50376
1987
11

N°d'ordre : 404

50376
1987
11

THÈSE

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

pour obtenir le titre de

DOCTEUR INGENIEUR

Spécialité : Automatique

par

LAKEL Rabah

Ingénieur E.N.P.A.

BIBLIOTHEQUE UNIVERSITAIRE
SECTION DE SCIENCES
LILLE

CONTRIBUTION A L'IDENTIFICATION DES MELANGES GAUSSIENS A L'AIDE D'UNE APPROCHE RECURSIVE.

Soutenue le 7 janvier 1987 devant la Commission d'Examen :

MM	P. VIDAL	Président
	C. VASSEUR	Rapporteur
	B. DUBUISSON	Examinateur
	D. WILLAEYS	Examinateur
	J.C.DIVERREZ	Examinateur

A la mémoire de mon père

A mes parents

A Saleha et Walid...

AVANT-PROPOS

Avant de présenter cette étude , je tiens à exprimer ma profonde gratitude à Monsieur le professeur P.VIDAL, Directeur du Centre d'Automatique de LILLE ,qui m'a accueilli dans son laboratoire. Que Monsieur VIDAL veuille bien accepter mes vifs remerciements pour l'honneur qu'il me fait en assurant la présidence du jury.

J'adresse ma plus vive reconnaissance à Monsieur le professeur C.VASSEUR pour m'avoir constamment aidé ,guidé , conseillé tout au long de l'élaboration de cette étude et de l'honneur qu'il me fait en acceptant de la rapporter. Que Monsieur VASSEUR veuille bien trouver ici l'expression de mon amitié .

Je remercie également Monsieur le professeur B.DUBUISSON , directeur du centre d'Automatique et Traitement du Signal de l'U.T de COMPIEGNE , pour l'intérêt qu'il a témoigné à cette étude et pour sa présence parmi les membres du jury.

Je suis très sensible à l'intérêt que Monsieur D.WILLAEYS professeur à l'universite de VALENCIENNES, porte à ce travail en acceptant de participer dans ce jury.

Mes remerciements vont aussi à Monsieur J.C. DIVERREZ P.D.G de SEROSI S.A VILLENEUVE D'ASCQ pour l'honneur qu'il me fait en acceptant de participer à ce jury

Je remercie également l'Université de ANNABA pour m'avoir aide financièrement et témoigne de la confiance , en m'accordant une bourse d'études .

PLAN GENERAL

CHAPITRE I

ANALYSE ET MODELISATION DES SYSTEMES COMPLEXES

I-1 INTRODUCTION

I-2 SELECTION DES PARAMETRES

I-3 METHODES DE CLASSIFICATION AUTOMATIQUE

- I-3-1 INTRODUCTION
- I-3-2 APPROCHE METRIQUE
- I-3-3 APPROCHE INFORMATIONNELLE
- I-3-4 APPROCHE STATISTIQUE
- I-3-5 APPROCHE GRAPHIQUE

I-4 MODELISATION ET IDENTIFICATION DES MODES DE FONCTIONNEMENT

- I-4-1 INTRODUCTION
- I-4-2 DECISION D'ATTRIBUTION NON PARAMETRIQUE
- I-4-3 DECISION D'ATTRIBUTION PARAMETRIQUE

I-5 APPROCHE RECURSIVE DANS L'ESTIMATION DES PARAMETRES STATISTIQUES

I-6 CONCLUSION

BIBLIOGRAPHIES

CHAPITREII

ESTIMATION RECURSIVE DES PARAMETRES STATISTIQUES D'UN MELANGE GAUSSIEN

II-1 INTRODUCTION

II-2 MISE EN EQUATION

II-3 DETERMINATION DE LA RECURRENCE SUR LA MATRICE DE COVARIANCE

II-4 DETERMINATION DE LA RECURRENCE SUR LES VALEURS PROPRES

II-5 DETERMINATION DE LA RECURRENCE SUR LES DIRECTIONS PROPRES

II-6 DETERMINATION DE LA RECURRENCE SUR LE VECTEUR MOYENNE ET LA PROBABILITE A PRIORI

II-7 ALGORITHME RAPIDE DE CALCUL DE λ^{-1}

II-8 RECAPITULATIF

II-9 SIMULATIONS ET PERFORMANCES
- II-9-1 METHODE
- II-9-2 SIMULATION 2D
- II-9-3 SIMULATION 3D
- II-9-4 ANALYSE DES RESULTATS OBTENUS

II-10 CONCLUSION

CHAPITRE III

FUSIONNEMENT DE DEUX COMPOSANTES

III-1 INTRODUCTION

III-2 DISTANCE MINIMALE ENTRE DEUX COMPOSANTES

III-3 DISPERSION DES OBSERVATIONS

III-4 ENTROPIE

III-5 DETECTION DES MODES

III-6 DETERMINATION DES POINTS M_{IK} ET M_{KI}

III-7 ALGORITHME RAPIDE DE RECHERCHE DES POINTS M_{IK} ET M_{KI}

III-8 EXEMPLES D'APPLICATION
- III-8-1 INTRODUCTION
- III-8-2 FONCTION BIMODALE
- III-8-3 UTILISATION DES PROBABILITES A POSTERIORI
- III-8-4 LES PARAMETRES DE LA COMPOSANTE SOMME

III-9 CONCLUSION

BIBLIOGRAPHIES

CHAPIRE IV

AUGMENTATION DU NOMBRE DE COMPOSANTES D'UN MELANGE GAUSSIEN

IV-1 INTRODUCTION

IV-2 MESURE DE LA DISPERSION DES COMPOSANTES

IV-3 DEFINITION D'UN CONTOUR DE CONFIANCE

IV-3-1 INTRODUCTION

IV-3-2 CAS OU LE NOMBRE D'OBSERVATIONS EST GRAND

IV-3-3 CAS OU LE NOMBRE D'OBSERVATIONS EST FAIBLE

IV-4 DEFINITION D'UN CONTOUR DE CONFIANCE POUR LES NOUVEAUX CENTRES DE MASSE

IV-5 PROCEDURE A SUIVRE DANS LE CAS OU LE NOMBRE DE COMPOSANTES D'UN MELANGE GAUSSIEN AUGMENTE

IV-5-1 INTRODUCTION

IV-5-2 APPARITION D'UNE NOUVELLE COMPOSANTE

IV-5-3 QUELQUES SITUATIONS PARTICULIERES

IV-5-4 ECLATEMENT D'NE COMPOSANTE EN DEUX

IV-6 EXEMPLES D'APPLICATION

IV-6-1 APPARITION D'UNE NOUVELLE COMPOSANTE

IV-6-2 ECLATEMENT D'UNE COMPOSANTE EN DEUX

IV-7 CONCLUSION

BIBLIOGRAPHIES

V-CONCLUSION

ANNEXE-1

ANNEXE-2

INTRODUCTION GENERALE

INTRODUCTION GENERALE

Le travail présenté dans ce mémoire s'inscrit dans le cadre des travaux de recherche menés actuellement au centre d'automatique de lille et visant à contribuer à l'identification des mélanges gaussiens modélisant les systèmes complexes.

Cette opération d'identification s'effectue selon une approche récursive qui permet d'intégrer d'une manière dynamique les nouveaux échantillons qui viennent enrichir la base d'apprentissage .

L'adoption d'une telle approche pour estimer les paramères statistiques des différentes composantes d'un mélange gaussien est rendue nécessaire d'une part par l'évolution tout à fait naturelle des systèmes étudiés et d'autre part par l'utilisation de bases d'apprentissage parfois incomplètes et donc ne contenant pas tous les modes de fonctionnement du système.

Ce mémoire est constitué de quatre chapitres .

Dans le premier chapitre , sont résumées les différentes phases qu'il faut observer pour analyser , identifier et optimiser un système complexe en utilisant uniquement l'analyse des données comme moyen d'investigation.

Dans le deuxième chapitre , sont établies toutes les équations récurrentes qui permettent d'estimer de manière dynamique les paramètres statistiques d'un mélange gaussien.

Dans le troisième chapitre ,quelques points privilégiés de l'espace d'observation , où s'exercent les influences mutuelles des différentes composantes d'un même mélange gaussien , sont présentés.

L'estimation des critères classiques de mesure de séparabilité des composantes en ces points permet de mettre en évidence éventuellement le début de fusionnement de certaines composantes.

Dans le quatrième chapitre , une définition du contour de confiance associé à chaque composante est présentée.

L'utilisation des contours de confiance comme seuils limites d'appartenance permet de déceler l'apparition de nouvelles distributions normales représentant les modes de fonctionnement du système étudié qui ont été ignorés lors de la construction de la base d'apprentissage.

Les chapitres deux , trois et quatre sont illustrés par des applications réalisées sur des données générées artificiellement.

CHAPITRE I

ANALYSE ET MODELISATION DES SYSTEMES COMPLEXES

CHAPITRE I

ANALYSE ET MODELISATION DES SYSTEMES COMPLEXES

I-1 INTRODUCTION 1

I-2 SELECTION DES PARAMETRES 5

I-3 METHODES DE CLASSIFICATION AUTOMATIQUE 7

I-3-1 INTRODUCTION 7
I-3-2 APPROCHE METRIQUE 9
I-3-3 APPROCHE INFORMATIONNELLE 11
I-3-4 APPROCHE STATISTIQUE 12
I-3-5 APPROCHE GRAPHIQUE 13

I-4 MODELISATION ET IDENTIFICATION DES MODES DE FONCTIONNEMENT 14

I-4-1 INTRODUCTION 14
I-4-2 DECISION D'ATTRIBUTION NON PARAMETRIQUE 15
I-4-3 DECISION D'ATTRIBUTION PARAMETRIQUE 17

I-5 APPROCHE RECURSIVE DANS L'ESTIMATION DES PARAMETRES STATISTIQUES .. 19

I-6 CONCLUSION 24

BIBLIOGRAPHIES 25

I-1 INTRODUCTION

Contrairement à certains systèmes où l'équation d'état est utilisée sous sa forme analytique pour mener à terme les opérations d'analyse,d'identification et d'optimisation,l'étude d'un système complexe décrit par une grande variété de variables ne permet pas d'appliquer les méthodes classiques de l'automatique.[TIT,79],[RIC,75],[STA,81]

A la suite des difficultés rencontrées dans l'utilisation des méthodes classiques,de nouvelles approches basées essentiellement sur l'analyse des données ont été développées. [RIS,78],[BRI,77],[BEN,77].

Le schéma bloc représenté par la figure1 résume parfaitement les différentes phases qu'il faut observer pour analyser un système complexe en utilisant uniquement les données qui le caractérisent.

C'est au cours de la troisième phase que la base d'apprentissage est construite.Elle est constituée de données recueillies sur un horizon fini,en observant le système étudié par l'intermédiaire de ses variables caractéristiques.

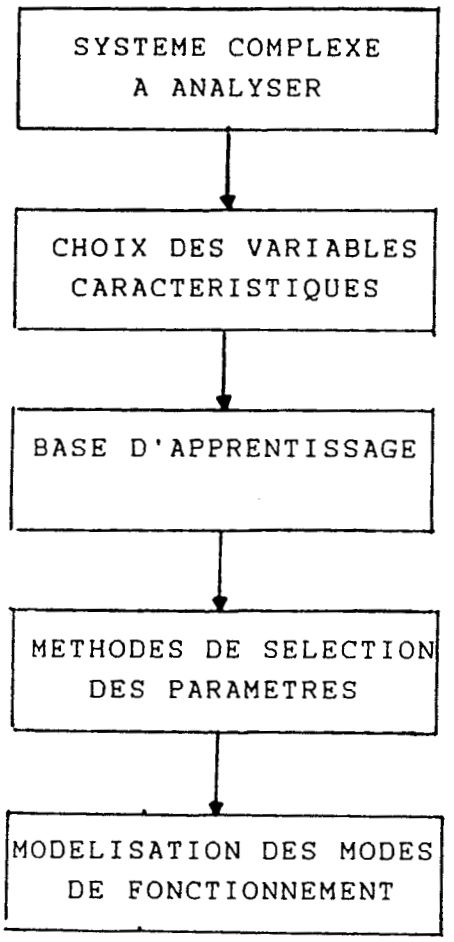


fig.1 schéma bloc représentant les phases à suivre pour observer un système complexe

BU LILLE

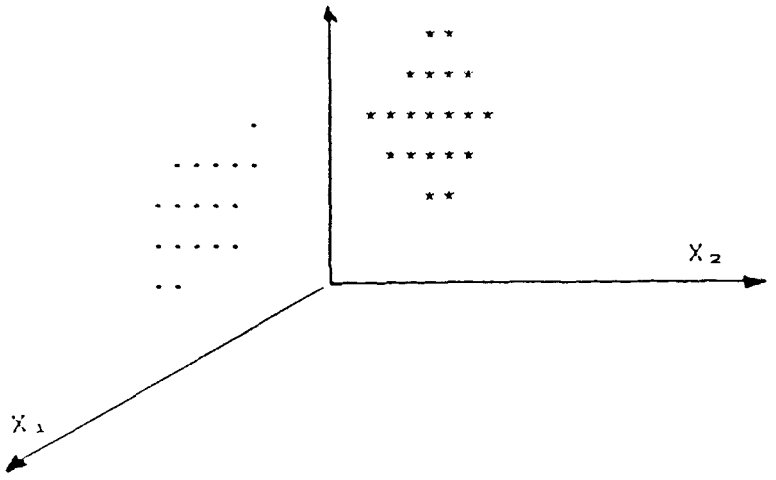


fig.2 la disposition des observations dans des régions distinctes traduit le nombre de modes

Il est important de noter que la base d'apprentissage, qui sert à analyser le système,est contituée uniquement des états connus de celui-ci.

Il arrive parfois que certains états ne soient pas observés,et il en résulte naturellement une base d'apprentissage incomplète.

Cette dernière remarque fera l'objet d'un développement aux chapitres III et IV

Dans ces deux chapitres, on étudie des méthodes qui permettent de déceler éventuellement toute variation dans le nombre de modes du système.

Au cours de l'analyse d'un système,le concours d'un expert est primordial dans les premières phases.

La qualité de la base d'apprentissage dépend essentiellement des variables caractéristiques retenues par l'expert

Ainsi l'objectif de modélisation des systèmes complexes impose de nouvelles méthodes d'analyse et ces dernières conditionnent les moyens de calcul permettant leur mise en oeuvre.

I-2 SELECTION DES CARACTERES

La sélection des paramètres decrivant le système est une étape importante.

Le nombre de publications et de travaux de recherche se rapportant à cette question est éloquent et traduit toute l'importance accordée à cette phase .[FU,80],[NAR,80], [FUK,72],[AND,72].

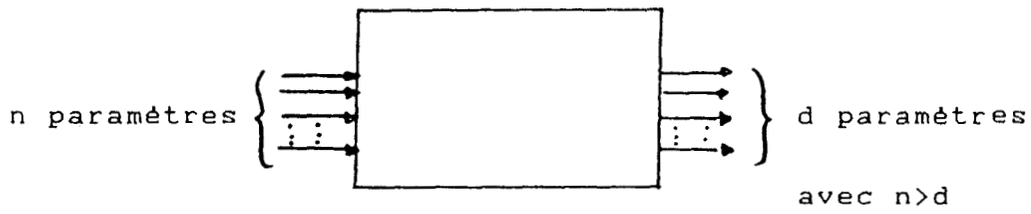


fig.3 sélection des paramètres

Le but essentiel de cette opération est la recherche des meilleurs paramètres afin de satisfaire deux objectifs:

_ le premier objectif est de réaliser,à partir de la base d'apprentissage,une partition de l'espace d'observation en composantes distinctes.

_ le second objectif est de réduire le nombre de paramètres,tout en conservant le premier résultat,afin d'éviter d'être pénalisé en coût de calcul lors de l'opération d'affectation de nouvelles observations.

Les techniques de sélection de caractères sont aussi nombreuses que variées.

Après l'opération de sélection,chaque observation est définie par un vecteur $X=(X_1,X_2,...,X_d)^T$,d représente la dimension de l'espace.

Le résultat de cette opération est naturellement fonction du choix initial des variables caractéristiques.

I-3 METHODES DE CLASSIFICATION AUTOMATIQUE

I-3-1 INTRODUCTION

L'objectif de la classification automatique est de réaliser,à partir de l'ensemble d'apprentissage , une partition de l'espace d'observation en composantes distinctes .

Cette séparation automatique des composantes est importante en reconnaissance des formes ; elle représente un outil précieux dans les problèmes d'apprentissage lorsque aucune information n'est donnée a priori sur les observations les plus représentatives de chaque composante.

Il existe actuellement une panoplie de techniques de classification automatique , et elles ont toutes pour objectif de réaliser une partition acceptable à partir de la base d'apprentissage.

Cette diversité des méthodes de classification automatique provient essentiellement de deux évolutions:

- la première évolution est le besoin de comprendre , d'analyser et finalement de maitriser les systèmes complexes.

-la seconde évolution est le résultat d'une progression des moyens informatiques qui se traduit d'une part par une plus grande capacité de mémorisation d'informations sur le fonctionnement des systèmes et d'autre part par l'augmentation de la puissance des moyens de calcul.

Toutefois on peut classer les méthodes de classification automatique en quatre groupes:

- approche métrique
- approche informationnelle
- approche statistique
- approche graphique

I-3-2 APPROCHE METRIQUE

L'approche métrique est basée sur l'utilisation d'une métrique qui permet de définir des indices de similarité ou de dissimilarité.

Dans cette approche on place:

i- les méthodes de classification hiérarchiques qui permettent d'établir une hiérarchie de partitions sous forme d'un arbre de classification.[CAI,76].

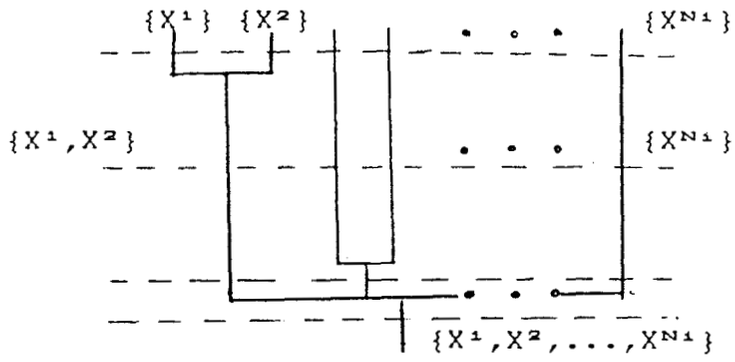


fig.4 arbre de classification

A la racine de l'arbre il y a une seule composante qui est en fait la base d'apprentissage ; les feuilles sont autant de composantes que d'observations dans la base.

Avec une procédure agglomérative on part des feuilles pour aboutir à la racine , alors qu'avec une procédure divisive on suit le chemin inverse.

Ces deux procédures sont arrêtées lorsque la partition obtenue correspond à la partition souhaitée.

ii la méthode du type nuées dynamiques est un algorithme itératif permettant d'obtenir la partition qui minimise un critère fixé W.[DID,79].

iii les méthodes utilisant les concepts flous permettent de définir des indices de dissemblance et de construire des sous-ensembles flous à l'aide d'une fonction d'appartenance. [KAU,75],[BER,86].

Le problème commun à toute les méthodes métriques est naturellement le choix de la métrique .

Il est évident que les résultats obtenus varient fortement en fonction de la métrique adoptée.

I-3-3 APPROCHE INFORMATIONNELLE

La méthode utilisant l'approche informationnelle permet de mettre en évidence le concept d'information entre variables appelé transinformation.[SHA,48],[ASH,65],[STA ,83].

Ce concept est très utile lorsque le couplage entre variables est du à des mécanismes non linéaires.

L'utilisation de quantités telles que l'entropie ou l'information mutuelle entre variables aléatoires nécéssite la discrétisation du domaine des valeurs de chaque variable.

Naturellement les résultats obtenus sont fonction de cette quantification.

I-3-4 APPROCHE STATISTIQUE

Les méthodes statistiques qui sont évoquées dans ce paragraphe ont pour objectif de déterminer les paramètres statistiques caractérisant chaque composante.

Parmi les techniques utilisées on peut citer:

- méthodes bayesiennes
- méthodes du maximum de vraisemblance
- méthodes utilisant les moments

Jusqu'à présent les solutions proposées ne sont applicables que sous certaines conditions:

-matrices de covariances égales [COO ,64],[DAY,69]
-seul la probabilité a priori est inconnue [KAZ,77]
-problème réduit à deux composantes [MAR,77],[FUK,83]
-seules les vecteurs moyennes et les matrices de covariances sont inconnues [MIZ,77]

D'autres solutions proposées nécessitent pour être appliquées des informations a priori difficiles à obtenir quand on analyse une base totalement inconnue.

I-3-5 APPROCHE GRAPHIQUE

La méthode graphique proposée par J.G. POSTAIRE et C.VASSEUR [POS,81] , permet de résoudre ce type de problème sans nécessiter aucune information a priori sur l'ensemble d'apprentissage , avec toutefois l'unique hypothèse que les observations suivent une distribution normale.

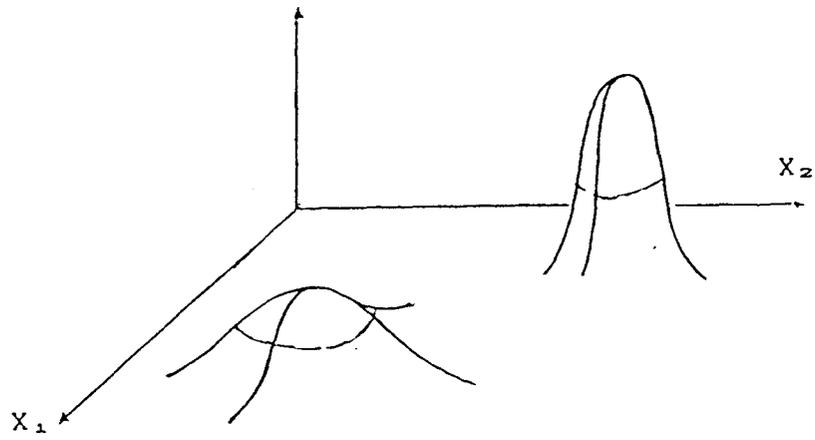


fig.5 fonction de distribution des observations

En utilisant l'algorithme rapide développé par J.G.POSTAIRE et C.VASSEUR [VAS,81] , on arrive à déterminer les modes de la fonction de distribution et les observations les plus représentatives de ces modes.

La méthode est quelque peu sensible à certaines particularités locales de la fonction de distribution.

I-4 MODELISATION DES MODES DE FONCTIONNEMENT

I-4-1 INTRODUCTION

L'objectif de la dernière phase de la figure 1 est de modéliser chacune des composantes afin de pouvoir identifier toute nouvelle observation en la classant à l'aide d'une règle d'attribution adaptée.

Mais avant de réaliser cette opération ,le résultat du partitionnement de l'espace d'observation est analysé par un expert du système étudié.

Généralement le résultat obtenu correspond à la partition souhaitée,dans le cas contraire cette situation est génératrice de nouvelles hypothèses de travail.

Chaque composante est modélisée soit:

-par un jeu de paramètres et dans ce cas l'attribution de nouvelles observations se fait à l'aide d'une décision paramétrique.

-par un ensemble d'observations considérées comme des prototypes et l'identification des nouvelles observations est réalisée à l'aide d'une décision non paramétrique.

I-4-2 DECISION D'ATTRIBUTION NON PARAMETRIQUE

En utilisant la régle des K plus proches voisins,toute nouvelle observation est attribuée à la composante la mieux représentée parmi ses K plus proches voisins.[FUK,84],[TOM,76].

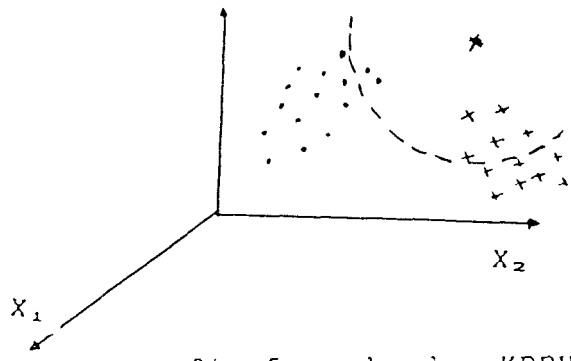


fig.5 regle des KPPV

L'avantage de cette méthode est de ne pas nécéssiter l'utilisation d'hypothèses sur la distribution des observations. Par contre elle présente deux handicaps majeurs:

-le premier vient du fait que la méthode nécéssite de gros efforts de mémorisation des observations.

-le deuxième est le coût de calcul trés élevé nécéssaire pour classer une nouvelle observation.

Des travaux ont été menés pour d'une part réduire le nombre de prototypes représentatifs de chaque composante,et d'autre part pour accélérer les calculs lors de l'attribution d'une observation.[CHIN,74],[GRE,84].

Lorsque les modes sont linéairement séparables,on recherche les meilleurs hyperplans séparateurs.[GU,85].

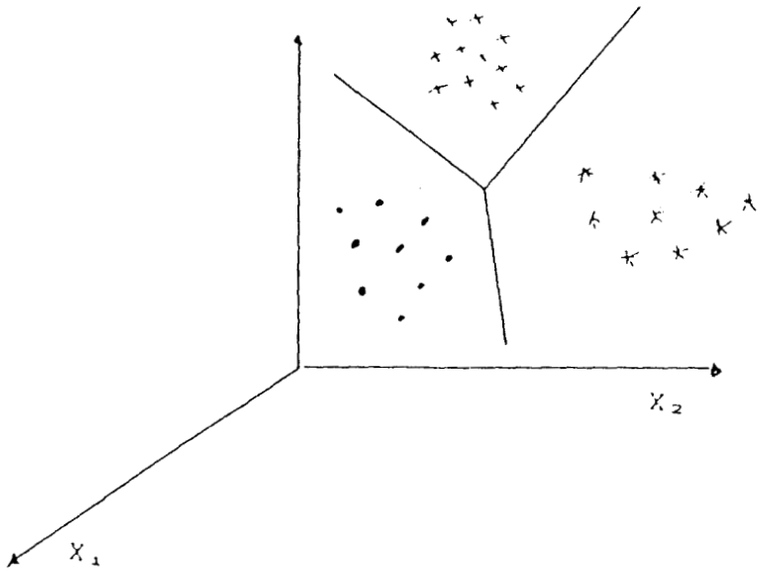


fig.5 composantes linéairement séparables

Les coefficients de ces hyperplans sont déterminés en utilisant les prototypes définissant chaque composante.

I-4-3 DECISION D'ATTRIBUTION PARAMETRIQUE

Une régle d'attribution basée sur une décision bayèsienne est généralement adoptée,car elle minimise l'erreur de classement.

Evidemment cela revient à poser a priori la forme de la loi et à estimer ses paramètres caractèristiques à partir des prototypes dont on dispose.

Cette loi est en général la loi de GAUSS multidimensionnelle.

$$P(X)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\ EXP\{-(X-\bar{X})^T\Sigma^{-1}(X-\bar{X})/2\}$$

avec $X=(X_1,X_2,...,X_d)^T$ vecteur aléatoire

$\overline{X}=(X_1,X_2,...,X_d)^T$ vecteur moyenne

Σ matrice de covariance

Soit C le nombre total de composantes du mélange gaussien.

Chaque composante w_K est définie par ses prototypes:

$w_K=\{X^i\}$ $i=1,N_K$ N_K est le nombre de prototypes de w_K
$K=1,C$

Les paramètres caractèristiques de w_K sont:

-probabilité à priori

$$P^K=\frac{N_K}{\sum_{i=1}^{C} N_i} \quad \text{avec les contraintes} \quad \begin{array}{l} 0\leqslant P^K \leqslant 1 \\ \sum_{K=1}^{C} P^K=1 \end{array}$$

-vecteur moyenne

$$\overline{X}^K = \frac{1}{N_K} \sum_{i=1}^{N_K} X^i$$

-matrice de covariance

$$\Sigma = \frac{1}{N_K - 1} \sum_{i=1}^{N_K} (X^i - \overline{X}^K)(X^i - \overline{X}^K)^T$$ cette matrice est symétrique, définie et positive

-valeurs et vecteurs propres

$$\lambda^K{}_j \qquad K=1,C$$
$$U^K{}_j \qquad j=1,d$$
$$\Sigma U^K{}_j = \lambda^K{}_j U^K{}_j \quad \forall\ K=1,C\ ;j=1,d$$

L'estimation des paramètres valeurs et vecteurs propres revêt un intéret certain dans la mesure où ces paramètres fournissent une image précise de chaque composante dans l'espace d'observation.

Les vecteurs propres fournissent les directions des axes d'inertie de la composante,et les valeurs propres fournissent la dispersion des observations autour de ces directions propres.[CAI,76].

I-5 APPROCHE RECURSIVE DANS L'ESTIMATION DES PARAMETRES STATISTIQUES

Il est fondamental de prévoir une approche récursive dans l'estimation des paramètres statistiques d'un mélange gaussien , ce qui permet d'intégrer d'une manière dynamique les nouvelles observations qui viennent enrichir la base d'apprentissage.

On a vu dans le premier paragraphe de ce chapitre que la base utilisée peut être dans certain cas incomplète et qu'il est fondamental de l'enrichir par l'apport de nouvelles observations.

Cette approche récursive permet d'affiner les estimations des paramètres statistiques fournies par les méthodes de classification automatique développées précédemment qui demeurent essentiellement caractérisées par une approche statique.

La démarche récursive dans l'identification d'un mélange gaussien est développée selon le schéma suivant:

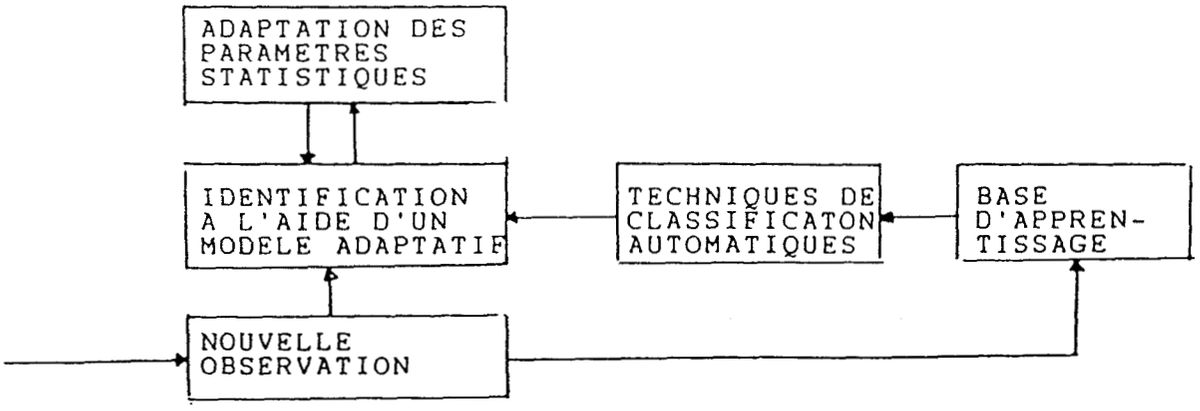


fig.7 approche récursive dans l'analyse d'un système

-les valeurs initiales des paramètres sont estimées à partir d'une base d'apprentissage selon la méthode graphique.
-chaque nouvelle observation est intégrée selon le schéma de la figure 7

Cette démarche est analogue à celle proposée par T.GU et B.DUBUISSON.[GU,85].

Le chapitre II est entièrement consacré à la question de l'adaptation des paramètres en fonction de l'arrivée des observations.

Si l'estimation récursive des paramètres tels que probabilité a priori , vecteur moyenne et matrice de covariance est relativement simple à effectuer par contre l'estimation des valeurs et vecteurs propres est plus délicate.

L'estimation de ces derniers paramètres est importante dans la mesure où ils fournissent une image précise sur l'évolution de chaque composante en fonction de l'enrichissement de la base d'apprentissage.

Ce suivi permanent de la réprésentation spatiale de chaque composante permet de déceler tout changement dans la structure du mélange.

Les chapitres III et IV sont essentiellement centrés sur cette question.

Les figures 8,9,10,11,12 et 13 illustrent selon les travaux de WALLISER [WAL,77] ,le changement de structure qui peut intervenir dans un mélange gaussien.

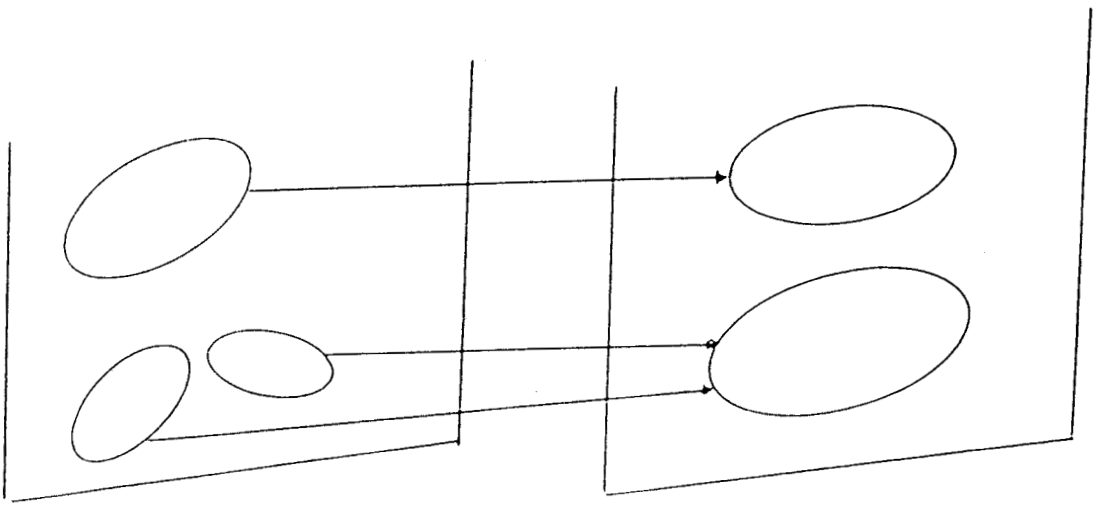


fig.8 fusionnement de deux composantes.

pour un cas monovariable,ce fusionnement correspond à la situation suivante:

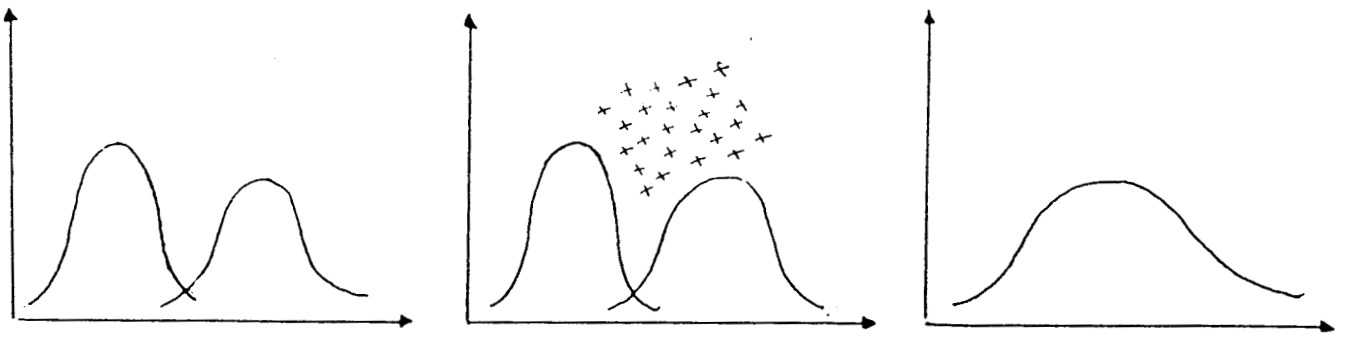


fig.9 fusionnement dans le cas monovariable

Les figures 10 et 11 représentent l'éclatement d'une composante en deux

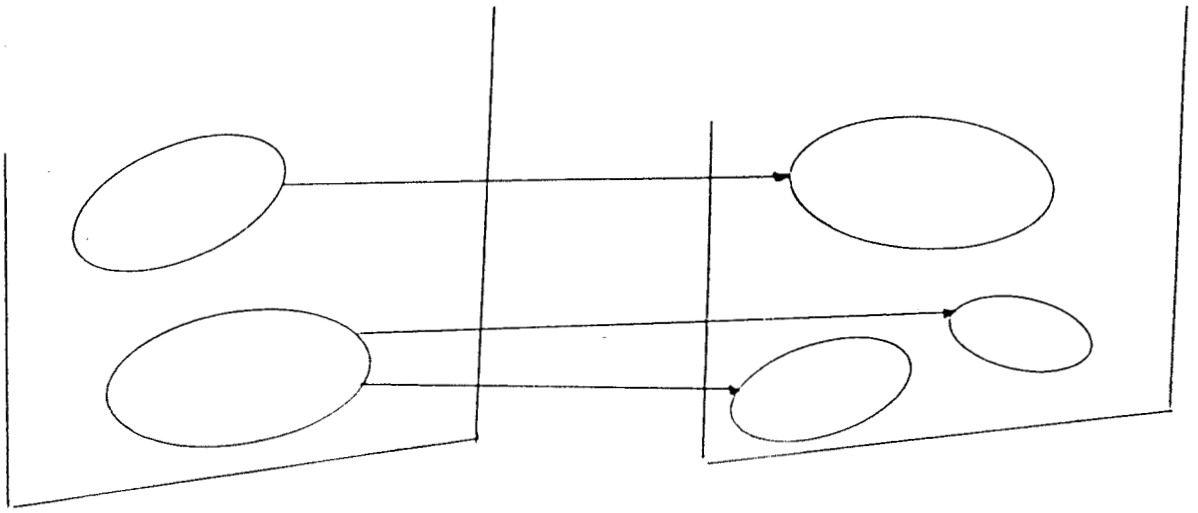


fig.10 scission d'une composante en deux.

Dans le cas monodimensionnel,cette scission est traduite par la figure 11

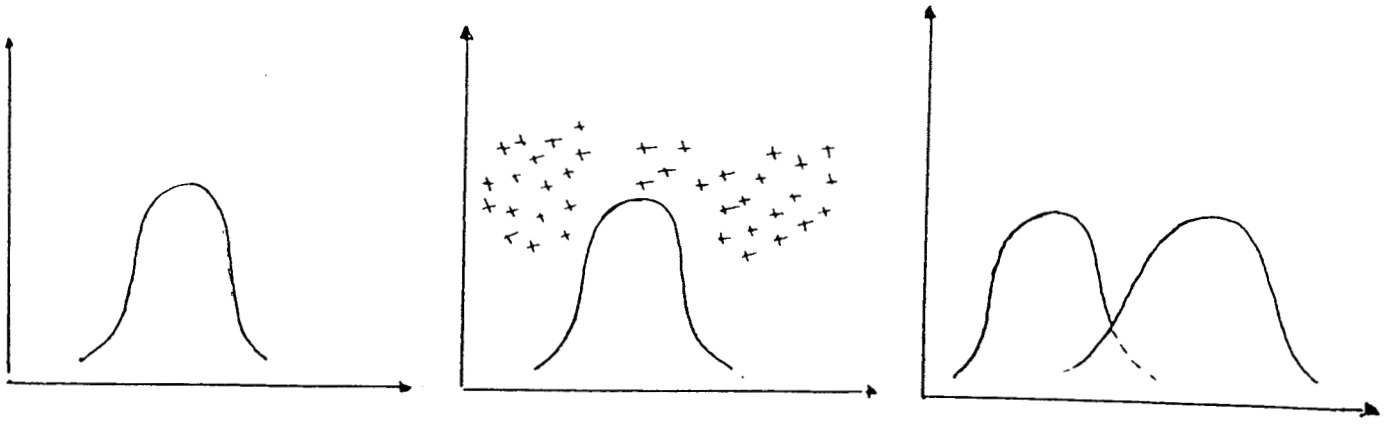


fig.11 scission d'une composante dans le cas monovariable

Le troisième et dernier cas correspond à la situation où une nouvelle composante apparait.

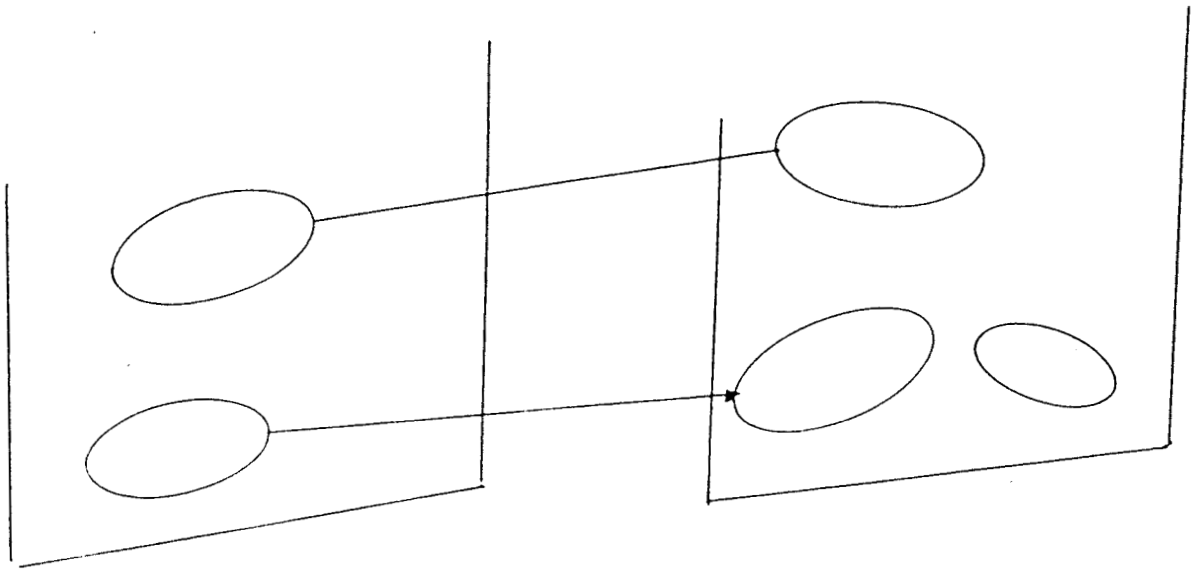


fig. 12 apparition d'une nouvelle composante

et pour le cas monodimensionnel,on a:

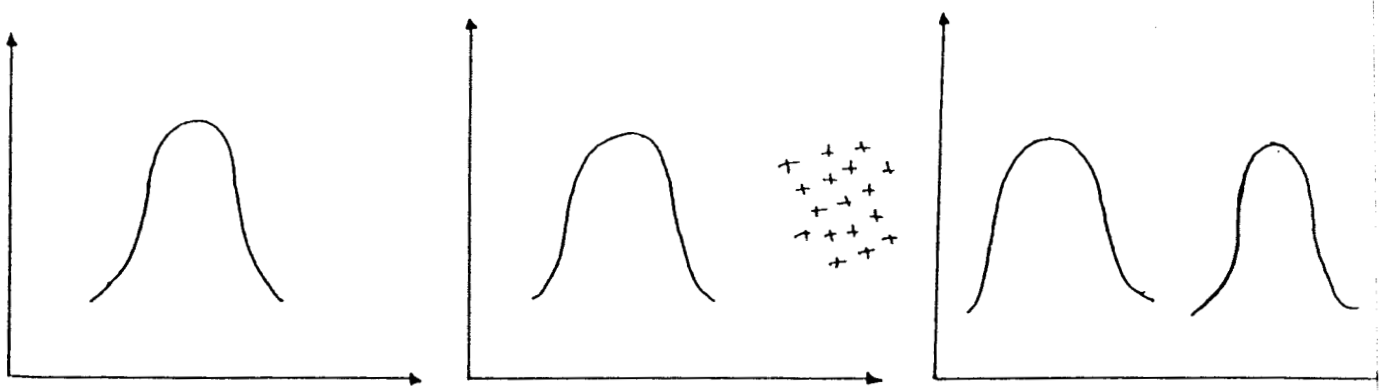


fig.13 apparition d'une composante dans le cas monovariable.

BU LILLE

I-6 CONCLUSION

L'étude des systèmes complexes par l'analyse des données est de plus en plus utilisée . Cette nouvelle approche est le fruit de plusieurs évolutions convergentes.

L'approche récursive dans l'estimation des paramètres d'un mélange gaussien permet d'apporter des améliorations aux résultats fournis par une quelconque méthode statique.

La qualité de la base d'apprentissage dépend d'une part du choix des variables caractéristiques qui décrivent le système et d'autre part de la durée de l'horizon d'observation.

Comme on ne peut pas allonger indéfiniment l'horizon d'observation , il importe de déceler tout changement dans la structure du modèle dans le cas où la base d'apprentissage ne contient pas tous les modes de fonctionnement . Cette étude fait l'objet des chapitres III et IV .

REFERENCES

A.TITLI ET AL
" ANALYSE ET COMMANDE DES SYSTEMES COMPLEXES"
CEPADUES EDITION 1979

M.RICHETIN
" ANALYSE STRUCTURALE DES SYSTEMES COMPLEXES EN VUE D'UNE COMMANDE HIERARCHISEE"
THESE DE DOCTORAT D'ETAT,UNIVERSITE DE TOULOUSE JUILLET 1975

M.STAROSWIEKI
" CONTRIBUTION A L'ANALYSE ET A LA COMMANDE DE SYSTEMES COMPLEXES A CRITERES MULTIPLES"
THESE DE DOCTORAT D'ETAT,UNIVERSITE DE LILLE 1 JANVIER 1979

J.RISSANEN
" MODELING BY SHORTEST DATA DESCRIPTION"
AUTOMATICA 1978

E.H.BRISTOL
" PATTERN RECOGNITION :AN ALTERNATIVE TO PARAMETER IDENTIFICATION IN ADAPTATIVE CONTROL"
AUTOMATICA 1977

J.P BENZECRI
" ANALYSE DES DONNEES TOME 2"
DUNOD 1980

K.S.FUKUNAGA
" RECENT DEVELOPPEMENT IN PATTERN RECOGNITION"
IEEE C.29 N°10 OCTOBRE 1980

P.M.NARENDA AND K.S.FUKUNAGA
" A CLASS OF FEATURE EXTRACTION CRITERIA AND ITS RELATION TO THE BAYES RISK"
IEEE IT-26 N°1 1980

P.M.NARENDA AND K.S.FUKUNAGA
" NON LINEAR FEATURE EXTRACTION WITH A GENERAL CRITERIA FUNCTION"
IEEE IT-24 N°5 1978

K.S.FUKUNAGA
" INTRODUCTION TO STATISTICAL PATTERNS RECOGNITION"
ACADEMIC PRESS NEW-YORK 1972

H.C.ANDREWS
" INTRODUCTION TO MATHEMATICAL TECHNICS IN PATTERN RECOGNITION"
WILEY-INTERSCIENCE NEW-YORK 1972

F.CAILLIEZ ET J.G.PAGE
" INTRODUCTION A L'ANALYSE DES DONNEES"
SMASH

A.KAUFMANN
" INTRODUCTION A LA THEORIE DES SOUS ENSEMBLES FLOUS TOME 3"
MASSON 1975

M.BEREAU
" CONTRIBUTION A LA THEORIE DES SOUS ENSEMBLES FLOUS A LA REGLE DE DISCRIMINATION DES K PLUS PROCHES VOISINS EN MODE PARTIELLEMENT SUPERVISE"
THESE DE DOCTORAT U.T DE COMPIEGNE JUILLET 1986

E.DIDAY
" OPTIMISATION EN CLASSIFICATION AUTOMATIQUE"
DUNOD 1979

C.E.SHANNON
" A MATHEMATICAL THEORY OF COMMUNICATION"
THE BELL SYSTEM TECHNICAL JOURNAL VOL XXVII JUILLET 1948

W.R.ASHBY
" MEASURING THE INTERNAL INFORMATION EXCHANGE IN A SYSTEM"
CYBERNITICA N°1 1965

M.STAROSWIEKI V.TORO M.SBAI
" STRUCTURAL ANALYSIS AND DECOMPOSITION OF COMPLEX SYSTEMS"
3RD IFAC IFORS SYMPOSIUM LSSTA VARSOVIE 1983

D.W.COOPER AND P.W.COOPER
" NON SUPERVISED ADAPTATIVE SIGNAL DETECTION AND PATTERN ANALYSIS"
INFORM CONT VOL 7 SEPTEMBRE 1964

N.E.DAY
" ESTIMATION THE COMPONENTS OF A MIXTURE OF NORMAL DISTRIBUTIONS"
BIOMETRICA VOL 56 DECEMBRE 1969

D.KAZAKOS
" RECURSIVE ESTIMATION OF A PRIOR PROBABILITIES USING A MIXTURE"
IEEE TRANS INFOR THEORY VOL 23 MARS 1977

U.E.MARKOV AND A;F.SMITH
" A QUASI BAYES UNSUPERVISED LEARNING PROCEDURE FOR PRIOR"
IEEE TRANS INFOR THEORY VOL 23 MARS 1977

K.S.FUKUNAGA AND T.E.FLICK
" ESTIMATION OF THE PARAMETERS OF A GAUSSIAN MIXTURE USING THE METHOD OF MOMENTS"
IEEE TRANS PAMI VOL 5 N°4 JULY 1983

T.Y.YONG AND C.CORALUPPI
" STOCHASTIQUE ESTIMATION OF A MIXTURE OF NORMAL DENSITE FUNCTION UNUSING AN INFORMATION THEORY CRITERION"
IEEE TRANS INFORM THEORY VOL 16 MAY 1970

R.MISOGUCHI AND M.SHUMIRA
" AN APPROACH TO UNSUPERVISED LEARNING CLASSIFICATION"
IEEE TRANS COMP VOL 24 OCTOBRE 1975

J.G.POSTAIRE ET C.VASSEUR
" AN APPROXIMATE SOLUTION TO NORMAL MIXTURE IDENTIFICATION WITH APPLICATION TO UNSUPERISED PATTERN CLASSIFICATION"
IEEE TRANS PAMI VOL 3 MARS 1981

J.G.POSTAIRE ET C.VASSEUR
" A FAST ALGORITHM FOR NONPARAMETRIC PROBABILITY DENSITY ESTIMATION "
IEEE TRANS PAMI VOL 4 N°6 NOV 1982

K.S.FUKUNAGA AND T.E.FLICK
" AN OPTIMAL GLOBAL NEAREST NEIGHBORS METRIC"
IEEE TRANS PAMI VOL 6 N°3 MARS 1984

I.TOMEL
" A GENERALISATION OF THE K-NN RULE"
IEEE TRANS SMC VOL 6 N°2 FEB 1976

C.L.CHANG
" FINDING PROTOTYPES FOR NEAREST NEIGHBOR CLASSIFIERS"
IEEE TRANS COMP VOL 23 N°11 NOV 1974

A.GRENIER
" EXTENSION DES METHODES STATISTIQUES DE RECONNAISSANCE DE FORMES POUR L'APPLICABILITE AU DIAGNOSTIC"
THESE DE DOCTEUR INGENIEUR,UNIVERSITE DE TECHNOLOGIE DE COMPIEGNE MARS 1985

T.GU ET B.DIBUISSON
" A GENERAL MODEL OF PATTERN CLASSIFICATION"
RAIRO PP 69-91 1985

B.WALLISER
" SYSTEMES ET MODELES,INTRODUCTION CRITIQUE A L'ANALYSE DES SYSTEMES"
EDITION SEUIL 1977

CHAPITREII

ESTIMATION RECURSIVE DES PARAMETRES STATISTIQUES D'UN MELANGE GAUSSIEN

CHAPITREII

ESTIMATION RECURSIVE DES PARAMETRES STATISTIQUES D'UN MELANGE GAUSSIEN

II-1 INTRODUCTION 1

II-2 MISE EN EQUATION 4

II-3 DETERMINATION DE LA RECURRENCE SUR LA MATRICE DE COVARIANCE 7

II-4 DETERMINATION DE LA RECURRENCE SUR LES VALEURS PROPRES 9

II-5 DETERMINATION DE LA RECURRENCE SUR LES DIRECTIONS PROPRES ...12

II-6 DETERMINATION DE LA RECURRENCE SUR LE VECTEUR MOYENNE ET LA PROBABILITE A PRIORI 13

II-7 ALGORITHME RAPIDE DE CALCUL DE λ^{+}_{1} 15

II-8 RECAPITULATIF 18

II-9 SIMULATIONS ET PERFORMANCES 19

II-9-1 METHODE 19

II-9-2 SIMULATION 2D 21

II-9-3 SIMULATION 3D 22

II-9-4 ANALYSE DES RESULTATS OBTENUS 23

II-10 CONCLUSION 24

II-1 INTRODUCTION

Soit un mélange gaussien multidimensionnel dont les paramètres statistiques sont les suivants:

-probabilité a priori	P^K ;	$K=1,C$
-vecteur moyenne	$\bar{X}^K$;	$K=1,C$
-matrice de covariance	Σ_K ;	$K=1,C$
-valeurs et vecteurs propres	λ^K_i ;	$K=1,C$
	U^K_i ;	$i=1,d$

-fonction de densité de probabilité conditionnelle

$$P^K(X)=\frac{1}{(2\pi)^{d/2}|\Sigma_K|^{1/2}}EXP\{-(X-\bar{X}^K)^T\Sigma^{-T}(X-\bar{X}^K)/2\}$$

avec
$$\begin{cases} \sum_{K=1}^{C} P^K P^K(X)=1 \\ C\text{:nombre de composantes} \\ d\text{:dimension de l'espace d'observation} \end{cases}$$

Soit alors un nouvel échantillon X attribué à la composante w_K selon une régle de décision qui minimise l'erreur de classement.[CHI,77],[TSO,78]

La figure 1 illustre,dans le cas d=2 les modifications de stucture dans la composante considérée induites par l'intégration d'une nouvelle observation.

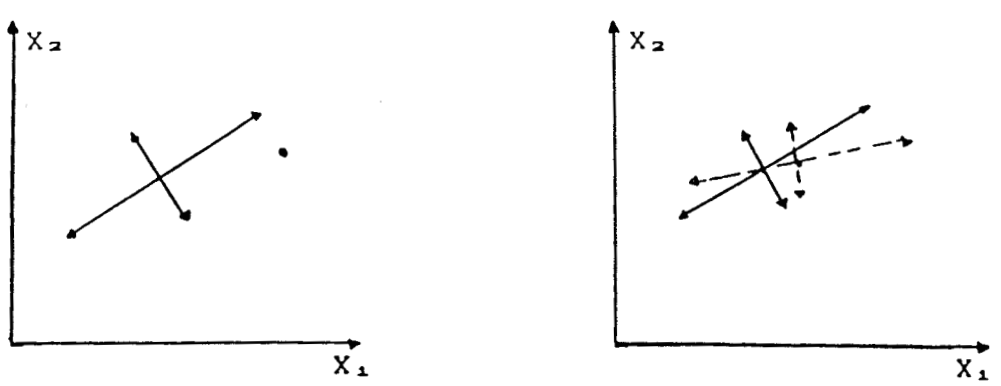


fig.1 modification de structure d'une composante

On peut décomposer les effets de l'arrivée d'une nouvelle observation selon les trois opérations illustrées par la figure2 .

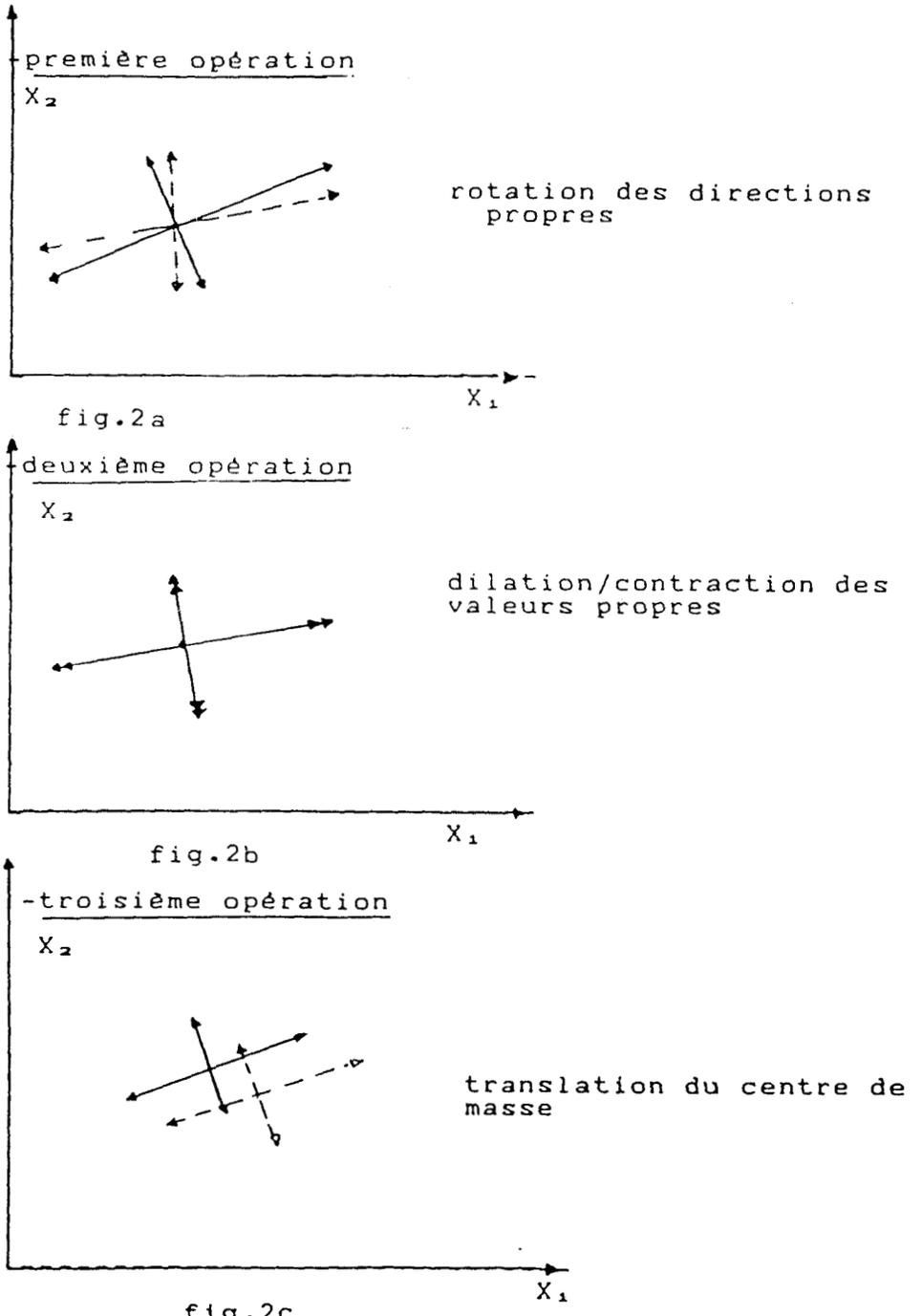


fig.1 décomposition de la modificationde structure

Il est clair que l'ordre de déroulement de ces trois opérations est indifférent.

La première opération (rotation des directions propres) illustre le passage des anciens vers les nouveaux vecteurs propres:

$$U^K_i \longrightarrow U^K_{i^*}$$

La deuxième opération (dilatation/contraction des valeurs propres) illustre le passage des anciennes aux nouvelles valeurs propres:

$$\lambda^K_i \longrightarrow \lambda^K_{i^*}$$

La troisième opération (translation du centre de masse) illustre le passage de l'ancien au nouveau vecteur moyenne:

$$\bar{X}^K \longrightarrow \bar{X}^{K*}$$

Pour toutes les autres composantes ($i \neq K$) les estimations des paramètres statistiques sont inchangées.Seules les probabilités a priori sont révisées en baisse

Chaque opération mentionnée précédemment est décrite en détail dans les développementsqui suivent.

II-2 MISE EN EQUATION

Pour la $K^{ième}$ composante, la matrice symétrique de covariance peut se mettre sous la forme suivante:

$$\Sigma_K = [U^K{}_1 \quad U^K{}_2 \ldots \quad U^K{}_d] \begin{bmatrix} \lambda^K{}_1 & & 0 \\ & \lambda^K{}_2 & \\ & & \ddots \\ 0 & & \lambda^K{}_d \end{bmatrix} [U^K{}_1 \; U^K{}_2 \; \ldots U^K{}_d]^T \quad (1)$$

Après intégration d'une nouvelle observation dans la composante w_K la matrice de covariance devient:

$$\Sigma_K{}^+ = [U^K{}_1{}^+ \; U^K{}_2{}^+ \ldots U^K{}_d{}^+] \begin{bmatrix} \lambda^K{}_1{}^+ & & 0 \\ & \lambda^K{}_2{}^+ & \\ & & \ddots \\ 0 & & \lambda^K{}_d{}^+ \end{bmatrix} [U^K{}_1{}^+ \quad U^K{}_2{}^+ \ldots U^K{}_d{}^+]^T \quad (2)$$

Les ensembles de vecteurs propres $(U^K{}_1, U^K{}_2, \ldots, U^K{}_d)$ et $(U^K{}_1{}^+, U^K{}_2{}^+, \ldots, U^K{}_d{}^+)$ constituent deux bases différentes de l'espace d'observation.

Le passage d'une base à l'autre se fait selon le tableau suivant:

	$U^{K}_{1}{}^{+}$	$U^{K}_{2}{}^{+}$	...	$U^{K}_{d}{}^{+}$
U^{K}_{1}	O^{K}_{11}	O^{K}_{12}	...	O^{K}_{1d}
U^{K}_{2}	O^{K}_{21}	O^{K}_{22}	...	O^{K}_{2d}
U^{K}_{d}	O^{K}_{d1}	O^{K}_{d2}	...	O^{K}_{dd}

(3)

Dans la suite du développement,pour ne pas alourdir les notations,l'indice K est supprimé;Ceci ne nuit pas à la compréhension.

Le tableau (3) contient la matrice [O] qui permet le passage d'une base vers l'autre.

$$[O]=\{ o_{ij}\} \qquad i,j=1,d \qquad (4)$$

avec $U_i{}^{+}= [O] \quad U_i \qquad \forall\ i=1,d \qquad (5)$

$$[O]^{-1}=[O]^{T}$$

Les élèments o_{ij} de la matrice [O] vérifient les relations suivantes:

$$\sum_{j=1}^{d}(o_{ij})^2 = \sum_{i=1}^{d}(o_{ij})^2 =1 \qquad \forall\ i,j \qquad (6a)$$

$$\sum_{j=1}^{d} o_{ij}.o_{mj} =0 \qquad \forall\ i\neq m \qquad (6b)$$

$$\sum_{i=1}^{d} o_{ij}.o_{im} =0 \qquad \forall\ j\neq m$$

Ces résultats découlent du fait que les bases sont orthonormées.

A partir de cette mise en équation la démarche de calcul est la suivante:

i) détermination de la récurrence pour la matrice de covariance
ii) " " " les valeurs propres
iii) " " " les vecteurs propres
iv) " " " le vecteur moyenne et la probabilité a priori

II-3 DETERMINATION DE LA RECURRENCE SUR LA MATRICE DE COVARIANCE

Soit $\{ X^i, i=1,N_k \}$ l'ensemble des prototypes représentant la composante w_k.

L'estimation de la matrice Σ est établie à partir des prototypes X^i , selon l'expression suivante:

$$\Sigma = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (X^i - \overline{X})(X^i - \overline{X})^T \qquad (7a)$$

L'intégration de l'observation X^{N_k+1} dans la base d'apprentissage conduit à une nouvelle estimation Σ^+ donnée par

$$\Sigma^+ = \frac{1}{N_k} \sum_{i=1}^{N_k+1} (X^i - \overline{X}^+)(X^i - \overline{X}^+)^T \qquad (7b)$$

Dès lors , la recurrence liant Σ^+ à Σ est donnée par un expression de la forme:

$$\Sigma^+ = \alpha\Sigma + \beta R \qquad (8)$$

avec $\alpha = \dfrac{N_k - 1}{N_k}$ $\qquad \beta = \dfrac{1}{(N_k + 1)}$

$$R = (X^{N_k+1} - \overline{X})(X^{N_k+1} - \overline{X})^T$$

En posant $(X^{Nk+1} - \bar{X})^T = (\Delta X_1, \Delta X_2, \ldots, \Delta X_d)^T$ la matrice R prend la forme développée suivante:

$$R = \begin{bmatrix} (\Delta X_1)^2 & \Delta X_1 . \Delta X_2 & & \Delta X_1 . \Delta X_d \\ \Delta X_2 . \Delta X_1 & (\Delta X_2)^2 & & \Delta X_2 . \Delta X_d \\ & & & \\ \Delta X_d . \Delta X_1 & \Delta X_d . \Delta X_2 & & \Delta X_d . \Delta X_d \end{bmatrix}$$

- on trouvera en annexe le détail des calculs conduisant à l'établissement de l'expression (8).

II-4 DETERMINATION DE LA RECURRENCE SUR LES VALEURS PROPRES

Par définition des valeurs propres, on a les égalités suivantes:

$$\Sigma^{*} U_i^{*} = \lambda^{*}_i \; U_i^{*} \qquad (10)$$

$$\Sigma^{*} = [U^{*}] \; [\lambda^{*}] \; [U^{*}]^T$$

En utilisant l'expressions (8) ,on obtient :

$$(\alpha\Sigma + \beta R) = [U^{*}][\lambda^{*}][U^{*}]^T$$

Et en se plaçant dans la base [U],il vient

$$[U]^T(\alpha\Sigma + \beta R)[U] = [U]^T[U^{*}][\lambda^{*}][U^{*}]^T[U]$$

En utilisant l'expression (5),on obtient

$$[U]^T(\alpha\Sigma + \beta R)[U] = [O][\lambda^{*}][O]^T \qquad (11)$$

Avec

$$[U]^T(\alpha\Sigma)[U] = \alpha[\lambda]$$

$$[U]^T(\beta R)[U] = \beta R^{*}$$

$$R^{*} = (X^{Nk+1} - \overline{X})^{*}(X^{Nk+1} - \overline{X})^{T*}$$

$$(X^{Nk+1} - \overline{X})^{*} = [U]^T(X^{Nk+1} - \overline{X})$$

On conserve pour R^{*} la même représentation que celle de R

A partir de (11) on tire:

$$\alpha[\lambda] + \beta R^{-} = [O][\lambda^{+}][O]^{T}$$

$$(\alpha[\lambda] + \beta R^{-})[O] = [O][\lambda^{+}] \qquad (12)$$

En extrayant la $i^{ème}$ colonne de l'égalité matricielle (12)

$$\alpha\lambda_1 O_{1i} + \beta\Delta X_1(O_{1i}.\Delta X_1 + O_{2i}.\Delta X_2 + \ldots + O_{di}.\Delta X_d) = O_{1i}\lambda^{+}_{i}$$

$$\alpha\lambda_2 O_{2i} + \beta\Delta X_2(O_{1i}.\Delta X_1 + O_{2i}.\Delta X_2 + \ldots + O_{di}.\Delta X_d) = O_{2i}\lambda^{+}_{i}$$

$$(13)$$

$$\alpha\lambda_d O_{di} + \beta\Delta X_d(O_{1i}.\Delta X_1 + O_{2i}.\Delta X_2 + \ldots + O_{di}.\Delta X_d) = O_{di}\lambda^{+}_{i}$$

A partir de(13) on tire:

$$\frac{O_{1i}(\lambda^{+}_{i} - \alpha\lambda_1)}{\Delta X_1} = \frac{O_{2i}(\lambda^{+}_{i} - \alpha\lambda_2)}{\Delta X_2} = \ldots = \frac{O_{di}(\lambda^{+}_{i} - \alpha\lambda_d)}{\Delta X_d} = F_i \quad (14)$$

Avec

$$F_i = \beta(O_{1i}.\Delta X_1 + O_{2i}.\Delta X_2 + \ldots + O_{di}.\Delta X_d) \qquad \forall\ i=1,d$$

Par aileurs pour la $i^{ème}$ égalite du systême (13),on peut écrire:

$$\lambda^{+}_{i} = \alpha\lambda_i + \beta\Delta X_i \sum_{m=1}^{d} \frac{O_{mi}}{O_{ii}} \Delta X_m \qquad (15)$$

les rapports tels que O_{mi}/O_{ii} apparaissant dans l'équation (15) peuvent être calculés à partir de l'équation (14) selon l'expression suivante:

$$\frac{O_{mi}}{O_{ii}} = \frac{\Delta X_m}{\Delta X_i} \frac{\lambda^+_i - \alpha \lambda_i}{\lambda^+_i - \alpha \lambda_m} \qquad \forall_{m=1,d} \qquad (16)$$

En reportant ces valeurs dans l'expression (15) on obtient l'équation finale en λ^+_i recherchée.

$$\beta \left[\frac{(\Delta X_1)^2}{\lambda^+_i - \alpha \lambda_1} + \frac{(\Delta X_2)^2}{\lambda^+_i - \alpha \lambda_2} + \ldots + \frac{(\Delta X_d)^2}{\lambda^+_i - \alpha \lambda_d} \right] = 1 \qquad (17)$$

Cette équation se ramène à la résolution d'une équation polynomiale en λ^+_i de degré d .La résolution de cette équation n'est en général pas simple.De plus le choix de la solution parmi les d racines n'est pas immédiat.

Pratiquement l'estimation des λ^+_i se fera selon une méthode algorithmique rapide qui est détaillée dans le paragraphe 7 de ce chapitre.

II-5 DETERMINATION DE LA RECURRENCES SUR LES VECTEURS PROPRES

L'estimation des nouveaux vecteurs propres $(U_1^*, U_2^*, \ldots, U_d^*)$ se ramène à la détermination de la matrice de rotation [O].

Pour cela on utilise les relations (16) et (17) qui permettent d'établir les expressions suivantes:

-d'aprés (16)

$$O_{ji} = \frac{\Delta X_j}{\Delta X_i} \quad \frac{\lambda^*_j - \alpha \lambda_i}{\lambda^*_i - \alpha \lambda_j} \quad O_{ii} \qquad (18)$$

-d'aprés (6a) et (18)

$$(O_{ii})^2 \sum_{j=1}^{d} \frac{(\Delta X_j)^2}{(\Delta X_i)^2} \frac{(\lambda^*_i - \alpha \lambda_i)^2}{(\lambda^*_i - \alpha \lambda_j)^2} = 1 \quad \forall\ i=1,d \qquad (19)$$

Sachant que l'intégration d'un nouveau échantillon ne peut pas provoquer une rotation supérieur à $\pi/2$, on choisit la racine positive pour O_{ii}.

$$O_{ii} = \left\{ \sum_{j=1}^{d} \frac{(\Delta X_j)^2}{(\Delta X_i)^2} \frac{(\lambda^*_i - \alpha \lambda_i)^2}{(\lambda^*_i - \alpha \lambda_j)^2} \right\}^{-1/2} \quad \forall i=1,d \qquad (20)$$

Les équations (20) et (18) sont utilisées pour déterminer les autres éléments de la matrice [O].

II-6 DETERMINATION DE LA RECURRENCE SUR LE VECTEUR MOYENNE ET LA PROBABILITE A PRIORI

a) Vecteur moyenne

Si N_K est le nombre de prototypes qui ont servi pour identifier la composante w_K , le vecteur moyenne $\overline{X}$ est doté d'un poids egal à N_K.

L'échantillon à intégrer est doté d'un poids egal à 1.

Le nouveau vecteur moyenne X^+ est donné par l'expression:

$$\overline{X}+ = \overline{X} + \frac{1}{N_K+1} (X^{N_K+1} - \overline{X}) \qquad (21)$$

b) probabilité a priori

Soit M le nombre total d'observations dans la base d'apprentissage.

$$M = \sum_{i=1}^{C} N_i$$

et P^i i=1,C les probabilités a priori des différentes composantes.

Aprés intégration d'une nouvelle observation dans w_K , les probabilités a priori P^i telles que $i \neq K$ sont révisée en baisse tandis que P^K est augmentée.

$$P^K = \frac{N_K}{M} \qquad \text{et} \qquad P^+ = \frac{N_{K+1}}{M+1}$$

A partir de ces deux expression on tire:

$$M.P^K + 1 = (M+1).P^{K+}$$

$$P^{K+} = P^K \frac{M}{M+1} + \frac{1}{M+1}$$

qu'on peut facilement mettre sous la forme

$$P^{K+} = P^K + \frac{(1 - P^K)}{M+1}$$

$$P^i = \frac{N_i}{M} \qquad \text{et} \qquad P^{i+} = \frac{N_i}{M+1} \qquad i \neq K$$

$$P^i.M = P^i(M+1)$$

$$P^{i+} = P^i \frac{M}{M+1}$$

$$P^{i+} = P^i - \frac{P^i}{M+1}$$

On peut écrire finalement:

$$P^{i+} = P^i + \frac{(F - P^i)}{M+1} \qquad \text{avec } F = \begin{cases} 1 & \text{si } i=K \\ 0 & \text{autrement} \end{cases} \qquad (22)$$

II-7 CALCUL RAPIDE DES λ^{+}_{i}

Formellement, le calcul des λ^{+}_{i} se ramène à la résolution de l'équation (17). La résolution de cette équation non linéaire conduit à des calculs lourds et longs.

le principe d'accélération du calcul de λ^{+}_{i} est fondé sur l'observation suivante:

- lorsque l'on intègre un nouvel échantillon dans une composante du mélange, la rotation des vecteurs propres est d'autant plus faible que le nombre de prototypes représentant cette dernière est grand.

Cette observation est illustrée par la figure 3

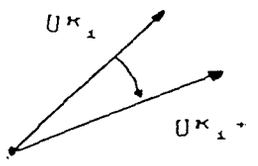


fig.3 faible rotation des directions propres

Par ailleurs, cette remarque se traduit sur les éléments de la matrice de rotation [O] par les approximations suivantes:

$$\left.\begin{array}{l} O_{ii} \simeq 1 \\ O_{ji} \simeq 0 \end{array}\right\} \forall\ j=1,d \qquad (23)$$

Compte tenu de ces remarques, l'expression (15) fournit une première estimation $\lambda^{+}_{i,\infty}$ pour λ^{+}_{i} selon l'expression:

$$\lambda^{+}_{i,\infty} = \alpha\,\lambda_{i} + \beta(\Delta X_{i})^{2} \qquad (24)$$

A partir de cette première estimation , le calcul de λ^{+}_{i} s'effectue selon le processus itératif suivant:

- on reporte la valeur $\lambda^{+}_{i,0}$ de l'expression (24) dans l'expression (16) en remplacement de λ^{+}_{i} et on calcule de cette manière les différents rapports O_{mi}/O_{ii}.

- en reportant les rapports O_{mi}/O_{ii} dans l'expression (15) on peut calculer une nouvelle estimation de $\lambda^{+}_{i,1}$ de λ^{+}_{i}.

- on répète les deux opérations précédentes jusqu'à ce que la différence des deux estimations de λ^{+}_{i} soit inférieure à un δ fixé.

- l'estimation choisie pour λ^{+}_{i} est la dernière approximation calculée.

Ce processus itératif se traduit par l'algorithme suivant:

procédure de calcul de λ^{+}_{i}

$\varepsilon = 2\delta$
$n = 0$
$\lambda^{+}_{i,0} = \alpha \lambda_i + \beta (\Delta X_i)^2$
WHILE $\varepsilon > \delta$
$n = n+1$
DO $\lambda^{+}_{i,n} = \lambda^{+}_{i,0}$
FOR $k=1,d \quad k \neq i$
DO

$$\frac{O_{mi}}{O_{ii}} = \frac{\Delta X_m}{\Delta X_i} \frac{\lambda^{+}_{i,n} - \alpha \lambda_i}{\lambda^{+}_{i,n} - \alpha \lambda_m}$$

$\lambda^{+}_{i,n} = \lambda^{+}_{i,n} + \beta \Delta X_i . \Delta X_m . O_{mi}/O_{ii}$
END
END
$\varepsilon = | \lambda^{+}_{i,n} - \lambda^{+}_{i,n-1} |$
END
END
$\lambda^{+}_{i} = \lambda^{+}_{i,n}$

END

La convergence rapide de cet algorithme a été vérifiée sur des exemples bi et tridimensionnels qui sont présentés et commentés dans les paragraphes qui suivent.

Ces exemples permettent d'établir une bonne estimation de la vitesse d'éxécution de cet algorithme sur un minicalculateur de type SEMS SOLAR.

II-8 RECAPITULATIF

Dans la pratique la méthode récursive d'estimation des paramètres d'un mélange gaussien se résume dans l'éxécution des étapes de calcul suivantes:

1 étape : calcul des λ^+_i solution de l'équation (17) en utilisant l'algorithme rapide développé dans le paragraphe II-7

2 étape : calcul des éléments O_{ij} de la matrice de rotation [O] en utilisant les équations (18) et (20)

3 étape : calcul des directions propres U^+_i selon l'expression (5)

4 étape : calcul du vecteur moyenne $\overline{X}^+_i$ selon la récurrence (21)

5 étape : calcul des nouvelles estimations des probabilités a priori selon la récurrence (22)

On voit que c'est la première étape qui présente les plus grandes difficultés de calcul ; l'algorithme rapide de calcul des λ permet de surmonter cette difficulté
.

II-9 SIMULATIONS NUMERIQUES ET PERFORMANCES

II-9-1 METHODE

Afin de valider la méthode et les performances de l'algorithme proposé, plusieurs simulations ont été effectuées, selon la procédure suivante:

Etape n°1 génération artificielle d'un ensemble E de S échantillons appartenant à un mélange gaussien caractérisé par les paramètres suivants:

- dimension de l'espace d
- nombre de composantes C
- caractéristiques de w_K $(K=1,C)$
  - .. moyenne $\overline{X}^K$
  - .. directions propres U^K_j $(j=1,d)$
  - .. valeurs propres λ^K_j
  - .. probabilité a priori P^K

Etape n°2 extraction à partir de l'ensemble E d'un sous ensemble e de s échantillons par tirage aléatoire.

Etape n°3 calcul des paramètres caractéristiques estimés du MG à partir du sous ensemble e . Ce calcul se fait en utilisant la méthode graphique présentée dans le chapitre I. Ces estimations constituent les valeurs initiales de l'algorithme récurrent avec les notations suivantes:

- pour chaque composante w_K $(K=1,C)$
  - .. vecteur moyenne $\overline{X}^K(s)$ $(j=1,d)$
  - .. directions propres $U^K_j(s)$
  - .. valeurs propres $\lambda^K_j(s)$
  - .. probabilité a priori $P^K(s)$

Etape n°4 calcul des paramètres caractéristiques estimés du MG en intégrant dans la base d'apprentissage e un sous-ensemble Ω de θ échantillons.

Ω E

$s \leq \theta \leq S$

Ce calcul se fait par recurrence à partir des valeurs initiales fournies par l'étape n°3

.Caractéristiques de la composante w_K après intégration de θ échantillons:

.. vecteur moyenne $X^K(\theta)$

.. directions propres $U^K{}_J(\theta)$

.. valeurs propres $\backslash^K{}_J(\theta)$

.. probabilité à priori $P^K(\theta)$

Etape n°5 comparaison des estimations courantes à l'étape n°4 avec les valeurs réelles associées au MG donné à l'étape n°1

Etape n°6 recommencer les étapes 2,3,4 et 5 pour différents sous ensembles e tirés de E.

Etape n°7 évaluation des résultats obtenus

II-9-2 SIMULATION 2D

Une première simulation a été effectuée pour d=2 à partir d'un ensemble E contenant S=400 échantillons , appartenant au MG constitué de trois composantes (C=3) dont les caractéristiques sont résumés dans le tableau-1.

	composante 1	composante 2	composante 3
vecteur moyenne	$\bar{X} = \begin{vmatrix} 250 \\ 200 \end{vmatrix}$	$\bar{X} = \begin{vmatrix} 140 \\ 180 \end{vmatrix}$	$\bar{X} = \begin{vmatrix} 200 \\ 100 \end{vmatrix}$
directions propres (radians)	U1= 0.25	U2= 0.5	U3= 0.64
valeurs propres	$\lambda_1=11$ $\lambda_2=7$	$\lambda_1=10$ $\lambda_2=6$	$\lambda_1=9$ $\lambda_2=6$
probabilité a priori	P1= 0.31	P2= 0.38	P3=0.31

tableau-1: caractéristiques du MG

Le tableau-2 fournit les résultats obtenus lorsque l'on fait fonctionner la récurrence à partir de divers sous-ensembles e , notées e_1, e_2, e_3, e_4 comportant respectivement $s_1=200$, $s_2=250$, $s_3=300$ et $s_4=400$ échantillons . La récurrence se déroule pour allant de $s_i (i=1,3)$ à S=400 et le tableau-2 fournit les valeurs finales estimées (θ =S=400).

Remarque : l'ensemble e_4 est égal à E et s_4 à S . Dans ces conditions il n'est pas nécessaire de faire fonctionner la récurrence et les paramètres estimés sont ceux fournis par la méthode statique (étape 3 seulement).

En complément des résultats consignés dans le tableau-2, les courbes de la figure 4 illustrent la dynamique d'évolution des estimations en fonction de θ.

II-9-3 SIMULATION 3D

La même procédure a été utilisée sur un exemple tridimensionnel . Les tableaux 3 et 4 représentent respectivement les caractéristiques du mélange gaussien et les résultats obtenus. (Le tableau 4 est analogue au tableau 2)

II-9-4 ANALYSE DES RESULTATS OBTENUS

Les écarts relatifs des estimations finales par rapport aux paramètres réels du MG sont évalués en comparant les tableaux 1 et 2 (exemple 2D) ou 3 et 4 (exemple 3D) . Ces écarts sont évalués pour des récurrence partant respectivement de e_1,e_2, e_3 et e_4=E.

la comparaison des résultats consignés dans le tableau 2-d (4-d pour le cas 3D) avec ceux des tableaux 2-a,2-b et 2-c (4-a,4-b et 4-c pour le cas 3D) permet d'évaluer la différrence de performance entre les méthodes statique et récursive.

Il apparait alors que la récurrence fournit des résultats comparable à ceux de la méthode statique.

On peut difficilement décrire cette analyse de performance de manière analytique . En effet , les évaluations dépendent de la structure du MG et de l'arrivée des échantillons .

Toutefois , il apparaît clairement que la recurrence permet d'obtenir des résultats d'une qualité équivalente à ceux obtenus par une méthode statique , avec l'avantage supplémentaire de pouvoir démarrer les traitements numériques à partir d'un ensemble relativement incomplet d'échantillons . Ce dernier aspect est très important car il permet une économie importante de mémoire lors de l'implantation sur calculateur.

Il importe de noter que le temps d'itération est indépendant du nombre d'échantillons déja traités , cette méthode est particulièrement bien adaptée à une implantation temps réel . A titre indicatif , pour l'exemple considéré , implanté sur mini calculateur SEMS SOLAR , le temps d'itération est égal à 0.076 secondes.

II-10 CONCLUSION

La méthode proposée , en complément à d'autres méthodes, permet d'aborder le problème de l'identification des MG sous un angle différent.

L'intérêt de cette démarche peut se situer à deux niveaux:

- niveau méthodologique
  - . possibilité de partir d'ensembles d'échantillons relativement incomplets
- niveau implantation informatique
  - . calcul d'itération rapide
  - . mobilisation mémoire faible

Tous ces éléments rendent la méthode bien adaptée à des mises en oeuvre temps réel pour des applications de type: identification des systèmes évolutifs ou prise de décision en ligne.

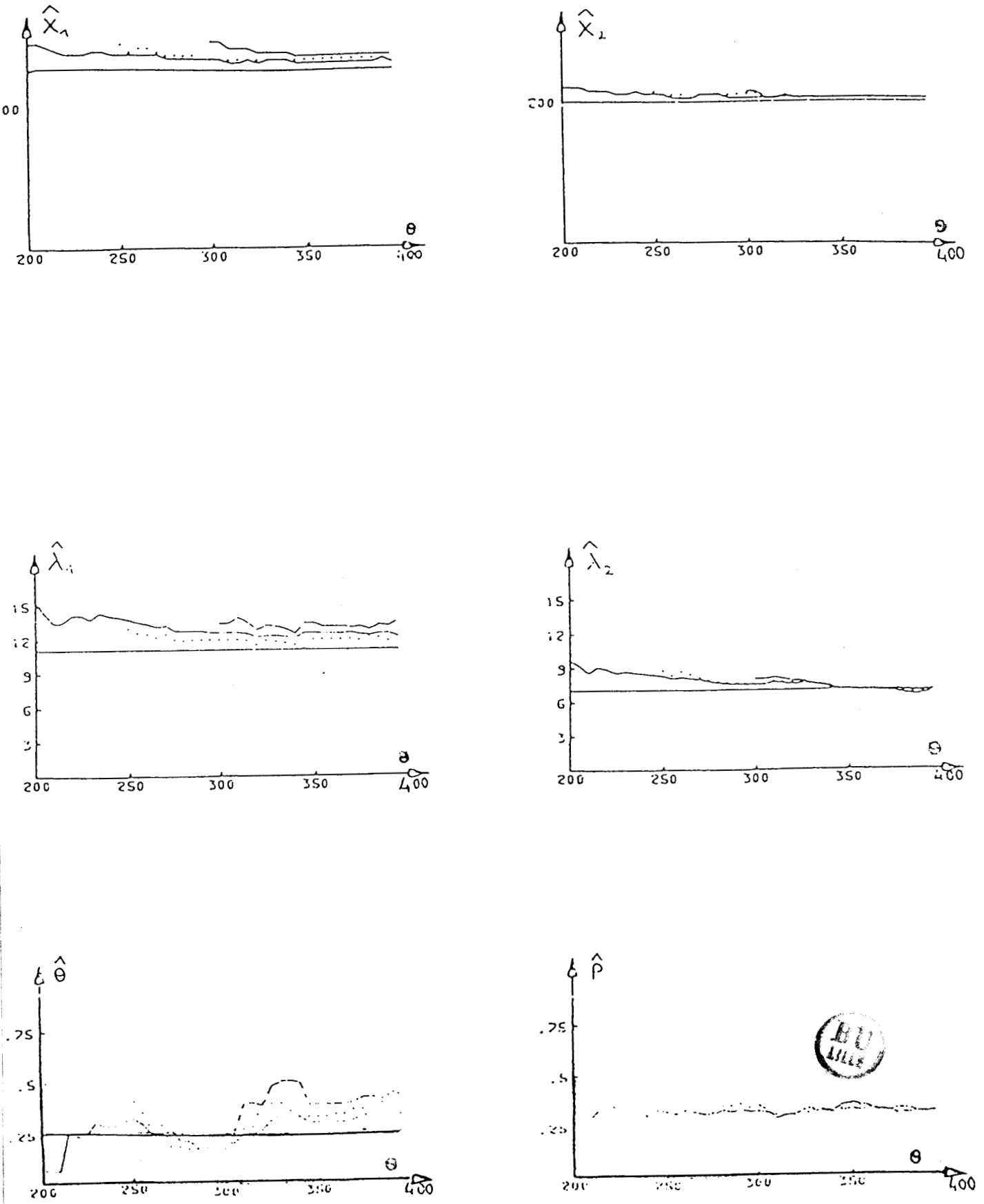
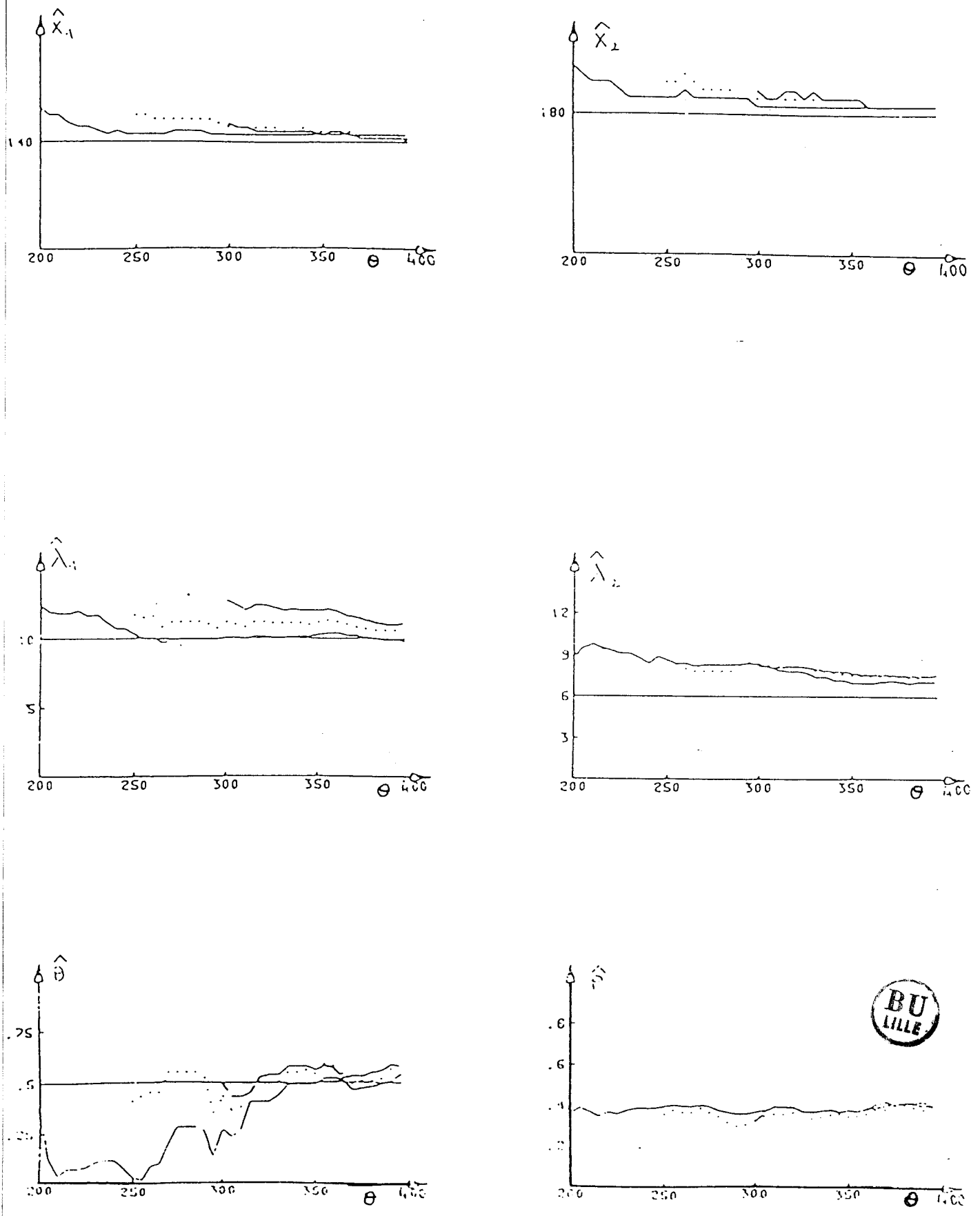


fig.4a évolution des différents paramètres de la composante 1
en fonction de l'arrivée des observations



BU LILLE

fig.4b évolution des différents paramètres de la composante 2 en fonction de l'arrivée des observations

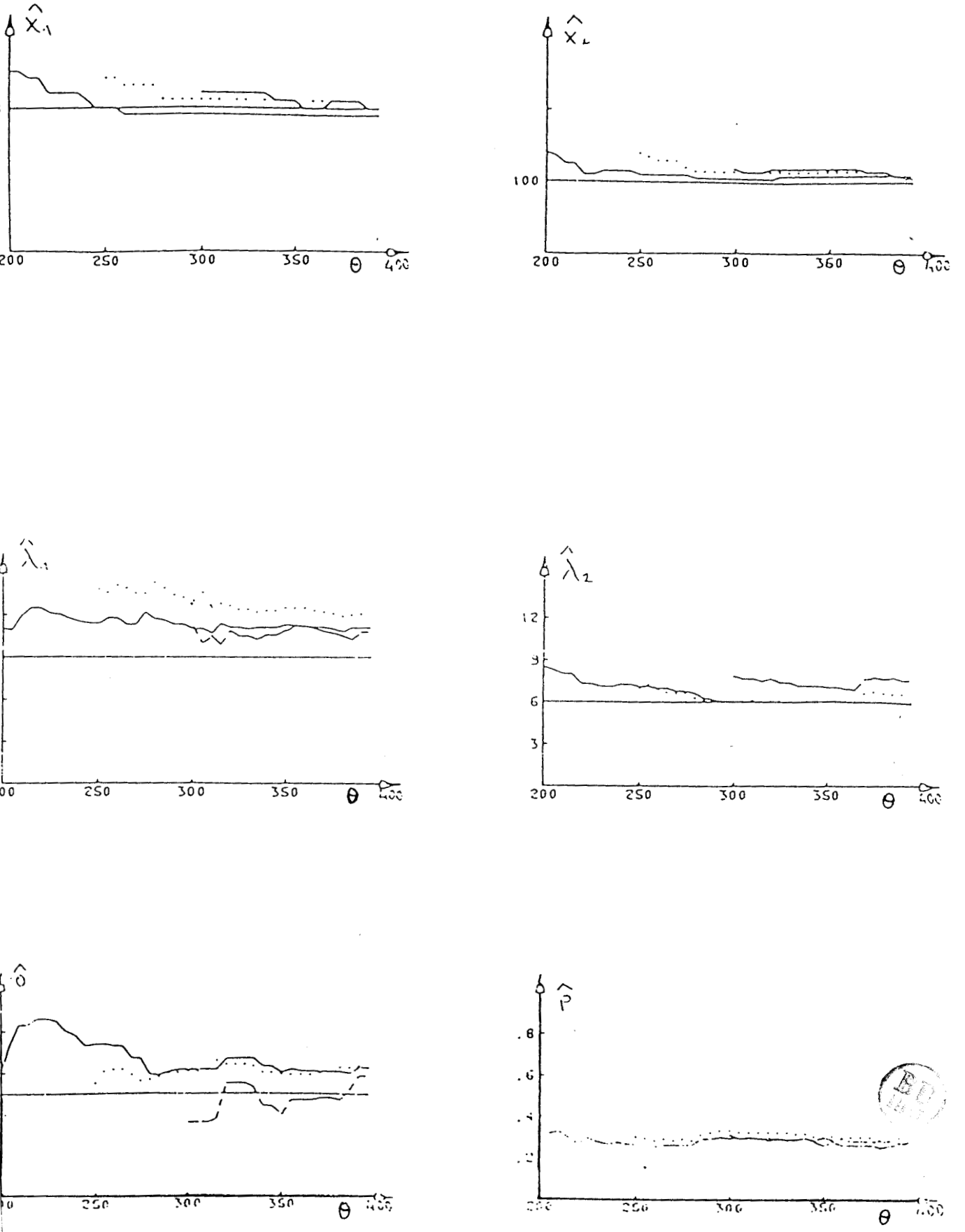


fig.4c évolution des différents paramètres de la composante 3 en fonction de l'arrivée des observations

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité à priori
COMPOSANTE 1	$\hat{\bar{X}}=\begin{bmatrix}252.7\\201.3\end{bmatrix}$	$\hat{\theta}=0.33$	$\hat{\lambda}_1=12.1$ $\hat{\lambda}_2=7.1$	$\hat{P}=0.32$
COMPOSANTE 2	$\hat{\bar{X}}=\begin{bmatrix}141.8\\181\end{bmatrix}$	$\hat{\theta}=0.54$	$\hat{\lambda}_1=10$ $\hat{\lambda}_2=7.6$	$\hat{P}=0.40$
COMPOSANTE 3	$\hat{\bar{X}}=\begin{bmatrix}199\\101.7\end{bmatrix}$	$\hat{\theta}=0.64$	$\hat{\lambda}_1=11.1$ $\hat{\lambda}_2=5.9$	$\hat{P}=0.28$

Tableau-2-a estimation des paramètres pour $s_1=200$ et $\sigma=400$

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité à priori
COMPOSANTE 1	$\hat{\bar{X}}=\begin{bmatrix}252.9\\201.4\end{bmatrix}$	$\hat{\theta}=0.37$	$\hat{\lambda}_1=11.1$ $\hat{\lambda}_2=7$	$\hat{P}=0.32$
COMPOSANTE 2	$\hat{\bar{X}}=\begin{bmatrix}142.8\\181\end{bmatrix}$	$\hat{\theta}=0.57$	$\hat{\lambda}_1=10.5$ $\hat{\lambda}_2=7.4$	$\hat{P}=0.38$
COMPOSANTE 3	$\hat{\bar{X}}=\begin{bmatrix}200.6\\102.3\end{bmatrix}$	$\hat{\theta}=0.64$	$\hat{\lambda}_1=12.1$ $\hat{\lambda}_2=6.6$	$\hat{P}=0.30$

Tableau-2-b estimation des paramètres pour $s_2=250$ et $\sigma=400$

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité à priori
OMPOSANTE 1	$\hat{\bar{X}}=\begin{bmatrix}254\\201.5\end{bmatrix}$	$\hat{\theta}=0.16$	$\hat{\lambda}_1=12.9$ $\hat{\lambda}_2=6.9$	$\hat{P}=0.37$
OMPOSANTE 2	$\hat{\bar{X}}=\begin{bmatrix}142.3\\181.2\end{bmatrix}$	$\hat{\theta}=0.59$	$\hat{\lambda}_1=11.1$ $\hat{\lambda}_2=7.1$	$\hat{P}=0.41$
OMPOSANTE 3	$\hat{\bar{X}}=\begin{bmatrix}204.4\\102.1\end{bmatrix}$	$\hat{\theta}=0.59$	$\hat{\lambda}_1=10.8$ $\hat{\lambda}_2=7.6$	$\hat{P}=0.6$

Tableau-2-c estimation des paramètres pour $s_3=300$ et $\sigma=400$

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité à priori
OMPOSANTE 1	$\hat{\bar{X}}=\begin{bmatrix}258\\205\end{bmatrix}$	$\hat{\theta}=0.16$	$\hat{\lambda}_1=16.9$ $\hat{\lambda}_2=9.2$	$\hat{P}=0.37$
OMPOSANTE 2	$\hat{\bar{X}}=\begin{bmatrix}148\\183\end{bmatrix}$	$\hat{\theta}=0.57$	$\hat{\lambda}_1=15.3$ $\hat{\lambda}_2=9.9$	$\hat{P}=0.32$
OMPOSANTE 3	$\hat{\bar{X}}=\begin{bmatrix}204\\105\end{bmatrix}$	$\hat{\theta}=0.47$	$\hat{\lambda}_1=17.9$ $\hat{\lambda}_2=5.5$	$\hat{P}=0.29$

Tableau-2-d estimation des paramètres pour $S=\sigma=400$

BU LILLE

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité à priori
OMPOSANTE 1	$\bar{X}=\begin{bmatrix}100\\100\\300\end{bmatrix}$	$\theta_1=\begin{bmatrix}1.02\\0.64\\1.32\end{bmatrix}$ $\theta_2=\begin{bmatrix}2.49\\1.96\\1.12\end{bmatrix}$ $\theta_3=\begin{bmatrix}1.31\\2.01\\0.53\end{bmatrix}$	$\lambda_1=8$ $\lambda_2=6$ $\lambda_3=4$	P=0.273
OMPOSANTE 2	$\bar{X}=\begin{bmatrix}300\\100\\100\end{bmatrix}$	$\theta_1=\begin{bmatrix}0.91\\0.87\\1.13\end{bmatrix}$ $\theta_2=\begin{bmatrix}2.27\\1.63\\0.73\end{bmatrix}$ $\theta_3=\begin{bmatrix}1.13\\2.40\\1.04\end{bmatrix}$	$\lambda_1=9$ $\lambda_2=6$ $\lambda_3=5$	P=0.363
OMPOSANTE 3	$\bar{X}=\begin{bmatrix}100\\300\\100\end{bmatrix}$	$\theta_1=\begin{bmatrix}1.43\\0.59\\1.05\end{bmatrix}$ $\theta_2=\begin{bmatrix}2.5\\1.71\\1.05\end{bmatrix}$ $\theta_3=\begin{bmatrix}1.05\\2.08\\0.79\end{bmatrix}$	$\lambda_1=9$ $\lambda_2=7$ $\lambda_3=5$	P=0363

Tableau-3 paramètres réels du MG

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité à priori
OMPOSANTE 1	$\hat{\bar{X}}=\begin{bmatrix}101.2\\103.4\\301.3\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}1.7\\0.2\\1.5\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}0.3\\1.4\\1.4\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.8\\1.7\\0.2\end{bmatrix}$	$\hat{\lambda}_1=8.6$ $\hat{\lambda}_2=7.5$ $\lambda_3=5.5$	$\hat{P}=0.288$
MPOSANTE 2	$\hat{\bar{X}}=\begin{bmatrix}304.1\\106.8\\101.8\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}1.2\\0.43\\1.51\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}2.7\\1.13\\1.71\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.7\\1.63\\0.14\end{bmatrix}$	$\hat{\lambda}_1=10.1$ $\hat{\lambda}_2=7.8$ $\hat{\lambda}_3=6.1$	$\hat{P}=0.329$
MPOSANTE 3	$\hat{\bar{X}}=\begin{bmatrix}100.4\\304.6\\102.7\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}0.59\\1.02\\1.37\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}2.10\\0.55\\1.71\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.81\\1.55\\0.24\end{bmatrix}$	$\hat{\lambda}_1=9.2$ $\hat{\lambda}_2=7.2$ $\hat{\lambda}_3=7.9$	$\hat{P}=0.381$

Tableau-4-a estimations des paramètres pour $s_1=200$ et $\sigma=400$

BU LILLE

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité à priori
MPOSANTE 1	$\hat{\bar{X}}=\begin{bmatrix}101.4\\102.9\\301.6\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}1.63\\0.19\\1.39\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}0.17\\1.53\\1.4\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.73\\1.76\\0.25\end{bmatrix}$	$\hat{\lambda}_1=9$ $\hat{\lambda}_2=8.3$ $\hat{\lambda}_3=5.6$	$\hat{P}=0.320$
MPOSANTE 2	$\hat{\bar{X}}=\begin{bmatrix}304.1\\106.3\\102.4\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}1.08\\0.54\\1.33\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}0.49\\2.03\\1.7\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.55\\1.84\\0.27\end{bmatrix}$	$\hat{\lambda}_1=11.6$ $\hat{\lambda}_2=7.6$ $\hat{\lambda}_3=6.1$	$\hat{P}=0.313$
MPOSANTE 3	$\hat{\bar{X}}=\begin{bmatrix}101.2\\304.6\\102.7\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}0.71\\0.89\\1.35\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}2.17\\0.66\\1.81\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.92\\1.49\\0.35\end{bmatrix}$	$\hat{\lambda}_1=9.9$ $\hat{\lambda}_2=7.3$ $\hat{\lambda}_3=6.4$	$\hat{P}=0.367$

Tableau-4-b estimations des paramètres pour $s_2=250$ et $\sigma=400$

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité à priori
MPOSANTE 1	$\hat{\bar{X}}=\begin{bmatrix}101.2\\102.6\\302\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}0.95\\0.65\\1.37\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}2.48\\0.91\\1.62\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.72\\1.69\\0.19\end{bmatrix}$	$\hat{\lambda}_1=9.6$ $\hat{\lambda}_2=8.2$ $\hat{\lambda}_3=5.6$	$\hat{P}=0.364$
POSANTE 2	$\hat{\bar{X}}=\begin{bmatrix}303\\106.2\\103.5\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}0.99\\0.61\\1.37\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}0.77\\2.21\\1.19\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.99\\1.50\\0.42\end{bmatrix}$	$\hat{\lambda}_1=12.2$ $\hat{\lambda}_2=8.3$ $\hat{\lambda}_3=6.4$	$\hat{P}=0.277$
POSANTE 3	$\hat{\bar{X}}=\begin{bmatrix}101.8\\304.9\\103.8\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}0.77\\0.99\\1.12\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}0.86\\2.40\\1.72\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.80\\1.97\\0.47\end{bmatrix}$	$\hat{\lambda}_1=10.5$ $\hat{\lambda}_2=7.7$ $\hat{\lambda}_3=6.4$	$\hat{P}=0.357$

Tableau-4-c estimations des paramètres pour $s_1=300$ et $\sigma=400$

BU LILLE

	vecteurs moyennes	directions propres (radians)	écarts types	probabilité a priori
OSANTE 1	$\hat{\bar{X}}=\begin{bmatrix}105.5\\108.0\\305.6\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}2.06\\0.52\\1.40\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}0.49\\1.09\\1.47\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.57\\1.76\\0.19\end{bmatrix}$	$\hat{\lambda}_1=12.4$ $\hat{\lambda}_2=10.2$ $\hat{\lambda}_3=6$	$\hat{P}=0.314$
OSANTE 2	$\hat{\bar{X}}=\begin{bmatrix}307.5\\110.7\\107.7\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}1.24\\0.33\\1.57\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}0.35\\1.90\\1.46\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.67\\1.52\\0.11\end{bmatrix}$	$\hat{\lambda}_1=11.7$ $\hat{\lambda}_2=10.0$ $\hat{\lambda}_3=7.1$	$\hat{P}=0.351$
OSANTE 3	$\hat{\bar{X}}=\begin{bmatrix}107.0\\306.8\\108.7\end{bmatrix}$	$\hat{\theta}_1=\begin{bmatrix}0.92\\0.73\\1.27\end{bmatrix}$ $\hat{\theta}_2=\begin{bmatrix}0.64\\2.18\\1.75\end{bmatrix}$ $\hat{\theta}_3=\begin{bmatrix}1.60\\1.92\\0.35\end{bmatrix}$	$\hat{\lambda}_1=12.3$ $\hat{\lambda}_2=9.3$ $\hat{\lambda}_3=7.4$	$\hat{P}=0.333$

Table-4.d estimation of parameters for $S=\sigma=400$

BU LILLE

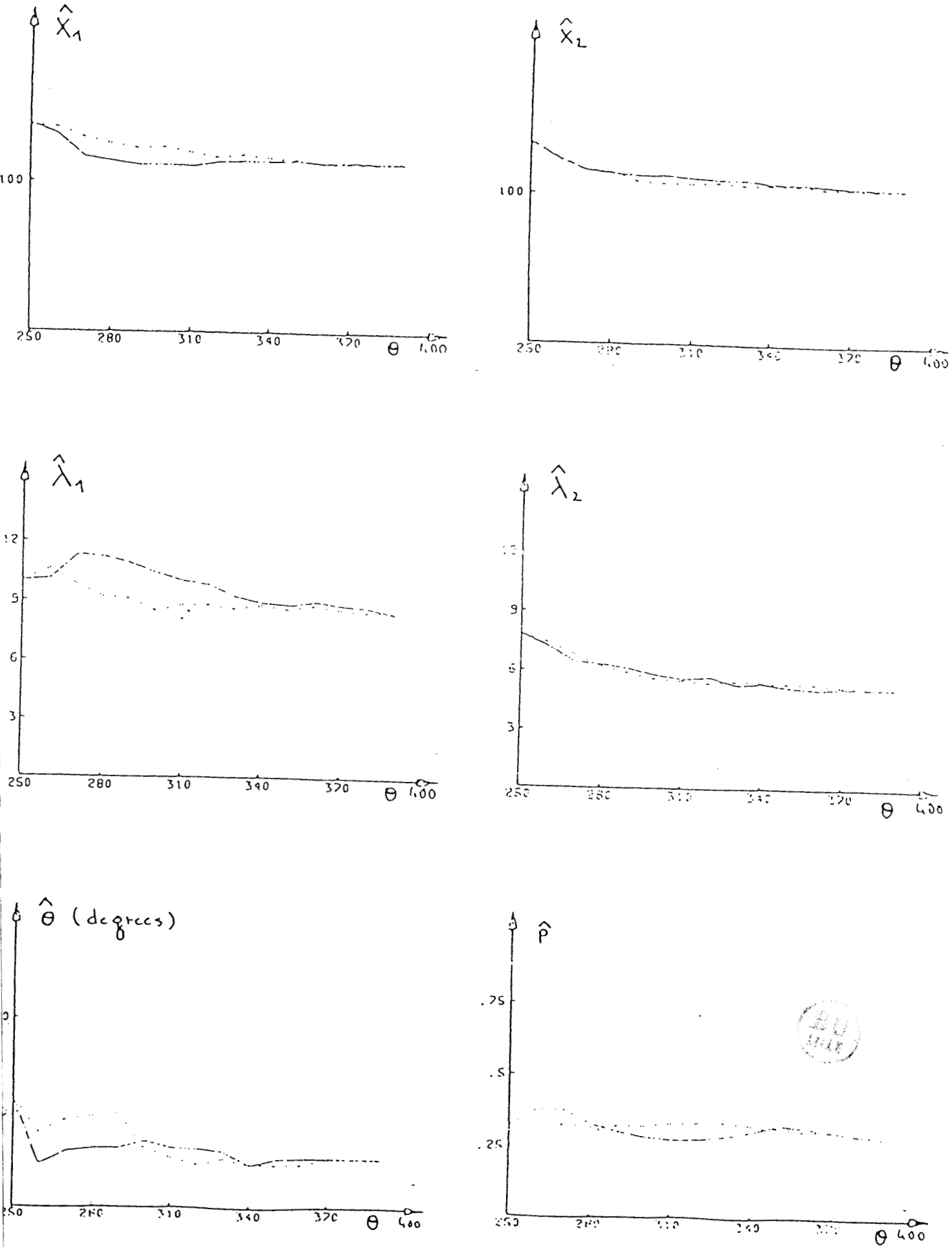


fig.5a intégration des observations arrivant dans deux ordres chronologiques différents (composante 1)

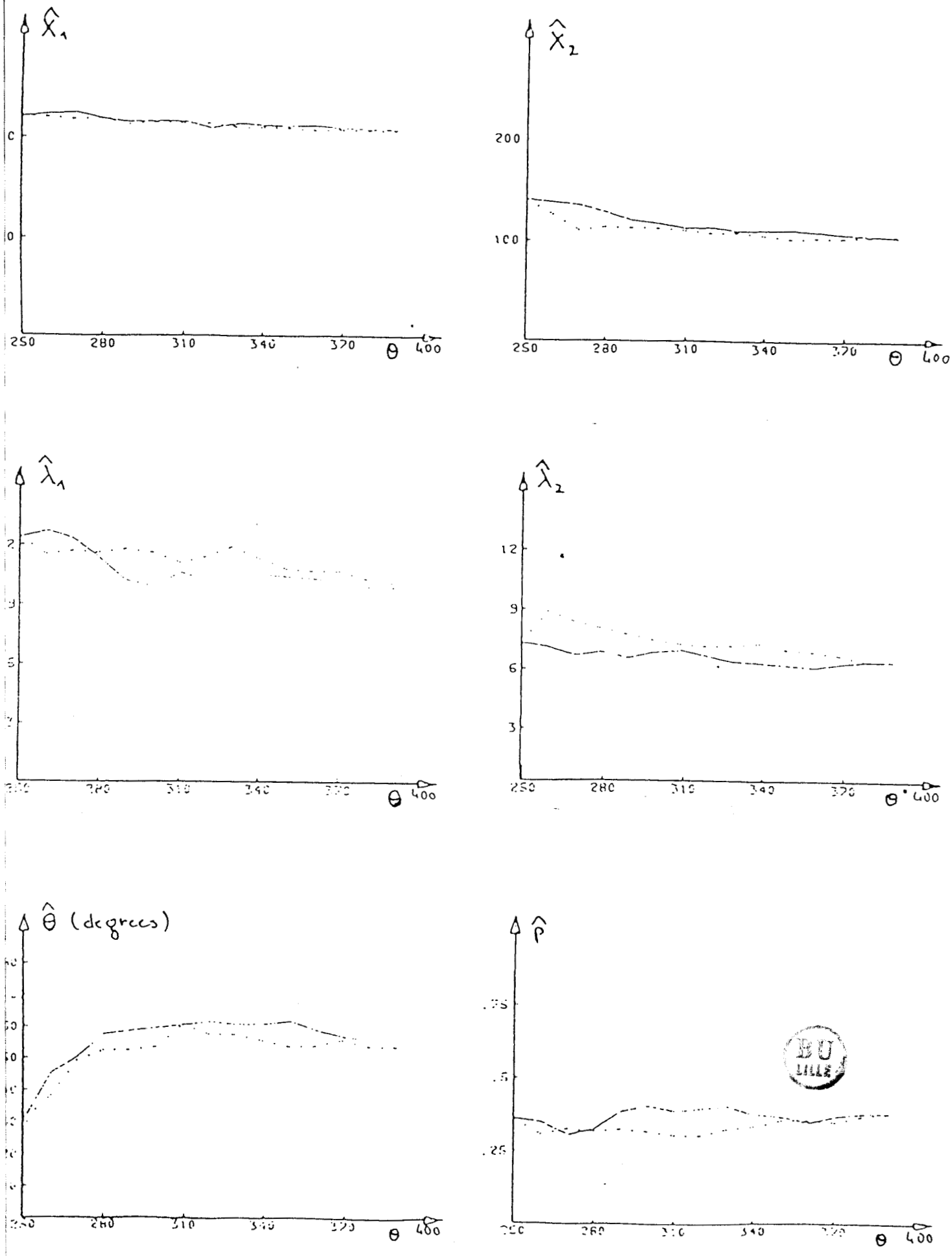
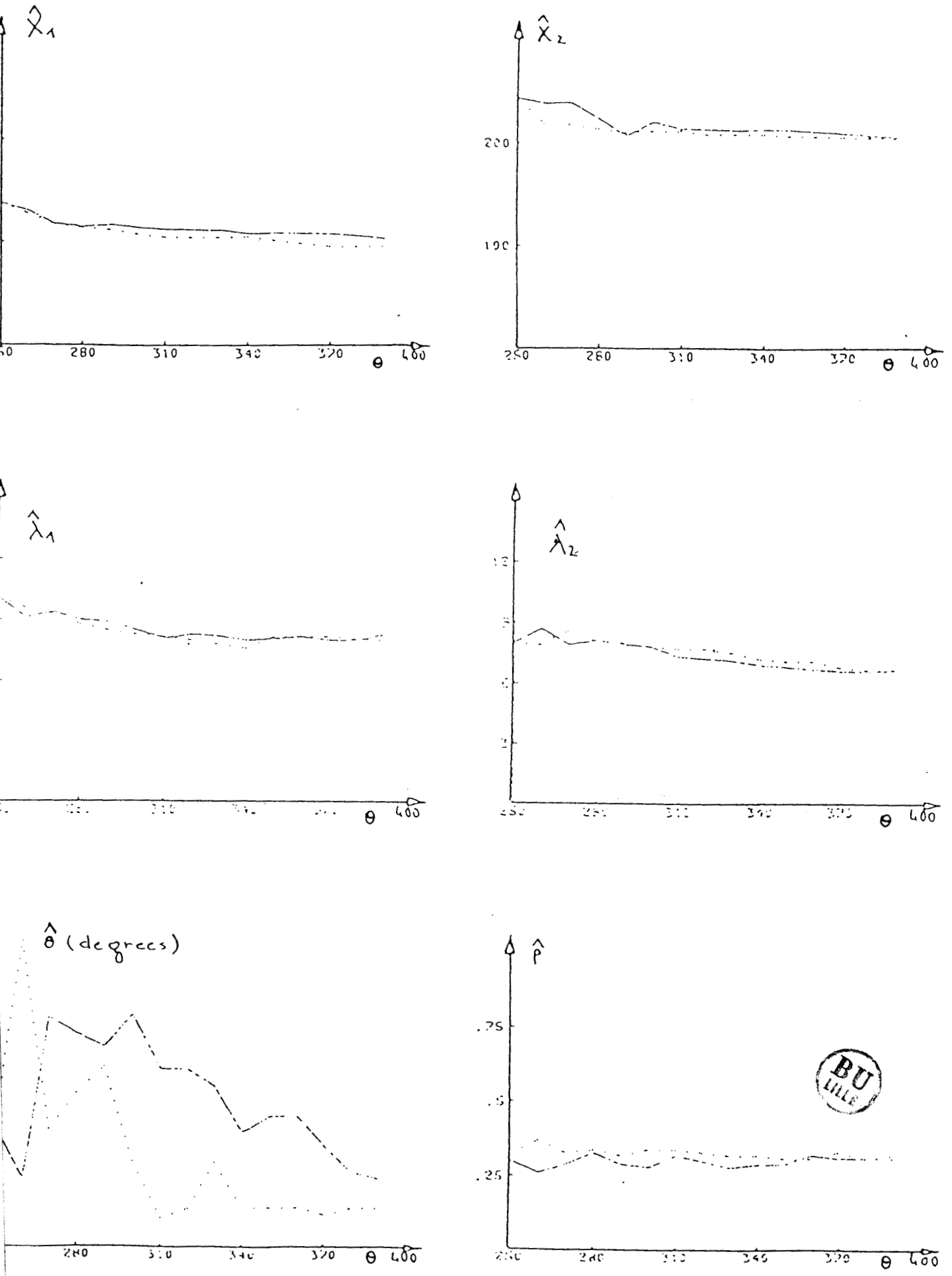


fig.5b intégration des observations arrivant dans deux ordres chronologiques différents (composante 2)

BU LILLE



BU LILLE

fig.5c intégration des observations arrivant dans deux ordres chronologiques différents (composante 3)

CHAPITRE III

FUSIONNEMENT DE DEUX COMPOSANTES

CHAPITRE III

FUSIONNEMENT DE DEUX COMPOSANTES

III-1 INTRODUCTION 1

III-2 DISTANCE MINIMALE ENTRE DEUX COMPOSANTES 2

III-3 DISPERSION DES OBSERVATIONS 3

III-4 ENTROPIE 5

III-5 DETECTION DES MODES 9

III-6 DETERMINATION DES POINTS M_{IK} ET M_{KI} 15

III-7 ALGORITHME RAPIDE DE RECHERCHE DES POINTS M_{IK} ET M_{KI} 21

III-8 EXEMPLES D'APPLICATION 25

- III-8-1 INTRODUCTION 25
- III-8-2 FONCTION BIMODALE 27
- III-8-3 UTILISATION DES PROBABILITES A POSTERIORI 35
- III-8-4 LES PARAMETRES DE LA COMPOSANTE SOMME 39

III-9 CONCLUSION 40

BIBLIOGRAPHIES 41

III-1 INTRODUCTION

Dans le chapitre I on a vu que certaines composantes peuvent fusionner dans le cas où la base d'apprentissage est incomplète.La figure 1 illustre cette situation.

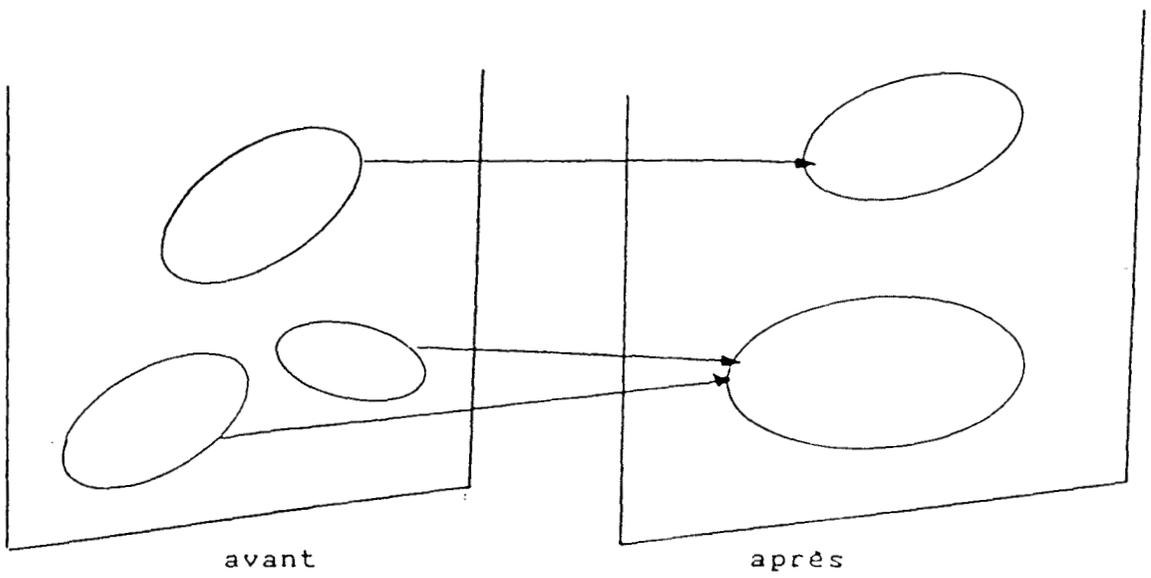


fig.1 fusionnement de deux composantes.

Il est important de définir des critères qui mesurent la séparabilité des composantes où dans le cas contraire le degré de recouvrement de ces dernières.

De tels critères permettent en effet de prendre la décision de fusionner ou non deux composantes.

Par ailleurs,les critères employés , pour être efficaces, doivent tenir compte de la géométrie des composantes.

III-2 DISTANCE MINIMALE ENTRE DEUX COMPOSANTES

Soient deux composantes w_I et w_K, un critère possible de séparabilité consiste à définir une distance euclidienne minimale séparant ces deux composantes de la manière suivante [GRE,84]:

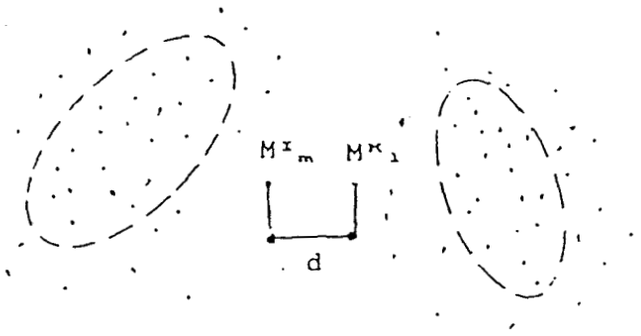


fig.2 mesure de la distance entre deux composantes

$$d = \min_{m,l} d(M^I_m, M^K_l) \qquad \begin{matrix} m=1,N_I \\ l=1,N_K \end{matrix}$$

Ce critère est rarement utilisé, car il ne tient pas compte de la géométrie des composantes.

III-3 DISPERSION DES OBSERVATIONS

Intuitivement, pour deux composantes w_I et w_K un critère de la forme $J=f(\overline{X}^I,\overline{X}^K,\Sigma_I,\Sigma_K)$ intègre mieux la géomètrie des composantes w_I et w_K.

Dans ce critère les vecteurs moyennes $\overline{X}^I$ et $\overline{X}^K$ fournissent les positions des centres de masse et les matrices de covariance Σ_I et Σ_K traduisent la dispersion des observations à l'intèrieur des composantes.

La forme quadratique suivante :

$$J_1=tr(S_w^{-1}.S_b) \qquad (1)$$

dans laquelle S_w et S_b représentent les expressions ci-dessous

$$S_w = P^I\Sigma_I + P^K\Sigma_K \qquad (2)$$

$$S_b = P^I(\overline{X}^I - \overline{X})(\overline{X}^I - \overline{X})^T + P^K(\overline{X}^K - \overline{X})(\overline{X}^K - \overline{X})^T \qquad (3)$$

$$\overline{X} = P^I\overline{X}^I + P^K\overline{X}^K \quad \text{centre de masse de } \overline{X}^I \text{ et } \overline{X}^K \qquad (4)$$

est une traduction du critère J à l'aide d'une fonction scalaire [FUK,72].

Ce critère quadratique J_1 est le plus couramment utilisé pour mesurer la séparabilité des composantes.

Le suivi permanent de ce critère en fonction de l'arrivée de nouvelles observations nous permet de suivre l'évolution de la séparabilité des composantes dans l'espace d'observation.

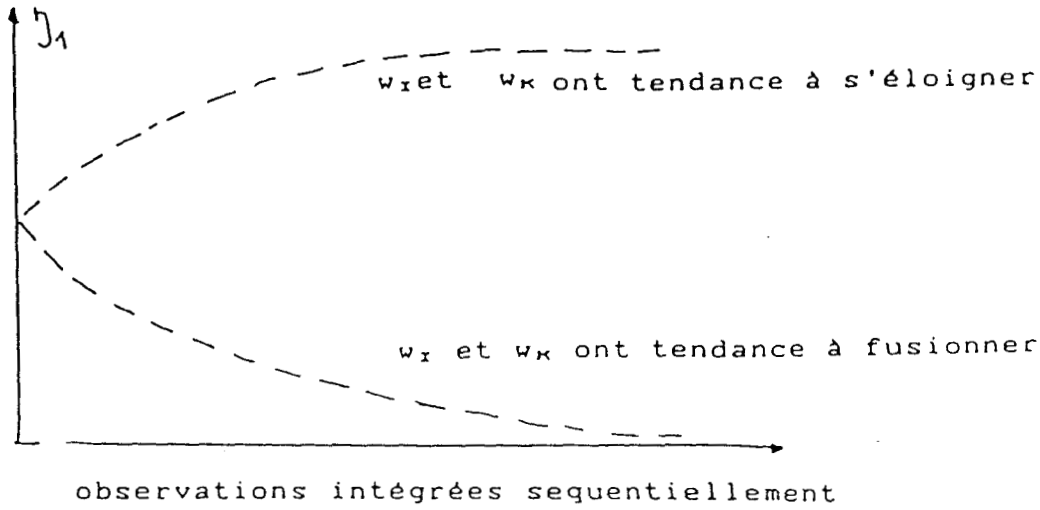


fig.3 évolution de la séparabilité des composantes

K.FUKUNAGA avec d'une part S.ANDO [FUK,77] et d'autre part R.D SHORT [FUK,78,a] [FUJK,78,b] a montré que le critère J_1 est équivalent à l'estimation des probabilités a posteriori $P(w_I/X)$ $P(w_K/X)$.

III-4 MESURE DE L'ENTROPIE

On a vu dans le paragraphe précédent que l'utilisation du critère $J=(\overline{X}^{I},\overline{X}^{K},\Sigma_{I},\Sigma_{K})$ qui intègre presque toutes les informations sur la géométrie des composantes est équivalent à l'estimation des probabilités à postériori $P(w_I/X)$ et $P(w_K/X)$.

Dans le cas où le mélange gaussien considéré contient uniquement deux composantes on a :

$$P(w_I/X) + P(w_K/X) = 1 \qquad (5)$$

Pour une une observation X à classer , plus la différence entre les probabilités a posteriori est grande plus l'incertitude sur son classement est faible.

Par extension deux composantes sont distinctes si l'incertitude de classement des observations ayant servies à les construire est faible.

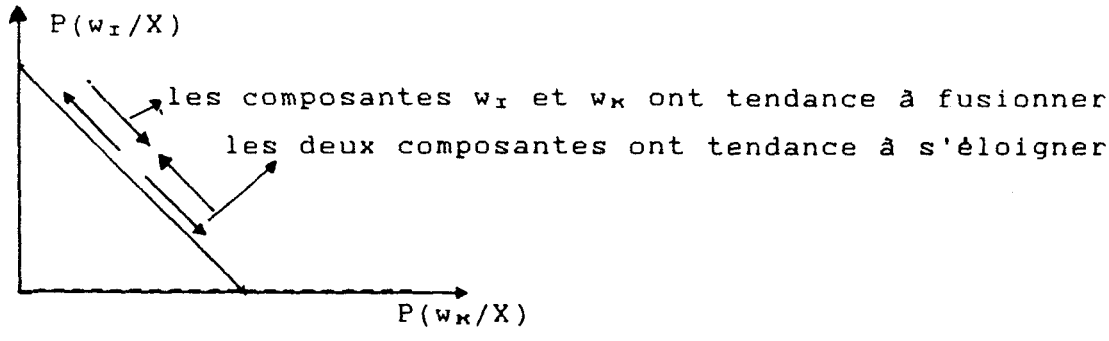


fig.4 les probabilités a posteriori comme critère de séparabilité

Si les valeurs des probabilités a posteriori ont tendance à se rapprocher , ceci traduit un recouvrement de plus en plus important des deux composantes.

En introduisant la notion d'entropie H définie par l'expression :

$$H=-P(w_I/X)\log(p(w_I/X)) -P(w_K/X)\log(P(w_K/X) \qquad (6)$$

On obtient une bonne mesure de séparabilité ou de recouvrement des composantes.[WAT,81]

Si les deux composantes w_I et w_K sont suffisamment distinctes l'entropie H met en évidence trois zones:

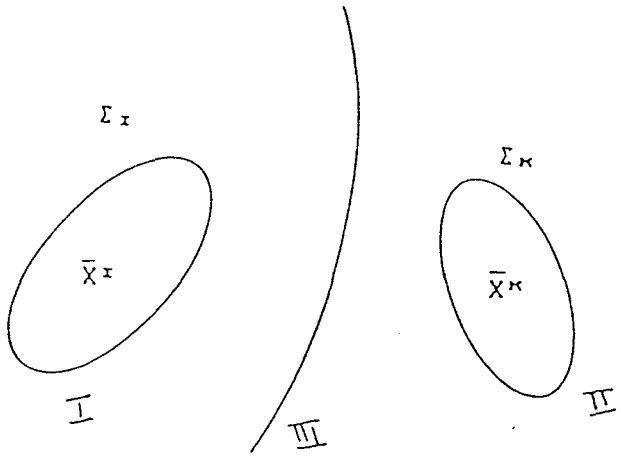


fig.5 séparation de l'espace d'observation en zones

L'entropie H est très faible dans les zones I et II , ce qui traduit l'absence d'incertitude de classement des observations dans ces zones.

Par contre l'entropie est très élevée dans la zone III , zone frontière . Comme les composantes sont distinctes très peu d'observations tombent dans cette partie de l'espace.

En suivant l'évolution des estimations de H ou des probabilités a posteriori uniquement en quelques points privilégiés , on peut avoir une idée précise sur la séparabilité des composantes.

Les points choisis sont: $\bar{X}^1, \bar{X}^K, M_{IK}, M_{KI}$. Ils sont représentés sur la figure 6

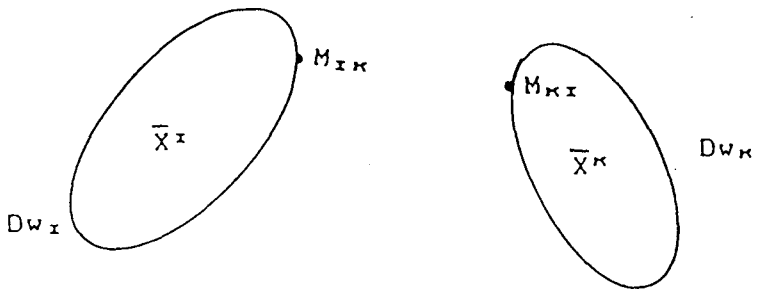


fig.6 présentation des points X^I, X^K, M_{IK}, M_{KI}

Les points M_{IK}, M_{KI} sont définis comme suit:

a)ils appartiennent respectivement aux frontières des domaines convexes Dw_I et Dw_K qui caractérisent les composantes w_I et w_K.

La frontière du domaine Dw_I est :

$$Dw_I = \{ X / (X - \bar{X}^I)^T \Sigma^{-1}_I (X - \bar{X}^I) = 1 \} \qquad (7)$$

De même pour Dw_K:

$$Dw_K = \{ X / (X - \bar{X}^K)^T \Sigma^{-1}_K (X - \bar{X}^K) = 1 \} \qquad (8)$$

$M_{IK} \in Dw_I$ et $M_{KI} \in Dw_K$

b)M_{IK} est le point le plus proche de w_K,au sens de la distance de MAHANALOBIS

Sa détermination peut être formulée de la manière suivante:

il s'agit de minimiser la distance

$$f_{KI}=(X - \bar{X}^K)^T\Sigma^{-1}_K(X - \bar{X}^K)$$

sous la contrainte (9)

$$(X - \bar{X}^I)^T\Sigma^{-1}_I(X - \bar{X}^I)=1$$

M_{KI} est définie de la même manière que M_{IK}

il s'agit de minimiser la distance

$$f_{IK}=(X - \bar{X}^I)^T\Sigma^{-1}_I(X - \bar{X}^I)$$

sous la contrainte (10)

$$(X - \bar{X}^K)^T\Sigma^{-1}_K(X - \bar{X}^K)=1$$

Si l'incertitude sur le classement des points tels que $(\bar{X}^I,M_{IK})$ et $(\bar{X}^K,M_{KI})$ qui caractérisent par définition respectivement les composantes w_I et w_K augmente alors le recouvrement des composantes devient important.

Il est en effet certain que l'influence d'une composante sur l'autre s'exerce en premier lieu sur les points tels que M_{IK} et M_{KI} et qu'elle y est maximale.

Un algorithme rapide de détermination des points M_{IK} et M_{KI} est développé dans le paragraphe 7 de ce chapitre

III-5 LES MODES DE LA FONCTION DE DENSITE CONJOINTE

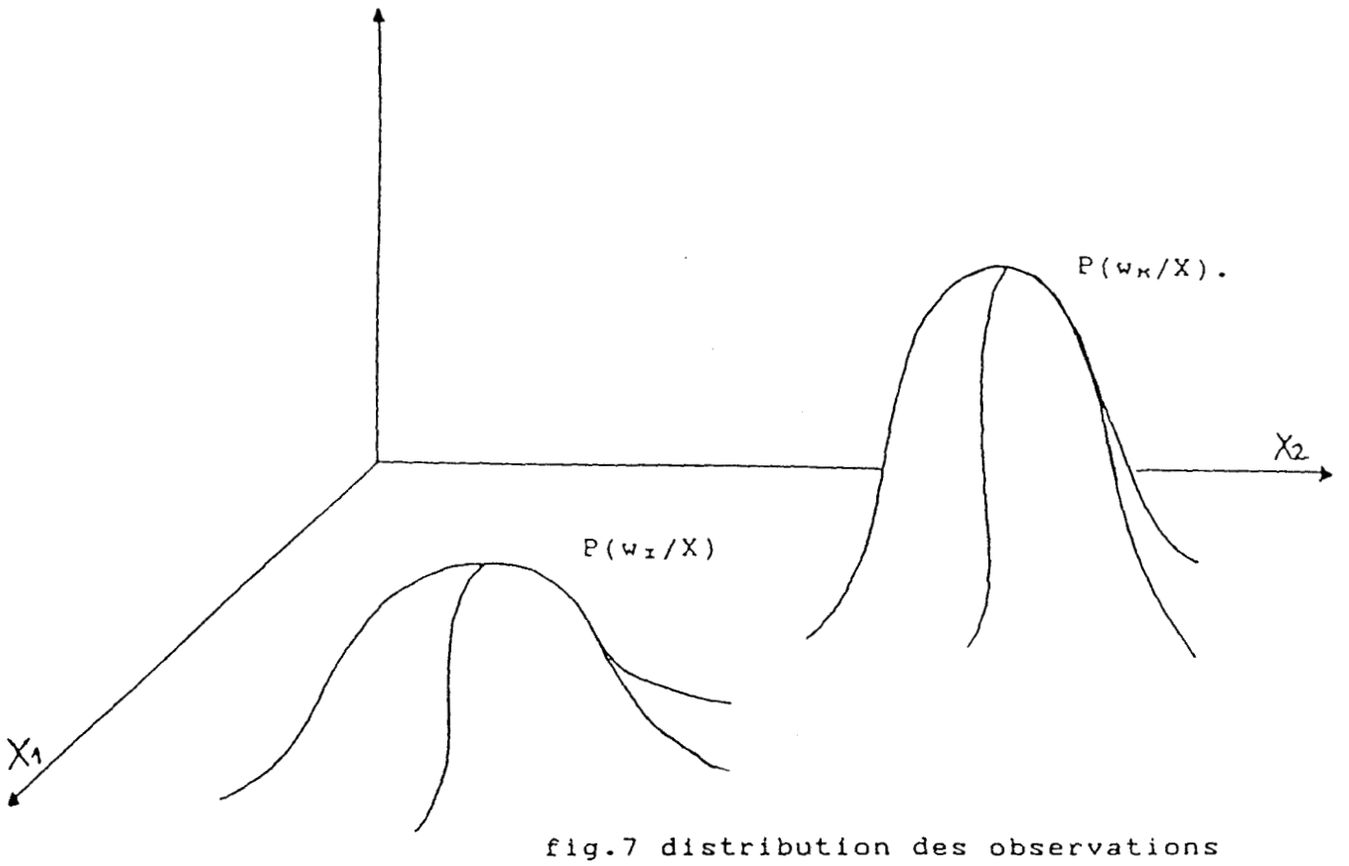


fig.7 distribution des observations

Les fonctions de densité de probabilité conditionnelle $P^I(X/w_I)$ sont des fonctions unimodales.

Dans le cas où les composantes w_I et w_K sont distinctes, la fonction de densité conjointe:

$$f_{IK}(X)=P^I.P^I(X/w_I) + P^K.P^K(X/w_K) \qquad (11)$$

est une fonction bimodale.[OEF,83]

Dans le cas contraire où w_I et w_K ont un fort taux de recouvrement ,là fonction $f_{IK}(X)$ devient une fonction unimodale.

Le suivi du nombre de modes est réalisé en estimant la fonction $f_{IK}(X)$ aux quatres points précédemment définis à savoir:$\bar{X}^I,\bar{X}^K,M_{IK},M_{KI}$

Les trois figures qui suivent illustrent alors les différentes possibilités qui peuvent se présenter pour deux composantes w_I et w_K.

Pour simplifier la représentation graphique on a choisi deux composantes telles que: $\bar{X}^I \neq \bar{X}^K$ et $\Sigma_I = \Sigma_K$

Premier cas:composantes distinctes

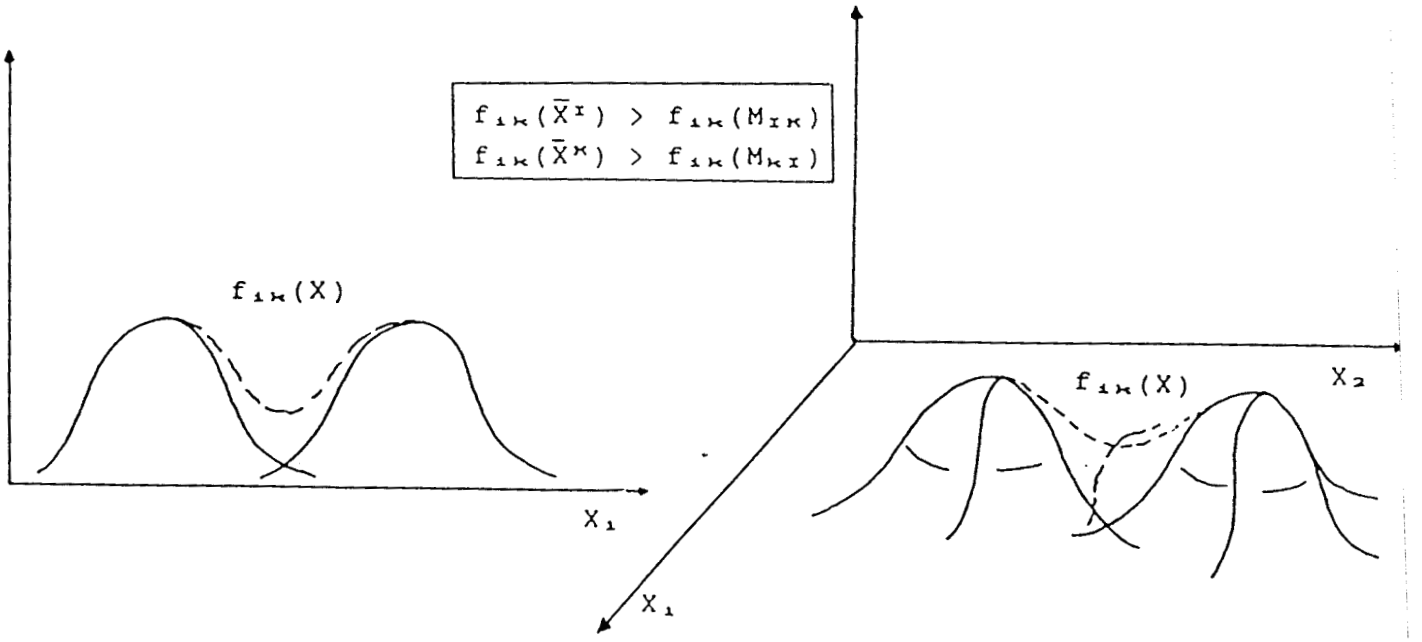


fig.8 composantes distinces pour les cas mono et bidimensionnel

On voit que: $f_{ik}(\bar{X}^I) > f_{ik}(M_{IK})$
$f_{ik}(\bar{X}^K) > f_{ik}(M_{KI})$

Ce qui traduit l'existence d'une vallée de la fonction de densité conjointe $f_{ik}(X)$ entre les deux centres de masse.

Deuxième cas:composantes à la limite du fusionnement

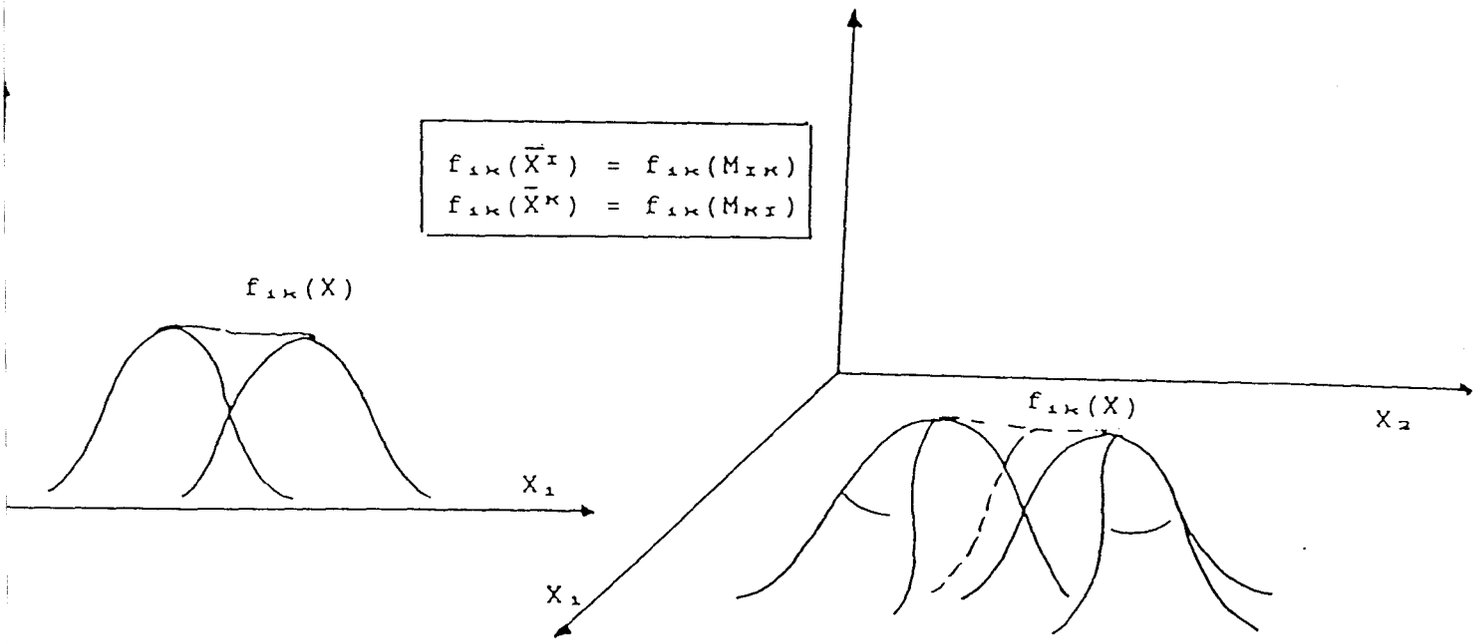


fig.9 composantes à la limite du fusionnement
pour les cas mono et bidimensionnel

$f_{ik}(\bar{X}^I) \cong f_{ik}(M_{IK})$ avec $M_{IK} = M_{KI}$
$f_{ik}(\bar{X}^K) \cong f_{ik}(M_{KI})$

Entre les deux centres de masse,$f_{ik}(X)$ présente un palier

Troisième cas: composantes en fusionnement

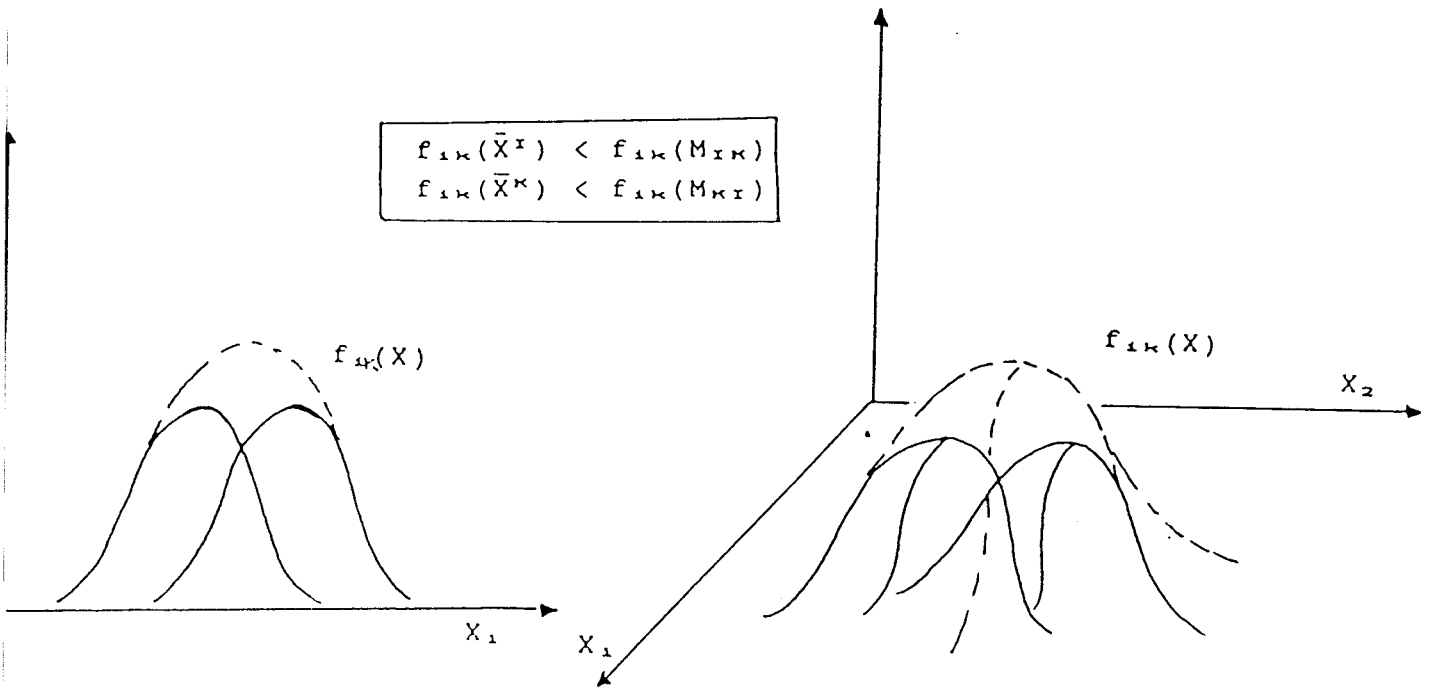


fig.10 composantes en fusionnement

Ainsi l'estimation de la fonction $f_{IK}(X)$ aux quatres points $\bar{X}^I, \bar{X}^K, M_{IK}, M_{KI}$ permet de mesurer le nombre de modes de la fonction de densité conjointe.

Dans le paragraphe 8 de ce chapitre, des exemples traitant le cas le plus général sont présentés: $\bar{X}^I \neq \bar{X}^K$ et $\Sigma_I \neq \Sigma_K$.

Dans le cas général, la comparaison de $f_{ik}(\overline{X}^I)$ à $f_{ik}(M_{IK})$ est fournie par les calculs suivants:

En utilisant l'équation (11) et en remplaçant les fonctions de densité de probabilité par leurs expressions on obtient

$$f_{ik}(\overline{X}^I)=\frac{P^I}{(2\pi)^{d/2}|\Sigma_I|^{1/2}}EXP(-(\overline{X}^I-\overline{X}^I)^T\Sigma^{-1}{}_I(\overline{X}^I-\overline{X}^I)/2)$$
$$+\frac{P^K}{(2\pi)^{d/2}|\Sigma_K|^{1/2}}EXP(-(\overline{X}^I-\overline{X}^K)^T\Sigma^{-1}{}_K(\overline{X}^I-\overline{X}^K)/2)$$

En posant: $(\overline{X}^I-\overline{X}^K)^T\Sigma^{-1}{}_K(\overline{X}^I-\overline{X}^K)=K_xi$

$$f_{ik}(\overline{X}^I)=\frac{P^i}{(2\pi)^{d/2}|\Sigma_I|^{1/2}}EXP(0)+\frac{P^K}{(2\pi)^{d/2}|\Sigma_K|^{1/2}}EXP(-K_xi/2) \quad (12)$$

De même pour $f_{ik}(M_{IK})$, on établit une expression analogue à (12)

En prenant: $(M_{IK}-\overline{X}^I)^T\Sigma^{-1}{}_I(M_{IK}-\overline{X}^K)=1 \qquad M_{IK}\in Dw_I$
$(M_{IK}-\overline{X}^K)^T\Sigma^{-1}{}_K(M_{IK}-\overline{X}^K)=K_{Mik}$

$$f_{ik}(M_{IK})=\frac{P^I}{(2\pi)^{d/2}|\Sigma_I|^{1/2}}EXP(-1/2)+\frac{P^K}{(2\pi)^{d/2}|\Sigma_K|^{1/2}}EXP(-K_{Mi}/2) \quad (13)$$

La comparaison de $f_{IK}(X)$ aux points $\overline{X}^I$ et M_{IK} conduit aux résultats suivants:

Composantes distinctes:

$$f_{IK}(\overline{X}^I) > f_{IK}(M_{IK}) = f_{IK}(X^I) - f_{IK}(M_{IK}) > 0$$

$$\frac{P^I.(1-0.3)}{|\Sigma_I|^{1/2}} > \frac{P^K.(EXP(-K_{Mi}/2)-EXP(-K_xi/2))}{|\Sigma_K|^{1/2}}$$

$$0.7\ \frac{P^{I}}{P^{K}}\ \frac{|\Sigma_{K}|^{1/2}}{|\Sigma_{I}|^{1/2}} > EXP(-K_{Mi}/2)\ -EXP(-K_{x}i/2) \qquad (14)$$

Composantes à la limite du fusionnement:

$$0.7\ \frac{P^{I}}{P^{K}}\ \frac{|\Sigma_{K}|^{1/2}}{|\Sigma_{I}|^{1/2}} = EXP(-K_{Mi}/2)\ -EXP(-K_{x}i/2) \qquad (16)$$

Composantes en fusionnement

$$0.7\ \frac{P^{I}}{P^{K}}\ \frac{|\Sigma_{K}|^{1/2}}{|\Sigma_{I}|^{1/2}} < EXP(-K_{Mi}/2)\ -EXP(-K_{x}i/2) \qquad (17)$$

III-6 DETERMINATION DES POINTS M_{IK} ET M_{KI}

La détermination des points M_{KI} et M_{IK} peut être envisagée de plusieurs manières.

Première méthode:

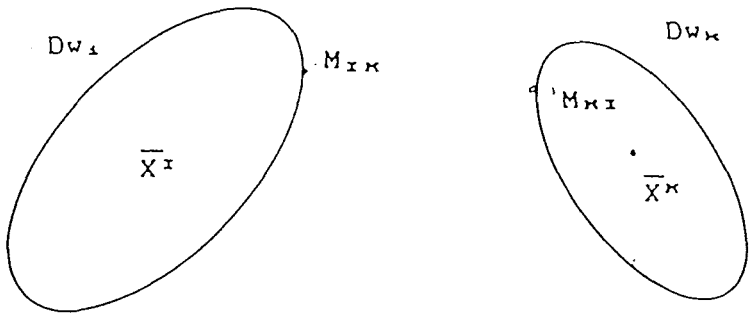


fig.11 contours Dw_I et Dw_K

$$Dw_I = \{ X / (X - \overline{X}^I)^T \Sigma^{-1}_I (X - \overline{X}^I) = 1 \}$$
$$Dw_K = \{ X / (X - \overline{X}^K)^T \Sigma^{-1}_K (X - \overline{X}^K) = 1 \}$$

La première méthode de recherche de ces deux points consiste à définir en fonction d'un paramètre K un faisceau de contours équiprobables tels que:

$$Dw_I(K) = \{ X / (X - \overline{X}^I)^T \Sigma^{-1}_I (X - \overline{X}^I) = K \} \qquad (17)$$

avec $K \in [0, \infty[$

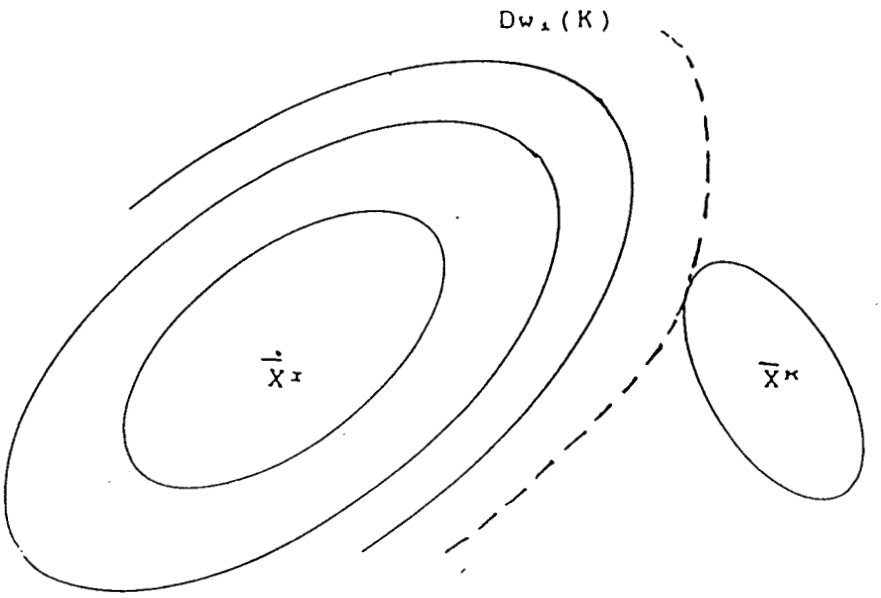


fig.12 faisceaux de contours équiprobables

Le point M_{KI} représente le point de tangence entre les contours $Dw_1(K)$ et Dw_K

La traduction analytique du problème posé est alors la suivante:

soit l'ensemble des points X tels que:

$$(X - \bar{X}^I)^T \Sigma^{-1}_I (X - \bar{X}^I) = K$$
$$\text{et} \qquad (19)$$
$$(X - \bar{X}^K)^T \Sigma^{-1}_K (X - \bar{X}^K) = 1$$

A partir de (19) on obtient (20)

$$(X - \bar{X}^I)^T \Sigma^{-1}_I (X - \bar{X}^I) = K(X - \bar{X}^K)^T \Sigma^{-1}_K (X - \bar{X}^K) \quad (20)$$

Dans ces conditions le point M_{KI} correspond à une valeur de K pour laquelle l'ensemble des points X solution du problème se réduit à un seul point.

On met l'expression (20) sous la forme suivante:

$$A(X) + B(X) + C = 0 \qquad (21)$$

A(X) fontionnelle quadratique
B(X) " linéaire
C constante scalaire

avec la matrice A égale à $A = \Sigma^{-1}_{I} - K\Sigma^{-1}_{K}$
le vecteur B " $B = -\Sigma^{-1}_{I}\bar{X}^{I} + K\Sigma^{-1}_{K}\bar{X}^{K}$ (22)
la constante C " $C = (\bar{X}^{I})^{T}\Sigma^{-1}_{I}\bar{X}^{I} + (\bar{X}^{K})^{T}\Sigma^{-1}_{K}\bar{X}^{K}$

Intuitivement,on voit que le problème présente deux solutions:

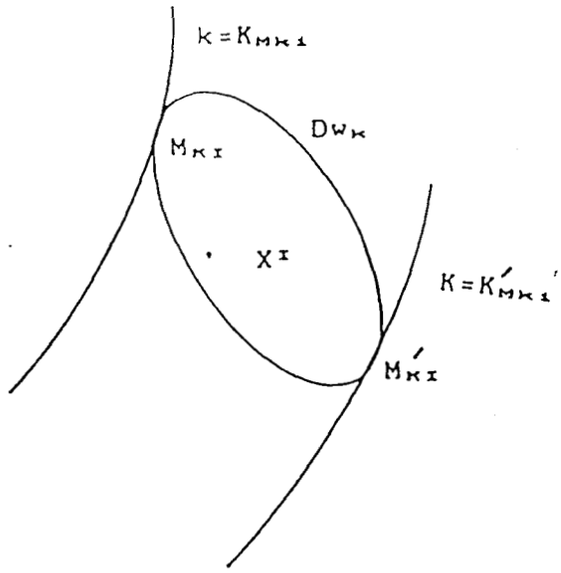
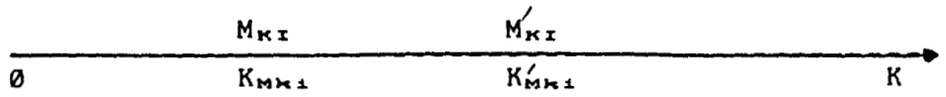


fig.13 M_{KI} et M'_{KI} points de tangence entre $Dw_I(K)$ et Dw_K

M_{KI} est le point le plus proche de w_I et appartenant à w_K
M_{KI} " " éloigné " "

Dans le cas où d=3 on a les résultats suivants:



Entre M_{KI} et M'_{KI} l'expression (2Ø) founit un faisceau de formes quadratiques.[GAN,66]

Pour $k=K_{MKI}$, la forme quadratique représentée par (21) a un seul point comme solution et il est aussi le centre de cette forme quadratique.[POS,79]

Soit $\overline{X}_o$ le centre cherché , en posant $Y=X - \overline{X}_o$ l'expression (21) devient:

$$(Y + \overline{X}_o)^T A(Y + X_o) + (Y + \overline{X}_o)^T B + C = 0 \qquad (23)$$

en développant(23) on obtient:

$$Y^T AY + Y^T(B + 2A\overline{X}_o) + C + \overline{X}_o^T A\overline{X}_o + \overline{X}_o^T B = 0 \qquad (24)$$

(24) représente une forme quadratique centrée si elle se met sous la forme: $Y^T AY + C' = 0$

le centre $\overline{X}_o$ solution du problème est fourni par l'expression:

$$B + 2A\overline{X}_o = 0 \Longrightarrow \overline{X}_o = A^{-1}B/2 \qquad (25)$$

Du fait que K intrevient dans l'établissement de la matrice A et du vecteur B , on ne peut malheureusement pas déterminer $\overline{X}_o$ d'une manière simple.

deuxième méthode

Une autre manière d'aborder le problème de la recherche des points M_{KI} et M_{IK} a été déja présentée dans le paragraphe 3 de ce chapitre.

Elle consiste à minimiser des distances au sens de MAHANALOBIS.

pour le point M_{KI} .

minimiser la fonction

$$f_I(X)=(X-\bar{X}^I)^T\Sigma^{-1}_I(X-\bar{X}^I)$$

sous la containte (26)

$$(X-\bar{X}^K)^T\Sigma^{-1}_K(X-\bar{X}^K)=1$$

En utilisant la technique des multiplicateurs de LAGRANGE on obtient:

$$g_i(X)=f_I(X)+\mu[(X-\bar{X}^K)^T\Sigma^{-1}_K(X-\bar{X}^K)-1]=0$$

$$\frac{g_i(X)}{\partial X}=\Sigma^{-1}_i(X-X^I)+\mu\Sigma^{-1}_K(X-X^K)=0 \qquad (27)$$

$$\frac{g_i(X)}{\partial \mu}=(X-\bar{X}^K)^T\Sigma^{-1}_K(X-\bar{X}^K)-1=0$$

avec μ multiplicateur de LAGRANGE

En multipliant la deuxième égalité des expressions (27) par le vecteur $(X-\bar{X}^K)^T$, on obtient la valeur de μ.

$$(X-\bar{X}^K)^T\Sigma^{-1}_I(X-\bar{X}^I)+\mu(X-\bar{X}^K)^T\Sigma^{-1}_K(X-\bar{X}^K)=0 \qquad (28)$$

$$\mu=-(X-\bar{X}^K)^T\Sigma^{-1}_I(X-\bar{X}^I) \qquad (29)$$

La valeur de μ est placée dans la deuxième égalité des expression (27).

$$\Sigma^{-1}_{I}(X - X^{I}) = \Sigma^{-1}_{K}(X - X^{K}).\underbrace{(X - X^{K})^{T}\Sigma^{-1}_{I}(X - X^{I})}_{\mu} \qquad (30)$$

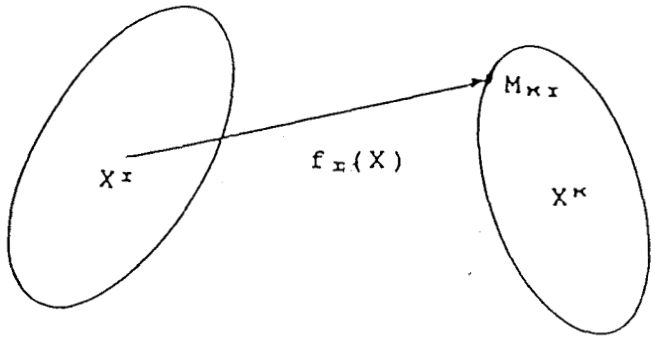


fig.14 minimisation de la distance $f_{IK}(X)$

Il n'y a pas de solution analytique simple à ce problème [OTT,78]

Le paragraphe suivant propose une solution numérique utilisant un algorithme rapide.

BU LILLE

III-7 ALGORITHME RAPIDE DE RECHERCHE DES POINTS M_{IK} ET M_{KI}

Pour trouver les points M_{IK} et M_{KI},on utilise un algorithme simple s'appuyant sur une méthode de relaxation.

Il consiste à générer deux suites de points $(M_{IK,0},M_{IK,1},\ldots,M_{KI,n})$ et $(M_{KI,0},M_{KI,1},\ldots,M_{KI,m})$ convergeant vers les solutions M_{KI} et M_{IK}.

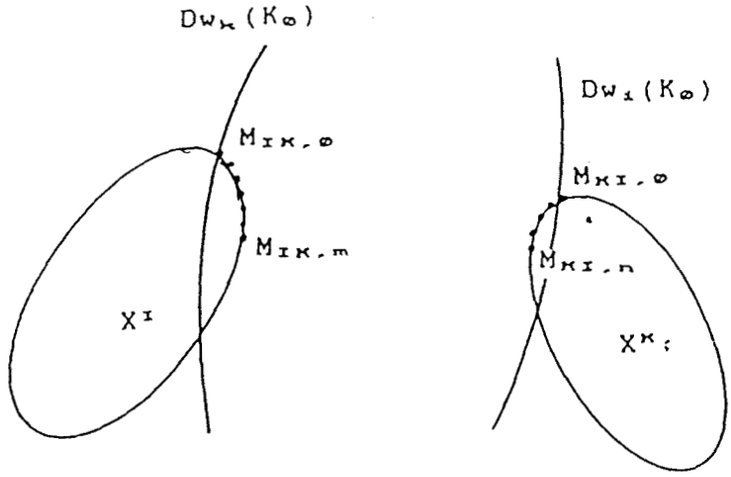


fig.15 définition de deux suites de points convergentes vers M_{IK} et M_{KI}

L'algorithme adopté se décompose en plusieurs étapes:

Première étape

Pour initialiser l'algorithme,on choisit sur le contour Dw_K un point initial $M_{KI,0}$

Pratiquement ce point est l'extrémité de la direction propre de w_k la plus proche de w_i au sens de la distance généralisée.

Le domaine $Dw_i(K_\emptyset)$ portant le point $M_{kI,\emptyset}$ a pour équation

$$Dw_i(K_\emptyset)=\{ X/(X - X^I)^T\Sigma^{-1}_I(X - X^I)=K_\emptyset\} \qquad (31)$$

avec
$$K_\emptyset=(M_{kI,\emptyset} - X^I)^T\Sigma^{-1}_I(M_{kI,\emptyset} - X^I)$$
$$1=(M_{kI,\emptyset} - X^k)^T\Sigma^{-1}_k(M_{kI,\emptyset} - X^k) \qquad (32)$$

Deuxième étape

Les domaines $Dw_i(K_\emptyset)$ et Dw_k étant convexes leur intresection est convexe.[CIA,85]

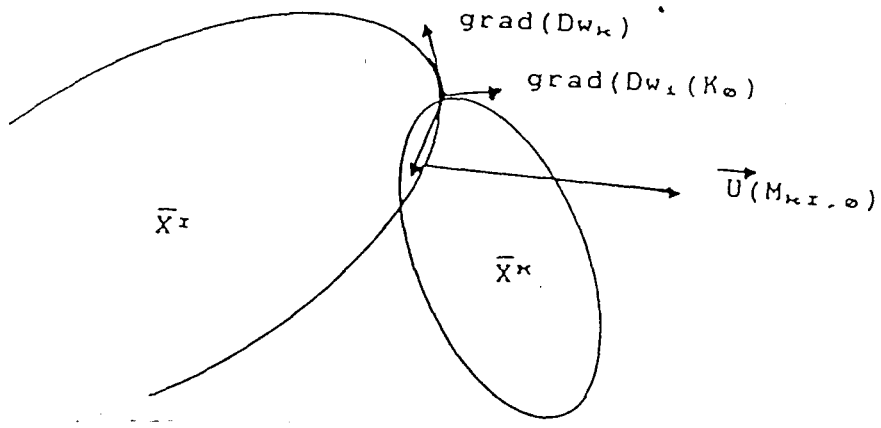


fig.16 intersection de deux domaines convexes

En calculant au point $X=M_{kI,\emptyset}$ les gradients de $Dw_i(K_\emptyset)$ et Dw_k et en composant ces gradients , on obtient une direction $U(M_{kI,\emptyset})$ dirigée vers le centre du domaine convexe intersection.

$$U(M_{kI,\emptyset})=- (\mathrm{grad}(Dw_i(K_\emptyset)) + \mathrm{grad}(Dw_k))_{X=Mki,\emptyset} \qquad (33)$$

est choisi de telle sorte que l'extrémité du vecteur U soit à l'intérieur du domaine convexe intersection.

Troisième étape

Soit $M_{KI,0}$ l'extrêmité du vecteur $\overrightarrow{U(_{KI,0})}$,dans ces conditions on appelle $M_{KI,1}$ l'intersection de la direction $\overrightarrow{\bar{X}^K M_{KI,0}}$ avec la frontière du domaine Dw_K.

Compte tenu de la convexité des domaines mis en jeu,$M_{KI,1}$ est plus proche de Point cherché M_{KI} que $M_{KI,0}$.

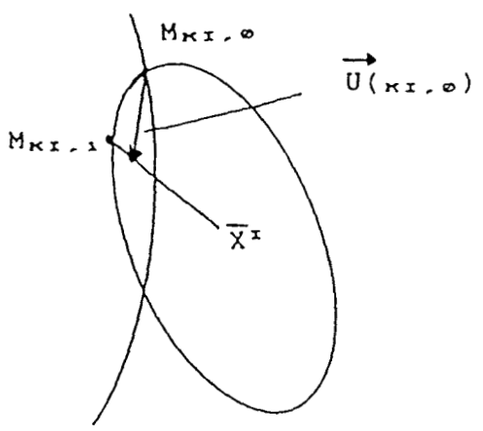


fig.17 définition de la suite de points $M_{K1,0}, M_{K1,1}, \ldots$

Quatrième étape

On définit alors le contour $Dw_1(K_1)$ portant le deuxième point de la suite par les expressions suivantes:

$$Dw_1(K_1)=\{ X/(X - \bar{X}^I)^T \Sigma^{-1}_I (X - \bar{X}^I)=K_1 \qquad (34)$$

avec $K_1=(M_{KI,1} - \bar{X}^I)^T \Sigma^{-1}_I (M_{KI,1} - \bar{X}^I)$

$$1=(M_{KI,1} - \bar{X}^K)^T \Sigma^{-1}_K (M_{KI,1} - \bar{X}^K) \qquad (35)$$

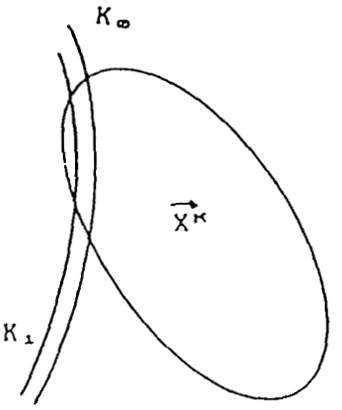


fig.18 définition des contours $K_0, K_1, \ldots$

Cinquième étape

En réitérant les étapes 2,3 et 4 on génère une suite de points $M_{KI,0}, M_{KI,1}, \ldots, M_{KI,n}$ convergeant vers le point M_{KI} cherché.

La même procédure est ensuite appliquée pour la recherche du point M_{IK}.

La détermination d'un point prend en moyenne 210 ms pour un exemple traité dans un espace bidimensionnel (d=2) et 314 ms pour un exemple tridimensionnel (d=3).

III-8 EXEMPLES D'APPLICATION

III-8-1 INTRODUCTION

L'exemple présenté a été éffectué dans un espace à deux dimensions (d=2) . On génère un MG composé de deux composantes dont les caractéristiques statistiques sont résumées dans le tableau suivant:

	vecteurs moyennes	écarts types	directions propres	probabilité a priori
composante 1	$\bar{X}_1=102$ $\bar{X}_2=200.1$	$\lambda_1=6.3$ $\lambda_2=4.6$	U=-8°	P=0.51
composante 2	$\bar{X}_1=123.1$ $\bar{X}=183.3$	$\lambda_1=5.9$ $\lambda_2=4.9$	U=2°	P=0.49

Tableau-1 paramètres du MG

Une troisième composante située entre les deux premières est crée , ses paramètres sont dans le tableau-2

vecteur moyenne	écarts types	directions propres
$\bar{X}_1=112.5$ $\bar{X}_2=192.5$	$\lambda_1=5$ $\lambda_2=3$	U=-45°

Tableau-2 paramètres de la troisième composante

BU LILLE

On prend un ensemble e_1 , contenant 200 observations tirées de la troisième composante , qu'on intègre dans le mélange initial.

Les deux critères de séparabilité , probabilité a posteriori et modes de la fonction de densité de probabilité conjointe , sont utilisés ; les résultats obtenus sont analysés.

III-8-2 FONCTION BIMODALE

Les courbes,représentées dans la figure 19,illustrent l'évolution des estimations de la fonction bimodale $f_{IK}(X)$ aux quatres points $\overline{X}^{I}$,$\overline{X}^{K}$,M_{KI} et M_{IK} en fonction de l'intégration des observations tirées de l'ensemble e_1.

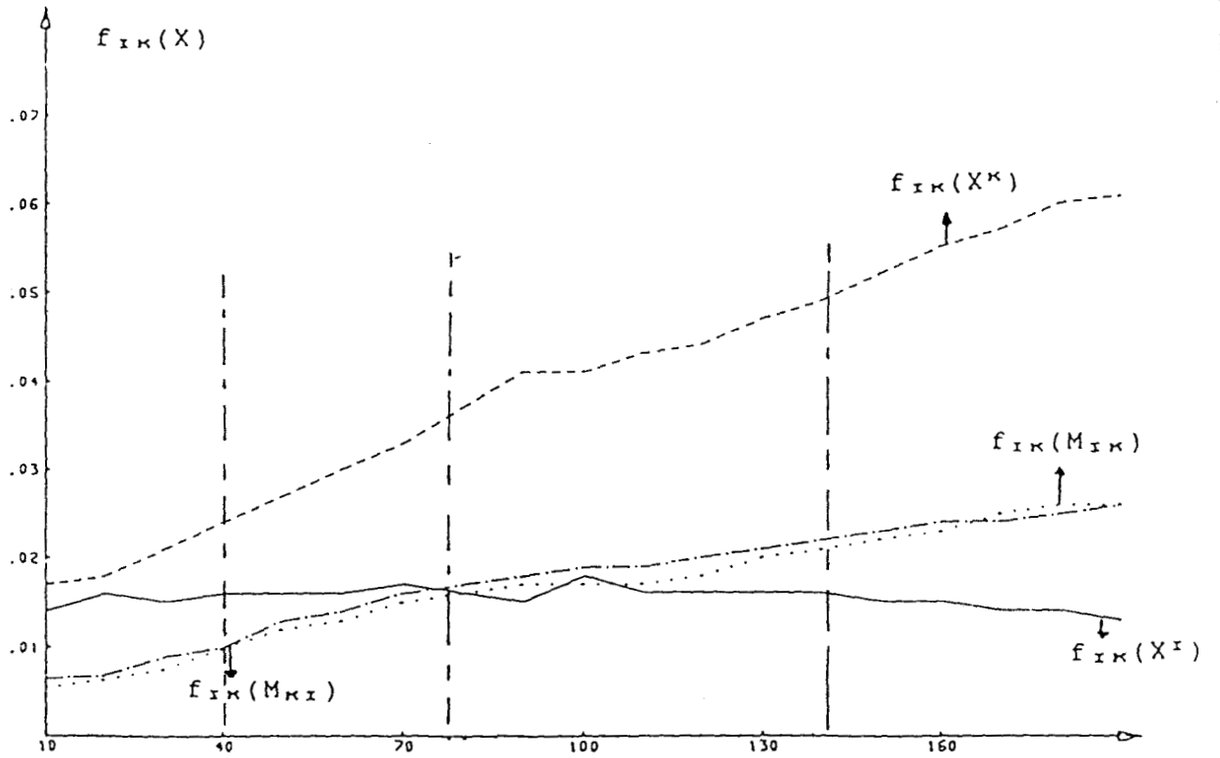


fig.19 évolution des estimations de $f_{IK}(X)$

Les courbes de la figures 19 font ressortir nettement l'existence de quatres zones.

la première zone correspond à la situation où les composantes du MG sont encore distinctes.

Pour donner une image du MG dans cette zône,on considère l'état du MG obtenu après intégration de 40 observations tirées e_1.

Le tableau-3 contient toutes les informations sur l'état du MG à ce moment précis de son évolution.

	vecteurs moyennes	écarts types	directions propres	probabilités a priori
composante 1	$\bar{X}_1=107.5$ $\bar{X}_2=199.5$	$\lambda_1=7$ $\lambda_2=3.8$	U=-28°	P=0.4
composante 2	$\bar{X}_1=116.7$ $\bar{X}_2=188.8$	$\lambda_1=7.3$ $\lambda_2=3.5$	U=-36°	P=0.6

Tableau-3 état du MG après intégration de 40 observations

$f_{IK}(X^I)=(0.16).10^{-1} \quad > \quad f_{IK}(M_{IK})=(0.99).10^{-2}$
$f_{IK}(X^K)=(0.4).10^{-1} \quad > \quad f_{IK}(M_{KI})=(0.1).10^{-1}$

La figure 20 , qui représente l'image du MG , montre la vallée qui sépare les deux centres de masse.

La deuxième zone correspond au cas critique où les composantes sont sur le point de fusionner.

le tableau-4 et la figure 21 illustrent l'état du MG après avoir intégré 80 observations tirées de e_1.

	vecteurs moyennes	écarts types	directions propres	probabilités a priori
composante 1	$\bar{X}_1=106.8$ $\bar{X}_2=199$	$\lambda_1=6.3$ $\lambda_2=3.4$	U=-27°	P=0.35
composante 2	$\bar{X}_1=115.5$ $\bar{X}_2=190.2$	$\lambda_1=6.4$ $\lambda_2=2.9$	U=-38	P=0.65

Tableau-4 état du MG après l'intégration de 80 observations

$$f_{IK}(\bar{X}^{I})=(0.16).10^{-1} \simeq f_{IK}(M_{IK})=(0.16).10^{-1}$$
$$f_{IK}(\bar{X}^{K})=(0.37).10^{-1} > f_{IK}(M_{KI})=(0.17).10^{-1}$$

On remarque sur la figure 21 que la fonction $f_{IK}(X)$ est sur le point de devenir unimodale.

La troisième zone correspond au cas interressant du fusionnement.

Après l'arrivée de la 140ième observation l'état du MG est comme suit:

	vecteurs moyennes	écarts types	directions propres	probabilités a priori
composante 1	$\bar{X}_1=107$ $\bar{X}_2=199$	$\lambda_1=5.9$ $\lambda_2=3.1$	U=-25	P=0.37
composante 2	$\bar{X}_1=114.8$ $\bar{X}_2=191$	$\lambda_1=5.73$ $\lambda_2=2.9$	U=-39	P=0.63

tableau-5 état du MG après integration de 140 observations

$$f_{IK}(\bar{X}^{I})=(0.16).10^{-1} < f_{IK}(M_{IK})=(0.2).10^{-1}$$
$$f_{IK}(\bar{X}^{K})=(0.49).10^{-1} > f_{IK}(M_{KI})=(0.22).10^{-1}$$

On note que la fonction $f_{IK}(X)$ ne présente plus qu'un seul môde.

La quatrième et dernière zône de la figure 19 correspond au cas où le degré de fusionnement des composantes est très important.

Après l'intégration de 180 observations on a les résultats suivants:

	vecteurs moyennes	écarts types	directions propres	probabilités a priori
composante 1	$\bar{X}_1=107.3$ $\bar{X}_2=199$	$\lambda_1=5.8$ $\lambda_2=3.1$	U=-25°	P=0.25
composante 2	$\bar{X}_1=114.3$ $\bar{X}_2=191.6$	$\lambda_1=5.4$ $\lambda_2=2.6$	U=-41°	P=0.75

Tableau-6 état du MG dans le cas où le taux de fusionnement est important

$$f_{IK}(\bar{X}^I)=(0.14).10^{-1} \quad < \quad f_{IK}(M_{IK})=(0.25).10^{-1}$$
$$f_{IK}(\bar{X}^K)=(0.57).10^{-1} \quad > \quad f_{IK}(M_{KI})=(0.24).10^{-1}$$

On remarque que $f_{IK}(M_{IK})$ devient plus grand que $f_{IK}(M_{KI})$, ceci provient du fait que M_{IK} se trouve à l'intérieur de la composante w_K . La figure 23 représente cette situation.

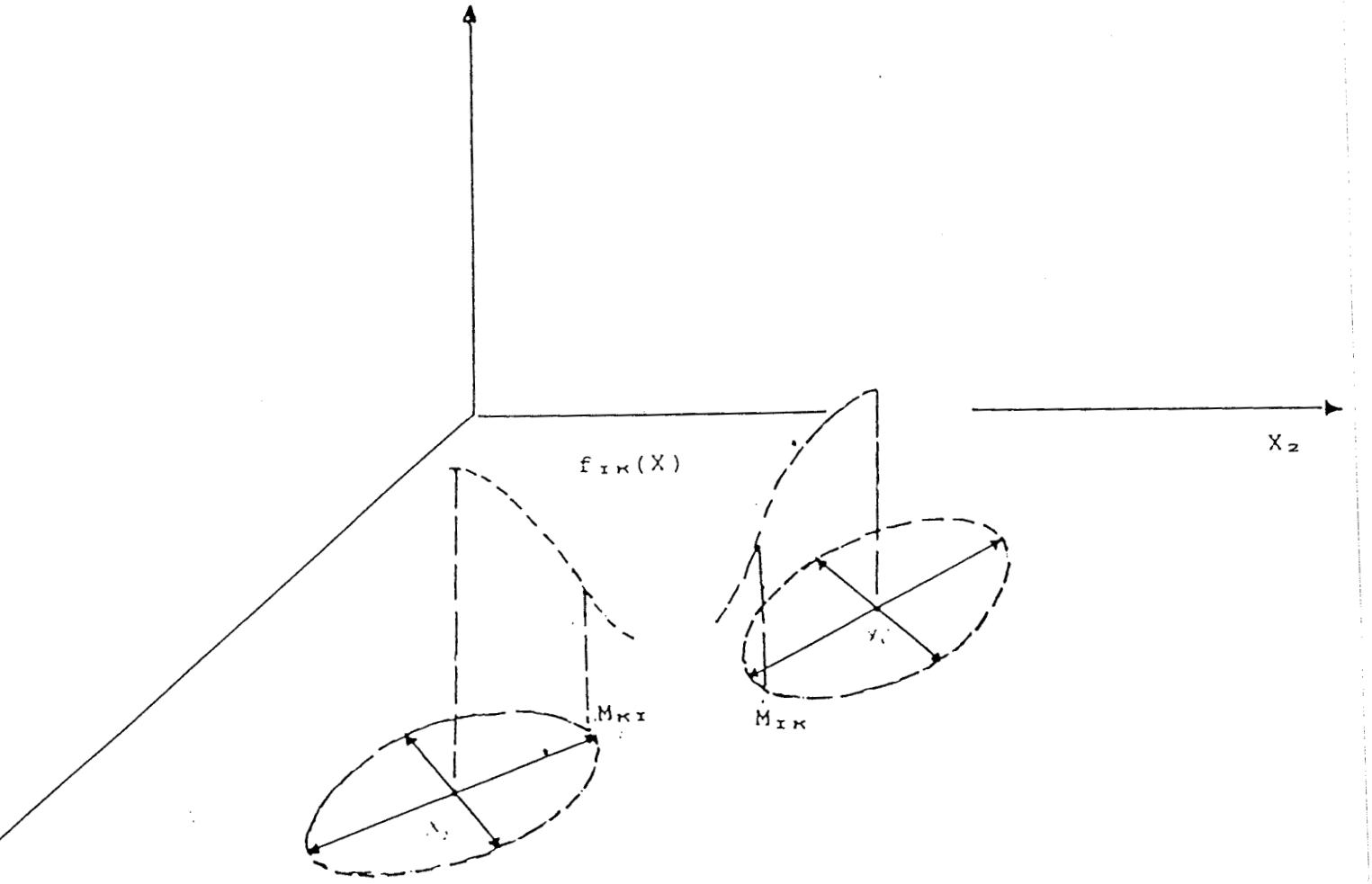


fig.20 image du MG obtenue apres l'intégration de 40 observations

BIBLIOTHEQUE

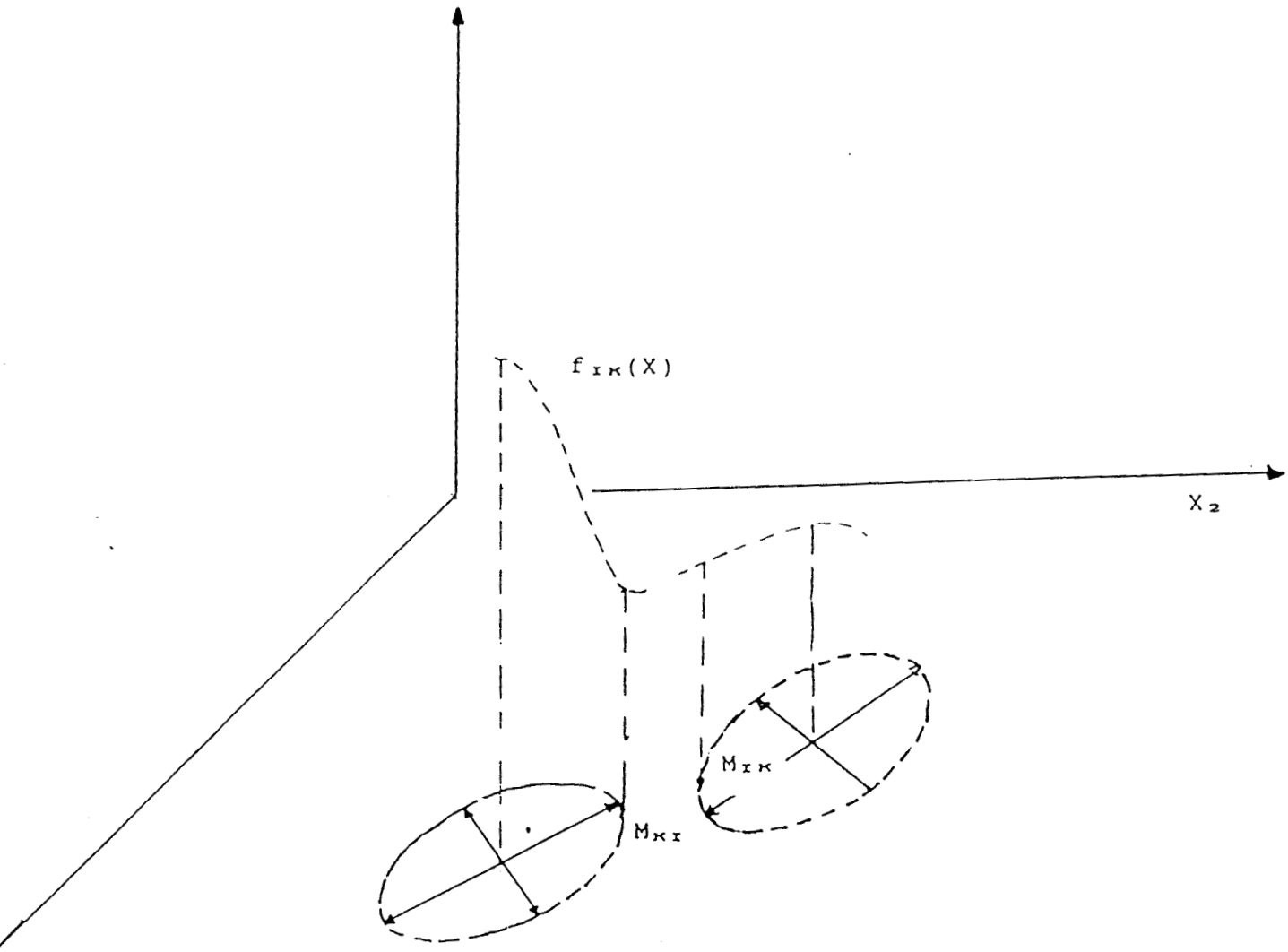


fig.21 image du MG obtenue apres l'intégration de 80 observations

BU LILLE

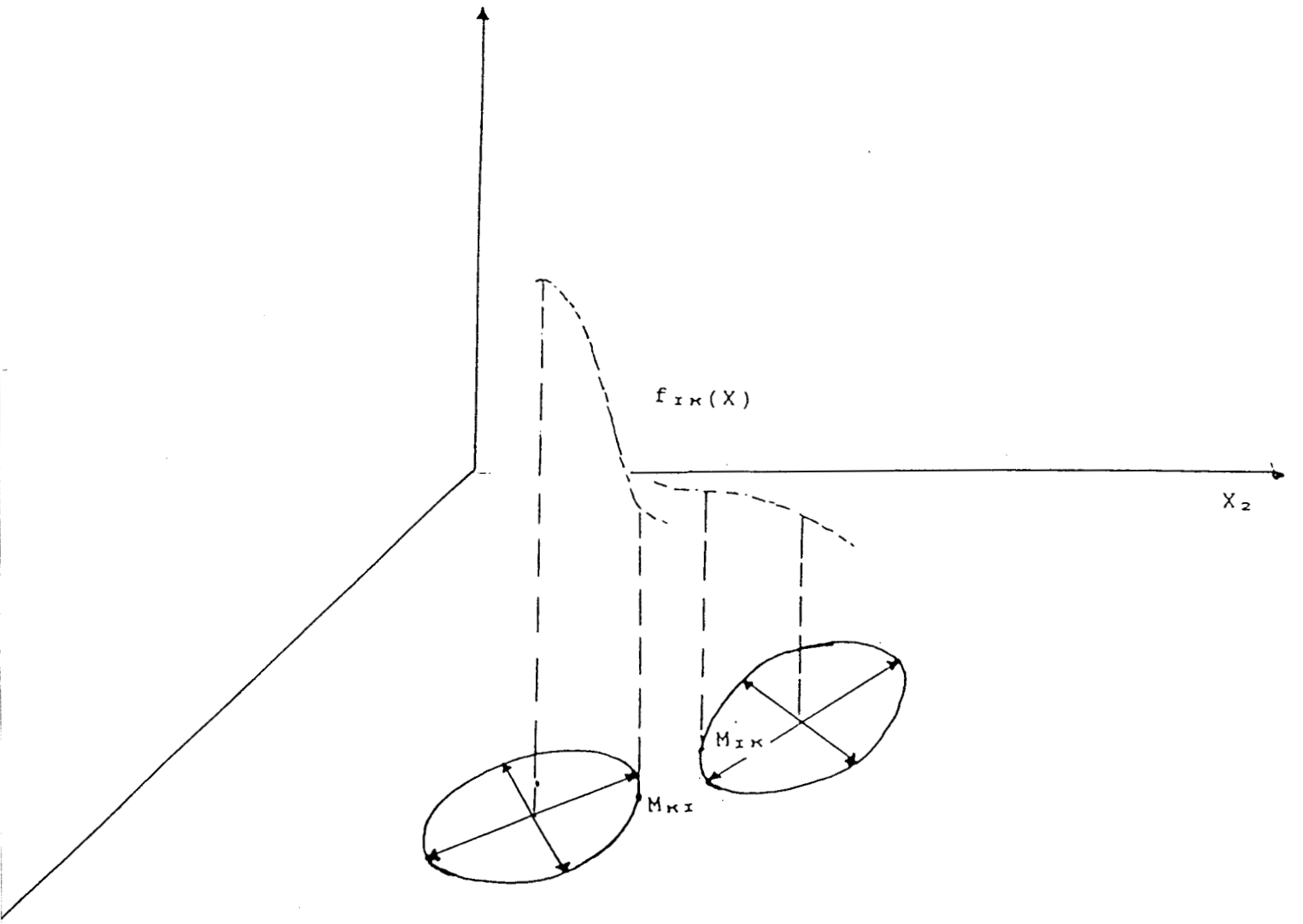


fig.22 image du MG obtenue apres l'intégration de 140 observations

BU LILLE

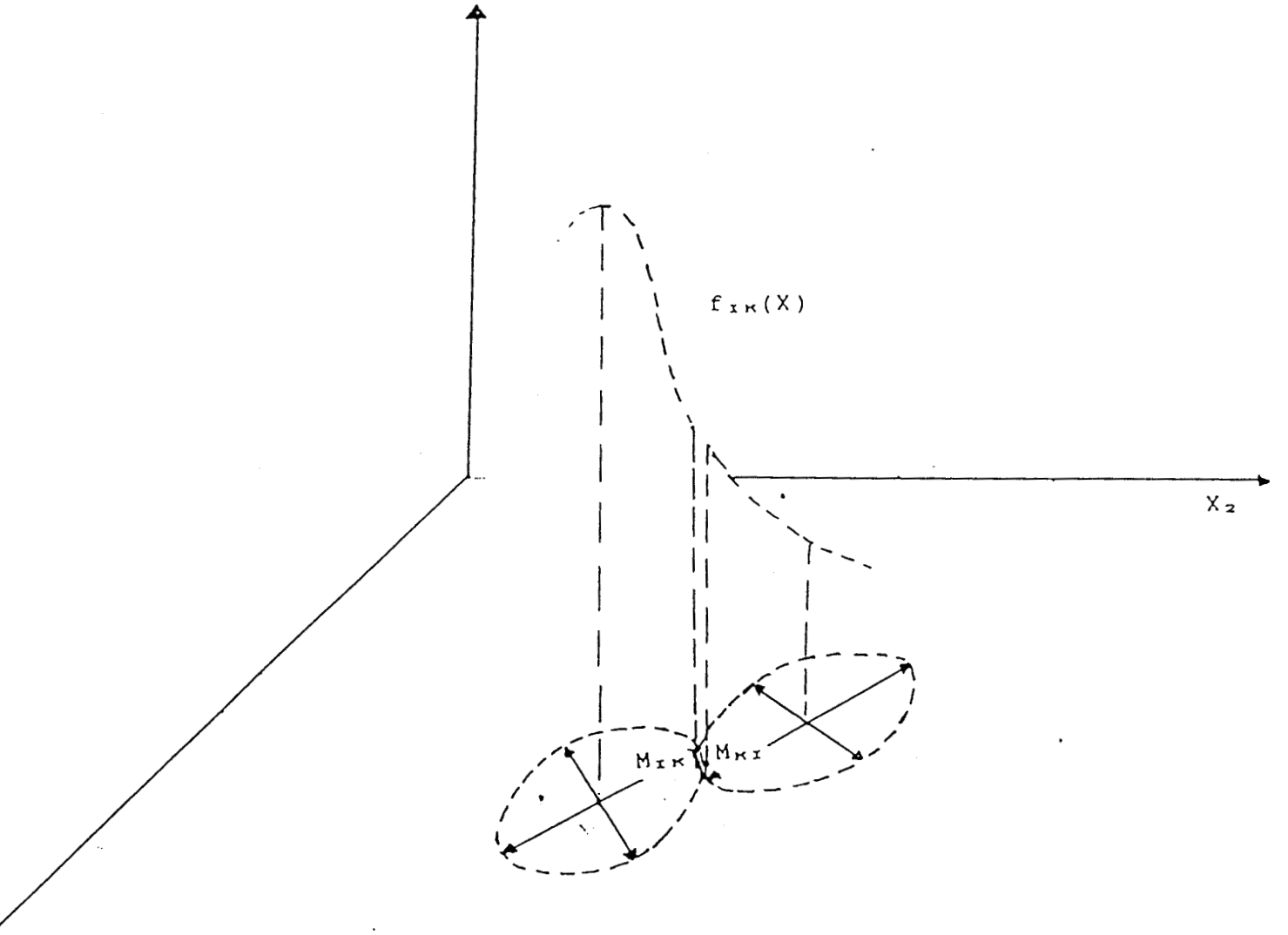


fig.23 image du MG obtenue apres l'intégration de 180 observations

III-8-3 UTILISATION DES PROBABILITES A POSTERIORI

le suivi du classement des points tels M_{IK} et M_{KI} , en fonction de l'intégration des nouvelles observations , est important dans la mesure où il permet d'une part de décéler le début d'influence d'une composante sur l'autre et d'autre part de pouvoir estimer le degré de cette influence.

Les figures 24,25 et 26 représentent l'évolution des probabilités a posteriori des points M_{IK} et M_{KI}.

Sur la figure 24 , on note que le point M_{IK} , qui appartient par définition à w_I , est constamment attribué à w_K à partir de l'intégration de 50 observations . Ce qui traduit l'influence de plus en plus prépondérante de w_K sur w_I.

La figure 26 , qui regroupe les figures 24 et 25 , met en évidence plusieurs zones.

Dans la première zone , les composantes w_I et w_K demeurent distinctes mais leur influence mutuelle commence à se faire sentir.

$$P^I(w_I/M_{IK}) > P^K(w_K/M_{IK})$$
$$P^K(w_K/M_{KI}) > P^I(w_I/M_{KI})$$

les points M_{IK} et M_{KI} sont correctement classés.

Dans la deuxième zone , le point M_{KI} continue d'être classé correctement tandis que le point M_{IK} est mal classé .

$$P^I(w_I/M_{IK}) < P^K(w_K/M_{IK})$$
$$P^K(w_K/M_{KI}) > P^I(w_I/M_{KI})$$

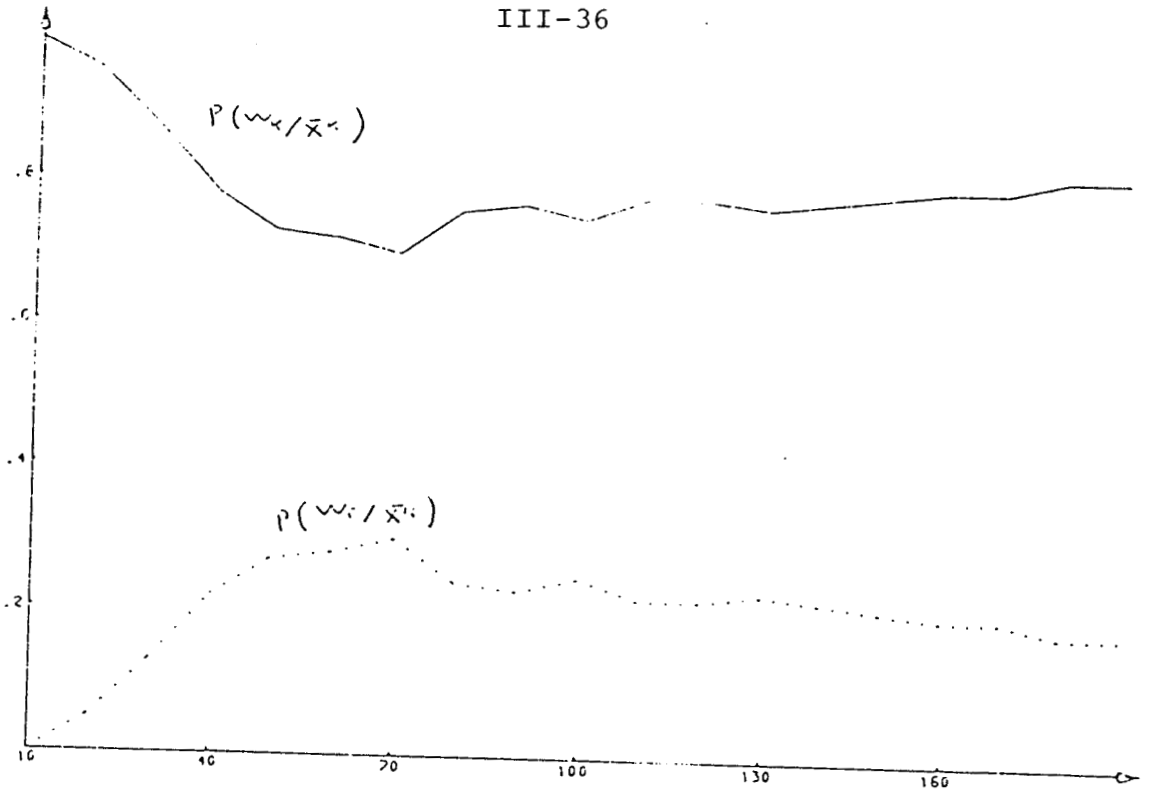


fig.25 évolution du classement du point M_{KI}

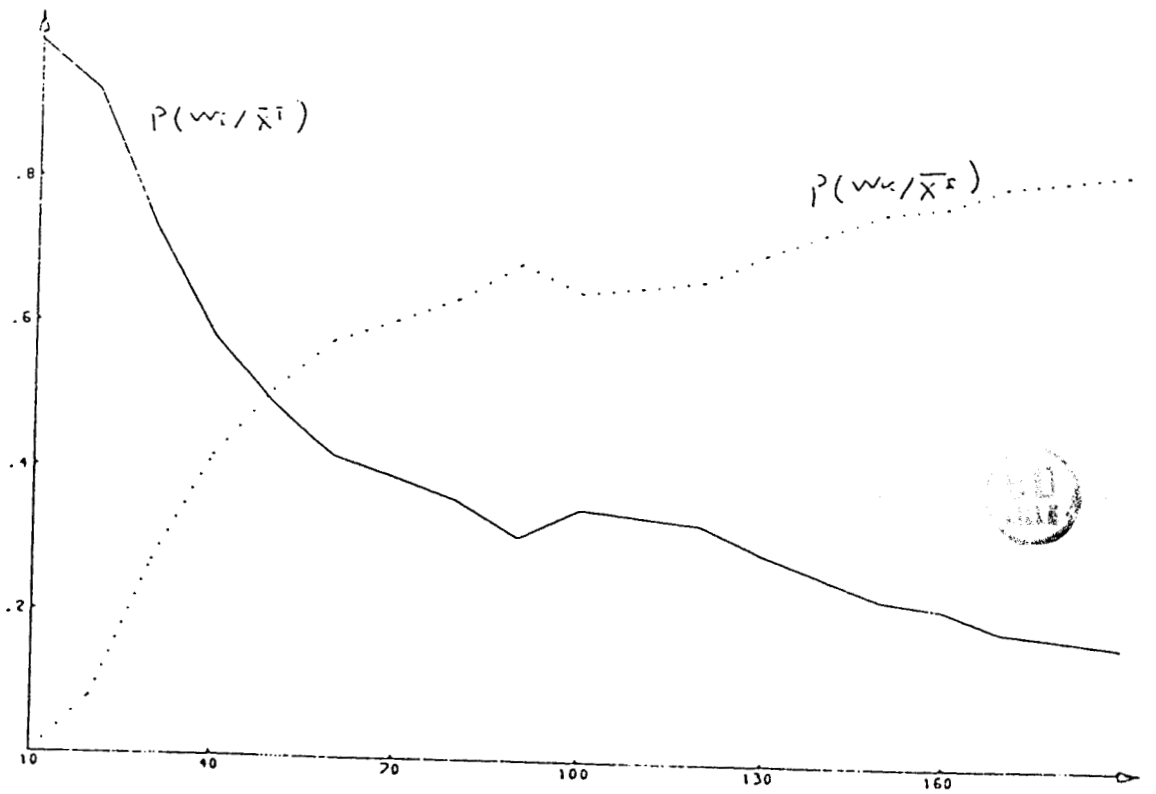


fig.24 évolution du classement du point M_{IK}

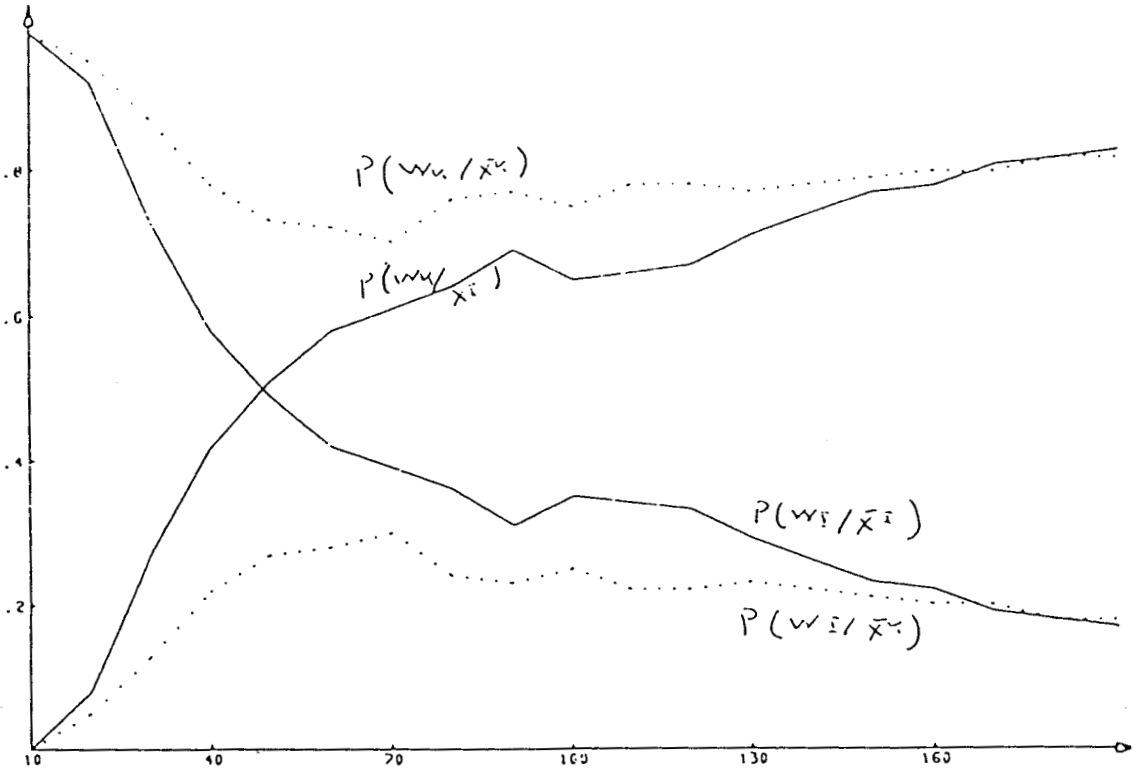
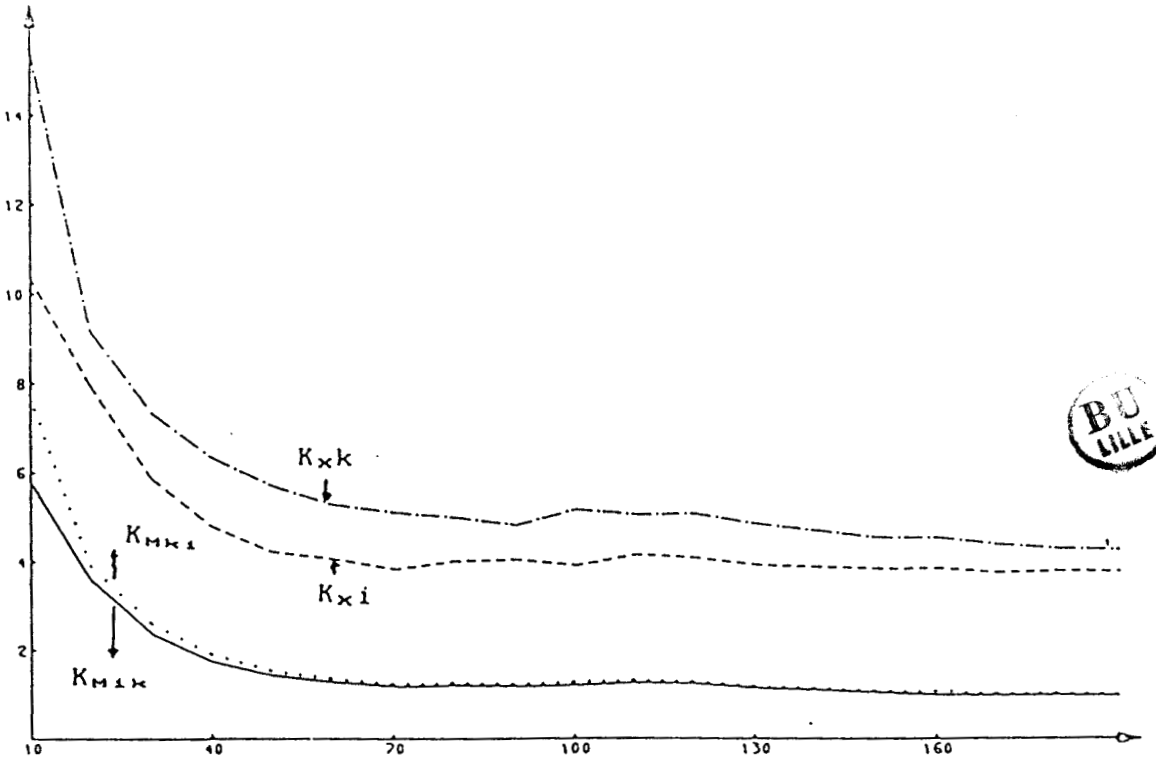


fig.26 évolution du classement des points M_{KI} et M_{IK}



BU LILLE

fig.28 évolution des différentes distances

Du fait que les composantes ont tendance à fusionner les points M_{IK} et M_{KI} se rapprochent de plus en plus comme le montre la figure 28.Ce qui explique la tendance des points M_{IK} et M_{KI} à avoir les mêmes probabilités a posteriori par rapport aux deux composantes.

Aprés l'intégration de 170 observations M_{IK} est confondu avec M_{KI}.

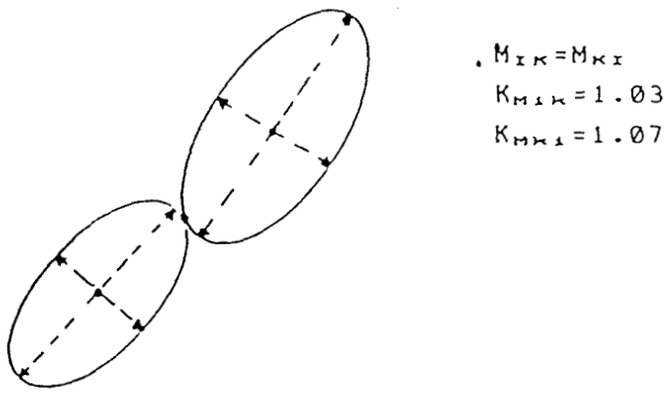


fig.29 état du mélange aprés avoir intégré 170 observations

La troisième zone traduit le degré de fusionnement des composantes.

Le point M_{IK},qui par définition caractérise w_I,se trouve à l'intérieur de w_K et possède par conséquent plus de chance d'appartenir à w_K que le point M_{KI} .La même remarque peut être faite pour le point M_{KI}.

III-8-4 PARAMETRE DE LA COMPOSANTE SOMME

les paramètres de la composante somme résultant du fusionnement de w_K et w_I sont les suivants:

- centre de masse $(P^I + P^K)X = P^I\bar{X}^I + P^K\bar{X}^K$

  $\bar{X}^1_1 = 111.85$

  $\bar{X}^1_2 = 193.4$

- matrice de covariance $(P^I + P^K)\Sigma = P^I\Sigma^I + P^K\Sigma^K$

-écarts types

$\lambda_1 = 5.34$

$\lambda_2 = 2.69$

- direction propre

  $U = -42°$

En créant un mélange gaussien constituè des trois composantes utilisées dans cet exemple (voir tableaux 1 et 3), le résultat fourni par la méthode graphique lorsqu'on prend un ensemble e_2 contenant 300 observations est le suivant:

centre de masse	écarts types	direction propre
$\bar{X}_1 = 110$ $\bar{X}_2 = 194$	$\lambda_1 = 12.8$ $\lambda_2 = 4.58$	$U = -36°$

III-9 CONCLUSION

Les critères retenus dans ce chapitre sont très utilisés pour mesurer la séparabilité des composantes d'un mélange gaussien .

L'estimation de ces critères uniquement en quelques points privilégiés de l'espace d'observation n'enlève rien à leurs efficacités , car les points choisis traduisent toute l'influence que peut avoir une composante donnée du mélange gaussien sur les autres .

REFERENCES

K.S.FUKUNAGA
"INTRODUCTION TO STATISTICAL PATTERN RECOGNITION"
ACADEMIC PRESS NEW-YORK 1972

K.S.FUKUNAGA AND S.ANDO
" THE OPTIMUM NON LINEAR FOR A SCATTER CRITERION IN DISCRIMINANT ANALYSIS"
IEEE TRANS INFORM THEORY VOL 23 N°4 JULY 1977

K.S.FUKUNAGA AND R.D.SHORT
" A CLASS OF FEATURE EXTRACTION CRITERION AND ITS RELATION TO THE BAYES RISK ESTIMATE"
IEEE TRANS INFORM THEORY VOL 26 N°1 JUNUARY 1978

K.S.FUKUNAGA AND R.D.SHORT
" NON LINEAR FEATURE EXTRACTION WITH A GENERAL CRITERION FUNCTION"
IEEE TRANS INFORM THEORY VOL 24 N°5 SEPTEMBE 1978

S.WATANABE
" PATTERN RECOGNITION AS A QUEST FOR MINIMUM ENTROPY"
PATTERN RECOGNITION VOL 13 N°5 1985

M.P.OEFFELIN AND P.G.VOS
" AN ALGORITHM FOR PATTERN DESCRIPTION ON THE LEVEL OF RELATIVE PROXIMITY"
PATTERN RECOGNITION VOL 16 N°3 1983

F.R.GANSMATCHER
" THEORIE DES MATRICES TOME 2"
DUNOD 1966

P.J.VAN OTTERLOO
" A DISTRIBUTION FREE GEOMETRIC UPPER BOUND FOR THE PROBABILITY OF ERROR OF A MINIMUM DISTANCE CLASSIFIER"
PATTERN RECOGNITION VOL 10 1978

P.G CIARLET
" INTRODUCTION A L'ANALYSE NUMERIQUE MATRICIELLE ET A L'OPTIMISATION"
MASSON 1985

M.POSTNIKOV
" LECONS DE GEOMETRIE"
EDITIONS DE MOSCOU 1979

CHAPIRE IV

AUGMENTATION DU NOMBRE DE COMPOSANTES D'UN MELANGE GAUSSIEN

CHAPIRE IV

AUGMENTATION DU NOMBRE DE COMPOSANTES D'UN MELANGE GAUSSIEN

IV-1 INTRODUCTION 1

IV-2 MESURE DE LA DISPERSION DES COMPOSANTES 3

IV-3 DEFINITION D'UN CONTOUR DE CONFIANCE 7

IV-3-1 INTRODUCTION 7

IV-3-2 CAS OU LE NOMBRE D'OBSERVATIONS EST GRAND 7

IV-3-3 CAS OU LE NOMBRE D'OBSERVATIONS EST FAIBLE 12

IV-4 DEFINITION D'UN CONTOUR DE CONFIANCE POUR LES NOUVEAUX CENTRES DE MASSE 16

IV-5 PROCEDURE A SUIVRE DANS LE CAS OU LE NOMBRE DE COMPOSANTES D'UN MELANGE GAUSSIEN AUGMENTE 18

IV-5-1 INTRODUCTION 18

IV-5-2 APPARITION D'UNE NOUVELLE COMPOSANTE 18

IV-5-3 QUELQUES SITUATIONS PARTICULIERES 23

IV-5-4 ECLATEMENT D'NE COMPOSANTE EN DEUX 25

IV-6 EXEMPLES D'APPLICATION 29

IV-6-1 APPARITION D'UNE NOUVELLE COMPOSANTE 29

IV-6-2 ECLATEMENT D'UNE COMPOSANTE EN DEUX 31

IV-7 CONCLUSION 37

BIBLIOGRAPHIES 38

IV AUGMENTATION DU NOMBRE DE COMPOSANTES D'UN MELANGE GAUSSIEN

IV-1 INTRODUCTION

Nous avons présenté dans le premier chapitre un aperçu sur les différentes situations dans lesquelles le nombre de composantes modélisant un système peut augmenter.

Les figures 1 et 2 nous rapellent ces différents cas.

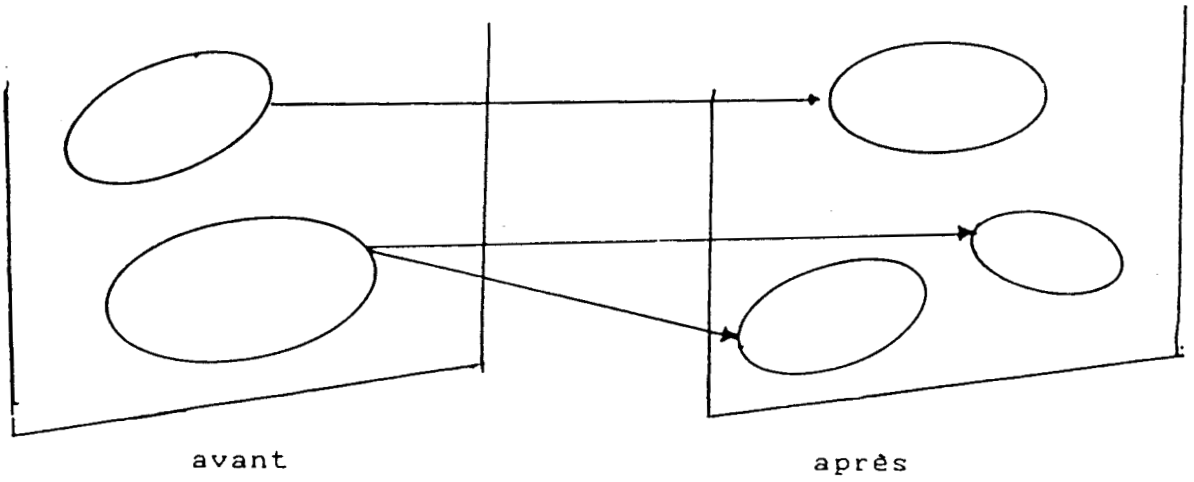


fig.1 scission d'une composante en deux.

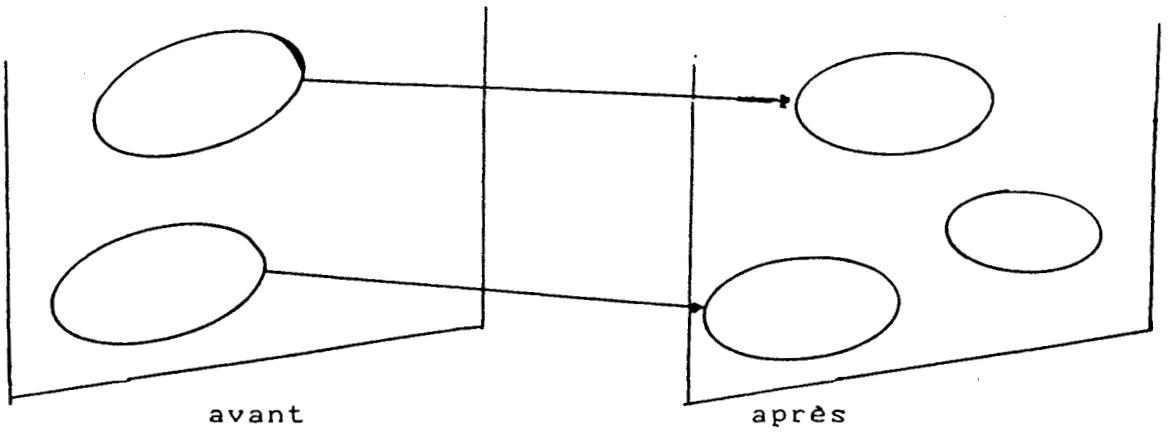


fig.2 apparition d'une nouvelle composante.

Le changement de structure du melange gaussien modélisant un système complexe peut avoir pour origine essentiellement deux causes.

La première cause vient du fait que tout système complexe peut , au cours de son évolution , voir sa structure se modifier d'une manière tout à fait naturelle.

La seconde cause provient tout simplement du fait de l'utilisation d'une base d'apprentissage initialement incomplète.

Les critères utilisés pour mettre en évidence ce changement de structure sont basés principalement sur l'étude de la structure interne de chaque composante entrant dans le MG.

Les paragraphes 2,3 et 4 sont consacrés à la présentation de quelques critères employés dans cette partie.

le premier critère , basé sur l'étude de la dispersion des observations autour des axes principaux d'inertie , sert surtout à mettre en évidence la scission des composantes.

Le second critère , basé sur la définition d'un contour de confiance englobant un pourcentage déterminé d'observations appartenant à la même composante , permet de déceler l'apparition de toute nouvelle composante.

Le troisième critère , développé par SMOLART [SMO,82], s'appui sur la définition d'un contour de confiance qui englobe les nouveaux centres de masse X^{+}.

Les critères 2 et 3 , comme on le verra par la suite , sont tout à fait équivalents.

IV-2 MESURE DE LA DISPERSION DES COMPOSANTES

L'objectif de la reconnaissance des formes est de tirer, à partir d'un nuage d'observations servant de base d'apprentissage, des formes représentatives de la base considérée.[WAT,81]

fig.3 nuage d'observations

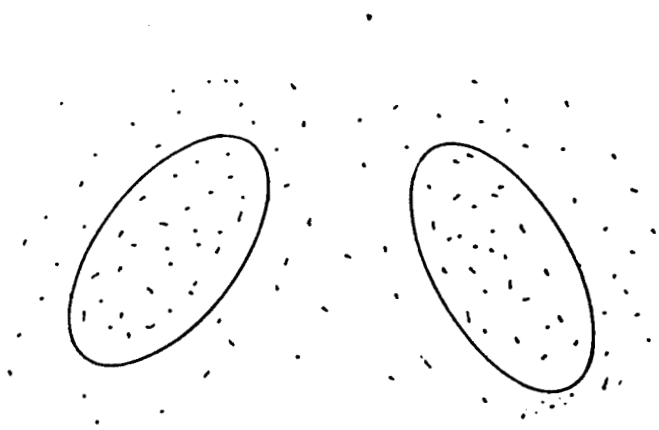


fig.4 formes retenues à partir du nuage d'observations

L'estimation récursive des paramètres statistiques développée dans le chapitre II permet d'améliorer l'estimation de ces paramètres et par conséquent d'améliorer la représentation spatiale des formes retenues.

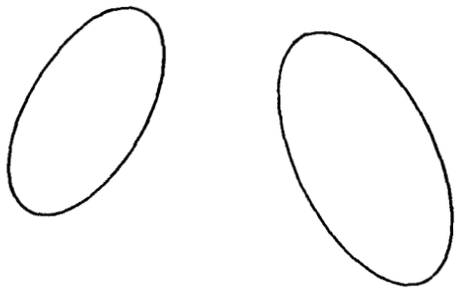
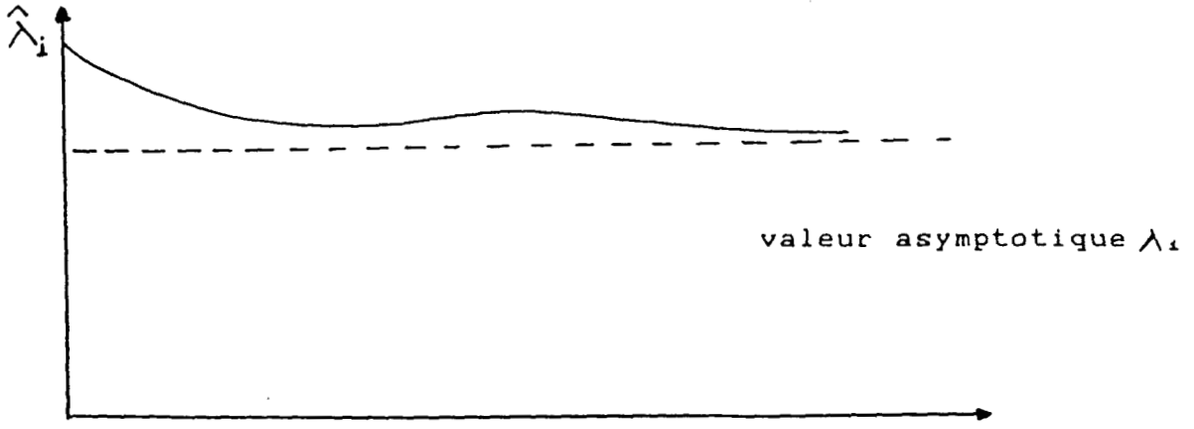


fig.5 concentration des formes dans des régions distinctes

Dans le cas où le choix des formes tirées du nuage d'observations est confirmé par l'intégration des nouvelles observations, les paramètres statistiques tels que les valeurs propres estimées $\hat{\lambda}_i$ qui représentent la dispersion des observations autour des directions propres U_i tendent asymptotiquement vers les paramètres réels λ_i définissant les formes retenues.



intégration séquentielle des observations

fig.6 évolution asymptotique des paramètres

Mais dans le cas où les nouvelles observations ne confirment pas le choix des formes initialement retenues, la dispersion des observations autour des directions propres se met à varier et les paramètres de la composante considérée évoluent vers d'autres valeurs asymptotiques.

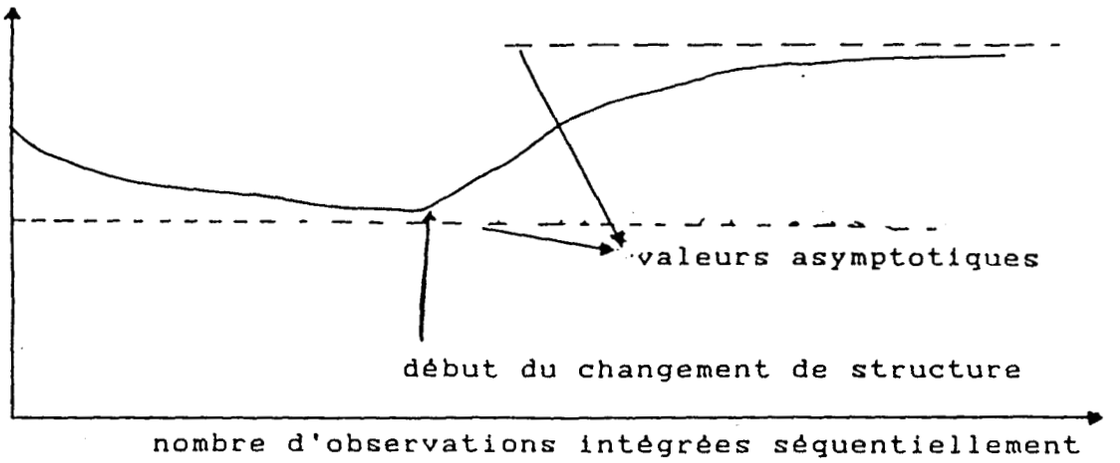


fig.7 évolutions asymptotiques des paramètres

La figure 8 illustre cette situation dans le cas de l'éclatement d'une composante en deux.

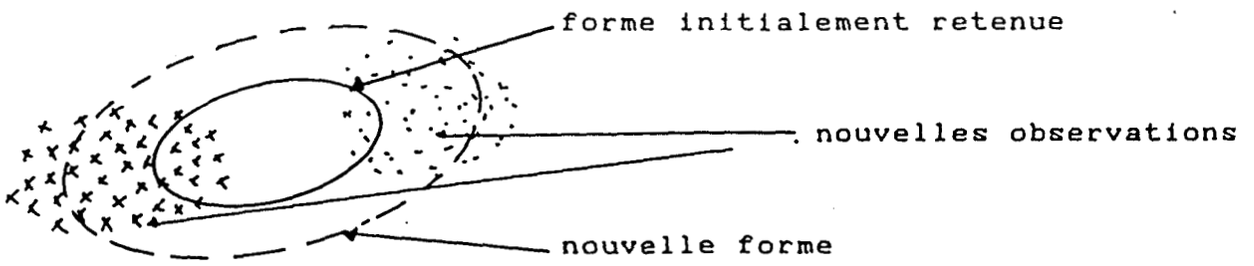


fig.8 évolution de la forme en fonction de l'intégration des observations

En analysant les figures 7 et 8 , on peut conclure qu'une nouvelle forme est obtenue lorsque les paramètres statistiques λ_i atteignent leurs nouvelles valeurs asymptotiques.

Ainsi le suivi de l'évolution des paramètres λ_i qui traduisent la dispersion des observations autour des directions propres est utilisé pour mettre en évidence l'apparition de nouvelles formes à partir de celles retenues initialement.

Ce critère est surtout utilisé pour déceler le début de scission d'une composante.

Le cas du glissement naturelle d'une composante est à distinguer de celui de la scission.Dans le premier cas les observations à intégrer sont situées en majorité d'un même coté de la composante considérée et ce phénomène se traduit par un glissement de son centre de masse; par contre,dans le second cas,les nouvelles observations sont distribuées de part et d'autre de la composante de telle sorte que son centre de masse ne glisse pas.

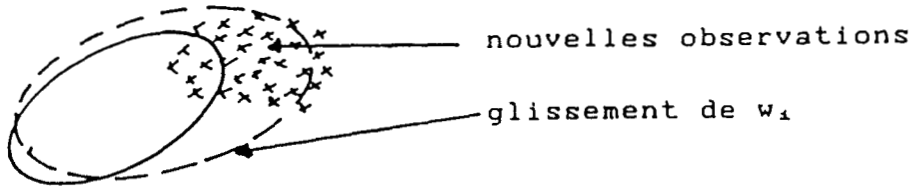


fig.9 glissement naturel d'une composante

IV-3 DEFINITION D'UN CONTOUR DE CONFIANCE

IV-3-1 INTRODUCTION

On définit ,pour une composante donnée w , des contours Dw(K) homothétiques dépendant du paramètre K et pouvant englober entre 0% et 100% des observations appartenant à la composante w .[LEH,59]

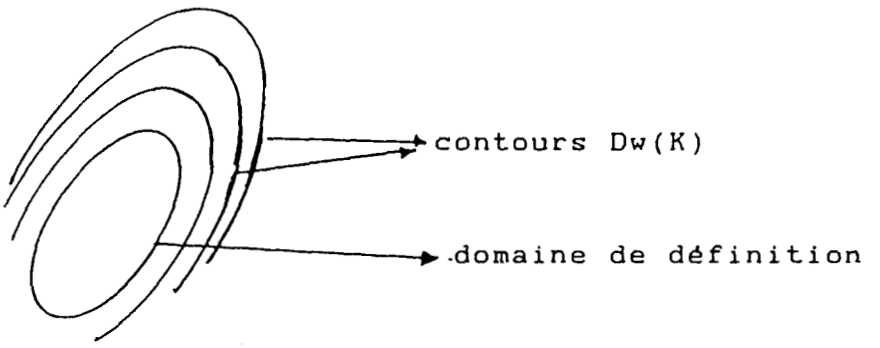


fig.10 définition de contours englobant un pourcentage déterminé d'observations

IV-3-2 CAS OU LE NOMBRE D'OBSERVATIONS EST GRAND

Soit { X^i } i=1,N l'ensemble des observations ayant servi à construire une composante w .

Dans le cas où le nombre N est grand,YOUNG dans [YOU,78] a montré que les contours Dw(K) suivent une distribution du χ^2 centrée à d degrés de liberté.

Pour atteindre ce résultat,il applique à toutes les observations X^i une transformation linéaire ,ce qui permet de définir une nouvelle variable Y^i.

$$Y^i = W(X^i - \overline{X}) \qquad (1)$$

Le centre de masse et la matrice de covariance des variables Y^i sont donnés respectivement par les expressions (2) et (3).

$$\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} W(X^i - \overline{X}) = \frac{1}{N} W \sum_{i=1}^{N} (X^i - \overline{X}) = 0 \qquad (2)$$

$$\Sigma_Y = \frac{1}{N-1} \sum_{i=1}^{N} Y^i . Y^{iT} = \frac{1}{N} \sum_{i=1}^{N} W(X^i - \overline{X}).(X^i - \overline{X})^T W^T$$

$$= W \frac{1}{N-1} \sum_{i=1}^{N} (X^i - \overline{X}).(X^i - \overline{X})^T W^T$$

$$= W \Sigma_X W^T \qquad (3)$$

Pour les variables X^i ,le contour de valeur K est défini par les expressions suivantes:

$$Dw(K) = \{ X / (X - \overline{X})^T \Sigma_X^{-1} (X - \overline{X}) = K \} \qquad (4)$$

$$= \{ X / (W^{-1} Y)^T \Sigma_X^{-1} (W^{-1} Y) = K \}$$

Ce qui est équivalent à :

$$Dw(K) = \{ Y / (Y)^T (W^{-1})^T \Sigma_X^{-1} (W^{-1}) Y = K \} \qquad (5)$$

A partir des expressions (4) et (5) on tire l'égalité suivante:

$$Dw_1(K) = \{ X, Y / (X - \bar{X})^T \Sigma_X^{-1} (X - \bar{X}) = (Y)^T \Sigma_Y^{-1} Y = K \} \quad (6)$$

La matrice de transformation W est choisie de la manière suivante :

$$W = [\lambda]^{-1/2} [U] \qquad (7)$$

Les matrices $[\lambda]$ et $[U]$ sont tirées de Σ_X et correspondent respectivement aux matrices de valeurs et vecteurs propres.

Dans ces conditions la matrice Σ_Y prend la forme suivante:

$$\Sigma_Y = W \Sigma_X W^T = [\lambda]^{-1/2} [U] \Sigma_X [U]^T [\lambda]^{-1/2}$$

$$= [I] \qquad (8)$$

Finalement le contour K suit une loi de distribution du χ^2 centrée à d degrés de liberté.

$$Dw(K) = \{ Y / (Y)^T [I] Y = \sum_{j=1}^{d} (Y_j)^2 = K \} \qquad (9)$$

Dans l'expression (9), les Y_j représentent les composantes de la variable Y.

La transformation W permet de réduire les domaines Dw(K) à des hypersphères centrées à l'origine et de rayon $\sqrt{K}$.

A partir des propriétés d'une distribution du χ^2 centrée avec d degrés de liberté on tire les égalités suivantes:

$$E\ [\ Dw(K)\ ]=d \qquad (10)$$

$$VAR\ [\ Dw(K)\ ]=2d \qquad (11)$$

Le contour standard est donné par l'expression (12)

$$Dw(K)= E\ [\ Dw(K)\ ] + \left(VAR\ [\ Dw(K)\ ]\right)^{1/2}$$

$$=d + \sqrt{2d} \qquad (12)$$

La figure 11 représente une distribution du $\chi^2(\alpha)$.

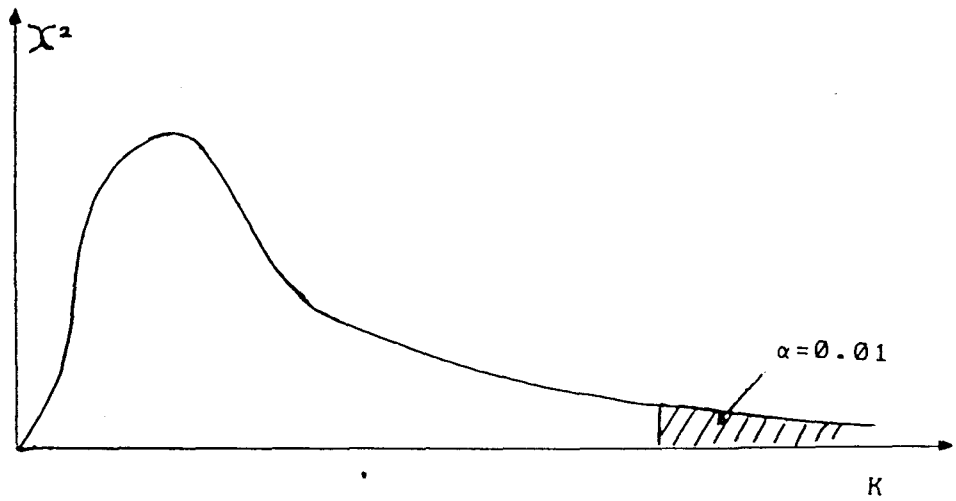


fig.11 distribution du $\chi^2(\alpha)$

Dans le cas classique de l'hypothèse Ho,c'est à dire que toutes les observations appartiennent à la même distribution normale,l'erreur de première espèce représentée par α traduit la probabilité de refuser de classer la nouvelle observation dans une composante donnée alors qu'elle lui appartient.

En imposant à α la valeur 0.01 , on détermine le contour Dw* qui englobe 99% des observations appartenant à la même distribution normale.

$$Dw^* = \chi_d^2\ (\alpha=0.01) \qquad (13)$$

Par exemple, pour un espace d'observation à trois dimensions (d=3) et pour α fixé à 0.01 (α=0.01) , la valeur du contour de confiance est fournie par :

$$Dw^{*}=\chi_{3}^{2}(\alpha=0.01)=11.34$$

Si l'observation X à classer est portée par un contour :

$$Dw(K)=(X-\bar{X})^{T}\Sigma_{X}^{-1}(X-\bar{X})=K \text{ ,tel que } K>Dw \quad \text{alors}$$

l'observation X est rejetée avec une probabilité d'erreur inférieure à 0.01 ;dans le cas contraire ,correspondant à $K<Dw^{*}$, l'observation X est intégrée dans la composante.

IV-3-3 CAS OU LE NOMBRE D'OBSERVATIONS EST FAIBLE

Dans le cas où le nombre d'observations N est faible les contours Dw(K) suivent une distribution de FISCHER à (d,N-d) degrés de liberté : $F_{d,N-d}(\alpha)$.

Pour cela, on utilise le résultat obtenu par ANDERSON dans [AND,58] ,qui stipule que la variable scalaire définie comme suit :

$$T^2 = P^T \Sigma^{-1} P \qquad (14)$$

(la variable P suit une distribution normale centrée identique à celle des observations X^i ,les variables X^i et P appartiennent à des sous-ensembles disjoints).

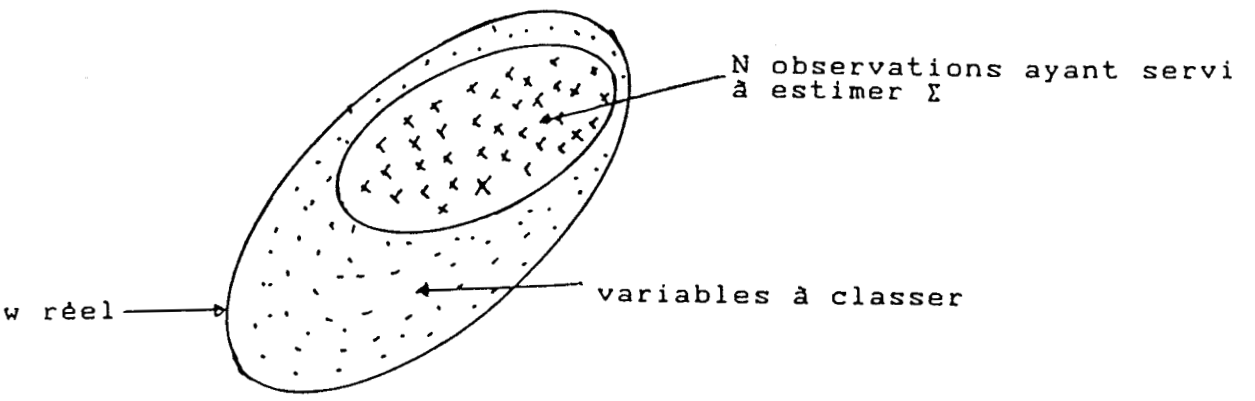


fig.12 représentation des variables X et P

a une distribution de FISCHER avec (d,N -d) degrés de liberté.

$$T^2 \frac{(N-d)}{(N-1).d} = F_{d,N-d}(\alpha) \qquad (15)$$

Anderson a montré que la région critique est donnée par l'expression suivante:

$$T^2 \geq \frac{d.(N-1)}{(N-d)} F_{d,N-d}(\alpha) \qquad (16)$$

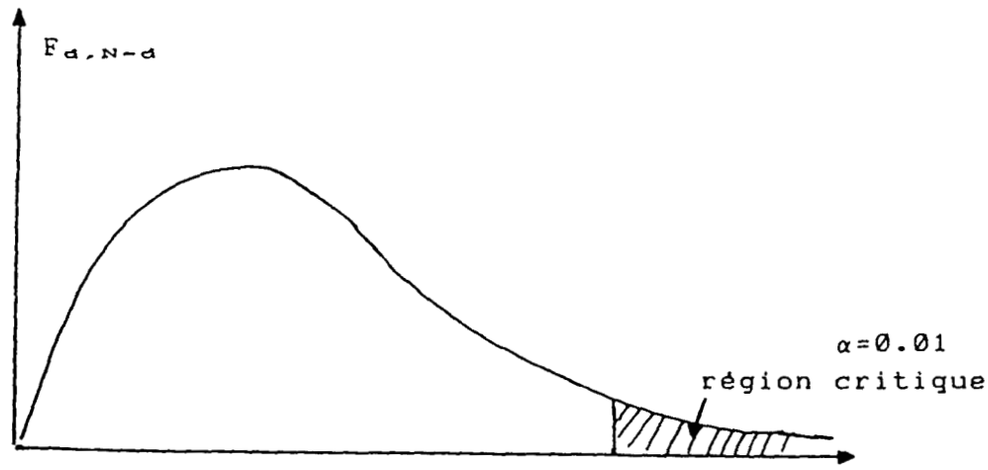


fig. 13 distribution de FISCHER

$F_{d,N-d}(\alpha)$ représente la valeur d'une variable de FISCHER centrée pour l'erreur de première espèce avec α fixé à 0.01.

En prenant la variable P égale à :

$$P = \left(\frac{N}{(N+1)}\right)^{1/2} (X - \overline{X}) \qquad (17)$$

BU LILLE

L'egalité (14) devient :

$$T^2 = \frac{N}{N+1} (X-\bar{X})^T \Sigma^{-1} (X-\bar{X}) \qquad (18)$$

$$T^2 = \frac{N}{N+1} Dw(K) \qquad (19)$$

A partir des propriétés d'une distribution de FISCHER on tire les résultats suivants :

$$E [Dw(K)] = \frac{N+1}{N} E [T^2] \qquad (20)$$

Et en utilisant la relation (16) on obtient :

$$E [Dw(K)] = \frac{(N+1)}{N} \cdot \frac{(N-1)}{N-d} . d . E [F_{d,N-d}(\alpha)] \qquad (21)$$

$$VAR [Dw(K)] = \frac{(N+1)^2}{N} VAR [T^2] \qquad (22)$$

$$= \frac{(N+1)^2}{(N)^2} \cdot \frac{(N-1)^2}{(N-d)^2} . d^2 . VAR [F_{d,N-d}(\alpha)] \qquad (23)$$

Avec $E [F_{d,N-d}(\alpha)] = \frac{N-d}{N-d-2}$ pour $N>d+2$ (24)

$$VAR [F_{d,N-d}(\alpha)] = \frac{2 (N-d)^2 . (N-2)}{d . (N-d-2)^2 . (N-d-4)} \quad \text{pour } N>d+4 \qquad (25)$$

Pour déterminer le contour Dw^* qui englobe 99% des observations qui appartiennent à la composante w il suffit de prendre pour α une valeur égale à 0.01.

$$Dw_i^* = \frac{(N+1)\ (N-1)}{N\ (N-d)}\ d\ F_{d,N-d}(\alpha=0.01) \qquad (26)$$

Par exemple pour N=100 et d=6 on obtient un contonr de confiance égal à 19.15 ; Il constitue en fait un seuil raisonnable à partir duquel on peut accepter ou rejeter avec une probabilité d'erreur inférieure à 0.01 , toute nouvelle observation.

Ce seuil est évidemment le même pour toutes les composantes ayant été construites à partir de 100 observations et évoluant dans un espace d'observation de dimension égale à 6

IV-4 DEFINITION D'UN CONTOUR DE CONFIANCE POUR LES NOUVEAUX CENTRES DE MASSE $\overline{X}^+$

D'une manière identique SMOLART dans [SMO,82] a défini un contour de confiance pour les nouveaux centres de masse $\overline{X}^+$.

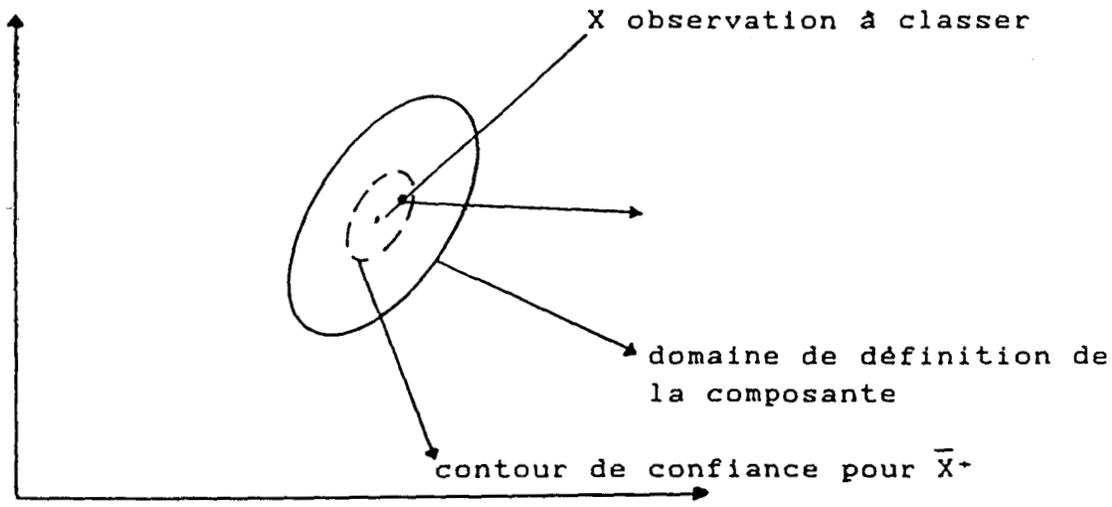


fig.14 contour de confiance pour $\overline{X}^+$

Si le nouveau centre $\overline{X}^+$,obtenu aprés l'intégration de l'observation X,se trouve à l'intérieur du contour de confiance alors l'observation X est acceptée;Dans le cas contraire elle suit une procédure qui sera détaillée dans les paragraphes suivants.

La relation entre les différents contours de confiance est facile à mettre en évidence.

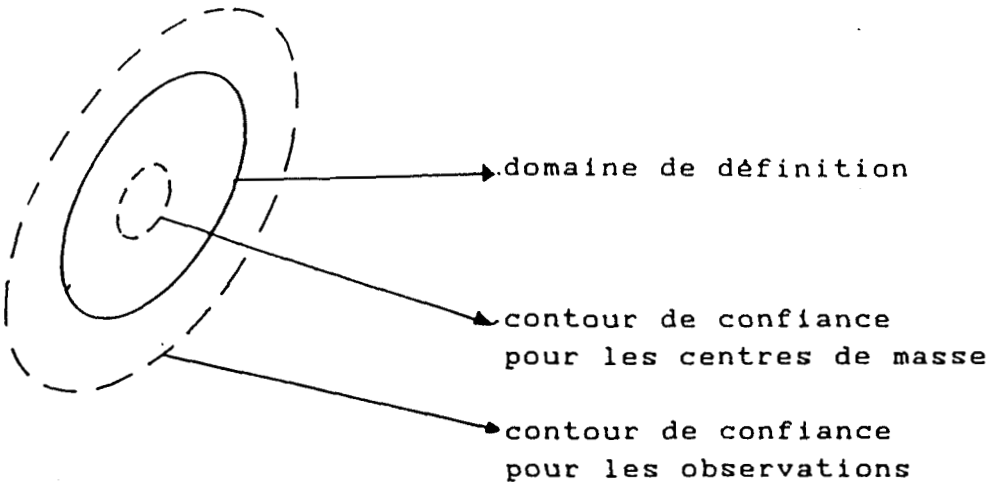


fig.15 relation entre les contours de confiance

Si une observation à classer se trouve à l'intérieur du contour de confiance qui englobe 99% des observations ,le nouveau centre de masse $\overline{X}^+$ est naturellement à l'intérieur du contour de confiance qui lui correspond.

La relation entre les deux contours de confiance est fournie par les expressions suivantes:

$$(N+1)\overline{X}^+ = N\overline{X} + X \qquad (27)$$

X^+ suit une distribution normale caractérisée par un centre de masse X et une matrice de covariance Σ_o égale à $\Sigma/(N+1)$.

$$Dw(K) = (X-X)^T \Sigma^{-1} (X-X) \qquad (28)$$

$$Dw^o(K) = (X^+-X)^T \Sigma_o^{-1} (X^+-X) \qquad (29)$$

En tirant de l'expression (27) , la relation $(X-X)=(N+1)(X^+-X)$ et en la plaçant dans (28) on obtient :

$$Dw(K) = (N+1)^2(X^+-X)^T \Sigma^{-1}(X^+-X)$$

$$\text{avec} \quad \Sigma^{-1}=\Sigma_o^{-1}/(N+1) \qquad (30)$$

Dw(K) devient :

$$Dw(K)=(N+1)^2(X^+-X)^T \Sigma_o^{-1} (X^+-X)/(N+1)$$

$$Dw(K)=(N+1)Dw^o(K)$$

Et finalement on a:

$$Dw^* = (N+1)\, Dw_1^{*o}$$

Dw^* et Dw^{*o} représentent respectivement les contours qui portent les nouveaux centres de masse et les contours de confiance associés à ces derniers.

IV-5 PROCEDURE A SUIVRE DANS LE CAS OU LE NOMBRE DE COMPOSANTES D'UN MELANGE GAUSSIEN AUGMENTENT

IV-5-1 INTRODUCTION

Dans le cas où l'application des critères développés précédemment s'avère positive , ce qui correspond soit à une grande variation de la dispersion autour des directions propres dans le cas d'une scission d'une composante en deux ou bien à une augmentation du nombre d'observations se trouvant à l'extérieur du domaine de confiance dans le cas de l'apparition d'une nouvelle composante , la procédure à suivre dépend du cas considéré.

Dans le cas de scission d'une composante , celle ci disparait en donnant naissance à deux composantes choisies à priori identiques avant d'avoir atteint leur nouvelles formes.

Dans le cas de l'apparition d'une nouvelle composante , on associe à chaque composante w_i concernée par ce changement de structure une composante w_i' dans laquelle sont intégrées , en plus des observations appartenant à w_i , toutes les observations ayant été attribuées à w_i et se trouvant en dehors de son domaine de confiance.

chaque composante w_i' contient alors deux distributions normales distinctes.

IV-5-2 APPARITION D'UNE NOUVELLE COMPOSANTE

Tout en associant à chaque composante w_i impliquée dans le changement de structure une composante w_i', on observe les règles suivantes dans les décisions d'attribution des nouvelles observations:

- on continue d'utiliser les paramètres de la composante w_i dans la fonction de classement .

-chaque fois qu'une observation est classée dans w_i,les paramètres de w_i' sont systématiquement adaptés selon les équations récurrentes développées dans le chapitre II,alors que les paramètres de w_i ne sont adaptés que lorsque l'observation se trouve à l'intérieur de son domaine de confiance.

$$\left.\begin{array}{l} X \text{ observation classée dans } w_i \\ \qquad\qquad \text{et} \\ (X-\bar{X}^i)^T \Sigma_i^{-1} (X-\bar{X}^i) = K > Dw_i^{\alpha} \end{array}\right] \Longrightarrow \begin{array}{l}\text{seuls les paramètres} \\ \text{de } w_i' \text{ sont ajustés}\end{array}$$

$$\left.\begin{array}{l} X \text{ observation classée dans } w_i \\ \qquad\qquad \text{et} \\ (X-\bar{X}^i)^T \Sigma_i^{-1} (X-\bar{X}^i) = K < Dw_i^{\alpha} \end{array}\right] \Longrightarrow \begin{array}{l}\text{les paramètres de } w_i \\ \text{et } w_i' \text{sont ajustés}\end{array}$$

Finalement ,la composante w_i' va contenir deux distributions normales ; la première a servi à construire w_i et la seconde représente la nouvelle composante.

La figure 16 représente les deux distributions de w_i'

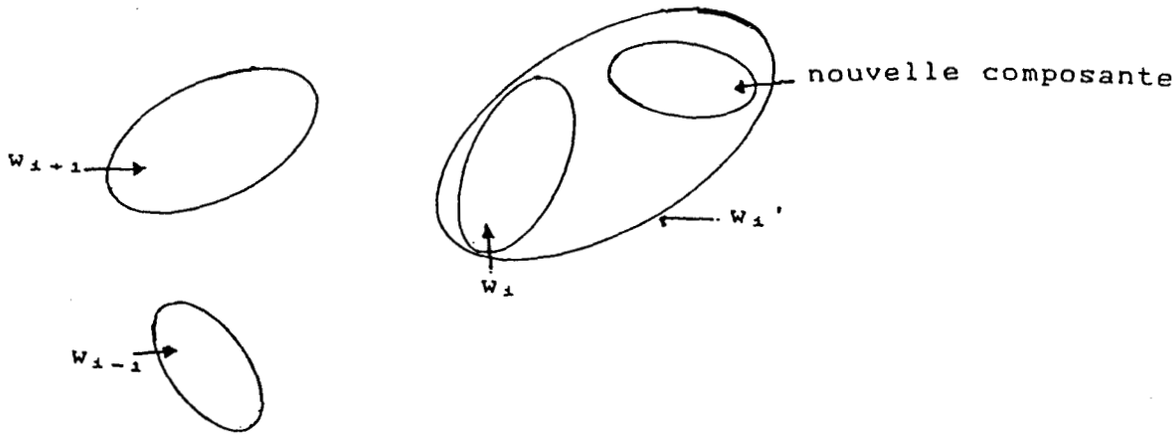


fig.16 représentation spatiale de la nouvelle composante

Dans certains cas simples , la procédure adoptée permet , en tenant compte d'une part des informations apportées par les observations non classées et d'autre part de la disposition spatiale des composantes déja existantes , de fournir les paramètres de la nouvelle composante .

Cette démarche est préférable à celle qui consiste à rejeter tout simplement les observations se trouvant à l'extérieur des différents domaines de confiance en les stockant dans une banque de données et de refaire la classification automatique lorsque le nombre de ces dernières atteint un certain seuil.

Dans les cas difficiles , cette démarche permet de fournir des informations a priori sur les nouvelles composantes.

Dans le cas où la décision de créer une nouvelle composante est prise , les paramètres de celle-ci sont fournis par les M_i observations attribuées à w_i et se trouvant en dehors de son domaine de confiance.

Le paramètre M_i est important dans la mesure où il représente le nombre minimal d'observations à partir duquel on peut construire une composante de manière fiable.[REE,74]

Les relations (31) et (32) , conformément aux calculs éffectués dans le paragraphe précédent , fournissent les valeurs prises par le contour moyen en fonction du nombre d'observations.

$$\text{si } M_i \longrightarrow \infty \qquad E\,[\,Dw_i(K)\,] = d \tag{31}$$

$$\text{si } M_i \text{ est faible } E\,[\,Dw_i(K)\,] = d.\frac{(M_i+1)(M_i-1)}{M_i(M_i-d-2)} \tag{32}$$

Le nombre minimal d'observations nécéssaire pour construire une composante est déterminé en fonction du risque β de deuxième éspece . Ce risque est lié à la probabilité d'accepter l'hypothèse Ho alors qu'elle est fausse.

$$\lim_{M_i \to \infty} E\,[\,Dw_i(K)\,] = d \qquad (33)$$

$$\lim_{M_i \to \infty} d\frac{(M_i+1)(M_i-1)}{M_i(M_i-d-2)} = d \qquad (34)$$

L'expression (35) donne le nombre d'observations nécessaire pour obtenir le contour moyen variant par rapport à sa vraie valeur d avec un pourcentage égale à 100β , β étant fixé.

$$E\,[\,Dw_i(K)\,] = d.\frac{M_i+1}{M_i}.\frac{M_i-1}{M_i-d-2} = d.(1 \pm \beta) \qquad (35)$$

En considérant la limite supérieure de l'expression (35) on obtient:

$$\frac{(M_i+1)}{M_i}.\frac{(M_i-1)}{(M_i-d-2)} = 1+\beta \qquad (36)$$

Ceci nous permet d'avoir une relation fournissant le paramètre M_i en fonction de (d et β)

$$M_i = \frac{(1+\beta)(d+2) \pm \sqrt{(1+\beta)^2(d+2) - 4\beta}}{2\beta} \qquad (37)$$

Pour des faibles valeurs de β ,une approximation de (34) est fournie par l'expression suivante :

$$M_i = (1+\beta^{-1})(d+2) \qquad (38)$$

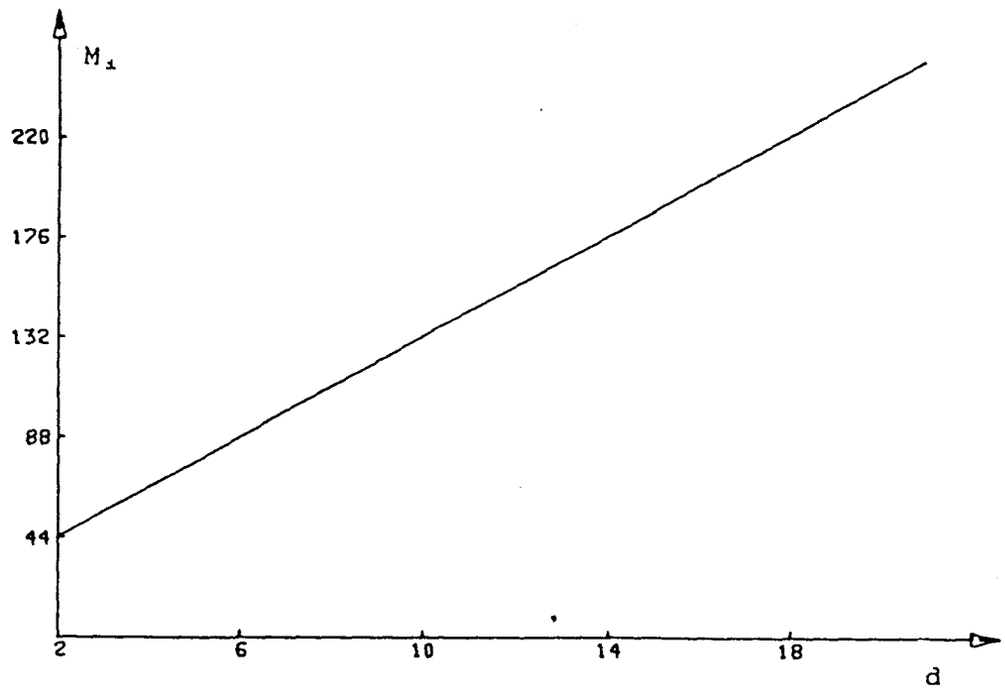


fig.17 nombre d'observations en fonction de d

Si les composantes w_i et w_i' ont les paramètres suivants:

	w_i	w_i'
centre de masse	$\overline{X}^i$	$\overline{X}^i$
matrice de covariance	Σ_i	Σ_i'
nombre d'observations	N_i	$N_i + M_i$

BU LILLE

alors la nouvelle composante est définie par les paramètres suivants

centre de masse $$\overline{X}^{i''} = \frac{(N_i+M_i).\overline{X}^{i'} - N_i.\overline{X}^{i}}{M_i} \qquad (39)$$

matrice de covariance $$\Sigma_i'' = \frac{(N_i+M_i).\Sigma_i' - N_i.\Sigma_i}{M_i} \qquad (40)$$

probabilité à priori $$P^{i''} = \frac{M_i}{M_i + \sum_{j=i}^{C} N_j} \qquad (41)$$

nombre d'observations M_i

Toutes les probabilités a priori des composantes déja existantes sont révisées en baisse; elles sont multipliées par un coefficient inférieur à 1 donné par l'expression suivante:

$$COEFF = \frac{\sum_{j=1}^{C} N_j}{M_i + \sum_{j=1}^{C} N_j} \qquad (42)$$

C représente le nombre de composantes initiales

IV-5-3 QUELQUES SITUATIONS PARTICULIERES

Il peut arriver que la nouvelle distribution soit partagée entre deux ou plusieurs composantes .

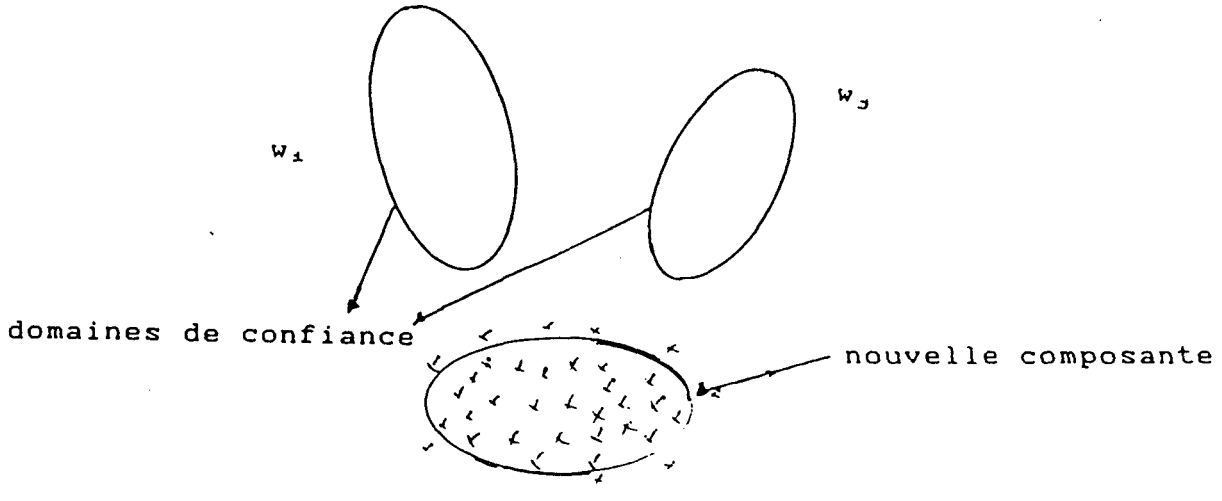


fig.18 nouvelle distribution partagée entre plusieurs composantes.

La procédure à observer est évidemment la même que celle détaillée précédemment;dés que les composantes w_i'' et w_j'' sont crées elles vont aussitôt fusionner.

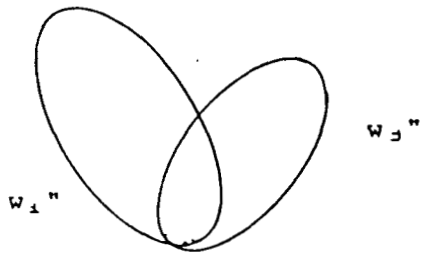


fig.19 fusionnement des composantes nouvellement crées

Naturellement dans le cas où une nouvelle distribution est partagée entre plusieurs composantes, sa détection est plus longue à mettre en évidence.

On peut aussi envisager le cas où deux distributions nouvelles apparaissent simultanément et soient situées prés d'une même composante.

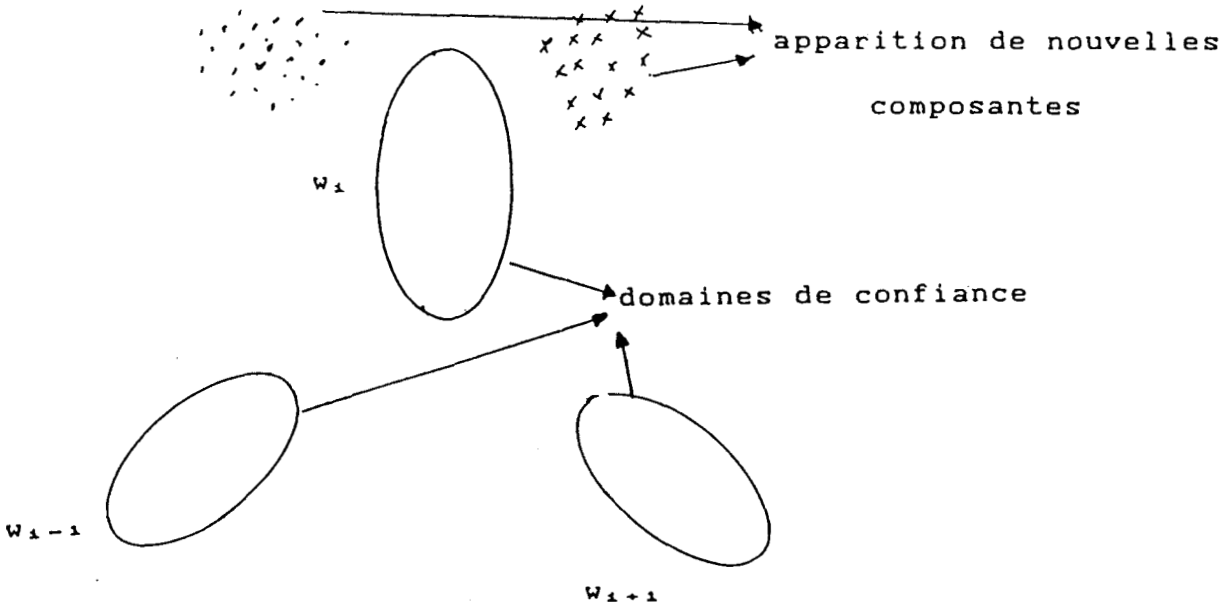


fig.20 apparition simultanée de deux nouvelles composantes situées prés de w_i

IV-5-4 ECLATEMENT D'UNE COMPOSANTE EN DEUX

La procédure à suivre dans ce cas est de diviser la composante considérée w en deux composantes a priori identiques ayant mêmes matrices de covariance et des centres de masse symétriques par rapport à celui de w et portés par le plus grand axe d'inertie .

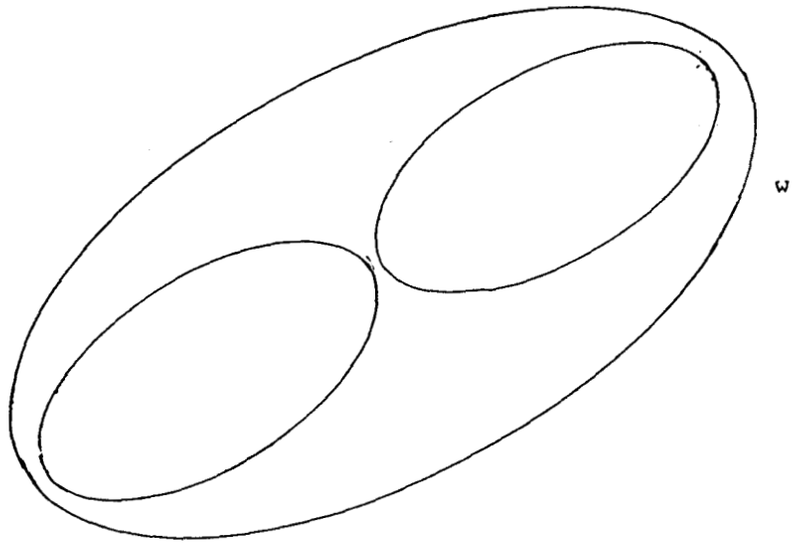


fig.21 division d'une composante en deux

BU LILLE

Le fait de diviser la composante w_1 en deux composantes a priori identiques nous place dans le cas limite entre le fusionnement et l'éclatement.

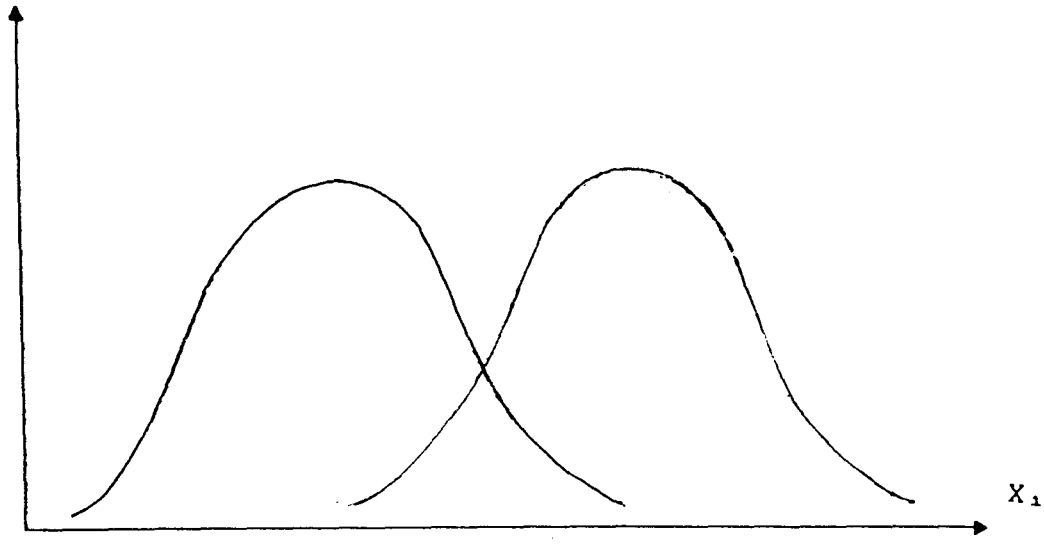


fig.22 cas limite entre le fusionnement et la scission.

Si l'hypothèse de la scission est confirmée par l'intégration des nouvelles observations , les deux composantes issues de w vont de plus en plus s'éloigner l'une de l'autre.

Les paramètres des deux nouvelles composantes sont définis comme suit:

Les deux centres de masse sont portés par la direction propre qui correspond au plus grand axe d'inertie de w_1.

$\bar{X}^1 \pm\sqrt{\lambda_1}U_1/2$

Les deux matrices de covariance sont égales et valent:
$\Sigma_1/2$

Les deux probabilités sont égales et valent:

$P^1/2$

λ_1 et U_1 correspondent au plus grand axe d'inertie

En plus de l'utilisation du premier critère présenté dans ce chapitre et qui permet de suivre la variation de la dispersion des observations ,la comparaison permanente des contours moyens théorique et calculé fournis par les deux expressions ci-dessous offre des informations interréssantes pour la mise en évidence de la scission d'une composante en deux.

$$Dw(K)_{moy.cal} = \frac{1}{N} \sum_{i=1}^{N} (X^i - \bar{X})^T \Sigma^{-1} (X^i - \bar{X})$$

$$Dw(K)_{moy.theo} = \frac{N+1}{N} \frac{N-1}{N-d-2} .d$$

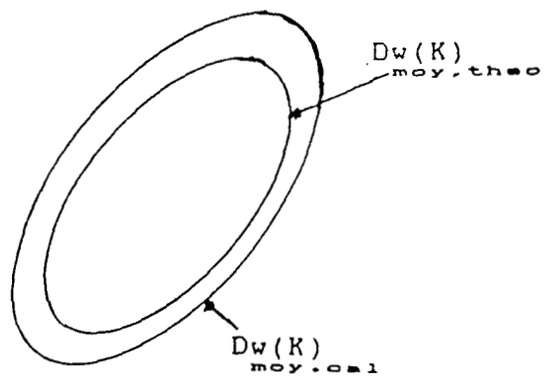


fig.23 comparaison des contours moyens calculé et théorique

En utilisant les trois expressions suivantes ,on arrive à determiner le contour moyen calculé d'une manière itérative en fonction de l'arrivée des observations.

$$Dw^{+}(K)_{moy.cal} = \frac{1}{N+1} \sum_{i=1}^{N+1} (X^{i} - \overline{X}^{+})^{T} \Sigma^{-1+} (X^{i} - \overline{X}^{+})$$

$$\overline{X}^{+} = \overline{X} - \frac{1}{N+1}(X^{N_1+1} - \overline{X})$$

$$\Sigma^{-1+} = \frac{N\ \Sigma^{-1}}{N-1} - \frac{A.\Sigma^{-1}(X^{N+1}-\overline{X})(X^{N+1}-\overline{X})^{T}\Sigma^{-1}}{1 + A.(X^{N+1}-\overline{X})^{T}\ \Sigma^{-1}\ (X^{N+1}-\overline{X})}$$

$$\text{avec } A = \frac{(N)^{2}}{(N)^{2} - 1}$$

La démonstration de la dernière expression se trouve dans [SMO,82].

En remplaçant $\overline{X}^{+}$ et Σ^{-1+} par leurs expressions on obtient $Dw^{+}(K)_{moy.cal}$ en fonction de $Dw(K)_{moy.cal}$

$$Dw^{+}(K)_{moy.cal} = A.Dw(K)_{moy.cal} - \frac{B.(((X^{N+1}-\overline{X})^{T}\ \Sigma^{-1}\ (X^{N+1}-\overline{X}))^{2}}{1+A.(X^{N+1}-\overline{X})^{T}\ \Sigma^{-1}\ (X^{N+1}-\overline{X})} - \frac{C.(X^{N+1}-\overline{X})^{T}\ \Sigma^{-1}\ (X^{N+1}-\overline{X})}{1+A.(X^{N_1+1}-\overline{X})^{T}\ \Sigma^{-1}\ (X^{N_1+1}-\overline{X})}$$

$$B = \frac{2(N)^{5}}{(N+1)^{5}(N-1)^{2}} \quad ; \quad C = \frac{(N)^{2}}{(N+1)^{2}} \cdot \left(\frac{1}{(N+1)^{2}} - \frac{N}{(N+1)^{2}} \right)$$

la démonstration de la dernière expression se trouve dans l'annexe 2

IV-6 EXEMPLES D'APPLICATION

IV-6-1 APPARITION D'UNE NOUVELLE COMPOSANTE

L'exemple présenté dans cette partie est effectué dans un espace à deux dimensions . On génère artificiellement un mélange gaussien constitué de deux composantes , et on tire à partir d'un ensemble e_1 contenant 200 observations les caractéristiques statistiques des différentes composantes.

Le tableau suivant résume les caractéristiques statistiques obtenues .

	vecteur moyenne	écarts types	direction propre	probabilité a priori
comp.1	$\overline{X}_1=154.5$ $\overline{X}_2=104.4$	$\lambda_1=9.5$ $\lambda_2=5.7$	$U=-9°$	$P_1=0.42$
comp.2	$\overline{X}_1=205.7$ $\overline{X}_2=123.6$	$\lambda_1=11.4$ $\lambda_2=6.6$	$U=49°$	$P_2=0.58$

Une troisième composante , située près de la deuxième composante comme le montre la figure 24 , est crée.

Ces caractéristiques sont les suivants:

vecteur moyenne	écarts types	direction propre
$\overline{X}_1=200$ $\overline{X}_2=190$	$\lambda_1=6$ $\lambda_2=4$	$U=45°$

BU LILLE

On prend un ensemble e_2 des observations , issues des trois composantes précédemment présentées , qu'on intègre dans le mélange gaussien initial.

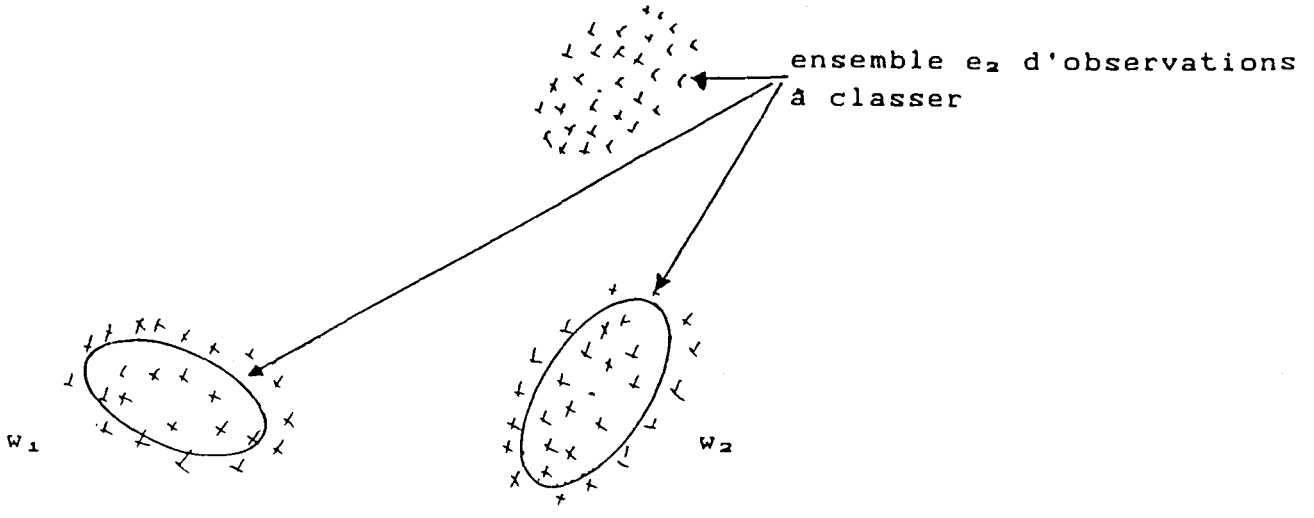


fig.24 représentation du MG et de l'ensemble e_2

Aprés l'intégration de quelques observation tirées de l'ensemble e_2 , la composante w'_2 associée à w_2 est crée . Et lorsque le nombre d'observations M_2 attribué à w_2 et se trouvant en dehors de son domaine de confiance atteint le seuil fixé alors une nouvelle composante est crée.

Ses caractéristiques statistiques sont les suivantes:

vecteur moyenne	écarts types	direction propre
$\overline{X}_1=200.7$ $\overline{x}_2=188.9$	$\lambda_1=3.8$ $\lambda_2=3.5$	U= 80°

IV-6-2 ECLATEMENT D'UNE COMPOSANTE EN DEUX

On génère artificiellement un mélange gaussien composé de deux composantes dont les caractéristiques statistiques , tirées à partir d'un ensemble e_3 , sont consignées dans le tableau suivant.

	vecteur moyenne	écarts types	direction propre	probabilité a priori
comp.1	$\overline{X}_1=103.6$ $\overline{X}_2=207$	$\lambda_1=11.3$ $\lambda_2=7.4$	U=-30°	$P_1=0.55$
comp.2	$\overline{X}_1=207.5$ $\overline{X}_2=103.7$	$\lambda_1=9.1$ $\lambda_2=7.4$	U=-45°	$P_2=0.45$

A ce mélange gaussien , on ajoute deux autres composantes situées de part et d'autre de la première composante et se trouvant à l'intérieur de son domaine de confiance.

	vecteur moyenne	écarts types	direction propre
comp.1	$\overline{X}_1=115$ $\overline{X}_2=185$	$\lambda_1=6$ $\lambda_2=4$	U=-30°
comp.2	$\overline{X}_1=85$ $\overline{X}_2=215$	$\lambda_1=6$ $\lambda_2=4$	U=-30°

la figure 25 donne une image sur la disposition spatiale des différentes composantes.

BU LILLE

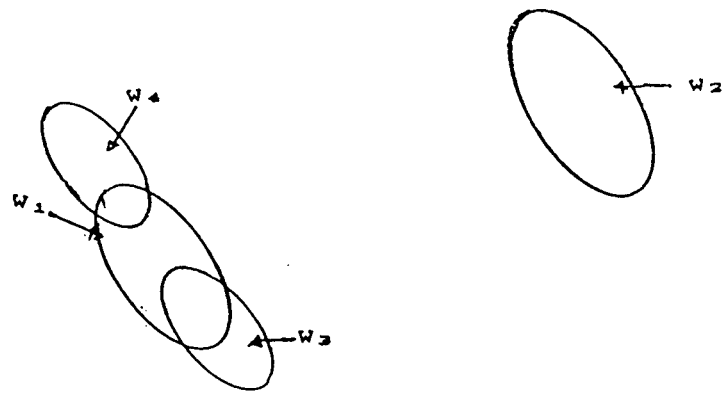


fig.25 disposition spatiale des différentes composantes.

Les figures 27 a et b représentent l'évolution de l'estimation des valeurs propres de la première composante.

On note que pour les 100 premières observations intégrées appartenant effectivement à la première composante , les paramètres estimés tels que valeurs propres évoluent vers les paramètres réels définissant cette dernière.

Au delà des 100 premières observations , les paramètres estimés évoluent vers d'autres valeurs asymptotiques que ceux de la première composante du fait que les observations intégrées appartiennent aux composantes 2 et 3.

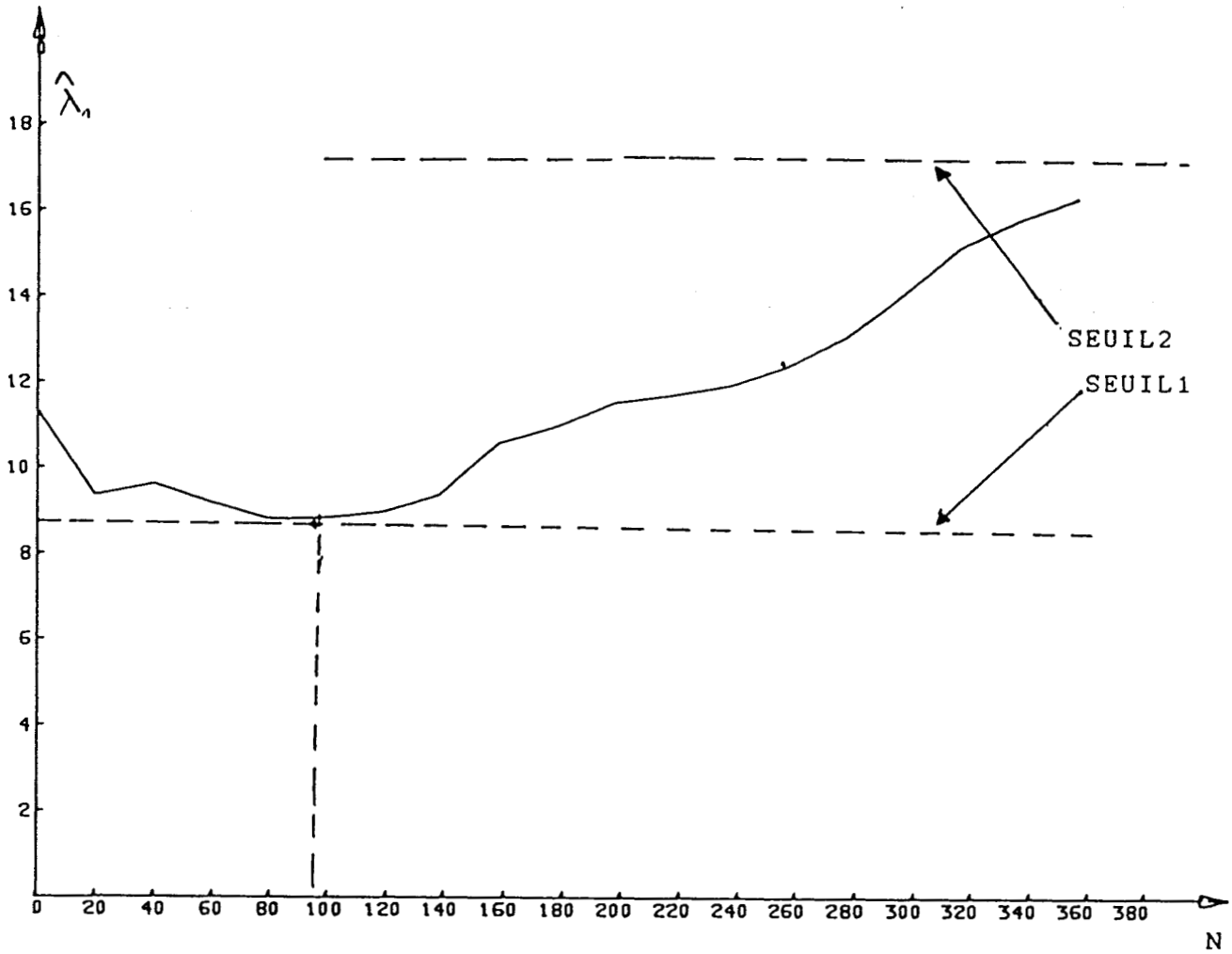
La figure 27 représente l'évolution de l'estimation du vecteur moyenne de la première composante , elle montre clairement qu'on est en présence d'un changement de structure et non pas d'un glissement naturel de la composante.

En prenant comme base d'apprentissage un ensemble e_4 , constitué d'observations tirées à partir des quatres composantes précédemment définies , le résultat est le suivant:

	vecteur moyenne	écarts types	direction propre
comp.1	$\overline{X}_1=92$ $\overline{X}_2=213.3$	$\lambda_1=9.4$ $\lambda_2=3.6$	U=-40°
comp.2	$\overline{X}_1=116.7$ $\overline{X}_2=189.9$	$\lambda_1=5.3$ $\lambda_2=3.9$	U=-34°
comp.3	$\overline{X}_1=205.7$ $\overline{X}_2=104.1$	$\lambda_1=3.84$ $\lambda=2.38$	U=-76°

On note que la première composante du mélange initial a disparu

BU LILLE



observations intégrés séquentiellement

fig.26a évolution de l'estimation de λ_1

BU LILLE

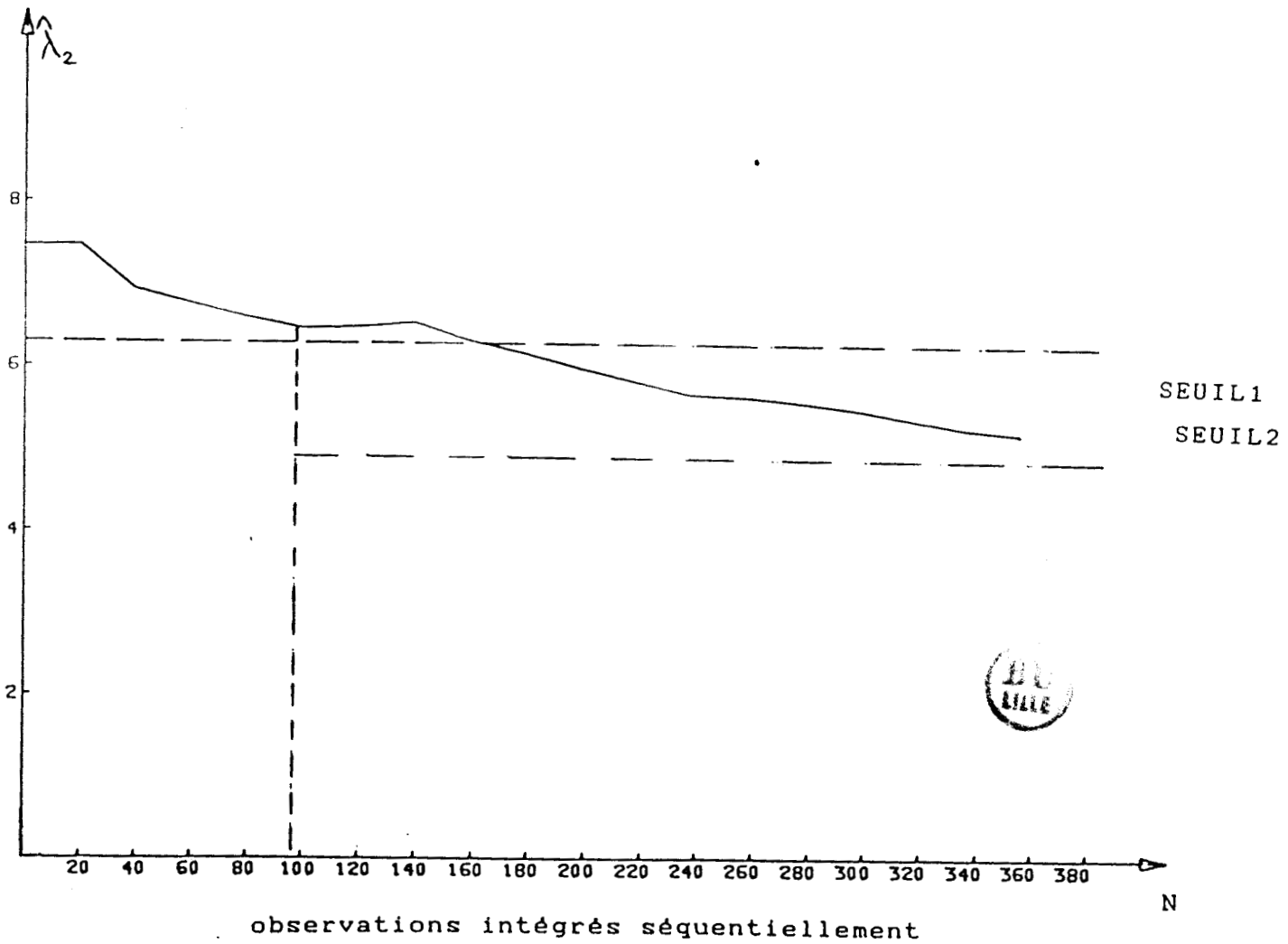


fig.26b évolution de l'estimation de λ_2

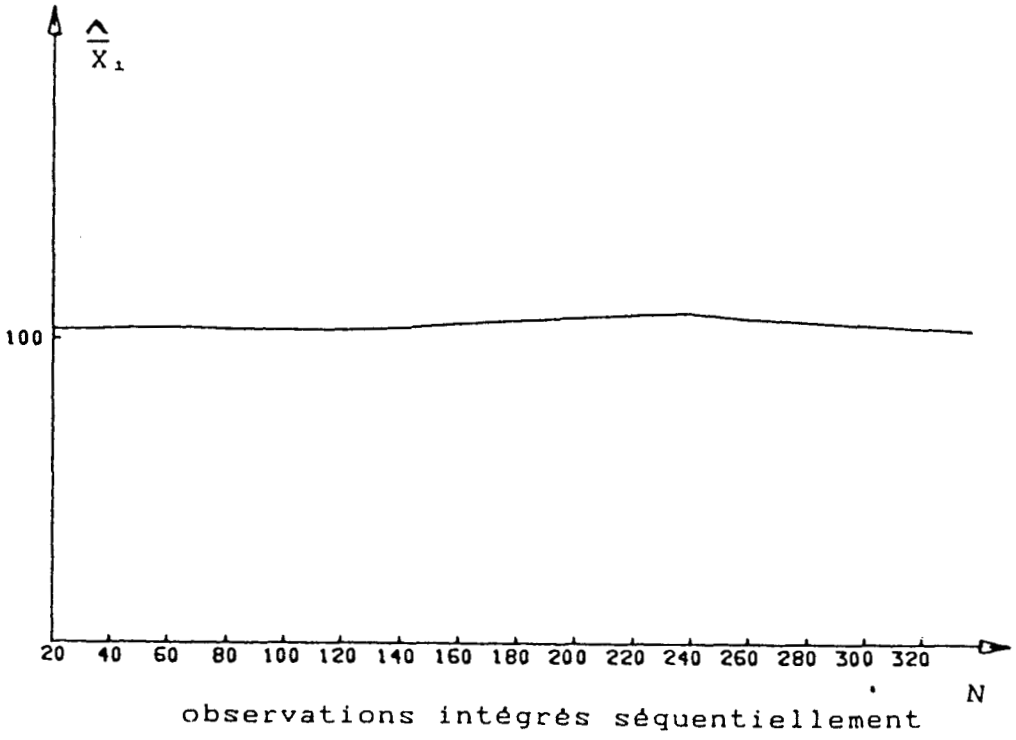


fig.27a évolution de l'estimation de $\bar{X}_1$

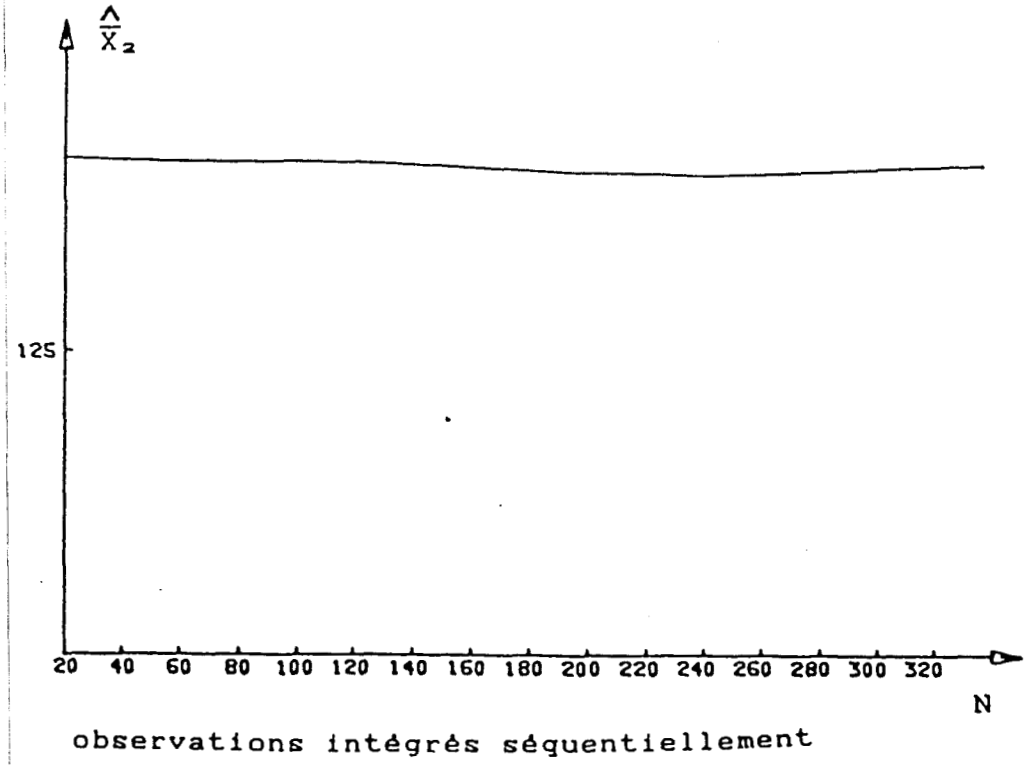


fig.27b évolution de l'estimation de $\bar{X}_2$

BU LILLE

IV-7 CONCLUSION

L'utilisation des différents contours de confiance permet dans les cas simples de fournir directement les paramètres statistiques des nouvelles composantes responsables du changement de structure du mélange gaussien.

Par contre dans les cas difficiles,illustrés par un ensemble de situations particulières,l'utilisation de ces critères ne peut que fournir des informations a priori sur la disposition relative des nouvelles composantes par rapport à celles déja existantes.

Evidemment ces informations sont utilement employées lors de l'opération d'actualisation du modèle gaussien.

REFERENCES

S.WATANABE
" PATTERN RECOGNITION AS A QUEST FOR MINIMUM ENTROPY "
PATTERN RECOGNITION VOL 13 N°5 1981

I.YOUNG
" FUTHER CONSIDERATION OF SAMPLES AND FEATURE SIZE "
IEEE TRANS.INFOR.THEORY 1978

E.L.LEHMAN
" TESTING STATISTICAL HYPOTHESE "
WILEY 1959 NEW-YORK

T.S.FERGUSSON
" A DECISION THEORIC APPROACH MATHEMATICAL STATISTICS "
ACADEMIC PRESS 1967

D.H.FOLEY
" CONSIDERATION OF SAMPLES AND FEATURE SIZE "
IEEE TRANS.INFOR.THEORY SEPTEMBRE 1972

A.SMOLART
" DEVELOPPEMENT D'UNE METHODE DE CLASSIFICATION AVEC REJET , APPLIQUEE AU DIAGNOSTIC DE FONCTIONNEMENT DU COEUR D'UNE CENTRALE NUCLEAIRE . TRAITEMENT DES SIGNAUX THERMIQUES ISSUES D'UN ESSAI DE SIMULATION "
THESE DE DOCTEUR TROISIEME CYCLE COMPIEGNE 1982

K.FUKUNAGA AND T.FLICK
" A TEST OF GAUSSIAN-NESS OF DATA SET USING CLUSTERING "
IEEE TRANS.PAMI MARS 1982

I.S.REED AND AL
" RAPID CONVERGENCE RATE IN ADAPTATIVE ARRAYS "
IEEE TRANS.AEROSP.ELEC.SYSTEME NOVEMBRE 1974

T.M.COVER
" GEOMITRICAL AND STATISTICAL PROPERTIES OF LINEAR INEQUALITIES WITH APPLICATIONS WITH PATTERN "
IEEE TRANS.ELEC.COMPUT 1965

G.PAASS
" CONSISTENT EVALUATION OF UNCERTAIN REASONING SYSTEMS "
6-m- JOURNEES INTERNATIONNALES DES S.E ET LEURS APPLICATIONS
AVRIL 1986 AVIGNON FRANCE

V-CONCLUSION

CONCLUSION

L'étude qui fait l'objet de ce mémoire se décompose essentiellement de trois parties.

La première partie représente la méthode récursive d'estimation des paramètres caractéristiques d'un mélange gaussien multidimensionnel.

On arrive à évaluer , de manière dynamique , l'effet produit par l'introduction de nouveaux échantillons dans la base d'apprentissage sur le modèle gaussien.

Cette approche dynamique , dans l'identification d'un mélange gaussien , est rendue nécessaire par l'évolution naturelle des systèmes étudiés et par la qualité de la base d'apprentissage parfois incomplète.

La deuxième partie traite du fusionnement des composantes du mélange gaussien.

L'estimation des critères de séparabilité des composantes en quelques points privilégiés , définis dans le chapitre III , permet de déceler le début de changement de stucture dans le modèle gaussien.

La définition de ces quelques points est importante dans la mesure où l'influence d'une composante donnée du mélange gaussien s'y exerce en premier lieu et en plus elle y est maximale.

La troisième partie considère le cas de l'augmentation du nombre de composantes d'un mélange gaussien.

La définition de la notion de contour de confiance associé à chaque composante et son utilisation comme seuil limite d'appartenance permet de mettre en évidence l'apparition de nouvelles distributions normales.

Ces trois parties qui semblent apparemment indépendantes sont en réalité complémentaires ,voire même indissociables dans le cas intérressant de la détection de l'apparition d'un mode de fonctionnement qui n'a pas été observé lors de la phase d'apprentissage.

La mise en évidence d'un mode de fonctionnement qui a été ignoré au cours de la phase d'apprentissage s'effectue chronologiquement comme suit :

- utilisation de la troisième partie pour déceler l'apparition d'une ou de plusieurs composantes (représentant naturellement le même mode de fonctionnement)

- utilisation de la deuxième partie pour suivre l'évolution spaciale de ces nouvelles composantes

- utilisation de la troisième partie pour d'une part déceler le début du fusionnement et d'autre part donner les paramètres caractéristiques de la composante résultante qui représente réellement le nouveau mode de fonctinnement.

Les trois parties développées dans cette étude constituent les fondements du module auto-organisation d'une chaine de commande ,comme l'illustre la figure 1.

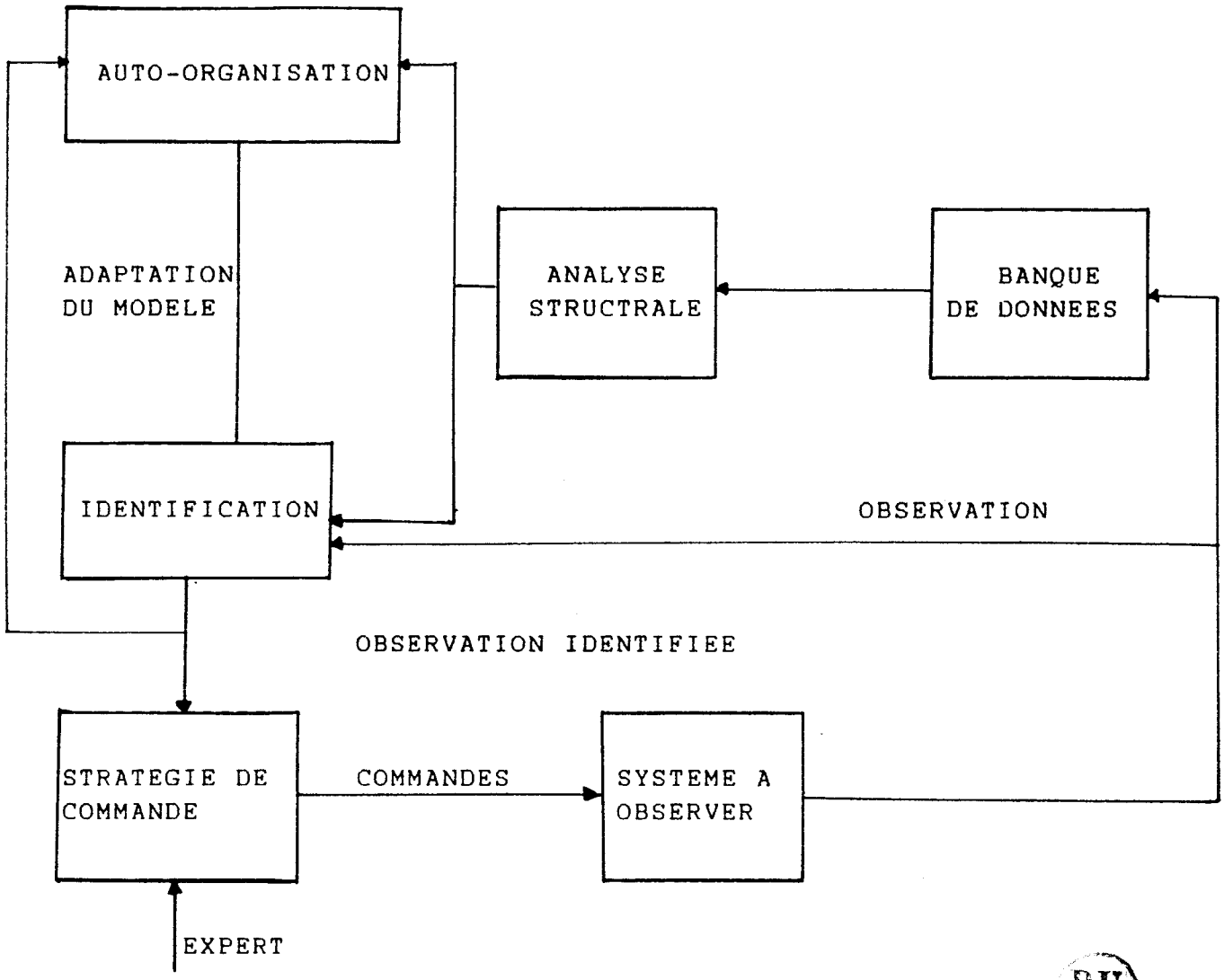


fig.1 chaine d'analyse et de commande

BU LILLE

Une fois l'observation identifiée , il importe de définir une stratégie de prise de décision .

C'est dans le sens d'une étude du module " stratégie de commande " que nous envisageons le développement de nos recherches

ANNEXES

ANNEXE 1

CALCUL DE L'EXPRESSION $\Sigma^+ = \alpha\Sigma + \beta R$

En prenant les expressions (6a) et (6b) du chapitre II obtenues à partir de l'ensemble des chantillons initiaux d'apprentissage (X^i ,i=1,N) ,on peut réécrire Σ^+ sous la forme:

$$\Sigma^+ = \frac{1}{N}\left[\sum_{i=1}^{N}\underbrace{(X^i - \bar{X}^+)(X^i - \bar{X}^+)^T}_{A} + \underbrace{(X^{N+1} - \bar{X}^+)(X^{N+1} - \bar{X}^+)^T}_{B}\right] \quad (A1)$$

$$A=\sum_{i=1}^{N}(X^i - \bar{X}^+)(X^i - \bar{X}^+)^T \qquad (A2)$$

$$B=(X^{N+1} - \bar{X}^+)(X^{N+1} - \bar{X}^+)^T \qquad (A3)$$

Par ailleurs d'après l'expression (20) du même chapitre

$$\bar{X}^+ = \bar{X} + \frac{1}{N+1}(X^{N+1} - \bar{X}) \qquad (A4)$$

En reportant (A4) dans respectivement (A2) et (A3), il vient:

CALCUL DE A

$$A=\sum_{i=1}^{N}\left[X^i - \bar{X} - \frac{1}{N+1}(X^{N+1} - \bar{X})\right]\left[X^i - \bar{X} - \frac{1}{N+1}(X^{N+1} - \bar{X})\right]^T$$

$$A=\sum_{i=1}^{N}(X^i - \bar{X}^+)(X^i - \bar{X})^T - \frac{2}{N+1}\sum_{i=1}^{N}(X^i - \bar{X})(X^{N+1} - \bar{X})^T$$

$$+ \frac{N}{(N+1)^2}(X^{N+1} - \bar{X})(X^{N+1} - \bar{X})^T$$

En remarquant que : $\sum_{i=1}^{N}(X^i - \bar{X}) = 0$

L'expression de A devient :

$$= (N-1)\Sigma + \frac{N}{(N+1)^2}(X^{N+1} - \bar{X})(X^{N+1} - \bar{X})^T \qquad (A5)$$

CALCUL DE B

$$B = [X^{N+1} - \bar{X} - \frac{1}{N+1}(X^{N+1} - \bar{X})][X^{N+1} - \bar{X} - \frac{1}{N+1}(X^{N+1} - \bar{X})]^T$$

$$B = \frac{N^2}{(N+1)^2}(X^{N+1} - \bar{X})(X^{N+1} - \bar{X})^T \qquad \cdot (A6)$$

En reportant ces résultats dans l'expression (A1) il vient:

$$\Sigma^+ = \frac{1}{N}[(N-1)\Sigma + \frac{N}{N+1}(X^{N+1} - \bar{X})(X^{N+1} - \bar{X})^T]$$

Ce qui peut s'exprimer sous la forme recherchée:

$$\Sigma^+ = \alpha\Sigma + \beta R$$

avec $\alpha = \frac{N-1}{N}$ $\beta = \frac{1}{(N+1)}$; $R = (X^{N+1} - \bar{X})(X^{N+1} - \bar{X})^T$

ANNEXE 2

A partir des relations (B1) et (B2) représentant respectivement les contours moyens calculés $Dw(K)_{moy.cal}$ et $Dw^{+}(K)_{moy.cal}$

$$Dw(K)_{moy.cal} = \frac{1}{N} \sum_{i=1}^{N} (X^{i} - \bar{X})^{T} \Sigma^{-1} (X^{i} - \bar{X}) \qquad (B1)$$

$$Dw^{+}(K)_{moy.cal} = \frac{1}{N+1} \sum_{i=1}^{N+1} (X^{i} - \bar{X}^{+})^{T} \Sigma^{-1+} (X^{i} - \bar{X}^{+}) \qquad (B2)$$

et en mettant (B2) sous la forme suivante:

$$Dw^{+}(K)_{moy.cal} = \frac{1}{N+1} \sum_{i=1}^{N} \underbrace{(X^{i} - \bar{X}^{+})^{T} \Sigma^{-1+} (X^{i} - \bar{X}^{+})}_{C} + \frac{1}{N+1} \underbrace{(X^{N+1} - \bar{X}^{+})^{T} \Sigma^{-1+} (X^{N+1} - \bar{X}^{+})}_{D} \qquad (B3)$$

on établit une relation permettant de déterminer $Dw^{+}(K)_{moy.cal}$ en fonction de $Dw(K)_{moy.cal}$

Les relations (B4) et (B5) fournissent respectivement les expressions de X^{+} et Σ^{-+}

$$\bar{X}^{+} = \bar{X} - \frac{1}{N+1} (X^{N+1} - \bar{X}) \qquad (B4)$$

$$\Sigma^{-1+} = \frac{N}{N-1} \Sigma^{-1} - \frac{N}{N-1} \cdot \frac{KOEF.\Sigma^{-1}(X^{N+1} - \bar{X})(X^{N+1} - \bar{X})^{T}\Sigma^{-1}}{1 + KOEF.(X^{N+1} - \bar{X})^{T}\Sigma^{-1+}(X^{N+1} - \bar{X})} \qquad (B5)$$

avec $$KOEF = \frac{N^2}{N^2-1}$$

En reportant d'une part (B4) et (B5) dans (B3),et en s'inspirant d'autre part de la démarche adoptée dans l'annexe1,on obtient le résultat suivant:

CALCUL DE C

$$C = \frac{1}{N+1} \sum_{i=1}^{N} [(X^i - \bar{X}) - \frac{1}{N+1}(X^i - \bar{X})]^T \Sigma^{-1} [(X^i - \bar{X}) - \frac{1}{N+1}(X^i - \bar{X})]$$

$$C = \frac{1}{N+1} \sum_{i=1}^{N} (X^i - \bar{X})^T \Sigma^{-1} (X^i - \bar{X}) + \frac{1}{(N+1)^3} \sum_{i=1}^{N} (X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})$$

$$- \frac{2}{(N+1)^2} \sum_{i=1}^{N} (X^i - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X}) \quad .$$

Le dernier terme de cette expression est nul.

$$D = \frac{1}{(N+1)^3} (X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})$$

$$C+D = \frac{1}{N+1} \sum_{i=1}^{N} (X^i - \bar{X})^T \Sigma^{-1} (X^i - \bar{X}) + \frac{1}{(N+1)^2} (X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})$$

$$= \frac{1}{N+1} \cdot \frac{N}{N-1} \sum_{i=1}^{N} (X^i - \bar{X})^T \Sigma^{-1} (X^i - \bar{X})$$

$$- \frac{1}{N+1} \cdot \frac{N}{N-1} \sum_{i=1}^{N} \frac{KOEF.(X^i - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^i - \bar{X})}{1 + KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}$$

$$+ \frac{1}{(N+1)^2} \cdot \frac{N}{N-1} (X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})$$

$$- \frac{1}{(N+1)^2} \cdot \frac{N}{N-1} \cdot \frac{KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}{1 + KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}$$

$$C+D = KOEF.Dw(K) - \frac{1}{(N+1)^2} \cdot \frac{N}{N-1} \frac{(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}{1 + KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}$$

$$- \frac{1}{N+1} \cdot \frac{N}{N-1} \sum_{i=1}^{N} \frac{KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^i - \bar{X})(X^i - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}{1 + KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}$$

$$C+D = KOEF.Dw(K) - \frac{1}{(N+1)^2} \cdot \frac{N}{N-1} \frac{(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}{1 + KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}$$

$$- \frac{1}{N+1} \cdot \frac{N}{N-1} \frac{KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} \left(\sum_{i=1}^{N} (X^i - \bar{X})(X^i - \bar{X})^T \right) \Sigma^{-1} (X^{N+1} - \bar{X})}{1 + KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}$$

Finalement l'expression du contour moyen est la suivante:

$$C+D = KOEF.Dw(K) - \frac{1}{(N+1)^2} \cdot \frac{N}{N-1} \frac{(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}{1 + KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}$$

$$- \frac{N (KOEF)^2 . (X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}{1 + KOEF.(X^{N+1} - \bar{X})^T \Sigma^{-1} (X^{N+1} - \bar{X})}$$

LILLE

BIBLIOTHEQUE INTERUNIVERSITAIRE
SECTION DE SCIENCES
LILLE

RESUME

Ce travail vise à définir une approche récursive pour identifier les mélanges gaussiens modélisant des systèmes dynamiques.

Le premier chapitre du mémoire décrit les phases nécessaires qu'il faut observer pour réaliser les opérations d'analyse, d'identification et d'optimisation sur les systèmes complexes en utilisant uniquement les données qui les caractérisent.

Le deuxième chapitre présente un ensemble d'équations récurrentes qui permettent d'affiner les estimations des paramètres statistiques des différentes composantes d'un mélange gaussien en fonction de l'enrichissement de la base de données par intégration de nouveaux échantillons.

Le troisième chapitre définit un ensemble de points dans l'espace d'observation où les influences mutuelles des différentes composantes d'un mélange gaussien atteignent leur maximum. L'estimation des critères de séparabilité des composantes en ces points permet de déceler éventuellement le début du fusionnement des composantes.

Finalement le quatrième chapitre propose une définition du contour de confiance associé à chaque composante. Son utilisation comme seuil limite d'appartenance permet de mettre en évidence l'apparition de nouvelles distributions normales représentant les modes de fonctionnement du système étudié qui n'ont pas été observés pendant la phase d'apprentisaage.