



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة عمار ثلجي بالأغواط
Université Amar Thelidji (Laghouat)
كلية العلوم و التكنولوجيا
Faculté de Science et Technologie
Département d'Electronique

Mémoire de MASTER

Domaine : Science Technologie.
Filière : Télécommunication.
Option : Réseaux et Télécommunication.

Présenté par

- HAMAI Katia.
- KELFA Ryma.

THEME

Analyse en Composantes Indépendantes (ACI) :
Signal parole

<u>Membre de jury</u>	<u>Grade</u>	<u>Qualité</u>
Mr. REGUIGUE Mourad	(MCB)	Président
Mr. CHELALI Safouane	(MCB)	Examineur
Mr. REGGAB Mourad	(MAA)	Encadreur

Année Universitaire 2019/2020

Remerciements

En tout premier lieu, on remercie le bon Dieu, qui nous a donné la force et la patience d'accomplir ce Modeste travail

On voudra adresser toute nos gratitudee au l'encadreur de notre mémoire, *Mr REGGAB Mourad*, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion.

On désira aussi à remercier les professeurs du département d'Électronique de l'université de Laghouat, pour leurs efforts constants pour nous enseigner, leurs précieuses informations et conseils, ainsi que leurs soutiens et encouragements.

Comme on exprimer notre appréciation envers nos amis et collègues qui ont pris en charge notre soutien moral et intellectuel tout au long de notre préparation de cette mémoire.

Enfin, je remercie tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Dédicaces

Je dédie ce travail à

Mes très chers parents,

Comme un gage d'appréciation et d'amour pour tous leurs sacrifices qu'ils ont faits pour moi, ainsi que leurs soutiens et encouragements tout au long de mes études, espérant qu'il soit une source de fierté et de bonheurs pour eux.

Ma merveilleuse sœur Keyssa,

Pour ses soutiens, encouragements et conseils, et surtout pour son support les moments de détresse. Sans oublier mes petits anges et la source de mon bonheur

FATÉH et KAHINA

Mon binôme Ryma,

La fille qui a vraiment travaillé avec moi, m'a soutenu et m'a encouragé et m'a aidée tous ces années que ce soit dans les études ou la vie personnelle.

Zahra et Hadjer et imane et Hamza,

Les personnes qui m'ont toujours conseillé, qui m'ont écouté et qui étaient avec moi dans les bons et les mauvais moments.

Que dieu vous garde tous pour moi.

H. Katia



Dédicaces

Je dédie ce travail à :

À mes chers parents,

Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être. Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours. Que ce modeste travail soit l'exaucement de vos vœux tant formulés, le fruit de vos innombrables sacrifices, bien que je ne vous en acquitterai jamais assez.

À mon très cher frère Aziz

Qui a été présent dans tous mes moments d'examens par son soutien moral et ses belles Surprises sucrées.

À ma petite sœur adorable Racha,

Qui sait toujours comment parcourir la joie et le bonheur pour toute la famille.

À mon binôme Katia,

Qui m'a assisté dans les moments difficiles et m'a pris doucement par la main pour traverser ensemble des épreuves pénible. Je te suis très reconnaissante, et je ne te remercierai jamais assez pour ton amabilité, ta générosité, ton aide précieuse.

À mon cher ami Yacine,

Qui m'a épaulé moralement tous les jours dans la construction de ce mémoire.

À mes chers amis Manel Kayssa Zahra Hadjer Iman,

En souvenir de notre sincère et profonde amitié et des moments agréables que nous avons passés ensemble. Je ne peux trouver les mots justes et sincères pour vous exprimer mon affection et mes pensées, vous êtes pour moi des sœurs et des amis sur qui je peux compter.

À mes chères cousines Yasmín Razika Lila Khaoula,

Veillez trouver dans ce travail l'expression de mon respect le plus profond et mon affection la plus sincère.

K. Rima



تلخيص

تحليل المكون المستقل (ACI) و طريقة مستخدمة لحل مشكلة الفصل الأعمى للمصادر (SAS) ، حيث يتمثل الهدف في إيجاد تمثيل خطي للبيانات غير الغوسية بحيث تكون المكونات مستقلة إحصائياً أو مستقلة قدر الإمكان. استخدامات تحليل المكون المستقل موجود في العديد من المجالات مثل: الطب الحيوي، الاتصالات، التصوير وغيرها. Fast-ICA هي واحدة من خوارزميات تحليل المكون المستقل لأكثر كفاءة لحساب مصفوفة المزج العكسي لاستخراج إشارات المصدر. تفصل هذه الخوارزمية المصادر المستقلة عن خليطها بقياس القيم غير الغوسية. في هذه المذكرة، سنطبق هذه الخوارزمية لتحقيق الفصل المصدر في Matlab.

كلمات المفتاحية: التحليل لمكونات مستقلة، الفصل الأعمى للمصادر، مصادر مستقلة إحصائياً ، مصادر غير غاوسية، خوارزمية Fast-ICA.

Résumé

L'analyse en composantes indépendantes est une méthode qui sert à résoudre le problème de la séparation aveugle des sources SAS, où le but est de trouver une représentation linéaire des données non-gaussiennes afin que les composantes soient statistiquement indépendantes, ou aussi indépendantes que possible. L'utilisation de la technique ACI est présente dans plusieurs domaines tel que : domaine biomédicaux, télécommunication, imageries et plusieurs d'autres. Fast-ICA est l'une des algorithmes ACI les plus efficaces pour calculer la matrice de mélange inverse pour extraire les signaux sources. L'algorithme Fast-ICA sépare les sources indépendantes de leurs mélanges en mesurant des valeurs non gaussiennes. Dans ce mémoire, nous appliquons cet algorithme pour réaliser la séparation des sources sous Matlab.

Mots clés : Analyse en composantes indépendantes, Séparation aveugle des sources, Indépendances statistiques, La non-gaussianité des sources, Algorithme Fast-ICA.

Abstract:

Independent component analysis is a method used to solve the problem of blind sources separation SAS, where the goal is to find a linear representation of non-gaussian data so that the components are statistically independent, or as independent as possible. The use of the ICA technique is present in several fields such as: biomedical, telecommunications, imaging and many others. Fast-ICA is one of the most efficient ACI algorithms for calculating the inverse mixing matrix for extracting source signals. The Fast-ICA algorithm separates independent sources from their mixtures by measuring non-gaussian values. In this thesis, we apply this algorithm to achieve source separation in Matlab.

Keywords : Independent component analysis, Blind sources separation, Statistical independence, Non-gaussian of sources, Fast-ICA algorithm

Sommaire

Remerciement

Dédicaces

Résumé

Introduction générale

1. Introduction	2
2. La problématique	3
3. Les hypothèses	4
4. La motivation	4
5. Organisation de la thèse	5

Chapitre1 : Analyse en composantes indépendantes

1.1 Introduction	8
1.2 ACI au fil des années	8
1.3 Définition d'ACI	9
1.4 L'indépendance statique	13
1.5 Théorème de centre limite	14
1.6 Prétraitement des données pour ACI	15
1.6.1 Centrage	15
1.6.2 Blanchiment	16
1.7 Fonction de contraste	18
1.7.1 La mesure de la non-gaussianité:	18
1.7.1.1 Kurtosis	20
1.7.1.2 La néguentropie	21
1.7.2 Minimisation des informations mutuelles:	24
1.7.3 Maximum de vraisemblance	25
1.8 Les algorithmes d'ACI	25
1.8.1 Algorithme JADE	26
1.8.2 Infomax	26
1.8.3 Algorithme Fast-ICA	26
1.9 Conclusion	27

Chapitre2 : Algorithme Fast-ICA

2.1 Introduction	29
2.2 Choix du Fast-ICA	29
2.3 Fast-ICA	33
2.3.1 Fast-ICA pour une seule unité	34
2.3.2 Fast-ICA pour plusieurs unités	35
2.4 Orthonormalisation de l'algorithme Fast-ICA	36
2.4.1 Méthode déflation	37
2.4.2 Méthode symétrique	37

2.5 Propriétés du FastICA	38
2.6 Conclusion	39

Chapitre 3 : Mise en œuvre de l'algorithme Fast-ICA

3.1 Introduction	41
3.2 Mise en œuvre du centrage	41
3.3 Mise en œuvre du blanchiment	42
3.4 Calcul de norme	45
3.5 Mise en œuvre de la kurtosis	45
3.6 Conclusion	49

Chapitre 4 : Résultat et discussion

4.1 Introduction	51
4.2 Simulation de l'algorithme Fast-ICA (à l'aide de Matlab) :	51
4.2.1 Chargement des données	51
4.2.2 Génération des données :	51
4.2.3 Prétraitement	55
4.2.4 Séparation des signaux	57
4.3 Comparaison entre les signaux de sources et les signaux séparés	58
4.3 Conclusion	60

Conclusion générale

1. Conclusion générale	62
2. Les travaux futurs et perspectives	63

Les références.	64
-----------------	----

Liste de figure

Figure 1.1 : La récupération des signaux sources " s " n à partir de leurs mélanges en utilisant ACI	9
Figure 1.2 : La distribution conjoints de sources s_1 et s_2 indépendantes	10
Figure 1.3 : La distribution conjoints de mélange x_1 et x_2	10
Figure 1.4 : Illustration d'une distribution centrée	14
Figure 1.5 : Illustration d'une distribution conjointe des mélanges avant le blanchiment	16
Figure 1.6 : Illustration d'une distribution conjointe des mélanges blanchis	16
Figure 1.7 : la distribution conjoints de sources s_1 et s_2 et de mélanges x_1 et x_2	17
Figure 1.8 : Super-gaussienne, gaussienne et sub-gaussienne	18
Figure 1.9 : Les distributions des variables obtenu par le calcul de kurtosis	19
Figure 2.1 : Erreur E_1 en fonction de la précision	28
Figure 2.2 : Erreur E_1 en fonction de la charge informatique	29
Figure 2.3 : Erreur en fonction de nombre de sources	30
Figure 2.4 : Réseau de neurones de 1' ACI	31
Figure 2.5 : Un seul neurone de 1' ACI	32
Figure 3.1 : Schéma fonctionnel de la mise en œuvre du centrage	42
Figure 3.2 : Schéma fonctionnel de la mise en œuvre de blanchiment	44
Figure 3.3 : Diagramme sur calculer de la norme	45
Figure 3.4 : Diagramme de la mise en œuvre de $Per_w(k)$	47
Figure 3.5 : Diagramme de la mise en œuvre de $w(k)$	48
Figure 4.1 : Signal source 1.	52
Figure 4.2 : Signal source 2.	52
Figure 4.3 : La distribution des signaux sources S_1 et S_2 .	53
Figure 4.4 : Signal mixte X_1 .	54
Figure 4.5 : Signal mixte X_2 .	54
Figure 4.6 : La distribution des signaux mixtes X_1 et X_2 .	55
Figure 4.7 : Avant et après le centrage.	56
Figure 4.8 : Avant et après le blanchiment.	56
Figure 4.9 : signal source estimé 1.	57
Figure 4.10 : signal source estimé 2.	58
Figure 4.11 : Signal d'erreur entre S_1 et signal estimé1.	59
Figure 4.12 : Signal d'erreur entre S_1 et signal estimé2.	59

Liste d'abréviations

ACI : Analyse en Composantes Indépendantes.

SAS : Séparation aveugle des sources.

EEG : L'enregistrement l'électroencéphalogramme

MEG : La magnétoencéphalographie.

ECG : L'enregistrement électrocardiogramme.

CDMA : techniques d'accès multiple par répartition en code

ACP : l'Analyse en composantes principales.

IS : Indépendante Statique.

HOS : Statistiques d'ordres plus élevés.

ML: Maximum of likelihood.

JADE: Joint Approximation Diagonalization of Eigenmatrices.

FP : Algorithme du point fixe simple.

FPsym : Algorithme du point fixe avec la méthode déflation.

FPsymth : Algorithme du point fixe avec la méthode symétrique.

ACY : Algorithme naturel du gradient.

WACY : White ACY.

BS : Algorithme de Bell et Sejnowski.

ExtBS : algorithme de gradient basé sur le critère de l'infomax.

EASI: Equivariant Adaptive Separation via Independence.

NPCA-RLS: Algorithm Nonlinear Principal Component Analysis by Recursive Least Squares).

Liste des signes

A : Coefficient de mélange.

W : Inverse de A .

m_x : La moyenne de X_k .

σ_k : La variance de X_k .

V : Matrice des mélanges blanchis V .

U : La matrice de blanchiment.

C : La matrice de covariance.

D_C : La matrice des valeurs propres.

W^T : La matrice de séparation dans l'espace blanchi.

$Kurt(y)$: Le cumuland d'ordre 4.

$H(y)$: Fonction d'entropie.

J : La néguentropie.

G_i : Fonction non quadratique.

I : Information mutuelle.

L : Maximum de vraisemblance.

φ_i : Fonctions scalaires non linéaires.

D : Matrice diagonale.

E : Matrice diagonale

Introduction générale

1. Introduction :

Dans les différents domaines, nous sommes toujours face à traiter d'énormes quantités des données. La plupart de temps, nous voulons utiliser une petite quantité d'informations utiles qui se trouve à l'intérieur de cette énorme quantité de données [1].

Depuis les temps anciens, des nombreuses méthodes d'analyse des données ont été utilisées, et au fil de temps, elles ont été développées pour faciliter le traitement des informations. L'analyse des données recouvre deux ensembles de techniques : certains de ces techniques sont utilisés pour extraire des informations utiles des données. Autres méthodes ont pour but de regrouper les données pour les clarifier et les mieux comprendre [2].

Dans cette mémoire, nous nous concentrerons sur la d'analyse des données qui vise à extraire des informations utiles. Parmi les méthodes d'analyse des données les plus utilisé et les plus connu c'est " l'Analyse en Composantes Indépendantes ACI " (En anglais : *Independent Component Analyse ICA*).

L'analyse en composantes indépendantes est connue comme une méthode qui sert à résoudre le problème de la séparation aveugle des sources SAS (En anglais : *Blind Source Separation BSS*), qui cherche à séparer les données en composants d'information de base où ces données peuvent prendre la forme d'images, parole, sons, canaux de télécommunication ou cours de bourse ...etc.[1]

Parmi l'un des scénarios de la séparation des sources est quand des capteurs (microphone par exemple) captent des signaux sources où chaque capteur contient un mélange de ces signaux sources. À la sortie de ces capteurs, le mélange est traité pour récupérer les signaux sources (voir la figure1.1). [1]

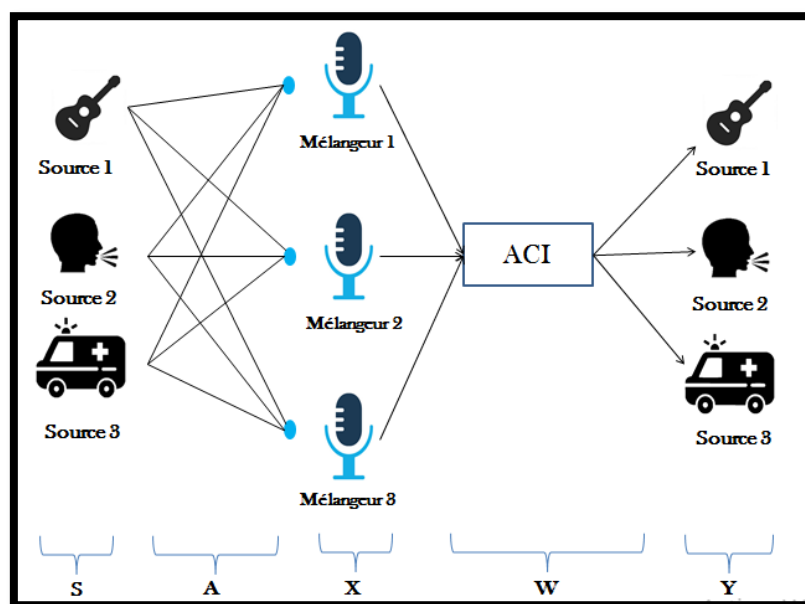


Figure 1: Exemple sur la séparation des sources.

Le terme "aveugle" (*Blind*) veut dire que la séparation se fait sans aucune information ni sur les sources ni sur le mécanisme du mélange [3].

Plusieurs algorithmes d'ACI sont connus et qui sert à extraire les signaux sources à partir d'un mélange. Parmi les algorithmes les plus utilisés c'est "FAST ICA".

2. La problématique :

➤ **Illustration de la soirée cocktail (*cocktail party*) :**

Imaginez que vous êtes à une fête où il y a de la musique et beaucoup de gens qui parlent en même temps. Vos oreilles vont capter plusieurs sons : la voix de votre amis, la voix des autres gens, la musique, les téléphone qui sons ...etc. Les oreilles humaines ont une capacité extraordinaire à différencier les voix l'une de l'autre, par exemple si vous voulez entendre ce que votre ami veut vous dire, il suffit juste de se concentrer avec lui. Donc on peut distinguer n'importe qu'elle voix ou sons on veut l'entendre même à la présence de plusieurs sources. La capacité de se concentrer et de reconnaître une source spécifique appelée l'effet cocktail party [4].

Pour but de séparer les sources (Nous les symbolisons ici "S"), nous plaçons plusieurs microphones dans la pièce. Plus que il y a des microphones dans la pièce plus la possibilité d'entendre tous dans la pièce.

Chaque microphone va recevoir un mélange (symboliser ici "X") des sources. Nous pourrions exprimer cela comme suit :

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S}$$

Où A est le coefficient de mélange et sans tenir compte de facteurs supplémentaires (sans bruit). Dans ce cas, on peut résoudre ce problème (extraire la source) si nous connaissons le paramètre A, nous pouvons résoudre l'équation linéaire par des méthodes classiques:

$$\mathbf{S} = \mathbf{W} \cdot \mathbf{X}$$

D'où W est l'inverse de A : $\mathbf{W} = \mathbf{A}^{-1}$

Le problème sera difficile si nous ne connaissons pas A. La question qui se pose est : _
Comment peut-on calculer (récupérer) les signaux sources S et les proportions A à partir seulement d'un signal mélange X ?

3. Les hypothèses :

Étant donné le problème posé, la solution au problème semble impossible, mais ce problème peut être résolu en utilisant certaines informations sur les caractéristiques statistiques des signaux sources.

Deux hypothèses qui sont liées l'une à l'autre et qui sont supposées sur les signaux sources pour résoudre le problème :

1) L'indépendance des signaux sources :

Hypothèse simple et réaliste, car si les signaux source sont provient des différents processus physiques (comme dans l'exemple de soirée cocktail), il n'y a aucun raison que les sources sera dépendantes l'un de l'autre. Par conséquent, les signaux sources sont 'statiquement indépendantes'' [5].

2) La non-gaussianité des signaux sources :

Un signal source qui n'a pas de bruit on aura une représentation qui n'est pas gaussienne. Par contre, si on fait un mélange des signaux qui ne sont pas gaussiens (on parle ici sur un signal mesuré), donc on aura une représentation gaussienne (théorème de la limite centrale).

Cela nous permet de mesurer la pureté d'un signal, si c'est pur on va avoir une représentation non-gaussienne, si non, on va avoir une représentation gaussienne.

4. La motivation :

L'utilisation de la technique ACI pour la séparation aveugle des sources a suscité un grand intérêt des chercheurs et des scientifiques où elle est présente dans plusieurs domaines. Parmi ces domaines d'applications :

1) Domaine biomédicaux : à l'aide de l'enregistrement l'électroencéphalogramme (EEG), et la magnétoencéphalographie (MEG) on peut détecter l'activité électrique du cerveau. L'algorithme ACI est utilisé pour séparer les signaux mesurés par les signaux bruités tel que les mouvements des yeux, le bruit musculaire...etc. L'ACI est utilisée aussi dans la séparation des battements cardiaques du fœtus des battements cardiaques de la mère à partir l'enregistrement électrocardiogramme (ECG) [4,6].

2) Données financiers : ACI aide à extraire les facteurs qui prédisent les mouvements financiers futurs. Par exemple, les cours des actions changent en fonction de l'évolution des facteurs de causalité au fil du temps (les derniers chiffres des ventes au détail, les taux de chômage, etc.). Le prix de chaque action est une combinaison différente de ces facteurs. Ces facteurs peuvent être supposés presque indépendants les uns des autres [1].

- 3) **Télécommunication** : Plusieurs applications des méthodes d'ICA et de séparation aveugle des sources (BSS) aux télécommunications (comme dans techniques d'accès multiple par répartition en code CDMA), pour but de détecter les signaux source reçus souhaité, mais il faut estimés en premier les quantités intermédiaires telles que les distorsions de canal, l'effet doppler ou la mauvaise synchronisation émetteur/récepteur, le bruit de canal [6,7].
- 4) **Imageries** : L'analyse par ACI est utilisée pour supprimer les bruits des images corrompues afin d'extraire une image naturelle. Cette technique est utilisée par exemple dans l'application "Google earth", où les satellites captent l'image d'une région de différents angles. Afin d'avoir une image sans bruit (dans ce cas les nuages) et nous permet de voir clairement les caractéristiques de la région, l'ACI est responsable pour éliminer ce bruit.

On peut notamment citer autre applications d'ACI comme:

- la sismologie.
- la localisation et la poursuite de cibles en radar et sonar.
- la reconnaissance des visages (face recognition).
- séparation d'extraits musicaux stéréophoniques.
- Et autre.

5. Organisation de la thèse :

Cette thèse est focalisée sur l'étude de la séparation aveugle de sources en particulier, des sources de signaux de parole en utilisant l'analyse en composantes indépendantes ACI.

L'objectif de ACI est d'analyser un signal mesuré et chercher les signaux sources qui sont caché à l'intérieur d'un signal mesuré à l'aide des capteurs sans savoir ce qu'il a comme signal source.

Cette mémoire est articulée sur quatre chapitres qui seront organisée comme suit :

Le premier chapitre sera une généralité sur l'analyse en composantes indépendantes, où nous parlerons tout d'abord sur le développement de cette technique au fil du temps et comment est même devenu efficace dans tous les domaines. Après cela, nous verrons la définition de cette technique ainsi les deux principales hypothèses qui en dépendent pour réaliser la séparation aveugle (l'indépendance statique et la non-gaussianité). Enfin, nous mentionnerons les différents algorithmes que l'ACI peut utiliser, puis concluons ce chapitre par une conclusion.

Dans le deuxième chapitre, nous parlerons sur l'algorithme Fast-ICA qui nous avons choisi de mettre en œuvre le processus de séparation dans la partie pratique de cette mémoire où nous verrons dans ce chapitre la définition et la méthode de base de cette algorithme et ces propriétés, et nous finirons par une conclusion.

Le troisième chapitre sera sur l'implémentation de différentes étapes (centrage, blanchiment, kurtosis) de l'algorithme Fast-ICA à partir des blocks de diagramme.

Dans le dernier chapitre qui sera qui sera la partie pratique de ce mémoire, où nous utiliserons Matlab pour réaliser la séparation de deux signaux audio en appliquant l'algorithme Fast-ICA.

Et nous concluons ce mémoire par une conclusion générale, qui sera un résumé de tout ce qui a été mentionné dans les chapitres précédents.

Chapitre 1

A *nalyse en composantes* *indépendantes*

1.1 Introduction :

Le traitement de signal connaît un grand développement qui continue jusqu'à l'heure actuelle. Tous les outils mathématiques, algorithmiques calculatoires utilisées ont conduit à une révolution technologique dans plusieurs domaines d'applications.

Le traitement de signal a pour but d'extraire l'information véhiculée par le signal. Ceci s'appuie sur la connaissance soit du signal, soit du système, soit des deux. Le signal porteur des informations qui nous intéressent subi à une transformation au cours de leur propagation telle qu'une superposition avec d'autres signaux de l'environnement.

La grande majorité des technologies modernes, dans la plupart des disciplines, utilisent des capteurs pour mesurer des grandeurs utiles pour l'étude ou le contrôle d'un phénomène physique. En pratique, malheureusement, les signaux obtenus grâce à ces capteurs sont souvent des mélanges de plusieurs contributions issues de signaux originels appelés sources.

La séparation aveugle de sources SAS consiste à estimer les signaux sources à partir des mélanges obtenus sur chaque capteur où cette séparation a été inspirée, à son commencement, du cerveau humain et de son organisation en réseaux neuronaux. L'analyse en composantes indépendantes ACI est une technique utilisée pour effectuer cette séparation aveugle où elle est connue comme une technique pour résoudre ce problème. Dans ce chapitre, nous parlerons sur tous ce qui concerne l'ACI et voir l'efficacité de cette technique à récupérer les signaux sources seulement à partir des signaux observé (mélanges).

1.2 ACI au fil des années:

Avant l'apparition d'ACI, la séparation aveugle de sources a été formalisée pour la première fois au début des années 1980, par des chercheurs français (*J. Herault, C. Jutten, and B. Ans*) en neurosciences et traitement du signal, pour modéliser biologiquement le codage du mouvement : un modèle simplifié de codage du mouvement dans la contraction musculaire, $x_1(t)$ et $x_2(t)$ sont deux types de signaux sensoriels mesurant la contraction musculaire. Ces derniers sont un mélange de deux facteurs $s_1(t)$ et $s_2(t)$ qui sont la position angulaire et la vitesse d'une articulation mobile. Le système nerveux doit pouvoir en quelque sorte déduire les signaux de position et de vitesse $s_1(t)$ et $s_2(t)$ à partir des réponses mesurées $x_1(t)$ et $x_2(t)$ [8].

Ces recherches ont développé un algorithme de séparation qui cherche à résoudre le problème du mélange instantané sans bruit avec une matrice de mélange carrée et des signaux à valeurs réelles. Depuis, une multitude d'algorithmes a été proposée : mélange instantané ou mélange convolutif, linéaire ou non linéaire, statistiques d'ordre deux ou d'ordre supérieur [7].

Tout au long des années 80, l'ACI est essentiellement resté un domaine de recherche plutôt français, avec une influence limitée au niveau international, où jusqu'au milieu des années 90, il est resté un effort de recherche assez petit et étroit.

Le premier atelier international sur l'ACI a été organisé en 1989 par *J.F. Cardoso* et *P. Comon* dans un article consacré à l'analyse spectrale d'ordre supérieur (En anglais: *higher-order spectral analysis*). *Cardoso* a introduit des méthodes algébriques exploitant la structure du tenseur de cumulants d'ordre quatre qui ont conduit au célèbre algorithme JADE (En anglais: *Joint Approximate Diagonalization of Eigenmatrices*) tout en s'inspirant des travaux de *C. Jutten*. L'utilisation de cumulants de quatrième ordre a déjà été proposée par *J.-L. Lacoume* en 1988 [3, 7,8].

En 1994, les travaux de *P. Comon* ont permis de formaliser le lien entre le problème de séparation aveugle de sources et l'ACI dans le cadre du mélange linéaire sans bruit additif [3]. L'ACI est en vérité une amélioration de la technique précédente l'Analyse en composantes principales ACP (En anglais: *Principal Component Analysis PCA*) basé sur la dé-corrélation (non corrélation). Par contre, l'ACI a changé complètement son intérêt où elle a remplacé la dé-corrélation des observations par une hypothèse plus forte qui est l'indépendance statique des sources.

L'ACI a attiré une attention plus large et un intérêt croissant après *A.J. Bell* et *T.J. Sejnowski* ont publié leur approche basée sur le principe infomax au milieu des années 90. Quelques années plus tard, *Hyvärinen* a présenté l'algorithme à virgule fixe ou FastICA, qui a contribué à l'application de l'ACI à des problèmes à grande échelle en raison de son efficacité de calcul [3].

Depuis lors, beaucoup de chercheurs dans le monde entier se sont intéressés vu la diversité de ses domaines d'application où une vague croissante d'articles, d'ateliers et de sessions spéciales consacrées à l'ACI. Une conférence internationale portant spécifiquement sur le sujet existe depuis 1999. Devant avoir initialement lieu tous les 18 mois, elle est désormais annuelle [8].

1.3 Définition d'ACI :

L'Analyse en Composantes Indépendantes ACI est une méthode, peut-être la plus largement utilisée, pour effectuer une séparation aveugle des sources. Pierre Comon a défini l'ACI dans son article (Comon 1994) comme suite :

« *L'analyse en composantes indépendantes (ACI) d'un vecteur aléatoire consiste en la recherche d'une transformation linéaire qui minimise la dépendance statistique en ses composantes* » [11].

Le but de cette séparation avec ACI est d'extraire des signaux utiles s_n à partir juste d'observer un mélange x_n de ces signaux sources.

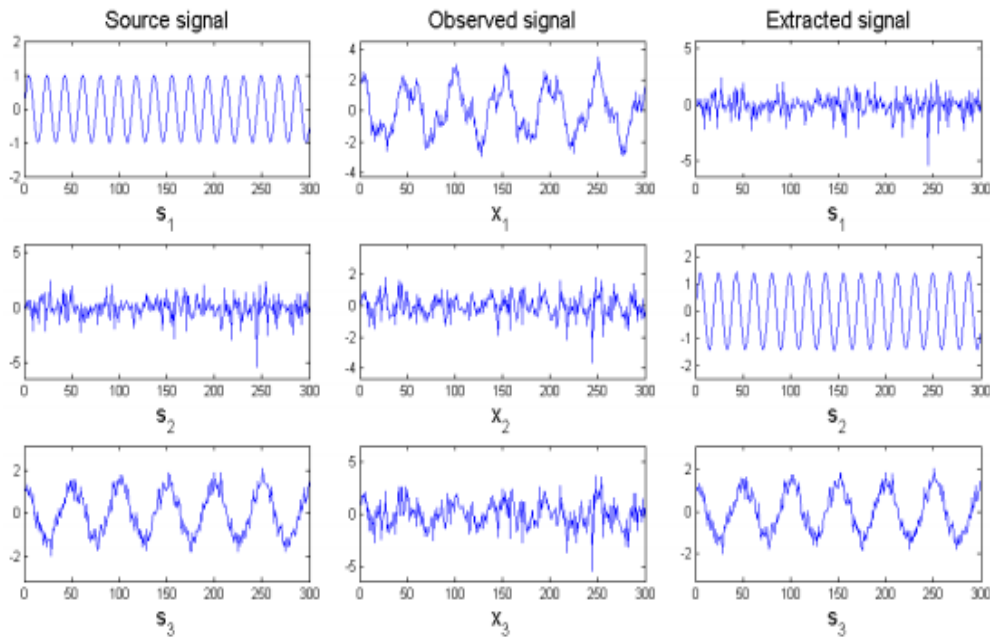


Figure 1.1 : La récupération des signaux sources s_n à partir de leurs mélanges en utilisant ACI [11].

Dans la figure 1.1, les signaux de sources s_1, s_2, s_3 qui sont inconnus et que l'on mélange de façon inconnue A sont estimés à partir des mélanges x_1, x_2, x_3 qui sont en revanche connus. Le problème de la séparation se fait à l'aveuglette, c'est-à-dire que l'on veut estimer deux inconnus (signaux s_n et matrice de mélange A) à partir d'un seul connu x_n . L'ACI a pu résoudre le problème à travers des informations statistiques des sources qui est basé sur l'indépendance.

- **Illustration statique :**

Considérons que deux composants indépendants ont les distributions uniformes suivantes :

$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{si } |s_i| \leq \sqrt{3} \\ 0 & \text{ailleurs} \end{cases} \quad (1.1)$$

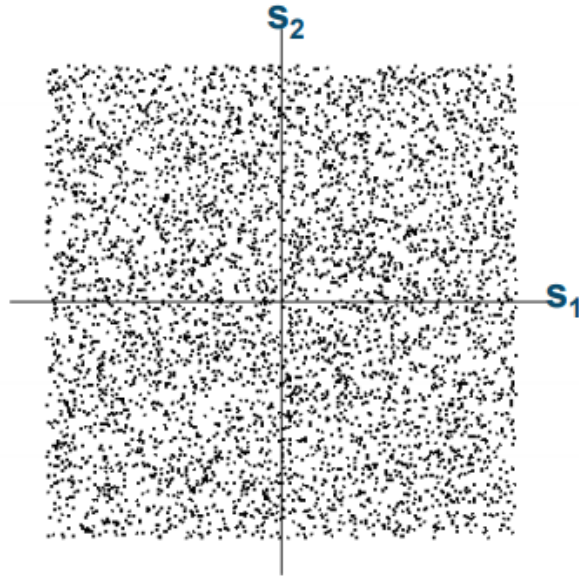


Figure 1.2 : La distribution conjointes de sources s_1 et s_2 indépendantes [9].

Supposons que les deux composants indépendants sont mélangés par une matrice A , un mélange x est générées à l'aide de modèle ACI. La distribution conjointe de ce mélange est comme suit :

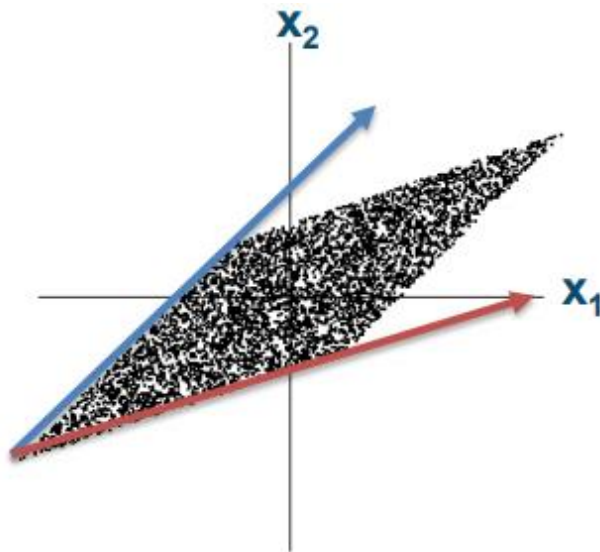


Figure 1.3 : La distribution conjointes de mélange x_1 et x_2 [9].

Les bords du parallélogramme dans la Figure 1.3 sont les directions des colonnes de la matrice A . Cela signifie que le problème semble avoir une solution [5].

On peut donner la définir de base d'ACI comme suite [7, 8]:

- A l'aide de plusieurs capteurs, on observe un mélange $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$ de n sources indépendantes $\mathbf{s}_n = (s_1, s_2, \dots, s_n)$. Nous supposons que chaque mélange \mathbf{x}_n ainsi que chaque composant indépendant s_n est un vecteur aléatoire :

$$\mathbf{x}_n = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{1.2}$$

$$\mathbf{s}_n = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} \tag{1.3}$$

- Le mélange a lieu à travers une matrice inconnue on la note A où : $A = \{a_{ij}\}$. la matrice A doit être égale au nombre de sources.

$$A = \{a_{ij}\} = \begin{bmatrix} a_{11} & \dots & a_{1j} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{ij} \end{bmatrix} \tag{1.4}$$

En utilisant cette notation à matrice vectorielle, le modèle de mélange linéaire instantané non bruité est écrit comme suit [8] :

$$\left. \begin{array}{l} x_1 = a_{11} \cdot s_1 + a_{12} \cdot s_2 + \dots + a_{1j} \cdot s_n \\ x_2 = a_{21} \cdot s_1 + a_{22} \cdot s_2 + \dots + a_{2j} \cdot s_n \\ \dots \\ x_n = a_{i1} \cdot s_1 + a_{i2} \cdot s_2 + \dots + a_{ij} \cdot s_n \end{array} \right\} \mathbf{x} = \mathbf{A} \cdot \mathbf{s} \tag{1.5}$$

La définition donnée ici est la plus fondamentale. Pour simplifier, nous omettons tout terme de bruit dans le modèle de base, car l'estimation du modèle sans bruit est déjà assez difficile. Cependant, on peut rencontrer certaines généralisations et modifications de la définition comme par exemple:

- Dans de nombreuses applications, il serait plus réaliste de supposer qu'il y a du bruit dans les mesures, ce qui signifierait l'ajout d'un terme de bruit dans le modèle (modèle linéaire instantané bruité) : $\mathbf{x} = \mathbf{A} \cdot \mathbf{s} + \boldsymbol{\sigma}$, d'où $\boldsymbol{\sigma}$ est le bruit.
- Dans de nombreux cas, le nombre de composantes indépendantes et de mélanges observés peut ne pas être égal.
- Dans un autre cas, les observateurs résultent d'une transformation non linéaire des sources (mélange non linéaire) : $\mathbf{x} = \mathbf{F}(\mathbf{s})$, d'où $\mathbf{F}(\mathbf{s})$ est une fonction non linéaire quelconque [4]

1.4 L'indépendance statique:

L'hypothèse de l'indépendance statique (En anglais : *Statistical Independence IS*) des sources est un concept clé dans l'analyse indépendante des composants. L'hypothèse dit que si les sources se produisent de différents processus physiques donc les sources ont rien avoir l'un de l'autres c'est-à-dire qu'ils sont indépendantes.

D'une autre manière, nous considérons deux variables scalaires différentes X et Y , On dit que X est indépendant de Y si la connaissance de la valeur de X ne donne aucune information sur la valeur de Y et vice versa. Prenons un exemple réel tel que le jeu de pile ou face et jeté d'un dé, d'où toute information qui nous apparaît n'indique pas à une autre information [10].

Il faut tenir en compte que les signaux sources s_n sont indépendants, mais les mélanges (les signaux observés) x_n ne sont pas, car chacun de ces mélanges contient un peu des signaux sources donc chacun est liés avec l'autre d'où il est possible de connaître la valeur de l'un d'eux à partir de l'autre [9].

Mathématiquement, l'indépendance statique est introduite par le concept de fonction de densité de probabilité (pdf) d'un événement. En statistiques, lorsque deux événements A et B sont étudiés, la probabilité conditionnelle $P(B/A)$, est définie comme la probabilité que l'événement B se produise étant donné que l'événement A se produit, et la probabilité conjointe, $P(A, B)$ est définie comme la probabilité qu' A et B se produisent.

Ils sont liés par la règle (loi de Bayes) :

$$P(B/A) = \frac{P(A,B)}{P(A)}. \quad (1.6)$$

D'où $P(A)$ est la densité marginale de A . Si les deux événements sont statistiquement indépendants $P(B/A) = P(B)$. Donc si A et B sont indépendante alors :

$$P(B/A) = P(B) = \frac{P(A,B)}{P(A)} \Rightarrow P(A, B) = P(A) P(B) \quad (1.7)$$

Pour résumé, on peut dire que pour deux variables aléatoires X et Y sont indépendants si et seulement si la densité conjointe $P(X, Y)$ est factorisé à des densités marginales de $P(X)$ et $P(Y)$:

$$P(X, Y) = P(X)P(Y). \quad (1.8)$$

La probabilité marginale de X et Y est donnée par :

$$P(X) = \int P(X, Y) dY \quad \text{et} \quad P(Y) = \int P(X, Y) dX. \quad (1.9)$$

Dans le cas de n variables aléatoires, cette définition s'étend naturellement où la densité conjointe doit être un produit de n termes.

Étant donné deux fonctions intégrables h_1 et h_2 de X et Y , les variables aléatoires indépendantes satisfont la propriété de base :

$$E \{h_1(X) h_2(Y)\} = E \{h_1(X)\} E \{h_2(Y)\} \quad (1.10)$$

• **La non-corrélation et l'indépendance :**

Il faut bien savoir que la non-corrélation ne signifie pas l'indépendance, mais le contraire est vrai. On dit que deux variables aléatoires X et Y sont non-corrélé si leur covariance est nulle :

$$\begin{aligned} C(X, Y) &= E \{(X - E\{X\})(Y - E\{Y\})\} \\ &= E\{XY\} - E\{X\}E\{Y\} = 0 \end{aligned} \quad (1.11)$$

On remarque alors que si les variables sont indépendantes ils sont aussi non-corrélé :

$$E\{XY\} = E\{X\}E\{Y\} \Rightarrow E\{XY\} - E\{X\}E\{Y\} = 0 \quad (1.12)$$

Par contre la non-corrélation n'implique pas l'indépendance comme par exemple, si on prend $(X, Y) = (1, 0)$ et ont une distribution de probabilité égal $1/4$, on trouve que :

$$E\{X^2 Y^2\} = 0 \neq 1/4 = E\{X^2\} E\{Y^2\}. \quad (1.13)$$

1.5 Théorème de centre limite:

Théorème de centre limite (en anglais center limite théorie) c'est un théorème fourni par A. Liapunov en 1901, ce théorème joue un rôle intéressant dans l'analyse de composant indépendante où on nous a permis de savoir si le signal reçu est un signal constitué de plusieurs signaux indépendantes.

Ce théorème dit que :

-La distribution d'une somme de variables aléatoires indépendantes tend vers une distribution gaussienne. Ainsi, une somme de deux variables aléatoires indépendantes a généralement une distribution plus proche de la gaussienne que n'importe laquelle des deux variables aléatoires originales [9].

On expriment ce théorème mathématiquement on supposant Z_1, Z_2, \dots, Z_i une série de variables aléatoires réelles définies sur le même espace de probabilité, indépendantes et identiquement. La somme de cette série est donnée par :

$$X_k = \sum_i^n Z_i \quad (1.14)$$

Parce que la moyenne et la variance de X_k peuvent augmenter sans être lié par $k \rightarrow \infty$, considérons le standard réalisable y_k au lieu de X_k .

$$y_k = \frac{X_k - m_x}{\sigma_k} \quad (1.15)$$

Où m_x et σ_k sont la moyenne et la variance de X_k . La distribution de y_k converge vers une distribution gaussienne avec une variance moyenne et unitaire nulle lorsque $k \rightarrow \infty$.

Ce résultat est appelé le théorème de la limite centrale. Il existe plusieurs formes différentes de théorèmes dans lesquels les hypothèses sur l'indépendance et la même distribution sont affaiblies.

Le théorème de la limite centrale est la principale raison de la modélisation de nombreux phénomènes aléatoires en tant que variables aléatoires gaussiennes. Par exemple, un bruit additif peut généralement être considéré comme la somme d'un grand nombre de petits effets de base, il peut donc naturellement être modélisé comme une variable aléatoire gaussienne.

1.6 Prétraitement des données pour ACI:

Pour but de simplifier et améliorer les algorithmes ACI, il est très utile de faire un prétraitement des données (les mélanges x_i de la matrice X) qui consiste à les centrer et les blanchir. Ces deux derniers permettent de normaliser les moments d'ordre 1 et d'ordre 2 à la matrice identité, puis traiter des moments d'ordre supérieur qui rendent l'estimation du problème de l'ACI plus simple et mieux conditionnée.

1.6.1 Centrage :

Le prétraitement le plus élémentaire et nécessaire est de centrer les données en soustrayant la moyenne de tous les signaux $\mu = E\{x\}$ de façon à rendre la variable x de moyenne nulle, le centrage peut être calculé comme suit:

$$x' = x - E\{x\} \quad (1.16)$$

Le but de centrage est pour simplifier les algorithmes ACI, ce que ne veut pas dire que la moyenne ne peut pas être estimée. Après avoir estimé la matrice de mélange A , le vecteur moyen de s peut être rajouté aux estimations centrées de s pour compléter estimation où la valeur de moyen de s est exprimé par :

$$S = A^{-1} \mu. \quad (1.17)$$

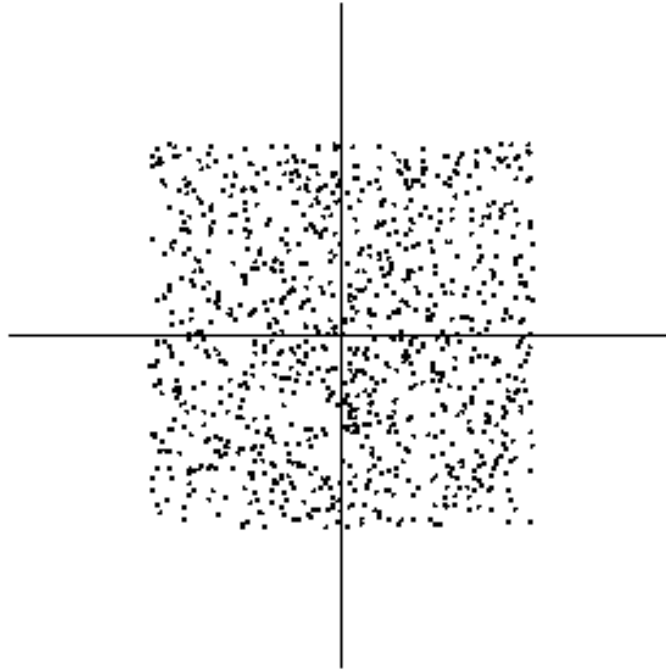


Figure 1.4 : illustration d'une distribution centrée [8].

1.6.2 Blanchiment :

Une deuxième étape très importante dans le prétraitement est le blanchiment qui consiste à dé-corréler les mélanges, c'est-à-dire qu'avant l'application de l'algorithme ICA nous transformons le vecteur X observé linéaire de sorte que nous obtenons un nouveau vecteur V qui est blanchi où la variance de chaque mélange blanchi est égale à 1 et leur matrice de covariance devient donc l'identité.

Pratiquement, pour passer de X à V , on introduit la matrice de blanchiment U , de taille $[n, m]$, telle que :

$$V = UX. \quad (1.18)$$

Cette matrice est obtenue par les étapes suivantes :

- 1- Calcul de la matrice de covariance C de X :

$$C = E[XX^T]. \quad (1.19)$$

- 2- Calcul de la matrice des valeurs propres D_C et des vecteurs propres E de C , telles que $CE = ED_C$ avec:

$$\begin{cases} C = ED_C E^{-1} \\ D_C = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \\ E = (E_{\lambda_1}, E_{\lambda_2}, \dots, E_{\lambda_n}) \end{cases} \quad (1.20)$$

3- Calcul de U par l'opération matricielle :

$$U = D_C^{1/2} E^T. \quad (1.21)$$

4- Calcul de V par :

$$V = U X. \quad (1.22)$$

L'intérêt de cette matrice des mélanges blanchis V est justement de passer d'un espace quelconque à un espace blanchi, défini principalement par une réduction de la dimensionnalité (de m à n), c'est-à-dire le nombre de mélanges passent de m (nombre initial de mélanges) à n (nombre de composantes indépendantes). Ce nouvel espace permet donc une simplification globale de l'ACI, où le blanchiment améliore la convergence des algorithmes, et réduit la dimensionnalité des données et le bruit. Donc, on peut dire que le blanchiment résout la moitié du problème d'ACI.

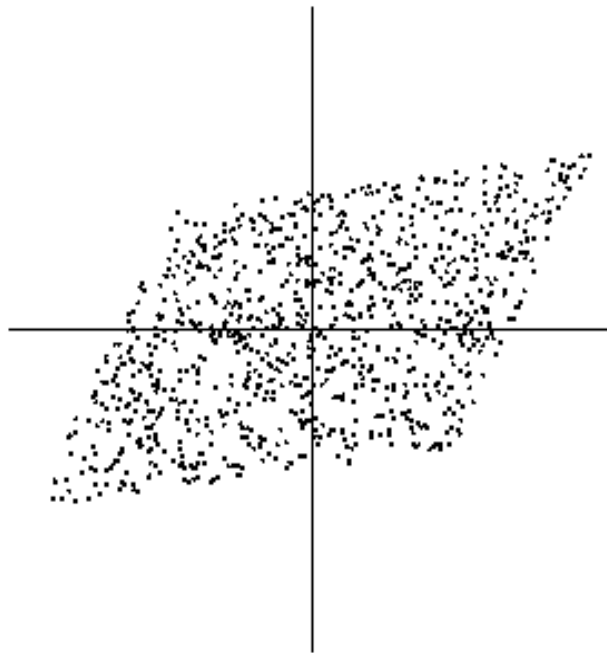


Figure 1.5 : illustration d'une distribution conjointe des mélanges avant le blanchiment [8].

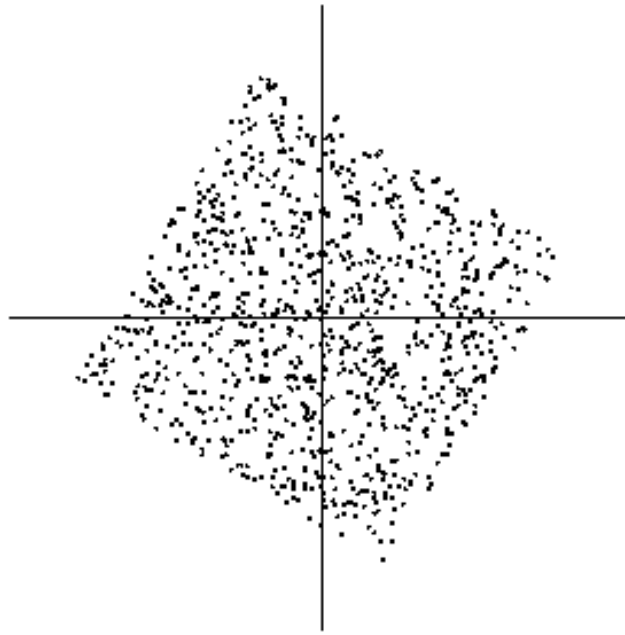


Figure 1.6 : illustration d'une distribution conjointe des mélanges blanchis [8].

1.7 Fonction de contraste :

Une fonction de contraste (appelée aussi contraste) qui sert à résoudre le problème de séparation aveugle des sources où est un critère avec son optimisation (maximisation) nous pouvons obtenir la matrice de démixtion W à partir de la non-gaussianité des sources (kurtosis, la négentropie) et la minimisation des informations mutuels. Elle est définie de la façon suivant :

Une fonction de contraste (ou contraste) est une fonction réelle des sorties du système global et qui satisfait les deux conditions suivantes :

- ✓ $C((y(n))_{n \in \mathbb{Z}}) \leq C((s(n))_{n \in \mathbb{Z}})$,
- ✓ Si $C((y(n))_{n \in \mathbb{Z}}) = C((s(n))_{n \in \mathbb{Z}})$, alors la séparation des sources est effective,

Où $(y(n))_{n \in \mathbb{Z}}$ est la sortie du système global, résultant d'un mélange par un filtre quelconque d'un tirage du vecteur $(s(n))_{n \in \mathbb{Z}}$ de sources indépendantes [12].

1.7.1 La mesure de la non-gaussianité:

Parmi les critères qui doivent être disponible pour séparer les sources dans l'ACI, une condition doit être remplie est la non-gaussianité des signaux sources. Dans le premier, nous allons voir pour quoi les variables gaussiennes sont interdites dans l'ACI.

Pour voir cela, nous supposons que la matrice du mélange est orthogonale et les signaux s_i sont gaussiens, et x_1 et x_2 sont gaussiens, non corrélés, de variance unitaire. Leur densité conjointe est donnée par :

$$p(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{x_1^2 + x_2^2}{2}} \quad (1.23)$$

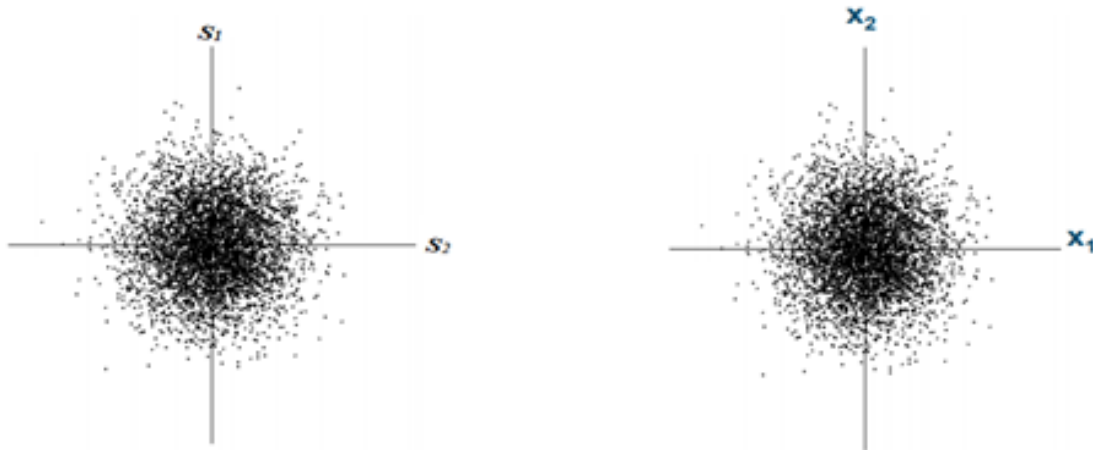


Figure 1.7 : la distribution conjointes de sources s_1 et s_2 et de mélanges x_1 et x_2 [9].

La figure 1.7 montre que la densité est complètement symétrique. Par conséquent, il ne contient aucune information sur les directions des colonnes de la matrice de mélange A . C'est pourquoi A ne peut pas être estimé.

Pour estimer une des composantes indépendantes, nous considérons une combinaison linéaire du x_i , notons ceci par y :

$$y = W^T x = W^T A \cdot s \quad (1.24)$$

Où W^T est une matrice de séparation dans l'espace blanchi à déterminer, et c'est l'inverse de A . Cette combinaison linéaire serait en fait égale à l'un des composants indépendants, où la maximisation de la non-gaussianité de $W^T x$ nous donnera l'un des composants indépendants à partir de deux méthodes : *kurtosis* ou de *la néguentropie* [5].

1.7.1.1 Kurtosis :

En statistique, la dé-corrélation qui travaille avec des moments d'ordres 1 (moyenne) et d'ordre 2 (variance) n'est pas suffisante pour résoudre l'ACI et pour avoir plus d'information statistique et pouvoir ainsi récupérer des composantes, c'est pour cela ils sont passés à des moments d'ordres plus élevés pour récupérer des composantes, indépendantes et pas seulement dé-corrélées. On parle de "statistiques d'ordres plus élevés" (En anglais : *Higher-Order Statistics HOS*). Par conséquent, les cumulants ont été introduits et notamment le kurtosis, qui est le cumulants d'ordre 4.

On détermine le kurtosis d'une variable aléatoire y centrée (si elle ne l'est pas, lui soustraire sa moyenne au préalable) comme le cumulants d'ordre 4, $K_4(y)$:

$$Kurt(y) = K_4(y) = E[y^4] - 3(E[y^2])^2 \quad (1.25)$$

Puisque y est une variance unitaire, on peut simplifier l'équation comme suit :

$$Kurt(y) = E[y^4] - 3 \quad (1.26)$$

Par définition, le kurtosis d'une variable gaussienne est nul et pour d'autres variables aléatoires est non nul. Ainsi, le kurtosis apparaît naturellement comme une fonction de contraste possible, car discriminatoire par rapport à la gaussianité des variables.

Le kurtosis peut être positif où il est appelé super-gaussien ayant une distribution de probabilité pointue ou il peut être négatif où il est appelé sub-gaussien ayant une distribution de probabilité allongée.

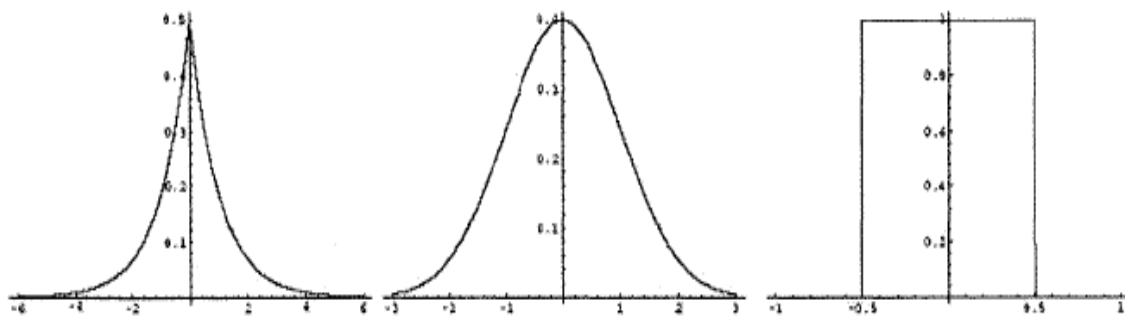


Figure 1.8 : Super-gaussienne, gaussienne et sub-gaussienne [11].

Ainsi, le but est de maximiser un kurtosis positif ou de minimiser un kurtosis négatif. Pour ce faire, on recherche itérativement les directions W , telles que $y = W^T x$.

Pour trouver l'inverse de la matrice de mélange afin d'identifier les sources, on procède source par source. On cherche un vecteur W tel que : $y = W^T x$ ait le kurtosis le plus différent de 0 possible :

- On part d'un vecteur W pris au hasard.
- On calcule la direction où le kurtosis croît le plus vite s'il est positif ou décroît le plus vite s'il est négatif, par des techniques de type «gradient», on peut trouver un optimum local. On a alors identifié une source.
- Pour chaque source, en fait les mêmes étapes précédentes.

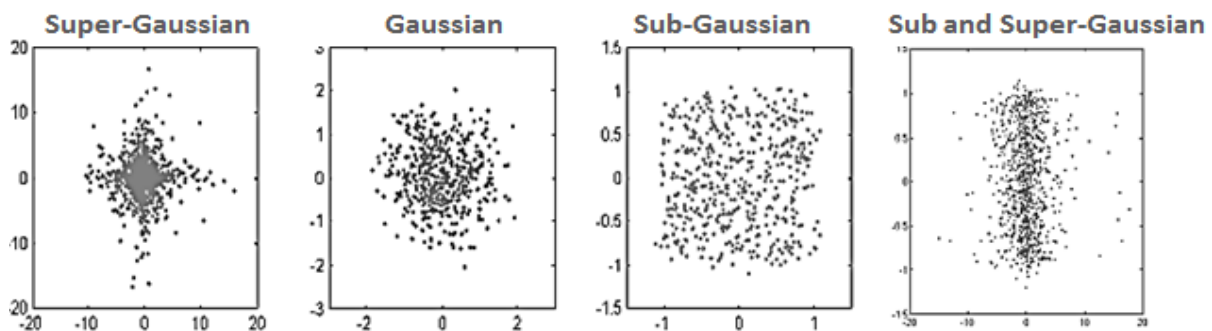


Figure 1.9 : Les distributions des variables obtenu par le calcul de kurtosis [9].

Le kurtosis présente un énorme désavantage : celui d'être très sensible aux valeurs aberrantes. En effet, la puissance 4 dans sa définition $E[y^4]$ entraîne que si l'un des mélanges ou l'un des coefficients de la matrice de séparation extrait est erroné, alors la puissance 4 amplifie cet effet. Ainsi, le kurtosis, bien que simple, apparaît insuffisamment efficace comme fonction de contraste ; on lui préfère la « négentropie », qui se veut une généralisation du kurtosis [7,11].

1.7.1.2 La négentropie

Dans l'analyse en composantes indépendantes (ACI), la théorie de l'estimation et la théorie de l'information offrent les deux principales approches théoriques pour caractériser les variables aléatoires : *l'entropie* qui est un concept de base de la théorie de l'information et une version légèrement modifiée de cette entropie, l'entropie normalisée négative ou *la négentropie*. Cette dernière est la deuxième méthode de la mesure de la non-gaussianité.

- **Rappel sur l'entropie**

L'entropie est une notion utilisée en théorie du codage où est une mesure du désordre c'est-à-dire plus la variable est « aléatoire », non structurée et imprévisible et plus l'entropie est grande.

En effet, la définition de l'entropie H d'une variable aléatoire discrète y de k valeurs possibles est :

$$H(y) = - \sum_{i=1}^k p_i \log p_i \quad (1.27)$$

Où $p_i = P(y = y_i)$ et y_i sont les valeurs possibles de y .

Dans le cas d'un vecteur Y de variables aléatoires de densité de probabilité $f(y)$, l'entropie devient l'entropie différentielle, définie par :

$$H(Y) = - \int f(y) \log f(y) dy = - E [\log f(y)]. \quad (1.28)$$

Le résultat essentiel de l'entropie est : une variable aléatoire gaussienne a l'entropie la plus grande parmi toutes les variables aléatoires possédant la même variance c'est-à-dire la distribution gaussienne présente l'entropie la plus élevée. L'entropie est petite pour les distributions qui sont concentrées sur certaines valeurs (un pdf qui est très pointu). Ceci montre que l'entropie peut être utilisée pour mesurer la non-gaussianité [7,11].

- **Définition la néguentropie :**

On définit la néguentropie J par la différence d'entropies entre une variable aléatoire gaussienne y_{gauss} et une variable quelconque y de même variance par l'équation suivante :

$$J(y) = H(y_{gauss}) - H(y). \quad (1.29)$$

L'intérêt de cette définition est que $J(y) = 0$ pour une variable gaussienne et $J(y)$ est strictement positif dans tous les autres cas. Alors le but est donc de maximiser $J(y)$.

Dans le cas d'un vecteur aléatoire Y , la néguentropie s'exprime sous même forme précédente. Où ce ne sont plus les variances qui sont égales, mais les matrices de covariance de Y et Y_{gauss} vecteur aléatoire gaussien.

L'avantage d'utiliser la néguentropie comme mesure de la non-gaussianité est qu'elle est bien justifiée par la théorie statistique. Mais le problème de l'utilisation de la néguentropie est cependant qu'elle est très difficile à calculer, où son estimation nécessite une estimation de pdf. Par conséquent, des approximations plus simples de la néguentropie sont très utiles [5,11]

- **Approximation de la néguentropie :**

L'expression de la néguentropie est très satisfaisante théoriquement parlant, par rapport aux statistiques. Cependant, elle est difficile à mettre en œuvre. En pratique, des approximations ont été utilisées pour parer ce problème.

La méthode classique de l'approximation de la néguentropie est l'utilisation des moments d'ordre supérieur (intervenir les cumulants d'ordre 3 et 4) :

$$J(y) \approx \frac{1}{12} k_3^2 + \frac{1}{48} k_4^2 = \frac{1}{12} E \{y^3\}^2 + \frac{1}{48} \text{kurt}(y)^2 \quad (1.30)$$

D'où y est un variable aléatoire centrée, de variance unitaire et moyenne nul. Cependant, cette approximation présente mêmes problèmes que kurtosis.

De nouvelles approximations ont ont été développées par *Hyvärinen* pour éviter les problèmes rencontrés avec les approximations précédentes. Ces approximations étaient basées sur le principe d'entropie maximale. Une nouvelle forme est introduite :

$$J(y) \approx \sum_{i=1}^p c [E \{G_i(y)\} - E \{G_i(v)\}]^2 \quad (1.31)$$

Où : c une constante positif et y et v sont deux variables respectivement quelconque et gaussienne, centrées et de variances unitaires et G_i est une fonction non quadratique et bien choisie :

- 1- $G_1(y) = y^4$.
- 2- $G_2(y) = 1/a_i \log(\cosh(ay))$ avec $1 \leq a \leq 2$.
- 3- $G_3(y) = -\exp(\frac{y^2}{2})$.

Pour le choix numéro (1), on tombe sur une approximation basée sur le kurtosis (l'estimateur classique). Pour les deux dernier choix (2) et (3) sont prouvé très utile.

À l'utilisation d'une seule fonction non quadratique G , l'approximation devient :

$$J(y) \approx c [E \{G(y)\} - E \{G(v)\}]^2 \quad (1.32)$$

Ces approximations de néguentropie qui sont simples, rapides pour calculer, nous donnent un très bon compromis entre les propriétés des deux mesures classiques de la non-gaussianité qui est données par kurtosis et néguentropie [5,7, 11].

1.7.2 Minimisation des informations mutuelles:

La théorie de l'information nous donne une autre méthode qui nous permettons l'utiliser comme critère pour trouver une estimation du modèle ACI, c'est la minimisation des informations mutuelles. L'une des principales utilités de l'information mutuelle est qu'elle sert à combiner entre de nombreux principes d'estimation, en particulier l'estimation du maximum de vraisemblance (En anglais : Maximum Likelihood ML) et la maximisation de la non-gaussianité.

1.7.2.1 L'information mutuelle :

Dans la théorie de l'information, l'information mutuelle est une mesure naturelle de la dépendance des variables aléatoires. On définit l'information mutuelle I entre N variables aléatoires y_i avec $i = 1 \dots N$, on utilisant le concept de la dérivée de l'entropie comme suit :

$$I(y_1, y_2, \dots, y_N) = \sum_{i=1}^N H(y_i) - H(y). \quad (1.33)$$

L'information mutuelle est toujours non-négative et nul si et seulement si les variables sont statistiquement dépendantes.

Une propriété importante de l'information mutuelle est que nous avons pour une transformation linéaire inversible $y = Wx$:

$$I(y_1, y_2, \dots, y_N) = \sum_{i=1}^N H(y_i) - H(x) - \log|\det W|. \quad (1.34)$$

Où la matrice W est déterminée de sorte que les informations mutuelles des composants y_i soient minimisées.

Cette information mutuelle est également exprimable en fonction de la néguentropie, avec les y_i de variances unitaires, sous la forme :

$$I(y_1, y_2, \dots, y_N) = C - \sum_{i=1}^N J(y_i) \quad (1.35)$$

À partir de l'équation (1.35), nous voyons que trouver une transformation linéaire inversible W qui minimise l'information mutuelle équivaut à peu près à trouver des directions dans lesquelles la néguentropie est maximisée. Nous avons vu précédemment que la néguentropie est une mesure de la non-gaussianité. Ainsi, (1.35) montre que l'estimation ACI par minimisation des informations mutuelles équivaut à maximiser la somme des non-gaussianités des estimations des composantes indépendantes, lorsque les estimations sont contraintes de ne pas être corrélées.

En pratique, il existe également des différences importantes entre ces deux critères (la néguentropie et minimiser l'information mutuelle).

La néguentropie et d'autres mesures de la non-gaussianité permettent l'estimation déflationniste, c'est-à-dire l'estimation une par une, des composantes indépendantes, car nous pouvons rechercher les maxima de la non-gaussianité d'une seule projection $W^T x$. Cette caractéristique n'est pas possible avec des informations mutuelles ou la plupart des autres critères, comme maximum de vraisemblance [8].

1.7.2.2 Maximum de vraisemblance :

Une autre approche très populaire pour l'estimation d'un modèle ACI est l'estimation du maximum de vraisemblance (En anglais : *Maximum of likelihood ML*). Estimer un paramètre par le maximum de vraisemblance consiste à proposer la valeur pour laquelle la probabilité de l'observation dans le modèle est la plus forte. Maximum de vraisemblance est désignée traditionnellement par la lettre "L" en raison de l'appellation anglo-saxonne de la vraisemblance, qui est "Likelihood".

En notant $W = (w_1, \dots, w_n)^T$ la matrice A^{-1} , le log de la vraisemblance est calculé comme suite :

$$L = \sum_{t=1}^T \sum_{i=1}^n \log f_i(w_i^T x(t)) + T \log |\det W| \quad (1.36)$$

Où :

- f_i représentent les fonctions de densité de s_i (ici supposés connus) et $x(t)$,
- $t = 1, \dots, T$ sont les réalisations de x ,
- le terme $\log |\det W|$ dans la vraisemblance vient des règles classiques pour la transformation linéaire ($y = Wx$) des variables aléatoires et leurs densités [5].

1.8 Les algorithmes d'ACI :

ACI possède de nombreux algorithmes, c'est-à-dire qu'il n'existe pas une seule méthode pour résoudre le problème d'ACI mais il existe différents méthodes, où le bon choix de la méthode ACI forme donc un tout. Comme l'affirmé Cardos :

«Il n'y a pas d'algorithme ACI universel».

Le but principal de ces algorithmes est d'extraire des composants indépendants. De nombreuses études traitant la technique ACI comme une boîte noire sans comprendre les détails internes.

L'algorithme ACI choisie est l'association de deux éléments qui sont importantes : la *fonction de contraste* (maximisant la non-gaussianité, minimisant les informations mutuelles, ou utilisant méthode d'estimation du maximum de vraisemblance) et *l'algorithme d'optimisation*.

1.8.1 Algorithme JADE :

JADE est l'abréviation en anglais de *Joint Approximation Diagonalization of Eigenmatrices*, c'est un algorithme développé par Cardoso en 1993 pour extraire les sources indépendantes et non-gaussiennes d'un signal mixte. Cette algorithme est basée sur le calcul des tenseurs des cumulants d'ordre quatre.

Ces cumulants peuvent être représentés par un tenseur (un cube). Le tenseur est une généralisation de la notion d'une matrice, c'est-à-dire que la matrice est une forme simple d'un tenseur. Ainsi, un tenseur de cumulants est une généralisation d'une matrice de covariance.

Pour récupérer les composantes indépendantes (les signaux sources), on diagonalise le tenseur de cumulants, parce que si les signaux sont indépendants, les cumulants à l'extérieur de la diagonale sont égaux à zéro. Et si les signaux ne sont pas indépendants, les cumulants à l'extérieur de la diagonale ne sont pas nuls. Alors l'algorithme JADE calcule les cumulants du signal observé, puis diagonalise le tenseur des cumulants où cette diagonalisation change les signaux observés en signaux purs [13,14].

1.8.2 Infomax :

L'algorithme d'infomax est un principe d'estimation pour l'ACI basé sur la maximisation de l'entropie de sortie d'un réseau de neurones avec des sorties non-linéaires.

Supposons que x est l'entrée du réseau neuronal dont les sorties sont données par :

$$y_i = \varphi_i(w_i^T x) \quad (1.37)$$

Où φ_i sont des fonctions scalaires non linéaires et les w_i sont les vecteurs de poids des neurones. On veut alors maximiser l'entropie des sorties comme suit :

$$H(y) = H(\varphi_1(w_1^T x), \dots, \varphi_n(w_n^T x)) \quad (1.38)$$

Si les φ sont bien choisis, ce cadre permet également d'estimer le modèle ICA [8,11].

1.8.3 Algorithme Fast-ICA :

L'algorithme Fast-ICA inventé par Aapo Hyvärinen à l'Université de technologie d'Helsinki, est l'une des méthodes les plus utilisées pour résoudre les problèmes de BSS. Les résultats expérimentaux établissent qu'il surpasse la plupart des algorithmes ICA en vitesse de convergence.

Après l'étape de blanchiment, Fast-ICA cherche à maximiser la non-gaussianité des y_i on utilisant souvent la néguentropie, chacune de ses composantes tendra vers une estimation des sources indépendantes

Pour éviter que toutes les estimations convergent vers la même source, il est nécessaire d'imposer l'orthogonalité des y_i . Deux méthodes existent pour imposer cette orthogonalité : *déflation et symétrique*. FastI-CA est une approche de type déflation c'est à dire que les composantes du mélange sont extraites l'une après l'autre.

1.9 Conclusion :

L'ACI est une technique connue depuis des années et développée au fil du temps pour régler le problème de séparation aveugle des sources, dans laquelle les données aléatoires observées sont exprimées sous la forme d'une transformation linéaire de composants statistiquement indépendants les uns des autres.

À côté de l'indépendance des sources qui est l'une des principes pour l'estimation de l'ACI, une approche fondamentale de l'ACI est donnée par le principe de non-gaussianité. Les composantes indépendantes peuvent être trouvées en trouvant des directions dans lesquelles les données sont au maximum non-gaussiennes

Avant l'estimation, deux procédures qui sont le centrage et le blanchiment sont appliqué sur les données pour but de simplifier le problème, puis à l'aide de des fonctions de contraste qui est soit basée sur la recherche des maxima de non-gaussianité ou l'estimation du maximum de vraisemblance ou bien par minimisé les informations mutuelles on extraire nos signaux sources. Plusieurs algorithmes tell que Fast-ICA, Infomax, JADE sont basés sur l'utilisation de l'une de ces fonctions de contraste.

Dans le prochaine chapitre, nous ce concentrons sur l'algorithme de Fast-ICA qui est l'algorithme de base de nos travaux, où tout d'abord, nous expliquerons le choix de l'algorithme Fast-ICA, puis nous passerons à la présentation de cette algorithme et son principe de fonctionnement.

Chapitre 2

*A*lgorithme Fast-ICA

2.1 Introduction :

Le principale but de la technique ACI est d'estimer les composantes indépendantes d'un signal mixte, et pour cela il existe plusieurs méthodes d'estimation où le bon choix facilite et accélère le calcul et donne des résultats plus précis. Un choix de méthode optimale peut donc se résumer en deux questions :

- *Comment veut-on estimer les composantes indépendantes?*
- *Et avec quel type d'algorithme?*

Dans l'ACI, il existe deux façons de calculer les composantes indépendantes,

- la première est l'estimation une par une qui est le mode *déflation*,
- l'autre est l'estimation toutes d'un coup qui est le mode *symétrique*.

Pour les types d'algorithmes existents, il y a trois types :

- a. **Les algorithmes par « lots »** : leurs capacités actuelles s'avèrent limitées. Ils sont basés sur la minimisation ou maximisation de fonctions de contraste, ne faisant intervenir que les cumulants d'ordre plus élevés.
- b. **Les algorithmes adaptatifs** : sont basés sur des méthodes de gradient stochastiques. Ces algorithmes présentent le défaut de converger lentement vers la solution et également d'être sensibles aux valeurs aberrantes.
- c. **Les algorithmes du point fixe** : le nom (point fixe) vient du fait que la règle récursive. Ils sont basés sur le kurtosis ou la néguentropie.

L'algorithme Fast-ICA est l'un des algorithmes les plus populaires dans le domaine de l'analyse en composantes indépendantes ACI. Il est une forme de l'algorithme du point fixe [11].

Dans ce chapitre, nous concentrons sur l'algorithme Fast-ICA où dans les titres suivants, nous allons citer la raison de choisir l'algorithme Fast-ICA, puis nous passons à la présentation de l'algorithme avec son principe

2.2 Choix du Fast-ICA :

Une étude de comparaison est faite sur différents algorithmes pour but de mesurer les propriétés de précision et de vitesse de calcul : la précision, charge informatique (le nombre d'opérations qu'il nécessite) et erreur en fonction du nombre de sources. Les algorithmes sont comparés sur des données générées artificiellement, constituées de mélanges de signaux de base (sinusoïdes, dents de scie, ...).

Les proportions des mélanges sont connues, en d'autres termes, la matrice de mélange A est connue et on connaît initialement S . Il est alors facile de comparer les composantes indépendantes rendues par chaque algorithme avec les signaux de s_i initial.

La comparaison est faite sur les algorithmes suivants :

- Les algorithmes du point fixe : simple (**FP**), avec la méthode déflation (**FPsym**), avec symétrique (**FPsymth**).
- L'algorithme naturel du gradient **ACY** (abréviation de : *Amari Cichocki-Yang*)
- L'algorithme **WACY** (White ACY), qui est l'ACY mais dans le cas blanc
- L'algorithme **BS** (de *Bell et Sejnowski*).
- L'algorithme étendu de Bell et Sejnowski, **ExtBS** (algorithme de gradient basé sur le critère de l'infomax).
- L'algorithme qui réalise la séparation adaptative équivariante de sources indépendantes **EASI** (abréviation de : *Equivariant Adaptive Separation via Independence*).
- L'algorithme **NPCA-RLS** (abréviation de : *Nonlinear Principal Component Analysis by Recursive Least Squares*).

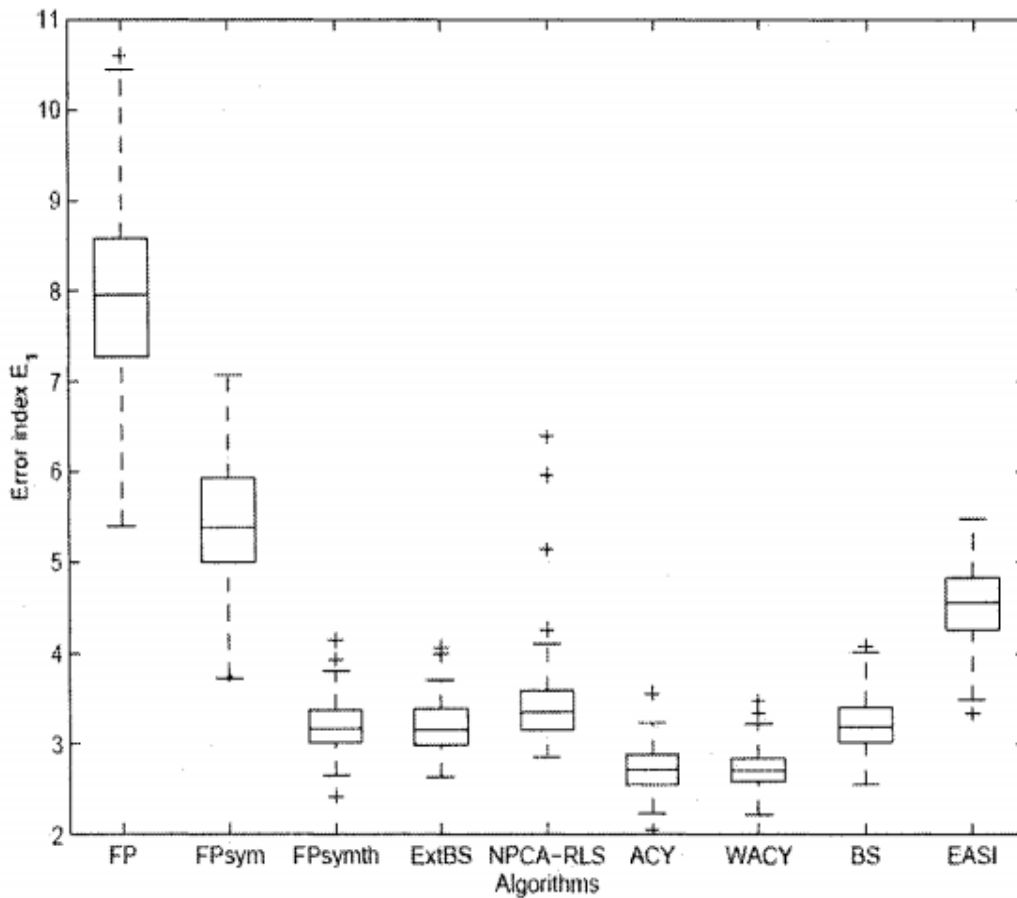


Figure 2.1 : Erreur E_1 en fonction de la précision [11].

La figure 2.1 montre les résultats pour la précision des algorithmes ci-dessous. Plus la valeur de E_1 est élevée plus l'algorithme est moins précis. Les intervalles pointés des boîtes représentent les valeurs possibles d' E_1 . Ainsi, les croix représentent les valeurs aberrantes.

On remarque d'après cette figure que l'algorithme du point fixe avec la méthode déflation **FPth** est dans le même ordre de performance de précision que les algorithmes **ExtBS** et **NPCA-RLS** et **BS**. On remarque aussi que tous les algorithmes FP ne présentent pas beaucoup de valeurs aberrantes malgré que l'algorithme à point fixe simple **FP** soit le plus mauvais dans la précision.

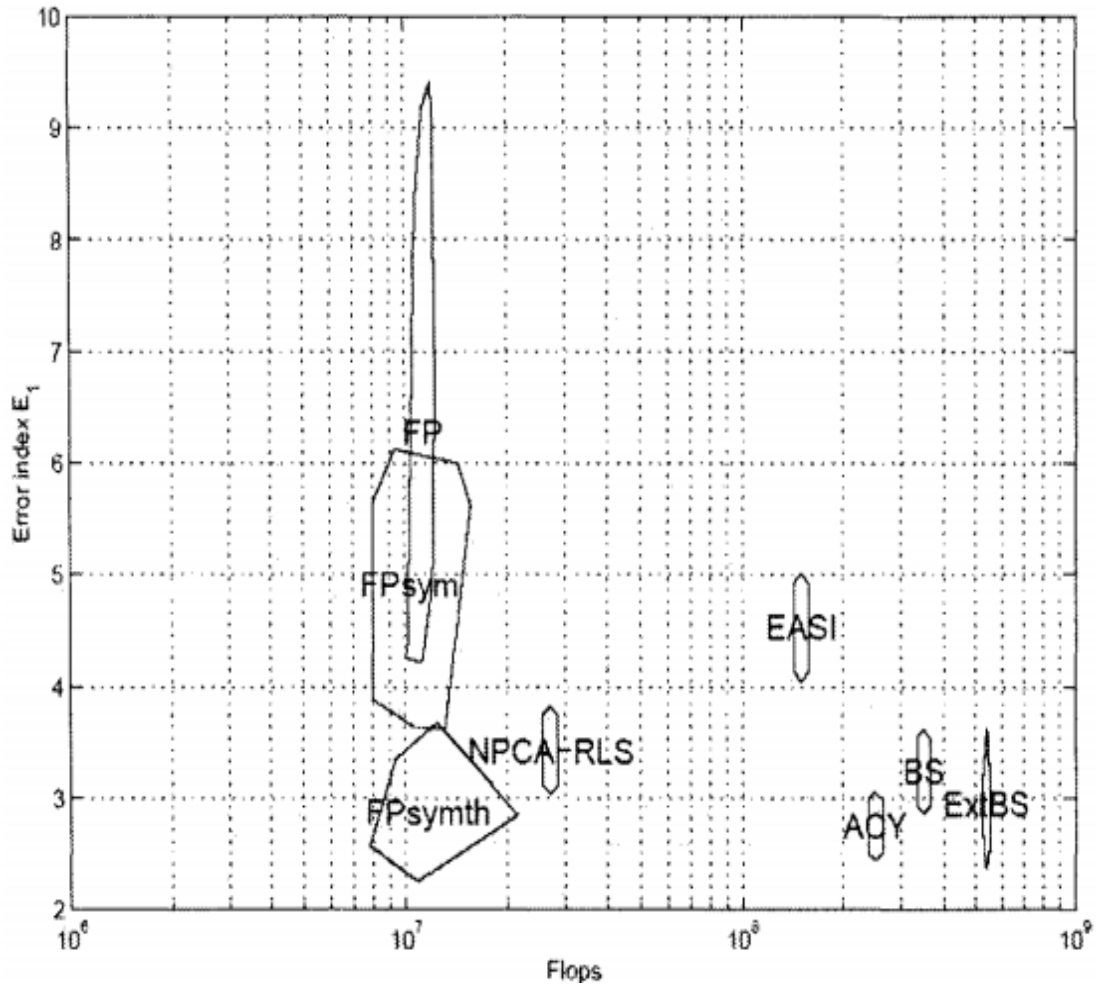


Figure 2.2 : Erreur E_1 en fonction de la charge informatique [11].

La figure 2.2 montre la mesure d'erreur E_1 et le nombre de calculs sur les flottants nécessaires à chaque algorithme pour converger vers les composantes indépendantes.

On remarque que les algorithmes à point fixe, quel qu'ils soient, convergent avec le moins de ressources informatiques nécessaires malgré qu'ils présentent un large intervalle des valeurs de E_1 . Pour les autres algorithmes **ACY**, **BS** et **ExtBS** offrent une erreur faible mais nécessitent plus d'opérations sur les flottants que les FP pour converger.

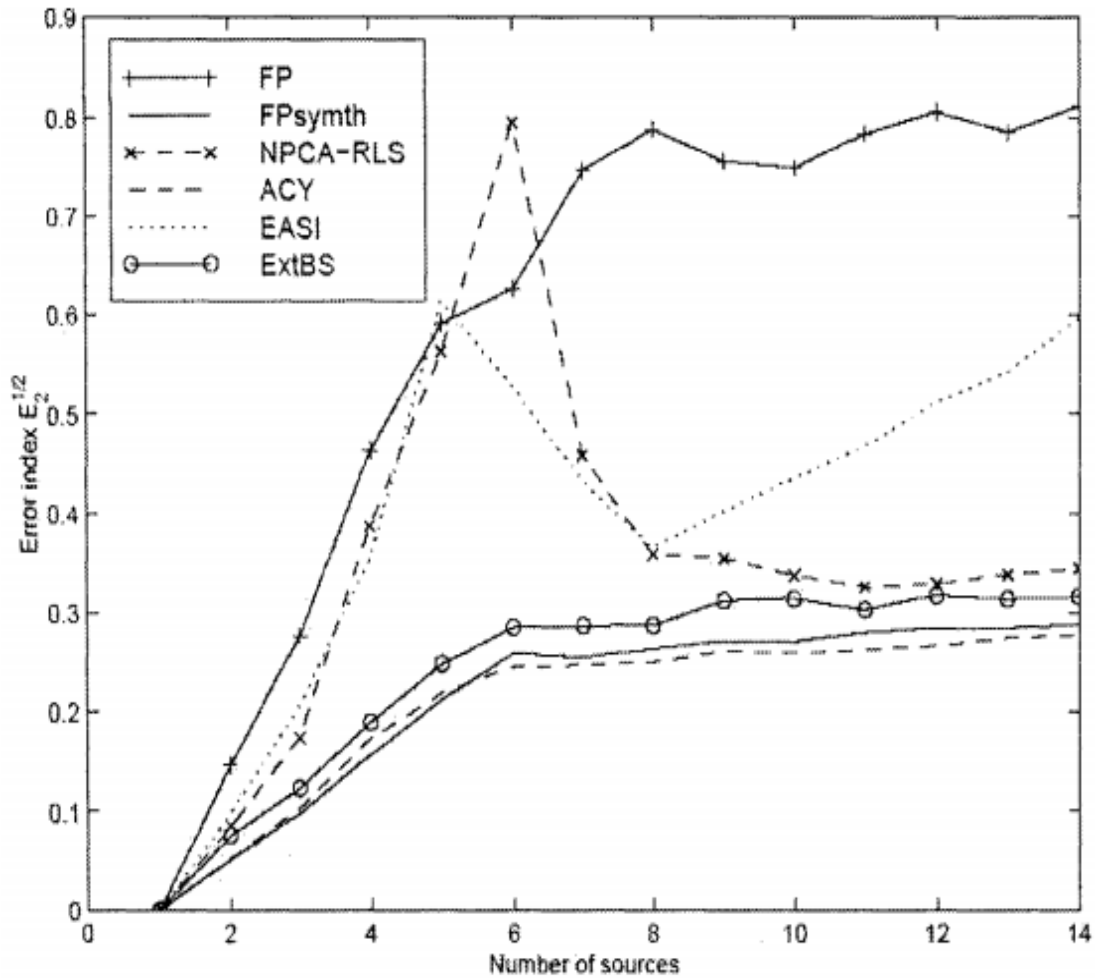


Figure 2.3 : Erreur en fonction de nombre de sources [11].

Dans la figure 2.3, l'erreur est calculée avec un autre indice E_2 qui est l'équivalent quadratique d' E_1 . On la mesure pour différentes valeurs du nombre de sources.

On constate globalement que les algorithmes **ACY**, **ExtBS** et **FPsymth** offrent le meilleur comportement et la meilleure précision avec l'augmentation de nombre de sources. Mais l'algorithme à point-fixe simple (FP) se comporte le plus mal. En outre, la précision de NPCA-RLS est proche des meilleurs, alors que celle de EASI se dégrade plus ou moins linéairement avec le nombre de sources.

- **Conclusion de l'étude de comparaison :**

D'après les trois études sur les différents algorithmes, on en arrive à la conclusion que, globalement, le meilleur algorithme est l'algorithme à point fixe. Son concepteur *Hyviriinen* lui a donné l'appellation « FastICA ».

Le Fast-ICA est un l'algorithme le plus rapide pour converger vers les composantes indépendantes, nécessitant le moins d'opérations et relativement peu sensible à l'augmentation du nombre de sources. Ces trois raisons majeures justifient donc notre choix à cet algorithme pour nos prochains travaux [11].

2.3 Fast-ICA :

Fast-ICA qui est connue sous le nom d'algorithme à virgule fixe, est une méthode très efficace pour maximiser les mesures de la non-gaussianité. Nous avons introduit différentes mesures de la non-gaussianité dans le chapitre précédent, tel que la néguentropie et kurtosis.

Fast-ICA une version qui peut également être décrite comme un réseau de neurones [9].

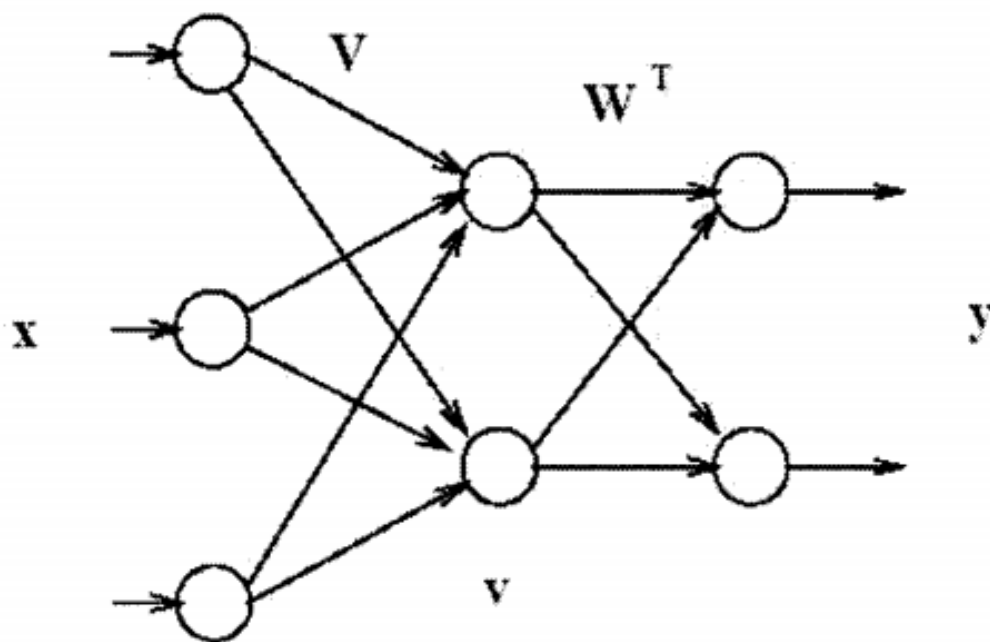


Figure 2.4 : Réseau de neurones de l' ACI [11].

On remarque que ce réseau possède trois couches de nœuds. La première est constituée des m nœuds qui représente les mélanges, c'est-à-dire la matrice X , la deuxième constituée des n mélanges blanchis V (où $n \leq m$), et enfin, la dernière des n composantes indépendantes Y [11].

On suppose ici que les données sont prétraitées par centrage et blanchiment. Regardons un seul neurone dans ce réseau :

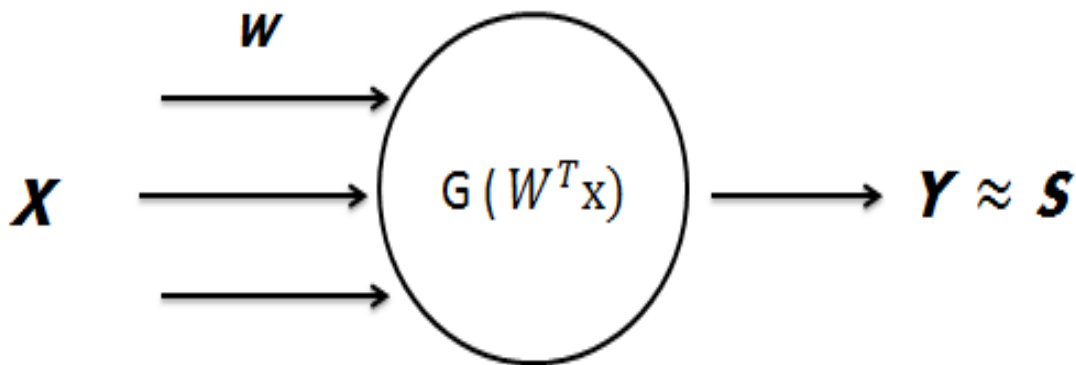


Figure 2.5 : Un seul neurone de l'ACI [9].

Nous allons commencer par un problème à une seule unité, puis généraliser à plusieurs unités.

2.3.1 Fast-ICA pour une seule unité :

Le Fast-ICA est basé sur un schéma d'itération à virgule fixe pour trouver un maximum de la non-gaussianité de $(w^T x)$. La non-gaussianité est ici mesurée par l'approximation de la négentropie $J(w^T x)$ puis kurtosis. La règle d'apprentissage Fast-ICA qui est à partir d'elle le neurone est capable de mettre à jour, trouve une direction, c'est-à-dire un vecteur unitaire w tel que la projection $(w^T x)$ maximise la non-gaussianité [5]. La méthode de base de l'algorithme Fast ICA est la suivante [9,11]:

- **Par de la négentropie :**

1. Choisissez un vecteur initial w_0 (par exemple, aléatoire).
2. Soit :

$$\mathbf{w}(k) = E\{xg(\mathbf{w}^T x)\} - E\{g'(\mathbf{w}^T x)\}\mathbf{w} \quad (2.1)$$

- Où g' est la dérivée de g , tel que soit :

$$g_1(y) = 1/a_i \log(\cosh(ay)) \text{ avec } 1 \leq a \leq 2 \quad , \quad g_2(y) = -\exp\left(\frac{y^2}{2}\right).$$

$$g'_1(y) = \tanh(a_i y). \quad , \quad g'_2(y) = y \exp\left(-\frac{y^2}{2}\right).$$

3. Diviser $w(k)$ par sa norme :

$$\mathbf{w} = \frac{\mathbf{w}(k)}{\|\mathbf{w}(k)\|} \quad (2.2)$$

4. S'il n'est pas convergé, retournez à l'étape 2. Sinon l'algorithme est convergent et génère $w(k)$.

Le vecteur final $w(k)$ qui est un vecteur de démixtion, sépare l'un des signaux de source non gaussiennes dans le sens que $w(k)^T x(t)$ où $t = 1, 2, \dots$ est égal à l'un des signaux source.

- **Par de kurtosis :**

1. Choisissez un vecteur initial w_0 (par exemple, aléatoire).
2. Soit :

$$\mathbf{w}(k) = E\{\mathbf{Z}[\mathbf{Z}^T \mathbf{w}(k-1)]^3\} - 3\mathbf{w}(k-1) \quad (2.3)$$

3. Diviser $w(k)$ par sa norme $\|\mathbf{w}(k)\|$.
4. S'il n'est pas convergé, retournez à l'étape 2. Sinon l'algorithme est convergent et génère $w(k)$.

2.3.2 Fast-ICA pour plusieurs unités:

L'algorithme Fast-ICA pour une seule unité, converge vers une seule colonne (vecteur) de la matrice de démixtion W . Pour trouver plus d'un vecteur pour estimer plusieurs composants indépendants, nous exécutons l'unité FastICA avec plusieurs unités w_1, w_2, \dots, w_n .

Pour empêcher que l'algorithme ne converge deux fois vers le même vecteur, il suffit d'ajouter une simple projection orthogonalisante à l'intérieur de la boucle, c'est à dire nous devons décorréliser les sorties $w_1^T x, w_2^T x, \dots, w_n^T x$ après chaque itération. Un moyen simple de réaliser la décorrélation est un schéma de déflation basé sur une décorrélation de type Gram-Schmidt : à chaque étape, le i -ème vecteur estimé doit être perpendiculaire aux $i-1$ vecteurs précédemment trouvés, car la matrice de démixtion doit être orthogonale

Dans certaines applications, une orthogonalisation symétrique peut être utile, on travaille directement à partir de W , qui à chaque itération, est orthogonalisée par Gram-Schmidt et actualisée. Les explications de ces deux méthodes sont données en 2.4 [5,15, 16].

2.4 Orthonormalisation de l'algorithme Fast-ICA :

Dans algorithmes ACI nous savons que les vecteurs de solution (vecteurs de base ACI) sont théoriquement orthogonaux ou orthonormaux, mais les algorithmes (comme point fixe) itératifs ne génèrent pas toujours automatiquement l'orthogonalité. Il peut alors être nécessaire d'orthogonaliser les vecteurs après chaque étape d'itération par une orthogonalisation de *Gram-Schmidt*.

Le procédé d'orthonormalisation de *Gram-Schmidt* est un algorithme permettant de fabriquer une famille orthonormée à partir d'une famille libre dans un espace euclidien.

On définit le théorème de *Gram-Schmidt* :

Si $(x_n)_{n \in N}$ est une famille libre d'un espace préhilbertien, il existe une et une seule famille orthonormée $(e_n)_{n \in N}$ telle que :

- $\text{Vect}(e_0, \dots, e_n) = \text{Vect}(x_0, \dots, x_n)$ pour tout n .
- Les produits scalaires $(e_n | x_n)$ sont strictement positifs pour tout n .

L'intérêt de cette ortho-normalisation est pour éviter que toutes les estimations convergent vers la même source, il est nécessaire d'imposer l'orthogonalité des y_i [8,11, 15].

L'orthonormalisation peut être accomplie essentiellement de deux manières :

- Méthode par déflation.
- Méthode par symétrique.

2.4.1 Méthode déflation :

La version dite « par déflation » estime itérativement les sources sont obtenues une par une et orthogonalise chaque estimation au moyen d'un procédé de *Gram-Schmidt* pour s'assurer de son orthogonalité avec eux est donnée par :

$$w_{i+1} = w_{i+1} - \sum_{j=1}^i (w_{i+1} w_j) w_j \quad (2.4)$$

$$w_{i+1} = -WW^T w_{i+1} \quad (2.5)$$

Où $W = (w_1, w_2, \dots, w_i)$ se construit à chaque nouveau vecteur calculé [11].

2.4.2 Méthode symétrique:

Une autre méthode peut être souhaitable c'est utiliser une dé-corrélation symétrique, dans laquelle aucun vecteur n'est "privilegié" par rapport aux autres; Cela signifie que les vecteurs w_i ne sont pas estimés un par un, Ils sont plutôt estimés en parallèle. L'une des raisons en est que la méthode de déflation a l'inconvénient que les erreurs d'estimation dans les premiers vecteurs sont ensuite accumulées par orthonormalisation. Une autre raison est que les méthodes d'orthonormalisation symétrique permettent le calcul parallèle de composants indépendants.

Où elle consiste à partir directement d'une matrice W quelconque, à l'orthogonaliser, selon *Gram-Schmidt*, par l'opération:

$$W = (WW^T)^{-\frac{1}{2}} W \quad (2.6)$$

Où W est la matrice avec des lignes W_i^T , Cela signifie que l'étape de mise à jour est effectuée en premier pour tous les n vecteurs des poids, alors la matrice W est orthogonalisée.

L'orthonormalisation symétrique est effectuée en effectuant d'abord l'étape itérative de l'algorithme à une unité pour chaque vecteur W_i en parallèle, puis orthogonaliser tous les W_i par des méthodes symétriques spéciales exprime dans l'étape suivant [11,15] :

1. Initialiser le W_i , $i=1,2,3,\dots,n$ (longueurs d'échantillon).
2. Faites une itération d'un algorithme d'une unité sur chaque W_i en parallèle.
3. Faire une orthogonalisation symétrique de la matrice W .
4. S'il ne converge pas, revenez à l'étape 3.

2.5 Propriétés du FastICA :

L'algorithme Fast-ICA présente un certain nombre de propriétés souhaitables par rapport aux méthodes existantes pour ICA, parmi les [15] :

1. La convergence est cubique (ou du moins quadratique), contrairement aux descentes de gradient qui sont linéaires. La conséquence est une convergence très rapide, d'où l'appellation « FastICA ».
2. L'algorithme Fast-ICA est facile à utiliser où il n'y pas de paramètre d'apprentissage à régler au préalable, contrairement aux algorithmes du gradient.
3. L'algorithme trouve directement les composantes indépendantes de n'importe quelle distribution non gaussienne.
4. Fast-ICA est moins sensible aux valeurs aberrantes, simple à calculer et nécessite peu d'espace mémoire
5. Les composantes peuvent être estimées une à une (déflation) ou
6. Simultanément (symétrique).

2.6 Conclusion :

L'algorithme FastICA qui utilise un schéma d'itération à virgule fixe est une méthode de calcul hautement efficace où elle a été trouvée dans plusieurs expériences, qu'elle est une méthode de maximisation très rapide de autres méthodes et fiable pour maximiser les mesures de non-gaussianité (soit en utilisant la kurtosis ou la négentropie) utilisées pour l'estimation ICA, en convergeant vers les composantes indépendantes.

Pour éviter que toutes les estimations convergent vers la même source, il est nécessaire d'imposer l'orthogonalité des y_i (sous contrainte de blanchiment, cela revient à dé-corréler les données)

Deux méthodes existent pour imposer cette orthogonalité (versions de Fast-ICA) : Fast-ICA monobloc (déflation) et Fast-ICA symétrique. Comme son nom l'indique, Fast-ICA à une unité correspond à la séparation à une unité, qui estime une ligne à la fois de la matrice de démixture W , tandis que Fast-ICA symétrique correspond à la séparation simultanée qui estime W dans son ensemble.

Chapitre 3

*M*ise œuvre de l'algorithme *Fast-ICA*

3.1 Introduction :

Le calcul complexe de Fast-ICA est l'une des principales barricades dans l'implémentation matérielle, où la vitesse de traitement est extrêmement souhaitable dans la mise en œuvre Fast ICA, et surtout dans les applications en temps réel telles que l'amélioration du signal vocal et l'extraction des fonctionnalités essentielles EEG/MEG et plusieurs d'autres.

Par conséquent, le processus ACI sera divisé en modules de sous traitement. Ces sous-modules sont développés indépendamment, puis intégrés ensemble.

Dans le chapitre précédant (chapitre 02), nous avons défini la méthode de base de l'algorithme Fast ICA, et dans ce chapitre, nous allons voir l'implémentation de l'algorithme Fast-ICA à partir des blocks de diagramme.

3.2 Mise en œuvre du centrage:

Le processus de centrage implique la soustraction de moyennes des signaux mixtes μ_1 et μ_2 de x_1 et x_2

Premièrement, les composantes de signaux mixtes sont accumulées les unes après les autres. Après avoir obtenu la somme de x , μ est obtenu en divisant la somme par la longueur de l'échantillon. Afin d'accélérer la vitesse de traitement, Utilisez la multiplication (multipliez par 1 / la longueur de l'échantillon) au lieu de la division.

Deuxièmement, la moyenne est soustraite des données de signaux mixtes pour réaliser le centrage. La formule d'opération est :

$$x(i) = x(i) - \left(\sum_{j=1}^{\text{la longueur de l'échantillon}} x(i) \right) \times (1 / \text{la longueur de l'échantillon}) \quad (3.1)$$

Où $i = 1, 2, \dots, \text{longueur d'échantillon}$ [15].

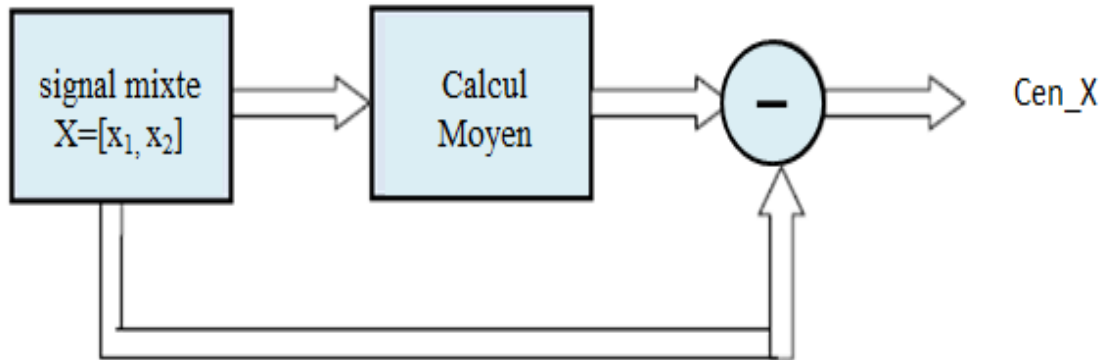


Figure3.1: Schéma fonctionnel de la mise en œuvre du centrage [15].

3.3 Mise en œuvre du blanchiment:

La première étape du blanchiment consiste à trouver la matrice de blanchiment P . P est donné par [15] :

$$P = D^{-1/2} \times E^T \quad (3.2)$$

Où $D = \text{diag}(\lambda_1, \lambda_2) =$ matrice diagonale des valeurs propres de la matrice de covariance C_x

$E = (e_1, e_2) =$ Matrice orthogonale des vecteurs propres de C_x .

$C_x = E\{XX^T\}$ est une matrice 2×2 .

Trois multiplicateurs sont nécessaires pour effectuer le calcul de XX^T :

Le multiplicateur-1 est utilisé pour $x_1 \times x_1$

Le multiplicateur-2 est utilisé pour $x_1 \times x_2$

Le multiplicateur-3 est utilisé pour $x_2 \times x_2$

$$\text{Où } X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1(1)x_1(2), \dots, x_1(\text{longueur de l'échantillon}) \\ x_2(1)x_2(2), \dots, x_2(\text{longueur de l'échantillon}) \end{pmatrix}$$

Puisque $x_2 \times x_1$ est égal à $x_2 \times x_1$, il suffit de réaliser l'un d'entre eux. Par conséquent, C_x peut être obtenu en multipliant en multipliant la somme par 1 / longueur d'échantillon.

Cette opération peut être exprimée comme :

$$C_{X_{00}} = \left(\sum_{j=1}^{\text{la longueur de l'échant}} x_1(j) \times x_1(j) \right) \times \left(\frac{1}{\text{longueur d'échant}} \right) \quad (3.3)$$

$$C_{X_{01}} = C_{X_{10}} = \left(\sum_{j=1}^{\text{longueur l'échant}} x_1(j) \times x_2(j) \right) \times \left(\frac{1}{\text{longueur d'échant}} \right) \quad (3.4)$$

$$C_{X_{11}} = \left(\sum_{j=1}^{\text{la longueur de l'échantillon}} x_2(j) \times x_2(j) \right) \times \left(\frac{1}{\text{longueur d'échant}} \right) \quad (3.5)$$

Après avoir calculé la matrice de covariance, l'étape suivante consiste à déterminer la matrice orthogonale des vecteurs propres $C_x(E)$ et la matrice diagonale des valeurs propres $C_x(D)$.

$$\sigma = \frac{C_{X_{22}} - C_{X_{11}}}{2 \times C_{X_{12}}} \quad (3.6)$$

$$T = \frac{\text{sign}(\sigma)}{\text{abs}(\sigma) + \sqrt{1 + (\sigma \times \sigma)}} \quad (3.7)$$

$$C = \frac{1}{\sqrt{1 + (T \times T)}} \quad (3.8)$$

$$S = T \times C \quad (3.9)$$

$$E = \begin{bmatrix} C & S \\ -S & C \end{bmatrix} \quad (3.10)$$

$$D = E^T \times [C_x \times E] \quad (3.11)$$

$$\text{Si } D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \text{ alors } D^{-1/2} = \begin{bmatrix} \lambda_1^{-1/2} & 0 \\ 0 & \lambda_2^{-1/2} \end{bmatrix} \quad (3.12)$$

$$\text{Où } \lambda_1^{-1/2} = \frac{1}{\sqrt{\lambda_1}} \text{ et } \lambda_2^{-1/2} = \frac{1}{\sqrt{\lambda_2}}$$

Par conséquent, la matrice de blanchiment P est obtenue en multipliant $D^{-1/2}$ par E^T :

$$P = D^{-1/2} \times E^T \quad (3.13)$$

Enfin, les données blanches Z sont obtenues après multiplication de P par X comme suivant :

$$Z = PX \quad (3.14)$$

Le schéma d'implémentation fonctionnelle est illustré dans la figure suivante :

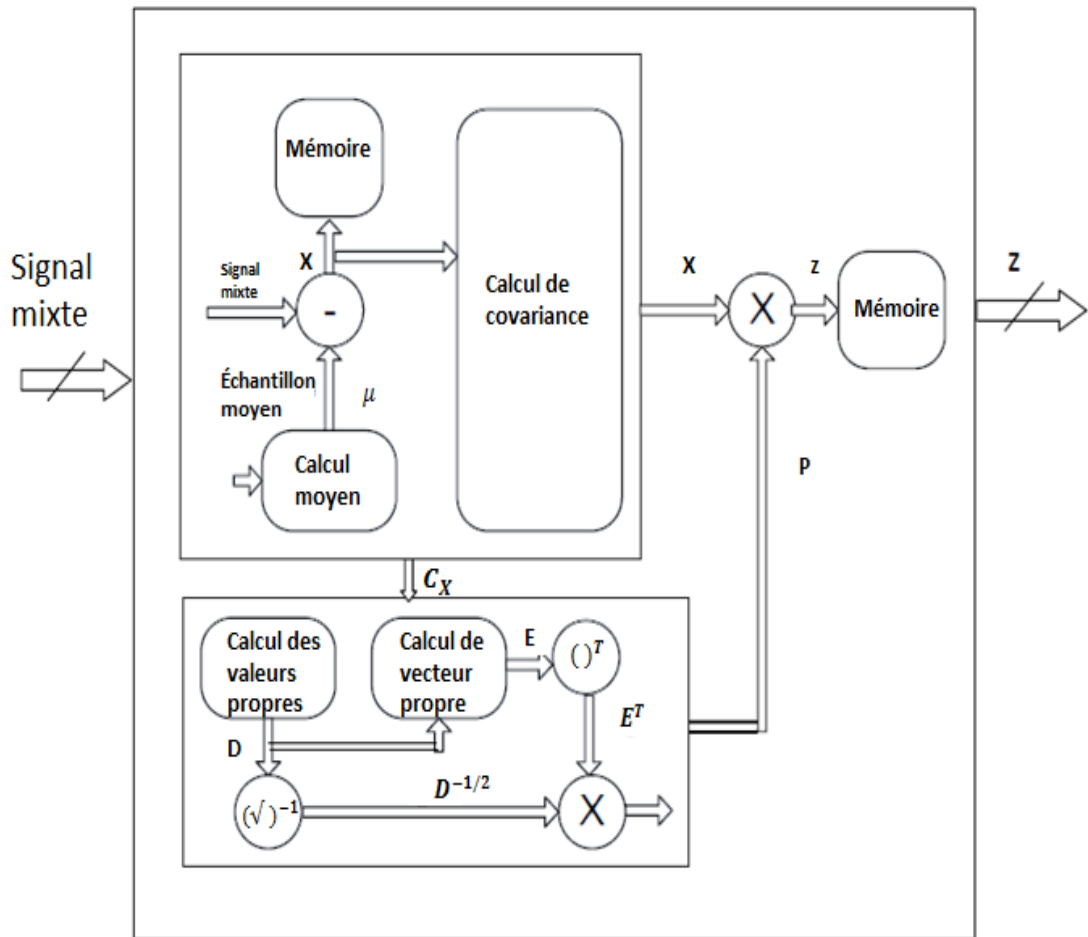


Figure 3.2 : Schéma fonctionnel de la mise en œuvre de blanchiment [15].

3.4 Calcul de norme :

Pour calculer la norme $\|w\|$, choisissons un vecteur aléatoire initial $w = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$ et calculons sa norme comme suite [15]:

$$\|w\| = \sqrt{(W_1 \times W_1) + (W_2 \times W_2)} \quad (3.15)$$

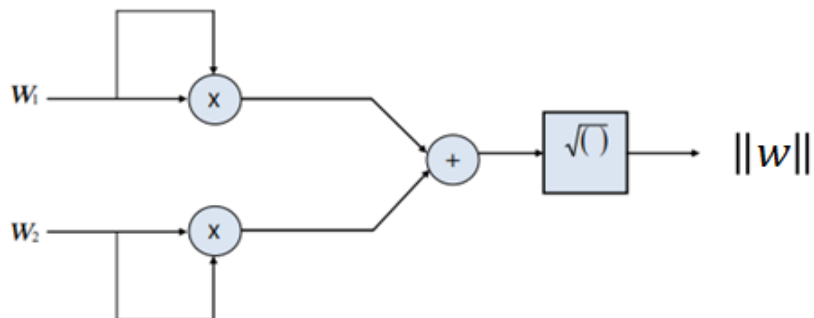


Figure 3.3 : Diagramme sur calculer de la norme [15].

Nous divisons W_1 et W_2 par sa norme, nous aurons :

$$w = \begin{bmatrix} W_1 / \|w\| \\ W_2 / \|w\| \end{bmatrix} \quad (3.16)$$

3.5 Mise en œuvre de la kurtosis :

Pour calculer le vecteur de séparation w , nous avons choisi de le calculer en utilisant la kurtosis, où l'équation est exprimée comme suit :

$$W(k) = E \{ Z [Z^T w(k-1)]^3 \} - 3w(k-1) \quad (3.17)$$

Nous séparons cette équation en deux pour appliquer les blocks l'implémentation de kurtosis :

$$W(k) = E \{ \underbrace{Z [Z^T w(k-1)]^3}_{(1)} \} - 3w(k-1) \quad (2)$$

Le calcul de (1) est d'abord mis en œuvre où l'opération peut être formulée comme suit :

$$Per_w(k) = \sum_{j=1} (Z(j) \times [Z(j) \times w(k-1)]^3) \quad (3.18)$$

Où $Per_w(k)$ est le résultat du calcul de $Z [Z^T w(k-1)]^3$.

$$Z [Z^T w(k-1)]^3 = Per_w(k) = (Per_w_1(k), Per_w_2(k)) \quad (3.19)$$

L'implémentation de (1) est présentée dans la figure suivante :

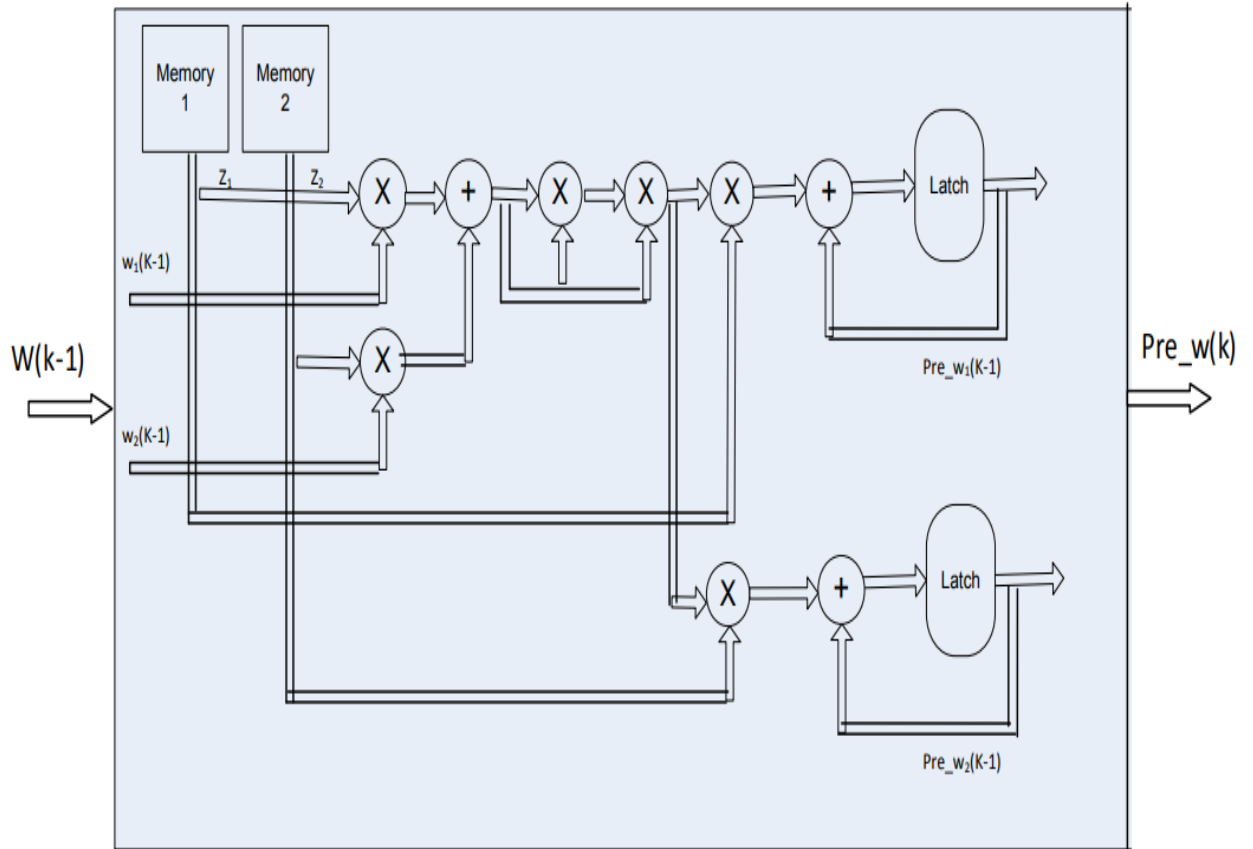


Figure 3.4 : Diagramme de la mise en œuvre de $Per_w(k)$ [15].

Après l'implémentation de (1), le calcul de (2) est mise en œuvre dans la Figure 3.5,

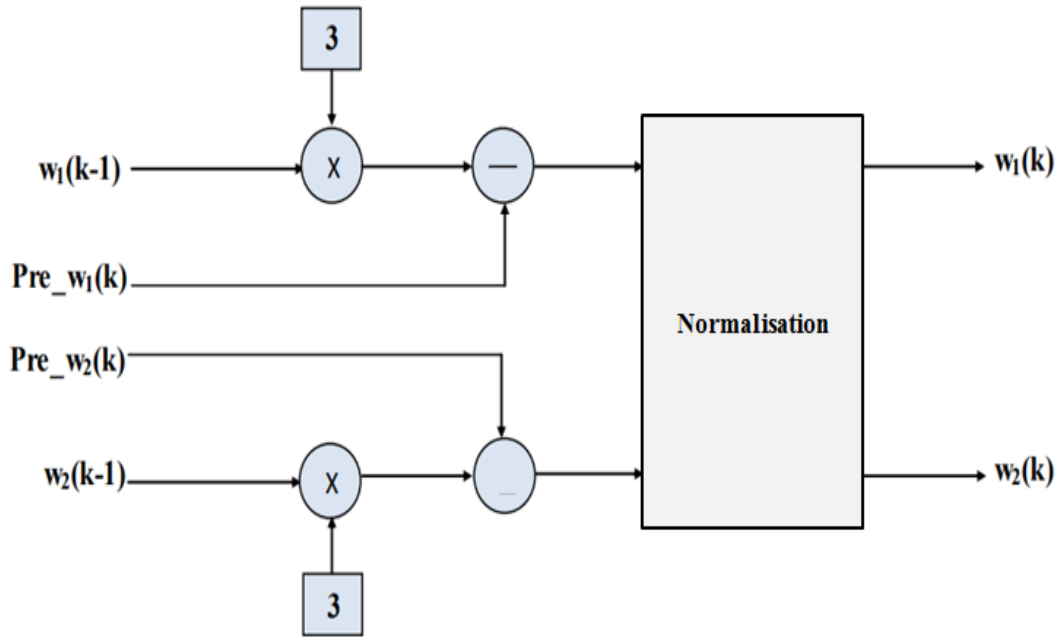


Figure 3.5 : Diagramme de la mise en œuvre de $w(k)$ [15].

Le block de normalisation dans la figure 3.4 a pour but de comparer entre la nouvelle valeur de w avec l'ancienne valeur de w , et si les valeurs ne correspondent pas, alors la nouvelle w est renvoyé à l'entrée du bloc et également stocké en tant que l'ancienne w dans un registre à des fins de comparaison.

Lorsque la nouvelle $w = l'ancien_w$, alors cette valeur est donnée à la sortie comme le vecteur convergé w qui donne une composante indépendante.

Pour trouver l'autre composante indépendante, un nouveau vecteur aléatoire w est supposé et il est dé-corrélé avec le w précédent et est à nouveau soumis au processus d'itération pour obtenir une valeur convergente optimisée pour $i = 1$ et 2 (pour deux signaux mixtes) [15].

$$\widetilde{A}_1^{-1} = \begin{bmatrix} w_1 & 0 \\ w_2 & 0 \end{bmatrix} \quad \text{Pour le premier composant indépendant}$$

$$\widetilde{A}_2^{-1} = \begin{bmatrix} w_1 & w_{1\text{ nouv}} \\ w_2 & w_{2\text{ nouv}} \end{bmatrix} \quad \text{Pour le deuxième composant indépendant}$$

Où $\begin{bmatrix} w_{1\text{ nouv}} \\ w_{2\text{ nouv}} \end{bmatrix}$ sont les nouveaux vecteurs aléatoires pour le deuxième composant indépendant.

3.6 Conclusion

Dans le but d'extraire les signaux sources d'un signal mixte en utilisant la l'algorithme Fast-ICA, nous avons présenté dans ce chapitre théoriquement les blocks principaux pour la mise en œuvre de l'algorithme Fast-ICA :

- Blocks de diagrammes pour le prétraitement d'un signal.
- Blocks de diagrammes pour le calcul de matrice de séparation W en utilisant le kirtosis.

À partir de cette étude théorique, nous pouvons passer à l'implémentation pratique sous MATLAB dans le chapitre suivant.

Chapitre 4

Résultat et discussion

4.1 Introduction :

Dans les chapitres précédents, nous avons présenté théoriquement la séparation aveugle avec la technique ACI ainsi nous avons appris les différents algorithmes que cette technique peut utiliser pour le processus de séparation des sources où dans ce mémoire, nous avons choisi l'algorithme Fast-ICA pour implémenter notre application.

Dans ce chapitre et à l'aide de Matlab, nous concevons un système de séparation de source aveugle par la méthode Fast-ICA à l'aide de deux signaux aléatoire, et Les résultats obtenus seront discutés.

4.2 Simulation de l'algorithme Fast-ICA (à l'aide de Matlab) :

4.2.1 Chargement des données

À l'aide de deux fichiers audio qui sont chargés depuis le répertoire "datafun" de Matlab (les sons standard de Matlab), nous avons appliqué l'algorithme Fast-ICA, où

Source 1 (S_1) : un son de rire.
Source 2 (S_2) : un son d'un train.

```
load('laughter.mat','y');%chargement de 1er source
s1=y(1:10000);
load('train.mat','y'); %chargement de 2ème source
s2=y(1:10000);

s=[s1 s2]; %former la matrice des sources
```

4.2.2 Génération des données :

➤ Les signaux sources :

```
figure
plot(s1)
title(['source ',num2str(1)])

figure
plot(s2)
title(['source ',num2str(2)])
```

```
figure
plot(s1,s2, '.')
title('distribution des sources')
xlabel('S1')
ylabel('S2')
```

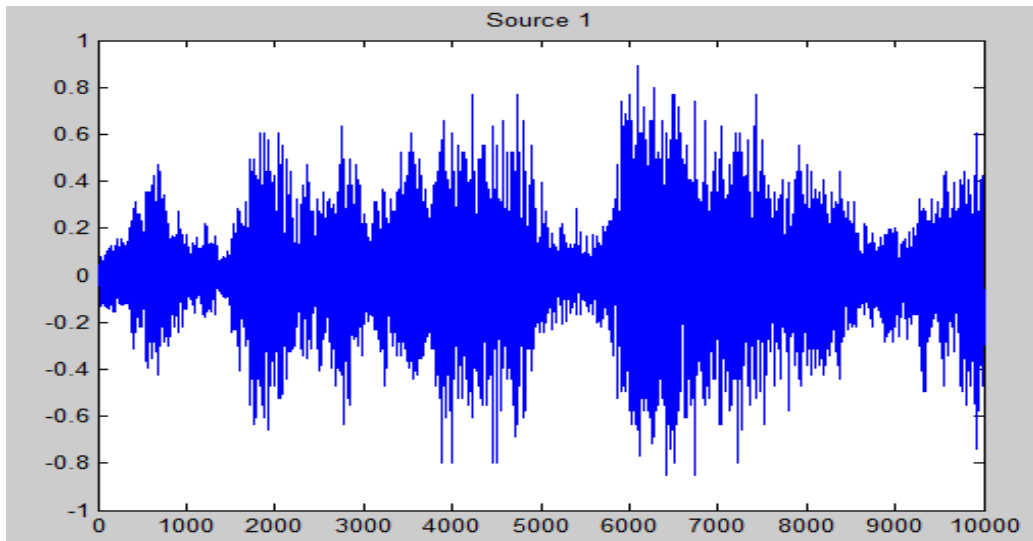


Figure 4.1 : Signal source 1.

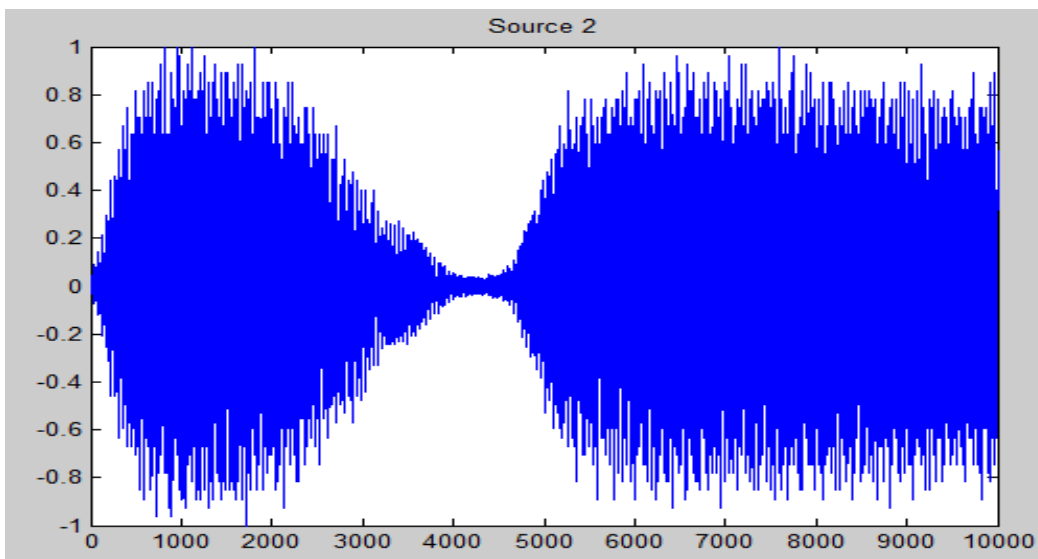


Figure 4.2 : Signal source 2.

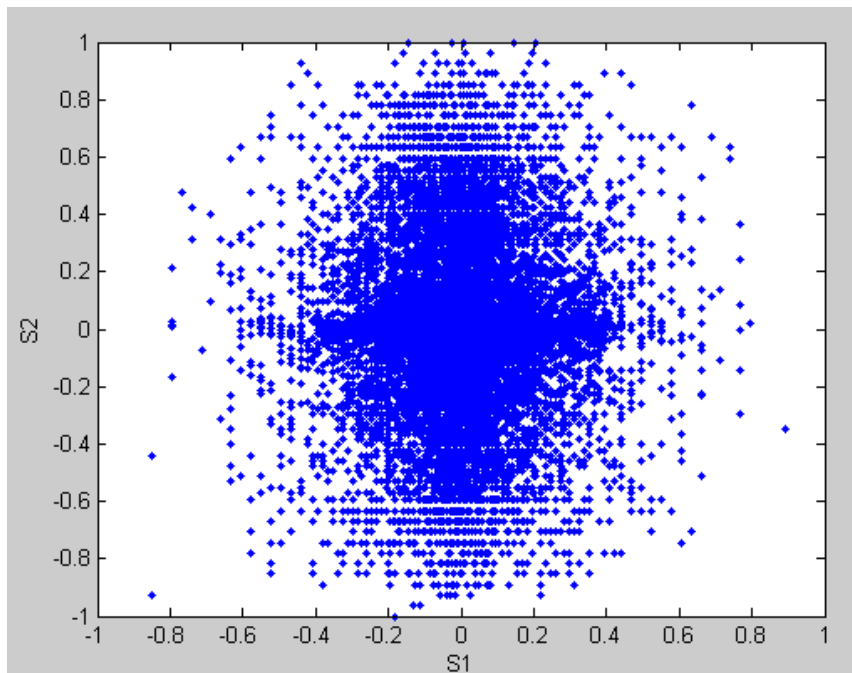


Figure 4.3 : La distribution des signaux sources S_1 et S_2 .

➤ Les signaux mixtes:

Nous avons mélangé les signaux S_1 et S_2 ensemble en utilisant une matrice de mélange aléatoire A :

```
A=[0.3 0.7; 0.8 0.3]; % matrice de mélange
aléatoire
x1=(0.3*s1)+(0.7*s2);% signal mélange1
x2=(0.8*s1)+(0.3*s2);% signal mélange2
x=[x1,x2]; %matrice de mélange
```

Pour tracer les signaux :

```
figure
plot(x1)
title(['signal mixte ', num2str(1)])

figure
plot(x2)
title(['signal mixte ', num2str(2)])
```

```
figure
plot(x1,x2, '.')
title('distribution des de mélange')
xlabel('X1')
ylabel('X2')
```

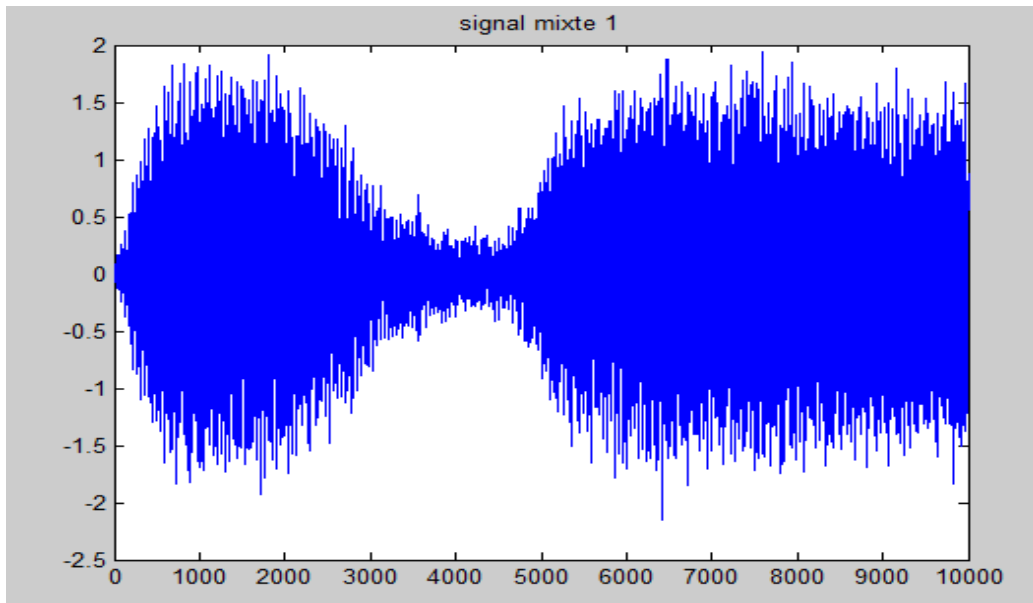


Figure 4.4 : Signal mixte X_1 .

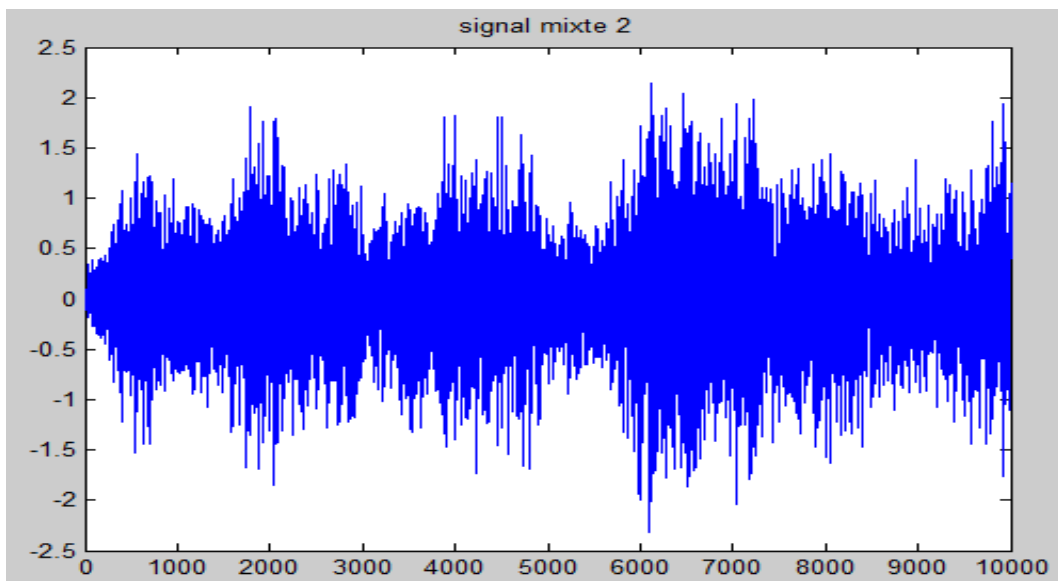


Figure 4.5 : Signal mixte X_2 .

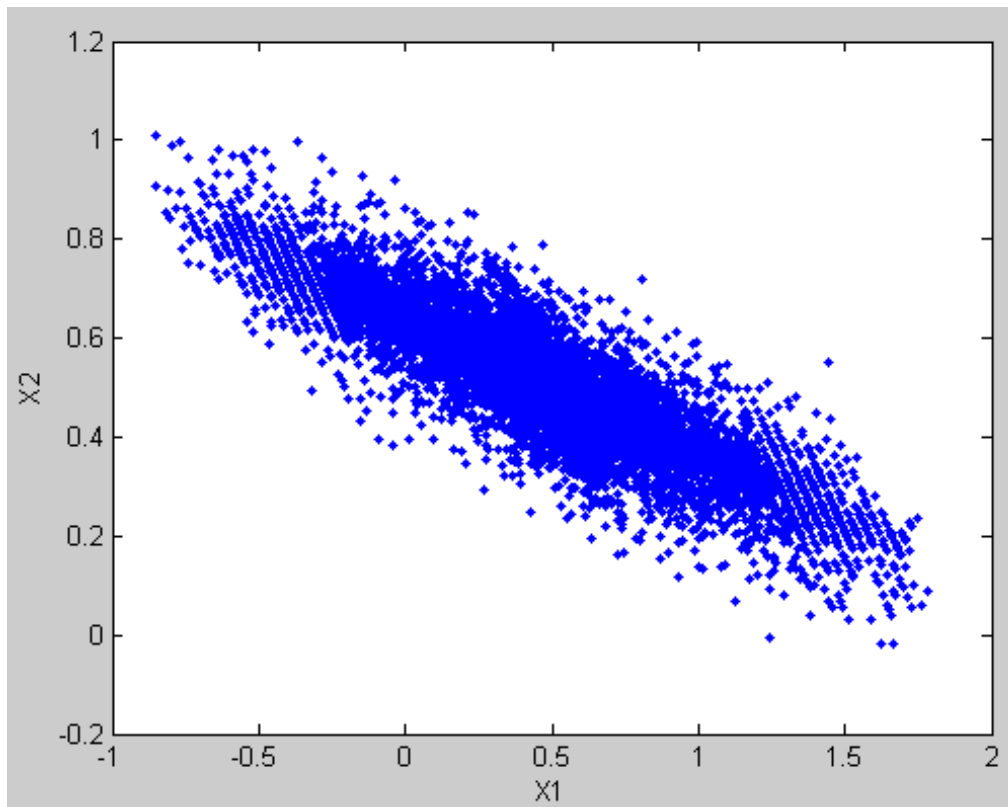


Figure 4.6 : La distribution des signaux mixtes X_1 et X_2 .

Commentaire :

On remarque que les signaux mixtes ont une certaine intensité par rapport aux signaux sources, et on le remarquera plus chaque fois que le nombre de sources augmente (s_1, s_2, s_3, \dots), c'est-à-dire que plus il y a des sources, plus il y a de bruit (théorème de centre limite).

Pour écouter les sons originaux ou les sons mixtes, nous pouvons utiliser la commande 'sound', par exemple :

```
>> sound(x1)
```

```
>> sound(x2)
```

4.2.3 Prétraitement :

Pour séparer efficacement les signaux sources, il faut appliquer un prétraitement sur les signaux mixtes.

➤ **Centrage**

Nous avons centré les données en soustrayant la moyenne de tous les signaux de façon à rendre la variable x de moyenne nulle.

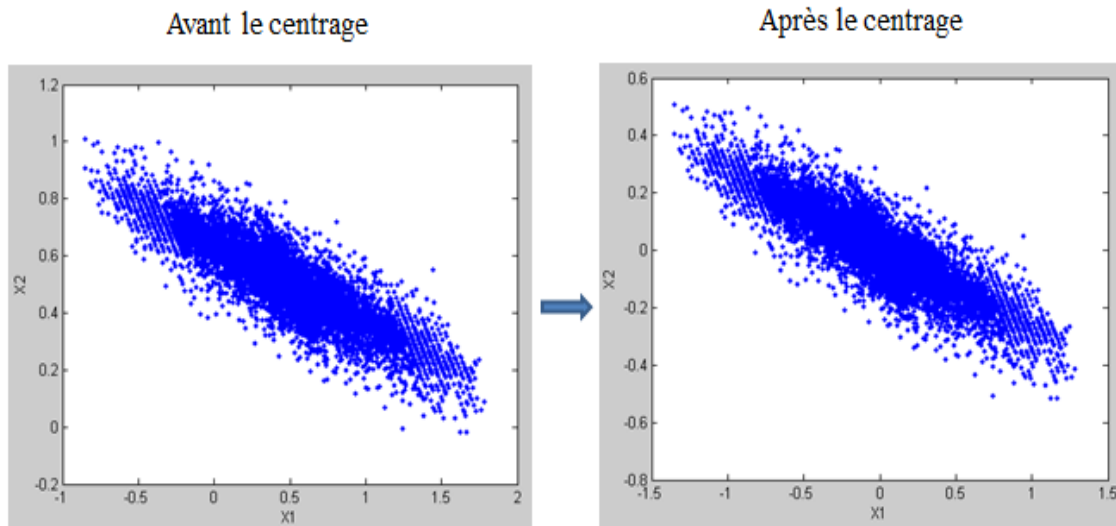


Figure 4.7 : Avant et après le centrage.

➤ **Blanchiment**

Après le centrage des données, nous avons blanchi tous les données mixtes de manière à ce qu'elle ait une moyenne nulle et une covariance d'identité.

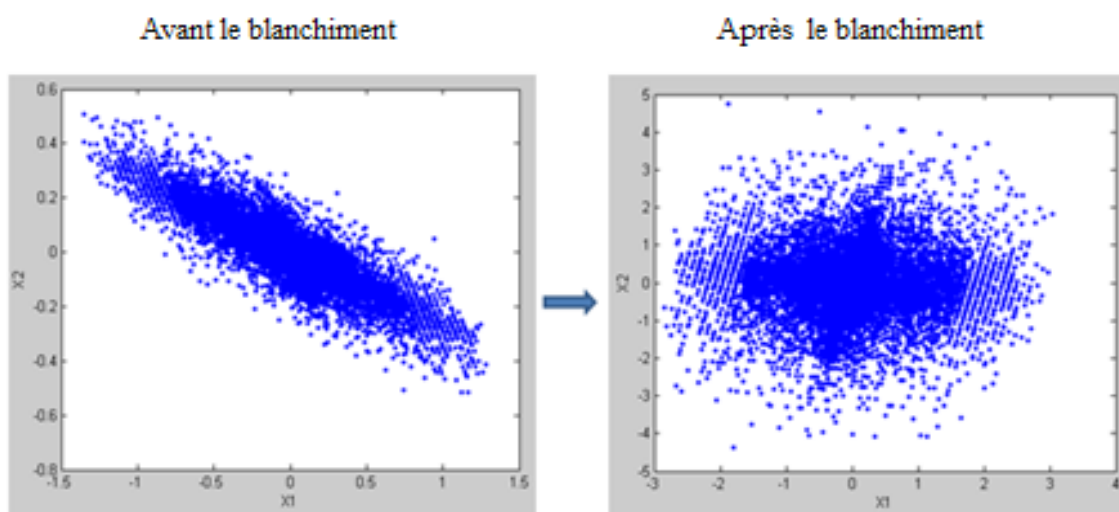


Figure 4.8 : Avant et après le blanchiment.

4.2.4 Séparation des signaux :

Pour extraire les signaux source, nous avons tapé les commandes suivantes :

```
mixing; %faire le mixage
[outsig, A, W] = fastica(x);% appliquer fastica sur x, A matrice mélange estimé,W matrice séparation estimé, outsig matrice contient les signaux séparé
```

Pour tracer les signaux :

```
figure
plot(outsig(1,:))
title(['signal estimé ',num2str(1)])

figure
plot(outsig(2,:))
title(['signal estimé ',num2str(2)])
```

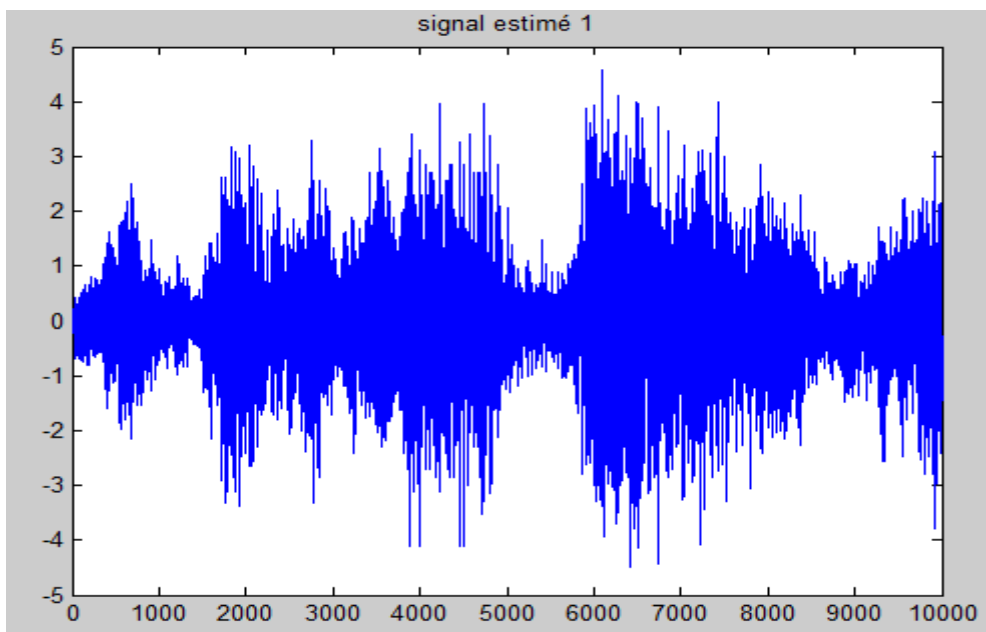


Figure 4.9 : signal source estimé 1.

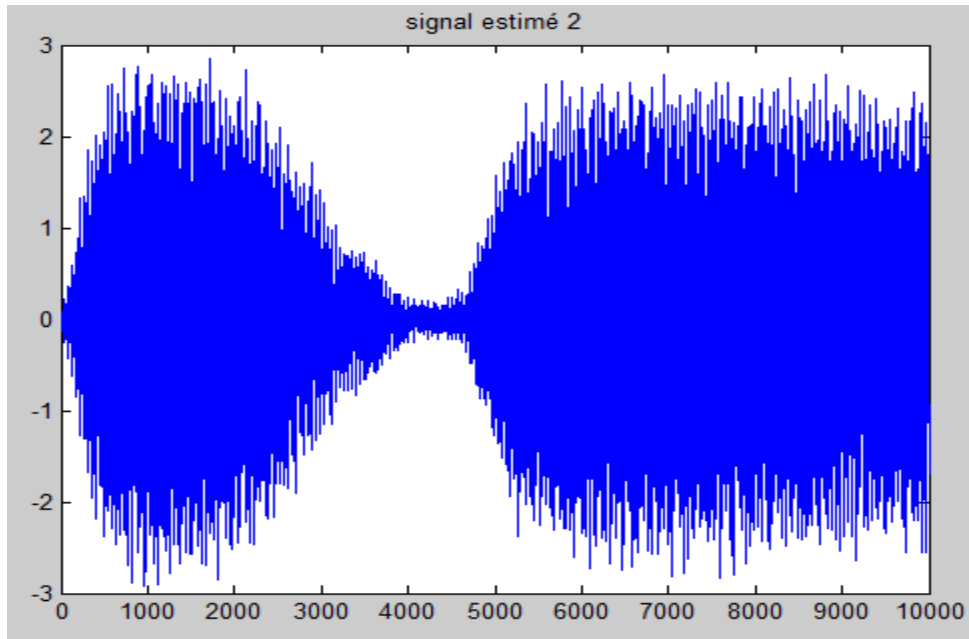


Figure 4.10 : Signal source estimé 2.

Commentaire

Généralement, l'ordre des signaux après le processus de séparation peut différer de l'arrangement d'origine. Il est nécessaire de réorganiser les colonnes pour faire correspondre les signaux estimés avec les signaux d'origine correspondants.

Les signaux estimés ressemblent aux signaux d'origine où la covariance des signaux : $1.11022e-15$.

4.3 Comparaison entre les signaux de sources et les signaux séparés :

Pour but de faire une comparaison entre les signaux sources et les signaux séparés, nous avons calculé la corrélation entre le signal source et signal séparé pour obtenir une mesure de la ressemblance entre eux. Et pour cela, nous avons tapé les deux commandes suivantes :

```
e= xcorr(s1',outsig(1,:)) ;  
figure  
plot(e)  
  
e= xcorr(s2',outsig(2,:)) ;  
figure  
plot(e)
```

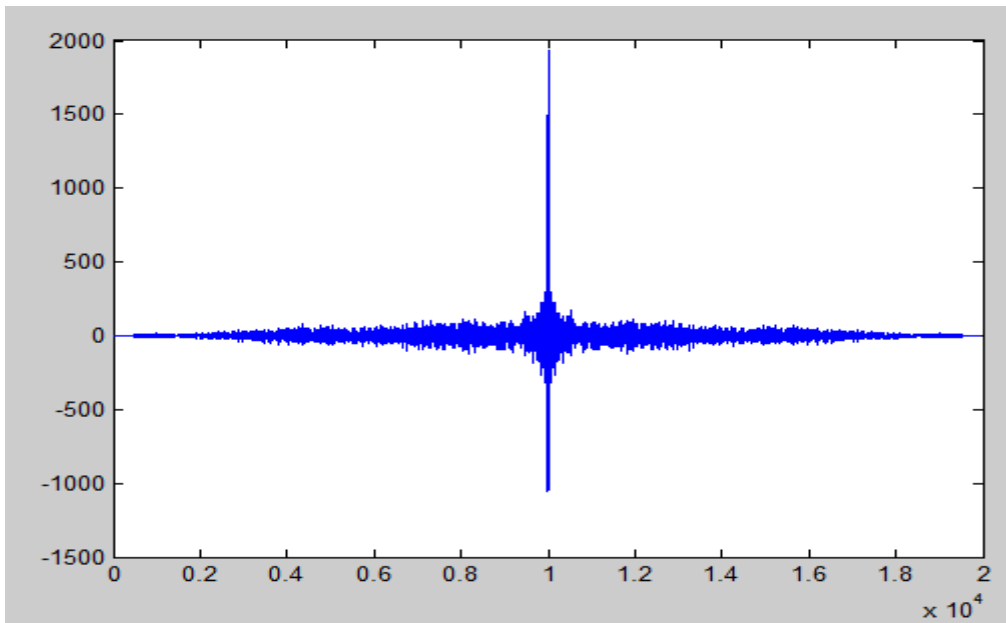


Figure 4.11 : Signal de corrélation entre S_1 et signal estimé 1.

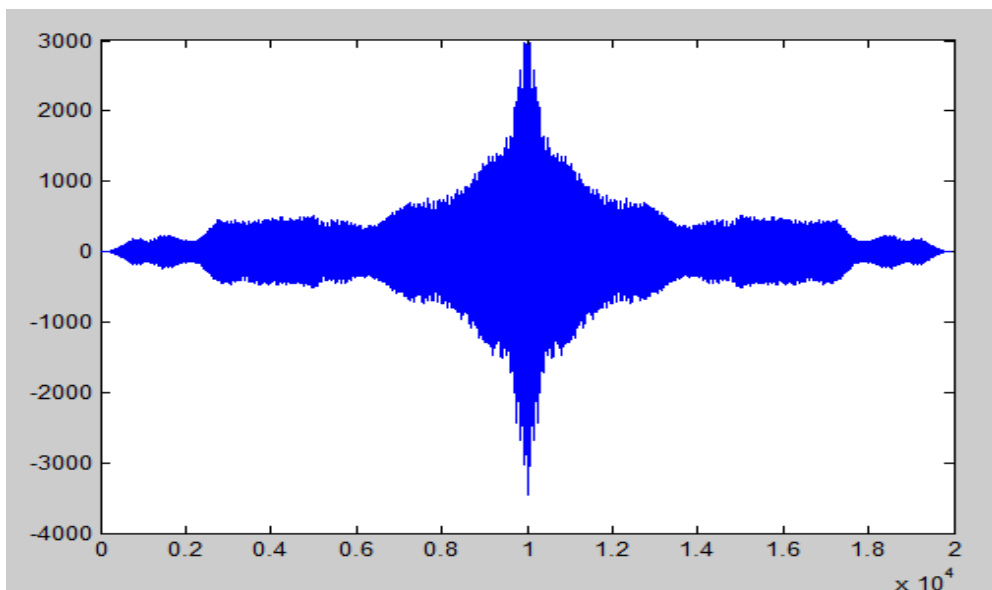


Figure 4.12 : Signal de corrélation entre S_2 et signal estimé 2.

Commentaire :

Chaque des figures (4.11) et (4.12) indiquent la ressemblance entre S_1 et le signal estimé 1 et S_2 et le signal estimé 2 où le pic se produit lorsque les éléments de S et signal estimé correspondent exactement (à la valeur 1) et les autres tend vers de zéro.

4.4 Conclusion :

À application de l'algorithme Fast-ICA, nous voyons bien qu'après que les deux signaux étaient complètement mélangés entre eux, nous avons réussi à réaliser une bonne séparation des deux signaux avec cette approche où il permet une séparation avec un calcul rapide et moins d'erreur.

La bonne séparation des signaux est basée sur le prétraitement des données avant appliqué l'algorithme Fast-ICA, car le blanchiment et le centrage préparent les données avant la séparation pour avoir des bons résultats.

*C*onclusion générale

1. Conclusion générale

Dans cette thèse, nous nous sommes intéressés au problème de la séparation aveugle avec la méthode d'analyse en composantes indépendantes ACI en utilisant des signaux de parole. Cette technique qui applique la séparation juste en utilisant le signal mixte n'a aucune information ni sur les signaux sources ni sur le mécanisme du mélange.

L'ACI repose essentiellement sur deux facteurs principaux qui rendent possible cette séparation : l'indépendance statique et la non gaussianité des signaux sources. Ces deux dernières sont des hypothèses liées entre elles: le premier dit que les signaux issus de processus différents sont certainement des signaux indépendants les uns des autres. D'autre part, la deuxième théorie dit que les signaux qui contiennent d'information (sans bruit) donnent une distribution non gaussienne et leur somme donnent une distribution gaussienne.

Avant de commencer le processus de séparation des signaux source provenant de signaux mixtes et pour but d'avoir des bons résultats après la séparation, il faut que les signaux mixtes passent par un prétraitement des données qui consiste à les centrer et les blanchir.

Pour estimer l'une des composantes indépendantes, une matrice de séparation W doit être calculée où W est l'inverse de matrice de mélange A , et pour cela, une fonction de contraste doit être choisie pour résoudre le problème de séparation. Parmi ces fonctions qui cherchent à maximiser la non gaussianité pour séparer les signaux sources dans l'ACI, cela se fait à partir de deux méthodes : *kurtosis* ou de *la néguentropie*. D'autres fonctions utilisent d'autres critères dans la théorie de l'information pour l'estimation de W , certains d'entre eux minimisent les informations mutuelles des sources et d'autres utilisent le maximum de vraisemblance des sources.

Au cours des années, l'ACI a acquis de nombreux algorithmes (JADE, Infomax, Fast-ICA...etc.), où chaque méthode est différente de l'autre selon la fonction de contraste utilisée (maximisant la non-gaussianité, minimisant les informations mutuelles, maximum de vraisemblance) et l'algorithme d'optimisation (algorithmes du point fixe, algorithme naturel du gradient...etc.).

Fast-ICA est un algorithme à un schéma d'itération à virgule fixe qui maximise une mesure de non-gaussianité des composants indépendants. Il est l'une des algorithmes efficace et populaire et nous avons choisi d'utiliser dans la partie pratique de notre mémoire où nous avons bien réussi de réaliser la séparation des signaux sources.

2. Les travaux futurs et perspectifs

Une architecture de l'algorithme Fast-ICA peut être développée pour le traitement séquentiel de signaux mixtes en temps réel, où une implémentation étendue de l'algorithme Fast ICA basé sur les modules proposés peut être effectuée pour une dimension supérieure (plus de de nombres des sources et mélanges) comme dans un le canal MIMO (Multiple-Input Multiple-Output) et le traitement de réception dans un système MIMO.

Le modèle du signal de la séparation aveugle de sources est donc sensiblement le même que celui d'un système de transmission MIMO, son application aux systèmes MIMO serait très prometteuse [7,15].

- [1] **James V. Stone**, INDEPENDENT COMPONENT ANALYSIS A Tutorial Introduction, Cambridge, A Bradford Book la presse du MIT.
- [2] **Samuel AMBAPOUR**, Introduction à l'analyse des données, Brazzaville, bureau d'application des méthodes statistiques et informatiques BAMSI, avril 2003, p7.
- [3] **Hanany Ould-Baba**, Analyse en Composantes Indépendantes Multidimensionnelles via des cumulants d'ordres variés, thèse de Doctorat en Sciences, Université de Technologie de Compiègne, 2014.
- [4] **TOUAIBIA Moufida et AKKACHE Ouardia**, Implémentation de la méthode de séparation de source aveugle FAST ICA sous XSG, mémoire de magister en traitement de l'information et Systèmes Électroniques, Université SAAD DAHLAB de Blida, 2017-2018
- [5] **Aapo Hyvärinen et Erkki Oja**, Independent Component Analysis: Algorithms and Applications, Neural Networks Research Centre Helsinki University of Technology.
- [6] **RABAH Fiad**, La séparation aveugle de sources par les techniques neuronales, mémoire de magister en électronique, Université de Batna.
- [7] **Yasmina BENABDERRAHMANE**, séparation aveugle de signaux de parole utilisant des statistiques d'ordre supérieur et la décomposition en sous-espaces, thèse de Doctorat en télécommunications, Université de Québec, 2011.
- [8] **Aapo Hyvärinen et Juha Karhunen et Erkki Oja**, Independent Component Analysis, Edition Wiley Interscience, version finale du 7 mars 2001.
- [9] **Shireen Elhabian et Aly Farag**, Independent Component Analysis (ICA), Université de Louisville, Septembre 2009.
- [10] **Yaseen Unnisa et Danh Tran and Fuchun Huang**, Statistical Independence and Independent Component Analysis, Australia, Collège d'ingénierie et des sciences, Université Victoria, 04 octobre 2016.
- [11] **Cyril VIRON**, Analyse en composantes indépendantes pour la caractérisation d'images hyper spectrales en télédétection, mémoire de magister en génie de la production automatisée, Université de Québec, 23 mai 2006.
- [12] **Marc Castella**, Séparation de sources non linéaires dans le cas des mélanges convolutifs, mémoire doctorat en science, Université de Marne la Vallée, 20 décembre 2004.
- [13] **Eric MOREUA**, Une généralisation de l'algorithme JADE, Université de Toulon et du Var, Dix-septième colloque GRETSI, Vannes, 13-17 septembre 1999.
- [14] **L. Senhadji et A. Kachenoura et L. Albera et P. Comon**, Revue générale sur l'exploitation des approches d'analyse en composantes indépendantes dans les interfaces cerveau machine, IRBM 30 (2009) 211–217

[15] **Sasmita KUMARI BEHERA**, Fast-ICA for blind source séparation and its implimentation, mémoire de magister, National Institute of Technology, 2009.

[16] **Tianwen Wei**, Analyse de la convergence de l'algorithme Fast-ICA: Echantillon de taille finie et infinie, mémoire doctorat en science, Université des Sciences et Technologies de Lille, 10 Juin 2013.

[17] <https://research.ics.aalto.fi/ica/fastica/code/dlcode.html>, le 3 mars 2020 à 20h.

[18] <https://fr.mathworks.com/help/stats/extract-mixed-signals.html#d120e57289>, le 22 mai 2020 à 13h:30.