

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR**

**ET DE LA RECHERCHE SCIENTIFIQUE**

**UNIVERSITE DE BATNA**

**FACULTE DE TECHNOLOGIE**

**MEMOIRE**

Présenté au

**DEPARTEMENT D'ELECTRONIQUE**

Pour l'obtention du diplôme de

**MAGISTER EN ELECTRONIQUE**

**Option : Traitement du signal**

Par

**FIAD RABAH**

Ingénieur d'état en électronique - Université de Constantine

**Thème**

**La séparation aveugle de sources par les techniques neuronales**

Mémoire soutenu publiquement le :

Devant le jury composé de :

**Pr. Moussa BENYOUCEF**

**Professeur U de Batna**

**Président**

**Dr . Djemai ARAR**

**M.C. A U de Batna**

**Rapporteur**

**Dr. Redha BENZID**

**M.C. A U de Batna**

**Examineur**

**Dr. Amar MEZAACHE**

**M.C.A U de M'sila**

**Examineur**

## Remerciements

Je tiens à remercier tout d'abord **ALLAH** le tout puissant qui nous a donné, durant toutes ces années, la santé, le courage et la foi pour avoir finaliser ce travail.

Je tiens à exprimer ma reconnaissance à mon encadreur **Mr. ARAR Djemai**, Maître de conférences à l'université de BATNA pour toute sa confiance qui m'a fait, son aide, son encouragement et ses conseils durant la préparation de ce travail.

Je tiens à remercier vivement monsieur **Mr. BENYOUCEF Moussa**, Professeur au département d'électronique de l'université de BATNA, d'avoir accepté de présider le jury de ce mémoire.

J'adresse mes vifs remerciements à **Mr. BENZID Redha**, Maître de conférences au département d'électronique de l'université de Batna, pour avoir accepté d'examiner ce travail et pour sa totale sympathie.

Je voudrais également exprimer mes vifs remerciements à monsieur **Mr. MEZAACHE Amar** maître de conférences au département d'électronique de l'université de M'sila pour l'honneur d'assister à ce jury et d'enrichir le manuscrit.

# *Sommaire*

# Sommaire

Introduction générale.....	2
----------------------------	---

## **Chapitre I : Généralités sur la Séparation Aveugle de Sources**

I-1 Introduction.....	6
I-2 Bref historique.....	6
I-3 Analyse en composantes indépendantes.....	6
I-3.1 Définition.....	7
I-4 La séparation aveugle de source.....	7
I-4.1 Quelques hypothèses de base.....	7
I-5. Information mutuelle :	10
I-6. Le critère de maximum de vraisemblance.....	10
I-7. Moments et cumulantes.....	10
I-7.1 Première fonction caractéristique.....	10
I-7.2 Deuxième fonction caractéristique.....	11
I-8. Fonction de contraste .....	11
I-8.1 Définition.....	11
I-9. Propriétés de l'indépendance.....	12
I-10. Les types de mélanges.....	13
I-10.1 Mélange linéaire.....	13
I-10.1.1 Instantané.....	13
I-10.1.2 Convolutif.....	13
I-10.2 Mélange non linéaire.....	13
I-10.3 Mélange sous déterminé.....	13
I-10.4 Mélange sur déterminé.....	13
I-11. Analyse en composantes principales.....	14
I-12. Blanchiment de données.....	14
I-13. Critères de performances de la SAS.....	14

1-13.1 Erreur quadratique moyenne (EQM).....	14
1-13.2Rapport sinal sur résidus (SNR).....	15
I-14. Conclusions.....	15

## **Chapitre II : Les Réseaux de Neurones Artificiels**

II-1. Introduction.....	17
II-2. Bref historique.....	17
II-3. Neurone biologique.....	18
II-4. Définition.....	19
II-5. Composition d'un neurone artificiel.....	19
II-5.1 Structure.....	19
II-5.2 le poids.....	20
II-5.3 Les différents types de fonctions d'activation.....	20
II-5.3.1 Fonction binaire.....	20
II-5.3.2 Fonction linéaire.....	21
II-5.3.3 Fonction sigmoïde.....	21
II-5.3.4 Fonction gaussienne.....	21
II-5.4 Structure d'interconnexion .....--.....	22
II-6. Apprentissage des réseaux de neurones artificiels.....	23
II-6.1 Apprentissage supervisé.....	23
II-6.1 Apprentissage non supervisé.....	24
II-7. Perceptron.....	24
II-7.1 Apprentissage du perceptron.....	25
II-8 Perceptron multicouche MLP.....	26

II-8.1 Apprentissage du MPL .....	27
II-8.2 Problème de minimum local .....	27
II-9. Le réseau à fonction de base radiale RBF.....	27
II-10. Conclusion.....	28

## **Chapitre III : Les Approches Neuronales**

### **III-1 Approche de Herault –Jutten**

III-1.1 Introduction .....	31
III-1.2 Architecture du réseau.....	31
III-1.3 La méthode du gradient.....	32
III-1.4 Réseau direct (feedforward).....	33
III-1.1 1 <sup>ère</sup> solution.....	34
III-1.2 2 <sup>ème</sup> solution.....	34
III-1.5 Réseau récurrent (feedback).....	34
III-1.6 La règle d'adaptation.....	35
III-7 Rôle de la non-linéarité.....	35

### **III.2 Approche de Moreau-Macchi**

III-2.1 Introduction.....	36
III-2.1.1 Réseau récurrent (feedback).....	37
III-2.1.2 Réseau direct (feed forward) .....	38
III-2.2 Loi d'adaptation.....	38
III-2.3 Réseau mixte.....	38
III-2.4 Solution.....	39
III-2.5 Loi d'adaptation du réseau mixte.....	40

### **III-3 Approche de Clemente-Acha**

III-3.1 Introduction.....	40
III-3.2 Hypothèses.....	40
III-3.3 Les solutions.....	41
III-3.4 Nouveau critère de séparation.....	42
<b>III.4 Approche de Cichocki-Amari</b>	
III-4.1 Modification de l'algorithme de H-J .....	43
III-4.2. Formalisation du problème.....	43
III-4.2.1 Réseau direct (feedforward).....	44
III-4.2.2 Réseau récurrent (feedback).....	45
III-4.3 Extension de la base d'apprentissage.....	45
<b>III-5 Approche de Cardoso-Laheld</b>	
III-5.1 Introduction.....	46
III-5.2 Série d'adaptation.....	47
III-5.3 Adaptation de la matrice de blanchiment.....	47
III-5.4 Adaptation de la matrice U.....	48
III-5.5 Hypothèse.....	48
III-5.6 Gradient Relatif.....	48
III-5.7 L'algorithme E.A.S.I.....	49
III-6 Conclusion.....	49

## **Chapitre IV: Discussions et Interprétation des Résultats**

IV-1.1 Exemple 1.....	52
IV-1.2 Exemple 2.....	55
IV-1.3 Exemple 3.....	57

IV-1.4 Exemple 4.....	60
IV-1.5 Exemple 5.....	63
IV-2 Conclusion.....	66

## **Conclusion Générale**

Conclusion générale.....	68
--------------------------	----



***Introduction***  
***Générale***

---

## Introduction générale

Le cerveau de l'être humain est l'unité centrale de traitement de l'information, et à base de ces neurones plusieurs signaux proviennent séparément que ce soit par voie visuelle ou auditive. La modélisation de la transmission des signaux d'information de mouvement d'une articulation par exemple au niveau du bras ou du pied, tels que la vitesse ou la position, a donné naissance à un problème appelé le problème de la séparation aveugle de sources.

La séparation aveugle de sources consiste à estimer un ensemble de signaux inconnus dits sources à partir des signaux connus dits observations (ou mélanges). Les observations sont des mélanges des signaux sources reçues par des capteurs (microphone, sondes, cameras, antennes...).

La séparation de sources est considérée comme achevée lorsque chaque signal source est restitué à un facteur scalaire près, on ne peut prévoir sur quelle sortie du séparateur un certain signal est restitué. La séparation est effectuée en exploitant les informations portant principalement sur les signaux eux-mêmes.

Le terme aveugle signifie que la séparation est faite sans information a priori sur les signaux sources, la seule information exigée est l'indépendance statistique de ces signaux.

La séparation aveugle de sources (SAS) peut être exploitée dans des nombreuses applications telles que :

- **Le domaine de la télécommunication** : l'application du signal émis, qui a une forme particulière, est une représentation codée. Le canal de transmission, caractérisé par une fonction de transfert, qui définit la distorsion apportée à la transmission par le canal (filtre) et en plus les problèmes particuliers comme l'effet doppler ou la mauvaise synchronisation émetteur/ récepteur, le bruit de canal [11 ,16].
- **Le domaine médical** : La séparation de sources est utilisée pour l'extraction non invasive de l'Electrocardiogramme (ECG), l'Electromyogramme (EMG) ou l'Electroencéphalogramme (EEG). On note par exemple l'extraction non invasive des battements cardiaques du fœtus à partir des signaux ECG enregistrés à l'aide d'électrodes placés sur le ventre de la mère est possible grâce à la séparation des sources [17].

- **le domaine sismique** : le signal émis est provoqué par explosion ou vibration, la fonction de transfert représente les données aux géologues, géophysiciens, pétroliers [2].

La difficulté de la séparation de sources est reliée directement à la nature des signaux séparés et la manière dont les sources sont mélangées.

Dans ce travail, on présente une étude rigoureuse du problème de la séparation aveugle de sources. Plusieurs techniques ou approches sont présentées également par des chercheurs pour apporter une solution à ce problème à savoir des techniques basées sur les réseaux de neurones artificiels, sur la maximisation et minimisation de l'information, sur l'information mutuelle, et sur les statistiques d'ordres supérieurs. On s'est focalisé sur le choix des approches neuronales dites également méthodes connexionnistes. Ce choix est légitime car le premier algorithme proposé était basé essentiellement sur une méthode connexionniste d'un coté, d'un autre coté, plusieurs chercheurs se sont concentrés sur cet axe.

Dans ce manuscrit, on expose quelques algorithmes adaptatifs de séparation aveugle de sources. Afin de démontrer les performances et les qualités des différents algorithmes neuronaux utilisés, un programme en MATLAB étant implémenté pour faciliter cette démarche.

Nous pouvons résumer les objectifs de ce travail dans les points suivants :

1. Exposer le problème de la séparation de sources, définition, concept et utilité ainsi que ces applications diverses.
2. Donner des solutions de ce problème en se basant principalement sur les méthodes neuronales ou connexionnistes.
3. Exposer quelques approches neuronales.
4. Implémenter un programme en MATLAB pour les exemples de simulation.

Ce mémoire est articulé sur quatre chapitres :

Dans le premier chapitre on rappelle la définition de la séparation aveugle de sources, les critères de séparation des signaux à savoir l'information mutuelle, le maximum de vraisemblance, les statistiques d'ordre supérieur et les différents types de mélanges.

Le deuxième chapitre donne un aperçu sur les réseaux de neurones artificiels, définition et concepts, leurs différents types et leur domaine d'utilisation.

Le chapitre trois traite plusieurs approches de séparation basées principalement et essentiellement sur les techniques neuronales, en effet l'approche de Héroult-Jutten est exposé en détail. Cette approche constitue la première tentative de solution du problème de la séparation de sources. D'autres approches constituent une amélioration de l'efficacité de l'algorithme de base H-J.

Le quatrième chapitre expose les résultats de quelques exemples de simulation pour chaque algorithme étudié afin de donner les avantages et les limitations de chaque approche.

Une conclusion générale donnera une synthèse en ce qui concerne les différentes démarches suivies dans ce travail et expose également les perspectives envisagées pour les travaux futurs.

# **CHAPITRE I**

## ***Généralités sur la Séparation de Sources***

## **I-1. Introduction**

La séparation aveugle de sources occupe un volume très important dans le traitement du signal, elle utilise de nombreuses parties de la théorie de signal, dans ce chapitre on donne quelques notions de base de la séparation de sources ainsi que les différentes définitions exploitées dans la séparation.

## **I-2. Bref historique**

Le premier travail dans le domaine de séparation aveugle de source a été présenté par Héroult et Jutten dans les années 80 [18].

Taleb et Jutten décrivent le problème biologique qui consiste à des réponses musculaires émises à l'essai de différentes sorties d'excitation [52].

Les travaux de P. Comon en 1994 ont permis de formaliser le lien entre la méthode d'analyse en composantes indépendantes (ICA) et le problème de la séparation aveugle de source (BSS) dans le cadre d'un mélange linéaire.[8][9].

En 1999, Taleb et Jutten ont introduit un nouveau type de mélange, les mélanges post non linéaire. [53]

Les travaux de J.F. Cardoso proposent d'envisager une technique appelée « the three easy routes to independent component analysis » pour les mélanges linéaires [8].

La séparation de source est un problème général dans le traitement du signal dont le principe consiste à retrouver les signaux sources à partir des signaux mélanges. La plupart des méthodes de la séparation se basent essentiellement sur l'indépendance statistique des sources [5][8][9], qui reste une condition nécessaire et suffisante.

## **I-3. Analyse en composantes indépendantes (ICA)**

La séparation à l'ordre 2 est essentiellement fondée sur la décorrélation des signaux observés. Pour le cas de signaux sources gaussiens, l'utilisation de la décorrélation n'exploite pas complètement l'indépendance qui est une propriété bien plus forte que la décorrélation. C'est dans ce contexte qu'apparaît l'intérêt de l'usage des statistiques d'ordre supérieur en séparation sources.

En utilisant des statistiques d'ordre supérieur à 2 on exploite d'avantage l'indépendance des sources, on obtient ainsi des critères permettant de réaliser la séparation sans information supplémentaire. Ceci conduit à un nouveau concept : l'Analyse en Composantes Indépendantes (en anglais Independent Component Analysis ou ICA) introduite par Jutten et Héroult dans leur travaux [44]. Ces méthodes ont été développées après par Comon [7]. Cette nouvelle approche vient compléter l'Analyse en composantes principales (ACP) fondée sur la seule décorrélation. L'ICA est une des voies majeures de la SAS. Son principe, sous l'hypothèse d'indépendance mutuelle des sources, consiste à transformer linéairement le vecteur des signaux observés en un vecteur dont les composantes sont statistiquement indépendantes. Il s'avère que, dans un contexte non bruité, cela est équivalent à l'estimation des paramètres du modèle instantané.

Plus précisément Comon a donné la définition de l'ICA suivante :

**I-3.1 Définition** L'ICA d'un vecteur aléatoire  $\mathbf{x}(\mathbf{n}) := (\mathbf{x}_1(\mathbf{n}), \dots, \mathbf{x}_p(\mathbf{n}))^T$  est défini comme la donnée d'une matrice de « séparation »  $\mathbf{B}$  telle que le vecteur  $\mathbf{y}(\mathbf{n}) := (\mathbf{y}_1(\mathbf{n}), \dots, \mathbf{y}_p(\mathbf{n}))^T := \mathbf{B}\mathbf{x}(\mathbf{n})$  soit à composantes indépendantes, dans le sens de la maximisation d'un « contraste »  $C(\mathbf{y}_1(\mathbf{n}), \dots, \mathbf{y}_p(\mathbf{n}))$  qui mesure l'indépendance des variables aléatoires  $\mathbf{y}_1(\mathbf{n}), \dots, \mathbf{y}_p(\mathbf{n})$ .

#### **I-4. La séparation aveugle de source**

La séparation aveugle des sources, dite BSS (Blind Source Separation) est un domaine de recherche très intéressant dans le traitement du signal. Les approches de solution de ce problème se décomposent essentiellement en deux grandes parties, statistiques et neuronales [21][28][33]. Ces méthodes sont proposées pour trouver les signaux sources reçus par des capteurs qui reçoivent un mélange de signaux sources. On dit que la séparation est aveugle en raison d'absence d'information propre sur les signaux sources et la nature de matrice de mélange des signaux reçus.

##### **I-4.1 Quelques hypothèses de base**

La SAS consiste à estimer les signaux sources à partir des signaux observés, éventuellement à certaines indéterminations (permutation et échelle) près, comme nous l'avons mentionné précédemment. On parle de la SAS lorsque nous ne disposons d'aucune information a priori ni sur les sources ni sur les mélanges. Cependant pour simplifier et permettre de résoudre le problème, on admet quelques hypothèses.

**Hypothèse 1 :** les signaux sources sont mutuellement statistiquement indépendants. D'un point de vue mathématique, cela signifie que la densité de probabilité conjointe des  $\mathbf{p}$  sources peut se factoriser comme le produit de leurs densités marginales

$$f(s_1(n-m_1), s_2(n-m_2), \dots, s_p(n-m_p)) = \prod_{i=1}^p f_i(s_i(n-m_i))$$

pour tout instant  $n$  et pour tout décalage  $\mathbf{m}_i, i=1, \dots, p$

Cette condition d'indépendance se simplifie dans le cas des mélanges linéaires instantanés en posant  $\mathbf{m}_i=0$ . Cette hypothèse fondamentale est commune à la plupart des méthodes de séparation à l'hypothèse de non corrélation des sources lors de l'utilisation de méthodes de séparation à l'ordre 2 pour des signaux stationnaires colorés ou pour des signaux non-stationnaires.

**hypothèse 2 :** Dans le cas où les signaux sources sont mutuellement statistiquement indépendants et les échantillons de chaque source indépendants et identiquement distribués (i.i.d), on suppose qu'au plus un seul signal source est gaussien dans le cadre de l'hypothèse 2 dans le cas d'un mélange linéaire instantané, les signaux  $\mathbf{y}=\mathbf{B}\mathbf{x}$  sont indépendants si et seulement si la matrice séparant  $\mathbf{B}$  est de la forme :

$$\mathbf{B}=\mathbf{D}\mathbf{P}\mathbf{A}^{-1},$$

Où  $\mathbf{D}$  représente une matrice diagonale et  $\mathbf{P}$  une matrice de permutation. Donc l'indépendance est équivalente à la séparation à un facteur d'échelle et à une permutation près. Dans le cadre de l'hypothèse 2, s'il y a plus d'une source gaussienne, l'indépendance ne conduit pas forcément à la séparation [17]

**Hypothèse 3 :** le nombre des signaux observés est égal au nombre des signaux sources ( $r=p$ ). Cette hypothèse est nécessaire dans la plupart des algorithmes existants. Cependant, certaines méthodes traitent le cas où le nombre d'observations est inférieur au nombre de sources  $r<p$  (mélanges sous déterminés).

**Hypothèse 4 :** la plupart des méthodes de la SAS supposent que les mélanges sont linéaires. Cette hypothèse est largement considérée dans la plupart des méthodes existantes. Cependant certains algorithmes traitent le cas où les mélanges sont non-linéaires [45]. Par ailleurs, d'autres hypothèses supplémentaires peuvent être faites sur les sources. En général ces hypothèses permettent la conception de nouveaux algorithmes.



Soit  $S$  un vecteur d'entrée  $S(s_1, s_2, \dots, s_k)$  avec  $k=1, 2, \dots, n$

Sont  $n$  signaux statiquement indépendants dites sources.

$X$  est le vecteur de sortie  $X(x_1, x_2, \dots, x_l)$  avec  $l=1, 2, \dots, m$

A la matrice de mélange, on a :

$$\mathbf{X} = \mathbf{AS} \quad \text{I.1}$$

La séparation de sources se résume à trouver une matrice  $B$  telle que :

$$\mathbf{Y} = \mathbf{BX} = \mathbf{S} \quad \text{I.2}$$

La figure suivante représente un schéma simplifié pour la séparation aveugle de sources :

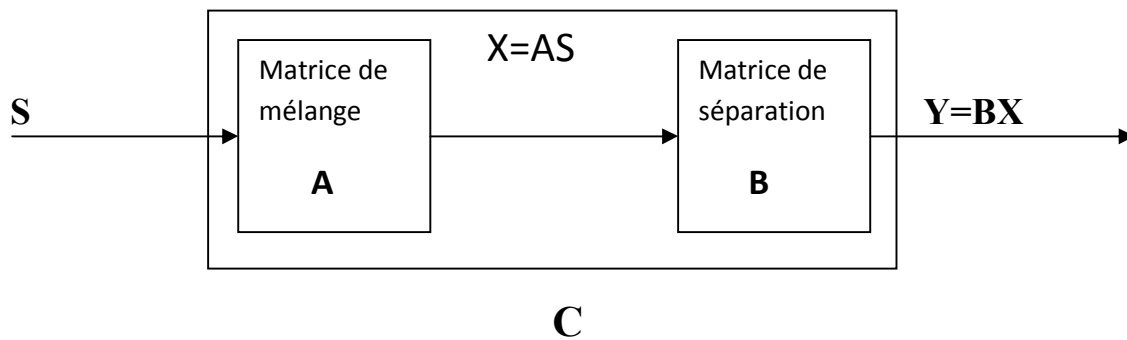


Fig. I.1: Schéma simplifié pour la séparation aveugle de sources.

$$\mathbf{C} = \mathbf{AB} \quad \text{I.3}$$

Pour que la séparation soit idéale, il faut que  $C=I$  (la matrice identité).

Nous rappelons quelques critères d'indépendance entre les signaux :

Un vecteur aléatoire  $x=[x_1, x_2, \dots, x_3]$  avec une densité de probabilité  $p(x_i)$  a des composantes indépendantes si :

$$\mathbf{P(x)} = \prod p_i(x_i) \quad \text{I.4}$$

### I-5. L'Information mutuelle

Soit  $f_x$  et  $g_y$  deux densités de probabilité. La divergence de kulback-lebner de  $f_x(x)$  et  $g_y(x)$  notée  $\delta(f_x/g_y)$  est donnée par [20] :

$$\delta(f_x/g_y) = \int f(x) \log \frac{f_x(x)}{g_y(x)} dx \quad \text{I.5}$$

La divergence de K-L est une grandeur non négative elle s'annule si les vecteurs  $X$  et  $Y$  sont de la même distribution.

Soit  $X$  un vecteur aléatoire de densité de probabilité conjointe  $f_x$  et de densité marginale  $f_{x_i}$  l'information mutuelle est :

$$I(x) = \delta(f_x(x) \prod f_{x_i}(x_i)) = \int f_x(x) \log \left( \frac{f_x(x)}{\prod f_{x_i}(x_i)} \right) dx_1 \dots dx_n \quad \text{I.6}$$

L'information mutuelle est une grandeur positive et elle s'annule si le vecteur  $X$  à des composantes indépendantes.

### I-6. Le critère de maximum de vraisemblance

A partir de l'équation matricielle de mélange I.1, on peut exprimer la densité de probabilité du vecteur en fonction des densités de probabilité des sources et déterminer l'inverse  $B$  de la matrice de mélange  $A$ :

$$P_x(x) = |det B| \prod p_i(b_i x(n)) \quad \text{I.7}$$

La difficulté majeure du critère devient du fait qu'on ne connaît pas la densité de probabilité des sources et des observations [26] .

### I-7. Moments et Cumulants

Plusieurs algorithmes utilisent les statistiques d'ordre supérieur :

#### I.7.1 Première fonction caractéristique [9][26]

Soit  $x$  un vecteur aléatoire de densité de probabilité  $f_x$   
 $\Phi_x(u)$  la valeur moyenne de la fonction  $h(x) = \exp(ju_x^T)$

$$\Phi_{\mathbf{x}}(\mathbf{u}) = E[\exp(\mathbf{j}\mathbf{u}\mathbf{x}^T)] = \int \exp(\mathbf{j}\mathbf{u}\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \quad \text{I.8}$$

### I.7.2 Deuxième fonction caractéristique

$$\Psi_{\mathbf{x}}(\mathbf{u}) = \ln(\Phi_{\mathbf{x}}(\mathbf{u})) \quad \text{I.9}$$

Le moment d'ordre  $k$  est donné par :

$$E(x_1, x_2, \dots, x_k) = (-j)^k \frac{d^k \Phi_{\mathbf{x}}(\mathbf{u})}{du_1 \dots du_k}$$

Le cumulants d'ordre  $k$  est donné par:

$$\text{Cum}(x_1, x_2, \dots, x_k) = (-j)^k \frac{d^k \Psi_{\mathbf{x}}(\mathbf{u})}{du_1 \dots du_k}$$

-Si  $x_i$  sont des variables indépendantes, alors les cumulants croisés d'ordre positif de  $x_i$  sont nuls.

-Si  $x$  est une variable aléatoire gaussien, alors tous ses cumulants d'ordre  $k > 2$  sont nuls.

### I-8. Fonction de contraste

Le concept des fonctions de contraste pour la séparation de sources a été introduit par P. Comon [12]. Une fonction de contraste, qui peut être vue comme une mesure d'indépendance, constitue un critère de séparation dans la mesure où sa maximisation résout le problème de la séparation.

Elle est définie de la façon suivante :

**I-8.1 Définition** une fonction de contraste  $\psi(\cdot)$  est une application à valeurs dans  $\mathbb{R}$  définie sur l'espace de vecteurs aléatoire  $\mathbf{y}$  de  $\mathbb{R}^p$ ; ne dépend que de la loi de probabilité de  $\mathbf{y}$  et qui vérifie les propriétés suivantes :

- ✓ Pour toute matrice de permutation  $P$ ,  $\psi(P\mathbf{y}) = \psi(\mathbf{y})$  ;
- ✓ Pour toute matrice diagonale  $D$ ,  $\psi(D\mathbf{y}) = \psi(\mathbf{y})$  ;
- ✓ Pour tout vecteur  $\mathbf{x}$ , de composantes indépendantes, et pour toute matrice  $S$ , on a :

$$\psi(S\mathbf{y}) \leq \psi(\mathbf{y}) \text{ et } [\psi(S\mathbf{y}) = \psi(\mathbf{y})] \Leftrightarrow [S = DP].$$

Où  $D$  est une matrice diagonale et  $P$  est une permutation quelconques. Les deux premiers points de la définition signifient que l'on ne discrimine pas une solution parmi

l'ensemble des solutions possibles. Ils imposent donc un contraste indépendant des indéterminations inhérentes à la séparation. Le troisième point indique qu'il s'agit de maximiser la fonction ( $S_y$ ) et que tous les maxima globaux sont des solutions de séparation. D'ailleurs, P. Comon [17] a proposé entre autres, comme contraste dans la méthode COM2, la maximisation d'un contraste défini comme la somme des modules au carré des kurtosis des sources estimées. Autres comme E. Moreau et J.C. Pesquet [46] ont introduit une classe de contrastes applicables aux mélanges convolutifs de sources centrées, indépendantes et identiquement distribuées (i.i.d), statistiquement mutuellement indépendantes et vérifiant certaines propriétés. E. Moreau et al. [47][48] ont également défini des contrastes non-symétriques qui permettent de relaxer certaines contraintes des contrastes traditionnels et de pouvoir intégrer de l'information a priori sur les sources. De son côté, P. Comon trouve une solution analytique nommée COM1 (Contraste Maximisation 1)[49] au problème d'optimisation du contraste défini à un signe près, comme la somme des kurtosis des sources estimées. Ce contraste a été initialement présenté par E. Moreau et al. dans [42][50][51]. En outre, E. Moreau montre que ce critère est un contraste à la condition que les kurtosis des sources soient de même signe pour séparer les sources

### I-9. Propriétés de l'indépendance

Si deux sources  $S_i$  et  $S_j$  sont indépendantes alors :

- $S_j$  et  $S_i$  sont aussi indépendantes.
- $K_i S_i$  et  $K_j S_j$ , ou  $K_i$  et  $k_j$  sont deux scalaires quelconques sont indépendants.
- $f_i(S_i)$  et  $f_j(S_j)$  ou  $f_i$  et  $f_j$  sont deux fonctions non linéaires quelconques sont aussi indépendantes.
- $H_i(S_i)$  et  $H_j(S_j)$  ou  $H_i$  et  $H_j$  sont deux filtres quelconques sont indépendants [26].

Selon le milieu de propagation des signaux et les types de capteurs (fonction de transfert de capteur) plusieurs types de mélanges sont exposés:

## **I-10. Les types de mélanges**

### **I-10.1 Mélange linéaire**

#### **I-10.1.1 Instantané**

Mélange sans mémoire dite instantané [12]. ce type de mélange définit la propagation de signal dans un canal de fonction de transfert constante sur sa bande passante, et la fonction de transfert est indéterminée d'ailleurs de cette bande.

Le mélange est donné par l'équation :

$$\mathbf{x}(t)=\mathbf{H}_0\mathbf{S}(t) +\mathbf{b}(t) \quad \text{I.10}$$

#### **I-10.1.2 Convolutif**

Si la transmission dans le canal fait intervenir des phénomènes de propagation, que l'on peut modéliser par des filtres linéaires, les mélanges s'écrivent :

$$\mathbf{X}(t)=\mathbf{H}(t)*\mathbf{S}(t)=\sum\mathbf{h}(k)\mathbf{S}(t-k) \quad \text{I.11}$$

La modélisation de mélanges convolutifs réalistes (par exemple, mélanges de sources audiophoniques dans une salle avec réverbération) [12].

#### **I-10.1.3 Mélange non linéaire**

Ce type de mélange est très général, la relation de mélange est alors de la forme :

$$\mathbf{X}(t)=\mathbf{H}(\mathbf{S}(t)) \quad \text{I.12}$$

Telle que H est une fonction non linéaire.

#### **I-10.1.4 Mélange sous déterminé**

Le nombre de mélanges est moins que celui des sources, dans ce cas pas de séparation sans priori supplémentaire.

#### **I-10.1.5 Mélange sur déterminé**

Le nombre de mélanges est plus grand que celui des sources, on peut faire un pré-traitement par PCA au blanchiment pour projeter dans l'espace du signal avant séparation.

### I-11. Analyse en composantes principales

Le principe de PCA est de trouver des variables décorrélées. L'intérêt de l'analyse en composantes principales réside dans la réduction de dimensions de représentation de donnée ou la projection des données sur un plan représentatif, ce plan est de maximum d'information[9].

### I.12. Blanchiment de données

Un vecteur est dit "blanc" si ses composantes sont décorrélées et de puissance égale à un. Ainsi, pour décorréler les composantes du vecteur  $X(t)$  tout en imposant en plus que leur puissance soit unitaire, on doit chercher une matrice  $W$  dite de "blanchiment" permettant de rendre le vecteur  $Xb(t) = WX(t)$  blanc. En d'autres termes, la matrice de covariance du vecteur  $Xb(t)$  doit être égale à l'identité. On dit que les composantes de  $Xb(t)$  sont blanchies. Cette étape de blanchiment ou de normalisation est nécessaire à de nombreuses méthodes de SAS.

### I-13. Critères de performances de la SAS

Afin d'étudier, par simulation, les qualités de séparation de nos critères et les performances de nos algorithmes, il est possible de mesurer la précision de chaque source estimée  $y_i$  en fonction de la vraie source si par critères de type erreur quadratique moyenne ou rapport signal sur résidus.

#### I-13.1 Erreur quadratique moyenne (EQM)

L'erreur quadratique moyenne (EQM) mesure la moyenne du carré de l'écart entre le signal source  $S_i$  et le signal estimé  $Y_i$ ,  $i=1, \dots, p$  ce terme s'écrit comme suit :

$$EQM_i = \hat{E}\{(s_i - y_i)^2\} := \frac{1}{N} \sum_{n=1}^N [s_i(n) - y_i]^2 \quad I.13$$

Où  $\hat{E}$  désigne la moyenne temporelle et  $N$  représente le nombre d'échantillons utilisés. La valeur moyenne de l'EQM sur toutes les sorties est

$$EQM := \frac{1}{p} \sum_{i=1}^p EQM_i = \frac{1}{p} \sum_{i=1}^p \frac{1}{N} \sum_{n=1}^N [s_i(n) - y_i(n)]^2 \quad I.14$$

La qualité de séparation s'apprécie naturellement avec une EQM la plus petite possible.

### I-13.2 Rapport signal sur résidus (SNR)

Le rapport signal sur résidus (noté SNR) est la mesure de performance la plus répandue dans la séparation de sources. Dans le cas de mélanges instantanés, le rapport signal sur résidus (SNR) est défini comme le logarithme du rapport en décibel (dB) de la puissance de la source  $s_i$  sur celle de l'écart entre la source  $s_i$  et son estimée  $y_i$  : Il s'écrit donc comme suit:

$$SNR_i = 10 \log_{10} \frac{\hat{E}(s_i^2)}{\hat{E}((y_i - s_i)^2)}, i = 1, \dots, p \quad \text{I.15}$$

Pour les mélanges convolutifs, le critère de performance, de l'estimation  $s_i$  par  $y_i$ , est défini comme suit :

$$SNR_i = 10 \log_{10} \frac{\hat{E}(s_i^2)}{\hat{E}(y_i^2 | s_i = 0)}, i = 1, \dots, p \quad \text{I.16}$$

Où  $\langle y_i | s_i = 0 \rangle$  est la sortie  $y_i$  lorsque la source  $s_i$  est rendue nulle (en supposant qu'il n'y a pas de permutation). La qualité de séparation s'apprécie avec une valeur du SNR la plus grande possible. Ce qui signifie qu'il n'y a pas une contribution importante d'autres sources ( $s_j, j=1, \dots, p$  et  $j \neq i$ ) à cette sortie  $y_i$ .

### I-14. Conclusion

Dans ce chapitre, nous avons donné un bref historique sur le problème de la séparation aveugle de sources (SAS) et nous avons ainsi présenté le principe de résolution du problème de la SAS, les différents critères d'indépendance utilisés, les propriétés de l'indépendance dont le choix d'utiliser le meilleur critère compatible et efficace à chaque mélange, et même d'exploiter des combinaisons entre elles, et également les types de mélanges existants.

**CHAPITRE II**

***Réseaux de Neurones***

***Artificiels***



## II-1. Introduction

Le cerveau humain est une machine très compliquée de point de vue résolution des problèmes, l'apprentissage est le cœur du cerveau, qui donne la manière correcte de la réaction, cette fiabilité crée l'idée d'un réseau de neurones artificiels.

## II-2. Bref historique

En 1943, McCulloch et Pitts proposent le modèle du neurone formel inspiré des observations neurophysiologiques et anatomiques du neurone biologique [15].

Les neurones constituent des unités simples (automates linéaires à seuils) mais assemblées sont capables de résoudre des fonctions logiques.

En 1949, Donald Hebb propose une règle d'apprentissage pour les réseaux de neurones artificiels [40].

En 1962, Rosenblatt avec le perceptron et Widrow avec l'adaline proposent un modèle de réseau à une couche et une méthode pour modifier les connexions. Rosenblatt démontre la convergence d'un algorithme itératif d'adaptation des poids pour un perceptron monocouche pour apprendre certaines fonctions logiques [15].

En 1969, Minsky et Papert mettent en évidence les limitations du perceptron. Ils apportent un éclairage théorique sur certains problèmes impossibles à résoudre avec ce modèle, comme le OU exclusif (XOR) [23].

En 1982, Hopfield présente un modèle de réseau de neurones entièrement connecté et doté d'un formalisme plus riche. Hopfield introduit la notion de fonction d'énergie et associe la mémoire à des attracteurs dans les systèmes dynamiques. Par ailleurs, "la machine de Boltzmann" est le premier modèle à lever les limitations du perceptron par la méthode du circuit simulé [19].

Cependant, l'avancée la plus marquante est l'algorithme de rétropropagation découvert tout d'abord par Werbos en 1974, Parker et Le Cun en 1985, puis par Rumelhart en 1986. Cette approche, bien que non biologiquement plausible, lève la limitation du perceptron de Rosenblatt. Cette méthode est destinée à des modèles multicouches statiques (MLP) (sans récurrences) [18].

Depuis, l'algorithme de rétropropagation a été étendu à des réseaux de neurones admettant des connexions récurrentes.

L'algorithme de rétropropagation dynamique permet de modéliser les processus dynamiques tels que les séries temporelles. Il repose sur le principe de minimisation d'une

fonction de l'erreur commise en sortie du réseau et effectue un calcul et une descente du gradient pour trouver le minimum de cette fonction coût. La preuve de la convergence de l'algorithme de rétropropagation dynamique vers un minimum global a été établie.

### II-3. Neurone biologique

Un réseau de neurones artificiels est une structure composée d'entités capables de calcul et interagissent entre eux. Il permet de traiter, par le biais de l'outil informatique, des problèmes de différentes natures par rapport aux outils classiques qui ont du mal à résoudre. En effet, son fonctionnement s'inspire de celui des cellules neuronales animales, et est donc différent des méthodes de calcul analytiques que l'on utilise ordinairement. Il s'avère très puissant dans des problèmes de reconnaissance, de classification, d'approximation ou de prévision.

En biologie, un neurone est une cellule nerveuse dont la fonction est de transmettre un signal électrique dans certaines conditions. Il agit comme un relai entre une couche de neurones et celle qui la suit. Les caractéristiques des neurones sont encore mal connues (et font l'objet de recherches) mais on connaît leur principe d'action.

Le corps d'un neurone est relié d'une part à un ensemble de dendrites (entrées du neurone) et d'autre part à un axone, partie étirée de la cellule, qui représentera pour nous sa sortie. Le neurone étudié est connecté aux neurones qui l'entourent : il reçoit au niveau de ses dendrites les signaux électriques des neurones "en amont", propagés par les axones de ces derniers. Les charges électriques s'accumulent dans le neurone jusqu'à dépasser un certain seuil : à ce moment la transmission du signal électrique se déclenche via son axone vers d'autres neurones "en aval".

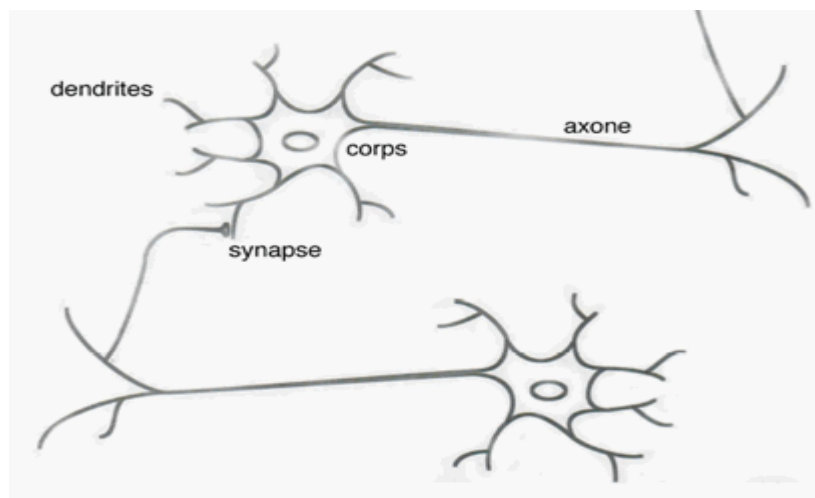


Fig. II.1 Neurone biologique

On remarque que les liaisons axone/dendrite entre deux neurones (connexions synaptiques) ne sont pas toutes de la même efficacité. Ainsi, l'entrée associée à une certaine dendrite du neurone pourra avoir plus d'influence qu'une autre sur la valeur de sortie. On peut représenter la qualité de la liaison par un poids, sorte de coefficient s'appliquant au signal d'entrée. Le poids sera d'autant plus grand que la liaison est bonne. Un poids négatif aura tendance à inhiber une entrée, tandis qu'un poids positif viendra l'accentuer [41][42].

#### II-4. Définition d'un RNA

Les réseaux de neurones artificiels (RNA) sont des réseaux fortement connectés de processeur élémentaire fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseau est évidemment un réseau[18][22][15].

#### II-5. Composition d'un neurone artificiel

##### II-5.1 Structure

Comme l'on a dit dans la définition chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre d'entrées qui est associé à un poids  $W$  (Weight),

Chaque processeur élémentaire est doté d'une sortie unique (figure II.1).

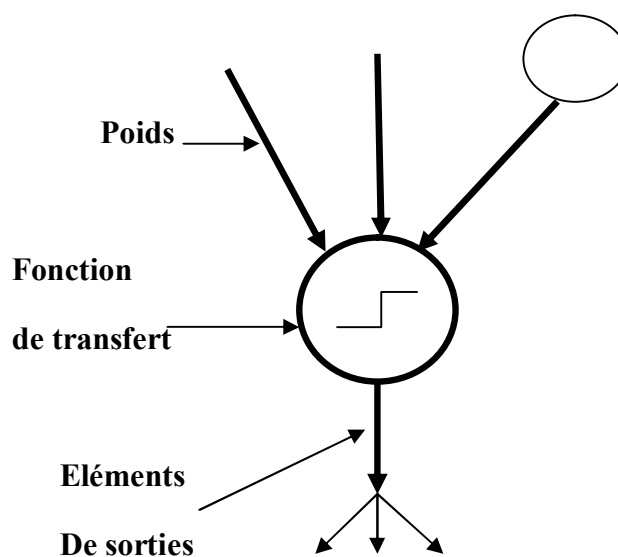


Fig. II.2 Synoptique d'un neurone artificiel

La valeur de sortie est un calcul simple (généralement la somme pondérée des entrées) dont le résultat est soumis à une transformation souvent non linéaire.

$$Y = \sum F(W_i * X_i) \quad \text{II.1}$$

Y : sortie du système.

X<sub>i</sub> : entrée de processeur

W<sub>i</sub> : poids relié à chaque entrée.

F : fonction d'activité ou fonction d'activation.

### II-5.2 Le poids (Weight)

Les poids sont des paramètres adaptatifs dont la valeur varie à chaque étape d'apprentissage pour que le système (réseau) converge vers les valeurs désirées via un algorithme d'apprentissage.

### II-5.3 Les différents types de fonction d'activation

Il existe plusieurs types de fonction d'activation dont les sorties soient linéaires ou non linéaires.

#### II-5.3.1 Fonction binaire

La fonction binaire correspond à un modèle proche de la réalité (figure II.3).

$$s = -1 \quad \text{si} \quad p \leq \theta$$

$$s = 1 \quad \text{si} \quad p \geq \theta$$

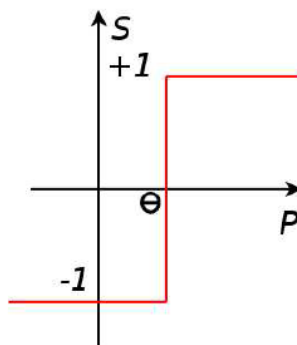


Fig. II.3 Fonction binaire

### II-5.3.2 Fonction linéaire

C'est une fonction dérivable, elle est utilisée pour des algorithmes d'apprentissage performants. (figure II.4)

$$S = AP$$

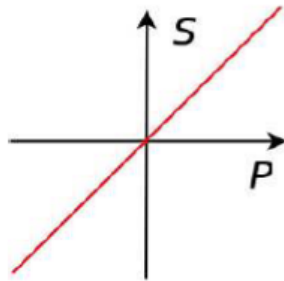


Fig. II.4 Fonction linéaire

### II-5.3.3 Fonction sigmoïde

Ce type de fonction d'activation est très utilisé parce qu'il combine l'avantage de seul et de la dérivabilité. (figure II.5)

$$S = \text{th}(p)$$

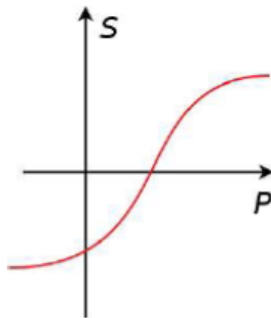


Fig. II.5 Fonction sigmoïde

### II-5.3.4 Fonction gaussienne

Fonction utilisée pour le réseau de neurones RBF (réseau à fonction de base radial).

(figure II.6)

$$S = \exp(-p^2 / \sigma)$$

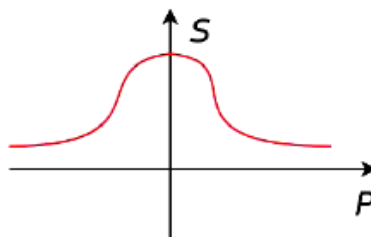


Fig. II.6 Fonction gaussienne

### II-5.4 Structure d'interconnexion[23]

Plusieurs types de réseaux à savoir :

#### II-5.4.1 Réseau multicouche

Il n'y a pas de connexions entre les neurones de même couche et les connexions ne se font qu'avec les neurones des couches suivantes.

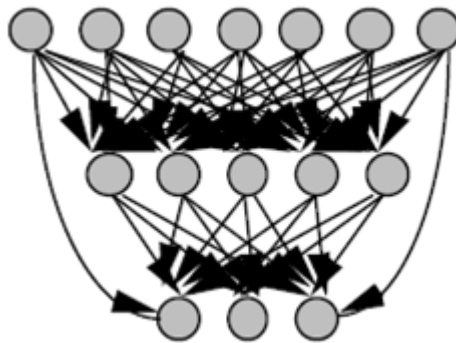


Fig. II.7 Structure multicouche

#### II-5.4.2 Réseau à connexion locale

Chaque neurone entretient des relations avec la seule couche suivante :

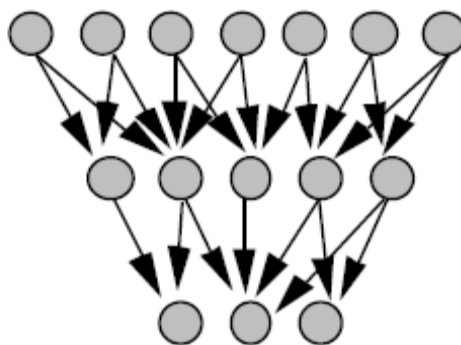


Fig. II.8 Structure à connexion locale

### II-5.4.3 Réseau à connexion récurrentes

L'information ramenée en arrière par rapport au sens de propagation

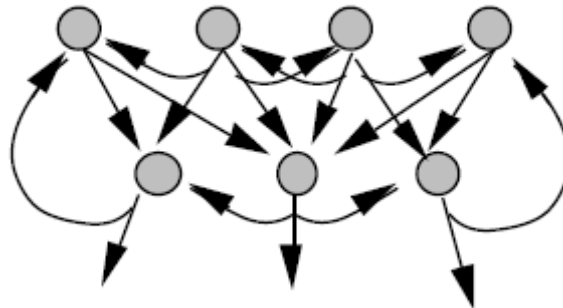


Fig. II.9 Structure récurrente

### II-5.4.4 Réseau à connexion complète

La structure la plus générale, chaque neurone peut se connecter à n'importe quel autre neurone.

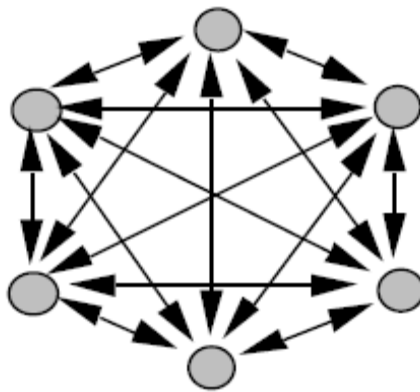


Fig. II.10 Structure à connexion complète

## II-6. Apprentissage des réseaux de neurones artificiels

### II.6.1 Apprentissage supervisé

L'apprentissage supervisé exige l'existence d'un expert qui possède une connaissance de l'environnement [15].

Les réponses désirées fournies par l'expert décrivent la fonction du réseau, le rôle de l'algorithme d'apprentissage est d'amener le réseau à remplir cette équation.

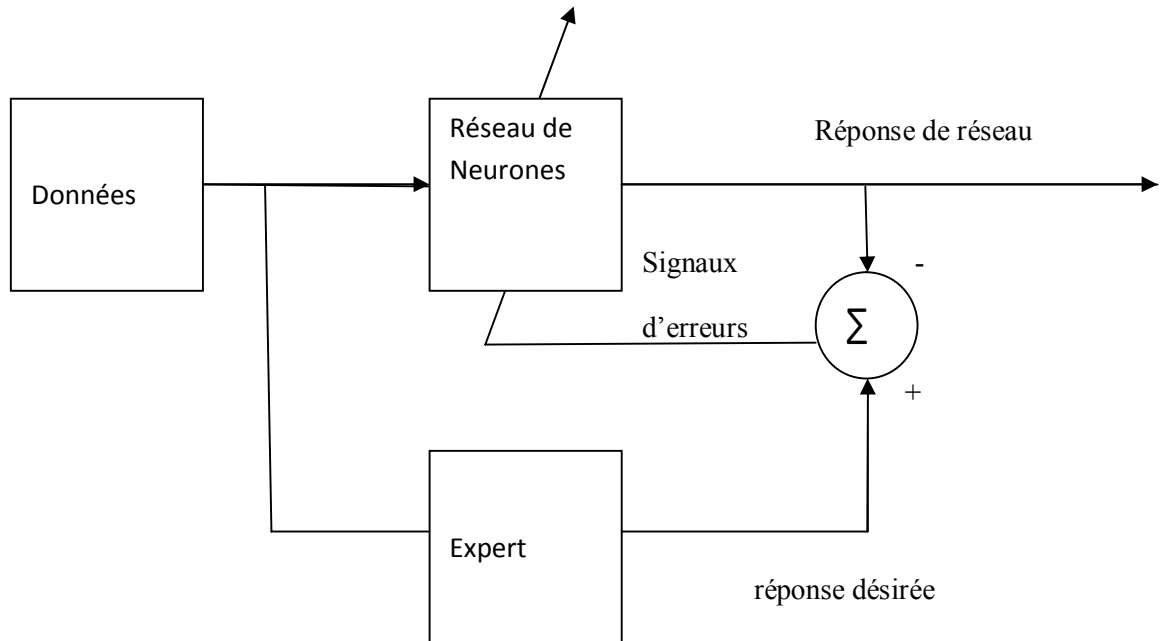


Fig. II.11 Schéma d'apprentissage supervisé

### II-6.2 Apprentissage non supervisé

Aucune réponse désirée n'est associée et l'algorithme d'apprentissage mis à jour les paramètres du réseau [15].

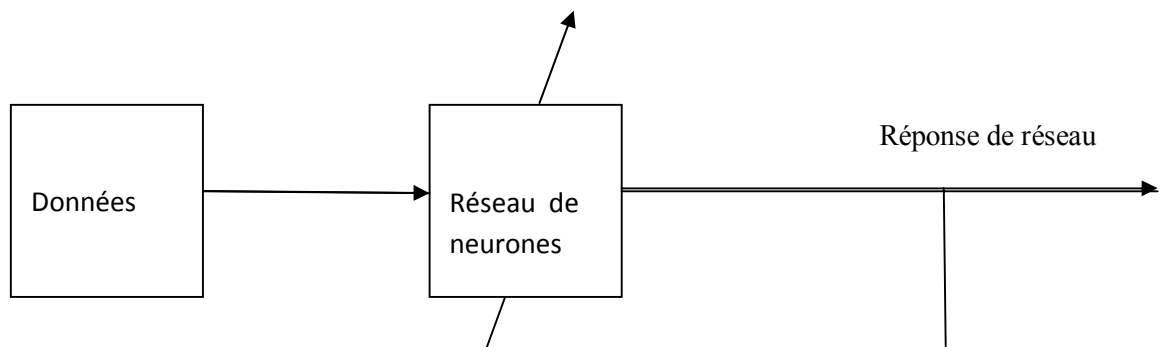


Fig. II.12 Schéma d'apprentissage non supervisée

### II-7. Perceptron

La forme la plus simple d'un réseau de neurones artificiels permet de classer correctement les objets appartenant à deux classes linéairement séparables.



Le perceptron possède  $n$  informations à l'entrée et il est composé de  $p$  neurones. Chacun neurone est connecté aux  $n$  informations d'entrée.

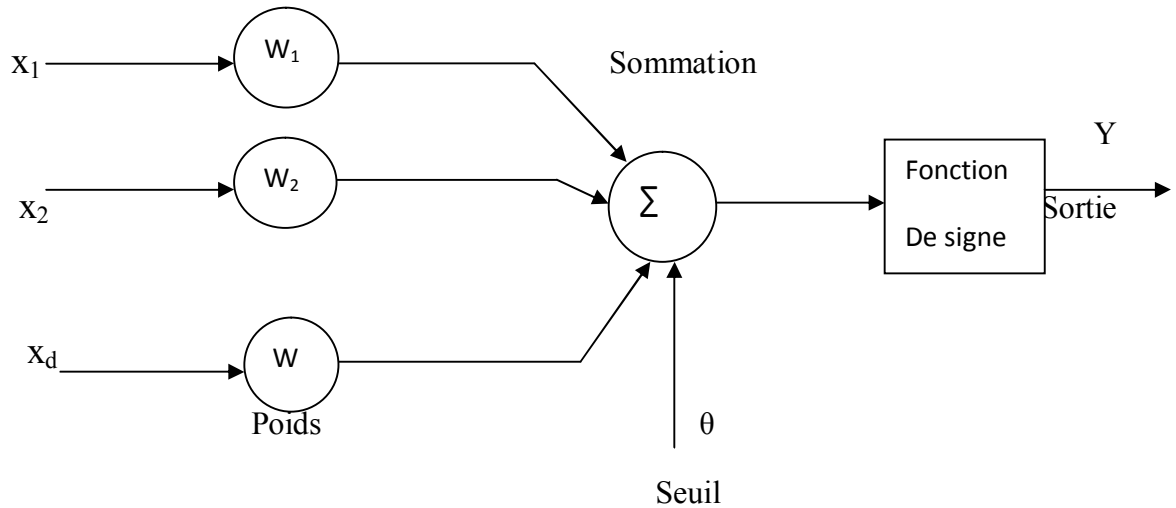


Fig. II.13 Schéma de perceptron

La sortie du réseau peut être donnée par la formule :

$$Y = f(\sum W_i X_i + \Theta) \quad \text{II.2}$$

Il existe deux types de perceptron direct et récurrent. Les perceptrons récurrents sont ceux qui alimentent leurs entrées avec leurs sorties alors que le perceptron direct non.

### II-7.1 Apprentissage du perceptron

La méthode élaborée par Widrow-Hoff[1] consiste à modifier les poids par la minimisation des erreurs de manière précise :

$$W_i(k+1) = W(k) + \Delta w_i \quad \text{II.3}$$

$$\Delta w_i = \mu(t - o) x_i \quad \text{II.4}$$

Où  $\mu$  est le coefficient d'apprentissage.

L'application de cet algorithme à maintes reprises permettra d'affiner la correction des erreurs et d'obtenir un perceptron en plus performant. Mais il existe le risque de surapprentissage « over fitting », c'est à dire le réseau devient très performant sur l'exemple utilisé pour l'apprentissage et pas généraliser pour des informations quelconques.

## II-8. Perceptron multicouche MLP

le MLP (Multi layer Perceptron) est un type de réseau de neurones le plus utilisé. Il est une face générale de perceptron, on peut disposer les neurones en plusieurs couches, ainsi l'information en entrée (Input) acheminée aux neurones de la première couche. Tous les neurones de la 1<sup>ère</sup> couche connectés à tous les neurones de 2<sup>ème</sup> couche, de la même manière jusqu'à la dernière couche, qui est appelée couche de sortie (Output) avec des fonctions d'activation similaire pour toutes les couches cachées sauf la couche de sortie peut appliquer une fonction d'activation différente adaptée au problème[25]..

On applique une fonction d'activation linéaire pour les problèmes de régression et la fonction sigmoïde pour les problèmes de classification.

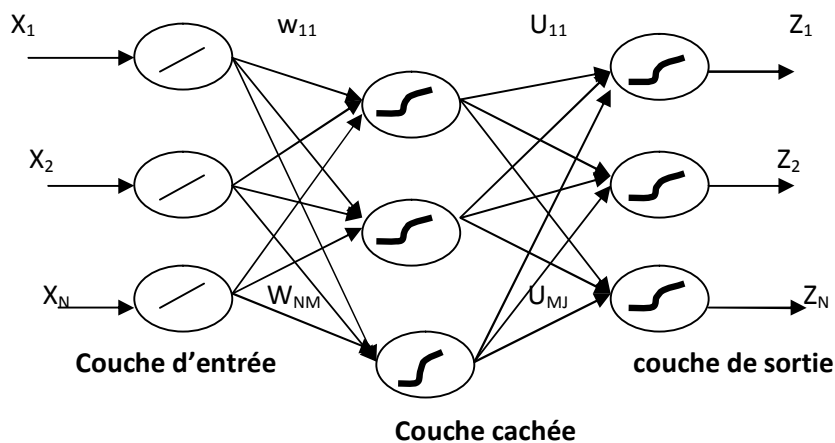


Fig. II.14 Schéma de réseau MLP

$W_{nm}$  : représente les poids reliant la couche d'entrée avec la couche cachée;

$U_{mj}$  : représente les poids reliant la couche cachée avec celle de la sortie.

Les composants du vecteur d'entrée  $X$  du MLP seront pondérés par les poids  $W_{nm}$  et ensuite transmis à la couche cachée selon les équations suivantes :

$$Y_k^{i+1}(\mathbf{x}) = \sum_{i=1}^n w_{ik}^{L+1} y_i^L(\mathbf{x}) \quad \text{II.5}$$

---

### II-8.1 Apprentissage de MPL

Il existe un algorithme permettant de corriger les poids vis-à-vis d'un ensemble d'exemples donnés. Cet algorithme est appelé algorithme de rétropropagation du gradient (Back propagation). Il utilise la même règle de modification des poids que l'algorithme de Window-Hoff[1]. Il est également donné dans sa version plus générale, c'est-à-dire avec plusieurs couches cachées. Une démonstration de l'efficacité de cet algorithme exige de la fonction d'activation qu'elle soit indéfiniment dérivable.

### II-8.2 Problème de minimums locaux

Par fois, on rencontre des problèmes de minimums locaux, le minimum trouvé dépend du point de départ de la recherche (initialisation des poids).

Pour résoudre ce problème, il faut effectuer différentes initialisations pour trouver plusieurs minimisations et retenir la meilleure et même le choix pour le coefficient d'apprentissage qu'il est grand dans le début d'apprentissage et après il décroît.

### II-9. Le réseau à fonction de base radiale RBF

Le réseau à fonction de base radiale RBF (radial basis function) est basé sur une architecture qui s'organise en deux couches seulement à savoir une couche cachée et une couche de sortie.

La couche cachée donne une transformation non linéaire de l'espace d'entrée, la couche de sortie calcule une combinaison linéaire des sorties de la couche cachée. Chaque neurone élémentaire calcule la distance entre l'entrée et le centre qu'il passe ensuite dans une non linéarité concrétisée par une fonction d'activation généralement gaussienne, la valeur d'entrée la plus proche de centre prend une valeur importante.

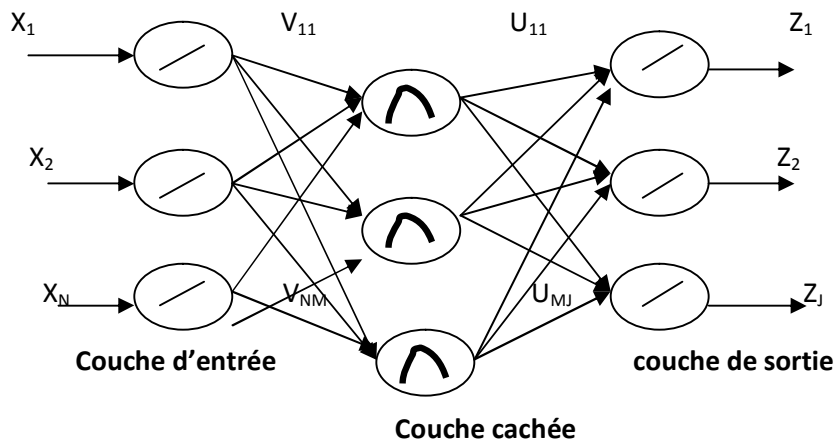


Fig. II.15 Schéma de réseau RBF

Le réseau RBF est constitué de N neurones d'entrée, M neurones cachés et J neurones de sortie.

La sortie du m<sup>ième</sup> neurone de la couche cachée est donnée par :

$$Y_m^{(q)} = \exp(-\|X^{(q)} - V_m\|^2 / (2\sigma_m^2)) \quad \text{II.6}$$

$V_m$  est le centre du m<sup>ième</sup> neurone de la couche cachée où du m<sup>ième</sup> neurone gaussien et  $\sigma_m$  est la largeur du m<sup>ième</sup> gaussienne.

La sortie du j<sup>ième</sup> neurone de la couche de sortie est donnée par:

$$Z_j^{(q)} = (1/M) [\sum W_{mj} Y_m^{(q)}] \quad \text{II.7}$$

$m=1, \dots, M$  et  $j=1, \dots, J$ .

$W_{mj}$  sont des poids reliant la couche cachée à celle de sortie [25][19].

### II-10. Conclusion

Dans ce chapitre on a étudié en détail le concept des réseaux de neurones artificiels qui constituent un outil très puissant dans la modélisation. En effet, les RNA jouissent d'une grande capacité d'apprentissage, d'approximation et de généralisation qui leur a permis de résoudre plusieurs problèmes dans différents domaines. On a donné également étudier les différents types de réseaux, les différentes structures d'interconnexion ainsi que les différentes manières d'apprentissage. En plus la caractéristique essentielle des réseaux de neurones artificiels est qu'ils peuvent capter les dépendances non linéaires de haut niveau entre les

variables explicatives, ce qui est possible grâce à la présence d'une transformation, elle même non linéaire, dans le calcul de la valeur prédite. Les réseaux de neurones sont de puissants outils de modélisation et de prédiction. Ils ont été adoptés dans divers champs d'application plus ou moins variés.

**CHAPITRE III**

**Approches Neuronales**

### III-1. Approche de Herault –Jutten [18]

#### III-1.1 Introduction

On considère dans tous les exemples de simulation un modèle simple de mélanges linéaires et que le nombre de capteurs est égale au nombre de sources.

Sur la figure suivante, on représente le schéma block d'un mélange de sources à travers une matrice de mélange A linéaire :

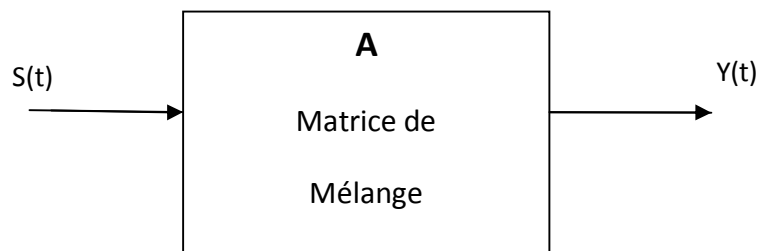


Fig. III.1 Schéma de mélange

$S(t)$  sont signaux sources et  $y(t)$  sont des observations tirées par les capteurs.

J. Herault et C. Jutten [18] dans leur algorithme proposent une solution neuronale pour séparer les sources désirées à partir des observations.

On suppose donc les signaux  $E_i(t)$  de la sortie de n captures

$$y_i(t) = \sum_{j=1}^n a_{ij} s_j(t) \quad 1 \leq i \leq n \quad \text{III.1}$$

Ou sous la forme matricielle :

$$Y(t) = AS(t) \quad \text{III.2}$$

- A : matrice de mélange
- S : sources indépendantes.

#### III-1.2 Architecture du réseau

La structure du réseau est un modèle récurrent. Pour extraire une source parmi les n sources présentées dans le mélange  $E_i(t)$ , il faut annuler (n-1) sources par des soustractions pondérées (fig. III.2)

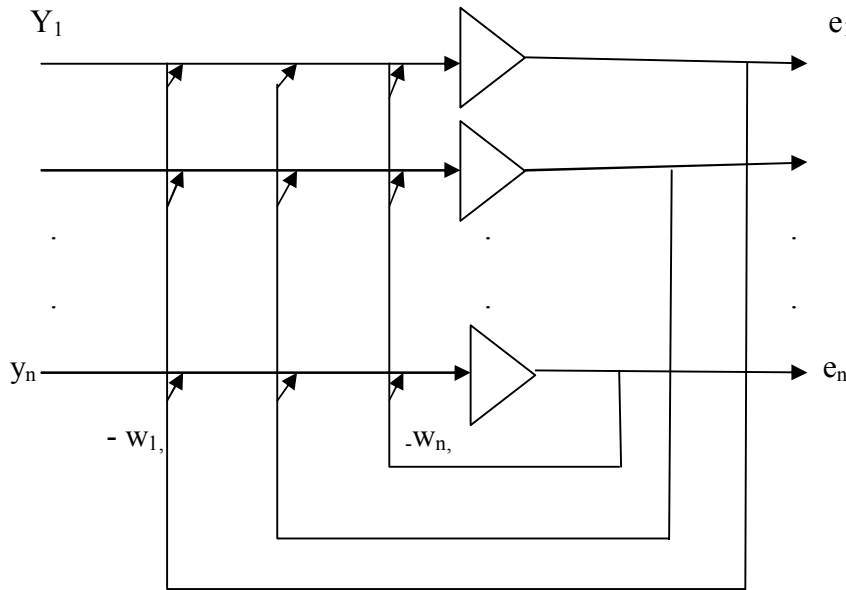


Fig. III.2 Architecture de réseau

On peut écrire la sortie de réseau comme suit :

$$E_i(t) = y_i(t) - \sum_{i \neq j}^n w_{ij} e_j(t) \tag{III.3}$$

$$E_n(t) = \sum (a_{nk} - w_{nk} a_{kk}) S_k(t) + a_{nn} S_n(t) \tag{III.4}$$

Pour les sources \$S(t)\$ indépendantes on peut écrire :

$$\langle E_n^2(t) \rangle = \sum (a_{nk} - w_{nk} a_{kk})^2 \langle S_k^2(t) \rangle + a_{nn}^2 \langle S_n^2(t) \rangle$$

La \$n^{\text{ème}}\$ sortie tend vers \$S\_n(t)\$ quand les coefficients \$(a\_{nk} - w\_{nk} a\_{kk})\$ sont minimisés.

### III-1.3 La méthode du gradient

La dérivée partielle de \$E\_n^2(t)\$ par rapport \$w\_{ij}\$ est :

$$\frac{\partial E^2}{\partial w} = 2E(t) \partial E(t) / \partial w \tag{III.5}$$

Le signal \$E(t)\$ à moyenne nulle :

$$e(t) = (I - C)^{-1} s(t)$$

$$\frac{\partial E}{\partial w} = -(I + W)^{-1} \partial(I + W) / \partial w \times (I + W)^{-1} s(t)$$



Pour les composants  $E_m(t)$  :

$$\frac{\partial E_m}{\partial W} = -q_{mi} E_k(t)$$

Et :

$$\frac{\partial E^2}{\partial w} = -2q_{mi} E_m(t)E_k(t)$$

L'application de la méthode de gradient donne la règle d'adaptation suivante :

$$\frac{dw}{dt} = aqE_i(t)E_k(t) \quad \text{III.6}$$

### III-1.4 Réseau direct (Feedforward)

Dans ce cas, le calcul devient plus simple et le schéma suivant présente la structure du réseau direct dans le cas de deux sources [18]:

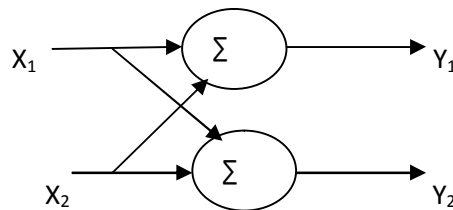


Fig. III.3 Réseau direct

La sortie  $e(t)$  du système est exprimée par :

$$E(t) = (I - W) y(t)$$

$$E(t) = (I - W) AS(t)$$

On a :

$$e_1(t) = y_1(t) - W_{21}y_2(t)$$

$$e_2(t) = y_2(t) - W_{12}y_1(t)$$

Donc :

$$e_1(t) = (a_{11} - w_{21}a_{21}) S_1 + (a_{12} - w_{12}a_{22}) S_2 \quad \text{III.7}$$

$$e_2(t) = (a_{21} - w_{12}a_{21}) S_1 + (a_{22} - w_{21}a_{12}) S_2 \quad \text{III.8}$$

Il y a deux solutions :

#### III-1.4.1 1<sup>ère</sup> solution:

$$W_{12} = a_{11}/a_{21} \quad e_1(t) = (a_{12} - a_{11}a_{22}/a_{21}) S_2$$

$$W_{21} = a_{22}/a_{12} \quad e_2(t) = (a_{21} - a_{22}a_{11}/a_{12}) S_1$$

#### III.1.4.2 2<sup>ème</sup> solution:

$$W_{12} = a_{12}/a_{22} \quad e_1(t) = (a_{11} - a_{12}a_{12}/a_{22}) S_1$$

$$W_{21} = a_{21}/a_{11} \quad e_2(t) = (a_{22} - a_{21}a_{12}/a_{11}) S_2$$

Dans le cas où les coefficients de la matrice de mélange sont proches la stabilité du système (convergence) est perturbée.

### III-1.5 Réseau récurrent (feedback)

Considérons le réseau du Herault-Jutten présenté par la figure ci dessous :

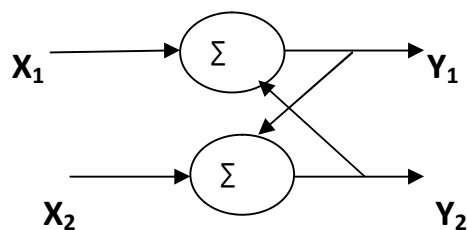


Fig. III.4 Réseau récurrent

Les signaux de sortie de réseau s'écrivent sous la forme :

$$e_1(t) = y_1(t) - w_{12}e_2(t)$$

$$e_2(t) = y_2(t) - w_{21}e_1(t)$$

Les poids de réseau  $w_{12}$  et  $w_{21}$  sont adaptés de manière à ce que les signaux de sortie proportionnels à ceux des signaux sources (annulation de l'influence du mélange).

$$e_1(t) = \frac{1}{1-w_{12}w_{21}}((a_{11} - w_{12}a_{21})S_1 + (a_{12} - w_{12}a_{22})S_2) \quad \text{III.9}$$

$$e_2(t) = \frac{1}{1-w_{12}w_{21}}((a_{21} - w_{21}a_{11})S_1 + (a_{22} - w_{21}a_{12})S_2) \quad \text{III.10}$$

### III-1.6. La règle d'adaptation

La règle d'adaptation proposée par Herault-Jutten[1] est donnée par la forme suivante :

$$\frac{dw}{dt} = \mu f(s_1(t))g(s_2(t)) \quad \text{III.11}$$

$\mu$  : gain d'adaptation positif.

$f$  et  $g$  sont deux fonctions non linéaires d'activation.

### III-1.7 Rôle de la non-linéarité

Supposons que les fonctions impaires  $f$  et  $g$  soient développables en séries au voisinage de 0 :

$$f(s_i) = \sum f_{2k+1} \times s_i^{2k+1}$$

$$g(s_i) = \sum g_{2l+1} \times s_i^{2l+1}$$

$$dw_{ij} = a \sum f_{2k+1} g_{2l+1} s_i^{2k+1} s_i^{2l+1}$$

on aura une stabilisation des poids, si l'incrément d'adaptation est en moyenne nulle, c'est-à-dire si :

$$\langle dw_{ij} \rangle = a \sum f_{2k+1} g_{2l+1} \langle s_i^{2k+1} \cdot s_i^{2l+1} \rangle$$

C'est-à-dire, si tous les moments croisés  $\langle s_i^{2k+1} \cdot s_i^{2l+1} \rangle$  sont nuls, et non pas seulement quand la covariance  $\langle s_i \cdot s_j \rangle$  est nulle. Les fonctions  $f$  et  $g$  introduisent ainsi des moments croisés d'ordre élevé, et la règle peut être considérée comme une approximation d'un test d'indépendance.

Pour ce calcul, nous avons fait l'hypothèse de l'existence d'un développement en série de MacLaurin de  $f$  et  $g$ . Cette condition n'est pas nécessaire, et l'on peut le vérifier expérimentalement. Par exemple, la fonction  $g(\cdot) = \text{signe}(\cdot)$ , qui n'est pas développable en série au voisinage de 0, a été testée avec succès[33][18].

## III-2. Approche de Moreau - Macchi[42].

### III-2.1 Introduction

Cette approche est une combinaison entre deux structures directes et récurrentes.

L'algorithme suppose les hypothèses suivantes :

- 1- Les sources non gaussiennes.
- 2- Les sources sont à moyenne nulle.
- 3- Les sources sont mutuellement indépendantes.
- 4- Les mélanges sur déterministes.

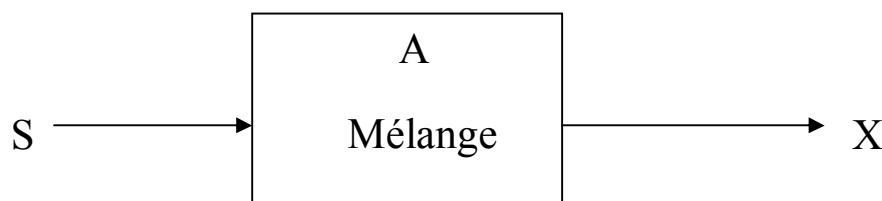


Fig. III.5 Signaux mélangés

Pour extraire les sources premières  $S$ , on a le système (réseau de séparation) suivant :

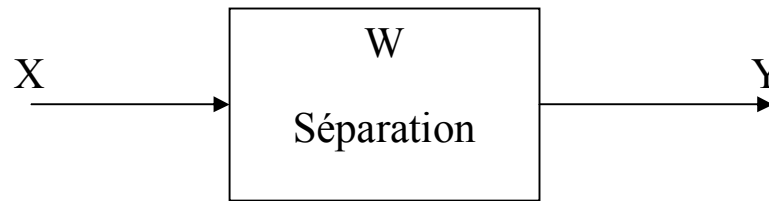


Fig. III.6 Réseau Séparateur

### III-2.1.1 Réseau récurrent (Feed back) :

La figure présente le cas de réseau récurrent (feed back) avec deux sources, la sortie d'un neurone est une entrée de l'autre [18][42].

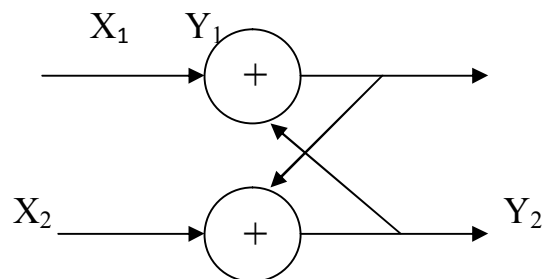


Fig. III.7 Réseau récurrent

Il y a deux solutions :

$$\begin{cases} w_{12} = -a_{12}/a_{22} \\ w_{21} = -a_{21}/a_{11} \end{cases} \quad \begin{cases} w_{12} = -a_{11}/a_{21} \\ w_{21} = -a_{22}/a_{12} \end{cases}$$

Et la sortie sera :

$$\begin{cases} y_1 = a_{11} S_1 \\ y_2 = a_{22} S_2 \end{cases} \quad \begin{cases} y_1 = a_{12} S_2 \\ y_2 = a_{21} S_1 \end{cases}$$

### III-2.1.2 Réseau direct (Feed forward)

Comme il est présenté sur la figure si dessous,

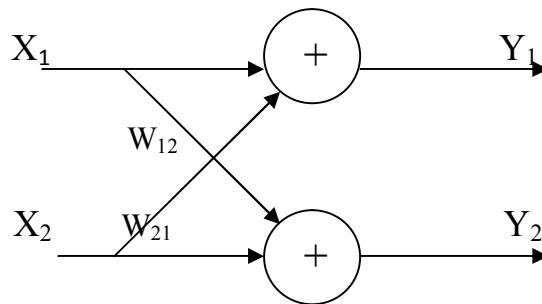


Fig.III.8 Réseau direct

$$Y = (I + w)X \quad \text{III.12}$$

Les solutions sont les mêmes avec l'exemple 1I (Herault\_Jutten)

### III-2.2 Loi d'adaptation

Même loi avec l'algorithme I (herault\_jutten) :

$$w(n) = w(n-1) + \Delta w(n) \quad \text{III.13}$$

$$\Delta w_{ij} = -\mu f(y_i) g(y_j) \quad \text{III.14}$$

$\mu$ : coefficient d'apprentissage positif.

$f, g$ : Deux fonctions d'activation non linéaires.

### III-2.3 Réseau mixte

L'idée de cet algorithme est de créer un ensemble avec les deux structures participantes direct et récurrent pour l'amélioration des performances des signaux de sortie (sources estimés) [42].

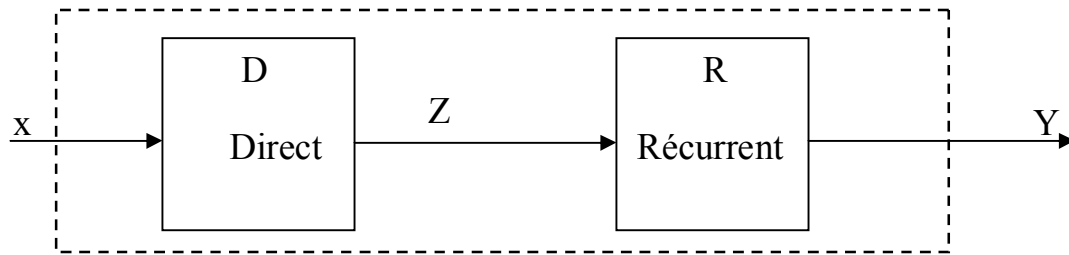


Fig.III.9 Réseau mixte

$$Y_1 = X_1 + m_{12}(X_2 + Y_2)$$

$$Y_2 = X_2 + m_{21}(X_1 + Y_1)$$

On a :

$$\frac{1}{1 - m_{12}m_{21}}(I + M^2) = k'(I + D)$$

$$m_{12} = \frac{1 \pm \sqrt{\Delta(D)}}{d_{21}}$$

$$m_{21} = \frac{1 \pm \sqrt{\Delta(D)}}{d_{12}}$$

Avec :  $\Delta(D) = 1 - d_{12}d_{21}$

$$k' = \frac{\pm 1}{\sqrt{\Delta(D)}}$$

### III-2.4 Solution

Le vecteur M obtenu par l'insertion de  $D = R_1$  et  $R_2$  dans l'équation précédente, il y a trois possibilités :

$\Delta(R_1)$  positive et  $\Delta(R_2)$  négative :

$$M = -g_{22}g_{11}(1 - \sqrt{\Delta(D)})U(G)$$

$\Delta(R_1)$  négative et  $\Delta(R_2)$  positive :

$$M = -g_{12}g_{21}(1 - \sqrt{\Delta(D)})U(G)$$

$\Delta(\mathbf{R}_1)$  et  $\Delta(\mathbf{R}_2)$  sont positives, il y a quatre paires de M selon les égalités précédentes

$$U(\mathbf{G}) = \left( \frac{\mathbf{1}}{g_{21}g_{22}}, \frac{\mathbf{1}}{g_{11}g_{12}} \right)$$

### III-2.5 Loi d'adaptation de réseau mixte :

La structure mixte est une cascade de deux systèmes identiques, et la loi d'adaptation du système est donnée par :

$$\Delta m_{ij} = -\mu y_i^3 y_j \quad \text{III.15} \quad \text{pour } i \neq j$$

Cette nouvelle architecture combine en cascade les deux anciennes structures avec la même loi d'adaptation de l'algorithme de Herault-Jutten. Dans le cas où les deux structures directe et récurrente ont de même paramètres, le réseau mixte propose deux ou bien quatre possibilités de séparer les mélanges, c'est le cas des deux structures avec des paramètres différents, la convergence du réseau direct est plus rapide que le réseau récurrent mais le réseau mixte converge plus rapide par rapport aux deux et avec une meilleure performance.

## III-3. Approche de Clemente-Acha

### III-3.1 Introduction

Cet Algorithme [29] est un développement de l'algorithme de base de Hearlt-Jutten. La stabilité de l'algorithme est reliée directement au kurtosis (signe). Si le kurtosis[8] est de même signe la stabilité est assurée.

### III-3.2 Hypothèse

$S_1(t)$  et  $S_2(t)$  sont :

- Sources stationnaires
- De moyennes nulles
- Non gaussiennes.
- De variance unitaire.

$Y_1(t)$  et  $y_2(t)$  sont deux sources obtenues par le mélange de  $S_1(t)$  et  $S_2(t)$  par la matrice A  
 $Y=AS$ .



$w_1(t)$  et  $w_2(t)$  deux signaux obtenus par la decorrélation des signaux d'observation  $y_1(t)$  et  $y_2(t)$ , est une étape de blanchiment des signaux par la matrice  $w$  :

$$W=A \cdot X ; \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix}$$

$W$  trouvé par la variation de la matrices de corrélation

$$R_y=[y \cdot y]$$

L'estimation des sources  $S_1$  et  $S_2$  par la combinaison de la décorrélation des signaux  $w(t)$  (blanchiment) et la matrice de séparation  $B$  dans un réseau à structure directe (feed forward structure)

$$E= B \cdot W ; \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} = \begin{bmatrix} 1 & b_{21} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & b_{21} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix}$$

La séparation est assurée quand  $B$  est estimée

$$A \cdot B = P \cdot D ,$$

$P$  matrice de permutation et  $D$  matrice diagonale.

### III-3.3 Les solutions

Deux solutions proposées :

$$\left\{ \begin{array}{l} b_{12} = -a \\ b_{12} = 1/a \end{array} \right. \rightarrow \begin{array}{l} \rightarrow b_{21} = a \\ b_{21} = -1/a \end{array}$$

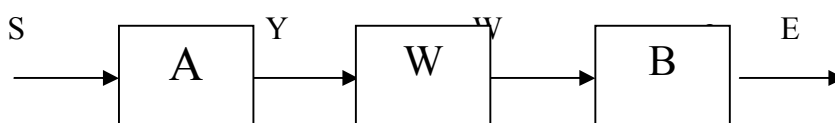


Fig. III.10 Structure directe

$$C_{mki}(w_1(t), w_2(t)) = C_{ki}$$

Et le kurtosis de source  $x_i(n)$  est :

$$K_{xi} = E[S_i^4(n)] - 3E^2[S_i(n)] \quad \text{III.16}$$

Il y a :

$$C_{13} = -a K_{S1} + a^3 k_{S2}$$

$$C_{22} = a^2(k_{S1} + K_{S2})$$

$$C_{w1} = C_{40} = K_{S1} + a^4 K_{S2}$$

### III-3.4 Nouveau critère de séparation

La règle d'adaptation sera la suivante :

$$\mathbf{B}(n+1) = \mathbf{b}(n) + \mu[\mathbf{e}_1(n) \cdot \mathbf{e}_2(n)]^T \quad \mu > 0 \quad \text{III.17}$$

Tel que  $r_1(n)$  et  $r_2(n)$  sont défini par :

$$r_1 = e_1 e_2 w_2^2 - b_{21} \delta w_1^2 \delta w_2^2 - 3b_{12} (\delta^2 w_2)^2$$

$$r_2 = e_1 e_2 w_1^2 w_2^2 - b_{12} \delta w_1^2 \delta w_2^2 - 3b_{21} (\delta^2 w_1)^2$$

$\delta w_1^2 = E[w_1^2]$ , car  $e_1(n)$  et  $e_2(n)$  sont une combinaison linéaire de  $w_1(n)$  et  $w_2(n)$  obtenue par :

$$E[r_1] = b_{12} k w_2 + b_{12} c_{22} + (1 b_{12} b_{21}) c_{13}$$

$$E[r_2] = b_{21} k w_1 + b_{12} c_{22} + (1 b_{12} b_{21}) c_{31}$$

On estime  $\delta^2 w_1$  et  $\delta^2 w_2$  d'abord et après on fait l'adaptation.

### III-4. Approche de Cichocki-Amari

#### III-4.1 Modification de l'algorithme de H-J

La modification a été appliquée sur le réseau proposée par Herault-Jutten pour résoudre les problèmes que l'ancien algorithme n'a pas pu résoudre.

$$X_i(t) = \sum h_{ij} s_j(t) \quad \text{III.18}$$

$H_{ij}$  : paramètre de mélange.

$S_i$  : source première, de moyenne nulle.

Mathématiquement l'indépendance de  $Y_i$  et  $Y_j$  est décrite par la relation suivante :

$$P(Y_i, Y_j) = P(Y_i)P(Y_j) \quad \text{III.19}$$

Pour des signaux indépendants avec des moyennes nulles, la matrice de covariance des variables  $f(Y_i)$  et  $g(Y_i)$  est donnée par la formule suivante :

$$R_{fg} = E[f(y)g^T(t)] - E[f(t)]E[g(t)] \quad \text{III.20}$$

Avec  $f$  et  $g$  sont des fonctions d'activation non linéaires.

Le critère d'adaptation est donné par H-J :

$$\frac{d w}{d t} = \mu(t) f(y_i) g(y_j(t)) \quad \text{III.21}$$

Non efficace dans plusieurs cas (matrice de mélange mal conditionnée,...)

L'idée de cet algorithme est de normaliser la variance  $E[f(y_i)g(y_i)]$  (égale à l'unité), S A D le signal de sortie  $y_i(t)$  est adapté :

$$E[f(y_i)g(y_i)] = \lambda_i$$

#### III-4.2 Formalisation du problème

Cet algorithme est nécessaire pour développer un réseau de neurones avec une loi d'adaptation pour résoudre les problèmes des signaux mélangés à grande différence d'échelle et lorsque la matrice de mélange est mal conditionnée.

## III-4.2.1 Algorithme direct (Feedforward)

$$Y_i(t) = \sum W_{ij}(t) x_i(t)$$

$$Y(t) = W(t) x(t)$$

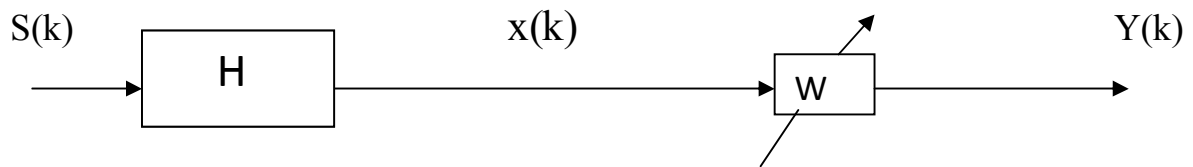


Fig. III.11 Réseau direct

$W(t)=[w_{ij}(t)] \in \mathbb{R}^{n^*n}$  matrice d'adaptation ajustable .

$Y(t)$  : source destinée.

$X(t)$  : observation.

$$W(k+1) = w(k) + \mu(k)[1 - f(y(k))g^T(y(k))] \quad \text{III.22}$$

Avec  $w(0) \neq 0$

$\mu(k)$  décrémente vers des petites valeurs pour la stabilité de l'algorithme.

III-4.2.2 Réseau récurrent (feedback)

Le réseau récurrent présenté sur la figure III.9 est plus efficace[33].

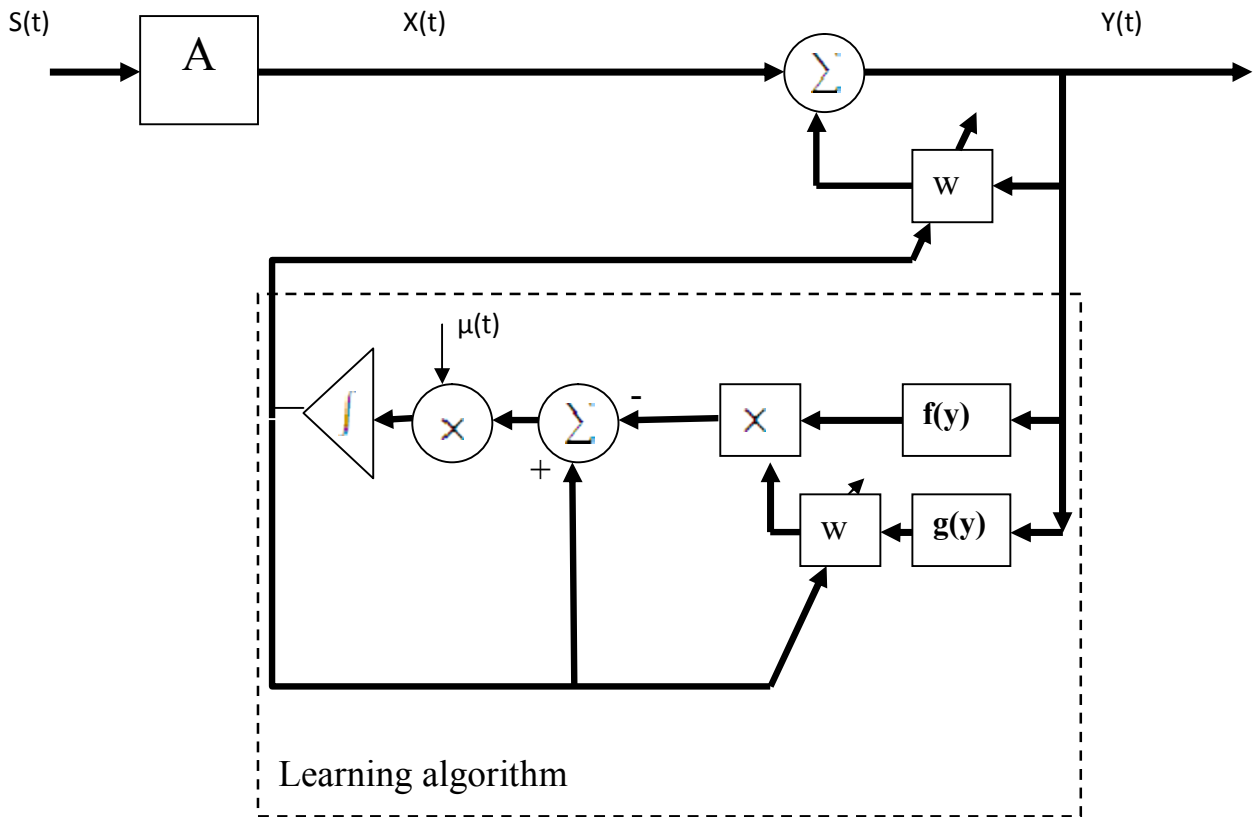


Fig. III.12 Réseau récurrent

$$W(k+1) = w(k) + \mu(k)[I+w(k)][1-f(y(k))g^T(y(k))] \quad \text{III.23}$$

III-4.3 Extension de la base d'apprentissage

Il y a plusieurs modifications et extensions de la base d'apprentissage de l'algorithme, par exemple :

$$\frac{dw}{dt} = -\mu(t)(w + I)G(y(t))$$

Avec  $G(y)$  prend les formes suivantes :

$$G_1(y) = I - f(y)g(y)$$

$$G_2(y) = \Lambda - yy^T - f(y)g^T(y) + f^*(y)g(y)$$

$$G_3(y) = \Lambda - f(y)g^T(y) - \text{diag}[yy^T - f(y)g(y)]$$

$$G_4(y) = \Lambda - \sum y(t)y(t - KT)$$

Le choix de la fonction  $G(y)$  dépend de la forme des signaux.

### III-5. Approche de Cardoso-Laheld

#### III-5.1 Introduction

La séparation des sources consiste à construire un modèle qui a la possibilité de trouver  $n$  signaux indépendants à partir de  $m$  signaux d'observations.

Dans cette partie nous étudions l'approche « Equivariant adaptive sources séparation » proposée par Cardoso-Laheld [28]

$$X_i = AS_i \quad i=1,2,\dots, n$$

$X_i$  et  $S_i$  sont les colonnes de taille  $m, n$  respectivement, et  $A$  est de  $m \times n$  dite matrice de mélange, l'idée est de trouver une matrice  $B_t$  de  $m \times n$  telle que :

$$Y_t = B_t X_t$$

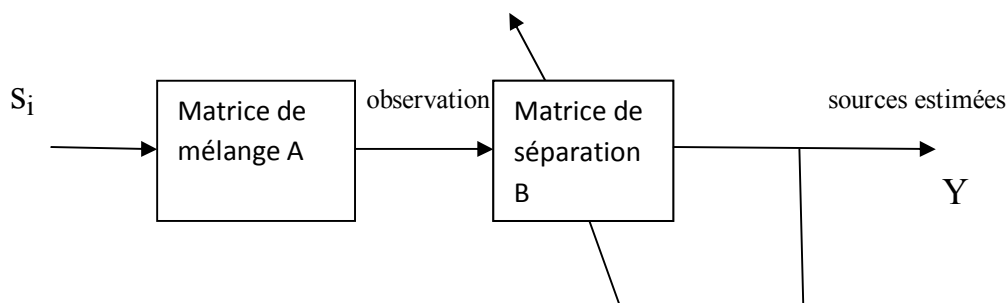


Fig. III.13 Adaptation de la matrice de séparation

On considère le système global  $C_i$  obtenu par une série de transferts de signaux (matrice de mélange, matrice de séparation)

$$C_i = B_i A$$

La convergence du système est basée sur la variation de la matrice de séparation  $B_i$  :

$$B_i A = I$$

$C_t$  converge vers matrice d'identité  $I$  (séparation idéale).

### III-5.2 Série d'adaptation

Le point essentiel de ce travail est d'exploiter un algorithme « série d'adaptation » qui a le rôle de donner le développement de la matrice  $B$  avec une manière correcte.

$$B_{t+1} = B_t + \lambda H(y_t) B_t \quad \text{III.24}$$

La séparation faite en série d'adaptations de deux étapes  $B = WU$ , la première étape est d'adapter la matrice de blanchiment, et la deuxième sur une matrice  $U$  dite matrice d'orthogonalité.

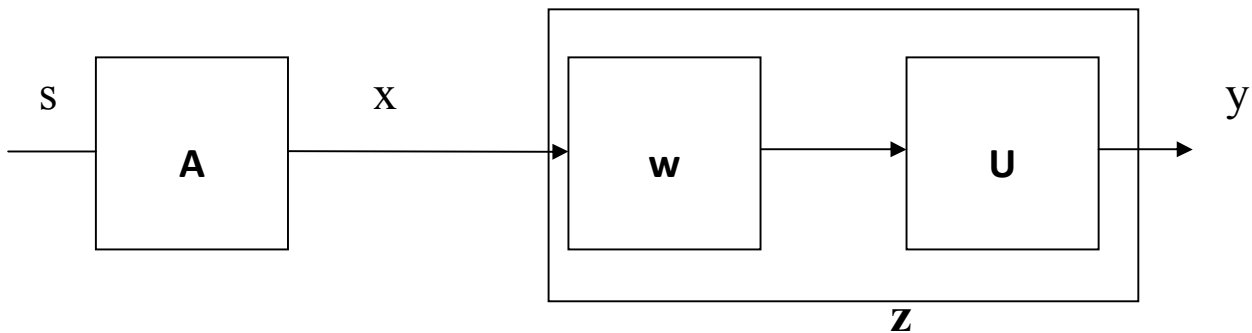


fig. III.14 Adaptation en deux étapes

### III-5.3 L'adaptation de la matrice de blanchiment

L'adaptation de la matrice  $W$  converge  $RY = EYY^T = I$  est obtenue par la minimisation de la distance entre  $RZ$  et  $I$ . La divergence de kulback-leibner entre  $RZ$  et  $I$  donnée par :

$$K(RZ) = \frac{1}{2} [\text{trace}(RZ) - \log \det(RZ) - n]$$

La propriété de la divergence mesurée est :  $K(RZ) > 0$ , pour avoir  $K(RZ) = 0$ , est minimisée  $\phi_2(w)$ , telle que :

$$\phi_2(w) = K(WR_Z W)$$

Le gradient relatif de la fonction  $\phi_2$  est exprimé par :

$$\nabla \phi_2 = R_Z - I = E[Z_t Z_t^T - I]$$

La règle d'adaptation de la matrice de blanchiment est :

$$W_{k+1} = W_k - \lambda [Z_t Z_t^T - I] W_k \quad \text{III.25}$$

### III-5.4 L'adaptation de la matrice U

Pour la minimisation de la fonction caractéristique, l'adaptation de la matrice U est donnée par l'équation suivante :

$$U_{k+1} = U_k - \lambda_k [f'(y_k) y_k^T - y_k f'(y_k)^T] u_k \quad \text{III.26}$$

### III-5.5 Hypothèses

- 1- A matrice inversable  $n \leq m$
- 2- Les composants de  $S_t$  de moyenne nulle.
- 3- Les composants de  $S_t$  sont mutuellement indépendants.
- 4- Les composants de  $S_t$  sont à des variances unitaires.

L'hypothèse 3 est la clé de la séparation aveugle de sources (absence d'information priori), l'hypothèse 4 propose une normalisation de variance :

$$R_s = E[S_t S_t^T] = I$$

### III-5.6 Gradient relatif

$\Phi(B)$  fonction d'objective de la matrice B, le gradient de  $\phi$  dans le point B :  $\frac{\partial \phi}{\partial B}(B)$

$$\Phi(B + \xi B) = \phi(B) + \langle \nabla \phi(B) / \xi \rangle + o(\xi)$$

$$f(y + \nabla y) = f(y) + f'(y) \nabla y + o(\nabla y)$$

$f'(y)$  est le gradient de f en y

$$\Phi(B + \xi B) = E f((B + \xi b)X) = E f(y + \xi y)$$

$$= E f(y) + E f'(y)^T \xi y + o(\xi)$$

$$= \phi(y) + \langle E f'(y) y^T / \xi \rangle + o(\xi)$$



Il y a :  $\nabla\phi(B) = \nabla E f(y) = \nabla E f(BX) = E(f'(y)y^T)$

Pour des petites valeurs positives de  $\lambda$  ( $\nabla\phi(B) \neq 0$ ),  $\phi$  est diminué, quand B est modifié vers  $B - \lambda \nabla\phi(B)$  la valeur de :

$$B_{t+1} = B_t - \lambda_t f'(y_t)y_t^T B_t \quad \text{III.27}$$

La minimisation de la fonction  $E f(y)$  pour des valeurs  $\lambda_t$  positives.

### III-5.7 L'algorithme E A S I :

Le point de stationnarité de l'algorithme EASI (Equivariant Adaptive source Independence) est de trouver une matrice B qui assure  $E H(y) = 0$ , ceci est décomposé en deux points :

$$E(yy^T) = I \quad (1)$$

$$E[f'(y)y^T - y f'(y)^T] = 0 \quad (2)$$

La 1<sup>ère</sup> équation normalise la variance et elle n'est pas suffisante pour la séparation, les composants de y sont mutuellement indépendants si :

$$E[y_i f'_j(y_i)] = E y_i E f'_j(y_j) = 0.$$

Avec des signaux de moyenne nulle, l'adaptation de B est donnée par :

$$B_{t+1} = B_t + \lambda_t [y_t y_t^T - I + g(y_t)y_t^T + y_t g(y_t)^T] B_t \quad \text{III.28}$$

$g(.)$  est une fonction non linéaire.

### III-6. Conclusion

La méthode de Héarult-Jutten basée sur les techniques neuronales constitue la première approche pour solutionner le problème de la séparation de sources, et comme cette technique souffre de quelques inconvénients notamment dans le cas de plusieurs sources à séparer ou lorsque la matrice de mélange est mal conditionné par exemple, l'algorithme ne converge pas, c'est dans optique que plusieurs chercheurs ont penché leur nez pour proposer et développer d'autres méthodes pour remédier à ces obstacles.

Ce chapitre a donné un aperçu général sur les algorithmes de séparation basés sur les approches neuronales appelées aussi approches connexionnistes avec les différentes lois d'adaptation.

# **CHAPITRE IV**

## ***Discutions et Interprétation des Résultats***

## IV-1. Exemples de Simulation

### IV-1.1 Exemple 1

Dans cet exemple, les signaux sources sont présentés par la figure IV.1

La source 1 est un signal déterministe et la source 2 est un bruit gaussien.

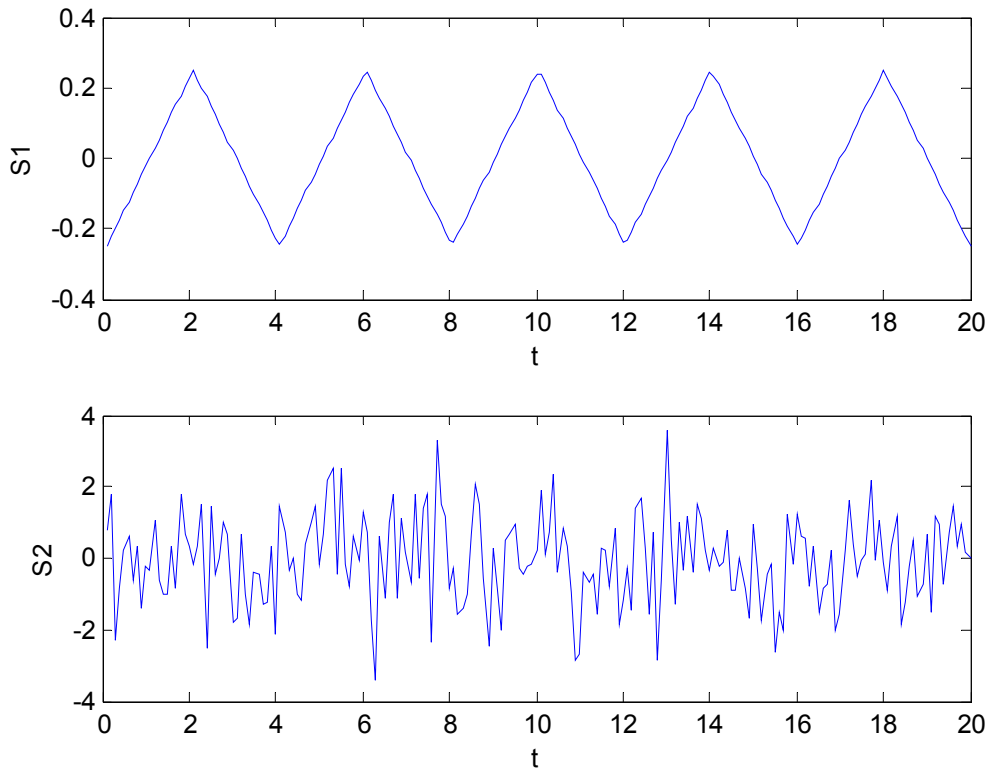


Fig IV.1 Les signaux sources

La matrice de mélange est donnée par :

$$A \begin{pmatrix} 0.4 & 0.7 \\ 0.2 & 0.5 \end{pmatrix}$$

Les observations (mélanges) reçues par capteurs sont présentées par la figure IV.2.

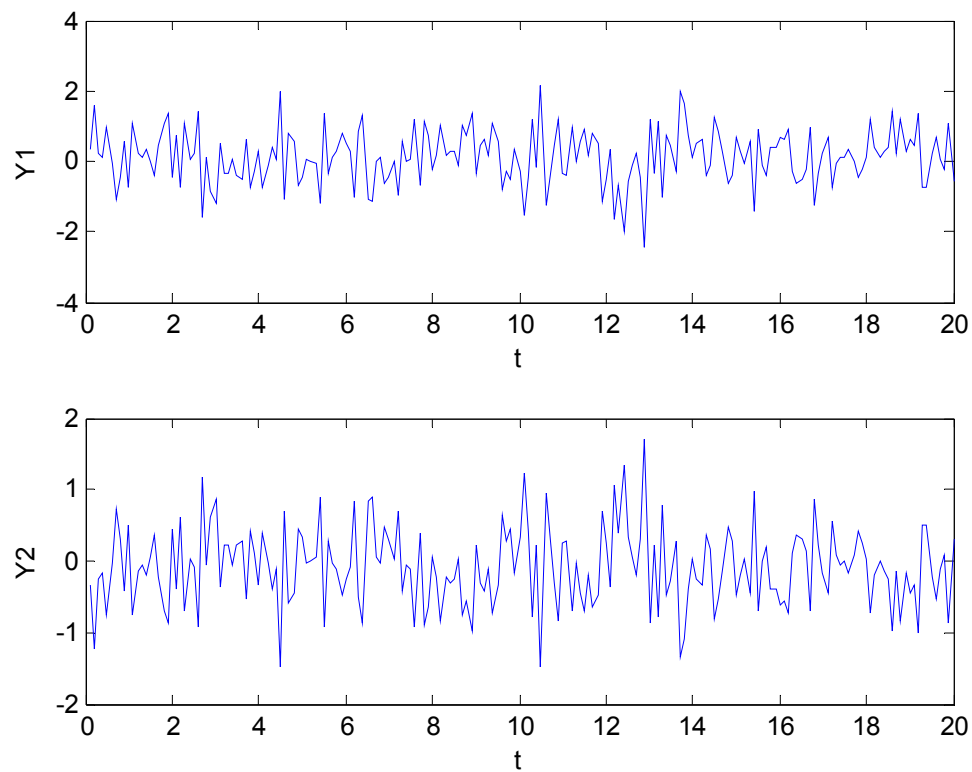


Fig IV.2 : Les signaux mélanges

On applique l'algorithme H-J avec :

- Les fonctions d'activations suivantes :

$$f(y) = y^3$$

$$g(y) = \sin(y)$$

- $\mu=0.01$  (gain d'adaptation ou pas d'apprentissage)
- $W_{ij}(0)=0$ (les poids initialisés a zéro)

Les signaux estimés sont présentées par la figure IV.3.

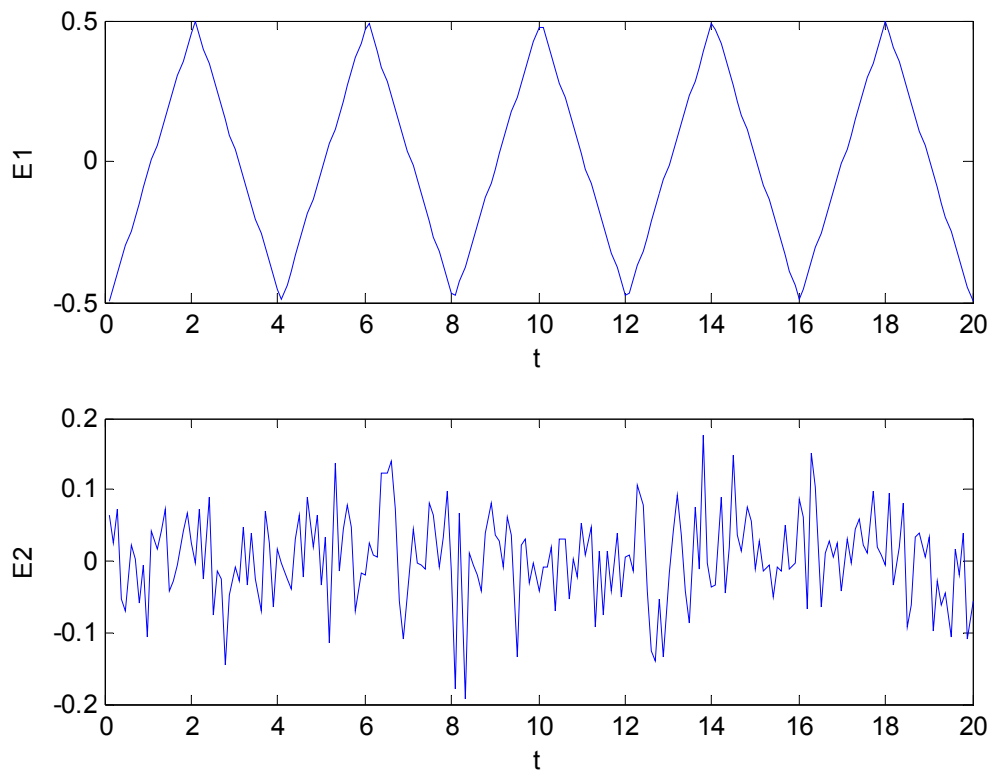


Fig IV.3 Les signaux estimés

On remarque que la vitesse de convergence de l'algorithme H-J pour ce cas d'exemple varie en fonction des paramètres initiaux de poids  $W_{ij}(0)$ . Cet algorithme ne converge pas dans le cas de matrice de mélange mal conditionnée ou avec des signaux sources à différentes échelles.

## IV-1.2 Exemple 2

Les signaux sources sont visualisés par la figure IV.4 :

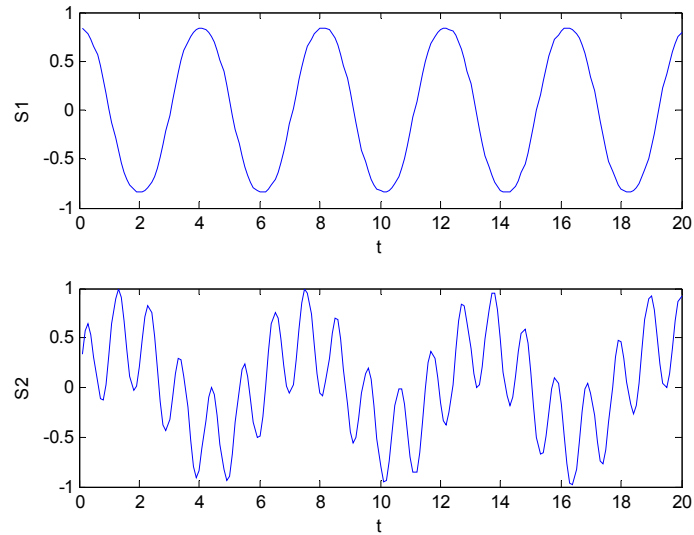


Fig IV.4 Les signaux sources

La matrice de mélange est donnée par :

$$A = \begin{pmatrix} 1 & 1.6 \\ 1.3 & 1 \end{pmatrix}$$

Les observations (mélanges) reçues par capteurs sont présentées par la figure IV.5 :

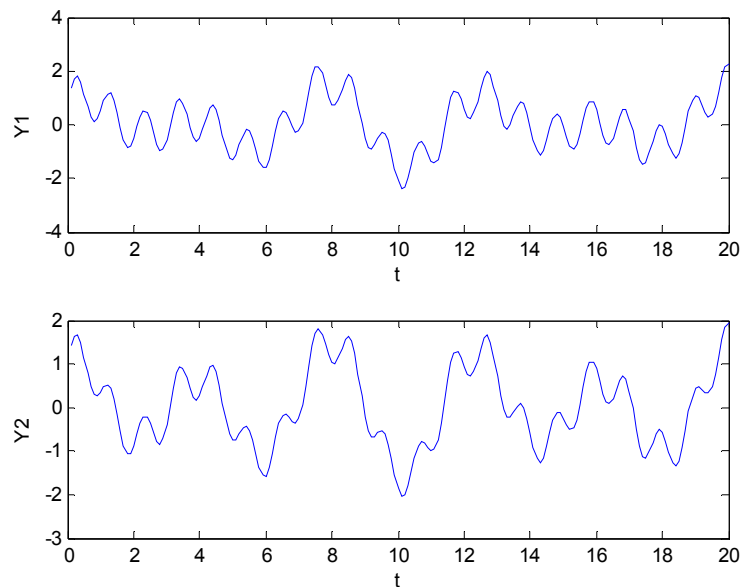


Fig IV.5 : Les signaux mélanges

On applique l'algorithme de Moreau-Macchi avec :

- Les fonctions d'activations suivantes :

$$f(y) = y^3$$

$$g(y) = y$$

- $\mu=0.001$  (gain d'adaptation ou pas d'apprentissage)

Les signaux estimés sont présentés par la figure IV.6 :

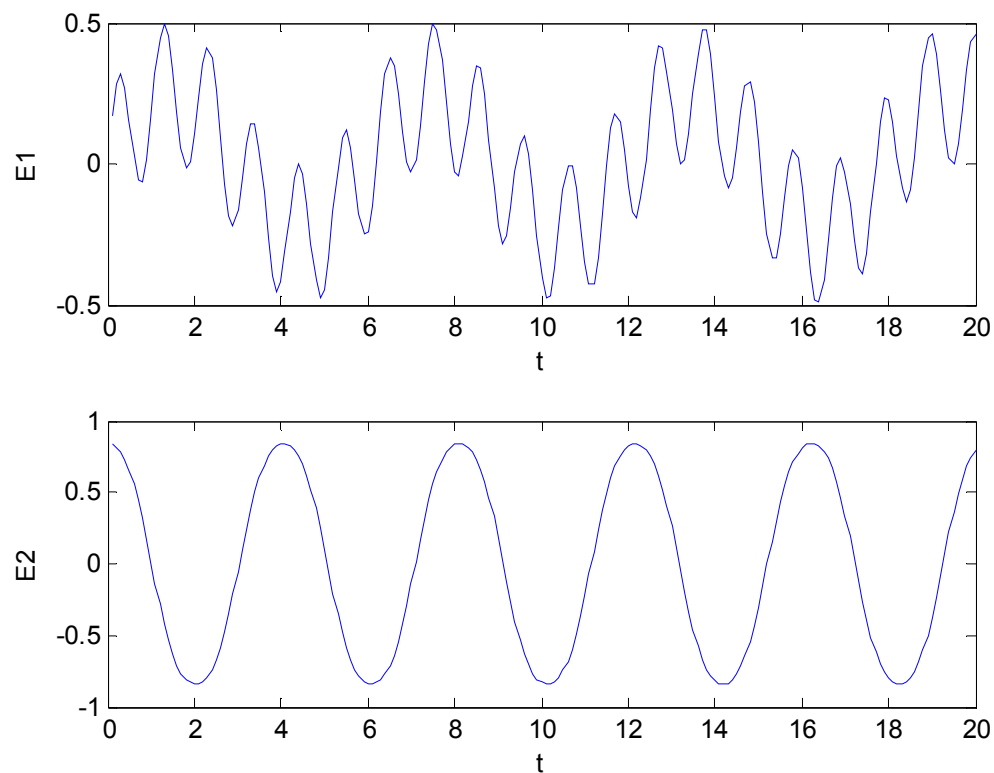


Fig IV.6 Les signaux Estimés

Cet algorithme donne des bons résultats, il converge après 200 itérations, à cause de son architecture, avec une erreur  $3.691 \cdot 10^{-4}$  pour la première source estimée et  $2.712 \cdot 10^{-4}$  pour la deuxième source estimée.



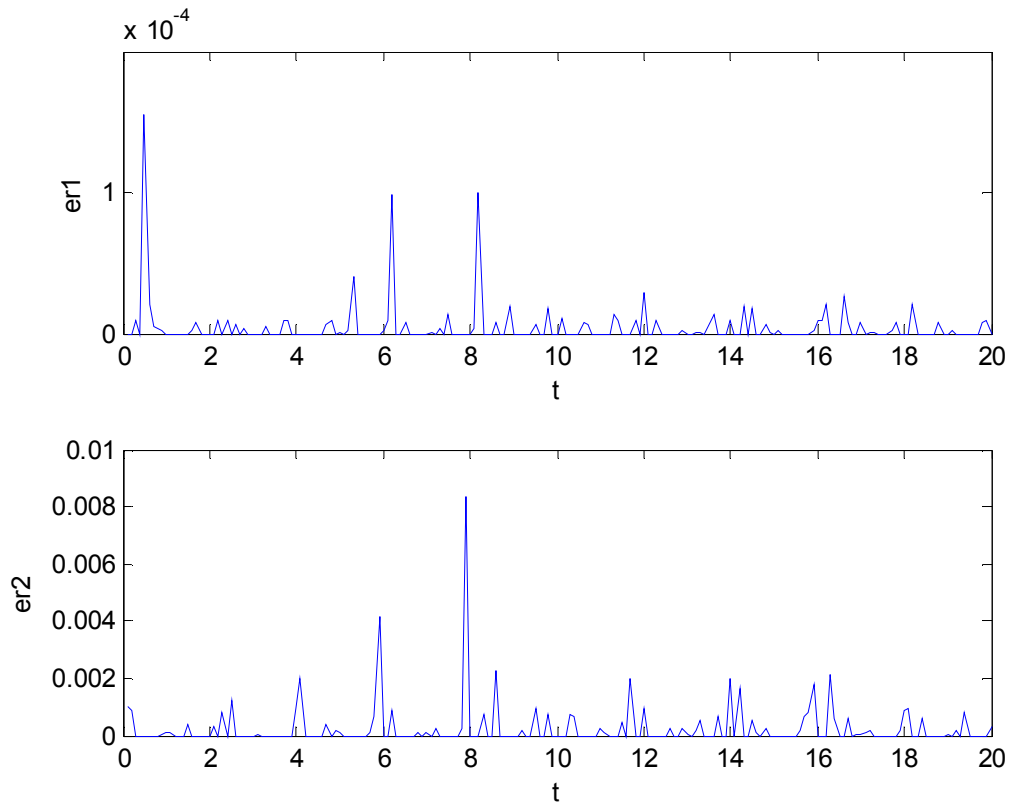


Fig IV.7 Les erreurs de séparation

### IV-1.3 Exemple 3

On a les signaux sources présentés par la figure suivante :

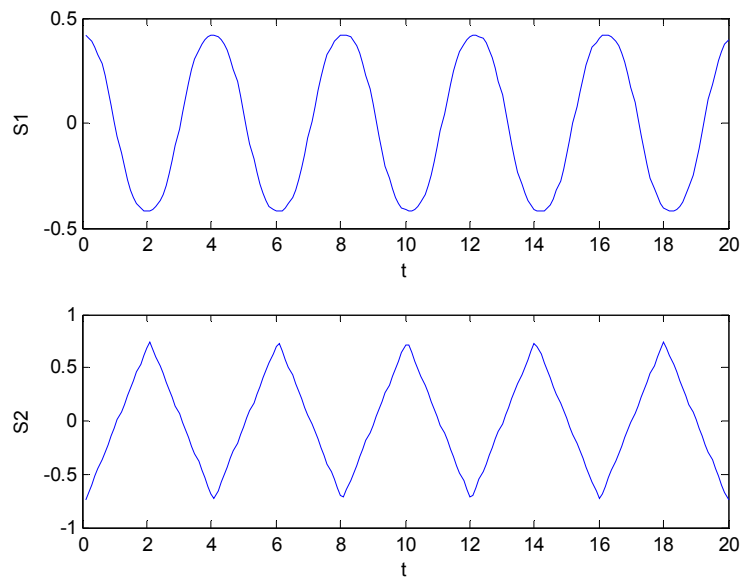


Fig IV.8 Les signaux sources

La matrice de mélange est donnée par :

$$A = \begin{pmatrix} 0.973 & 0.546 \\ 0.221 & 0.476 \end{pmatrix}$$

Les observations (mélanges) reçues par capteurs sont présentées par la figure IV.9 :

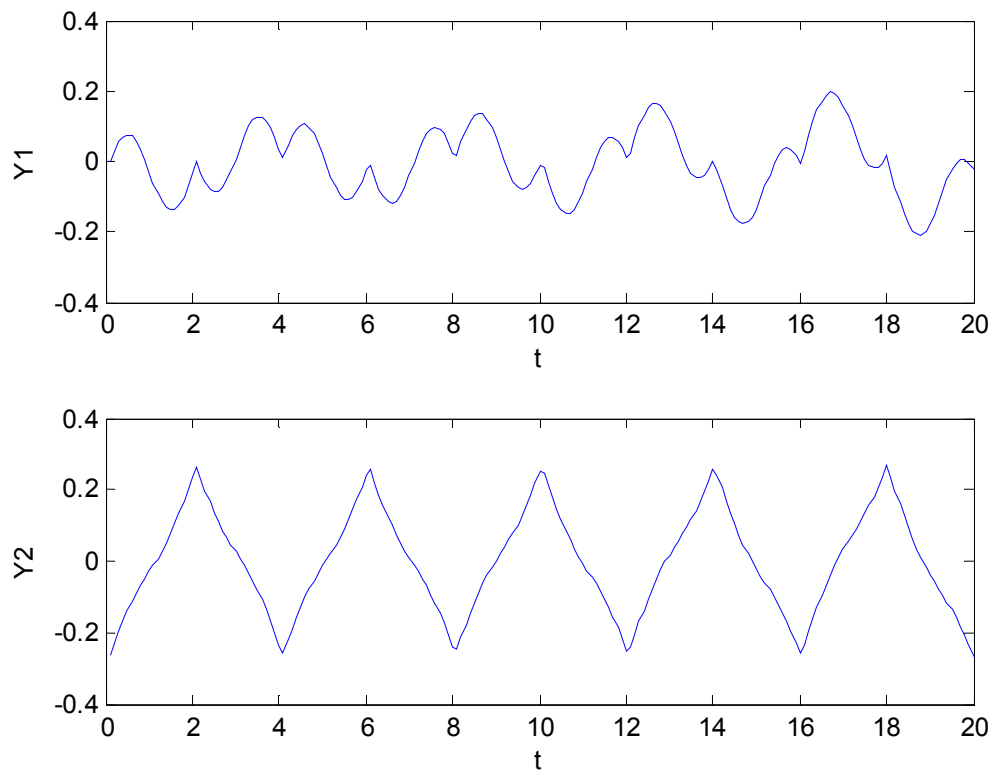


Fig IV.9 : Les signaux mélanges

Après  $\approx 1500$  itérations, la matrice de blanchiment a les valeurs suivantes :

$$A = \begin{pmatrix} 1 & -1.666 \\ 1.666 & 1 \end{pmatrix}$$

La matrice de séparation d'après l'exemple 3 prend les valeurs suivantes :

$$B = \begin{pmatrix} 1 & -0.6 \\ 0.6 & 1 \end{pmatrix}$$

Les signaux estimés sont présentés par la figure IV.10 :

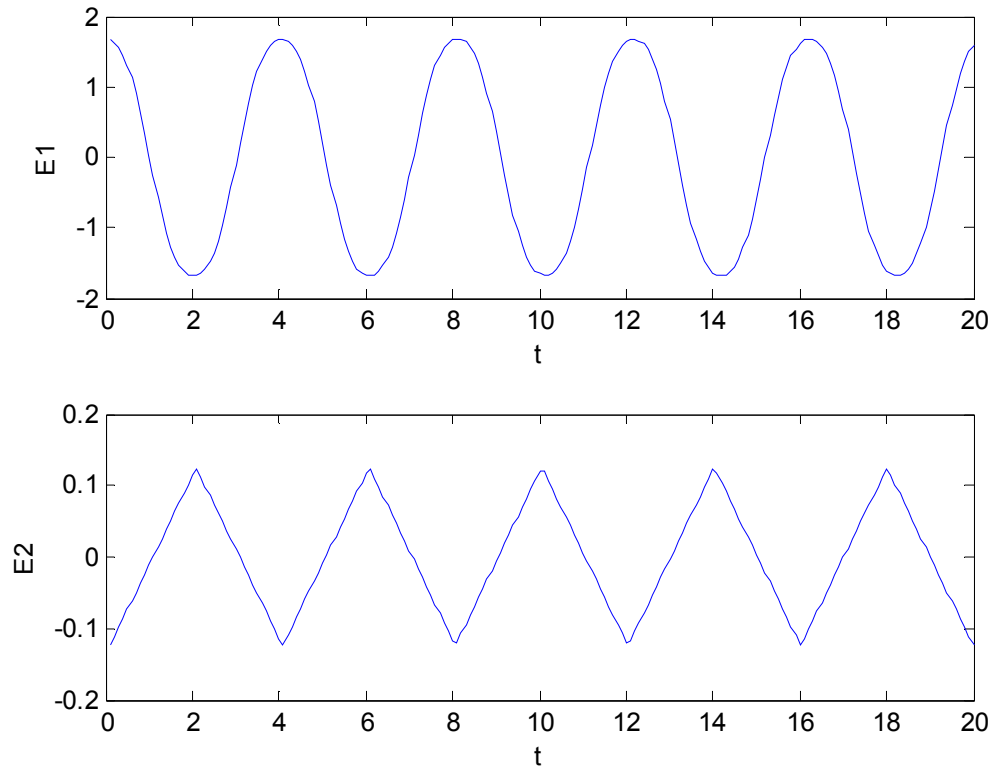


Fig IV.10 Les signaux Estimés

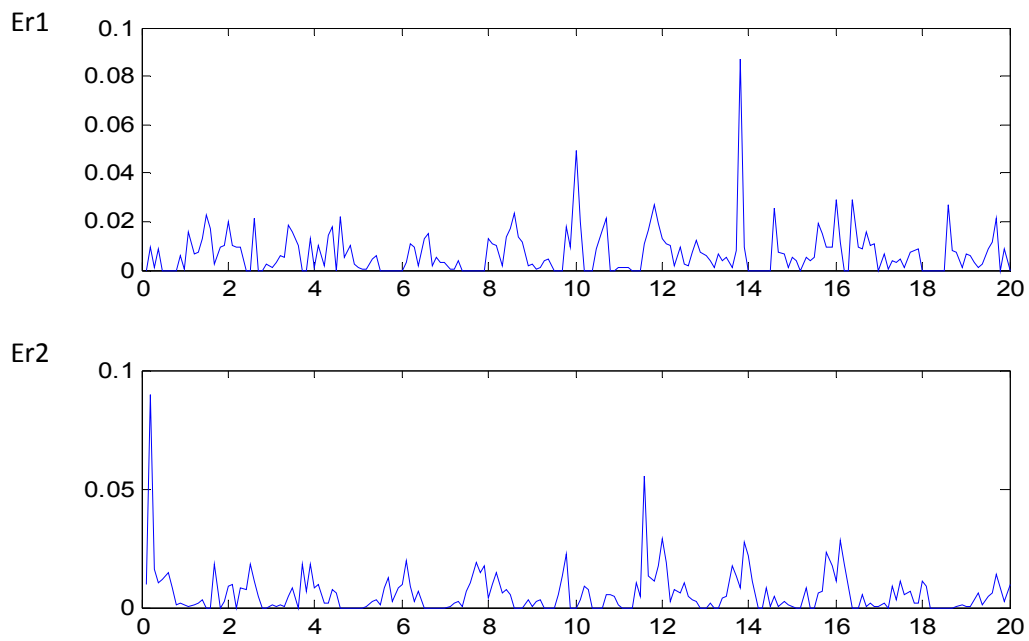


Fig IV.11 Les erreurs de séparation

Cet algorithme donne des bons résultats mais il converge après 1500 itérations, à cause de son architecture, avec une erreur  $1.456 \cdot 10^{-4}$  pour la première source estimée et  $1.788 \cdot 10^{-5}$  pour la deuxième source estimée.

#### IV-1.4 Exemple 4

Les signaux sources sont des signaux à grande différence d'échelle. Dans ce cas, nous confirmons l'efficacité de l'algorithme de Cichocki-Amari. On présente les signaux sources par la figure IV.12

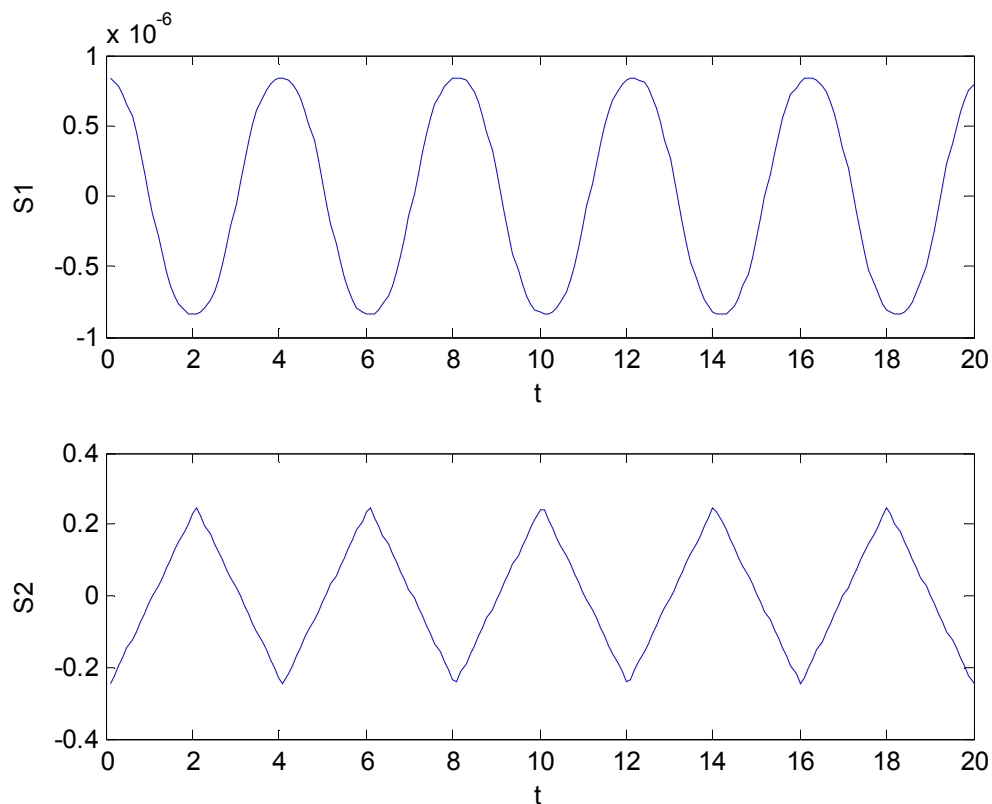


Fig IV.12 Les signaux sources

La matrice de mélange est donnée par :

$$A = \begin{pmatrix} 0.57 & -0.43 \\ 0.89 & -0.92 \end{pmatrix}$$

On obtient les observations qui sont présentées par la figure suivante :

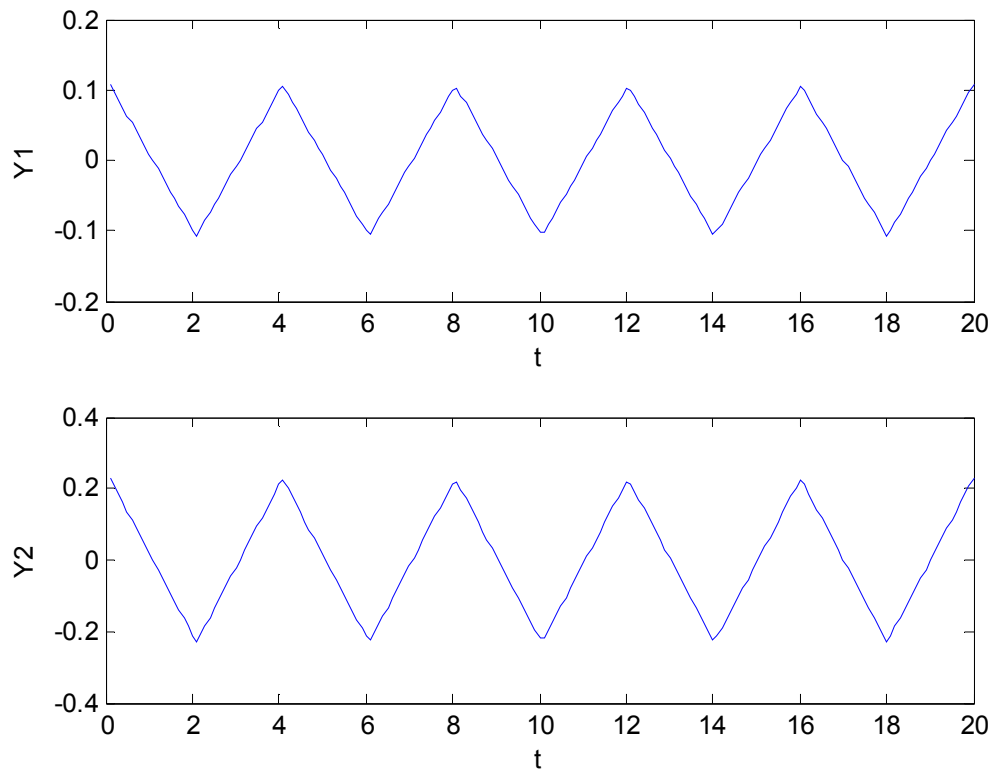


Fig IV.13 Les signaux observations

On applique l'algorithme de Cichocki-Amari avec :

- Les fonctions d'activations suivantes :

$$f(y) = y^3$$

$$g(y) = \tanh(y)$$

- $\mu=0.01$  (gain d'adaptation ou pas d'apprentissage)

Les signaux estimés sont présentés par la figure IV.14 :

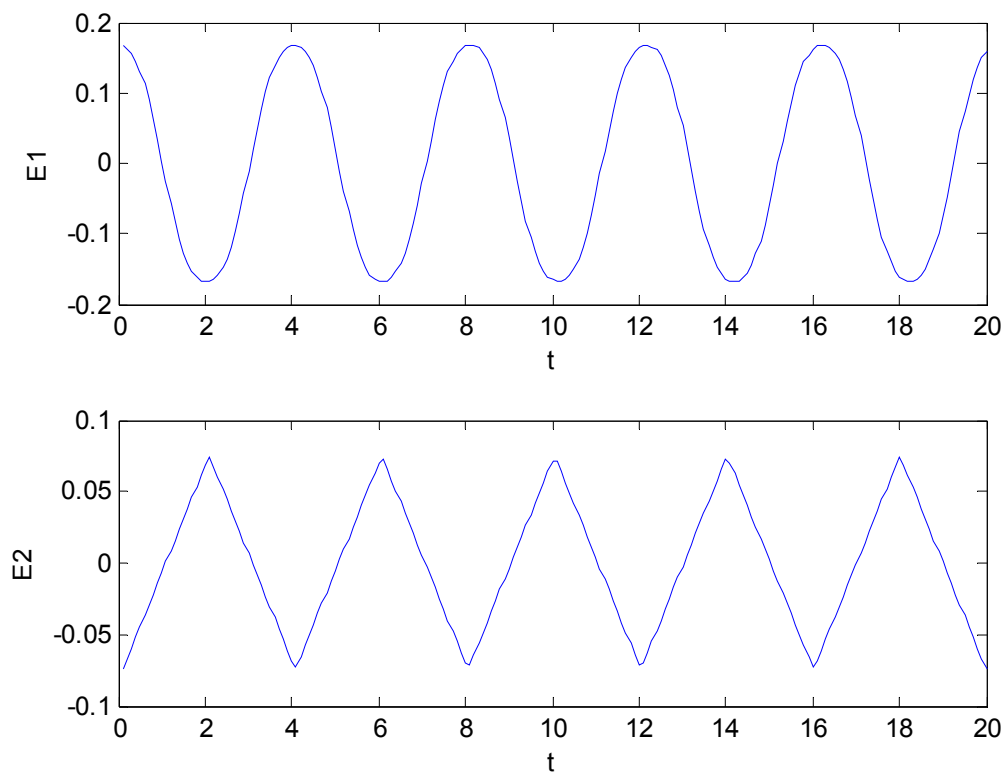


Fig IV.14 Les signaux Estimés

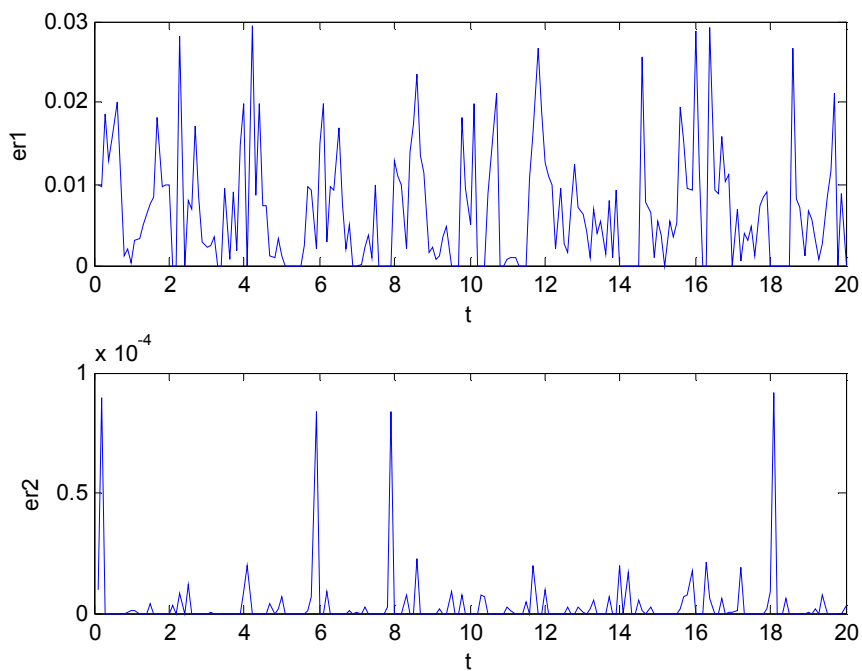


Fig IV.15 Les erreurs de séparation

Cet algorithme donne une bonne séparation même dans le cas où la matrice de mélange est mal conditionnée. La convergence de cet algorithme est obtenue après 20 itérations et avec des erreurs :  $3.8765 \cdot 10^{-4}$  et  $3.5468 \cdot 10^{-4}$  pour la première et la deuxième source estimés respectivement.

### IV-1.5 Exemple 5

Cet exemple est donné pour tester la résistivité de l'algorithme présenté par Cardoso-Laheld pour les signaux à grande différence d'échelle et avec une matrice de mélange mal conditionnée :

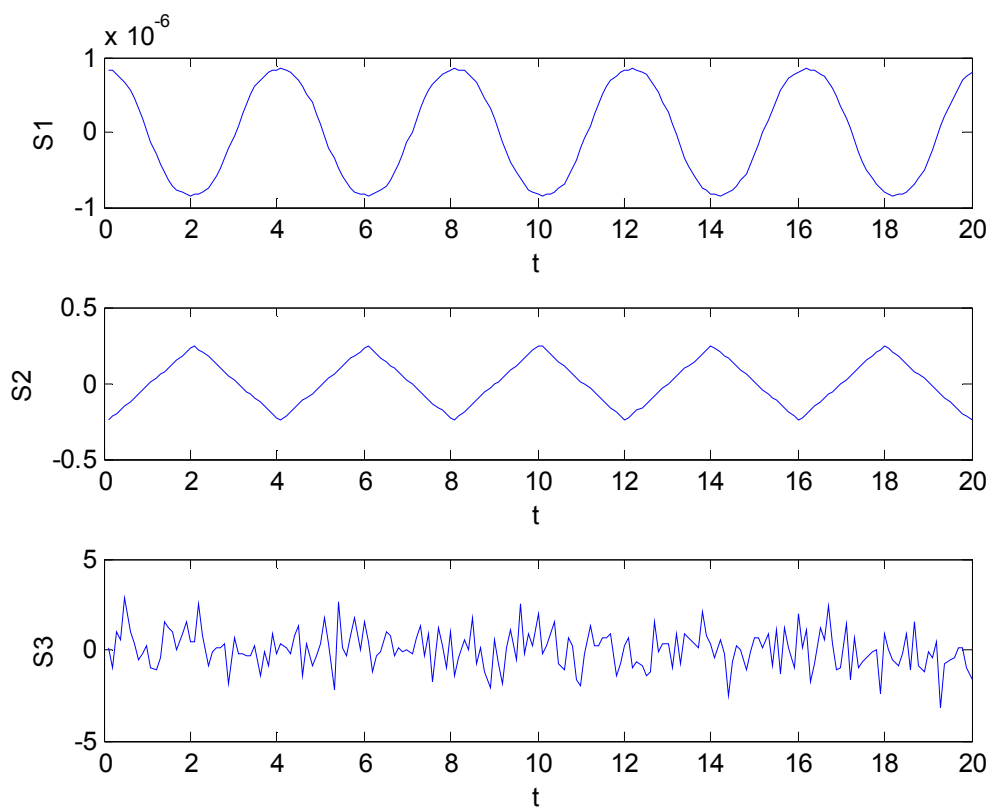


Fig IV.16 Les signaux sources

La matrice de mélange est donnée par :

$$A \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2.001 & 3 \\ 1 & 2 & 3.002 \end{pmatrix}$$

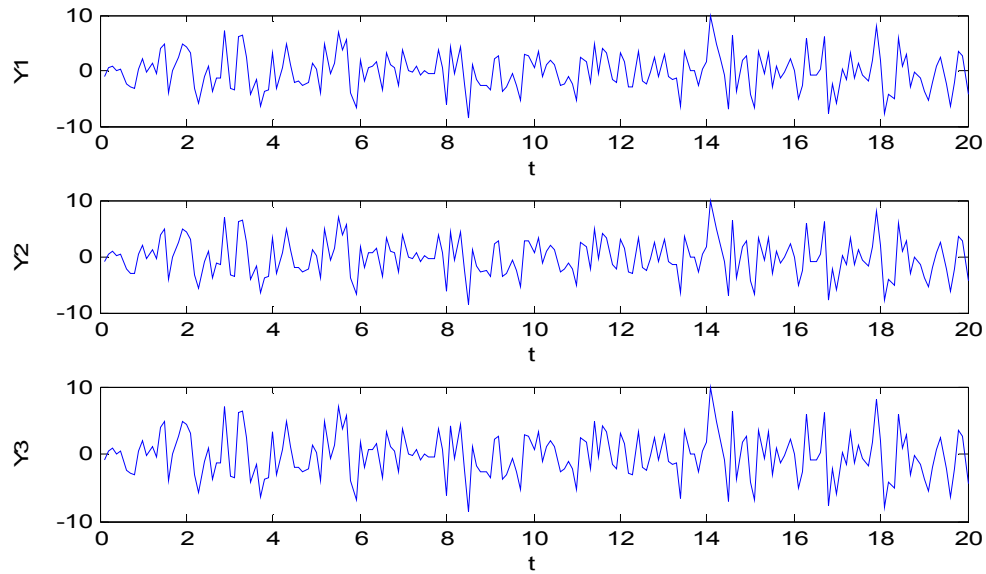


Fig IV.17 Les signaux mélanges

On applique l'algorithme (E.A.S.I) avec :

- La fonction d'activation suivante :  
$$g(y) = |y|^2 y$$
- $\mu=0.03$  (gain d'adaptation au pas d'apprentissage)

Les signaux estimés sont présentés par la figure IV.18 :



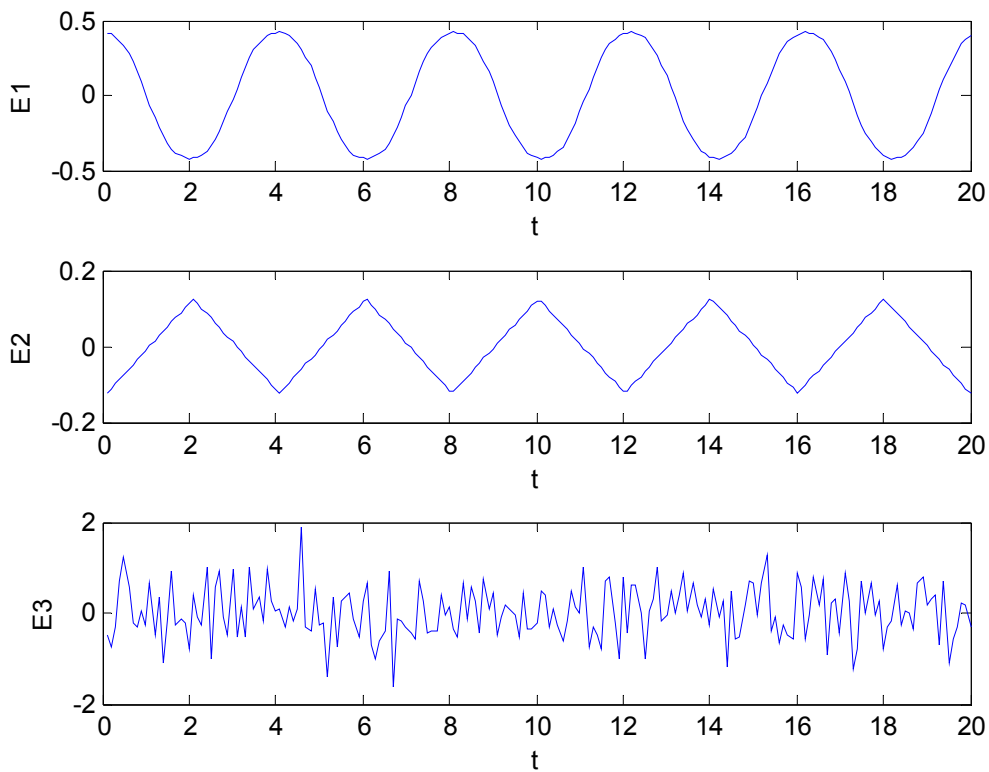


Fig IV.18 Les signaux estimés

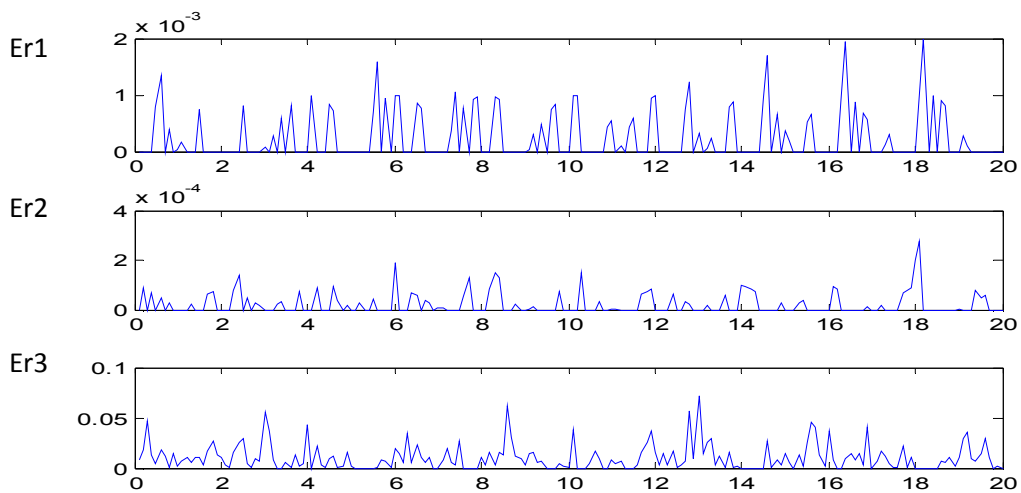


Fig IV.19 Les erreurs de séparation

Les signaux estimés sont identiques à celles d'entrées, l'algorithme commence à identifier les sources après 80 itérations, avec les erreurs  $4.5675 \cdot 10^{-4}$ ,  $1.7658 \cdot 10^{-5}$  et  $2.456 \cdot 10^{-2}$  pour la première et la deuxième et la troisième source estimés respectivement.

## IV.2. Conclusion

Les résultats de simulation donnent une vue générale sur la capacité de chaque algorithme d'extraire les signaux primaires. L'algorithme proposé par Herault-Jutten a la possibilité de séparer deux sources seulement et l'étude de cet algorithme est la base de tous les algorithmes proposés par la suite.

Pour les autres algorithmes, il y a une différence remarquable dans les résultats trouvés à cause des différentes lois d'adaptation propres à chacun. Les algorithmes de Cichocki-Amari et Clemente-Acha ont une difficulté provoquée par les matrices de mélanges mal conditionnées. L'algorithme Cardoso-Laheld est capable de traiter le problème de mélange mal conditionnée et même pour des signaux à grande différence d'échelle. Il donne de meilleures performances.

# **Conclusion**

## **Générale**

### Conclusion générale

Dans ce mémoire, nous avons abordé la problématique de la séparation aveugle de sources (SAS). En particulier, nous nous sommes intéressés au problème de la séparation de mélanges linéaires instantanés de sources indépendantes. La richesse bibliographique du problème de la SAS ainsi que son champ d'application sont immenses et remarquables.

Après une étude rigoureuse du problème de la séparation aveugle de sources, sa définition, son concept et ses diverses applications, plusieurs techniques ont été proposées dans la littérature pour apporter une solution à ce problème. Dans notre travail, nous nous sommes basées sur les approches neuronales, appelées aussi approches connexionnistes. En effet, on a étudié en détail la structure de l'algorithme de base de Héroult-Jutten qui représente la première approche qui a été donnée pour la solution du problème de la SAS, d'autres techniques qui s'inspirent de cet algorithme, et qui représentent une amélioration de l'algorithme de base de Héroult-Jutten (appelées également extensions), ont été également exposées.

Pour tester l'application de ces différents algorithmes étudiés, nous avons implémenté un programme en MATLAB qui permet de visualiser les signaux sources, les signaux mélanges ainsi que les signaux estimés. Les performances de ces différents algorithmes sont montrées par des exemples de simulation pour les différents cas d'application.

Au cours de l'étude des exemples de simulation, nous avons remarqué que plusieurs facteurs influent sur la convergence des algorithmes de séparation tels que le choix de la règle d'adaptation qui permet d'accélérer la séparation des signaux mélanges et assure une meilleure stabilité, et également le choix de l'architecture du réseau neuronal qui fait la différence entre les exemples de simulation ainsi que le pas d'apprentissage.

Dans le cas où nous avons un manque d'information sur l'indépendance statistique des signaux sources, cela rend difficilement voire impossible la séparation des mélanges.

Pour les travaux futurs, plusieurs autres techniques peuvent donner des solutions efficaces au problème de la SAS à savoir les statistiques d'ordres supérieurs, l'information mutuelle, la maximisation et la minimisation de l'information...etc pour les différents types de mélanges.

# ***Bibliographie***

## Références Bibliographiques

- 1- S. Breton, Une approche neuronale du contrôle robotique utilisant la vision binoculaire par reconstruction tridimensionnelle, thèse de doctorat, université de Haut Alsace, 2001.
- 2- J. EL Mhamdi, F. Regragui & M. Harnafi, Traitement adaptatif appliqué au signal sismique, Bulletin de l'Institut Scientifique, Rabat, section Sciences de la Terre, 2008, n°30, 13-22.
- 3- Z. Malouche, O Macchi et E Moreau, séparation adaptative de sources binaire a l'aide d'un réseau de neurones non lineare boucles, quinziesme colloque gretsi-jaun-les-pins-du 18 au 21 septembre 1995 .
- 4- H.L. Nguyen THI, J. Caelen et Ch. Jutten, Réhaussement de la parole par la séparation de sources dans un mélange convolutif, Colloque C5, supplément au Journal de Physique III, Volume 4, mai 1994
- 5- S. Chitroub, Analyse en Composantes Indépendantes D'images , Laboratoire de traitement du signal et d'images, Faculté d'électronique et d'informatique, Université des Sciences et Technologie , Revue Télédétection, 2007, vol. 7, n° 1-2-3-4, p. 541-555.
- 6- A Pérez-Urbe, Réseaux de neurones artificiels et apprentissage supervisé , haut ecole de l'igieniere et de gestion du canton de vaud 2002
- 7- P. Comon, "independent component analysis, A new concept". signal processing, vol 36, N°3 pp 287-314, 1994
- 8- A. Hyvannenn " Survey on independent componenet Analysis", Neural computing surveye, vol 2, pp 94-128, 1999
- 9- S. Amari, A. Cichocki "Adaptative Blind Signal Proveling- Neural Network ". proceedings of the IEEE, Vol 86 N° 10, October 1998
- 10 - Bell A.J, Sejnowski T.J., "the independent Component of Natural Scences are Edge Filter". Vision Research, vol 37, N°23, pp 3327-3338, 1997

- 11- L. Benaroya, X. Rodet, A. Robel, Éric le Carpentier, Frédéric BIMBOT, Comment évaluer les algorithmes de séparation de sources audio ,In Proceeding of ICA'99 ,1999.
- 12- P. Comon, C. Jutten , Séparation de sources. Principes et algorithmes, CNRS, Univ. of Nice, Sophia-Antipolis : GIPSA-lab, CNRS, Univ de Grenoble2002
- 13- Alexey Ozerov, Adaptation de modèles statistiques pour la séparation de sources mono-captur, thèse de doctorat universite de rennes 1, 2005
- 14 – S. Lagrange,Contributions aux méthodes d'estimation en aveugle Thèse de doctorat, Université Joseph Fourier 2005.
- 15- C. Touzet , les réseaux de neurones artificiels, pp 22-33, juillet 1992  
[www.cavalex.com/pdf/livre\\_touzet](http://www.cavalex.com/pdf/livre_touzet)
- 16- Svetlana Bozinoski, Anisse Taleb, Jean-Claude Guizzo, et Christian Jutten, Séparation de sources Application à la séparation de signaux et de brouilleurs dans un satellite de télécommunications ,IEEE Trans.on Signal processing 47(10) : 2807-2820.1990.
- 17- A. Kachenoura , Traitement Aveugle de Signaux Biomédicaux thèse de doctorat, L'Université de Rennes. 1 06 juillet 2006
- 18- Ch. Jutten, J. Herault, A neuromimetic solution for the problem of sources discrimination , Traitement du Signal. Grenoble série 5 Volume no 6 - 1988
- 19-K. Volpi ; Réseau de neurones PP 4-22 ,2005,  
[tel.archives-ouvertes.fr/docs/00/73/11/47/PDF/Boscuti\\_thA\\_se.pdf](http://tel.archives-ouvertes.fr/docs/00/73/11/47/PDF/Boscuti_thA_se.pdf)
- 20 – J-F. Cardoso, Analyse en composantes independantes, CNRS/ENST Paris, France,1998
- 21- Ch. Jutten, J. Herault, detection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé ,de l'Académie des sciences.Paris série III, page 525-528 ,1984
- 22- [www.obs.univ-bpclermont.fr/atmos/.../cours.../cours\\_RN\\_2006.pdf](http://www.obs.univ-bpclermont.fr/atmos/.../cours.../cours_RN_2006.pdf)
- 23- G. Dreyfus,Les Réseaux de Neurones, École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI),1998

- 24- P. Paquet, utilisation de RN en finance Faculté de droit d'économe et de gestion France.2000
- 25 – M. Fatima, S. Hamid « Comparaison de méthodes de classification réseau RBF, MLP et RVFLNN1 », Damascus University Journal Vol. (25) - No. (2) 2009
- 26 – O. Bermond, J-f. Cardoso, Méthodes de separation de sources dans le cas sous-déterminé , Dix-septième colloque GRETSI, Vannes, 13-17 septembre 1999
- 27- S. Amari, Natural Gradient Learning for Over- and Under-Complete Bases in ICA , RIKEN Brain Science Institute, Wako-shi, Hirosawa, Saitama 351-01, Japan
- 28- J-F. Cardoso, Beate Hvam Laheld, Equivariant adaptive source séparation ,IEEE tren. On signal processing, vol 44, no.12,(1996),3017-3031.
- 29-R.clement, J I. acha «A new algorithm for the aaptive separation of sources «,In Proceeding of ICA'99,1999.
- 30-P . Chevalier , P. Comon « séparation aveugle de sources » in Elsevier Linear Algebra and its Applications, vol. 391C, pp. 3-30, November 2004
- 31-A. Cichocki, R. rubust « neural network whith on-line learning for blind identification and blind separation of sources », IEEE Volume: 43 Page(s): 894 - 906 Unbehauen(1996)
- 32-G. Burel, Blind Separation of Sources, A Non-Linear Neural Algorithm, Neural Networks, Vol. 5, No. 6, Nov-Dec 1992, pp. 937-947
- 33-S. Amari , A .Cichocki, «Adaptive Blind Signal Processing—Neural Network Approaches«, IEEE, VOL. 86, NO. 10, October 1998
- 34-A. Meyer-Base, P. Gruber, F. Theis, S. Foo, « Blind source separation based on self-organizing neural network, Engineering Applications of Artificial Intelligence « Journal Engineering Applications of Artificial Intelligence Volume 19 Issue 3, April, 2006 Pages 305-311
- 35- Z. Bouabida, Z-E. Hadj Slimane et F-B. Reguig, « Séparation du Signal Electrocardiogramme Fœtale de ses Interférents par l'ACI de Point Fixe«. SETIT conférence sousse 21-24 marsh 2012



- 36- A- N Tripathi ,R. Pratap Singh , «A Neural Network Algorithm based Blind source Separation using Fast Fixed Point Independent Component Analysis», International Journal of Electronics and Computer Science Engineering, ISSN 2277-1956/V1N1-15-21
- 37- P-Y Robert, « Analyse et Classification de Signaux Neuronaux Modélisés », Mémoire présenté en vue de l'obtention du diplôme de maîtrise es sciences appliquées. Université de Montréal 2007
- 38- ApSTAT Technologies « Introduction Aux Réseaux De Neurones », Montréal (Québec) H3T 2B1 2002.
- 39- A-Cichocki,S I-Amari, « adaptive blind signal and image processing» tome2,pp273-297(2002)
- 40- Dr. Nabil Benoudjit, « Les réseaux de neurones artificiels ». Support de cours pour les Post-gradués. Département d'Electronique, Université de Batna (2010)
- 41- G. DREYFUS, « Les Réseaux de Neurones ». École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris ( ESPCI ), Laboratoire d'électronique.1998
- 42- E- MOREAU, O-MACCHI, « Novel Architectures for the Self Adaptive Separation of Signals», IEEE Transactions on Signal Processing, Vol. 45, No. 4, pp 918-926, April 1997.
- 43- R-M. Clemente, J-I. Acha, « A New Algorithm for the Adaptive Separation of Source Université de Sevilla, Spain 2001.
- 44- C. JUTTEN, J. HERAULT, Independent Component Analysis versus PCA. In proc. EUSIPCO Grenoble, pages 643-646, 1988.
- 45- R.Virgario. «Extraction of ocular artifacts from eeg using independent component analysis». Electroenceph. Clin. Neurophysiol, 103:395-404, 1997.
- 46-H.Akaike. « A New look at the statistical model identification».IEEE trans. Automatic control,AC-19:716-723,1974;System identification.and time-series analysis.
- 47- W.Hardle. Smoothing technique with implementation in s. Springer-Verlag,1991
- 48-E.Moreau ,N.Thiron-Mereau. « Non symmetrical contrasts for source separation». In IEEE transactions on signal processing, 47(8) : 2241-2252,1991
- 49- P.Comon. «From source separation to blind equalization contrast-based approaches». In ICISP 01. Int Conf. on image and signal processing, Agdir, Morocco, pages 20-32, 2001

- 50- P.Comon and E. Moreau , « improved contrast dedicated to blind separation in communications in proc», ICASSP, Munich , Germany, pages 3453-3456, 1997
- 51- T.Oostendrop. modelling the fetal ecg. PH.D . dissertation , K.U . Nijmegen, the Netherlands 1989 .
- 52- A. Taleb and C. Jutten. «On underdetermined source separation». In Proceedings ICASSP 99, volume 3, pages 1445–1448, Arizona, Phoenix, March 1999.
- 53- A. Taleb and C. Jutten. «Source separation in post non linear mixture». IEEE Transactions on signal Processing, 47(10):2807–2820, October 1999.

## Résumé :

Dans ce travail, nous avons étudié le problème de la séparation aveugle de sources ainsi que quelques méthodes pour sa solution. Ces méthodes sont basées essentiellement sur l'application des réseaux de neurones artificiels. L'algorithme de Base de Héroult-Jutten a été exposé en détail ainsi que d'autres approches appelées extensions.

## Mots clés :

Séparation aveugle de sources, algorithme de Héroult-Jutten, réseaux de neurones artificiels (RNA).

---

## Abstract:

In this work, we studied the problem of blind sources separation and the techniques used to solve it based essentially by using the artificial neural networks. The Héroult-Jutten approach constitutes the first algorithm proposed solving this problem. Other methods have been proposed as an extension of the basic H-J algorithm.

## Key words:

Blind sources separation, algorithm of Héroult-Jutten, artificial neural networks (ANN).

---

## ملخص

في هذا العمل, قمنا بدراسة المشكلة الخاصة بفصل مصادر الإشارات فضلا عن بعض الأساليب لحلها وتستخدم هذه الطرق أساسا على تطبيق الشبكات العصبية الاصطناعية. حيث تطرقنا بالتفصيل إلى خوارزمية Héroult-Jutten إضافة إلى أساليب أخرى كإمتداد لهذه الخوارزمية .