il apparaît néanmoins que les résultats obtenus demeurent très sensibles à l'ajustement de ce paramètre /22/.

Nous allons montrer comment l'utilisation d'une technique de filtrage non linéaire permet de conserver les modes réels des fonctions de densité estimées tout en éliminant ces variations non significatives, souvent dues à un trop petit pas de discrétisation. L'ajustement du pas deviendra, de ce fait, beaucoup moins critique.

Sachant que toute l'information disponible pour assurer la classification des observations est contenue dans la version discrète de la fonction de densité de probabilité sous-jacente :

$$\hat{f}(P_r)$$
,  $r = 1, 2, ..., R$ 

on conçoit que ce filtrage est fondamental pour la suite des opérations.

# **III.3. FILTRAGE NON LINEAIRE**

L'estimation non paramètrique avec noyau cubique se révèle très sensible aux irrégularités de la distribution des observations. La présence du bruit ne peut qu'accentuer l'apparition de variations spatiales locales importantes, mais non significatives, de l'estimateur (cf. figure III.4b, paragraphe III.3.4. du même chapitre).

Avant d'aborder la détection des modes proprement dite, il s'avère donc nécessaire de procéder à un filtrage de la fonction de densité estimée. Les techniques du traitement des images numériques nous enseignent que les filtres linéaires (passe-bas), qui consistent à remplacer le niveau de gris en un point par une moyenne pondérée des niveaux de gris dans son voisinage, assurent un lissage efficace mais, en contre-partie, atténuent fortement les contours présents dans l'image /23/, /24/. Par contre, le filtre non linéaire de type "médian" est plus fréquemment utilisé car il présente la propriété remarquable de diminuer le bruit d'une image tout en préservant ses contours /25/, /26/, /27/.

Cette constation est à la base de la généralisation de cette technique de filtrage à des fonctions multidimensionnelles présentée ci-après et utilisée dans une phase de prétraitement, précédant la détection des contours des modes présents dans une fonction de densité.

Ce filtre médian utilisé en traitement d'image est très simple à réaliser, mais comme tous les filtres non linéaires, il se prête mal à l'analyse théorique.

# III.3.1. Principe du filtre médian

Le filtre médian est d'abord apparu comme une technique de traitement de signal /28/. Opérant sur une fenêtre glissante le long du signal, le filtre médian donne en sortie la valeur de l'échantillon qui, dans la fenêtre, a autant de voisins plus grand que lui que de voisins plus petits. La figure III.2. permet de comparer les effets d'un filtre médian de largeur 3 à ceux d'un filtre à moyenne glissante de même largeur. On constate que le filtre médian élimine le bruit de manière comparable au filtre à moyenne glissante, mais que les transitions importantes sont moins bien préservées par ce dernier.

L'utilisation de ce filtre en traitement des images revient à remplacer la valeur du niveau de gris en chaque point par le niveau de gris encadré par autant de valeurs qui lui sont supérieures que de valeurs inférieures dans un voisinage donné du point considéré /29/, /30/, /31/.



- Figure 111.2. :
- (a) Signal brut constitué d'un front sur lequel est superposé un bruit
- (b) Signal résultant après passage du filtre médian n=3
- (c) Signal résultant après passage du filtre passe-bas n=3.





# III.3.2. Analyse statistique

Soit une variable aléatoire X, de densité de probabilité f(X), sa moyenne  $\overline{X}$ , sa variance  $v_{\overline{X}}$  et son écart type  $\sigma_{\overline{X}}$  sont définis statistiquement par :

$$\overline{X} = E(X) = \int_{-\infty}^{+\infty} X f(X) dX$$
$$v_{X} = E(X - \overline{X})^{2} = \int_{-\infty}^{+\infty} (X - \overline{X})^{2} f(X) dX$$
$$\sigma_{X} = \sqrt{v_{X}}$$

Les quartiles donnent des informations semblables sur la variable aléatoire X. Le quartile  $q_n$ , n = 1, 2, 3 est la valeur de cette variable pour laquelle la fonction de répartition F(X) atteint la valeur n/4.

On a:  

$$F(q_n) = \frac{n}{4}$$
  
où:  
 $F(X) = \int_{-\infty}^{X} f(Y) dY$ 

A la moyenne  $\overline{X}$  correspond le quartile  $q_2$ .

$$q_2 = M(X) = F^{-1}(\frac{1}{2})$$

Les quartiles  $q_2$  et  $q_3$  peuvent servir, comme  $v_X$  et  $\sigma_X$ , à caractériser la dispersion de la distribution de la variable X.

On sait que la moyenne  $\overline{X}$  et la médiane M(X) tendent vers la même valeur asymptotique lorsque la variable aléatoire X obéit à une loi statistique symétrique /32/. Cependant, lorsqu'on dispose d'un nombre limité d'observations, une mesure ératique sur la variable X fausse l'estimation de la moyenne  $\overline{X}$  d'une manière non négligeable. L'effet d'une telle valeur aberrante est, par contre, peu sensible sur l'estimation des quartiles. L'estimateur de la médiane est ainsi généralement considéré comme plus robuste que celui de la moyenne /33/, /34/.

## III.3.3. Principe de programmation

L'implantation d'un filtre médian multidimensionnel nécessite de définir avec précision le voisinage de chaque point d'échantillonnage  $P_r$  dans le réseau d'hypercubes de l'espace normalisé  $R_2^N$  défini au paragraphe III.2.4(b).

Soient  $p_{r,1}$ ,  $p_{r,2}$ , ...,  $p_{r,n}$ , ...,  $p_{r,N}$  les coordonnées entières d'un point  $P_r$ . Son voisinage de taille m, noté  $U_m(P_r)$  est constitué des points  $P_r$ , tels que :

$$U_{m}(P_{r}) = \left[ (p_{r',1}, \dots, p_{r',n}, \dots, p_{r',N}) / p_{r,n} - m \leq p_{r',n} \leq p_{r,n} + m \right]$$

où :

$$r = 1, 2, ..., R$$
 et  $n = 1, 2, ..., N$ 

La figure (III.3.) représente les voisinages  $U_1(P_r)$  constitués de 9 voisins dans un espace bidimensionnel et de 27 voisins dans un espace tridimensionnel.

Le filtre médian, utilisé en traitement des images numériques et traitement du signal, est basé sur une opération mathématique simple qui est très aisément généralisable à un espace multidimensionnel.

Le filtre médian remplace l'estimateur  $\hat{f}(P_r)$  au point d'échantillonage  $P_r$  par la valeur médiane des  $(2m+1)^N$  valeurs de la fonction estimée aux points situés dans le voisinage  $U_m(P_r)$ . Par analogie avec l'estimation par la méthode du noyau, on appelle le voisinage  $U_m(P_r)$  la "fenêtre du filtre médian" de taille m.



Soit  $\hat{f}(P_r)$  la valeur de l'estimateur filtré au point  $P_r$ , qui est donnée sous la forme :

$$\hat{f}'(P_r) = valeur mediane \left[\hat{f}(P_r)/P_r \in U_m(P_r)\right]$$

La procédure de progammation consiste à ordonner l'ensemble des valeurs estimées de la fonction de densité de probabilité aux points d'échantillonnage  $P_r$ , situés à l'intérieur de la fenêtre  $U_m(P_r)$ , de la plus petite à la plus grande. La

valeur médiane est celle de rang  $(\frac{2m + 1)^{N} + 1}{2}$  dans cette série ordonnée.

Notons qu'une variante du filtre médian, appelé "filtre médian séparable", peut également être utilisée /35/. Cette variante consiste en une succession d'applications d'un filtre médian à une dimension, parallèlement aux N axes de l'espace des données. Les résultats obtenus par cette méthode ne sont pas identiques à ceux résultant de l'application du filtre médian multidimensionnel utilisé, car ce filtre séparable ne prend pas en compte le voisinage hypercubique complet des points d'échantillonnage.

L'avantage du filtre médian séparable est sa très grande facilité d'implantation, il suffit de programmer le filtre à une dimension suivant une direction puis, par le biais de simples rotations, de l'appliquer parallèlement à chacun des axes de l'espace des données.

Il est cependant apparu que les résultats obtenus par utilisation d'un filtre séparable sont moins bons que ceux résultant de l'utilisation du filtre multidimensionnel présenté. Ce dernier sera donc utilisé dans toute la suite de ce travail.

#### III.3.4. Résultats expérimentaux

Le filtre médian ainsi défini a été testé sur des données multidimensionnelles générées artificiellement. On utilise les données constituées de vecteurs aléatoires provenant de distributions normales.

## EXEMPLE 1 :

Pour permettre une visualisation simple des performances obtenues, on considère en premier lieu un exemple bidimensionnel, constitué de trois distributions normales dont les paramètres statistiques sont consignés dans le tableau 1. Ces distributions sont équiprobables et représentées chacune dans l'échantillon analysé par 200 observations (cf. Figure (III.4.(a)).

La figure (III.4.(b)) montre la grande variabilité spatiale de l'estimateur non paramètrique obtenu pour un ajustement du paramètre  $h_Q$  tel que  $(1/h_Q)$  soit au mileu de sa plus grande plage de variation pour laquelle le nombre de modes détectés en fin de procédure demeure constant.

Après filtrage (cf. Figure III.4.(c)), il ne subsiste que les modes significatifs de la fonction de densité de probabilité sous-jacente. Le résultat présenté a été obtenu pour la taille de la fenêtre du filtre m = 1.

#### EXEMPLE 2 :

Après le premier exemple qui a permis de tester les effets du filtre médian sur un exemple de trois distributions normales non bruitées, nous proposons un second exemple destiné à mettre en évidence l'intérêt de ce filtre pour éliminer le bruit. L'exemple proposé est constitué des trois distributions équiprobables du tableau 1, représentées chacunes par 200 observations, auxquelles on a ajouté un

Distribution	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\overline{X}_1 = \begin{bmatrix} 0\\0 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 18 & 0 \\ 0 & 18 \end{bmatrix}$	Pb <sub>1</sub> =1/3
2	$\overline{X}_2 = \begin{bmatrix} 0 \\ 14 \end{bmatrix}$	$S_{2} = \begin{bmatrix} 18 & 0 \\ 0 & 18 \end{bmatrix}$	Pb <sub>2</sub> =1/3
3	$\overline{\mathbf{X}}_{3} = \begin{bmatrix} 1 & 3 \\ 7 \end{bmatrix}$	$S_{3} = \begin{bmatrix} 18 & 0 \\ 0 & 18 \end{bmatrix}$	Pb <sub>3</sub> =1/3

Tableau 1: Les valeurs des paramètres statistiques<br/>de la distribution de l'exemple 1.

Distribution	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori	
1	$\overline{\mathbf{X}}_{1} = \begin{bmatrix} 4\\ 4 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	Pb <sub>1</sub> =i/3	
2	$\overline{X}_2 = \begin{bmatrix} 15\\4 \end{bmatrix}$	$S_2 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	Pb <sub>2</sub> =1/3	
3	$\overline{\mathbf{X}}_{3} = \begin{bmatrix} 10\\15 \end{bmatrix}$	$S_3 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	Pb <sub>3</sub> =1/3	

Tableau 2 : Les valeurs des paramètres statistiquesde la distribution de l'exemple 2.

. .



Figure III.4.(a) : Représentation graphique des 600 observations de l'exemple 1.





BU



Figure III.4.(c) : Version lissée de la fonction de densité de l'exemple 1, par application du filtre médian de taille m = 1.



bruit constitué d'observations uniformément distribuées sur tout l'espace. En notant b le nombre d'observations constitutant ce bruit, on peut estimer la probabilité a priori du bruit dans l'échantillon par :

$$P(b) = \frac{b}{b+Q}$$

où Q est le nombre d'observations provenant des classes en présence.

Le niveau du bruit dans l'échantillon analysé est alors défini par :

$$\Theta = P(b) \times 100 \%$$

Dans l'exemple présenté, pour lequel b = 800 et Q = 600, le niveau de bruit s'établit à 57 % (figure (III.5.(a))). La figure (III.5.(b)) montre l'estimation non paramétrique brute obtenue pour un ajustement de  $h_Q$ , suivant les mêmes conditions que pour l'exemple 1 ( $h_Q = 1/30$ ).

Le résultat du filtrage est présenté sur la figure (III.5.(c)), la taille m de la fenêtre est prise égale à 1. On peut voir qu'il ne subsiste que les modes significatifs de la fonction de densité de probabilité sous jacente.

# **III.4. CONCLUSION**

Dans ce chapitre, un rappel sur les techniques d'estimation non paramètriques et leur implantation nous a permis de préciser le modèle utilisé pour la représentation des données dans un espace multidimensionnel normalisé.

Nous avons généralisé l'utilisation du filtre médian au prétraitement des estimations brutes des fonctions de densité sous-jacentes à la distribution de données multidimensionnelles. La mise en oeuvre de ce filtre non linéaire est de la toute première importance pour la suite de l'analyse des données. Il sera systématiquement appliqué aux fonctions de densité estimées car il permet de préserver les contours des modes tout en éliminant les effets du bruit. Ce n'est que



Figure III.5.(a) : Distribution des observations dans l'espace à deux dimensions de l'exemple 2.



Figure III.5.(a) : Estimation brute de la fonction de densité sous-jacente de l'exemple 2.







Figure III.5.(c) : Version lissée de la fonction de densité de l'exemple 2 par application du filtre médian de taille m = 1.

sur les versions lissées des estimations que nous appliquerons les opérateurs de détection des contours présentés au chapitre suivant. Pour alléger les notations,

$$\hat{f}(P_r), r = 1, 2, ...., R$$

représentera, dans tout ce qui suit, la fonction de densité sous-jacente estimée et filtrée sur laquelle reposent toutes les procédures de classification exposées dans ce mémoire.

. . . .

# REFERENCES III

- /1/ WEGMAN, E.J.
   "Nonparametric probability density estimation"
   I.A summary of available methods"
   Technometrics, Vol. 14, pp 533-546, 1972
- /2/ DUDA, R.et HART, P.E. "Pattern Classification and Scene Analysis" J. Wiley, New-York, Chapitre 4, 1973
- /3/ SILVERMAN, L. "Asymptotic theory of statistical tests and estimation" Chakravarti, Ed., New-York, Academic P., 1980
- /4/ BANON, G. "Estimation nonparamétrique de densité de probabilité pour les processus de Markov" Thèse d'Etat, Univ. P. Sabatier, Toulouse, lère Partie, 1977
- /5/ ASSELIN DE BEAUVILLE, J.P. "Estimation non paramétrique de la densité et du mode. Exemple de la distribution Gamma" Rev. Statist. Appl., Vol. XXVI, N° 3, pp 47-70, 1978
- /6/ LOFTSGAARDEN, D.O. et QUESENBERNY, C.P. "A non parametric estimate of a multivariate density function" Ann. Math. Stat., Vol. 36, pp 1049-1051, 1965
- /7/ COVER, T.M. et HART, P.E. "Nearest neighbor pattern classification" I.E.E.E. Trans. Info. Theory, Vol. IT-13, pp 21-27, 1967
- /8/ FUKUNAGA, K. "Introduction to statistical Pattern Recognition" Academic Press, New-York, 1972
- /9/ WEGMAN, E.J.
   "Non parametric probability estimation : II A Comparison of
   density estimation methods"
   J. Statist. Comput. Simul., Vol. 1, pp 225-245, 1972
- /10/ TITTERINGTON, D.M.
   "A comparative study of kernel-based density estimates for
   categorical data"
   Technometrics, 22, pp 259-270, 1980

- /11/ GAILLAT, G. "Méthodes statistiques de reconnaissance des formes" E.N.S.T.A., Paris, 1983
- /12/ ROSENBLATT, M.
   "Remarks on some nonparametric estimates of a density function"
   Ann. Math. Stat., Vol. 27, pp 832-837, 1956
- /13/ PARZEN, E. "On estimation of a probability density function and mode" Ann. Math. Stat., Vol. 33, pp 1065-1076, 1962
- /14/ CACOULLOS, T.
   "Estimation of multivariate density"
   Ann. Inst. Stat. Math., Vol. 18, pp 179-189, 1962
- /15/ MURTHY, V.K.
   "Nonparametric estimation of multivariate densities with
   applications"
   Multivariate Analysis Academic Press, New-York, pp 43-56, 1966
- /16/ DEHEUVELS, P. "Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque\_sûre des estimateurs de la densité" C.R. Ac. Sc., T-278, Série A, pp 1217-1220, 1974
- /17/ DEVROYE, L. et MACHELL, F.
   "Data Structure in kernel density estimation"
   I.E.E.E. Trans. Pattern Analysis & Machine Intell., Vol. PAMI-7,
   N° 3, pp 360-366, 1985
- /18/ POSTAIRE, J.G. et VASSEUR, C.P.A. "A fast Algorithm for nonparametric probability density estimation" I.E.E.E. Trans., Pattern Analysis & Machine Intell., Vol. PAMI-4, N° 6, pp 663-666, 1982
- /19/ MEISEL, W.S.
   "Computer-oriented approaches to pattern recognition"
   Academic Press, New-York, 1972
- /20/ EIGEN, D.J., FROMM, F.R. et NORTHOUSE, R.A. "Cluster analysis based on dimensional information with applications to feature selection and classification" I.E.E.E. Trans., Syst. Man and Cybernetics, Vol. SMC-4, N° 3, pp 284-294, 1974
- /21/ TOMEK, I.
   "A modification of a clustering method"
   I.E.E.E. Trans., Syst. Man and Cybernetics, Vol. SMC-5, N° 3,
   pp 394-396, 1975

/22/ POSTAIRE J.G. "De l'Image à la Décision. Analyse des images numériques et théorie de la décision". DUNOD, Paris, 1987

- /23/ HAMMING, R.W.
   "Digital Filters"
   Englewood Cliffs, N.J., Prentice Hall, 1977
- /24/ VELLEMAN, P.F.
   "Definition and comparison of robust nonlinear data smoothing
   algorithms"
   J. Amer. Stat. Assoc., Vol. 75, N° 371, pp 609-615, 1980
- /25/ VELLEMAN, P.F.
   "Robust nonlinear data smoothers : Definitions and Recommendations"
   Proceedings, National Academy of Sciences, 74, pp 434-436, 1977
- /26/ TUKEY, J.W.
   "Exploratory data analysis"
   Addison-Wesley, Reading, MA, 1977
- /27/ JUSTUSSON, B.I.
   "Noise reduction by median filtering"
   Proc. I.E.E.E. Conf. Image Processing, pp 502-504, 1978
- /28/ RABINEE, L.R. , SAMBUR, M.R. et SCHMIDT, C.E. "Applications of nonlinear smoothing algorithm to speech processing" I.E.E.E. Trans. Acoust. Speech & Signal Proc., Vol. ASSP-23, N° 6, pp 552-557, 1975
- /29/ PRATT, W.K.
   "Digital image processing"
   Wiley-Interscience, New-York, 1978
- /30/ JUSTUSSON, B.I.
   "Median Filtering : Statistical Properties"
   Topics in Applied Physics, Vol. 43, pp 161-196, 1981
- /31/ BOVIK, A.C., HUANG, T.S. et MUNSON, D.C.
   "A generalization of median filtering using linear combinations of order statistics"
   I.E.E.E. Trans. Acoust., Speech & Signal Proce., Vol. ASSP-31, N° 6, pp 1342-1349, 1983
- /32/ DAVID, H.A. "Order Statistics" J. Wiley, New-York , 1970
/33/ BICKEL, P.J. et LEHMANN, E.L.
 "Descriptive statistics for nonparametric models, II location"
 Annal. of Statistics, Vol. 3, N° 5, pp 1045-1069, 1975

/34/ DESSIMOZ, J.D.
 "Traitement des contours en reconnaissance de formes visuelles :
 Application en robotique"
 Thèse de Docteur ès-Sciences Techniques, E.P.F. de Lausanne,
 Suisse, 1980

/35/ NARENDRA, P.M.
 "A seperable median filter for image noise smoothing"
 I.E.E.E. Trans. Pattern Anal. & Machine Intell., Vol. PAMI-3,
 N° 1, pp 20-29, 1981.

# **CHAPITRE IV**

# DETECTION ET EXTRACTION DES CONTOURS DES FONCTIONS DE DENSITE DE PROBABILITE

## **IV.1. INTRODUCTION**

Les outils présentés dans le chapitre précédent nous ont permis d'estimer la fonction de densité de probabilité sous-jacente par un algorithme rapide utilisant la méthode du noyau. Une généralisation du filtre médian au cas multidimensionnel a été utilisée pour lisser ces fonctions estimées, tout en préservant leurs contours.

Au cours de ce chapitre, on montre qu'en généralisant des concepts bien établis en traitement des images numériques /1/, /2/, /3/, la mise en évidence des modes peut être décomposée en deux étapes.

La première, appelée détection des contours, consiste à identifier les points d'échantillonnage au voisinage desquels la fonction de densité sous-jacente présente des variations locales importantes. Cette phase est basée sur l'utilisation d'opérateurs différentiels sensibles aux variations locales d'amplitude de la fonction de densité estimée et filtrée.

L'extraction des contours constitue la seconde phase de la procédure présentée dans ce chapitre. Il s'agit de reconstituer les contours à partir des points d'échantillonnage caractérisés par les fortes variations de la fonction de densité.

## **IV.2. DETECTION DES CONTOURS**

#### IV.2.1. Principe

En traitement des images numériques, les opérateurs différentiels sont largement utilisés pour localiser les contours des régions présentes dans une scène. Ces opérateurs donnent en fait des approximations du gradient de la fonction image et sont donc très sensibles aux variations locales des niveaux de gris. Il existe de nombreux opérateurs de type gradient, connus sous différents noms : L'opérateur de Roberts /4/, de Prewitt /5/, de Sobel /6/, etc....

D'autres auteurs ont proposé des opérateurs basés sur le concept de Laplacien /7/ qui est une double dérivation de la fonction du niveau de gris de l'image.

En général, ces opérateurs n'ont été utilisés que sur des données bidimensionnelles. La généralisation de certains de ces opérateurs pour la détection des contours au cas multidimensionnel a été stimulée par le développement des méthodes de traitement des images numériques tridimensionnelles dans le domaine médical /8/, /9/.

Ces opérateurs de type gradient peuvent être adaptés à des données multidimensionnelles pour détecter les contours des modes des fonctions de densité.

#### IV.2.2. Généralisation de l'opérateur de Roberts

L'opérateur différentiel symétrique établi par Roberts /4/ permet d'estimer le gradient en tout point d'une image bidimensionnelle en utilisant les lignes de balayage.

Plus précisèment, soit  $P_r = (p_{r,1} p_{r,2})$  un point image où le niveau de gris est  $f(P_r)$ . L'estimation du gradient de la fonction "niveau de gris" par application

de l'opérateur de Roberts est donnée sous la forme :

(4.1.) 
$$\tilde{G}(P_r) = |f(p_{r,1}, p_{r,2}) - f(p_{r,1} + 1, p_{r,2} + 1)|$$
  
+  $|f(p_{r,1}, p_{r,2} + 1) - f(p_{r,1} + 1, p_{r,2})|$ 

Cet opérateur simple, illustré sur la figure IV.1., a été généralisé pour la détection des contours sur des images tridimensionnelles obtenues en radiologie par des moyens optiques ou ultrasonores /10/, /11/.



Figure IV.1. : Procédure d'estimation du gradient à deux dimensions par application de l'opérateur de Roberts.

.

Cette généralisation peut être poussée dans l'espace à N dimensions où les points  $P_r$ , r = 1, 2, ..., R, représentent les points d'échantillonnage d'une fonction de densité estimée et filtrée  $\hat{f}(P_r)$ .

L'opérateur différentiel multidimensionnel nécessite le calcul de N(N-1)/2 gradients élémentaires définis par :

$$(4.2.) \quad \hat{G}_{\alpha,\beta}(P_{r}) = |\hat{f}(P_{r,1}, \dots, P_{r,\alpha}, \dots, P_{r,\beta}, \dots, P_{r,N})|$$

$$\stackrel{\alpha \neq \beta}{=} - \hat{f}(P_{r,1}, \dots, P_{r,\alpha} + 1, \dots, P_{r,\beta} + 1, \dots, P_{r,N})|$$

$$+ |\hat{f}(P_{r,1}, \dots, P_{r,\alpha}, \dots, P_{r,\beta} + 1, \dots, P_{r,N})|$$

$$- |\hat{f}(P_{r,1}, \dots, P_{r,\alpha} + 1, \dots, P_{r,\beta}, \dots, P_{r,N})|$$

L'estimation du gradient au point  $P_r$  peut alors être obtenue sous la forme :

(4.3) 
$$\hat{G}(P_r) = \sum_{\alpha=1}^{N} \sum_{\beta=1}^{N} |\hat{G}_{\alpha,\beta}(P_r)|$$
 avec  $\alpha \neq \beta$ 

Cette formulation est une extension directe de l'opérateur défini par Roberts pour les images à deux dimensions. Notons que pour déterminer le module du gradient en chaque point de l'espace des données, on a utilisé la somme des valeurs absolues dont l'implantation sur calculateur numérique est beaucoup plus aisée et rapide que la norme euclidienne.

Les précautions prises pour accélèrer l'estimation non paramétrique de la fonction de densité sont également nécessaires pour déterminer le gradient de cette fonction. Les techniques de calcul utilisées dans l'algorithme d'estimation rapide sont immédiatement adaptables. L'opérateur différentiel utilise le réseau d'hypercubes assurant la discrétisation de l'espace des données. Il n'est appliqué que dans les régions non vides de l'espace des données.

### IV.2.3. Généralisation de l'opérateur de Prewitt

Un autre type d'opérateur différentiel couramment utilisé en traitement des images numériques a été obtenu en modélisant localement le niveau du gris d'une image par une fonction analytique au sens des moindres carrés /5/. Le gradient déterminé à partir de ce modèle local fournit une approximation du module du gradient de la fonction image elle-même. La modélisation par un hyperplan conduit à une généralisation de cet opérateur, connu sous le nom d'opérateur de Prewitt /12/.

L'adaptation de cette approche à l'estimation du gradient de la fonction de densité de probabilité sous-jacente à la distribution des observations est immédiate.

Soit  $H_{\delta}(P_r)$  r = 1, 2, ..., R le voisinage hypercubique du point d'échantillonnage  $P_r$  défini sous la forme :

(4.4.)  $H_{\delta}(P_{r}) = \{(x_{1}, \ldots, x_{N}) / p_{r,n} - \delta_{n} \leq x_{n} \leq p_{r,n} + \delta_{n}, n = 1, 2, \ldots, N\}$ où  $\delta = (\delta_{1}, \delta_{2}, \ldots, \delta_{n}, \ldots, \delta_{N})$  indique la taille du voisinage.

La translation du point  $P_r$  à l'origine ne modifie pas le gradient en ce point. Dans ce cas,  $P_r$  devient  $P_0 = (0, 0, ...., 0)$  et le voisinage  $H_{\delta}(P_r)$  s'écrit sous la forme :

$$(4.5.) H_{\delta}(P_0) = \{(x_1, x_2, \dots, x_N) / -\delta_n \leq x_n \leq +\delta_n, n = 1, 2, \dots, N\}$$

L'hyperplan

(4.6) 
$$\pi(x_1, x_2, \dots, x_N) \equiv a_0 + \sum_{n=1}^N a_n x_n$$

qui modélise la fonction  $f(x, x_2, ..., x_N)$  au voisinage de P<sub>0</sub> est déterminé en minimisant l'expression :

(4.7.) 
$$E(a_0, a_1, \dots, a_N) = \int_{H_{\delta}(P_0)} \left[ \pi(x_1, x_2, \dots, x_N) - \dots + \pi_{\delta}(P_0) - f(x_1, \dots, x_N) \right]^2 dx_1 dx_2 \dots dx_N$$

Les valeurs des coefficients  $a_n$  qui minimisent l'erreur  $E(a_0, a_1, ..., a_N)$  doivent vérifier les conditions nécessaires d'extrémalité:

(4.8.) 
$$\frac{\partial E(a_0, a_1, \dots, a_N)}{\partial a_n} = 0$$
,  $n = 0, 1, 2, \dots, N$ 

Tenant compte de l'équation (4.7.), on peut écrire :

(4.9.) 
$$\frac{\partial E}{\partial a_n} = \frac{\partial}{\partial a_n} \int_{H_{\delta}(P_0)} [\pi(x_1, \ldots, x_N) - f(x_1, \ldots, x_N)]^2 dx_1 dx_2 \ldots dx_N$$

Pour  $a_0$ , la solution de l'équation (4.8.) est  $a_0 = f(0)$ , de telle sorte qu'au point  $P_r$ ,  $a_0$  est donné par l'équation :

$$a_0 = f(P_r)$$

Les composantes du gradient de la fonction  $\pi(x_1, x_2, ..., x_N)$  sont données par la solution des équations

$$\frac{\partial E}{\partial a_n} = 0$$
,  $n = 1, 2, \dots, N$ 

L'équation (4.9.) s'écrit :

(4.10.) 
$$\int_{H_{\delta}(P_0)} x_n \left[ \pi(x_1, \dots, x_N) - f(x_1, \dots, x_N) \right] dx_1 dx_2 \dots dx_N = 0$$

Posons :

(4.11.) 
$$\gamma(x_1, \ldots, x_{n-1}, x_{n+1}, \ldots, x_N) = \pi(x_1, \ldots, x_N) - a_n x_n$$

où  $\forall$ (.) est une fonction réelle qui, par construction, ne dépend pas de  $x_n$ .

L'équation (4.10) se met alors sous la forme :

/

$$(4.12.) \int_{H_{\delta}(P_{0})} x_{n} | \gamma(x_{1}, \dots, x_{n-1}, x_{n+1}, \dots, x_{N}) + a_{n} x_{n} - \dots$$

$$\dots - f(x_{1}, \dots, x_{N}) | dx_{1} dx_{2} \dots dx_{N} = 0$$
ou encore
$$(4.13.) \int_{H_{\delta}(P_{0})} a_{n} x_{n}^{2} dx_{1} \dots dx_{N} = \int_{H_{\delta}(P_{0})} x_{n} f(x_{1}, \dots, x_{N}) dx_{1} dx_{2} \dots dx_{N} - A$$

$$uvec: A = \int_{H_{\delta}(P_{0})} x_{n} \gamma'(x_{1}, \dots, x_{n-1}, x_{n+1}, \dots, x_{N}) dx_{1} dx_{2} \dots dx_{N}$$

Cette dernière intégrale est identiquement nulle. On peut le montrer en la décomposant en un produit d' intégrales, sous la forme :

$$A = \int_{-\delta_1}^{+\delta_1} dx_{1} \cdots \int_{-\delta_{n-1}}^{+\delta_{n-1}} dx_{n+1} \cdots \int_{-\delta_N}^{+\delta_N} \int_{-\delta_n}^{+\delta_N} dx_{1} \cdots \int_{-\delta_n}^{+\delta_n} dx_{n} dx_{n}$$

Comme :

$$\int_{-\delta_n}^{+\delta_n} x_n \, dx_n = 0$$

on obtient finalement A = 0.

Les coefficients  $a_n$  sont finalement donnés par les équations :

(4.14.) 
$$a_n = \frac{\int_{H_{\delta}(P_0)} x_n f(x_1, \dots, x_N) dx_1 dx_2 \dots dx_N}{\int_{H_{\delta}(P_0)} x_n^2 dx_1 dx_2 \dots dx_N}$$

Dans l'espace discrétisé, les valeurs des fonctions de densité ne sont connues qu'aux points d'échantillonnage  $P_r$ , r = 1, 2, ..., R et le voisinage H  $(P_r)$ devient le voisinage  $U_m(P_r)$  défini au paragraphe III.3.3. du chapitre III, où m représente la taille du voisinage du point  $P_r$ , tel que

$$U_{m}(P_{r}) = \left[ (p_{r',1}, \dots, p_{r',n}, \dots, p_{r',N}) / p_{r,n} - m \leq p_{r',n} \leq p_{r,n} + m \right]$$
  

$$n = 1, 2, \dots, N \right]$$

Les coefficients a<sub>n</sub> peuvent alors s'écrire sous la forme :

(4.15.) 
$$a_n = \sum_{U_m(P_r)} p_{r',n} \hat{f}(p_{r',1}, \dots, p_{r',N}) / \sum_{U_m(P_r)} (p_{r',n})^2$$

avec

(4.16.) 
$$\sum_{U_{m}(P_{r})} (p_{r',n})^{2} = 2(1 + 2^{2} + ... + m^{2})$$

Rappelons que ces coefficients ont été obtenus en supposant le point  $P_r$  à l'origine. Dans le cas général où  $P_r$  est un point quelconque de  $\mathbb{R}^N$ , le coefficient  $a_n$  est donc proportionnel à une combinaison linéaire des valeurs des fonctions  $\hat{f}(P_r)$  aux points  $P_r \in U_m$  ( $P_r$ ) pondérées par les coordonnées de leurs positions par rapport au point  $P_r$ .

Le gradient au point P<sub>r</sub> peut alors être estimé sous la forme :

(4.17.) 
$$\hat{G}(P_r) = \sum_{n=1}^{N} |a_n|$$

Comme pour l'opérateur de Roberts, l'opérateur de Prewitt n'est appliqué que dans les régions non vides de l'espace des données, ce qui assure des gains de temps substanciels par rapport à une application systématique sur tout l'espace discrétisé. Pour illustrer ce qui précède, examinons le cas bidimensionnel, en utilisant le voisinage  $U_1(P_r)$  du point  $P_r = (p_{r,1}, p_{r,2})$ , (cf.figure IV.2.).

Dans ces conditions on a :

$$\sum_{U_1(P_r)} (p_{r',n})^2 = 6$$



Figure IV.2. : Procédure d'estimation du gradient bidimensionnel par application de l'opérateur de Prewitt.

Les fonctions de densité  $\widehat{f}(P_r)$  sont ainsi pondérées par les positions du point  $P_r$ , par rapport au point  $P_r$  selon le schéma de la figure IV.2.bis.



Figure IV.2.bis : (a) Pondération des fonctions de densité pour obtenir (a<sub>1</sub>)

(b) Pondération des fonctions de densité pour obtenir (a<sub>2</sub>)

La valeur approchée du gradient obtenue au point  $P_r$  est alors :

$$\hat{G}(P_r) = \frac{1}{6} |\hat{f}(p_{r,1} + 1, p_{r,2} + 1) + \hat{f}(p_{r,1} + 1, p_{r,2}) + \hat{f}(p_{r,1} + 1, p_{r,2} - 1) - \hat{f}(p_{r,1} - 1, p_{r,2} - 1) - \hat{f}(p_{r,1} - 1, p_{r,2}) - \hat{f}(p_{r,1} - 1, p_{r,2} + 1)| + \hat{f}(p_{r,1} + 1, p_{r,2} + 1) + \hat{f}(p_{r,1} - 1, p_{r,2} - 1)| + \hat{f}(p_{r,1} + 1, p_{r,2} - 1) - \hat{f}(p_{r,1} - 1, p_{r,2} - 1)| + \hat{f}(p_{r,1} - 1, p_{r,2} - 1)|$$

## IV.3. COMPORTEMENT DES OPERATEURS DIFFERENTIELS MULTIDIMENSIONNELS

Les performances des deux opérateurs différentiels généralisés dans le cadre de la détection des contours des fonctions de densité de probabilité sont analysées en utilisant des données générées artificiellement.

#### Exemple 1

Le comportement des deux opérateurs est d'abord analysé sur des données bidimensionnelles constituées des 600 observations de la figure (IV.3.), issues d'un mélange de trois composantes normales équiprobables. A ces données on ajoute un bruit uniforme qui corrrespond à un niveau de 57 %. Le tableau (IV.1.) résume les paramètres statistiques de ces observations. Cet exemple a été traité au chapitre précédent où on a présenté la fonction de densité estimée brute et sa version filtrée (cf. figure III.5.(a) et III.5.(b)).

La figure (IV.4) indique la réponse des deux opérateurs proposés. On constate que les contours des modes sont matérialisés par des zones où ces réponses sont très sensiblement supérieures à celles obtenues dans les autres zones.

La comparaison systématique des performances des opérateurs différentiels a fait l'objet d'un certain nombre de travaux en traitement des images numériques /13/, /14/, /15/, /16/. Sans faire appel à ces méthodes de quantification, un simple examen visuel des résultats présentés sur la figure (IV.4) montre que l'opérateur de Prewitt permet d'obtenir des contours des modes plus larges et plus réguliers que ceux résultant de l'utilisation de l'opérateur de Roberts qui apparaît plus sensible aux irrégularités de la distribution des observations. Ces résultats concordent avec ceux obtenus par les praticiens en traitement des images numériques /16/, /17/, /18/. L'opérateur de Roberts, qui ne met en jeu que 4 voisins, à tendance à fournir des points de contour non significatifs associés aux bruits et aux irrégularités locales, alors que l'opérateur de Prewitt, qui prend en compte un voisinage plus étendu permet, en général, d'isoler les points de contour de leur environnement.

L'exemple tridimensionnel suivant montre que cette plus grande robustesse de l'approche de Prewitt par rapport à celle de Roberts se retrouve lors de la détection des contours des fonctions de densité multidimensionnelles.

Distribution:	Vecteurs Moyenne	Probabilités a priori	
1	$\overline{\mathbf{X}}_{1} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$	$S_1 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	$Pb_1 = 1/3$
2	$\overline{X}_2 = \begin{bmatrix} 15\\4 \end{bmatrix}$	$S_2 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	$Pb_2 = 1/3$
3	$\overline{\mathbf{x}}_{3} = \begin{bmatrix} 10\\15 \end{bmatrix}$	$S_3 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	$Pb_3 = 1/3$

Tableau 1: Paramètres statistiques de la distribution<br/>de l'exemple 1.





Figure IV.3. : Représentation graphique des observations de l'exemple 1.











## Exemple 2

Le second exemple utilisé pour évaluer les performances des deux opérateurs différentiels est constitué d'observations tridimensionnelles provenant de deux distributions normales dont les paramètres statistiques sont consignés dans le tableau 2. A ces deux groupements, représentés chacun 500 observations, on ajoute un bruit uniforme qui correspond à un niveau de 50 % (cf. Figure IV.5.).

Dans l'espace tridimensionnel, l'estimation du gradient par l'opérateur de Roberts au point  $P_r = (p_{r,1}, p_{r,2}, p_{r,3}), r = 1, 2, ..., R$ , est donnée sous la forme :

$$\hat{G}(P_{r}) = |\hat{G}_{1,2}(P_{r})| + |\hat{G}_{2,3}(P_{r})| + |\hat{G}_{1,3}(P_{r})|$$

avec

$$\hat{G}_{1,2}(P_r) = |\hat{f}(p_{r,1}, p_{r,2}, P_{r,3}) - \hat{f}(p_{r,1}^{+1}, p_{r,2}^{+1}, p_{r,3})| + |\hat{f}(p_{r,1}^{+1}, p_{r,2}, p_{r,3}) - \hat{f}(p_{r,1}, p_{r,2}^{+1}, p_{r,3})| \hat{G}_{2,3}(P_r) = |\hat{f}(p_{r,1}, p_{r,2}, p_{r,3}) - \hat{f}(p_{r,1}, p_{r,2}^{+1}, p_{r,3}^{+1})| + |\hat{f}(p_{r,1}, p_{r,2}^{+1}, p_{r,3}) - \hat{f}(p_{r,1}, p_{r,2}, p_{r,3}^{+1})| \hat{G}_{1,3}(P_r) = |\hat{f}(p_{r,1}, p_{r,2}, p_{r,3}) - \hat{f}(p_{r,1}^{+1}, p_{r,2}, p_{r,3}^{+1})| + |\hat{f}(p_{r,1}^{+1}, p_{r,2}, p_{r,3}) - \hat{f}(p_{r,1}, p_{r,2}, p_{r,3}^{+1})|$$

Les réponses obtenues par l'opérateur de Roberts sont présentées figure (IV.6(a)) sous la forme d'une série de plans parallèles et équidistants, perpendiculaires à l'axe  $0x_3$ . Pour bien faire apparaître les contours, seules les valeurs du gradient au-dessus d'un certain seuil sont représentées. Par superpositions de ces plans on peut reconstituer les contours fermés des modes.

Distribution	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori	
1	$\overline{\mathbf{X}}_{1} = \begin{bmatrix} 4\\ 4\\ 4 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$	Pb <sub>1</sub> = 1/2	
2	$\overline{X}_{2} = \begin{bmatrix} 12\\12\\8\\ \end{bmatrix}$	$S_{2} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$	$Pb_2 = 1/2$	

<u>Tableau 2</u> : Paramètres statistiques de la distribution de l'exemple 2.









En faisant appel à l'opérateur de Prewitt, les coefficients  $a_n$  donnés par l'équation (4,15) sont obtenus en utilisant un voisinage  $U_1(P_r)$  de telle sorte que :

$$\sum_{U_1(P_r)} (p_{r',n})^2 = 18$$

Les coefficients de pondération nécessaires au calcul de l'estimateur sont obtenus à partir des coordonnées des voisins par rapport au point  $P_r$ , suivant les trois directions de l'espace (cf. figure IV.7.).

D'une manière analogue à l'opérateur de Roberts, la figure (IV.6(b)) illustre les réponses :

$$\widehat{G}(P_r) = \sum_{n=1}^{3} |a_n|$$

de l'opérateur de Prewitt dans une série de plans équidistants et orthogonaux à l'axe  $0x_3$ . Par superposition de cette série de plans, on peut reconstituer les contours fermés des deux modes détectés.

L'examen visuel de ces différents résultats confirme les conclusions issues de l'exemple précédent. Avec l'opérateur de Roberts, l'estimateur du gradient présente des variations importantes au niveau des contours alors que les réponses de l'opérateur de Prewitt sont plus homogènes et d'une plus grande régularité tout au long des contours.

Si en traitement des images numériques, l'opérateur de Roberts est fréquemment utilisé pour des contraintes de calcul en temps réel, une telle préoccupation ne concerne pas en général les problèmes de classification. C'est la raison pour laquelle nous ne ferons appel dans la suite qu'à l'opérateur de Prewitt qui assurera une meilleure fiabilité à la procédure d'extraction des contours présentée au paragraphe suivant.







Figure IV.7. :

- (a) Pondération des fonctions de densité pour obtenir a<sub>1</sub>
- (b) Pondération des fonctions de densité pour obtenir a<sub>2</sub>
- (c) Pondération des fonctions de densité pour obtenir a 3





Figure IV.6.(a)



## **IV.4. EXTRACTION DES CONTOURS**

On peut définir le contour d'un mode comme une hyper-surface fermée constituée de points connexes caractérisés par de fortes réponses des opérateurs différentiels présentés au paragraphe (IV.2.) /19/, /20/, /21/. Il s'agit donc d'extraire de l'ensemble des points d'échantillonnage où le gradient de la fonction de densité sous-jacente a été estimé, des sous-ensembles connexes répondant à cette définition.

## IV.4.a. Principe d'extraction

L'extraction des contours est basée sur une procédure d'agrégation séquentielle omnidirectionnelle initialisée au point d'échantillonnage  $P^0$  où le gradient est le plus élevé. L'algorithme d'extraction recherche dans le voisinage  $U_m(P^0)$  d'autres points appartenant au même contour que  $P^0$ . Si m = 1, seuls les voisins immédiats de  $P^0$  sont pris en considération. Des valeurs de m supérieur à l'unité permettent de tolérer la présence de "trous" dans les contours.

Soit  $P_r$  un point courant d'un contour obtenu au cours de la procédure séquentielle d'extraction. On examine les valeurs du gradient de la fonction de densité au niveau des points  $P_r$ , qui appartiennent au voisinage  $U_m(P_r)$ . Seuls les points  $P_r$ , pour lesquels le gradient est suffisamment élevé appartiendront au même contour que  $P_r$ . Plus précisèment, soit  $\widehat{G}(P_r)$  l'estimation de la valeur du gradient au point  $P_r$  et soit  $\alpha$  un coefficient ajustable tel que  $\alpha \in \left[0, 1\right]$ .

Seuls les points  $P_r$ ,  $\in U_m(P_r)$  qui satisfont la condition :

$$\hat{G}(P_r) \ge \alpha. \hat{G}(P^0)$$

avec

$$\widehat{G}(P^0) = \max_{r} \{\widehat{G}(P_r)\}$$

sont rattachés au contour auquel appartient le point  $P_r$ . Ces points forment un nouvel ensemble de points courants sur lesquels la procédure d'agrégation est réitérée à l'étape suivante.

Dans cette procédure d'agrégation, un point courant dont on examine les voisins est nécessairement un voisin d'un point examiné dans une étape précédente. Tout nouveau point accepté ne doit pas être un point déjà affecté à un contour dans une étape antérieure. A chaque étape, tous les points courants disponibles doivent être pris en considération afin de sélectionner le plus grand nombre de points courants possibles pour l'étape suivante.

Lorsqu'à une étape donnée il n'est pas possible de trouver un seul nouveau point à rattacher au contour en cours d'extraction, le processus d'agrégation est terminé pour ce contour.

La recherche d'un nouveau contour s'effectue alors à partir du point d'échantillonnage présentant le plus fort gradient parmi les points non encore assignés à un contour. Le processus d'extraction se poursuit jusqu'à ce qu'il n'existe plus de points répondant au critère d'agrégation. Un nouveau contour est alors défini et l'extraction est réitérée à partir du point du gradient le plus élevé parmi les points non encore sélectionnés.

Ce processus d'extraction des contours est itéré jusqu'à ce que tous les points de l'espace où le gradient a été estimé soient incorporés à des contours. Cette technique d'extraction tend à générer des contours non significatifs constitués d'agrégats de points d'échantillonnage où le gradient est relativement élevé. Après avoir imposé une valeur minimale du gradient du point initial de chaque contour, tous les agrégats obtenus par la procédure d'extraction sont analysés. En rappelant que les points d'échantillonnage sont les centres d'hypercubes de côté unité introduits à des fins de discrétisation, il est facile d'évaluer le nombre d'observations situées à l'intérieur de chaque agrégat. Seuls sont considérés comme des contours les agrégats peur lesquels le nombre d'observations situées dans les hypercubes qui les constituent représente une fraction significative de l'ensemble des observations disponibles. Les agrégats non significatifs, caractérisés par un petit nombre d'observations, sont facilement rejetés.

#### IV.4.2. Exemple

L'application de l'algorithme d'extraction des contours est illustrée sur des données bidimensionnelles constituées de trois distributions normales nonéquiprobables. Les paramètres statistiques de chacune des trois distributions, en présence sont indiqués dans le tableau 3. A ces trois distributions ont a ajouté un bruit constitué d'observations uniformément distribuées sur tout l'espace. Le niveau du bruit dans cet échantillon est de 50 % (cf. figure IV.8.).

La figure (IV.9.) représente les contours obtenus en indiquant leur ordre d'extraction par l'algorithme. La taille du voisinage  $U_m(P_r)$  utilisé est fixée à m=1 et le coefficient  $\propto$  introduit dans la procédure d'extraction a été ajusté à la valeur  $\propto = 0.5$ . Pour tous les exemples présentés dans ce mémoire, l'algorithme d'extraction a été appliqué sur les valeurs du gradient estimé par l'opérateur de Prewitt avec  $\propto = 0.5$  et m = 1. Malgré la variété des exemples traités, aucun réajustement de ces paramètres n'a donc été nécessaire.

Distribution	Vecteurs Moyenne	Matrices de Covariance,	Probabilités a priori	
1	$\overline{\mathbf{X}}_{1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	Pb <sub>1</sub> = 0,23	
2	$\overline{\mathbf{X}}_2 = \begin{bmatrix} 0 \\ 12 \end{bmatrix}$	$S_2 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	Pb <sub>2</sub> = 0,30	
3	$\overline{X}_3 = \begin{bmatrix} 12\\5 \end{bmatrix}$	$S_3 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	Pb <sub>3</sub> = 0,47	

Tableau 3	:	Les	paramètres	statistiques	de	la	distribu-
		tion	ı de l'exemp	ple 3.			





Figure.IV.8.: Distribution des observations dans l'espace à deux dimensions de l'exemple 3.





Figure IV.9. : Les contours des fonctions de densité obtenus par l'algorithme d'extraction suivant l'ordre de leur apparition.

Les chiffres indiquent les valeurs estimées du gradient, codées entre 0 et 10 •



(1)

#### **IV.5. CONCLUSION**

Dans ce chapitre nous avons mis en évidence les contours de la fonction de densité sous-jacente à la distribution d'un ensemble d'observations, par application de deux opérateurs différentiels, initialement développés en traitement des images numériques.

La généralisation de ces opérateurs a permis de détecter les modes des fonctions de densité, par l'intermédiaire de leurs contours.

Nous allons maintenant montrer comment la mise en évidence de ces contours permet d'aborder des problèmes de classification automatique tant sous l'hypothèse paramétrique que sous l'hypothèse non paramétrique.

## REFERENCES IV

- /1/ ROSENFELD, A. et THURSTON, N. "Edge and Curve detection for visual scene analysis" I.E.E.E., Trans. Comput., Vol. C-20, pp 562-569, 1971
- /2/ DAVIS, L.S.
   "A survey of detection techniques"
   Comput. Graphics and Image Processing, N° 4, pp 248-270, 1975
- /3/ BALLARD, D.H. et BROWN, C.M. "Computer Vision" Prentice Hall, Englewood Cliffs, NJ, 1982
- /4/ ROBERTS, L.G.
   "Machine Perception of three dimensional solids, in optical and electro-optical information processing"
   T.J. Tippet, et al. Eds, M.I.T. Press, Cambridge, Mass., 1965
- /5/ PREWITT, J.M.S.
   "Object enhancement and extraction, in Picture Processing and Psychopictorics"
   B.S. Lipkin et A. Rosenfeld, Eds, Academic Press, New-York, pp 75-149, 1970
- /6/ IANNINO, A. et SHAPIRO, S.D. "An iterative generalized of the Sobel edge detection operator" Proc. I.E.E.E., Conf. Pattern Recognition and Image Processing, pp 130-137, 1979
- /7/ MODESTINO, J.W. et FRIES, R.W. "Edge detection in noisy images using recursive digital filtering" Comput. Graphics Image Processing, N° 6, pp 409-433, 1977
- /8/ ZUCKER, S.W. et HUMMEL, R.A. "A three-dimensional edge operator" I.E.E.E. Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, N° 3, pp 324-331, 1981
- /9/ KENDALL, P.J.
   "Application of pattern recognition to medical data analysis"
   Applications of Pattern Recognition, K.S. Fu Editor, CRC Press,
   Inc., pp 169-195, 1982

- /10/ TOUZANI, A. et POSTAIRE, J.G.
   "Clustering by mode boundary detection"
   Article soumis à I.E.E.E. Trans. Pattern Anal. Machine
   Intell. 1987
- /11/ LIU, H.K.
   "Two and three-dimensional boundary detection"
   Comput. Graphics and Image Processing, Vol. 6, pp 123-134, 1977
- /12/ HERMAN, G.T. et LIU, H.K.
   "Dynamic boundary surface detection"
   Comput. Graphics and Image Processing, Vol. 7, pp 130-138, 1978
- /13/ MORGENTHALER, D.G. et ROSENFELD, A. "Multidimensional edge detection by hypersurface fitting" I.E.E.E.-Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, N° 4 pp 482-486, 1981
- /14/ ABDOU, I.E. et PRATT, W.K.
   "Quantitative desing and evaluation of enhancement/Thresholding
   edge detection"
   Proc. I.E.E.E., Vol. 67, N° 5, pp 753-763, 1979
- /15/ DI ZENZO, S.
   "A note on the gradient of multi-image"
   Computer Vision, Graphics, and Image Processing, Vol. 33,
   pp 116-125, 1986
- /16/ KITCHEN, L. et ROSENFELD, A.
   "Edge evaluation using local edge coherence"
   I.E.E.E., Trans. Syst. Man. Cybern., Vol. SMC-11, N° 9,
   pp 597-605, 1981
- /17/ PELI, T. et MALAH, D.
   "A study of edge detection algorithms"
   Comput. Graphics and Image Processing, Vol. 20, pp 1-21, 1982
- /18/ PREVOL, C.
   "Reconnaissance artificielle des formes sur les méthodes de
   détection de contours"
   Le Nouvel Automatisme, pp 47-53, mars 1982
- /19/ MORGENTHALER, D.G. et ROSENFELD, A. "Surfaces in three-dimensional Digital Images" Information and Control, 51, pp 227-247, 1981
- /20/ UDUPA, J.K., SRIHARI, S.N. et HERMAN, G.T. "Boundary detection in multidimensions" I.E.E.E. Trans. Pattern Anal. Machine Intell., Vol. PAMI-4, N° 1, pp 41-50, 1982
- /21/ ARTZY, E., FRIEDER, G. et HERMAN, G.T.
   "The theory, desing, implementation and evaluation of a three dimensional surface detection algorithm"
   Computer Graphics and Image Processing, Vol. 15, pp 1-24, 1981.

# CHAPITRE V

# APPLICATION DES TECHNIQUES DE RECHERCHE DES CONTOURS DES MODES EN CLASSIFICATION AUTOMATIQUE

## V.1. - INTRODUCTION

Nous avons introduit, dans la première partie de ce mémoire, la notion de contour d'une fonction de densité de probabilité.

Dans une première partie (V.A.) de ce chapitre, nous allons montrer comment exploiter la procédure de détection et d'extraction du contour d'une fonction de densité pour identifier les paramètres statistiques d'une distribution normale d'observations multivariables, en n'utilisant aucune autre information que celle apportée par les observations. L'adaptation de cette méthode d'identification pour déterminer les paramètres des mélanges gaussiens permettra d'aborder le problème de la classification automatique sous l'hypothèse paramètrique normale en approchant la distribution des observations disponibles par un modèle probabiliste.

Dans une deuxième partie (V.B.) du chapitre on adapte la procédure de détection et d'extraction des contours à la recherche des modes sous l'hypothèse non paramètrique pour déboucher sur une nouvelle méthode de recherche des groupements.

# V.A. - CLASSIFICATION PARAMETRIQUE

# V.A.1. DETERMINATION DES PARAMETRES D'UNE DISTRIBUTION NORMALE MULTIDIMENSIONNELLE

Nous avons montré au chapitre II que le contour d'une fonction de densité de probabilité normale est un hyperellipsoïde dont les caractéristiques géomètriques sont liées par des relations simples aux paramètres de la distribution, c'est-à-dire au vecteur moyenne et à la matrice de covariance.

L'utilisation du schéma, décrit en 4 étapes, pour l'identification d'une distribution normale à partir d'un échantillon d'observations nécessite, en premier lieu, la localisation du contour de la fonction de densité sous-jacente.

## V.A.1.1. Localisation du contour d'une fonction de densité normale

Rappelons que le gradient d'une fonction de densité peut être déterminé par l'application de l'opérateur différentiel multidimensionnel de Prewitt, défini au chapitre précédent, sur une version estimée et filtrée de cette fonction. Une fois déterminé, le gradient en chaque point de l'espace des données discrétisé au pas  $h_Q$ , un algorithme d'extraction permet de reconstituer les contours sous la forme d'ensemble de points de discrétisation connexes où le gradient présente des valeurs élevées.

Dans la pratique, en appliquant cette technique de détection des contours à des observations distribuées selon une loi normale (cf. figure V.1.), on n'obtient qu'une approximation ( $\hat{\Gamma}$ ) de l'hyperellipsoïde idéal ( $\Gamma$ ) définissant ce contour, constituée de l'ensemble des points ( $\hat{\Gamma}$ ) isolés par la procédure d'extraction (cf. figure V.2.(a)). En effet, comme nous l'avons déjà montré, les opérateurs différentiels sont sensibles aux irrégularités de la distribution des observations. A ces sources d'imprécision viennent de plus s'ajouter les effets d'une discrétisation spatiale dont la finesse ne permet pas toujours de rendre compte de la forme exacte de l'hyperellipsoïde constituant le contour.







Figure V.2.(b)

Figure V.2. :

- Contour réel de la fonction de densité de la figure V.1. ([):
- $(\hat{\Gamma})$  : Contour extrait
- $(\Gamma^*)$  : Modèle ellipsoîdal du contour  $(\widehat{\Gamma})$



Afin d'obtenir un modèle aussi proche que possible du contour réel ( $\int$ ) définissant la distribution des observations, nous allons procéder à une modélisation de l'ensemble des points constituant ( $\hat{f}$ ).

#### V.A.1.2. Modélisation du contour

Soient  $B_j$ ,  $j = 1, 2, ..., J_{\uparrow\uparrow}$ , les centres des hypercubes adjacents constituant l'approximation du contour ( $\hat{\Gamma}$ ) de la fonction de densité :

$$p(X) = \frac{1}{(2\pi)^{\frac{N}{2}} |S|^{1/2}} \exp \left[-\frac{1}{2} (X - \bar{X})^{T} S^{-1} (X - \bar{X})\right]$$

On considère l'ensemble ( $\hat{I}$ ) des  $J_{\underline{I}}$  hypercubes intérieurs au contour ( $\hat{\Gamma}$ ). Cet ensemble est obtenu par une procédure de chaînage initialisée au barycentre de ( $\hat{\Gamma}$ ). L'ensemble ( $\hat{I}$ ) ainsi obtenu est fusionné avec l'ensemble ( $\hat{\Gamma}$ ) pour constituer un domaine ( $\hat{D}$ ) qui n'est autre qu'une approximation du domaine de concavité(D) de la fonction de densité /1/.

On sait que les coordonnées du centre d'inertie  $\hat{O}$  du domaine ( $\hat{D}$ ) constituent une approximation du vecteur moyenne  $\overline{X}$  de la distribution normale.

Soient  $B_j$ , j = 1, 2, ..., J tel que  $J = J_{\Gamma} + J_I$ , les centres des hypercubes connexes constituant le domaine ( $\hat{D}$ ). Le calcul des produits d'inertie pour l'ensemble des points  $B_{j,j} = 1, 2, ..., J$ , par rapport au centre d'inertie  $\hat{O}$  permet de déterminer le tenseur central d'inertie du domaine ( $\hat{D}$ ), considéré comme un solide constitué de particules de masse unité situées aux points  $B_{j,j} = 1, 2, ..., J$ .

La recherche des axes principaux d'inertie du domaine ( $\hat{D}$ ) se ramène alors à la recherche des vecteurs propres de ce tenseur d'inertie. Les vecteurs propres  $\overrightarrow{V_n}$ , n = 1, 2, ..., N de l'inverse de la matrice de covariance peuvent ainsi
être approchés par les vecteurs unitaires  $\overline{V_1}, \overline{V_2}, ..., \overline{V_n}, ..., \overline{V_N}$  des axes principaux d'inertie.

La longueur  $\hat{L}_n$  du domaine( $\hat{D}$ ) dans la nième direction principale est alors obtenue simplement en projetant les points  $B_j$ , j = 1, 2, ..., J, sur le nième axe principal.

$$\widehat{\mathbf{L}}_{n} = \max_{j} \overline{\mathbf{OB}}_{j}^{*} \overline{\mathbf{V}}_{n}^{*} - \min_{j} \overline{\mathbf{OB}}_{j}^{*} \overline{\mathbf{V}}_{n}^{*}$$

Une approximation  $\widehat{\lambda}_n$  de la nième valeur propre de l'inverse de la matrice de covariance est donnée par :

$$\hat{\lambda}_n = 4/\hat{L}_n^2$$

La matrice



avec  $\widehat{\mathbf{v}} = \left[\widehat{\mathbf{v}}_1, \, \widehat{\mathbf{v}}_2, \, \dots, \, \widehat{\mathbf{v}}_N\right]$ 

constitue une approximation de l'inverse de la matrice de covariance.

Le vecteur moyenne et la matrice de covariance d'une distribution normale peuvent donc être approchés en assimilant le contour détecté ( $\bigcap$ ) et son intérieur à un modèle hyperellipsoïdal ( $\bigcap$ <sup>\*</sup>). Cette approche est maintenant testée sur un exemple.

### V.A.1.3. - Exemple d'application

Le comportement de la méthode d'identification des paramètres d'une distribution normale développé ci-dessus est testé sur des données générées artificiellement.

Une distribution de fonction de densité bivariable de vecteur moyenne :

$$\overline{\mathbf{X}} = \begin{bmatrix} 5 \\ -5 \\ -5 \end{bmatrix}$$

et de matrice de covariance :

 $S = \begin{bmatrix} 7.0 & 0.0 \\ 0.0 & 3.0 \end{bmatrix}$ 

est représentée successivement par 100, 200 et 500 observations (cf. figure V.3.).

Les contours ( $\widehat{\Gamma}$ ) détectés après filtrage et application de l'opérateur de Prewitt sont assimilés à des modèles ellipsoïdaux ( $\Gamma^*$ ) définis par :

$$(\mathbf{X} - \mathbf{\hat{\overline{X}}})^{\mathrm{T}} \mathbf{\hat{S}}^{-1} (\mathbf{X} - \mathbf{\hat{\overline{X}}}) = 1$$

où  $\hat{\overline{X}}$  et  $\hat{\overline{S}}$  sont respectivement le vecteur moyenne et la matrice de covariance obtenus selon la méthode développée au V.A.1.2.





Figure V.3. Représentation graphique des observations :

(a) Q = 100(b) Q = 200(c) Q = 300



La figure V.4. montre que ces modèles ellipsoïdaux ( $[ ]^*$ ) sont d'autant plus proche du véritable contour réel ([ ]) que le nombre d'observations disponibles est important.



Ces résultats ont été obtenus pour des pas de discrétisation satisfaisant la relation :

$$h_Q = h_0 - \frac{1}{\sqrt{Q}}$$

qui assure la convergence uniforme de l'estimation lorsque le nombre Q d'observations disponibles augmente (cf. Tableau A.1.).

Nombre d'observation Q	100	200	500
Pas de discréti- sation h <sub>Q</sub>	1/12	1/17	1/28

#### Tableau A.1.

L'intérêt de cette méthode d'identification d'une distribution réside essentiellement dans ses possibilités d'application à l'analyse des mélanges gaussiens.

## V.A.2. - IDENTIFICATION DES MELANGES GAUSSIENS

Soit f(X) la fonction de densité de probabilité sous-jacente à la distribution d'un échantillon d'observations d'équation :

$$f(X) = \sum_{k=1}^{K} p(X/C_k) P(C_k)$$

$$p(X/C_{k}) = \frac{1}{(2\pi)^{\frac{N}{2}}|S_{k}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X-\bar{X}_{k})^{T}S_{k}^{-1}(X-\bar{X}_{k})\right]$$

est la fonction de densité conditionnelle normale de la classe  $C_k$  de vecteur moyenne  $\overline{X}_k$  et de matrice de covariance  $S_k \cdot P(C_k)$  est la probabilité d'apparition a priori de la classe  $C_k$  de telle sorte que :

$$P(C_k) \ge 0$$

$$\sum_{k=1}^{K} P(C_k) =$$

Le problème de l'identification des mélanges gaussiens se ramène à l'utilisation des observations distribués selon la loi f(X) pour déterminer :

- le nombre de classes K présentes dans l'échantillon

- la probabilité a priori  $P(C_k)$ 

1

- le vecteur moyenne  $\overline{X_k}$
- la matrice de covariance S<sub>k</sub>

de chaque classe

Si les composantes du mélange ne présentent pas un fort degré de chevauchement, il existe une correspondance bijective entre les modes de la fonction de densité sous-jacente et les classes en présence /2//3//4/. Dans ces conditions, la fonction de densité conditionnelle associée à chaque composante est prépondérante sur les autres composantes. On peut donc assimiler son contour à celui du mode correspondant.

En assimilant le contour réel ( $\bigcap_k$ ) de la composante  $P(X/C_k)$  et le contour ( $\bigcap_k'$ ) du mode du mélange correspondant, on pourra obtenir des valeurs

où :

approchées  $\widehat{X}_k$  du vecteur moyenne  $\overline{X}_k$  et  $\widehat{S}_k$  de la matrice de covariance  $S_k$  par la méthode d'analyse des contours exposés au  $\S$ .V.A.1.. Les résultats présentés au  $\S$ .V.A.3. montreront que les résultats ainsi obtenus seront d'autant meilleurs que le degré de chevauchement des classes en présence est réduit

Les derniers paramètres à déterminer restent donc les probabilités a priori les différentes composantes constituant le mélange.

Il a été montré que, la probabilité pour qu'une observation d'un échantillon normal distribué selon la fonction de densité sous-jacente :

$$P(X) = \frac{1}{(2\pi)\frac{N}{2}|S|\frac{1}{2}} \exp\left[-\frac{1}{2}(X - \overline{X})^{T}S^{-1}(X - \overline{X})\right]$$

appartienne au domaine intérieur(D)au contour réel ( $\square$ ) de la fonction de densité correspondante est une constante indépendante de la matrice de covariance de la distribution /5/.

Considérons maintenant le mélange :

$$f(X) = \sum_{k=1}^{K} p(X|C_k) P(C_k)$$

La probabilité pour qu'une observation d'un échantillon distribué selon la loi f(X) provienne de la classe  $C_k$  et appartienne au domaine  $(D_k)$  délimité par le contour  $(\bigcap_k)$  est proportionnelle à  $P(C_k)$ .

La fonction de densité  $p(X/C_k)$  étant supposée fortement prépondérante sur les autres composantes du mélange à l'intérieur du contour ( $\lceil k \rangle$ ), la plupart des observations appartenant à  $(D_k)$  proviennent de la classe  $C_k$ . La probabilité pour qu'une observation appartienne, quelle que soit la classe dont elle provient, au domaine  $(D_k)$ , peut-être approximée par la probabilité  $P(C_k)$ . Des valeurs approchées  $\hat{P}(C_k)$  des probabilités a priori  $P(C_k)$  peuvent alors être obtenues en assimilant les contours des composantes à ceux des modes du mélange et en résolvant le système d'équations :

$$\begin{pmatrix} \frac{\hat{P}(C_1)}{p(X \in D_1')} = \cdots = \frac{\hat{P}(C_k)}{p(X \in D_k')} = \cdots = \frac{\hat{P}(C_k)}{p(X \in D_k')} \\ \hat{P}(C_1) + \cdots + \hat{P}(C_k) + \cdots + \hat{P}(C_k) = 1 \end{pmatrix}$$

où (D'<sub>k</sub>), k = 1, 2, ..., K sont des domaines intérieurs aux contours des modes du mélange noté ( $\prod_{k}'$ ), k = 1, 2, ..., K.

En remplaçant le contour  $(\bigcap'_k)$  de chaque mode du mélange par le contour détecté  $(\bigcap'_k)$ , on peut remplacer les domaines  $(D'_k)$  par les domaines  $(\widehat{D}_k)$  intérieurs aux contours  $(\bigcap'_k)$ .

Soit  $q_k$  le nombre d'observations situées à l'intérieur de l'ensemble des hypercubes constituant le domaine ( $\hat{D}_k$ ) associé à la classe  $C_k$ . Le rapport  $q_k/Q$  est un estimateur classique de la probabilité  $p(X \in \hat{D}_k)$ . Dans ce cas le système d'équation à résoudre devient :

$$\left\{\begin{array}{c} \widehat{P}(C_1) \\ q_1 \\ \widehat{P}(C_1) + \dots + \widehat{P}(C_k) + \dots + \widehat{P}(C_k) + \dots + \widehat{P}(C_k) = 1\end{array}\right.$$

A ce stade, on dispose de valeurs approchées de tous les paramètres définissant le mélange. L'identification d'un mélange d'un nombre inconnu de composantes gaussiennes peut donc être entreprise selon le schéma directeur suivant :

- Estimation et filtrage des fonctions de densité de la distribution des observations disponibles
- Détection et extraction des contours des modes du mélange et identification du nombre de classes en présence dans l'échantillon.
- Modélisation des contours par des hyperellipsoïdes. Détermination des valeurs approchées du vecteur moyenne et de la matrice de covariance de chacune des classes mises en évidence.
- Détermination des probabilités a priori des classes constituant le mélange à partir du nombre d'observations situées à l'intérieur de chaque contour.

Les performances de cette méthode sont maintenant évaluées sur quelques exemples.

#### V.A.3 - RESULTATS EXPERIMENTAUX

Les observations utilisées pour les expériences sont des données générées artificiellement par calculateur. L'avantage de l'utilisation des données artificielles comme moyen de test, est que l'expérimentateur connaît la structure de ses données, en garde le contrôle et peut évaluer la qualité des résultats obtenus.

### Exemple A.1.

Les données utilisées dans ce premier exemple sont des observations issues de trois classes normales non sphériques et équiprobables, dont les paramètres statistiques sont consignés dans le tableau A.2.. La figure V.6. présente l'échantillon disponible, constitué de 200 observations par classe.

L'algorithme d'identification a été appliqué pour différentes valeurs du paramètre de discrétisation  $h_Q$  de l'espace des données  $R_1^N$ . Le choix de ce type de paramètre au milieu de la plus grande plage de variation pour laquelle le nombre de modes détectés demeure constant a déjà été utilisé pour optimiser certains algorithmes de classification /6/ /7/.

Pour l'exemple considéré le nombre de modes détectés est resté stable et égal à trois pour des valeurs de  $(1/h_Q)$  comprises entre [22-71] (cf. figure V.7.). Cette plage est exceptionnellement étendue si on la compare aux résultats obtenus avec d'autres algorithmes de classification /6/ /8/. Les résultats présentés figure V.8. et tableau A.2. ont été obtenus pour  $(1/h_Q) = 46$  c'est-à-dire au milieu de cette plage.





Les valeurs approximatives des probabilités a priori, vecteurs moyenne et matrices de covariance des trois composantes sont utilisées pour mettre en oeuvre une procédure de classement optimal, qui minimise le taux d'erreur /9/. En effet, le contexte probabiliste dans lequel nous nous sommes placés permet d'optimiser le processus de classification en partitionnant l'espace en région  $\mathcal{G}_k$ telle que :

$$x \in \mathcal{Q}_k \implies P(C_k/X) \ge P(C_i/X), i = 1, 2, \dots, K$$

Les fonctions de décision qui optimisent la procédure de classement peuvent être choisies de la forme /10//11/:

$$g_k(X) = p(X/C_k) P(C_k)$$

Pour les mélanges gaussiens, on choisit :

$$g_{k}(X) = -\frac{1}{2} \operatorname{Log} \left| S_{k} \right| - \frac{1}{2} \left( X - \overline{X}_{k} \right)^{T} S_{k}^{-1} \left( X - \overline{X}_{k} \right) + \operatorname{Log} P(C_{k})$$

L'optimisation du processus de classification consiste à remplacer, dans l'expression de  $g_k(X)$ , les valeurs exactes du vecteur moyenne  $\overline{X}_k$ , de la matrice de covariance  $S_k$  et de la probabilité a priori de la classe  $C_k$ ,  $P(C_k)$ , par leurs valeurs approchées  $\widehat{\overline{X}}_k$ ,  $\widehat{S}_k$  et  $\widehat{P}(C_k)$  obtenues par la procédure d'identification des mélanges.

Pour ce premier exemple le taux d'erreur minimum de classification obtenu avec les paramètres statistiques déterminés à partir des contours des mélanges est égal à 1,2 % pour un ajustement du paramètre du discrétisation  $h_Q$  à la valeur (1/46). La figure VI.9. indique quelques valeurs du taux d'erreur en fonction du paramètre  $h_Q$ . On constate que l'on reste très proche de la valeur théorique minimale pour  $(1/h_Q)$  variant de 38 à 65, ce qui assure une très grande robustesse de l'algorithme par rapport à l'ajustement de ce paramètre.

Les valeurs du taux d'erreur inférieures au taux théorique minimum proviennent de la méthode d'estimation du taux d'erreur réel. Celui-ci est en fait estimé par le rapport du nombre d'observations mal classées sur le nombre total d'observations constituant l'échantillon. Dans le cas d'échantillon fini, cet estimateur peut ne pas être très précis.







Classes	Vecteurs Moyenne	Matrices de Covariance	Probabilités à priori
1	$\overline{X}_{1} = \begin{bmatrix} 0\\ 0\\ 0 \end{bmatrix} ;  \widehat{\overline{X}}_{2} = \begin{bmatrix} 0, 51\\ 0, 30 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 10, 0 & 0, 0 \\ 0, 0 & 2, 5 \end{bmatrix},  \widehat{S}_{1} = \begin{bmatrix} 10, 56 & -0, 2 \\ -0, 2 & 2, 12 \end{bmatrix}$	P(C <sub>1</sub> )=0,33;P(C <sub>1</sub> )=0,34
2	$\overline{X}_{2} = \begin{bmatrix} 0\\\\10 \end{bmatrix} ;  \widehat{\overline{X}}_{2} = \begin{bmatrix} 0,70\\\\9,16 \end{bmatrix}$	$S_{2} = \begin{bmatrix} 10, 0 & 0, 0 \\ 0, 0 & 2, 5 \end{bmatrix},  \widehat{S}_{2} = \begin{bmatrix} 9, 8 & 0, 08 \\ 0, 08 & 2, 9 \end{bmatrix}$	P(C <sub>2</sub> )=0,33,P(C <sub>2</sub> )=0,33
3	$\overline{\mathbf{X}}_{3} = \begin{bmatrix} 11\\ \\ 5 \end{bmatrix} ;  \widehat{\overline{\mathbf{X}}}_{3} = \begin{bmatrix} 12\\ \\ 5,8 \end{bmatrix}$	$S_{3} = \begin{bmatrix} 2,5 & 0,0 \\ 0,0 & 10,0 \end{bmatrix}; \ \hat{S}_{3} = \begin{bmatrix} 2,25 & 0,12 \\ 0,12 & 11,1 \end{bmatrix}$	P(C <sub>3</sub> )=0,33;P(C <sub>3</sub> )=0,33
	Taux d'erreur minimum théorique	1,8 % Taux d'erreur ré	éel 1,2 %

Tableau A.2. : Valeurs exactes et approchées des paramètres statistiques de la distribution de l'exemple A.1.





Figure V.6. : Représentation graphique des 600 observations de l'example A.1.





.

Figure V.8.(a) : Réponse de l'opérateur différentiel de Prewitt aux données de l'exemple A.1.

•

.





Figure V.8.(b) : Modèles ellipsoïdaux des contours du mélange des trois lois normales de l'exemple A.1.





k, k = 1, 2, 3 Contours extraits des composantes du mélange

Contours réels des composantes du mélange

3 Modèles ellipsoïdaux des contours du mélange.

#### Exemple A.2.

Le second exemple est constitué de données bidimensionnelles extraites de trois classes équiprobables dont les paramètres statistiques apparaissent dans le tableau A.3.. Les 600 observations constituant l'échantillon sont représentées sur la figure V.10.. La réponse de l'opérateur de Prewitt appliquée à la fonction de densité estimée et filtrée et les contours correspondants sont représentés sur la figure V.11. Les valeurs approximatives des paramètres statistiques du mélange obtenues par la procédure d'identification sont consignées au tableau A.3.. L'écart entre le taux d'erreur obtenu et la valeur optimale reste faible, bien qu'il soit plus grand que pour l'exemple A.1., conséquence du chevauchement plus important des trois composantes.







Classes	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\overline{X}_{1} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} ;  \widehat{\overline{X}}_{1} = \begin{bmatrix} -0, 32 \\ -0, 59 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 13 & 0 \\ 0 & 4 \end{bmatrix};  \hat{S}_{1} = \begin{bmatrix} 13,97 & -0,5 \\ -0,5 & 4,12 \end{bmatrix}$	P(C <sub>1</sub> )=0,33;P(C <sub>1</sub> )=0,32
2	$\overline{\mathbf{X}}_{2} = \begin{bmatrix} 0 \\ 10 \end{bmatrix} \qquad \widehat{\overline{\mathbf{X}}}_{2} = \begin{bmatrix} -0, 62 \\ 9, 97 \end{bmatrix}$	$S_{2} = \begin{bmatrix} 13 & 0 \\ 0 & 4 \end{bmatrix},  \widehat{S}_{2} = \begin{bmatrix} 14,96 & -0,05 \\ 0 & -0,05 & 5,0 \end{bmatrix}$	P(C <sub>2</sub> )=0,33,P(C <sub>2</sub> )=0,35
3	$\overline{\mathbf{X}}_{3} = \begin{bmatrix} 11\\ \\ 5 \end{bmatrix} \qquad \widehat{\overline{\mathbf{X}}}_{3} = \begin{bmatrix} 10,88\\ \\ 5,23 \end{bmatrix}$	$S_{3} = \begin{bmatrix} 4 & 0 \\ 0 & 13 \end{bmatrix};  \widehat{S}_{3} = \begin{bmatrix} 4,88 & 1,1 \\ 1,1 & 15,21 \end{bmatrix}$	P(C <sub>3</sub> )=0,33,P(C <sub>3</sub> )=0,33
	Taux d'erreur minimum théorique	e 3,5 % Taux d'erreur r	réel 3,17 %

Tableau A.3. : Valeurs exactes et approchées des paramètres statistiques de la distribution de l'exemple A.2.





Figure V.11.(a) : Réponse de l'opérateur différentiel de Prewitt aux données de l'exemple A.2.







Figure V.11. : Contours et modèles ellipsoîdaux des contours du mélange des trois lois normales de l'exemple A.2.

k, k = 1, 2, 3 Contours extraits des composantes du mélange.

 $-\frac{1}{k}, k = 1, 2, 3$ 

Contours réels des composantes du mélange.

Modèles ellipsoîdaux des contours du mélange.

#### Exemple A.3.

L'exemple A.3. présenté est constitué de 1200 observations tridimensionnelles provenant de trois distribution normales dont les caractéristiques sont consignées dans le tableau A.4.. La figure V.12. représente l'échantillon ainsi constitué.

Pour visualiser les résultats de la procédure de détection et d'extraction des contours, on utilise une série de plans parallèles, équidistants et orthogonaux à l'axe  $Ox_3$  de l'espace des données. Les sections des contous par ces plans permettent une visualisation des modes détectés (cf. figure V.13.). Le taux d'erreur obtenu en utilisant les paramètres statistiques déduits de ces modèles (cf. tableau A.4.) reste très proche du taux d'erreur théorique optimal.

Probabilités à priori	$P(c_1)=0, 33, \hat{P}(c_1)=0, 32$	$P(c_2)=0,33,\widehat{P(c_2)}=0,34$	$P(c_3)=0,33,\widehat{P}(c_3)=0,33$	1,5%
Matrices de Covariance	$\begin{bmatrix} 0 & 0 \\ 10 & 0 \\ 0 & 10 \end{bmatrix}, \hat{S}_{1} = \begin{bmatrix} 9,79 & -1,81 & -0,10 \\ -1,81 & 9,09 & 0,59 \\ -0,10 & 0,59 & 10,87 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 10 & 0 \\ 0 & 0 \\ 0 & 10 \end{bmatrix} \xrightarrow{52} = \begin{bmatrix} 10, 43 & 0, 59 & -0, 03 \\ -0, 59 & 11, 41 & 1, 05 \\ -0, 03 & 1, 05 & 11, 61 \end{bmatrix}$	$ \begin{array}{ccc} 0 & 0 \\ 10 & 0 \\ 0 & 10 \end{array} \right  \begin{array}{c} 8,97 & 0,27 & -0,94 \\ 0,27 & 11,7 & 1,4 \\ -0,94 & 1,4 & 10,23 \end{array} $	; Taux d'erreur réel
	$S_1 = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$	$S_2 = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$	$S_3 = \begin{bmatrix} 10\\ 0 \end{bmatrix}$	e 1,7 %
Vecteirs Moyenne	$\overline{\mathbf{X}}_{1} = \begin{bmatrix} 0, 0 \\ 7, 5 \\ 6, 0 \end{bmatrix},  \mathbf{\widehat{X}}_{1} = \begin{bmatrix} 0, 30 \\ 7, 20 \\ 5, 90 \end{bmatrix}$	$\overline{\mathbf{X}}_{2} = \begin{bmatrix} 13, 0 \\ 0, 0 \\ 3, 0 \end{bmatrix},  \widehat{\mathbf{X}}_{2} = \begin{bmatrix} 13, 39 \\ 0, 06 \\ 2, 80 \end{bmatrix}$	$\overline{X}_{3} = \begin{bmatrix} 13, 0 \\ 15, 0 \\ 0, 0 \end{bmatrix},  \widehat{X}_{3} = \begin{bmatrix} 13, 76 \\ 14, 8 \\ -0, 07 \end{bmatrix}$	Taux d'erreur minimum théoriqu
Classes	<b></b>	2	3	-

Tableau A.4. : Valeurs exactes et approchées des paramètres statistiques de la distribution tridimensionnelle de l'exemple A.3.







Figure V.13.

.



BU



Figure V.13. (suite 2)



Contours obtenus par la procédure d'extraction dans 11 sections planes orthogonales à l'axe Ox<sub>3</sub> pour l'exemple A.3.

#### Exemple A.4.

La procédure a été appliquée avec succès à des données de dimensions 4 dont les paramètres statistiques sont consignés dans le tableau A.5.. Les résultats obtenus ne peuvent être visualisés, mais on peut noter que le taux d'erreur obtenu reste proche du taux d'erreur théorique optimal.

On peut faire remarquer que les contours sont plus aisément détectés et identifiés dans des espaces de dimension élevée que dans ceux de plus faible dimension. Ce résultat, a priori paradoxal, provient sans doute du fait que plus la dimension des données est élevée, plus les transitions entre les régions modales et le reste de l'espace sont brutales. En effet, pour un même nombre d'observations, l'espace interclasse est d'autant plus vide d'observations que la dimension des données est plus élevée.

Classes	Valeurs Moyenne	Matrices de Covariance	Probabilité à priori
-	$\bar{X}_{1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix};  \tilde{X}_{1} = \begin{bmatrix} 0, 10 \\ -0, 09 \\ 0, 36 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}, S_{1} = \begin{bmatrix} 5,72 & 0,42 & 0,09 & -0,3 \\ 5,72 & 0,42 & 6,13 & 0,60 & 1,27 \\ 0,09 & 0,60 & 5,61 & -1,00 \\ -0,30 & 1,27 & -1,00 & 6,24 \end{bmatrix}$	$P(c_1)=0,33,\hat{P}(c_1)=0,34$
7	$\overline{X}_{2} = \begin{bmatrix} 0 \\ 13 \\ 8 \\ 4 \end{bmatrix} ;  \overline{X}_{2} = \begin{bmatrix} 0, 21 \\ 0, 20 \\ 9, 06 \\ 9, 06 \end{bmatrix}$	$S_{2} = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 6 & 0 \end{bmatrix}; \begin{array}{c} \hat{S}_{2} = \begin{bmatrix} 6,11 & 1,52 & -0,87 & -1,31 \\ 1,52 & 5,88 & 0,91 & 0,60 \\ -0,87 & 0,91 & 6,80 & -2,21 \\ -1,31 & 0,60 & -2,21 & 7,2 \end{bmatrix}$	$P(C_2)=0, 33, \widehat{P}(C_2)=0, 35$
e	$\overline{\mathbf{X}}_{3} = \begin{bmatrix} 12 \\ 6 \\ 4 \\ 8 \end{bmatrix} ;  \widehat{\mathbf{X}}_{3} = \begin{bmatrix} 12, 61 \\ 6, 22 \\ 4, 13 \\ 7, 65 \end{bmatrix}$	$S_{3} = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 6 & 0 \\ \end{bmatrix}, \begin{array}{c} 5,86 & -0,00 & -1,12 & 0,53 \\ -1,12 & 0,19 & 5,93 & 0,21 \\ 0,53 & 0,00 & 0,21 & 6,09 \\ \end{array}$	P(c <sub>3</sub> )=0,33, P(c <sub>3</sub> )=0,31
	Taux d'erreur minimum théorique	1,16 % Taux d'erreur réel 4,2.	86
	Tableau 4.5. : Valeurs exa	ctes et annrochées des normôtros cretinis.	

,

.

Valeurs exactes et approchées des paramètres statistiques de la distribution multidimensionnelle de l'exemple A.4.

BU

# V.B. - CLASSIFICATION NON PARAMETRIQUE

# **V.B.1. RECHERCHE DES GROUPEMENTS PAR DETECTION DES CONTOURS** DES MODES

Lorsque la distribution des observations des classes en présence est très éloignée de la loi normale, le problème de la détermination des classes, se ramène à la recherche des groupements. Dans cette seconde partie du chapitre, nous allons adapter la procédure de détection des contours des modes des fonctions de densité au problème de la détection des modes d'une distribution de type quelconque. Cette procédure débouche sur des méthodes heuristiques qui permettent d'établir l'existence des classes dans des lots de données qui apparaissent sous la forme des groupements d'observations.

Nous considérons que chaque mode est caractérisé par le contour correspondant.

Pour évaluer les performances de la technique de recherche des groupements proposée, la procédure de détection et d'extraction des contours développés au chapitre IV est appliquée à des données générées artificiellement.

#### V.B.2. - RESULTATS EXPERIEMENTAUX

#### Exemple B.1.

Les données bidimensionnelles utilisées dans cet exemple sont analogues à celles utilisées dans /12/. Cet échantillon représenté figure V.14. est constitué de deux classes en forme de croissants. La figure V.15.(a) montre l'estimateur non paramétrique obtenu pour un ajustement de  $(1/h_Q)$  au milieu de sa plus grande plage de variation pour laquelle le nombre de modes détectés en fin de procédure demeure constant ( $h_Q = 1/32$ ). La version lissée de cet estimateur est illustrée par la figure V.15.(b). Les deux modes de la fonction de densité sous-





A second s





Figure V.15.(c) : Réponse de l'opérateur différentiel de Prewitt aux données de l'exemple B.1.



Figure V.16. : Classification des données de l'exemple B.1.

(a) Contours obtenus par la procédure d'extraction

(b) Résultat de la classification.

jacente sont parfaitement mis en évidence par leurs contours représentés figure V.15.(c).

En assignant chaque observation à la classe associée au contour le plus proche, on obtient les groupements de la figure V.16.. Un simple examen visuel de ce résultat met en évidence les performances de la méthode. Pour cet exemple comme pour les suivants, la taille m du voisinage  $U_m(P_r)$  utilisé pour le filtre médian et l'extraction des contours a été fixée à m = 1. Le coefficient  $\alpha$ , introduit dans la procédure d'extraction est resté fixe à la valeur  $\alpha = 0.5$ , comme pour tous les exemples précédents.

#### Exemple B.2.

L'échantillon utilisé pour cet exemple est constitué de trois classes fortement imbriquées (cf. figure V.17.). La technique de détection des contours des modes a été appliquée à cet échantillon. La fonction de densité et sa version lissée sont représentées sur la figure V.18.. Les contours des modes sont matérialisés sur la figure 18(c) et les groupements résultant de la procédure de classification (figure V.19.) montrent une nouvelle fois, l'intérêt de l'approche proposée. Le paramètre de discrétisation pour cet exemple a été ajusté à la valeur  $h_0 = 1/40$ .

Pour cet exemple, comme pour le précédent, la meilleure manière de juger la qualité des résultats est certainement l'examen visuel des groupements finaux des observations. Les trois classes représentées concordent, à quelques exceptions près, avec les conclusions d'un examen visuel de l'échantillon.




Figure V.18.(a) : Estimation brute de la fonction de densité sous-jacente de l'exemple B.2.





BU





Figure V.19. : Classification des données de l'exemple B.2.

(a) Contours obtenus par la procédure d'extraction

(b) Résultat de la classification.

# Exemple B.3.

Les deux exemples bidimensionnels précédents ont permis d'analyser le comportement de la procédure de recherche de groupements par détection des contours des modes, grâce à une visualisation aisée des résultats de simulation.

Nous présentons maintenant un exemple tridimensionnel dont les paramètres statistiques sont consignés dans le tableau B.1.. La figure V.20. représente l'échantillon ainsi constitué.

Pour visualiser les résultats de la procédure de recherche de groupements, on utilise une série de plans parallèles, équidistants et orthogonaux à l'axe  $0x_3$  de l'espace des données. Les sections des contours par ces plans permettent une visualisation des modes détectés (cf. figure V.21.(a)). La classification obtenue dans chacune de ces sections de l'espace des données par assignation de chaque observation au contour le plus proche au sens de la distance euclidienne, est indiquée par la figure V.21.(b), qui illustre le comportement satisfaisant de la procédure dans un espace multidimensionnel.

Classes	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\bar{X}_{1} = \begin{bmatrix} 0, 0 \\ 0, 0 \\ 0, 0 \\ 0, 0 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 16 \end{bmatrix}$	Pb <sub>1</sub> = 0,33
2	$\overline{X}_{2} = \begin{bmatrix} 0, 0 \\ 15, 0 \\ 3, 0 \end{bmatrix}$	$S_2 = \begin{bmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 16 \end{bmatrix}$	Pb <sub>2</sub> = 0,33
3	$\bar{\mathbf{X}}_{3} = \begin{bmatrix} 13,0\\7,5\\6,0 \end{bmatrix}$	$S_{3} = \begin{bmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 16 \end{bmatrix}$	$Pb_3 = 0,33$

Tableau B.1. : Les valeurs des paramètres statistiques de l'exemple B.3.

2







$$z_{1} = z_{x}$$

- X 🖌

 $\iota_x \blacktriangleleft$ 







 $z_{\mathbf{x}}$ 

7<sub>x</sub> ,

 $z_{\mathbf{x}}$ 





 $\psi$ ,  $\xi - = \xi x$ 



.



Figure V.21.(a) suite 1

BU



Figure V.21.(a) suite 2

► <sup>x</sup>1



Figure V.21.(a) suite 3.

Contours obtenus par la procédure d'extraction dans 14 sections planes orthogonales à l'axe Ox<sub>3</sub>.





 $x_3 = -12,0$ 

 $x_3 = -9,8$ 



 $x_3 = -8,70$ 

Figure V.21.(b)



BU

Figure V.21.(b) Suite 1



•

 $x_3 = -3,4$ 

 $x_3 = -2,3$ 

 $x_3 = -1, 2$ 

BU

 $\dot{x_3} = -0, 2$ 

Figure V.21.(b) suite 2



 $x_3 = 0,8$ 

$$x_3 = 1,9$$

 $x_3 = 3,0$ 

BU



Figure V.21.(b) suite 3



BL



Figure V.21.(b) suite 4



BU



BU

×1

#### V.2. - CONCLUSION

Dans une première partie de ce chapitre, nous avons proposé une méthode d'identification des distributions normales basée sur l'analyse des contours des fonctions de densité sous-jacente. La procédure utilise les relations simples qui existent entre les paramètres d'une distribution normale et les caractéristiques géomètriques du contour.

Sous l'hypothèse de normalité et sans aucune information a priori sur la distribution des observations, l'analyse des contours et la procédure d'identification sont mis à profit pour ramener le problème de l'optimisation du processus de classification automatique à celui de l'analyse des mélanges en déterminant, le nombre de classes, les valeurs approximatives du vecteur-moyenne, de la matrice de covariance et la probabilité a priori de chacune des classes.

Les résultats présentés dans la première partie montrent que le taux d'erreur de la classification, basé sur les valeurs approximatives des paramètres de la distribution, reste proche du taux minimal théorique.

En abandonnant l'hypothèse de normalité et en l'absence de tout modèle, l'analyse local des contours des fonctions de densité de probabilité a permis de tester les groupements en présence quelque soit leur taille et forme.

# REFERENCES V

/1/ POSTAIRE, J.G. "Fonctions convexes et optimisation du processus de classification automatique : I. Optimisation par analyse de la convexité des fonctions de densité multivariables" RAIRO/Automatique, Vol. 16, Nº 4, pp 357-379, 1982 /2/ HENRICHON, E.G. et FU, K.S. "On mode estimation in pattern recognition" Proc. 7th Symp. Adaptive Processes UCLA, 1968 /3/ FLEISS, J.L. et ZUBIN, J. "On the method and theory of clustering" Multiv. Behavioral Research, pp 235-250, 1969 /4/ KITTLER, J. "A locally sensitive method for cluster analysis" Pattern Recognition, Pergamon Press, Vol. 8, pp 23-33, 1976 15/ POSTAIRE, J.G. "Optimisation du processus de classification automatique par analyse de la convexité des fonctions de densité de probabilité. Application à la reconnaissance de la forme des arbres en dendrométrie" Thèse d'Etat, Univ. des Sciences et Techn. de Lille 1, 1981 /6/ EIGEN, D.J., FROMM, F.R. et NORTHOUSE, R.A. "Cluster analysis based on dimensional information with applications to feature selection and classification" IEEE, Trans. Syst. Man , Cybern., Vol. SMC-4, Nº 3, pp 284-294, 1974 /7/ TOMEK, I. "A modification of a clustering method" IEEE, Trans. Syst. Man , Cybern., Vol. SMC-5, Nº 3, pp 394-396, 1975 /8/ POSTAIRE J.G. et VASSEUR, C.P.A. "An approximate solution to normal mixture identification with application to unsupervised pattern classification" IEEE, Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, pp 163-179, 1981 DUDA, R.O. et HART, P.E. /9/ "Pattern Classification and scene analysis" J. WILEY, NEW YORK, pp 10-39, 1973

/10/ MARILL, T. et GREEN, D.M.
 "Statistical recognition functions and design of patterns recognizers"
 IRE Tran. Elec. Comp. Vol. EC-9, pp 472-477, 1960

- /11/ DEVIJER, P.A. et KITTLER, J.
   "Pattern recognition : A statiscal approach"
   Prentice/Hall Int. Englewood Cliffs, 1982
- /12/ KOONTZ, W.L.G., NARENDRA, P.M. et FUKUNAGA, K. "A graph theoritic approach to nonparametric cluster analysis" IEEE, Trans. Comput., Vol. C-25, N° 9, pp 936-944, 1976

# SECTION III

# APPLICATION DES TECHNIQUES D'ETIQUETAGE PROBABILISTE A LA RECHERCHE DES GROUPEMENTS

# CHAPITRE VI

# EXTRACTION DES CONTOURS PAR ETIQUETAGE PROBABILISTE ITERATIF

# **VI.1. INTRODUCTION**

Les méthodes de classification paramètriques et non paramètriques présentées dans les deux premières parties de ce mémoire sont essentiellement basées, sur une procédure de détection et d'extraction des contours des fonctions de densité sous-jacentes à la distribution des observations.

Cependant, malgré tout le soin apporté à la mise en oeuvre de la procédure d'estimation, malgré l'efficacité du filtre médian qui atténue fortement le bruit tout en préservant les contours, ces derniers ne sont pas toujours faciles à extraire.

Il s'agit en effet de reconstituer des ensembles de points connexes où les réponses des opérateurs différentiels utilisés indiquent des variations locales importantes des fonctions de densité. De nombreux essais ont montré qu'un simple seuillage de ces réponses ne permet presque jamais de reconstituer les contours fermés des modes en présence.

Si le seuil est trop élevé, on risque d'ignorer certains points des contours où la réponse de l'opérateur différentiel sera localement atténuée, par suite d'irrégularités dans la distribution des observations ou sous l'effet de la discrétisation de l'espace des données.

A l'inverse, si le seuil est trop bas, on risque d'obtenir des contours trop épais qui ne s'individualiseraient pas bien les uns aux autres.

.

La figure VI.1. constitue un exemple typique de telles situations.



Cette difficulté de reconstitution des contours est la principale raison pour laquelle nous avons introduit, au chapitre IV, une procédure d'extraction qui permet de suivre les contours, même lorsque la réponse de l'opérateur différentiel présente trop de variabilité spatiale pour être seuillée avec succès.

Dans ce chapitre, nous proposons une autre solution à ce problème d'extraction des contours. On peut en effet considérer que la probabilité pour qu'un point d'échantillonnage soit un point de contour est fonction de l'intensité de la réponse de l'opérateur différentiel en ce point.

Nous allons montrer comment cette probabilité peut être renforcée ou atténuée en fonction de l'information disponible dans le voisinage de chaque point en faisant appel à une procédure d'étiquetage probabiliste itérative /1//2//3/. On réduit ainsi considérablement les ambiguités rencontrées habituellement dans les procédures d'extraction des contours /4//5//6/.

### VI.2. - ETIQUETAGE INITIAL

On suppose que l'on dispose de l'ensemble des réponses  $\hat{G}(P_r)$ , r = 1, 2, ..., R de l'opérateur différentiel appliqué à la version filtrée de l'estimation de la fonction de densité sous-jacente.

Ces réponses peuvent être utilisées pour définir la probabilité initiale :

$$P^{(0)}(C/P_r), r = 1, 2, ..., R$$

qu'un point P<sub>r</sub> appartienne à un contour. En désignant par :

Max 
$$\left[ \widehat{G}(P_{r_{r}}) \right]$$

la valeur maximale de la réponse de l'opérateur différentiel sur tout l'espace des données, cette probabilité peut être définie par :

$$P^{(0)}(C / P_r) = \frac{\hat{G}(P_r)}{Max|\hat{G}(P_r)|}$$
,  $r = 1, 2, ..., R$ 

La probabilité initiale  $P^{(0)}(\bar{C}/P_r)$  qu'un point  $P_r$  n'appartienne pas à un contour est liée de façon évidente à  $P^{(0)}(C/P_r)$  par la relation :

$$P^{(0)}(\bar{C}/P_r) = 1 - P^{(0)}(C/P_r)$$

La probabilité  $P^{(0)}(C/P_r)$  est considérée, dans tout ce qui suit, comme la probabilité initiale pour que l'étiquette "contour", notée "C" soit exacte pour le point  $P_r$ .

Ces probabilités initiales sont réajustées de manière itérative par une procédure opérant en parallèle sur tous les points  $P_r$ , r = 1, 2, ..., R. A chaque itération de nouvelles probabilités sont calculées en chaque point  $P_r$  en fonction des probabilités assignées aux voisins de  $P_r$  à l'itération précédente. Si les probabilités dans le voisinage de  $P_r$  sont compatibles avec celles assignées au point  $P_r$ , alors cette dernière est augmentée. Par contre, en cas d'incompatibilité, elle est diminuée.

Afin de quantifier ces compatibilités entre voisins, on introduit des relations de compatibilité entre les probabilités d'exactitude des étiquettes "contour" et "non contour".

#### **VI.3. - COEFFICIENTS DE COMPATIBILITE**

Les probabilités d'exactitude de l'étiquette "contour", notée "C" et de l'étiquette "non contour", notée " $\overline{C}$ " seront remises à jour en utilisant les quatre

coefficients de compatibilités suivants :

$$\begin{array}{l} - \ C_{r,r'} \ (C, \ C) \\ - \ C_{r,r'} \ (\overline{C}, \ C) \\ - \ C_{r,r'} \ (\overline{C}, \ \overline{C}) \\ - \ C_{r,r'} \ (\overline{C}, \ \overline{C}) \end{array}$$

Le coefficient  $C_{r,r}$ ,  $(C, \overline{C})$  quantifie la compatibilité de l'étiquette "contour" (C) au point  $P_r$  avec l'étiquette "non contour" ( $\overline{C}$ ) au point  $P_r$ . Les trois autres coefficients correspondent à des étiquetages différents des points  $P_r$  et  $P_r$ .

Pour chaque paire de points  $\{P_r, P_r\}$ , on définit la compatibilité entre l'étiquette  $\gamma$  assignée au point  $P_r$  et l'étiquette  $\gamma$ 'assignée au point  $P_r$ , où  $\gamma$  et  $\gamma'$ sont indifféremment soit l'étiquette "contour" (C), soit l'étiquette "non contour" ( $\overline{C}$ ):

$$C_{\mathbf{r},\mathbf{r}'}(\mathbf{Y},\mathbf{Y}') = \frac{\left[\frac{P(\mathbf{Y}/P_{\mathbf{r}'}) - \overline{P(\mathbf{Y})}\right] \left[P(\mathbf{Y}'/P_{\mathbf{r}'}) - \overline{P(\mathbf{Y}')}\right]}{\left[\frac{P_{\max}(\mathbf{Y}) - \overline{P(\mathbf{Y})}\right] \left[P_{\max}(\mathbf{Y}) - \overline{P(\mathbf{Y}')}\right]} \quad \text{conduction}$$

où:

 $P(\gamma/P_r)$  est la probabilité d'exactitude de l'étiquette  $\gamma$  au point  $P_r$ .

$$\overline{P(\gamma)} = \frac{1}{R} \sum_{r=1}^{R} P(\gamma/P_r)$$

$$P_{max}(\gamma) = \frac{Max}{r} \left[ P(\gamma/P_r) \right], r = 1, 2, \dots, R$$

$$\gamma = C \quad \text{ou} \quad \gamma = \overline{C}$$

Ces coefficients de compatibilité sont calculés seulement pour des paires de points connexes. Ils sont supposés nuls pour des paires de points non voisins. On vérifie aisément qu'ils appartiennent à l'intervalle  $\begin{bmatrix} -1, +1 \end{bmatrix}$  et satisfont aux propriétés suivantes :

1) Si l'étiquette  $\gamma$  au point  $P_r$  est compatible avec l'étiquette  $\gamma'$  au point  $P_r$ , alors :

$$C_{r,r'}(\gamma, \gamma') > 0$$

2) Si l'étiquette  $\gamma$  au point  $P_r$  est incompatible avec l'étiquette  $\gamma'$  au point  $P_{r'},$  alors :

 $C_{r,r'}(\gamma, \gamma') < 0$ 

3) S'ils n'existent aucune relation de dépendance entre l'étiquette  $\gamma$  au point P<sub>r</sub> et l'étiquette  $\gamma$ 'au point P<sub>r</sub>, alors :

$$C_{r,r'}(\gamma, \gamma') = 0$$

# VI.4. - ETIQUETAGE PROBABILISTE ITERATIF

L'étiquetage initial et les coefficients de compatibilités proposés cidessus nous permettent maintenant d'envisager la mise en oeuvre d'une procédure d'étiquetage probabiliste itératif /7//8/.

La nouvelle estimation de la probabilité d'exactitude de l'étiquette assignée au point  $P_r$  à la (k+1) ième itération est fonction à la fois de l'estimation

précédente  $P^{(k)}(\gamma / P_r)$  de cette probabilité et de la contribution du voisinage du point  $P_r$ :

$$P^{(k+1)}(\gamma/P_{r}) = \frac{P^{(k)}(\gamma/P_{r}) \left[1 + q^{(k)}(\gamma/P_{r})\right]}{P^{(k)}(C/P_{r}) \left[1 + q^{(k)}(C/P_{r})\right] + P^{(k)}(\overline{C}/P_{r}) \left[1 + q^{(k)}(\overline{C}/P_{r})\right]}$$

où :

$$q^{(k)}(\delta/P_{r}) = \frac{1}{(3^{N}-1)} \sum_{\substack{P_{r} \in U_{1}(P_{r}) \\ r' \neq r}} \left[ \rho(\gamma, C) C_{r,r'}(\gamma, C) P^{(k)}(C/P_{r}) + \dots + \rho(\gamma, C) C_{r,r'}(\gamma, C) P^{(k)}(C/P_{r}) \right]$$

La somme intervenant dans la détermination de  $q^{(k)}(\delta/P_r)$  est calculée sur le voisinage immédiat du point  $P_r : U_1(P_r)$ , qui est le voisinage défini au paragraphe 3.3. du chapitre III de ce mémoire.  $(3^N - 1)$  est le nombre de voisins du point  $P_r$ .

Les coefficients  $P(\gamma, \gamma')$  permettent de pondérer l'influence qu'un point  $P_r$ , étiqueté  $\gamma'$  peut avoir sur un point  $P_r$  étiqueté  $\gamma$ . Ces coefficients doivent satisfaire la condition /9//10/:

$$\rho(\mathbf{C},\mathbf{C}) + \rho(\mathbf{C},\overline{\mathbf{C}}) + \rho(\overline{\mathbf{C}},\mathbf{C}) + \rho(\overline{\mathbf{C}},\overline{\mathbf{C}}) = 1$$

Le processus d'étiquetage probabiliste est itéré jusqu'à ce que les changements intervenus entre deux itérations successives, quantifiés sous la forme :

$$\mu(\mathbf{k}) = \frac{1}{R} \sum_{r=1}^{R} \left[ P^{(k+1)}(C/P_r) - P^{(k)}(C/P_r) \right]$$

atteignent une valeur suffissament petite  $\mu_o$  qui assure la stabilisation des résultats. L'ajustement de ce seuil  $\mu_b$  résulte, dans la pratique, d'un compromis entre la qualité recherchée et le temps admissible pour atteindre le résultat final /11/ /12/.

En général, quand le processus d'étiquetage itératif est stabilisé au sens du critère précédent, les probabilités d'exactitudes des étiquettes atteignent pour la plupart des valeurs extrêmes.

 $P_r^{(k)}(C/P_r) = 1$  indique que l'étiquette "contour" peut être attribuée au point  $P_r$  sans ambiguité. De manière analogue,  $P_r^{(k)}(\overline{C}/P_r) = 1$  indique sans ambiguité l'exactitude de l'étiquette "non contour".

Les résultats présentés dans la dernière partie de ce chapitre montrent que l'écrasante majorité des probabilités atteignent ces valeurs extrêmes en quelques itérations et que les résultats restent très stables, même si on continue à itérer le processus. Il peut cependant arriver qu'à la dernière itération certaines probabilités restent strictement comprises entre 0 et 1. Toutefois, compte-tenu de la réduction très importante des ambiguités, un simple seuillage des probabilités finales permet d'extraire les contours de manière sûre et satisfaisante.

# VI.5. - RESULTATS EXPERIMENTAUX

Pour analyser le comportement du processus d'étiquetage probabiliste, et montrer son intérêt en classification automatique, on considère quelques exemples d'analyse de données générées artificiellement.

# Exemple 1

Le premier exemple présenté est constitué de trois distributions gaussiennes équiprobables dont les paramètres statistiques sont consignés dans le tableau 1. Chaque classe est représentée par 200 observations. A ces 600 observations, on ajoute 400 observations distribuées uniformément sur tout l'espace afin de montrer la robustesse de la procédure par rapport au bruit (cf. figure VI.2,).

La figure VI.3. présente l'estimation brute de la fonction de densité sous-jacente. L'effet du filtre médian, appliqué à cette fonction, est clairement visible sur la figure VI.4.

La figure VI.1. montre comment la réponse de l'opérateur différentiel de Robets appliqué à cette fonction de densité estimée filtrée conduit à un estimateur du gradient à très forte variabilité spatiale. Bien que l'opérateur de Prewitt (figure VI.5.) donne, comme nous l'avons déjà indiqué au chapitre IV, des résultats beaucoup moins sensibles au bruit et aux irrégularités de la distribution analysée, on constate que les contours sont parsemés de dépressions et de pics qui rendent leur extraction délicate.

La figure VI.6. montre le résultat de la procédure d'étiquetage itératif après 6 itérations. Aucune modification de ces résultats n'a pu être décelée en itérant d'avantage le procédé. On constate que toutes les probabilités ont convergé soit vers 0, soit vers 1, ce qui ne laisse subsister aucune ambiguité quand à la détermination des contours qui apparaissent nettement individualisés. Les coefficients de pondération  $\rho(\gamma, \gamma')$  ont été tous pris égaux à 1.

Le paramètre  $h_Q$  qui définit la discrétisation de l'espace des données, a été ajusté en faisant appel au concept de stabilité du nombre de contours des modes détectés /13/ /14/. On constate (cf. figure VI.7.) que l'algorithme est peu sensible à l'ajustement de ce paramètre puisque le nombre de modes détectés reste stable et égal à trois pour  $(1/h_Q) \in [21 - 46]$ .

Distribution	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\vec{\mathbf{x}}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 2, 5 & 0 \\ 0 & 2, 5 \end{bmatrix}$	$Pb_1 = 1/3$
2	$\vec{X}_2 = \begin{bmatrix} 0 \\ 10 \end{bmatrix}$	$S_2 = \begin{bmatrix} 2, 5 & 0 \\ 0 & 2, 5 \end{bmatrix}$	$Pb_2 = 1/3$
3	$\overline{\mathbf{x}}_3 = \begin{bmatrix} 9\\5 \end{bmatrix}$	$S_3 = \begin{bmatrix} 2,5 & 0\\ 0 & 2,5 \end{bmatrix}$	$Pb_3 = 1/3$

Tableau 1 : Valeurs des paramètres statistiques de la distribution de l'exemple 1.





Figure VI.2. : Représentation graphique des 1 000 observations de l'exemple 1.



Figure VI.3. : Estimation brute de la fonction de densité sousjacente de l'exemple 1.



.

.

Figure VI.4. : Version lissée de la fonction de densité de l'exemple 1 par application du filtre médian de taille m = 1.




Figure VI.5. : Réponse de l'opérateur de Prewitt pour distribution de l'exemple 1.



.









#### Exemple 2

Effectuer une classification en présence d'un bruit uniforme ne présente guère d'intérêt. En effet, une fois les classes mises en évidence, on ne dispose d'aucune information pour différencier les observations provenant réellement des sources aléatoires constituant les classes de celles provenant du bruit.

Pour illustrer l'intérêt de l'étiquetage probabiliste à des fins de classification, on considère un second échantillon bidimensionnel (cf. figure VI.8.) dont les caractéristiques statistiques sont données dans le tableau 2.

Bien que les réponses à l'opérateur différentiel de Prewitt ( figure VI.9.) présentent des variations importantes le long des modes, les 10 itérations conduisent aux contours très nets représentés dans la figure VI.10-Ces contours peuvent être isolés par un simple seuillage (cf. figure VI.11.(a)).

Les résultats de la classification présentés figure VI.11(b) ont été obtenus en assignant chaque observation au contour le plus proche (au sens des distances euclidiennes). Malgré l'emploi de cette règle de classement extrêmement simple, les 3 classes ont été parfaitement individualisées.

Distribution	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\vec{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$S_1 = \begin{bmatrix} 13 & 0 \\ 0 & 4 \end{bmatrix}$	$Pb_1 = 1/3$
2	$\bar{\mathbf{X}}_{2} = \begin{bmatrix} 0 \\ 10 \end{bmatrix}$	$S_2 = \begin{bmatrix} 13 & 0 \\ 0 & 4 \end{bmatrix}$	$Pb_2 = 1/3$
3	$\overline{\mathbf{X}}_{3} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$	$S_3 = \begin{bmatrix} 4 & 0 \\ 0 & 13 \end{bmatrix}$	$Pb_3 = 1/3$

Tableau 2 : Valeurs des paramètres statistiques des trois distributions de l'exemple 2.



Figure VI.8. : Représentation graphique des 600 observations de l'exemple 2.















Figure VI.11. : Identification des groupements par simple seuillage des probabilités d'étiquette "Contour".

#### Exemple 3

and and for the million of a carden the set of a second and a second second and a second second as

Les deux exemples bidimensionnels précédents ont permis d'analyser le comportement de la procédure d'extraction des contours par étiquetage probabiliste grâce à une visualisation aisée des résultats.

On s'intéresse maintenant à des données à trois dimensions dont les caractéristiques statistiques sont consignées dans le tableau 3. Les résultats sont visualisés en représentant les différentes phases de l'algorithme de classification dans neuf plans parallèles, équidistants et orthogonaux à l'axe  $0x_3$  de l'espace de représentation des données.

La figure VI.13. représente les réponses de l'opérateur différentiel généralisé de Prewitt dans 9 sections de l'espace des données. Les résultats de la procédure d'étiquetage probabiliste itératif sont présentés sur la figure VI.14. En seuillant les probabilités ainsi obtenues au bout de 7 itérations au niveau 0,8, on obtient les sections des contours de la figure VI.15..

Ces contours servent finalement de "noyau" pour chacune des classes mises en évidence. Ces observations sont alors assignés à la classe dont le contour est le plus proche au sens de la distance euclidienne. Les résultats ainsi obtenus sont visualisés sur la figure VI.16.

Distribu- tion	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\overline{X}_{1} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 12 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 12 \end{bmatrix}$	$Pb_1 = 1/2$
2	$\overline{X}_{2} = \begin{bmatrix} 10\\10\\2 \end{bmatrix}$	$S_{2} = \begin{bmatrix} 12 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 12 \\ & & & \end{bmatrix}$	$Pb_2 = 1/2$

# Tableau 3 : Paramètres statistiques des deux distributions tridimensionnelles de l'exemple 3.



Figure VI.12. : Distribution des observations dans l'espace à trois dimensions.





















Probabilité d'exactitude de l'étiquette "Contour"









## VI.6. - CONCLUSION

Nous avons montré comment la procédure d'étiquetage itératif permet de lever les ambiguités rencontrées quand il s'agit d'identifier les points de l'espace de représentation des données situées sur des contours de classes.

Cette procédure facilite grandement l'extraction des contours présents dans une fonction de densité et conduit à de bons résultats de classification dans le cadre des approches non paramètriques.

Il est cependant apparu que l'utilisation de cette approche sous l'hypothèse paramètrique gaussienne conduisait à des résultats peu significatifs, les contours ayant souvent tendance à "s'empater" et à s'éloigner du modèle hyperellipsoïdal attendu.

L'intérêt de cette nouvelle approche dans l'hypothèse non paramètrique nous a poussé à explorer d'autres possibilités d'utilisation de la méthode d'étiquetage probabiliste en classification. C'est dans ce sens que nous présentons, dans le chapitre suivant, une nouvelle méthode de recherche des groupements faisant appel au même concept d'étiquetage itératif et parallèle.

#### REFERENCES VI

- /1/ ZUCKER, S.W.
   "Relaxation labeling and the reduction of local ambiguities"
   In Proc. 3rd Int. Joint Conf. Pattern Recognition, San Diego, CA,
   pp 852-861, Nov. 1976
- /2/ ROSENFELD, A., HUMMEL. R.A. et ZUCKER, S.W.
   "Scene labeling by relaxation operations"
   IEEE Trans. Syst. Man Cybern., Vol. SMC-6, pp 420-433, 1976
- /3/ ROSENFELD, A.
   "Iterative methods in image analysis"
   In Proc. IEEE Conf. Pattern Recognition and Image Processing,
   Troy, N.Y., pp 14-18, june 1977
- /4/ ZUCKER, S.W., HUMMEL, R.A. et ROSENFELD, A. "An application of relaxation labeling to line and curve enhancement" IEEE Trans. Computer, Vol. C-26, Nº 4, pp 394-403, 1977
- /5/ DANKER, A.J. et ROSENFELD, A. "Blob detection by relaxation" IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, Nº 1, pp 79-92, 1981
- /6/ ROSENFELD, A. et SMITH, R.C.
   "Thresholding using relaxation"
   IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-5, N° 5,
   pp 598-606, 1981
- /7/ POSTAIRE,J.G. et TOUZANI, A. "Mode boundary detection by relaxation with application to cluster analysis" Article soumis à Pattern Recognition, 1987
- /8/ PELEG, S. et ROSENFELD, A. "Determining compatibility coefficients for curve enhancement relaxation processes" IEEE Trans. Syst. Man, Cybern., Vol. SMC-8, Nº 7, pp 548-555, INSI 45
- /9/ SCHACHTER, B.J., LEV, A., ZUCKER, S.W. et ROSENFELD, A.
   "An application of relaxation methods to edge reinforcement" IEEE Trans. Syst. Man, Cybern., Vol. SMC-7, Nº 11, pp 813-816, 1977
- /10/ LEV, A., ZUCKER, S.W. et ROSENFELD, A.
   "Iterative enhancement of noisy images"
   IEEE Trans. Syst. Man, Cybern., Vol. SMC-7, pp 435-442, juin 1977

- /11/ ZUCKER, S.W., KRISHNAMURTHY et HAAR, R. "Relaxation processes for scene labeling : Convergence, speed, and stability" IEEE Trans. Syst. Man, Cybern., Vol. SMC-8, pp 41-48, 1978
- /12/ FEKETE, G., EKLUNDH, J.O. et ROSENFELD, A. "Relaxation : evaluation and applications" IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, N° 4, pp 459-469, 1981.
- /13/ EIGEN, D.J., FROMM, F.R. et NORTHOUSE, R.A. "Cluster analysis based on dimensional information with applications to feature selection and classification" IEEE Trans. Syst. Man, Cybern., Vol. SMC-4, pp 284-294, 1974
- /14/ POSTAIRE J.G. et VASSEUR, C.P.A. "An approximate solution to normal mixture identification with application to unsupervised pattern classification" IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, pp 163-179, 1981.

# CHAPITRE VII

# DETECTION DIRECTE DES MODES PAR ETIQUETAGE PROBABILISTE ITERATIF

#### **VII.1. INTRODUCTION**

Nous avons vu, dans le premier chapitre de ce mémoire, qu'en classification automatique statistique, la détection des modes est souvent abordée par des méthodes de recherche de maxima locaux basées sur la notion de gradient /1//2//3//4//5//6/ ou par des méthodes d'analyse de convexité locale /8//9/. Dans ce travail, nous avons introduit et exploité le concept de contour qui permet également de mettre en évidence les modes présents dans une fonction de densité /10//11//12/.

La nécessité de faire appel à ces moyens détournés pour faire apparaître les modes sous-jacents à la distribution d'un ensemble d'observations totalement inconnues résulte du fait qu'il est extrêmement difficile de déterminer les maxima locaux d'une fonction de densité multivariable par analyse directe des valeurs de son estimateur /13//14//15//16//17//18/.

Dans ce dernier chapitre, nous proposons de reprendre cette approche directe qui revient, en fait, à détecter un mode dès que la valeur estimée de la fonction de densité de probabilité sous-jacente dépasse un certain seuil, en introduisant le processus d'étiquetage probabiliste présenté au chapitre précédent /19/.

## VII.2. - DETERMINATION DES PROBABILITES INITIALES

Le principe de base des procédures d'étiquetage probabilistes peut être adapté au problème de la détection des modes en assignant à chaque point de l'espace de représentation des données une étiquette "mode" ou "vallée" dont l'exactitude n'est pas initialement garantie. Les risques d'erreurs associés à cet étiquetage sont quantifiés par la probabilité d'exactitude de chacune des étiquettes en chaque point de l'espace.

Pour être plus explicite, on suppose que l'on dispose de l'ensemble des valeurs estimées et éventuellement filtrées,  $\hat{f}(P_r)$  de la fonction de densité sous-jacente à la distribution des observations à analyser aux points d'échantillonnage  $P_r$ ,  $r = 1, 2, ..., R, R \leq Q$ , où Q est le nombre d'observations disponibles.

Ces valeurs peuvent être utilisées pour définir la probabilité initiale :

$$P^{U}(M/P_{r})$$
,  $r = 1, 2, ..., R$ 

qu'un point P<sub>r</sub> appartienne à un mode.

En désignant par :

$$\operatorname{Max}\left[\widehat{f}(P_r)\right]$$
 et  $\operatorname{Min}\left[\widehat{f}(Pr)\right]$ 

Les valeurs maximales et minimales de l'estimateur sur tout l'espace des données, cette probabilité peut être définie par :

$$P^{O}(M/P_{r}) = \frac{\widehat{f}(P_{r}) - Min[\widehat{f}(P_{r})]}{Max[\widehat{f}(P_{r})] - Min[\widehat{f}(P_{r})]} \qquad r = 1, 2, \dots, R$$

La probabilité initiale  $P^{(0)}(V/P_r)$  pour qu'un point  $P_r$  n'appartienne pas à un mode et soit donc situé dans une vallée est liée de façon simple à  $P^{(0)}(M/P_r)$ . En effet, en définissant

$$P^{(0)}(V/P_r) = \frac{Max[\hat{f}(P_r)] - \hat{f}(P_r)}{Max[\hat{f}(P_r)] - Min[\hat{f}(P_r)]}$$

on constate que :

$$P^{(0)}(V/P_r) = 1 - P^{(0)}(M/P_r)$$

Les probabilités  $P^{(0)}(M/P_r)$  et  $P^{(0)}(V/P_r)$  sont considérées, par la suite, comme les probabilités que l'étiquette "mode", notée "M", et l'étiquette "vallée", notée "V", soient exactes pour le point  $P_r$ .

Afin de réajuster ces probabilités de manière itérative et en parallèle sur l'ensemble des points  $P_r$ , r = 1, 2, ..., R, il s'agit maintenant de définir des coefficients de compatibilité entre les probabilités d'exactitude des étiquettes "M" et "V" pour des points voisins.

# VII.3. - COEFFICIENTS DE COMPATIBILITE ET ETIQUETAGE ITERATIF

Pour chaque paire de points voisins  $\{P_r, P_r, \}$ , on définit une mesure  $C_{r,r'}(\lambda, \lambda')$  de la compatibilité entre l'étiquette  $\lambda$  assignée au point  $P_r$  et l'étiquette  $\lambda'$  assignée au point  $P_r$ .

$$C_{\mathbf{r},\mathbf{r}'}(\lambda,\lambda') = \frac{\left[P(\lambda/P_{\mathbf{r}}) - \overline{P(\lambda)}\right] \left[P(\lambda'/P_{\mathbf{r}'}) - \overline{P(\lambda')}\right]}{\left[P_{\max}(\lambda) - \overline{P(\lambda)}\right] \left[P_{\max}(\lambda') - \overline{P(\lambda')}\right]}$$

où:

 $P(\lambda/P_r)$  est la probabilité d'exactitude de l'étiquette  $\lambda$  au point  $P_r$ .

$$\overline{P(\lambda)} = \frac{1}{r} \sum_{r=1}^{R} P(\lambda/P_r)$$

$$P_{max}(\lambda) = \underset{r}{Max} \left[ P(\lambda/P_r) \right] \qquad r = 1, 2, ..., R$$

 $\lambda$  et  $\lambda'$  peuvent être indifféremment l'étiquette "M" ou l'étiquette "V".

Ces coefficients de compatibilité sont supposés nuls pour des paires de points non connexes. Ils appartiennent à l'intervalle  $\begin{bmatrix} -1, +1 \end{bmatrix}$  et satisfont aux propriétés suivantes :

1) si l'étiquette  $\lambda$  au point  $P_r$  est compatible avec l'étiquette  $\lambda^i$  au point  $P_{r'}$ , alors :

$$C_{r,r'}(\lambda,\lambda') > 0$$

2) si l'étiquette  $\lambda$  au point  $P_r$  est incompatible avec l'étiquette  $\lambda'$  au point  $P_{r'}$ , alors :

$$C_{r,r'}(\lambda,\lambda') < 0$$

3) s'il n'y a pas d'influence mutuelle entre l'étiquette  $\lambda$  au point P<sub>r</sub> et l'étiquette  $\lambda'$  au point P<sub>r</sub>, alors :

$$C_{r,r'}(\lambda,\lambda') = 0$$

La nouvelle valeur  $P^{(k+1)}(\lambda/P_r)$  de la probabilité d'exactitude de l'étiquette  $\lambda$  au point  $P_r$  à l'itération (k+1) est fonction de la valeur précédente  $P^{(k)}(\lambda/P_r)$  et des probabilités assignées à l'itération (k) aux voisins immédiats de  $P_r$ :

$$P^{(k+1)}(\lambda/P_{r}) = \frac{P^{(k)}(\lambda/P_{r}) \left[1 + q^{(k)}(\lambda/P_{r})\right]}{P^{(k)}(M/P_{r}) \left[1 + q^{(k)}(M/P_{r})\right] + P^{(k)}(V/P_{r}) \left[1 + q^{(k)}(V/P_{r})\right]}$$

où:

$$q^{(k)}(\lambda/P_{r}) = \frac{1}{(3^{N}-1)} \sum_{\substack{P_{r'} \in U_{1}(P_{r}) \\ r' \neq r}} \left[ \rho(\lambda, M) C_{r,r'}(\lambda/M) P^{(k)}(M/P_{r'}) + \dots \right]$$

... + 
$$\rho(\lambda, V) C_{r,r'}(\lambda/V) P^{(k)}(V/P_r)$$

 $U_1(P_r)$  représente les  $(3^{N-1})$  voisins immédiats du point  $P_r$ .

Les coefficients  $\rho(\lambda, \lambda')$ , qui satisfont la relation :

$$\rho(M,M) + \rho(V,M) + \rho(M,V) + \rho(V,V) = 1$$

permettent de pondérer les influences des voisins de  $P_r$  dans le processus de réajustement en fonction des étiquettes qui leur sont assignées.

Ce processus d'étiquetage est itéré jusqu'à ce que

$$\frac{1}{R}\sum_{r=1}^{R} \left[P^{(k+1)}(M/P_r) - P^{(k)}(M/P_r)\right] \leqslant \epsilon_0$$

où le nombre E, est ajusté en fonction du degré de convergence souhaité.

L'expérience montre que les probabilités atteignent très vite les valeurs 0 et 1, ce qui fait que l'ajustement du paramètre  $\mathcal{E}_{o}$  influe très peu la procédure itérative.

Les quelques exemples qui suivent sont destinés à mettre en évidence les performances de cette procédure d'étiquetage itératif qui permet, lorsque les ambiguités ont disparu, de détecter très aisément les modes de la fonction de densité sous-jacente et de procéder à la classification des observations disponibles.

## VII.4. - RESULTATS EXPERIMENTAUX

Le comportement de la procédure de détection des modes et vallées par étiquetage itératif est illustré sur quelques ensembles de données générées artificiellement.

#### Exemple 1

Le premier exemple est constitué de trois classes gaussiennes équiprobables dont les paramètres statistiques sont présentés dans le tableau 1. Chacune des classes est représentée dans l'échantillon de la figure VII.1. par 200 observations.

La figure VII.2. représente l'estimation brute de la fonction de densité sous -jacente à cet échantillon bidimensionnel. Compte-tenu de la grande variabilité sptatiale de l'estimateur, il semble difficile de définir automatiquement un seuil permettant d'isoler les modes de façon fiable.

La figure VII.3. montre les probabilités d'exactitude de l'étiquette "mode" après 25 itérations du procesus d'étiquetage itératif. Sur ce diagramme on constate que les modes apparaissent sous la forme de trois régions connexes où la probabilité prend la valeur 1, éliminant ainsi toute ambiguité. Le reste de l'espace est affecté à la probabilité 0, qui indique sans équivoque qu'il s'agit de la région étiquetée "vallée".

Pour cet exemple, les coefficients de pondération  $\rho(\lambda, \lambda')$  ont été ajustés de la manière suivante :

$$\rho(M, M) = 0,80$$
  $\rho(M,V) = 0,15$ 

 $\rho(V,M) = 0,025$ 

La valeur de  $\rho(M,M)$  élevée permet de favoriser l'apparition des modes. A l'opposé, une valeur élevée de  $\rho(V,V)$  aurait favorisé l'apparition des vallées, ce qui pourrait être utile quand des classes présentent un degré de chevauchement très important.

 $\rho(V,V) = 0,025$ 

Un simple seuillage des valeurs atteintes par les probabilités permet d'isoler les modes qui sont reconstitués par analyse de connexité des points d'échantillonnage sélectionnés.

La classification de l'échantillon se ramène alors à un problème de classement. On peut considérer les observations contenues dans les hypercubes constituant chaque mode comme des prototypes de chaque classe et utiliser une méthode de classement classique /20/.

Les résultats présentés sur la figure VII.4. ont été obtenus en assignant les observations non contenues dans les modes à la classe représentée par le prototype le plus proche, au sens de la distance euclidienne.

Distribution	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\vec{\mathbf{X}}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$S_{1} = \begin{bmatrix} 27 & 0 \\ 0 & 8 \end{bmatrix}$	$Pb_1 = 1/3$
2	$\overline{X}_2 = \begin{bmatrix} 0\\ 14 \end{bmatrix}$	$S_2 = \begin{bmatrix} 27 & 0 \\ 0 & 8 \end{bmatrix}$	$Pb_2 = 1/3$
3	$\overline{\mathbf{X}}_{3} = \begin{vmatrix} 13.0 \\ 7,5 \end{vmatrix}$	$S_3 = \begin{bmatrix} 8 & 0 \\ 0 & 27 \end{bmatrix}$	$Pb_3 = 1/3$

Tableau 1 : Les paramètres statistiques de la distribution de l'exemple 1.





Figure VII.1. : Distribution des observations dans l'espace bidimensionnel correspondant à l'exemple 1.





Figure VII.3. : Représentation des probabilités d'exactitude de l'étiquette "mode" obtenues à la dernière itération du processus d'étiquetage probabiliste.





Figure VII.4. : Résultat de classification des données de l'exemple 1.



Pour cet exemple, l'ajustement du paramètre  $h_Q$  qui définit la finesse de la discrétisation de l'espace des données, a été obtenu en faisant appel au concept de stabilité du nombre de modes détectés /21/ /22/ /23/.

La figure VII.5. indique les variations du nombre de modes mis en évidence à la fin de la procédure d'étiquetage itérative en fonction du paramètre  $h_Q$ . On constate que l'algorithme est peu sensible à l'ajustement de ce paramètre puisque le nombre de modes détecté reste stable et égal à trois pour  $(1/h_Q) \in [28, 58]$ . Ce résultat peut être favorablement comparé à ceux obtenus pour d'autres algorithmes pour lesquels la largeur de la plage de stabilité est beaucoup plus réduite /23/.





#### Exemple 2

Le second exemple présenté a pour principale finalité de mettre en évidence les possibilités offertes par l'algorithme de traiter des échantillons au sein desquels les classes présentent des poids inégaux.

Les trois groupements gaussiens constituant ce second échantillon de 600 observations sont générés artificiellement à partir des données statistiques du tableau 2 (cf. figure VII.6.). Sur la figure VII.7., qui représente l'estimation de la fonction de densité sous-jacente, on constate qu'aucune procédure de seuillage ne peut être mise en oeuvre pour isoler les 3 modes présents. En effet, un seuil trop haut ne mettrait en évidence que 1 ou 2 modes. En baissant le seuil pour faire apparaître le mode le plus petit, les deux autres fusionneraient et ne seraient plus discernables. Dans de telles situations, on peut éventuellement faire appel à une procédure à seuils multiples. En effet, en analysant les fusionnements des modes qui apparaissent en faisant décroître le seuil, les méthodes d'analyse hiérarchiques par chaînage simple /24/ permettent de retrouver la structure des données.

Les résultats de la figure VII.8., obtenus au bout de 32 itérations, montrent que la méthode d'étiquetage probabiliste apporte une solution simple à ce problème.

La classification finale, présentée figure VII.9., a été obtenue en assignant chaque observation à son prototype le plus proche, en pondérant les distances par le nombre de prototypes situés dans chaque mode mis en évidence.

Distribution	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\overline{X}_{1} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$S_1 = \begin{bmatrix} 27 & 0 \\ 0 & 8 \end{bmatrix}$	$Pb_1 = 2/10$
2	$\overline{X}_2 = \begin{bmatrix} 0\\ 14 \end{bmatrix}$	$S_2 = \begin{bmatrix} 27 & 0 \\ 0 & 8 \end{bmatrix}$	$Pb_2 = 3/10$
3	$\overline{\mathbf{X}}_{3} = \begin{bmatrix} 13.0\\ 7.5 \end{bmatrix}$	$S_3 = \begin{bmatrix} 8 & 0 \\ 0 & 27 \end{bmatrix}$	$Pb_3 = 5/10$

Tableau 2 : Les paramètres statistiques de la distribution de l'exemple 2.

.





Figure VII.6. : Représentation graphique de la distribution de l'exemple 2.





•

Figure VII.7. : La fonction de densité sous-jacente à la distribution de l'exemple 2.


Figure VII.8. : Représentation graphique des probabilités d'exactitude de l'étiquette "mode".



Figure VII.9. : Résultat de classification des données de l'échantillon de l'exemple 2.

LILL

#### Exemple 3

Les deux premiers exemples de type bidimensionnel ont permis de présenter aisément les étapes intermédiaires du processus de classification.

Nous présentons maintenant un exemple tri-dimensionnel dont les paramètres statistiques sont consignés dans le tableau 3. Seule la distribution des observations, à raison de 300 points par classe, est représentée dans l'espace à 3 dimensions (cf. figure VII.10).

La fonction de densité estimée est présentée sur la partie gauche de la figure VII.11. 10 sections planes allant de  $x_3 = -1.5$  à  $x_3 = 3.0$ . On constate les grandes disparités des valeurs de l'estimateur de section à section.

On peut faire remarquer que les fonctions de densité associées à chacune des classes en présence présentent un fort degré de chevauchement au niveau des sections  $x_3 = 0.5$  et  $x_3 = 1.0$ .

Sur la partie droite de la figure VII.11. apparaissent les résultats de la procédure d'étiquetage itératif. Les modes y sont caractérisés par des régions connexes où la probabilité d'exactitude de l'étiquette "mode" atteint la valeur 1. La séparation des deux modes au niveau des sections  $x_3 = 0.5$  et  $x_3 = 1.0$  est nette.

Distribu- tion	Vecteurs Moyenne	Matrices de Covariance	Probabilités a priori
1	$\overline{X}_{1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\mathbf{S}_{1} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 18 & 0 \\ 0 & 0 & 27 \end{bmatrix}$	$Pb_1 = 1/3$
2	$\overline{X}_{2} = \begin{bmatrix} 0\\ 14\\ 0 \end{bmatrix}$	$S_{2} = \begin{bmatrix} 18 & 0 & 0 \\ 0 & 27 & 0 \\ 0 & 0 & 9 \end{bmatrix}$	$Pb_2 = 1/3$
3	$\overline{X}_{3} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}$	$S_3 = \begin{bmatrix} 27 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 18 \end{bmatrix}$	$Pb_3 = 1/3$

Tableau 3 : Paramètres statistiques de la distribution de l'exemple 3.



Figure VII.10. Distribution des observations de l'exemple 3 dans l'espace à trois dimensions.





Figure VII.11.



Figure VII.11. (suite)







Probabilité d'exactitude de l'étiquette "mode"







Figure VII.11. (suite)

### VII - CONCLUSION

Dans ce dernier chapitre, nous avons proposé une approche directe à la détection des modes en classification automatique, qui ne fait appel ni à des concepts différentiels, ni à des concepts de convexité. Grâce à la procédure d'étiquetage probabiliste, la discrimination "mode" / "vallée" revient à une simple procédure de seuillage qui permet d'isoler les zones où la fonction de densité de probabilité sous-jacente présente des valeurs localement élevées, même lorsque les classes ne sont pas équiprobables et lorsqu'elles se chevauchent notablement.

#### REFERENCES VII

- /1/ FUKUNAGA, K. et HOSTETLER, L.O. "The estimation of the gradient of a density function, with applications in pattern recognition" IEEE, Trans. Inform. Theory, Vol. IT-21, Nº 1, pp 32-40, 1975
- /2/ KOONTZ, W.L., NARENDRA, P. et FUKUNAGA, K.
  "A graph theoretic approach to nonparametric cluster analysis" IEEE Trans. Comput., Vol. C-25, No 9, pp 936-944, 1976
- /3/ KOONTZ, W.L. et FUKUNAGA, K. "Asymptotic analysis of a nonparametric clustering technique" IEEE Trans. Comput., Vol. C-21, Nº 9, pp 967-974, 1972
- /4/ KATZ, J.O. et ROHLF, F.J. "Function-point cluster analysis" Systematic Zoology, Vol. 22, pp 295-301, 1973
- /5/ KOONTZ, W.L. et FUKUNAGA, K. "A nonparametric Valley-seeking technique for cluster analysis" IEEE Trans. Comput., Vol. C-21, N° 2, pp 171-178, 1972
- /6/ KITTLER, J.
  "A locally sensitive method for cluster analysis"
  Pattern Recognition, Pergamon Press, Vol. 8, pp 23-33, 1976
- /7/ DEVIJVER, P.A. et KITTLER, J. "Pattern Recognition : A statistical approach" Englewood Cliffs, N.J. Prentice-Hall, 1982
- /8/ VASSEUR, C.P.A. et POSTAIRE,J.G. "A convexity testing method for cluster analysis" IEEE Trans. Syst. Man, Cybern., Vol. SMC-10, N° 3, pp 145-149, 1980
- /9/ POSTAIRE, J.G. et VASSEUR, C.P.A. "Analyse locale de la forme des fonctions multivariables - Application en classification automatique" Congrés AFCET-IRIA, Reconnaissance des Formes et Intelligence Artificielle, Toulouse, pp 348-355, 1979
- /10/ TOUZANI,A. et POSTAIRE, J.G.
  "Clustering by mode boundary detection"
  Article soumis à IEEE Trans. Pattern Anal. Machine Intell., 1987

/11/ POSTAIRE, J.G. et TOUZANI, A. "Détection des modes par recherche des contours en classification automatique" Article soumis à A.P.I.I. - RAIRO, AFCET, PARIS, 1987 /12/ POSTAIRE, J.G. et TOUZANI, A. "Mode boundary detection by relaxation with application cluster analysis" Article soumis à Pattern Recognition-Pergamon Press, 1987 /13 / PARZEN, E. "On estimation of a probability density function and mode" Ann. Math. Stat., Vol. 33, pp 1065-1076, 1962 /14/ RYZIN. J.V. "On strong consistency of density estimates" Annu. Math. Statist., Vol. 40, pp 1765-1772, 1969 /15/ HARTIGAN, J.A. "Clusters as modes" Proc. 1st Int. Symp. on data analysis and informations, Versailles, 1977 /16/ HARTIGAN, J.A. "Distribution problems in clustering" Classification & Clustering, J. Van Rysin, Ed., pp 45-71, Academic Press, New-York, 1977 /17/ KOPP, B. "Ein neuer verfarhren zur cluster analyse" Biometrishe Meitshrift, Vol. 18, pp 473-476, 1976 /18/ BOCK, A.H. "Automatishe Klassifikation" Vandenhoeck et Rubrecht, Göttingen, 1974 /19/ TOUZANI, A. et POSTAIRE, J.G. "Mode detection by relaxation" Article soumis à IEEE Trans. Pattern Anal. Machine Intell., 1987 /20 / DUDA, R.O. et HART, P.E. "Pattern Classification and scene analysis" J. Wiley, New-York, 1973 /21/ EIGEN, D.J., FROMM, F.R. et NORTHOUSE, R.A. "Cluster analysis based on dimensional information with applications to feature selection and classification" IEEE, Trans. Syst. Man Cybern. Vol. SMC-4, Nº 3, pp 284-294, 1974

/22/ TOMEK, I.
 "A modification of a clustering method"
 IEEE Trans. Syst. Man Cybern., Vol. SMC-5, N° 3, pp 394-396, 1975

/23/ POSTAIRE, J.G.
 "De l'image à la décision. Analyse des images numériques et théorie
 de la décision"
 DUNOD, Paris, 1987.

# **CONCLUSION GENERALE**

Dans ce mémoire, nous avons abordé les problèmes de classification automatique dans le cas le plus délicat, c'est-à-dire lorsque l'on ne dispose d'aucune information a priori sur les données à classer. Dans ces conditions, l'information nécessaire pour effectuer la classification doit être extraite des observations à classer elles-mêmes.

Dans cette optique, nous avons introduit, en nous basant sur un modèle analytique, le concept de contour d'une fonction de densité de probabilité. Cette approche fournit un cadre général pour aborder l'aspect statistique de la classification automatique (Chapitre II). La même approche mathématique permet en effet de résoudre de manière originale et efficace des problèmes d'analyse des mélanges aussi bien que des problèmes de recherche de groupements.

Ce concept s'est révélé très riche de retombées, car la mise en évidence des contours de la fonction de densité sous-jacente à la distribution d'un ensemble d'observations (Chapitre III et Chapitre IV) permet d'aborder aussi bien des problèmes de classification paramètrique que des problèmes non paramètriques (Chapitre V).

C'est sans doute dans le cadre de l'analyse des mélanges que l'utilisation du concept de contour apparaît le plus féconde sur le plan formel. Sans aucune hypothèse restrictive, la détermination des contours de la fonction de densité d'un mélange à partir des observations disponibles permet d'abord de détecter les classes en présence. L'analyse des contours permet ensuite de déterminer, pour chaque composante mise en évidence, des valeurs approchées du vecteur moyenne, de la matrice de covariance et de la probabilité a priori. Ces informations sont utilisées pour calculer des fonctions de décision qui permettent d'optimiser le processus de classification. Cependant, si l'approche paramètrique débouche sur la notion de classification optimale, en ce sens qu'elle permet de minimiser le risque d'erreur, il faut mettre en garde l'analyste contre son emploi abusif. En effet, lorsque les données ne suivent par le modèle analytique sur lequel est basée cette approche, la procédure aura tendance à imposer une structure aux données au lieu d'aider à découvrir cette structure.

Pour de telles situations, l'approche non paramètrique permet, en l'absence de modèle, d'établir l'existence de groupements d'observations dans des lots de données. Le caractère local de l'analyse du contour des fonctions de densité permet alors de mettre en évidence des groupements de forme et de taille très variées.

Finalement, nous avons montré comment les procédures d'étiquetage itératif peuvent être adaptés avec profit au problème de la recherche des groupements d'observations sous hypothèse non paramètrique. Qu'il s'agisse d'améliorer l'extraction des contours (Chapitre VI) ou de faire ressortir la présence de maxima locaux significatifs des fonctions de densités estimées (Chapitre VII), ces méthodes itératives réduisent considérablement les ambiguités rencontrées dans ces problèmes de classification non supervisés.

L'utilisation de techniques non paramètriques pour la détection des contours des fonctions de densité (Chapitres III et Chapitre IV) a posé des problèmes difficiles, d'ailleurs bien connus des praticiens.

L'utilisation de la méthode d'estimation du noyau et son implantation sous la forme d'un algorithme rapide présente le grand avantage de permettre une discrétisation simple de l'espace de représentation des données. Le réseau hypercubique de discrétisation ainsi introduit facilite l'implantation des algorithmes d'extraction des contours et de modélisation. Cependant, les résultats de cette technique d'estimation se dégradent dès que l'on se trouve en présence de groupements d'observations de données très différentes. Pour de telles situations, il est préférable de faire appel à la technique d'estimation dite des "k plus proches voisins". Cette technique présente l'énorme avantage d'ajuster la fenêtre utilisée pour déterminer l'estimateur en fonction de la densité locale des observations disponibles. Cette fenêtre, variable, étendue dans les zones de faible densité et réduite dans les zones de densité élevée, assure une plus grande représentativité à l'estimateur dans le cas d'échantillons finis /1//2//3/.

La mise en oeuvre de cette procédure et surtout son exploitation ultérieure semblent cependant plus délicates. Des travaux sont actuellement en cours dans ce sens et les premiers résultats sont encourageants /4/. Il semble, dès à présent, que le recours à cet estimateur permettra d'obtenir des résultats fiables avec des échantillons de taille moindre que ceux nécessaires pour les méthodes présentées dans ce mémoire.

Nous avons également mis en évidence les difficultés rencontrées lors de l'extraction des contours des fonctions multivariables. Le processus d'étiquetage itératif apporte une première solution à ce problème, mais ses effets ont tendance à faire grossir les contours, de telle sorte qu'ils deviennent inexploitables dans l'optique paramètrique.

Une solution pour rendre la méthode d'étiquetage itératif, plus fine, consisterait à définir les probabilités initiales sur des critères locaux non globaux comme nous l'avons fait jusqu'à présent. En effet, le fait de comparer les valeurs de l'estimateur ou de l'opérateur différentiel en un point d'échantillonnage au maxima et minima déterminés sur tout l'espace de représentation des données peut conduire à faire disparaître certaines informations pertinentes. Cette détermination locale des probabilités initiales est actuellement à l'étude. Elle devrait permettre de diminuer le nombre d'itérations nécessaires et surtout de mieux pondérer certaines influences mutuelles entre voisins, affinant ainsi la détermination des contours. Cette technique d'étiquetage itératif avec ajustement local des probabilités initiales pourra d'ailleurs être étendue à la détection directe des modes par analyse des valeurs prises par l'estimateur des fonctions de densité. Cette amélioration permettra sans doute de traiter avec succès des cas où les pondérations entre les différentes classes en présence présenteront de grandes disparités.

Une dernière application de l'étiquetage probabiliste concerne l'amélioration de la détermination de la convexité des fonctions de densité sousjacente. Ce travail, qui a été conduit parallèlement à celui présenté dans ce mémoire /5/, est sur le point de déboucher. Comme pour l'extraction des contours et la détection directe des modes, il apportera plus de robustesse aux techniques de classification basées sur le concept de convexité /6/.

Ce travail, ainsi que ses prolongements en cours de développement, montre que la recherche des contours des fonctions de densité de probabilité associée aux techniques d'étiquetage probabiliste itératif constitue une voie d'approche fructueuse aux problèmes de classification automatique non supervisée.

## REFERENCES - CONCLUSION

- /1/ DUDA, R.O. et HART, P.E. "Pattern Classification and scene analysis" Wiley, New-York, 1973
- /2/ COVER, T.M. et HART, P.E. "Nearest neighbor pattern classification" IEEE, Trans. Info. Theory, Vol. IT-13, pp 21-27, 1967
- /3/ LOFTSGAARDEN, D.O. et QUESENBERNY, C.P. "A nonparametric estimate of a multivariable density function" Ann. Math. Stat., Vol. 36, pp 1049-1051, 1965
- /4/ CZESNALOWICZ, E. "Application de l'estimateur non parametrique des K.P.P.V. à la détection des contours en classification automatique" Mémoire de D.E.A. - Juin 1987
- /5/ OLEJNIK, S. "Analyse de la convexité d'une fonction de densité de probabilité par étiquetage probabiliste : Application à la classification automatique non supervisée" Thèse de 3ème Cycle - à soutenir prochainement
- /6/ POSTAIRE, J.G. "Une approche unique pour l'analyse des mélanges et la détection des modes en classification automatique" Rev. Stat. Appl., Vol. XXXJ, Nº 4, pp 17-36, 1983

