

**Application de l'analyse en composantes indépendantes sur des
données boursières**

par

Maryeme MAZINE

mémoire présenté au Département de mathématiques
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, juin 2010



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-70744-9
Our file *Notre référence*
ISBN: 978-0-494-70744-9

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Le 28 juin 2010

*le jury a accepté le mémoire de Madame Maryème Mazine
dans sa version finale.*

Membres du jury

**Professeur Ernest Monga
Directeur de recherche
Département de mathématiques**

**Professeur André Mayers
Codirecteur de recherche
Département d'informatique**

**Professeure Sylvie Béquet
Membre
Université Bishop's
Williams School of Business**

**Professeur Bernard Colin
Président rapporteur
Département de mathématiques**

Résumé

Quels sont les facteurs qui influencent les fluctuations des actions boursières ? Voici une question qui préoccupe au plus haut niveau les investisseurs, ainsi qu'un nombre considérable de chercheurs s'intéressant à ce domaine très complexe et toujours d'actualité. En effet, pour prendre une décision d'investissement, les concernés ont recours à des méthodes telles que la valeur du bilan, la méthode des multiples et la méthode d'actualisation. Ces dernières contiennent des facteurs sous-jacents liés, entre autres, à l'économie, au secteur d'activité et à la psychologie de l'investisseur.

Afin d'être en mesure d'identifier de tels facteurs, nous proposons de procéder à une analyse en composantes indépendantes (ACI) sur les prix à la fermeture des titres boursiers. Une approche qui appartient à la méthode de Séparation Aveugle des Sources (SAS). L'ACI vise à extraire des composantes indépendantes inconnues uniquement à partir des données observées.

Ce mémoire est constitué de quatre chapitres : le premier chapitre présente en général la méthode de la séparation aveugle des sources (SAS), le deuxième décrit l'ACI, le troisième définit l'approche de l'ACI basée sur le critère de maximisation de la non-gaussianité. Finalement, le quatrième illustre les résultats de l'application de l'ACI sur les titres boursiers.

REMERCIEMENTS

Qu'il me soit permis de présenter ici mes remerciements à l'ensemble des personnes qui ont rendu possible le présent mémoire et qui ont contribué à son élaboration sous quelque forme que ce soit.

J'exprime mes profonds remerciements à mes directeurs, Ernest Monga et André Mayers pour l'aide compétente qu'ils m'ont apportée. Leur oeil critique m'a été très précieux pour structurer le travail et pour améliorer la qualité des différentes sections.

Je remercie également Bernard Colin d'avoir accepté de présider mon jury et d'avoir partagé sa passion de la statistique, lors de ses enseignements ou dans de simples discussions.

J'adresse mes sincères remerciements à Sylvie Béquet d'avoir accepté de participer à mon jury, mais surtout pour l'aide précieuse qu'elle m'a apportée tout au long de mes recherches. J'ai apprécié sa sympathie, sa disponibilité et surtout son enthousiasme à partager ses connaissances.

J'exprime ma gratitude envers ma famille pour leur amour et leur soutien. Merci de m'avoir donné les moyens de réussir.

Mes remerciements s'étendent également à tous mes proches et amis, qui m'ont soutenue et encouragée durant la réalisation de ce mémoire.

Enfin, je remercie tout le personnel du département de mathématiques et du département d'informatique.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
TABLE DES MATIÈRES	iv
LISTE DES TABLEAUX	viii
LISTE DES FIGURES	x
INTRODUCTION	1
CHAPITRE 1 — Séparation aveugle des sources	3
1.1 Présentation de la problématique	4
1.1.1 Formulation mathématique	4
1.1.2 Modélisation des sources et du système de mélange	5
1.1.3 Réalisation de la séparation	6
1.2 Hypothèses	7
1.3 Indéterminations	9

1.3.1	Indétermination de permutation	9
1.3.2	Indétermination du facteur d'échelle	9
1.4	Critère de la séparation	10
1.4.1	Insuffisance de la décorrélation	10
1.4.2	L'indépendance	11
1.5	Optimisation dans la séparation des sources	12
1.5.1	La descente du gradient	12
1.5.2	Méthode de Newton	13
1.5.3	La fonction de Lagrange	14
1.5.4	Déflation	15
1.5.5	Orthogonalisation symétrique	16
CHAPITRE 2 — Analyse en composantes indépendantes		18
2.1	Définition de l'analyse en composantes indépendantes	19
2.1.1	Présentation du modèle	19
2.1.2	Contraintes et restrictions du modèle	20
2.2	Propriétés statistiques de l'ACI	21
2.2.1	Lien entre la décorrélation et le blanchiment	21
2.2.2	La gaussianité dans l'ACI	23
2.2.3	Indépendance et gaussianité	25
2.3	Réduction de la dimension de données	26

2.3.1	Analyse en composantes principales	27
2.3.2	Blanchiment	30
2.4	Mesures de séparation dans l'ACI	31
2.4.1	Entropie	32
2.4.2	Information mutuelle	34
2.4.3	Le coefficient d'aplatissement	36
2.4.4	Négentropie	37
CHAPITRE 3 — ACI par maximisation de la non-gaussianité		39
3.1	Quelques critères de séparation	40
3.1.1	Maximiser la vraisemblance	40
3.1.2	Minimiser l'information mutuelle	41
3.1.3	La décorrélation non linéaire	43
3.1.4	L'annulation des cumulants croisés d'ordre supérieur	44
3.2	Maximiser la non-gaussianité	46
3.2.1	Mesurer la non-gaussianité dans l'ACI	47
3.2.2	Analyse de l'estimateur et choix de la fonction F	51
3.2.3	Algorithme de séparation des sources avec la négentropie	59
CHAPITRE 4 — ACI sur des données boursières		64
4.1	Jeu de données	67

4.1.1	Présentation	67
4.1.2	Motivation	68
4.2	Classification des composantes indépendantes	69
4.2.1	Méthodologie	69
4.2.2	Analyse des résultats	71
4.3	Prévision avec l'ACI	90
4.3.1	Méthodologie	90
4.3.2	Analyse des résultats	93
CONCLUSION		98
Annexe A		100
4.4	Reconstitution des titres	101
4.4.1	Reconstitution avec le critère EAM	101
4.4.2	Reconstitution avec le critère séquentiel	105
4.4.3	Analyse sectorielle	108
4.5	Estimation des titres	109
Annexe B		112
4.6	La matrice du mélange A	112
BIBLIOGRAPHIE		114

LISTE DES TABLEAUX

4.1	Tableau de l'erreur absolue moyenne de chaque composante dans la reconstitution d'IMO	75
4.2	Tableau des poids des m composantes indépendantes sélectionnées pour la reconstitution du titre i	81
4.3	Tableau des poids de chaque composante indépendante dans la reconstitution des titres IMO, NXY et TLM	82
4.4	Tableau des prévalences non pondérées et des prévalences pondérées de chaque composante dans le secteur de l'énergie	84
4.5	Tableau des répartitions séquentielles	85
4.6	Tableau de calcul de la statistique χ^2	86
4.7	Tableau des composantes retenues selon leur χ^2	86
4.8	Comparaison de la concordance et du taux de prévision entre notre modèle et le modèle classique AR.	93
4.9	Comparaison de la concordance et du taux de prévision entre notre modèle et le modèle classique AR.	95
4.10	Tableau des quotients de concordance et de prévision	97

4.11	Tableau des IC sélectionnées selon le critère EAM pour les titres MFC, BNS et KFS	101
4.12	Tableau des composantes retenues selon leur χ^2	105
4.13	Tableau des poids de chaque composante dans la reconstitution des titres MFC, BNS et KFS	108
4.14	Tableau des prévalences non pondérées et des prévalences pondérées de chaque composante dans le secteur financier	108
4.15	Comparaison de la concordance et le taux de prévision entre notre modèle et le modèle classique AR.	109

LISTE DES FIGURES

2.1	Distribution conjointe du mélange blanchi de deux composantes indépendantes uniformes s_1 et s_2	31
2.2	Distribution conjointe de deux composantes indépendantes uniformes s_1 et s_2	31
2.3	Exemple de distribution sur-gaussienne : la distribution de laplace	36
2.4	Exemple de distribution sous-gaussienne : la distribution uniforme	36
4.1	Répartition sectorielle de l'indice S&P / TSX en date du 19 août 2009 . .	67
4.2	Tableaux des variables	68
4.3	Illustration des composantes indépendantes IC1, IC16 et IC10 obtenues par les fonctions de contrastes F_1, F_2 et F_3	72
4.4	Illustration des composantes indépendantes IC4, IC1 et IC16 obtenues par les fonctions de contrastes F_2 et F_3	73
4.5	IMO	74
4.6	IMO par rapport aux années	74
4.7	Indice de proximité pour le titre IMO	76

4.8	Reconstitution du titre IMO avec l'approche EAM	77
4.9	Comparaison entre IMO et sa reconstitution avec l'approche EAM	77
4.10	Composante indépendante numéro 23	78
4.11	Composante indépendante numéro 8	79
4.12	Composante indépendante numéro 4	79
4.13	Composante indépendante numéro 20	80
4.14	Composante indépendante numéro 7	83
4.15	Composante indépendante numéro 24	83
4.16	Composante indépendante numéro 3	87
4.17	Composante indépendante numéro 13	88
4.18	Composante indépendante numéro 19	88
4.19	Reconstitution de IMO avec l'approche séquentielle	89
4.20	Comparaison entre IMO et sa reconstitution avec l'approche séquentielle	89
4.21	Illustration du modèle de prédiction : les sources sont obtenues par la transformation de séparation $S = \mathbf{W}X$. Les \hat{S}_i correspondent aux composantes prédites par le modèle AR. \hat{X}_i est la prédiction de la série X_i , elle a été calculée par le modèle de mélange $X = \mathbf{A}S$	91
4.22	Variations des prix du titre IMO en 2009	93
4.23	Prédiction d'IMO avec l'ACI : la reconstitution du titre à été effectuée avec l'approche EAM	94
4.24	Prédiction d'IMO directement avec AR	94

4.25	Comparaison entre les modèles de prévision	95
4.26	Prédiction d'IMO avec l'ACI : la reconstitution du titre à été effectuée avec l'approche séquentielle	96
4.27	Comparaison entre les modèles de prévision	96
4.28	Composante indépendante numéro 7	102
4.29	Composante indépendante numéro 9	102
4.30	Composante indépendante numéro 17	102
4.31	Composante indépendante numéro 23	103
4.32	Composante indépendante numéro 25	103
4.33	Composante indépendante numéro 14	103
4.34	Le titre BNS et sa reconstitution avec le critère EAM	104
4.35	Le titre MFC et sa reconstitution avec le critère EAM	104
4.36	Le titre KFS et sa reconstitution avec le critère EAM	105
4.37	Composante indépendante numéro 10	106
4.38	Composante indépendante numéro 21	106
4.39	Le titre BNS et sa reconstitution avec le critère séquentiel	107
4.40	Le titre MFC et sa reconstitution avec le critère séquentiel	107
4.41	Le titre original MFC et sa prédiction de 2009	109
4.42	Le titre original BNS et sa prédiction de 2009	110
4.43	Le titre original KFS et sa prédiction de 2009	110

4.44 Les coefficients de la matrice du mélange A	112
---	-----

Notations mathématiques

Majuscule grasse	Matrice
Majuscule	Vecteur colonne
Minuscule	Scalaire
$(.)^T$	Opération de transposition
$\det(A)$	Déterminant
$E\{X\}$	Espérance mathématique de la variable X
$\text{Var}(X)$	Variance mathématique de X
$\text{Cov}(X, Y)$	Covariance mathématique entre X et Y
$E\{SS^T\}$	La matrice de variance covariance du vecteur aléatoire X
I	La matrice identité

INTRODUCTION

Quotidiennement, les investisseurs en bourse se posent des questions concernant l'évolution des cours de la bourse. La valeur d'une action correspond à son cours en bourse. Ce dernier est le résultat de la confrontation de l'offre et de la demande de titres. Les investisseurs se servent de plusieurs méthodes pour prendre une décision sur les actions qui les intéressent. Les méthodes font appel à des indicateurs qui se basent sur :

- La valeur de bilan : cette valeur est estimée en fonction des fonds propres de la société (Le capital, la prime d'émission, ...).
- La méthode des multiples : L'approche des multiples se base sur l'évaluation d'un ensemble d'entreprises comparables. Les indicateurs les plus utilisés sont les multiples boursiers (capitalisation boursière/résultat net ...) et les multiples de transaction (déterminés à partir des transactions observées sur le marché).
- La méthode d'actualisation (Discounted Cash Flow) : cette méthode consiste à calculer la valeur actuelle de la société sur la base des entrées nettes de fonds futurs.

Il est évident que ces indicateurs contiennent plusieurs facteurs sous-jacents liés par exemple à l'économie, au secteur d'activité, à la psychologie de l'investisseur, ... C'est dans cette optique que l'idée d'extraire de tels facteurs est intéressante.

L'analyse en composante indépendante (ACI) est une méthode de séparation aveugle des sources basée sur l'indépendance. Cette méthode vise à estimer les n signaux inconnus et supposés indépendants (les sources) d'après la seule connaissance des m mélanges de

ces signaux (les observations). Le terme aveugle signifie que les sources et le processus du mélange ne sont pas connus.

Dans ce mémoire, nous avons extrait grâce à l'ACI, des structures sous-jacentes à partir des prix des actions. Ces structures ont été utilisées pour interpréter des facteurs indépendants qui influencent le marché boursier. De plus, nous avons construit un modèle de prédiction basé sur ces structures.

Nous étudierons dans le chapitre 1 le mélange de base de la séparation aveugle des sources (SAS) : le mélange linéaire instantané. Nous discuterons des indéterminations et de l'identifiabilité de la SAS. Nous verrons qu'il n'est possible de résoudre un problème de SAS qu'au prix de certaines contraintes. Enfin, nous exposerons certaines approches d'optimisation utilisées dans la SAS.

Dans le chapitre 2 nous allons décrire la méthode de l'analyse en composantes indépendantes (ACI). Cette méthode de SAS est basée sur l'indépendance statistique des sources. Ainsi, dans ce chapitre, nous discuterons des différents principes statistiques de cette méthode.

Plusieurs algorithmes ont été développés pour résoudre l'ACI, le chapitre 3 fera l'objet de l'algorithme que nous avons utilisé en pratique et qui porte sur le critère de la maximisation de la non-gaussianité.

Le chapitre 4 quant à lui est réservé à l'application de l'ACI sur les données boursières. Nous nous sommes intéressés aux prix de certaines actions de l'indice S&P / TSX de la bourse de Toronto. Notre étude est divisée en deux parties :

- Une première partie montrant l'apport descriptif de l'ACI sur un tel jeu de données.
- Une seconde partie qui décrit notre modèle de prédiction basé sur les composantes indépendantes.

Enfin, le bilan de nos travaux et quelques perspectives de recherches seront proposés en conclusion.

CHAPITRE 1

Séparation aveugle des sources

La séparation aveugle des sources consiste à élaborer des méthodes permettant d'identifier les variables inobservables dites **sources** ou **latentes**, à partir de mélanges de ces dernières que l'on appelle observations. Le terme aveugle désigne le fait que nous n'avons aucune information sur le processus de ce mélange ni sur les sources.

Le problème de la séparation des sources est un problème classique qui a été initié par Jutten et Héroult dans les années 80 [HA84, HJA85b, HJA85a].

À son origine, la séparation des sources se place dans le cadre de mélange linéaire instantané. Cardoso [CS93] a proposé d'utiliser les cumulants d'ordre quatre pour effectuer la séparation. Par la suite, il a publié l'algorithme JADE en 1993. Durant cette même année, Comon a introduit la notion de fonctions de contraste, il a proposé comme contraste de maximiser la somme des cumulants d'ordre 4 [Com94]. Quant à Delfosse et Loubaton [DL95], ils ont développé une approche par déflation : c'est-à-dire que les sources sont extraites une par une. Toujours dans le même contexte non bruité, Cardoso et Laheld ont développé des algorithmes adaptatifs basés sur la notion de gradient matriciel relatif [CL92]. Dans [PG97], Pham et Garrat ont développé une méthode basée sur le principe

du maximum de vraisemblance.

Face à cette expansion de méthodes dans ce domaine, le cas du mélange convolutif a été introduit. La majorité de ces algorithmes ont été généralisés à partir de ceux déjà développés dans le cadre instantané [Hay94, Nan99].

Ce premier chapitre présente les différents principes de la séparation aveugle des sources : tout d'abord, nous allons introduire le modèle de base de cette méthode, celui des mélanges linéaires instantanés stationnaires. Ensuite, nous abordons les problèmes d'identification et d'indétermination de la séparation. Puis, nous citons les principales hypothèses de cette méthode. Ensuite, nous déterminons les critères de séparation, en particulier celui de l'indépendance. Enfin, nous présentons quelques approches d'optimisation utilisées pour la construction des algorithmes de séparation.

1.1 Présentation de la problématique

Il est nécessaire de préciser les hypothèses que nous considérons tout au long de ce mémoire. Nous nous plaçons dans le cas de mélanges linéaires sans bruit additif où les sources sont supposées indépendantes. De plus, nous traitons uniquement le cas où, le nombre des sources est égal au nombre de mélange de ces dernières.

1.1.1 Formulation mathématique

L'illustration classique de la séparation des sources est le problème appelé *the cocktail party problem*. En effet, supposons que nous soyons dans une salle dense où plusieurs personnes discutent, nous disposons de plusieurs microphones qui nous permettent d'en-

registrar la parole. Ces microphones fournissent alors n signaux appelés observations où chaque signal est une combinaison linéaire des sources originelles. Les différents coefficients de ces sommes dépendent de la position des différents microphones vis-à-vis des convives.

Le problème de la séparation des sources peut être modélisé sous la forme générale suivante :

$$\begin{aligned}
 X_1(t) &= a_{11}S_1(t) + a_{12}S_2(t) + \dots + a_{1m}S_m(t) \\
 X_2(t) &= a_{21}S_1(t) + a_{22}S_2(t) + \dots + a_{2m}S_m(t) \\
 &\dots \dots \dots \\
 X_n(t) &= a_{n1}S_1(t) + a_{n2}S_2(t) + \dots + a_{nm}S_m(t)
 \end{aligned} \tag{1.1}$$

où

- $S(t) = [S_1(t), S_2(t), \dots, S_m(t)]^T$ représente le vecteur des signaux sources.
- $X(t) = [X_1(t), X_2(t), \dots, X_n(t)]^T$ est le vecteur des observations.

La formulation matricielle du problème s'écrit comme suit :

$$X(t) = \mathbf{A}S(t) \tag{1.2}$$

\mathbf{A} est la matrice de mélange, elle est de dimension $n \times m$ et elle est constituée des coefficients inconnus a_{ij} .

1.1.2 Modélisation des sources et du système de mélange

Concernant les signaux sources, ceux-ci peuvent correspondre à des valeurs réelles ou complexes, ils peuvent également être stationnaires ou non-stationnaires.

La séparation du mélange peut appartenir à deux classes : d'une part, la classe des transformations linéaires instantanées, les capteurs reçoivent des combinaisons linéaires

des sources à chaque instant. Et d'autre part, celle des transformations par convolution, c'est-à-dire que les signaux captés dépendent des sources et de leurs versions retardées. Nous laissons le lecteur se référer par exemple à [Hay94, AC98] pour plus de détails sur cette dernière.

La séparation du mélange peut être plus complexe, notamment, dans le cas où les combinaisons des sources sont non linéaires.

1.1.3 Réalisation de la séparation

L'objectif de la séparation de sources est de reconstruire les sources $S_i(t)$ à partir des observations. Pour ce faire, supposons que la matrice de mélange \mathbf{A} soit une matrice carrée de plein rang. Ainsi, il existe une solution à ce problème tel que :

$$S(t) = \mathbf{A}^{-1}X(t) \quad (1.3)$$

L'idée principale de la séparation des sources est de trouver une transformation \mathbf{W} qui permette d'obtenir des signaux statistiquement indépendants, ou en d'autres termes de calculer une estimation $\hat{\mathbf{W}}$ de la matrice inverse de \mathbf{A} . Nous avons alors :

$$Y(t) = \hat{S}(t) = \mathbf{W}X(t) \quad (1.4)$$

les Y_i correspondent aux estimations des sources S_i . La matrice \mathbf{W} est la matrice de séparation, c'est une matrice carrée qui permet de récupérer les signaux sources à un facteur d'échelle et une permutation près. On en déduit que \mathbf{W} peut s'écrire sous la forme :

$$\mathbf{W} = \mathbf{P}\mathbf{\Lambda}\mathbf{A}^{-1} \quad (1.5)$$

ou encore :

$$\mathbf{W}\mathbf{A} = \mathbf{P}\mathbf{\Lambda} \quad (1.6)$$

où \mathbf{P} est une matrice de permutation et Λ est une matrice diagonale non singulière correspondant au facteur d'échelle. Lorsque la séparation est parfaite nous obtenons l'équivalence suivante :

$$\mathbf{WA} = \mathbf{P}\Lambda = \mathbf{I} \quad (1.7)$$

1.2 Hypothèses

Au vu des différents types de mélanges, il est nécessaire de faire un rapide tour d'horizon de ces possibilités avant de parler plus amplement des hypothèses considérées dans la séparation.

Le processus du mélange diffère selon le domaine d'application :

1. Le mélange linéaire instantané est la forme la plus simple, chaque observation est une combinaison linéaire des sources et le problème consiste à identifier la matrice du mélange.
2. Le mélange linéaire à atténuation et retard, est rencontré lorsque les contributions des sources dans les observations sont des versions pondérées et décalées de ces sources.
3. Le mélange convolutif est le plus compliqué à résoudre. Dans ce cas, les contributions des sources sur les capteurs sont des versions filtrées des sources originelles.
4. Le mélange non linéaire est celui qui est le moins exploré, à cause de sa complexité. Cependant, des classes particulières de ce type ont été étudiées telles que : les

mélanges post non-linéaires ou les mélanges linéaires quadratiques.

Pour estimer la matrice de séparation \mathbf{W} , plusieurs hypothèses ont été considérées. En effet, chaque hypothèse émise sur la distribution des sources correspond à une famille de séparation des sources, parmi ces familles citons :

1. L'Analyse en Composantes Indépendantes (ACI). C'est une méthode qui suppose l'indépendance entre les sources, c'est-à-dire que lorsque $i \neq j$, $S_i(t)$ et $S_j(t)$ sont indépendantes quel que soit l'instant t . Cette approche possède une solution si et seulement si au plus une source est gaussienne. En considérant l'indépendance et la non-gaussianité, nous pouvons donc estimer la matrice \mathbf{W} . Cette famille de méthodes fera l'objet de notre étude et sera plus détaillée dans les chapitres suivants.
2. La parcimonie est un exemple d'information à priori très efficace pour déterminer le problème de la séparation aveugle des sources. Les approches de l'ACI basées sur la représentation parcimonieuse [Hul99, ZP01] permettent de résoudre le problème sous-déterminé (nombre de sources est supérieur au nombre d'observations). Ces approches reposent sur le principe que, dans une représentation appropriée des signaux, les composantes des sources sont indépendantes et parcimonieuses. Ainsi, chaque observation est due essentiellement à une seule source par ce fait, le problème de séparation revient à affecter chaque composante à la bonne source.
3. La Factorisation en Matrices Positives. Elle suppose que les sources et les coefficients de mélange sont tous les deux positifs. On cherche alors à identifier deux matrices positives dont le produit est égal à la matrice des observations. Cette méthode est utilisée par exemple en astrophysique pour la séparation d'espèces chimiques dans des nuages de poussière interstellaires.

1.3 Indéterminations

En considérant les hypothèses citées plus haut, la séparation est bel et bien possible. En revanche, la non-connaissance de la matrice \mathbf{A} ne garantit pas l'égalité entre l'estimation de la matrice de séparation \mathbf{W} et l'inverse de la matrice du mélange \mathbf{A}^{-1} .

1.3.1 Indétermination de permutation

Comme expliqué plus haut, l'estimation de la matrice \mathbf{W} produit des signaux Y_i indépendants. Cependant, nous n'avons aucune information sur l'ordre de ces nouveaux signaux. Ceci se traduit par le fait que dans le modèle de mélange, nous pouvons facilement substituer la matrice de permutation \mathbf{P} avec son inverse : $X = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{S}$.

Les éléments $\mathbf{P}\mathbf{S}$ sont les sources d'origine S_i dans un autre ordre. La matrice $\mathbf{A}\mathbf{P}^{-1}$ est la nouvelle matrice de mélange à déterminer.

1.3.2 Indétermination du facteur d'échelle

Une autre ambiguïté de la séparation aveugle des sources est liée à l'impossibilité d'estimer la variance des signaux sources. Les S_i et la matrice \mathbf{A} sont inconnus, par conséquent, toute source S_i peut être multipliée par un scalaire quelconque α_i . En divisant la colonne a_i correspondante par le même scalaire, la forme des observations reste invariante à cette transformation.

$$x = \sum_i \begin{pmatrix} 1 \\ \alpha_i a_i \end{pmatrix} (s_i \alpha_i) \quad (1.8)$$

Pour résoudre ce problème, les sources sont souvent supposées de matrice variance unité $E\{SS^T\} = \mathbf{I}$.

1.4 Critère de la séparation

Avant d'introduire la propriété de l'indépendance, nous allons expliquer la raison pour laquelle la décorrélation est insuffisante dans le cadre de la séparation des sources.

1.4.1 Insuffisance de la décorrélation

Dans plusieurs cas pratiques, l'information donnée par les statistiques d'ordre 2 est souvent utilisée. Cependant, la décorrélation appelée aussi le blanchiment n'entraîne pas l'indépendance. Supposons que nos sources aient une variance unité, ainsi :

$$E[SS^T] = \mathbf{I} \quad (1.9)$$

La décomposition en valeurs singulières de la matrice \mathbf{A} peut s'écrire :

$$\mathbf{A} = \mathbf{U}\Sigma^{1/2}\mathbf{V} \quad (1.10)$$

où Σ est une matrice diagonale composée des valeurs propres de la matrice $\mathbf{A}\mathbf{A}^T$, \mathbf{U} et \mathbf{V} sont deux matrices unitaires. La matrice de variance des observations s'exprime comme : $E[XX^T] = E[\mathbf{A}SS^T\mathbf{A}^T] = \mathbf{A}E[SS^T]\mathbf{A}^T = \mathbf{U}\mathbf{U}^T$.

On constate qu'à l'ordre 2 (matrice de variance-covariance) la matrice \mathbf{V} a disparu ainsi il est impossible de l'estimer. La matrice \mathbf{A} ne pourra être estimée qu'à une matrice orthogonale \mathbf{V} près.

Par conséquent, dans le cadre de la séparation aveugle des sources, la décorrélation ne suffit pas pour avoir une séparation exacte, l'utilisation d'un critère plus puissant tel que l'indépendance est plus appropriée. Toutefois, la décorrélation peut être utilisée comme un prétraitement permettant de simplifier le problème de la séparation aveugle des sources.

1.4.2 L'indépendance

Considérons le vecteur $X = [X_1, X_2, \dots, X_n]^T$ et désignons respectivement par $p(x)$ et $p(x_i)$ la densité conjointe de X et les densités marginales. Le vecteur X est dit à composantes statistiquement indépendantes si et seulement si :

$$p(x) = \prod_{i=1}^n p(x_i) \quad (1.11)$$

En effet, les composantes du vecteur X sont mutuellement indépendantes si la densité de probabilité conjointe de X , $p(x)$ est égale au produit des densités marginales $p(x_i)$.

Ainsi, pour mesurer l'indépendance il suffit de vérifier que la loi jointe de X est égale au produit des lois marginales.

Considérons une variable aléatoire X et deux distributions possibles de cette variable $p(x)$ et $q(x)$, la divergence de Kullback [CT91] s'exprime comme suit :

$$D(p(x), q(x)) = \int p(x) \log \frac{p(x)}{q(x)} dx \quad (1.12)$$

La divergence de Kullback [CT91] est une fonction positive et elle s'annule que si et seulement si les deux distributions $p(x), q(x)$ sont égales.

Considérons maintenant la divergence de Kullback entre $p(x)$ et $\prod_{i=1}^n p(x_i)$, nous obtenons l'information mutuelle de X :

$$I(X) = \int p(x) \log \frac{p(x)}{\prod_{i=1}^n p(x_i)} dx \quad (1.13)$$

En effet, l'information mutuelle est une mesure naturelle de l'indépendance, elle est positive et s'annule lorsque (1.11) est vérifiée.

Il existe d'autres mesures de l'indépendance, qui se basent sur la théorie de l'information, ces dernières seront détaillées dans le chapitre suivant.

1.5 Optimisation dans la séparation des sources

Une fois que le critère de séparation est déterminé, plusieurs facteurs sont pris en considération pour effectuer la séparation.

Tout d'abord, il est important de rappeler que l'étape du blanchiment est une bonne initialisation dans la séparation des sources. Cette étape est réalisée soit en blanchissant les observations en premier, soit en ajoutant une contrainte de blanchiment aux critères de la séparation. Ensuite, les sources peuvent être estimées simultanément ou par extractions successives. Enfin, l'optimisation des critères de séparation se résout soit en utilisant des techniques d'optimisation tels que la descente du gradient, notamment, le gradient relatif ou naturel [CL96, AC98], soit en appliquant des techniques de diagonalisation conjointe de matrice.

Dans cette section, nous nous intéresserons aux techniques du gradient. Pour ce qui est de la diagonalisation conjointe, nous laisserons le lecteur se référer à [Deg01] pour plus d'informations.

1.5.1 La descente du gradient

Dans la majorité des problèmes d'optimisation telle que la séparation de sources, la recherche de solutions s'effectue en minimisant ou maximisant une fonction avec ou sans certaines contraintes. Dans cette section, nous considérons le problème de minimisation, plus précisément nous allons exposer l'approche de la descente du gradient [Hay94].

Nous souhaitons estimer le vecteur W par minimisation d'une fonction $J(W)$ de la forme $J(W) = E[f(W)]$. En effet, on part d'un vecteur initial $W(0)$, ensuite on calcule le gradient pour ce vecteur, puis on se déplace avec une distance appropriée dans la direction opposée du gradient.

Pour $t = 1, 2, \dots$, la règle à mettre à jour à chaque itération est de la forme suivante :

$$W(t) = W(t-1) - \alpha(t) \frac{\partial J(W)}{\partial W} \Big|_{W=W(t-1)} \quad (1.14)$$

où α est le pas de déplacement. À chaque instant t , W est remis à jour en fonction de la direction du gradient pondéré par α .

Une solution de W est obtenue à la convergence de (1.14), c'est-à-dire lorsque les valeurs de $W(t)$ sont égales aux valeurs de $W(t-1)$, la distance entre ces deux vecteurs est nulle et par conséquent, l'équation n'évolue plus. La solution obtenue peut correspondre à un minimum local ou global.

Le problème avec ce genre d'approche réside dans le fait que, la vitesse de convergence dépend du choix de α : en effet, un α trop petit conduit à une convergence lente. Aussi, un α trop grand mène à une certaine instabilité qui rend la convergence impossible. L'utilisation d'un pas variable permet de résoudre en partie ce problème en calculant un pas optimal permettant d'accélérer la convergence. La méthode de Newton est souvent une meilleure alternative.

1.5.2 Méthode de Newton

En analyse numérique, un grand nombre de méthodes ont été introduites et ont été jugées plus performantes que la descente du gradient. L'avantage d'utiliser ces méthodes dans la séparation de sources est dû à leur rapidité de convergence en terme de nombre d'itérations requis. Cependant, elles souffrent d'une grande complexité dans le calcul par itération.

Considérons le développement de Taylor de la fonction $J(W)$:

$$J(W') = J(W) + \left[\frac{\partial J(W)}{\partial W} \right]^T (W' - W) + \frac{1}{2} (W' - W)^T \frac{\partial^2 J(W)}{\partial W^2} (W' - W) + \dots \quad (1.15)$$

Pour minimiser la fonction $J(W)$, nous cherchons un nouveau vecteur W' tel que la valeur de $J(W)$ soit petite. Notons $W' - W$ par ΔW , nous allons minimiser la fonction suivante :

$$J(W') - J(W) = \left[\frac{\partial J(W)}{\partial W} \right]^T \Delta W + \frac{1}{2} \Delta W^T \frac{\partial^2 J(W)}{\partial W^2} \Delta W$$

Le gradient de cette fonction est égal à $\left[\frac{\partial J(W)}{\partial W} \right] + \frac{\partial^2 J(W)}{\partial W^2} \Delta W$. Notons que la matrice hessienne est symétrique, si cette dernière est aussi définie positive alors le graphique de la fonction a une forme parabolique, ainsi, le minimum est atteint lorsque le gradient est nul :

$$\Delta W = - \left[\frac{\partial^2 J(W)}{\partial W^2} \right]^{-1} \frac{\partial J(W)}{\partial W} \quad (1.16)$$

Par conséquent, l'itération du second ordre est de la forme suivante :

$$W' = W - \left[\frac{\partial^2 J(W)}{\partial W^2} \right]^{-1} \frac{\partial J(W)}{\partial W} \quad (1.17)$$

où, l'on calcule le gradient et le hessien au vecteur initial W .

Cet algorithme est appelé la méthode de Newton, cette méthode est souvent efficace lorsqu'il s'agit d'un problème de minimisation, elle est caractérisée par sa rapidité et sa convergence quadratique. Cependant, son calcul est plutôt complexe. En effet, l'inverse de la matrice hessienne n'est pas souvent calculable, pour remédier à cela, certains paramètres sont ajoutés à l'algorithme avant le calcul de la matrice hessienne inverse [AU94]. En pratique, nous utilisons souvent des approximations de la méthode de Newton, par exemple dans l'ACI, l'algorithme FastICA fait appel à une approximation de Newton qui permet d'extraire les signaux sources et qui apporte une convergence rapide comparée au nombre de calculs requis [Hay94].

1.5.3 La fonction de Lagrange

Dans les approches d'optimisation citées plus haut, nous avons parlé du cas de minimisation sans contraintes. Cependant, il existe plusieurs techniques visant à résoudre le cas

avec contraintes.

Ce problème est modélisé comme suit :

$$\min J(W) \quad \text{sous contraintes : } H_i(W) = 0 \quad i = 1, \dots, k \quad (1.18)$$

où $H_i(W)$ est l'ensemble des contraintes.

Parmi les techniques les plus répandues dans ce domaine figure la méthode de Lagrange.

La fonction de Lagrange [Hay94] est de la forme suivante : $L(W, \lambda_1, \dots, \lambda_k) = J(W) + \sum_{i=1}^k \lambda_i H_i(W)$

$\lambda_1, \dots, \lambda_k$ sont appelés les multiplicateurs de Lagrange. Le nombre k correspond au nombre des contraintes.

Si W est une solution recherchée, il existe un vecteur λ_i tel que le gradient de L est nulle :

$$\frac{\partial J(W)}{\partial W} + \sum_{i=1}^k \lambda_i \frac{\partial H_i(W)}{\partial W} = 0 \quad (1.19)$$

Pour résoudre ce système d'équations, il suffit d'appliquer la méthode de Newton ou toute autre technique semblable.

1.5.4 Déflation

Les méthodes d'optimisation présentées plus haut permettent d'extraire une source S_i à partir du vecteur du mélange X par l'estimation d'un vecteur W_i^T , fournissant une estimation Y_i :

$$Y_i = W_i^T X \quad (1.20)$$

Où W_i^T est la i ème ligne de la matrice de séparation \mathbf{W} .

En général, nous nous intéressons à estimer plusieurs composantes indépendantes. L'exécution de cet algorithme plusieurs fois, avec différents vecteurs initiaux nous permet d'obtenir différents vecteurs de séparation W_i^T , cela dit certains de ces vecteurs peuvent

converger vers le même minimum.

Pour remédier à cela, Hyvärinen [Hay94] propose une approche de déflation orthogonale. Tout d'abord, on préconditionne le mélange par l'étape de blanchiment $Z = \mathbf{W}X$, ensuite la séparation est effectuée comme décrit précédemment.

La déflation orthogonale utilise la méthode de Gram-Schmidt : supposons que nous avons estimé p vecteurs W_1, \dots, W_p , l'étape suivante consistera à calculer le vecteur W_{p+1} en minimisant le critère de séparation $J(W_{p+1})$, ainsi après chaque itération, nous soustrayons la projection $(W_{p+1}^T W_j)W_j, j = 1, \dots, p$ de W_{p+1} :

$$W_{p+1} \leftarrow W_{p+1} - \sum_{j=1}^p (W_{p+1}^T W_j)W_j \quad (1.21)$$

Puis, nous normalisons le vecteur tel que $W_{p+1} \leftarrow \frac{W_{p+1}}{\|W_{p+1}\|}$.

1.5.5 Orthogonalisation symétrique

Cet exemple de déflation permet d'estimer au fur et à mesure les vecteurs de la matrice de séparation \mathbf{W} . Ceci dit, si des erreurs ont été commises sur les premières lignes de la matrice \mathbf{W} elles dégraderont l'estimation des lignes suivantes, ce qui affaiblira la séparation.

Toutefois, Hyvärinen [Hay94] propose une approche alternative appelée l'orthogonalisation symétrique. L'orthogonalisation symétrique de la matrice \mathbf{W} est effectuée de la façon suivante :

$$\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{-1/2}\mathbf{W} \quad (1.22)$$

La matrice $(\mathbf{W}\mathbf{W}^T)^{-1/2}$ est obtenue en effectuant la décomposition spectrale en valeurs propres de $\mathbf{W}\mathbf{W}^T = \mathbf{E}\text{diag}(d_1, \dots, d_n)\mathbf{E}^T$.

Pour effectuer l'orthogonalisation symétrique, il existe certaines méthodes itératives qui ne nécessitent pas la décomposition en valeurs propres. Un exemple simple de ce fait est l'algorithme suivant :

1. $\mathbf{W} \leftarrow \mathbf{W}/\|\mathbf{W}\|$

2. $\mathbf{W} \leftarrow \frac{3}{2}\mathbf{W} - \frac{1}{2}\mathbf{W}\mathbf{W}^T\mathbf{W}$

3. Retour à l'étape 2 si $\mathbf{W}\mathbf{W}^T$ n'est pas proche de la matrice identité \mathbf{I}

La convergence de cette itération est prouvée dans [Hyv99].

CHAPITRE 2

Analyse en composantes indépendantes

Parmi les différentes approches de la séparation des sources, nous nous sommes principalement intéressés à l'analyse en composantes indépendantes. Comme dans la majorité des travaux sur l'ACI, nous avons adopté le modèle linéaire instantané non bruité. Dans ce chapitre nous allons présenter les hypothèses du modèle, puis nous étudierons les différentes propriétés statistiques utilisées par cette méthode. Enfin, les mesures de séparation utilisées par cette approche seront abordées.

2.1 Définition de l'analyse en composantes indépendantes

2.1.1 Présentation du modèle

L'analyse en composantes indépendantes consiste en la recherche d'une transformation des données observées en composantes indépendantes. L'illustration classique de cette méthode est le problème de la soirée cocktail (the cocktail party), décrit dans le chapitre précédent. Rappelons que la problématique de la soirée cocktail conduit à résoudre l'équation suivante :

$$X_i = a_{i1}S_1 + a_{i2}S_2 + \dots + a_{im}S_m \quad \text{pour } i = 1, \dots, n \quad \text{et } j = 1, \dots, m \quad (2.1)$$

Les a_{ij} sont les coefficients de la matrice de mélange, les S_j sont les composantes indépendantes, ce sont des variables dites latentes. En d'autres termes, ce sont des facteurs qui influencent les observations, ces facteurs ne sont pas mesurables. Par l'ACI, nous tentons de les extraire et de leur attribuer une signification physique.

Dans ce modèle, seulement les observations x_i sont connues, ainsi, en s'appuyant sur certaines hypothèses, nous cherchons à estimer les composantes indépendantes et les coefficients du mélange a_{ij} .

Dans la séparation des sources, le modèle défini ci-dessus est considéré comme modèle de base. Les X_i et les S_j sont supposés être des variables aléatoires plutôt que des séries temporelles ou des signaux, de plus, la notion du temps est négligée dans cette formulation, d'où l'appellation **modèle linéaire instantané**.

L'écriture matricielle du modèle s'exprime sous la forme :

$$X = AS \quad (2.2)$$

où \mathbf{A} est la matrice du mélange.

Dans le cadre de ce mémoire, nous nous limitons à ce cas général. Toutefois, selon le domaine d'application, certaines contraintes, telles que le bruit, peuvent être ajoutées pour ajuster le modèle.

2.1.2 Contraintes et restrictions du modèle

Pour estimer le modèle, certaines hypothèses doivent être prises en considération, notamment :

1. L'indépendance des variables sources S_j . Il s'agit de l'hypothèse principale de l'analyse en composantes indépendantes. Soit un vecteur aléatoire $X = [X_1, \dots, X_n]^T$. Les composantes du vecteur sont dites statistiquement indépendantes si la densité conjointe $p(x_1, \dots, x_n)$ s'écrit sous forme du produit des densités marginales $p(x_i)$:

$$p(x) = \prod_{i=1}^n p(x_i) \quad (2.3)$$

2. Les composantes indépendantes doivent suivre une distribution non-gaussienne. En effet, la distribution gaussienne se caractérise par la nullité de ses cumulants d'ordre supérieur à deux, ainsi, l'indépendance est équivalente à la décorrélation et par conséquent, l'ACI revient à faire une ACP. Cela dit, le modèle reste fiable si au plus une composante est gaussienne.
3. La matrice de mélange est une matrice carrée. En effet, nous supposons que le nombre d'observations n est égal à celui des composantes indépendantes m . Cette condition simplifie le modèle et implique l'existence de la matrice \mathbf{B} , qui est l'inverse de \mathbf{A} . Dans le cas où $\det(\mathbf{A}) \neq 0$ nous avons :

$$S = \mathbf{B}X \quad (2.4)$$

Il faut noter qu'il est possible d'extraire plus de sources que d'observations [OF96, OF97, OL99, LS00, HI02]. Réciproquement, dans le cas où le nombre d'observations est supérieur à celui des sources, nous pouvons utiliser des techniques de réduction de dimension tels que l'ACP.

En plus de ces conditions d'identification, il existe deux indéterminations dans le modèle. Ces dernières sont l'indétermination de permutation et celle du facteur d'échelle expliquées dans le chapitre précédent.

2.2 Propriétés statistiques de l'ACI

Dans cette section, nous exposons les différentes propriétés statistiques utilisées dans la séparation, en particulier dans l'analyse en composantes indépendantes.

2.2.1 Lien entre la décorrélation et le blanchiment

Soit le vecteur $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ la moyenne de ses composantes est définie par :

$$\mu_i = E\{X_i\} = \int x_i p(x_i) dx_i, \quad i = 1, 2$$

La variance de X_i est donnée par :

$$\begin{aligned} \text{Var}(X_i) &= E\{(X_i - \mu_i)^2\} \\ &= \int (x_i - \mu_i)^2 p(x_i) dx_i \end{aligned}$$

où $p(x_i)$ est la densité de la variable X_i pour $i = 1, 2$.

La covariance entre les composantes du vecteur X est définie par :

$$\text{Cov}(X_1, X_2) = E\{(X_1 X_2)\} - E\{X_1\}E\{X_2\}$$

Les variables aléatoires X_1 et X_2 sont dites décorrélées si et seulement si leur covariance est nulle, on écrit alors :

$$\text{Cov}(X_1, X_2) = E\{(X_1 X_2)\} - E\{X_1\}E\{X_2\} = 0 \quad (2.5)$$

Ainsi, la matrice de covariance est une matrice diagonale telle que :

$$\mathbf{C}_X = \begin{bmatrix} \text{Var}(X_1) & 0 \\ 0 & \text{Var}(X_2) \end{bmatrix}$$

Supposons maintenant que X soit un vecteur décorrélé (ses composantes ont une covariance nulle), de moyenne nulle et de variance unité. Il en découle que sa matrice de covariance est égale à la matrice identité :

$$\mathbf{C}_X = \mathbf{I} \quad (2.6)$$

Ce vecteur X est appelé vecteur blanchi, cette propriété est invariante par rapport aux transformations orthogonales. Soit l'équation suivante :

$$Y = \mathbf{P}X \quad (2.7)$$

où \mathbf{P} est une matrice orthogonale ($\mathbf{P}^T \mathbf{P} = \mathbf{P} \mathbf{P}^T = \mathbf{I}$) et X un vecteur blanchi.

$$E\{Y\} = E\{\mathbf{P}X\} = \mathbf{P}E\{X\} = 0 \quad (2.8)$$

et :

$$\mathbf{C}_Y = E\{\mathbf{P}X(\mathbf{P}X)^T\} = \mathbf{P}E\{XX^T\}\mathbf{P}^T \quad (2.9)$$

$$= \mathbf{P}\mathbf{C}_X\mathbf{P}^T = \mathbf{I} \quad (2.10)$$

Le vecteur Y est bel et bien blanchi. En effet, le blanchiment est une simple transformation linéaire d'un vecteur X en un vecteur Z tel que :

$$Z = \mathbf{V}X \quad (2.11)$$

Une manière d'estimer la matrice \mathbf{V} est la décomposition en valeur et vecteurs propres : Posons, $E\{XX^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^T$ où \mathbf{E} est la matrice orthogonale des vecteurs propres de $E\{XX^T\}$, \mathbf{D} est la matrice diagonale des valeurs propres. Le blanchiment peut s'établir par la matrice \mathbf{V} tel que : $\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T$.

2.2.2 La gaussianité dans l'ACI

Comme nous l'avons énoncé plus haut, la non-gaussianité est une hypothèse fondamentale pour l'estimation du modèle, dans cette section nous allons détailler les raisons de cette restriction.

Considérons le vecteur gaussien centré réduit $S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$, S_1 et S_2 deux variables aléatoires indépendantes, leur densité conjointe est exprimée comme suit :

$$\begin{aligned} p(s_1, s_2) &= \frac{1}{2\pi} \exp\left(-\frac{s_1^2 + s_2^2}{2}\right) \\ &= \frac{1}{2\pi} \exp\left(-\frac{\|s\|^2}{2}\right) \end{aligned}$$

Rappelons que dans le cas d'une transformation linéaire non singulière, telle que $X = \mathbf{A}S$, la fonction de densité s'écrit sous la forme :

$$p(x) = \frac{1}{|\det\mathbf{A}|} p_S(\mathbf{A}^{-1}x)$$

Ainsi, la fonction de densité conjointe des mélanges X_1 et X_2 est exprimée comme :

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{A}^T x\|^2}{2}\right) |\det(\mathbf{A}^T)|$$

Dû à l'orthogonalité de la matrice \mathbf{A} nous avons :

$$- \mathbf{A}^{-1} = \mathbf{A}^T$$

$$- \|\mathbf{A}^T x\|^2 = \|x\|^2$$

$$- |\det(\mathbf{A})| = 1$$

Ce qui implique que :

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{\|x\|^2}{2}\right)$$

La fonction de densité des mélanges est semblable à celle des composantes indépendantes, nous constatons aussi que la matrice de mélange \mathbf{A} n'intervient pas dans la formule de la fonction de densité conjointe des mélanges. Et donc, la résolution de ce problème est impossible, car la matrice de mélange ne pourra pas être estimée.

Une autre façon de voir cela est d'introduire le lien entre l'indépendance et la décorrélation dans le cas gaussien :

Soit $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ un vecteur gaussien de moyenne nulle. Sa matrice de covariance \mathbf{C}_X est définie comme :
$$\begin{bmatrix} \sigma_{X_1}^2 & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \sigma_{X_2}^2 \end{bmatrix}$$

La densité conjointe de X est :

$$p(x_1, x_2) = \frac{1}{(2\pi)\sqrt{\det\mathbf{C}_X}} \exp\left(-\frac{1}{2}X^T\mathbf{C}_X^{-1}X\right)$$

On a vu précédemment que lorsque les variables X_1 et X_2 sont décorrélées, leur covariance est égale à zéro, ainsi, $\det \begin{bmatrix} \sigma_{X_1}^2 & 0 \\ 0 & \sigma_{X_2}^2 \end{bmatrix} = \sigma_{X_1}^2 \sigma_{X_2}^2$.

Par conséquent :

$$\begin{aligned} p(x_1, x_2) &= \frac{1}{\sqrt{2\pi}\sigma_{X_1}} \exp\left(-\frac{x_1^2}{2\sigma_{X_1}^2}\right) \times \\ &\quad \frac{1}{\sqrt{2\pi}\sigma_{X_2}} \exp\left(-\frac{x_2^2}{2\sigma_{X_2}^2}\right) \\ &= p(x_1)p(x_2) \end{aligned}$$

Il en résulte donc que X_1 et X_2 sont indépendants.

Dans le cadre d'une distribution gaussienne, l'indépendance est équivalente à la décorrélation, par conséquent, l'ACI peut être remplacée par une simple analyse en composantes principales.

2.2.3 Indépendance et gaussianité

L'estimation de la matrice de mélange peut s'effectuer en se basant sur la relation existante entre la gaussianité et l'indépendance. L'idée est simple, pour extraire une source indépendante, il suffit de chercher sa distribution la plus éloignée possible d'une distribution gaussienne. Cette façon de procéder tire sa justification du théorème de la limite centrale.

Théorème de la limite centrale : Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes définies sur le même espace de probabilité, suivant la même loi E de moyenne μ et de variance σ^2 . Considérons la somme $S_X = X_1 + X_2 + \dots + X_n$. Alors, lorsque n est grand, S_X suit approximativement une loi normale d'espérance $n\mu$ et de variance $n\sigma^2$.

Reprenons le modèle de l'ACI :

$$X = \mathbf{A}S \quad (2.12)$$

les composantes sont indépendantes et identiquement distribuées, nous souhaitons, estimer une composante indépendante Y_i , l'équation devient :

$$Y_i = B_i^T X = B_i^T \mathbf{A}S \quad (2.13)$$

où $\mathbf{B} = \mathbf{A}^{-1}$, le vecteur B_i est la i ème colonne de la matrice \mathbf{B} et X représente le vecteur du mélange.

En remplaçant $B_i^T \mathbf{A}$ par le vecteur Q_i^T , nous obtenons l'expression suivante :

$$Y_i = Q_i^T S = \sum_j q_{ij} S_j \quad (2.14)$$

Suite à ce développement nous remarquons que Y_i est une combinaison linéaire de composantes indépendantes S_j . La composante Y_i est unique ce qui signifie que le vecteur Q_i^T correspond à un vecteur ayant un seul élément non nul ($Q_i^T = [0, \dots, 1, \dots, 0]$ tous les éléments de ce vecteur sont nuls sauf le i ème qui correspond à 1).

En pratique, nous ne pouvons pas trouver la valeur exacte du vecteur B_i , néanmoins, nous pouvons fournir une bonne estimation de ce vecteur. D'après le théorème de la limite centrale, la somme de variables indépendantes tend vers une loi normale. Par conséquent, $Y_i = Q_i^T S$ est plus gaussien que S_j et il devient moins gaussien si le vecteur Q_i^T a un seul élément non nul. Ceci revient à dire que **maximiser la non-gaussianité** de $B_i^T X$ correspond à trouver **une composante indépendante**.

L'idée est donc d'estimer les composantes Y_i de manière à ce que leur densité de probabilité soit la plus éloignée possible d'une distribution gaussienne. On dit qu'on cherche à maximiser leur non-gaussianité.

2.3 Réduction de la dimension de données

L'ACI est basée sur le critère d'indépendance entre les sources, leur matrice de covariance est diagonale. La décorrélation ou l'analyse en composantes principales a pour objectif d'annuler la corrélation existante entre les observations, ceci revient à annuler la covariance. Cependant, la décorrélation n'entraîne pas l'indépendance. L'analyse en composantes principales a été essentiellement développée pour la réduction de la dimension, nous allons donc voir d'une part, les principes de cette méthode, et d'autre part, l'utilité d'une telle méthode dans la séparation des sources.

2.3.1 Analyse en composantes principales

L'analyse en composantes principales (ACP) [CC80] est une technique d'analyse de données multidimensionnelles qui permet d'examiner les liens entre les individus et les variables. Soit le vecteur $X = [X_1, \dots, X_p]^T$ de dimension p , de moyenne μ et de matrice de covariance \mathbf{C}_X . Le but de l'ACP est de déterminer des nouvelles variables Y_1, \dots, Y_p , combinaisons linéaires des variables initiales, tout en maximisant la variance. Les Y_i sont appelées des composantes principales et elles sont disposées dans l'ordre décroissant de la variance.

Chaque composante Y_i est une combinaison linéaire des X_i telle que :

$$Y_i = a_{1i}X_1 + a_{2i}X_2 + \dots + a_{pi}X_p \quad (2.15)$$

$A_i = [a_{1i}, \dots, a_{pi}]^T$ est un vecteur de constantes.

La variance de Y_i s'écrit comme :

$$\begin{aligned} \text{Var}(Y_i) &= \text{Var}(A_i^T X) \\ &= A_i^T \text{Var}(X) A_i \\ &= A_i^T \mathbf{C}_X A_i \end{aligned}$$

La première composante principale Y_1 est choisie de telle sorte que cette dernière ait la plus grande variance possible, cette contrainte revient à résoudre le problème d'optimisation suivant :

$$\begin{array}{ll} \text{Maximiser} & A_1^T \mathbf{C}_X A_1 \\ \text{Sous la contrainte} & A_1^T A_1 = 1 \end{array}$$

ceci revient à maximiser la quantité : $A_1^T \mathbf{C}_X A_1 - \lambda(A_1^T A_1 - 1)$, en dérivant par rapport au vecteur A_1 et en égalant le gradient à 0 il vient :

$$(\mathbf{C}_X - \lambda \mathbf{I})A_1 = 0 \quad (2.16)$$

Cette équation a une solution A_1 non nulle si et seulement si la matrice $(\mathbf{C}_X - \lambda \mathbf{I})$ est singulière. Ainsi λ doit vérifier :

$$\det(\mathbf{C}_X - \lambda \mathbf{I}) = 0 \quad (2.17)$$

Une solution non nulle de cette équation existe si et seulement si λ est la valeur propre de \mathbf{C}_X .

La variance de la composante Y_1 peut s'écrire comme :

$$\begin{aligned} \text{Var}(A_1^T X) &= A_1^T \mathbf{C}_X A_1 \\ &= A_1^T \lambda \mathbf{I} A_1 \\ &= \lambda \end{aligned}$$

Étant donné que le but est de maximiser la variance, A_1 n'est autre que le vecteur propre de \mathbf{C}_X correspondant à la plus grande valeur propre $\lambda = \lambda_1$.

La seconde composante est obtenue de la même manière tout en rajoutant la condition d'orthogonalité $A_2^T A_1 = 0$.

Soit \mathbf{A} la matrice $(p \times p)$ des vecteurs propres et Y le vecteur de composantes principales, nous avons :

$$Y = \mathbf{A}^T X \quad (2.18)$$

la matrice de variance de Y est de la forme :

$$\mathbf{C}_Y = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

Puisque les composantes principales sont décorréélées leur matrice de variance est diagonale. Notons que la somme des variances des Y_i est donnée par : $\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \lambda_i = \text{trace}(\mathbf{C}_Y)$

Nous avons aussi : $\text{trace}(\mathbf{C}_Y) = \text{trace}(\mathbf{A}^T \mathbf{C}_X \mathbf{A}) = \sum_{i=1}^p \text{Var}(X_i)$.

Il en résulte une propriété importante. En effet, la somme des variances des observations initiales est égale à celle des composantes principales.

L'ACP est une transformation orthogonale basée sur la maximisation de la variance des nouvelles composantes et entraînant la réduction de la dimensionalité. Le principe de cette transformation est plus au moins similaire à celui de l'ACI. Toutefois, il existe quelques différences que nous énumérons comme suit :

- L'ACP et l'ACI sont deux méthodes d'analyse de données multidimensionnelles permettant la transformation des observations selon certains critères. L'ACP s'appuie sur la corrélation entre les variables, par conséquent elle mesure les moments d'ordre deux. Tandis que l'ACI repose sur l'indépendance, entre autre, elle fait appel à des moments d'ordre supérieur à deux.
- Dans le cadre de données gaussiennes, l'application de l'ACI ne rapporte aucune information de plus que l'ACP. Nous pouvons dire que l'ACI permet d'introduire de l'information non gaussienne dans la recherche des composantes.
- Dans l'ACP, les composantes sont ordonnées selon leur degré d'importance : la première composante correspond à celle qui rapporte le maximum d'information et ainsi de suite. En revanche, les composantes indépendantes ne disposent d'aucun ordre.
- L'ACP est toujours réalisable alors qu'une ACI réussie repose sur l'existence de variables latentes que l'on approxime par le modèle $X = AS$. Le défi de l'ACI est de trouver des composantes indépendantes significatives et faciles à interpréter.

2.3.2 Blanchiment

Maintenant, nous allons voir comment cette technique peut être introduite dans l'ACI. Soit le modèle général de l'ACI :

$$X = \mathbf{A}S \quad (2.19)$$

Supposons que le vecteur des observations X est blanchi tel que :

$$Z = \mathbf{V}X = \mathbf{V}\mathbf{A}S = \tilde{\mathbf{A}}S \quad (2.20)$$

Selon les propriétés du vecteur Z et l'hypothèse émise sur les sources S_j 1.9, nous pouvons écrire :

$$E\{ZZ^T\} = \tilde{\mathbf{A}}E\{SS^T\}\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I} \quad (2.21)$$

L'utilité du blanchiment réside dans le fait que la nouvelle matrice du mélange $\tilde{\mathbf{A}}$ est une matrice orthogonale. Par conséquent, nous estimons $n(n+1)/2$ paramètres au lieu de n^2 . Notons que l'ACP peut être utilisée comme un moyen de blanchiment de données. Dans ce cas, la matrice \mathbf{V} se définit comme :

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^T \quad (2.22)$$

\mathbf{E} est la matrice des vecteurs propres de la matrice de covariance des \mathbf{C}_X et \mathbf{D} est la matrice diagonale des valeurs propres.

Il est facile de vérifier que Z est bel et bien blanchi :

$$\begin{aligned} E\{ZZ^T\} &= \mathbf{V}E\{XX^T\}\mathbf{V}^T \\ &= \mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{E}\mathbf{D}\mathbf{E}^T\mathbf{E}\mathbf{D}^{-1/2} \\ &= \mathbf{I} \end{aligned}$$

Comme le montre la figure 2.1 ci-dessous, les composantes blanchies sont des composantes orthogonales. Par ce fait, les rendre indépendantes (figure 2.2) peut se faire simplement

par le choix d'une rotation appropriée.

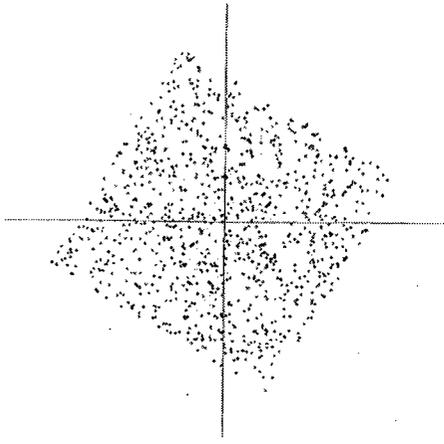


Figure 2.1 – Distribution conjointe du mélange blanchi de deux composantes indépendantes uniformes s_1 et s_2

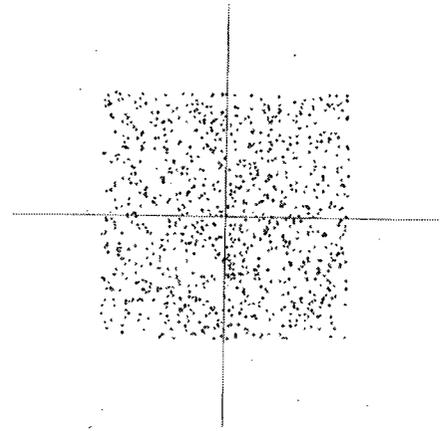


Figure 2.2 – Distribution conjointe de deux composantes indépendantes uniformes s_1 et s_2

Ces techniques de réduction de dimensions sont utilisées comme prétraitement afin de faciliter la séparation. Dorénavant, l'appellation Z correspondra au vecteur d'observation blanchi.

2.4 Mesures de séparation dans l'ACI

Après avoir posé le modèle de l'ACI, en avoir exposé les différentes hypothèses, nous allons nous pencher sur les mesures utilisées pour effectuer la séparation. Ces dernières se décomposent en deux catégories, une première famille basée sur l'approche de la théorie de l'information et une seconde qui repose sur le calcul de certaines statistiques descriptives.

2.4.1 Entropie

La notion de l'entropie a été introduite dans les années 50 par Shannon [CT91, Iha93]. Soit X une variable aléatoire discrètes ayant une probabilité $p(x)$, $X \in \mathcal{X}$ avec \mathcal{X} un ensemble dénombrable. Alors, l'entropie de X , notée $H(X)$ est définie par :

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log_2 \left(\frac{1}{p(x)} \right) \quad (2.23)$$

Dans le cas continu, l'entropie s'exprime comme :

$$H(X) = \int p(x) \log_2 \left(\frac{1}{p(x)} \right) dx \quad (2.24)$$

L'entropie d'une variable aléatoire est une mesure quantitative de l'incertitude, ou de la quantité d'information, associée aux valeurs prises par la variable aléatoire.

L'entropie possède de nombreuses propriétés traitées dans les livres de théorie de l'information [CT91, Iha93], ici nous énumérons celles dont nous avons besoin dans la séparation :

– $H(X) \geq 0$.

L'égalité aura lieu si $p(x_i)$ est nulle en tous éléments i sauf en un point i' pour lequel $p(x_{i'}) = 1$. En d'autres termes, l'entropie est maximale lorsque les éléments de X sont équiprobables (distribution uniforme) et nulle lorsque l'événement est certain.

– L'entropie de $Y = \mathbf{M}X$, combinaison linéaires de X , est :

$$H(Y) = H(X) + \log |\det \mathbf{M}| \quad (2.25)$$

Notons que l'entropie est conservée par les transformations orthogonales.

– Soit X un vecteur gaussien $N(0, \mathbf{C}_X)$, son entropie $H(X)$ est donnée par :

$$H_N(X) = \frac{n}{2} \log(2\pi e) + \frac{1}{2} \log(\det(\mathbf{C}_X)) \quad (2.26)$$

Où n est la dimension de X . En effet, rappelons que la densité d'une loi gaussienne $N(0, \mathbf{C}_X)$ s'écrit :

$$p(x) = (2\pi)^{-n/2} (\det \mathbf{C}_X)^{-1/2} \exp\left(-\frac{1}{2} X^T \mathbf{C}_X^{-1} X\right) \quad (2.27)$$

Le logarithme de cette densité vaut :

$$\log(p(x)) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det \mathbf{C}_X) - \frac{1}{2} X^T \mathbf{C}_X^{-1} X \quad (2.28)$$

$X^T \mathbf{C}_X^{-1} X$ peut s'écrire comme $\text{trace}(X^T X \mathbf{C}_X^{-1})$, on a :

$$E(X^T \mathbf{C}_X^{-1} X) = E(\text{trace}(X^T X \mathbf{C}_X^{-1})) = \text{trace}(E(X X^T) \mathbf{C}_X^{-1}) = \text{trace}(\mathbf{I}) = n .$$

En remplaçons l'expression de la densité gaussienne dans la formule de l'entropie nous obtenons l'équation (2.26) .

- L'entropie d'une distribution gaussienne a la plus grande valeur parmi toutes les distributions de même matrice de covariance.

Pour voir cela, nous allons considérer la divergence de Kullback entre la densité gaussienne $g(x)$ et la densité $p(x)$ de même matrice de covariance que $g(x)$:

$$\begin{aligned} 0 &\leq D(p, g) \\ &= \int p \log\left(\frac{p}{g}\right) \\ &= -H(p) - \int p \log(g) \\ &= -H(p) - \int g \log(g) \\ &= -H(p) + H_N(g) \end{aligned}$$

$\int g \log(g) = \int p \log(g)$ puisque g et p ont les mêmes moments de la forme quadratique $\log(g(x))$ [CT91].

Une notion importante dans l'approximation de l'entropie est celle de l'entropie maximale. Soit X une variable aléatoire, supposons que l'information disponible sur cette variable est exprimée comme suit :

$$\int p(x) F^i(x) dx = c_i, \quad \text{pour } 1 \leq i \leq m \quad (2.29)$$

L'entropie maximale cherche à maximiser l'entropie de cette variable tout en respectant cette information fournie. Ainsi, nous cherchons à résoudre le système d'équations suivant :

$$\begin{array}{l} \text{Maximiser :} \\ \text{Sous les contraintes :} \end{array} \quad \begin{array}{l} H(p) \\ \int p(x) dx \\ \int p(x) F^i(x) dx \end{array} \quad = \quad \begin{array}{l} - \int p(x) \log p(x) dx \\ 1 \\ c_i \text{ pour } 1 \leq i \leq m \end{array}$$

Le lagrangien à maximiser est :

$$H(p) + \lambda_0 \left(\int p(x) dx - 1 \right) + \sum_{i=1}^m \lambda_i \left(\int p(x) F^i(x) dx - c_i \right) \quad (2.30)$$

Les coefficients $\lambda_0, \lambda_1, \dots, \lambda_m$ sont les multiplicateurs de Lagrange.

L'entropie maximale est de la forme :

$$f(x) = \exp(\lambda_0 - 1 + \sum_{i=1}^m \lambda_i F^i(x)) \quad (2.31)$$

Admettons que les contraintes du système d'équations soient de la forme suivante :

$$\begin{aligned} \int xp(x)dx &= 0 \\ \int x^2p(x)dx &= \sigma^2 \end{aligned}$$

L'entropie maximale s'écrit :

$$f(x) = \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2) \quad (2.32)$$

En remplaçant les constantes par des valeurs appropriées, nous constatons que cette densité correspond à la densité d'une distribution gaussienne.

Dans la séparation des sources, l'utilisation directe de l'entropie est très difficile d'où le recours à certaines approximations. L'approximation par l'entropie maximale fournit des estimateurs robustes, cette approche apporte une borne supérieure de l'entropie et étant donné que le but de la séparation des sources est de minimiser l'entropie, la minimisation de son maximum est équivalente dans ce cas à la minimisation de la vraie entropie.

2.4.2 Information mutuelle

L'information mutuelle d'un vecteur aléatoire X s'écrit :

$$I(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i) - H(X) \quad (2.33)$$

Nous pouvons interpréter l'information mutuelle en utilisant la notion de l'entropie. En effet, il s'agit de mesurer la différence entre l'entropie donnée par le vecteur X et celle donnée par ses composantes séparément.

L'information mutuelle peut être exprimée au sens de la divergence de Kullback comme suit :

$$D(p, q) = \int p(x) \log \frac{p(x)}{q(x)} dx \quad (2.34)$$

où $p(x)$ est la fonction de densité du vecteur X et $q(x)$ est le produit des densités marginales $\prod_{i=1}^n p(x_i)$. Par ce fait, l'information mutuelle peut être utilisée pour mesurer l'indépendance entre les variables aléatoires. Mentionnons quelques propriétés de cette quantité :

- L'information mutuelle est symétrique $I(X, Y) = I(Y, X)$, au sens usuellement admis voir [CT91].
- $I(X) \geq 0$ l'égalité n'ayant lieu que si et seulement si les variables X_i sont indépendantes.

La preuve découle principalement de l'inégalité de Jensen :

$$\begin{aligned} -D(p, q) &= \int p(x) \log \frac{q(x)}{p(x)} dx \\ &\leq \log \int p(x) \frac{q(x)}{p(x)} dx \\ &= \log \int q(x) dx \\ &\leq \log 1 = 0 \end{aligned}$$

L'information mutuelle, telle qu'elle est définie, est une mesure naturelle de l'indépendance. Malheureusement, elle l'est au niveau théorique. En pratique, d'autres mesures, qui font éventuellement appel à des approximations de la densité, sont utilisées.

2.4.3 Le coefficient d'aplatissement

Le coefficient d'aplatissement est le cumulante d'ordre 4 d'une variable aléatoire et il est défini par :

$$\text{kurtosis}(X) = K(X) = \frac{E\{(X - \mu)^4\}}{(E\{(X - \mu)^2\})^2} - 3 \quad (2.35)$$

où, μ est la moyenne de la variable aléatoire X . Pour une variable centrée et de variance unité, le coefficient d'aplatissement s'écrit :

$$K(X) = E\{X^4\} - 3 \quad (2.36)$$

Pour une distribution gaussienne le coefficient d'aplatissement est nul. En revanche, un coefficient négatif indique que la distribution est sous-gaussienne, alors qu'un coefficient positif indique une distribution sur-gaussienne.

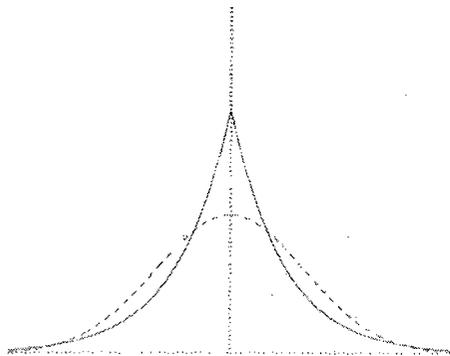


Figure 2.3 – Exemple de distribution sur-gaussienne : la distribution de laplace

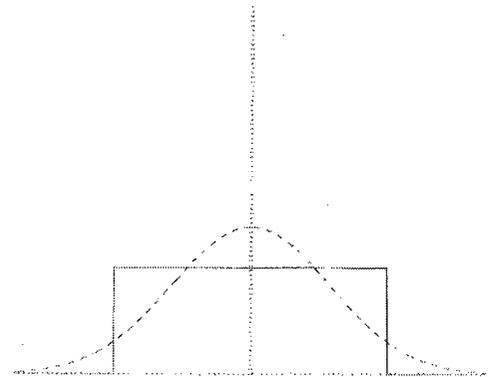


Figure 2.4 – Exemple de distribution sous-gaussienne : la distribution uniforme

De plus, le coefficient d'aplatissement est additif, $\kappa(X + Y) = \kappa(X) + \kappa(Y)$ et pour tout coefficient β , nous pouvons l'écrire comme, $\kappa(\beta X) = \beta^4 \kappa(X)$.

La question qui se pose désormais, est comment appliquer ce coefficient dans la séparation ?

Considérons tout d'abord le modèle $Y_i = W_i^T Z$ où Z est le vecteur des mélanges. Rappelons que, comme vu plus haut, la maximisation de la non-gaussianité nous permet d'estimer les composantes indépendantes. Ainsi, estimer la composante Y_i revient à résoudre le problème d'optimisation suivant :

$$\begin{aligned} \text{Max } |\kappa(Y_i)| &= |\kappa(W_i^T Z)| \\ \text{sous la contrainte } E\{SS^T\} &= \mathbf{I} \end{aligned}$$

Malgré sa simplicité, le coefficient d'aplatissement demeure très sensible aux points extrêmes. Toutefois, il existe d'autres mesures, plus robustes, de la non-gaussianité telles que la négentropie.

2.4.4 Négentropie

Soit X un vecteur aléatoire admettant $g(x)$ pour densité. Notons par $p(x)$ la densité d'une variable gaussienne de même matrice de covariance que X . La négentropie associée à X est :

$$J(X) = \int g(x) \log \frac{g(x)}{p(x)} dx \quad (2.37)$$

Il est facile de voir que la négentropie mesure l'écart entre une distribution quelconque et celle de la distribution gaussienne. Une façon plus explicite d'exprimer cette mesure est d'utiliser l'entropie, nous avons :

$$J(X) = H_N(X) - H(X) \quad (2.38)$$

où H_N est l'entropie de la densité de la distribution normale $p(x)$.

Puisque la négentropie s'exprime sous la forme de la divergence de Kullback $D(g, p)$, nous pouvons conclure que la négentropie est toujours positive et elle s'annule si et seulement si le vecteur X est gaussien.

La négentropie est invariante par rapport aux transformations linéaires inversibles. En effet, considérons la transformation linéaire $Y = \mathbf{M}X$, avec $E\{YY^T\} = \mathbf{M}\mathbf{C}_X\mathbf{M}^T$ la négentropie est exprimée comme suit :

$$\begin{aligned}
 J(\mathbf{M}X) &= \frac{1}{2} \log |\det \mathbf{M}\mathbf{C}_X\mathbf{M}^T| + \frac{n}{2}[1 + \log 2\pi] - (H(X) + \log \det |\mathbf{M}|) \\
 &= \frac{1}{2} \log |\det \mathbf{C}_X| + 2\frac{1}{2} \log |\det \mathbf{M}| + \frac{n}{2}[1 + \log 2\pi] - (H(X) + \log \det |\mathbf{M}|) \\
 &= \frac{1}{2} \log |\det \mathbf{C}_X| + \frac{n}{2}[1 + \log 2\pi] - H(X) \\
 &= H_N(X) - H(X) = J(X)
 \end{aligned}$$

Où n est la dimension du vecteur X .

Dans ce mémoire, nous avons étudié le critère de séparation basé sur la maximisation de la non-gaussianité, en particulier celui de la maximisation de la négentropie. Ainsi, le chapitre suivant fera l'objet d'une étude détaillée de cette approche.

CHAPITRE 3

ACI par maximisation de la non-gaussianité

Depuis les années 1980, plusieurs algorithmes de l'ACI ont été proposés. Chacun a ses avantages et ses inconvénients. Les algorithmes que nous allons présenter dans ce chapitre ont la propriété de supprimer ou d'affaiblir la dépendance entre les composantes d'un vecteur aléatoire. Cependant, nous allons plutôt privilégier ceux qui facilitent le calcul de la matrice inverse de mélange, même si une légère dépendance demeure entre les variables aléatoires.

Il existe deux grandes familles d'algorithmes :

- Une famille de méthodes paramétriques, notamment, les méthodes à noyau, les méthodes à contrastes exacts et les méthodes directes. Ce sont des méthodes basées sur une fonction de contraste exact, ainsi le critère de l'indépendance est véridique et donc la matrice inverse de \mathbf{A} est plus précise [SJG07, M03, ST04].
- Une famille de méthodes classiques basées sur des mesures telles que le maximum de vraisemblance, la négentropie, l'information mutuelle, \dots . Cette famille est composée

de quatre catégories :

- La maximisation du caractère non gaussien (Hyvärinen, 1999).
- La maximisation de la vraisemblance notamment, l’algorithme Infomax (Bell et Senejowski, 1995).
- Les méthodes tensorielles, telles que l’algorithme JADE (Cardoso, 1990 – 1994) .
- La décorrélation non linéaire (Jutten et Héroult, 1987 et Cichocki-Unbehauen, 1992).

En général, ces méthodes classiques de séparation aveugle des sources réunissent deux aspects : une méthode statistique pour mesurer le critère de base de la séparation qui est l’indépendance et d’autre part une méthode algorithmique pour optimiser la mesure statistique. Dans ce chapitre, nous passerons en revue certains critères de séparation. Ensuite nous décrivons l’approche, que nous avons privilégiée dans ce mémoire et qui porte sur la maximisation de la non-gaussianité.

3.1 Quelques critères de séparation

3.1.1 Maximiser la vraisemblance

En considérant l’équation du modèle de l’ACI $X = \mathbf{A}S$. La densité du vecteur X s’écrit :

$$p(x) = |\det \mathbf{B}| p(s) \quad (3.1)$$

Sous l’hypothèse de l’indépendances des S_i nous avons :

$$p(x) = |\det \mathbf{B}| \prod_{i=1}^n p(s_i) \quad (3.2)$$

où \mathbf{B} est l’estimée de la matrice inverse du mélange \mathbf{A}^{-1} .

En prenant $S = \mathbf{B}X$, on obtient :

$$p(x) = |\det \mathbf{B}| \prod_{i=1}^n p(B_i^T x) \quad (3.3)$$

où B_i^T est la i ème ligne de la matrice \mathbf{B} .

On suppose que nous avons M réalisations du vecteur X . La fonction du maximum de vraisemblance est donnée par :

$$L(\mathbf{B}) = \prod_{m=1}^M |\det \mathbf{B}| \prod_{i=1}^n p_i(B_i^T x_m) \quad (3.4)$$

En appliquant le logarithme sur cette fonction, le critère à maximiser en terme de maximum de vraisemblance est :

$$J_{MV}(\mathbf{B}) = \log L(\mathbf{B}) = \sum_{m=1}^M \sum_{i=1}^n \log p_i(B_i^T x_m) + M \log |\det \mathbf{B}| \quad (3.5)$$

En terme d'estimation statistique, la fonction $J_{MV}(\mathbf{B})$ est un estimateur optimal [Hyv99], cependant dans cette expression, il est nécessaire que les densités marginales $p(s_i)$ soient connues. Dans le cas contraire, il faut approximer ces dernières ce qui engendre des erreurs. Pour ce faire, souvent la densité est approximée d'une façon à ce que cette dernière possède un nombre réduit de paramètres. Ce choix provoque un faible nombre d'erreurs ce qui n'influence pas l'estimation de la matrice \mathbf{B} .

Dans la pratique, plusieurs algorithmes ont été développés pour optimiser cette approche :

- Pham et Garat ont développé le gradient de la fonction J_{MV} . Cette méthode c'est avéré très coûteuse puisque l'inverse de la matrice \mathbf{B} est calculé à chaque itération [PG97].
- Dans [Car98], Cardoso a défini le maximum de vraisemblance comme la divergence de Kullback entre la vraie densité des sources p et celle que l'on suppose \hat{p} . De plus, il a proposé d'utiliser le gradient relatif de cette expression [CL96].

Par surcroît, la maximisation de la vraisemblance peut se faire par l'algorithme FastICA tel que décrit dans le chapitre 9 du [HKO01].

3.1.2 Minimiser l'information mutuelle

L'information mutuelle est une mesure naturelle de l'indépendance. Ainsi, il suffit d'estimer la matrice de mélange \mathbf{B} telle que l'information mutuelle de $\mathbf{B}X$ soit minimale.

En effet, cette approche est toujours valable même si les observations ne suivent pas le modèle de l'ACI puisque la minimisation de l'information mutuelle donne toujours des composantes indépendantes.

De plus, cette mesure a un lien avec plusieurs principes d'estimation, en particulier le maximum de vraisemblance et la maximisation de la non-gaussianité.

Information mutuelle et non gaussianité : Rappelons l'expression de l'information mutuelle :

$$I(Y) = \sum_{i=1}^n H(Y_i) - H(Y) \quad (3.6)$$

En considérant la transformation linéaire $Y = \mathbf{B}X$, l'information mutuelle peut s'écrire comme :

$$I(Y) = \sum_{i=1}^n H(Y_i) - H(X) - \log |\det \mathbf{B}| \quad (3.7)$$

On suppose que les composantes de Y sont décorréelées et de variance unité. Par conséquent $\det \mathbf{B} = (\det E[XX^T])^{-1/2}$. De plus $H(X)$ ne dépend pas de la matrice \mathbf{B} . Alors $I(Y)$ peut se simplifier en :

$$I(Y) = \sum_{i=1}^n H(Y_i) + Cte \quad (3.8)$$

Par définition la négentropie est de la forme :

$$J(Y) = H_N(Y) - H(Y) \quad (3.9)$$

Ainsi :

$$\sum_{i=1}^n H(Y_i) = \sum_{i=1}^n H_N(Y_i) - \sum_{i=1}^n J(Y_i) \quad (3.10)$$

$H_N(Y_i)$ est l'entropie d'une gaussienne centrée réduite, cette quantité est constante. On remarque donc, que sous la contrainte de décorrélation et de normalisation à l'unité,

maximiser la somme de la non-gaussianité des variables Y_1, Y_2, \dots, Y_n revient à minimiser leur information mutuelle.

Information mutuelle et maximum de vraisemblance : Reprenons l'équation du maximum de vraisemblance, en remplaçant la somme des observations par l'estimation de leur moyenne nous avons :

$$\frac{1}{M} \log L(\mathbf{B}) = E\left\{ \sum_{i=1}^n \log p_i(B_i^T x) \right\} + \log |\det \mathbf{B}| \quad (3.11)$$

En faisant l'hypothèse que les densités de probabilité des composantes $p(s_i)$ sont égales aux densités estimées $p(y_i)$, il vient que le premier terme de l'équation précédente est égale à $-\sum_{i=1}^n H(B_i^T x)$. Ainsi maximiser la vraisemblance (3.11) revient à minimiser (3.7) si l'on connaît la probabilité des sources.

3.1.3 La décorrélation non linéaire

La décorrélation n'implique pas l'indépendance, l'idée est de considérer la décorrélation non linéaire afin de satisfaire l'indépendance. Soient n variables aléatoires Y_1, \dots, Y_n , ces variables sont mutuellement indépendantes si :

$$E\{f_1(Y_1) \cdots f_n(Y_n)\} = E[f_1(Y_1)] \times \cdots \times E[f_n(Y_n)] \quad (3.12)$$

où f_i sont des fonctions intégrables quelconques.

Par définition les variables aléatoires Y_i sont dites décorrélées non linéairement quand :

$$E\{f_1(Y_1) \cdots f_n(Y_n)\} = 0 \quad (3.13)$$

où au moins une fonction f_i est non linéaire.

Pour que les variables soient indépendantes, une condition suffisante pour que la propriété (3.13) soit vérifiée et qu'au moins une de ces fonctions non linéaires soit impaire et vérifie $E[f_i(Y_i)] = 0$.

L'algorithme de Jutten et Héroult [HKO01] est basé sur ce principe. En effet, la séparation est effectuée en annulant la corrélation non-linéaire entre les variables Y_i définie :

$$Y = X - MY \quad (3.14)$$

où X est le vecteur du mélange défini par le modèle de l'ACI et M est une matrice de diagonale nulle.

$$Y = (\mathbf{I} + M)^{-1}X \quad (3.15)$$

où $(\mathbf{I} + M)^{-1}$ est l'estimée de la matrice de séparation B .

Cet algorithme souffre de quelques inconvénients : la convergence n'est pas toujours garantie [Sor91] et le nombre estimé de sources est limité.

À partir de cette étude, plusieurs approches similaires ont été proposées tels que, l'algorithme de Cichoki et Unbehauen [CU96, CUR94, CM92] et la notion de fonction d'estimation par Amari [AC97].

3.1.4 L'annulation des cumulants croisés d'ordre supérieur

Nous avons déjà vu que la maximisation de la valeur absolue du moment d'ordre 4 permet de maximiser la non-gaussianité. Cette méthode n'exploitait pas le fait que tous les cumulants croisés de variables indépendantes sont nuls. Par contre, l'annulation des cumulants croisés n'engendre l'indépendance que si aucune des variables n'est gaussienne. En considérant le modèle de l'ACI : $X = AS$, le cumulants d'ordre 4 des observations

s'écrit :

$$\text{Cum}(X_i, X_j, X_k, X_l) = \sum_{n=1}^N K(S_n) a_{in} a_{jn} a_{kn} a_{ln} \quad (3.16)$$

Rappelons que $K(S)$ est le coefficient d'aplatissement de S .

Annuler tous les cumulants demande beaucoup de temps de calcul. Dans [Com94], Comon montre que l'indépendance peut être obtenue en annulant les cumulants croisés d'ordre quatre tel que :

$$J_{\text{comon}}(Y) = \sum_{ijkl \neq iiii} \text{Cum}_4(Y_i, Y_j, Y_k, Y_l)^2 \quad (3.17)$$

D'autres critères annulant certaines tranches de cumulants ont été introduits, notamment celui de Cardoso appelé le critère JADE (Joint Approximate Diagonalization of Eigenmatrices) est donné par :

$$J_{\text{JADE}}(Y) = \sum_{ijkl \neq iikl} \text{Cum}_4(Y_i, Y_j, Y_k, Y_l)^2 \quad (3.18)$$

La minimisation de ce critère est ramenée à une diagonalisation simultanée d'un ensemble de matrices de cumulants. Pour plus de détails nous laissons le lecteur se référer à [Car99].

En pratique, l'algorithme JADE est très performant. En effet, le calcul n'est pas complexe et la séparation est efficace. Toutefois, il reste sensible dans un espace de grandes dimensions.

Bien entendu, toutes les méthodes décrites plus haut permettent l'estimation des composantes indépendantes. Chacune de ses méthodes tend à optimiser un critère statistique spécifique. Toutefois, plusieurs travaux ont montré des liens entre ces approches. En effet, il a été prouvé que l'approche de la vraisemblance coïncide avec celle de l'information mutuelle [Car99, Com94]. Cette même fonction de contraste a été mise en relation avec la négentropie. Cette ressemblance met en évidence le lien entre les approches basées

sur l'approximation de l'information mutuelle et l'approximation avec la négentropie. Cependant, il existe quelques différences entre ces différentes approches, nous citons donc :

- Le principe consistant à maximiser la non-gaussianité permet d'extraire une seule composante alors que les autres différents critères estiment toutes les composantes indépendantes simultanément.
- Certaines approches approximent les densités des sources tandis que d'autres utilisent l'approximation des cumulants.
- Le maximum de vraisemblance requiert la connaissance a priori de la densité des sources. Dans le cas contraire, une mauvaise approximation de cette dernière donne des estimations erronées.

3.2 Maximiser la non-gaussianité

L'approche que nous avons utilisée dans ce mémoire est la maximisation de l'aspect non-gaussien des sources proposée par Hyvärinen [HKO01]. Le principe de cette approche est basé sur le théorème de la limite centrale (section 2.2.3), l'idée est d'estimer les sources de manière à ce que leur densité de probabilité soit la plus éloignée possible d'une densité de probabilité gaussienne. Dans cette section, nous allons tout d'abord, décrire les mesures de non-gaussianité utilisées par cette approche, ensuite, nous exposons les propriétés statistiques de ses estimateurs. Enfin, nous présentons l'algorithme de cette approche basé sur la méthode du point fixe.

3.2.1 Mesurer la non-gaussianité dans l'ACI

Le coefficient d'aplatissement :

On considère le vecteur aléatoire $Z = [Z_1, \dots, Z_n]^T$ et le i ème vecteur de la matrice de séparation $W_i = [W_{i1}, \dots, W_{in}]^T$. D'après les propriétés du coefficient d'aplatissement , nous avons :

$$K(W_i^T Z) = \sum_{j=1}^n w_{ij}^4 K(Z_j)$$

Nous avons vu que minimiser ou maximiser le $K(W_i^T Z)$ revient à trouver une composante indépendante S_i . Une façon simple pour résoudre ceci est d'appliquer la méthode du gradient descendant ou ascendant.

Sous la contrainte $E\{SS^T\} = \mathbf{I}_n$, le coefficient d'aplatissement de $W_i^T Z$ s'écrit :

$$\begin{aligned} K(W_i^T Z) &= E\{(W_i^T Z)^4\} - 3[E\{(W_i^T Z)^2\}]^2 \\ &= E\{(W_i^T Z)^4\} - 3\|W_i\|^4 \end{aligned}$$

La fonction objective est de la forme :

$$J(W_i) = E\{(W_i^T Z)^4\} - 3\|W_i\|^4 + F(\|W_i\|^2)$$

F est la pénalité associée à la contrainte, l'itération du gradient proposée par Hyvärinen [HKO01] est définie par :

$$W_i(t+1) = W_i(t) \pm \mu(t) [Z(t)(W_i(t)^T Z(t))^3 - 3\|W_i(t)\|^2 W_i(t) + f(\|W_i(t)\|^2) W_i(t)] \quad (3.19)$$

les deux premiers termes de cette itération correspondent au gradient du $K(W_i^T Z)$, le dernier terme est obtenu par le gradient de $F(\|W_i\|^2)$. Aussi, le signe positif avant les crochets s'applique au cas où l'on cherche un maximum local, réciproquement le signe négatif correspond à un minimum local.

La convergence de ce type d'algorithme est souvent très sensible au choix du paramètre $\mu(t)$. Pour ce faire, Hyvärinen [HKO01] a développé l'approche du point fixe. Ce dernier est obtenu lorsque :

$$E\{Z(W_i^T Z)^3\} - 3\|W_i\|^2 W_i + f(\|W_i\|^2)W_i = 0$$

Pour obtenir une itération plus rapide, cette équation est reformulée comme suit :

$$W_i = \text{scalaire} \times (E\{Z(W_i^T Z)^3\} - 3\|W_i\|^2 W_i)$$

Après chaque itération le vecteur W_i est divisé par sa norme $\|W_i\|$, par conséquent, $\|W_i\|^2 = 1$. De plus, puisque nous nous intéressons seulement à la direction du membre de droite de l'équation, le scalaire peut être négligé. L'itération devient :

$$W_i = E\{Z(W_i^T Z)^3\} - 3W_i \quad (3.20)$$

Cette première version de l'algorithme souffre de la non-robustesse du coefficient d'aplatissement. Dans la pratique, une deuxième version plus performante, basée sur la négentropie est souvent utilisée.

La négentropie :

La négentropie est une mesure plutôt théorique basée sur la théorie de l'information, dans la pratique nous utilisons des approximations de cette mesure. Le calcul exact de la négentropie se fait à l'aide de la densité de la variable aléatoire. Ce calcul exact est en général ardu et il est souvent nécessaire de procéder à une approximation qui consiste essentiellement à remplacer la fonction de densité par une fonction polynomiale ou non polynomiale découlant du développement en série de type Taylor.

Approximation de la négentropie par des fonctions polynomiales : Cette approximation trouve ses bases dans le développement de Gram-Charlier. Soit X une variable aléatoire de moyenne 0 et de variance 1. De plus, on suppose que sa distribution est proche d'une distribution gaussienne standardisée de densité $\varphi(x)$. Par définition, le développement de Gram-Charlier s'écrit comme :

$$p(x) \approx \hat{p}(x) = \varphi(x) \left(1 + k_3(x) \frac{H_3(x)}{3!} + k_4(x) \frac{H_4(x)}{4!} \right) \quad (3.21)$$

$k_3(x) = E\{X^3\}$ et $k_4(x) = E\{X^4\} - 3$ représentent les cumulants d'ordre 3 et 4 respectivement. H_3 et H_4 sont les polynômes de Hermite.

Considérons le développement logarithmique suivant : $\log(1 + \varepsilon) \approx \varepsilon - \varepsilon^2/2$, en reportant ces approximations dans la formule de la négentropie, on obtient l'approximation de cette dernière. Toutefois, il est nécessaire de considérer la propriété suivante des polynômes de Hermite :

$$\int \varphi(x) H_i(x) H_j(x) dx = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases} \quad (3.22)$$

où i et j sont des entiers positifs.

Après quelques manipulations algébriques, nous obtenons :

$$J(X) \approx \frac{1}{12} [E\{X^3\}]^2 + \frac{1}{48} [K(X)]^2 \quad (3.23)$$

Dans ce cas, il en résulte que l'approximation de la négentropie par des fonctions polynomiales revient à approcher cette mesure par les cumulants d'ordre 3 et 4.

Approximation de la négentropie par des fonctions non polynomiales : Une autre manière d'estimer la négentropie est d'utiliser l'entropie maximale. L'idée est d'approximer la densité d'entropie maximale $p(x)$. On suppose toujours que cette densité

est proche d'une distribution gaussienne standardisée $\varphi(x)$, $p(x)$ s'écrit sous la forme suivante :

$$p(x) = A \exp\left(\frac{-x^2}{2} + a_{n+1}x + (a_{n+2} + \frac{1}{2})x^2 + \sum_{i=1}^n a_i F^i(x)\right) \quad (3.24)$$

où les coefficients a_i , $i = 1, 2, \dots, n+1$, permettent de rapprocher cette densité $p(x)$ à une densité gaussienne.

Rappelons que l'intégrale de l'entropie maximale est définie comme :

$$\int p(x) F^i(x) dx = c_i \quad \text{pour } i = 1, \dots, n \quad (3.25)$$

F^i est une fonction non polynomiale. Réciproquement, l'approximation correspondra à celle obtenue avec les cumulants d'ordre supérieur.

Nous adoptons le développement limité suivant : $\exp(\epsilon) \approx 1 + \epsilon$.

On obtient donc :

$$p(x) \approx \bar{A} \varphi(x) (1 + a_{n+1}x + (a_{n+2} + 1/2)x^2 + \sum_{i=1}^n a_i F^i(x)) \quad (3.26)$$

où $\bar{A} = \sqrt{2\pi}A$.

Pour $i, j = 1, \dots, n$, l'expression finale de l'approximation de la densité nécessite les propriétés suivantes :

$$\int \varphi(x) F^i(x) F^j(x) dx = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases} \quad (3.27)$$

$$\int \varphi(x) F^i(x) x^k dx = 0 \quad \text{pour } k = 0, 1, 2 \quad (3.28)$$

Les coefficients \bar{A} , a_{n+1} , a_{n+2} et a_i sont déterminés à partir des intégrales suivantes :

$$\begin{aligned} \int p(x) dx &= \bar{A}(1 + (a_{n+1} + 1/2)) &= 1 \\ \int p(x)x dx &= \bar{A}a_{n+1} &= 0 \\ \int p(x)x^2 dx &= \bar{A}(1 + 3(a_{n+2} + 1/2)) &= 1 \\ \int p(x)F^i(x) dx &= \bar{A}a_i &= c_i, \quad i = 1, \dots, n \end{aligned}$$

L'approximation de la densité $p(x)$ de l'entropie maximale est exprimée comme :

$$\hat{p}(x) \approx \varphi(x) \left(1 + \sum_{i=1}^n c_i F^i(x) \right) \quad (3.29)$$

En remplaçant l'approximation de la densité dans la formule de la négentropie nous obtenons :

$$J(X) \approx \frac{1}{2} \sum_{i=1}^n E\{F^i(X)\}^2 \quad (3.30)$$

La fonction F^i est une fonction non quadratique, où i correspond à un indice.

Le choix de la fonction F^i peut correspondre par exemple, à la sélection de deux fonctions F^1 et F^2 qui sont respectivement impaire et paire. Ce système de fonctions permet de mesurer l'asymétrie et l'aplatissement de la distribution des sources.

En pratique, nous utilisons l'approximation suivante :

$$J(X) = [E\{F(X)\} - E\{F(V)\}]^2 \quad (3.31)$$

où V est une variable gaussienne centrée réduite.

La fiabilité de cette approche de la séparation dépend d'une part, de la fonction objective choisie et d'autre part, de la méthode d'optimisation utilisée. En effet, pour aboutir à une bonne estimation du vecteur W_i , le choix de la fonction F doit répondre à certaines exigences statistiques, concernant, entre autres, la variance asymptotique et la robustesse. De plus, les propriétés de l'algorithme, telles que la convergence, la mémoire utilisée et la stabilité numérique sont nécessaires pour une meilleure optimisation de cette fonction.

3.2.2 Analyse de l'estimateur et choix de la fonction F

D'un point de vue statistique, l'estimation de la fonction objective se réduit à un choix d'une fonction de contraste F non quadratique.

Dans cette section nous allons analyser les propriétés statistiques des fonctions F utilisées

dans l'approximation de la négentropie. Pour simplifier, nous nous plaçons dans le cadre où l'on souhaite estimer une seule composante indépendante $Y_i = W_i^T Z$, le problème à optimiser est le suivant :

$$\begin{array}{ll} \text{Maximiser} & E\{F(W_i^T Z)\} \\ \text{sous la contrainte} & : E\{(W_i^T Z)^2\} = 1 \end{array}$$

Deux propriétés statistiques du vecteur estimé \hat{W}_i sont étudiées à savoir la variance asymptotique et la robustesse.

Variance asymptotique

En pratique, on dispose de M observations de Z , et l'on estime l'espérance à maximiser par la moyenne de l'échantillon. Cette façon de procéder induit des erreurs dans l'estimation \hat{W}_i du vecteur W_i . Une manière de mesurer l'erreur commise par cette approximation consiste à déterminer la matrice de variance-covariance de $\sqrt{M}\hat{W}_i$ et d'en déterminer la limite lorsque $M \rightarrow \infty$. La précision de l'estimation n'est autre que la trace de cette matrice de variance-covariance asymptotique.

Résultat : La trace de la matrice de variance-covariance asymptotique de \hat{W}_i est égale à :

$$V_F = C(\mathbf{A}) \frac{E\{f^2(S_i)\} - (E\{S_i f(S_i)\})^2}{(E\{S_i f(S_i)\} - f'(S_i))^2} \quad (3.32)$$

où f est le gradient de la fonction F , f' est le hessien de F et $C(\mathbf{A})$ est une constante qui dépend seulement de la matrice \mathbf{A} .

Preuve :

Posons le modèle de base de l'ICA : $Z_{(n,1)} = \mathbf{A}_{(n,n)} S_{(n,1)}$, nous cherchons à résoudre le problème d'optimisation suivant :

$$\begin{aligned} & \text{Maximiser } E\{F(W_i^T Z)\} \\ & \text{Sous la contrainte } E\{(W_i^T Z)^2\} = 1 \end{aligned}$$

où W_i est le i ème vecteur de la matrice de séparation \mathbf{W} .

On considère le changement de variables suivant : $Q_{(n,1)} = \mathbf{A}_{(n,n)}^T W_{(n,1)}$, l'équation qui définit la solution optimale devient :

$$\sum_m S_m f(\hat{Q}^T S_m) = \lambda \sum_m S_m S_m^T \hat{Q} \quad m = 1, \dots, M \quad (3.33)$$

où M est la taille de l'échantillon et λ est le multiplicateur de Lagrange.

En effet, posons le modèle de base de l'ICA : $Z = \mathbf{A}S$, nous cherchons à résoudre le problème d'optimisation précédent. Or, avec le changement de variable nous avons :

$$\begin{aligned} W^T Z &= W^T \mathbf{A}S \\ &= Q^T S \end{aligned}$$

Sans perte de généralité, nous supposons que \hat{Q} est proche de la solution idéale c'est à dire $Q = (1, 0, 0, \dots)$.

Considérons l'écriture vectorielle suivante $S = \begin{pmatrix} S_1 \\ S_{-1} \end{pmatrix}$ et $\hat{Q} = \begin{pmatrix} \hat{Q}_1 \\ \hat{Q}_{-1} \end{pmatrix}$ où, S_1 et \hat{Q}_1 représentent la première composante des vecteurs S et \hat{Q} respectivement. S_{-1} et \hat{Q}_{-1} sont les vecteurs des autres composantes.

Pour simplifier, nous supprimons l'indice m de l'équation 3.33, on obtient :

$$\sum_m \begin{pmatrix} S_1 \\ S_{-1} \end{pmatrix} f(\hat{Q}^T S) = \lambda \sum_m \begin{pmatrix} S_1^2 + S_1 S_{-1}^T \hat{Q}_{-1} \\ S_{-1} S_1 + S_{-1} S_{-1}^T \hat{Q}_{-1} \end{pmatrix}$$

Considérons le vecteur \hat{Q}^T sans sa première composante \hat{Q}_1^T , nous avons :

$$\sum_m S_{-1} f(\hat{Q}^T S) = \lambda \sum_m [S_{-1} S_1 + S_{-1} S_{-1}^T \hat{Q}_{-1}] \quad (3.34)$$

Le développement en séries d'ordre 1 de la fonction $f(\hat{Q}^T S)$ au tour du point S_1 est :

$$\begin{aligned} f(\hat{Q}^T S) &= f(\hat{Q}_1^T S_1) + f'(\hat{Q}_1^T S_1) \hat{Q}_{-1}^T S_{-1} \\ &= f(S_1) + f'(S_1) \hat{Q}_{-1}^T S_{-1} \end{aligned}$$

L'application de ce développement à l'équation 3.34 donne :

$$\sum_m S_{-1}[f(S_1) + f'(S_1)\hat{Q}_{-1}^T S_{-1}] = \lambda \sum_m [S_{-1}S_1 + S_{-1}S_{-1}^T \hat{Q}_{-1}] \quad (3.35)$$

Après quelques manipulations algébriques, nous obtenons :

$$\frac{1}{\sqrt{M}} \sum_m S_{-1}[f(S_1) - \lambda S_1] = \frac{1}{M} \sum_m S_{-1}[-S_{-1}^T f'(S_1) + \lambda S_{-1}^T] \hat{Q}_{-1} \sqrt{M} \quad (3.36)$$

Si on considère la solution optimale $Q = (1, 0, 0, \dots)$ nous avons $\lambda = E\{S_1 f(S_1)\}$.

L'équation 3.36 peut s'écrire sous la forme $u = v \hat{Q}_{-1} \sqrt{M}$.

$$\begin{aligned} v &= \frac{1}{M} \sum_m S_{-1}[E\{S_1 f(S_1)\} S_{-1}^T - S_{-1}^T f'(S_1)] \\ &= \frac{1}{M} \sum_m S_{-1} S_{-1}^T [E\{S_1 f(S_1)\} - f'(S_1)] \end{aligned}$$

et

$$u = \frac{1}{\sqrt{M}} \sum_m S_{-1}[f(S_1) - E\{S_1 f(S_1)\} S_1]$$

Supposons que la variable $Y = S_{-1}[f(S_1) - E\{S_1 f(S_1)\} S_1]$ soit de moyenne nulle nous avons :

$$\begin{aligned} \text{Cov}(Y) &= E\{Y Y^T\} \\ &= E\{(f(S_1) - E\{S_1 f(S_1)\} S_1)^2 S_{-1} S_{-1}^T\} \\ &= E\{f^2(S_1)\} - (E\{S_1 f(S_1)\})^2 \end{aligned}$$

Nous constatons que le premier terme de l'équation 3.36, désigné par u converge vers une variable aléatoire ayant une distribution normale de moyenne zéro et de matrice de variance-covariance égale à la matrice identité multipliée par $E\{f^2(S_1)\} - (E\{S_1 f(S_1)\})^2$.

Le second terme de cette même équation (v) converge vers la matrice identité multipliée

par $E\{S_1 f(S_1)\} - E\{f'(S_1)\}$.

Puisque $\hat{Q}_{-1} = (A_{-1})^T \hat{W}$ et en égalant les traces des matrices de variance-covariance de u et $v\hat{Q}_{-1}\sqrt{M}$ nous obtenons le résultat ci-dessus.

Ainsi, la comparaison de la matrice de variance-covariance asymptotique de deux estimateurs ayant deux fonctions de contrastes différentes F revient à comparer la trace de leurs matrices de variance-covariance V_F . Un bon estimateur est celui qui minimise V_F d'où le résultat suivant :

Résultat : La trace de la matrice de variance-covariance asymptotique de l'estimateur \hat{W}_i , est minimisée lorsque la fonction F est de la forme suivante :

$$F_{opt}(y) = c_1 \log(p_i(y)) + c_2 y^2 + c_3 \quad (3.37)$$

où p_i est la fonction de densité des S_i et c_1, c_2, c_3 sont des constantes choisies d'une manière arbitraire.

Souvent, la version simplifiée suivante : $F_{opt}(y) = \log p_i(y)$ est utilisée. Notons que cette même fonction a été obtenue par plusieurs approches de séparation de sources telle que la méthode du maximum de vraisemblance [PG97].

La robustesse

La robustesse traduit la résistance de l'estimation aux données aberrantes. La robustesse de l'estimateur \hat{W}_i dépend d'une part, du choix de la fonction F et d'autre part, de la méthode d'estimation utilisée. Supposons que l'estimation a été effectuée d'une manière robuste, la question qui se pose est comment le choix de la fonction F peut influencer la robustesse de \hat{W}_i ?

La robustesse de l'estimateur \hat{W}_i peut être étudiée en utilisant la théorie des M-estimateurs.

Par définition, un M-estimateur est un estimateur défini comme solution $\hat{\theta}$ de θ dans l'équation :

$$E\{\psi(Z, \theta)\} = 0 \quad (3.38)$$

où Z est un vecteur aléatoire, et ψ est une fonction.

Il est facile de voir que l'estimateur \hat{W}_i peut être formulé comme solution de l'équation (3.38). En effet, l'application des conditions de Khun-Tucker sur le problème d'optimisation de base permet d'obtenir :

$$\psi(Z, \theta) = \begin{pmatrix} Zf(W_i^T Z) + c\lambda W_i \\ \|W_i\|^2 - 1 \end{pmatrix} \quad (3.39)$$

où $c = (E\{F(W_i^T Z)\} - E\{F(V)\})^{-1}$, $\theta = (W_i, \lambda)$ (λ est le multiplicateur de Lagrange) et f correspond à la dérivée de la fonction F .

L'analyse de la robustesse d'un M-estimateur consiste à calculer sa fonction d'influence $IF(Z, \hat{\theta})$, cette dernière mesure l'influence des points extrêmes sur l'estimation.

D'une manière intuitive, la fonction d'influence doit être bornée par rapport à Z , de ce fait, l'influence du point extrême est bornée ce qui n'affecte pas l'estimation. Cette condition, nous mène à la notion de B-robustesse.

On dit qu'un estimateur est B-robuste, lorsque sa fonction d'influence est bornée en fonction de Z , c'est-à-dire que $\lim sup_Z \|IF(Z, \hat{\theta})\|$ est finie pour tout Z . Dans le cas contraire, le cas où la fonction d'influence n'est pas bornée, cette dernière doit croître plus lentement que Z .

Pour un M-estimateur, la fonction d'influence est donnée par :

$$IF(Z, \hat{\theta}) = B\psi(Z, \hat{\theta}) \quad (3.40)$$

B est une matrice inversible négligeable qui ne dépend pas de Z . En utilisant la définition précédente de ψ nous avons :

$$\|\psi(Z, (W_i, \lambda))\|^2 = c_1 \frac{1}{\gamma^2} h^2(W_i^T Z) + c_2 h(W_i^T Z) + c_3 \quad (3.41)$$

Avec $\gamma = \frac{W_i^T Z}{\|Z\|}$, c_1, c_2 et c_3 sont des constantes qui ne dépendent pas de Z et $h(y) = yf(y)$.
 Suivant cette expression, la robustesse de \hat{W}_i dépend essentiellement du choix de la fonction $h(y)$. L'estimateur \hat{W}_i n'est pas B-robuste puisque γ fait en sorte que sa fonction d'influence ne soit pas bornée pour tout Z . Toutefois, le choix d'une fonction $h(y)$ qui croît lentement par rapport à Z permet d'augmenter la robustesse de \hat{W}_i .

Résultat : Soit Z un vecteur blanchi d'une manière robuste, sa fonction d'influence n'est pas bornée pour tout Z . Toutefois, si $h(y) = yf(y)$ est bornée alors la fonction d'influence est bornée pour l'ensemble $\{Z | \hat{W}_i^T Z / \|Z\| > \epsilon\}$ pour tout $\epsilon > 0$ où f est la dérivée de la fonction F [HKO01, Hyv97].

En résumé, pour obtenir un estimateur \hat{W}_i robuste, la fonction $F(y)$ doit être bornée. Dans le cas contraire, $F(y)$ doit croître plus lentement que $|y|$. Ceci approuve la sensibilité du coefficient d'aplatissement ($F_k(y) = y^4$), par rapport aux points extrêmes.

Le choix de la fonction F

Après avoir défini les notions de variance asymptotique et de robustesse de l'estimateur \hat{W}_i , nous allons voir comment dans la pratique ces propriétés peuvent être utiles quant au choix de la fonction de contraste F .

Considérons la famille de densité suivante :

$$p_\alpha(s) = c_1 \exp(c_2 |s|^\alpha) \quad (3.42)$$

α est une constante positive et c_1 et c_2 sont des constantes de normalisation qui font en sorte que p_α soit une fonction de densité de variance unité. Selon α , p_α peut prendre plusieurs formes :

- Si $0.5 < \alpha < 2$, alors la densité est sur-gaussienne (kurtosis > 0)

- Si $\alpha = 2$, la distribution est gaussienne (kurtosis = 0)
- Si $\alpha > 2$, il s'agit d'une distribution sous-gaussienne (kurtosis < 0)

En terme de variance asymptotique, la fonction optimale F s'écrit :

$$F_{opt}(y) = |y|^\alpha \quad (3.43)$$

Par ailleurs, en vertu des propriétés de robustesse de l'estimateur \hat{W}_i , la fonction de contraste doit plutôt correspondre à cette forme :

$$F_{opt}(y) = |y|^\alpha \quad \text{avec} \quad \alpha < 2 \quad (3.44)$$

Puisque pour $\alpha \leq 1$ ce type de fonction n'est pas différentiable à 0, des approximations différentiables de cette fonction sont utilisées :

- Si $\alpha = 1$, la fonction de densité correspond à une densité de la loi double exponentielle.

La fonction de contraste est $F_1(y) = \log \cosh(a_1 y)$.

- Si $\alpha < 1$, la fonction de contraste utilisée est $F_2(y) = -\exp(-a_2 y^2/2)$. Cette dernière fournit un estimateur robuste.

où a_1 et a_2 sont deux constantes. L'analyse théorique précédente donne certaines idées quant au choix de la fonction de contraste. Toutefois en pratique, d'autres critères sont pris en considération, notamment, la simplicité du calcul de la fonction de contraste.

En résumé, 3 types de fonctions de contraste sont utilisés :

$$F_1(y) = \frac{1}{a_1} \log \cosh(a_1 y), \quad f_1(y) = \tanh(a_1 y) \quad (3.45)$$

$$F_2(y) = -\frac{1}{a_2} \exp(-a_2 \frac{y^2}{2}), \quad f_2(y) = y \exp(-a_2 \frac{y^2}{2}) \quad (3.46)$$

$$F_3(y) = \frac{1}{4} y^4, \quad f_3(y) = y^3 \quad (3.47)$$

où $1 \leq a_1 \leq 2$ et $a_2 = 1$.

En général, la fonction F_1 est la plus utilisée en pratique. Toutefois, lorsque les composantes indépendantes suivent une distribution sur-gaussienne ou quand le critère de

robustesse doit être pris en compte, l'utilisation de la fonction F_2 est fortement suggérée. La fonction F_3 est souvent utilisée sur un ensemble de composantes indépendantes sous gaussiennes jugées sans points extrêmes [HKO01, Hyv97].

3.2.3 Algorithme de séparation des sources avec la négentropie

Dans la section précédente, nous avons abordé les différentes propriétés statistiques de la fonction de contraste basée sur la maximisation de la négentropie. Dans la pratique, l'optimisation de cette fonction nécessite l'implémentation d'un algorithme adéquat.

Algorithme du point fixe

Sous la contrainte de normalisation $E(W_i^T Z) = \|W_i\|^2 = 1$, le gradient de l'approximation de la négentropie (3.31) est donné par :

$$\frac{\partial J(W_i^T Z)}{\partial W_i} = [E\{F(W_i^T Z)\} - E\{F(V)\}]E\{Zf(W_i^T Z)\} \quad (3.48)$$

où f est la dérivée de la fonction F .

L'itération du gradient est :

$$W_i(t+1) = W_i(t) \pm \mu(t)[E\{F(W_i^T Z)\} - E\{F(V)\}]E\{Zf(W_i^T Z)\}$$

Comme pour le coefficient d'aplatissement, la maximisation de la négentropie est beaucoup plus rapide en utilisant la méthode du point fixe. Cette approche est basée sur une approximation de la méthode de Newton. Rappelons que maximiser la négentropie de $W_i^T Z$ revient à résoudre le problème d'optimisation suivant :

$$\begin{array}{l} \text{Maximiser} \\ \text{sous la contrainte} \end{array} : \quad \begin{array}{l} E\{F(W_i^T Z)\} \quad W_i = [w_{i1}, \dots, w_{in}] \\ E\{(W_i^T Z)^2\} = \|W_i\|^2 = 1 \end{array}$$

où au maximum les vecteurs W_i correspondent aux lignes de la matrice de séparation \mathbf{W} . Le point critique est obtenu lorsque le gradient de Lagrange est égale à 0 :

$$E\{Zf(W_i^T Z)\} - \beta W_i = 0 \quad (3.49)$$

En définissant par H le membre de droite de l'équation, nous avons :

$$\frac{\partial H}{\partial W_i} = E\{ZZ^T f'(W_i^T Z)\} - \beta I$$

Pour simplifier l'inversion de cette matrice, la quantité $E\{ZZ^T f'(W_i^T Z)\}$ peut s'approcher comme suit :

$$\begin{aligned} E\{ZZ^T f'(W_i^T Z)\} &\approx E\{ZZ^T\}E\{f'(W_i^T Z)\} \\ &= E\{f'(W_i^T Z)\}I \end{aligned}$$

Par ce fait, la matrice est diagonale, ce qui simplifie l'inversion de cette dernière. L'approximation de Newton est donnée par :

$$\begin{aligned} W_i^+ &= W_i - [E\{Zf(W_i^T Z)\} - \beta W_i] / [E\{f'(W_i^T Z)\} - \beta] \\ W_i^* &= W_i^+ / \|W_i^+\| \end{aligned}$$

où W_i^* est la nouvelle valeur de W_i et $\beta = E\{W_i^T Z f(W_i^T Z)\}$.

En multipliant, la première équation par $\beta - E\{f'(W_i^T Z)\}$ nous obtenons l'itération de l'algorithme FastICA :

$$\begin{aligned} W^+ &= E\{Zf(W^T Z)\} - E\{f'(W^T Z)\}W \\ W^* &= W^+ / \|W^+\| \end{aligned}$$

L'algorithme FastICA est un algorithme très performant basé sur le principe d'apprentissage itératif du point fixe. Le but est de trouver la direction du vecteur W_i tel que $F(W_i^T Z)$ maximise l'aspect non gaussien. La convergence de cet algorithme a été prouvée dans [HKO01].

L'algorithme du point fixe pour estimer une seule composante est décrit de la façon suivante :

Algorithme FastICA pour estimer une composante indépendante :

- Centrage des données
 - Blanchiment (équation 2.22) pour obtenir Z
 - Choix initial du vecteur unitaire W (choix aléatoire par exemple)
 - **Répéter**
 - $W \leftarrow E\{Zf(W^T Z)\} - E\{f'(W^T Z)\}W$ où f' est la dérivée de f
 - $W \leftarrow \frac{W}{\|W\|}$ (normalisation de W)
 - **Jusqu'à** convergence de W
-

La convergence signifie que les anciennes valeurs du vecteur W se rapprochent des nouvelles valeurs de W .

L'algorithme précédent permet d'estimer une seule composante indépendante. Afin de pouvoir estimer plusieurs composantes, il faut s'assurer que ces dernières ne soient pas identiques. Pour se faire, il suffit de décorrélérer les vecteurs W_i à chaque itération. Cela peut se faire de deux façons : la décorrélation déflationniste ou la décorrélation symétrique :

Algorithme FastICA avec la décorrélation déflationniste

- Centrage des données
- Blanchiment (équation 2.22) pour obtenir Z
- **Pour tout** $i = 1, \dots, r$ **Faire**
 - Choix initial du vecteur aléatoire W_i
 - **Répéter**

- $W_i \leftarrow E\{Zf(W_i^T Z)\} - E\{f'(W_i^T Z)\}W_i$ où f' est la dérivée de f .
 - Faire l'orthogonalisation suivant l'équation 1.21
 - $W_i \leftarrow \frac{W_i}{\|W_i\|}$ (normalisation de W_i)
 - **Jusqu'à** convergence de W_i
 - **Fin Pour**
-

r est le nombre de composantes indépendantes à extraire.

L'algorithme FastICA avec l'orthogonalisation symétrique est décrit comme suit :

Algorithme FastICA avec la décorrélation symétrique :

- Centrage des données
 - Blanchiment (équation 2.22) pour obtenir Z
 - **Pour tout** $i = 1, \dots, r$ **Faire**
 - Choix initial de la matrice W
 - Faire l'orthogonalisation suivant l'équation 1.22
 - **Répéter**
 - **Pour tout** $i = 1, \dots, r$ **Faire**
 - $W_i \leftarrow E\{Zf(W_i^T Z)\} - E\{f'(W_i^T Z)\}W_i$ où f' est la dérivée de f
 - **Fin Pour**
 - Faire l'orthogonalisation suivant l'équation 1.22
 - **Jusqu'à** convergence de W
 - **Fin Pour**
-

Propriétés de l'algorithme FastICA

Dans le cadre de l'ACI, l'algorithme du point fixe a un certain nombre de propriétés que nous énumérons ainsi :

- Sous les contraintes du modèle de l'ACI, la convergence est cubique ou du moins quadratique. Contrairement à la méthode du gradient où la convergence est linéaire. La rapidité de la convergence du point fixe a été prouvée par de nombreuses simulations [HKO01].
- Contrairement à la méthode du gradient, la version finale de l'algorithme du point fixe n'exige aucun choix de paramètres ce qui facilite son utilisation.
- La performance de la méthode du point fixe est optimisée par un choix approprié de la fonction F .
- Le nombre de composantes indépendantes extraites peut être contrôlé par l'utilisateur.

Il est légitime de s'interroger sur la fiabilité de la maximisation de la non-gaussianité. En effet notre choix de cette approche est justifié d'une part, par les propriétés statistiques de son estimateur et d'autre part, par la simplicité de son algorithme. De plus, en pratique la majorité des phénomènes étudiés ne suivent pas une distribution gaussienne, ainsi, cette distribution est souvent jugée la moins intéressante.

Le code de l'algorithme FastICA est disponible gratuitement sur le site :

<http://www.cis.hut.fi/projects/ica/fastica/code/dlcode.html>

CHAPITRE 4

ACI sur des données boursières

Souvent, la performance des différents algorithmes de l'ACI cités dans le chapitre précédent, a été testée sur des signaux générés et mélangés artificiellement. Pour évaluer cette méthode sur des données réelles, l'ACI a été utilisée dans de nombreux domaines, citons :

- La séparation des signaux de la parole : Il s'agit du modèle de base de la séparation aveugle des sources, d'où le problème de la **soirée cocktail**. Le modèle de séparation linéaire instantané n'est pas approprié dans ce cas puisque les signaux ont tendance à être convolutifs. On pourra se reporter à [Tor99] pour une revue des méthodes convolutives.
- Imagerie médicale : les enregistrements magnéto-encéphalographiques (MEG) forment une technique non invasive, qui mesure les activités neuronales du cortex. Les MEG permettent l'extraction des signaux neuromagnétiques malgré la présence des artefacts. Dans [VHO⁺98] les auteurs de cet article ont proposé une approche utilisant l'ACI pour séparer les activités cérébrales des artefacts. Cette approche est basée sur la supposition que les signaux cérébraux et ceux des artefacts, suivent un processus de mélange et qu'ils sont statistiquement indépendants.

L'ACI peut être appliquée sur d'autres techniques de mesure de l'activité cérébrale. Notamment, l'imagerie par résonance magnétique fonctionnelle (IRMf). Cette technique permet de détecter les zones actives du cerveau lors de l'exécution de tâches spécifiques. Les signaux émis par l'IRMf comprennent différentes sources de variabilité, y compris éventuellement les artefacts de machines, les pulsations physiologiques, les mouvements de la tête et les modifications hémodynamiques induites par les différentes conditions expérimentales. L'ACI permet de séparer les zones actives du cerveau spatialement indépendantes qui correspondent à des zones fonctionnelles [BS04].

- Télécommunication : dans le contexte de L'Accès Multiple par Répartition en Code (AMRC) (par exemple : le système de communication wireless) en télécommunication, l'ACI présente un potentiel suscitant un intérêt croissant. L'objectif principal d'un tel système est de permettre à chaque utilisateur de communiquer d'une façon fiable malgré l'occupation des mêmes ressources par d'autres utilisateurs. L'intérêt de l'ACI est de séparer les différents utilisateurs qui partagent les mêmes sources fréquentielles et temporelles. (chapitre 23 de [HKO01])
- Finance : une première application de l'ACI sur des données boursières a été réalisée par [BW97]. Le but de cette étude est d'extraire des facteurs indépendants permettant d'expliquer les structures sous-jacentes des marchés boursiers. Cette étude utilise les prix journaliers des actions des 28 plus importantes entreprises dans la bourse de Tokyo entre 1986 et 1989. Afin de rendre les données stationnaires, ils se sont intéressés à la différence entre deux valeurs successives du prix de l'action, l'extraction des sources a été effectuée avec l'algorithme JADE. Les résultats obtenus ont montré que seulement l'utilisation des 4 premières composantes, jugées les plus dominantes, a permis de résumer la totalité des fluctuations des prix de la banque Tokyo-Mitsubishi. De plus, une étude comparative entre l'ACI et l'ACP sur le même jeu de données a montré que les composantes indépendantes expliquent des phénomènes plus intéressants que les composantes principales.

Toujours en finance, l'ACI a été effectuée sur les flux de liquidité de 40 magasins, appartenant à la même chaîne, sur une période de 140 semaines. 5 composantes indépendantes ont été estimées avec l'algorithme FastICA. Les premières composantes révèlent les pics de vente de Noël ou bien les baisses pendant la saison estivale. La dernière composante est plutôt délicate quant à son interprétation. Celle-ci peut concerner, par exemple la position concurrentielle de la chaîne de magasins [KO98].

De plus en raison de son apport descriptif, l'ACI peut être utilisée pour des fins prévisionnelles. Le modèle de prédiction est utilisé sur les composantes indépendantes au lieu d'être effectué sur les observations brutes. Le pré-traitement par l'ACI donne des meilleurs résultats de prédiction des séries temporelles [MKO99].

Ces premières applications dans le domaine financier ont fourni des résultats assez prometteurs, ce qui a éveillé notre intérêt à ce domaine. Ainsi, dans ce chapitre nous présenterons notre application de l'ACI sur les prix à la fermeture de quelques titres boursiers appartenant à l'indice S&P / TSX.

Notre étude est divisée en deux parties. Une première partie descriptive qui porte sur la reconstitution des observations avec les composantes indépendantes les plus dominantes et une seconde partie qui décrit notre modèle de prédiction basé sur l'ACI.

4.1 Jeu de données

4.1.1 Présentation

Notre jeu de données porte sur le prix journalier à la fermeture de 25 entreprises qui font partie de l'indice S&P / TSX de la bourse de Toronto. La période étudiée couvre 8 ans (du 05/01/2000 au 31/12/2008).

L'indice S&P / TSX mesure la performance de la bourse de Toronto, il est calculé à partir des capitalisations boursières d'un grand nombre d'entreprises cotées sur cette bourse. Il peut se décomposer en sous-indices, chaque secteur regroupe des entreprises ayant des caractéristiques semblables. Les indices sectoriels sont décrits par la figure 4.1 .

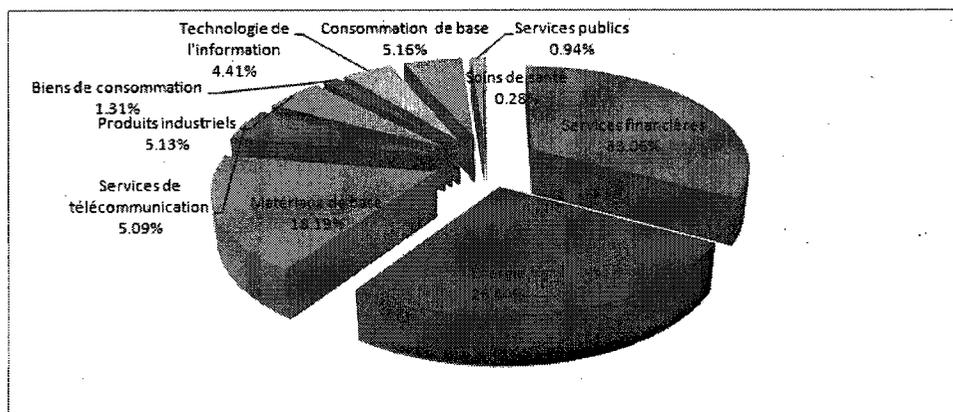


Figure 4.1 – Répartition sectorielle de l'indice S&P / TSX en date du 19 août 2009

Étant donné que la contribution de chaque secteur sur le marché boursier est différente, le nombre de titres considérés dans notre échantillon diffère d'un secteur à l'autre. Tandis que, les secteurs les plus influents sur le marché sont représentés par 3 titres, les secteurs les plus faibles sont décrits par 2 ou voir 1 titre comme c'est le cas pour les services publics (0.28 % du marché boursier canadien) figure 4.2.

Energie	Matières Premières	Industrielles	Consommation Discrétionnaire	Consommation
Imperial Oil IMO	Kinross K	Canadian National Railways CNR	Cogeco Distribution CCA	Saputo SAP
Nexen NXY	Cameco CCO	Cascades CAE	Amica Mature Lyfestyles ACC	Loblaw Cie L
Talisman Energy TLM	Inmet Mining IMN	Bombardier BBB	Shoppers Drug Mart SC	Forzani Group FGL
Santé/Biotechnologie	Financières	Information Technologie	Telecommunications	Services Publics
Medical Devices Inc MDS	Manulife MFC	Open Text OTC	Rogers RGI.B	Emera Cop EMA
WED	Bank of Nova Scotia BNS	Research in Motion RIMM	Manitoba Telecom Services MBT	
	Kinsway Financial Services KFS			

Figure 4.2 – Tableaux des variables

Notre jeu de données comporte 2260 observations et 25 actions, la variable étudiée étant la variation relative du prix donnée par :

$$x_t = \frac{p(t)}{p(t-1)} - 1 \quad (4.1)$$

Où $p(t)$ est le prix de l'action à la fermeture au jour t .

4.1.2 Motivation

Notre étude consiste à appliquer l'ACI sur des titres boursiers. Le but est d'utiliser les composantes indépendantes afin d'extraire des facteurs sous-jacents qui influencent les titres boursiers. Ces structures sous-jacentes pourront ensuite être utilisées dans la prédiction.

En effet, les fluctuations des marchés boursiers peuvent être causées par des changements internes à l'entreprise (rachat d'actions par la société, annonce de bons résultats financiers, ...) et d'autres externes liés par exemple à l'économie mondiale, la guerre, les catastrophes naturelles, ... L'analyse de ces facteurs permettra de fournir les informations nécessaires aux investisseurs quant aux choix de leurs investissements boursiers. De plus, elle permettra d'anticiper certains changements drastiques du marché boursier (crise financière, ...).

Nous nous sommes basés sur le principe descriptif de l'ACI proposé dans [BW97, LCX99]. Ainsi, nous avons développé deux approches permettant la reconstitution de chaque titre avec un nombre quelconque de composantes jugées dominantes. Ces composantes dites dominantes seront utilisées par la suite pour prédire les prix des titres boursiers .

4.2 Classification des composantes indépendantes

4.2.1 Méthodologie

Contrairement à l'ACP, les composantes indépendantes ne sont pas ordonnées selon le degré d'information donnée par chaque dernière. En d'autres termes, la contribution de chaque composante n'est pas directement mesurée avec l'ACI.

Plusieurs approches ont été utilisées pour établir un classement de ces composantes. Dans [BW97] l'ordre déterminé est basé sur le poids de chaque composante, celui-ci est déterminé en multipliant les composantes par les lignes de la matrice de mélange \mathbf{A} , ensuite le poids est trié selon la norme L_∞ . Nous exposons dans cette section deux méthodes de classification des composantes indépendantes.

Classification selon le critère de l'erreur absolue moyenne

Considérons le modèle de l'ACI :

$$X = \mathbf{A}S \tag{4.2}$$

Notons par i la i ème série d'observations $X_{ij} = [x_{i1}, \dots, x_{in}]$, admettons que nous ayons k composantes indépendantes. On définit la contribution de la l ème composante par :

$$\hat{X}_{ijl} = a_{il}\hat{S}_l \quad l = 1, \dots, k \tag{4.3}$$

\hat{X}_{ijl} est la reconstitution de la série originale X_{ij} par la l ème composante indépendante. L'erreur absolue moyenne (EAM) entre X_{ij} et \hat{X}_{ijl} est définie comme :

$$EAM(X_{ij}, \hat{X}_{ijl}) = \frac{1}{n} \sum_{j=1}^n |x_{ij} - \hat{x}_{ijl}| \quad (4.4)$$

L'EAM permet de quantifier la différence entre la série originale et la reconstitution de cette série avec chaque composante indépendante. Une fois que cette mesure est calculée, la question que nous devons nous poser est : comment choisir le nombre de composantes dites influentes dans la reconstitution de la série originale ?

En effet, nous avons adopté un critère de coupure par proximité. Une fois les composantes sont triées par ordre croissant de l'erreur absolue moyenne de la reconstitution (EAMR), nous calculons la différence d'EAMR entre deux composantes successives :

$$m(l) = EAM(X_{ij}, \hat{X}_{ijl}) - EAM(X_{ij}, \hat{X}_{ijl-1}) \quad l = 2, \dots, k \quad (4.5)$$

Ensuite, nous traçons la courbe des $m(l)$ par rapport aux composantes l , le nombre de composantes dominantes correspond au point minimum global m^* .

Les composantes indépendantes dominantes sont les composantes qui reconstituent la série des observations originales avec la plus petite erreur absolue moyenne. En effet, ces composantes correspondent aux facteurs indépendants qui contrôlent la tendance de la série originale et donc nous pouvons considérer le reste des composantes non dominantes comme composantes résiduelles.

Classification selon un test de similarité

La seconde approche de classification est basée sur un test de similarité des séquences. L'estimation directionnelle des prix est le concept clef de cette approche. En effet, sur le marché boursier, il est souvent question de connaître la direction du prix de l'action plutôt que sa valeur. Nous avons donc jugé intéressant d'étudier la direction de la variation du

prix.

Nous avons calculé la différence $x(t) - x(t - 1)$ entre les variations des prix ensuite nous avons transformé cette différence en variable directionnelle $z(t)$ tel que :

$$z(t) = \begin{cases} 1, & \text{si } [x(t) - x(t - 1)] \geq 0 \\ -1, & \text{si } [x(t) - x(t - 1)] < 0 \end{cases} \quad (4.6)$$

La semaine boursière ayant cinq jours, nous avons décidé d'utiliser des séquences composées de cinq directions consécutives. Chaque séquence évalue par conséquent les directions des prix enregistrées sur une période de cinq jours.

Par la suite nous avons considéré la somme des directions de chaque séquence de cinq jours ce qui nous a permis de réduire le problème à la considération de 6 cas :

- -5 ou (5), ce cas correspond à celui où toutes les directions sont négatives (positives).
- -3 ou (3), une seule direction est positive(négative).
- -1 ou (1), 3 directions sont négatives (positives).

Notre idée est de tester la similarité entre les observations mélangées et les composantes indépendantes en se basant sur les propriétés séquentielles calculées de chacune. En effet, chaque variable est caractérisée par le nombre des sommes séquentielles calculées dans la série. Les composantes indépendantes ayant la même répartition séquentielle que le mélange peuvent être considérées comme influentes sur ce dernier.

4.2.2 Analyse des résultats

Pour appliquer l'ACI sur notre jeu de données, nous avons utilisé le programme FastICA sous Matlab. En ce qui concerne la fonction de contraste utilisée, nous avons procédé à une étude comparative entre les différentes fonctions proposées par le programme FastICA :

$$F_1 = \frac{1}{a_1} \log \cosh(a_1 y) , F_2 = -\frac{1}{a_2} \exp(-a_2 y^2), F_3 = \frac{1}{4} y^4 \text{ et } F_4 = \frac{1}{3} y^3 .$$

Par défaut a_1 et a_2 sont égales à 1. Puisque l'algorithme n'a pas convergé avec F_4 , nous sommes contents des 3 premières fonctions.

D'après la figure 4.3, on voit que les composantes indépendantes (IC) obtenues sont quasiment semblables à une différence près qui est l'ordre de l'extraction de ces dernières.

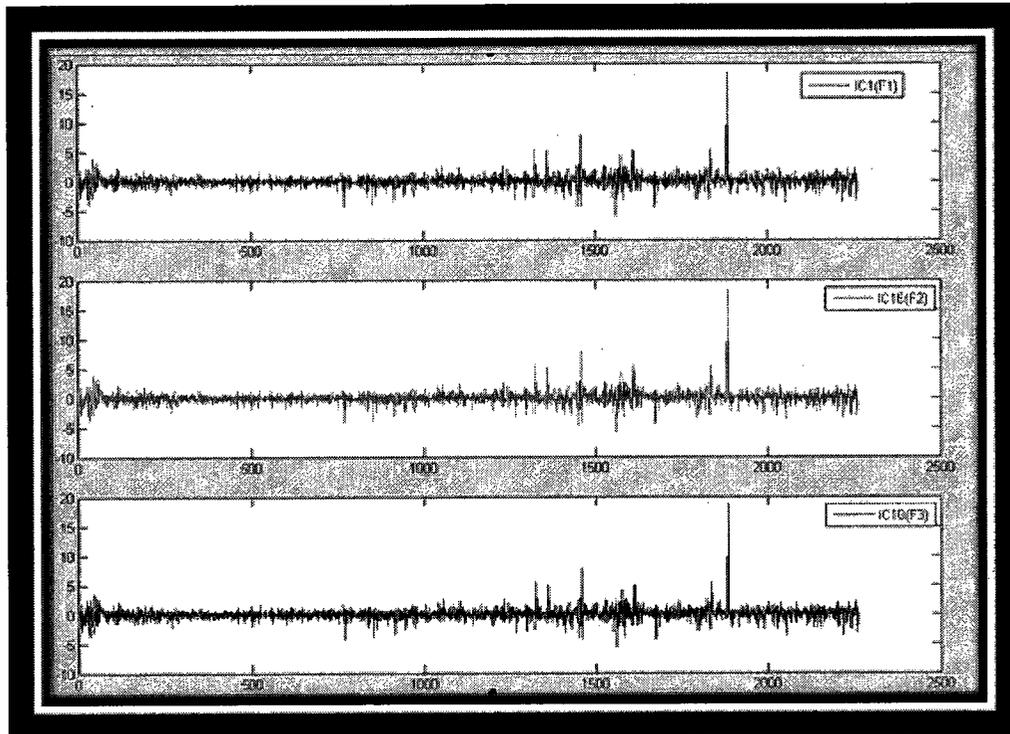


Figure 4.3 – Illustration des composantes indépendantes IC1, IC16 et IC10 obtenues par les fonctions de contrastes F_1 , F_2 et F_3

Les abscisses de la figure 4.3 représentent les observations, l'axe des ordonnées décrit les variations relatives des prix de 2008 à 2000.

On peut voir que la première IC obtenue par la fonction de contraste F_1 est similaire à la 16 ème ou à la 10 ème obtenues respectivement par les fonctions F_2 et F_3 .

Toutefois, nous avons remarqué que selon la fonction de contraste utilisée, un même phénomène peut être décrit par une ou plusieurs composantes indépendantes. Dans la figure 4.4, nous obtenons avec F_3 l'IC 1 qui décrit la fluctuation enregistrée le 16 mai 2006.

L'IC16 quant à elle décrit la fluctuation vue le 14 mai 2004. On peut voir que ces deux fluctuations sont enregistrées au mois de mai et donc on peut soupçonner une similarité entre ces deux phénomènes. Avec la fonction F_2 ces deux fluctuations sont regroupées dans la même composante indépendante IC1.

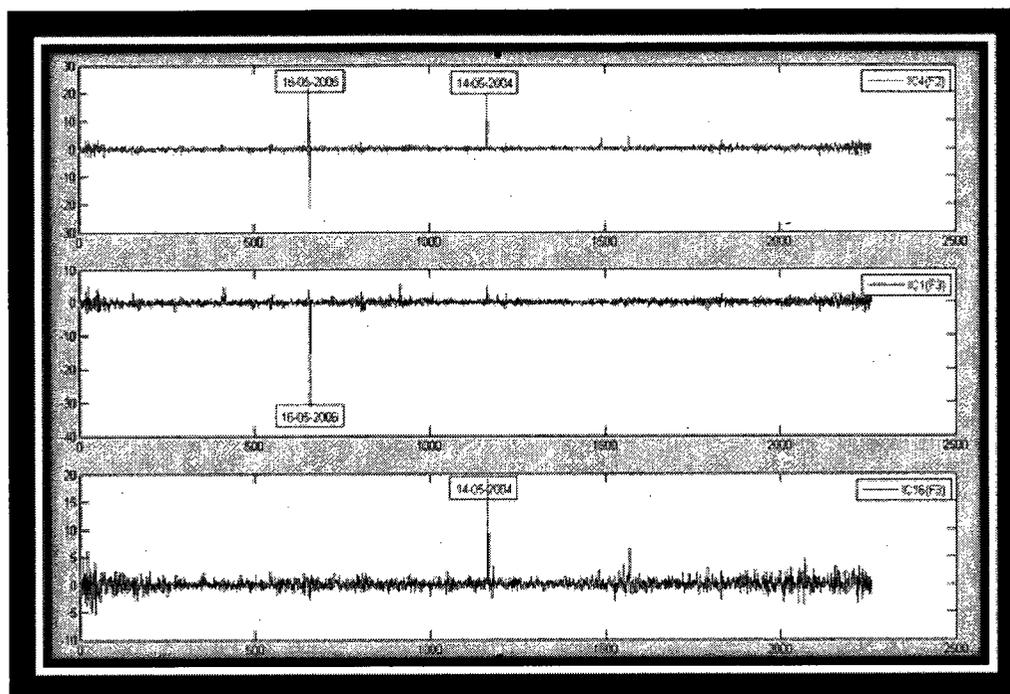


Figure 4.4 – Illustration des composantes indépendantes IC4, IC1 et IC16 obtenues par les fonctions de contrastes F_2 et F_3

En terme d'erreur absolue moyenne, la séparation avec la fonction F_2 nous donne la plus petite erreur de reconstitution comparée aux différentes fonctions de contraste proposées par FastICA. Compte tenu de ces résultats pratiques, nous avons décidé d'effectuer l'ACI en utilisant la fonction de contraste $-\frac{1}{a_2} \exp(-a_2 y^2)$.

Dans cette partie nous allons présenter seulement l'analyse effectuée sur le titre IMO, d'autres résultats concernant certains titres seront présentés en annexes.

IMO est le symbole du titre boursier de la société Imperial Oil. Cette compagnie pétrolière canadienne produit du pétrole brut et du gaz naturel. La compagnie produit actuellement environ 6% de toute l'énergie primaire consommée au Canada.

La figure 4.5 décrit les fluctuations du titre IMO depuis le 5 janvier 2000 (observation 2260) jusqu'au 31 décembre 2008 (observation 0).

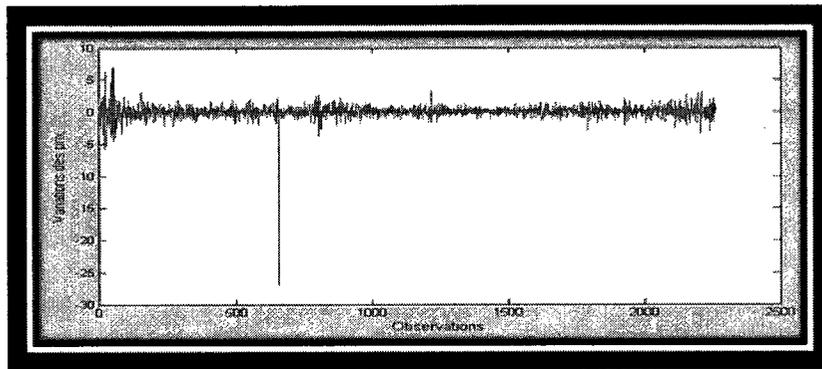


Figure 4.5 – IMO

La figure 4.6 regroupe les variations par année. On voit que les variations sont plutôt semblables d'une année à l'autre excepté pour les années 2000, 2006 et 2008.

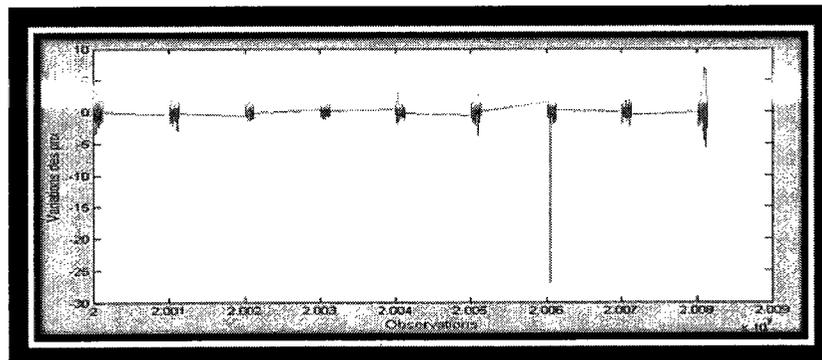


Figure 4.6 – IMO par rapport aux années

Par le biais de l'ACI nous allons tenter d'expliquer ces changements de variation des prix.

Résultats de la reconstitution par le critère d'EAM :

Dans le tableau 4.1 nous indiquons les erreurs absolues moyennes obtenues en estimant le titre IMO par chaque composante indépendante. Les statistiques $EAM(X_i, \hat{X}_{ij})$ et $m(j)$ sont calculées telles que décrites plus haut.

ICs	$EAM(X_i, \hat{X}_{ij})$	$m(j)$
IC23	0.387716238	
IC8	0.515888964	0.12817273
IC4	0.538455116	0.02256615
IC20	0.577352059	0.03889694
IC5	0.577954742	0.00060268
IC24	0.578068941	0.00011420
IC11	0.578089985	0.00002104
IC22	0.578127592	0.00003761
IC16	0.57833102	0.00020343
IC7	0.578538997	0.00020798
IC1	0.578690408	0.00015141
IC2	0.578869354	0.00017895
IC10	0.578873568	0.00000421
IC15	0.578962644	0.00008908
IC18	0.57896552	0.00000288
IC3	0.578968757	0.00000324
IC14	0.578979458	0.00001070
IC17	0.578990677	0.00001122
IC19	0.578990729	0.00000005
IC13	0.579150715	0.00015999
IC12	0.57917004	0.00001933
IC9	0.579182754	0.00001271
IC25	0.579373995	0.00019124
IC6	0.579442098	0.00006810
IC21	0.579532136	0.00009004

Tableau 4.1 – Tableau de l'erreur absolue moyenne de chaque composante dans la reconstitution d'IMO

En ce qui concerne le nombre choisi de composantes, nous avons tracé la figure 4.7. Cette

figure illustre le critère de coupure par proximité, ce dernier nous permettra de mesurer la perte de contribution dans la reconstitution d'IMO.

On remarque qu'à partir de la 4^{ème} composante, l'erreur absolue moyenne devient stable et presque similaire. Ainsi, le nombre de composantes retenues est 4.

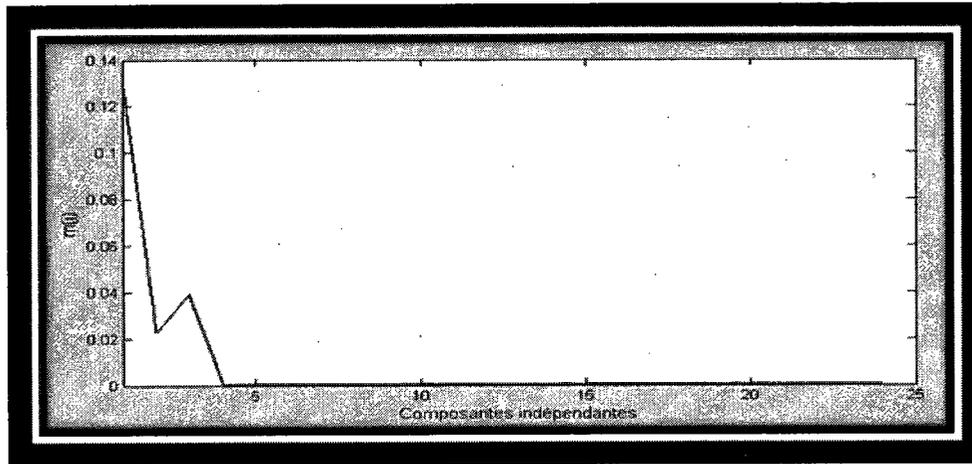


Figure 4.7 – Indice de proximité pour le titre IMO

En d'autres termes, le titre IMO peut être reconstitué par les 4 composantes indépendantes IC23, IC8, IC4 et IC20. D'après le modèle de l'ACI, la fonction de reconstitution du titre IMO peut s'écrire comme :

$$IMO = -0.717898968S_{23} + 0.489349158S_8 + 0.461045793S_4 + 0.06257865S_{20} \quad (4.7)$$

Les coefficients de cette équation correspondent aux constantes a_{1l} pour $l = 23, 8, 4$ et 20 (voir les coefficients de la matrice **A** dans l'annexe B).

Dans la figure 4.8 nous présentons la reconstitution du titre IMO par ces 4 composantes indépendantes.

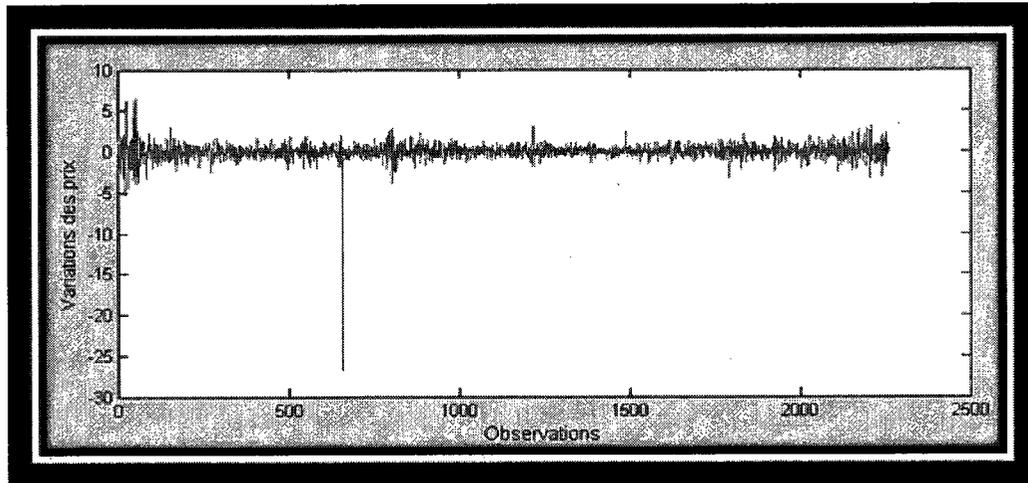


Figure 4.8 – Reconstitution du titre IMO avec l’approche EAM

Nous pouvons voir dans la figure 4.9 que la reconstitution du titre IMO avec ces 4 composantes est presque parfaite.

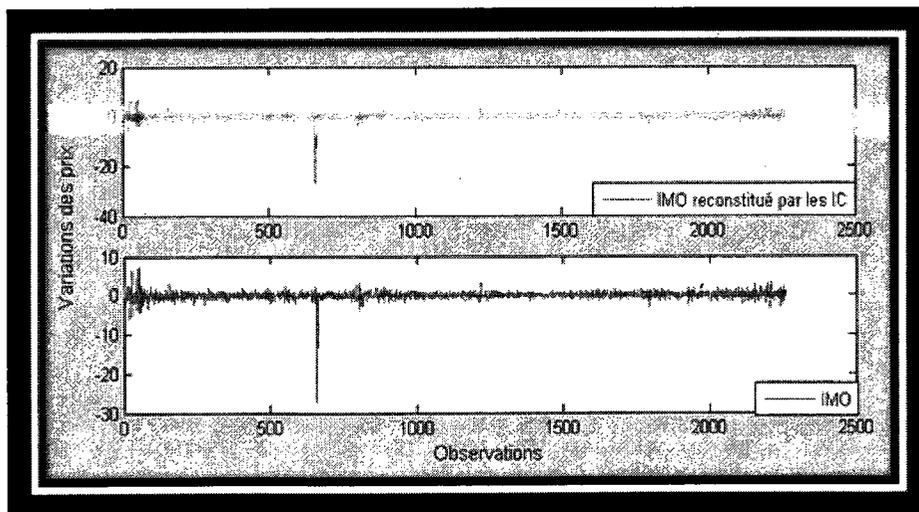


Figure 4.9 – Comparaison entre IMO et sa reconstitution avec l’approche EAM

Interprétation : l'interprétation des composantes indépendantes est une étape importante dans notre analyse. En effet, cette analyse descriptive nous permettra de comprendre le comportement du titre IMO et de déceler les facteurs sous-jacents qui influencent ce titre. Notre analyse des composantes a été guidée par une spécialiste en finance de marché, nous nous sommes intéressés à analyser les tendances apparentes dans les séries des composantes indépendantes d'une part, et d'autre part nous avons examiné les observations correspondant à un changement considérable du prix.

D'après les figures 4.10, 4.11 et 4.12 on peut voir que les fluctuations importantes sont souvent enregistrées durant le mois de mai. En effet, c'est ce que l'on appelle la phase de transition entre l'hiver et l'été, durant cette période la consommation de l'énergie baisse et par conséquent la production diminue.

En outre la variable production, l'IC23 reflète le phénomène du coût de stockage qui augmente en période de crise puisque la demande diminue.

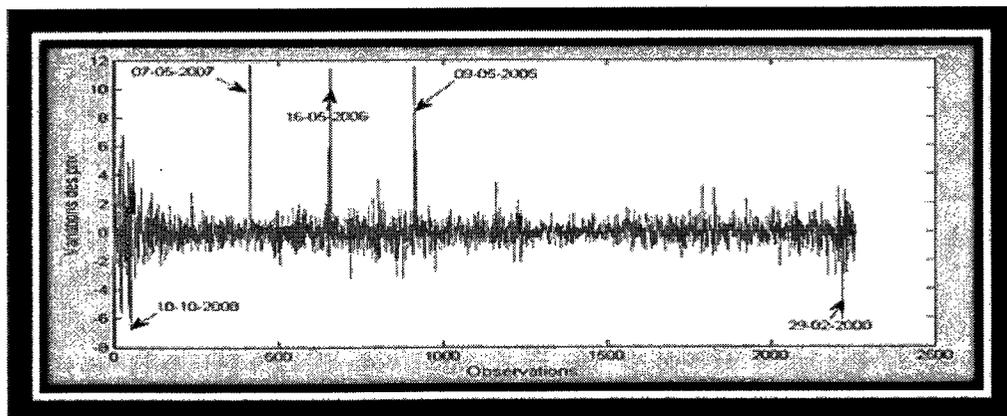


Figure 4.10 – Composante indépendante numéro 23

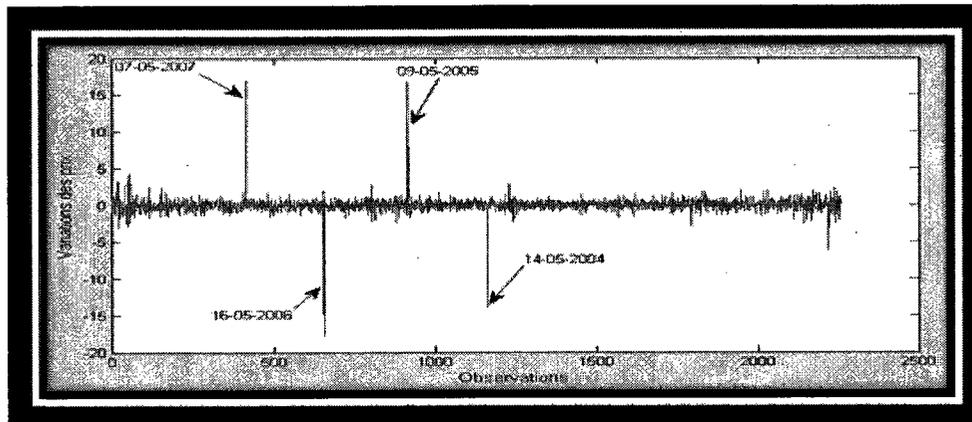


Figure 4.11 – Composante indépendante numéro 8

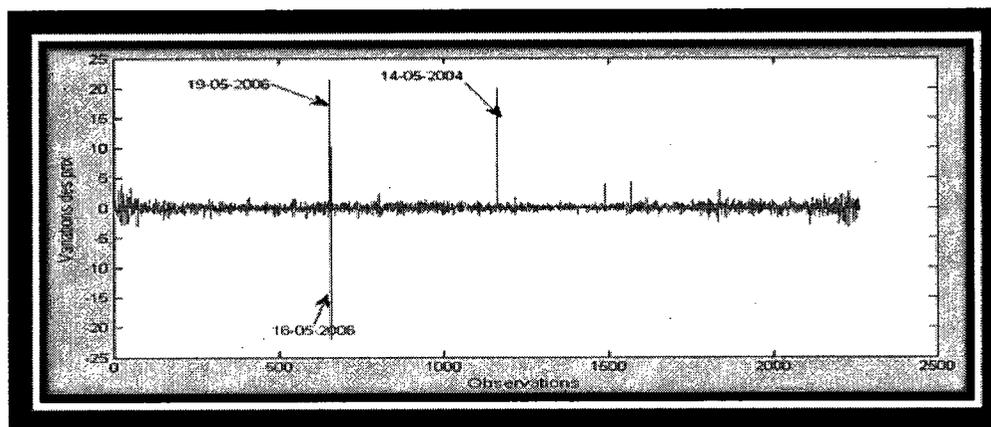


Figure 4.12 – Composante indépendante numéro 4

La figure 4.13 nous illustre deux dates correspondant à l'appréciation du dollar canadien comparé au dollar américain. Toutefois, il ne faut pas négliger le lien existant entre le taux de change et le taux d'intérêt. Les années 2004 et 2006 ont été marquées par une forte augmentation du dollar canadien par ce fait, les taux d'intérêt américains ont augmenté d'une façon considérable.

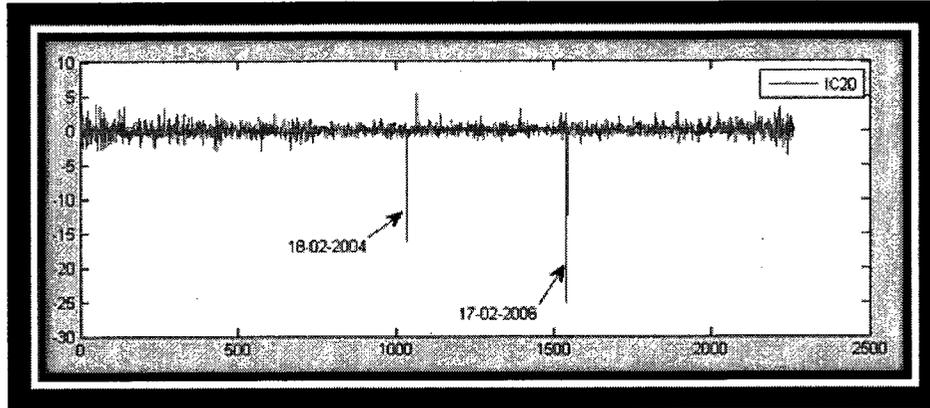


Figure 4.13 – Composante indépendante numéro 20

D'après ces composantes indépendantes, nous pouvons conclure que le titre IMO est sensible aux changements de température qui influencent la demande. De plus, l'appréciation du dollars canadien a un impact négatif sur ce titre. En effet, comme le prix du baril de pétrole se négocie en dollar américain, toute hausse du dollar canadien entraîne une baisse dans les bénéfices de la société.

En réalité, les activités de la société Imperial Oil touchent les 4 coins du monde ce qui explique sa sensibilité à l'international tel que, les changements du taux de change. Aussi, la société est très spécialisée par ce fait, il est tout à fait normal d'avoir détecté seulement les facteurs de la production et du taux de change.

Analyse sectorielle

Après une analyse individuelle de chaque titre, nous avons pensé à effectuer une analyse sectorielle. Cette analyse s'appuie sur un modèle progressif basé sur les rangs des composantes : considérons le tableau 4.2, le i ème titre est caractérisé par m composantes indépendantes influentes que nous classons par leur ordre d'importance. À chaque composante IC_j nous attribuons un rang inverse ($R_{ij} = m - j + 1$). Ensuite nous calculons le poids de chaque composante $\frac{R_{i1}}{\sum_{j=1}^m R_{ij}}$, ce dernier est le ratio entre le rang de la compo-

sante R_{ij} et le total des rangs de toutes les composantes influentes sur le titre $\sum_{j=1}^m R_{ij}$.

<i>i</i> ème titre		
IC	Rangs	Poids
IC1	R_{i1}	$\frac{R_{i1}}{\sum_{j=1}^m R_{ij}}$
IC2	R_{i2}	$\frac{R_{i2}}{\sum_{j=1}^m R_{ij}}$
...
IC <i>l</i>	R_{il}	$\frac{R_{il}}{\sum_{j=1}^m R_{ij}}$
...
IC <i>m</i>	R_{im}	$\frac{R_{im}}{\sum_{j=1}^m R_{ij}}$
Total	$\sum_{j=1}^m R_{ij}$	1

Tableau 4.2 – Tableau des poids des m composantes indépendantes sélectionnées pour la reconstitution du titre i

Enfin, nous calculons la prévalence de chaque composante, cette mesure permet de quantifier la contribution de chaque composante dans le secteur. En effet, considérons un secteur A constitué de k titres et admettons que ces titres ont la composante IC_j en commun. Pour chaque titre, le poids de cette dernière est exprimé comme :

$$P_{ij} = \frac{R_{ij}}{\sum_{j=1}^{m_i} R_{ij}} \quad (4.8)$$

où i est le titre du secteur, j désigne la composante indépendante et m_i est le nombre de composantes du titre i . Notons que ce nombre varie d'un titre à l'autre.

La prévalence non pondérée (PNP) de IC_j est :

$$PNP = \frac{\sum_{i=1}^k P_{ij}}{k} \quad (4.9)$$

La prévalence pondérée (PP) de IC_j est :

$$PP = \frac{\sum_{i=1}^k P_{ij} m_i (m_i + 1) / 2}{\sum_{i=1}^k m_i (m_i + 1) / 2} \quad (4.10)$$

Nous présentons dans ce paragraphe les résultats de l'analyse du secteur de l'énergie. Pour ce secteur notre échantillon est constitué de 3 titres IMO, NXY et TLM :

- Le titre pétrolier IMO expliqué plus haut.
- Le titre NXY appartient à la société Nexen. Cette entreprise offre une diversité de produits et de services. En plus de sa production de pétrole brut et de gaz naturel, elle produit le pétrole synthétique, des produits chimiques et énergétiques.
- Le titre TLM appartient à la société d'énergie Talisman Inc., cette entreprise est spécialisée dans l'exportation, le développement et la production de pétrole et de gaz.

IMO			NXY			TLM		
IC	Rangs	Poids	IC	Rangs	Poids	IC	Rangs	Poids
IC23	4	0.4	IC23	2	0.67	IC23	5	0.33
IC8	3	0.3	IC8	1	0.33	IC4	4	0.27
IC4	2	0.2				IC8	3	0.2
IC20	1	0.1				IC7	2	0.13
						IC24	1	0.07
Total	10	1	Total	3	1	Total	15	1

Tableau 4.3 – Tableau des poids de chaque composante indépendante dans la reconstitution des titres IMO, NXY et TLM

D'après le tableau 4.3, nous constatons que les composantes dominantes sont souvent les mêmes pour les 3 titres.

Outre le facteur de la production, nous remarquons que les composantes 7 et 24 contribuent dans la reconstitution du titre TLM. La composante 7 (figure 4.28) décrit la production et l'appréciation du dollar canadien.

En ce qui concerne la composante 24 (figure 4.15), celle-ci illustre deux dates importantes. Il s'agit d'une part du 24 décembre 2004 où le prix du pétrole a vécu une chute considérable et d'autre part, du 14 février 2006 où le dollar canadien a augmenté.

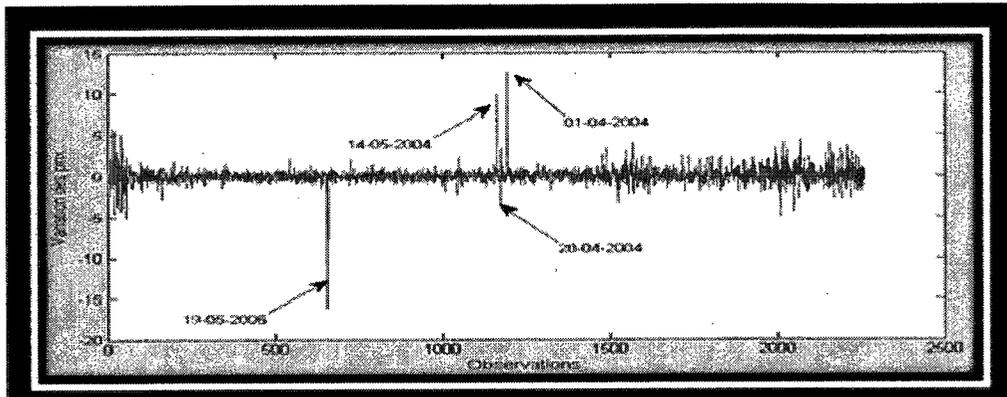


Figure 4.14 – Composante indépendante numéro 7

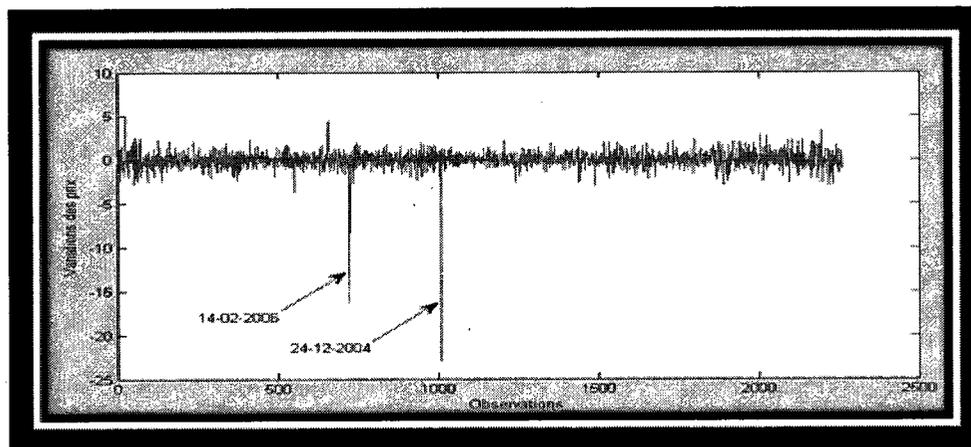


Figure 4.15 – Composante indépendante numéro 24

Dans le tableau 4.4, la valeur 0.47 est la prévalence non pondérée de la composante 23 dans le secteur de l'énergie. La composante 23 contient 47% de l'information sur ce secteur.

En terme de prévalence non pondérée, nous constatons que IC23 et IC8 représentent 75% de l'information donnée par la reconstitution du secteur de l'énergie.

IC	Prévalences non pondérées PNP	Prévalences pondérées PP
IC23	0.47	0.39
IC8	0.28	0.25
IC4	0.16	0.21
IC7	0.04	0.07
IC20	0.03	0.04
IC24	0.02	0.04

Tableau 4.4 – Tableau des prévalences non pondérées et des prévalences pondérées de chaque composante dans le secteur de l'énergie

D'après l'analyse avec les composantes indépendantes, nous pouvons conclure que le secteur d'énergie est sensible à la variation de la production (enregistrée durant les changements de la température, par exemple). De plus, les variations des prix du pétrole et des taux de change influencent ce secteur.

Les fluctuations reflétées par la composante IC23 en 2000 et 2008, laisse croire que ce secteur a subi quelques instabilités avec la crise économique et la crise financière respectivement.

En réalité, IMO et TLM sont deux titres très spécialisés, ils sont exposés au marché mondial ce qui explique leurs sensibilités à ce niveau (fluctuation de taux de change, production, exportation, ...)(IC23, IC8, IC24). Le titre NXY est très diversifié ainsi il est plus sensible que ce soit face à l'environnement économique ou mondial (IC23).

Reconstitution séquentielle

Cette approche de reconstitution est basée sur le test du khi2. Il s'agit de comparer la répartition de séquences de l'échantillon à celle qui existe dans la population. Dans notre cas, la population correspond aux titres boursiers tandis que l'échantillon représente les composantes indépendantes.

Le tableau 4.5 représente les caractéristiques séquentielles du titre IMO et des compo-

santes IC3 et IC16.

Profil de la population IMO		Profil de l'échantillon IC3		Profil de l'échantillon IC16	
Séquences	Nombre	Séquences	Nombre	Séquences	Nombre
-5	1	-5	2	-5	5
-3	163	-3	163	-3	183
-1	905	-1	951	-1	995
1	969	1	927	1	874
3	212	3	207	3	195
5	5	5	5	5	3
Total	2255	Total	2255	Total	2255

Tableau 4.5 – Tableau des répartitions séquentielles

Effectuons le test entre IMO et IC3, les hypothèses statistiques à vérifier sont :

H_0 : IC3 est répartie selon IMO

H_1 : La répartition séquentielle de IC3 n'est pas représentative du titre IMO

Rappelons que la statistique χ^2 est calculée comme :

$$\chi^2 = \sum_{i=1}^k \frac{(f_{oi} - f_{ti})^2}{f_{ti}} \quad (4.11)$$

f_{oi} sont les fréquences observées dans la population et f_{ti} sont les fréquences théoriques en supposant que H_0 (IC3 est répartie selon IMO) est vraie.

La statistique χ^2 est distribuée selon une loi du χ^2 à $k - 1$ degrés de liberté où k est le nombre de modalités.

D'après le tableau 4.6, nous ne pouvons pas rejeter l'hypothèse H_0 au seuil $\alpha = 0.05$ ($\chi^2 = 4.28257 < \chi_{0.05,4}^2 = 9.48772$).

Modalités	f_{oi}	f_{ti}	$(f_{oi} - f_{ti})^2 / f_{ti}$
Somme de séquence ≤ -3	165	164	0.00610
Somme de séquence = -1	951	905	2.33812
Somme de séquence = 1	927	969	1.82043
Somme de séquence = 3	207	212	0.11792
Somme de séquence = 5	5	5	0
			$\chi^2 = 4.28257$

Tableau 4.6 – Tableau de calcul de la statistique χ^2

La répartition séquentielle de la composante 3 est représentative de celle du titre IMO au seuil de signification $\alpha = 0.05$.

Nous avons effectué ce test pour les 25 composantes indépendantes. Le tableau 4.7 représente les composantes retenues et la valeur de la statistique χ^2 de chacune.

IC	χ^2
IC3	4.28258
IC8	4.66058
IC24	4.65789
IC19	5.80469
IC13	7.72880

Tableau 4.7 – Tableau des composantes retenues selon leur χ^2

Grâce aux tests de similarité, nous pouvons déduire que la répartition des séquences des composantes indépendantes 3, 8, 24, 19 et 13 est similaire à celle du titre IMO.

Interprétation : en plus de l'influence des facteurs, production (IC8), prix du pétrole (IC24) et taux de change (IC24) sur IMO, nous remarquons l'existence d'un nouveau facteur décrit par la composante 3 dans la figure 4.16. Cette composante illustre le prix de l'or qui a augmenté jusqu'à 62% de sa valeur en 2003. Ce phénomène est dû à la crise technologique (2001 – 2002) puisque durant cette période le marché est assez instable ce qui pousse les investisseurs vers des titres plus sûrs tels que les aurifères.

La conformité entre cette composante et IMO met en évidence le lien entre le prix du pétrole et celui de l'or. En effet, le jeu de données que nous analysons couvre plusieurs crises (la conjoncture économique 2000, la crise technologique 2001 et la crise financière 2008). Une telle succession de chocs rend les investisseurs plus réticents à l'idée de faire une rotation sectorielle et donc ils se focalisent sur des secteurs fiables tels que l'énergie et les matières premières.

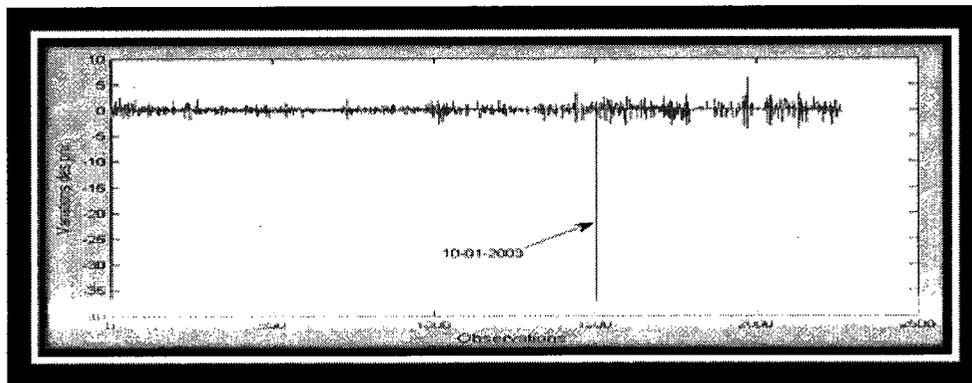


Figure 4.16 – Composante indépendante numéro 3

D'après les figures 4.17 et 4.18 nous observons plusieurs changements dans les composantes 13 et 19, cette instabilité ne nous permet pas de donner une interprétation judicieuse et précise aux composantes. Toutefois, nous pouvons dire que cette instabilité est causée par les crises : économique, technologique et financière.

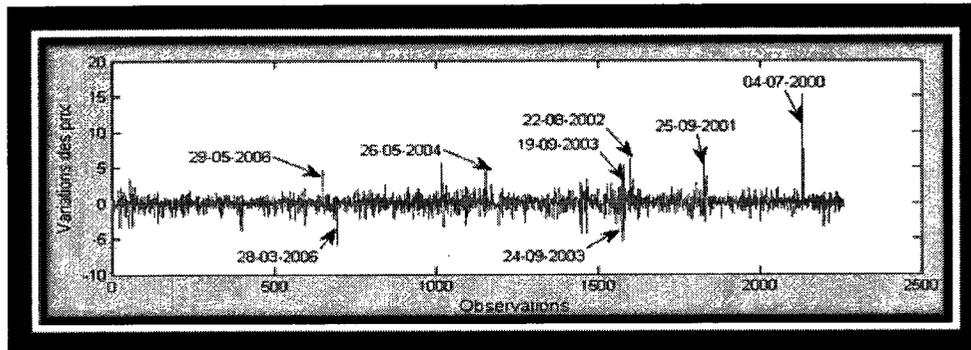


Figure 4.17 – Composante indépendante numéro 13

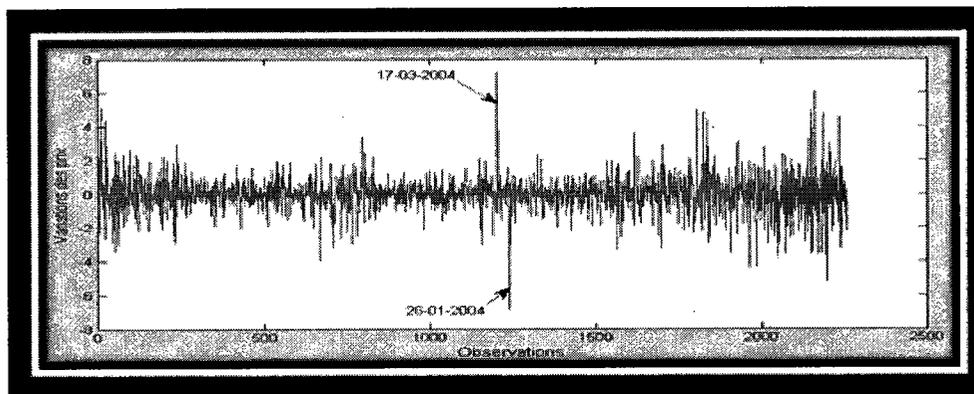


Figure 4.18 – Composante indépendante numéro 19

La fonction de la reconstitution est représentée comme :

$$\begin{aligned}
 IMO = & -0.002169617S_3 + 0.489349158S_8 + 0.014782312S_{24} \\
 & + (-0.022600289)S_{19} + (-0.042663247)S_{13}
 \end{aligned}
 \tag{4.12}$$

Comme le montre la figure 4.20, la reconstitution n'est pas tout à fait parfaite. Toutefois, avec cette analyse nous avons abouti à des résultats similaires à ceux obtenus par l'approche de l'erreur absolue moyenne. De plus, nous avons décelé le lien existant entre le marché de l'or et le pétrole.

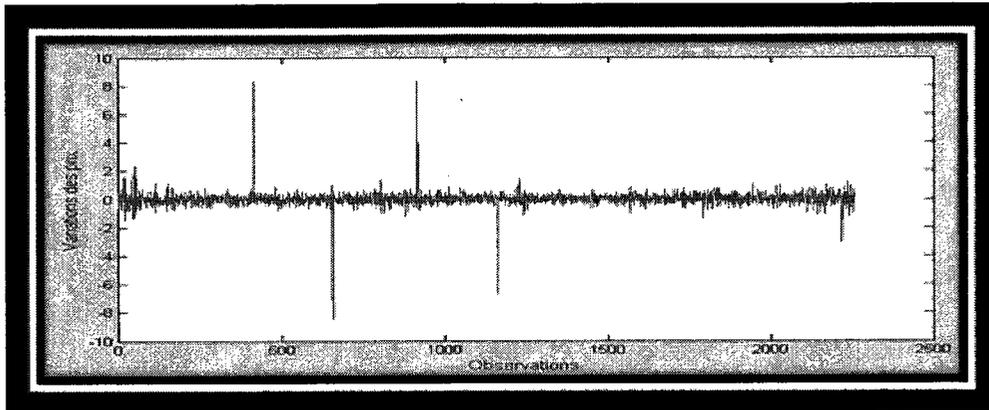


Figure 4.19 – Reconstitution de IMO avec l’approche séquentielle

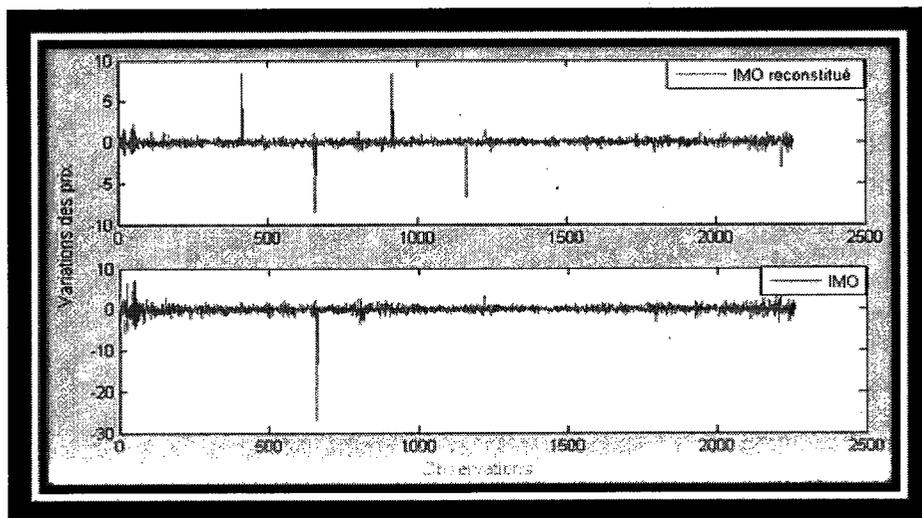


Figure 4.20 – Comparaison entre IMO et sa reconstitution avec l’approche séquentielle

D’après ces résultats, nous affirmons que les composantes indépendantes permettent de révéler certaines structures intéressantes dans le marché boursier. Néanmoins, l’interprétation de ces composantes n’est pas souvent évidente. Cette complexité d’interprétation peut être liée à la nature des observations et à la méthode de séparation utilisée.

4.3 Prévision avec l'ACI

4.3.1 Méthodologie

D'un point de vue descriptif, nous avons vu que l'interprétation des composantes indépendantes nous a permis d'avoir une idée globale sur certains facteurs qui influent sur le prix des titres boursiers. Nous nous sommes donc, interrogés sur l'apport de cette méthode au niveau prévisionnel. Dans [MKO99] les auteurs ont montré des résultats intéressants en utilisant les composantes indépendantes dans la prédiction. Les modèles de prédictions proposés sont soit des modèles autorégressifs (AR), des modèles de régression ou des modèles basés sur les réseaux de neurones. En nous basant sur ces idées, nous avons appliqué un modèle AR sur les composantes indépendantes afin de prédire les variations des prix pour l'année 2009. Comme le montre la figure 4.21, les étapes de notre démarche sont :

1. Application de l'ACI sur les variations des prix des titres.
2. Selon le critère de classification utilisé, nous sélectionnons le nombre de composantes indépendantes dites dominantes pour chaque titre.
3. Application d'un modèle AR sur les composantes dominantes.
4. Chaque composante prédite est multipliée par le coefficient de mélange correspondant a_{ij} . Ainsi, nous obtenons la prédiction \hat{x}_i de l'observation d'origine x_i .

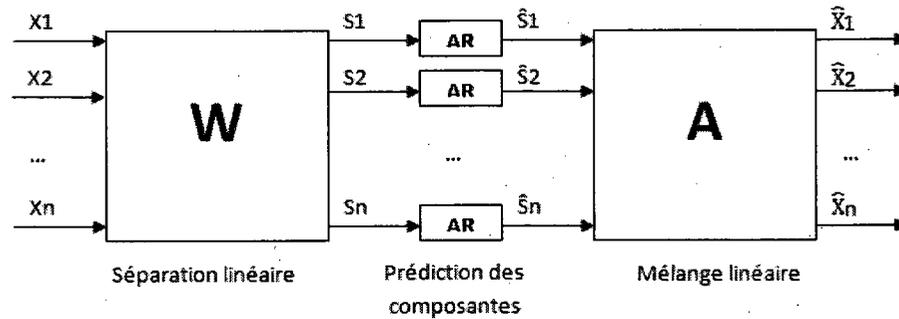


Figure 4.21 – Illustration du modèle de prédiction : les sources sont obtenues par la transformation de séparation $S = \mathbf{W}X$. Les \hat{S}_i correspondent aux composantes prédites par le modèle AR. \hat{X}_i est la prédiction de la série X_i , elle a été calculée par le modèle de mélange $X = \mathbf{A}S$.

Rappelons que le modèle autorégressif d'ordre p de la façon suivante :

$$x_t = c + \sum_{i=1}^p \varphi_i x_{t-i} + \xi_t \quad (4.13)$$

où φ_i sont les paramètres du modèle, c est une constante et ξ_t est un bruit blanc.

Pour qualifier notre prédiction, nous avons calculé deux mesures : la concordance et le taux de prévision.

Concordance : la concordance entre la série prédite et la série d'origine est une variable qui mesure la similarité directionnelle des prix entre le titre d'origine et sa prédiction. Pour chaque série nous calculons la différence entre les observations $x(t) - x(t-1)$, cette différence est transformée en variable binaire qui mesure la direction de la fluctuation du prix. Cette variable directionnelle est égale à :

- -1 si la différence est négative.
- 1 si la différence est positive.

Ensuite, nous comparons la variable directionnelle de la série prédite avec celle de la série d'origine. Une concordance égale à 1 signifie que la direction des deux séries est identique,

dans le cas contraire la concordance est égale à 0.

Taux de prévision : le taux de prévision nous permet de mesurer la précision de notre modèle. Pour chaque valeur prédite, nous calculons sa différence avec la valeur observée dans la série d'origine. Ceci se fait en utilisant la valeur de la concordance :

- Si la concordance est égale à 1, alors nous soustrayons de la plus grande valeur absolue la seconde valeur. Désignons par a_i la valeur absolue de la différence entre les variations relatives des prix prédites et par b_i la valeur absolue de la différence entre les variations relatives observée. La prévision T_i pour l'observation i est définie comme suit :
 - Si $a_i > b_i$ alors $T_i = a_i - b_i$
 - Sinon $T_i = b_i - a_i$
- Si la concordance est égale à 0, alors nous additionnons les valeurs absolues des deux séries. En d'autres termes on aura : $T_i = a_i + b_i$.

Le taux de prédiction est le rapport entre la somme des différences lorsque la concordance est égale à 0 et le total des différences. Ainsi, lorsque la prédiction est identique à la réalité, le taux de prédiction sera égale à 100%.

Pour valider notre approche, nous avons utilisé un modèle autorégressif directement sur les données d'origine (sans introduire les composantes indépendantes). Les résultats de ce modèle ont été utilisés à titre comparatif avec notre modèle.

4.3.2 Analyse des résultats

Dans nos simulations, nous avons considéré un échantillon test de 90 observations du titre IMO enregistrées en 2009. Le modèle AR utilisé est d'ordre 90.

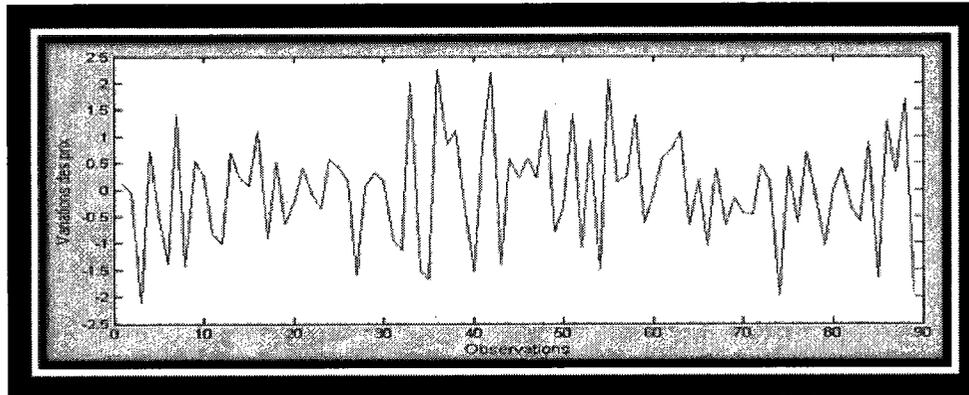


Figure 4.22 – Variations des prix du titre IMO en 2009

La prédiction avec notre modèle, nous permet d'avoir $\approx 56\%$ de concordance avec la série d'origine. Le taux de prévision est de $\approx 54\%$. D'après le tableau 4.8 nous remarquons que la concordance du modèle de comparaison est légèrement meilleure que celle obtenue avec notre modèle. En revanche, le taux de prévision est un peu meilleur avec notre modèle.

	Concordance	Taux de prévision
Notre modèle	0.55682	0.54415
Modèle de comparaison	0.57955	0.52263

Tableau 4.8 – Comparaison de la concordance et du taux de prévision entre notre modèle et le modèle classique AR.

La figure 4.23 représente la prédiction du titre IMO avec l'ACI que l'on calcule comme suit :

$$\hat{IMO}_{2009} = -0.717898968\hat{S}_{23} + 0.489349158\hat{S}_8 + 0.461045793\hat{S}_4 + 0.06257865\hat{S}_{20} \quad (4.14)$$

où \hat{S}_i est la prédiction de S_i calculée avec le modèle AR.

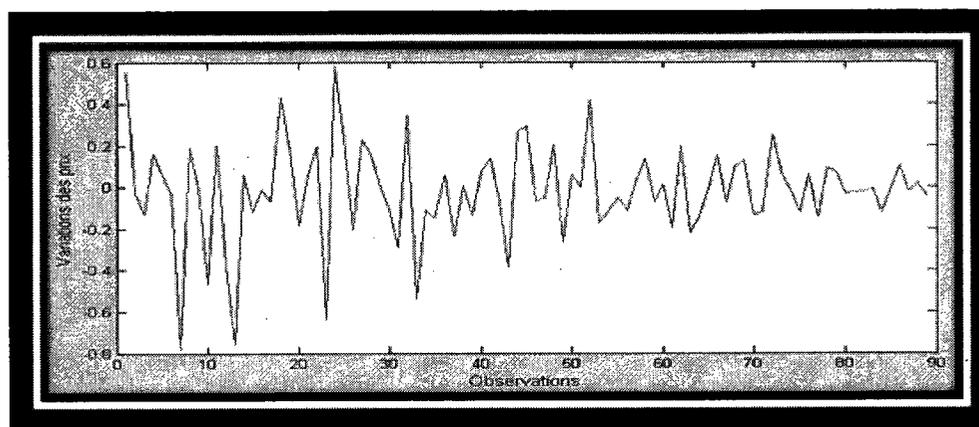


Figure 4.23 – Prédiction d'IMO avec l'ACI : la reconstitution du titre à été effectuée avec l'approche EAM

L'application du modèle AR directement sur la série originale nous permet d'avoir la prédiction illustrée par la figure 4.24.

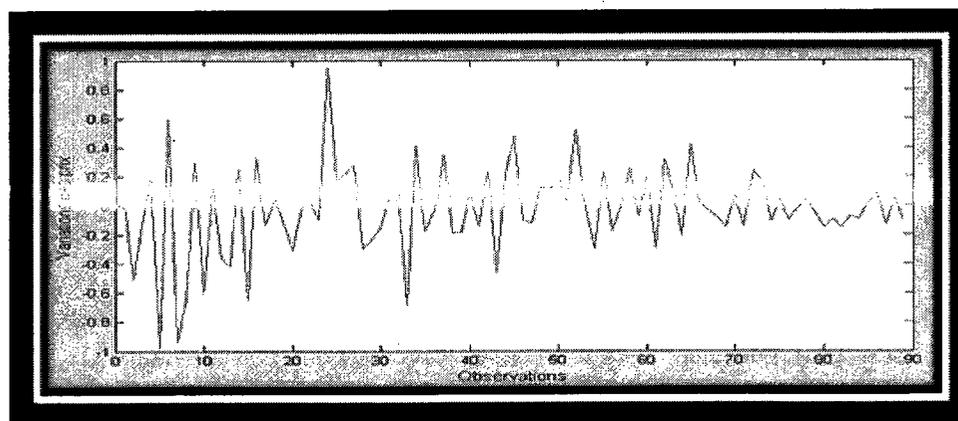


Figure 4.24 – Prédiction d'IMO directement avec AR

La figure 4.25 illustre la série originale IMO et sa prédiction avec les deux modèles.

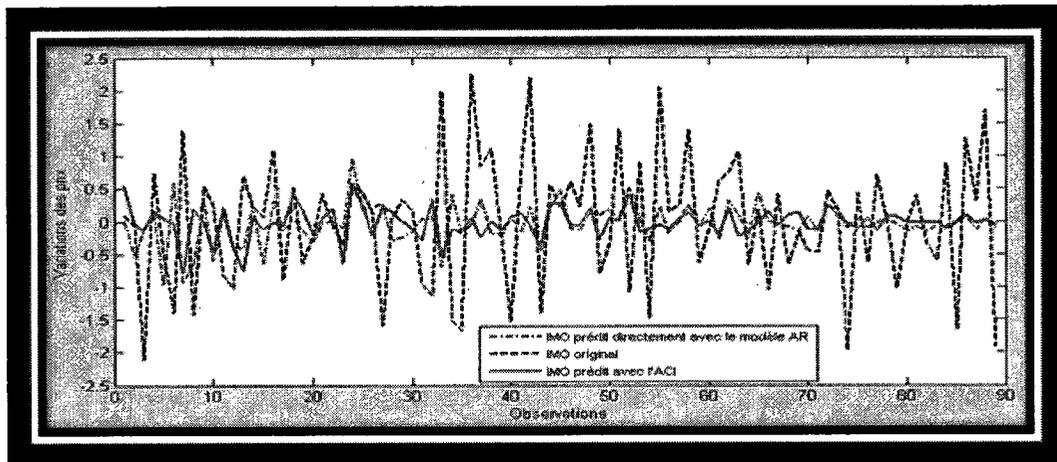


Figure 4.25 – Comparaison entre les modèles de prévision

En ce qui concerne la prévision avec le modèle de reconstitution séquentielle, nous avons obtenu les résultats suivants :

	Concordance	Taux de prévision
Notre modèle	0.40909	0.65269
Modèle de comparaison	0.57955	0.52263

Tableau 4.9 – Comparaison de la concordance et du taux de prévision entre notre modèle et le modèle classique AR.

Nous constatons qu'avec la reconstitution séquentielle le taux de concordance diminue comparé à notre modèle précédent. Toutefois, le taux de prévision calculé est légèrement supérieur.

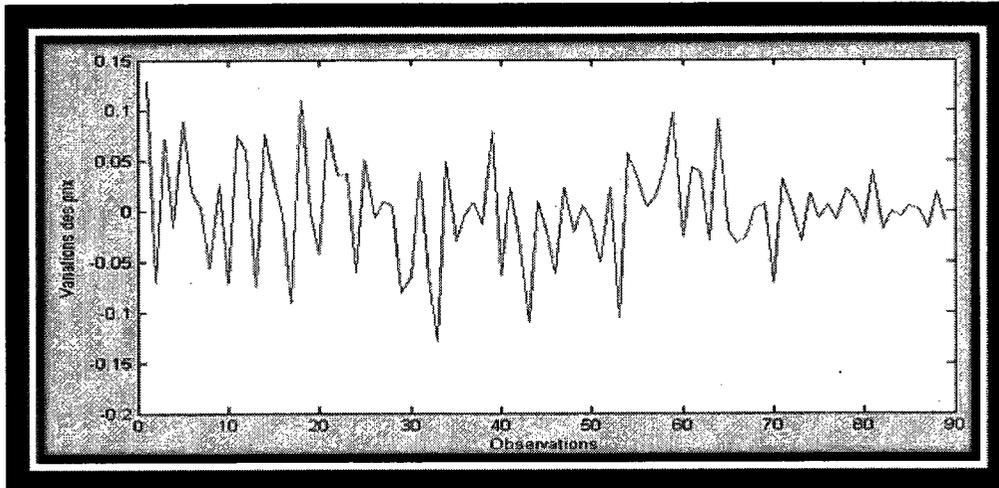


Figure 4.26 – Prédiction d'IMO avec l'ACI : la reconstitution du titre à été effectuée avec l'approche séquentielle

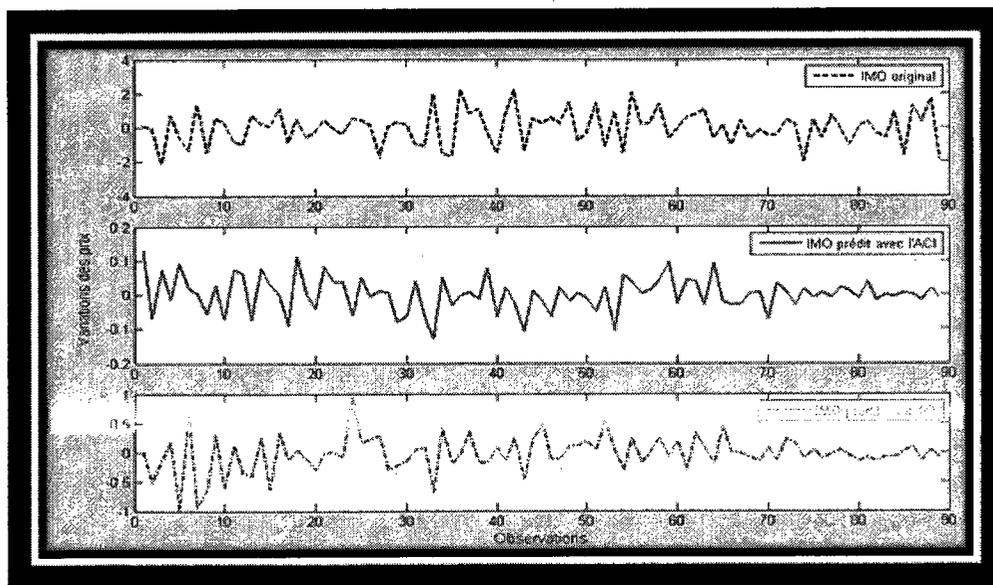


Figure 4.27 – Comparaison entre les modèles de prévision

Nous remarquons qu'avec les deux critères de reconstitution (EAM et séquentielle), notre modèle de prédiction donne des meilleurs résultats en terme de taux de prévision. En revanche, au niveau de la concordance, notre modèle est moins performant que le modèle de comparaison.

En effet, ces deux statistiques ne mesurent pas le même aspect. La concordance est plutôt une mesure de similarité directionnelle tandis que le taux de prévision mesure la précision entre la valeur prédite et la valeur observée. Toutefois, nous proposons un quotient qui prend en compte ces deux mesures. Désignons par C la concordance et par P le taux de prévision d'un modèle quelconque les quotients sont exprimés tels que :

- Le quotient de la concordance est $\frac{C}{C+P}$
- Le quotient de la prévision est $\frac{P}{C+P} = 1 - \frac{C}{C+P}$

L'utilisation de l'un ou l'autre de ces quotients dépend de l'accent que l'on désire mettre soit sur la similarité directionnelle soit sur la précision de la valeur prédite.

Le tableau 4.10 montre les quotients des modèles de prévision avec les deux critères de reconstitution.

	Quotient de la concordance	Quotient de prévision
Notre modèle avec le critère EAM	0.50575	0.49424
Notre modèle avec le critère séquentiel	0.38528	0.61471
Modèle de comparaison	0.52582	0.47417

Tableau 4.10 – Tableau des quotients de concordance et de prévision

CONCLUSION

Dans les chapitres 1, 2 et 3 de ce mémoire nous avons présenté, dans le cadre du mélange linéaire instantané, une approche généralisée de la séparation aveugle des sources. Le chapitre 1, introduit la méthode de séparation aveugle des sources (SAS) et présente les hypothèses de cette méthode et les modèles découlant de ces dernières. Le chapitre 2 décrit l'analyse en composante indépendante (ACI) et expose les hypothèses et les restrictions de cette méthode. Selon le critère considéré pour garantir l'indépendance, plusieurs approches de l'ACI peuvent exister. Le chapitre 3 décrit brièvement quelques-unes de ces approches et souligne certaines ressemblances entre ces dernières. Nous avons adopté l'approche de la maximisation de l'aspect non-gaussien des sources. Ce chapitre présente ses aspects théoriques et ses caractéristiques statistiques. Cette méthode est implantée par l'algorithme du point fixe FastICA.

Enfin, la contribution de ce mémoire est décrite dans le chapitre 4. Les actions boursières sont relativement complexes et très variables. Il est donc difficile d'identifier ces facteurs influents. Cette difficulté est liée entre autres au nombre de facteurs externes et internes à l'entreprise.

Notre méthode est basée sur l'ACI, nous avons proposé deux critères permettant de classer les composantes indépendantes selon leur contribution dans la reconstitution du titre d'origine. Cette façon de procéder nous a permis d'identifier les composantes influentes sur lesquelles nous nous sommes basés, pour faire l'analyse des titres boursiers.

Nous avons montré que l'interprétation de ces composantes permet de souligner certaines structures sous-jacentes influentes sur les variations des titres boursiers.

Une seconde contribution de ce mémoire est le développement d'un modèle de prédiction basé sur les composantes indépendantes. Nous avons donc appliqué un modèle autorégressif sur les composantes indépendantes. La validation quantitative de ce modèle a été effectuée par deux mesures : la concordance et le taux de prévision. Celles-ci sont simples à calculer et permettent d'évaluer la performance du modèle de prévision.

Ce modèle de prévision basé sur l'ACI, a été comparé avec un modèle autorégressif appliqué directement sur les observations. Les résultats obtenus en terme de taux de prévision sont meilleurs que ceux obtenus avec le modèle de comparaison. En revanche, au niveau de la concordance notre modèle est moins performant que le modèle de comparaison.

Une limite de notre approche réside dans la linéarité du modèle de l'ACI et du modèle de prévision. La linéarité est une simplification parfois suffisante, mais elle peut être parfois très limitatrice. Des travaux futurs devraient étudier l'extension de l'ACI au cas non linéaire. En effet, d'après les figures des variations des prix, présentées dans le chapitre 4, nous constatons que le comportement des marchés boursiers est discontinu, les fluctuations des prix sont souvent caractérisées par des hausses et des baisses soudaines avec quelques périodes de stabilité. Ce mouvement non linéaire des prix nous laisse confiant quant aux résultats de cette approche.

De plus, ce comportement des prix boursiers est une caractéristique d'un processus de causalité, par conséquent l'alternance entre les périodes de grandes fluctuations et les périodes stables ne se produit pas par hasard. Des travaux futurs devraient développer un modèle non linéaire basé, par exemple, sur les arbres de décision. Ce modèle pourra prendre en considération d'une part, les facteurs sous-jacents extraits par l'ACI et d'autre part, certains indicateurs utilisés par les spécialistes pour anticiper les variations des prix des titres boursiers.

ANNEXE A

Dans cette section, nous allons présenter les résultats de l'application de l'ACI sur le secteur financier. Notre échantillon comporte 3 titres financiers :

- Manulife Financial Corp.(société financière Manuvie) : cette société de services financiers établie au Canada offre un éventail de produits de protection financière. Elle détient le titre boursier MFC.
- Bank of Nova Scotia (banque Scotia) : cette société financière est l'une des plus importantes institutions en Amérique du Nord et la plus internationale des banques canadiennes. Elle offre une gamme diversifiée de produits et de services bancaires. Elle détient le titre boursier BNS.
- Kinsway Financial Services (Kinsway services financiers) : cette compagnie d'assurance offre des assurances automobiles et autres produits spécialisés. Elle détient le titre boursier KFS.

Dans cette section, nous allons voir comment l'application de l'ACI sur ces titres nous a permis d'une part, de déterminer les facteurs influents sur le secteur financier, et d'autre part, d'estimer les variations des prix de ces titres en 2009.

4.4 Reconstitution des titres

4.4.1 Reconstitution avec le critère EAM

Le critère EAM nous a permis de considérer les composantes citées dans le tableau 4.11 :

MFC	BNS	KFS
IC7	IC9	IC25
IC9	IC23	IC14
IC17		

Tableau 4.11 – Tableau des IC sélectionnées selon le critère EAM pour les titres MFC, BNS et KFS

D'après les composantes indépendantes, nous constatons que les titres MFC et BNS sont affectés par l'appréciation du dollar canadien et par les changements de production (IC9 et IC7). De plus, les composantes indépendantes IC17 et IC23 sont classées les dernières dans la reconstitution de MFC et BNS respectivement. Nous pouvons donc conclure que ces titres sont moins sensibles à la crise économique et à la crise financière. En ce qui concerne le titre KFS, nous remarquons que ce dernier est plutôt sensible aux crises enregistrées durant la période étudiée (IC25 et IC14).

En effet, MFC et BNS sont exposés sur le plan international ce qui explique leur sensibilité par rapport aux variations de taux de change et de la production. Aussi, ces deux titres sont réglementés d'où leur faible sensibilité à la crise financière comparée au titre KFS.

En période de crise, les taux d'intérêt varient d'où la sensibilité de KFS en particulier, à la conjoncture économique et à la crise financière. Toujours en raison de la conjoncture économique le secteur de transport a enregistré une baisse considérable en 2007 (période après crise). Étant donné que KFS opère dans ce domaine, nous remarquons qu'il a été affecté par cet incident(IC14).

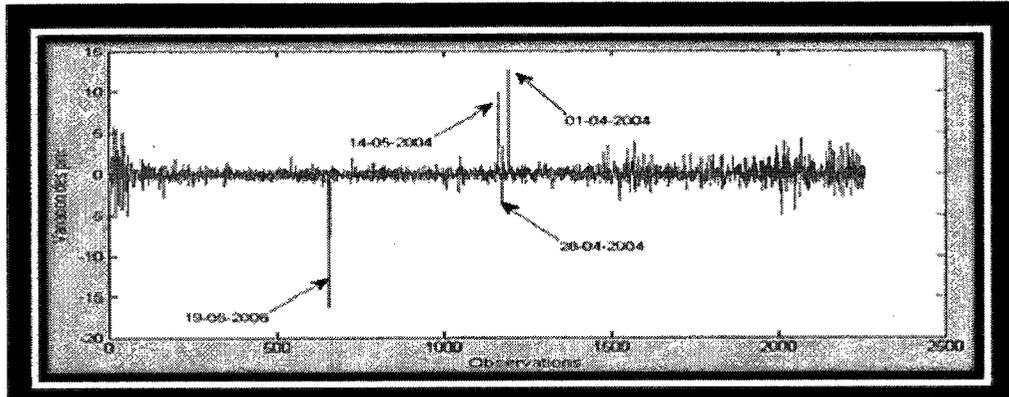


Figure 4.28 – Composante indépendante numéro 7

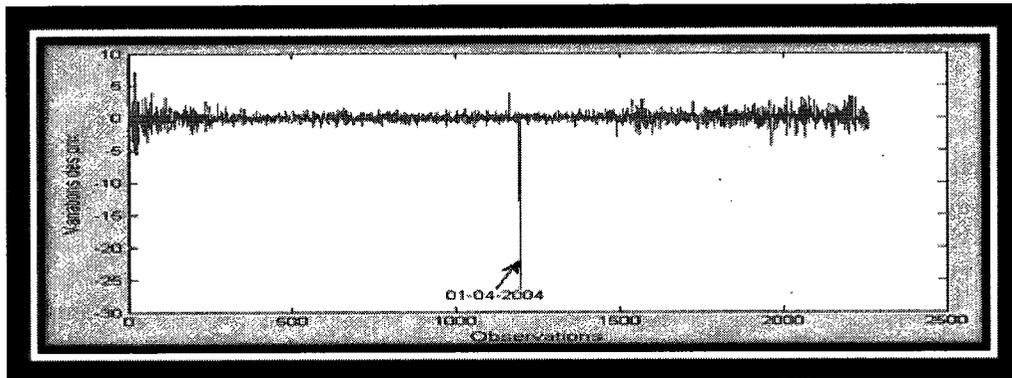


Figure 4.29 – Composante indépendante numéro 9

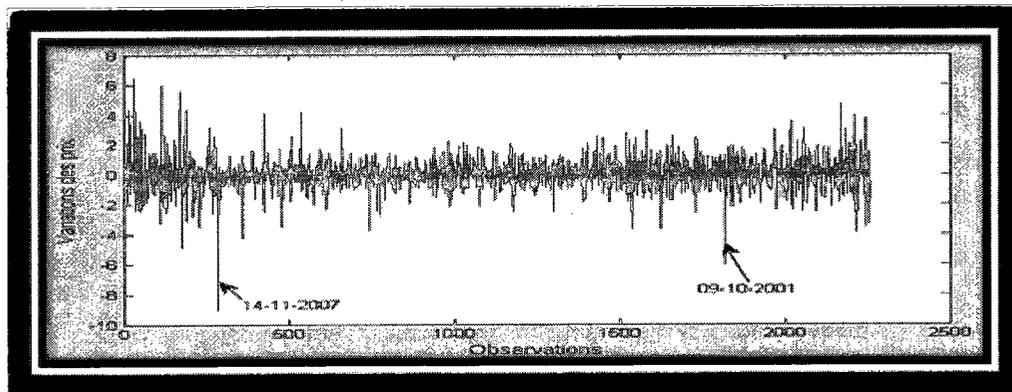


Figure 4.30 – Composante indépendante numéro 17

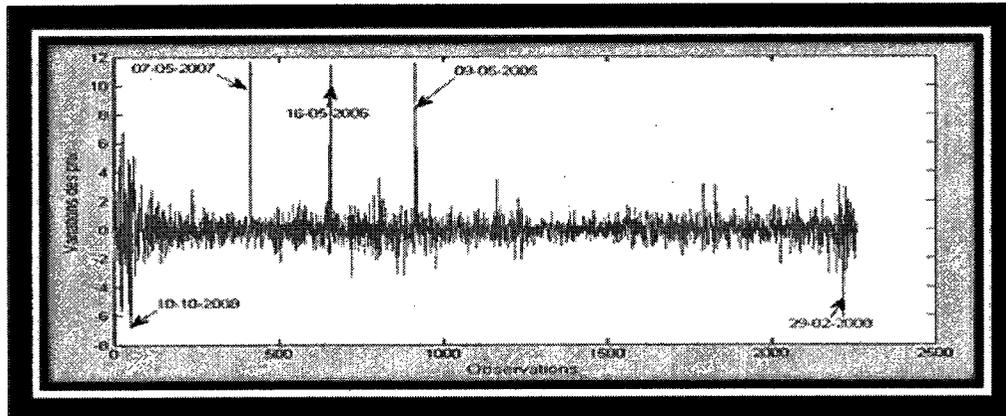


Figure 4.31 – Composante indépendante numéro 23

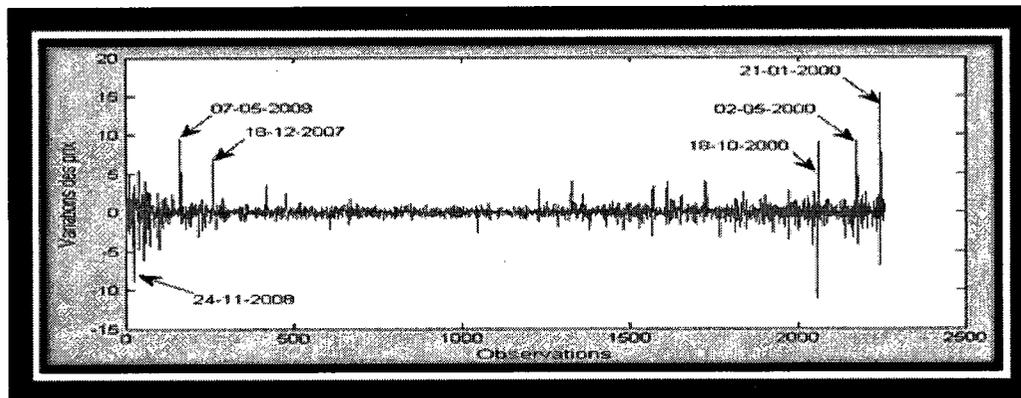


Figure 4.32 – Composante indépendante numéro 25

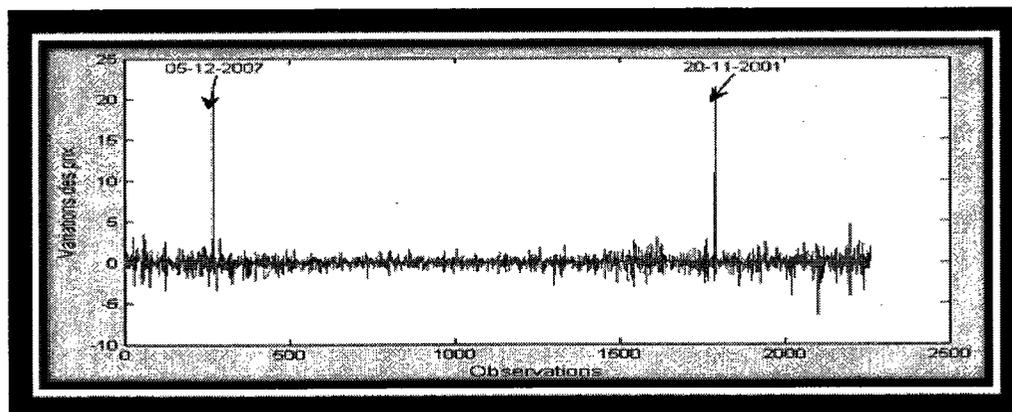


Figure 4.33 – Composante indépendante numéro 14

La reconstitution des titres avec les IC sélectionnées est présentée dans les figures 4.34, 4.35 et 4.36.

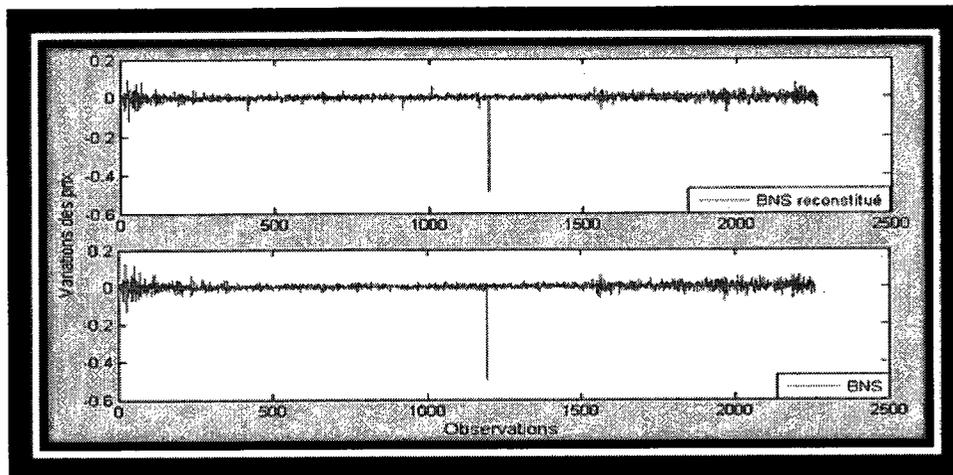


Figure 4.34 – Le titre BNS et sa reconstitution avec le critère EAM

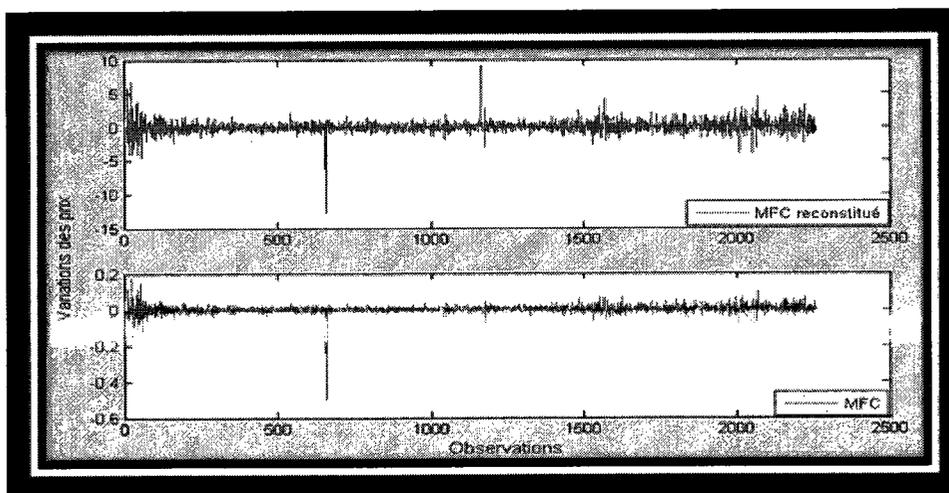


Figure 4.35 – Le titre MFC et sa reconstitution avec le critère EAM

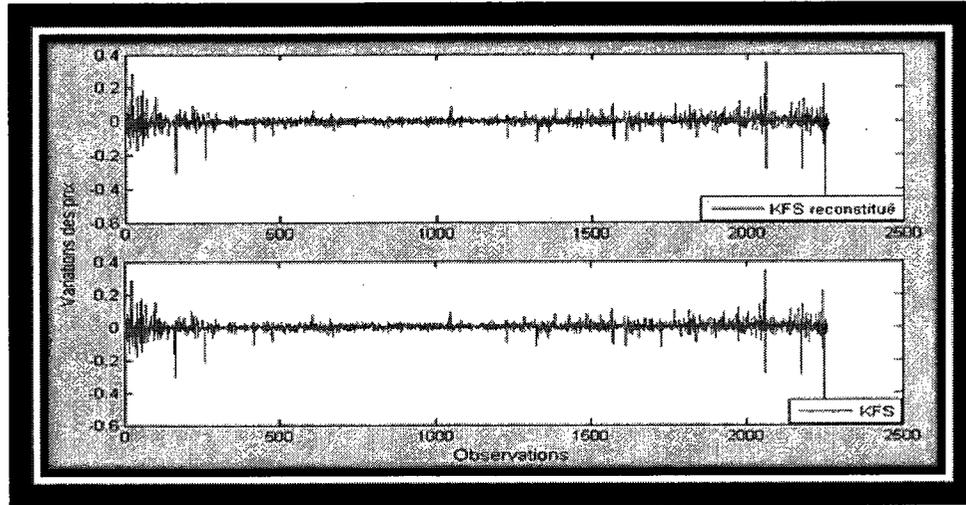


Figure 4.36 – Le titre KFS et sa reconstitution avec le critère EAM

4.4.2 Reconstitution avec le critère séquentiel

La comparaison des séquences entre les 3 titres et les 25 composantes selon le test du khi2 a permis de sélectionner les composantes indiquées dans le tableau 4.12.

MFC		BNS	
IC	χ^2	IC	χ^2
IC3	0.12	IC8	4.13
IC10	3.72	IC21	3.54
IC4	5.42		
IC14	5.43		
IC24	5.46		

Tableau 4.12 – Tableau des composantes retenues selon leur χ^2

Les composantes IC8, IC4, IC24, IC14 et IC3 ont été expliquées dans le 4ème chapitre de ce mémoire. Les figures 4.37 et 4.38 montrent les variations des prix des composantes IC10 et IC21.

La reconstitution séquentielle nous explique les mêmes phénomènes soulignés par la re-

constitution avec le critère de l'erreur absolue moyenne. En effet, MFC et BNS sont affectés par la production (IC4,IC8 et IC21), les variations des taux de changes (IC24 et IC21), la crise économique et financière (IC10 et IC14) et l'augmentation du prix du pétrole en 2003 (IC3 et IC21).

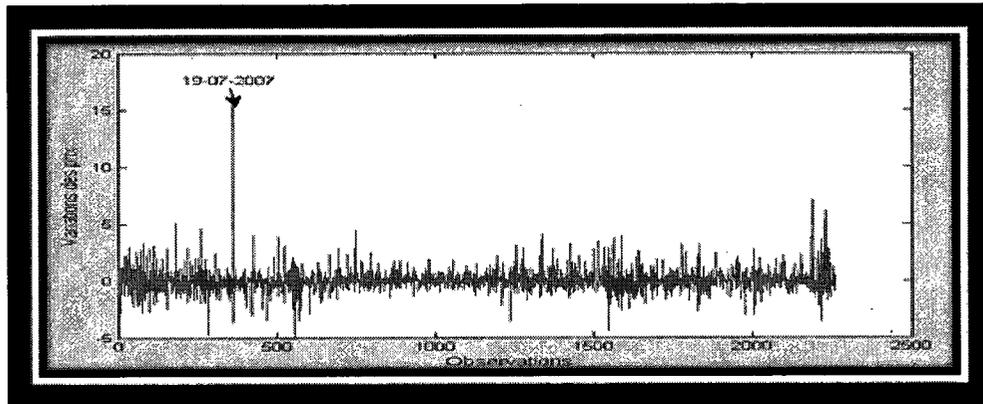


Figure 4.37 – Composante indépendante numéro 10

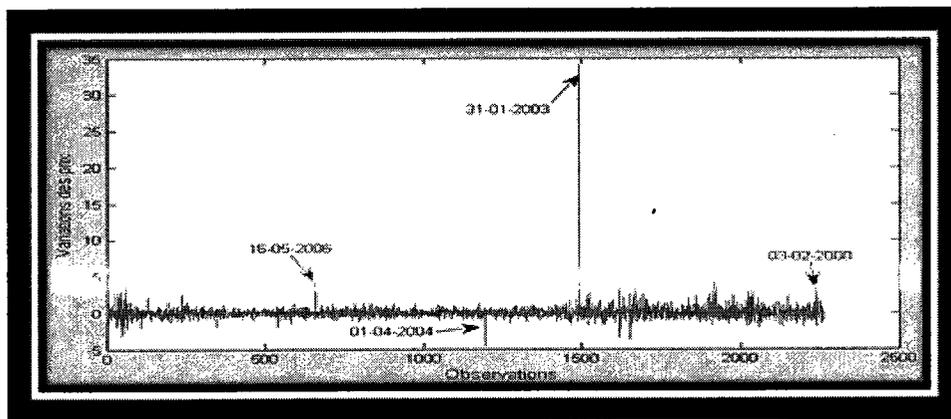


Figure 4.38 – Composante indépendante numéro 21

La reconstitution des titres avec les IC sélectionnées selon le critère séquentiel est présentée dans les figures 4.39 et 4.40.

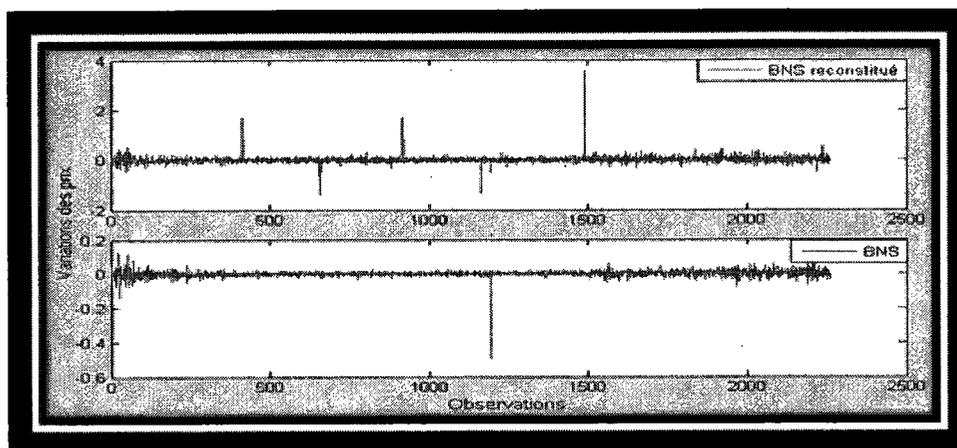


Figure 4.39 – Le titre BNS et sa reconstitution avec le critère séquentiel

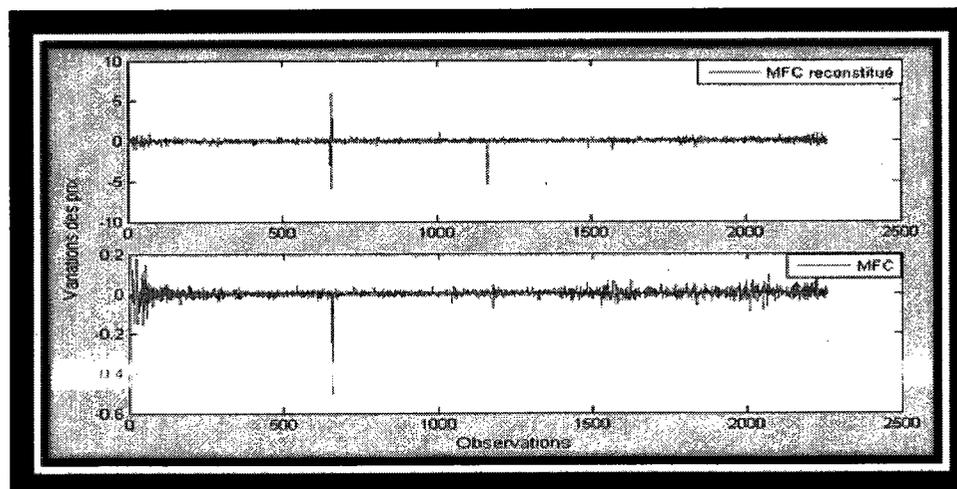


Figure 4.40 – Le titre MFC et sa reconstitution avec le critère séquentiel

Notons que pour le titre KFS, nous avons rejeté l'hypothèse H_0 (La répartition séquentielle des composantes est similaire à celle du titre KFS) pour toutes les composantes indépendantes. La répartition séquentielle de KFS ne concorde pas avec la répartition séquentielle des composantes indépendantes.

4.4.3 Analyse sectorielle

Les tableaux 4.13 et 4.14 présentent les résultats de notre analyse du secteur financier. Cette analyse permet de révéler le lien existant entre le secteur financier et celui de l'énergie.

MFC			BNS			KFS		
IC	Rangs	Poids	IC	Rangs	Poids	IC	Rangs	Poids
IC7	3	0.5	IC9	2	0.67	IC25	2	0.67
IC9	2	0.33	IC23	1	0.33	IC14	1	0.33
IC17	1	0.17						
Total	6	1	Total	3	1	Total	3	1

Tableau 4.13 – Tableau des poids de chaque composante dans la reconstitution des titres MFC, BNS et KFS

IC	Prévalences non pondérées PNP	Prévalences pondérées PP
IC9	0.33	0.33
IC25	0.22	0.17
IC7	0.17	0.25
IC23	0.11	0.08
IC14	0.11	0.08
IC17	0.06	0.08

Tableau 4.14 – Tableau des prévalences non pondérées et des prévalences pondérées de chaque composante dans le secteur financier

4.5 Estimation des titres

Dans cette section nous présentons les résultats de la prévision des titres MFC, BNS et KFS avec l'ACI. Le critère de classification utilisé est le critère EAM.

	MFC		BNS		KFS	
	Concor.	Taux de prév.	Concor.	Taux de prév.	Concor.	Taux de prév.
Modèle ACI	0.45454	0.57355	0.43182	0.73097	0.5	0.70925
Modèle AR	0.52273	0.58907	0.47728	0.70585	0.55682	0.6506

Tableau 4.15 – Comparaison de la concordance et le taux de prévision entre notre modèle et le modèle classique AR.

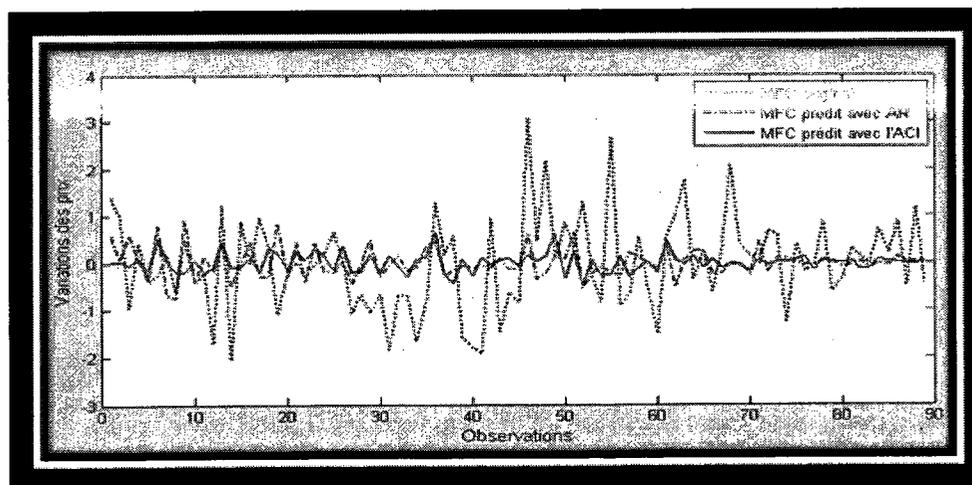


Figure 4.41 – Le titre original MFC et sa prédiction de 2009

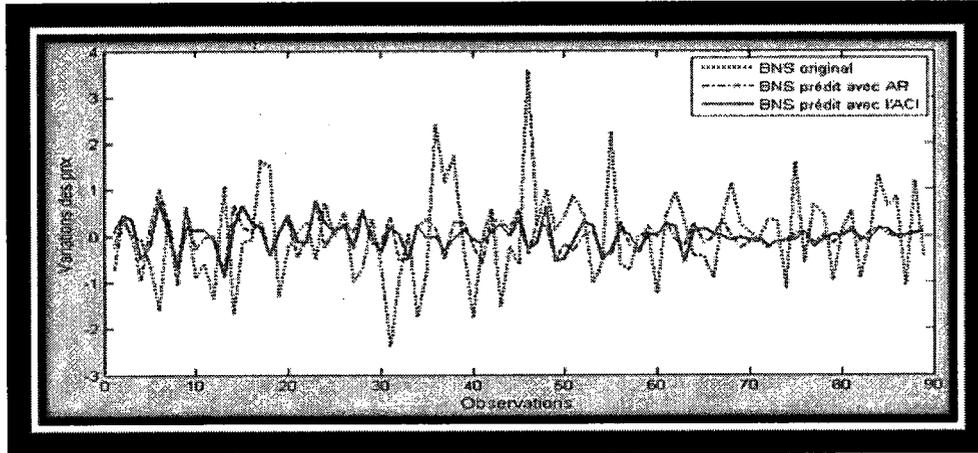


Figure 4.42 – Le titre original BNS et sa prédiction de 2009

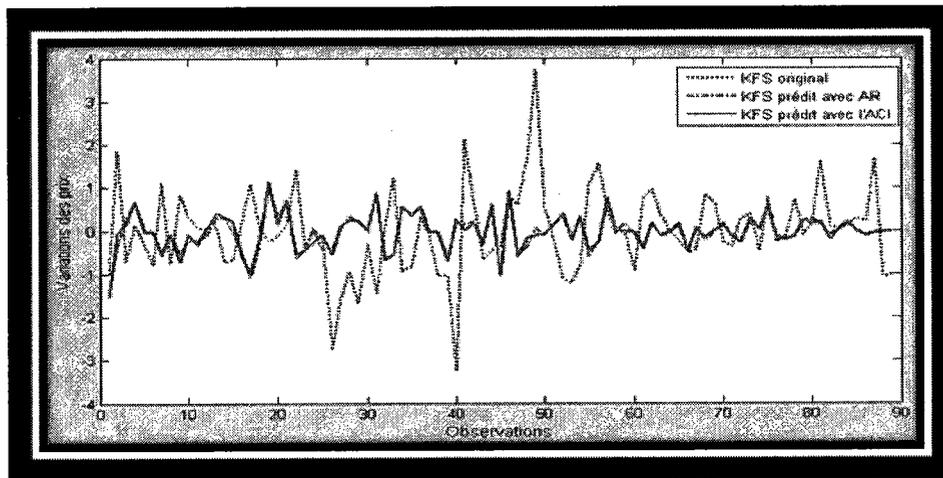


Figure 4.43 – Le titre original KFS et sa prédiction de 2009

ANNEXE B

4.6 La matrice du mélange A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	20	21	22	23	24	25
AD	0.016	-0.01	-0.002	0.041	0.051	-0.055	-0.01	0.028	0.02	0.01	0.045	-0.012	-0.043	0.008	-0.0001	-0.031	-0.0562	0.0576	-0.057	0.08	0.72	0.015	-0.02429
ADY	0.04	0.03	-0.003	-0.08	0.072	-0.029	-0.03	-0.35	-0.04	0.057	-0.032	-0.064	0.003	-0.0399	-0.014	-0.0177	0.0484	0.0181	0.12	-0.83	0.061	-0.03389	
ALM	0.02	0.029	-0.02	-0.61	0.079	-0.053	-0.05	0.431	-0.02	0.069	-0.016	-0.05	0.0105	0.0205	-0.02	-0.0088	0.0549	0.0599	0.09	-0.63	0.068	-0.02976	
AL	-0.02	0.02	0.007	0.056	0.007	-0.033	0.047	0.056	-0.1	-0.01	0.002	-0.03	0.001	-0.0185	-0.0012	0.007	0.0014	0.0297	0.9542	0.16	-0.16	0.031	0.002429
CCO	0.022	0.01	-0.021	0.007	0.052	-0.047	0.063	0.075	0.05	0.05	0.075	-0.035	-0.033	-0.0032	0.0222	-0.019	0.0361	0.0368	0.0668	0.16	-0.31	0.917	-0.01646
DAN	0.011	0.021	-0.03	-0.03	0.027	-0.033	0.047	0.071	0.02	0.02	0.037	-0.005	-0.022	0.0964	0.0218	-0.036	0.0315	0.0215	0.0515	0.92	-0.18	0.06	-0.00684
ENR	0.025	-0.0038	-0.01	0.1	-0.034	0.057	0.058	0.18	0.03	0.091	-0.034	-0.017	-0.0273	-0.0204	-0.052	0.0374	0.9415	-0.018	0.11	-0.12	0.078	-0.07346	
CAE	0.056	0.052	-0.024	1E-03	0.058	-0.011	0.016	0.044	0.03	-0.01	0.086	-0.023	-0.038	-0.0157	0.0445	-0.965	0.0998	0.0317	-0.018	0.08	-0.13	0.041	-0.07458
BSB	0.056	0.034	-0.025	-0.0105	-0.078	0.014	0.046	0.14	0.01	0.034	-0.028	-0.92	-0.0417	0.0955	-0.174	0.0753	0.0387	0.0414	0.13	-0.14	0.036	-0.06334	
ECA	0.072	0.027	0.003	-0.01	0.076	0.007	0.01	0.006	0.03	0.06	0.074	-0.026	-0.033	-0.0011	0.9753	-0.048	0.0419	0.061	-0.022	0.07	-0.05	0.011	-0.06326
ACC	0.011	0.01	-0.938	-0.003	-0.027	-0.003	0	0	0.004	-0.014	-0.005	0.0089	0.0034	-5E-04	0.0191	-0.0175	0.0279	0	-0	-0.01	-0.01373		
SC	0.013	0.019	-0.037	0.019	0.028	-0.066	-0.02	0.05	-0.01	-0.02	-0.992	-0.002	-0.0051	0.02	0.002	-0.0126	-0.016	0.0124	0.02	0.002	0.031	-0.02536	
SAP	0.022	0.049	0.003	5E-04	0.046	-0.069	0.011	0.026	0.04	-0.01	0.015	-0.031	0.0029	-0.9881	0.032	-0.03	0.0289	0.0163	-0.023	0.01	-0.05	5E-04	0.011562
FGL	0.032	0.975	0.034	0.032	0.01	-0.03	0.026	0.037	0.06	-0.01	-0.047	-0.039	-0.0112	0.007	-0.028	0.0327	0.0601	0.0438	0.06	-0.07	0.01	-0.11758	
MDS	0.025	0.014	0.005	0.01	0.025	-0.02	0.037	0.031	0.07	0.04	0.978	-0.025	-0.077	-0.0418	0.0232	-0.039	0.0391	0.0225	-7E-04	0	-0.1	0.029	-0.06432
VED	0.106	0.042	0.007	-0.01	0.042	-0.003	0.003	0.004	0.01	0.98	0.041	-0.035	0.0058	0.0327	0.0222	-0.077	0.0018	0.0113	-0.006	0.04	-0.02	0.019	-0.02078
MFC	0.007	0.011	-0.014	-0.27	0.071	-0.058	0.781	0.22	0.35	0.03	0.062	-0.046	-0.06	-0.0257	0.0422	-0.092	0.1046	0.0652	0.013	0.03	-0.3	-0.04	-0.0366
BNS	-0.03	0.028	0.005	-0.05	0.074	-0.003	-0.14	0.095	0.32	0.02	0.13	-0.034	-0.01	-0.0043	0.0303	-0.046	0.1203	0.0241	0.1031	0.07	-0.2	1E-03	-0.10916
UTC	0.101	0.048	0.002	-0.03	0.054	-0.005	0.019	0.012	0.05	0.03	0.073	-0.014	-0.055	0.0094	0.0579	-0.064	0.0563	0.0014	0.0027	0.05	-0.06	-0	-0.05136
MMM	0.963	0.012	-0.01	0.009	0.052	-9E-04	0.068	0.029	0.13	0.02	0.095	-0.031	-0.043	0.0055	0.0445	-0.095	0.0516	0.024	0.0106	0.04	-0.03	0.046	8.46E-03
BGLB	0.07	0.057	0.024	0.017	0.555	-0.034	0.03	0.027	0.05	0.01	0.105	-0.009	-0.03	0.0289	0.147	-0.075	0.0122	0.0284	-0.003	0.07	-0.05	0.015	0.004794
BSY	0.057	0.072	-0.007	-0.01	0.012	-0.047	0.025	0.041	0.04	0.04	0.016	-0.016	0.0295	0.0015	0.0239	0.01	0.0904	-0.0134	-0.007	0.1	-0.08	0.038	0.002544
EMA	0.046	-0.0111	-0.01	0.02	-0.981	0.031	0.014	0	0.02	0.051	-0.026	0.0358	0.0352	0.0372	-0.024	0.0955	0.0072	-9E-04	0.04	-0.09	0.03	-0.03052	
ES	0.025	0.008	0.002	-0.053	-0.035	0.004	0.037	0.02	0.06	0.032	0.0002	-0.02	-0.0327	-0.0002	-0.017	0.0225	0.0042	0.0103	0.04	-0.05	0.027	-0.98883	

Figure 4.44 – Les coefficients de la matrice du mélange A

Bibliographie

- [AC97] S.-I. Amari and J.-F. Cardoso. Cardoso Blind Source Separation Semiparametric Statistical Approach. *IEEE Trans. on Signal Processing*, 45(11) :2692–2700, 1997.
- [AC98] S.I. Amari and A. Cichocki. Adaptive Blind Signal Processing Neural Network Approaches. In *Proc. of the IEEE*, volume 86, October 1998.
- [AU94] A.Cichocki and R. Unbehauen. *Neural Networks for Signal Processing and Optimisation*. Wiley, 1994.
- [BS04] C. F. Beckmann and S. M. Smith. Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Trans. on Medical Imaging*, 23(2) :137–152, 2004.
- [BW97] A. D. Back and A. S. Weigend. A First Application of Independent Component Analysis to Extracting Structure from Stock Returns. *Neural Systems*, 8(5), 10 1997.
- [Car98] J. F. Cardoso. Blind Signal Separation : Statistical Principles. *Proceedings of the IEEE*, 9(10) :2009–2025, 10 1998.
- [Car99] J.-F. Cardoso. High-Order Contrast for Independent Component Analysis. *Neural Computation*, 11 :157–192, 1999.

- [CC80] C. Chatfield and A. J. Collins. *Introduction to Multivariate Analysis*. Chapman and Hall, 1980. ISBN : 0-412-16030-7.
- [CL92] J.-F. Cardoso and B. Laheld. Adaptive Blind Source Separation for Channel Spatial Equalization. In *Proc. of COST 229 Workshop on Adaptive Signal Processing*, pages 19–26, 1992.
- [CL96] J. F. Cardoso and B. Laheld. Equivariant Adaptive Source Separation. *IEEE Trans. Signal Processing*, 44(12) :3017–3030, 12 1996.
- [CM92] A. Cichocki and L. Moszczynski. A New Learning Algorithm for Blind Separation of Sources. *Electronics Letters*, 28(21) :1986–1987, 1992.
- [Com94] P. Comon. Independent Component Analysis : A New Concept? *Signal Processing*, 36(3) :287–314, 1994.
- [CS93] J.-F. Cardoso and A. Souloumiac. Blind Beamforming for Non-Gaussian Signals. *IEE Proceedings-F*, 140 :362–370, 1993.
- [CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, NY, USA, 1991.
- [CU96] A. Cichocki and R. Unbehauen. Robust Neural Networks with On-Line Learning for Blind Identification and Blind Separation of Sources. *IEEE Trans. on Circuit and Systems*, 439(11) :894–906, 1996.
- [CUR94] A. Cichocki, R. Unbehauen, and E. Rummert. Robust Learning Algorithm for Blind Separation of Signals. *Electronics Letters*, 30(17) :1386–1387, 1994.
- [Deg01] S. Degerine. Algorithmes de diagonalisation conjointe approchée. In C. Jutten and A. Guérin-Dugué, editors, *Actes De l'école de printemps : De la séparation des sources à l'analyse en composantes indépendantes : Méthodes, Algorithmes et Applications*, pages 129–144, Villard de Lans, France, 2001.
- [DL95] N. Delfosse and Ph. Loubaton. Adaptive Blind Separation of Independent Sources : A Deflation Approach. *Signal Processing*, 45 :59–83, 1995.

- [HA84] J. Héroult and B. Ans. Circuit neuronaux à synapses modifiables : décodage de messages composites par apprentissage non supervisé. *C.R. de l'Académie des Sciences*, 299(3-13) :525-528, 1984.
- [Hay94] S. Haykin. *Blind Deconvolution*. Prentice Hall, 1994.
- [HI02] A. Hyvarinen and M. Inki. Estimating Overcomplete Independent Component Bases for Image Windows. *Mathematical Imaging and Vision*, 17 :139-152, 2002.
- [HJA85a] J. Héroult, C. Jutten, and B. Ans. Adaptive Neural Architectures : Detection of Primitives. In *COGNITIVA 85*, pages 593-597, Paris, 1985.
- [HJA85b] J. Héroult, C. Jutten, and B. Ans. Détection de grandeurs à primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *Actes du Xème colloque GRETSI*, pages 1017-1022, 1985.
- [HKO01] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley and Sons, 2001.
- [Hul99] M. Van Hulle. Clustering Approach to Square and Non-Square Blind Source Separation. *Neural Networks for Signal Processing*, pages 315-323, 1999.
- [Hyv97] A. Hyvärinen. One-Unit Contrast Functions for Independent Component Analysis : A Statistical Analysis. *Neural Networks for Signal Processing*, 24 :388-397, 1997.
- [Hyv99] A. Hyvarinen. Fast and Robust Fixed Point Algorithms for Independent Component Analysis. *Electronics Letters*, 10(3) :626-634, 1999.
- [Iha93] S. Ihara. *Information Theory for Continuous Systems*. World Scientific, 1993.
- [KO98] K. Kiviluoto and E. Oja. Independent Component Analysis for Parallel Financial Time Series. In s. Usui and T. Omori, editors, *International Conference*

- on *Neural Information Processing*, volume 2, pages 895–898, Tokyo, Japan, 1998.
- [LCX99] Z. Lai, Y. Cheung, and L. Xu. Independent Component Ordering in ICA Analysis of Financial Data, 1999.
- [LS00] S. Lewicki and T. J. Sejnowski. Learning Overcomplete Representation. *Neural Computation*, 12(2) :337–365, 2000.
- [M03] E. G. Miller and J. W. Fisher 3. Independent Components Analysis by Direct Entropy Minimization. Technical Report CSD-03-1221, University of California, 2003.
- [MKO99] S. Malaroiu, K. Kiviluoto, and E. Oja. Time Series Prediction with Independent Component Analysis. In *Advances Investment Technologies*, pages 20–21, Gold Coast, Australia, 12 1999.
- [Nan99] A. Nandi. *Blind Estimation Using Higher Order Statistics*. Kluwer, 1999.
- [OF96] B.A Olshausen and D.J. Fields. Emergence of Simple-cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature*, 381 :607–609, 1996.
- [OF97] B.A Olshausen and D.J. Fields. Sparse Coding with an Overcomplete Basis Set : A Strategy Employed by v1. *Vision Research*, 37(23) :3311–3325, 1997.
- [OL99] B.A Olshausen and M.S. Lewicki. A Probabilistic Framework for the Adaptation and Comparison of Image Codes. *Optical Society of America*, A 16 :1587–1601, 1999.
- [PG97] D.-T. Pham and Ph. Garat. Blind Separation of Mixtures of Independent Sources Through a Quasi Maximum Likelihood Approach. *IEEE Trans. Signal Processing*, 45(7) :1712–1725, 1997.
- [SJK07] H. Shen, S. Jegelka, and A. Gretton. Fast Kernel ICA using an Approximate Newton Method. In M. Meila and X. Shen, editors, *Proceedings of the 11th*

International Conference on Artificial Intelligence and Statistics (AISTATS), pages 476–483, Brookline, MA, USA, 03 2007. Microtome.

- [Sor91] E. Sorouchyari. Blind Separation of Sources, Part iii : Stability Analysis. *Signal Processing*, 24 :21–29, 1991.
- [ST04] A. Samarov and A. Tsybakov. Nonparametric Independent Component Analysis. *Bernoulli*, 10(4) :565–582, 08 2004.
- [Tor99] K. Torkkola. Blind Separation For Audio Signals : Are We There Yet? In *Actes of ICA*, pages 239–244, Aussois, France, 1999.
- [VHO⁺98] R. Vigário, M. Hämmäläinen, E. Oja, R. Hari, and V. Jousmäki. Independent Component Analysis for Identification of Artifacts in Magnetoencephalographic. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Neural Information Processing Systems 10*, Cambridge, MA, USA, 1998. MIT Press.
- [ZP01] M. Zibulevsky and B.A. Pearlmutter. Blind Source Separation by Sparse Decomposition in a Signal Dictionary. *Neural Computations*, 13(4) :863–882, 2001.