THÈSE DE DOCTORAT DE l'UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

Spécialité : Génie Informatique, Automatique & Traitement du Signal
Option : Traitement du Signal

présentée par

Mohamed Salem OULD MOHAMED

pour obtenir le grade de

DOCTEUR DE l'UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

Sujet de la thèse :

Contribution à la séparation aveugle de sources par utilisation des divergences entre densités de probabilité : application à l'analyse vibratoire

Soutenue le 08 novembre 2010 devant le jury composé de :

Président	M. Laurent DUMAS	Professeur à l'UVSQ de Versailles
Rapporteurs	M. Jérôme ANTONI	Professeur à l'UTC de Compiègne
	M. Eric MOREAU	Professeur à l'USTV de Toulon
Examinateur	M. Mohammed EL RHABI	Responsable scientifique, Realeyes 3D
Directeur de thèse	M. Georges DELAUNAY	Professeur émérite à l'URCA
Co- $encadrants$	M. Hassan FENNIRI	Maître de conférences à l'URCA
	M. Amor KEZIOU	Maître de conférences à l'URCA

Remerciements

Tout d'abord, je tiens à remercier très vivement mon Directeur de thèse Monsieur Georges Delaunay et mes deux encadrants Hassan Fenniri et Amor Keziou pour leurs soutien constant, leurs orientations, leurs conseils et leurs encouragements qui m'ont permis de mener ce travail. Qu'ils trouvent ici l'expression de toute ma reconnaissance.

Je remercie très chaleureusement Messieurs Jérôme Antoni et Eric Moreau pour avoir accepté d'être rapporteurs de cette thèse. Je leur suis très reconnaissant du temps qu'ils ont consacré à l'évaluation de ce travail. J'apprécie à sa juste valeur leur présence dans le jury.

J'adresse mes sincères remerciements à Messieurs Laurent Dumas et Mohammed El Rhabi pour l'intérêt qu'ils ont accordé à mon travail et pour avoir accepté de participer à ce jury.

Je tiens à remercier également Monsieur Janan Zaytoon, Directeur du Laboratoire CReSTIC pour m'avoir accueilli dans son Laboratoire ainsi que tous les membres du CReSTIC.

Je voudrais remercier très chaleureusement Elhaj Iyane Thiam et Jean-Michel Mavoungou (Issa) qui font preuve de leur gentillesse et patience. Je voudrais également adresser un salut amical à mes amis avec qui j'ai partagé de bons moments.

Je termine par un grand merci à toute ma famille pour le soutien qu'elle m'a apporté tout au long de la préparation de cette thèse.

Résumé :

Dans cette thèse, nous proposons un nouvel algorithme de séparation aveugle de sources, basé sur l'optimisation de l'information mutuelle sous contraintes. Le problème d'optimisation sous contraintes est résolu par passage au problème dual. L'estimateur proposé du gradient stochastique utilise l'estimation des densités par maximum de vraisemblance dans des modèles de lois exponentielles choisis par minimisation du critère AIC. Ensuite, la méthode a été généralisée à l'ensemble des divergences entre densités de probabilité. Nous avons montré que l'algorithme utilisant la divergence particulière de Hellinger a de bonnes propriétés d'efficacité et robustesse en présence du bruit en comparaison avec l'information mutuelle. Dans le cadre de signaux cyclostationnaires, les méthodes précédentes de séparation ont été adaptées en utilisant des statistiques du second ordre. Nous illustrons les performances des algorithmes proposés pour des signaux simulés et pour des signaux réels issus de machines tournantes.

Mots-clés :

Séparation aveugle de sources ; Information mutuelle ; AIC ; Sélection de modèles ; Maximum de vraisemblance ; Famille de lois exponentielles ; Optimisation sous contraintes ; Lagrangien ; Divergences entre densités de probabilité ; Distance de Hellinger ; Cyclostationnarité ; Statistiques d'ordre 2 ; Analyse vibratoire.

Abstract :

In this thesis, we propose a new blind source separation algorithm based on the optimization of mutual information under constraints. The optimization problem is solved by using the dual problem. The estimator of stochastic gradient is based on the estimation of the densities by maximum likelihood method. The densities are chosen from exponential families using the AIC criterion. Then, we propose a new algorithm for blind source separation based on the minimization of divergences witch generalizes the Mutual Information (MI) approach. We show that the algorithm using Hellinger's divergence has better properties in terms of efficiency-robustness, for noisy data. In the context of cyclostationary signals, the above methods of separation were adapted using second order statistics. We illustrate the performances of the proposed algorithms through simulations and on real rotating machine vibration signals.

Keywords :

Blind source separation; Mutual information; AIC; Model selection; Maximum likelihood; Exponential family laws; Optimization under constraints; Lagrangian; Divergences; Hellinger distance; Cyclostationarity; Second-order statistic; Vibratory analysis.

Table des matières

Introduction générale

1	La séparation aveugle de sources (SAS)			5
1.1 Introduction			luction	5
	1.2	Applie	cations	6
	1.3	3 Modélisation		
		1.3.1	Modèle de mélanges linéaires instantanés	11
		1.3.2	Modèle de mélanges linéaires convolutifs	15
	1.4	1.4 Principe de résolution de la SAS		19
		1.4.1	Le cas de mélanges linéaires instantanés	19
		1.4.2	Le cas de mélanges linéaires convolutifs	22
		1.4.3	Analyse en Composantes Indépendantes (ACI)	24
	1.5 Quelques hypothèses de base		ues hypothèses de base	25
1.6 Critères utilisés en SAS		es utilisés en SAS	27	
		1.6.1	Information mutuelle (IM)	27
		1.6.2	Maximum de vraisemblance	30
		1.6.3	Statistiques d'ordre deux (SO2)	32
		1.6.4	Statistiques croisées d'ordre supérieur	35
		1.6.5	Fonctions de contraste	38
	1.7	Critèr	es de performances de la SAS	40

1

		1.7.1	Erreur quadratique moyenne (EQM)	. 40
		1.7.2	Rapport signal sur résidus (SNR)	. 40
	1.8	8 Conclusion		
Ι	SA	AS sta	ationnaires	43
2	Mél	anges	linéaires instantanés	45
	2.1	Introd	luction	. 45
	2.2	Métho	ode basée sur la minimisation de l'IM sous contraintes	. 46
		2.2.1	Introduction	. 46
		2.2.2	Définitions	. 48
		2.2.3	IM sous contraintes	. 50
		2.2.4	Selection de modèles	. 53
		2.2.5	Résultats de simulations	. 55
	2.3	Métho	odes basées sur la minimisation des α -divergences \ldots \ldots	. 59
		2.3.1	Divergences entre lois de probabilité	. 60
		2.3.2	Approche 1 : méthode "décorrélation-rotation"	. 65
		2.3.3	Approche 2 : méthode "directe"	. 73
	2.4	Conclu	usion	. 77
3	Mél	anges	linéaires convolutifs	83
	3.1	Introd	luction	. 83
	3.2	Modèl	e convolutif	. 84
	3.3	Nouve	eaux critères de séparation via les α -divergences \ldots \ldots	. 85
	3.4	Calcul	l du gradient	. 86
	3.5	Résult	tats de simulations	. 88
	3.6	Etude	s expérimentales	. 90
	3.7	Conclu	usion	. 92

TABLE	DES	MATIÈRES
TADLL	$\mathbf{D}\mathbf{D}\mathbf{O}$	MALLED

II	$\mathbf{S}_{\mathbf{r}}$	AS cy	vclostationnaires	95
4	Mél	anges	linéaires instantanés de signaux cyclostationnaires	97
	4.1	Introd	luction	. 97
	4.2	Cyclos	stationnarité : définitions et propriétés	. 98
		4.2.1	Cyclostationnarité et applications	. 98
		4.2.2	Caractéristiques cyclostationnaires	. 100
	4.3 Méthodes utilisant les statistiques cycliques d'ordre 2			. 101
		4.3.1	Introduction	. 101
		4.3.2	Description du critère de séparation de sources cyclostation-	
			naires	. 103
		4.3.3	Extension du critère au cas de plusieurs sources $(p\geq 3)$. 105
		4.3.4	Résultats de simulations	. 106
	4.4	Métho	odes basées sur la minimisation des divergences pénalisées	. 111
		4.4.1	Introduction	. 111
		4.4.2	Méthodes basées sur la minimisation des α -divergences	. 112
		4.4.3	Méthode utilisant les statistiques d'ordre 2	. 114
		4.4.4	Méthodes basées sur la minimisation des α -divergences pénalisé	es
			(combinaison des deux critères 4.21 et 4.24)	. 116
	4.5	Concl	usion	. 120
	Con	clusio	n générale	125
	Ann	iexe		129
\mathbf{Li}	ste d	es tabl	eaux	143
Ta	ble d	les figu	ıres	144
Bi	bliog	graphie	9	148

Introduction générale

Cette thèse est consacrée au développement et à l'implantation de nouvelles techniques pour la résolution du problème de la Séparation Aveugle de Sources (SAS). La SAS consiste à estimer un ensemble de signaux inconnus dits "sources", à partir d'un ensemble de signaux connus dits "observations". Ces observations sont généralement des mélanges des sources et proviennent de capteurs comme des microphones, sondes, accéléromètres, antennes, caméras, ... etc. Vu les nombreuses applications que permet la SAS, son exploitation est devenue primordiale dans de nombreuses disciplines scientifiques telles que les télécommunications, l'acoustique, le génie biomédical, l'astrophysique, la mécanique, ... etc. La plupart des méthodes de SAS supposent l'indépendance statistique des sources, elles consistent donc à rendre les observations indépendantes au sens statistique. Nous retrouvons dans la littérature différentes méthodes exploitant cette idée dont certaines utilisent les statistiques d'ordre deux, d'autres les statistiques d'ordre supérieur alors que d'autres encore exploitent l'Information Mutuelle (IM). De ce fait, et compte tenu de la diversité de ces méthodes, il y a une recherche incessante de nouvelles méthodes plus performantes. D'ailleurs, les méthodes de la SAS que nous proposons dans cette thèse vont dans ce sens : Nous proposons de nouvelles méthodes qui exploitent les statistiques d'ordre deux, l'information mutuelle sous contraintes, les divergences entre densités de probabilité, et les divergences pénalisées, avec pour cadre applicatif l'analyse vibratoire.

Avant d'entamer un exposé des nouveaux outils de SAS stationnaires (partie I) et cyclostationnaires (partie II) que nous proposons, il nous a semblé utile de passer en revue au préalable l'état de l'art du problème de la SAS (chapitre 1) qui ne peut être exhaustif. Nous rappelons le problème de la SAS et les éléments qui s'y rattachent, les différents domaines d'applications de la SAS, la modélisation du mélange, l'objectif et les difficultés de la résolution de ce problème.

La suite de la thèse est décomposée en deux parties. La partie I traite le cas des mélanges de signaux sources stationnaires. Dans le chapitre 2, et dans le cadre de mélanges linéaires instantanés, nous proposons deux nouvelles méthodes de la SAS. Une première méthode basée sur la minimisation de l'Information Mutuelle (IM) sous contraintes estimée via l'utilisation du critère AIC (Akaike Information Criterion) pour la sélection de modèles de densités dans une famille croissante de lois exponentielles et estimation des paramètres par maximum de vraisemblance. La deuxième méthode exploite les divergences entre densités de probabilité. Cette méthode s'applique également dans le cas convolutif (chapitre 3). Elle généralise l'approche classique de l'IM qui correspond ici à un choix particulier de la divergence de Kullback-Leibler modifiée. Nous verrons que le choix de la divergence de Hellinger fournit une méthode qui améliore celle de l'IM en termes d'éfficacité-robustesse pour des signaux bruités, elle présente de plus le meilleur compromis parmi toutes les méthodes utilisant les α -divergences y compris la méthode de l'IM. Dans le paragraphe 3.6 nous avons traité des signaux réels vibratoires en utilisant le critère de Hellinger.

La partie II traite le cas des mélanges linéaires instantanés de sources cyclostationnaires. Dans ce cadre, nous proposons deux méthodes. La première est basée sur des statistiques cycliques d'ordre deux, elle est adaptée au cas des sources cyclostationnaires de fréquences cycliques connues et différentes. La deuxième utilise à la fois les divergences entre densités et les statistiques cycliques d'ordre deux. Cette dernière méthode améliore l'efficacité et la vitesse de convergence des algorithmes de séparation des signaux appropriés.

Chapitre 1

La séparation aveugle de sources (SAS)

1.1 Introduction

La séparation aveugle de sources (SAS) est un domaine de recherche qui n'a cessé d'attirer l'attention d'un très grand nombre de chercheurs. Cet enthousiasme profond pourrait s'expliquer par l'intérêt que la SAS représente dans plusieurs situations et applications diverses. Celle-ci est devenue une discipline générique dans de nombreux domaines tels que les télécommunications, l'acoustique, le génie biomédical, l'astrophysique, la mécanique, ... etc.

La SAS est un problème général en traitement du signal, dont le principe consiste à retrouver un ensemble de signaux inobservables dits "signaux sources", à partir d'un ensemble de signaux observables dits "observations". Ces observations sont souvent des mélanges de ces sources et proviennent de capteurs comme par exemple des microphones, sondes, accéléromètres, antennes, caméras, ... etc. Nous pouvons observer sur chaque capteur, la sortie d'un système réalisant le "mélange" des signaux sources. La nature du mélange et le milieu de propagation de ces sources sont généralement inconnus. Aucune information n'est donc disponible sur les sources ni sur les mélanges. Vu ces ambiguïtés, il est difficile voire impossible de retrouver les sources sans faire quelques hypothèses. Nous allons présenter brièvement les différents domaines d'applications de la SAS, la modélisation du mélange, l'objectif et les difficultés de la résolution de ce problème.

Le problème de la SAS a été introduit et formulé par Hérault et Jutten au milieu des années 1980 [94, 96]. Depuis, ce thème a fait l'objet de nombreuses recherches dans divers domaines d'applications.

1.2 Applications

Pour mieux appréhender le problème de la SAS, et sa résolution, nous essayerons de présenter dans ce paragraphe quelques applications de la SAS et les travaux initiateurs.

1 - Mansour [130] utilise la SAS pour traiter des signaux de paroles. Plus particulièrement, elle a été utilisée dans le domaine du rehaussement de la parole [144],
[157], [21] et [25].

2 - Pham *et al.* [156] exploitent la SAS dans le cadre de signaux audio. Torkkola [177] et Mitianoudis et Davies [133], se sont aussi intéressés à la séparation des signaux audio. Dans [83, 84], Gelle *et al.* séparent deux signaux sonores émis simultanément par deux haut-parleurs; voir aussi Westner et Bove [184, 183].

3 - Dans le domaine des radio-communications, Charkani [43], utilise la SAS dans les systèmes radio-mobiles, dans les techniques de type SDMA (Spatial Division Multiple Access), ou dans le système du téléphone à mains-libres. Xavier *et al.* [187] adoptent la SAS pour les GSM MIMO à architecture SDMA. En système de communications sans fil, en particulier le système CDMA (Code Division Multiple Access), Ristaniemi *et al.* [164] et Dimitri [62] ont fait appel à la technique de la SAS pour éliminer les interférences entre symboles reçues au niveau de chaque antenne. Nous trouvons aussi d'autres applications en communications numériques dans les références [109, 108].

4 - De nombreux auteurs se sont tournés vers le diagnostic de machines tournantes. Pour contrôler le fonctionnement de chaque moteur, Capdevielle [38] sépare les vibrations des machines tournantes. Sur une plateforme où deux moteurs fonctionnent simultanément, Gelle *et al.* [84, 85] et El Rhabi *et al.* [66], appliquent la SAS aux signaux vibratoires pour diagnostiquer l'état de fonctionnement de chaque moteur lors qu'ils fonctionnent simultanément. Afin d'extraire la contribution des vibrations générées par un passage des billes sur un défaut localisé sur la bague interne d'un roulement, Randall *et al.* [163], Bouguerriou *et al.* [23] et Boustany et Antoni [29] appliquent la SAS pour traiter le problème.

5 - Concernant le domaine du génie biomédical, la SAS a été considérée par De Lathauwer *et al.* [58] pour le traitement de signaux Electro-Cardio-Grammes (ECG), en se basant sur les travaux de [158] et [147]. Les auteurs l'utilisent pour séparer le signal correspondant aux battements du cœur d'un fœtus de celui de sa mère. Dans [59], les auteurs appliquent la SAS en considérant le cas de trouble du rythme cardiaque chez le fœtus et aussi la présence de jumeaux. Boustany et Antoni [30] utilisent la SAS pour l'extraction de l'électrocardiogramme d'un fœtus à partir des mesures prises sur l'abdomen de la mère ; voir aussi Gao *et al.* [77]. Pour réduire les artefacts en ECG et faciliter l'analyse des signaux ECG, Wisbeck *et al.* [185] recourent à la SAS en utilisant la technique de l'"Analyse en Composantes Indépendantes" (ACI) [50]. Vigärio [178] a appliqué aussi la technique de la SAS pour éliminer les artefacts rencontrés dans des signaux Electro-Encéphalo-Graphiques (EEG). De même, pour des signaux Magnéto-Encéphalo-Gramme (MEG), Vigärio *et al.* [179, 180] ont exploité la technique de la SAS pour éliminer les différents artefacts qui se combinent avec les informations MEG utiles; voir aussi les travaux de [13, 14], [116] et [159].

6 - En astronomie, Nuzillard et Bijaoui [145] ont examiné les possibilités offertes par les méthodes de la SAS en tant qu'outil exploratoire sur un jeu d'images multispectrales d'une galaxie, le but étant d'aider les astronomes à améliorer la connaissance des corps célestes.

7 - La SAS a aussi été appliquée dans le traitement des signaux de radar provenant de plusieurs aires pour le contrôle du trafic aérien d'un aéroport; voir Desodt et Muller [60] et Chaumette *et al.* [45].

8 - En acoustique sous-marine, le signal reçu par un ensemble de capteurs est un mélange des différentes sources élémentaires filtrées par l'environnement. Gaeta *et al*.
[74] ont exploité la SAS pour isoler chaque source à partir des signaux reçus.

9 - Afin de contrôler la dégradation de l'écran thermique d'un réacteur nucléaire,
d'Urso et Cai [65] ont appliqué différents algorithmes de séparation de sources pour la séparation des signaux d'un réacteur nucléaire.

10 - Enfin, dans les domaines des signaux et ondes sismiques, citons, parmi d'autres, les travaux de Thirion *et al.* [173], LeBihan *et al.* [126], Vrabie [181] et Larue [125].

Sur le plan pratique, notre contribution s'inscrit plus particulièrement dans le domaine de l'éléctromécanique et plus précisément le diagnostic de l'état de fonctionnement de machines tournantes.

Nous allons tout d'abord aborder le problème de la modélisation de la relation signaux observés-signaux sources. Les signaux observés sont des mélanges des signaux sources. Le paragraphe suivant présente des modèles mathématiques pour ces mélanges.

1.3 Modélisation

Les signaux sources se propagent dans un environnement appelé communément milieu de propagation. Ils y subissent naturellement des transformations. Les signaux observés (mesurés) sont donc des mélanges plus ou moins complexes des signaux sources. Comme ceux-ci sont inaccessibles (inconnus), il est impossible de déterminer les mélanges par la seule connaissances des signaux mesurés. Nous allons donc proposer des modèles pour ces mélanges.

- Mélanges linéaires : Ils sont utilisés lorsque la relation signaux observéssignaux sources est linéaire : on considère que chaque observation est constituée d'une somme pondérée des sources, ou des versions filtrées de celles-ci.
- Mélanges non-linéaires : Ils sont utilisés si la relation entre les signaux observés et les signaux sources est non linéaire. Ces modèles ont été peu étudiés à cause de leur complexité. Des classes particulières de ces modèles, par exemple les mélanges "post non-linéaires" et les mélanges linéaires "quadratiques" ont cependant donné lieu à des travaux.

Nous pouvons classer les mélanges en trois catégories selon le nombre de sources relativement au nombre d'observations :

- Les mélanges sous-déterminés : lorsque le nombre d'observations est inférieur au nombre de sources.
- Les mélanges sur-déterminés : si le nombre d'observations est supérieur au nombre de sources.
- Les mélanges dits déterminés : lorsque le nombre d'observations et le nombre de sources sont égaux.

Les mélanges peuvent être "instantanés" ou "convolutifs". Ils peuvent être aussi variants ou invariants dans le temps. En général, ce sont les modèles linéaires invariants dans le temps qui sont considérés dans la plupart des applications. Cependant, il faut noter que plusieurs travaux ont traité le cas non linéaire [171], [11], [112], [2], [168] et [61]. Dans le cadre de cette thèse, nous allons nous restreindre aux mélanges linéaires invariants déterminés. Nous présenterons donc en détail, dans le paragraphe suivant, les modèles des mélanges linéaires "instantanés" et "convolutifs".

La figure 1.1 résume différentes approches de séparation à partir des informations issues de mélanges de sources que l'on suppose au départ : linéaires, invariants, instantanés ou convolutifs. La plupart des outils disponibles dans la littérature pour aboutir aux sources "estimées" sont résumés Fig. 1.1. Les approches adoptées dont certaines plus particulièrement développées dans cette thèse (qui utilisent les α divergences et les statistiques cycliques d'ordre deux) apparaissent en grisé dans les différents blocs.



FIG. 1.1 – Procédés de séparation de sources.

1.3.1 Modèle de mélanges linéaires instantanés

Le processus de mélange entre les sources et les observations est modélisé par l'équation vectorielle suivante

$$\mathbf{x}(t) := \mathbf{A}\mathbf{s}(t),\tag{1.1}$$

où

 $\mathbf{s}(t) := (s_1(t), \dots, s_p(t))^T$ est le vecteur des signaux sources que l'on cherche à estimer; il est composé de p signaux.

 $\mathbf{x}(t) := (x_1(t), \dots, x_r(t))^T$ est le vecteur des signaux observés; il est constitué de r composantes.

 $\mathbf{A} \in \mathbb{R}^{r \times p}$ est la matrice de mélange (inconnue); elle est composée de coefficients scalaires.

En cas de présence de bruit additif dans l'environnement, le modèle mathématique de mélange relatif à cette situation s'écrit

$$\mathbf{x}(t) := \mathbf{As}(t) + \mathbf{b}(t), \tag{1.2}$$

où $\mathbf{b} := (b_1(t), \dots, b_r(t))^T$ est le vecteur des signaux bruits supposé être centré et dont les composantes sont généralement supposées mutuellement indépendantes, et indépendantes des signaux sources.

Le schéma de modélisation de la relation sources-observations peut être illustré par la figure 1.2.



FIG. 1.2 – Schéma de propagation pour un mélange instantané.

Séparer aveuglement les sources, à partir de leurs mélanges, est entaché d'indétermination. En effet, la séparation peut être obtenue avec une infinité de solutions [176], [32]. La contrainte d'indépendance des sorties n'est cependant pas suffisante pour résoudre complètement le problème. Comon présente dans [50] l'ensemble des indéterminations pour le cas linéaire instantané. Il montre que sans condition supplémentaire, les sources ne peuvent être estimées qu'à une permutation et un facteur d'échelle près. Dans la suite, nous allons présenter ce problème et son incidence sur la reconstruction des sources.

L'indétermination de permutation

La première indétermination liée au problème de la SAS est celle de la permutation. En effet, l'ordre de restitution des signaux est arbitraire car toute permutation appliquée sur le vecteur \mathbf{s} et sur les colonnes de la matrice \mathbf{A} correspondante donne naissance au même vecteur \mathbf{x} . Dans ce cas, la relation qui lie les deux vecteurs \mathbf{s} et \mathbf{x} sous forme vectorielle s'écrit

$$\mathbf{x}(t) = \sum_{j=1}^{p} \mathbf{a}_{j} s_{j}(t), \quad \text{où} \quad \mathbf{a}_{j} := (a_{1j}, \dots, a_{pj})^{T}.$$
 (1.3)

En utilisant la propriété de commutativité de l'opération d'addition, la relation cidessus peut s'écrire comme suit

$$\mathbf{x}(t) = \mathbf{a}_1 s_1(t) + \mathbf{a}_2 s_2(t) + \dots + \mathbf{a}_p s_p(t) = \mathbf{a}_2 s_2(t) + \mathbf{a}_1 s_1(t) + \dots + \mathbf{a}_p s_p(t), \quad (1.4)$$

c'est-à-dire,

$$\mathbf{x}(t) = \begin{bmatrix} a_{12}s_2(t) \\ \vdots \\ a_{p2}s_2(t) \end{bmatrix} + \begin{bmatrix} a_{11}s_1(t) \\ \vdots \\ a_{p1}s_1(t) \end{bmatrix} + \dots + \begin{bmatrix} a_{1p}s_p(t) \\ \vdots \\ a_{pp}s_p(t) \end{bmatrix}$$
(1.5)

 \Leftrightarrow

$$\mathbf{x}(t) = \begin{bmatrix} a_{12} & a_{11} & \dots & a_{1p} \\ a_{22} & a_{21} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p2} & a_{p1} & \dots & a_{pp} \end{bmatrix} \begin{bmatrix} s_2(t) \\ s_1(t) \\ \vdots \\ s_p(t) \\ \end{bmatrix} = \widetilde{\mathbf{A}}\widetilde{\mathbf{s}}(t), \qquad (1.6)$$

où $\widetilde{\mathbf{A}}$ et $\widetilde{\mathbf{s}}(t)$ sont respectivement la nouvelle matrice de mélange et le nouveau vecteur source. Cette permutation peut être modélisée par une matrice de permutation \mathbf{P} telle que

$$\mathbf{x}(t) := \mathbf{A}\mathbf{s}(t) = \mathbf{A}\mathbf{I}\mathbf{s}(t) = (\mathbf{A}\mathbf{P})\left(\mathbf{P}^{-1}\mathbf{s}(t)\right) =: \widetilde{\mathbf{A}}\widetilde{\mathbf{s}}(t).$$
(1.7)

Nous aboutirons donc au même vecteur d'observations $\mathbf{x}(t)$ si on considère $\widetilde{\mathbf{A}} := \mathbf{AP}$ et $\widetilde{\mathbf{s}}(t) := \mathbf{P}^{-1}\mathbf{s}(t)$. Nous concluons que la multiplication à droite de la matrice de mélange \mathbf{A} par une matrice de permutation \mathbf{P} ne change pas les mélanges, mais implique que l'ordre des sources est modifié.

L'indétermination du facteur d'échelle

Nous avons vu dans le paragraphe précédent que la permutation de deux colonnes dans la matrice de mélange et de deux sources, ne change en rien les observations. Nous allons voir dans ce paragraphe, que la multiplication d'une colonne (de la matrice de mélange) et la division d'une source par un scalaire non nul ne changera pas le vecteur mélange. En effet, pour tout $\alpha_i \neq 0$, on a

$$\mathbf{x}(t) = \sum_{i=1}^{p} \mathbf{a}_{i} s_{i}(t) = \sum_{i=1}^{p} (\alpha_{i} \mathbf{a}_{i}) \left(\frac{s_{i}(t)}{\alpha_{i}}\right),$$

i.e.,

$$\mathbf{x}(t) = \begin{bmatrix} \alpha_1 a_{11} & \alpha_2 a_{12} & \dots & \alpha_p a_{1p} \\ \alpha_1 a_{21} & \alpha_2 a_{22} & \dots & \alpha_p a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 a_{p1} & \alpha_2 a_{p2} & \dots & \alpha_p a_{pp} \end{bmatrix} \begin{bmatrix} \frac{s_1(t)}{\alpha_1} \\ \frac{s_2(t)}{\alpha_2} \\ \vdots \\ \frac{s_p(t)}{\alpha_p} \end{bmatrix}.$$
(1.8)

En notant Λ la matrice diagonale d'éléments diagonaux $\alpha_1, \ldots, \alpha_p$, $\hat{\mathbf{A}}$ la nouvelle matrice de mélange et $\hat{\mathbf{s}}$ le nouveau vecteur des sources, on peut écrire

$$\mathbf{x}(t) := \mathbf{A}\mathbf{s}(t) = \mathbf{A}\mathbf{I}\mathbf{s}(t) = (\mathbf{A}\mathbf{\Lambda})\left(\mathbf{\Lambda}^{-1}\mathbf{s}(t)\right) =: \hat{\mathbf{A}}\hat{\mathbf{s}}(t).$$
(1.9)

Cela montre que l'amplitude des sources est indéterminée.

Il est à noter que le principal inconvénient des méthodes de la SAS réside dans l'indétermination liée à la puissance des sources estimées. Divers travaux ont proposé des solutions pour lever cette contrainte d'indétermination, voir par exemple [114], [155], [134] et [67].

1.3.2 Modèle de mélanges linéaires convolutifs

Le processus de mélange est modélisé par

$$x_i(t) := \sum_{j=1}^p a_{ij}(t) * s_j(t), \quad i = 1, \dots, r,$$
(1.10)

où

* est l'opérateur de convolution.

 $s_j(t)$ est le signal émis par la j^{ieme} source $j = 1, \ldots, p$.

 $x_i(t)$ est le signal observé à la sortie du capteur $i, i = 1, \ldots, r$.

 $a_{ij}(t)$ est la réponse impulsionnelle du filtre caractérisant la propagation du signal de la j^{eme} source jusqu'au i^{eme} capteur, à l'instant t.

Ceci se traduit, en temps continu, sous forme vectorielle par l'équation suivante

$$\mathbf{x}(t) := \mathbf{A} * \mathbf{s}(t), \tag{1.11}$$

où

 $\mathbf{s} := (s_1(t), \dots, s_p(t))^T$ est le vecteur des signaux sources. $\mathbf{x} := (x_1(t), \dots, x_r(t))^T$ est le vecteur des signaux mélanges.

A représente ici la matrice des filtres de mélange qui regroupe les réponses impulsionnelles des filtres.

La matrice \mathbf{A} peut s'écrire alors sous la forme

$$\mathbf{A}(t) := \begin{bmatrix} a_{11}(t) & a_{12}(t) & \dots & a_{1p}(t) \\ a_{21}(t) & a_{22}(t) & \dots & a_{2p}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1}(t) & a_{r2}(t) & \dots & a_{rp}(t) \end{bmatrix}.$$
 (1.12)

En cas de présence de bruit additif dans l'environnement, le modèle mathématique de mélange relatif à cette situation s'écrit comme suit

$$\mathbf{x}(t) = \mathbf{A} * \mathbf{s}(t) + \mathbf{b}(\mathbf{t}), \tag{1.13}$$

où $\mathbf{b} := (b_1(t), \dots, b_r(t))^T$ est le vecteur des signaux bruits supposé être centré et dont les composantes sont généralement supposées mutuellement indépendantes, et indépendantes des signaux sources.

La version discrète de (1.11) s'écrit

$$\mathbf{x}(n) := [\mathbf{A}(n)] * \mathbf{s}(n), \tag{1.14}$$

où \ast est l'opérateur de convolution discret. La transformée en z donne

$$\mathbf{x}(z) = [\mathbf{A}(z)]\mathbf{s}(z), \tag{1.15}$$

où $\mathbf{A}(z)$ est la transformée en z de $\mathbf{A}(n)$, elle représente la matrice de fonctions de transfert, de dimension $r \times p$. La matrice $\mathbf{A}(z)$ peut s'écrire dans le cas général de la manière suivante

$$\mathbf{A}(z) = \begin{bmatrix} \sum_{k=-\infty}^{+\infty} a_{11}(k) z^{-k} & \sum_{k=-\infty}^{+\infty} a_{12}(k) z^{-k} & \dots & \sum_{k=-\infty}^{+\infty} a_{1p}(k) z^{-k} \\ \sum_{k=-\infty}^{+\infty} a_{21}(k) z^{-k} & \sum_{k=-\infty}^{+\infty} a_{22}(k) z^{-k} & \dots & \sum_{k=-\infty}^{+\infty} a_{2p}(k) z^{-k} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=-\infty}^{+\infty} a_{r1}(k) z^{-k} & \sum_{k=-\infty}^{+\infty} a_{r2}(k) z^{-k} & \dots & \sum_{k=-\infty}^{+\infty} a_{rp}(k) z^{-k} \end{bmatrix}$$

où

 $A_{ij}(z) := \sum_{k=-\infty}^{+\infty} a_{ij}(k) z^{-k}$ est la fonction de transfert de la j^{eme} source relativement au i^{eme} capteur.

 z^{-1} est le terme qui représente l'opérateur de retard.

Pour que le mélange soit réellement convolutif, il faut que pour au moins un couple d'indice (i, j), la fonction de transfert $A_{ij}(z) = \sum_{k=-\infty}^{+\infty} h_{ij}(k) z^{-k}$ ait au moins deux coefficients $a_{ij}(k)$ non nuls. La plupart des travaux dans le cas des mélanges linéaires convolutifs supposent un modèle faisant intervenir des filtres causaux à réponses impulsionnelles finies (RIF). Ainsi, la matrice de filtres $\mathbf{A}(z)$ peut s'écrire

$$\mathbf{A}(z) = \begin{bmatrix} \sum_{k=0}^{K} a_{11}(k) z^{-k} & \sum_{k=0}^{K} a_{12}(k) z^{-k} & \cdots & \sum_{k=0}^{K} a_{1p}(k) z^{-k} \\ \sum_{k=0}^{K} a_{21}(k) z^{-k} & \sum_{k=0}^{K} a_{22}(k) z^{-k} & \cdots & \sum_{k=0}^{K} a_{2p}(k) z^{-k} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=0}^{K} a_{r1}(k) z^{-k} & \sum_{k=0}^{K} a_{r2}(k) z^{-k} & \cdots & \sum_{k=0}^{K} a_{rp}(k) z^{-k} \end{bmatrix} . (1.16)$$

Le schéma général de la relation sources-observations dans ce cas peut être illustré par la figure 1.3. Ce type de mélange constitue un modèle satisfaisant dans des domaines variés tels que les télécommunications [89], l'acoustique [129] ou la mécanique [189]. Il permet de modéliser correctement les effets induits par la propagation réelle des signaux.



FIG. 1.3 – Schéma de propagation pour un mélange convolutif.

Comme pour les mélanges linéaires instantanés, les indéterminations subsistent et s'aggravent dans le cas de mélanges linéaires convolutifs (pour plus des détails voir par exemple [43], [100] et [167]). L'estimation des sources $s_j(n)$, j = 1, ..., p, n'est possible qu'à une indétermination de permutation et de filtrage près. En effet, si l'on prend un filtre quelconque G(z), la contribution d'une source d'indice j sur le capteur i, voir (1.10), peut se réécrire par transformée en z sous la forme

$$x_i(z) = \frac{A_{ij}(z)}{G(z)}(G(z)s_j(z)).$$
(1.17)

La réécriture de (1.15), en remplaçant la source $s_j(z)$ du vecteur $\mathbf{s}(z)$ par $G(z)s_j(z)$ et en remplaçant les éléments $A_{ij}(z)$ de la colonne j de $\mathbf{A}(z)$ par $\frac{A_{ij}(z)}{G(z)}$, est toujours possible. Sans d'autres hypothèses, il n'y a aucun moyen de lever cette indétermination. Cette dernière est relativement complexe et présente une grande gêne à la reconstruction des sources. Notons que la séparation de mélanges convolutifs de sources peut être ramenée à un ensemble de séparations instantanées. Ceci est réalisé par le passage dans le domaine fréquentiel, où nous obtenons un modèle de mélange instantané pour chaque bande de fréquences [124], [39], [166], [186], [150] et [25]. Comme nous l'avons vu, des indéterminations peuvent ainsi se produire à chaque bande. En effet, à cause des permutations, rien ne garantit *a priori* que les contributions fréquentielles extraites sur une sortie donnée pour les différentes bandes de fréquences proviennent de la même source. De plus l'indétermination sur l'amplitude amenée par les méthodes normalisant les signaux de sortie va également perturber la reconstruction des différentes bandes de fréquences. Ainsi l'approche fréquentielle séduisante par sa simplicité possède elle aussi un réel problème d'indétermination pouvant amener à une reconstruction erronée des sources.

Dans la section suivante, nous allons aborder le principe général de la résolution du problème de la SAS.

1.4 Principe de résolution de la SAS

Sous l'hypothèse d'indépendance statistique des sources, la résolution du problème de la SAS consiste à appliquer des transformations aux observations pour obtenir des signaux statistiquement indépendants.

1.4.1 Le cas de mélanges linéaires instantanés

Le but est d'estimer les sources \mathbf{s} à partir des observations \mathbf{x} . L'estimateur s'écrit

$$\mathbf{y}(n) := \mathbf{B}\mathbf{x}(n),\tag{1.18}$$

où \mathbf{B} est la matrice séparante. Le problème consiste à déterminer cette matrice scalaire \mathbf{B} , à partir des observations \mathbf{x} , qui conduit à la "meilleure" estimation possible des sources **s**. Il existe différentes types de méthodes pour résoudre ce problème réparties en deux grandes familles : les méthodes par bloc et les méthodes adaptatives.

Méthode de traitement par bloc

Les méthodes de séparation par bloc sont basées sur le schéma de la figure 1.4.



FIG. 1.4 – Modèle de séparation par bloc pour un mélange linéaire instantané.

Pour ce modèle nous pouvons distinguer deux approches :

- Les approches à une seule étape

Ces approches consistent, à partir des observations, à identifier directement la matrice de mélange ou à trouver une matrice **B** qui, appliquée au vecteur des observations $\mathbf{x} := (x_1, \ldots, x_p)^T$, permet d'obtenir des signaux $(y_1, \ldots, y_p)^T$ statistiquement indépendants.

- Les approches à deux étapes

La première étape consiste à "blanchir" (i.e., décorréler) les signaux observés $\mathbf{x}(n)$; cela se fait par multiplication de $\mathbf{x}(n)$ par la matrice de "blanchiment" \mathbf{W} : les signaux blanchis (décorrélés) s'écrivent donc

$$\mathbf{z}(n) := \mathbf{W}\mathbf{x}(n). \tag{1.19}$$

La deuxième étape consiste à trouver une matrice unitaire (de rotation) \mathbf{U} qui, appliquée aux signaux $\mathbf{z}(n)$, rend ces derniers statistiquement indépendants. Cette dernière étape se fait généralement par optimisation d'un critère d'indépendance.

Méthodes adaptatives

Les méthodes adaptatives de séparation sont généralement schématisées comme il est illustré sur la figure 1.5.



FIG. 1.5 – Modèle de séparation adaptative.

Le principe de ces méthodes consiste à modifier de manière itérative la matrice séparante **B** pour rendre les signaux, reconstitués à partir des observations, statistiquement indépendants. Ces méthodes ont vu le jour grâce à l'algorithme proposé par Jutten [111] et Hérault *et al.* [96], présenté à l'aide d'un réseau de neurones récursif.

Il est à noter que le principe général utilisé pour déterminer la matrice séparante **B** est basé sur l'optimisation d'un critère d'indépendance.

1.4.2 Le cas de mélanges linéaires convolutifs

L'estimateur des signaux sources, dans ce cas, est de la forme

$$\mathbf{y}(n) = \mathbf{B} * \mathbf{x}(n), \tag{1.20}$$

où * est l'opérateur de convolution et **B** est la matrice des filtres séparante. Le problème consiste à trouver cette matrice des filtres **B**, à partir des observations **x**, qui conduit à une bonne estimation des sources **s**. Dans ce but deux structures bien connues sont disponibles : la structure "directe" et la structure "récurrente".

La structure directe

Lorsque chaque sortie du système de séparation est une simple combinaison linéaire des observations, le schéma de la séparation que l'on qualifie de directe [99], est illustré par la figure 1.6. Si $\mathbf{A}(z)$ est une matrice inversible à gauche et si les sources sont statistiquement indépendantes, la solution du problème de séparation est possible à une permutation et à un filtrage près

$$\mathbf{y}(n) = [\mathbf{B}(z)]\mathbf{x}(n) = [\mathbf{B}(z)\mathbf{A}(z)]\mathbf{s}(n), \qquad (1.21)$$

où la matrice des filtres $\mathbf{B}(z)$ vérifie $[\mathbf{B}(z)\mathbf{A}(z)] = [\mathcal{PH}(z)]$, avec \mathcal{P} une permutation et \mathcal{H} un opérateur de filtrage.

La structure récurrente

Elle s'applique lorsque chaque sortie du système de séparation $y_i(n)$ peut être considérée comme une combinaison linéaire d'une observation $x_i(n)$ et des autres sorties $y_j(n)$ (figure 1.7). Pour ce modèle, les signaux estimés sont donnés, sous la forme vectorielle par l'équation suivante

$$\mathbf{y}(z) = \mathbf{x}(z) + \mathbf{V}(z)\mathbf{y}(z), \qquad (1.22)$$



FIG. 1.6 – Structure de séparation directe (dans le cas où p = 2 et r = 2).

ce qui donne

$$\mathbf{y}(z) = [1 - \mathbf{V}(z)]^{-1} \mathbf{x}(z),$$
 (1.23)

à condition que $[1 - \mathbf{V}(z)]^{-1}$ existe et que tous ses pôles soient dans le cercle unité. La matrice de séparation pour la structure directe équivalente est alors

$$\mathbf{B}(z) = [1 - \mathbf{V}(z)]^{-1}.$$
 (1.24)

La figure 1.7 représente la structure de séparation récurrente pour un système de deux sources-deux observations. Dans ce cas les signaux estimés sont donnés comme suit

$$y_1(z) = B_{11}(z)x_1(z) - B_{12}(z)y_2(z),$$

$$y_2(z) = B_{22}(z)x_2(z) - B_{21}(z)y_1(z),$$

et la matrice séparante $\mathbf{B}(z)$ s'écrit

$$\mathbf{B}(z) = \frac{1}{1 - B_{12}(z)B_{21}(z)} \begin{bmatrix} B_{11}(z) & -B_{12}(z) \\ -B_{21}(z) & B_{22}(z) \end{bmatrix}.$$
 (1.25)



FIG. 1.7 – Structure de séparation récurrente (p = 2 et r = 2).

La structure directe et la structure récurrente peuvent être associées pour donner une structure mixte, voir par exemple [63], [137] et [46]. Dans le chapitre 3, nous allons nous restreindre au cas de la structure directe.

1.4.3 Analyse en Composantes Indépendantes (ACI)

La séparation à l'ordre 2 est essentiellement fondée sur la décorrélation des signaux observés. Sauf dans 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 ici qu'apparaissent l'intérêt de l'usage des statistiques d'ordre supérieur en séparation de sources. En utilisant des statistiques d'ordre supérieur à 2, on exploite davantage 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 (ACI) introduite par Jutten et Hérault dans [113], et développée par Comon dans [49, 50]. Cette nouvelle approche vient completer l'Analyse en Composantes Principales (ACP) fondée sur la seule décorrélation, qui est une méthode très utilisée en statistique. L'ACI 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'ACI suivante.

Définition 1.1. L'ACI d'un vecteur aléatoire $\mathbf{x}(n) := (x_1(n), \dots, x_p(n))^T$ est défini comme la donnée d'une matrice de "séparation" **B** telle que le vecteur $\mathbf{y}(n) :=$ $(y_1(n), \dots, y_p(n))^T := \mathbf{B}\mathbf{x}(n)$ soit à composantes indépendantes, dans le sens de la maximisation d'un "contraste" $C(y_1(n), \dots, y_p(n))$ qui mesure l'indépendance des variables aléatoires $y_1(n), \dots, y_p(n)$.

1.5 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 le mélange. 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 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 m_i , i = 1, ..., p. Cette condition d'indépendance se simplifie dans le cas des mélanges linéaires instantanés en posant $m_i = 0$. Cette hypothèse fondamentale est commune à la plupart des méthodes. Elle est réduite 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; voir le paragraphe 1.6.3.

Hypothèse 2. Dans le cas où les signaux sources sont mutuellement statistiquement indépendants et les échantillons de chaque source indépendants et identiquement distribués (iid), on suppose qu'au plus un seul signal source est gaussien.

Comon [49, 50], dans le cadre de l'hypothèse 2 et dans le cas d'un mélange linéaire instantané, a montré que les signaux $\mathbf{y} := \mathbf{B}\mathbf{x}$ sont indépendants si et seulement si la matrice séparante \mathbf{B} est de la forme

$$\mathbf{B} = \mathbf{D}\mathbf{P}\mathbf{A}^{-1},\tag{1.26}$$

où **D** représente une matrice diagonale et **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 forcement à la séparation ; voir [49, 50].

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 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 [112], [2] et [61].

Par ailleurs, d'autres hypothèses supplémentaires peuvent être faites sur les sources. En général ces hypothèses permettent la conception de nouveaux algorithmes, par exemple les algorithmes qui seront présentés aux chapitre 2 et 3 dans lesquels les signaux sources sont supposés stationnaires et au chapitre 4 dans lequel nous traitons le cas des signaux sources cyclostationnaires.

Nous présenterons dans le paragraphe suivant une vue d'ensemble des divers critères utilisés en SAS.

1.6 Critères utilisés en SAS

Comme nous l'avons mentionné précédemment, la plupart des méthodes de la SAS utilisent l'hypothèse 1. Différents critères d'indépendance peuvent être trouvés dans la littérature (voir par exemple [131] et [115]). Par définition, deux variables aléatoires sont dites indépendantes lorsque leur densité de probabilité conjointe est égale au produit de leurs densités marginales. Nous rappelons, dans ce qui suit, quelques critères qui permettent de vérifier l'indépendance des sources et qui sont utilisés en SAS.

1.6.1 Information mutuelle (IM)

L'information mutuelle (IM) est un critère permettant la mesure de la dissimilarité entre la densité conjointe d'un ensemble de signaux et le produit des densités marginales de chacune de ses composantes. Pour un vecteur aléatoire $\mathbf{y} := (y_1, \ldots, y_p)^T$, l'IM $I(\mathbf{y})$, s'écrit

$$I(\mathbf{y}) := \int_{\mathbb{R}^p} f_{\mathbf{y}}(t) \log \frac{f_{\mathbf{y}}(t)}{\prod_{i=1}^p f_{y_i}(t_i)} dt, \qquad (1.27)$$

où $t := (t_1, \ldots, t_p)^T$, $f_{\mathbf{y}}$ est la densité jointe du vecteur aléatoire \mathbf{y} , et f_{y_i} la densité marginale de la variable aléatoire y_i , $i = 1, \ldots, p$. La divergence de Kullback-Leibler modifiée (KL_m) entre les deux densités $\prod_{i=1}^p f_{y_i}$ et $f_{\mathbf{y}}$ est définie par

$$KL_{m}\left(\prod_{i=1}^{p} f_{y_{i}}, f_{\mathbf{y}}\right) := -\int_{\mathbb{R}^{p}} f_{\mathbf{y}}(t) \log \frac{\prod_{i=1}^{p} f_{y_{i}}(t_{i})}{f_{\mathbf{y}}(t)} dt = \int_{\mathbb{R}^{p}} f_{\mathbf{y}}(t) \log \frac{f_{\mathbf{y}}(t)}{\prod_{i=1}^{p} f_{y_{i}}(t_{i})} dt.$$
(1.28)

D'où la relation suivante entre l'information mutuelle et la divergence de Kullback-Leibler modifiée

$$I(\mathbf{y}) = KL_m\left(\prod_{i=1}^p f_{y_i}, f_{\mathbf{y}}\right).$$
(1.29)

Ainsi, en utilisant les propriétés de la fonction logarithme, on peut montrer que l'IM est toujours positive ou nulle. En particulier, elle est nulle si et seulement si $f_{\mathbf{y}} = \prod_{i=1}^{p} f_{y_i}$, i.e., si et seulement si les composantes du vecteur \mathbf{y} sont statistiquement indépendantes

$$I(\mathbf{y}) = 0$$
 ssi y_1, \dots, y_p sont statistiquement indépendantes. (1.30)

Nous pouvons montrer également que l'IM est égale à la différence entre la somme des entropies des variables aléatoires y_1, \ldots, y_p et l'entropie du vecteur aléatoire **y**. Rappelons d'abord la définition de l'entropie d'un vecteur aléatoire ou d'une variable aléatoire. Nous allons utiliser la notation \mathbb{E} pour désigner l'espérance mathématique. L'entropie du vecteur aléatoire **y** s'écrit

$$H(\mathbf{y}) := -\mathbb{E}\left(\log f_{\mathbf{y}}(\mathbf{y})\right) = -\int_{\mathbb{R}^p} f_{\mathbf{y}}(t) \log f_{\mathbf{y}}(t) \, dt.$$
(1.31)

L'entropie de la variable aléatoire y_i , $i = 1, \ldots, p$, s'écrit

$$H(y_i) := -\mathbb{E}\left(\log f_{y_i}(y_i)\right) = -\int_{\mathbb{R}} f_{y_i}(t_i) \log f_{y_i}(t_i) \, dt_i.$$
(1.32)

En utilisant les définitions (1.27), (1.31) et (1.32), un calcul simple conduit à la relation suivante

$$I(\mathbf{y}) = \sum_{i=1}^{p} H(y_i) - H(\mathbf{y}).$$
(1.33)

Vu la relation (1.30), on peut utiliser l'IM comme critère de la SAS. En effet, en définissant la SAS comme une transformation linéaire inversible du vecteur observé $\mathbf{x}(n)$,

$$\mathbf{y}(n) = \mathbf{B}\mathbf{x}(n),\tag{1.34}$$

où **B** est une matrice de dimension $p \times p$; la séparation est obtenue par minimisation de $I(\mathbf{y}) := I(\mathbf{B}\mathbf{x})$ en **B** :

$$\min_{\mathbf{B}} I(\mathbf{B}\mathbf{x}). \tag{1.35}$$

Si on note $\widetilde{\mathbf{B}}$:= $\operatorname{argmin}_{\mathbf{B}} I(\mathbf{Bx})$, le vecteur $\widetilde{\mathbf{y}}$:= $\widetilde{\mathbf{Bx}}$ est donc de composantes statistiquement indépendantes. En utilisant la relation (1.34), on peut montrer que

$$H(\mathbf{y}) = H(\mathbf{x}) + \log |\det(\mathbf{B})|, \qquad (1.36)$$

puisque nous avons

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{f_{\mathbf{x}}(\mathbf{x})}{|\det(\mathbf{B})|}.$$
(1.37)

L'application du logarithme à l'équation (1.37) nous donne

$$\log f_{\mathbf{y}}(\mathbf{y}) = \log f_{\mathbf{x}}(\mathbf{x}) - \log |\det(\mathbf{B})|.$$
(1.38)

D'où, en prenant l'espérance, on obtient

$$H(\mathbf{y}) = H(\mathbf{x}) + \log |\det(\mathbf{B})|.$$
(1.39)

L'équation (1.33) peut donc s'écrire

$$I(\mathbf{y}) = \sum_{i=1}^{p} H(y_i) - H(\mathbf{x}) - \log |\det(\mathbf{B})|.$$
(1.40)

L'IM a été introduite pour la première fois comme un critère de séparation de sources par Comon [50]. Comme le calcul de la relation (1.27) est impossible car il dépend de densités inconnues; la séparation de signaux est réalisée en minimisant un "estimateur" de (1.27) ou de (1.40). Nous pouvons citer à ce sujet, parmi d'autres, les travaux de [152], [9], [66] et [149]. Dans ces travaux le critère (1.27) a été estimé par
différentes méthodes : la méthode à noyau, la méthode des espacements, la méthode polynomiale, ... etc

Notons qu'il existe une relation très forte entre l'IM et la méthode d'estimation statistique par maximum de vraisemblance.

1.6.2 Maximum de vraisemblance

C'est une méthode qui est très utilisée en estimation statistique; voir par exemple [91] et [122]. L'objectif est de chercher les paramètres du mélange qui maximisent la probabilité d'occurrence des observations. Dans un mélange linéaire instantané, l'expression du vecteur signaux mélanges est

$$\mathbf{x}(n) := \mathbf{As}(n). \tag{1.41}$$

Nous pouvons exprimer la densité de probabilité du vecteur $\mathbf{x}(n)$ en fonction des densités de probabilité des sources $\mathbf{s}(n)$ et du déterminant de la matrice inverse **B** de la matrice de mélange **A**. On a

$$f_{\mathbf{x}}(\mathbf{x}(n)) = |\det(\mathbf{B})| f_{\mathbf{s}}(\mathbf{s}(n)) = |\det(\mathbf{B})| \prod_{i=1}^{p} f_{s_i}(s_i(n)).$$
(1.42)

Cette densité peut aussi être donnée en fonction des lignes $\mathbf{b_i}$ de la matrice \mathbf{B} par la relation

$$f_{\mathbf{x}}(\mathbf{x}(n)) = |\det(\mathbf{B})| \prod_{i=1}^{p} f_{s_i}(\mathbf{b}_i \mathbf{x}(\mathbf{n})).$$
(1.43)

Considérons un ensemble de N échantillons indépendants et de même loi du vecteur $\mathbf{x} : \mathbf{x}(n), n = 1, \dots, N$. La vraisemblance V de l'obtention de cet ensemble peut s'écrire comme le produit de cette densité évaluée à ces N points

$$V(\mathbf{B}) := \prod_{n=1}^{N} \left[|\det(\mathbf{B})| \prod_{i=1}^{p} f_{s_i}(\mathbf{b_ix}(n)) \right].$$
(1.44)

Il est souvent plus pratique d'utiliser le logarithme de la vraisemblance qui s'écrit

$$\log V(\mathbf{B}) = \sum_{n=1}^{N} \sum_{i=1}^{p} \log f_{s_i}(\mathbf{b_i x}(n)) + N \log |\det(\mathbf{B})|.$$
(1.45)

En divisant par le nombre d'échantillons N, nous avons

$$\frac{1}{N}\log V(\mathbf{B}) = \sum_{i=1}^{p} \frac{1}{N} \sum_{n=1}^{N} \log f_{s_i}(\mathbf{b_ix}(n)) + \log |\det(\mathbf{B})|.$$
(1.46)

D'autre part, par la loi des grands nombres, on a

$$\frac{1}{N}\sum_{n=1}^{N}\log f_{s_i}(\mathbf{b_ix}(n)) \approx \mathbb{E}\left(\log f_{s_i}(\mathbf{b_ix})\right)$$

lorsque N est suffisamment grand. D'où

$$\frac{1}{N}\log V(\mathbf{B}) \approx \sum_{i=1}^{p} \mathbb{E}\{\log f_{s_i}(\mathbf{b_ix})\} + \log |\det(\mathbf{B})|, \qquad (1.47)$$

ce qui donne

$$\frac{1}{N}\log V(\mathbf{B}) \approx -\sum_{i=1}^{p} H(\mathbf{b}_{i}\mathbf{x}) + \log |\det(\mathbf{B})| =: \sum_{i=1}^{p} H(y_{i}) + \log |\det(\mathbf{B})|. \quad (1.48)$$

Si nous faisons une comparaison entre les équations (1.48) et (1.40), et comme $H(\mathbf{x})$ représente une constante par rapport à la matrice de séparation **B**, nous avons

$$I(\mathbf{y}) \approx -\frac{1}{N} \log V(\mathbf{B}) + C^{te}.$$
(1.49)

Le critère de l'IM est égal à une constante près à l'opposé du critère de log-vraisemblance. Gaeta et Lacoume [75, 76] ont montré les premiers la possibilité d'utiliser le principe du maximum de vraisemblance pour réaliser la séparation de sources dans le cas de mélanges linéaires instantanés ou convolutifs. Ensuite, Pham *et al.* dans [155, 154] ont proposé des méthodes pour résoudre le problème de séparation de source basées sur le principe du maximum de vraisemblance. De nombreuses autres contributions considérant le maximum de vraisemblance pour la SAS ont été présentées, citons par example Belouchrani et Cardoso [16, 17], Cardoso [41], Amari [5] et Moulines *et al.* [141].

1.6.3 Statistiques d'ordre deux (SO2)

La décorrélation (l'indépendance à l'ordre 2) est la première étape vers l'indépendance (à tout ordre) des signaux estimés. Pour parvenir à l'indépendance des signaux estimés nous utilisons la matrice de covariance des observations afin d'identifier la matrice de mélange. La matrice de covariance des signaux observés s'écrit

$$R_{\mathbf{x}} := \mathbb{E}\{\mathbf{x}(t)\mathbf{x}(t)^{T}\} = \mathbb{E}\{(\mathbf{A}\mathbf{s}(t))(\mathbf{A}\mathbf{s}(t))^{T}\} = \mathbf{A}\mathbb{E}\{\mathbf{s}(t)\mathbf{s}(t)^{T}\}\mathbf{A}^{T} =: \mathbf{A}R_{\mathbf{s}}\mathbf{A}^{T}.$$
(1.50)

Vu que les sources sont supposées indépendantes, la matrice R_s est donc diagonale. Si nous supposons en plus que les sources sont normalisées, cela conduit la matrice R_s à devenir une matrice identité $R_s = \mathbf{I}$. Dans ce cas la matrice de covariance devient

$$R_{\mathbf{x}} = \mathbf{A}\mathbf{A}^T.$$

En effet, nous constatons que la matrice \mathbf{A} est une racine carrée de la matrice d'autocovariance $R_{\mathbf{x}}$. Cette propriété ne permet pourtant pas à elle seule d'identifier la matrice de mélange \mathbf{A} puisqu'il n'y a pas unicité des racines carrées d'une matrice. En effet, il est possible de montrer que les racines carrées de $R_{\mathbf{x}}$ seront toutes de la forme $\mathbf{A}\mathbf{U}$, où \mathbf{U} est une matrice unitaire quelconque (c'est-à-dire une matrice carrée vérifiant $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$). Par conséquent, nous constatons que nous ne pouvons pas trouver une inverse à gauche de la matrice \mathbf{A} , la séparation à l'ordre deux devenant impossible sans d'autres hypothèses supplémentaires. Autrement dit, la décorrélation des signaux de sortie ne suffit pas à séparer les sources. En effet, si on multiple le vecteur source $\mathbf{s}(t)$ par une matrice orthogonale \mathbf{A} quelconque, alors les composantes du vecteur $\mathbf{x}(t)$ résultant sont décorrélées elles aussi. Cependant, on peut quand même effectuer la séparation à l'aide des statistiques d'ordre 2 des signaux en imposant l'une des hypothèses 5 ou 6 suivantes.

Hypothèse 5. Les sources sont colorées.

Si l'hypothèse 5 est vérifiée, alors il est possible de séparer des sources en imposant l'intercorrélation temporelle des signaux de sortie décalés nulle

$$\exists \tau / \mathbb{E}\{y_i(t)y_j(t-\tau)\} = 0, \ \forall i \neq j.$$

$$(1.51)$$

En effet, les matrices d'autocovariance des sources centrées $\mathbf{s}(t)$ définies par

$$R_{\mathbf{s}}(\tau) := \mathbb{E}\{\mathbf{s}(t)\mathbf{s}(t-\tau)^T\},\tag{1.52}$$

sont réelles et diagonales. Les matrices d'autocovariance des observations $\mathbf{x}(t)$ s'écrivent alors

$$R_{\mathbf{x}}(\tau) := \mathbb{E}\left\{\mathbf{x}(t)\mathbf{x}(t-\tau)^{T}\right\} = \mathbf{A}R_{\mathbf{s}}(\tau)\mathbf{A}^{T},$$
(1.53)

et sont simultanément diagonalisables pour divers décalages τ . Parmi les méthodes statistiques d'ordre 2 connues qui supposent les sources colorées, citons AMUSE [176] ou SOBI [15]. AMUSE (pour Algorithm for Multiple Unknown Signals Extraction) suppose que les sources sont mutuellement non-corrélées avec des autocorrélations normalisées différentes non nulles, i.e.,

$$\exists t_1, t_2 / \mathbb{E}\{s_i(t_1)s_j(t_2)\} = 0, \ \forall i \neq j$$
(1.54)

 et

$$\exists \tau > 0 / \frac{\mathbb{E}\{s_i(t)s_i(t-\tau)\}}{\mathbb{E}\{s_i(t)^2\}} \neq \frac{\mathbb{E}\{s_j(t)s_j(t-\tau)\}}{\mathbb{E}\{s_j(t)^2\}}, \ \forall i \neq j.$$
(1.55)

Après avoir blanchi les observations (le vecteur des observations blanchies est noté $\mathbf{z}(t)$), la méthode utilise la fonction d'autocovariance $R_{\mathbf{z}}(\tau)$ pour une seule valeur du décalage τ non-nulle. La matrice de mélange est estimée en reprenant le formalisme de l'équation (1.53) appliquée aux observations blanchies. Récemment, une version convolutive temporelle d'AMUSE a été proposée par Teixeira *et al.* [172]. SOBI (pour Second Order blind Identification) est une extension d'AMUSE. La méthode SOBI consiste à diagonaliser plusieurs matrices $R_{\mathbf{z}}(\tau)$ pour plusieurs valeurs différentes et non-nulles de τ . Il suffit que la condition (1.55) soit satisfaite pour l'un des décalages temporels τ choisis pour que l'on puisse séparer les sources, ce qui représente un avantage par rapport à AMUSE qui n'utilise qu'un seul décalage. Une version convolutive temporelle de SOBI a été proposée par Bousbia-Salah *et al.* [26].

Hypothèse 6. Les sources sont non-stationnaires.

Si l'hypothèse 6 est vérifiée, alors les statistiques qui varient dans le temps fournissent de l'information supplémentaire. Dans ce cas, la décorrélation des sorties entre elles à plusieurs instants suffit pour effectuer la séparation. En effet, en raison de la non-stationnarité à l'ordre 2, la covariance de $\mathbf{y}(t)$ dépend de l'instant t, ce qui engendre de nouvelles contraintes par rapport au simple blanchiment. Dans [169], Souloumiac propose de couper le signal temporel en deux et de calculer la matrice de covariance $R_{\mathbf{x}}(0)$ sur chaque zone temporelle, puis de diagonaliser conjointement ces deux matrices. Pham et Cardoso [154] ont étendu la méthode de Souloumiac en prenant plus de deux zones pour calculer la covariance. Choi et al. ont proposé une méthode, SONS (pour Second Order Non-stationary source Separation) [47], qui peut être vue comme une extension des méthodes de Pham et Cardoso et de SOBI : comme dans [154], le signal temporel est découpé en plusieurs zones, et les observations sont d'abord blanchies puis des matrices de covariance $R_{\mathbf{z}}(\tau)$ $(\tau \neq 0)$ sont calculées sur chacune de ces zones avant d'être conjointement diagonalisées. Dans [92], Hosseini et Deville proposent deux méthodes basées sur les propriétés statistiques dans le domaine fréquentiel. Ces méthodes sont appelées méthodes à corrélation spectrale. La première est basée sur la diagonalisation d'une matrice calculée à partir des statistiques d'ordre 2 des signaux exprimés dans le domaine de Fourier. Elle est utilisée en particulier dans le cadre de signaux cyclostationnaires. La deuxième méthode, étudiée aussi dans [93], fait l'hypothèse de signaux sources blancs et non-stationnaires et travaille dans le domaine fréquentiel pour obtenir des sources autocorrélées et stationnaires. Elle fait ensuite appel à des méthodes temporelles existantes, comme AMUSE ou SOBI, transposées dans le domaine fréquentiel. D'autres approches à l'ordre 2 linéaires instantanées ou convolutives ont été proposées par Féty [73], Kawamoto *et al.* [118], Parra et Spence [150], Pham *et al.* [157], Rahbar *et al.* [162], Boustany et Antoni [28], Buchner *et al.* [35], Bouguerriou *et al.* [22], Ohata *et al.* [146], Keziou *et al.* [121] et Ould *et al.* [148]. Un état de l'art plus complet sur les méthodes statistiques d'ordre 2 peut être trouvé dans les thèses de Boumaraf [24], Thomas [175] et Puigt [160].

1.6.4 Statistiques croisées d'ordre supérieur

Dans le cas général, deux signaux sont spatialement indépendants si et seulement si tous leurs cumulants croisés sont nuls. Ainsi, comme nous l'avons déjà mentionné, l'indépendance entre deux signaux se situe bien au-delà de la simple décorrélation. Cependant, l'annulation de tous les cumulants, i.e., à tous les ordres et pour tous les instants disponibles est impossible car la charge de calcul nécessaire serait infinie. L'indépendance est alors approchée à un certain ordre. Dans la plupart des cas, l'ordre 4 est considéré comme étant satisfaisant. Malgré la complexité accrue, les méthodes aux ordres supérieurs semblent donc plus indiquées que la décorrélation pour le problème de la SAS.

Moments croisés

Les moments croisés non-linéaires sont définis à partir de deux fonctions g et h (qui ne doivent pas être toutes les deux linéaires) par la formule

$$C_{(g,h)}(i,j) = \mathbb{E}\{g(y_i(n))h(y_j(n))\}, \quad \forall i \neq j,$$

$$(1.56)$$

et ont donné lieu aux premiers algorithmes de la SAS développés par Hérault et Ans [94], Hérault *et al.* [96]. Pour annuler la fonction $C_{(g,h)}(i, j)$, ils utilisent l'algorithme

adaptatif

$$c_{ij}(n+1) = c_{ij}(n) + \mu_n g(y_i(n)) h(y_j(n)), \qquad (1.57)$$

où $i \neq j$, μ_n est le pas d'adaptation et $c_{ij}(n)$ est le coefficient de séparation évalué à l'instant n.

Pour la séparation de mélanges convolutifs, l'algorithme de Hérault et Jutten a été transposé dans le domaine fréquentiel par Back et Tsoi [12]. Nguyen-Thi et Jutten [143] ont aussi étendu l'algorithme d'origine au cas convolutif, dans le domaine temporel, pour le cas de deux sources et deux observations. Ils supposent que les filtres croisés de mélanges A_{12} et A_{21} sont à réponse impulsionnelle finie (RIF), linéaires et causaux. Les filtres A_{11} et A_{22} sont, eux supposés égaux à 1 et le mélange est à phase minimale. Ils utilisent une structure de séparation récursive qui recherche l'annulation des moments croisés, i.e.,

$$\mathbb{E}\{g(y_i(n))h(y_j(n-k))\} = 0, \quad \forall i \neq j, \quad \forall k = 0, \dots, L,$$

$$(1.58)$$

où L est l'ordre du filtre. Cette méthode a été étudiée de façon approfondie par Servière [165] et une étude sur la séparation des mélanges convolutifs par les réseaux de type Hérault-Jutten [114, 95] a été faite par Berthommier et Choi [20]. Les fonctions g et h les plus utilisées sont

$$g(x) = x^3,$$
$$h(x) = x.$$

En utilisant ces fonctions, le critère revient à l'annulation du moment croisé (3,1) des signaux de sortie. On montre, par ce procédé, que seuls les mélanges, où les sources sont globalement sous-gaussiennes, sont inversibles. Les principaux avantages de cette méthode sont le coût de calcul réduit et la faible complexité d'implémentation. Cependant, en raison de l'utilisation d'une structure récursive, cette technique souffre de grosses limitations sur le mélange qui doit impérativement être à phase minimale. L'optimisation de ces fonctions g et h est donc très importante pour la qualité de la convergence et dépend du type de signaux présents. Pour des mélanges convolutifs à phase minimale, Charkani et Deville [44] ont montré qu'en optimisant ces fonctions, il est possible de séparer des sources sur-gaussiennes et sous-gaussiennes. Des implantations en situations réelles ont alors été réalisées avec succès dans la thèse de Charkani [43].

Cumulants croisés

Le cumulant croisé d'ordre 4, $Cum_4(y_i, y_j, y_j, y_j)$ est un autre critère d'indépendance utilisé pour la première fois par Lacoume et Ruiz [123, 124] et Comon [48, 49]. On peut généraliser les méthodes basées sur les cumulants croisés par une approche tensorielle.

Définition 1.2. Un tenseur de cumulant **F** est un opérateur défini par les cumulants croisés d'ordre 4 des données, i.e.,

$$\mathbf{F}_{ijkl} := Cum_4(y_i, y_j, y_k, y_l). \tag{1.59}$$

 \mathbf{F} est diagonal dans le cas de signaux de sortie y_i , y_j , y_k et y_l indépendants. La diagonalisation de ce tenseur \mathbf{F} est réalisée en pratique via l'identification de matrices propres pour la transformation linéaire

$$\mathbf{F}_{ij}(\mathbf{M}) := \sum_{kl} m_{kl} Cum_4(y_i, y_j, y_k, y_l), \qquad (1.60)$$

où m_{kl} est un élément de la matrice **M** qui est transformée. Comme tout opérateur linéaire symétrique, le tenseur **F** a une décomposition en valeurs propres. Une matrice propre du tenseur est, par définition, une matrice vérifiant

$$\mathbf{F}(\mathbf{M}) = \lambda \mathbf{M}.\tag{1.61}$$

Les matrices propres du tenseur seront notées par la suite \mathbf{M}_i . Cette décomposition en valeurs propres correspond à une diagonalisation conjointe des matrices $\mathbf{F}(\mathbf{M}_i)$. L'algorithme JADE [42], en choisissant pour matrices \mathbf{M}_i les matrices propres du tenseur, se ramène à la maximisation du critère

$$J_{jade} = \sum_{i,k,l=1}^{p} |Cum_4(y_i, y_i, y_k, y_l)|^2.$$
(1.62)

Dans [188], Yeredor a proposé une diagonalisation conjointe non-orthogonale au sens des moindres carrés et l'a appliquée à JADE. Ferréol *et al.* ont étendu la méthode SOBI à l'ordre quatre sous le nom de FOBIUM (Fourth Order Blind Identification of Underdetermined Mixtures of sources) [69]. En outre, l'algorithme FOBIUM permet de traiter des mélanges sous-déterminés de sources. D'autres approches à l'ordre 4 ont été proposées par Cardoso [40], Comon [52], Pedersen et Nielsen [151]. Un état de l'art plus complet sur les méthodes statistiques d'ordre supérieur peut être trouvé dans les références [131] et [115].

Il est à noter que les statistiques d'ordre 3 sont très peu utilisées dans la séparation de sources car elles décrivent la dissymétrie des densités de probabilité. Or, la plupart des signaux réels traités ont des densités de probabilité symétriques.

1.6.5 Fonctions de contraste

Le concept des fonctions de contraste pour la séparation de sources a été introduit par Comon [49, 50]. 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 séparation. Elle est définie de la façon suivante

Définition 1.3. Une fonction de contraste $\Psi(\cdot)$ est une application à valeurs dans \mathbb{R} définie sur l'espace de vecteurs aléatoires \mathbf{y} de \mathbb{R}^p , ne dépendant que de la loi de probabilité de \mathbf{y} et qui vérifie les propriétés suivantes

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

$$\Psi(\mathbf{S}\mathbf{y}) \le \Psi(\mathbf{y})$$

et

$$[\Psi(\mathbf{Sy}) = \Psi(\mathbf{y})] \iff [\mathbf{S} = \mathbf{DP}]$$

où \mathbf{D} est une matrice diagonale et \mathbf{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 $\Psi(\mathbf{Sy})$ et que tous les maxima globaux sont des solutions de séparation. D'ailleurs, Comon [50] 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. Par ailleurs, E. Moreau et J.C. Pesquet [138] 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. [139, 140] 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 Comon trouve une solution analytique, nommée COM1 (COntrast Maximization 1) [51], au problème d'optimisation du contraste défini, au signe près, comme la somme des kurtosis des sources estimées. Ce contraste a été initialement présenté par Moreau et al. dans [136, 53, 174]. En outre, Moreau montre que ce critère est un contraste à la condition que les kurtosis des sources soient de même signe.

Notons que les propriétés de la définition (1.3) sont aussi vérifiées pour les critères basés sur les "divergences" que nous seront amenés à exploiter par la suite pour séparer les sources.

1.7 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 s_i par des critères de type erreur quadratique moyenne ou rapport signal sur résidus.

1.7.1 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, ..., p. Ce terme s'écrit comme suit

$$EQM_i := \widehat{\mathbb{E}}\{(s_i - y_i)^2\} := \frac{1}{N} \sum_{n=1}^N [s_i(n) - y_i(n)]^2, \qquad (1.63)$$

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

$$EQM := \frac{1}{p} \sum_{i=1}^{p} EQM_i = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{N} \sum_{n=1}^{N} [s_i(n) - y_i(n)]^2.$$
(1.64)

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

1.7.2 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 (en supposant qu'il n'y a pas de permutation)

$$SNR_i := 10 \log_{10} \frac{\widehat{\mathbb{E}}\{s_i^2\}}{\widehat{\mathbb{E}}\{(y_i - s_i)^2\}}, \ i = 1, \dots, p.$$
(1.65)

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

$$SNR_i := 10 \log_{10} \frac{\widehat{\mathbb{E}}[y_i^2]}{\widehat{\mathbb{E}}[y_i^2 \mid_{s_i=0}]}, \ i = 1, \dots, p,$$
 (1.66)

où $[y_i |_{s_i=0}]$ 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 provenant d'autres sources $(s_j, j = 1, ..., p \text{ et } j \neq i)$ à cette sortie y_i .

1.8 Conclusion

Dans ce chapitre qui ne peut être exhaustif, nous avons présenté le problème de la SAS et les éléments qui s'y rattachent. Nous avons commencé par donner quelques exemples d'application de la SAS dans divers domaines. Après avoir répertorié les deux modèles de mélanges (linéaire et non-linéaire), nous avons classé le type de mélange linéaire en deux catégories : le modèle de mélange linéaire instantané, qui utilise une matrice de coefficients scalaires pour reformuler la relation entre les observations et les sources et le modèle de mélange linéaire convolutif qui est modélisé par un filtrage entre les sources et les observations. Après avoir présenté cette modélisation des mélanges, nous avons exposé le principe de résolution du problème de la SAS dans le cas instantané puis dans le cas convolutif. Pour le cas de mélange linéaire instantané, nous avons rappelé différentes méthodes pour résoudre le problème de la SAS réparties en deux familles : les méthodes par bloc (s) et les méthodes adaptatives. Nous avons aussi distingué deux structures bien connues et disponibles pour le mélange convolutif : la structure récurrente et la structure directe. Nous avons expliqué que la solution fournie par les algorithmes de SAS n'est pas unique car les sources sont estimées à une permutation et à un facteur d'échelle près (dans le cas instantané) ou à une permutation et à un filtrage près (dans le cas convolutif). Nous avons rappelé le principe de l'analyse en composantes indépendantes, des hypothèses de base nécessaires à la séparation de sources, et plusieurs critères permettant de réaliser la séparation. Enfin, nous avons établi des mesures de performances afin d'étudier les qualités d'estimation des sources par les critères que nous proposons et la performance de nos algorithmes à suivre. Nous avons plus particulièrement insisté sur les algorithmes de Jade et Sobi car ils nous serviront de référence pour juger des améliorations (en terme de séparation) apportées par les algorithmes que nous serons amenés à développer. Dans la suite, nous proposerons de nouvelles méthodes de séparation aveugle de sources dans le cadre des mélanges linéaires des signaux stationnaires (partie I). Tandis que, dans la partie II, nous proposerons de nouveaux algorithmes dans le cadre des signaux non-stationnaires, plus précisément des signaux cyclostationnaires.

Première partie

SAS stationnaires

Chapitre 2

Mélanges linéaires instantanés

2.1 Introduction

Nous rappelons que les premières méthodes de SAS ont été élaborées au milieu des années 80 [94, 96]. Ensuite, Comon [49, 50] a défini rigoureusement le cadre mathématique de l'Analyse en Composantes Indépendantes, à partir de laquelle dérivent plusieurs applications. La SAS est une technique de traitement qui permet de restituer un ensemble de signaux non observables appelés sources à partir d'un ensemble de signaux mesurés appelés observations. La plupart des méthodes de séparation aveugle de sources supposent l'indépendance statistique de celles-ci; elles consistent donc à rendre les observations indépendantes au sens statistique. Certaines de ces méthodes se contentent de rendre les observations indépendantes à l'ordre deux (décorrélation ou blanchiment) [15], d'autres utilisent des statistiques d'ordres supérieurs (voir par exemple [135]). D'autres encore sont basées sur la minimisation de l'Information Mutuelle (voir par exemple [50, 171, 4, 152, 67]). De ce fait, et compte tenu de la diversité de ces méthodes et des applications qui en découlent, on observe une recherche incessante de nouvelles méthodes plus performantes. Dans ce qui suit, dans le cadre des mélanges linéaires instantanés, nous présentons deux nouvelles méthodes de la SAS que nous avons mises au point. Une première méthode est basée sur la minimisation de l'information mutuelle (IM) sous contraintes et utilisation du critère AIC (Akaike Information Criterion) pour la sélection de modèles des densités et estimation des paramètres par maximum de vraisemblance. La deuxième méthode est basée sur la minimisation des divergences entre densités de probabilité. Sans perte de généralité, nous allons considérer le cas de deux sources-deux observations (p = 2).

2.2 Méthode basée sur la minimisation de l'IM sous contraintes

2.2.1 Introduction

La plupart des méthodes de SAS supposent l'indépendance statistique des sources; sous cette dernière hypothèse, ces méthodes consistent à rendre les observations indépendantes au sens statistique. Dans le cas linéaire instantané, le mélange de deux sources est de la forme

$$\mathbf{x} = \mathbf{As},$$

où $\mathbf{s} := (s_1, s_2)^T$, $\mathbf{x} := (x_1, x_2)^T$ et \mathbf{A} sont respectivement les signaux sources, les signaux mélanges observés et la matrice de mélange de dimension 2×2 . Le but est d'estimer les sources \mathbf{s} à partir des observations \mathbf{x} . L'estimateur, dans ce cas, est de la forme

$$\mathbf{y} = \mathbf{B}\mathbf{x},$$

où **B** est une matrice 2×2 . Le problème consiste donc à chercher un estimateur $\widehat{\mathbf{B}}$, à partir des observations **x**, qui conduit à la meilleure estimation possible des sources **s**,

$$\mathbf{y} = \widehat{\mathbf{B}}\mathbf{x} \simeq \mathbf{s}$$

Si A est inversible, si chacun des signaux sources est une suite i.i.d, si l'hypothèse d'indépendance mutuelle des sources s_1 et s_2 est vérifiée et si au plus une source est gaussienne, une solution à ce problème, à une indétermination du facteur d'échelle et à une permutation près, consiste à chercher une matrice $\widehat{\mathbf{B}}$ qui rende les composantes de y indépendantes au sens statistique, c.f. [42] et [50]. Dans ce qui suit on suppose que ces hypothèses sont vérifiées. La minimisation de l'IM du vecteur y est considérée comme le critère idéal pour l'estimation de **B**. Cela conduit à l'estimation des densités marginales f_{y_1} et f_{y_2} et des fonctions score marginales associées que nous définirons dans le paragraphe à suivre. Plusieurs méthodes d'estimation paramétriques et non paramétriques des densités ont été proposées dans la littérature, c.f. [153] et [9]. Nous proposons dans ce travail, une autre approche paramétrique d'estimation basée sur la sélection d'un modèle, via la minimisation du critère d'information de Akaike (AIC) [3], parmi une famille croissante de modèles de lois exponentielles de dimensions quelconques. Les paramètres du modèle sélectionné sont estimés par maximum de vraisemblance. Les estimateurs des fonctions scores sont obtenus par dérivation des densités estimées. Enfin pour remédier au problème de l'indétermination du facteur d'échelle, et stabiliser l'algorithme de minimisation, nous proposons de minimiser l'information mutuelle sous la contrainte de normalisation des puissances des sources estimées.

Avant d'expliciter la méthode, il convient de revenir sur la définition de l'information mutuelle et de rappeler les définitions de l'algorithme du gradient et des fonctions scores.

2.2.2 Définitions

Information mutuelle

Définition 2.1. Soient $\mathbf{y} := (y_1, y_2)^T$ un vecteur aléatoire, $f_{\mathbf{y}}$ la densité de probabilité jointe de \mathbf{y} et $f_{y_i}, i \in \{1, 2\}$, les densités de probabilité marginales des composantes $y_i, i \in \{1, 2\}$, du vecteur \mathbf{y} . L'information mutuelle du vecteur aléatoire \mathbf{y} est définie comme étant la divergence de Kullback-Leibler modifiée entre les densités $\prod_{i=1}^2 f_{y_i}$ et $f_{\mathbf{y}}$, *i.e.*,

$$I(\mathbf{y}) := KL_m\left(\prod_{i=1}^2 f_{y_i}, f_{\mathbf{y}}\right) := -\mathbb{E}\left(\log\frac{\prod_{i=1}^2 f_{y_i}(y_i)}{f_{\mathbf{y}}(\mathbf{y})}\right) = -\int_{\mathbb{R}^2} f_{\mathbf{y}}(t)\log\frac{\prod_{i=1}^2 f_{y_i}(t_i)}{f_{\mathbf{y}}(t)}\,dt$$

$$(2.1)$$

 $o\hat{u} t := (t_1, t_2)^T.$

À partir des propriétés de la divergence de Kullback-Leibler modifiée, nous déduisons que l'information mutuelle est toujours positive et n'est nulle que si et seulement si $f_{\mathbf{y}} = \prod_{i=1}^{2} f_{y_i}$, c'est-à-dire, ssi les composantes y_1, y_2 sont indépendantes. Autrement dit,

$$\min_{\mathbf{B}} I(\mathbf{y}) = 0, \tag{2.2}$$

et le minimum est atteint en $\mathbf{B} = \mathbf{D}\mathbf{P}\mathbf{A}^{-1}$, où \mathbf{D} représente une matrice diagonale et \mathbf{P} une matrice de permutation.

Algorithme du gradient

Supposons que dans un algorithme de séparation des sources, l'information mutuelle des sources estimées ait été choisie comme critère d'indépendance. Dans ce cas, si nous utilisons l'algorithme de descente du gradient pour la minimisation, nous n'avons pas besoin d'estimer $I(\mathbf{y})$ elle-même, mais seulement sa dérivée par rapport aux coefficients de la matrice séparante **B**. L'IM dépend de la densité jointe $f_{\mathbf{y}}$, d'où, sa dérivée par rapport à **B** dépend des dérivées partielles de $f_{\mathbf{y}}$. La matrice de séparation **B** doit être estimée de manière à ce que l'information mutuelle $I(\mathbf{y})$ de $\mathbf{y} = \mathbf{B}\mathbf{x}$ soit minimisée. L'algorithme de descente du gradient

$$\mathbf{B} \leftarrow \mathbf{B} - \mu \frac{dI(\mathbf{y})}{d\mathbf{B}},\tag{2.3}$$

,

nécessite donc l'estimation de la dérivée $\frac{dI(\mathbf{y})}{d\mathbf{B}}$. Cette étape passe par l'utilisation des fonctions scores.

Fonctions scores

Tout d'abord, nous rappelons la définition de la fonction score d'une variable aléatoire réelle.

Définition 2.2. La fonction score d'une variable aléatoire scalaire y est égale à l'opposé de la dérivée du logarithme de sa densité, i.e.,

$$\psi(y) := -\frac{d}{dy} \log f_y(y) = -\frac{f_y(y)'}{f_y(y)}$$

où f_y désigne la densité de probabilité de y.

Dès lors, nous déduisons de cette définition différents types de fonctions scores pour un vecteur aléatoire $\mathbf{y} := (y_1, y_2)^T$.

Définition 2.3. La fonction score marginale (FSM) d'un vecteur aléatoire **y** est définie comme étant le vecteur des fonctions scores de ses composantes, i.e.,

$$\boldsymbol{\psi}_{\boldsymbol{y}}(\mathbf{y}) := \left(\psi_1(y_1), \psi_2(y_2)\right)^T,$$

оù

$$\psi_i(y_i) := -\frac{d}{dy_i} \log f_{y_i}(y_i) = -\frac{f_{y_i}(y_i)'}{f_{y_i}(y_i)} , \quad i = 1, 2.$$

Définition 2.4. La fonction score jointe (FSJ) du vecteur \mathbf{y} est définie comme étant le vecteur gradient de $(-\log f_{\mathbf{y}}(\mathbf{y}))$, i.e.,

$$\boldsymbol{\phi}_{\boldsymbol{y}}(\mathbf{y}) := \left(\phi_1(\mathbf{y}), \phi_2(\mathbf{y})\right)^T,$$

 $o \dot{u}$

$$\phi_i(\mathbf{y}) := -\frac{\partial}{\partial y_i} \log f_{\mathbf{y}}(\mathbf{y}) = \frac{-\frac{\partial f_{\mathbf{y}}(\mathbf{y})}{\partial y_i}}{f_{\mathbf{y}}(\mathbf{y})} \quad , \quad i = 1, 2.$$

Définition 2.5. La fonction score différentielle (FSD) du vecteur **y** est définie comme étant la différence entre les fonctions FSM et FSJ, i.e.,

$$\boldsymbol{\beta}_{\boldsymbol{y}}(\mathbf{y}) := \boldsymbol{\psi}_{\boldsymbol{y}}(\mathbf{y}) - \boldsymbol{\phi}_{\boldsymbol{y}}(\mathbf{y}).$$

Pour calculer les fonctions scores nous devons donc estimer les densité de probabilité qui sont *a priori* inconnues.

2.2.3 IM sous contraintes

L'approche que nous proposons est basée sur la minimisation d'un "estimateur" de l'information mutuelle

$$I(\mathbf{y}) := \mathbb{E}\left(\log\frac{f_{\mathbf{y}}(\mathbf{y})}{f_{y_1}(y_1)f_{y_2}(y_2)}\right) = \int_{\mathbb{R}^2} f_{\mathbf{y}}(t_1, t_2)\log\frac{f_{\mathbf{y}}(t_1, t_2)}{f_{y_1}(t_1)f_{y_2}(t_2)} dt_1 dt_2 \qquad (2.4)$$

sous les contraintes

$$(\operatorname{var}(y_1) - 1)^2 = 0$$
 et $(\operatorname{var}(y_2) - 1)^2 = 0$,

où

$$\operatorname{var}(y_i) := \mathbb{E}\left[(y_i - \mathbb{E}(y_i))^2 \right]$$
 est la variance de $y_i, \quad i = 1, 2.$

Ce qui revient à considérer que la puissance estimée des sources restituées est identique pour les deux sources et unitaire.

Pour résoudre le problème (2.4), nous allons utiliser le Lagrangien associé. Il s'écrit

$$L(\mathbf{B}, \lambda_1, \lambda_2) := I(\mathbf{y}) + \lambda_1 \left(\operatorname{var}(y_1) - 1 \right)^2 + \lambda_2 \left(\operatorname{var}(y_2) - 1 \right)^2.$$
(2.5)

Notons

$$C(\mathbf{y}, \lambda_1, \lambda_2) := \lambda_1 (\operatorname{var}(y_1) - 1)^2 + \lambda_2 (\operatorname{var}(y_2) - 1)^2.$$

Selon le principe de dualité, la solution du problème avec contrainte (2.4) est donnée par celle du problème d'optimisation sans contrainte suivant

$$\max_{\lambda_1,\lambda_2} \min_{\mathbf{B}} L(\mathbf{B},\lambda_1,\lambda_2) = \max_{\lambda_1,\lambda_2} \min_{\mathbf{B}} \left\{ I(\mathbf{y}) + C(\mathbf{y},\lambda_1,\lambda_2) \right\}.$$
 (2.6)

Comme ce dernier est sans contrainte, la solution peut être calculée par utilisation de l'algorithme de descente du gradient. À l'optimum, on a

$$\frac{dL}{d\mathbf{B}} = 0, \quad \frac{\partial L}{\partial \lambda_{\mathbf{i}}} = 0, \ i = 1, 2,$$

où

$$\frac{dL}{d\mathbf{B}} = \frac{dI(\mathbf{y})}{d\mathbf{B}} + \frac{dC(\mathbf{y}, \lambda_1, \lambda_2)}{d\mathbf{B}}$$
(2.7)

 et

$$\frac{\partial L}{\partial \lambda_i} = \left(\operatorname{var}(y_i) - 1 \right)^2, \ i = 1, 2.$$
(2.8)

En utilisant la relation $\mathbf{y} = \mathbf{B}\mathbf{x}$, nous avons montré, voir (1.36), que l'entropie du vecteur \mathbf{y} peut s'écrire comme suit

$$H(\mathbf{y}) = H(\mathbf{x}) + \log |\det(\mathbf{B})|, \qquad (2.9)$$

où

$$H(\mathbf{y}) := -\mathbb{E}\left(\log f_{\mathbf{y}}(\mathbf{y})\right) \text{ et } H(\mathbf{x}) := -\mathbb{E}\left(\log f_{\mathbf{x}}(\mathbf{x})\right).$$

D'où (2.1) peut s'écrire

$$I(\mathbf{y}) = \sum_{i=1}^{p} H(y_i) - H(\mathbf{x}) - \log |\det(\mathbf{B})|.$$
(2.10)

Dans cette équation, le terme $H(\mathbf{x})$ est indépendant de la matrice **B**, d'où

$$\frac{dI(\mathbf{y})}{d\mathbf{B}} = \sum_{i=1}^{p} \frac{dH(y_i)}{d\mathbf{B}} - \frac{d}{d\mathbf{B}} \log |\det(\mathbf{B})|.$$
(2.11)

Or le terme $\sum_{i=1}^{p} \frac{dH(y_i)}{d\mathbf{B}}$ ne dépend que des densités marginales et de leurs dérivées. En effet,

$$\sum_{i=1}^{p} \frac{dH(y_i)}{d\mathbf{B}} = -\sum_{i=1}^{p} \mathbb{E}\left[\frac{d\log f_{y_i}(y_i)}{d\mathbf{B}}\right] = -\sum_{i=1}^{p} \mathbb{E}\left[\frac{f'_{y_i}(y_i)}{f_{y_i}(y_i)}\frac{dy_i}{d\mathbf{B}}\right]$$
$$= \mathbb{E}\left[\boldsymbol{\psi}(\mathbf{y})\mathbf{x}^T\right],$$
(2.12)

où $\boldsymbol{\psi}(\mathbf{y}) := (\psi_1(y_1), \psi_2(y_2))^T$ et $\psi_i(\cdot) := -\frac{f'_{y_i}(\cdot)}{f_{y_i}(\cdot)}$, la fonction score marginale de y_i , i = 1, 2.

D'autre part, on montre que (voir Annexe; A.1)

$$\frac{d}{d\mathbf{B}}\log|\det(\mathbf{B})| = \mathbf{B}^{-T}.$$
(2.13)

En regroupant les deux termes du gradient (2.12) et (2.13), l'expression du gradient de $I(\mathbf{y})$ par rapport à **B** est

$$\frac{dI(\mathbf{y})}{d\mathbf{B}} = \mathbb{E}\left[\boldsymbol{\psi}(\mathbf{y})\mathbf{x}^{T}\right] - \mathbf{B}^{-T}.$$
(2.14)

En supposant que $\mathbb{E}(y_i) = 0$, i = 1, 2, on peut montrer que l'expression du gradient du terme $C(\mathbf{y}, \lambda_1, \lambda_2)$ par rapport à la matrice **B** est donnée comme suit

$$\frac{dC(\mathbf{y},\lambda_1,\lambda_2)}{d\mathbf{B}} = 4 \operatorname{diag}(\lambda_1,\lambda_2)\mathbb{E}\left(\mathbf{w}\mathbf{x}^T\right), \qquad (2.15)$$

où $\mathbf{w} := (w_1, w_2)^T$ avec $w_i := (\mathbb{E}[y_i^2] - 1) y_i$, i = 1, 2, et $\operatorname{diag}(\lambda_1, \lambda_2)$ est la matrice diagonale de termes diagonaux λ_1 et λ_2 .

D'où la possibilité d'exprimer le gradient de (2.6) par la relation suivante

$$\frac{dL}{d\mathbf{B}} = \mathbb{E}\left[\boldsymbol{\psi}(\mathbf{y})\mathbf{x}^{T}\right] - \mathbf{B}^{-T} + 4 \operatorname{diag}(\lambda_{1}, \lambda_{2})\mathbb{E}\left(\mathbf{w}\mathbf{x}^{T}\right), \qquad (2.16)$$

 et

$$\frac{\partial L}{\partial \lambda_i} = \left(\operatorname{var}(y_i) - 1 \right)^2, \ i = 1, 2.$$
(2.17)

L'expression du gradient (2.16) du critère (2.6) nécessite une estimation des densités de probabilité marginales f_{y_i} et des fonctions scores marginales associées $\psi_i(y_i)$. Ceci est l'objet du paragraphe suivant.

2.2.4 Selection de modèles

Soit $\mathbf{y}(1) := (y_1(1), y_2(1))^T, \dots, \mathbf{y}(N) := (y_1(N), y_2(N))^T$ un échantillon du vecteur aléatoire $\mathbf{y} := (y_1, y_2)^T$. Pour estimer les densités marginales $f_{y_1}(\cdot)$ et $f_{y_2}(\cdot)$, nous allons modéliser ces densités de probabilité par des familles de densités exponentielles de dimensions (nombre de paramètres) $d_1, d_2 \in \mathbb{N}^*$. Le modèle retenu est celui qui minimise le critère d'information de Akaike (AIC) [3]. Dans ce qui suit, pour simplifier, on note $p_y(\cdot)$ la densité marginale de y_1 ou y_2 . Une densité de loi exponentielle (de dimension d) s'écrit

$$f_d(y,\theta,a,b) := \frac{\exp\left\{\theta_1 y + \dots + \theta_d y^d\right\}}{\int_a^b \exp\left\{\theta_1 y + \dots + \theta_d y^d\right\} \, dy}, \ a,b \in \mathbb{R}, \theta \in \mathbb{R}^d \ \text{et } d = 1,2,\dots$$
(2.18)

Notons AIC(d) le critère AIC associé au modèle f_d et à l'échantillon $y(1), \ldots, y(N)$, i.e.,

$$AIC(d) := \min_{\theta \in \mathbb{R}^d} \left\{ -\frac{1}{N} \sum_{n=1}^N \log f_d(y(n), \theta, a, b) + \frac{d}{N} \right\}.$$
 (2.19)

Notons que le terme d/N est dû au biais de l'estimation de l'entropie de f_y , il dépend de la dimension d du modèle; la quantité AIC(d) est donc un estimateur corrigé du biais de l'entropie de $f_y(\cdot)$ à l'intérieur du modèle f_d (c.f. e.g. [3]). Définissons

$$d^* := \arg\min_{d \in \mathbb{N}^*} AIC(d).$$
(2.20)

Ainsi f_{d^*} , la famille exponentielle de dimension d^* , est le modèle retenu. Soit $\hat{\theta}$ la solution de (2.19). Notons que $\hat{\theta}$ est l'estimateur du maximum de vraisemblance du paramètre θ pour le modèle exponentiel retenu. Lorsque le support [a, b] est inconnu, on l'estime par $[\hat{a}, \hat{b}] := [\min_{i=1,...,N} y(i), \max_{i=1,...,N} y(i)]$. Cette démarche conduit donc au choix du meilleur modèle et à une estimation optimale des paramètres de celui-ci. La densité $f_y(\cdot)$ est donc estimée par

$$\widehat{f}_y(y) = f_{d^*}(y,\widehat{\theta},\widehat{a},\widehat{b}).$$
(2.21)

Ce qui permet de déduire l'estimateur suivant de la fonction score marginale $\psi_y(\cdot)$

$$\widehat{\psi}_{y}(y(n)) = -\frac{\widehat{f}_{y}(y(n))'}{\widehat{f}_{y}(y(n))} = -\sum_{j=1}^{d^{*}} j\widehat{\theta}_{j}y(n)^{j-1}, \quad n = 1, \dots, N.$$
(2.22)

Notons que l'estimateur de la fonction score proposé ici est différent de celui proposé dans [8] par Babaie-Zadeh. En effet, l'estimateur proposé dans [8] utilise une méthode de type moindres carrés et une modélisation de la fonction score par un polynôme de degré d fixé a priori. Notre approche à l'avantage de fournir la dimension adéquate d^* du modèle et des estimateurs optimaux $\hat{\theta}$, des paramètres du modèle choisi, par maximum de vraisemblance.

Par ailleurs le choix des familles exponentielles, pour modéliser les densités marginales f_{y_i} , est également justifié par les propriétés suivantes :

- Les familles exponentielles sont des fonctions de densités qui s'adaptent pour la modélisation de la plupart des densités usuelles (uniforme, normale, exponentielle, ... etc).
- Le choix optimal f_{d*}(y) pour modéliser f_{yi} est indépendant des valeurs de B; autrement dit d* est invariant par changement d'échelle B; ainsi dans l'algorithme ci-après la sélection des modèles des densités f_{y1} et f_{y2} est réalisé une seule fois à l'étape d'initialisation.

Algorithme

Nous résumons la méthode de séparation proposée par l'algorithme suivant

Initialisation de B, λ₁, λ₂, y = Bx, choix de d*.
Itérer :

y = Bx.
Calcul de dL/dB = Ê [ψ(y)x^T] - B^{-T} + 4diag(λ₁, λ₂)Ê (wx^T).

B ← B - μdL/dB, (λ₁, λ₂)^T ← (λ₁, λ₂)^T + μ ((var(y₁) - 1)², (var(y₂) - 1)²)^T.
Répéter jusqu'à convergence.

L'estimateur $\widehat{\psi}(\cdot)$ dans l'étape (2) de l'algorithme précédent est fourni par (2.22) : $\widehat{\psi}(\mathbf{y}) = \left(\widehat{\psi}_{y_1}(y_1), \widehat{\psi}_{y_2}(y_2)\right)^T$.

2.2.5 Résultats de simulations

Pour illustrer les performances de l'algorithme proposé, nous présentons trois exemples de résultats de simulations. Dans le premier, on traite le cas de deux signaux sourcesdeux signaux mélanges. Dans les deuxième et troisième exemples, nous traitons le cas de trois signaux sources-trois signaux mélanges.

Exemple 2.1. Nous considérons deux signaux sources indépendants $s_1(n)$ et $s_2(n)$, n = 1, ..., N, *i.i.d* de loi uniforme sur [0, 1] (centrés et normalisés). Le nombre d'échantillons est fixé à N = 2000. On utilise la matrice de mélange suivante

$$A := \left(\begin{array}{cc} 1 & 0.5 \\ 0.5 & 1 \end{array} \right).$$

La qualité de séparation est évaluée par le rapport signal sur résidus défini par l'équation (1.65). La présente méthode est comparée avec la méthode proposée par Babaie-Zadeh [8] et celle proposée par Pham [153]. Cette dernière est basée sur des estimateurs des fonctions scores utilisant la méthode d'estimation non paramétrique à noyau; le noyau utilisé est approché par des splines d'ordre trois

$$K(u) = \begin{cases} 3/4 - u^2 & si & |u| \le 1/2; \\ (3/2 - |u|)^2/2 & si & 1/2 \le |u| \le 3/2; \\ 0 & sinon. \end{cases}$$
(2.23)

Les simulations montrent une bonne performance de l'algorithme proposé par rapport à ceux de Babaie-Zadeh [8] et Pham [153]. Ces résultats sont présentés sur les figures 2.1 et 2.2. Les simulations sont répétées 100 fois avec un pas fixe. Aussi, nous présentons sur la figure 2.3 les plans d'espace des signaux sources, des signaux mélanges et des sources estimées fourni par la méthode proposée. On constate que notre méthode permet d'estimer les sources inconnues avec un meilleur SNR (près de 2 dB par rapport à la méthode proposée par Pham et près de 4 dB par rapport à celle proposée par Babaie-Zadeh).



FIG. 2.1 – Rapport signal sur résidus (SNR) de notre algorithme comparé avec celui de Pham en fonction du nombre d'itérations (SNR1 : source 1; SNR2 : source 2).

Exemple 2.2. Nous considérons trois signaux sources indépendants $s_1(n)$, $s_2(n)$ et $s_3(n)$, n = 1, ..., N. Les signaux $s_1(n)$ et $s_3(n)$, n = 1, ..., N, sont des suites i.i.d de loi uniforme sur [0,1] (centrés et normalisés), et $s_2(n)$, n = 1, ..., N, est une suite i.i.d de loi gaussienne centrée réduite. Le nombre d'échantillons est fixé à



FIG. 2.2 – Rapport signal sur résidus (SNR) de notre algorithme comparé avec celui de Babaie-Zadeh en fonction du nombre d'itérations (SNR1 : source 1; SNR2 : source 2).

N = 2000. On utilise la matrice de mélange suivante

$$\mathbf{A} := \left(\begin{array}{rrrr} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{array} \right).$$

Nous présentons sur la figure 2.4 les SNRs des trois sources estimées fournis par la méthode proposée. Les simulations sont répétées 100 fois avec un pas fixe. Nous observons que la méthode donne des résultats satisfaisants (SNR environ 28 dB). Nous présentons dans la figure 2.5 les signaux sources, les observations, les sources estimées et la différence entre les sources et leurs estimées fournies par l'algorithme proposé.

Exemple 2.3. Nous considérons trois signaux sources indépendants $s_1(n)$, $s_2(n)$ et $s_3(n)$, n = 1, ..., N. Le signal s_1 est de loi exponentielle, s_2 de loi gaussienne et s_3



FIG. 2.3 – De gauche à droite, l'espace des signaux sources, signaux mélanges et sources estimées.

de loi uniforme sur [0, 1]. Le nombre d'échantillons est fixé à N = 2000. On utilise la matrice de mélange suivante

$$\mathbf{A} := \left(\begin{array}{rrrr} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{array} \right).$$

Nous présentons sur la figure 2.6 les SNRs des trois sources estimées fournis par la méthode proposée. Les simulations sont répétées 100 fois avec un pas fixe. Nous observons que la méthode donne des résultats satisfaisants (SNR d'environ 28 dB). Nous présentons dans la figure 2.7 les signaux sources, les observations, les sources estimées et la différence entre les sources et leurs estimées fournies par l'algorithme proposé.

Dans cette section, nous avons présenté une méthode de SAS, alternative à celles proposées par Babaie-Zadeh [8] et Pham [153], basée sur la minimisation de l'IM sous contraintes et l'estimation des densités marginales par critère AIC et maximum de vraisemblance dans des familles de lois exponentielles. Nous allons présenter dans la section suivante de nouveaux critères de résolution du problème de SAS basés sur



FIG. 2.4 – Les SNRs de trois sources données par l'algorithme de l'IM sous contraintes.

la minimisation des divergences entre densités de probabilité et qui généralisent le critère de l'information mutuelle.

2.3 Méthodes basées sur la minimisation des α divergences

Nous venons de voir que la minimisation de l'IM sous contrainte améliorait l'estimation des sources grâce notamment à l'emploi de fonctions scores estimées via la sélection du meilleur modèle de densités par utilisation du critère AIC. La méthode de SAS que nous allons développer dans cette section est basée sur la minimisation des divergences entre densités de probabilité. Nous allons considérer le cas des mélanges linéaires instantanés. La méthode que nous proposons généralise celle utilisant le critère de l'information mutuelle. Nous montrons que la méthode utilisant



FIG. 2.5 – De haut en bas, les signaux sources, les mélanges, les sources estimées et l'écart entre les sources et leurs estimées.

la divergence particulière de Hellinger possède de bonnes propriétés en terme d'efficacité-robustesse. Dès lors, il semble opportun de rappeler brièvement, la notion de divergences entre densités de probabilité et leurs propriétés en donnant des motivations à l'utilisation de celles-ci en séparation aveugle de sources.

2.3.1 Divergences entre lois de probabilité

Les divergences entre lois de probabilité ont été introduites par Csiszár [55]. Soit φ une fonction convexe quelconque définie sur $[0, +\infty]$ à valeurs dans $[0, +\infty]$ telle que $\varphi(1) = 0$. Pour toutes lois de probabilité Q et P définies sur \mathbb{R}^p telles que Q est absolument continue par rapport à P, la I_{φ} -divergence entre Q et P est définie par

$$I_{\varphi}(Q,P) := \int_{\mathbb{R}^p} \varphi\left(\frac{dQ(x)}{dP(x)}\right) \ dP(x).$$
(2.24)



FIG. 2.6 – Les SNRs de trois sources données par l'algorithme de l'IM sous contraintes.

Si Q n'est pas absolument continue par rapport à P, on pose $I_{\varphi}(Q, P) = +\infty$. Le rapport $\frac{dQ(x)}{dP(x)}$ est la dérivée de Radon-Nikodym de Q par rapport à P. Notons que si Q et P admettent respectivement des densités $q(\cdot)$ et $p(\cdot)$ sur \mathbb{R}^p , alors $\frac{dQ(x)}{dP(x)} = \frac{q(x)}{p(x)}$ et on a

$$I_{\varphi}(Q,P) := I_{\varphi}(q,p) = \int_{\mathbb{R}^p} \varphi\left(\frac{q(x)}{p(x)}\right) p(x) \, dx.$$
(2.25)

Pour toute loi de probabilité P, l'application $Q \mapsto I_{\varphi}(Q, P)$ est convexe et est positive. Si Q = P, alors $I_{\varphi}(Q, P) = 0$. De plus, si la fonction $x \mapsto \varphi(x)$ est strictement convexe sur un voisinage de x = 1, on a la propriété fondamentale suivante

$$I_{\varphi}(Q, P) = 0$$
 si et seulement si $Q = P.$ (2.26)

Toutes ces propriétés sont présentées et démontrées dans [55] et [127]. La propriété (2.26) est fondamentale pour résoudre le problème de la SAS que nous allons considérer. Les I_{φ} -divergences sont invariantes par changement d'échelle. Cepen-



FIG. 2.7 – De haut en bas, les signaux sources, les mélanges, les sources estimées et l'écart entre les sources et leurs estimées.

dant, elles ne sont en général pas symétriques; $I_{\varphi}(Q, P)$ et $I_{\varphi}(P, Q)$ diffèrent. Elles coïncident si la fonction convexe φ vérifie $\varphi(x) - x\varphi(\frac{1}{x}) = c(x-1)$ où c est une constante (voir Liese et Vajda [127] Théorème 1.13).

Largement utilisée en théorie de l'information, la divergence de Kullback-Leibler, notée KL-divergence, est associée à la fonction convexe réelle $\varphi(x) = x \log x - x + 1$; elle est définie donc par

$$KL(Q,P) := \int_{\mathbb{R}^p} \log\left(\frac{dQ(x)}{dP(x)}\right) \ dQ(x).$$
(2.27)

La divergence de Kullback-Leibler modifiée, notée KL_m -divergence, est associée à la fonction convexe $\varphi(x) = -\log x + x - 1$, i.e.,

$$KL_m(Q,P) := \int_{\mathbb{R}^p} -\log\left(\frac{dQ(x)}{dP(x)}\right) \ dP(x).$$
(2.28)

D'autres divergences, largement utilisées en statistique sont les divergences de χ^2 et

 $\chi^2\text{-modifiée}~(\chi^2_m)$

$$\chi^{2}(Q,P) := \frac{1}{2} \int_{\mathbb{R}^{p}} \left(\frac{dQ(x)}{dP(x)} - 1 \right)^{2} dP(x)$$
(2.29)

 et

$$\chi_m^2(Q, P) := \frac{1}{2} \int_{\mathbb{R}^p} \frac{\left(\frac{dQ(x)}{dP(x)} - 1\right)^2}{\left(\frac{dQ(x)}{dP(x)}\right)} \, dP(x) \tag{2.30}$$

qui sont associées aux fonctions convexes $\varphi(x) = \frac{1}{2}(x-1)^2$ et $\varphi(x) = \frac{1}{2}(x-1)^2 / x$, respectivement. La distance de Hellinger (*H*) est également une divergence; elle est associée à la fonction convexe $\varphi(x) = 2(\sqrt{x}-1)^2$,

$$H(Q,P) := \int_{\mathbb{R}^p} 2\left(\sqrt{\frac{dQ(x)}{dP(x)}} - 1\right)^2 dP(x).$$
(2.31)

Tous les exemples précédents des divergences font partie de la classe des "divergences de puissance" introduite par [54] qui est définie par la classe des fonctions convexes

$$x \in \mathbb{R}^*_+ \mapsto \varphi_{\alpha}(x) := \begin{cases} \frac{x^{\alpha} - \alpha x + \alpha - 1}{\alpha(\alpha - 1)} & \text{si } \alpha \neq 0, \alpha \neq 1; \\ -\log x + x - 1 & \text{si } \alpha = 0; \\ x \log x - x + 1 & \text{si } \alpha = 1. \end{cases}$$
(2.32)

Le tableau (2.1) présente, suivant le choix de la fonction convexe φ_{α} , la divergence associée. Notons que les divergences présentées ci-dessus ont été utilisées en esti-

La fonction φ_{α}	φ_1	$arphi_0$	φ_2	φ_{-1}	$\varphi_{\frac{1}{2}}$
La divergence associée	KL	KL_m	χ^2	χ^2_m	H

TAB. 2.1 – Exemples de fonctions convexes et leurs divergences associées.

mation statistique par [119], [33], [120] et [34]. Il a été montré que leur utilisation généralise la méthode classique du maximum de vraisemblance, et conduit à des estimateurs ayant des propriétés similaires, voire meilleures dans certains cas; par

exemple l'estimateur utilisant la divergence de Hellinger améliore celui du maximum de vraisemblance en terme d'efficacité-robustesse pour des données bruitées. Sur la figure 2.8, nous présentons les fonctions convexes φ_{α} des divergences KL, KL_m et H.



FIG. 2.8 – Les fonctions convexes φ_{α} des divergences KL, KL_m et H.

Après avoir présenté les divergences entre lois de probabilité, nous allons voir comment exploiter ces notions pour construire des critères aptes à résoudre le problème de la SAS et ce via deux approches :

- Approche 1 (à deux étapes) : la séparation est composée de deux étapes. La première étape est celle du blanchiment des observations \mathbf{x} . Elle consiste à décorréler les observations. La deuxième étape consiste à estimer une matrice unitaire de rotation $\mathbf{U}(\theta)$ qui rend les signaux blanchis \mathbf{z} statistiquement indépendants. Cette dernière consiste à déterminer le paramètre optimal " θ " qui minimise un critère et permet d'obtenir une matrice unitaire unique de rotation.

- Approche 2 (à une seule étape) : la séparation est réalisée par la minimisation d'un critère par rapport aux coefficients de la matrice séparante **B**.

2.3.2 Approche 1 : méthode "décorrélation-rotation"

Description et principe de la méthode

Sans perte de généralité, nous considérons le cas de deux sources-deux observations (p = 2). Nous considérons d'abord l'étape de blanchiment des mélanges **x**. Cette dernière consiste à décorréler les signaux observés pour obtenir les signaux blanchis que l'on note **z**. Ensuite, on applique l'étape de rotation des signaux blanchis afin de restituer les sources inconnues via la minimisation d'un critère basé sur le choix d'une divergence. Pour ce faire, nous appliquons aux signaux blanchis **z** une matrice de rotation **U** vérifiant det(**U**) = 1. Cette matrice est paramétrée par une rotation de Givens et s'écrit comme suit

$$\mathbf{U}(\theta) := \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix},$$

où θ est l'angle de rotation des signaux blanchis \mathbf{z} tel que $\theta \in [-\pi, \pi]$. D'où la possibilité d'écrire les égalités suivantes

$$y_1(n) = \cos(\theta)z_1(n) + \sin(\theta)z_2(n),$$

$$y_2(n) = -\sin(\theta)z_1(n) + \cos(\theta)z_2(n).$$

Il s'agit de trouver l'angle $\hat{\theta}$ optimal tel que les sources estimées y_1 et y_2 soient le plus indépendantes possible en utilisant cette fois les α -divergences.

Soit $\mathbf{y} := (y_1, y_2)^T$ un vecteur aléatoire, et notons $f_{\mathbf{y}}$ la densité jointe du vecteur \mathbf{y} et f_{y_i} la densité marginale de y_i , la i^{eme} composante de \mathbf{y} . La $I_{\varphi_{\alpha}}$ -divergence entre
le produit $f_{y_1}f_{y_2}$ des densités marginales et la densité jointe $f_{\mathbf{y}}$ s'écrit (voir 2.24)

$$I_{\varphi_{\alpha}}(f_{y_{1}}f_{y_{2}},f_{\mathbf{y}}) := \int_{\mathbb{R}^{2}} \varphi_{\alpha} \left(\frac{f_{y_{1}}(t_{1})f_{y_{2}}(t_{2})}{f_{\mathbf{y}}(t_{1},t_{2})} \right) f_{\mathbf{y}}(t_{1},t_{2}) dt_{1} dt_{2}$$
$$= \mathbb{E} \left[\varphi_{\alpha} \left(\frac{f_{y_{1}}(y_{1})f_{y_{2}}(y_{2})}{f_{\mathbf{y}}(y_{1},y_{2})} \right) \right].$$
(2.33)

D'après la propriété (2.26), la quantité (2.33) est positive et s'annule si et seulement si $f_{y_1}f_{y_2} = f_y$, i.e., ssi les composantes y_1, y_2 sont mutuellement indépendantes. Cette propriété peut nous servir de critère. Celui-ci peut être estimé par

$$\widehat{I}_{\varphi_{\alpha}}(\mathbf{y}) := \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha} \left(\frac{\widehat{f}_{y_1}(y_1(n)) \widehat{f}_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n))} \right),$$
(2.34)

où $\hat{f}_{y_1}, \hat{f}_{y_2}$ et \hat{f}_y sont respectivement les estimateurs à noyau des densités f_{y_1}, f_{y_2} et f_y . Dans ce cas, pour séparer les sources nous avons besoin de rendre indépendants les signaux y_1 et y_2 et la séparation sera obtenue par la minimisation du critère (2.34) par rapport à l'angle θ

$$\min_{\theta} \widehat{I}_{\varphi_{\alpha}}(\mathbf{y}) = \min_{\theta} \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha} \left(\frac{\widehat{f}_{y_1}(y_1(n)) \widehat{f}_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n))} \right).$$
(2.35)

Pour calculer ce minimum nous allons utiliser la méthode de descente du gradient. Nous avons besoin de calculer le gradient du critère (2.34). Nous donnons dans le paragraphe suivant, la forme explicite du gradient.

Calcul du gradient $\frac{d\widehat{I}_{\varphi_{\alpha}}(\mathbf{y})}{d\theta}$

On a

$$\frac{dI_{\varphi_{\alpha}}(\mathbf{y})}{d\theta} = \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha}' \left(\frac{\widehat{f}_{y_1}(y_1(n))\widehat{f}_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n))} \right) \frac{d\left(\frac{f_{y_1}(y_1(n))f_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n))}\right)}{d\theta}.$$
 (2.36)

Pour calculer cette quantité $\frac{dI_{\varphi_{\alpha}}(\mathbf{y})}{d\theta}$, il suffit donc de calculer les deux termes suivants

$$\varphi'_{\alpha}\left(\frac{\widehat{f}_{y_1}(y_1(n))\widehat{f}_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n),y_2(n))}\right) \quad \text{et} \quad \frac{d\left(\frac{f_{y_1}(y_1(n))f_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n),y_2(n))}\right)}{d\theta}.$$

Cependant, pour faire ce calcul, nous avons besoin au préalable de calculer les estimateurs \hat{f}_{y_1} , \hat{f}_{y_2} et \hat{f}_{y} . Ceci est l'objet du paragraphe suivant.

Estimation de la densité par la méthode à noyau

En statistique, l'estimation par noyau (ou encore méthode de Parzen-Rosenblatt) est une méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. Le principe de cette méthode est le suivant.

Définition 2.6. Soit (y(n), n = 1, ..., N) une suite de variables aléatoires à valeurs dans \mathbb{R} indépendantes et identiquement distribuées (i.i.d). Soit f_y la densité commune aux variables (y(n), n = 1, ..., N). L'estimateur à noyau, noté \hat{f}_y , de la densité f_y est défini comme suit

$$\widehat{f}_y(y) := \frac{1}{Nh} \sum_{n=1}^N K\left(\frac{y - y(n)}{h}\right),$$

où N est la taille du signal y(n), n = 1, ..., N, la fonction $K(\cdot)$ est appelée "noyau" de l'estimateur et h est la largeur de la fenêtre du noyau ou le paramètre de lissage.

Pour que $\hat{f}_y(y)$ soit une densité de probabilité, le noyau doit satisfaire les deux conditions suivantes :

- 1. $\forall y, \quad K(y) \ge 0.$
- 2. $\int_{-\infty}^{+\infty} K(y) \, dy = 1.$

Les cas les plus usuels sont la densité gaussienne, celle uniforme sur [-1, 1] ou triangulaire. La forme du noyau n'est pas très déterminante pour la qualité de l'estimation contrairement à la valeur de h. Bien souvent, la fonction K est choisie comme étant la densité gaussienne standard (espérance nulle et variance unitaire). Notons que le noyau gaussien s'exprime de la façon suivante

$$K: y \in \mathbb{R} \mapsto K(y) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right).$$
(2.37)

C'est ce noyau que nous allons utiliser pour l'ensemble de nos simulations. Le paramètre de lissage h détermine la régularité et la précision de l'estimation de la

densité (voir la figure 2.9). Théoriquement [97], l'estimateur à noyau est une estimation biaisée de la densité. Le biais de l'estimation tend vers zéro quand $h \rightarrow 0$, et la variance de l'estimation tend vers zéro quand $Nh \rightarrow +\infty$. Ainsi, dans la pratique, quand un nombre limité d'observations est disponible, le choix de h est important. La figure 2.9 montre l'estimation de la densité d'une réalisation normale centrée unitaire de 4000 échantillons avec une valeur "optimale" de $h \approx 1.06N^{-\frac{1}{5}} = 0.2$, puis h = 2 et enfin h = 0.02. La loi théorique est représentée par des courbes discontinues et les densités estimées sont représentées en continu. Si h est trop faible on a des irrégularités dans l'estimation de la densité ce qui pose des problèmes lors du calcul des fonctions scores par dérivation des densités estimées. Mais un choix de h trop grand conduit à une mauvaise estimation de la densité. Une règle de base pour le choix de h, lorsque le noyau utilisé est gaussien, est [97]

$$h = \widehat{\sigma}_y \left(\frac{4}{3N}\right)^{\frac{1}{5}} \approx 1.06 \widehat{\sigma}_y N^{-\frac{1}{5}},$$

où $\hat{\sigma}_y$ désigne l'écart-type (cette formule est optimale lorsque les densités à estimer sont gaussiennes).

Les estimateurs respectifs \hat{f}_{y_1} , \hat{f}_{y_2} et $\hat{f}_{\mathbf{y}}$ de f_{y_1} , f_{y_2} et $f_{\mathbf{y}}$ s'écrivent

$$\widehat{f}_{y_1}(y_1) = \frac{1}{Nh} \sum_{n=1}^N K\left(\frac{y_1 - y_1(n)}{h}\right), \ \widehat{f}_{y_2}(y_2) = \frac{1}{Nh} \sum_{n=1}^N K\left(\frac{y_2 - y_2(n)}{h}\right)$$

 et

$$\widehat{f}_{\mathbf{y}}(y_1, y_2) = \frac{1}{Nh^2} \sum_{n=1}^{N} K\left(\frac{y_1 - y_1(n)}{h}\right) K\left(\frac{y_2 - y_2(n)}{h}\right),$$

où K est le noyau (2.37).

Une fois les fonctions de densité estimées, nous pouvons facilement calculer l'estimateur du gradient $\frac{dI_{\varphi_{\alpha}}(\mathbf{y})}{d\theta}$ donné par l'équation (2.36).

Pour donner l'expression du premier terme $\varphi'_{\alpha}\left(\frac{\widehat{f}_{y_1}(y_1(n))\widehat{f}_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n),y_2(n))}\right)$, nous avons besoin de revenir à la définition des fonctions convexes évoquées précédemment (2.32). Ainsi



FIG. 2.9 – Influence de la largeur du noyau gaussien pour une distribution gaussienne.

il convient d'étudier chaque fonction suivant le choix des valeurs de α . Dans ce cas, on a (voir Annexe; **A.3**)

$$\varphi_{\alpha}'\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{y}(y_{1}(n),y_{2}(n))}\right) = \begin{cases} \frac{\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{y}(y_{1}(n),y_{2}(n))}\right)^{\alpha-1} & \text{si } \alpha \neq 0, \alpha \neq 1; \\ -\frac{1}{\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{y}(y_{1}(n),y_{2}(n))}\right)} + 1 & \text{si } \alpha = 0; \quad (2.38) \\ \log\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{y}(y_{1}(n),y_{2}(n))}\right) & \text{si } \alpha = 1. \end{cases}$$

Le second terme de la relation (2.36) s'écrit

$$\frac{d\left(\frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))}{\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))}\right)}{d\theta} = \left[\frac{\left[\frac{d\hat{f}_{y_1}(y_1(n))}{d\theta}\hat{f}_{y_2}(y_2(n)) + \frac{d\hat{f}_{y_2}(y_2(n))}{d\theta}\hat{f}_{y_1}(y_1(n))\right]\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))}{\left(\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))\right)^2}\right] - \left[\frac{\left[\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))\right]\frac{d\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))}{d\theta}}{d\theta}}{\left(\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))\right)^2}\right], \quad (2.39)$$

où

$$\frac{d\widehat{f}_{y_1}(y_1)}{d\theta} = \frac{1}{Nh} \sum_{n=1}^N K'\left(\frac{y_1 - y_1(n)}{h}\right) \frac{d}{d\theta} \left(\frac{y_1 - y_1(n)}{h}\right),$$
$$\frac{d\widehat{f}_{y_2}(y_2)}{d\theta} = \frac{1}{Nh} \sum_{n=1}^N K'\left(\frac{y_2 - y_2(n)}{h}\right) \frac{d}{d\theta} \left(\frac{y_2 - y_2(n)}{h}\right)$$

 et

$$\begin{aligned} \frac{d\widehat{f}_{\mathbf{y}}(y_1, y_2)}{d\theta} &= \frac{1}{Nh^2} \sum_{n=1}^N K'\left(\frac{y_1 - y_1(n)}{h}\right) K\left(\frac{y_2 - y_2(n)}{h}\right) \frac{d}{d\theta} \left(\frac{y_1 - y_1(n)}{h}\right) \\ &+ K'\left(\frac{y_2 - y_2(n)}{h}\right) K\left(\frac{y_1 - y_1(n)}{h}\right) \frac{d}{d\theta} \left(\frac{y_2 - y_2(n)}{h}\right), \end{aligned}$$

avec

$$\frac{d}{d\theta} \left(\frac{y_1 - y_1(n)}{h} \right) = \frac{-\sin(\theta)(z_1 - z_1(n)) + \cos(\theta)(z_2 - z_2(n))}{h}, \\ \frac{d}{d\theta} \left(\frac{y_2 - y_2(n)}{h} \right) = \frac{-\cos(\theta)(z_1 - z_1(n)) - \sin(\theta)(z_2 - z_2(n))}{h}.$$

Algorithme

Nous résumons la méthode de séparation proposée par l'algorithme suivant

- Blanchiment des mélanges $\mathbf{z} = \mathbf{W}\mathbf{x}$.
- Centrage et normalisation de **z**.
- Initialisation de $\theta,\ \mu.$
- Itérer :

 y = U(θ)z.

 2. Calculer dÎ_{φα}(y)/dθ.
 3. θ ← θ − μ dÎ_{φα}(y)/dθ.
 Répéter jusqu'à convergence.

Résultats de simulations

Nous présentons dans ce paragraphe un exemple de simulations pour illustrer la performance de ces méthodes. Dans cet exemple, on traite le cas de deux sources et deux mélanges.

Exemple 2.4. (le cas non bruité)

Nous considérons deux signaux indépendants $s_1(n)$ et $s_2(n)$, n = 1, ..., N, i.i.d de loi uniforme sur [0, 1] (centrés et normalisés). Le nombre d'échantillons est fixé à N = 1000. La matrice de mélange est

$$\mathbf{A} := \left(\begin{array}{cc} 1 & 0.85 \\ 0.65 & 1 \end{array} \right)$$

L'angle optimal obtenu pour les critères KL_m , H et KL est respectivement $\hat{\theta}_{KL_m} = -0.0480 \ rd$, $\hat{\theta}_H = -0.0482 \ rd$ et $\hat{\theta}_{KL} = -0.0484 \ rd$. Toutes ces valeurs sont calculées par utilisation de l'algorithme de descente du gradient avec un pas fixe ($\mu = 0.02$), pour 60 itérations, en prenant comme point initial $\theta = 0$. Pour mesurer la qualité de séparation, nous utilisons le critère de performance (1.65) défini dans le chapitre 1. Nous illustrons sur la figure 2.10 les courbes de la convergence de l'angle θ vers ses valeurs optimales en utilisant les divergences KL_m ($\alpha = 0$), H ($\alpha = 0.5$) et KL($\alpha = 1$). Nous présentons dans le tableau (2.2), en dB, le rapport signal sur résidus (SNR) pour les critères KL_m ($\alpha = 0$), H ($\alpha = 0.5$) et KL($\alpha = 1$), que l'on compare avec les résultats de séparation obtenus en utilisant les algorithmes de Jade et Sobi; pour des signaux de taille N = 1000. Notons que (\mathbf{s}, \mathbf{x}), dans le tableau (2.2), est défini par la relation suivante

$$SNR_i := 10 \log_{10} \frac{\widehat{\mathbb{E}}\{s_i^2\}}{\widehat{\mathbb{E}}\{(x_i - s_i)^2\}}, \ i = 1, \dots, p.$$
(2.40)

Les simulations sont répétées 50 fois. Nous constatons que les méthodes KL_m , Het KL donnent de bien meilleurs résultats que les algorithmes de Jade et Sobi. Par contre, nous ne constatons aucune différence significative entre les méthodes utilisant les divergences. Nous présentons également dans la figure 2.11 les plans d'espace des signaux sources, observations, signaux blanchis et sources estimées par la méthode basée sur la divergence de Hellinger (H).



FIG. 2.10 – Courbes de l'angle θ suivant le nombre d'iterations pour les critères KL_m , H et KL.

	(\mathbf{s}, \mathbf{x})	$Sobi(\mathbf{s}, \mathbf{y})$	$Jade(\mathbf{s}, \mathbf{y})$	$I_0(\mathbf{s},\mathbf{y})$	$I_{0.5}(\mathbf{s},\mathbf{y})$	$I_1(\mathbf{s},\mathbf{y})$
source s_1	3.23	16.04	35.48	37.95	37.60	37.25
source s_2	4.92	17.25	37.75	39.06	39.03	39.05

TAB. 2.2 – SNR en dB pour Sobi, Jade, KL_m , H et KL.

Exemple 2.5. (le cas bruité)

Dans cet exemple, on contamine les signaux de mélange $\mathbf{As}(n)$ de l'exemple (2.4) avec un bruit gaussien $\mathbf{b}(n) := (b_1(n), b_2(n))^T$. Sur la figure 2.12, nous présentons les SNRs moyens (des deux sources) pour les trois critères, KL_m , H et KL en fonction du rapport signal sur bruit (RSB). Rappelons que le RSB du signal observé $x_i(n), n = 1, ..., N$, est défini par la relation suivante

$$RSB_i := -10\log_{10}\frac{P_{x_i}}{P_{b_i}}, \ i = 1, 2,$$
(2.41)

où $P_{x_i} := \mathbb{E}(x_i(n)^2)$ est la puissance du signal mélange observé x_i et $P_{b_i} := \mathbb{E}(b_i(n)^2)$ est la puissance du bruit b_i .

Le RSB varie entre $-60 \, dB \, et -20 \, dB$. Nous constatons que KL_m est plus robuste pour certaines valeurs du RSB entre $-50 \, et -27 \, dB$, par contre entre $-60 \, et -50 \, dB$, et entre $-27 \, et -20 \, dB$, KL_m est moins robuste. On observe le phénomène contraire pour KL. Le critère basé sur Hellinger présente un bon compromis entre les critères KL et KL_m et ce quelque soit le degré de contamination des données ; voir figure 2.12. Ceci peut s'expliquer de la façon suivante. Lorsque les données sont bruitées, deux cas peuvent se produire (voir critère 2.34).

$$\begin{array}{l} cas \ 1 : \ \frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))}{\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))} <<1, \ on \ l'appelle \ "outlier".\\ cas \ 2 : \ \frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))}{\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))} >>1, \ on \ l'appelle \ "inlier". \end{array}$$

D'après la figure 2.8, il est clair que KL est la plus robuste par rapport aux outliers, et la moins robuste par rapport aux inliers. A contrario, KL_m est la moins robuste par rapport aux outliers et la plus robuste par rapports aux inliers. Comme la fonction convexe associée à Hellinger est comprise entre celles associées à KL et KL_m , le critère basé sur la divergence de Hellinger présente un bon compromis en terme de robustesse dans le cas des outliers et inliers. Comme dans la pratique, nous ne connaissons pas si le bruit conduit à un outlier ou à un inlier, il convient donc d'utiliser le critère basé sur la divergence de Hellinger.

2.3.3 Approche 2 : méthode "directe"

À la différence de la méthode précédente, celle-ci procède en une seule étape.

Description du critère

Nous présentons la méthode pour un mélange de deux sources-deux observations (p = 2):

$$y_1(n) = b_{11}x_1(n) + b_{12}x_2(n),$$

 $y_2(n) = b_{21}x_1(n) + b_{22}x_2(n).$

Dans cette approche, la séparation sera obtenue par la minimisation du critère (2.34) mais cette fois-ci par rapport aux coefficients b_{pq} de la matrice séparante **B**, i.e.,

$$\min_{\mathbf{B}} \widehat{I}_{\varphi_{\alpha}}(\mathbf{y}) = \min_{\mathbf{B}} \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha} \left(\frac{\widehat{f}_{y_1}(y_1(n)) \widehat{f}_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n))} \right).$$
(2.42)

Nous allons utiliser la méthode de descente du gradient pour calculer ce minimum. Ceci nécessite le calcul du gradient du critère (2.42). Nous allons donner la forme explicite du gradient dans le paragraphe suivant. Nous utilisons la méthode à noyau pour estimer les densités f_{y_1} , f_{y_2} et f_y .

Calcul du gradient

On a

$$\frac{\partial \widehat{I}_{\varphi_{\alpha}}(\mathbf{y})}{\partial b_{pq}} = \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha}^{'} \left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{\mathbf{y}}(y_{1}(n), y_{2}(n))} \right) \frac{\partial \left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{\mathbf{y}}(y_{1}(n), y_{2}(n))} \right)}{\partial b_{pq}}.$$
 (2.43)

Nous constatons donc que pour calculer le gradient (2.43), il suffit de calculer les deux termes suivants

$$\varphi_{\alpha}^{'}\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{\mathbf{y}}(y_{1}(n),y_{2}(n))}\right) \quad \text{et} \quad \frac{\partial\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{\mathbf{y}}(y_{1}(n),y_{2}(n))}\right)}{\partial b_{pq}}.$$

Concernant le premier terme $\varphi'_{\alpha}\left(\frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))}{\hat{f}_{(\mathbf{y}y_1(n),y_2(n))}}\right)$, le calcul a été déjà fait au paragraphe (2.3.2). Le deuxième terme est donné comme suit

$$\frac{\partial \left(\frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))}{\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))}\right)}{\partial b_{pq}} = \left[\frac{\left[\frac{\partial \hat{f}_{y_1}(y_1(n))}{\partial b_{pq}}\hat{f}_{y_2}(y_2(n)) + \frac{\partial \hat{f}_{y_2}(y_2(n))}{\partial b_{pq}}\hat{f}_{y_1}(y_1(n))\right]\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))}{\left(\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))\right)^2}\right] - \left[\frac{\left[\frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))}{\hat{f}_{y_2}(y_2(n))}\right]\frac{\partial \hat{f}_{\mathbf{y}}(y_1(n),y_2(n))}{\partial b_{pq}}}{\left(\hat{f}_{\mathbf{y}}(y_1(n),y_2(n))\right)^2}\right], \quad (2.44)$$

où

$$\frac{\partial \widehat{f}_{y_1}(y_1)}{\partial b_{pq}} = \frac{1}{Nh} \sum_{n=1}^{N} K'\left(\frac{y_1 - y_1(n)}{h}\right) \frac{\partial}{\partial b_{pq}}\left(\frac{y_1 - y_1(n)}{h}\right),$$
$$\frac{\partial \widehat{f}_{y_2}(y_2)}{\partial b_{pq}} = \frac{1}{Nh} \sum_{n=1}^{N} K'\left(\frac{y_2 - y_2(n)}{h}\right) \frac{\partial}{\partial b_{pq}}\left(\frac{y_2 - y_2(n)}{h}\right)$$

 et

$$\frac{\partial \widehat{f}_{\mathbf{y}}(y_1, y_2)}{\partial b_{pq}} = \frac{1}{Nh^2} \sum_{n=1}^N K'\left(\frac{y_1 - y_1(n)}{h}\right) K\left(\frac{y_2 - y_2(n)}{h}\right) \frac{\partial}{\partial b_{pq}}\left(\frac{y_1 - y_1(n)}{h}\right) + K'\left(\frac{y_2 - y_2(n)}{h}\right) K\left(\frac{y_1 - y_1(n)}{h}\right) \frac{\partial}{\partial b_{pq}}\left(\frac{y_2 - y_2(n)}{h}\right).$$

avec

$$\frac{\partial}{\partial b_{pq}} \left(\frac{y_1 - y_1(n)}{h}\right)_{p=1,q=1,2} = \frac{x_q - x_q(n)}{h}, \quad \frac{\partial}{\partial b_{pq}} \left(\frac{y_1 - y_1(n)}{h}\right)_{p=2,q=1,2} = 0,$$

$$\frac{\partial}{\partial b_{pq}} \left(\frac{y_2 - y_2(n)}{h}\right)_{p=2,q=1,2} = \frac{x_q - x_q(n)}{h} \quad \text{et} \quad \frac{\partial}{\partial b_{pq}} \left(\frac{y_2 - y_2(n)}{h}\right)_{p=1,q=1,2} = 0,$$

K est le noyau gaussien standard (2.37) et K' sa dérivée.

Algorithme

La méthode proposée est schématisée par l'algorithme suivant

Initialisation de B = I, μ.
Itérer :

y = Bx, centrage et normalisation de y.
Calculer dÎ_{φα}(y)/dB.
B ← B − μ dÎ_{φα}(y)/dB.

Répéter jusqu'à convergence.

Résultats de simulations

Exemple 2.6. (le cas non bruité)

Nous présentons dans ce paragraphe un exemple de simulations pour illustrer les performances des méthodes proposées. Nous utilisons les mêmes données que celles traitées dans l'exemple (2.4). La détermination de la matrice séparante **B** a été calculée par utilisation de l'algorithme de descente du gradient par rapport aux coefficients de la matrice séparante **B**. Pour mesurer la qualité de séparation, nous utilisons le critère de performance (1.65). Sur la figure 2.13, nous présentons les SNRs moyens en fonction du nombre d'itérations pour les trois critères IM, H et KL. Les simulations sont répétées 50 fois pour des données non-bruitées. Nous ne constatons aucune différence sensible entre les trois méthodes utilisant les divergences. Nous présentons dans la figure 2.14 le plan d'espace des signaux sources, observations et sources estimées par la méthode basée sur la divergence de Hellinger. Nous constatons que cette méthode permet de restituer les sources originales avec un SNR satisfaisant (proche de 40 dB).

Exemple 2.7. (le cas bruité)

Dans cette exemple, nous traitons les mêmes données que celles traitées dans l'exemple (2.5). Nous présentons sur la figure 2.15 les SNRs moyens en fonction du RSB pour les trois critères, KL_m , H et KL. Le RSB varie entre -60 dB et -20 dB. Les simulations sont répétées 50 fois. Nous constatons que KL_m est plus robuste pour l'intervalle entre -50 et -27 dB, cependant entre -60 et -50 dB, et entre -27 et -20 dB, KL_m est moins robuste. On observe le phénomène contraire pour la divergence KL. Le critère basé sur la divergence de Hellinger présente encore un bon compromis entre les critères KL et KL_m , indépendamment du degré de contamination des données. Il convient donc d'utiliser le critère basé sur la divergence de Hellinger ($\alpha = 0.5$) en l'absence d'information sur le degré de contamination des données.

Dans le chapitre suivant, nous allons étendre cette dernière méthode au cas de mélanges convolutifs.

2.4 Conclusion

Dans ce chapitre nous avons développé des nouvelles méthodes de la SAS stationnaires pour des mélanges linéaires instantanés. Nous avons tout d'abord proposé une méthode basée sur l'optimisation de l'information mutuelle sous contraintes par un passage au problème dual et estimation du gradient stochastique par maximum de vraisemblance dans des familles de lois exponentielles. L'algorithme proposé s'applique pour une large classe de lois de probabilité pouvant être modélisées par une famille de lois exponentielles choisies selon le critère AIC. De plus, la pénalisation permet de s'affranchir du problème de l'indétermination du facteur d'échelle et stabilise l'algorithme. Nous avons illustré les performances de cet algorithme, en terme d'efficacité, avec des signaux simulés en comparaison avec deux autres méthodes présentées par Babaie-Zadeh [8] et Pham [153]. Le critère s'applique quelque soit le nombre de sources; par exemple le cas de trois sources a été illustré. Nous avons présenté ensuite des algorithmes de séparation aveugle de signaux utilisant les α divergences. Celles-ci généralisent l'approche de l'information mutuelle. Nous avons montré que la méthode basée sur la distance de Hellinger présente de bonnes propriétés de robustesse en présence du bruit en comparaison avec les autres méthodes.



FIG. 2.11 – Approche 2. De gauche à droite et de haut en bas, l'espace des signaux sources, signaux mélanges, signaux blanchis et sources estimées avec le critère basé sur la divergence de Hellinger.



FIG. 2.12 – Approche 1. En haut : SNR des trois critères : $IM \ (\alpha = 0), H \ (\alpha = 0.5)$ et $KL \ (\alpha = 1)$. En fonction du RSB - en bas (de gauche à droite) : zoom sur SNR (autour de -55 dB, -39 dB et -20 dB).



FIG. 2.13 – SNR en fonction du nombre d'itérations pour KL_m , H et KL (cas non bruité).



FIG. 2.14 – Approche 2. De gauche à droite, l'espace des signaux sources, signaux mélanges et sources estimées par le critère utilisant la divergence de Hellinger.



FIG. 2.15 – Approche 2. En haut : SNR des trois critères : $IM \ (\alpha = 0), H \ (\alpha = 0.5)$ et $KL \ (\alpha = 1)$, en fonction du RSB. En bas (de gauche à droite) : zoom sur SNR (autour de -52 dB, -39 dB et -20 dB).

Chapitre 3

Mélanges linéaires convolutifs

3.1 Introduction

La méthode de la SAS que nous proposons dans ce chapitre, pour traiter le cas des mélanges convolutifs, est également basée sur la minimisation des α -divergences entre densités de probabilité. Notons que d'autres classes de divergences ont été utilisées dans la littérature. Dans [68], les auteurs présentent des algorithmes de séparation de sources basé sur la minimisation des divergences de Renyi; voir aussi [182]. L'utilisation des α -divergences que nous proposons est justifiée par les raisons suivantes. Elles généralisent l'approche de l'IM. Elles s'appliquent dans les cas de mélanges instantanés ou convolutifs. La classe des α -divergences contient la divergence de Hellinger qui a des bonnes propriétés d'efficacité-robustesse en estimation statistique, comme il a été démontré dans [19], [128] et [110]. Par ailleurs, l'estimation de ces α -divergences est faite par injection des estimateurs à noyau des densités de probabilité dans l'expression de leurs définitions. Notons que dans ce chapitre, nous allons utiliser les propriétés des divergences présentées au chapitre précédant dans le paragraphe 2.3.1.

3.2 Modèle convolutif

Le mélange, dans ce cas, s'écrit sous la forme

$$\mathbf{x}(t) = \mathbf{A} * \mathbf{s}(t) + \mathbf{b}(\mathbf{t}), \tag{3.1}$$

où * est le produit de convolution, **s** est le vecteur des signaux sources et **A** est la matrice des filtres de mélange. **b** est le vecteur des signaux bruits, dont les éléments sont supposés centrés, mutuellement indépendants, et indépendants des signaux sources. La présence du bruit additif dans le modèle de mélange complique le problème de la séparation de sources. Pour simplifier le problème, le bruit est supposé négligeable (après prétraitement). Le but est d'estimer les sources **s** à partir des observations **x**. L'estimateur, dans ce cas, est de la forme

$$\mathbf{y}(t) = \mathbf{B} * \mathbf{x}(t), \tag{3.2}$$

où **B** est la matrice des filtres séparante. Le problème consiste donc à chercher un estimateur $\hat{\mathbf{B}}$ de **B**, à partir des observations **x**, qui conduise à une estimation

$$\widehat{\mathbf{y}}(t) = \widehat{\mathbf{B}} * \mathbf{x}(t),$$

des sources s. La transposition dans le domaine discret du modèle (3.2) conduit à

$$\mathbf{y}(n) = [\mathbf{B}(z)]\mathbf{x}(n) = \sum_{k} \mathbf{B}_{k}\mathbf{x}(n-k) = \sum_{k} \mathbf{B}_{k}\mathbf{A}_{k}\mathbf{s}(n-k), \quad (3.3)$$

où \mathbf{A}_k et \mathbf{B}_k correspondent respectivement à la transformée en z des matrices des filtres \mathbf{A} et \mathbf{B} . Si \mathbf{A} est une matrice inversible à gauche et si les sources sont statistiquement indépendantes, la solution du problème de séparation est possible à une permutation et à un filtrage près

$$\mathbf{y}(n) = [\widehat{\mathbf{B}}(z)]\mathbf{x}(n) = [\widehat{\mathbf{B}}(z)\mathbf{A}(z)]\mathbf{s}(n), \qquad (3.4)$$

où la matrice des filtres $\widehat{\mathbf{B}}(z)$ vérifie $[\widehat{\mathbf{B}}(z)\mathbf{A}(z)] = [\mathcal{PH}(z)]$, avec \mathcal{P} une permutation et \mathcal{H} un opérateur de filtrage.

3.3 Nouveaux critères de séparation via les α divergences

Le critère (2.33) généralisé à p sources, peut être estimé par

$$\widehat{I}_{\varphi_{\alpha}}(\mathbf{y}) := \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha} \left(\frac{\widehat{f}_{y_1}(y_1(n)) \cdots \widehat{f}_{y_p}(y_p(n))}{\widehat{f}_{\mathbf{y}}(y_1(n), \dots, y_p(n))} \right),$$
(3.5)

où $\widehat{f}_{y_1}, \ldots, \widehat{f}_{y_p}$, et $\widehat{f}_{\mathbf{y}}$ sont respectivement les estimateurs à noyau des densités f_{y_1} , \ldots, f_{y_p} , et $f_{\mathbf{y}}$. Cependant, dans le cas de mélanges convolutifs, l'indépendance des signaux $y_1(n), y_2(n), \ldots, y_p(n)$ (pour tout n) ne suffit pas pour séparer les sources. En d'autres termes, dans le cas de deux sources-deux observations, par exemple, pour séparer les sources s_1 et s_2 nous avons besoin de rendre indépendants les signaux $y_1(n)$ et $y_2(n-m)$ pour tout n et tout décalage m; voir [10, 66]. Par conséquent, nous remplaçons le critère (3.5) par

$$\widehat{J}_{\alpha} := \sum_{\mathbf{m}} \widehat{I}_{\varphi_{\alpha}}(\mathbf{y}^{\mathbf{m}}) = \sum_{\mathbf{m}} \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha} \left(\frac{\widehat{f}_{y_1}(y_1(n)) \widehat{f}_{y_2}(y_2(n-m_2)) \cdots \widehat{f}_{y_p}(y_p(n-m_p))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n-m_2) \dots, y_p(n-m_p))} \right),$$
(3.6)

où

$$\mathbf{y}^{\mathbf{m}}(n) := (y_1(n-m_1), y_2(n-m_2), \dots, y_p(n-m_p))^T,$$

 $m_1 = 0 \text{ et } \mathbf{m} := (m_1, m_2, \dots, m_p)^T.$

La séparation sera obtenue par minimisation de celui-ci sur la matrice des filtres ${f B}$

$$\min_{\mathbf{B}} \widehat{J}_{\alpha} = \min_{\mathbf{B}} \sum_{\mathbf{m}} \widehat{I}_{\varphi_{\alpha}}(\mathbf{y}^{\mathbf{m}}).$$
(3.7)

Nous allons utiliser la méthode de descente du gradient pour calculer ce minimum. Nous avons donc besoin de calculer le gradient du critère \hat{J}_{α} par rapport aux coefficients de la matrice des filtres **B**. Sans perte de généralité, nous considérons le cas de deux sources (p = 2). Nous donnons dans le paragraphe suivant, la forme explicite du gradient. Le cas de plusieurs sources se traite de façon analogue.

3.4 Calcul du gradient

Notons $\mathbf{B}_k(p,q)$, p,q = 1, 2, les éléments de la matrice \mathbf{B}_k pour tout retard k. Un calcul direct donne

$$\frac{\partial \widehat{J}_{\alpha}}{\partial \mathbf{B}_{k}(p,q)} = \sum_{m_{2}} \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha}^{\prime} \left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n-m_{2}))}{\widehat{f}_{\mathbf{y}}(y_{1}(n), y_{2}(n-m_{2}))} \right) \frac{\partial \left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n-m_{2}))}{\widehat{f}_{\mathbf{y}}(y_{1}(n), y_{2}(n-m_{2}))} \right)}{\partial \mathbf{B}_{k}(p,q)},$$
(3.8)

où $\varphi'_{\alpha}(\cdot)$ est la dérivée de la fonction convexe $\varphi_{\alpha}(\cdot)$ introduite dans (2.32). Ainsi, pour $\alpha = 0$, i.e., dans le cas de l'information mutuelle, nous obtenons

$$\frac{\partial \widehat{J}_0}{\partial \mathbf{B}_k(p,q)} = \sum_{m_2} \frac{1}{N} \sum_{n=1}^N \left(1 - \frac{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n-m_2))}{\widehat{f}_{y_1}(y_1(n))\widehat{f}_{y_2}(y_2(n-m_2))} \right) \frac{\partial \frac{f_{y_1}(y_1(n))f_{y_2}(y_2(n-m_2))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n-m_2))}}{\partial \mathbf{B}_k(p,q)}.$$
(3.9)

Pour $\alpha = 1$, i.e., le cas de la divergence de Kullback-Leibler, on a

$$\frac{\partial \widehat{J}_1}{\partial \mathbf{B}_k(p,q)} = \sum_{m_2} \frac{1}{N} \sum_{n=1}^N \log \left(\frac{\widehat{f}_{y_1}(y_1(n)) \widehat{f}_{y_2}(y_2(n-m_2))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n-m_2))} \right) \frac{\partial \frac{\widehat{f}_{y_1}(y_1(n)) \widehat{f}_{y_2}(y_2(n-m_2))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n-m_2))}}{\partial \mathbf{B}_k(p,q)}.$$
(3.10)

Enfin, pour $\alpha \neq 0, 1$, nous obtenons

$$\frac{\partial \widehat{J}_{\alpha}}{\partial \mathbf{B}_{k}(p,q)} = \sum_{m_{2}} \frac{1}{N(\alpha-1)} \sum_{n=1}^{N} \left(\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n-m_{2}))}{\widehat{f}_{\mathbf{y}}(y_{1}(n), y_{2}(n-m_{2}))} \right)^{\alpha-1} - 1 \right) \times \frac{\partial \frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n-m_{2}))}{\widehat{f}_{\mathbf{y}}(y_{1}(n), y_{2}(n-m_{2}))}}{\partial \mathbf{B}_{k}(p,q)}.$$
(3.11)

Notons que

$$\frac{\partial \frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n-m))}{\hat{f}_{\mathbf{y}}(y_1(n),y_2(n-m))}}{\partial \mathbf{B}_k(p,q)} = \left[\frac{\left(\frac{\partial \hat{f}_{y_1}(y_1(n))}{\partial \mathbf{B}_k(p,q)}\hat{f}_{y_2}(y_2(n-m)) + \frac{\partial \hat{f}_{y_2}(y_2(n-m))}{\partial \mathbf{B}_k(p,q)}\hat{f}_{y_1}(y_1(n))\right)\hat{f}_{\mathbf{y}}(y_1(n),y_2(n-m))}{(\hat{f}_{\mathbf{y}}(y_1(n),y_2(n-m)))^2}\right]$$

$$-\left[\frac{\left(\widehat{f}_{y_1}(y_1(n))\widehat{f}_{y_2}(y_2(n-m))\right)\frac{\partial\widehat{f}_{\mathbf{y}}(y_1(n),y_2(n_m))}{\partial \mathbf{B}_k(p,q)}}{(\widehat{f}_{\mathbf{y}}(y_1(n),y_2(n-m)))^2}\right],$$

où $\widehat{f}_{y_1}(y_1(n))$, $\widehat{f}_{y_2}(y_2(n-m))$ et $\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n-m))$ sont les estimateurs respectivement de $f_{y_1}(y_1(n))$, $f_{y_2}(y_2(n-m))$ et $f_{\mathbf{y}}(y_1(n), y_2(n-m))$, définis par

$$\widehat{f}_{y_1}(y_1(n)) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{y_1(n) - y_1(i)}{h}\right),$$

$$\widehat{f}_{y_2}(y_2(n-m)) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{y_2(n-m) - y_2(i)}{h}\right)$$

 et

$$\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n-m)) = \frac{1}{Nh^2} \sum_{i=1}^N K\left(\frac{y_1(n) - y_1(i)}{h}\right) K\left(\frac{y_2(n-m) - y_2(i)}{h}\right).$$

Notons qu'on a également

$$\frac{\partial \widehat{f}_{y_1}(y_1(n))}{\partial \mathbf{B}_k(p,q)} = \frac{1}{Nh} \sum_{i=1}^N K' \left(\frac{y_1(n) - y_1(i)}{h} \right) \frac{\partial}{\partial \mathbf{B}_k(p,q)} \left(\frac{y_1(n) - y_1(i)}{h} \right),$$
$$\frac{\partial \widehat{f}_{y_2}(y_2(n-m))}{\partial \mathbf{B}_k(p,q)} = \frac{1}{Nh} \sum_{i=1}^N K' \left(\frac{y_2(n-m) - y_2(i)}{h} \right) \frac{\partial}{\partial \mathbf{B}_k(p,q)} \left(\frac{y_2(n-m) - y_2(i)}{h} \right)$$

 et

$$\begin{aligned} \frac{\partial \widehat{f}_{\mathbf{y}}(y_1(n), y_2(n-m))}{\partial \mathbf{B}_k(p, q)} &= \frac{1}{Nh^2} \sum_{i=1}^N K' \left(\frac{y_1(n) - y_1(i)}{h} \right) K \left(\frac{y_2(n-m) - y_2(i)}{h} \right) \\ &\times \frac{\partial}{\partial \mathbf{B}_k(p, q)} \left(\frac{y_1(n) - y_1(i)}{h} \right) \\ &+ K' \left(\frac{y_2(n-m) - y_2(i)}{h} \right) K \left(\frac{y_1(n) - y_1(i)}{h} \right) \\ &\times \frac{\partial}{\partial \mathbf{B}_k(p, q)} \left(\frac{y_2(n-m) - y_2(i)}{h} \right), \end{aligned}$$

où K est le noyau gaussien standard (2.37) et K' sa dérivée.

Algorithme

Nous résumons la méthode de séparation proposée par l'algorithme suivant

Initialisation de B_k = I, ∀k et μ.
Itérer :

y = B * x. Centrage et normalisation de y.
Calculer dÎ_{φα}(y)/dB.
B ← B − μ dÎ_{φα}(y)/dB.

Répéter jusqu'à convergence.

3.5 Résultats de simulations

Exemple 3.1. (le cas non bruité)

Dans cet exemple, les deux signaux sources s_1 et s_2 sont des variables aléatoires indépendantes et de même loi uniforme (sur [0,1]) centrées et réduites. La matrice des filtres de mélange considérée est

$$\mathbf{A}(\mathbf{z}) := \begin{bmatrix} 1 + 0.2z^{-1} + 0.1z^{-2} & 0.5 + 0.3z^{-1} + 0.1z^{-2} \\ 0.5 + 0.3z^{-1} + 0.1z^{-2} & 1 + 0.2z^{-1} + 0.1z^{-2} \end{bmatrix}.$$

La matrice séparante des filtres **B** a été calculée par utilisation des algorithmes de descente du gradient par rapport aux coefficients de celle-ci. Pour mesurer la qualité de séparation, nous utilisons le SNR comme critère de performance (1.66) défini au chapitre 1. Sur la figure 3.1, nous présentons les SNRs moyens (des deux sources) en fonction du nombre d'itérations pour les trois critères KL_m ($\alpha = 0$), H ($\alpha = 0.5$) et KL ($\alpha = 1$). Les simulations sont répétées 100 fois pour des données non-bruitées. Nous ne constatons aucune différence sensible entre les trois méthodes utilisant les divergences. Cependant, la méthode de l'IM reste légèrement supérieure par rapport aux autres méthodes.



FIG. 3.1 – Qualité de la séparation des trois critères : KL_m ($\alpha = 0$), Hellinger ($\alpha = 0.5$) et KL ($\alpha = 1$), en fonction du nombre d'itérations (cas convolutif).

Exemple 3.2. (le cas bruité)

Dans cet exemple, les deux signaux des mélanges x_1 et x_2 sont bruités. Nous comparons la qualité de séparation des critères (3.7), pour $\alpha = 0$, $\alpha = 0.5$ et $\alpha = 1$, en fonction du rapport signal sur bruit (RSB). La somme dans (3.7) est calculée pour

$$m_2 = -M, -M+1, \ldots, M-1, M;$$

avec M = 2p où p est le degré du filtre (p=2). Nous utilisons le SNR comme critère de performance (1.66) défini au chapitre 1. Pour estimer les densités et leurs dérivées, nous avons utilisé un noyau gaussien standard. Sur la figure 3.2, nous présentons les SNR moyens (des deux sources) en fonction du RSB pour les trois critères, KL_m $(\alpha = 0), KL (\alpha = 1)$ et Hellinger $(\alpha = 0.5)$. Le RSB varie entre -20 dB et 0 dB. La taille des signaux est N = 500, le nombre d'itérations est fixé à 150 et les simulations sont répétées 100 fois. Les intervalles de confiance (au niveau 95%) des différentes valeurs des SNRs sont représentés par des segments verticaux. Nous constatons que KL_m est plus robuste pour certaines valeurs du RSB (entre -20 et -12 dB, et entre -8 et 0 dB), par contre entre -12 et -8 dB, KL_m est moins robuste. On observe la propriété contraire pour KL. Le critère basé sur Hellinger présente un bon compromis entre les critères KL et KL_m et ce quelque soit le degré de contamination des données; voir figure 3.2. Ceci peut s'expliquer dans le cas général par la forme des fonctions convexes φ_{α} de la façon suivante, c.f. la figure 2.8. En effet, lorsque les données sont bruitées, deux cas peuvent se produire (voir critère 3.5).

$$\cos 1 : \frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))}{\hat{f}_{y}(y_1(n),y_2(n))} << 1, \text{ on l'appelle "outlier".} \\ \cos 2 : \frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n))}{\hat{f}_{y}(y_1(n),y_2(n))} >> 1, \text{ on l'appelle "inlier".}$$

D'après la figure 2.8, il est clair que KL est plus robuste par rapport aux outliers, et moins robuste par rapport aux inliers. A contrario, KL_m est moins robuste par rapport aux outliers et plus robuste par rapport aux inliers. Comme la fonction convexe associée à Hellinger (figure 2.8) est comprise entre celles associées à KL et KL_m , le critère basé sur Hellinger présente donc le meilleur compromis en termes de robustesse dans le cas des outliers et inliers. Comme dans la pratique, nous ne connaissons pas si le bruit conduit à un outlier ou à un inlier, il convient donc d'utiliser le critère basé sur Hellinger.

3.6 Etudes expérimentales

Exemple 3.3. Dans cet exemple nous exploitons le critère de Hellinger pour séparer des signaux réels. Nous considérons les signaux vibratoires issus du banc de mesure présenté dans la figure 3.3. Il est constitué de deux moteurs à courant continu (1.4 et 1.1 kW) qui fonctionnent avec des vitesses de rotation différentes. Ces deux moteurs sont fixés sur la même structure métallique. Un accéléromètre a été fixé sur chaque moteur afin d'enregistrer les vibrations. Le but de l'expérience est de diagnostiquer l'état de fonctionnement de chacun des deux moteurs lors qu'ils fonctionnent



FIG. 3.2 – Qualité de la séparation des trois critères : KL_m ($\alpha = 0$), Hellinger ($\alpha = 0.5$) et KL ($\alpha = 1$), en fonction du RSB (cas convolutif).

simultanément. Naturellement les signaux enregistrés (Densité Spectrale de Puissance, figure 3.5) sur chaque moteur sont affectés par les vibrations issues de l'autre. La SAS permet donc de trouver les signaux propres à chaque moteur (figure 3.6). Pour pouvoir mesurer la qualité de séparation du critère choisi (Hellinger), nous allons comparer les résultats avec les signaux références (signaux sources), i.e., ceux enregistrés sur chaque moteur lorsqu'ils fonctionnent séparément (figure 3.4). La fréquence de rotation du moteur 1 (1.1 kW) est de 48.5 Hz, et celle du moteur 2 (1.4 kW) est de 31.5 Hz. L'expérience a révélé la présence de défauts sur les quatre roulements à billes équipant le banc de mesure (figure 3.3). Les roulements à billes 1B, 2B et 2C ont les mêmes caractéristiques (2207 KTV C3) distinctes de celles du roulement à billes 2A (6203 RS C3). Les 4 défauts sont décrits par les 4 fréquences 134 Hz (défaut sur la bague externe du roulement 2C), 179 Hz (défaut sur la bague externe du roulement 2A), 207 Hz (défaut sur la bague externe du roulement 1B) et 210 Hz (défaut sur la bague interne du roulement 2B). Le milieu de propagation est modélisé par un filtre d'ordre 50 (temps de réponse de la structure à un choc impulsionnel). La longueur des échantillons est de 4096 points. La figure (3.6) montre clairement que les fréquences de rotation fondamentales et leurs harmoniques sont bien restituées sur chaque moteur. Après séparation, le défaut de roulement à billes du moteur 1, indiqué par la raie à 207 Hz est bien attribué au moteur 1 alors que les 3 autres fréquences 139 Hz, 179 Hz et 210 Hz sont attribuées aux signaux liés à la rotation du moteur 2. Nous pouvons aussi remarquer que toutes les raies sont correctement affectées (absence de permutations de raie pour des canaux de même fréquence).



FIG. 3.3 – Le banc de mesure.

3.7 Conclusion

Nous avons présenté dans ce chapitre de nouveaux algorithmes de la SAS basés sur la minimisation des α -divergences. Les méthodes proposées généralisent l'approche



FIG. 3.4 – DSP des signaux sources pris séparément.

de l'information mutuelle dans le cas des mélanges linéaires convolutifs. Nous avons montré par simulation que l'algorithme utilisant la divergence de Hellinger a de meilleures propriétés de robustesse en présence du bruit en comparaison avec l'information mutuelle et la divergence de Kullback-Leibler. Nous avons illustré les performances de cet algorithme, en terme d'efficacité et robustesse, sur des signaux simulés et sur des signaux réels issus de machines tournantes.



FIG. 3.5 – DSP des signaux mélanges.



FIG. 3.6 – DSP des signaux estimés.

Deuxième partie

SAS cyclostationnaires

Chapitre 4

Mélanges linéaires instantanés de signaux cyclostationnaires

4.1 Introduction

La plupart des techniques de séparation aveugle de sources visent à séparer des sources statistiquement indépendantes supposées stationnaires. Toutefois, dans de nombreux contextes applicatifs, tels que les radiocommunications, les sources rencontrées ont rarement les propriétés requises précédemment. En effet, certaines de ces sources, comme les sources à modulation analogique, sont généralement non stationnaires. En outre, parmi ces dernières, certaines sources sont cyclostationnaires comme les signaux à modulation numérique. La séparation aveugle de sources cyclostationnaires a fait l'objet de nombreux travaux. [7, 107, 70, 28] et [23] considèrent ce problème et proposent des méthodes s'appliquant dans des situations spécifiques particulières. Ainsi, Bouguerriou *et al.* [23] proposent une méthode permettant d'extraire un signal d'intérêt cyclostationnaire de fréquence cyclique connue en exploitant l'autocorrélation cyclique dans le cadre d'un mélange instantané. Les signaux cyclostationnaires sont d'un intérêt pratique dans les domaines des radiocommunications, du radar et aussi du diagnostic des machines tournantes qui nous interesse plus particulièrement. Dans un premier temps, nous allons donner la définition et les propriétés de la cyclostationnarité. Ensuite, nous étendons les travaux proposés dans [23], en proposant une autre méthode de séparation aveugle de sources cyclostationnaires à l'ordre 2. Cette méthode permet d'extraire plusieurs signaux sources cyclostationnaires de fréquences cycliques connues et différentes deux-à-deux. Enfin, nous proposons une méthode de la SAS cyclostationnaires basée sur la minimisation des α -divergences et pénalisation.

4.2 Cyclostationnarité : définitions et propriétés

Les signaux stationnaires sont caractérisés par des propriétés statistiques (moyenne, variance, ... etc) invariantes dans le temps. Cette définition est fondamentale dans la classification des signaux. C'est pourquoi tout signal dont les propriétés statistiques varient dans le temps est considéré dés lors comme non-stationnaire. Cette classe de signaux implique un certain nombre de difficultés liées aux méthodes d'analyse à définir. La classe des signaux non-stationnaires comprend une grande variété de signaux dont les signaux non-stationnaires périodiques appelés signaux cyclostationnaires qui constituent le centre d'intérêt de ce chapitre. Les signaux cyclostationnaires sont ceux dont les statistiques varient de façon périodique dans le temps. Leur intérêt est lié au fait qu'ils sont omniprésents dans plusieurs domaines.

4.2.1 Cyclostationnarité et applications

Bennett en 1958 [18] a observé la présence de signaux cyclostationnaires dans la mise en oeuvre des algorithmes de synchronisation de systèmes de communications. Depuis, plusieurs chercheurs ont introduit des concepts clés pour la représentation de processus cyclostationnaires. Dans [90], Gudzenko a présenté une étude sur l'es-

timation spectrale de processus cyclostationnaire. Un peu plus tard, dans [87, 88] Gladyshev a travaillé sur l'analyse spectrale et a introduit le concept de la quasipériodicité de la corrélation du processus. En 1963, Nedoma [142] a présenté la cycloérgodicité pour un processus cyclostationnaire avec une période unique et plus tard en 1983 Boyles et Gardner [31] ont généralisé ce principe aux signaux cyclostationnaires à plusieurs périodes. Après la première utilisation de Bennett de la cyclostationnarité dans le domaine des communications, Franks [72] et Gardner [78] ont développé une étude détaillée pour la cyclostationnarité dans le domaine des communications. En 1975 Gardner et Franks [81] ont étudié les avantages de l'exploitation des processus cyclostationnaires en particulier dans le contexte de filtrage optimal. En 1988, Gardner [79] présente sa théorie par une approche non-probabiliste appelée approche Fraction-Of-Time (FOT). En parallele Dandawate et Giannakis [57, 56] et Giannakis [86] ont présenté une alternative à l'approche de Gardner par une approche probabiliste. En 1992 Spooner [170] a considéré la théorie d'ordre supérieur de signaux cyclostationnaires. Izzo et Napoli [105, 102, 103, 101, 104] et Izzo et al. [106] ont également contribué à la recherche fondamentale dans le domaine du traitement de signaux cyclostationnaires. Ces recherches théoriques sur la question de la cyclostationnarité ont abouti à diverses applications dont la médecine [71], la climatologie [98], l'hydraulogie [117], l'océanologie [64], le domaine des communications [36], [79] et [86] et la mécanique qui connaît un regain d'intérêt chez les chercheurs [132], [37], [163], [190], [6] et [161] dans une perspective de diagnostic.

L'historique de la cyclostationnarité nous permet de mieux cerner l'évolution et le contexte dans lequel cette discipline a vu le jour. Mais pour en saisir d'avantage le sens, il est évidemment nécessaire d'en préciser les propriétés mêmes.

4.2.2 Caractéristiques cyclostationnaires

S'agissant d'une classe caractérisée par sa non stationnarité, la cyclostationnarité a par nature des propriétés statistiques variables dans le temps avec une spécificité essentielle : le caractère cyclique de ses statistiques variant dans le temps [79, 80, 82]. Par ailleurs, lorsque les propriétés statistiques au premier et au second ordre d'un signal sont périodiques, alors celui-ci est dit cyclostationnaire au sens large. Autrement dit, il y a périodicité de la moyenne et de la fonction d'autocorrélation. Caractériser la cyclostationnarité au sens large revient à se limiter au premier et au second ordre.

Définition 4.1. (Cyclostationnarité à l'ordre 1)

Soit t un instant fixe. Le signal aléatoire x(t) est dit cyclostationnaire à l'ordre 1 si sa moyenne varie périodiquement dans le temps. Ainsi nous avons

$$\mathbb{E}(x(t)) = \mathbb{E}(x(t+kT_0)), \ \forall k \in \mathbb{Z},$$
(4.1)

où T_0 définit la période cyclique fondamentale à l'ordre 1 et $\mathbb{E}(.)$ représente l'espérance statistique du processus x(t).

Définition 4.2. (Cyclostationnarité à l'ordre deux)

Soit t un instant fixe. Le signal aléatoire x(t) est dit cyclostationnaire à l'ordre deux, avec des fréquences cycliques ($k\alpha_0 := \frac{k}{T_0}; k \in \mathbb{Z}$), si sa fonction d'autocorrélation vérifie

$$R_x(t,\tau) := \mathbb{E}\left(x(t)x(t+\tau)\right) = \mathbb{E}\left(x(t+kT_0)x(t+kT_0+\tau)\right) =: R_x(t+kT_0,\tau), \ \forall k \in \mathbb{Z}, \forall \tau, t \in \mathbb{Z}, t \in \mathbb{Z}, \forall \tau, t \in \mathbb{Z}, t \in \mathbb{Z},$$

i.e., si la fonction

$$t \mapsto R_x(t,\tau)$$

est périodique de période T_0 , pour tout décalage τ .

Définition 4.3. (Cycloérgodicité à l'ordre deux)

Soit x(t) un signal aléatoire cyclostationnaire à l'ordre deux de fréquences cycliques $(k\alpha_0 := \frac{k}{T_0}, \ k \in \mathbb{Z})$. Le signal x(t) est cycloérgodique à l'ordre deux si la moyenne temporelle (en t) de

$$x(t)x(t+\tau)e^{-i2\pi\frac{k}{T_0}t}, \ k\in\mathbb{Z}, \ \forall \tau$$

est un nombre certain; i.e., si

$$\lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) x(t+\tau) e^{-i2\pi \frac{k}{T_0} t} dt$$

est un nombre certain $\forall k \in \mathbb{Z}, \ \forall \tau$.

Il a été démontré dans [80, 81], que la fonction d'autocorrélation de tout signal aléatoire cyclostationnaire (de fréquences cycliques $k\alpha_0, k \in \mathbb{Z}$) et cycloérgodique à l'ordre deux, peut se décomposer comme suit

$$R_x(t,\tau) := \mathbb{E}(x(t)x(t+\tau)) = \sum_{k=-\infty}^{+\infty} R_x^{k\alpha_0}(\tau)e^{i2\pi\frac{k}{T_0}t},$$
(4.3)

avec

$$R_x^{k\alpha_0}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) x(t+\tau) e^{-i2\pi \frac{k}{T_0} t} dt, \ k \in \mathbb{Z}.$$
 (4.4)

4.3 Méthodes utilisant les statistiques cycliques d'ordre 2

4.3.1 Introduction

Le mélange est de la forme

$$\mathbf{x} = \mathbf{As},\tag{4.5}$$

où $\mathbf{A} \in \mathbb{R}^{p \times p}$ est la matrice de mélange, de dimensions $p \times p$ et $\mathbf{s} := (s_1, \dots, s_p)^T$ représente les signaux sources, supposés indépendants et cyclostationnaires à l'ordre
2, de fréquences cycliques différentes deux-à-deux. Le but est donc d'estimer les sources **s** à partir des observations $\mathbf{x} := (x_1, \ldots, x_p)^T$. L'estimateur s'écrit donc

$$\mathbf{y} = \mathbf{B}\mathbf{x},\tag{4.6}$$

où $\mathbf{B} \in \mathbb{R}^{p \times p}$ est la matrice séparante. Le problème consiste donc, à partir des observations \mathbf{x} , à déterminer cette matrice \mathbf{B} , qui conduit à la meilleure estimation possible des sources \mathbf{s} , en exploitant la cyclostationnarité des signaux sources. La méthode que nous proposons se décompose en deux étapes :

La première étape consiste à blanchir les observations \mathbf{x} . Soit $\mathbf{W} \in \mathbb{R}^{p \times p}$ la matrice de blanchiment des signaux observés \mathbf{x} . Notons les signaux blanchis

$$\mathbf{z} := \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s},\tag{4.7}$$

avec $\mathbf{V} := \mathbf{W}\mathbf{A}$ qui est une matrice unitaire; c'est-à-dire $\mathbf{V}\mathbf{V}^{\mathbf{T}} = I$ et det $(\mathbf{V}) = 1$. Pour retrouver les signaux sources \mathbf{s} , il faut estimer la matrice inverse \mathbf{U} de \mathbf{V} , qui sera par conséquent une matrice unitaire. La deuxième étape consiste donc à faire une rotation des signaux blanchis \mathbf{z} afin de rendre ses composantes indépendantes. Soit \mathbf{U} une matrice $p \times p$ telle que $\mathbf{U}\mathbf{U}^{\mathbf{T}} = I_p$ et det $(\mathbf{U}) = 1$. Cette matrice est paramétrée par les rotations de Givens; elle s'écrit

$$\mathbf{U}(\theta) := \prod_{1 \le i < k \le p} G(i, k, \theta_m), \tag{4.8}$$

où

$$G(i,k,\theta_m)_{j,l} = \begin{cases} \cos(\theta_m) & \text{si} \qquad j=i, l=i \text{ ou } j=k, l=k;\\ \sin(\theta_m) & \text{si} \qquad j=i, l=k;\\ -\sin(\theta_m) & \text{si} \qquad j=k, l=i;\\ 1 & \text{si} \qquad j=l;\\ 0 & \text{sinon}, \end{cases}$$
(4.9)

pour tous $1 \le j, l \le p$, et $\theta_m \in]-\pi/2, \pi/2[$, $m = 1, \ldots, p(p-1)/2$, sont les composantes de θ , c'est-à-dire les angles de rotation du plan des observations blanchies **z**. Le nouveau processus devient

$$\mathbf{y} = \mathbf{U}(\theta)\mathbf{z} = \mathbf{U}(\theta)\mathbf{W}\mathbf{x} = \mathbf{B}\mathbf{x}.$$
(4.10)

La matrice de séparation \mathbf{B} s'écrit donc comme le produit des deux matrices \mathbf{U} et \mathbf{W}

$$\mathbf{B} = \mathbf{U}(\theta)\mathbf{W}$$

Dans la suite, nous présentons un critère de séparation permettant un choix optimal de θ , basé sur l'utilisation de la corrélation cyclique des sources. Nous montrons que ce critère conduit à la parfaite séparation et nous prouvons l'existence et l'unicité de la solution.

4.3.2 Description du critère de séparation de sources cyclostationnaires

Par souci de clarté, et sans perte de généralités, nous présentons la méthode pour un mélange de deux sources-deux observations (p = 2). L'extension au cas de plusieurs sources ($p \ge 3$) est présentée au paragraphe (4.3.3) par l'équation (4.16). Notons α_1 et α_2 respectivement les fréquences cycliques fondamentales de s_1 et s_2 . Nous supposons que s_1 et s_2 sont mutuellement indépendantes, $\alpha_1 \neq \alpha_2$ et s_1 et s_2 sont cyclostationnaires à l'ordre 2. De l'équation (4.10), il vient

$$y_{1}(n) = \cos(\theta)z_{1}(n) + \sin(\theta)z_{2}(n), y_{2}(n) = -\sin(\theta)z_{1}(n) + \cos(\theta)z_{2}(n).$$
(4.11)

Comme y_1 est un mélange de s_1 et s_2 qui sont de fréquences cycliques différentes α_1 et α_2 respectivement, il est naturel, pour que y_1 soit le plus proche possible de s_1 , de maximiser les coefficients $|R_{y_1}^{m\alpha_1}(0)|$ et minimiser $|R_{y_1}^{m\alpha_2}(0)|$, i.e., maximiser en θ , l'écart

$$\Gamma_1(\theta) = \sum_{m=1}^{L_1} \left(|R_{y_1}^{m\alpha_1}(0)|^2 - |R_{y_1}^{m\alpha_2}(0)|^2 \right), \qquad (4.12)$$

où

 $R_{y_1}^{m\alpha_1}(0)$ est la fonction d'autocorrélation cyclique du signal y_1 aux fréquences cycliques $m\alpha_1$ du signal s_1 pour un retard $\tau = 0$.

 $R_{y_1}^{m\alpha_2}(0)$ est la fonction d'autocorrélation cyclique du signal y_1 aux fréquences cycliques $m\alpha_2$ du signal s_2 pour un retard $\tau = 0$.

De même, comme y_2 est un mélange de s_2 et s_1 qui sont de fréquences cycliques différentes α_2 et α_1 respectivement, il est naturel, pour que y_2 soit le plus proche possible de s_2 , de maximiser les coefficients $|R_{y_2}^{m\alpha_2}(0)|$ et minimiser $|R_{y_2}^{m\alpha_1}(0)|$, i.e., maximiser en θ , l'écart

$$\Gamma_2(\theta) = \sum_{m=1}^{L_2} \left(|R_{y_2}^{m\alpha_2}(0)|^2 - |R_{y_2}^{m\alpha_1}(0)|^2 \right).$$
(4.13)

Enfin, pour restituer s_1 et s_2 à partir de y_1 et y_2 , en tenant compte de (4.12) et (4.13), on propose de maximiser en θ le contraste suivant

$$\Gamma(\theta) = \Gamma_1(\theta) + \Gamma_2(\theta) = \sum_{m=1}^{L_1} \left(|R_{y_1}^{m\alpha_1}(0)|^2 - |R_{y_1}^{m\alpha_2}(0)|^2 \right) + \sum_{m=1}^{L_2} \left(|R_{y_2}^{m\alpha_2}(0)|^2 - |R_{y_2}^{m\alpha_1}(0)|^2 \right).$$
(4.14)

Notons que d'après (4.11), il vient

$$\begin{aligned} R_{y_1}^{m\alpha_1}(0) &= \cos(\theta)^2 R_{z_1}^{m\alpha_1}(0) + \sin(\theta)^2 R_{z_2}^{m\alpha_1}(0) + 2\cos(\theta)\sin(\theta) R_{z_{12}}^{m\alpha_1}(0), \\ R_{y_1}^{m\alpha_2}(0) &= \cos(\theta)^2 R_{z_1}^{m\alpha_2}(0) + \sin(\theta)^2 R_{z_2}^{m\alpha_2}(0) + 2\cos(\theta)\sin(\theta) R_{z_{12}}^{m\alpha_2}(0), \\ R_{y_2}^{m\alpha_2}(0) &= \sin(\theta)^2 R_{z_1}^{m\alpha_2}(0) + \cos(\theta)^2 R_{z_2}^{m\alpha_2}(0) - 2\sin(\theta)\cos(\theta) R_{z_{12}}^{m\alpha_2}(0), \\ R_{y_2}^{m\alpha_1}(0) &= \sin(\theta)^2 R_{z_1}^{m\alpha_1}(0) + \cos(\theta)^2 R_{z_2}^{m\alpha_1}(0) - 2\sin(\theta)\cos(\theta) R_{z_{12}}^{m\alpha_1}(0). \end{aligned}$$

Ces formules sont utilisées pour calculer la fonction coût (4.14) pour tout θ . Les paramètres L_1 et L_2 sont les nombres de coefficients considérés dans la décomposition en série de Fourier de l'autocorrélation et de l'intercorrélation des signaux blanchis z_1 et z_2 . Le choix de ces paramètres se fait indépendamment de θ à l'aide d'un seuillage des coefficients des décompositions en série de Fourier de l'autocorrélation et de l'intercorrélation des signaux blanchis z_1 et z_2 .

Remarque 4.1. La méthode proposée reste valable dans le cas où une seule fréquence cyclique est connue (par exemple α_1) et si les autres sources sont stationnaires (par exemple s_2). Dans ce cas, nous prenons naturellement $\alpha_2 = 0$, et puisque L_1 et L_2 sont des constantes (ne dépendent pas de θ), le critère Γ peut s'écrire

$$\Gamma(\theta) = \sum_{m=1}^{L_1} \left(|R_{y_1}^{m\alpha_1}(0)|^2 - |R_{y_1}^0(0)|^2 \right) + \sum_{m=1}^{L_2} \left(|R_{y_2}^0(0)|^2 - |R_{y_2}^{m\alpha_1}(0)|^2 \right).$$
(4.15)

Le critère Γ peut se simplifier ainsi

$$\Gamma(\theta) = \sum_{m=1}^{L_1} |R_{y_1}^{m\alpha_1}(0)|^2 - \sum_{m=1}^{L_2} |R_{y_2}^{m\alpha_1}(0)|^2.$$

Remarque 4.2. Notons que le premier terme dans (4.15) correspond au critère proposé par Bouguerriou et al. dans [23].

4.3.3 Extension du critère au cas de plusieurs sources $(p \ge 3)$

La fonction coût Γ se généralise au cas de p sources de la manière suivante

$$\Gamma(\theta) = \sum_{i=1}^{p} \sum_{m=1}^{L_i} \left(|R_{y_i}^{m\alpha_i}(0)|^2 - \sum_{j=1; j \neq i}^{p} |R_{y_i}^{m\alpha_j}(0)|^2 \right).$$
(4.16)

La méthode proposée est justifiée par la proposition suivante (pour la démonstration voir Annexe; **A.2**).

Proposition 4.1. La maximisation du critère Γ conduit à la parfaite séparation des sources. En outre, la valeur maximale existe et est unique.

Remarque 4.3. Notons que le critère Γ est symétrique et donne les estimations de toutes les sources simultanément.

Algorithme

Nous résumons la méthode de séparation proposée par l'algorithme suivant

Blanchiment des mélanges z = Wx.
Initialisation de θ, μ.
Itérer :

y = Uz.
Calculer dΓ(θ)/dθ.
θ ← θ − μ dΓ(θ)/dθ.

Répéter jusqu'à la convergence.
Restitution des sources s = U(θ)z.

4.3.4 Résultats de simulations

Nous présentons dans ce paragraphe deux exemples de simulations pour illustrer la performance de la méthode. Dans le premier exemple, on traite le cas de deux sources cyclostationnaires de fréquences cycliques différentes. Dans le deuxième exemple, nous traitons le cas de trois sources cyclostationnaires de fréquences cycliques toutes différentes.

Exemple 4.1. Considérons les deux sources cyclostationnaires à l'ordre deux s_1 et s_2 respectivement de fréquences cycliques $\alpha_1 = 2F_1 = 80$ Hz et $\alpha_2 = 2F_2 = 30$ Hz

$$s_1(n) = a(n)\cos(2\pi F_1 n),$$

 $s_2(n) = b(n)\cos(2\pi F_2 n),$

où a(n) et b(n) sont des signaux blancs indépendants gaussiens centrés et réduits. Le nombre d'échantillons et la fréquence d'échantillonnage sont fixés respectivement à 4000 et 40 KHz. La matrice de mélange est

$$\mathbf{A} := \left(\begin{array}{cc} 1 & 0.9 \\ 0.75 & 1 \end{array} \right)$$

La figure 4.1 illustre la courbe representative du critère proposé; c'est-à-dire la fonction coût

$$\theta \in \left] \frac{-\pi}{2}, \frac{\pi}{2} \right[\mapsto \Gamma(\theta).$$
(4.17)

Sur ce même diagramme on a représenté l'évolution du critère proposé par Bouguerriou et al. dans [23] (pour le principe de cet algorithme voir Annexe; A.5) :

$$\theta \in \left] \frac{-\pi}{2}, \frac{\pi}{2} \right[\mapsto C_i(\theta) = \left| R_{y_i}^{\alpha_i}(0) \right|^2, i = 1, 2,$$

ainsi que l'erreur quadratique moyenne

$$\theta \in \left] \frac{-\pi}{2}, \frac{\pi}{2} \right[\mapsto EQM(\mathbf{s}, \mathbf{y}, \theta) = \sum_{i=1}^{2} \left[s_i(n) - y_i(\theta) \right]^2$$
(4.18)

entre les signaux sources **s** et les estimés **y**. Dans (4.17), on choisit $L_1 = L_2 =$ 1; les autres coefficients de la décomposition en série de Fourier sont considérés comme négligeables. On constate que notre critère est régulier; il admet un seul maximum $\hat{\theta} = -0.0995$ rd correspondant à l'estimation optimale (au sens de l'erreur quadratique moyenne) des sources s_1 et s_2 ; (voir figure 4.1). Les maxima de C_1 et C_2 sont respectivement $\hat{\theta}_1 = -0.0962$ rd et $\hat{\theta}_2 = -0.1027$ rd. Tous les maxima sont calculés par utilisation de l'algorithme de descente du gradient avec un pas fixe ($\mu =$ 0.01), pour 50 itérations, en prenant comme point initial $\theta = 0$. Nous observons que le maximum du critère proposé Γ est très proche de l'optimum théorique $\tilde{\theta} = -0.0996$ rd, qui correspond aussi au minimum de l'EQM.

Nous présentons dans le tableau (4.1), en dB, le SNR pour les trois algorithmes Sobi, Cyclosobi, Jade (pour le principe de ces algorithmes voir Annexe; A.5), et ceux exploitant les fonctions de coût C_1 , C_2 et Γ , où la taille des signaux est N = 4000. Les simulations sont répétées 500 fois. Nous constatons que la méthode Sobi est la moins performante parmi ces méthodes car elle a été conçue pour la séparation de signaux stationnaires. Nous constatons également que la méthode Cyclosobi proposée par Boustany et Antoni [27], et qui représente une extension de la méthode Sobi mais pour les signaux cyclostationnaires, donne des résultats mielleurs que celle de Sobi. La méthode Jade donne des résultats intermédiaires entre la méthode Cyclosobi et la méthode de Bougarriou C_1 et C_2 qui consiste à extraire un signal d'intérêt cyclostationnaire. Nous observons que la méthode utilisant le critère Γ que nous avons proposé donne les meilleurs résultats. Nous présentons dans la figure 4.2 les signaux sources, les observations et les sources estimées par maximisation de Γ .



FIG. 4.1 – Fonctions coût Γ , C_1 , C_2 et Erreur Quadratique Moyenne.

	(\mathbf{s}, \mathbf{x})	$Sobi(\mathbf{s}, \mathbf{y})$	$Cyclosobi(\mathbf{s},\mathbf{y})$	$Jade(\mathbf{s}, \mathbf{y})$	$C_1(\mathbf{s}, \mathbf{y})$	$C_2(\mathbf{s},\mathbf{y})$	$\Gamma(\mathbf{s}, \mathbf{y})$
s_1	3.99	22.25	31.58	35.41	38.78		41.08
s_2	4.91	22.34	30.65	34.44		39.05	41.94

TAB. 4.1 – SNR en dB pour Sobi, Cyclosobi, Jade, C_1 , C_2 et Γ .

Exemple 4.2. Considérons trois sources cyclostationnaires à l'ordre deux de fréquences



FIG. 4.2 – Les signaux sources (en haut), les mélanges (au milieu) et les sources estimées (en bas) obtenus par maximisation de Γ .

cycliques respectives $\alpha_1 = 2F_1 = 60$ Hz, $\alpha_2 = 2F_2 = 140$ Hz et $\alpha_3 = 2F_3 = 260$ Hz

$$s_1(n) = a(n)\cos(2\pi F_1 n),$$

$$s_2(n) = b(n)\cos(2\pi F_2 n),$$

$$s_3(n) = c(n)\cos(2\pi F_3 n),$$

où a(n), b(n) et c(n) sont des signaux blancs gaussiens indépendants centrés et réduits. Le nombre d'échantillons et la fréquence d'échantillonnage sont fixés respectivement à 4000 et 40 Khz. La matrice de mélange est

$$\mathbf{A} := \left(\begin{array}{cccc} 1 & 0.9 & 0.75 \\ 0.75 & 1 & 0.8 \\ 0.65 & 0.85 & 1 \end{array} \right).$$

Dans ce cas 3×3 de séparation de sources, après l'étape de blanchiment, nous aurons besoin de trois rotations θ_1 , θ_2 et θ_3 de l'espace des signaux blanchis pour pouvoir restituer les sources. Soit U une matrice 3×3 telle que $UU^{T} = I$ et det U = 1. Elle s'écrit

$$\mathbf{U}(\theta) = U_1(\theta_1) \cdot U_2(\theta_2) \cdot U_3(\theta_3), \tag{4.19}$$

où $U_1(\theta_1), U_2(\theta_2)$ et $U_3(\theta_3)$ sont respectivement les rotations de Givens suivantes

$$\left(\begin{array}{ccc}\cos\theta_1 & \sin\theta_1 & 0\\ -\sin\theta_1 & \cos\theta_1 & 0\\ 0 & 0 & 1\end{array}\right), \left(\begin{array}{ccc}\cos\theta_2 & 0 & \sin\theta_2\\ 0 & 1 & 0\\ -\sin\theta_2 & 0 & \cos\theta_2\end{array}\right), \left(\begin{array}{ccc}1 & 0 & 0\\ 0 & \cos\theta_3 & \sin\theta_3\\ 0 & -\sin\theta_3 & \cos\theta_3\end{array}\right).$$

Les paramètres θ_1 , θ_2 et θ_3 représentent les trois angles de rotation possibles de l'espace des signaux blanchis \mathbf{z} , où $-\pi/2 < \theta_1, \theta_2, \theta_3 < \pi/2$. Le critère $\Gamma(\theta)$ dans ce cas s'écrit (voir l'équation (4.16))

$$\begin{split} \Gamma(\theta) &= \sum_{m=1}^{L_1} \left(|R_{y_1}^{m\alpha_1}(0)|^2 - \left(|R_{y_1}^{m\alpha_2}(0)|^2 + |R_{y_1}^{m\alpha_3}(0)|^2 \right) \right) + \\ &+ \sum_{m=1}^{L_2} \left(|R_{y_2}^{m\alpha_2}(0)|^2 - \left(|R_{y_2}^{m\alpha_1}(0)|^2 + |R_{y_2}^{m\alpha_3}(0)|^2 \right) \right) + \\ &+ \sum_{m=1}^{L_2} \left(|R_{y_3}^{m\alpha_3}(0)|^2 - \left(|R_{y_3}^{m\alpha_1}(0)|^2 + |R_{y_3}^{m\alpha_2}(0)|^2 \right) \right). \end{split}$$

Le maximum de la fonction coût Γ est calculé par utilisation de l'algorithme de descente du gradient avec un pas fixe ($\mu = 0.001$), pour 100 itérations, en prenant comme point initial ($\theta_1, \theta_2, \theta_3$) = (0,0,0). Nous montrons sur le tableau (4.2), les SNRs entre les signaux sources et sources estimées par maximisation de $C_i, i = 1, 2, 3$ et Γ , où la taille des signaux est N = 4000. Les simulations sont répétées 500 fois. Nous constatons que la méthode Sobi ne donne pas de bons résultats. La méthode Jade donne de meilleurs résultats. La méthode de Bouguerriou C_1, C_2 et C_3 permet de séparer les signaux avec des résultats plus performants que ceux obtenus avec la méthode de Jade. Nous observons que la méthode proposée Γ donne les meilleurs résultats. Nous présentons dans la figure 4.3 les signaux sources, les observations et les sources estimés par maximisation de Γ .

	(\mathbf{s}, \mathbf{x})	$Sobi(\mathbf{s},\mathbf{y})$	$Jade(\mathbf{s}, \mathbf{y})$	$C_1(\mathbf{s}, \mathbf{y})$	$C_2(\mathbf{s}, \mathbf{y})$	$C_3(\mathbf{s}, \mathbf{y})$	$\Gamma(\mathbf{s},\mathbf{y})$
s_1	1.54	12.20	29.59	35.75			36.85
s_2	1.85	11.18	29.39		35.30		37.20
s_3	1.97	10.22	28.77			35.50	37.15

4.4 Méthodes basées sur la minimisation des divergences pénalisées 111

TAB. 4.2 – SNR en dB pour Sobi, Jade, C_1 , C_2 , C_3 et Γ .

4.4 Méthodes basées sur la minimisation des divergences pénalisées

4.4.1 Introduction

Dans le cas linéaire instantané, le mélange de deux sources est de la forme

$$\mathbf{x} = \mathbf{As},$$

où $\mathbf{s} := (s_1, s_2)^T$, $\mathbf{x} = (x_1, x_2)^T$ et \mathbf{A} sont respectivement les signaux sources, les signaux mélanges et la matrice de mélange de dimension 2×2 . Le but est d'estimer les sources \mathbf{s} à partir des observations \mathbf{x} . L'estimateur, dans ce cas, est de la forme

$$\mathbf{y} = \mathbf{B}\mathbf{x},\tag{4.20}$$

où \mathbf{B} est une matrice 2 × 2. Le problème consiste donc à chercher un estimateur \mathbf{B} , à partir des observations \mathbf{x} , qui conduit à la meilleure estimation possible des sources \mathbf{s} ,

$$\mathbf{y} = \widehat{\mathbf{B}}\mathbf{x}$$
.

Une solution à ce problème, à une indétermination du facteur d'échelle et à une permutation près, consiste à chercher une matrice $\widehat{\mathbf{B}}$ qui rend les composantes de **y** indépendantes au sens statistique. Nous avons montré dans le chapitre 2 (paragraphe 2.3.3) et dans le contexte stationnaire que la minimisation des divergences,



FIG. 4.3 – Les signaux sources (en haut), les mélanges (au milieu) et les sources estimées (en bas) obtenus par maximisation de Γ .

entre $f_{y_1}f_{y_2}$ et $f_{\mathbf{y}}$, conduit à la résolution du problème de la SAS et donne des résultats satisfaisants. Nous allons voir comment ces critères se manifestent dans le contexte cyclostationnaire. Ensuite, nous allons considérer une méthode utilisant des statistiques cycliques d'ordre 2 afin d'estimer la matrice séparation **B**. Enfin, comme nous sommes dans un contexte cyclostationnaire et pour intégrer la notion de cyclostationnarité, nous proposons de combiner les deux critères précédents; i.e., pénaliser les critères des divergences (2.3.3) par des statistiques cycliques d'ordre 2.

4.4.2 Méthodes basées sur la minimisation des α -divergences

Cette méthode a été développée, dans le chapitre 2 (section 2.3), dans le contexte stationnaire. Nous allons aborder ces critères dans le contexte cyclostationnaire. Dans ce cas, la séparation sera obtenue par la minimisation du critère (2.34) par rapport aux coefficients b_{pq} de la matrice séparante **B**, i.e.,

$$\min_{\mathbf{B}} \widehat{I}_{\varphi_{\alpha}}(\mathbf{y}) = \min_{\mathbf{B}} \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha} \left(\frac{\widehat{f}_{y_1}(y_1(n)) \widehat{f}_{y_2}(y_2(n))}{\widehat{f}_{\mathbf{y}}(y_1(n), y_2(n))} \right).$$
(4.21)

Résultats de simulations

Nous présentons dans ce paragraphe un exemple de simulations pour illustrer les performances de ces méthodes pour des signaux cyclostationnaires. On traite le cas de deux sources indépendantes cyclostationnaires de fréquences cycliques différentes.

Exemple 4.3. Considérons les deux sources cyclostationnaires à l'ordre deux s_1 et s_2 respectivement de fréquences cycliques $\alpha_1 = 2F_1 = 80$ Hz et $\alpha_2 = 2F_2 = 30$ Hz

$$s_1(n) = a(n)\cos(2\pi F_1 n),$$

 $s_2(n) = b(n)\cos(2\pi F_2 n),$

où a(n) et b(n) sont des signaux indépendants blancs gaussiens centrés et réduits. Le nombre d'échantillons et la fréquence d'échantillonnage sont fixés respectivement à 1000 et 5 KHz. La matrice de mélange est

$$\mathbf{A} := \left(\begin{array}{cc} 1 & 0.85\\ 0.65 & 1 \end{array}\right)$$

La détermination de la matrice séparante **B** a été calculée par utilisation de l'algorithme de descente du gradient par rapport aux coefficients de la matrice séparante. Nous présentons dans la figure 4.4, en dB, le SNR moyen en fonction du nombre d'itérations pour les trois critères IM ($\alpha = 0$), Hellinger ($\alpha = 0.5$) et KL ($\alpha = 1$). Les simulations sont répétées 100 fois pour des données non-bruitées. Nous remarquons que les trois critères IM, H et KL nécessitent environ 60 itérations pour atteindre la convergence dans le cas cyclostationnaire (voir la figure 4.4); cependant dans le cas stationnaire la convergence de ces trois critères ne nécessite qu'environ 40 itérations; voir la figure 2.13 dans le chapitre 2. Nous constatons également que les performances de ces méthodes se dégradent dans le cas cyclostationnaire (les SNRs moyens des trois critères IM, H et KL dans le cas cyclostationnaire sont d'environ 35 dB; voir la figure 4.4, tandis que dans le cas stationnaire nous obtenons près de 40 dB; voir la figure 2.13 dans le chapitre 2). Nous constatons que les méthodes basées sur la minimisation des α -divergences sont plus performantes avec des signaux stationnaires.



FIG. 4.4 – SNRs moyens en fonction du nombre d'itérations pour les trois critères IM, Hellinger et KL.

4.4.3 Méthode utilisant les statistiques d'ordre 2

Dans l'article [1], Abed-Meraim *et al.* ont prouvé que pour des sources cyclostationnaires possédant des fréquences cycliques distinctes, la matrice \mathbf{B} est une matrice de séparation si et seulement si

$$R_{y_i}^{\alpha_i}(0) = 1$$
 et $R_{y_{ij}}^{\alpha_i}(\tau) = 0$, pour tout $1 \le i \ne j \le p$, (4.22)

4.4 Méthodes basées sur la minimisation des divergences pénalisées 115

où $R_{y_i}^{\alpha_i}(0)$ et $R_{y_{ij}}^{\alpha_i}(\tau)$ sont respectivement l'autocorrélation cyclique de y_i et l'intercorrélation cyclique de y_{ij} , et τ est un décalage temporel. Les auteurs ont également montré que dans le cas où les deux conditions précédentes (avec $\tau = 0$) sont satisfaites, i.e., si $R_{y_i}^{\alpha_i}(0) = 1$ et $R_{y_{ij}}^{\alpha_i}(0) = 0$, la matrice **B** demeure la solution du problème (4.20). Rappelons que dans [1], Abed-Meraim *et al.* ont formulé le critère (4.22), pour $\tau = 0$, et ont abouti à la fonction de contraste suivante

$$G(\mathbf{y}) := \sum_{i=1}^{r} |R_{y_i}^{\alpha_i}(0) - 1|^2 + \sum_{1 \le i < j \le p} |R_{y_{ij}}^{\alpha_i}(0) + R_{y_{ji}}^{\alpha_j}(0)|^2 + |R_{y_{ij}}^{\alpha_i}(0) - R_{y_{ji}}^{\alpha_j}(0)|^2.$$
(4.23)

La matrice séparante **B** est obtenue par minimisation de la fonction de contraste $G(\mathbf{y})$. Dans le même esprit, nous proposons de minimiser en **B** le critère suivant, obtenu après réduction du 2^{eme} terme de l'expression $G(\mathbf{y})$ et dans le cas de deux sources,

$$C(\mathbf{y}) = \sum_{i=1}^{2} |R_{y_i}^{\alpha_i}(0) - 1|^2 + 2|R_{y_{12}}^{\alpha_1}(0)|^2 + 2|R_{y_{21}}^{\alpha_2}(0)|^2, \ i = 1, 2,$$
(4.24)

où $R_{y_i}^{\alpha_i}(0)$ est l'autocorrélation cyclique du signal y_i et $R_{y_{12}}^{\alpha_1}(0)$ et $R_{y_{21}}^{\alpha_2}(0)$ sont les intercorrélations cycliques de y_1 et y_2 par rapport aux fréquences cycliques α_1 et α_2 . En d'autre termes, la minimisation de cette fonction de contraste conduit à rendre maximum les autocorrélations cycliques $R_{y_i}^{\alpha_i}(0)$, i = 1, 2, et minimum les intercorrélations cycliques $R_{y_{12}}^{\alpha_1}(0)$ et $R_{y_{21}}^{\alpha_2}(0)$.

Nous allons utiliser la méthode de descente du gradient pour calculer la matrice séparante **B**. Nous avons donc besoin de calculer le gradient du critère $C(\mathbf{y})$. Nous en donnons, dans le paragraphe suivant, la forme explicite.

Calcul du gradient

Notons b_{pq} , p, q = 1, 2, les éléments de la matrice **B**. Le gradient de ce critère par rapport aux coefficients b_{pq} de la matrice séparante peut s'écrire comme suit

$$\frac{\partial C(\mathbf{y})}{\partial b_{pq}} = \frac{\partial}{\partial b_{pq}} \left(\sum_{i=1}^{2} |R_{y_i}^{\alpha_i}(0) - 1|^2 + 2|R_{y_{12}}^{\alpha_1}(0)|^2 + 2|R_{y_{21}}^{\alpha_2}(0)|^2 \right).$$

Pour simplifier les résultats présentés nous optons pour les notations suivantes

$$R_{y_1}^{\alpha_1}(0) := R_{y_1}^{\alpha_1}, \quad R_{y_2}^{\alpha_2}(0) := R_{y_2}^{\alpha_2}, \quad R_{y_{12}}^{\alpha_1}(0) := R_{y_{12}}^{\alpha_1} \quad \text{et} \quad R_{y_{21}}^{\alpha_2}(0) := R_{y_{21}}^{\alpha_2}$$

Un calcul direct donne les expressions suivantes

$$\frac{\partial C(\mathbf{y})}{\partial b_{pq}} = 2 \frac{\partial R_e(R_{y_1}^{\alpha_1})}{\partial b_{pq}} [R_e(R_{y_1}^{\alpha_1}) - 1] + 2I_m(R_{y_1}^{\alpha_1}) \frac{\partial I_m(R_{y_1}^{\alpha_1})}{\partial b_{pq}}
+ 2 \frac{\partial R_e(R_{y_2}^{\alpha_2})}{\partial b_{pq}} [R_e(R_{y_2}^{\alpha_2}) - 1] + 2I_m(R_{y_2}^{\alpha_2}) \frac{\partial I_m(R_{y_2}^{\alpha_2})}{\partial b_{pq}}
+ 4R_e(R_{y_{12}}^{\alpha_1}) \frac{\partial R_e(R_{y_{12}}^{\alpha_1})}{\partial b_{pq}} + 4I_m(R_{y_{12}}^{\alpha_1}) \frac{\partial I_m(R_{y_{12}}^{\alpha_1})}{\partial b_{pq}}
+ 4R_e(R_{y_{21}}^{\alpha_2}) \frac{\partial R_e(R_{y_{21}}^{\alpha_2})}{\partial b_{pq}} + 4I_m(R_{y_{21}}^{\alpha_2}) \frac{\partial I_m(R_{y_{21}}^{\alpha_2})}{\partial b_{pq}}.$$
(4.25)

Les détails du calcul du gradient (4.25) sont donnés en Annexe (voir A.4).

Résultats de simulations

Nous présentons dans ce paragraphe un exemple de simulations pour illustrer les performances de cette méthode. On traite le cas de deux sources cyclostationnaires de fréquences cycliques différentes.

Exemple 4.4. Dans cette exemple, nous utilisons les mêmes données que celles traitées dans l'exemple (4.3). Nous présentons sur la figure 4.5 le SNR moyen des signaux sources estimés par utilisation du critère (4.24). Nous constatons que celuici converge plus rapidement, par contre les résultats obtenus sont un peu moins performants que ceux présentés sur la figure 4.4.

4.4.4 Méthodes basées sur la minimisation des α -divergences pénalisées (combinaison des deux critères 4.21 et 4.24)

Dans ce paragraphe, nous nous proposons de combiner les deux critères 4.21 et 4.24. Cette combinaison nous permet de proposer un nouveau critère de SAS que



FIG. 4.5 - SNR moyen en fonction du nombre d'itérations pour le critère C.

l'on appelle critère basé sur la minimisation des α -divergences "pénalisées", que l'on note $IP_{\varphi_{\alpha}}(f_{y_1}f_{y_2}, f_{\mathbf{y}}))$ et qui peut donc s'écrire, dans le cas de deux sources-deux observations, comme suit

$$IP_{\varphi_{\alpha}}(f_{y_{1}}f_{y_{2}}, f_{\mathbf{y}}) := \frac{1}{N} \sum_{n=1}^{N} \varphi_{\alpha} \left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n))}{\widehat{f}_{\mathbf{y}}(y_{1}(n), y_{2}(n))} \right) \\ + \left[\sum_{i=1}^{2} |R_{y_{i}}^{\alpha_{i}}(0) - 1|^{2} + 2|R_{y_{12}}^{\alpha_{1}}(0)|^{2} + 2|R_{y_{21}}^{\alpha_{2}}(0)|^{2} \right], \quad (4.26)$$

où le premier terme représente les divergences entre densités de probabilité et le deuxième terme représente une pénalisation. Cette pénalisation a été introduite pour tenir compte de la cyclostationnarité des signaux sources. La séparation sera obtenue par minimisation du critère (4.26) par rapport aux coefficients b_{pq} de la matrice séparante **B**, i.e.,

$$\min_{\mathbf{B}} IP_{\varphi_{\alpha}}(\mathbf{y}) = \min_{\mathbf{B}} \left[I_{\varphi_{\alpha}}(\mathbf{y}) + C(\mathbf{y}) \right].$$
(4.27)

Nous allons utiliser la méthode de descente du gradient pour calculer ce minimum.

Calcul du gradient

Le gradient de ce critère, par rapport aux coefficients b_{pq} , p, q = 1, 2 de la matrice séparante **B**, peut s'écrire comme suit

$$\frac{\partial I P_{\varphi_{\alpha}}(\mathbf{y})}{\partial b_{pq}} = \frac{\partial I_{\varphi_{\alpha}}(\mathbf{y})}{\partial b_{pq}} + \frac{\partial C(\mathbf{y})}{\partial b_{pq}},\tag{4.28}$$

les deux termes $\frac{\partial I_{\varphi_{\alpha}}(\mathbf{y})}{\partial b_{pq}}$ et $\frac{\partial C(\mathbf{y})}{\partial b_{pq}}$ sont déjà calculés; voir (2.43) et (4.25).

Algorithme

Nous résumons la méthode de séparation proposée par l'algorithme suivant

Résultats de simulations

Nous présentons dans ce paragraphe trois exemples de simulations pour illustrer les performances de ces méthodes. Dans le premier, on traite le cas de deux sources cyclostationnaires de fréquences cycliques différentes. Dans le deuxième exemple, nous traitons le cas où les données sont contaminées. Ensuite, dans le troisième exemple, nous comparons la méthode basée sur la divergence de Hellinger pénalisée $\widehat{IP}_{\varphi_{0.5}}(\mathbf{y})$ avec la méthode (4.14) qui utilise les statistiques cycliques d'ordre 2 présentées dans la section (4.3).

Exemple 4.5. (cas non bruité)

Dans cette exemple, nous utilisons les mêmes données que celles traitées dans l'exemple
(4.3). La matrice séparante B est calculée par utilisation de l'algorithme de descente

du gradient par rapport aux coefficients de la matrice séparante **B**. Nous présentons dans le figure 4.6, en dB, le SNR moyen en fonction du nombre d'itérations pour les trois critères KL_m pénalisée $(\widehat{IP}_{\varphi_0})$, H pénalisée $(\widehat{IP}_{\varphi_{0.5}})$ et KL pénalisée $(\widehat{IP}_{\varphi_1})$. Les simulations sont répétées 100 fois pour des données non bruitées. Nous remarquons que les trois critères \widehat{IP}_{φ_0} , $\widehat{IP}_{\varphi_{0.5}}$ et \widehat{IP}_{φ_1} pénalisés nécessitent environ 20 itérations pour atteindre la convergence. Nous constatons donc que la pénalisation améliore la vitesse de convergence et l'efficacité (environs 2 dB en moyenne) par rapport aux méthodes basées sur la minimisation des α -divergences sans pénalisation (figure 4.4). Donc la combinaison des deux critères permet d'améliorer les performances. Nous ne constatons également aucune différence sensible entre les trois critères utilisant les divergences pénalisées. Nous présentons sur la figure 4.7 les signaux sources, les signaux mélangés et les sources estimées par la méthode utilisant la divergence de Hellinger pénalisée. Nous constatons que cette méthode permet de faire la séparation et de restituer les sources originales avec un bon SNR (proche de 38 dB).

Exemple 4.6. (cas bruité)

Dans cette exemple, nous contaminons les mêmes données que celles traitées dans l'exemple (4.3). Nous présentons sur la figure 4.8 les SNR moyens en fonction du RSB pour les trois critères, \widehat{IP}_{φ_0} , $\widehat{IP}_{\varphi_{0.5}}$ et \widehat{IP}_{φ_1} . Le RSB varie entre -60 dB et -20 dB. Les simulations sont répétées 50 fois. Nous constatons que \widehat{IP}_{φ_0} est plus robuste pour l'intervalle entre -60 et -26 dB, cependant entre -26 et -20 dB, \widehat{IP}_{φ_0} est moins robuste. On observe le phénomène contraire pour \widehat{IP}_{φ_1} . Le critère basé sur la divergence de Hellinger pénalisée ($\widehat{IP}_{\varphi_{0.5}}$) présente toujours un bon compromis entre les critères \widehat{IP}_{φ_1} et \widehat{IP}_{φ_0} , indépendamment du degré de contamination des données. Il convient donc d'utiliser le critère basé sur la divergence de Hellinger pénalisée ($\widehat{IP}_{\varphi_{0.5}}$) en l'absence d'information sur le degré de contamination des données.

Exemple 4.7. (comparaison des performances des critères $IP_{\varphi_{0.5}}$ et Γ)



FIG. 4.6 – SNRs moyens en fonction des itérations pour les trois critères $\widehat{IP}_{\varphi_0}, \widehat{IP}_{\varphi_{0.5}}$ et \widehat{IP}_{φ_1} .

Dans cet exemple, nous utilisons les mêmes données (non bruitées) que celles traitées dans l'exemple (4.3). Nous montrons sur le tableau (4.3), les SNRs entre les signaux sources et sources estimées par la méthode utilisant la divergence de Hellinger pénalisée ($\widehat{IP}_{\varphi_{0.5}}$) et celle utilisant les statistiques d'ordre deux Γ (4.14). Les simulations sont répétées 100 fois. Nous constatons que la méthode utilisant la divergence de Hellinger pénalisée donne de meilleures performances que celles obtenues par les statistiques d'ordre 2; on constate en effet d'après le tableau (4.3) que l'amélioration apportée est très proche de 1 dB.

4.5 Conclusion

Dans ce chapitre, nous avons proposé un nouveau critère de séparation aveugle de sources cyclostationnaires au second ordre de fréquences cycliques supposées connues et différentes deux-à-deux, dans le cadre d'un mélange linéaire instantané.



FIG. 4.7 – Les signaux sources (en haut), les mélanges (au milieu) et les sources estimées (en bas) par la méthode basée sur la divergence de Hellinger pénalisée $(\widehat{IP}_{\varphi_{0.5}}).$

La méthode exploite les caractéristiques cyclostationnaires à l'ordre deux des signaux sources. La méthode a été testée pour le cas de deux sources mais peut être étendue à un nombre quelconque de sources par généralisation. Elle est facile à mettre en oeuvre, et très peu coûteuse en temps de calcul; la fonction coût proposée est régulière, elle présente un seul maximum et ce quelque soit la dimension du paramètre de rotation. Ceci permet un calcul rapide du maximum par utilisation des algorithmes de type descente du gradient. Nous avons également proposé des algorithmes de séparation basés sur les α -divergences pénalisées. Nous avons constaté que la pénalisation améliore la vitesse de convergence et l'efficacité. Nous avons également constaté que la méthode basée sur la divergence de Hellinger pénalisée a de bonnes propriétés d'efficacité-robustesse en présence du bruit en comparaison avec les algorithmes utilisant l'IM pénalisée ou la divergence de KL pénalisée. Nous

	(\mathbf{s}, \mathbf{x})	$\Gamma(\mathbf{s},\mathbf{y})$	$IP_{\varphi_{0.5}}(\mathbf{s},\mathbf{y})$
s_1	3.99	36.32	37.30
s_2	4.91	37.17	38.18

TAB. 4.3 – SNR en dB pour le critère Γ (4.14) et celui utilisant la divergence de Hellinger pénalisée $IP_{\varphi_{0.5}}$.

avons aussi constaté que pour des données non bruitées le critère utilisant la divergence de Hellinger pénalisée conduisait à de meilleurs résultats que celui basé sur les statistiques cycliques d'ordre 2.



FIG. 4.8 – En haut : SNR des trois critères : \widehat{IP}_{φ_0} , $\widehat{IP}_{\varphi_{0.5}}$ et \widehat{IP}_{φ_1} , en fonction du RSB. En bas (de gauche à droite) : zoom sur SNR (autour de -50 dB et autour de -20 dB).

Conclusion générale

Tout au long de la thèse, nous nous sommes intéressés au problème de la séparation aveugle de mélanges linéaires instantanés et convolutifs de sources stationnaires ou de sources cyclostationnaires. De nombreuses approches ont été développées et étudiées : elles dépendent du mélange et de la nature des signaux sources considérés. Dans le premier chapitre de cette thèse, nous avons mis en avant l'importance de la SAS dans la résolution de nombreux problèmes physiques concrets. Nous avons fait le point sur les différentes techniques les plus courantes mises en oeuvre pour assurer la séparation en fonction de certaines caractéristiques des sources et des mélanges ; ceci pour mieux appréhender les améliorations apportées par les algorithmes que nous avons été amenés à développer par la suite.

Dans la première partie de cette thèse, nous avons développé de nouvelles méthodes de la SAS stationnaires. Dans le chapitre 2 et pour des mélanges linéaires instantanés, nous avons tout d'abord proposé une méthode basée sur la minimisation de l'information mutuelle sous contraintes par un passage au problème dual et estimation du gradient stochastique par maximum de vraisemblance dans des familles de lois exponentielles. L'algorithme proposé s'applique pour une large classe de lois de probabilité pouvant être modélisées par une famille de lois exponentielles. Nous avons illustré les performances de cet algorithme, en terme d'efficacité, avec des signaux simulés en comparaison avec deux autres méthodes proposées par Babaie-Zadeh et Pham. Le critère s'applique quelque soit le nombre de sources ; par exemple le cas de trois sources a été illustré. Ensuite, nous avons présenté des algorithmes de séparation aveugle de signaux utilisant les α -divergences pour des mélanges linéaires instantanés. Celles-ci généralisent l'approche de l'information mutuelle qui est obtenue pour le choix particulier de la divergence de Kullback-Leibler modifiée. Nous avons montré que la méthode basée sur la divergence de Hellinger présente de bonnes propriétés de robustesse en présence du bruit en comparaison avec l'information mutuelle et la divergence de Kullback-Leibler.

Dans l'objectif d'étendre les méthodes précédentes au cas des mélanges linéaires convolutifs, nous avons présenté dans le chapitre 3 de nouveaux algorithmes de la SAS basés sur la minimisation des α -divergences pour des mélanges linéaires convolutifs. Les méthodes proposées généralisent l'approche de l'information mutuelle dans le cas des mélanges linéaires convolutifs. Nous avons montré par simulation que l'algorithme utilisant la divergence de Hellinger a de meilleures propriétés de robustesse en présence du bruit en comparaison avec l'information mutuelle et la divergence de Kullback-Leibler. Nous avons illustré les performances de cet algorithme, en termes d'efficacité et robustesse, sur des signaux simulés. Enfin, une étude expérimentale est décrite sur l'application de l'algorithme utilisant la divergence de Hellinger aux mesures vibratoires d'une machine d'essai de roulements. Les résultats obtenus sont de qualité satisfaisante ; toutes les raies sont correctement affectées (absence de permutations de raies pour des canaux de même fréquence) et les défauts sont correctement identifiés et attribués.

Dans la deuxième partie de cette thèse, nous avons développé de nouvelles méthodes de la SAS cyclostationnaires en utilisant la connaissance de leurs fréquences cycliques fondamentales. Nous avons tout d'abord proposé un nouveau critère de séparation aveugle de sources cyclostationnaires au second ordre de fréquences cycliques supposées connues et différentes deux-à-deux, dans le cadre d'un mélange linéaire instantané. La méthode exploite les caractéristiques cyclostationnaires à l'ordre deux

Conclusion générale

des signaux sources. La fonction coût proposée est régulière, elle présente un seul maximum et ce quelque soit la dimension du paramètre de rotation. Ceci permet un calcul rapide du maximum par utilisation des algorithmes de type descente du gradient. Ensuite, nous avons également proposé des algorithmes de séparation basés sur les α -divergences "pénalisées". Nous avons constaté que la pénalisation améliore la vitesse de convergence et l'efficacité. Nous avons également constaté que la méthode basée sur la divergence de Hellinger pénalisée a de bonnes propriétés de robustesse en présence du bruit. Enfin, nous avons montré que le critère basé sur la divergence de Hellinger pénalisé que la celui utilisant seulement les statistiques cycliques d'ordre deux.

Des perspectives se dégagent selon que l'on traite de la technique de la séparation de sources dans son cadre général ou que l'on s'intéresse à son application.

- Durant notre étude expérimentale, nous avons remarqué une certaine lenteur de la méthode basée sur la divergence de Hellinger du fait de notre méthode d'estimation de densité, la méthode à noyau. Par conséquent, il serait intéressant de tester une autre méthode plus rapide (par exemple, des méthodes utilisant les ondelettes pour l'estimation des densités).
- Nous n'avons pas réalisé d'essais sur la base de trois sources réelles. Nous souhaitons tester l'algorithme utilisant la divergence de Hellinger pour illustrer l'essai de plusieurs sources réelles.
- Les algorithmes proposés, pour séparer des sources cyclostationnaires, ne traitent à ce jour que du cas des mélanges instantanés. Une étude sur le cas convolutif est en cours mais n'a pas encore formellement abouti, l'obstacle résidant principalement dans le choix d'une fonction de pénalisation appropriée.

Annexe

A.1

Lemme 4.1. Soit B une matrice carrée de dimension M inversible. On a

$$\frac{d}{d\mathbf{B}}\log|\det(\mathbf{B})| = \mathbf{B}^{-T}.$$
(4.29)

Démonstration :

Soit M_{ij} le déterminant de la matrice **B** à laquelle on a retiré la ligne *i* et la colonne *j*. La comatrice de **B**, **Com**(**B**), est donnée par

$$\left((-1)^{i+j}M_{ij}\right)_{1\leq i,j\leq M}.$$

On sait de plus que si ${\bf B}$ est inversible on a

$$\mathbf{B}^{-1} = \frac{\left[\mathbf{Com}(\mathbf{B})\right]^T}{\det(\mathbf{B})}.$$

On a donc pour les termes b_{ij}

$$\frac{\partial}{\partial b_{ij}} \log |\det(\mathbf{B})| = \frac{1}{\det(\mathbf{B})} \frac{\partial}{\partial b_{ij}} \det(\mathbf{B}) = \frac{(-1)^{i+j} M_{ij}}{\det(\mathbf{B})} = \left(\frac{\mathbf{Com}(\mathbf{B})}{\det(\mathbf{B})}\right)_{ij} = \left(\mathbf{B}^{-T}\right)_{ij}.$$

D'où

$$\frac{d}{d\mathbf{B}}\log|\det(\mathbf{B})| = \mathbf{B}^{-T}.$$

A.2

Démonstration de la Proposition 4.1

Nous nous restreignons au cas de deux sources (p = 2). Le cas de plusieurs sources $(p \ge 3)$ se démontre de manière analogue. Supposons que les signaux sources s_1 et s_2 sont centrés, de variance unité et statistiquement indépendants. Les signaux blanchis peuvent être écrits $\mathbf{z} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{V}\mathbf{s}$ où $\mathbf{V} = \mathbf{W}\mathbf{A}$ est une matrice unitaire, c'est-à-dire, $\mathbf{V}\mathbf{V}^T = \mathbf{I}$ et det $(\mathbf{V}) = 1$. Donc il existe $\hat{\alpha} \in] - \pi/2, \pi/2[$ tel que

$$\mathbf{V} = \begin{pmatrix} \cos(\widehat{\alpha}) & \sin(\widehat{\alpha}) \\ -\sin(\widehat{\alpha}) & \cos(\widehat{\alpha}) \end{pmatrix}$$

Les signaux estimés sont donc

$$\mathbf{y} = \mathbf{U}\mathbf{z} = \mathbf{U}\mathbf{V}\mathbf{s}$$
 où $\mathbf{U} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$. (4.30)

D'où,

$$y_{1}(n) = a(\theta)s_{1}(n) + b(\theta)s_{2}(n)$$

$$y_{2}(n) = -b(\theta)s_{1}(n) + a(\theta)s_{2}(n)$$
(4.31)

avec

$$a(\theta) = \cos(\theta)\cos(\widehat{\alpha}) - \sin(\theta)\sin(\widehat{\alpha});$$

$$b(\theta) = \cos(\theta)\sin(\widehat{\alpha}) + \sin(\theta)\cos(\widehat{\alpha}).$$

La valeur $\hat{\theta}$ de θ conduisant à la séparation est telle que $\mathbf{U}\mathbf{V} = \mathbf{I}$, c'est-à-dire, $\hat{\theta} = -\hat{\alpha}$. En fait, le calcul direct, montre que $a(\hat{\theta}) = 1$, $b(\hat{\theta}) = 0$. Nous montrerons que $\hat{\theta} = -\hat{\alpha}$ est l'unique valeur de θ qui maximise la fonction coût Γ . En utilisant la formule (4.31), et le fait que

$$R_{s_1}^{m\alpha_2}(0) = R_{s_2}^{m\alpha_1}(0) = R_{s_{12}}^{m\alpha_1}(0) = R_{s_{12}}^{m\alpha_2}(0) = 0,$$

nous pouvons écrire le critère sous la forme

$$\Gamma(\theta) = a(\theta)^4 \sum_{m=1}^{L_1} \left| R_{s_1}^{m\alpha_1}(0) \right|^2 - b(\theta)^4 \sum_{m=1}^{L_1} \left| R_{s_2}^{m\alpha_2}(0) \right|^2 - b(\theta)^4 \sum_{m=1}^{L_2} \left| R_{s_1}^{m\alpha_1}(0) \right|^2 + a(\theta)^4 \sum_{m=1}^{L_2} \left| R_{s_2}^{m\alpha_2}(0) \right|^2$$

$$(4.32)$$

 soit

$$\Gamma(\theta) = a(\theta)^4 \left[\sum_{m=1}^{L_1} \left| R_{s_1}^{m\alpha_1}(0) \right|^2 + \sum_{m=1}^{L_2} \left| R_{s_2}^{m\alpha_2}(0) \right|^2 \right] - b(\theta)^4 \left[\sum_{m=1}^{L_1} \left| R_{s_1}^{m\alpha_1}(0) \right|^2 + \sum_{m=1}^{L_2} \left| R_{s_2}^{m\alpha_2}(0) \right|^2 \right].$$
(4.33)

La dérivée de (4.33) peut être simplifiée en

$$\frac{d\Gamma(\theta)}{d\theta} = \left(4a(\theta)^3 a'(\theta) - 4b(\theta)^3 b'(\theta)\right) \left(\sum_{m=1}^{L_1} \left|R_{s_1}^{m\alpha_1}(0)\right|^2 + \sum_{m=1}^{L_2} \left|R_{s_2}^{m\alpha_2}(0)\right|^2\right). \quad (4.34)$$

Comme

$$\left(\sum_{m=1}^{L_1} \left| R_{s_1}^{m\alpha_1}(0) \right|^2 + \sum_{m=1}^{L_2} \left| R_{s_2}^{m\alpha_2}(0) \right|^2 \right) \neq 0,$$

on peut écrire

$$\frac{d\Gamma(\theta)}{d\theta} = 0 \iff a(\theta)^3 a'(\theta) - b(\theta)^3 b'(\theta) = 0 \iff$$

$$\cos(\theta + \hat{\alpha}) = 0 \quad \text{ou} \quad \sin(\theta + \hat{\alpha}) = 0. \tag{4.35}$$

1) Si
$$0 < \widehat{\alpha} < \pi/2$$
, (4.35) $\iff \theta = -\widehat{\alpha}$ où $\theta = -\widehat{\alpha} + \pi/2$.

2) Si $-\pi/2 < \hat{\alpha} < 0$, (4.35) $\iff \theta = -\hat{\alpha}$ où $\theta = -\hat{\alpha} - \pi/2$.

Dans les deux cas, la première solution de $\theta = -\hat{\alpha}$ est un maximum et la seconde est un minimum. En fait, un calcul explicite montre que

$$\left. \frac{d^2 \Gamma(\theta)}{d\theta^2} \right|_{\theta = -\hat{\alpha}} < 0$$

 et

$$\left. \frac{d^2 \Gamma(\theta)}{d\theta^2} \right|_{\theta = -\widehat{\alpha} \pm \pi/2} > 0.$$

Cela prouve que la fonction coût Γ admet un maximum unique $\hat{\theta} = -\hat{\alpha}$ qui mène à la parfaite séparation des signaux sources.

A.3

Calcul de $\varphi_{\alpha}^{'}(\cdot)$

Pour calculer le terme $\varphi'_{\alpha}\left(\frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n-m))}{\hat{f}_{\mathbf{y}}(y_1(n),y_2(n-m))}\right)$, on revient à la définition des fonctions convexes. On a

Si $\alpha = 0$, alors

$$\varphi_{\alpha}(x) = -\log x + x - 1$$
 et $\varphi'_{\alpha}(x) = -\frac{1}{x} + 1.$

Si $\alpha = 1$, alors

$$\varphi_{\alpha}(x) = x \log x - x + 1$$
 et $\varphi'_{\alpha}(x) = \log x$

Si $\alpha \neq \{0, 1\}$, alors

$$\varphi_{\alpha}(x) = \frac{x^{\alpha} - \alpha x + \alpha - 1}{\alpha(\alpha - 1)}$$
 et $\varphi'_{\alpha}(x) = \frac{x^{\alpha - 1} - 1}{\alpha - 1}$

Lorsque on pose $x = \frac{\hat{f}_{y_1}(y_1(n))\hat{f}_{y_2}(y_2(n-m))}{\hat{f}_y(y_1(n),y_2(n-m))}$ on peut donc écrire

$$\varphi_{\alpha}^{'}\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n-m))}{\widehat{f}_{\mathbf{y}}(y_{1}(n),y_{2}(n-m))}\right) = \begin{cases} \frac{\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n-m))}{\widehat{f}_{\mathbf{y}}(y_{1}(n),y_{2}(n-m))}\right)^{\alpha-1} & \text{si} \quad \alpha \neq 0, \alpha \neq 1; \\ -\frac{1}{\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n-m))}{\widehat{f}_{\mathbf{y}}(y_{1}(n),y_{2}(n-m))}\right)} & \text{si} \quad \alpha = 0; \\ \log\left(\frac{\widehat{f}_{y_{1}}(y_{1}(n))\widehat{f}_{y_{2}}(y_{2}(n-m))}{\widehat{f}_{\mathbf{y}}(y_{1}(n),y_{2}-m(n))}\right) & \text{si} \quad \alpha = 1. \end{cases}$$

A.4

Calcul du gradient du critère (4.24)

On a

$$C(\mathbf{y}) = \left(R_{y_1}^{\alpha_1}(0) - 1\right)^2 + \left(R_{y_2}^{\alpha_2}(0) - 1\right)^2 + 2R_{y_{12}}^{\alpha_1}(0)^2 + 2R_{y_{21}}^{\alpha_2}(0)^2.$$

Nous cherchons à calculer le terme $\frac{\partial C(\mathbf{y})}{\partial b_{pq}}$. Pour simplifier, nous utilisons les notations

$$R_{y_i}^{\alpha_i} := R_{y_i}^{\alpha_i}(0)$$
 et $R_{y_{ij}}^{\alpha_i} := R_{y_{ij}}^{\alpha_i}(0), \ i = 1, 2$ et $j = 2$.

On peut écrire le critère $C(\mathbf{y})$ comme suit

$$C(\mathbf{y}) = (R_{y_1}^{\alpha_1} - 1) (R_{y_1}^{\alpha_1} - 1)^* + (R_{y_2}^{\alpha_2} - 1) (R_{y_2}^{\alpha_2} - 1)^* + 2R_{y_{12}}^{\alpha_1} (R_{y_{12}}^{\alpha_1})^* + 2R_{y_{21}}^{\alpha_2} (R_{y_{21}}^{\alpha_2})^*$$

$$= R_{y_1}^{\alpha_1} (R_{y_1}^{\alpha_1})^* - ((R_{y_1}^{\alpha_1})^* + R_{y_1}^{\alpha_1}) + 1 + R_{y_2}^{\alpha_2} (R_{y_2}^{\alpha_2})^* - ((R_{y_2}^{\alpha_2})^* + R_{y_2}^{\alpha_2}) + 1$$

$$+ 2R_{y_{12}}^{\alpha_1} (R_{y_{12}}^{\alpha_1})^* + 2R_{y_{21}}^{\alpha_2} (R_{y_{21}}^{\alpha_2})^*.$$

Nous avons alors le gradient de $C(\mathbf{y})$ donné par

$$\begin{split} \frac{\partial C}{\partial b_{pq}} &= \frac{\partial R_{y_1}^{\alpha_1} \left(R_{y_1}^{\alpha_1}\right)^*}{\partial b_{pq}} - \frac{\partial \left(\left(R_{y_1}^{\alpha_1}\right)^* + R_{y_1}^{\alpha_1}\right)}{\partial b_{pq}} + \frac{\partial R_{y_2}^{\alpha_2} \left(R_{y_2}^{\alpha_2}\right)^*}{\partial b_{pq}} - \frac{\partial \left(\left(R_{y_2}^{\alpha_2}\right)^* + R_{y_2}^{\alpha_2}\right)}{\partial b_{pq}} \\ &+ 2\frac{\partial R_{y_{12}}^{\alpha_1} \left(R_{y_{12}}^{\alpha_1}\right)^*}{\partial b_{pq}} + 2\frac{\partial R_{y_{21}}^{\alpha_2} \left(R_{y_{21}}^{\alpha_2}\right)^*}{\partial b_{pq}} \\ &= \frac{\partial |R_{y_1}^{\alpha_1}|^2}{\partial b_{pq}} - 2\frac{\partial Re\left(R_{y_1}^{\alpha_1}\right)}{\partial b_{pq}} + \frac{\partial |R_{y_2}^{\alpha_2}|^2}{\partial b_{pq}} - 2\frac{\partial Re\left(R_{y_2}^{\alpha_2}\right)}{\partial b_{pq}} + 2\frac{\partial |R_{y_{12}}^{\alpha_1}|^2}{\partial b_{pq}} + 2\frac{\partial |R_{y_{21}}^{\alpha_2}|^2}{\partial b_{pq}} \\ &= \frac{\partial \left(Re\left(R_{y_1}^{\alpha_1}\right)^2 + Im\left(R_{y_1}^{\alpha_1}\right)^2\right)}{\partial b_{pq}} - 2\frac{\partial Re\left(R_{y_1}^{\alpha_1}\right)}{\partial b_{pq}} + \frac{\partial \left(Re\left(R_{y_2}^{\alpha_2}\right)^2 + Im\left(R_{y_2}^{\alpha_2}\right)^2\right)}{\partial b_{pq}} \\ &- 2\frac{\partial Re\left(R_{y_2}^{\alpha_2}\right)}{\partial b_{pq}} + 2\frac{\partial \left(Re\left(R_{y_{11}}^{\alpha_1}\right)^2 + Im\left(R_{y_{12}}^{\alpha_1}\right)^2\right)}{\partial b_{pq}} + 2\frac{\partial \left(Re\left(R_{y_{21}}^{\alpha_2}\right)^2 + Im\left(R_{y_{21}}^{\alpha_2}\right)^2\right)}{\partial b_{pq}} , \end{split}$$

ce qui donne

$$\frac{\partial C}{\partial b_{pq}} = 2Re\left(R_{y_{1}}^{\alpha_{1}}\right)\frac{\partial\left(Re\left(R_{y_{1}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} + 2Im\left(R_{y_{1}}^{\alpha_{1}}\right)\frac{\partial\left(Im\left(R_{y_{1}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} - 2\frac{\partial\left(Re\left(R_{y_{1}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} + 2Re\left(R_{y_{2}}^{\alpha_{2}}\right)\frac{\partial\left(Re\left(R_{y_{2}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}} + 2Im\left(R_{y_{2}}^{\alpha_{2}}\right)\frac{\partial\left(Im\left(R_{y_{2}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}} - 2\frac{\partial\left(Re\left(R_{y_{2}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}} + 4Re\left(R_{y_{12}}^{\alpha_{1}}\right)\frac{\partial\left(Re\left(R_{y_{21}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} + 4Im\left(R_{y_{12}}^{\alpha_{1}}\right)\frac{\partial\left(Im\left(R_{y_{12}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} + 4Re\left(R_{y_{21}}^{\alpha_{2}}\right)\frac{\partial\left(Re\left(R_{y_{21}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}} + 4Im\left(R_{y_{21}}^{\alpha_{2}}\right)\frac{\partial\left(Im\left(R_{y_{21}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}},$$

 et

$$\frac{\partial C}{\partial b_{pq}} = 2 \frac{\partial \left(Re\left(R_{y_{1}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} \left[Re\left(R_{y_{1}}^{\alpha_{1}}\right) - 1\right] + 2Im\left(R_{y_{1}}^{\alpha_{1}}\right) \frac{\partial \left(Im\left(R_{y_{1}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}}
+ 2 \frac{\partial \left(Re\left(R_{y_{2}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}} \left[Re\left(R_{y_{2}}^{\alpha_{2}}\right) - 1\right] + 2Im\left(R_{y_{2}}^{\alpha_{2}}\right) \frac{\partial \left(Im\left(R_{y_{2}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}}
+ 4Re\left(R_{y_{12}}^{\alpha_{1}}\right) \frac{\partial \left(Re\left(R_{y_{12}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} Re\left(R_{y_{12}}^{\alpha_{1}}\right) + 4Im\left(R_{y_{12}}^{\alpha_{1}}\right) \frac{\partial \left(Im\left(R_{y_{12}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}}
+ 4Re\left(R_{y_{21}}^{\alpha_{2}}\right) \frac{\partial \left(Re\left(R_{y_{21}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}} Re\left(R_{y_{21}}^{\alpha_{2}}\right) + 4Im\left(R_{y_{21}}^{\alpha_{2}}\right) \frac{\partial \left(Im\left(R_{y_{21}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}},$$

 soit

$$\frac{\partial C(\mathbf{y})}{\partial b_{pq}} = 2 \frac{\partial \left(Re\left(R_{y_{1}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} \left[Re\left(R_{y_{1}}^{\alpha_{1}}\right) - 1\right] + 2Im\left(R_{y_{1}}^{\alpha_{1}}\right) \frac{\partial \left(Im\left(R_{y_{1}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}}
+ 2 \frac{\partial \left(Re\left(R_{y_{2}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}} \left[Re\left(R_{y_{2}}^{\alpha_{2}}\right) - 1\right] + 2Im\left(R_{y_{2}}^{\alpha_{2}}\right) \frac{\partial \left(Im\left(R_{y_{2}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}}
+ 4Re\left(R_{y_{12}}^{\alpha_{1}}\right) \frac{\partial \left(Re\left(R_{y_{12}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}} Re\left(R_{y_{12}}^{\alpha_{1}}\right) + 4Im\left(R_{y_{12}}^{\alpha_{1}}\right) \frac{\partial \left(Im\left(R_{y_{12}}^{\alpha_{1}}\right)\right)}{\partial b_{pq}}
+ 4Re\left(R_{y_{21}}^{\alpha_{2}}\right) \frac{\partial \left(Re\left(R_{y_{21}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}} Re\left(R_{y_{21}}^{\alpha_{2}}\right) + 4Im\left(R_{y_{21}}^{\alpha_{2}}\right) \frac{\partial \left(Im\left(R_{y_{21}}^{\alpha_{2}}\right)\right)}{\partial b_{pq}}.$$

Nous prenons en compte les relations suivantes

$$y_1 = b_{11}x_1 + b_{12}x_2,$$

$$y_2 = b_{21}x_1 + b_{22}x_2,$$

nous avons

$$\begin{split} R_{y_1}^{\alpha_1}(0) &= b_{11}^2 R_{x_1}^{\alpha_1}(0) + b_{12}^2 R_{x_2}^{\alpha_1}(0) + 2b_{11}b_{12}R_{x_{12}}^{\alpha_1}(0), \\ R_{y_2}^{\alpha_2}(0) &= b_{21}^2 R_{x_1}^{\alpha_2}(0) + b_{22}^2 R_{x_2}^{\alpha_2}(0) + 2b_{21}b_{22}R_{x_{21}}^{\alpha_2}(0), \\ R_{y_{12}}^{\alpha_1}(0) &= b_{11}R_{x_1}^{\alpha_1}(0)b_{21} + b_{11}R_{x_{12}}^{\alpha_1}(0)b_{22} + b_{12}R_{x_{21}}^{\alpha_1}(0)b_{21} + b_{12}R_{x_2}^{\alpha_2}(0)b_{22}, \\ R_{y_{21}}^{\alpha_2}(0) &= b_{21}R_{x_1}^{\alpha_2}(0)b_{11} + b_{21}R_{x_{12}}^{\alpha_2}(0)b_{12} + b_{22}R_{x_{21}}^{\alpha_2}(0)b_{11} + b_{22}R_{x_2}^{\alpha_2}(0)b_{12}. \end{split}$$

Enfin, nous obtenons les expressions suivantes

$$\begin{aligned} \frac{\partial C(\mathbf{y})}{\partial b_{11}} &= 4[b_{11}R_e(R_{x_1}^{\alpha_1}) + b_{12}R_e(R_{x_{12}}^{\alpha_1})][R_e(R_{y_1}^{\alpha_1}) - 1] + 4[b_{11}I_m(R_{x_1}^{\alpha_1}) + b_{12}I_m(R_{x_{12}}^{\alpha_1})]I_m(R_{y_1}^{\alpha_1}) \\ &+ 8[b_{21}R_e(R_{x_1}^{\alpha_1}) + b_{22}R_e(R_{x_{12}}^{\alpha_1})]R_e(R_{y_{12}}^{\alpha_1}) + 8[b_{21}I_m(R_{x_1}^{\alpha_1}) + b_{22}I_m(R_{x_{12}}^{\alpha_1})]I_m(R_{y_{12}}^{\alpha_1}) \\ &+ 8[b_{21}R_e(R_{x_1}^{\alpha_2}) + b_{22}R_e(R_{x_{12}}^{\alpha_2})]R_e(R_{y_{21}}^{\alpha_2}) + 8[b_{21}I_m(R_{x_1}^{\alpha_2}) + b_{22}I_m(R_{x_{12}}^{\alpha_2})]I_m(R_{y_{21}}^{\alpha_2}),\end{aligned}$$

$$\frac{\partial C(\mathbf{y})}{\partial b_{12}} = 4[b_{12}R_e(R_{x_2}^{\alpha_1}) + b_{11}R_e(R_{x_{12}}^{\alpha_1})][R_e(R_{y_1}^{\alpha_1}) - 1] + 4[b_{12}I_m(R_{x_2}^{\alpha_1}) + b_{11}I_m(R_{x_{12}}^{\alpha_1})]I_m(R_{y_1}^{\alpha_1})
+ 8[b_{21}R_e(R_{x_{21}}^{\alpha_1}) + b_{22}R_e(R_{x_2}^{\alpha_1})]R_e(R_{y_{12}}^{\alpha_1}) + 8[b_{21}I_m(R_{x_{21}}^{\alpha_1}) + b_{22}I_m(R_{x_2}^{\alpha_1})]I_m(R_{y_{12}}^{\alpha_1})
+ 8[b_{21}R_e(R_{x_{12}}^{\alpha_2}) + b_{22}R_e(R_{x_2}^{\alpha_2})]R_e(R_{y_{21}}^{\alpha_2}) + 8[b_{21}I_m(R_{x_{12}}^{\alpha_2}) + b_{22}I_m(R_{x_2}^{\alpha_2})]I_m(R_{y_{21}}^{\alpha_2}),$$

$$\begin{aligned} \frac{\partial C(\mathbf{y})}{\partial b_{21}} &= 4[b_{21}R_e(R_{x_1}^{\alpha_2}) + b_{22}R_e(R_{x_{12}}^{\alpha_2})][R_e(R_{y_2}^{\alpha_2}) - 1] + 4[b_{21}I_m(R_{x_1}^{\alpha_2}) + b_{22}I_m(R_{x_{12}}^{\alpha_1})]I_m(R_{y_2}^{\alpha_2}) \\ &+ 8[b_{11}R_e(R_{x_1}^{\alpha_1}) + b_{12}R_e(R_{x_{21}}^{\alpha_1})]R_e(R_{y_{12}}^{\alpha_1}) + 8[b_{11}I_m(R_{x_1}^{\alpha_1}) + b_{12}I_m(R_{x_{21}}^{\alpha_1})]I_m(R_{y_{12}}^{\alpha_1}) \\ &+ 8[b_{11}R_e(R_{x_1}^{\alpha_2}) + b_{12}R_e(R_{x_{12}}^{\alpha_2})]R_e(R_{y_{21}}^{\alpha_2}) + 8[b_{11}I_m(R_{x_1}^{\alpha_1}) + b_{12}I_m(R_{x_{12}}^{\alpha_2})]I_m(R_{y_{21}}^{\alpha_2}),\end{aligned}$$

 et

$$\begin{aligned} \frac{\partial C(\mathbf{y})}{\partial b_{22}} &= 4[b_{22}R_e(R_{x_2}^{\alpha_2}) + b_{21}R_e(R_{x_{12}}^{\alpha_2})][R_e(R_{y_2}^{\alpha_2}) - 1] + 4[b_{22}I_m(R_{x_2}^{\alpha_2}) + b_{21}I_m(R_{x_{12}}^{\alpha_2})]I_m(R_{y_2}^{\alpha_2}) \\ &+ 8[b_{11}R_e(R_{x_{12}}^{\alpha_1}) + b_{12}R_e(R_{x_2}^{\alpha_1})]R_e(R_{y_{12}}^{\alpha_1}) + 8[b_{11}I_m(R_{x_{12}}^{\alpha_1}) + b_{12}I_m(R_{x_2}^{\alpha_1})]I_m(R_{y_{12}}^{\alpha_1}) \\ &+ 8[b_{11}R_e(R_{x_{21}}^{\alpha_2}) + b_{12}R_e(R_{x_2}^{\alpha_2})]R_e(R_{y_{21}}^{\alpha_2}) + 8[b_{11}I_m(R_{x_{21}}^{\alpha_2}) + b_{12}I_m(R_{x_2}^{\alpha_2})]I_m(R_{y_{21}}^{\alpha_2}).\end{aligned}$$

A.5

Algorithme de Bouguerriou et al. [23]

Cet algorithme est appliqué pour extraire un signal cyclostationnaire mélangé avec d'autres signaux stationnaires. Il se constitue de deux étapes ; l'étape de blanchiment des signaux mélanges \mathbf{x} . Cette dernière permet de décorréler les signaux mesures pour obtenir les signaux mélanges blanchis \mathbf{z} . Il reste donc, par la suite, une étape de rotation des signaux blanchis afin de restituer les sources inconnues au départ. Nous résumons les différentes étapes décrivant cette méthode de la sorte :

- 1. Blanchiment des signaux mélanges.
- 2. Estimation des différentes autocorrélations et intercorrélations des mélanges blanchis pour un décalage en temps $\tau = 0$.
- 3. Estimation des coefficients du critère sur une durée T_i , la période cyclique fondamentale de la source cyclostationnaire d'intérêt.
- 4. Détermination du critère C_i en fonction de l'angle de rotation θ .
- 5. Recherche de l'argument du maximum du critère : $\hat{\theta} = \arg \max(C_i(\theta))$.
- 6. Application de la rotation optimale aux mélanges blanchis.
- 7. Restitution de la source cyclostationnaire.

Algorithme SOBI [15]

L'algorithme SOBI est basé sur des statistiques d'ordre deux. Sous l'hypothèse de signaux sources colorés, de spectres différents. Le principe de l'algorithme SOBI est résumé de la sorte :

- 1. Calculer la matrice de covariance $R_{\mathbf{x}}(\tau)$ des observations.
- 2. Estimer la matrice de blanchiment **W**.
- 3. Blanchir les observations : $\mathbf{z}(n) = \mathbf{W}\mathbf{x}(n)$.
- 4. Estimer la matrice \mathbf{V} par la diagonalisation conjointe d'un ensemble de ma-

trices issues de la fonction de corrélation des observations blanchies à différents retard $\tau \neq 0$.

- 5. Estimer la matrice de mélange par $\mathbf{A} = \mathbf{W}^{\sharp}\mathbf{V}$ où le symbole \sharp dénote la matrice pseudo-inverse.
- 6. Estimer les signaux sources par $\mathbf{s}(n) = \mathbf{A}^H R_{\mathbf{z}}(0)^{-1} \mathbf{x}(n)$.

Algorithme CYCLOSOBI de Boustany et Antoni [27]

L'algorithme CYCLOSOBI (SOBI pour les signaux cyclostationnaires) représente une extension de la méthode SOBI [15]. Il utilise une méthode de diagonalisation conjointe des matrices de covariance cyclique. Cet algorithme utilise seulement des statistiques à l'ordre deux. Le principe de celui-ci est comme suit :

- 1. Estimer de matrice de blanchiment **W**.
- 2. Blanchir les observations : $\mathbf{z}(n) = \mathbf{W}\mathbf{x}(n)$.
- 3. Estimer la matrice unitaire **V** par la diagonalisation conjointe d'un ensemble de matrices hermitiennes $R_{\mathbf{z}}^{\alpha_i}(\tau)R_{\mathbf{z}}^{\alpha_i}(\tau)^H$ sur des ensembles de fréquences cycliques α_i et/ou les retards τ .

Algorithme JADE [42]

L'algorithme JADE utilise un critère à base des cumulants d'ordre quatre. Son principe est résumé de la sorte :

- 1. Constituer la matrice de covariance $R_{\mathbf{x}}$ et calculer une matrice de blanchiment \mathbf{W} .
- 2. Constituer les cumulants d'ordre quatre $Cum_4(\mathbf{z})$ du processus blanchi $\mathbf{z} = \mathbf{W}\mathbf{x} = \mathbf{U}\mathbf{s} + \mathbf{W}\mathbf{b}$, (\mathbf{U} , \mathbf{W} et \mathbf{b} sont respectivement matrice unitaire, matrice de blanchiment et bruit blanc), et calculer l'ensemble de ses p plus significatives (ordre croissant en valeur absolu) valeurs et matrices propres { λ_r, M_r }; $1 \leq r \leq p$, (p est le nombre des sources).
3. Diagonaliser conjointement l'ensemble des matrices $\lambda_r M_r$ par une matrice unitaire $\widehat{\mathbf{U}}$, où la diagonalisation a été définie par maximisation du critère suivant :

$$C(\mathbf{U}, \mathcal{N}) = \sum_{r=1}^{p} |\text{diag}(\mathbf{U}^* N_r \mathbf{U})|^2$$

où $|\text{diag}(\cdot)|$ est la norme du vecteur composé des éléments diagonaux de la matrice argument et $\mathcal{N} = \{N_r | 1 \leq r \leq p\}$ est l'ensemble à diagonaliser.

4. Estimer la matrice de séparation $\mathbf{A} = \mathbf{W}^{\sharp} \mathbf{U}$ où le symbole \sharp dénote la matrice pseudo-inverse de Moore-Penrose.

L'auteur utilise la technique de Jacobi étendue pour maximiser son critère de diagonalisation conjointe.

Abréviations

- EQM : Erreur Quadratique Moyenne
- RSB : Rapport Signal sur Bruit
- SAS : Séparation Aveugle de Sources
- RIF : Réponse Impulsionnelle Finie
- ACI : Analyse en Composantes Indépendantes
- IM: Information Mutuelle
- AMUSE : Algorithm for Multiple Unknown Signals Extraction
- FOBIUM : Fourth Order Blind Identification of Underdetermined Mixtures of sources
- JADE : Joint Approximate Diagonalization of Eigen matrices
- SOBI : Second Order Blind Identification
- SONS : Second Order Non-stationary source Separation
- CYCLOSOBI : Cyclostationary Second Order Blind Identification
- AIC : Akaike Information Criterion (Critère d'Information d'Akaike)
- FSM : Fonction Score Marginale
- FSJ : Fonction Score Jointe
- FSD : Fonction Score Différentielle
- SO2 : Statistique d'Ordre Deux
- SOS : Statistique d'Ordre Supérieur

Notations

- SNR: Rapport signal sur résidus
- $\widehat{\mathbb{E}}$: Moyenne temporelle
- \mathbb{E} : Espérance mathématique
- A : Matrice de mélange
- **B** : Matrice séparante ou matrice de séparation
- I : Matrice identité
- \mathbf{W} : Matrice de blanchiment
- \mathbf{s} : Vecteur de sources
- \mathbf{x} : Vecteur des observations
- \mathbf{z} : Vecteur des observations blanchies
- y : Vecteur de sources estimées
- $R_x(\cdot)$: Autocorrélation du signal x(t)
- $R_x^{\alpha}(\cdot)$: Autocorrélation cyclique du signal x(t)
- $R^{\alpha}_{x_{12}}(\cdot)$: Intercorrélation cyclique entre les signaux $x_1(t)$ et $x_2(t)$
- p: Nombre de sources
- N: Taille du signal
- τ : Décalage temporel
- $|\cdot|$: Module ou valeur absolue

$Re(\cdot)$: Partie réelle

- $Im(\cdot)$: Partie imaginaire
- Cum_4 : Cumulant d'ordre 4
- $K {\cal L}_m\,$: Divergence de Kullback-Leibler modifiée
- $KL\,$: Divergence de Kullback-Leibler
- ${\cal H}\,$: Divergence de Hellinger

Publications

- New blind separation criterion for second order cyclostationary sources, M. Ould Mohamed, A. Keziou, H. Fenniri et G. Delaunay, Arima-Office, INRIA, Revue ARIMA, 12 :1-14, 2010.
- Séparation aveugle de sources par minimisation des alpha/divergences,
 A. Keziou, H. Fenniri, M. Ould Mohamed et G. Delaunay, XXII Colloque
 GRETSI 2009, 8-11 septembre 2009, Dijon France, septembre 2009.
- 3. Séparation aveugle de sources par minimisation des alpha-divergences : application au diagnostic et à la surveillance des machines tournantes, M. Ould Mohamed, A. Keziou et H. Fenniri, Second International Conference on Industrial Risk Engineering (CIRI 2009), Reims, France 13-15 mai 2009, 2009.
- 4. A new criterion for second order cyclostationary blind source separation, A. Keziou, M. Ould Mohamed, H. Fenniri et G. Delaunay, 2nd International Conference on Signals, Circuits and Systems (SCS'08) IEEE, Hammamet Tunisia, novembre 2008.
- 5. Séparation aveugle de sources par optimisation de l'information mutuelle sous contraintes, estimation et selection de modèles, Ould Mohamed, A. Keziou, H. Fenniri et G. Delaunay, 9th African Conference on Research in Computer Science and Applied Mathematics (CARI'2008), Proceedings CARI'2008, Rabat, Morroco, octobre 2008.
- 6. Séparation de sources et caractérisation de processus cyclostation-

naires, M. Ould Mohamed, 1^{ere} journée des doctorants du CReSTIC, Juin 2008, membre du comité d'organisation et participant, Reims, France.

7. Séparation aveugle de sources cyclostationnaires par utilisation des statistiques de second ordre, M. Ould Mohamed, A. Keziou, H. Fenniri et G. Delaunay, MAnifestation des JEunes Chercheurs en Sciences et Technologies de l'Information et de la Communication (MajecSTIC 2007), fasicule N°2, pp 191-198. Schedae, Prépublications de l'Université de Caen Basse Normandie, Caen, France, octobre 2007.

Liste des tableaux

2.1	Exemples de fonctions convexes et leurs divergences associées 63
2.2	SNR en dB pour Sobi, Jade, KL_m , H et KL
4.1	SNR en dB pour <i>Sobi</i> , <i>Cyclosobi</i> , <i>Jade</i> , C_1 , C_2 et Γ
4.2	SNR en dB pour Sobi, Jade, C_1, C_2, C_3 et Γ
4.3	SNR en dB pour le critère Γ (4.14) et celui utilisant la divergence de
	Hellinger pénalisée $IP_{\varphi_{0.5}}$

Table des figures

1.1	Procédés de séparation de sources	11
1.2	Schéma de propagation pour un mélange instantané	12
1.3	Schéma de propagation pour un mélange convolutif	18
1.4	Modèle de séparation par bloc pour un mélange linéaire instantané	20
1.5	Modèle de séparation adaptative	21
1.6	Structure de séparation directe (dans le cas où $p = 2$ et $r = 2$)	23
1.7	Structure de séparation récurrente $(p = 2 \text{ et } r = 2)$	24
2.1	Rapport signal sur résidus (SNR) de notre algorithme comparé avec	
	celui de Pham en fonction du nombre d'itérations (SNR1 : source 1 ;	
	SNR2 : source 2)	56
2.2	Rapport signal sur résidus (SNR) de notre algorithme comparé avec	
	celui de Babaie-Zadeh en fonction du nombre d'itérations (SNR1 :	
	source 1; SNR2 : source 2). \ldots	57
2.3	De gauche à droite, l'espace des signaux sources, signaux mélanges et	
	sources estimées.	58
2.4	Les SNRs de trois sources données par l'algorithme de l'IM sous	
	contraintes.	59
2.5	De haut en bas, les signaux sources, les mélanges, les sources estimées	
	et l'écart entre les sources et leurs estimées.	60

2.6	Les SNRs de trois sources données par l'algorithme de l'IM sous	
	contraintes.	61
2.7	De haut en bas, les signaux sources, les mélanges, les sources estimées	
	et l'écart entre les sources et leurs estimées	62
2.8	Les fonctions convexes φ_{α} des divergences KL , KL_m et H	64
2.9	Influence de la largeur du noyau gaussien pour une distribution gaus-	
	sienne	69
2.10	Courbes de l'angle θ suivant le nombre d'iterations pour les critères	
	$KL_m, H $ et KL	72
2.11	Approche 2. De gauche à droite et de haut en bas, l'espace des signaux	
	sources, signaux mélanges, signaux blanchis et sources estimées avec	
	le critère basé sur la divergence de Hellinger	79
2.12	Approche 1. En haut : SNR des trois critères : IM ($\alpha = 0$), H ($\alpha =$	
	0.5) et $KL \ (\alpha = 1)$. En fonction du RSB - en bas (de gauche à droite) :	
	zoom sur SNR (autour de -55 dB , -39 dB et -20 dB)	80
2.13	SNR en fonction du nombre d'itérations pour KL_m , H et KL (cas	
	non bruité)	81
2.14	Approche 2. De gauche à droite, l'espace des signaux sources, signaux	
	mélanges et sources estimées par le critère utilisant la divergence de	
	Hellinger.	81
2.15	Approche 2. En haut : SNR des trois critères : IM ($\alpha = 0$), H ($\alpha =$	
	0.5) et KL ($\alpha = 1$), en fonction du RSB. En bas (de gauche à droite) :	
	zoom sur SNR (autour de -52 dB , -39 dB et -20 dB)	82
3.1	Qualité de la séparation des trois critères : KL_m ($\alpha = 0$), Hellinger	
	$(\alpha$ = 0.5) et KL $(\alpha$ = 1), en fonction du nombre d'itérations (cas	
	$\operatorname{convolutif}). \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots $	89

TABLE DES FIGURES

3.2	Qualité de la séparation des trois critères : KL_m ($\alpha = 0$), Hellinger	
	$(\alpha = 0.5)$ et KL $(\alpha = 1)$, en fonction du RSB (cas convolutif)	91
3.3	Le banc de mesure	92
3.4	DSP des signaux sources pris séparément.	93
3.5	DSP des signaux mélanges	94
3.6	DSP des signaux estimés	94
4.1	Fonctions coût Γ , C_1 , C_2 et Erreur Quadratique Moyenne	108
4.2	Les signaux sources (en haut), les mélanges (au milieu) et les sources	
	estimées (en bas) obtenus par maximisation de Γ	109
4.3	Les signaux sources (en haut), les mélanges (au milieu) et les sources	
	estimées (en bas) obtenus par maximisation de Γ	112
4.4	SNRs moyens en fonction du nombre d'itérations pour les trois critères	
	IM, Hellinger et KL	114
4.5	SNR moyen en fonction du nombre d'itérations pour le critère C	117
4.6	SNRs moyens en fonction des itérations pour les trois critères \widehat{IP}_{φ_0} ,	
	$\widehat{IP}_{\varphi_{0.5}}$ et \widehat{IP}_{φ_1}	120
4.7	Les signaux sources (en haut), les mélanges (au milieu) et les sources	
	estimées (en bas) par la méthode basée sur la divergence de Hellinger	
	pénalisée $(\widehat{IP}_{\varphi_{0.5}})$	121
4.8	En haut : SNR des trois critères : $\widehat{IP}_{\varphi_0}, \widehat{IP}_{\varphi_{0.5}}$ et \widehat{IP}_{φ_1} , en fonction	
	du RSB. En bas (de gauche à droite) : zoom sur SNR (autour de -50	
	dB et autour de -20 dB)	123

Bibliographie

- K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua. Blind source separation using second-order cyclostationary statistics. *IEEE Trans. Signal* processing, 49(4):694–701, 2001.
- [2] S. Achard, D-T. Pham, and C. Jutten. Criteria based on mutual information minimization for blind source separation in post nonlinear mixtures. *Signal Processing*, 85 :965–974, 2005.
- [3] 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.
- [4] L.B. Almeida. Ica of linear and nonlinear mixtures based on mutual information. in International Joint Conference on Neural Networks, Washington, DC, USA, July 2001.
- [5] S. I. Amari. Neural learning in structured parameter spaces-natural riemannian gradient. Neural Information Processing System Natural and Synthetic, Colorado, USA, pages 127–133, December 1996.
- [6] J. Antoni. Cyclic spectral analysis of rolling-element bearing signals : facts and fictions. Journal of Soundard Vibration, 304(3-5) :497–529, 2007.

- [7] J. Antoni, J. Danière, F. Guillet, and R.B. Randall. Effective vibration analysis of ic engines using cyclostationarity. partii : new results on the reconstruction of the cylinder pressure. *Journal of Soundard Vibration*, 257(5) :839–856, 2002.
- [8] M. Babaie-Zadeh. mutual information minimization. Ph.D INPG, France et Sharif university of technology, Iran, 2002.
- M. Babaie-Zadeh and C. Jutten. A general approach for mutual information minimization and its application to blind source separation. *Signal Processing*, 85 :975–995, 2005.
- [10] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. Separating convolutive mixtures by mutual information minimization. in Proceedings of IWANN'2001, Granada, Spain, pages 834–842, 2001.
- [11] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. Using multivariate score functions in source separation : Application to post non-linear mixtures. *Scientia-Iranica*, 9(4) :409–418, 2002.
- [12] A. D. Back and A. C. Tsoi. Blind deconvolution of signals using a complex recurrent network. Ds Actes Intl. Work. on Neural Networks for Signal Processing (NNSP'94), Ermioni, grèce, pages 565–574, 1994.
- [13] A.K. Barros and N. Ohnishi. Removal of quasi-periodic sources from physiological measurements. in ICA'99, Aussois, France, pages 11–15, January 1999.
- [14] A.K. Barros, R. Vigärio, V. Jousmäki, and N. Ohnishi. Extraction of eventrelated signals from multichannel bioelectrical measurements. *IEEE Transactions* on Biomedical Engineering, 47(5):583–588, May 2000.

- [15] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind separation technique using second-order statistics. *IEEE Trans. Signal pro*cessing, 45(2):434–444, 1997.
- [16] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation for discrete sources. *in Proc. EUSIPCO*, pages 768–771, 1994.
- [17] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation by the expectation-maximization technique : deterministic and stochastic implementation. in Proc. International Symposium on Non-Linear Theory and Applications NOLTA, Las Vegas, NV, USA, pages 49–53, 1995.
- [18] W.R. Bennett. Statistics of regenerative digital transmission. Bell System Technical Journal, 37 :1501–1542, 1958.
- [19] R. Beran. Minimum Hellinger distance estimates for parametric models. Ann. Statist., 5(3):445–463, 1977.
- [20] F. Berthommier and S. Choi. Several improvements of the herault-jutten model for speech segregation. Ds Actes Intl. Conf. (ICA'03), nara, Japon, pages 1089–1094, 2003.
- [21] W. Bobillet, E. Grivel, R. Guidorzi, and M. Najim. Cancelling convolutive and additive coloured noises for speech enhancement. In Proceedings of ICASSP 2004, Montreal, Canada, May 17-21, 2004.
- [22] N. Bouguerriou, C. Capdessus, and A. K. Nandi. New source separation method for cyclostationary sources and its application to roller bearing vibrations. *ICA Research Network workshop, Liverpool, UK*, 2006.
- [23] N. Bouguerriou, M. Haritopoulos, C. Capdessus, and L. Allam. Novel cyclostationarity-based blind source separation algorithm using second order

statistical properties : Theory and application to the bearing defect diagnosis. Mechanical Systems and Signal Processing, 19(6) :1260–1281, 2005.

- [24] H. Boumaraf. Séparation aveugle de mélanges convolutifs de sources. Thèse de l'Université Joseph Fourier, INPG, France, 2005.
- [25] H. Boumaraf, D.-T. Pham, and C. Servière. Blind separation of convolutive mixture of speech signals. *Proceeding of the EUSIPCO 2005 Conference. Antalya, Turkey*, September 2005.
- [26] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Maraim. Jacobi-like algorithme for blind signal separation of convolutive mixtures. *Electronic Lettres*, 37 :1049–1050, 2001.
- [27] R. Boustany and J. Antoni. Separation of rotating machines vibration signals. Twelfth International Congress on Sound and Vibration ICSV, Lisbon, Turkey, 2005.
- [28] R. Boustany and J. Antoni. A subspace method for the blind extraction of a cyclostationary source : Application to rolling element bearing diagnostics. *Mechanical Systems and Signal Processing*, 19(6) :1245–1259, 2005.
- [29] R. Boustany and J. Antoni. Blind extraction of a cyclostationary signal using reduced-rank cyclic regression a unifying approach. *Mechanical Systems and Signal Processing*, 22(3) :520–541, 2008.
- [30] R. Boustany and J. Antoni. Une méthode de sous espace pour l'extraction aveugle d'une source cyclostationnaire. 20ème Colloque GRETSI sur le Traitement du Signal et des Images, Louvain-la-Neuve, Belgique, Septembre 2005.
- [31] R. Boyles and W.A. Gardner. Cycloergodic properties of discrete-parameter

nonstationary stochastic processes. *IEEE Transactions on Information Theory*, 29(1) :105–114, 1983.

- [32] I. Bradaric, A. P. Petropulu, and K. I. Diamantaras. On resolving the column permutation ambiguity in the estimates of mimo system response. 2000 conference on Information Sciences and Systems, 2000.
- [33] M. Broniatowski and A. Keziou. Minimization of φ-divergences on sets of signed measures. Studia Sci. Math. Hungar., 43(4):403–442, 2006.
- [34] M. Broniatowski and A. Keziou. Parametric estimation and tests through divergences and the duality technique. J. Multivariate Anal., 100(1) :16–36, 2009.
- [35] H. Buchner, A Aichner, and W. Hellermann. A generalization of blind source separation algorithm for convolutive mextures based on second-order statistics. *IEEE Transactions on Speech and Audio Processing*, 13(1):120–134, 2005.
- [36] J. Campbell, A. Gibbs, and B. Smith. The cyclostationary nature of crosstalk interference from digital signals in multipair cable-part i : Fundamentals. *IEEE Transactions on Communications*, 31(5) :629–637, May 1983.
- [37] C. Capdessus, M. Sidahmed, and J.-L. Lacoume. Cyclostationary processes :application in gear faults early diagnosis. *Mechanical Systems and Signal Processing*, 14(3):371–385, 2000.
- [38] V. Capdevielle. Séparation de sources large bande à l'aide des moments d'ordre supérieur. PhD thesis, INP Grenoble, France, 1996.
- [39] V. Capdevielle, C. Servière, and J.-L. Lacoume. Blind separation of wideband sources : application to rotating machine signals. In Proceedings of Eusipco'96, Triestle, Italy, pages 2085–2088, 1996.

- [40] J.-F. Cardoso. Source separation using higher order moments. in Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP, Glasgow, Royaume-Uni, 4 :2109–2112, 1989.
- [41] J.-F. Cardoso. Infomax and maximum likelihood for blind source separation. IEEE Letters on Signal processing, 4(4) :112–114, April 1997.
- [42] J.-F. Cardoso and A. Souloumiac. Blind signal beamforming for non gaussian signals. *Proceedings of the IEEE*, 140(6) :362–370, 1993.
- [43] N. Charkani. Séparation auto-adaptative de sources pour les mélanges convolutifs. application à la téléphonie mains-libres dans les voiture. *PhD thesis*, *INP Grenoble, France*, 1996.
- [44] N. Charkani and Y. Deville. Self-adaptive separation of convolutively mixed signals with a recursive structure. part ii : Theoretical extensions and application to synthetic and real signals. *Signal Processing*, 75(2) :117–140, juin 1999.
- [45] E. Chaumette, P. Comon, and D. Muller. An ica based technique for radiating sources estimation; application to airport surveillance. *IEE Proceedings part F. Special issue on Applications of High-Order Statistics*, 140(6) :395–401, 1993.
- [46] S. Choi and A. Cichocki. A hybrid learning approach to blind deconvolution of linear mimo systems. *Electronic Letters*, 35(7) :1429–1430, Août 1999.
- [47] S. Choi, A. Cichocki, and A. Belouchrani. Second order non-stationary source separation. *journal of VLSI Signal Processing*, 32(1-2):93–104, 2002.
- [48] P. Comon. Separation of sources using higher-order cumulants. In Proceedings of SPIE, pages 170–181, 1989.

- [49] P. Comon. Independent component analysis. Proc. Int. Workshop on Higher-Order Statistics, Chamrousse, France, pages 111–120, 1991.
- [50] P. Comon. Independent component analysis, a new concept? Signal Processing, IEEE, 36(3) :287–314, 1994.
- [51] P. Comon. From source separation to blind equalization, contrast-based approaches. in ICISP 01, Int. Conf. on Image and Signal Processing, Agadir, Morocco, pages 20–32, 2001.
- [52] P. Comon. Blind identification and source separation in 2 × 3 underdetermined mixtures. *IEEE Transactions on Signal Processing, Agadir, Mo*rocco, 52(1) :11–22, 2004.
- [53] P. Comon and E. Moreau. Improved contrast dedicated to blind separation in communications. *in Proc. ICASSP, Munich, Germany*, pages 3453–3456, 1997.
- [54] N. Cressie and T. R. C. Read. Multinomial goodness-of-fit tests. J. Roy. Statist. Soc. Ser. B, 46(3) :440–464, 1984.
- [55] I. Csiszár. Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten. Magyar Tud. Akad. Mat. Kutato Int. Kozl, 8 :85–108, 1963.
- [56] A.V. Dandawate and G.B. Giannakis. Modeling (almost) periodic moving average processes using cyclic statistics. *IEEE Transactions on Signal Processing*, 44(3):673–684, March 1996.
- [57] A.V. Dandawate and G.B. Giannakis. Statistical tests for presence of cyclostationarity. *IEEE Transactions on Signal Processing*, 42(9) :2355–2369, September 1994.

- [58] L. De Lathauwer, D. Callaerts, B. De Moor, and J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation,. *Girona, Spain*, pages 134–138, 12-14 June 1995.
- [59] L. De Lathauwer, B. De Moor, and J. Vandewalle. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Transactions on Biomedical Engineering*, 47(5):567–572, 2000.
- [60] G. Desodt and D. Muller. Complex independent components analysis applied to the separation of radar signals. In L. Torres, E. Masgrau, and M. A. Lagunas, editors, Signal Processing V, Theories and Applications, Barcelona, Espain, pages 665–668, 1994.
- [61] S. Deville, Y. Hosseini. Recurrent networks for separating extractabletarget nonlinear mixtures.parti : Non-blind configurations. *Signal Processing*, 89 :378–393, 2009.
- [62] N. Dimitri. Méthodes parafac généralisées pour l'extraction aveugle de sources. application aux systèmes ds-cdma. Ph.D. dissertation, Université de Cergy-Pontoise, France, 2007.
- [63] A. Dinc and Y. Bar-Ness. Convergence and performance comparison of the three different structures of boostrap blind adaptive algorithm for multisignal co-channel separation. *In proc. MILCOM'92*, pages 913–918, 1992.
- [64] Y.P. Dragan and I.N. Yavorsky. Rythmics of sea waves and underwater acoustic signals. *Naukova Dumka, Kiev, Ukraine*, 1982 (in Russian).
- [65] G. d'Urso and L. Cai. Sources separation method applied to reactor monitoring. In Proc. Workshop Athos working group, Girona, Spain, June 1995.

- [66] M. El Rhabi, H. Fenniri, G. Gelle, and G. Delaunay. Blind separation of rotating machine signals using penalized mutual information criterion and minimal distortion principle. *Mechanical Systems and Signal Processing*, 19(6) :1282– 1292, 2005.
- [67] M. El Rhabi, G. Gelle, H. Fenniri, and G. Delaunay. A penalized mutual information criterion for blind separation of convolutive mixtures. *Signal Pro*cessing, 84 :1979–1984, 2004.
- [68] D. Erdogmus, K.E. Hild, and J. Principe. Blind source separation using renyi's mutual information. *IEEE Signal Processing Letters*, 8(6) :174–176, 2001.
- [69] A. Ferréol, L. Alberta, and P. Chevalier. Fourth order blind identification of underdetermined mixtures of sources (fobium). *IEEE Trans. Signal Proces*sing, 53(5) :1640–1653, May 2005.
- [70] A. Ferréol, P. Chevalier, and L. Alberta. Second and higher order blind separation of first and second order cylostationary sources - application to am, fsk, cpfsk and deterministic sources. *IEEE Trans. Signal Processing*, 52(4) :845– 861, April 2004.
- [71] C.J. Finelli and J.M. Jenkins. A cyclostationary least mean squares algorithm for discrimination of ventricular tachycardia from sinus rhythm. International Conference of the IEEE Engineering in Medicine and Biology Society, Baltimore, MD, 13:740–741, 1991.
- [72] L.E. Franks. Signal theory, prentice-hall, englewood cliffs, nj. 1969.
- [73] L. Féty. Méthodes de traitement d'antenne adaptées aux radiocommunications. Thèse de doctorat. Télécom, Paris, France, June 1988.

- [74] M. Gaeta, F. Briolle, and F. Mondoloni. Comparaison de méthodes de séparation de sources appliquées en acoustique sous-marine. Seizième Colloque GRETSI Grenoble, France, 15-19 Septembre 1997.
- [75] M. Gaeta and J.-L. Lacoume. Estimateurs du maximum de vraisemblance étendus à la séparation de sources non gaussiennes. Traitement du Signal, 7(5):419–434, 1990.
- [76] M. Gaeta and J.-L. Lacoume. Source separation without a priori knowledge : the maximum likelihood solution. In EUSIPCO, pages 621–624, 1990.
- [77] P. Gao, E.-C. Chang, and L. Wyse. Blind separation of fetal ecg from single mixture using svd and ica. *IEEE Proceedings*, 3 :1418–1422, 2003.
- [78] W. A. Gardner. Introduction to random processes with applications to signals and systems. *Macmillan, New York*, 1985.
- [79] W. A. Gardner. Statistical spectral analysis : A non-probabilistic theory. Prentice-Hall., Englewood Cliffs, NJ, 1988.
- [80] W. A. Gardner. Exploitation of the spectral redundancy in cyclostationary signals. *IEEE Mag. Signal Processing*, 8(2) :14–26, 1991.
- [81] W. A. Gardner and L. E. Franks. Characterization of cyclostationary random signal processes. *IEEE Trans. Information Theory*, 21(1):4–14, 1975.
- [82] W. A. Gardner and C. M. Spooner. The cumulant theory of cyclostatioary time-series, part i : Foundation. *IEEE Trans. Signal Processing*, 42(12) :3387– 3408, 1994.
- [83] G. Gelle. Les statistiques d'ordre supérieur appliquées à la détection et à la séparation de sources. Thèse de l'Université de Reims Champagne Ardenne, France, 1998.

- [84] G. Gelle, M. Colas, and G. Delaunay. Separation of convolutive mixtures of harmonic signals with a temporal approach. application to rotating machine monitoring. *In Proceedings of ICA*'99, 1999.
- [85] G. Gelle, M. Colas, and C. Servière. Blind source separation : A tool for rotating machine monitoring by vibration analysis. *Journal of Sound and Vibration*, 248(5) :865–885, 2001.
- [86] G.B. Giannakis. Cyclostationary signal analysis, in : V.k. madisetti, d. williams (eds.), statistical signal processing section of digital signal processing handbook. CRC Press, Boca Raton, FL, 1999 (Chapter 17).
- [87] E.G. Gladyshev. Periodically correlated random sequences. Soviet Math. Dokl., 2:385–388, 1961.
- [88] E.G. Gladyshev. Periodically and almost periodically correlated with continuous time parameter. Theory of Probability and Its Applications, 8 :173–177, 1963.
- [89] O. Grellier and P. Comon. Blind equalization and source separation with msk inputs. In Proceedings of SPIE Conf. Adv. Sig .Proc. VIII, San Diego, pages 26–34, 1998.
- [90] L.I. Gudzenko. On periodic nonstationary processes. Radio Eng. Electron. Phys., 4 :220–224 (in Russian), 1959.
- [91] P. G. Hoel. Introduction to mathematical statistics. John Wiley and Sons, 1984.
- [92] S. Hosseini and Y. Deville. Blind separation of nonstationary sources by spectral decorrelation. In Springer-Verlag (editor) : Proceedings of the Fifth In-

ternational Conference on Independent Component Analysis and blind Signal Separation, Granada, Spain, LNCS 3195 :279–286, 2004.

- [93] S. Hosseini, Y. Deville, and H. Saylani. From time-domain separation of stationary temporally correleted sources to frequency-domain separation of nonstationary sources. In Proceedings of the Fourth International Conference on Physics in Signal and Image Processing (PSIP 2005), Toulouse, France, pages 55–59, 2005.
- [94] J. Hérault and B. Ans. Réseaux neuronaux à synapses modifiables : Décodage de messages sensoriels composites par apprentissage non supervisé et permanent. C. R. de l'Academie des Sciences, Paris série III, pages 525–528, 1984.
- [95] J. Hérault and C. Jutten. Réseaux neuraux et traitement du signal. Hermès, 1994.
- [96] J. Hérault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neurominétrique en apprentissage non supervisé. Xème colloque GRETSI, Nice, France, 20-24 May, pages 1017–1022, 1985.
- [97] W. Härdle. Smoothing techniques with implementation in s. Springer-Verlag, 1991.
- [98] H.L. Hurd, P. Bloomfield, and R.B. Lund. Periodic correlation in meteorological time series. *Fifth International Meeting on Statistical Meteorology*, pages 1–6, 1992.
- [99] A. Hyvarinen, J. Karhunen, and E. Oja. Independent component analysis. JohnWiely and Sons, 2001.

- [100] Y. Inouye and K. Hirano. Cumulant-based blind identification of linear multiinput-multi-ouput systems driven by colored inputs. *IEEE Transactions on Signal Processing*, 45(6) :1543–1552, June 1997.
- [101] L. Izzo and A. Napolitano. Linear time variant transformations of generalized almost cyclostationary signals. ii. development and applications. *IEEE Transactions on Signal Processing*, 50(12) :2962–2975, December 2002.
- [102] L. Izzo and A. Napolitano. Higher-order statistics for rices representation of cyclostationary signals. *Signal Processing*, 56(3) :279–292, February 1997.
- [103] L. Izzo and A. Napolitano. Multirate processing of time series exhibiting higher order cyclostationarity. *IEEE Transactions on Signal Processing*, 46(2):429– 439, February 1998.
- [104] L. Izzo and A. Napolitano. Sampling of generalized almostcyclostationary signals. *IEEE Transactions on Signal Processing*, 51(6):1546–1556, June 2003.
- [105] L. Izzo and A. Napolitano. Higher order cyclostationarity properties of sampled time series. Signal Processing, 54(3) :303–307, November 1996.
- [106] L. Izzo, L. Paura, and G. Poggi. An interference tolerant algorithm for localization of cyclostationary signal sources. *IEEE Transactions on Signal Processing*, 40(7) :1682–1686, July 1992.
- [107] M. G. Jafari, S. R. Alty, and J. A. Chambers. New natural gradiant algorithm for cyclostationary sources. *IEEE Proceedings*, 151(1):62–68, 2004.
- [108] P. Jallon and A. Chevreuil. Separation of instantaneous mixtures of cyclostationary sources. Signal Processing, 87 :2718–2732, 2007.

- [109] P. Jallon, A. Chevreuil, P. Loubaton, and P. Chevalier. Séparation de mélanges convolutifs de modulations cpm sur-échantillonées. *GRETSI'05*, September 2005.
- [110] R. Jiménez and Y. Shao. On robustness and efficiency of minimum divergence estimators. *Test*, 10(2) :241–248, 2001.
- [111] C. Jutten. Calcul neuromiméque et traitement du signal. analyse en composantes indépendantes. Thèse de doctorat d'état, INPG, Grenoble, France, 1987.
- [112] C. Jutten, M. Babaie-Zadeh, and S. Hosseini. Three easy ways for separating nonlinear mixtures? Signal Processing, 84 :217–229, 2004.
- [113] C. Jutten and J. Hérault. Independent component analysis versus pca. in Proc. EUSIPCO, Grenoble, France, pages 643–646, 1988.
- [114] C. Jutten and J. Hérault. Blind separation of sources, part i : An adaptative algorithm based on neuromimetic architecture. *Signal Processing*, 24(1) :1–10, 1991.
- [115] A. Kachenoura, L. Albera, and L. Senhadji. Séparation aveugle de sources en ingénierie biomédicale blind source separation in biomedical engineering. *ITBM-RBM*, 28 :20–34, 2007.
- [116] A. Kachenoura, H. Gauvrit, and L. Senhadji. Extraction and separation of eyes movements and the muscular tonus from a restricted number of electrodes using the independent component analysis. in 25th annual international conference of the IEEE Engineering in Medicine and Biology Society, Cancun, Mexico, pages 2359–2362, September 2003.

- [117] A.R. Kacimov, Y.V. Obnosov, and N.D. Yakimov. Groundwater flow in a medium with a parquet-type conductivity distribution. *Journal of Hydrology*, 226(3-4) :242–249, 1999.
- [118] M. Kawamoto, K. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi. Blind separation for convolutive mixtures of non-stationary signals. Ds Actes Intl. Conf. on neural information processing (ICONIP'98), Kitakyushu, Japon, Octobre 1998.
- [119] A. Keziou. Dual representation of φ-divergences and applications. C. R. Math. Acad. Sci. Paris, 336(10) :857–862, 2003.
- [120] A. Keziou and S. Leoni-Aubin. On empirical likelihood for semiparametric two-sample density ratio models. J. Statist. Plann. Inference, 138(4) :915– 928, 2008.
- [121] A. Keziou, M.S. Ould Mohamed, H. Fenniri, and G. Delaunay. A new criterion for second order cyclostationary blind source separation. 2nd International Conference on Signals, Circuits and Systems (SCS 2008). IEEE, Hammamet Tunisia, 07-09 November 2008.
- [122] J.-L. Lacoume, P.-O. Amblard, and P. Comon. Statistiques d'ordre supérieur pour le traitement du signal. *Masson*, 1997.
- [123] J.-L. Lacoume and P. Ruiz. Sources identification : a solution based on cumulants. Ds Actes Workshop. on Spectrum, Estimations and Modelling, pages 199–203, 1988.
- [124] J.-L. Lacoume and P. Ruiz. Separation of independent sources from correlated inputs. *IEEE Trans. on Signal Processing*, 40(12):3074–3078, December 1992.

- [125] A. Larue. Blancheur et non-gaussianité pour la déconvolution aveugle de données bruitées : Application aux signaux sismiques. Thèse de l'Institut National Polytechnique de Grenoble, France, 2006.
- [126] N. Le Bihan, J.I. Mars, and H. Pedersen. Wavefield separation on multicomponent sensors. 62nd meeting of european association of geoscientists and engineers, 2000.
- [127] F. Liese and I. Vajda. Convex statistical distances, volume 95 of Teubner-Texte zur Mathematik [Teubner Texts in Mathematics]. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987.
- [128] B. G. Lindsay. Efficiency versus robustness : the case for minimum Hellinger distance and related methods. Ann. Statist., 22(2) :1081–1114, 1994.
- [129] S. Léwy. Acoustique industrielle et aéroacoustique. *HERMES*, *Paris*, 2001.
- [130] A. Mansour. Contribution à la séparation aveugle de sources. Thèse de l'Institut National Polytechnique de Grenoble, France, 1997.
- [131] A. Mansour. Blind separation of sources : Methods, assumptions and applications. In Special Issue on Digital Signal Processing in IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E8(8) :1498–1512, August 2000.
- [132] A.C. McCormick and A.C. Nandi. Cyclostationarity in rotating machine vibrations. *Mechanical Systems and Signal Processing*, 12(2):225–242, 1998.
- [133] N. Mitianoudis and M. Davies. Audio source separation of convolutive mixtures. *IEEE Trans. on Speech and Audio Processing*, 2002.
- [134] E. Moreau. Apprentissage et adaptativité, séparation auto-adaptative de sources indépendantes. PhD Thesis, Université de Paris-Sud, France, 1995.

- [135] E. Moreau. A generalization of joint-diagonalization criteria for source separation. Trans. Signal Processing, IEEE, 49(3):530–541, 2001.
- [136] E. Moreau and O. Macchi. New self-adaptatif algorithms for source separation based on contrast functions. In Proc. HOS'93, SP Workshop on Higher-order Statistics, Lake Tahoe, USA, 2 :215–219, June 1993.
- [137] E. Moreau and O. Macchi. Two novel architectures for the self-adaptive separation of signals. In proc. ICC'93, International Conference on Communication, Geneva, Switzerland, 2 :1154–1159, May 1993.
- [138] E. Moreau and J. C. Pesquet. Generalized contrasts for multichannel blind deconvolution of linear systems. *IEEE Signal Processing Letters*, 4(6) :182– 183, 1997.
- [139] E. Moreau, J. C. Pesquet, and N. Thirion-Moreau. An equivalence between non symetrical contrasts and cumulant matching for blind source separtion. In International Conference on Independent Component Analysis, pages 301–306, 1999.
- [140] E. Moreau and N. Thirion-Moreau. Non symetrical contrasts for source separation. *IEEE Transactions on Signal Processing*, 47(8) :2241–2252, 1999.
- [141] E. Moulines, J.-F. Cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In Proceedings of ICASSP-97, Munich, Germany, April 21-24 1997.
- [142] J. Nedoma. Uber die ergodizitat and r-ergozitat stationarer wahrscheinlichkeitsmasse. Zeitschrift fur Wahrscheinlichkeitstheorie, 2 :90–97, 1963.
- [143] H. L. Nguyen Thi and C. Jutten. Blind source separation for convolutive

mixtures. *IEEE Transaction on Signal Processing*, 45(2) :209–229, March 1995.

- [144] H. L. Nguyen Thi, C. Jutten, and J. Caelen. Speech enhancement : Analysis and comparison of methods in various real situations. In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, Signal Processing VI, Theories and Applications, Brussels, Belgium. Elsevier, pages 303–306, August 1992.
- [145] D. Nuzillard and A. Bijaoui. Blind source separation and analysis of multispectral astronomical images. Astronomy and Astrophysics Supplement Series, Ser. 147 :129–138, 2000.
- [146] M. Ohata, K. Matsuoka, and T. Mukai. An adaptive blind separation method using para-hermitian whitening filter for convolutively mixed signals. *Signal Processing*, 87(1):33–50, January 2007.
- [147] T. Oostendorp. Modelling the fetal ecg. Ph.D. dissertation, K.U. Nijmegen, the Netherlands, 1989.
- [148] M.S. Ould Mohamed, A. Keziou, H. Fenniri, and G. Delaunay. Nouveau critère de séparation aveugle de sources cyclostationnaires au second ordre. *Revue ARIMA*, 12 :1–14, 2010.
- [149] M.S. Ould Mohamed, A. Keziou, H. Fenniri, and G. Delaunay. Séparation aveugle de sources par optimisation de l'information mutuelle sous contraintes, estimation et selection de modèles. 9th African Conference on Research in Computer Science and Applied Mathematics. Proceedings CARI'2008, Rabat, Morroco, octobre 2008.
- [150] L. Para and C. Spence. Convolutive blind separation of non-stationary sources. *IEEE Trans. Speech and audio processing*, 8(3) :320–327, 2000.

- [151] M. S. Pedersen and C. M. Nielsen. Gradient flow convolutive blind source separation. Ds Actes Intl. Work. on Machine Learning for Signal Processing (MLSP'04), Saint Louis, Brésil, pages 335–344, 2004.
- [152] D.-T. Pham. Mutual information approach to blind separation of stationary sources. *IEEE Trans. Information Theory*, 48(7), 2002.
- [153] D.-T. Pham. Fast algorithme for estimating mutual information, entropies and score functions. Proceedings of ICA 2003 Conference, Nara, Japan, 2003.
- [154] D.-T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of non stationary sources. *IEEE Trans. Signal Processing*, 49(9) :1837–1848, 2001.
- [155] D.-T. Pham, P. Garat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. *In Proc. EUSIPCO*, pages 771–774, 1992.
- [156] D.-T. Pham, C. Servière, and H. Boumaraf. Blind separation of convolutive audio mixtures using nonstationarity. *Proceeding of ICA 2003 Conference*, pp. 975-980, Nara, Japan, April 2003.
- [157] D.-T. Pham, C. Servière, and H. Boumaraf. Blind separation of speech mixtures based on nonstationarity. *Proceeding of ISSPA 2003 Conference, Paris, France*, pages 639–644, July 2003.
- [158] P. Plonsey. Bioelectric phenomena. McGraw-Hill, New Yor, 1969.
- [159] F. Poree, A. Kachenoura, H. Gauvrit, C. Morvan, G. Carrault, and L. Senhadji. Blind source separation for ambulatory sleep recording. on Information Technology in Biomedicine, 10(2):293–301, 2006.

- [160] M. Puigt. Méthodes de séparation aveugle de sources fondées sur des transformées temps-fréquence. application à des signaux de parole. Thèse de l'université Toulouse III - Paul Sabatier, France, 2007.
- [161] A. Raad, J. Antoni, and M. Sidahmed. Indicators of cyclostationarity : Theory and application to gear fault monitoring. *Mechanical Systems and Signal Processing*, 22(3) :547–587, 2008.
- [162] K. Rahbar, J. P. Reilly, and J. H. Manton. Blind identification of mimo fir systems driven by quasistationary sources using second-order statistics : A frequency domain approach. *IEEE Trans. on Signal Processing*, 52(2), 2004.
- [163] R.B. Randall, J. Antoni, and S. Chobsaard. The relationship between spectral correlation and envelope analysis for cyclostationary machine signals, application to ball bearing diagnostics. *Mechanical Systems and Signal Processing*, 15(5):945–962, 2001.
- [164] T. Ristaniemi, K. Raju, J. Karhunen, and E. Oja. Inter-cell interference cancellation cdma array systems by independent component analysis. in Proc. Independent Component Analysis(ICA'03), pages 739–744, April 2003.
- [165] C. Servière. Blind source separation of convolutive muxtures. Ds Actes Work. on statistical Signal and array processing, (SSAP'96), Corfou, Grèce, pages 316–319, juin 1996.
- [166] C. Servière and V. Capdevielle. Blind adaptive separation of wide-band sources. In Proceedings of ICASSP'96, Atlanta, Georgia, USA, 5:2698–2701, May 1996.
- [167] C. Simon. Séparation aveugle des sources en mélanges convolutifs. PhD Thesis, Université de Marne-la-Vallée Paris, France, 1999.

BIBLIOGRAPHIE

- [168] J. Solé-Casals, C. Jutten, and D.T. Pham. Fast approximation of nonlinearities for improving inversion algorithms of pnl mixtures and wiener systems. *Signal Processing*, 85 :1780–1786, 2005.
- [169] A. Souloumiac. Blind source detection and separation using second-order nonstationary. In proceedings IEEE International Conference on acoustics, speech and signal processing (ICASSP'95), Detroit, Michigan, USA, 3 :1912–1915, May 1995.
- [170] C. M. Spooner. Theory and application of higher-order cyclostationarity. Ph.D. Thesis, University of California, Davis, CA, June 1992.
- [171] A. Taleb and C. Jutten. Source separation in post nonlinear mixtures. IEEE Trans. on Signal Processing, 47(10) :2807–2820, 1990.
- [172] A. R. Teixeira, A. M. Tomé, E. W. Lang, and K. Standlthanner. Delayed amuse a tool for blind source separation and denoising. Ds Actes Intl. Conf. on Independent Component Analysis and Blind Source Separation (ICA'04), Grenade, Espagne, Septembre 2004.
- [173] N. Thirion, J. Mars, and J.-L. Lacoume. Séparation aveugle de signaux large bande : un nouveau challenge en prospection sismique. In Actes du XVeme colloque GRETSI, Juan-Les-Pins, France, pages 1335–1338, 18-21 September 1995.
- [174] N. Thirion and E. Moreau. New criteria for blind signal separation. in IEEE Workshop on Statistical Signal and Array Processing, Pennsylvania, US, pages 344–348, 2000.
- [175] J. Thomas. Algorithmes temporels rapides à point fixe pour la séparation aveugle de mélanges convolutifs et/ou sous-déterminés. Thèse de l'université Toulouse III - Paul Sabatier, France, 2007.

- [176] R. L. L. Tong and Y. H. V. C. Soon. Indeterminacy and identifiability of blind identification. *IEEE Trans. CS*, 38 :499–509, 1991.
- [177] K. Torkkola. Blind separation for audio signals are we there yet? In Proceedings of the Workshop on Independent Component Analysis and Blind Signal Separation, Aussois, France, January 11-15 1999.
- [178] R. Vigärio. Extraction of ocular artifacts from eeg using independent component analysis. *Electroenceph. clin. Neurophysiol*, 103 :395–404, 1997.
- [179] R. Vigärio, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. Proc. NIPS'97, Cambridge, MA, MIT Press, pages 229–235, 1998.
- [180] R. Vigärio, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent component approach to the analysis of eeg and meg recordings. *IEEE Transactions on Biomedical Engineering*, 47(5) :589–593, May 2000.
- [181] V. Vrabie. Statistiques d'ordre supérieur : applications en géophysique et électrotechnique. Thèse en cotutelle de l'Institut National Polytechnique de Grenoble, France et de l'université POLITEHNICA de Bucarest, Roumanie, 2003.
- [182] F. Vrins, D.-T. Pham, and M. Verleysen. Is the general form of Renyi's entropy a contrast for source separation?, volume 4666 of Lecture Notes in Computer Science. Springer, Berlin, 2007.
- [183] A. Westner and V. M. Bove. Applying blind source separation and deconvolution to real-world acoustic environments. *Proceeding 106th of Audio Engineering Society*, 1999.

- [184] A. Westner and V. M. Bove. Blind separation of real world audio signals using overdetermined mixtures. In Proceeding of 1st international conference on ICA and BSS, Aussois, France, January 11-15, 1999.
- [185] J. Wisbeck, A. K. Barroso, and R. Ojeda. Application of ica in the separation of breathing artefacts in ecg signals. in International Conference on Neural Information Processing, (ICONIP'98), pages 211–214, 1998.
- [186] H. C. Wu and J. Principe. Simultaneous diagonalization in the frequency domain (sdif) for source separation. In Proceedings of ICA 99, Aussois, France, January 1999.
- [187] J. M. F. Xavier, V. A. N. Barroso, and J. M. F. Moura. Closed-form blind channel identification and source separation in sdma systems through correlative coding. *IEEE Journal on Selected Areas in Communications*, 16 :1506–1517, 1998.
- [188] A. Yeredor. Non-orthogonal joint diagonalisation in the least-square sense with application in blind source separation. *IEEE Transactions on Signal Processing*, 50(7) :1545–1551, 2002.
- [189] A. Ypma. Learning methods for machine vibration analysis and health monitoring. PhD thesis, Pattern Recognition Group, Dept. of Applied Physics, Delft University of Technology, 2001.
- [190] Z. Zhu and F. Kong. Cyclostationarity analysis for gearbox condition monitoring : approaches and effectiveness. *Mechanical Systems and Signal Processing*, 19(3) :467–482, 2005.