

Algorithmes temporels rapides à point fixe pour la séparation aveugle de mélanges convolutifs et/ou sous-déterminés

THÈSE

présentée et soutenue publiquement le 12 décembre 2007

pour l'obtention du

Doctorat de l'Université de Toulouse
délivré par l'Université Toulouse III – Paul Sabatier
(spécialité traitement du signal)

par

Johan Thomas

Composition du jury

<i>Président :</i>	Jean-François Cardoso	Directeur de Recherche CNRS	LTCI-ENST, Paris
<i>Rapporteurs :</i>	Christian Jutten	Professeur	GIPSA-UJF, Grenoble
	Dinh-Tuan Pham	Directeur de Recherche CNRS	LJK, Grenoble
<i>Examinatrice :</i>	Christine Joblin	Directrice de Recherche CNRS	CESR, Toulouse
<i>Directeurs :</i>	Yannick Deville	Professeur	LATT-UPS, Toulouse
	Shahram Hosseini	Maître de Conférences	LATT-UPS, Toulouse

Mis en page avec la classe thloria.

Remerciements

En premier lieu, je tiens à remercier tout particulièrement Yannick Deville et Shahram Hosseini pour m'avoir proposé d'effectuer ma thèse dans leur équipe de séparation de sources. Leurs qualités scientifiques et humaines ainsi que leur expérience dans le domaine de la recherche ont représenté de réels atouts durant ces trois années de thèse.

Mes remerciements vont aussi à Christine Joblin et Olivier Berné du Centre d'Etude Spatiale des Rayonnements pour la coopération que nous avons mise en place pendant ma thèse et qui a permis l'application de certaines de nos méthodes de séparation de sources à l'étude des nuages de poussières interstellaires.

Je remercie M. Jean-François Cardoso, Mme Christine Joblin, M. Christian Jutten et M. Dinh-Tuan Pham pour avoir accepté d'être les membres de mon jury de soutenance ainsi que pour leurs remarques qui ont contribué à l'amélioration de la qualité de ce rapport.

J'exprime toute ma reconnaissance à mes parents pour m'avoir encouragé et soutenu tout au long de mes études.

Mes salutations à mes collègues de bureau Rima, Matthieu et Hicham avec qui il a été très agréable de travailler. Mes sentiments amicaux vont également à tous les membres de l'équipe S2I : Ali, Chahinez, Christophe, Denis, Eric, Guilhem, Hervé, Jean, Jean-François, Jean-Louis, José-Philippe, Karine, Laurent, Olivier, Paul, Rima, Sébastien, Sylvie et Truswin.

Je remercie tous les membres du LATT, permanents, doctorants ou stagiaires qui m'ont permis de travailler dans un contexte de travail idéal, ainsi que le service informatique du laboratoire pour son efficacité et sa disponibilité pour répondre aux attentes des chercheurs.

Mes remerciements aux développeurs des logiciels libres Fedora Core, Texmaker, SAOImage DS9, et de la classe Thloria qui m'ont été d'une grande utilité.

Je remercie enfin le Théâtre national du Capitole, la ville de Toulouse, ainsi que les stations de ski et refuges pyrénéens pour m'avoir permis de mener mes études dans un environnement idéal.

Développe en toi l'indépendance à tout moment, avec bienveillance, simplicité et modestie.

Marc Aurèle

Table des matières

Table des figures	ix
Liste des tableaux	xi
Notations	xiii
Introduction générale	1
 I Mélanges convolutifs (sur-)déterminés	 9
1 Mélanges convolutifs : état de l'art	11
1.1 Introduction	11
1.2 Position du problème	11
1.3 Hypothèses et critères de séparation	16
1.3.1 Statistiques croisées d'ordre supérieur	17
1.3.2 La vraisemblance	19
1.3.3 L'information mutuelle	21
1.3.4 La non-gaussianité	22
1.3.5 Statistiques croisées du second ordre	25
1.3.6 La parcimonie	26
1.3.7 La positivité	29
1.4 Classification des méthodes de SAS convolutives	30
1.4.1 Les méthodes fréquentielles	30
1.4.2 Les méthodes temporelles	32
1.5 Conclusion	33
 2 Extension des méthodes de SAS basées sur la non-gaussianité	 35

2.1	Introduction	35
2.2	Position du problème	35
2.3	Approches existantes	37
2.4	Nouvelles méthodes pour extraire un processus d'innovation	39
2.5	Reformulation du mélange sous forme linéaire instantanée .	44
2.6	Performances atteignables par le modèle linéaire instantané sous-déterminé	46
2.7	Méthode globale de SAS proposée	48
2.8	Conclusion	48
3	Résultats expérimentaux	51
3.1	Introduction	51
3.2	Séparation de signaux de télécommunications	51
3.3	Séparation de signaux audio	56
3.4	Tests statistiques	60
3.5	Conclusion	62
II	Mélanges sous-déterminés	65
4	Mélanges sous-déterminés : état de l'art	67
4.1	Estimation de la matrice de mélange	68
4.2	Reconstruction des sources	68
4.3	La séparation différentielle de sources	71
5	Méthode de SAS différentielle proposée pour les mélanges instantanés	73
5.1	Introduction	73
5.2	Un nouveau critère de SAS basé sur le kurtosis différentiel .	73
5.3	Résumé des méthodes proposées	78
5.4	Preuves de convergence	80
5.5	Conclusion	82
6	Méthode de SAS différentielle proposée pour les mélanges convolutifs	83
6.1	Introduction	83

6.2	Utilisation du kurtosis différentiel dans le cas convolutif . . .	83
6.3	Résumé de la méthode proposée	88
6.4	Reformulation du mélange sous forme linéaire instantanée .	90
6.5	Conclusion	91
7	Résultats expérimentaux	93
7.1	Introduction	93
7.2	Séparation de mélanges linéaires instantanés de signaux télécoms	93
7.3	Séparation de mélanges linéaires instantanés de signaux audio	96
7.4	Séparation de mélanges convolutifs de signaux télécoms . . .	98
7.5	Tests statistiques	100
7.6	Conclusion	105
III	Mélanges linéaires instantanés contenant un grand nombre d'échantillons	107
8	Algorithmes rapides optimisés pour les mélanges linéaires instantanés	109
8.1	Introduction	109
8.2	Position du problème	109
8.3	Un algorithme optimisé rapide pour la SAS (sur-) déterminée	111
8.3.1	Principes et limitations de l'approche standard . . .	111
8.3.2	Une nouvelle approche pour optimiser le kurtosis . .	112
8.3.3	Résumé de la méthode proposée	115
8.4	Un algorithme optimisé rapide pour la SAS sous-déterminée	115
8.4.1	Principe et limitations de notre approche DFICA . .	115
8.4.2	Une nouvelle approche pour optimiser le kurtosis différentiel	117
8.4.3	Résumé de la méthode proposée	118
8.5	Conclusion	119
9	Résultats expérimentaux	121
9.1	Introduction	121

9.2	Application à l'astrophysique	121
9.3	Tests statistiques	126
9.4	Conclusion	128
Conclusion générale		131
Annexes		137
Annexe A : Blanchiment		139
Annexe B : Estimation des contributions des sources		143
Annexe C : Définitions des critères SIR_{out} , SIR_{in} , $SIRI$ et SNR_{in}		145
Annexe D : Gradient du kurtosis différentiel		147
Annexe E : La corrélation différentielle		149
Annexe F : Définitions des critères SIR_{out} , SIR_{in} et $SIRI$ pour les méthodes de séparation partielle		151
Annexe G : Choix de la famille de vecteurs d'extraction		153
Bibliographie		155
Publications de l'auteur		173

Table des figures

1.1	Illustration d'un mélange convolutif acoustique à 3 sources et 3 capteurs.	12
1.2	Schéma-bloc du processus de mélange.	14
1.3	Schéma-bloc d'un dispositif de séparation direct.	14
1.4	Schéma-bloc d'un dispositif de séparation récurrent.	15
1.5	Kurtosis non-normalisé et normalisé.	23
2.1	Schéma-bloc du processus de déflation.	38
2.2	Schéma-bloc du processus de blanchiment convolutif.	40
3.1	Sources utilisées 1 et 2 correspondant à des signaux binaires. . .	52
3.2	Résultats de séparation d'un mélange convolutif de signaux binaires.	53
3.3	Sources utilisées 1 et 2 correspondant à des signaux télécoms AM à 4 états.	54
3.4	Résultats de séparation d'un mélange convolutif de signaux AM. . .	55
3.5	Réponses impulsionnelles de la matrice de filtres de mélange utilisée correspondant aux azimuts 15° et -30° (ordre=256). . . .	56
3.6	Résultats de séparation d'un mélange convolutif de signaux audio (1).	57
3.7	Réponses impulsionnelles des filtres de mélange utilisés correspondant aux azimuts 10° et 60° (ordre=256).	58
3.8	Résultats de séparation d'un mélange convolutif de signaux audio (2).	59
3.9	SIR_{out} des sources extraites en fonction de Q pour des sources colorées artificielles dans le cas de processus d'innovation uniformes et en utilisant le critère du kurtosis.	61
3.10	SIR_{out} des sources extraites en fonction de Q pour des sources colorées artificielles dans le cas de processus d'innovation laplaciens et en utilisant le critère de la néguentropie.	61
3.11	SIR_{out} des sources utiles extraites en fonction du SNR d'entrée. . .	63
3.12	Temps d'extraction moyen du premier processus d'innovation en fonction de Q pour une optimisation de type Tugnait-Newton et pour notre optimisation de type point fixe.	63

4.1	Schéma du principe d'estimation des sources à partir de la matrice de mélange en considérant qu'une seule source (celle d'indice 3) est active à l'instant n	69
4.2	Schéma du principe d'estimation des sources à partir de la matrice de mélange par la méthode du trajet le plus court.	70
7.1	Résultats de séparation d'un mélange linéaire-instantané de signaux télécoms.	95
7.2	Résultats de séparation d'un mélange linéaire-instantané de signaux audio.	97
7.3	Sources utilisées 1 et 2 correspondant à des signaux télécoms AM démodulés à 4 états.	98
7.4	Résultats de séparation d'un mélange convolutif de signaux télécoms.	99
7.5	SIR_{out} obtenus avec les versions à déflation de FastICA et DFICA en fonction du SNR d'entrée.	101
7.6	$Perf$ et SIR_{out} obtenus avec les versions symétriques de FastICA et DFICA en fonction du SNR d'entrée.	101
7.7	SIR_{out} obtenus avec C-FICA et C-DFICA en fonction du SNR d'entrée.	103
7.8	SIR_{out} obtenus avec C-DFICA en fonction du nombre d'échantillons.	103
7.9	Temps d'extraction moyen du premier processus d'innovation en fonction de Q pour une optimisation de type Newton et pour notre optimisation de type point fixe (algorithme C-DFICA). . .	104
9.1	Image de la nébuleuse NGC 2023 prise par le télescope Hubble à la longueur d'onde $\lambda_1=474$ nm.	124
9.2	Image de la nébuleuse NGC 2023 prise par le télescope Hubble à la longueur d'onde $\lambda_2=631$ nm.	124
9.3	Image séparée 1 obtenue avec l'algorithme FastICA en 115 sec. et avec l'algorithme O-FICA en 4.5 sec.	125
9.4	Image séparée 2 obtenue avec l'algorithme FastICA en 115 sec. et avec l'algorithme O-FICA en 4.5 sec.	125
9.5	Rapport t'/t des temps de calcul en fonction du nombre T d'échantillons.	127
9.6	Indice de performances $Perf$ en fonction du temps de calcul. . .	129
1	Schéma-bloc des étages de blanchiment et de normalisation. . .	140
2	Schéma-bloc du processus de blanchiment et de normalisation. .	140

Liste des tableaux

- 1.1 Tableau de synthèse des équations de mélange et de séparation. 16
- 1.2 Tableau récapitulatif des avantages et inconvénients respectifs
des méthodes de séparation temporelles et fréquentielles [150]. . 34

Notations

- SAS : Séparation Aveugle de Sources (*BSS : Blind Source Separation* en anglais).
- ACI : Analyse en Composantes Indépendantes (*ICA : Indep. Component Analysis*).
- $\mathbf{x}(n)$: vecteur colonne composé de signaux d'observation $x_i(n)$.
- $\mathbf{X}(z)$: transformée en \mathcal{Z} du vecteur $\mathbf{x}(n)$.
- \mathbf{A} : matrice de mélange instantané composée de coefficients scalaires réels a_{ij} .
- \mathbf{G} : matrice de performances définie par le produit des matrices de mélange et de séparation (\mathbf{G} est composée de coefficients g_{ij}).
- \mathbf{M}^t : matrice transposée de la matrice \mathbf{M} .
- $\mathbf{A}(z)$: matrice de mélange convolutif composée de filtres $A_{ij}(z)$.
- $\mathbf{H}(z)$: matrice de filtres modélisant la coloration des processus d'innovation des sources *et* le mélange convolutif des sources.
- N : nombre de sources présentes dans le mélange.
- P : nombre d'observations du mélange.
- T : nombre d'échantillons de chaque observation.
- Q : ordre de la matrice de filtres $\mathbf{H}(z)$.
- $\text{corr}_{xy}(n)$: coefficient de corrélation des signaux x et y à l'instant n .
- $\text{Dcorr}_{xy}(n)$: coefficient de corrélation différentielle des signaux x et y entre les instants n_1 et n_2 .
- $\mathbf{R}_{\mathbf{x}}(n)$: matrice d'autocorrélation du vecteur \mathbf{x} à l'instant n .
- $\mathbf{DR}_{\mathbf{x}}(n_1, n_2)$: matrice d'autocorrélation différentielle du vecteur \mathbf{x} entre les instants n_1 et n_2 .
- $\mathbf{r}_{xy}(n)$: vecteur d'intercorrélation des signaux x et y à l'instant n .
- filtre RIF : filtre à réponse impulsionnelle finie ou filtre à moyenne ajustée (MA).
- filtre RII : filtre à réponse impulsionnelle infinie ; filtre autorégressif (AR) ou filtre autorégressif à moyenne ajustée (ARMA).
- $\delta(n)$: symbole de Kronecker, défini par $\delta(0) = 1$ et $\delta(n \neq 0) = 0$.
- $*$: opérateur de convolution.
- *SIR* : Rapport signal à interférences (*Signal to Interference Ratio*).
- *SIRI* : amélioration du *SIR* (*Signal to Interference Ratio Improvement*).
- *SNR* : Rapport signal à bruit (*Signal to Noise Ratio*).
- $E\{y(n)\}$: espérance du signal $y(n)$.
- $\text{kurt}_y(n)$: kurtosis non-normalisé du signal y à l'instant n .
- $\text{Dkurt}_y(n_1, n_2)$: kurtosis différentiel du signal y entre les instants n_1 et n_2 .
- $\text{kurt}_y^{\text{nor}}(n)$: kurtosis normalisé du signal y à l'instant n .
- $\text{Dkurt}_y^{\text{nor}}(n_1, n_2)$: kurtosis normalisé différentiel du signal y entre n_1 et n_2 .
- $\text{pow}_y(n)$: puissance du signal y à l'instant n .
- $\text{Dpow}_y(n_1, n_2)$: puissance différentielle du signal y entre les instants n_1 et n_2 .
- i.i.d. : indépendant et identiquement distribué.

Introduction générale

La grande majorité des technologies modernes, dans la plupart des disciplines, utilisent des capteurs pour mesurer des grandeurs utiles pour l'étude ou le contrôle d'un phénomène physique. Les microphones par exemple captent les ondes sonores se propageant dans l'air, les antennes radio mesurent l'évolution d'un champ électromagnétique, les détecteurs photo-sensibles numérisent l'image obtenue par un système optique, enfin les électrodes placées sur le crâne d'un patient permettent d'enregistrer le champ électrique associé à son activité cérébrale.

En pratique, malheureusement, les signaux obtenus grâce à ces capteurs sont souvent des mélanges de plusieurs contributions issues de signaux originaux appelés sources. Dans le cas le plus général, ces contributions sont des signaux obtenus par filtrage non-linéaire des sources physiques. Dans d'autres cas plus simples, elles peuvent être des versions pondérées et/ou décalées temporellement des sources.

La Séparation Aveugle de Sources (SAS) consiste à estimer les signaux sources (ou éventuellement leurs contributions) à partir des mélanges obtenus sur chaque capteur. Le corps humain possède un organe réalisant ce type d'opération : le cerveau. Celui-ci peut en effet se focaliser sur un phénomène extérieur particulier détecté par les sens. On peut ainsi écouter une personne s'exprimant dans une salle bruitée où d'autres personnes parlent en même temps (un hall de conférences par exemple). La zone du cerveau dédiée à la vision peut aussi fixer notre attention sur une zone précise du champ visuel.

La séparation de sources, en tant que discipline scientifique est relativement récente et s'est inspirée, à son commencement, du cerveau humain et de son organisation en réseaux neuronaux [88] (1984). Elle s'est aussi basée sur des résultats obtenus en déconvolution aveugle, en adaptant un critère particulier, le kurtosis ou autocumulant d'ordre 4 qui est aujourd'hui encore l'un des plus utilisés. La SAS a ensuite connu durant deux décennies des évolutions constantes, et a trouvé un vaste champ d'applications, ce qui en fait actuellement un domaine de recherche majeur en traitement du signal. On utilise les techniques de séparation aveugle de sources dans des domaines aussi divers que le traitement de sources audio, les télécommunications multicanales, l'analyse de signaux biomédicaux, l'imagerie astrophysique multispectrale, le traitement de données géophysiques, la détection et la localisation radar et bien d'autres [64, 65, 102, 103, 106].

Le processus de mélange entre sources et capteurs est modélisé en général par l'équation vectorielle suivante :

$$\mathbf{x}(t) = \mathcal{A}(\mathbf{s}(t)) + \mathbf{b}(t) \quad (1)$$

où

- $\mathbf{s}(t)$ est le vecteur des signaux sources que l'on cherche à estimer.

- $\mathbf{x}(t)$ est le vecteur des observations mesurées par les capteurs.
- $\mathbf{b}(t)$ est le vecteur de bruit qui modélise les erreurs de mesure des capteurs.
- \mathcal{A} est l'opérateur de mélange.

On peut classer les problèmes de SAS en fonction de la nature de l'opérateur \mathcal{A} .

a) Si cet opérateur est linéaire, on parle de *mélange linéaire*. L'équation (1) s'écrit alors

$$\mathbf{x}(t) = \mathbf{A}(t) * \mathbf{s}(t) + \mathbf{b}(t) \quad (2)$$

où

- $\mathbf{A}(t)$ est une matrice de réponses impulsionnelles de filtres.
- $*$ est l'opérateur de convolution continue.

Si les signaux mesurés sont à temps discret, ce qui est nécessairement le cas pour des signaux numériques enregistrés, la variable de temps t est remplacée par un indice temporel n entier et l'opérateur de convolution continue est remplacé par une convolution discrète. L'équation (2) devient alors

$$\mathbf{x}(n) = \mathbf{A}(n) * \mathbf{s}(n) + \mathbf{b}(n). \quad (3)$$

Parmi les mélanges linéaires, on peut isoler deux cas particuliers de mélanges :

- Les mélanges sont *linéaires instantanés* quand la matrice de réponses impulsionnelles de filtres $\mathbf{A}(n)$ est constituée d'échantillons unités δ centrés en 0. La matrice $\mathbf{A}(n)$ s'écrit alors sous la forme $\mathbf{A}(n) = \mathbf{A}\delta(n)$, où \mathbf{A} est une matrice, et l'opérateur de convolution devient une simple multiplication matricielle

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{b}(n). \quad (4)$$

- Les mélanges sont *linéaires à atténuations et retards* quand la fonction $\mathbf{A}(n)$ est constituée d'échantillons unités d'amplitudes différentes et centrés à des instants différents suivant le couple source-capteur considéré.

b) Les *mélanges non-linéaires* sont les plus généraux mais aussi les plus difficiles à traiter. Ils ont d'abord été étudiés dans un cas particulier : celui où la non-linéarité est introduite uniquement au niveau du capteur du fait par exemple d'un phénomène d'écrtage. Les observations consistent alors en une somme de contributions qui subit une non-linéarité au moment de la conversion en signal électrique. On parle alors de mélanges *post-non-linéaires*. Notons que la zone auditive du cerveau effectue la séparation de ce type de mélanges. En effet l'oreille humaine n'a pas une réponse parfaitement linéaire : l'influx nerveux du nerf auditif n'est pas proportionnel à la pression exercée sur le

tympan. Depuis, des configurations plus générales ont été traitées comme les mélanges linéaires quadratiques [70] ou certaines classes particulières de mélanges [76]. La classe des mélanges non-linéaires n'est pas été étudiée dans la suite de ce document.

On peut aussi classer les mélanges suivant le nombre d'observations relativement au nombre de sources. Les mélanges sont dits *sur-déterminés* quand le nombre d'observations est supérieur au nombre de sources. Si le nombre d'observations est inférieur au nombre de sources, on parle de mélanges *sous-déterminés*. Enfin, si les nombres de sources et d'observations sont égaux, le mélange est dit déterminé. Selon le cas, la matrice de mélange $\mathbf{A}(n)$ a un plus grand nombre de lignes (mélanges sur-déterminés), de colonnes (mélanges sous-déterminés), ou un nombre égal de lignes et de colonnes (mélanges déterminés).

Dans la première partie de cette thèse, nous nous sommes intéressés aux *mélanges convolutifs (sur-)déterminés*, pour lesquels nous proposons un algorithme à point fixe opérant dans le domaine temporel. Le chapitre 1 est consacré à un état de l'art de la séparation aveugle de sources. Nous décrivons l'ensemble des classes de critères utilisées dans la littérature jusqu'ici et évoquons pour chacun d'eux les extensions qui ont été proposées pour traiter les mélanges convolutifs. Nous analysons ensuite les avantages et inconvénients des deux grandes approches fréquentielle et temporelle de séparation aveugle pour mélanges convolutifs. Dans le chapitre 2, nous cherchons à étendre le très populaire algorithme FastICA aux mélanges convolutifs. Pour cela, nous introduisons un processus de blanchiment spatio-temporel non-causal, qui, en initialisant d'une certaine façon les paramètres d'extraction, permet d'utiliser une procédure d'optimisation de type point fixe analogue à celle de FastICA. L'estimation des contributions des sources dans les observations est réalisée au moyen d'un critère quadratique correspondant à la différence entre une observation donnée et une version filtrée du processus d'innovation extrait par l'optimisation à point fixe. L'optimisation de ce critère est réalisée au moyen d'un processus de filtrage de Wiener non-causal. Dans le chapitre 3, nous validons notre algorithme en effectuant des tests de séparation utilisant des signaux réels et simulés, puis nous effectuons une analyse statistique pour comparer ses performances en termes de séparation et de rapidité avec l'algorithme de Tugnait qui suppose les mêmes hypothèses de stationnarité et de non-gaussianité des sources. Cette première partie a fait l'objet d'une publication dans la revue IEEE SPL [203] (une première version de l'article a été présentée au Workshop ECMS qui a eu lieu à Toulouse en mai 2005 [202]).

Dans la deuxième partie, nous nous intéressons aux *mélanges sous-déterminés linéaires instantanés ou convolutifs*, pour lesquels nous cherchons là aussi à étendre l'algorithme FastICA. Nous présentons dans le chapitre 4 un état de l'art de la séparation aveugle de sources sous-déterminée et constatons que l'hypothèse de parcimonie des sources est utilisée pour la plupart des mé-

thodes existantes. Dans le chapitre 5 consacré aux mélanges *linéaires instantanés sous-déterminés*, en nous basant sur le concept de séparation différentielle de sources, nous proposons un processus de blanchiment différentiel des observations qui permet l'emploi d'itérations à point fixe pour optimiser un critère différentiel correspondant à la différence entre deux instants du kurtosis du signal de sortie. Deux versions de l'algorithme proposé, baptisé DFICA sont présentées : une version à déflation et une version parallèle. Dans le chapitre 6, nous utilisons le même concept de séparation différentielle de sources pour les mélanges *convolutifs sous-déterminés* pour lesquels nous proposons un processus de blanchiment spatio-temporel différentiel permettant d'extraire par une procédure d'optimisation à point fixe les processus d'innovation de sources dites utiles en présence de sources de bruit. L'estimation des contributions des sources utiles, superposées à des contributions de sources de bruit, est réalisée au moyen d'un critère quadratique différentiel dont l'optimisation est effectuée au moyen d'un processus de filtrage de Wiener dit différentiel. Nous appelons C-DFICA cet algorithme destiné aux mélanges convolutifs sous-déterminés. Le chapitre 7 démontre l'utilité de nos algorithmes en présentant des tests de séparation de mélanges de sources réelles et artificielles ; des tests de Monte-Carlo permettent d'évaluer la résistance de nos algorithmes différentiels DFICA et C-DFICA à la présence de sources de bruit. Nous comparons aussi la rapidité de l'algorithme C-DFICA avec celle d'une version différentielle de l'algorithme de Tugnait développée auparavant dans l'équipe. Le contenu de cette seconde partie a donné lieu à une publication dans la revue IEEE TSP [204].

La troisième partie est consacrée aux mélanges *linéaires instantanés* de sources contenant un grand nombre d'échantillons. Cette configuration est fréquente notamment en séparation aveugle d'images du fait de l'augmentation du nombre de pixels des capteurs photo-sensibles de types CCD ou CMOS. Le chapitre 8 met d'abord en évidence une limitation de l'algorithme FastICA applicable aux mélanges *linéaires instantanés (sur-)déterminés* qui, à chaque itération de l'algorithme d'optimisation à point fixe, effectue un calcul de statistiques sur l'ensemble des échantillons disponibles, ce qui demande de réaliser un grand nombre d'opérations élémentaires. FastICA demande pour cela le stockage en mémoire de signaux, qui, si le nombre d'échantillons disponibles est élevé, demande un espace mémoire important. Après avoir constaté que le critère du kurtosis est polynômial relativement aux coefficients d'extraction à estimer, nous proposons une procédure d'identification de ce polynôme qui permet ensuite d'appliquer des itérations de type point fixe dans un espace de calcul moins gourmand. Nous baptisons O-FICA cet algorithme applicable aux mélanges linéaires instantanés de longs signaux de mélange. Nous reprenons ensuite le principe d'identification polynômiale multivariable et l'appliquons à l'algorithme DFICA pour les mélanges *linéaires instantanés sous-déterminés*. Le kurtosis différentiel pouvant aussi être identifié comme un polynôme, une procédure d'identification similaire à celle d'O-FICA permet d'optimiser l'algorithme DFICA lorsque les mélanges contiennent un grand nombre d'échan-

tillons. Nous baptisons O-DFICA cet algorithme proposé pour la séparation sous-déterminée de signaux longs. Le chapitre 9 permet de mesurer le gain en temps de calcul obtenu avec nos algorithmes optimisés comparé à leur version standard associée. Un article au sujet de l'algorithme O-FICA a été présenté à la conférence GRETSI en Septembre 2007 [206] (une version anglaise a été présentée au Workshop ECMS en mai 2007 [205]). Un article de revue est actuellement en cours de révision pour la revue IEEE TSP concernant les deux algorithmes O-FICA et O-DFICA.

Première partie

Mélanges convolutifs
(sur-)déterminés

Chapitre 1

Mélanges convolutifs : état de l'art

1.1 Introduction

La séparation aveugle de mélanges convolutifs est un domaine de recherche récent et très prometteur. Elle étudie la séparation des *mélanges linéaires* généraux, les mélanges *linéaires instantanés* et à *atténuations et retards* étant des cas particuliers réalistes pour certaines applications uniquement. En effet, le temps de propagation entre une source et un capteur n'est jamais parfaitement nul, même s'il peut parfois être négligé. La sous-classe des mélanges à atténuations et retards modélise ces retards de propagation mais suppose que le signal n'est pas déformé par le milieu, ce qui peut ne pas être conforme à la réalité. Pour les applications audio par exemple, on ne rencontre ce type de mélanges que lorsque l'enregistrement est réalisé dans une chambre anéchoïque qui élimine les trajets multiples (ou réverbérations) entre les sources et les capteurs. Très souvent, donc, de nombreuses versions temporellement décalées d'une même source parviennent à chaque capteur en raison de réflexions multiples dans le milieu. Chaque version décalée possède sa propre atténuation qui correspond à un coefficient de la réponse impulsionnelle du filtre du couple source-capteur considéré. Le phénomène de réverbération acoustique est illustré de façon simplifiée par la figure 1.1. Dans ce chapitre, nous commençons par définir le modèle de mélange convolutif puis nous répertorions les différents types de critères d'abord utilisés pour la séparation de mélanges linéaires instantanés puis étendus à la classe des mélanges convolutifs. Nous divisons ensuite les méthodes convolutives en deux types : les méthodes fréquentielles utilisant la transformée de Fourier à court terme et les méthodes temporelles qui cherchent à estimer des filtres d'extraction formulés temporellement.

1.2 Position du problème

La classe générale des mélanges convolutifs tient compte de la déformation du signal propagé et la modélise par un filtrage entre la source et l'observation,

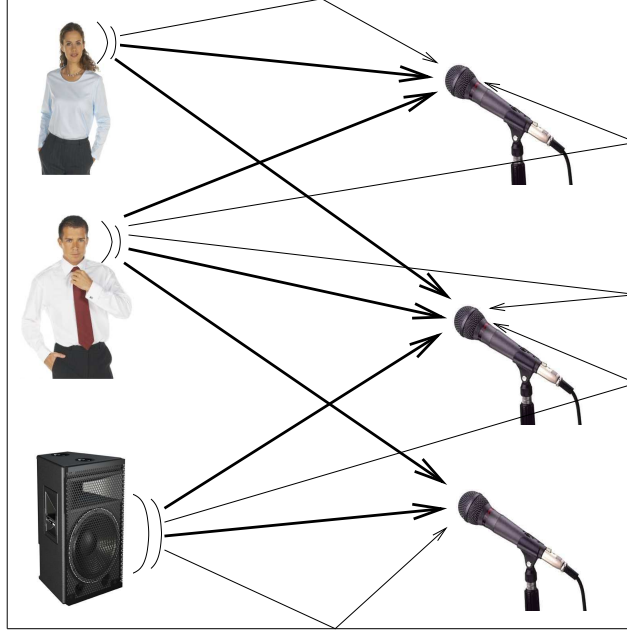


FIG. 1.1 – Illustration d'un mélange convolutif acoustique à 3 sources et 3 capteurs.

ce qui s'écrit mathématiquement sous la forme d'une convolution. La contribution d'une source d'indice j sur un capteur i s'exprime donc de la manière suivante :

$$x_{ij}(t) = a_{ij}(t) * s_j(t) \quad (1.1)$$

où $*$ est l'opérateur de convolution¹.

Toutes les équations reliant les P observations aux N sources (où $P \geq N$) peuvent s'écrire de manière synthétique par la relation matricielle suivante :

$$\mathbf{x}(t) = \mathbf{A}(t) * \mathbf{s}(t) \quad (1.2)$$

où $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^t$ et $\mathbf{x}(t) = [x_1(t), \dots, x_P(t)]^t$ sont respectivement les vecteurs source et observation et où

$$\mathbf{A}(t) = \begin{pmatrix} a_{11}(t) & \cdots & a_{1N}(t) \\ \vdots & & \vdots \\ a_{P1}(t) & \cdots & a_{PN}(t) \end{pmatrix} \quad (1.3)$$

regroupe les réponses impulsionnelles des filtres présents entre sources et capteurs. Lorsque l'on considère des signaux numériques, l'éq. (1.2) s'exprime à l'aide d'un indice temporel discret n au lieu de la variable de temps continue t :

$$\mathbf{x}(n) = \mathbf{A}(n) * \mathbf{s}(n) \quad (1.4)$$

¹Le terme de bruit de mesure $\mathbf{b}(n)$ n'est pas pris en compte ici.

où $*$ est cette fois l'opérateur de convolution discrète.

Dans le domaine en \mathcal{Z} , on obtient alors

$$\mathbf{X}(z) = \mathbf{A}(z)\mathbf{S}(z) \quad (1.5)$$

où $\mathbf{A}(z)$ est une matrice de fonctions de transfert de filtres qui est la transformée en \mathcal{Z} de $\mathbf{A}(n)$

$$\mathbf{A}(z) = \begin{pmatrix} \sum_{k=-\infty}^{+\infty} a_{11}(k)z^{-k} & \cdots & \sum_{k=-\infty}^{+\infty} a_{1N}(k)z^{-k} \\ \vdots & & \vdots \\ \sum_{k=-\infty}^{+\infty} a_{P1}(k)z^{-k} & \cdots & \sum_{k=-\infty}^{+\infty} a_{PN}(k)z^{-k} \end{pmatrix} \quad (1.6)$$

Pour que le mélange soit réellement convolutif (c.-à-d. pas seulement à atténuations et retards), il faut que pour au moins un couple d'indices (i, j) , la fonction $A_{ij}(z) = \sum_{k=-\infty}^{+\infty} a_{ij}(k)z^{-k}$ ait deux coefficients $a_{ij}(k)$ non nuls. La grande majorité des travaux de séparation aveugle de sources dans le cas convolutif supposent un modèle de mélange faisant intervenir des filtres causaux à Réponses Impulsionnelles Finies (RIF). La matrice de filtres $\mathbf{A}(z)$ se réécrit alors

$$\mathbf{A}(z) = \begin{pmatrix} \sum_{k=0}^K a_{11}(k)z^{-k} & \cdots & \sum_{k=0}^K a_{1N}(k)z^{-k} \\ \vdots & & \vdots \\ \sum_{k=0}^K a_{P1}(k)z^{-k} & \cdots & \sum_{k=0}^K a_{PN}(k)z^{-k} \end{pmatrix} \quad (1.7)$$

Notons toutefois les travaux de Cichocki qui supposent l'existence d'une composante auto-régressive dans le processus de mélange [52]. La figure 1.2 représente le processus de mélange sous forme de schéma-bloc.

Si toutes les fonctions de transfert ont un seul coefficient non nul, le mélange est simplement à atténuations et retards et si de plus les fonctions d'une même colonne de $\mathbf{A}(z)$ ont ce coefficient non nul au même indice k , le mélange peut être réécrit de manière linéaire instantanée en décalant si besoin les sources.

La séparation aveugle de sources cherche à estimer les sources $s_j(n)$ formant le vecteur $\mathbf{s}(n)$. Dans le cas d'un mélange convolutif, cette estimation n'est possible qu'à une indétermination de filtrage près. En effet, si l'on prend un filtre quelconque $F(z)$, la contribution d'une source j sur un capteur i (1.1) peut se réécrire dans le domaine en \mathcal{Z} sous la forme

$$X_{ij}(z) = \frac{A_{ij}(z)}{F(z)}(F(z)S_j(z)), \quad (1.8)$$

si bien que l'on peut réécrire (1.5) en remplaçant la source $S_j(z)$ du vecteur $\mathbf{S}(z)$ par $F(z)S_j(z)$ et en remplaçant les éléments $A_{ij}(z)$ de la colonne j de $\mathbf{A}(z)$ par $\frac{A_{ij}(z)}{F(z)}$. Sans d'autres hypothèses sur les sources, il n'y a aucun moyen de lever cette indétermination. Dans le cas de sources i.i.d. (indépendantes et

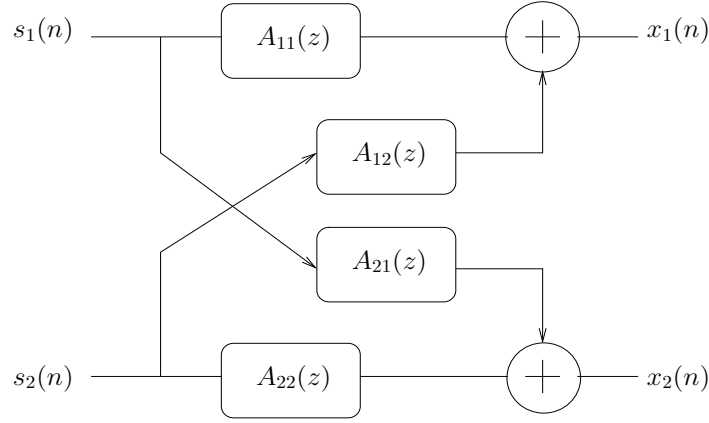


FIG. 1.2 – Schéma-bloc du processus de mélange.

identiquement distribuées), cette indétermination peut être levée en partie. On peut alors identifier les sources à un retard et à un scalaire multiplicatif près (éventuellement complexe) [212].

En fait il est parfois suffisant d'estimer un jeu de filtres de séparation $\mathbf{W}(z)$ qui suppriment les interférences des autres sources introduites par le processus de mélange. Ces filtres peuvent être à réponses impulsionnelles finies ou à réponses impulsionnelles infinies (RII). Le système de séparation RIF causal s'exprime par

$$\forall l = 1 \dots N, y_l(n) = \sum_{i=1}^P \sum_{k=0}^{K'} w_{li}(k) x_i(n-k) \quad (1.9)$$

ce qui peut s'écrire dans le domaine en \mathcal{Z}

$$\mathbf{Y}(z) = \mathbf{W}(z)\mathbf{X}(z). \quad (1.10)$$

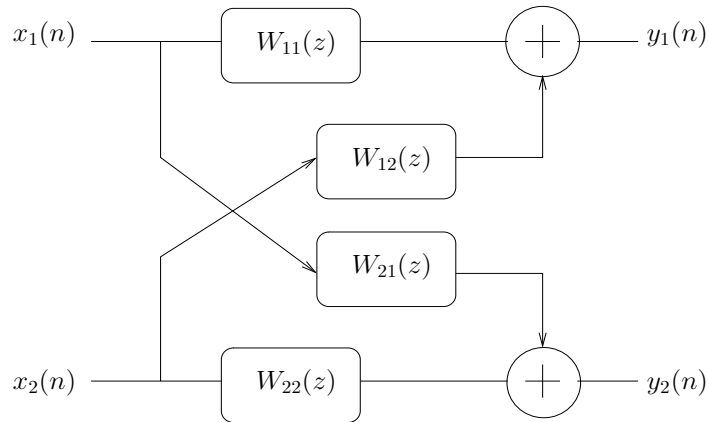


FIG. 1.3 – Schéma-bloc d'un dispositif de séparation direct.

Le système de séparation de la figure 1.3 est qualifié de direct [97]. Il existe aussi des systèmes de séparation de type récurrent dont les sources estimées sont données par les équations suivantes :

$$\forall l = 1 \dots N, y_l(n) = x_l(n) + \sum_{i=1}^P \sum_{k=0}^{K'} u_{li}(k) y_i(n-k) \quad (1.11)$$

ce qui peut être exprimé matriciellement dans le domaine en \mathcal{Z} par

$$\mathbf{Y}(z) = \mathbf{X}(z) + \mathbf{U}(z)\mathbf{Y}(z), \quad (1.12)$$

d'où on déduit

$$\mathbf{Y}(z) = (\mathbf{I} - \mathbf{U}(z))^{-1} \mathbf{X}(z) \quad (1.13)$$

sous réserve que $(\mathbf{I} - \mathbf{U}(z))^{-1}$ existe et que tous ses pôles soient dans le cercle unité. La matrice de séparation 'directe' équivalente est alors

$$\mathbf{W}(z) = (\mathbf{I} - \mathbf{U}(z))^{-1} \quad (1.14)$$

La figure 1.4 représente le dispositif de séparation récurrent sous forme de schéma-bloc².

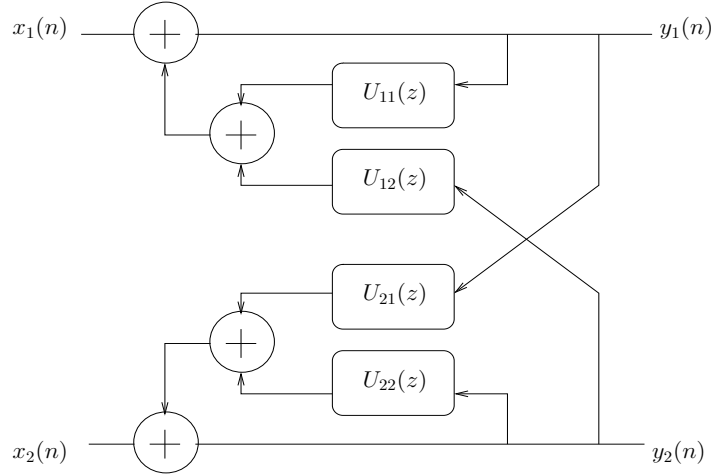


FIG. 1.4 – Schéma-bloc d'un dispositif de séparation récurrent.

Les réseaux directs et récurrents peuvent être combinés pour donner un réseau hybride où une structure directe est suivie d'une structure récurrente [50, 51]. Dans cette thèse, nous avons développé des méthodes basées uniquement sur les réseaux directs qui sont les plus utilisés en raison de leur stabilité, qui est toujours vérifiée contrairement aux réseaux récurrents.

Les équations de mélange et de séparation dans le domaine temporel et dans le domaine en \mathcal{Z} sont regroupées dans le tableau 1.1.

²Le plus souvent, les filtres directs $\mathbf{U}_{ii}(z)$ sont choisis égaux à des filtres nuls.

Processus	Domaine temporel	Domaine en \mathcal{Z}
Mélange RIF causal	$\forall i = 1..P, x_i(n) = \sum_{j=1}^N \sum_{k=0}^K a_{ij}(k) s_j(n-k)$	$\mathbf{X}(z) = \mathbf{A}(z)\mathbf{S}(z)$
Séparation directe	$\forall l = 1..N, y_l(n) = \sum_{i=1}^P \sum_{k=0}^{K'} w_{li}(k) x_i(n-k)$	$\mathbf{Y}(z) = \mathbf{W}(z)\mathbf{X}(z)$
Séparation récurrente	$\forall l = 1..N, y_l(n) = x_l(n) + \sum_{i=1}^P \sum_{k=0}^{K'} u_{li}(k) y_i(n-k)$	$\mathbf{Y}(z) = (\mathbf{I} - \mathbf{U}(z))^{-1} \mathbf{X}(z)$

TAB. 1.1 – Tableau de synthèse des équations de mélange et de séparation.

En pratique, on cherche souvent à estimer des versions filtrées particulières des sources qui sont les contributions sur les différents capteurs. Dans le cas de signaux i.i.d. (en général des signaux de télécommunications), on cherche à estimer les sources non filtrées à un retard et à un facteur d'échelle près.

L'autre indétermination concerne l'ordre des sources séparées : on ne sait pas à quel indice de source j correspond un signal séparé i . Cette indétermination n'existe en fait que si l'on a numéroté de façon arbitraire les signaux sources inconnus (dans le cas contraire, on peut ordonner les sources à posteriori de la même façon que les signaux séparés associés). Cette numérotation des sources à estimer est toutefois nécessaire pour les méthodes fréquentielles comme nous le verrons.

1.3 Hypothèses et critères de séparation

Les hypothèses généralement admises par les méthodes de séparation de mélanges convolutifs appartenant à la classe de l'Analyse en Composantes Indépendantes sont les suivantes :

Hypothèse 1 Les sources sont centrées. En pratique on peut toujours se ramener à un mélange de sources centrées en centrant les observations.

Hypothèse 2 Les signaux considérés sont stationnaires et ergodiques de sorte que l'on peut estimer les espérances par des moyennages temporels.

Hypothèse 3 La matrice de mélange $\mathbf{A}(z)$ est inversible à gauche et la matrice inverse peut être approchée à l'aide de filtres RIF éventuellement non causaux d'ordres suffisamment grands.

Hypothèse 4 Les sources sont statistiquement indépendantes, c.-à-d. la densité de probabilité conjointe du vecteur signal $(s_1(n-d_1), \dots, s_N(n-d_N))$ est égale au produit des densités marginales :

$$p(s_1(n-d_1), \dots, s_N(n-d_N)) = \prod_{i=1}^N p(s_i(n-d_i)), \quad \forall d_1, \dots, d_N \quad (1.15)$$

Hypothèse 5 Les sources à estimer sont non-gaussiennes, sauf une éventuellement.

Hypothèse 6 Les sources à estimer résultent de la coloration de processus d'innovation par filtrage RIF d'ordre suffisamment grand.

L'Hypothèse 3 exige que les sources soient situées à des endroits différents de l'espace (ou au moins qu'elles émettent dans des directions différentes). Elle implique aussi d'avoir un nombre d'observations P supérieur ou égal au nombre de sources N . Nous éviterons cependant cette hypothèse dans la seconde partie de ce manuscrit, où l'on cherchera à effectuer une séparation partielle de sources non-stationnaires en présence de sources de bruit stationnaires.

La majorité des critères de séparation de sources pour mélanges convolutifs ont été utilisés d'abord par des méthodes applicables aux mélanges linéaires instantanés. Nous détaillons ici les principaux critères qui peuvent être classés en 7 catégories. Les 4 premières catégories sont historiquement les plus anciennes et exigent les hypothèses classiques relatives à l'ACI décrites plus haut. Les trois suivantes regroupent des critères où certaines des hypothèses précédentes sont remplacées par des hypothèses spécifiques : non-corrélation, coloration des sources, non-stationnarité, parcimonie ou positivité.

1.3.1 Statistiques croisées d'ordre supérieur

L'annulation de statistiques croisées d'ordre supérieur est un critère très utilisé en séparation de sources. En effet, une façon d'exprimer l'indépendance des sorties du dispositif de séparation, si celles-ci sont centrées et à densités symétriques, est d'annuler tous les moments croisés d'ordre impair

$$E \{y_i(n)^\alpha y_j(n - \tau)^\beta\} = 0, \forall i \neq j, \forall \tau, \forall \alpha, \beta = 1, 3, 5, 7 \dots \quad (1.16)$$

En fait, il n'est pas nécessaire d'annuler tous ces moments croisés pour réaliser la séparation. En lien avec ce critère, on peut utiliser deux familles de statistiques qui quantifient l'indépendance des sorties :

1) Les moments croisés non-linéaires sont définis à partir de deux fonctions f et g par la formule³

$$C_{(f,g)}(i, j) = E \{f(y_i(n))g(y_j(n))\}, \forall i \neq j \quad (1.17)$$

et ont donné lieu aux premiers algorithmes de séparation de sources développés par J. Héroult, C. Jutten et B. Ans [88–90, 104].

³Pour ne pas se ramener à un simple critère de décorrélation, f et g ne doivent pas être toutes les deux linéaires.

Pour annuler la fonction $C_{(f,g)}(i, j)$, ils utilisent l'algorithme adaptatif

$$c_{ij}(n+1) = c_{ij}(n) + \mu_n f(y_i(n))g(y_j(n)) \quad (1.18)$$

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

Pour la séparation de mélanges convolutifs, l'algorithme de Héroult et Jutten a été transposé dans le domaine fréquentiel par A. Back et A. Tsoi [20], approche qui a ensuite été approfondie dans [111]. H. L. Nguyen-Thi et C. Jutten [105, 139] ont aussi proposé un algorithme convolutif dans le domaine temporel qui recherche l'annulation des moments croisés

$$E \{ f(y_i(n))g(y_j(n-k)) \}, \forall i \neq j, \forall k = 0 \dots Q. \quad (1.19)$$

Leur approche suppose d'avoir un mélange à phase minimale ayant des filtres unités entre les sources et observations de même indice. Cette méthode a été étudiée de façon approfondie par C. Servière [185] et une étude sur la séparation des mélanges convolutifs par les réseaux de type Héroult-Jutten a été faite par Berthommier et Choi [31].

Les fonctions f et g les plus utilisées sont $f(x) = x^3$ et $g(y) = y$ en raison de leur simplicité. L'utilisation de ces fonctions f et g revient à l'annulation du moment croisé (3,1) des signaux de sortie, qui ne permet que la séparation de sources *globalement sous-gaussiennes* [46].

Pour des mélanges convolutifs à phase minimale, N. Charkani et Y. Deville ont montré qu'en optimisant les fonctions f et g , il est possible de séparer des sources *sur-gaussiennes* et *sous-gaussiennes* [45–47].

La décorrélation non-linéaire est équivalente pour certains choix de f et g à d'autres critères de séparation. Ainsi, en choisissant dans l'équation 1.17 pour fonctions f et g les deux jeux de fonctions ($f = \text{score}_i$, $g = \text{id}$) et ($f = \text{id}$, $g = \text{score}_j$), où score_k , $k \in \{i, j\}$ est la fonction *score* de la source k , la décorrélation non-linéaire est équivalente à la maximisation de vraisemblance [97].

2) Le cumulant croisé d'ordre 4, $Cum(y_i, y_j, y_j, y_i)$ est un autre critère d'indépendance utilisé pour la première fois par J. L. Lacoume et P. Ruiz [112] puis par P. Comon [54]. Comme le moment croisé (3,1), il est nul dans le cas de deux signaux y_i et y_j indépendants mais aucune preuve de consistance du critère n'a été établie [189].

Une approche plus générale d'utilisation des cumulants croisés est basée sur l'analyse tensorielle. Considérons le tenseur \mathbf{T} défini par

$$\mathbf{T}_{ijkl} = Cum(y_i, y_j, y_k, y_l) \quad (1.20)$$

\mathbf{T} 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{T} est réalisée en pratique grâce à l'identification de matrices propres pour la transformation linéaire

$$\mathbf{F}_{ij}(\mathbf{M}) = \sum_{kl} m_{kl} Cum(y_i, y_j, y_k, y_l). \quad (1.21)$$

On peut montrer que les matrices propres associées à des composantes indépendantes s'écrivent sous la forme $\mathbf{w}_i \mathbf{w}_i^t$ où \mathbf{w}_i est égal à une colonne de la matrice de mélange blanchie, et que leur valeurs propres associées sont les kurtosis des composantes indépendantes.

Une façon rapide d'obtenir ces matrices propres est la méthode de puissance qui consiste à appliquer itérativement la transformation \mathbf{F} jusqu'à la convergence. On peut montrer que cette méthode est équivalente à l'algorithme FastICA basé sur le kurtosis.

Une autre méthode est basée sur la propriété suivante : si on suppose les sorties y_i, y_j, y_k et y_l indépendantes entre elles, l'image de toute matrice \mathbf{M} suivant (1.21) doit être diagonale ce qui permet d'effectuer la séparation par une diagonalisation conjointe approchée de plusieurs matrices $\mathbf{F}(\mathbf{M}_r)$ définies par (1.21). L'algorithme JADE [44], en choisissant pour matrices \mathbf{M}_r les matrices propres du tenseur \mathbf{T} [97], se ramène à la minimisation du critère

$$J_{JADE} = \sum_{(ijkl) \neq (iikl)} \text{Cum}(y_i, y_j, y_k, y_l)^2 \quad (1.22)$$

Dans [151] l'algorithme JADE a été appliqué dans le domaine fréquentiel pour séparer des mélanges convolutifs.

Une autre méthode, FOBI [41], très liée à JADE, consiste à diagonaliser la matrice de corrélation pondérée

$$\mathbf{\Gamma} = E \{ \mathbf{z}(n) \mathbf{z}(n)^t \| \mathbf{z}(n) \|^2 \}. \quad (1.23)$$

où $\mathbf{z}(n)$ est une version blanchie et normalisée des observations. On peut montrer qu'une matrice réalisant cette diagonalisation est nécessairement une matrice de séparation. Pour que cette méthode s'applique, il faut cependant que toutes les sources aient des kurtosis normalisés différents.

A. Ferréol et al. ont étendu la méthode SOBI à l'ordre 4 permettant ainsi l'identification de mélanges sous-déterminés (FOBIUM [77]).

1.3.2 La vraisemblance

M. Gaeta et J. L. Lacoume en 1990 [79] puis D. T. Pham et al. [162] deux années plus tard ont introduit en séparation de sources le critère du maximum de vraisemblance qui a été largement utilisé depuis en séparation instantanée [85, 86, 155, 159, 161] et convolutive [14, 62, 137, 148, 149].

Nous cherchons ici les paramètres de mélange qui maximisent la probabilité d'occurrence des observations. A partir de l'équation matricielle de mélange

$$\mathbf{x}(n) = \mathbf{A} \mathbf{s}(n), \quad (1.24)$$

on peut exprimer la densité de probabilité du vecteur $\mathbf{x}(n)$ en fonction des densités de probabilité des sources et du déterminant de l'inverse \mathbf{B} de la

matrice de mélange \mathbf{A}

$$p_x(\mathbf{x}) = |\det \mathbf{B}| \prod_i p_i(s_i) \quad (1.25)$$

ce qui peut aussi être exprimé en fonction du vecteur d'observation et des lignes \mathbf{b}_i de \mathbf{B} par

$$p_x(\mathbf{x}) = |\det \mathbf{B}| \prod_i p_i(\mathbf{b}_i \mathbf{x}(n)). \quad (1.26)$$

Si on dispose de T échantillons et que l'on considère qu'ils sont indépendants, la vraisemblance V de l'obtention du vecteur $\mathbf{x}(n)$ s'écrit

$$V = \prod_{n=1}^T |\det \mathbf{B}| \prod_i p_i(\mathbf{b}_i \mathbf{x}(n)). \quad (1.27)$$

En général, on considère plutôt la log-Vraisemblance définie par le logarithme de la vraisemblance

$$\log V = \sum_n \sum_i \log(p_i(\mathbf{b}_i \mathbf{x}(n))) + T \log |\det \mathbf{B}| \quad (1.28)$$

ce qui donne en divisant chaque terme par T et en supposant que le moyennage temporel est égal à l'espérance mathématique (ergodicité)

$$\begin{aligned} \frac{1}{T} \log V &= \sum_i E \{ \log(p_i(\mathbf{b}_i \mathbf{x}(n))) \} + \log |\det \mathbf{B}| \\ &= - \sum_i H(\mathbf{b}_i \mathbf{x}(n)) + \log |\det \mathbf{B}|, \end{aligned} \quad (1.29)$$

où $H(y(n))$ est l'entropie différentielle du signal $y(n)$ définie par

$$H(y(n)) = - \int p_y(\eta) \log(p_y(\eta)) d\eta. \quad (1.30)$$

Comme pour les critères d'information mutuelle et de néguentropie que nous allons présenter ensuite, la difficulté majeure du critère de vraisemblance vient du fait que l'on ne connaît pas la densité de probabilité des sources et des observations. On peut alors soit les supposer connues à priori, soit supposer qu'elles appartiennent à une famille donnée de distributions. Nous verrons plus loin que le critère du maximum de vraisemblance est équivalent au critère d'information mutuelle et que si on suppose que le déterminant de \mathbf{B} est rendu constant par une décorrélacion et une normalisation des observations, la vraisemblance est aussi équivalente à la néguentropie (voir critères de non-gaussianité).

J. F. Cardoso a montré [42] que la maximisation de vraisemblance est équivalente à la maximisation de l'entropie de $\mathbf{g}(\mathbf{y})$ où $g_1 \dots g_N$ réalisent l'approximation des fonctions de répartition des sources. Cette approche appelée *Infomax* a été utilisée [13, 23, 231] pour la séparation de mélanges linéaires instantanés et pour la séparation des mélanges convolutifs [115, 173, 192, 210, 211, 225].

1.3.3 L'information mutuelle

Pour des signaux indépendants, la densité de probabilité conjointe est égale au produit des densités marginales. L'information mutuelle est un critère qui mesure la similitude entre la densité conjointe et le produit des densités marginales.

On peut mesurer la similitude entre $p(\mathbf{y})$ et $\prod_i p_i(y_i)$ en calculant leur divergence de Kullback-Leibler définie par

$$I(y_1, \dots, y_N) = \text{div} K \left(p(\mathbf{y}), \prod_i p_i(y_i) \right) = \int_{\mathbf{y}} p(\mathbf{y}) \log \left(\frac{p(\mathbf{y})}{\prod_i p_i(y_i)} \right) d\mathbf{y} \quad (1.31)$$

Comme l'argument du logarithme dans (1.31) vaut 1 partout si et seulement si la densité de probabilité mutuelle est égale presque partout au produit des densités marginales, l'information mutuelle est nulle uniquement pour des variables y_i indépendantes [43].

La divergence de Kullback-Leibler de la densité conjointe et du produit des densités marginales du signal \mathbf{y} est égale à la différence entre la somme des entropies différentielles des variables marginales y_i et l'entropie différentielle du vecteur \mathbf{y} :

$$I(y_1, \dots, y_N) = \sum_i H(y_i) - H(\mathbf{y}) \quad (1.32)$$

Si cette information mutuelle est nulle, les variables ne portent aucune information commune.

Sachant que pour une transformation linéaire $\mathbf{y} = \mathbf{B}\mathbf{x}$

$$H(\mathbf{y}) = H(\mathbf{x}) + \log|\det \mathbf{B}|, \quad (1.33)$$

on obtient

$$I(y_1, \dots, y_N) = \sum_i H(y_i) - H(\mathbf{x}) - \log|\det \mathbf{B}|. \quad (1.34)$$

Comme $H(\mathbf{x})$ est constante par rapport à la matrice de séparation \mathbf{B} à estimer, la comparaison de (1.29) et (1.34) donne

$$I(y_1, \dots, y_N) = -\frac{1}{T} \log V + C^{te}. \quad (1.35)$$

L'information mutuelle I est donc égale à une constante près à l'opposé du critère de log-Vraisemblance.

Le critère d'information mutuelle a été introduit pour la première fois en ACI par P. Comon [57] puis par K. Matsuoka et M. Kawamoto [130] pour la séparation de mélanges linéaires instantanés. Ce critère a ensuite donné lieu à de nombreux travaux, en particulier de D. T. Pham [154, 156].

On peut citer aussi les travaux de K. E. Hild, D. Erdogmus et al. [75,91,92] ainsi que de S. Samadi, M. Babaie-Zadeh, C. Jutten et K. Nayebi [19,181].

Pour la séparation de mélanges convolutifs, l'information mutuelle a été peu utilisée. On peut citer toutefois quelques approches temporelles [93, 124, 153, 157, 174] et l'approche fréquentielle de D.T. Pham [158].

1.3.4 La non-gaussianité

Un autre critère très utilisé en séparation de sources est la non-gaussianité. Ce critère est fondamental en particulier dans le cas de sources blanches stationnaires puisque la non-gaussianité des sources est alors nécessaire pour pouvoir effectuer la séparation. En effet, les signaux résultant d'un mélange orthogonal de signaux gaussiens indépendants sont aussi indépendants ce qui interdit toute possibilité d'identification par un critère d'indépendance. La séparation de sources par maximisation de la non-gaussianité est basée sur un phénomène très intuitif : une combinaison linéaire de signaux non-gaussiens de même distribution est davantage gaussienne que les signaux d'origine, la gaussianité d'un signal étant définie comme l'écart de la distribution de ce signal par rapport à la distribution d'un signal gaussien de même puissance. On peut donc séparer les sources d'un mélange linéaire instantané en maximisant la non-gaussianité du signal de sortie obtenu par une combinaison linéaire des observations.

Le premier critère pour évaluer la non-gaussianité fut historiquement l'autocumulant d'ordre 4 ou kurtosis (non-normalisé) [61] qui a été utilisé auparavant en déconvolution aveugle de sources en faisant en sorte que le signal de sortie soit de puissance unité. Il est défini pour des signaux centrés à partir des moments d'ordres 2 et 4 par l'équation :

$$\text{cum}(y(n), y(n), y(n), y(n)) = E \{y^4(n)\} - 3E \{y^2(n)\}^2 \quad (1.36)$$

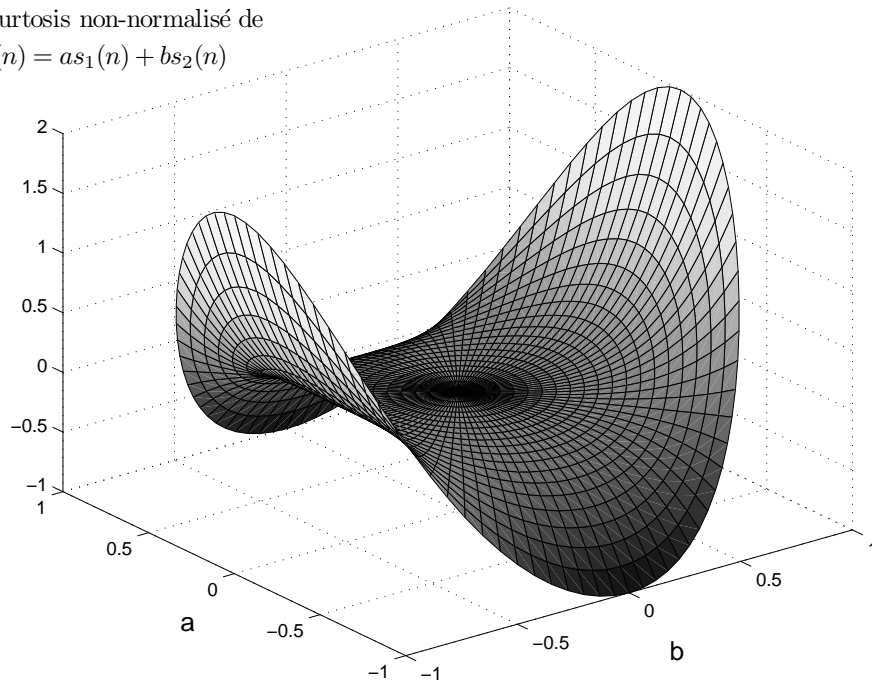
Nous avons représenté en haut de la figure 1.5 le kurtosis non-normalisé d'une combinaison linéaire $y(n) = as_1(n) + bs_2(n)$ de deux signaux $s_1(n)$ et $s_2(n)$ de puissance unité et de kurtosis non-normalisés valant respectivement -1 et 2 . Nous voyons que les extremums locaux du kurtosis de $y(n)$ sur le bord de la surface définie par l'équation $a^2 + b^2 = 1$ sont les 4 points $(a, b) \in \{(-1, 0), (0, -1), (1, 0), (0, 1)\}$ qui correspondent aux points séparants du cercle unité.

L'inconvénient majeur du kurtosis est que son estimation est très sensible aux valeurs extrêmes car le moment d'ordre 4 est alors mal estimé. On utilise donc prioritairement ce critère pour des sources sous-gaussiennes où les "outliers" sont peu présents.

Une famille de critères plus récente et qui évite ce problème est basée sur l'entropie différentielle définie par (1.30).

Un résultat fondamental de théorie de l'information établit que parmi toutes les distributions, la distribution gaussienne est celle qui a la plus grande

kurtosis non-normalisé de
 $y(n) = as_1(n) + bs_2(n)$



kurtosis normalisé de
 $y(n) = as_1(n) + bs_2(n)$

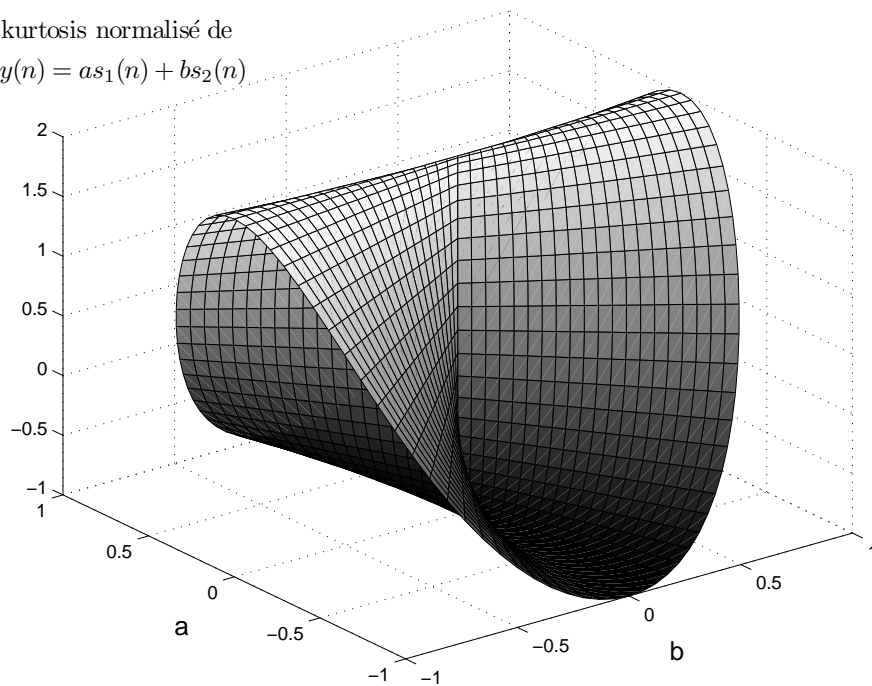


FIG. 1.5 – Kurtosis non-normalisé (en haut) et kurtosis normalisé (en bas) de la combinaison linéaire $y(n) = as_1(n) + bs_2(n)$ de deux signaux $s_1(n)$ et $s_2(n)$ de puissance unité et de kurtosis non-normalisés (ou normalisés) valant -1 et 2.
 -Abscisses et ordonnées : coefficients a et b de la combinaison linéaire.
 -Hauteur : valeur du kurtosis.

entropie et peut donc être considérée comme la plus aléatoire. Au contraire, à variance égale, les distributions très structurées (très plates ou très piquées) auront une entropie faible. Un critère de non-gaussianité peut donc être obtenu en comparant l'entropie du signal de sortie normalisé $y(t)$ avec l'entropie d'une variable gaussienne $y_{gauss}(t)$ de puissance unité. On nomme néguentropie ce critère défini par

$$neg(y(n)) = H(y_{gauss}(n)) - H(y(n)) \quad (1.37)$$

où $H(y(n))$ est l'entropie différentielle de $y(n)$ définie par (1.30). En considérant l'équation (1.29) on montre [97] que si le déterminant de \mathbf{B} est rendu constant par une décorrélation et une normalisation des observations (\mathbf{B} est alors orthogonale), la maximisation de la vraisemblance et donc la minimisation de l'information mutuelle (1.29) sont équivalentes à la maximisation de la somme des néguentropies des signaux de sortie. Cependant ces critères ne sont pas parfaitement équivalents car dans le cas de la néguentropie, que l'on peut définir pour une unique sortie, il est possible d'extraire les sources une à une, ce qui est impossible pour la log-Vraisemblance ou l'information mutuelle.

La difficulté de ce critère vient du fait qu'il est difficilement estimable, faisant intervenir les densités de probabilité des sources (estimées en pratique à partir des densités de probabilité des observations) inconnues a priori. On utilise donc en général une approximation définie à partir d'une fonction G non quadratique [95] :

$$neg(y(n)) \approx \left(E \left\{ G(y_{gauss}(n)) \right\} - E \left\{ G(y(n)) \right\} \right)^2 \quad (1.38)$$

Les méthodes basées sur l'optimisation de la non-gaussianité demandent une normalisation, soit du critère, soit du signal extrait. En effet le kurtosis par exemple augmente avec le facteur d'échelle que l'on applique devant un signal. Pour pallier ce problème on peut normaliser le critère du kurtosis et définir le kurtosis normalisé par

$$kurt_y^{nor}(n) = \frac{kurt_y(n)}{(E \{y^2(n)\})^2}. \quad (1.39)$$

Nous avons représenté sur la figure 1.5 en bas le kurtosis normalisé d'une combinaison linéaire $y(n) = as_1(n) + bs_2(n)$ de deux signaux $s_1(n)$ et $s_2(n)$ de puissance unité et de kurtosis non-normalisés (ou normalisés) valant respectivement -1 et 2 . Les extremums locaux du kurtosis normalisé correspondent aux droites $(a, 0)$ et $(0, b)$ où $a, b \in [-1, 1] \setminus 0$. Ce kurtosis normalisé, contrairement au kurtosis standard, est indépendant du facteur d'échelle.

On peut aussi imposer au signal extrait d'être de norme unité. Le procédé utilisé consiste alors à décorréler et à normaliser les observations afin que la normalisation de la sortie $y(n)$ puisse s'effectuer en normalisant simplement le vecteur d'extraction de source appliqué au vecteur d'observations décorrélé et normalisé.

La non-gaussianité a été utilisée pour la séparation de mélanges linéaires instantanés par P. Comon qui a proposé comme critère la somme des valeurs au carré des kurtosis des signaux de sortie ainsi qu'un algorithme pour l'optimiser (COM2 [55–57]). E. Moreau et O. Macchi ont ensuite proposé comme critère la somme des kurtosis des sorties [136], critère qui ne permet la séparation que dans le cas où les kurtosis des sources sont de même signe. P. Comon et E. Moreau ont par la suite donné une solution analytique à l'optimisation de ce critère (COM1 [59]).

N. Delfosse et P. Loubaton [61] ont ensuite proposé un algorithme à déflation qui estime itérativement chaque source en optimisant la valeur absolue ou le carré du kurtosis d'une unique sortie. Hyvärinen et Oja ont par la suite développé une famille d'algorithmes de type point-fixe appelée FastICA utilisant au choix le kurtosis ou la néguentropie [96, 98] pour estimer les sources soit simultanément par une approche symétrique, soit itérativement par une approche à déflation.

La non-gaussianité a aussi été largement utilisée pour la séparation de mélanges convolutifs, majoritairement en utilisant le critère du kurtosis, non-normalisé [122, 141, 190] ou normalisé [40, 191, 212–214]. En raison du nombre de coefficients de filtres d'extraction à estimer, les algorithmes d'optimisation proposés pour ces méthodes sont très lents en général, même en remplaçant l'algorithme à gradient par un algorithme de Newton [6].

Notons aussi l'existence de méthodes fréquentielles qui, grâce à l'utilisation de la transformée de Fourier à court terme, reformulent le mélange de manière linéaire instantanée [110, 165]. Nous verrons dans la prochaine section que ce type d'approche présente de nombreux inconvénients.

L'objectif de la première partie de cette thèse a été de développer un algorithme temporel rapide à point fixe convolutif utilisant au choix le kurtosis ou la néguentropie suivant le type de sources à séparer.

1.3.5 Statistiques croisées du second ordre

Ce type de critère exige d'autres hypothèses que celles évoquées en début de chapitre. Les hypothèses 4 et 5 sont en effet remplacées par l'hypothèse de non-corrélation alliée aux hypothèses de coloration ou de non-stationnarité des sources. Nous savons que la décorrélation de signaux de sortie entre eux n'est en général pas suffisante pour effectuer la séparation de sources [97]. En effet, si on multiplie le vecteur source par une matrice orthogonale quelconque (unitaire dans la cas complexe), le vecteur résultant est aussi composé de signaux décorrélés mais non séparés.

On peut cependant effectuer la séparation en supposant une des deux hypothèses suivantes :

1) Si on suppose que les sources sont colorées, il est possible de séparer les sources grâce à l'annulation de l'intercorrélacion des signaux de sortie plus

ou moins décalés entre eux :

$$E \{y_i(n)y_j(n-\tau)\} = 0, \forall i \neq j, \forall \tau \in \mathcal{T} \quad (1.40)$$

où \mathcal{T} est un ensemble de valeurs entières.

Ce critère n'est pas applicable aux sources blanches puisque dans ce cas, $E \{s_i(n)s_i(n-\tau)\}$ est nul quelle que soit la valeur de τ non nulle et donc $E \{\mathbf{y}(n)\mathbf{y}(n-\tau)^t\} = \mathbf{G}E \{\mathbf{s}(n)\mathbf{s}(n-\tau)^t\} \mathbf{G}^t = \mathbf{0}$ quelle que soit la matrice de performances \mathbf{G} définie par le produit des matrices de mélange et de séparation.

Les méthodes utilisant ce critère de séparation cherchent en pratique à diagonaliser simultanément les matrices

$$\mathbf{C}_\tau^\mathbf{x} = E \{\mathbf{x}(n)\mathbf{x}(n-\tau)^t\} \quad (1.41)$$

correspondant à plusieurs valeurs de τ .

Dans le cas de AMUSE [207, 208], on cherche à diagonaliser conjointement $\mathbf{C}_0^\mathbf{x}$ et $\mathbf{C}_\tau^\mathbf{x}$ où $\tau \neq 0$. Pour cela, l'algorithme AMUSE réalise d'abord un blanchiment des observations qui donne un nouveau vecteur d'observations $\mathbf{z}(n)$ vérifiant $\mathbf{C}_0^\mathbf{z} = \mathbf{I}$, puis diagonalise la matrice $\mathbf{C}_\tau^\mathbf{z}$. Une version de AMUSE pour mélanges convolutifs a été proposée [198].

L'algorithme SOBI développé par Belouchrani et al. [24] cherche lui à diagonaliser de façon approchée $\mathbf{C}_0^\mathbf{x}$ et plusieurs matrices $\mathbf{C}_\tau^\mathbf{x}$. Comme pour AMUSE, une étape de blanchiment précède la diagonalisation conjointe des matrices $\mathbf{C}_\tau^\mathbf{z}$. Une version convolutive temporelle de SOBI [37] ainsi qu'une version temps-fréquence [25] ont été proposées par la même équipe.

2) On peut aussi supposer que les sources étudiées sont non-stationnaires [131, 160, 163, 164, 184, 220, 224], en travaillant éventuellement dans le domaine fréquentiel.

Les statistiques qui varient dans le temps fournissent alors de l'information supplémentaire. Dans ce cas, la décorrélation des sources entre elles à plusieurs instants n permet d'effectuer la séparation :

$$E \{y_i(n)y_j(n)\} = 0, \forall i \neq j, \forall n \quad (1.42)$$

En effet, en raison de la non-stationnarité à l'ordre 2 la covariance de $\mathbf{y}(n)$ dépend de l'instant n ce qui engendre de nouvelles contraintes par rapport au blanchiment et permet même la séparation de sources gaussiennes. Ce critère (1.42) (éventuellement transposé dans le domaine fréquentiel) a été étendu au cas convolutif dans le cadre de méthodes temporelles [38, 107, 108] et surtout fréquentielles [38, 146, 147, 163, 164, 170–172].

1.3.6 La parcimonie

L'utilisation de la parcimonie des sources est très populaire actuellement en séparation aveugle de sources (on parle de *Sparse Component Analysis*) [84]. En

général, les observations sont transposées dans une représentation améliorant le caractère parcimonieux des sources, c.-à-d. leur nullité dans certaines zones de la représentation, et permettant donc une meilleure estimation des paramètres de mélange (on peut citer l'analyse temps-fréquence, l'analyse temps-échelle et l'analyse multi-résolution). Contrairement à beaucoup d'autres approches, elle permet l'estimation de sources gaussiennes et/ou corrélées ainsi que la séparation des mélanges sous-déterminés.

On distingue trois types d'approches tirant parti de la parcimonie des sources :

1) Une méthode consiste à identifier séparément les colonnes de la matrice de mélange grâce à des techniques de clustering appliquées au nuage de points des observations pour estimer la matrice de mélange [35,36,94,119,120]. Plusieurs techniques de clustering peuvent être employées [228] pour identifier les vecteurs colonnes (K-means [132], K-medians [200], C-means [228]). Une méthode de ce type pour mélanges linéaires à atténuations et retards a été proposée par P. Bofill [33].

2) Une autre méthode suppose l'orthogonalité disjointe des sources qui est une hypothèse relativement restrictive sur les sources. Elle suppose en effet que dans chaque case du plan temps-fréquence, une seule source soit active. On peut alors en calculant des rapports d'observations identifier les paramètres d'amplitude et de phase de la matrice de mélange puis reconstituer les sources à l'aide de masques temps-fréquence binaires. L'algorithme DUET utilise ce principe et permet la séparation de mélanges linéaires à atténuations et retards [21, 22, 101, 175, 177–179, 229]. Des versions convolutives de DUET ont été développées [32, 133, 134]. Notons l'existence d'approches similaires à DUET, par exemple l'utilisation par C. Choi [49] de l'algorithme de clustering K-means et l'utilisation de la "basis pursuit" basée sur la norme ℓ^q par R. Saab [180]. Récemment, une approche de type sous-espace a été proposée par Aissa-El-Bey et al., qui permet la reconstruction des sources lorsque le nombre de sources actives à chaque point temps-fréquence est strictement inférieur au nombre d'observations [17,18]. Une approche bayésienne de DUET a aussi été étudiée pour ce cas où plusieurs sources sont actives dans une même zone temps-fréquence [128,129].

3) D'autres méthodes prennent elles en compte la similitude des observations entre elles pour effectuer l'estimation de la matrice de mélange dans des zones privilégiées du plan temps-fréquence où une seule source est active. L'algorithme TIFROM [7,9,195] estime cette similitude en calculant des variances de rapports d'observations dans le plan temps-fréquence (n, ω) :

$$J_{TIFROM} = Var \left\{ \frac{X_i(n, \omega)}{X_j(n, \omega)} \right\} \quad (1.43)$$

Si cette variance est faible dans une zone du plan temps-fréquence, on peut

montrer que la zone en question est mono-source ce qui permet l'identification d'une colonne de la matrice de mélange par moyennage de rapports d'observations. Une version améliorée de cet algorithme a été développée spécifiquement pour les mélanges à atténuations et retards [166, 168].

L'algorithme TIFCORR [66, 71, 196] utilise le coefficient de corrélation⁴ entre 2 observations dans le plan temps-fréquence :

$$J_{TIFCORR} = \frac{|R_{x_i x_j}(n, \omega)|^2}{R_{x_i x_i}(n, \omega) R_{x_j x_j}(n, \omega)}, \quad (1.44)$$

où, si $n_1 \cdots n_L$ sont les instants de la fenêtre temporelle d'indice n ,

$$R_{x_i x_j}(n, \omega) = \frac{1}{L} \sum_{p=1}^L x_i(n_p, \omega) x_j^*(n_p, \omega). \quad (1.45)$$

Si le module de $J_{TIFCORR}$ est proche de 1 dans une zone temps fréquence, les deux observations contiennent majoritairement une source dans cette zone, ce qui permet là encore d'identifier une colonne de la matrice de mélange. Chaque coefficient de mélange est cette fois estimé en calculant le rapport entre l'intercorrélation de deux observations et l'autocorrélation de l'une d'elles. Une version de TIFCORR utilisant un clustering de type médiane (K-medians) pour l'identification de la matrice de mélange a été proposée [169] ainsi qu'une version pour mélanges à atténuations et retards [167].

TIFCOHERE [10, 11] utilise une fonction statistique appelée fonction de cohérence définie par le rapport entre le module au carré de la densité interspectrale de puissance et le produit des densités (auto-)spectrales de puissance segmentées temporellement.

$$J_{TIFCOHERE} = \frac{|S_{x_i x_j}(n, \omega)|^2}{S_{x_i x_i}(n, \omega) S_{x_j x_j}(n, \omega)} \quad (1.46)$$

Là encore, une valeur de $J_{TIFCOHERE}$ élevée (proche de 1), indique que l'on a une zone mono-source permettant l'identification d'une colonne de la matrice de mélange. TIFCOHERE identifie ensuite chaque coefficient de mélange par le calcul du rapport entre la densité interspectrale de puissance de deux observations et la densité autospectrale de l'une d'elles. Une version à atténuations et retards a aussi été proposée [10] ainsi qu'une version convolutive [12].

Notons que ces trois algorithmes TIFFROM, TIFCORR et TIFCOHERE sont transposables à d'autres domaines d'analyse que le domaine temps-fréquence. Une version temps-échelle de TIFCORR a ainsi été proposée [68] récemment.

⁴Les auteurs ont proposé deux versions de TIFCORR ; l'une centre les observations dans la fenêtre d'analyse au préalable, et l'autre non. Nous ne présentons ici que la version non centrée.

M. Xiao, S. Xie et Y. Fu ont proposé une méthode de détection des zones mono-sources basée sur une normalisation du vecteur d'observation à chaque instant et sur la dérivation temporelle de ce nouveau vecteur. Si la dérivée du vecteur normalisé est proche de 0 dans une zone du domaine d'analyse, cela indique qu'une seule source est présente dans cette zone et que l'on peut y identifier une colonne de la matrice de mélange (par une approche de type clustering ou par le calcul de rapports d'observations) [226].

1.3.7 La positivité

La positivité (ou non-négativité), applicable lorsque les matrices des échantillons des sources et de mélange sont à coefficients réels positifs ou nuls conduit à une technique récente de séparation aveugle de sources développée dans le milieu des années 90 [1, 2, 15, 113, 114, 145].

La factorisation en matrices non-négatives (*NMF : Non-negative Matrix Factorization* en anglais) consiste à décomposer la matrice des observations \mathbf{X} de dimension $P \times T$, P et T étant respectivement les nombres d'observations et d'échantillons, sous la forme

$$\mathbf{X} = \mathbf{W}\mathbf{H} \quad (1.47)$$

et sous contrainte de positivité des coefficients des matrices \mathbf{W} et \mathbf{H} , identifiées respectivement comme les matrices de mélange et de sources.

La factorisation en matrices non-négatives n'est en général pas unique. Toute matrice \mathbf{B} inversible permet en effet de réécrire l'équation (1.47) sous la forme :

$$\mathbf{X} = \mathbf{W}\mathbf{B}^{-1}\mathbf{B}\mathbf{H} \quad (1.48)$$

Si les matrices $\mathbf{W}_1 = \mathbf{W}\mathbf{B}^{-1}$ et $\mathbf{H}_1 = \mathbf{B}\mathbf{H}$ sont aussi à coefficients positifs ou nuls, on définit ainsi une nouvelle factorisation NMF de \mathbf{X} . Dans le cas le plus simple, \mathbf{B} s'écrit sous la forme $\mathbf{P}\mathbf{\Delta}$ où $\mathbf{\Delta}$ est une matrice diagonale positive constituée de facteurs d'échelle et \mathbf{P} est une matrice de permutation.

Il existe plusieurs types de NMF qui diffèrent par la mesure de distance utilisée pour évaluer l'écart entre \mathbf{X} et $\mathbf{W}\mathbf{H}$, et éventuellement par le type de régularisation employée. Deux distances relativement simples ont été proposées par Lee et Seung : la norme de Frobenius de la différence $\mathbf{X} - \mathbf{W}\mathbf{H}$, définie par

$$\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 = \sum_{ij} (X_{ij} - (WH)_{ij})^2 \quad (1.49)$$

et la divergence entre \mathbf{X} et $\mathbf{W}\mathbf{H}$ définie par

$$D(\mathbf{X} \mid \mathbf{W}\mathbf{H}) = \sum_{ij} (X_{ij} \log(\frac{X_{ij}}{(WH)_{ij}}) - X_{ij} + (WH)_{ij}). \quad (1.50)$$

En adaptant les coefficients de \mathbf{W} et \mathbf{H} de façon itérative [113, 114], ces deux critères donnent deux algorithmes de NMF différents. Les deux mises à jour

proposées par Lee et Seung pour ces critères basés sur la norme de Frobenius (1.49) et sur la divergence (1.50) sont respectivement

$$\begin{cases} H_{a\mu} \leftarrow H_{a\mu} \frac{(W^t X)_{a\mu}}{(W^t W H)_{a\mu}} \\ W_{ia} \leftarrow W_{ia} \frac{(X H^t)_{ia}}{(W H H^t)_{ia}} \end{cases} \quad (1.51)$$

et

$$\begin{cases} H_{a\mu} \leftarrow H_{a\mu} \frac{\sum_i W_{ia} X_{i\mu} / (W H)_{i\mu}}{\sum_k W_{ka}} \\ W_{ia} \leftarrow W_{ia} \frac{\sum_\mu H_{a\mu} X_{i\mu} / (W H)_{i\mu}}{\sum_\nu W_{a\nu}} \end{cases} \quad (1.52)$$

En interprétant les sources comme des spectres de puissance à séparer et en adaptant les équations de Lee et Seung (1.51-1.52), des extensions convolutives de la NMF ont été proposées récemment [193, 194, 219].

1.4 Classification des méthodes de SAS convolutives

On distingue deux grandes classes de méthodes pour la séparation de mélanges convolutifs : les méthodes fréquentielles qui utilisent la transformée de Fourier pour se ramener à un problème de séparation aveugle de sources instantanée et les méthodes temporelles qui effectuent la séparation au moyen de filtres dits d'extraction estimés dans le domaine temporel. Dans cette section, nous présentons les principes et les limitations respectives de ces deux types d'approche⁵.

1.4.1 Les méthodes fréquentielles

Les méthodes fréquentielles [125] simplifient le modèle de mélange convolutif en utilisant la transformée de Fourier. La relation de convolution (1.2) dans le domaine temporel devient dans le domaine fréquentiel

$$\dot{\mathbf{x}}(f) = \dot{\mathbf{A}}(f) \dot{\mathbf{s}}(f) \quad (1.53)$$

En pratique, on utilise la transformée de Fourier à court terme [7, 71, 99, 135, 146, 147, 163, 164, 170–172, 182, 183, 192, 229], pour obtenir plusieurs échantillons par fréquence mais aussi pour tirer parti de la variabilité temporelle des signaux sources, par exemple leur non-stationnarité [146, 147, 163, 164, 170–172] ou leur

⁵Le tableau 1.2 en fin de section récapitule les avantages et inconvénients des deux types de méthodes.

parcimonie dans le cadre d'approches déterministes [7, 71, 229] (on montre que la transformée de Fourier à court terme est optimale pour les signaux de parole [176]). On obtient alors l'équation de mélange suivante :

$$\dot{\mathbf{x}}(f, t) = \dot{\mathbf{A}}(f) \dot{\mathbf{s}}(f, t) \quad (1.54)$$

Par ces méthodes fréquentielles, on obtient donc pour chaque fréquence un mélange linéaire instantané complexe que l'on peut séparer en utilisant les mêmes critères qu'en séparation de sources instantanée. Ces méthodes fréquentielles présentent cependant trois difficultés majeures :

- Les indéterminations d'échelle et de permutation propres à la séparation de sources font qu'un vecteur source obtenu pour une fréquence donnée peut être pondéré et ordonné de façon différente de celui obtenu à une autre fréquence. La reconstruction des signaux dans le domaine temporel nécessite de corriger cet effet.
- En utilisant la transformée de Fourier à court terme, on obtient beaucoup moins de points pour chaque fréquence, ce qui rend difficile l'estimation des statistiques [16]. Pour certains algorithmes en effet, il est nécessaire que la taille de fenêtre utilisée par la transformée de Fourier à court terme soit beaucoup plus grande que l'ordre des filtres de mélange, ce qui engendre un faible nombre d'échantillons par fréquence. Une méthode qui pallie ce problème a été proposée par C. Servière [186].
- Un autre problème vient du fait que l'équation (1.53) n'est parfaitement valide que si le mélange est modélisé en utilisant l'opérateur de convolution circulaire alors que le mélange réel réalise une convolution linéaire. Ce problème a été étudié par plusieurs auteurs qui ont montré [16, 147] que l'approximation n'est acceptable que si la longueur des fenêtres temporelles est suffisamment grande par rapport à l'ordre des filtres de mélange (plus grande d'un facteur 2 au moins).

Notons que le problème des permutations évoqué plus haut peut être évité si une unique source active dans un intervalle de temps permet d'estimer les paramètres de mélange de cette source pour toutes les fréquences [10, 12].

Un algorithme fréquentiel de type point fixe directement adapté de FastICA a été développé par R. Prasad et al. [165] pour la séparation de mélanges convolutifs de signaux de parole.

Dans la première partie de cette thèse, nous considérons les mélanges de signaux stationnaires et cherchons à éviter le problème des indéterminations d'échelle et de permutation bande à bande en introduisant pour cela une méthode temporelle.

1.4.2 Les méthodes temporelles

Les méthodes temporelles à structure directe cherchent à estimer une source ou une de ses contributions à l'aide d'une combinaison convolutive des signaux d'observations :

$$y(n) = \sum_{k=1}^P h_k(n) * x_k(n). \quad (1.55)$$

Par rapport à la séparation de mélanges linéaires instantanés, on a donc remplacé l'estimation d'un vecteur d'extraction composé de scalaires par l'estimation d'un vecteur de P réponses impulsionnelles de filtres $h_k(n)$ associés aux P observations. L'estimation est donc rendue plus difficile en raison du grand nombre d'inconnues, en particulier si l'ordre élevé du mélange impose de choisir des filtres d'extraction $h_k(n)$ d'ordre élevé.

La séparation convolutive de sources dans le domaine temporel est très liée aux problèmes de déconvolution aveugle [72, 100, 187, 188, 221], la déconvolution cherchant à estimer à l'aide d'un filtre une source unique i.i.d. ayant subi un filtrage par le milieu de propagation.

De plus, historiquement, un des premiers critères de séparation de sources (pour mélanges instantanés), l'autocumulant d'ordre 4 ou kurtosis fut introduit par Wiggins en 1978 pour la déconvolution aveugle de sources sismiques [221]. Il s'est inspiré du théorème de la limite centrale et du fait qu'une combinaison linéaire de sources aléatoires indépendantes non-gaussiennes de même distribution est davantage gaussienne que les sources d'origine. Le critère qu'il propose pour mesurer cette non-gaussianité, le kurtosis normalisé (1.39), est maximum en valeur absolue lorsque le filtre appliqué à l'observation restitue le signal i.i.d. à l'entrée du canal de propagation (le filtre de déconvolution est alors égal à l'inverse du filtre du canal, à un retard et un facteur multiplicatif complexe près).

De la même façon, N. Delfosse et P. Loubaton ont montré [61] en séparation de mélanges linéaires instantanés que la valeur absolue ou le carré du kurtosis (non-normalisé) du signal de sortie normalisé est maximisée lorsque les coefficients d'extraction utilisés correspondent à un facteur près à une ligne de l'inverse de la matrice de mélange.

Ce lien entre séparation de sources et déconvolution a induit des méthodes temporelles [212–214], dédiées aux mélanges convolutifs et utilisant le critère du kurtosis normalisé. Cependant, les méthodes proposées étaient lentes du fait du nombre important de paramètres à estimer et de l'utilisation d'algorithmes de type gradient pour optimiser le critère. Un autre algorithme d'optimisation, utilisé par F. Abrard et Y. Deville [6, 8] et basé lui sur l'algorithme de Newton, est plus rapide mais reste très lent lorsque l'ordre choisi pour les filtres d'extraction est élevé. D'autre part, la consistance du critère de séparation n'a été prouvée que pour des filtres d'extraction infinis ou pour des filtres d'ordre

fini dans des cas particuliers (pour des filtres de mélange AR par exemple). S. Douglas et al. [73, 74] ont proposé en parallèle à nous un algorithme de type point fixe pour mélanges convolutifs fonctionnant dans le domaine temporel. L'article ne donne pas de preuve de convergence et décrit une procédure de blanchiment spatio-temporelle itérative dont la rapidité n'est pas étudiée. De plus ses performances n'ont pas été comparées à d'autres algorithmes convolutifs utilisant les mêmes hypothèses de stationnarité et de non-gaussianité des sources comme l'algorithme de Tugnait [212] par exemple.

1.5 Conclusion

Dans ce chapitre, nous avons présenté un état de l'art de la séparation aveugle de sources. Après avoir répertorié les grands types de critères existant pour la séparation aveugle de mélanges linéaires instantanés et convolutifs, nous avons classé les algorithmes convolutifs en deux grandes catégories. Les méthodes fréquentielles, qui utilisent la transformée de Fourier pour reformuler les mélanges sous forme linéaire instantanée, souffrent à chaque fréquence des indéterminations d'échelle et de permutation bande à bande. Les méthodes temporelles demandent elles l'estimation de filtres dits d'extraction contenant un grand nombre de coefficients à estimer, ce qui fait que ces méthodes sont lentes et moins utilisées en pratique que les méthodes fréquentielles bien qu'elles n'imposent pas de résoudre les indéterminations d'échelle et de permutation bande à bande. Nous proposons dans la suite de cette partie un algorithme temporel de type point fixe pour mélanges convolutifs permettant la séparation rapide de mélanges convolutifs RIF.

Méthodes temporelles		Méthodes fréquentielles	
Avantages	Inconvénients	Avantages	Inconvénients
<ul style="list-style-type: none"> • L'hypothèse d'indépendance est mieux vérifiée dans le domaine temporel. • La convergence vers le point optimal est plus précise. 	<ul style="list-style-type: none"> • Les performances sont dégradées en acoustique dans un environnement à forte réverbération. 	<ul style="list-style-type: none"> • Les mélanges convolutifs sont transformés en mélanges instantanés à chaque fréquence. • Grâce à la FFT, les calculs sont simplifiés par rapport à une implémentation dans le domaine temporel. 	<ul style="list-style-type: none"> • A chaque fréquence, il y a une ambiguïté d'échelle et de permutation qui doit être résolue. • L'utilisation de la transformée de Fourier à court terme réduit le nombre d'échantillons à chaque fréquence. • Lorsque l'on a peu d'échantillons à chaque fréquence, l'indépendance est mal estimée en pratique. • La modélisation sous forme de convolution circulaire dégrade les performances de séparation.

TAB. 1.2 – Tableau récapitulatif des avantages et inconvénients respectifs des méthodes de séparation temporelles et fréquentielles [150].

Chapitre 2

Extension des méthodes de SAS basées sur la non-gaussianité

2.1 Introduction

Ce chapitre présente une nouvelle méthode de séparation aveugle de sources pour les mélanges convolutifs RIF de processus MA. Elle consiste en une extension dans le domaine temporel de l'algorithme FastICA développé par Hyvärinen et Oja pour les mélanges linéaires instantanés. Nous proposons d'utiliser un blanchiment convolutif non-causal des observations ainsi qu'une initialisation particulière des paramètres d'extraction afin d'employer des algorithmes d'optimisation de type point fixe, associés aux critères de kurtosis ou de néguentropie, pour estimer les processus d'innovation des sources. Les contributions des sources sur les capteurs sont ensuite estimées en utilisant un critère quadratique correspondant à la puissance de la différence entre une observation donnée et une version filtrée non-causale du processus d'innovation d'une source.

2.2 Position du problème

Comme indiqué dans le chapitre précédent, la séparation aveugle de source (SAS) consiste à estimer un jeu de N sources inconnues à l'aide de P observations constituées de mélanges de ces sources quand les paramètres de mélange sont inconnus. Notons $\mathbf{s}(n) = [s_1(n), \dots, s_N(n)]^t$ le vecteur des sources et $\mathbf{x}(n) = [x_1(n), \dots, x_P(n)]^t$ le vecteur des observations. Nous considérons ici les mélanges convolutifs définis par un jeu de filtres inconnus dont les réponses impulsionnelles sont $a_{ij}(n)$, où $i = 1, \dots, P$ et $j = 1, \dots, N$. Les relations entre les sources et les observations peuvent être écrites dans le domaine temporel sous la forme

$$x_i(n) = \sum_{j=1}^N \sum_{k=-\infty}^{\infty} a_{ij}(k) s_j(n-k), \quad \forall i = 1, \dots, P. \quad (2.1)$$

La relation globale dans le domaine en \mathcal{Z} est alors

$$\mathbf{X}(z) = \mathbf{A}(z).\mathbf{S}(z) \quad (2.2)$$

où $\mathbf{X}(z)$ et $\mathbf{S}(z)$ sont respectivement les transformées en \mathcal{Z} de $\mathbf{x}(n)$ et $\mathbf{s}(n)$, et la matrice $\mathbf{A}(z)$ de dimension $P \times N$ est composée des fonctions de transfert $A_{ij}(z) = \mathcal{Z} [a_{ij}(n)]$ des filtres de mélange.

Dans cette partie, chaque source $s_j(n)$ est supposée être un processus MA et s'exprimer dans le domaine en \mathcal{Z} suivant

$$S_j(z) = F_j(z).U_j(z) \quad (2.3)$$

où $F_j(z)$ correspond à un filtre RIF causal et $U_j(z)$ est la transformée en \mathcal{Z} d'un processus $u_j(n)$ qui est le processus d'innovation de $s_j(n)$.

En notant $\mathbf{U}(z) = [U_1(z) \quad \dots \quad U_N(z)]^t$, nous pouvons exprimer l'équation de mélange (2.2) suivant

$$\mathbf{X}(z) = \mathbf{H}(z).\mathbf{U}(z) \quad (2.4)$$

où $\mathbf{H}(z) = \mathbf{A}(z).\mathbf{F}(z)$ avec $\mathbf{F}(z) = \begin{pmatrix} F_1(z) & & 0 \\ & \ddots & \\ 0 & & F_N(z) \end{pmatrix}$.

Nous faisons les hypothèses suivantes concernant le modèle de mélange présenté ci-dessus :

- Le processus $\mathbf{u}(n) = [u_1(n), \dots, u_N(n)]^t$ est réel, de moyenne nulle, temporellement i.i.d. et spatialement indépendant, c.-à-d. ses composantes $u_j(n)$ sont statistiquement indépendantes entre elles mais n'ont pas nécessairement la même distribution. Nous supposons qu'une composante au plus est gaussienne.
- Les matrices de filtres $\mathbf{F}(z)$, $\mathbf{A}(z)$ et donc $\mathbf{H}(z)$ sont causales, RIF et non singulières. Notons que les systèmes RII peuvent aussi être approximés par des modèles équivalents RIF (d'ordre élevé).

Le but de la SAS convolutive est typiquement d'estimer les contributions de toutes les sources dans chaque observation, c.-à-d. $A_{ij}(z).S_j(z)$. Dans les méthodes à déflation telles que [212], ceci est réalisé en utilisant la procédure suivante :

1. Extraire le processus d'innovation $u_j(n)$ d'une source $s_j(n)$ à partir des observations.
2. Identifier P filtres de coloration et les appliquer à $u_j(n)$ pour estimer les contributions de $s_j(n)$ dans chaque observation, c.-à-d. $A_{ij}(z).S_j(z)$.

3. Retenir la contribution la plus puissante⁶ comme sortie correspondant à l'une des sources.
4. Soustraire les contributions de toutes les observations pour obtenir un nouveau jeu d'observations appelées observations réduites.
5. Enlever l'observation réduite la moins puissante⁷. Si le nombre d'observations réduites subsistant est supérieur à 1, réitérer la procédure à partir de l'étape 1.

Cette procédure est représentée sous forme de schéma-bloc dans la figure 2.1.

Nous considérons ici les méthodes de SAS dans le domaine temporel qui utilisent la non-gaussianité comme critère pour réaliser la première étape de la procédure ci-dessus et qui sont donc basées sur l'Analyse en Composantes Indépendantes (ACI) [98]. Dans la prochaine section, nous analysons les principes et les limites des méthodes existantes et nous proposons dans les sections suivantes une approche pour les étendre afin d'obtenir pour les mélanges convolutifs des méthodes à convergence rapide basées sur le kurtosis ou la néguentropie. Les performances de séparation des méthodes proposées sont présentées dans le chapitre 3 et des conclusions sont ensuite tirées de ces travaux.

2.3 Approches existantes

Comme nous l'avons mentionné au chapitre 1, N. Delfosse et P. Loubaton [61] ont proposé la première méthode de SAS à déflation fondée sur le kurtosis pour les mélanges *linéaires instantanés*, où les filtres $H_{ij}(z)$ sont remplacés par des coefficients scalaires. Cette méthode consiste premièrement à obtenir une version spatialement blanchie et normalisée $\mathbf{z}(n)$ du vecteur d'observation $\mathbf{x}(n)$, c.-à-d. un jeu de combinaisons linéaires de ces observations qui sont mutuellement décorrélées à l'instant n et qui ont des variances unitaires⁸. Un premier signal de sortie est ensuite obtenu par une combinaison linéaire $y(n) = \mathbf{w}^t \mathbf{z}(n)$ des observations décorrélées et normalisées, où \mathbf{w} est un vecteur normalisé de coefficients, sélectionné de manière à maximiser le carré (ou la valeur absolue) du kurtosis *non-normalisé* de $y(n)$ défini par l'équation (1.36) pour un signal à moyenne nulle. Delfosse et Loubaton ont prouvé dans [61] que les maxima locaux de ce critère correspondent aux points de séparation. Ils ont utilisé une méthode de type gradient pour maximiser ce critère. Cela nécessite de sélectionner convenablement le pas d'adaptation et implique de toute façon une convergence lente. Hyvärinen et Oja ont résolu ce problème

⁶Nous choisissons de retenir la contribution la plus puissante car c'est celle qui est la moins perturbée par une interférence ou un bruit de puissance donnée.

⁷Nous choisissons d'enlever l'observation réduite la moins puissante car elle est susceptible de contenir des contributions faibles des sources restant à estimer.

⁸L'annexe A est consacrée à cette première étape de blanchiment et de normalisation des observations.

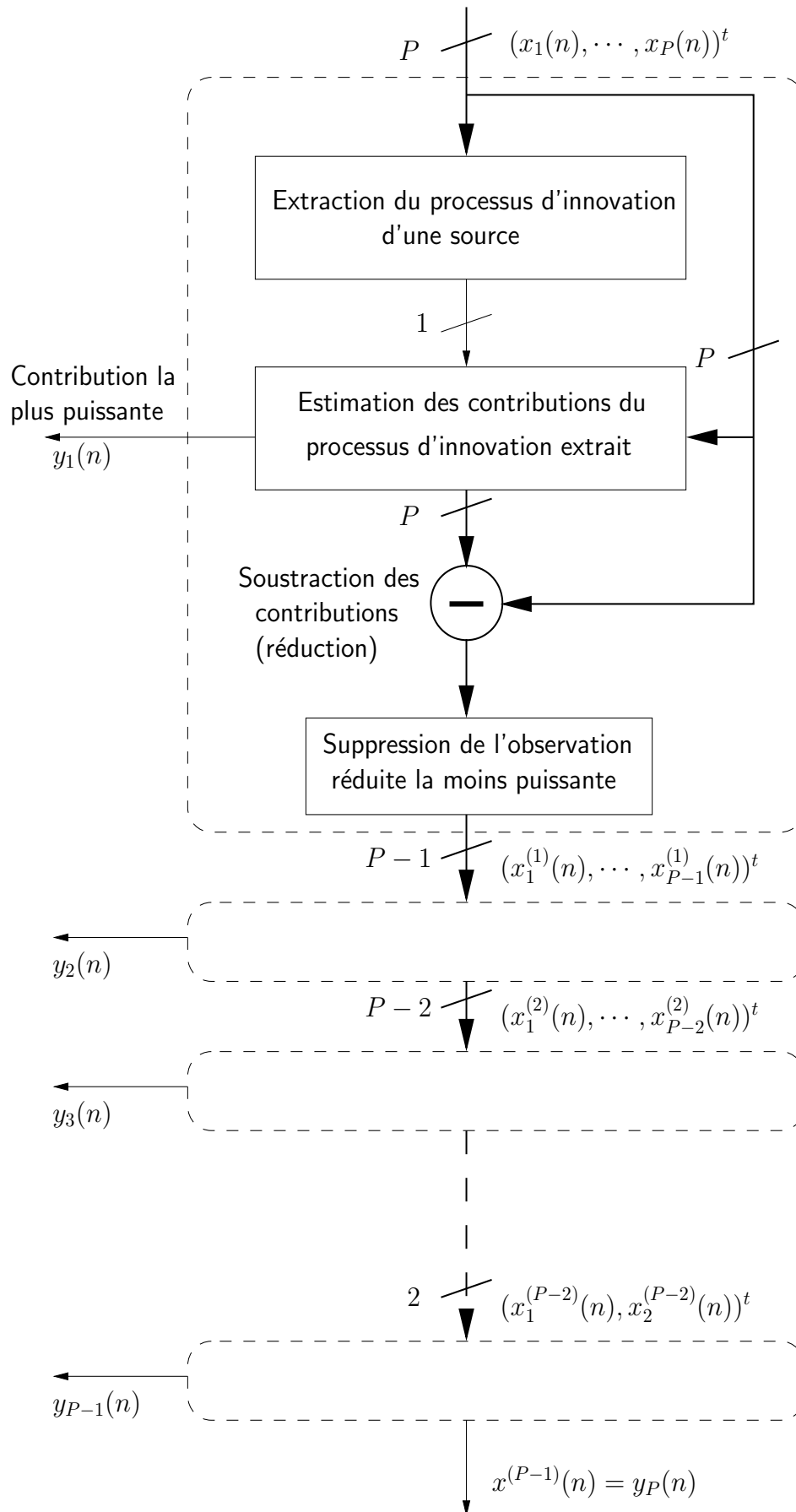


FIG. 2.1 – Schéma-bloc du processus de déflation.

en introduisant un algorithme de type point fixe pour optimiser le critère ci-dessus [98]. Brièvement, cet algorithme tire parti du fait que l'on réalise une optimisation sous contrainte du critère considéré et met à jour itérativement le vecteur \mathbf{w} des coefficients de la combinaison linéaire avec une version normalisée du gradient de ce critère sur la surface de contrainte. Cet algorithme ne comprend aucun paramètre à régler (tel que le pas d'adaptation pour les méthodes de type gradient) et sa convergence est très rapide.

Une approche différente fut proposée par J. K. Tugnait pour les mélanges *convolutifs* [212]. Elle opère directement sur les observations, c.-à-d. sans les blanchir ni les normaliser, mais utilise alors la valeur absolue du kurtosis *non-normalisé* du signal de sortie $y(n)$ défini par (1.39), comme critère de séparation. J. K. Tugnait a prouvé que les points séparants correspondent aux maxima locaux de ce critère quand on recombine les observations à l'aide de filtres d'extraction doublement infinis. Il a proposé d'optimiser ce critère en utilisant une approche de type gradient, ce qui entraîne là encore une convergence lente. Les tests réalisés dans notre équipe [6] ont montré que, même en utilisant un algorithme d'optimisation de type Newton modifié, la convergence reste lente, particulièrement pour les filtres d'ordre élevé.

Ce chapitre a pour but de combler le manque mis en évidence par cet historique, c.-à-d. d'introduire des méthodes rapides à base de kurtosis ou de néguentropie pour les mélanges convolutifs. Dans ce but, nous cherchons comment étendre aux mélanges convolutifs l'approche basée sur un blanchiment et une optimisation de type point fixe du kurtosis non-normalisé, proposée précédemment pour les mélanges linéaires instantanés.

2.4 Nouvelles méthodes pour extraire un processus d'innovation

Toutes les méthodes citées dans la section précédente demandent une normalisation, car le kurtosis non-normalisé de $y(n)$ tend vers l'infini quand la puissance de $y(n)$ tend vers l'infini. Dans l'approche de Tugnait, c'est le critère qui est normalisé car la méthode consiste à estimer un des processus d'innovation $u_j(n)$ à un retard et à un facteur d'échelle près en maximisant la valeur absolue du kurtosis normalisé d'une combinaison convolutive des observations définie par :

$$y(n) = \sum_{p=1}^P k_p(n) * x_p(n) = \sum_{p=1}^P \sum_{r=-R}^R k_p(r) x_p(n-r) \quad (2.5)$$

où $k_p(n)$, $p = 1 \dots P$ sont P filtres RIF non causaux.

Au contraire, les deux approches linéaires instantanées mentionnées dans la

section précédente utilisent le kurtosis non-normalisé et sont basées sur une normalisation de la puissance de $y(n)$:

$$E \{y(n)^2\} = 1. \quad (2.6)$$

Cela est facilement réalisé grâce à étape de blanchiment et de normalisation des observations fournissant le vecteur $\mathbf{z}(n)$, qui entraîne que $y(n) = \mathbf{w}^t \mathbf{z}(n)$ vérifie $E \{y(n)^2\} = \|\mathbf{w}\|^2$, si bien qu'en choisissant \mathbf{w} tel que $\|\mathbf{w}\|^2 = 1$, on garantit que $E \{y(n)^2\} = 1$. Nous étendons ici cette méthode aux mélanges convolutifs. Dans ce but, la première étape de notre approche réalise une orthonormalisation convolutive des observations, définie comme suit.

A chaque instant n , nous considérons le vecteur d'observations étendu défini par

$$\tilde{\mathbf{x}}(n) = [x_1(n+R), \dots, x_1(n-R), \dots, x_P(n+R), \dots, x_P(n-R)]^t \quad (2.7)$$

contenant $\tilde{P} = (2R+1)P$ éléments⁹. Nous en déduisons le vecteur colonne, de dimension \tilde{P} , $\tilde{\mathbf{z}} = [\tilde{z}_1(n), \dots, \tilde{z}_{\tilde{P}}(n)]^t$ défini par

$$\tilde{\mathbf{z}}(n) = \mathbf{B} \tilde{\mathbf{x}}(n) \quad (2.8)$$

où \mathbf{B} est une matrice $\tilde{P} \times \tilde{P}$ choisie de telle sorte que

$$E \{\tilde{z}_i(n) \tilde{z}_j(n)\} = \delta_{ij}, \forall i, j \in \{1, \dots, \tilde{P}\} \quad (2.9)$$

où δ est le symbole de Kronecker, ce qui est équivalent à

$$\mathbf{R}_{\tilde{\mathbf{z}}}(n) = \mathbf{I}. \quad (2.10)$$

où $\mathbf{R}_{\tilde{\mathbf{z}}}(n)$ est la matrice d'autocorrélation du vecteur signal $\tilde{\mathbf{z}}(n)$. La figure 2.2 illustre cet étage de blanchiment dit convolutif.



FIG. 2.2 – Schéma-bloc du processus de blanchiment convolutif.

Par rapport à $\tilde{\mathbf{x}}(n)$, l'opération (2.8) peut être considérée comme un blanchiment et une normalisation conventionnels, ce blanchiment consistant en une Analyse en Composantes Principales (ACP). Mais par rapport aux observations d'origine $x_i(n)$, cela peut être interprété différemment. En effet, les éq. (2.7) et (2.8) montrent que les signaux $\tilde{z}_i(n)$ sont des mélanges convolutifs des

⁹Certaines méthodes à sous-espaces utilisent ce principe d'étalement des vecteurs signaux [3–5, 83, 127] (nous utiliserons aussi dans la suite une version étalée du vecteur source).

observations $x_i(n)$. L'éq. (2.9) signifie alors que les signaux $\tilde{z}_i(n)$ sont créés de façon à être décorrés et à variance unité, ce qui peut être vu comme une orthonormalisation spatio-temporelle des observations $x_i(n)$. Notons $y(n)$ le signal extrait

$$y(n) = \mathbf{w}^t \tilde{\mathbf{z}}(n) = \sum_{m=1}^{\tilde{P}} w_m \tilde{z}_m(n). \quad (2.11)$$

En combinant les définitions (2.7), (2.8) et (2.11) de $y(n)$ avec la version temporelle de (2.4), on montre que $y(n)$ est une combinaison linéaire de versions décalées des processus $u_k(n)$ c.-à-d.

$$y(n) = \sum_{m=1}^{\tilde{P}} w_m \sum_{j=1}^{\tilde{P}} b_{mj} \tilde{x}_j(n) \quad (2.12)$$

$$= \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd} u_q(n-d) \quad (2.13)$$

où v_{qd} , D_{min} et D_{max} sont obtenus à partir des ordres des filtres RIF impliqués dans (2.4) et (2.5).

En utilisant les propriétés du kurtosis pour des variables aléatoires indépendantes, nous déduisons à partir de (2.12)

$$\begin{aligned} kurt_y(n) &= \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd}^4 kurt_{u_q}(n-d) \\ &= \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd}^4 \alpha_{qd}, \end{aligned} \quad (2.14)$$

où

$$\alpha_{qd} = kurt_{u_q}(n-d) \quad (2.15)$$

Calculons maintenant la puissance de $y(n)$ en tenant compte de la linéarité de l'opérateur de puissance pour des variables indépendantes

$$pow_y(n) = \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd}^2 pow_{u_q}(n-d). \quad (2.16)$$

Chaque processus $u_q(n)$ est supposé être identiquement distribué à tous les instants $n-d$ où d varie de D_{min} à D_{max} . Les termes $pow_{u_q}(n-d)$ dans (2.16) ne dépendent donc pas de d , c.-à-d. que (2.16) implique seulement un terme de puissance $pow_{u_q}(n)$ pour chaque source, valable quel que soit le décalage d appliqué au processus $u_q(n)$. Ces puissances peuvent en outre être remises à l'échelle avec un facteur positif grâce l'indétermination d'échelle propre à la

SAS. Ainsi, elles peuvent être supposées égales à 1 de telle sorte que (2.16) devient

$$pow_y(n) = \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd}^2. \quad (2.17)$$

Calculons maintenant d'une autre manière la puissance de $y(n)$:

$$\begin{aligned} pow_y(n) &= E \{y(n)^2\} \\ &= \mathbf{w}^t \mathbf{R}_{\tilde{\mathbf{z}}}(n) \mathbf{w} \\ &= \|\mathbf{w}\|^2 \end{aligned} \quad (2.18)$$

d'après (2.10), ce qui entraîne grâce à (2.17)

$$\|\mathbf{w}\|^2 = \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd}^2. \quad (2.19)$$

Imposer à \mathbf{w} d'être sur la sphère unité $\|\mathbf{w}\| = 1$ permet donc de contraindre de façon simple notre vecteur $\{v_{qd}, 1 \leq q \leq N, d \in [D_{min}, \dots, D_{max}]\}$ à être sur la sphère unité

$$\sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd}^2 = 1 \quad (2.20)$$

de dimension $N \times D$ où $D = D_{max} - D_{min} + 1$.

L'optimisation de la valeur absolue du kurtosis $kurt_y(n)$ défini par (2.14), sous la contrainte (2.6) (réalisable par la normalisation de \mathbf{w}) appartient donc au même type de problème que l'optimisation de

$$f(v_q, q = 1..N) = \sum_{q=1}^N v_q^4 \alpha_q \quad (2.21)$$

sous la contrainte

$$\sum_{q=1}^N v_q^2 = 1 \quad (2.22)$$

avec un jeu de variables ici notées v_{qd} , au lieu du jeu de variables v_q dans le cas instantané (l'indice des retards d est absent dans le cas des mélanges instantanés). Le type de fonction défini dans (2.21) a été largement étudié dans le contexte des méthodes de SAS instantanée.

En appliquant les résultats de [61, 97], on garantit que les maxima de $|kurt_y(n)|$ sur la sphère unité correspondent aux points tels qu'une seulement des variables considérées v_{qd} est non nulle (égale à ± 1)¹⁰. L'éq. (2.12) montre

¹⁰Les mélanges convolutifs entraînent une approximation concernant les points que l'on peut atteindre, comme on l'expliquera dans la section suivante.

que le signal se compose alors seulement d'une version pondérée et décalée $e_l(n)$ du processus $u_l(n)$ correspondant.

Des algorithmes performants pour réaliser une optimisation sous contrainte de la valeur absolue du kurtosis de $y(n)$ peuvent ensuite être directement déduits des approches précédentes pour les mélanges linéaires instantanés. En effet, la section suivante montre que les mélanges convolutifs $\tilde{\mathbf{x}}(n)$ étudiés peuvent être reformulés comme des mélanges instantanés sous certaines conditions. Nous proposons donc comme extension de FastICA [98] l'algorithme convolutif à point fixe et kurtosis suivant, basé sur notre vecteur modifié \mathbf{w} :

- Initialiser \mathbf{w} à une valeur \mathbf{w}_0 , par exemple en utilisant les approches présentées ci-dessous.
- Répéter les étapes 1) et 2) jusqu'à la convergence

$$1) \quad \mathbf{w} = E \{ \tilde{\mathbf{z}}(n) (\mathbf{w}^t \tilde{\mathbf{z}}(n))^3 \} - 3\mathbf{w} \quad (2.23)$$

$$2) \quad \mathbf{w} = \mathbf{w} / \|\mathbf{w}\| \quad (2.24)$$

La valeur initiale \mathbf{w}_0 de \mathbf{w} mentionnée précédemment peut être sélectionnée de façon aléatoire. Une approche améliorée peut être obtenue en tirant parti du fait qu'il existe une relation entre notre vecteur \mathbf{w} et les coefficients des filtres RIF de l'approche de Tugnait définie par (2.5). En effet, considérons les colonnes de notre matrice \mathbf{B} et indiquons les comme suit

$$\mathbf{B} = [\mathbf{b}_1^{-R}, \dots, \mathbf{b}_1^{+R}, \dots, \mathbf{b}_P^{-R}, \dots, \mathbf{b}_P^{+R}], \quad (2.25)$$

en nous basant sur (2.7) et (2.8). En utilisant (2.8), le signal (2.11) extrait par notre méthode s'écrit $y(n) = \mathbf{w}^t \mathbf{B} \tilde{\mathbf{x}}(n) = \mathbf{w}^t \sum_{p=1}^P \sum_{r=-R}^R \mathbf{b}_p^r x_p(n-r)$. En identifiant cette expression avec le signal de sortie (2.5) dans l'approche de Tugnait, nous obtenons

$$k_p(r) = \mathbf{w}^t \mathbf{b}_p^r, \quad \forall p \in \{1, \dots, P\}, \quad \forall r \in \{-R, \dots, R\} \quad (2.26)$$

et donc

$$\mathbf{K} = \mathbf{w}^t \mathbf{B} \quad (2.27)$$

où le vecteur ligne \mathbf{K} est composé des coefficients des réponses impulsionnelles des filtres $k_1(n), \dots, k_P(n)$. Cette relation nous permet d'initialiser notre vecteur \mathbf{w} comme dans la méthode de Tugnait, c.-à-d. avec des filtres unités $k_p(n) = \delta(n)$, de sorte que $y(n)$ soit la somme des observations $x_p(n)$. Cela revient à prendre $\mathbf{K} = \mathbf{K}_0$ défini par

$$\mathbf{K}_0 = \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{k_1}, \dots, \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{k_P}^t. \quad (2.28)$$

L'éq. (2.27) entraîne alors

$$\mathbf{w}_0^t = \mathbf{K}_0 \mathbf{B}^{-1}. \quad (2.29)$$

Cette initialisation de \mathbf{w} nous donne de meilleurs résultats expérimentaux qu'une initialisation aléatoire et est utilisée dans le chapitre 3.

Pour les mélanges linéaires instantanés, au lieu d'utiliser le kurtosis, une autre fonction de contraste basée sur la néguentropie a été proposée par Hyvärinen pour estimer la non-gaussianité [96]. Elle a démontré sa meilleure robustesse et sa plus faible variance par rapport à l'approche fondée sur le kurtosis. En particulier, elle est plus robuste aux valeurs extrêmes que celle basée sur le kurtosis qui fait appel à un moment d'ordre 4, dont l'estimation est très sensible aux valeurs extrêmes. De plus, un algorithme à point fixe rapide et fiable a aussi été développé par Hyvärinen pour ce type de fonction. Nous étendons cet algorithme aux mélanges convolutifs de la même manière que l'approche fondée sur le kurtosis, en utilisant la formule d'adaptation suivante à la place de (2.23) à chaque itération :

$$1) \quad \mathbf{w} = E \{ \tilde{\mathbf{z}}(n) g(\mathbf{w}^t \tilde{\mathbf{z}}(n)) \} - E \{ g'(\mathbf{w}^t \tilde{\mathbf{z}}(n)) \} \mathbf{w} \quad (2.30)$$

où g et g' sont les dérivées première et seconde d'une fonction non-quadratique G utilisée pour estimer la néguentropie.

L'algorithme à kurtosis proposé peut être optimisé en termes de temps de calcul nécessaire pour estimer l'espérance dans (2.23). En effet, (2.23) nécessite de calculer un produit matrice-vecteur $\tilde{\mathbf{z}} = \mathbf{B}\tilde{\mathbf{x}}(n)$ pour chaque indice d'échantillon n afin d'en déduire la somme des termes $\tilde{\mathbf{z}}(\mathbf{w}^t \tilde{\mathbf{z}})^3$ utilisée pour l'estimation de l'espérance. En calculant le vecteur $\mathbf{w}_1 = \mathbf{w}^t \mathbf{B}$ une seule fois pour l'itération donnée de l'algorithme et en remplaçant $\tilde{\mathbf{z}}$ par $\mathbf{B}\tilde{\mathbf{x}}$ dans l'équation de mise à jour (2.23), nous obtenons

$$1) \quad \mathbf{w} = \mathbf{B} E \{ \tilde{\mathbf{x}}(n) (\mathbf{w}_1^t \tilde{\mathbf{x}}(n))^3 \} - 3\mathbf{w} \quad (2.31)$$

Nous calculons ainsi seulement deux produits matrice-vecteur pour estimer l'espérance, c.-à-d. un pour $\mathbf{w}_1 = \mathbf{w}^t \mathbf{B}$ et un dans (2.31). Le même principe s'applique à l'algorithme (2.30) basé sur la néguentropie.

2.5 Reformulation du mélange sous forme linéaire instantanée

Les P observations $x_i(n)$ considérées s'expriment en fonction des N processus d'innovation $u_j(n)$ suivant (2.4). Elles sont donc des mélanges RIF causaux d'ordre Q de ces processus. Considérons maintenant le vecteur $\tilde{\mathbf{x}}(n)$ défini par (2.7) et composé d'observations décalées. L'analyse fournie par exemple dans [78] implique que $\tilde{\mathbf{x}}(n)$ peut être interprété comme un jeu de mélanges linéaires instantanés de sources qui sont ici des versions décalées et pondérées

des processus d'innovation $u_j(n)$. Ce mélange linéaire instantané s'écrit sous forme matricielle

$$\tilde{\mathbf{x}}(n) = \tilde{\mathbf{H}} \tilde{\mathbf{u}}(n) \quad (2.32)$$

où

- $\tilde{\mathbf{x}}(n)$ est défini par (2.7),
- $\tilde{\mathbf{u}}(n) = [u_1(n+R), \dots, u_1(n-R-Q), \dots, u_N(n+R), \dots, u_N(n-R-Q)]^t$,
- $\tilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{H}}_{11} & \dots & \tilde{\mathbf{H}}_{1N} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{H}}_{P1} & \dots & \tilde{\mathbf{H}}_{PN} \end{bmatrix}$,
- avec $\tilde{\mathbf{H}}_{ij} = \begin{bmatrix} h_{ij}(0) & \dots & h_{ij}(Q) & \dots & 0 \\ & \ddots & & \ddots & \\ 0 & & h_{ij}(0) & \dots & h_{ij}(Q) \end{bmatrix}$, matrice de Toeplitz de dimension $(2R+1) \times (2R+1+Q)$.

On montre alors facilement que si

$$PL \geq N(Q+L) \quad (2.33)$$

où $L = (2R+1)$ est le nombre de retards utilisés pour constituer le vecteur d'observations étendu $\tilde{\mathbf{x}}(n)$, alors le mélange $\tilde{\mathbf{x}}(n) = \tilde{\mathbf{H}} \tilde{\mathbf{u}}(n)$ a au moins autant d'observations que de sources et le mélange reformulé de façon linéaire instantanée est donc (sur-)déterminé.

Donc, si (2.33) est vérifiée, l'analyse pour les mélanges linéaires instantanés fournie dans [61] prouve rigoureusement que, en maximisant la valeur absolue du kurtosis non-normalisé du signal $y(n)$ défini par (2.11) sous la contrainte (2.6), nous extrayons une version décalée et pondérée du processus d'innovation, $\alpha_l u_l(n-r)$, dont l'estimée est notée $e_l(n)$ dans la suite.

Si (2.33) n'est pas vérifiée, le problème de SAS *reformulé en instantané* est sous-déterminé, c.-à-d. qu'il implique moins d'observations que de sources (notons que c'est en particulier le cas quand $P = N$). Cette sous-détermination est liée à l'ordre fini des filtres d'extraction équivalents appliqués aux observations par les étages de traitement (2.7), (2.8) et (2.11). Certaines approximations sont donc nécessaires. Néanmoins, quand le rapport $\frac{PL}{N(Q+L)}$ associé à (2.33) tend vers 1 (ce qui est le cas quand $P = N$ et quand L est grand), un processus d'innovation pondéré et décalé peut encore être estimé précisément à l'aide d'une combinaison linéaire des observations disponibles dont la valeur absolue du kurtosis non-normalisé est maximale sous la contrainte (2.6). L'impact en termes de SIR^{11} , du caractère sous-déterminé du mélange reformulé

¹¹Le rapport signal à interférences est défini par : $SIR = 10 \log_{10}(E\{y(n)^2\} / E\{(y(n) - \hat{y}(n))^2\})$, où $\hat{y}(n)$ est l'estimation de $y(n)$.

sous forme linéaire instantanée est étudié dans la section qui suit.

2.6 Performances atteignables par le modèle linéaire instantané sous-déterminé

Nous avons vu dans la section précédente qu'un mélange convolutif RIF causal peut être réinterprété comme un mélange linéaire instantané comportant $\tilde{P} = PL$ observations et $\tilde{N} = N(Q + L)$ sources où L est le nombre de retards de $\tilde{\mathbf{x}}(n)$ et Q est l'ordre du modèle défini par la somme des ordres de mélange et de coloration des processus d'innovation. Soit \mathbf{w} un vecteur d'extraction de processus d'innovation. Redéfinissons ici ce vecteur pour y inclure l'étape de blanchiment réalisée grâce à la matrice \mathbf{B} , c.-à-d. la notation \mathbf{w} utilisée ici correspond au produit $\mathbf{B}^t \mathbf{w}$ avec la notation précédente. Supposons que ce vecteur extrait approximativement la source de $\tilde{\mathbf{u}}(n)$ d'indice i .

Considérons maintenant Δ , norme de l'écart entre le vecteur de séparation global $\mathbf{w}^t \tilde{\mathbf{H}}$ et le vecteur ligne $\mathbf{b}_i = \underbrace{[0, \dots, 0]_{i-1}}_{i-1}, \underbrace{1, [0, \dots, 0]_{\tilde{N}-i}}_{\tilde{N}-i}$ composé d'un 1 à

l'indice i et de 0 ailleurs :

$$\begin{aligned} \Delta &= \|\mathbf{w}^t \tilde{\mathbf{H}} - \mathbf{b}_i\| \\ &= \|\tilde{\mathbf{H}}^t \mathbf{w} - \mathbf{b}_i^t\|. \end{aligned} \quad (2.34)$$

La formule des moindres carrés [123] définit le vecteur \mathbf{w} optimal minimisant la valeur de Δ :

$$\begin{aligned} \mathbf{w}_{opt} &= ((\tilde{\mathbf{H}}^t)^t \tilde{\mathbf{H}}^t)^{-1} \tilde{\mathbf{H}} \mathbf{b}_i^t \\ &= (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^t)^{-1} \tilde{\mathbf{H}} \mathbf{b}_i^t \end{aligned} \quad (2.35)$$

et donc

$$\mathbf{w}_{opt}^t = \mathbf{b}_i \tilde{\mathbf{H}}^t (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^t)^{-1}. \quad (2.36)$$

Le vecteur de séparation global optimal est donc

$$\mathbf{w}_{opt}^t \tilde{\mathbf{H}} = \mathbf{b}_i \tilde{\mathbf{H}}^t (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^t)^{-1} \tilde{\mathbf{H}} \quad (2.37)$$

et est égal à la ligne i de \mathbf{H}_1 où $\mathbf{H}_1 = \tilde{\mathbf{H}}^t (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^t)^{-1} \tilde{\mathbf{H}}$, de dimension $\tilde{P} \times \tilde{N}$ est le projecteur orthogonal sur $(Ker \tilde{\mathbf{H}})^\perp$ [123].

En appliquant la matrice \mathbf{H}_1 au vecteur source $\tilde{\mathbf{u}}(n)$, on obtient donc un vecteur signal donc le $i^{ème}$ élément correspond à la meilleure estimation au sens des moindres carrés du $i^{ème}$ signal du vecteur source $\tilde{\mathbf{u}}(n)$.

Déterminons maintenant les performances moyennes qui peuvent être obtenues par ce critère des moindres carrés. Soit $\tilde{\mathbf{u}}'(n)$ l'estimation du vecteur $\tilde{\mathbf{u}}(n)$ obtenue à l'aide du projecteur \mathbf{H}_1 par la formule $\tilde{\mathbf{u}}'(n) = \mathbf{H}_1 \tilde{\mathbf{u}}(n)$. Sachant que \mathbf{H}_1 est le projecteur orthogonal sur $(Ker \tilde{\mathbf{H}})^\perp$ de dimension \tilde{P} , la différence $\tilde{\mathbf{u}}(n) - \tilde{\mathbf{u}}'(n)$ est le projeté de $\tilde{\mathbf{u}}(n)$ sur $Ker \tilde{\mathbf{H}}$ qui est de dimension $\tilde{N} - \tilde{P}$.

Calculons la puissance moyenne de l'erreur commise pour l'estimation des \tilde{N} sources :

$$Perr_{moy} = \frac{1}{\tilde{N}} E \{ \|\tilde{\mathbf{u}}(n) - \tilde{\mathbf{u}}'(n)\|^2 \}. \quad (2.38)$$

Soit une \mathcal{B} une base orthonormée adaptée à la décomposition orthogonale

$$\mathbb{R}^{\tilde{N}} = (Ker \tilde{\mathbf{H}})^\perp \oplus Ker \tilde{\mathbf{H}} \quad (2.39)$$

et \mathbf{U} la matrice unitaire de passage de la base canonique à cette base \mathcal{B} .

Notons $\hat{\mathbf{u}}(n)$ le vecteur $\tilde{\mathbf{u}}(n)$ exprimé dans cette nouvelle base. Comme $\tilde{\mathbf{u}}(n) - \tilde{\mathbf{u}}'(n)$ est le projeté de $\tilde{\mathbf{u}}(n)$ sur $Ker \tilde{\mathbf{H}}$,

$$\tilde{\mathbf{u}}(n) - \tilde{\mathbf{u}}'(n) = \mathbf{U}[0, \dots, 0, \hat{u}_{\tilde{P}+1}(n), \dots, \hat{u}_{\tilde{N}}(n)]^t. \quad (2.40)$$

La puissance moyenne de l'erreur commise dans l'estimation des \tilde{N} sources vaut donc

$$\begin{aligned} Perr_{moy} &= \frac{1}{\tilde{N}} E \{ \|\tilde{\mathbf{u}}(n) - \tilde{\mathbf{u}}'(n)\|^2 \} \\ &= \frac{1}{\tilde{N}} E \{ \|[0, \dots, 0, \hat{u}_{\tilde{P}+1}(n), \dots, \hat{u}_{\tilde{N}}(n)] \mathbf{U}^t\|^2 \} \end{aligned} \quad (2.41)$$

et comme \mathbf{U} est unitaire,

$$\begin{aligned} Perr_{moy} &= \frac{1}{\tilde{N}} E \{ \|[0, \dots, 0, \hat{u}_{\tilde{P}+1}(n), \dots, \hat{u}_{\tilde{N}}(n)]\|^2 \} \\ &= \frac{1}{\tilde{N}} E \left\{ \sum_{i=\tilde{P}+1}^{\tilde{N}} \hat{u}_i^2(n) \right\} \\ &= \frac{1}{\tilde{N}} \sum_{i=\tilde{P}+1}^{\tilde{N}} E \{ \hat{u}_i^2(n) \}. \end{aligned} \quad (2.42)$$

Les signaux $\hat{u}_i(n)$ résultant des signaux $\tilde{u}_i(n)$ par changement de base orthonormale, en supposant que les $\tilde{u}_i(n)$ sont de puissance unité, les $\hat{u}_i(n)$ sont aussi de puissance unité et

$$Perr_{moy} = \frac{1}{\tilde{N}} (\tilde{N} - \tilde{P}) \quad (2.43)$$

Le Rapport Signal à Interférence (*SIR : Signal to Interference Ratio en anglais* défini par le rapport entre la puissance du signal à estimer et la puissance de l'erreur d'estimation) associé à cette puissance moyenne de l'erreur d'estimation des sources est alors

$$\begin{aligned} SIR &= \frac{1}{1 - \tilde{P}/\tilde{N}} \\ &= \frac{1}{1 - \frac{PL}{N(Q+L)}} \end{aligned} \quad (2.44)$$

Notons que cette performance moyenne est la performance maximale moyenne obtenue grâce à la matrice pseudo-inverse de $\tilde{\mathbf{H}}^t$ réalisant l'optimisation du critère des moindres carrés. On ne peut pas garantir en pratique que notre algorithme basé sur l'optimisation de la non-gaussianité permettra d'obtenir cette performance maximale théorique. Nous ne quantifions donc ici que la dégradation de la séparation due au caractère sous-déterminé du modèle de mélange linéaire instantané.

Notons toutefois que si L est grand, \tilde{P}/\tilde{N} devient proche de 1, ce qui entraîne un SIR possible élevé.

2.7 Méthode globale de SAS proposée

L'étape d'extraction de processus d'innovation proposée en section 2.3 nous donne une estimation $e_l(n)$ d'un processus d'innovation de source à un facteur d'échelle et à un retard près, qui peut ensuite être colorée pour obtenir chaque contribution de la l ième source dans la k ième observation $x_k(n)$. Ceci peut être fait en calculant les filtres de coloration non causaux

$$C_{kl}(z) = \sum_{r=-R'}^{R'} c_{kl}(r) z^{-r} \quad (2.45)$$

qui rendent les signaux $c_{kl}(n) * e_l(n)$ les plus proches des $x_k(n)$ au sens des moindres carrés [6]. Nous montrons en annexe A que cela peut être réalisé par des filtres RIF non causaux de Wiener, dont les coefficients de réponses impulsionnelles forment des vecteurs \mathbf{c}_{kl} définis par

$$\mathbf{c}_{kl} = \mathbf{R}_{e_l}(n)^{-1} \mathbf{r}_{e_l x_k}(n) \quad (2.46)$$

où $\mathbf{R}_{e_l}(n)$ est la matrice d'autocorrélation du signal $e_l(n)$ et $\mathbf{r}_{e_l x_k}(n)$ est le vecteur d'intercorrélations des signaux $e_l(n)$ et $x_k(n)$. Notons que la matrice d'autocorrélation a une structure de Toeplitz très régulière et qu'il existe de nombreuses méthodes rapides [215] pour résoudre l'équation matricielle linéaire (2.46).

Après avoir soustrait les contributions $c_{kl}(n) * e_l(n)$ de toutes les observations, nous obtenons une autre configuration de mélange avec $N - 1$ sources $s_i(n)$. La première étape doit être répétée comme expliqué dans la section 2.1 pour extraire le processus d'innovation d'une autre source.

2.8 Conclusion

Dans ce chapitre, nous avons introduit une nouvelle approche pour la séparation dans le domaine temporel de mélanges convolutifs de sources. Nos algorithmes sont basés sur un blanchiment spatio-temporel (ou convolutif) non-causal des observations et sur une initialisation particulière des paramètres

d'extraction qui permettent l'utilisation d'itérations de type point fixe pour extraire les processus d'innovation des sources par l'optimisation des critères du kurtosis ou de la néguentropie. La pertinence de notre méthode est notamment basée sur la réinterprétation du mélange sous forme linéaire instantanée. Pour obtenir les contributions des sources sur les capteurs, nous utilisons un critère quadratique correspondant à la puissance de la différence entre une observation et une version filtrée du processus d'innovation extrait et proposons d'utiliser un processus de Wiener non-causal pour l'optimiser. Nous baptisons notre algorithme C-FICA (extension convolutive de FastICA).

Chapitre 3

Résultats expérimentaux

3.1 Introduction

Dans les deux premières sections de ce chapitre, nous illustrons le fonctionnement de notre algorithme en effectuant quelques simulations de séparation de sources de télécommunications et de sources audio en utilisant soit des filtres aléatoires, soit des filtres réels mesurés au niveau des oreilles d'une tête de mannequin. Ensuite, nous effectuons une étude statistique des performances de notre algorithme en termes de *SIR*, de temps de calcul et de résistance au bruit. Nous comparons aussi ses performances et sa rapidité à celles de l'algorithme de J. K. Tugnait travaillant avec les mêmes hypothèses de stationnarité et de non-gaussianité des sources.¹²

3.2 Séparation de signaux de télécommunications

Nous avons testé notre algorithme C-FICA en utilisant le critère du kurtosis dans une configuration relative aux applications télécoms. Les deux premières sources utilisées sont ici deux sources binaires mélangées à l'aide d'un jeu de filtres RIF causaux à 10 coefficients générés par une loi uniforme entre 0 et 1. Les deux sources font 50000 échantillons chacune. Nous avons choisi ici $R = 10$ et $R' = 30$ où R et R' sont définis respectivement par (2.7) et (2.45).

¹²Une page internet [201] est consacrée à notre algorithme C-FICA. Elle comprend une archive téléchargeable contenant une implémentation sous Matlab de l'algorithme ainsi que deux programmes de démonstration illustrant la séparation de mélanges convolutifs de signaux artificiels et audio.

La figure 3.1 représente les deux signaux sources utilisés sur une longueur de 300 échantillons. La figure 3.2 représente de haut en bas les deux observations du mélange, les deux contributions des sources 1 et 2 respectivement sur les capteurs 1 et 2 (nous avons choisi de représenter la contribution de plus forte puissance pour chacune des sources), et les estimations de ces deux contributions. Les rapports signal à interférences (*SIR* : *Signal to Interference Ratio* en anglais)¹³ de sortie des contributions estimées sont respectivement 18.7 dB et 18.6 dB, et les améliorations de *SIR* (*SIRI*) de 16.8 dB et 16.9 dB.

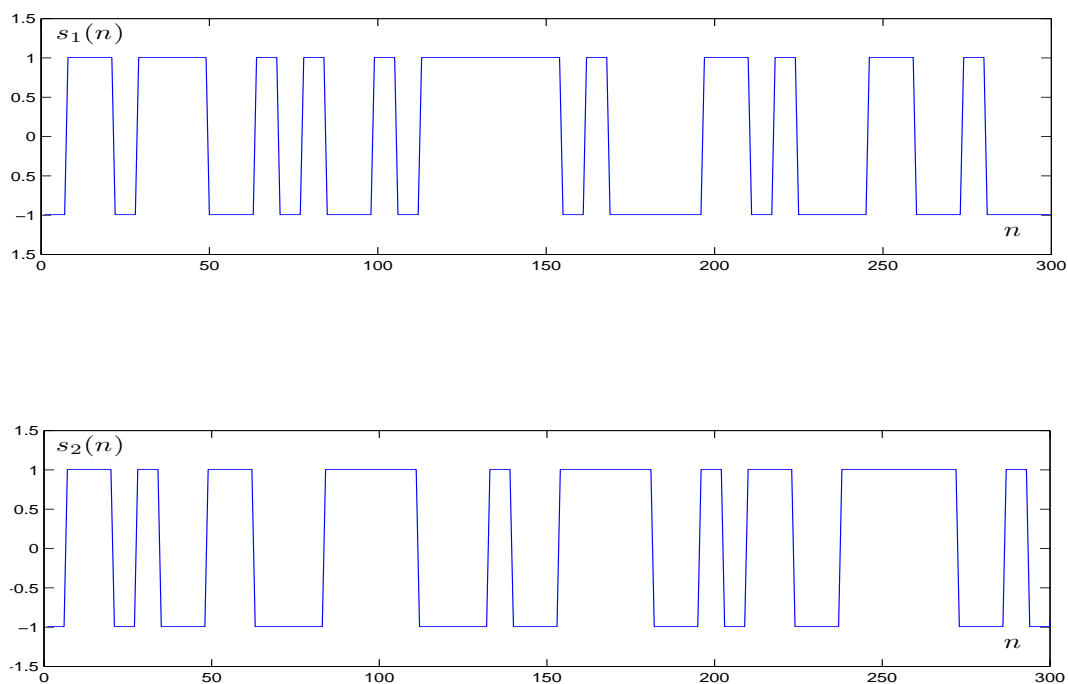


FIG. 3.1 – Sources utilisées 1 et 2 correspondant à des signaux binaires.

¹³Nous définissons les mesures de performances SIR_{out} , SIR_{in} et $SIRI$ dans l'annexe C.

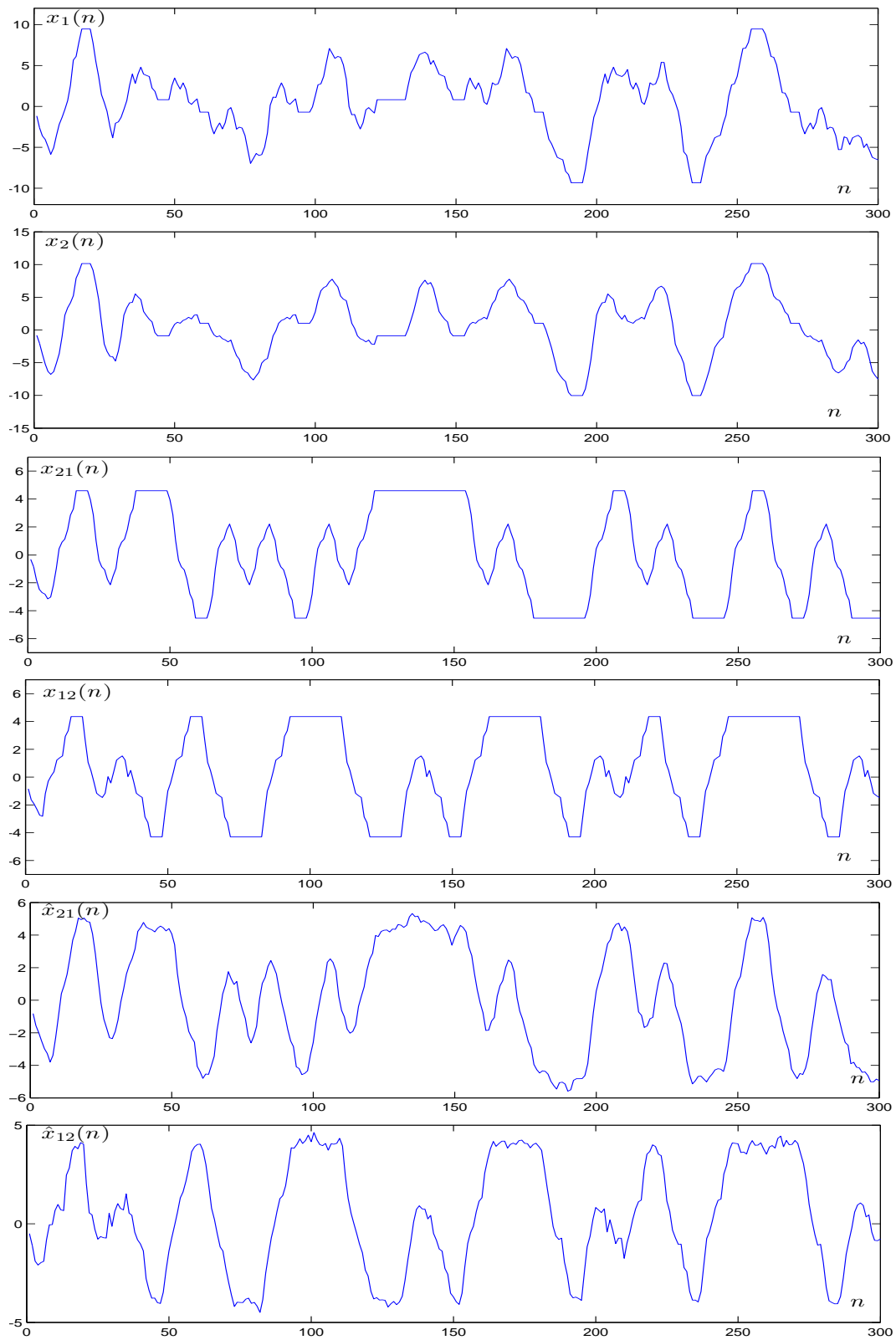


FIG. 3.2 – De haut en bas : observations 1 et 2, contributions réelles des sources 1 et 2 respectivement sur les capteurs 2 et 1, contributions estimées.

Les figures 3.3 et 3.4 représentent les résultats obtenus dans une configuration similaire à l'exception des sources qui sont cette fois des sources AM à 4 états.

Les SIR de sortie des contributions estimées sont cette fois respectivement 25.6 dB et 22.2 dB, et les améliorations de SIR ($SIRI$) de 19.1 dB et 24.2 dB.

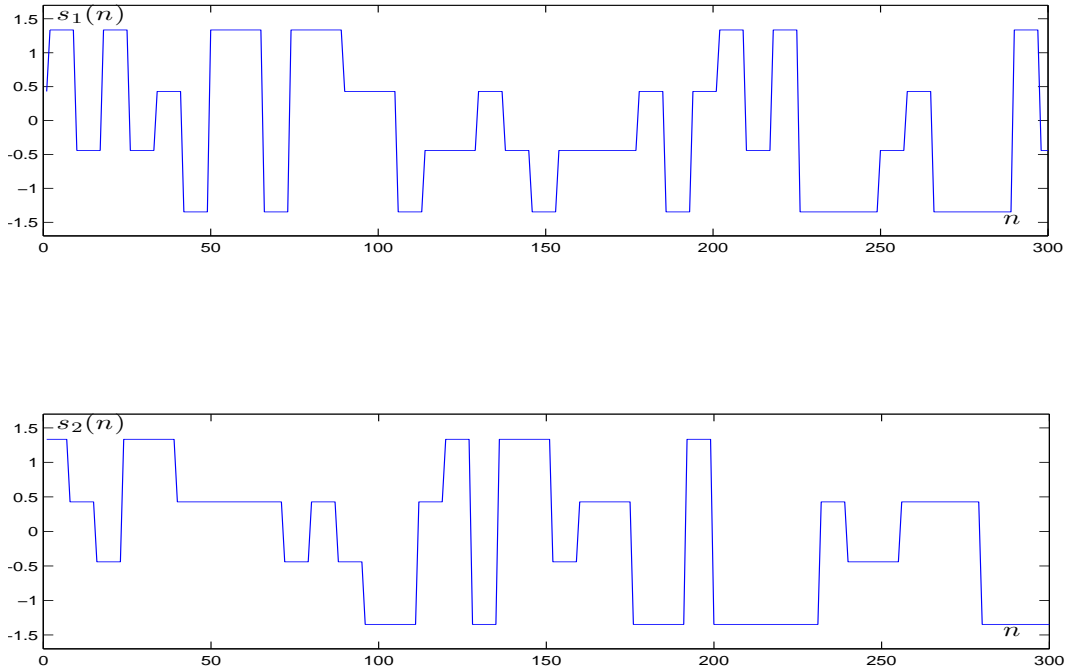


FIG. 3.3 – Sources utilisées 1 et 2 correspondant à des signaux télécoms AM à 4 états.

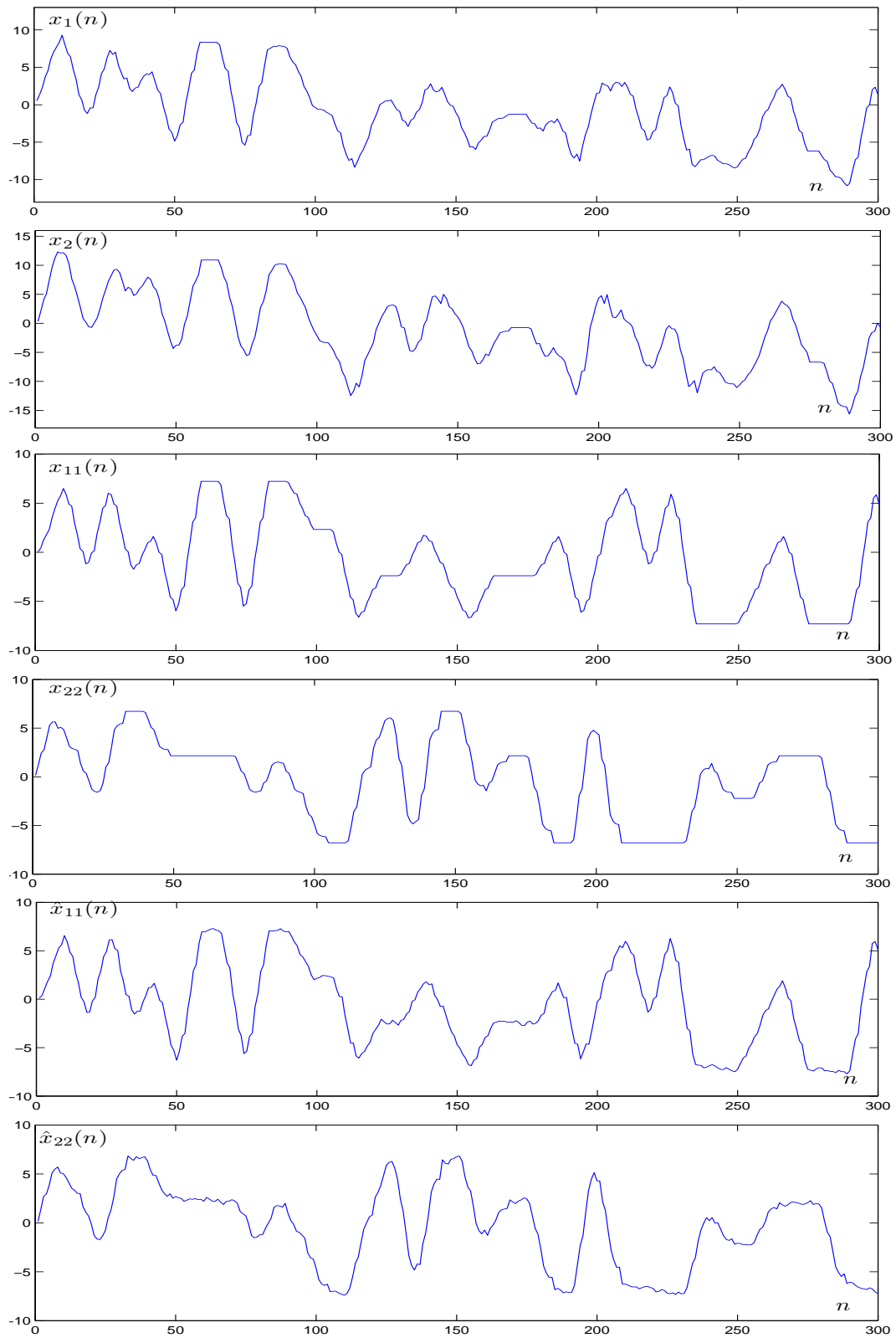


FIG. 3.4 – De haut en bas : observations 1 et 2, contributions réelles des sources 1 et 2 respectivement sur les capteurs 1 et 2, contributions estimées.

3.3 Séparation de signaux audio

Nous avons effectué deux tests de séparation dans le cas de mélanges de deux signaux de parole et de musique en utilisant des filtres réels d'ordre 256 (à une fréquence d'échantillonnage de 20 kHz) mesurés au niveau des oreilles d'une tête de mannequin [81] (pour le premier test nous avons sélectionné les filtres associés aux angles 15° et -30°). La séparation a été réalisée ici en utilisant la néguentropie avec $G(x) = e^{-\frac{x^2}{2}}$ et en choisissant $R = 40$ et $R' = 100$. Nous avons aussi choisi une fenêtre d'extraction de processus d'innovation où les deux sources sont stationnaires (les filtres obtenus ont ensuite été appliqués à la totalité des signaux). Pour choisir cette fenêtre d'extraction, nous avons développé une procédure qui consiste à effectuer la séparation successivement sur plusieurs fenêtres et à choisir l'une de celles donnant des sorties de contributions les plus corrélées. La figure 3.5 représente les 4 filtres réels utilisés et la figure 3.6 représente de haut en bas les deux observations du mélange, les deux contributions des sources 1 et 2 respectivement sur les capteurs 1 et 2 (nous avons choisi de représenter la contribution de plus forte puissance pour chacune des sources), et les estimations de ces deux contributions. Les SIR des contributions estimées sont de 12.8 dB et 14.3 dB et les améliorations de SIR ($SIRI$) de 5.7 dB et 11.7 dB.

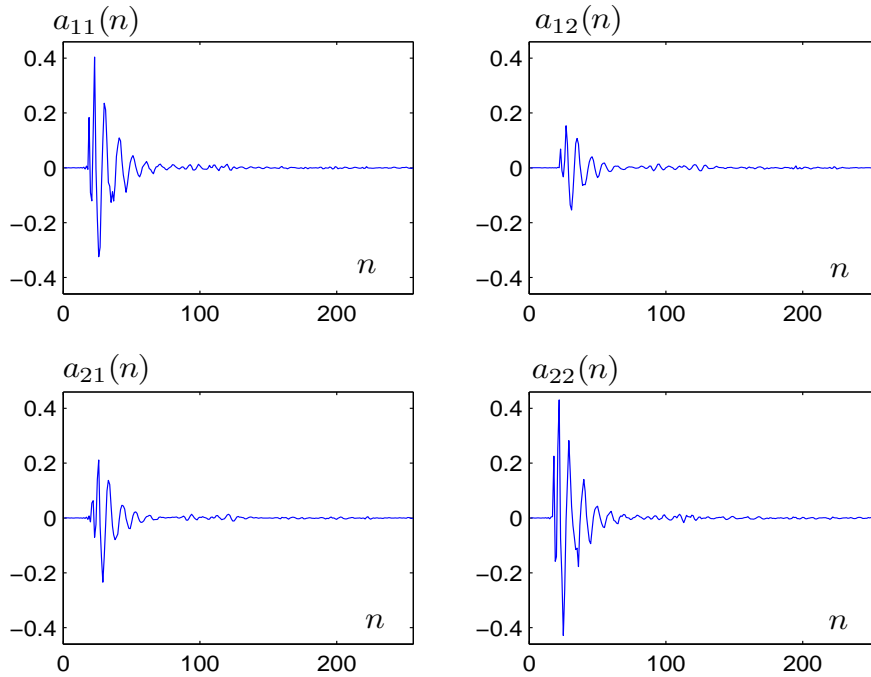


FIG. 3.5 – Réponses impulsionnelles de la matrice de filtres de mélange utilisée correspondant aux azimuts 15° et -30° (ordre=256).

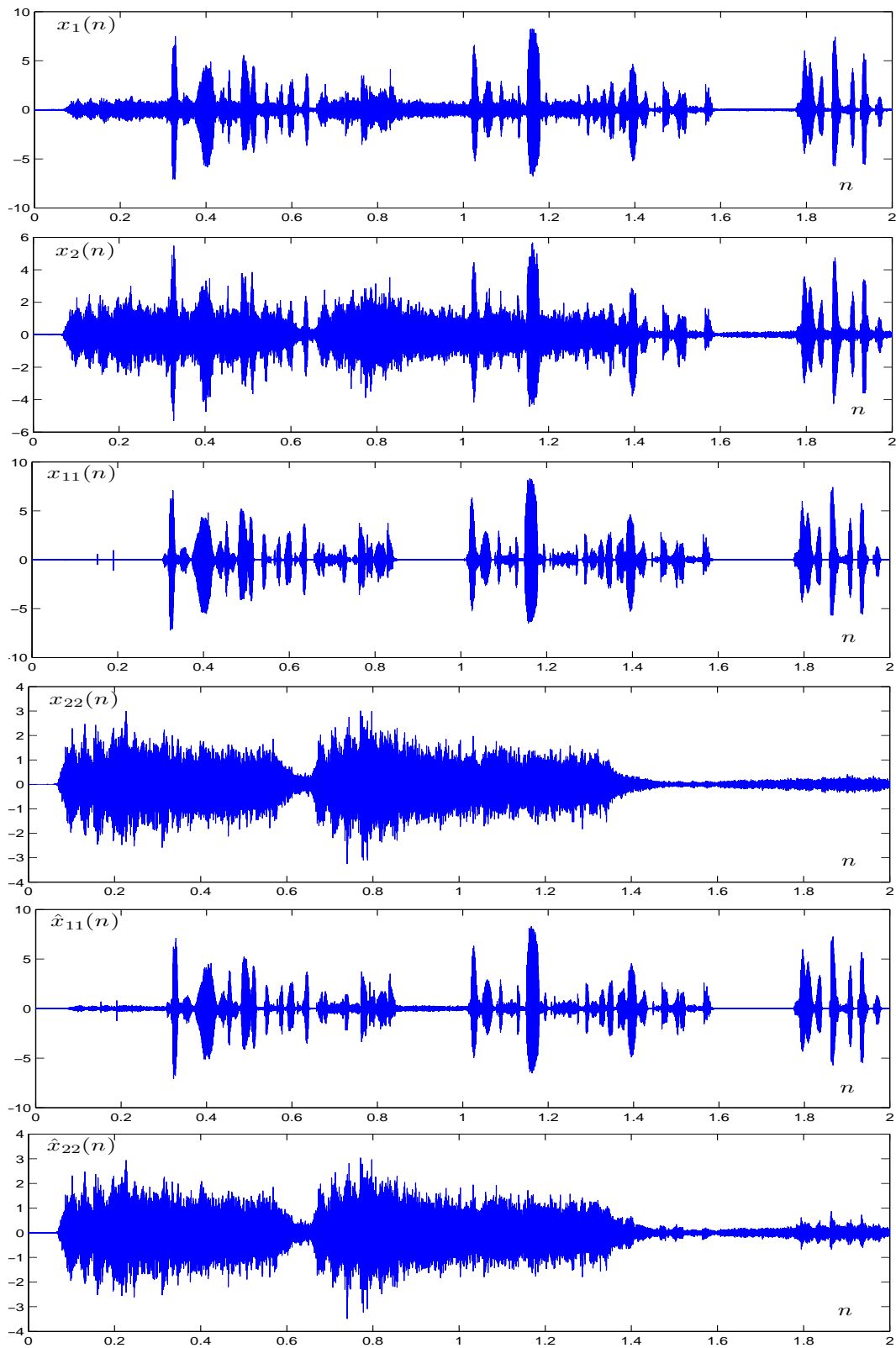


FIG. 3.6 – De haut en bas : observations 1 et 2, contributions réelles des sources 1 et 2 respectivement sur les capteurs 1 et 2, contributions estimées.

Les figures 3.7 et 3.8 représentent les résultats obtenus dans une configuration de mélange similaire à ce qui précède à l'exception des positions des sources qui ont cette fois été associées aux angles 10° et 60° . Nous sommes donc ici dans une configuration où les deux sources se trouvent du même côté de la tête de mannequin. Cette configuration peut poser des problèmes pour certains algorithmes comme celui de L. Parra qui n'a pas donné de résultats auditifs satisfaisants dans ce cas à la différence du cas précédent. Pour notre approche, les *SIR* des contributions estimées sont ici de 13.2 dB et 13.0 dB et les améliorations de *SIR* (*SIRI*) de 6.2 dB et 13.2 dB.

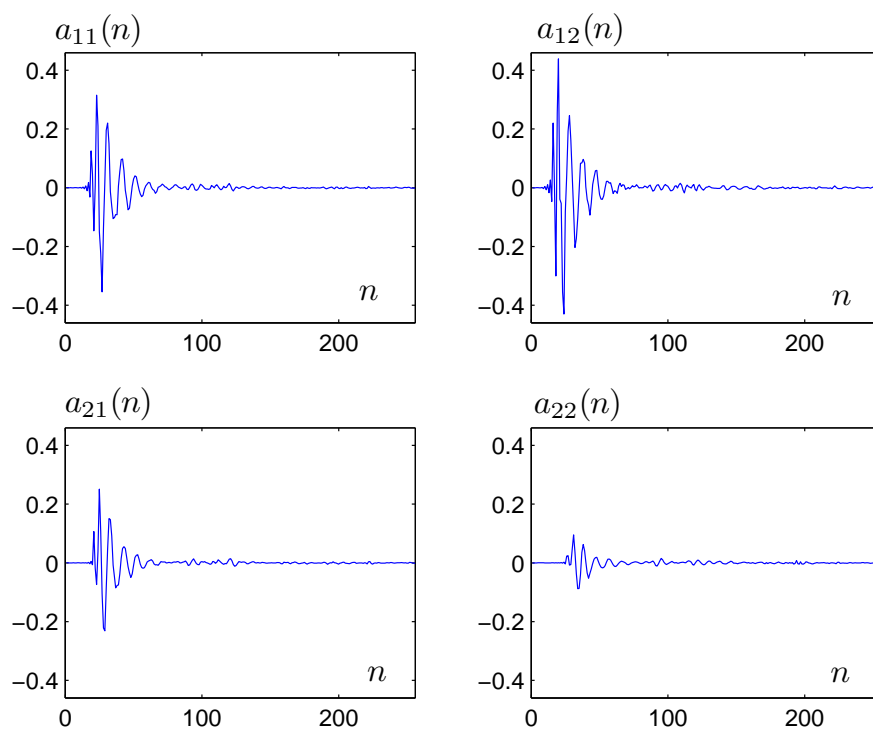


FIG. 3.7 – Réponses impulsionnelles des filtres de mélange utilisés correspondant aux azimuts 10° et 60° (ordre=256).

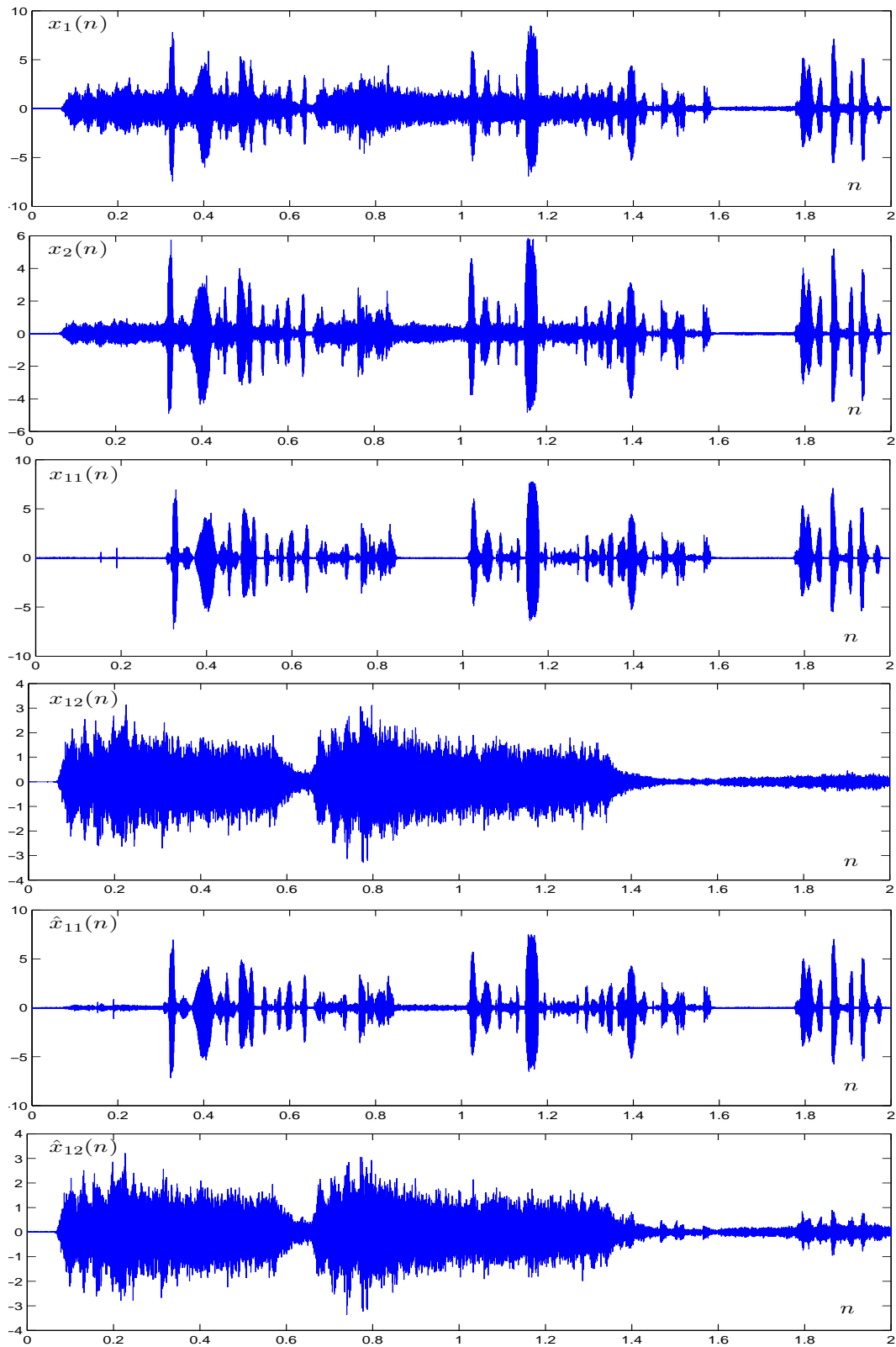


FIG. 3.8 – De haut en bas : observations 1 et 2, contributions réelles des sources 1 et 2 sur le capteur 1, contributions estimées.

3.4 Tests statistiques

Dans cette section, nous étudions les performances de nos méthodes dans plusieurs configurations. Les algorithmes sont d'abord testés pour $P = 2$ mélanges convolutifs de $N = 2$ sources normalisées colorées artificiellement contenant 100000 échantillons. Nous avons effectué des tests dans les deux cas de figure où les processus d'innovation $u_j(n)$ ont des distributions uniforme et laplacienne. Dans le cas de processus à distribution uniforme, nous avons utilisé le critère du kurtosis qui s'est avéré légèrement plus performant que la néguentropie (les SIR_{out} des contributions estimées sont supérieurs de 2 à 3 dB). Dans le cas de processus laplaciens, le critère de la néguentropie en utilisant la non-linéarité gaussienne $G(x) = e^{-x^2/2}$ a donné de meilleures performances et a donc été préféré au kurtosis. Nous avons représenté les SIR_{out} ¹⁴, en fonction de l'ordre du modèle Q que nous définissons comme la somme des ordres des filtres de mélange $A_{ij}(z)$ et de coloration d'innovation $F_j(z)$. Pour chaque valeur de Q , 100 simulations de Monte-Carlo ont été réalisées en faisant varier les coefficients des filtres de mélange et de coloration d'innovation avec une loi uniforme. Les moyennes et écarts-types des SIR_{out} résultants ont ensuite été calculés. L'ordre R dans (2.7) a été fixé à $R = Q$ car nos tests ont montré que cela entraîne un bon compromis entre le SIR_{out} et le temps de calcul. De la même manière, l'ordre des filtres non causaux $C_{kl}(z)$ utilisés pour colorer les processus d'innovation estimés a été fixé à $R' = 2Q$.

La figure 3.9 montre que dans le cas de processus d'innovation uniformes, pour des ordres de modèle Q inférieurs à 60, les moyennes des SIR_{out} sont comprises entre 17 et 9.5 dB pour la première source et un peu moins pour la seconde source, excepté pour le cas linéaire instantané, qui conduit à des performances nettement meilleures. Nous avons observé qu'en choisissant un seuil de critère d'arrêt plus petit (nous l'avons choisi égal à 10^{-12} ici), il est possible d'obtenir de meilleurs résultats pour des valeurs de Q élevées. Cependant, le temps de calcul est alors plus long, ce qui limite la possibilité d'une analyse statistique par des tests de Monte-Carlo. Notons que les coefficients de filtres utilisés ici ne sont pas décroissants dans l'ensemble comme ceux des filtres réels dont les derniers coefficients sont en général faibles. Cela explique que nous ne puissions pas ici atteindre avec de bonnes performances des valeurs de Q égales à l'ordre des filtres réels utilisés dans la section précédente.

La figure 3.10 montre que pour des processus d'innovation à distribution laplacienne, les SIR_{out} moyens sont compris entre 19 et 13 dB pour des valeurs de Q inférieures à 80. Ces performances sont meilleures que dans le cas d'une distribution uniforme, probablement en raison d'optimums du critère plus prononcés dans le cas laplacien, la distribution laplacienne étant plus éloignée de

¹⁴Dans cette section, nous avons utilisé des filtres artificiels de puissances à peu près égales; le SIR_{in} est donc proche de 0 car les deux contributions de sources normalisées sur un capteur donné sont alors de puissances à peu près égales.

la loi normale que la distribution uniforme (le kurtosis normalisé d'une variable uniforme est égal à -1.2 tandis que celui d'une variable laplacienne est égal à 3).

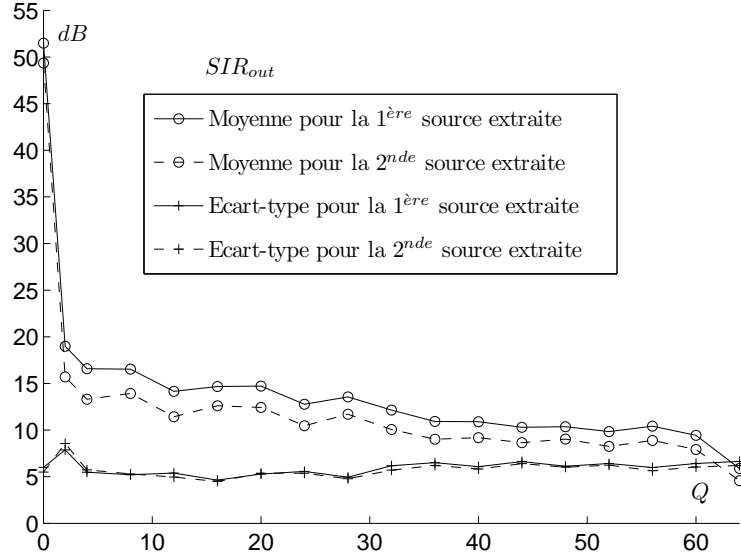


FIG. 3.9 – SIR_{out} des sources extraites en fonction de Q pour des sources colorées artificielles dans le cas de processus d'innovation uniformes et en utilisant le critère du kurtosis.

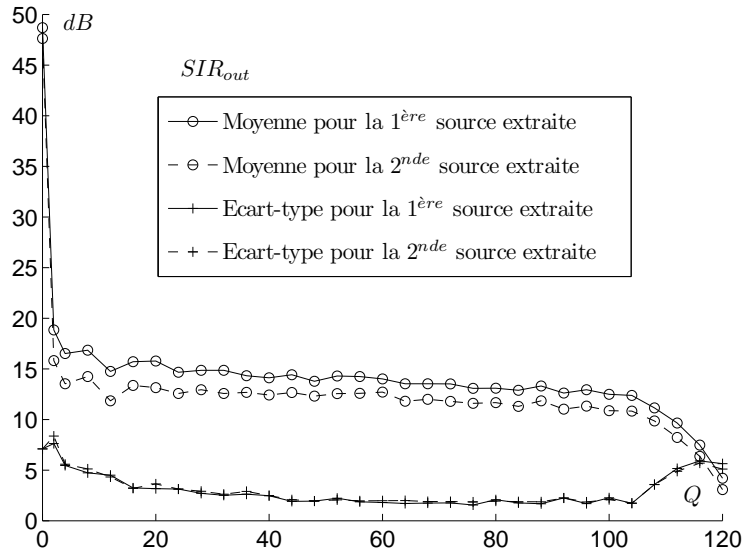


FIG. 3.10 – SIR_{out} des sources extraites en fonction de Q pour des sources colorées artificielles dans le cas de processus d'innovation laplaciens et en utilisant le critère de la néguentropie.

Dans la seconde série d'expériences, nous avons testé notre algorithme dans le cas sous-déterminé $N = 3$ et $P = 2$, quand les deux signaux observés contiennent une source de bruit blanc gaussien¹⁵ stationnaire en plus des deux sources utiles mentionnées précédemment (nous avons choisi ici des processus d'innovation de densité uniforme). Nous avons fixé l'ordre du modèle à $Q = 20$. En faisant varier la puissance de la source de bruit et donc le rapport signal à bruit d'entrée (SNR_{in}), nous avons étudié la robustesse de l'estimation de la matrice de mélange en présence de bruit. Dans ce but, nous avons calculé les SIR_{out} des deux sources utiles estimées dans lesquelles nous avons enlevé les contributions de la source de bruit. La figure 3.11 montre que notre méthode est assez robuste au bruit puisque les SIR_{out} de la première source restent supérieurs à 10 dB pour un SNR_{in} décroissant jusqu'à 8 dB.

Pour $N = P = 2$, nous avons comparé le temps de calcul de notre méthode avec une version modifiée de l'algorithme de Tugnait développée dans [6], qui permet déjà une plus grande rapidité que l'approche de Tugnait en utilisant un algorithme de Newton modifié. Nous avons fait varier l'ordre du modèle Q entre 0 et 30 pour des sources contenant $T = 10000$ échantillons ($T > 10000$ ou $Q > 30$ entraînent des temps de calcul trop grands pour 100 tests de Monte-Carlo de la méthode de Tugnait modifiée). Les résultats de la figure 3.12 représentent la durée de l'extraction du premier processus d'innovation, qui est de loin l'étape la plus lourde en temps de calcul. Cela montre que notre méthode est à peu près 100 fois plus rapide que la méthode Tugnait-Newton. De plus, elle donne des SIR_{out} légèrement supérieurs, d'à peu près 0.5 dB.

3.5 Conclusion

Dans ce chapitre, nous avons validé notre algorithme C-FICA en l'appliquant d'abord à des mélanges RIF de sources de télécommunications et de sources audio. Nous avons ensuite effectué des tests de Monte-Carlo pour apprécier la robustesse de notre méthode en fonction de l'ordre du mélange, de l'ordre de coloration des sources et du niveau de bruit présent dans les observations. Les performances en termes des Rapport Signal à Interférence sont légèrement meilleures qu'avec la méthode de Tugnait-Newton, tandis que le temps de calcul est de l'ordre de 100 fois plus faible.

¹⁵Le kurtosis d'un signal gaussien est nul mais en raison du blanchiment qui se fait à partir de statistiques du second ordre, ce bruit gaussien n'est pas transparent pour notre critère d'optimisation.

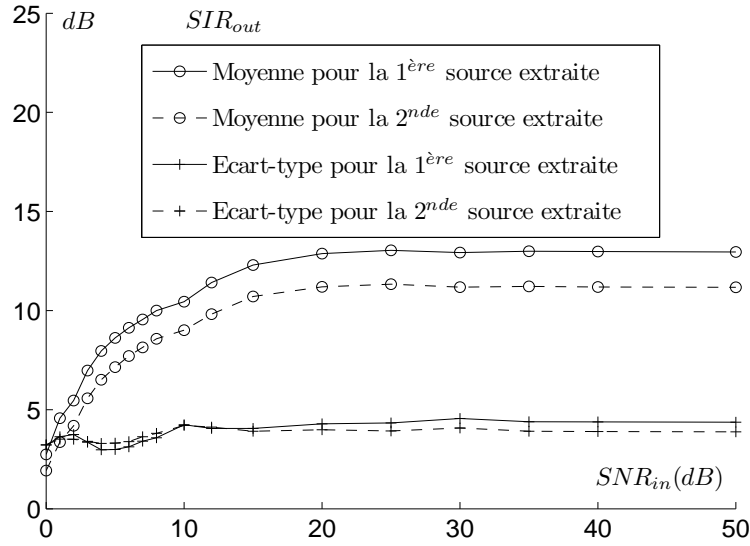


FIG. 3.11 – SIR_{out} des sources utiles extraites en fonction du SNR d'entrée.

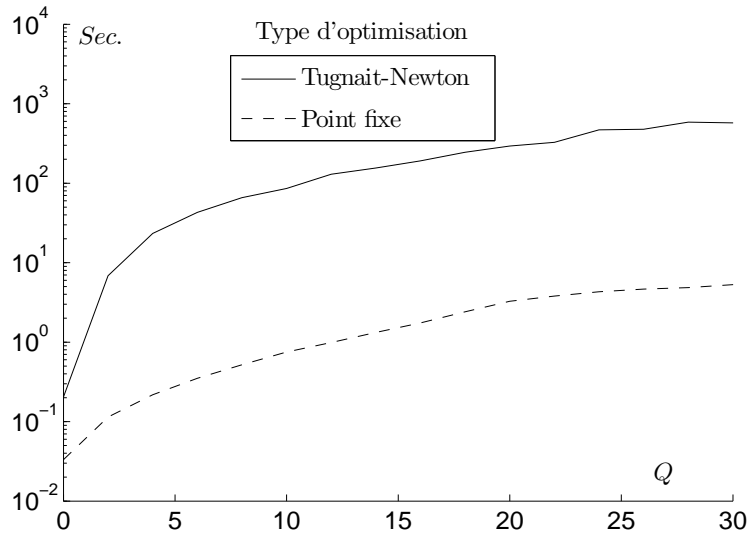


FIG. 3.12 – Temps d'extraction moyen du premier processus d'innovation en fonction de Q pour une optimisation de type Tugnait-Newton et pour notre optimisation de type point fixe.

Deuxième partie

Mélanges sous-déterminés

Chapitre 4

Mélanges sous-déterminés : état de l'art

La séparation aveugle de sources sous-déterminée cherche à estimer un jeu de N signaux sources inconnus $s_j(n)$ à partir d'un jeu de P signaux observés $x_i(n)$ lorsque le nombre d'observations est inférieur au nombre de sources ($P < N$). Les signaux observés sont dans la plupart des cas des mélanges linéaires instantanés ou convolutifs des signaux sources. Les mélanges convolutifs s'expriment par

$$\mathbf{x}(n) = \mathbf{A}(n) * \mathbf{s}(n) \quad (4.1)$$

où $\mathbf{s}(n) = [s_1(n), \dots, s_N(n)]^t$ et $\mathbf{x}(n) = [x_1(n), \dots, x_P(n)]^t$ sont les vecteurs des sources et des observations, $*$ représente l'opérateur de convolution, et la matrice de mélange $\mathbf{A}(n)$ est composée des réponses impulsionnelles de filtres de mélange inconnus. Ce contexte général inclut les mélanges linéaires instantanés. $\mathbf{A}(n)$ se réduit alors à une matrice composée de termes proportionnels à des diracs centrés à l'origine $\mathbf{A}\delta(n)$ et le vecteur d'observations s'exprime par un simple produit matriciel

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n). \quad (4.2)$$

Le nombre de sources N étant supérieur au nombre d'observations P , la matrice de mélange $\mathbf{A}(n)$ possède davantage de colonnes que de lignes et n'est donc pas inversible à gauche. Dans le cas général, il est donc impossible, même en connaissant parfaitement les paramètres du mélange, d'estimer correctement les sources par des méthodes de séparation linéaires.

La séparation de sources sous-déterminée a été abordée pour la première fois par Belouchrani et Cardoso [26] (1994). Elle se décompose en général en deux sous-problèmes : l'estimation de la matrice de mélange et l'estimation des sources.

4.1 Estimation de la matrice de mélange

La grande majorité des algorithmes de séparation pour mélanges sous-déterminés supposent que les sources sont parcimonieuses, c.-à-d. qu'il existe des zones du domaine d'analyse (domaines temps-fréquence, en ondelettes ou multi-résolution par exemple) où une source seulement est non nulle. Dans ce cas, l'analyse d'une zone mono-source permet l'identification d'une colonne de la matrice de mélange.

L'identification des coefficients de mélange peut être réalisée soit par l'étude du nuage de points des observations, grâce à des techniques de clustering [35, 36, 94, 119, 120], soit par le calcul de rapports d'observations, de coefficients d'intercorrélation ou de densités interspectrales des observations (resp. DUET [229] ou TIFROM [7], TIFCORR [71], TIFCOHERE [10]). Dans le cas de DUET, le calcul des rapports d'observations est couplé à une analyse d'histogramme à deux dimensions (amplitude, retard) pour déterminer le nombre de sources ainsi que les coefficients de mélange.

Les algorithmes basés sur la parcimonie peuvent être plus au moins contraignants sur les propriétés des sources. L'algorithme DUET par exemple suppose que les sources sont totalement disjointes dans le plan temps-fréquence : dans chaque case du plan temps-fréquence, une seule source doit être présente (ou présente de façon prépondérante). Les algorithmes TIFROM, TIFCORR et TIFCOHERE supposent eux que chaque source possède au moins une zone mono-source associée. Dans ce cas, l'existence de zones multi-sources est tolérée.

Notons que dans le cas de mélanges contenant un grand nombre de sources, l'existence de zones mono-sources est moins probable.

D'autres hypothèses que la parcimonie peuvent être utilisées pour la séparation de mélanges sous-déterminés. On peut par exemple supposer que les sources sont discrètes [26, 58, 197], de distributions connues [34, 152, 209], ou qu'elles vérifient un modèle donné [109, 230] (un article de E. Vincent et X. Rodet suppose par exemple que les sources de musique vérifient un modèle markovien [218]). D'autres méthodes formulent des hypothèses sur le processus de mélange [39].

La séparation de sources sous-déterminée a été très peu étudiée dans le cas de mélanges convolutifs. On peut citer toutefois [17, 32, 34, 144, 152]. Le modèle linéaire à atténuations et retards a été lui plus étudié, principalement par les approches temps-fréquence DUET [229], TIFROM [7], TIFCORR [71] et TIFCOHERE [10].

4.2 Reconstruction des sources

Une fois la matrice de mélange identifiée, il reste à séparer les sources entre elles, ce qui ne peut être fait par simple inversion matricielle puisque la matrice de mélange sous-déterminée n'est pas inversible à gauche. Une solution consiste

à calculer la matrice pseudo-inverse de Moore-Penrose définie par

$$\mathbf{A}^+ = \mathbf{A}^t(\mathbf{A}\mathbf{A}^t)^{-1}. \quad (4.3)$$

Dans ce cas, on obtient une séparation approximative des sources : chaque sortie contient une source prépondérante superposée à des résidus des autres sources.

Une autre solution consiste à annuler la présence de certaines sources en effectuant une séparation partielle. Pour un mélange à N sources et P observations, il est ainsi possible d'annuler la présence de $P - 1$ sources dans les signaux de sortie. Chaque signal de sortie contiendra alors $N - P + 1$ sources. Cette méthode peut être employée par exemple en cas de présence de sources parasites se superposant aux sources utiles que l'on veut séparer entre elles.

Une autre solution très simple [121, 216, 217] consiste, à chaque instant n , à associer le vecteur d'observations à la source dont le vecteur colonne de \mathbf{A} est le plus proche. Pour cela, on calcule les produits scalaires du vecteur d'observation avec les N colonnes normalisées de la matrice de mélange et on identifie l'indice de la source active comme celui ayant le plus grand produit scalaire en valeur absolue. On retient alors ce produit scalaire (signé) comme valeur pour la source active. La figure 4.1 illustre cette méthode de reconstruction de sources.

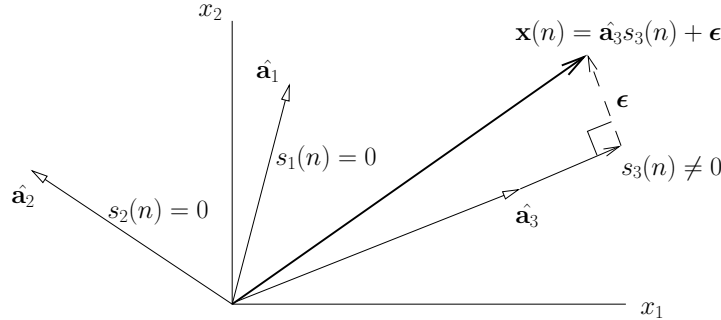


FIG. 4.1 – Schéma du principe d'estimation des sources à partir de la matrice de mélange en considérant qu'une seule source (celle d'indice 3) est active à l'instant n .

Une extension logique de cette approche consiste à assigner plusieurs sources à chaque observation. En général, on cherche alors à chaque instant n un vecteur source à P composantes non nulles vérifiant l'équation de mélange et de norme ℓ^1 minimale [35, 36, 118, 119, 199, 222]. Cela revient à résoudre à chaque instant n le problème d'optimisation suivant :

$$\min_{s_1(n) \dots s_N(n)} \left(\sum_{j=1}^N |s_j(n)| \right) \text{ t.q. } \sum_{j=1}^N \hat{\mathbf{a}}_j s_j(n) = \mathbf{x}(n) \quad (4.4)$$

où $\hat{\mathbf{a}}_j$ est l'estimée de la $j^{\text{ème}}$ colonne de la matrice de mélange et où P composantes seulement du vecteur $\mathbf{s}(t)$ sont supposées non nulles. Cette technique

dite du trajet le plus court exige cependant une forte parcimonie des sources, et ce d'autant plus que le rapport P/N est faible car elle suppose qu'à chaque instant, P sources seulement parmi N soient non nulles. La figure 4.2 illustre cette méthode du trajet le plus court.

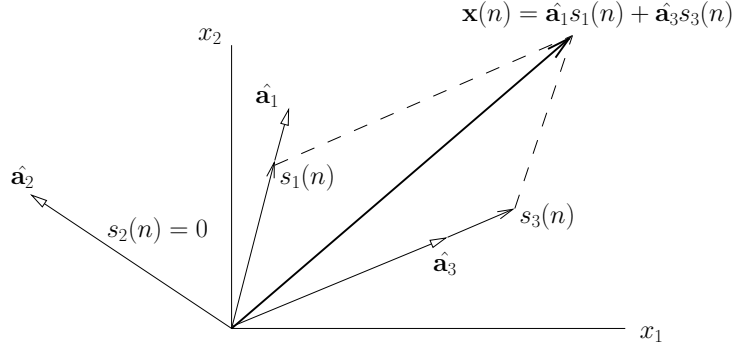


FIG. 4.2 – Schéma du principe d'estimation des sources à partir de la matrice de mélange par la méthode du trajet le plus court.

D'autres normes que la norme ℓ^1 ont été employées pour déterminer ce plus court trajet (norme ℓ^q avec $q < 1$ par exemple [179]).

Une approche basée sur les statistiques du second ordre et appelée principe de décomposition statistiquement parcimonieuse (*SSDP : statistically sparse decomposition principle*) a aussi été proposée par M. Xiao, S. Xie et Y. Fu [227] pour identifier les sources prépondérantes à séparer entre elles.

Une autre technique couramment employée consiste à utiliser un dictionnaire de formes d'onde auquel les sources sont supposées appartenir [27, 28, 232]. L'appartenance des sources au dictionnaire engendre de nouvelles contraintes qui rendent possible leur estimation. La décomposition n'étant pas unique, plusieurs méthodes peuvent être employées pour reconstituer les sources (Method Of Frames MOF [60], Best Orthogonal Basis BOB [53], Matching Pursuit MP [116, 117, 126], Basis Pursuit [48] BP).

Nous avons vu que les méthodes linéaires ne permettent pas une séparation exacte des sources. Dans le cas général, une combinaison linéaire des observations contiendra en effet au minimum $N - P + 1$ sources. Cependant la séparation exacte peut être réalisée au moyen d'un masquage binaire du plan temps-fréquence sous réserve que les sources soient totalement disjointes dans le plan temps-fréquence. L'algorithme DUET utilise cette technique qui consiste pour l'estimation d'une source donnée à masquer les cases temps-fréquences associées aux autres sources [229]. Récemment Aissa-El-Bey et al. ont proposé une méthode pour relâcher l'hypothèse d'orthogonalité des sources en se basant sur une projection en sous-espaces. Cette méthode suppose qu'à chaque point temps-fréquence, le nombre de sources actives est strictement inférieur au nombre d'observations. Deux articles ont été publiés respectivement pour les mélanges linéaires instantanés [18] et convolutifs [17].

Nous cherchons dans cette partie à éviter toutes ces hypothèses parfois très restrictives sur le mélange et les sources. Les méthodes de séparation partielle rapides présentées ici supposent uniquement que les sources mélangées sont divisées en deux catégories : certaines sont non-stationnaires et correspondent aux sources utiles que l'on cherche à séparer entre elles, d'autres sont stationnaires et correspondent à des sources de bruit qui peuvent rester présentes dans les sorties de l'algorithme de séparation.

Notons que la superposition d'un signal de bruit stationnaire sur un signal de parole est moins gênante que la superposition d'un autre signal de parole. En effet, le cerveau se focalise davantage sur la parole et l'on est davantage perturbé par une personne parlant en même temps que notre interlocuteur que par un bruit de fond neutre.

4.3 La séparation différentielle de sources

Dans [67], Deville, Benali et Abrard ont introduit un concept général de SAS différentielle pour traiter les mélanges sous-déterminés. Dans sa version standard, la séparation différentielle de sources considère la situation où (au plus) P des N sources mélangées sont non-stationnaires alors que les $N - P$ autres sources (au moins) sont stationnaires.

Les P sources non-stationnaires sont les signaux utiles dans cette approche, alors que les $N - P$ sources stationnaires sont considérées comme des sources de bruit. Notre concept de SAS différentielle réalise la séparation partielle des P sources utiles (le principe de la séparation partielle a été défini dans la section précédente) [67].

Néanmoins, nous faisons remarquer que cette séparation partielle peut-être considérée comme une séparation pseudo-complète dans les situations où l'on ne cherche pas à séparer les sources de bruit qui ne contiennent aucune information. Cette méthode peut être d'un intérêt pratique pour le scénario du "cocktail party" bruité. On pourrait aussi envisager notre approche dans les systèmes de communications MIMO, pour lesquels les signaux reçus sont souvent perturbés par des sources de bruit stationnaires dans les applications réelles. Cette méthode pourrait aussi être appliquée aux signaux biomédicaux, qui contiennent eux aussi diverses composantes de bruit stationnaire.

Bien que le concept de séparation différentielle soit défini de façon assez générale dans [67], il n'a été appliqué pour l'instant qu'à une méthode de SAS simple mais restrictive, limitée à $P = 2$ mélanges, impliquant seulement deux filtres à réponses impulsionnelles finies (RIF) strictement causaux (c.-à-d. pas de mélange instantané), et basée sur des algorithmes à convergence lente. Nous introduisons ici des méthodes de SAS beaucoup plus puissantes et leurs algo-

arithmes, basées sur la SAS différentielle, pour les mélanges linéaires instantanés et convolutifs¹⁶.

Notre méthode instantanée est obtenue en étendant aux mélanges sous-déterminés le critère de séparation appelé kurtosis [61] et l'algorithme FastICA associé, de type point fixe et à convergence rapide [98]. Dans le cas convolutif, nous étendons aux mélanges sous-déterminés l'algorithme temporel rapide à point fixe C-FICA restreint aux mélanges (sur-)déterminés (i.e. $P \geq N$) que nous avons proposé dans la partie 1 du manuscrit. Nous gardons ainsi les propriétés attractives de ce dernier algorithme.

¹⁶Un algorithme différentiel à point fixe légèrement différent a été proposé auparavant dans notre équipe [69] pour les mélanges linéaires instantanés. La convergence de cet algorithme n'a pas été prouvée rigoureusement et la version parallèle n'a pas été étudiée.

Chapitre 5

Méthode de SAS différentielle proposée pour les mélanges instantanés

5.1 Introduction

Ce chapitre concerne la séparation de mélanges linéaires instantanés sous-déterminés. Nous proposons une méthode de SAS partielle qui sépare P sources supposées non-stationnaires en gardant des composantes résiduelles des $N - P$ autres sources supposées être des sources de "bruit" stationnaire. Cette méthode est basée sur le concept général de séparation différentielle récemment introduit dans notre équipe. L'approche proposée consiste en une extension différentielle de l'algorithme FastICA et repose sur un blanchiment différentiel des observations permettant d'appliquer un algorithme de type point fixe pour l'optimisation du critère du kurtosis différentiel que nous définissons.

5.2 Un nouveau critère de SAS basé sur le kurtosis différentiel

La méthode FastICA standard [98], qui est seulement applicable aux mélanges linéaires instantanés avec $P = N$ (ou $P > N$), extrait une source au moyen d'une procédure en deux étapes. La première consiste à appliquer au vecteur d'observations $\mathbf{x}(n)$ une matrice réelle $P \times P$, ce qui donne le vecteur

$$\mathbf{z}(n) = \mathbf{B}\mathbf{x}(n). \quad (5.1)$$

Dans la méthode FastICA standard, \mathbf{B} est choisie de manière à décorrélérer spatialement et à normaliser les observations. La seconde étape de cette méthode standard consiste ensuite à obtenir un signal de sortie $y(n)$ en faisant une combinaison linéaire des signaux contenus dans $\mathbf{z}(n)$, c.-à-d.

$$y(n) = \mathbf{w}^t \mathbf{z}(n) \quad (5.2)$$

où \mathbf{w} est un vecteur, qui est contraint à être de norme unité. Ce vecteur \mathbf{w} est sélectionné de manière à optimiser le kurtosis (non-normalisé) de $y(n)$, défini comme l'autocumulant sans décalage d'ordre 4.

$$kurt_y(n) = Cum(y(n), y(n), y(n), y(n)). \quad (5.3)$$

Étudions maintenant le cas sous-déterminé, c.-à-d. $P < N$. Nous obtenons à nouveau un signal de sortie $y(n)$ en accord avec (5.1) et (5.2). Nous cherchons à définir de nouveaux critères pour sélectionner \mathbf{B} et \mathbf{w} , de façon à réaliser la SAS partielle, définie dans le chapitre précédent, des P sources utiles. À cette fin, nous appliquons le concept général de SAS différentielle décrit dans [67] au critère spécifique du kurtosis utilisé dans la méthode FastICA standard. Nous considérons donc deux instants n_1 et n_2 , et introduisons ensuite le kurtosis différentiel (non-normalisé) que nous associons à (5.3) pour ces deux instants. Nous définissons ce paramètre par

$$Dkurt_y(n_1, n_2) = kurt_y(n_2) - kurt_y(n_1). \quad (5.4)$$

Montrons maintenant que, bien que le paramètre standard $kurt_y(n)$ dépende de toutes les sources, sa version différentielle $Dkurt_y(n_1, n_2)$ dépend seulement des sources non-stationnaires. Les éq. (4.2), (5.1) et (5.2) donnent

$$y(n) = \mathbf{v}^t \mathbf{s}(n) \quad (5.5)$$

où le vecteur

$$\mathbf{v} = (\mathbf{B}\mathbf{A})^t \mathbf{w} \quad (5.6)$$

inclut les effets des étages de mélange et de séparation. En notant v_q , avec $q = 1, \dots, N$, les éléments de \mathbf{v} , l'éq. (5.5) implique que le signal de sortie $y(n)$ peut être exprimé en fonction de toutes les sources comme suit

$$y(n) = \sum_{q=1}^N v_q s_q(n). \quad (5.7)$$

En utilisant les propriétés des cumulants et l'indépendance supposée de toutes les sources, on en déduit facilement que

$$kurt_y(n) = \sum_{q=1}^N v_q^4 kurt_{s_q}(n) \quad (5.8)$$

où $kurt_{s_q}(n)$ est le kurtosis de la source $s_q(n)$, à nouveau défini par (5.3). Le kurtosis standard (5.8) de la sortie dépend donc réellement des kurtosis de toutes les sources. Le kurtosis différentiel associé de la sortie, défini par (5.4), peut être exprimé par

$$Dkurt_y(n_1, n_2) = \sum_{q=1}^N v_q^4 Dkurt_{s_q}(n_1, n_2) \quad (5.9)$$

où nous définissons le kurtosis différentiel $Dkurt_{s_q}(n_1, n_2)$ de la source $s_q(n)$ de la même façon que dans (5.4). Tenons compte maintenant de l'hypothèse suivant laquelle P sources sont non-stationnaires¹⁷, alors que les autres sources sont stationnaires.

Nous notons I l'ensemble contenant les P indices des sources non-stationnaires inconnues. Le kurtosis standard $kurt_{s_q}(n)$ de toute source $s_q(n)$ avec $q \notin I$ prend alors les mêmes valeurs pour les deux instants n_1 et n_2 , si bien que $Dkurt_{s_q}(n_1, n_2) = 0$ ¹⁸. L'éq. (5.9) se simplifie alors en

$$Dkurt_y(n_1, n_2) = \sum_{q \in I} v_q^4 Dkurt_{s_q}(n_1, n_2). \quad (5.10)$$

Cela montre explicitement que ce paramètre différentiel dépend seulement des sources non-stationnaires. De plus, pour des sources données et deux instants n_1 et n_2 , celui-ci peut être vu comme une fonction $f(\cdot)$ du jeu de variables $\{v_q, q \in I\}$, c.-à-d. que $Dkurt_y(n_1, n_2)$ est égal à

$$f(v_q, q \in I) = \sum_{q \in I} v_q^4 \alpha_q \quad (5.11)$$

où les paramètres α_q sont égaux aux kurtosis différentiels $Dkurt_{s_q}(n_1, n_2)$ des sources non-stationnaires. Le type de fonction défini dans (5.11) a été largement étudié dans le contexte des méthodes de SAS à kurtosis standard, c.-à-d. les méthodes pour le cas $P = N$, parce que le kurtosis standard utilisé comme critère de SAS dans ce cas peut être exprimé sous la forme (5.11)¹⁹.

Le résultat suivant a été établi (voir [97] p. 173 pour la configuration à 2 sources et [61] pour une preuve générale). Supposons que tous les paramètres α_q avec $q \in I$ sont non nuls, c.-à-d. que toutes les sources non-stationnaires ont des kurtosis différentiels non nuls pour les deux instants n_1 et n_2 . Considérons les valeurs de la fonction (5.11) sur la sphère unité de dimension P , c.-à-d. sur l'ensemble $\{v_q, q \in I\}$ tel que

$$\sum_{q \in I} v_q^2 = 1. \quad (5.12)$$

¹⁷Dans la suite, le nombre \tilde{N} de sources non-stationnaires est supposé être égal au nombre P d'observations.

¹⁸Notons que la stationnarité complète des sources $s_q(n)$ avec $q \notin I$ est suffisante mais pas nécessaire pour notre méthode : nous avons seulement besoin que leurs kurtosis différentiels (et leurs puissances différentielles introduites ci-après) soient nuls pour les instants considérés. En dehors du cas stationnaire, les kurtosis différentiels sont nuls par exemple quand l'aplatissement (mesuré par le kurtosis normalisé défini par $kurt_s(n)/E\{s(n)^2\}^2$) et la puissance au carré des sources varient de façon inversement proportionnelle.

¹⁹Dans les approches standard, la somme pour $q \in I$ dans (5.11) est réalisée pour toutes les $P = N$ sources et les paramètres α_q sont égaux aux kurtosis standard $kurt_{s_q}(n)$ de toutes les sources. Néanmoins, cela n'a pas d'influence sur l'argumentation qui suit, basée sur les propriétés générales du type de fonction défini par (5.11).

Les résultats obtenus dans [61], [97] impliquent dans notre cas que les maxima de la valeur absolue de $f(v_q, q \in I)$ sur la sphère unité sont tous les points tels qu'une seule variable v_q , avec $q \in I$, est non nulle. L'éq. (5.7) montre que le signal de sortie $y(n)$ contient alors une contribution d'une seule source non-stationnaire (et des contributions de toutes les sources stationnaires). Nous réalisons donc bien la SAS partielle voulue pour une des sources non-stationnaires.

Le dernier aspect de notre méthode qui doit être précisé est comment sélectionner la matrice \mathbf{B} et comment contraindre le vecteur \mathbf{w} (qui est le paramètre contrôlé en pratique, et non pas \mathbf{v}) pour que les variables $\{v_q, q \in I\}$ vérifient la condition (5.12). Nous définissons d'abord la puissance différentielle $Dpow_y(n_1, n_2)$ d'un signal y entre deux instants n_1 et n_2 par

$$Dpow_y(n_1, n_2) = E \{y(n_2)^2\} - E \{y(n_1)^2\}. \quad (5.13)$$

En utilisant l'indépendance des signaux sources, on peut montrer facilement que, comme pour le kurtosis différentiel (5.10), nous avons

$$Dpow_y(n_1, n_2) = \sum_{q \in I} v_q^2 Dpow_{s_q}(n_1, n_2). \quad (5.14)$$

L'indétermination d'échelle propre à la SAS permet de remettre à l'échelle les puissances différentielles des P sources non-stationnaires avec des facteurs positifs. Donc, sous réserve que ces P puissances différentielles soient strictement positives pour les instants n_1 et n_2 considérés (dans le cas de puissances différentielles toutes négatives, on peut échanger les instants n_1 et n_2), elles peuvent être supposées égales à 1 sans perte de généralité. L'éq. (5.14) s'écrit alors

$$Dpow_y(n_1, n_2) = \sum_{q \in I} v_q^2 \quad (5.15)$$

si bien que l'équation de contrainte (5.12) peut être exprimée en termes de puissance différentielle unité pour le signal de sortie $y(n)$. Nous introduisons ensuite une extension différentielle de l'étage de décorrélation et de normalisation des méthodes à kurtosis standard, c.-à-d. que nous cherchons à obtenir une matrice \mathbf{B} qui permette de normaliser la puissance différentielle de $y(n)$ en normalisant simplement le vecteur d'extraction \mathbf{w} . Dans ce but, nous définissons la matrice d'autocorrélation différentielle de $\mathbf{x}(n)$ par

$$\mathbf{DR}_\mathbf{x}(n_1, n_2) = \mathbf{R}_\mathbf{x}(n_2) - \mathbf{R}_\mathbf{x}(n_1) \quad (5.16)$$

où $\mathbf{R}_\mathbf{x}(n_i) = E \{\mathbf{x}(n_i)\mathbf{x}(n_i)^t\}$ est la matrice de corrélation standard de $\mathbf{x}(n)$ à l'instant n_i . Considérons maintenant la décomposition en vecteurs propres ([82] p. 393) de la matrice symétrique réelle $\mathbf{DR}_\mathbf{x}(n_1, n_2)$:

$$\mathbf{DR}_\mathbf{x}(n_1, n_2) = \mathbf{E}\mathbf{\Delta}\mathbf{E}^t \quad (5.17)$$

où \mathbf{E} est une matrice $P \times P$ orthogonale réelle et $\mathbf{\Delta}$ est une matrice $P \times P$ diagonale. On peut montrer comme suit que les éléments de $\mathbf{\Delta}$ sont strictement positifs. Les éq. (4.2) et (5.16) donnent

$$\begin{aligned} \mathbf{DR}_x(n_1, n_2) &= \mathbf{R}_x(n_2) - \mathbf{R}_x(n_1) \\ &= \mathbf{A}\mathbf{R}_s(n_2)\mathbf{A}^t - \mathbf{A}\mathbf{R}_s(n_1)\mathbf{A}^t \\ &= \mathbf{A}(\mathbf{R}_s(n_2) - \mathbf{R}_s(n_1))\mathbf{A}^t \\ &= \mathbf{A}\mathbf{DR}_s(n_1, n_2)\mathbf{A}^t \end{aligned} \quad (5.18)$$

où $\mathbf{DR}_s(n_1, n_2)$ est une matrice diagonale en raison de l'indépendance des sources et ses éléments d'indices $q \notin I$ sont nuls en raison de la stationnarité des sources correspondantes. On peut vérifier facilement que les colonnes de \mathbf{A} d'indices $q \notin I$ ne donnent aucun terme dans la partie droite de (5.18), si bien que nous avons aussi

$$\mathbf{DR}_x(n_1, n_2) = \check{\mathbf{A}}\mathbf{\Lambda}\check{\mathbf{A}}^t \quad (5.19)$$

où $\check{\mathbf{A}}$ est composée des P colonnes de \mathbf{A} d'indices $q \in I$ et la matrice diagonale $\mathbf{\Lambda}$ contient les puissances différentielles des P sources non-stationnaires pour les deux instants n_1 et n_2 . Alors que (5.18) met en jeu des matrices rectangulaires, (5.19) contient seulement des matrices $P \times P$.

Si nous supposons que $\check{\mathbf{A}}$ est inversible, le théorème 4.2.1 de [82] p. 141 implique que $\mathbf{DR}_x(n_1, n_2)$ est définie positive si et seulement si $\mathbf{\Lambda}$ est aussi définie positive. Donc, si nous supposons à nouveau que les puissances différentielles des sources non-stationnaires sont strictement positives (pour les instants n_1 et n_2), (5.17) et (5.19) montrent que toutes les valeurs de la diagonale de $\mathbf{\Delta}$ sont strictement positives. De plus, si ces puissances différentielles sont mises à l'échelle unité, alors $\mathbf{\Lambda} = \mathbf{I}$ et (5.19) devient

$$\mathbf{DR}_x(n_1, n_2) = \check{\mathbf{A}}\check{\mathbf{A}}^t. \quad (5.20)$$

Comme $\mathbf{\Delta}$ n'a que des valeurs positives, elle a une racine carrée réelle et nous pouvons définir la matrice de blanchiment différentiel \mathbf{B} par

$$\mathbf{B} = \mathbf{\Delta}^{-1/2}\mathbf{E}^t. \quad (5.21)$$

Donc, avec \mathbf{z} défini par (5.1) nous avons

$$\begin{aligned} \mathbf{DR}_z(n_1, n_2) &= \mathbf{B}\mathbf{DR}_x(n_1, n_2)\mathbf{B}^t \\ &= \mathbf{\Delta}^{-1/2}\mathbf{E}^t\mathbf{E}\mathbf{\Delta}\mathbf{E}^t\mathbf{E}\mathbf{\Delta}^{-1/2} \\ &= \mathbf{I}. \end{aligned} \quad (5.22)$$

Considérons maintenant le signal de sortie $y(n)$ défini par (5.2). Nous avons

$$\begin{aligned} Dpow_y(n_1, n_2) &= E\{y(n_2)^2\} - E\{y(n_1)^2\} \\ &= \mathbf{w}^t\mathbf{R}_z(n_2)\mathbf{w} - \mathbf{w}^t\mathbf{R}_z(n_1)\mathbf{w} \\ &= \mathbf{w}^t\mathbf{DR}_z(n_1, n_2)\mathbf{w} \\ &= \|\mathbf{w}\|^2. \end{aligned} \quad (5.23)$$

En tenant compte des relations (5.15) et (5.23), nous obtenons

$$\|\mathbf{w}\|^2 = \sum_{q \in I} v_q^2. \quad (5.24)$$

En forçant notre vecteur \mathbf{w} à être de norme unité, nous contraignons ainsi le vecteur $\{v_q, q \in I\}$ à être sur la sphère unité. Donc, comme expliqué ci-dessus, les maxima de la valeur absolue de $Dkurt_{\mathbf{w}^t \mathbf{z}}(n_1, n_2)$ sur la surface de contrainte $\|\mathbf{w}\| = 1$ correspondent rigoureusement aux points de séparation partielle.

5.3 Résumé des méthodes proposées

La première méthode qui résulte de l'analyse ci-dessus se décompose en plusieurs étapes :

Etape 1 Sélectionner deux intervalles temporels non recouvrants pour estimer les paramètres statistiques (kurtosis, autocorrélation et puissance) aux deux instants n_1 et n_2 . Dans la plupart des cas, nous avons obtenu de meilleurs résultats avec des intervalles tels que les puissances différentielles des sources (qui peuvent être estimées grossièrement par les puissances différentielles des observations) sont élevées. Ces intervalles doivent être tels que toutes les sources non-stationnaires²⁰ aient des puissances différentielles positives (définies par (5.13)) et des kurtosis différentiels non nuls²¹.

Les signes des puissances différentielles des sources peuvent être obtenus en estimant la matrice de corrélation différentielle \mathbf{DR}_x et en calculant sa décomposition en vecteurs propres $\mathbf{E}\mathbf{\Delta E}^t$. En effet, si toutes les valeurs propres sont strictement positives, nous avons prouvé ci-dessus que les P sources utiles ont toutes des puissances différentielles strictement positives (si toutes les valeurs propres sont strictement négatives, nous pouvons permuter les deux in-

²⁰"non-stationnaire" signifie ici "non-stationnaire à long terme". Plus précisément, toutes les sources doivent être stationnaires à l'intérieur de chacun des deux intervalles temporels "courts" associés aux instants n_1 et n_2 , de sorte que leurs statistiques puissent être estimées pour chacun de ces intervalles, par moyennage temporel. Cela correspond à la "stationnarité à court terme". Les sources "utiles" évoquées ci-dessus (resp. sources de "bruit") sont donc des signaux dont les statistiques doivent varier (resp. ne pas varier) d'un des intervalles considérés à l'autre, c.-à-d. des sources qui sont non-stationnaires à long terme (resp. stationnaires à long terme).

²¹Notons que l'hypothèse $Dkurt_{s_q}(n_1, n_2) \neq 0, \forall q \in I$ ne nécessite pas que les distributions des sources aient des aplatissements différents aux deux instants n_1 et n_2 . En effet, l'aplatissement se mesure par le kurtosis normalisé défini par $kurt_s(n)/E\{s(n)^2\}^2$, si bien qu'une source peut avoir des kurtosis non-normalisés différents mais le même aplatissement aux deux instants tant que cette source a des puissances différentes à ces instants. Donc, excepté le cas très spécifique où les kurtosis normalisés et les puissances au carré des sources varient de façon inversement proportionnelle, nous sommes sûrs que les sources ont des kurtosis différentiels non nuls.

tervalles temporels pour les rendre toutes positives).

[Étape 2] Calculer la matrice $\mathbf{B} = \mathbf{\Delta}^{-1/2} \mathbf{E}^t$ à partir des matrices \mathbf{E} et $\mathbf{\Delta}$ définies ci-dessus. Cette matrice réalise une décorrélation et une normalisation différentielles des observations, c.-à-d. qu'elle donne un vecteur $\mathbf{z}(n)$ défini par (5.1) qui vérifie (5.22).

[Étape 3] Créer un signal de sortie $y(n)$ défini par (5.2), où \mathbf{w} est un vecteur qui satisfait $\|\mathbf{w}\| = 1$ et qui est adapté pour maximiser la valeur absolue du kurtosis différentiel de $y(n)$, défini par (5.4). Divers algorithmes peuvent être utilisés pour réaliser cette optimisation, par exemple en développant des versions différentielles d'algorithmes précédemment proposés pour le cas $P = N$.

L'approche la plus classique est basée sur une optimisation de type gradient [97]. Nous préférons ici utiliser une méthode améliorée déduite de l'algorithme standard à point fixe FastICA [98], qui a deux avantages sur l'algorithme du gradient, c.-à-d. sa convergence rapide et son absence de paramètre réglable. Brièvement, comme l'algorithme standard FastICA, notre algorithme linéaire instantané différentiel à point fixe, noté DFICA par la suite, tire parti du fait que si un vecteur d'extraction \mathbf{w} optimise le critère $|Dkurt_y(n_1, n_2)|$, alors le gradient de $Dkurt_y(n_1, n_2)$ a la même direction que \mathbf{w} (voir [97] p. 178). Notre équation de mise à jour utilise donc le gradient du kurtosis différentiel qui est calculé à l'annexe D, c.-à-d. qu'elle utilise l'assignation

$$\mathbf{w} \propto \frac{\partial Dkurt_y(n_1, n_2)}{\partial \mathbf{w}}. \quad (5.25)$$

Plus précisément, en partant d'un vecteur à norme unité quelconque \mathbf{w} , notre algorithme DFICA consiste à réitérer les deux opérations suivantes :

1) Mise à jour différentielle de \mathbf{w}

$$\begin{aligned} \mathbf{w} &\leftarrow E \{ \mathbf{z}(n_2)(\mathbf{w}^t \mathbf{z}(n_2))^3 \} - E \{ \mathbf{z}(n_1)(\mathbf{w}^t \mathbf{z}(n_1))^3 \} - 3 [\mathbf{R}_z(n_1) + (1 + \mathbf{w}^t \mathbf{R}_z(n_1) \mathbf{w}) \mathbf{I}] \mathbf{w} \\ &\propto \frac{\partial Dkurt_y(n_1, n_2)}{\partial \mathbf{w}} \end{aligned} \quad (5.26)$$

où les paramètres statistiques sont estimés par moyennage temporel.

2) Normalisation de \mathbf{w} , pour obtenir l'égalité $\|\mathbf{w}\| = 1$, c.-à-d.

$$\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}. \quad (5.27)$$

[Étape 4] Le signal source non-stationnaire $y(n)$ extrait à l'étape 3 est ensuite utilisé pour soustraire ses contributions de tous les signaux d'observation. Les signaux résultants sont ensuite réutilisés avec la procédure complète

décrite ci-dessus, extrayant ainsi une autre source, et ainsi de suite jusqu'à ce que toutes les sources non-stationnaires soient extraites. Ceci correspond à une procédure à déflation [61], mais une version différentielle de cette procédure doit ici être définie.

Cette déflation différentielle opère en réalité comme la déflation standard, à l'exception du fait que les paramètres statistiques sont remplacés par leur version différentielle pour estimer les valeurs a_{kl} de la matrice de mélange \mathbf{A} . En effet, nous prouvons en annexe E que le facteur d'échelle de la contribution d'une source extraite $s_l(n)$ dans la $k^{\text{ème}}$ observation peut être obtenu en estimant la corrélation différentielle $Dcorr_{yx_k}(n_1, n_2)$ définie par

$$Dcorr_{yx_k}(n_1, n_2) = E \{y(n_2)x_k(n_2)\} - E \{y(n_1)x_k(n_1)\} \quad (5.28)$$

dont on montre qu'elle est égale à la valeur a_{kl} de la matrice de mélange \mathbf{A} .

À la place de la version à déflation de notre méthode DFICA, une extension différentielle de l'approche symétrique décrite dans [98] peut être utilisée. La méthode DFICA symétrique résultante est composée des étapes 1 et 2, suivies par la réitération des deux opérations :

- 1) Mise à jour différentielle des P vecteurs \mathbf{w}_i , avec $i = 1, \dots, P$, en utilisant (5.26).
- 2) Orthogonalisation symétrique des vecteurs \mathbf{w}_i , c.-à-d.

$$\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^t)^{-1/2} \mathbf{W} \quad (5.29)$$

où $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_P]^t$.

5.4 Preuves de convergence

Considérons les expressions de mise à jour (5.26)-(5.27) de notre algorithme différentiel à déflation. Le signal de sortie $y(n)$ peut être exprimé en fonction des sources suivant (5.5) avec $\mathbf{v} = \mathbf{M}^t \mathbf{w}$ et $\mathbf{M} = \mathbf{B}\mathbf{A}$. De plus, nous avons prouvé ci-dessus que le kurtosis différentiel de $y(n)$ dépend seulement des sources utiles (dont les indices appartiennent à I) et de la sous-matrice associée $\tilde{\mathbf{A}}$. Donc, en considérant la composante globale $\tilde{y}(n)$ de $y(n)$ correspondant aux sources non-stationnaires, on peut réaliser le changement de variable $\tilde{\mathbf{v}} = \tilde{\mathbf{M}}^t \mathbf{w}$ où la matrice $\tilde{\mathbf{M}} = \mathbf{B}\tilde{\mathbf{A}}$ combine les étages de mélange et de blanchiment différentiel pour les sources non-stationnaires. Nous avons donc

$$\tilde{y}(n) = \tilde{\mathbf{v}}^t \tilde{\mathbf{s}}(n) \quad (5.30)$$

où le vecteur colonne $\tilde{\mathbf{s}}(n)$ contient seulement les valeurs des sources non-stationnaires, c.-à-d. $s_q(n)$ avec $q \in I$. Pour la lisibilité, nous omettrons ici les

instants n_1 et n_2 dans la notation du kurtosis différentiel.

La loi de mise à jour

$$\mathbf{w} \leftarrow \frac{\partial Dkurt_{\mathbf{w}^t \mathbf{z}}}{\partial \mathbf{w}} \quad (5.31)$$

s'écrit donc

$$\mathbf{w} \leftarrow \frac{\partial \check{\mathbf{v}}}{\partial \mathbf{w}} \frac{\partial Dkurt_{\mathbf{w}^t \mathbf{z}}}{\partial \check{\mathbf{v}}}. \quad (5.32)$$

Comme $Dkurt_{\mathbf{w}^t \mathbf{z}} = Dkurt_{\mathbf{v}^t \mathbf{s}} = Dkurt_{\check{\mathbf{v}}^t \check{\mathbf{s}}}$, nous pouvons réécrire (5.32) suivant

$$\mathbf{w} \leftarrow \check{\mathbf{M}} \frac{\partial Dkurt_{\check{\mathbf{v}}^t \check{\mathbf{s}}}}{\partial \check{\mathbf{v}}}. \quad (5.33)$$

L'expression de mise à jour de $\check{\mathbf{v}}$ s'écrit donc

$$\check{\mathbf{v}} = \check{\mathbf{M}}^t \mathbf{w} \leftarrow \check{\mathbf{M}}^t \check{\mathbf{M}} \frac{\partial Dkurt_{\check{\mathbf{v}}^t \check{\mathbf{s}}}}{\partial \check{\mathbf{v}}}. \quad (5.34)$$

En combinant (5.1) et (5.20), on obtient

$$\begin{aligned} \mathbf{D}\mathbf{R}_{\mathbf{z}}(n_1, n_2) &= \mathbf{B}\mathbf{D}\mathbf{R}_{\mathbf{x}}(n_1, n_2)\mathbf{B}^t \\ &= \mathbf{B}\check{\mathbf{A}}\check{\mathbf{A}}^t\mathbf{B}^t \\ &= \check{\mathbf{M}}\check{\mathbf{M}}^t \end{aligned} \quad (5.35)$$

et en considérant (5.22)

$$\check{\mathbf{M}}\check{\mathbf{M}}^t = \mathbf{I}. \quad (5.36)$$

$\check{\mathbf{M}}$ est donc orthogonale et nous avons

$$\check{\mathbf{M}}^t\check{\mathbf{M}} = \mathbf{I}. \quad (5.37)$$

Ainsi l'expression de mise à jour (5.34) devient

$$\check{\mathbf{v}} \leftarrow \frac{\partial Dkurt_{\check{\mathbf{v}}^t \check{\mathbf{s}}}}{\partial \check{\mathbf{v}}}. \quad (5.38)$$

En considérant l'éq. (5.10) et en prenant en compte le fait que le vecteur $\check{\mathbf{v}}$ contient seulement les coefficients des sources non-stationnaires s_i , $i \in I$, nous pouvons exprimer l'équation de mise à jour de chaque composante de $\check{\mathbf{v}}$ dans (5.38) suivant

$$\check{v}_i \leftarrow 4Dkurt_{s_i}\check{v}_i^3 \propto Dkurt_{s_i}\check{v}_i^3 \quad (5.39)$$

En choisissant $j \in I$ de sorte que $Dkurt_{s_j} \neq 0$ et $\check{v}_j \neq 0$, nous obtenons

$$\frac{|\check{v}_i|}{|\check{v}_j|} \leftarrow \frac{|Dkurt_{s_i}|}{|Dkurt_{s_j}|} \left(\frac{|\check{v}_i|}{|\check{v}_j|} \right)^3, \quad \forall i \in I. \quad (5.40)$$

Cette équation de récurrence nous permet d'obtenir l'expression analytique de $\frac{|\check{v}_i|}{|\check{v}_j|}$:

$$\frac{|\check{v}_i(k)|}{|\check{v}_j(k)|} = \sqrt{\left| \frac{Dkurt_{s_j}}{Dkurt_{s_i}} \right|} \cdot \left(\sqrt{\left| \frac{Dkurt_{s_i}}{Dkurt_{s_j}} \right|} \left| \frac{\check{v}_i(0)}{\check{v}_j(0)} \right| \right)^{3^k}. \quad (5.41)$$

Toutes les composantes $\check{v}_i(k)$ sauf celle $\check{v}_j(k)$ qui vérifie $j = \operatorname{argmax}_p (\sqrt{|Dkurt_{s_p}|} |\check{v}_p(0)|)$ deviennent donc rapidement petites comparées à $\check{v}_j(k)$. Avec la normalisation $\|\mathbf{w}\| = 1$ qui implique $\|\check{\mathbf{v}}\| = 1$ en raison de l'orthogonalité de la matrice \mathbf{M} , nous concluons que $\check{v}_j(k) \rightarrow \pm 1$ et que $\check{v}_i(k) \rightarrow 0, \forall i \neq j$.

Nous avons donc prouvé que $\check{\mathbf{v}}$ converge vers un vecteur ayant seulement un élément non nul (égal à ± 1). Cela implique la séparation partielle des sources utiles, car $\check{y}(n) = \check{\mathbf{v}}^t \check{\mathbf{s}}(n)$. Ainsi, nous avons prouvé la convergence globale (c.-à-d. quel que soit $\check{\mathbf{v}}(0)$) de notre algorithme. De plus, cette convergence est cubique (comme pour l'algorithme (sur-)déterminé FastICA [98]), ce qui veut dire une convergence très rapide.

Pour notre algorithme symétrique, la même approche que [142] peut être utilisée dans le cas différentiel. En effet, dans le cas (sur-)déterminé, Oja a prouvé que quand les mises à jour obéissent aux lois

$$v_{pq} \leftarrow kurt_{s_q}(n) v_{pq}^3 \quad (5.42)$$

pour la $q^{\text{ème}}$ composante de chaque vecteur d'extraction \mathbf{v}_p , ces mises à jour combinées avec l'orthogonalisation symétrique $\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^t)^{-1/2} \mathbf{W}$ entraînent une convergence cubique vers les points de séparation et la stabilité de l'algorithme dans le voisinage des points séparants, comme dans l'algorithme à déflation.

En remarquant que notre algorithme différentiel donne (5.39), qui est équivalent à (5.42) avec $q = i$, excepté que les kurtosis standard $kurt_{s_q}(n)$ sont remplacés par leur version différentielle $Dkurt_{s_i}(n_1, n_2)$, nous prouvons rigoureusement la convergence cubique de notre algorithme différentiel symétrique.

5.5 Conclusion

Dans ce chapitre, nous avons proposé un algorithme de type point fixe pour la séparation partielle de mélanges linéaires instantanés sous-déterminés. Nous avons introduit le processus de blanchiment différentiel qui permet d'obtenir un algorithme d'optimisation à point fixe dont nous démontrons la convergence cubique et la stabilité. Deux versions de notre algorithme, baptisé DFICA, sont proposées : une version à déflation qui estime les sources itérativement et une version parallèle. Pour estimer les contributions de sources à soustraire des observations durant le processus de déflation, nous proposons de calculer la corrélation différentielle entre deux instants.

Chapitre 6

Méthode de SAS différentielle proposée pour les mélanges convolutifs

6.1 Introduction

Dans le cas convolutif, nous étendons ici notre algorithme temporel à point fixe dénommé C-FICA aux mélanges sous-déterminés. Notre approche est basée sur le concept de séparation différentielle de sources et sur un processus de blanchiment différentiel convolutif non-causal qui permet l'utilisation de mises à jour de type point fixe pour l'optimisation du kurtosis différentiel introduit dans le chapitre précédent. L'estimation des contributions des sources est réalisée au moyen d'un critère quadratique différentiel que nous définissons.

6.2 Utilisation du kurtosis différentiel dans le cas convolutif

Nous supposons ici avoir le même modèle de mélange que celui décrit dans le chapitre 1 (1.4) excepté que cette fois le nombre de sources est supérieur au nombre d'observations ($N > P$). Nous cherchons ici à étendre notre algorithme de séparation convolutif C-FICA au cas sous-déterminé, en utilisant le concept de SAS différentielle introduit dans [67]. L'algorithme résultant est donc appelé C-DFICA dans la suite. Définissons à nouveau $y(n)$ par (2.7), (2.8) et (2.11) où $\tilde{\mathbf{B}}$ et \mathbf{w} seront sélectionnés comme expliqué ci-dessous.

Comme dans le chapitre 5, nous notons $kurt_y(n_1)$ et $kurt_y(n_2)$ resp. les kurtosis non-normalisés de $y(n)$ pour deux instants n_1 et n_2 , et nous définissons son kurtosis différentiel non-normalisé par (5.4). En combinant les définitions (2.7), (2.8) et (2.11) de $y(n)$ avec (2.4), on montre que $y(n)$ est une combinaison

linéaire de versions décalées des processus $u_k(n)$ c.-à-d.

$$y(n) = \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd} u_q(n-d) \quad (6.1)$$

où D_{min} et D_{max} sont obtenus à partir des ordres des filtres RIF impliqués dans (2.4) et (2.7).

Les processus $u_q(n)$ sont ici encore supposés centrés, à valeurs réelles, mutuellement et temporellement indépendants. P d'entre eux sont encore supposés non-stationnaires à long terme (donc non identiquement distribués) et correspondent aux sources utiles, alors que les $N - P$ autres sont stationnaires et correspondent aux sources de "bruit". En utilisant la multi-linéarité du kurtosis pour des variables aléatoires indépendantes, nous en déduisons à partir de (5.4) et (6.1)

$$\begin{aligned} Dkurt_y(n_1, n_2) &= \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd}^4 [kurt_{u_q}(n_2 - d) - kurt_{u_q}(n_1 - d)] \\ &= \sum_{q=1}^N \sum_{d=D_{min}}^{D_{max}} v_{qd}^4 \alpha_{qd}, \end{aligned} \quad (6.2)$$

où

$$\alpha_{qd} = Dkurt_{u_q}(n_1 - d, n_2 - d) \quad (6.3)$$

est défini comme dans (5.4). Comme dans le cas linéaire instantané, notons I l'ensemble contenant les P indices des sources non-stationnaires inconnues. Les hypothèses ci-dessus de notre concept de SAS différentielle signifient que seuls les processus $u_q(n)$ avec $q \notin I$ sont supposés avoir le même kurtosis à des instants différents, c.-à-d.

$$\forall q \notin I, \forall d \in [D_{min}..D_{max}], \quad kurt_{u_q}(n_1 - d) = kurt_{u_q}(n_2 - d). \quad (6.4)$$

Ainsi nous avons

$$Dkurt_y(n_1, n_2) = \sum_{q \in I} \sum_{d=D_{min}}^{D_{max}} v_{qd}^4 \alpha_{qd}. \quad (6.5)$$

Nous voyons donc que les sources de bruit (dont les indices n'appartiennent pas à l'ensemble I) sont invisibles pour notre critère de séparation différentielle. En faisant une extension du cas instantané, nous considérons ici les valeurs de (6.5) sur la sphère unité de dimension²² $P \times D$ où $D = D_{max} - D_{min} + 1$, c.-à-d. l'ensemble $\{v_{qd}, q \in I, d \in [D_{min}, \dots, D_{max}]\}$ tel que

$$\sum_{q \in I} \sum_{d=D_{min}}^{D_{max}} v_{qd}^2 = 1. \quad (6.6)$$

²²Les mélanges convolutifs entraînent une légère approximation concernant les points de cette sphère que l'on peut atteindre, comme on l'expliquera dans la section 6.3.

On peut d'abord montrer facilement que la puissance différentielle de $y(n)$, à nouveau définie par (5.13), s'écrit ici

$$Dpow_y(n_1, n_2) = \sum_{q \in I} \sum_{d=D_{min}}^{D_{max}} v_{qd}^2 Dpow_{u_q}(n_1 - d, n_2 - d). \quad (6.7)$$

Chaque processus $u_q(n)$ avec $q \in I$ est supposé ici être non-stationnaire entre n_1 et n_2 . Néanmoins, il est supposé être identiquement distribué à tous les instants $n_i - d$ où d varie de D_{min} à D_{max} . En pratique, cela signifie que nous sélectionnons n_1 et n_2 tels que $n_2 - n_1 \gg D$, et que $u_q(n)$ est non-stationnaire à long terme entre les instants n_1 et n_2 mais stationnaire à court terme à l'intérieur de chaque fenêtre $[n_i - D_{min}, n_i - D_{max}]$ avec $i = 1$ ou 2 . Les termes $Dpow_{u_q}(n_1 - d, n_2 - d)$ dans (6.7) ne dépendent donc pas de d , c.-à-d. que (6.7) implique seulement une puissance différentielle $Dpow_{u_q}(n_1, n_2)$ unique pour chaque source non-stationnaire. Ces puissances différentielles peuvent être remises à l'échelle avec un facteur positif grâce à l'indétermination d'échelle propre à la SAS. Ainsi, si elles sont strictement positives, elles peuvent être supposées égales à 1 si bien que (6.7) devient

$$Dpow_y(n_1, n_2) = \sum_{q \in I} \sum_{d=D_{min}}^{D_{max}} v_{qd}^2. \quad (6.8)$$

Donc, si $\mathbf{DR}_{\tilde{\mathbf{x}}}(n_1, n_2) = \mathbf{E}\mathbf{\Delta}\mathbf{E}^t$ est la décomposition en vecteurs propres de $\mathbf{DR}_{\tilde{\mathbf{x}}}(n_1, n_2)$ définie pour $\tilde{\mathbf{x}}$ comme dans (5.16), nous définissons $\tilde{\mathbf{B}}$ par $\tilde{\mathbf{B}} = \mathbf{\Delta}^{-1/2}\mathbf{E}^t$. Ainsi, en définissant $\tilde{\mathbf{z}}$ par (2.8), nous pouvons prouver comme dans (5.22) que

$$\mathbf{DR}_{\tilde{\mathbf{z}}}(n_1, n_2) = \mathbf{I} \quad (6.9)$$

ce qui implique, en utilisant la même approche que dans (5.23) mais maintenant avec (2.11),

$$Dpow_y(n_1, n_2) = \|\mathbf{w}\|^2 \quad (6.10)$$

si bien que grâce à (6.8)

$$\|\mathbf{w}\|^2 = \sum_{q \in I} \sum_{d=D_{min}}^{D_{max}} v_{qd}^2. \quad (6.11)$$

Faire varier \mathbf{w} avec $\|\mathbf{w}\| = 1$ permet donc de contraindre de façon simple notre vecteur $\{v_{qd}, q \in I, d \in [D_{min}, \dots, D_{max}]\}$ à être sur la sphère unité (6.6) dans le cas convolutif sous-déterminé.

L'optimisation de la valeur absolue du kurtosis différentiel $Dkurt_y(n_1, n_2)$ défini par (6.2), sous la contrainte (6.6) appartient donc au même type de problème qu'avec (5.11)-(5.12), avec un jeu de variables notées ici v_{qd} au lieu de v_q . En appliquant à nouveau les résultats de [61], [97] qui étaient déjà utilisés

dans le chapitre 5, on garantit ici que les maxima de $|Dkurt_y(n_1, n_2)|$ sur la sphère unité correspondent aux points tels qu'une seule des variables considérées v_{qd} , avec $q \in I$, est non nulle. L'éq. (6.1) montre que le signal de sortie se compose alors seulement d'une version pondérée et décalée $e_l(n)$ du processus $u_l(n)$ associé à une source non-stationnaire (ainsi que des contributions des sources stationnaires).

Nous colorons ensuite $e_l(n)$ pour estimer ses contributions dans chaque observation. Ceci est réalisé ici en adaptant la solution de Wiener que nous avons utilisée dans la partie I du manuscrit, c.-à-d. en introduisant un filtrage de Wiener différentiel, que nous définissons comme suit.

Considérons la différence $x'_k(n)$ entre l'observation $x_k(n)$ et une version filtrée du signal extrait précédent $e_l(n)$ avec $l \in I$, c.-à-d.

$$x'_k(n) = x_k(n) - \sum_{r=-R'}^{R'} c(r)e_l(n-r) \quad (6.12)$$

où les $c(r)$ sont les coefficients d'un filtre RIF non causal. En raison des hypothèses de la section 2.1, $x_k(n)$ est un mélange RIF causal de tous les processus d'innovation, c.-à-d. que (2.4) nous donne

$$x_k(n) = \sum_{q=1}^N \sum_{r=0}^L h_{kq}(r)u_q(n-r). \quad (6.13)$$

De plus,

$$e_l(n) = \beta_l u_l(n - r_l) + \sum_{q \notin I} \sum_p \beta_{qp} u_q(n - p) \quad (6.14)$$

où les bornes de p ne sont pas précisés ici car le termes correspondants s'annulent plus bas dans (6.18). Donc, si R' est assez grand pour que $r_l - R' \leq 0$ et $r_l + R' \geq L$, (6.12)-(6.14) nous donnent

$$\begin{aligned} x'_k(n) &= \sum_{r=r_l-R'}^{r_l+R'} [h_{kl}(r) - \beta_l c(r - r_l)] u_l(n - r) \\ &+ \sum_{q \in I, q \neq l} \sum_{r=0}^L h_{kq}(r) u_q(n - r) + \sum_{q \notin I} \sum_p \gamma_{kq}(p) u_q(n - p) \end{aligned} \quad (6.15)$$

où les poids $\gamma_{kq}(p)$ dépendent de $c(r)$, $h_{kq}(r)$ et β_{qp} . Ces termes et les bornes sur p n'ont pas besoin d'être définis ici pour les mêmes raisons que plus haut. La puissance différentielle de $x'_k(n)$ est définie par

$$Dpow_{x'_k}(n_1, n_2) = E \{x'_k(n_2)^2\} - E \{x'_k(n_1)^2\}. \quad (6.16)$$

Sachant que tous les processus d'innovation sont centrés, mutuellement et temporellement indépendants, et que

$$Dpow_{u_q}(n_1 - p, n_2 - p) = 0, \forall p \in \{1..N\}, \forall q \notin I, \quad (6.17)$$

les éq. (6.15) et (6.16) donnent

$$\begin{aligned} Dpow_{x'_k}(n_1, n_2) &= \sum_{r=r_l-R'}^{r_l+R'} [h_{kl}(r) - \beta_l c(r - r_l)]^2 Dpow_{u_l}(n_1 - r, n_2 - r) \\ &+ \sum_{q \in I, q \neq l} \sum_{r=0}^L h_{kq}^2(r) Dpow_{u_q}(n_1 - r, n_2 - r). \end{aligned} \quad (6.18)$$

Comme nous avons supposé que toutes les puissances différentielles $Dpow_{u_l}(n_1 - r, n_2 - r)$ avec $l \in I$ sont strictement positives, l'éq. (6.18) montre que la minimisation de $Dpow_{x'_k}(n_1, n_2)$ en fonction de $c(r)$ correspond exactement à

$$h_{kl}(r) - \beta_l c(r - r_l) = 0, \quad \forall r \in [r_l - R', \dots, r_l + R']. \quad (6.19)$$

En comparant ce résultat avec (6.15), nous en déduisons que pour obtenir une observation modifiée $x'_k(n)$ où la contribution de la $l^{ème}$ source a été enlevée consiste à minimiser $Dpow_{x'_k}(n_1, n_2)$. Nous utilisons la méthode de Newton pour optimiser ce critère, qui est noté J dans la suite et peut être exprimé par

$$J = E \left\{ \left(x_k(n_2) - \sum_{r=-R'}^{R'} c(r) e_l(n_2 - r) \right)^2 \right\} - E \left\{ \left(x_k(n_1) - \sum_{r=-R'}^{R'} c(r) e_l(n_1 - r) \right)^2 \right\}. \quad (6.20)$$

L'équation de mise à jour du vecteur $\mathbf{c} = [c(-R'), \dots, c(R')]^t$ s'écrit

$$\mathbf{c} \leftarrow \mathbf{c} - \left(\frac{\partial^2 J}{\partial \mathbf{c}^2} \right)^{-1} \frac{\partial J}{\partial \mathbf{c}} \quad (6.21)$$

où $\frac{\partial J}{\partial \mathbf{c}}$ et $\frac{\partial^2 J}{\partial \mathbf{c}^2}$ sont respectivement le gradient et le Hessian du critère J . Calculons la $i^{ème}$ composante du gradient de J par rapport à \mathbf{c} :

$$\begin{aligned} \frac{\partial J}{\partial c_i} &= 2E \left\{ -e_l(n_2 - i) \cdot \left(x_k(n_2) - \sum_{r=-R'}^{R'} c(r) e_l(n_2 - r) \right) \right\} \\ &- 2E \left\{ -e_l(n_1 - i) \cdot \left(x_k(n_1) - \sum_{r=-R'}^{R'} c(r) e_l(n_1 - r) \right) \right\}. \end{aligned} \quad (6.22)$$

En particulier

$$\left. \frac{\partial J}{\partial c_i} \right|_{\mathbf{c}=0} = 2E \{ -e_l(n_2 - i) x_k(n_2) \} - 2E \{ -e_l(n_1 - i) x_k(n_1) \}. \quad (6.23)$$

Donc nous avons

$$\left. \frac{\partial J}{\partial \mathbf{c}} \right|_{\mathbf{c}=0} = -2 (\mathbf{r}_{e_l x_k}(n_2) - \mathbf{r}_{e_l x_k}(n_1)) \quad (6.24)$$

où $\mathbf{r}_{e_l x_k}(n_p)$ est le vecteur d'intercorrélation entre les signaux $e_l(n_p - i)$, avec $i \in [-R', R']$ et $x_k(n_p)$, avec $p = 1$ ou 2 . Calculons maintenant la $(i, j)^{\text{ème}}$ composante du Hessien de J :

$$\frac{\partial^2 J}{\partial c_i \partial c_j} = 2E \{e_l(n_2 - i)e_l(n_2 - j)\} - 2E \{e_l(n_1 - i)e_l(n_1 - j)\}. \quad (6.25)$$

Donc, nous avons

$$\frac{\partial^2 J}{\partial \mathbf{c}^2} = 2 (\mathbf{R}_{e_l}(n_2) - \mathbf{R}_{e_l}(n_1)) \quad (6.26)$$

où $\mathbf{R}_{e_l}(n_p)$ est la matrice d'autocorrélation associée aux valeurs de $e_l(n)$ autour de $n = n_p$. En initialisant \mathbf{c} avec $\mathbf{c} = 0$, la valeur de \mathbf{c} après une itération de l'algorithme de Newton s'écrit

$$\begin{aligned} \mathbf{c}' &= - \left(\frac{\partial^2 J}{\partial \mathbf{c}^2} \right)^{-1} \frac{\partial J}{\partial \mathbf{c}} \Big|_{\mathbf{c}=0} \\ &= -\frac{1}{2} (\mathbf{R}_{e_l}(n_2) - \mathbf{R}_{e_l}(n_1))^{-1} (-2) (\mathbf{r}_{e_l x_k}(n_2) - \mathbf{r}_{e_l x_k}(n_1)) \\ &= (\mathbf{R}_{e_l}(n_2) - \mathbf{R}_{e_l}(n_1))^{-1} (\mathbf{r}_{e_l x_k}(n_2) - \mathbf{r}_{e_l x_k}(n_1)). \end{aligned} \quad (6.27)$$

Sachant que l'algorithme de Newton atteint un point stationnaire en une itération pour les critères quadratiques et que notre critère est quadratique positif, nous prouvons que le vecteur \mathbf{c}' correspond au maximum global de J . Notons que cette expression est une version différentielle du filtre de Wiener standard défini par (2.46).

6.3 Résumé de la méthode proposée

Nous proposons la procédure suivante pour réaliser la séparation partielle rapide des mélanges convolutifs sous-déterminés. Cette procédure est basée sur une version adaptée d'algorithme à point fixe et sur notre processus de filtrage différentiel pour estimer les contributions des sources dans les observations dans le contexte d'une approche à déflation.

[Etape 1] Sélectionner deux intervalles temporels non recouvrants en utilisant le même type d'approche que dans le chapitre 5.

[Etape 2] Estimer la matrice de corrélation différentielle $\mathbf{DR}_{\tilde{\mathbf{x}}}(n_1, n_2)$ du vecteur d'observations étendu $\tilde{\mathbf{x}}$. Calculer ensuite la décomposition en vecteurs propres de cette matrice. Cela nous donne une matrice \mathbf{E} dont les colonnes sont les vecteurs propres de norme unité de l'estimée de $\mathbf{DR}_{\tilde{\mathbf{x}}}(n_1, n_2)$, ainsi qu'une matrice diagonale $\mathbf{\Delta}$ qui contient les valeurs propres de l'estimée de $\mathbf{DR}_{\tilde{\mathbf{x}}}(n_1, n_2)$. Calculer ensuite la matrice $\tilde{\mathbf{B}} = \mathbf{\Delta}^{-1/2} \mathbf{E}^t$. Nous obtenons ainsi un vecteur $\tilde{\mathbf{z}}(n)$ défini par (2.8) qui vérifie (6.9).

Etape 3 Initialiser le vecteur d'extraction \mathbf{w} , en utilisant (2.29) comme pour l'algorithme C-FICA mais avec la nouvelle matrice $\tilde{\mathbf{B}}$ définie à l'étape 2. Créer un signal de sortie $y(n)$ défini par (2.11), où \mathbf{w} est un vecteur normalisé qui est adapté pour maximiser la valeur absolue du kurtosis différentiel de $y(n)$. La même approche que dans l'étape 3 de la section 5.2 peut être utilisée. Cela nous donne l'algorithme C-DFICA défini comme suit :

1) Mise à jour différentielle de \mathbf{w}

$$\begin{aligned} \mathbf{w} \leftarrow & E \left\{ \tilde{\mathbf{z}}(n_2)(\mathbf{w}^t \tilde{\mathbf{z}}(n_2))^3 \right\} - E \left\{ \tilde{\mathbf{z}}(n_1)(\mathbf{w}^t \tilde{\mathbf{z}}(n_1))^3 \right\} - 3 \left[\mathbf{R}_{\tilde{\mathbf{z}}}(n_1) + (1 + \mathbf{w}^t \mathbf{R}_{\tilde{\mathbf{z}}}(n_1) \mathbf{w}) \mathbf{I} \right] \mathbf{w} \\ & \propto \frac{\partial Dkurt_y(n_1, n_2)}{\partial \mathbf{w}} \end{aligned} \quad (6.28)$$

2) Normalisation de \mathbf{w} , pour vérifier la condition $\|\mathbf{w}\| = 1$, c.-à-d.

$$\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}. \quad (6.29)$$

Etape 4 Le processus non-stationnaire $u_l(n)$ extrait par $y(n)$ dans l'étape 3 est ensuite coloré au moyen de notre filtrage de Wiener différentiel introduit dans la section précédente afin d'estimer les contributions de la source correspondante dans chaque observation. Ce filtre minimise la puissance différentielle de la différence entre une observation et une version filtrée non causale $c_{kl}(n) * y(n)$ du signal $y(n)$. Nous avons démontré la consistance de cette minimisation et avons établi l'expression des filtres de coloration qui la réalisent, c.-à-d.

$$\mathbf{c}_{kl} = (\mathbf{R}_{e_l}(n_2) - \mathbf{R}_{e_l}(n_1))^{-1} (\mathbf{r}_{e_l x_k}(n_2) - \mathbf{r}_{e_l x_k}(n_1)). \quad (6.30)$$

Comme avec les fonctions statistiques kurtosis et puissance, nous avons prouvé que ce processus de coloration ne dépend pas des sources de bruit. Les signaux résultants $c_{kl}(n) * y(n)$ sont ensuite soustraits des observations. Les étapes 2 à 4 sont ensuite réitérées jusqu'à ce que tous les processus d'innovation soient extraits.

Nous n'introduisons pas ici de version symétrique pour cet algorithme différentiel destiné aux mélanges convolutifs. En effet, comme nous le développerons dans la prochaine section, les mélanges convolutifs peuvent être exprimés sous forme de mélanges instantanés de multiples versions décalées des sources et augmentent donc considérablement la dimensionnalité de l'espace d'optimisation par rapport au chapitre 5. Nous économisons donc énormément de mémoire et de temps de calcul en extrayant les processus d'innovation les uns après les autres.

6.4 Reformulation du mélange sous forme linéaire instantanée

Dans chacun des signaux de mélange de $\tilde{\mathbf{x}}(n)$ définis par (2.7), nous considérons ici seulement la composante engendrée par les sources non-stationnaires, car le critère de SAS ci-dessus dépend seulement de cette composante. L'analyse de la section 2.5 concernant les mélanges convolutifs (sur-)déterminés montre que ces mélanges peuvent être interprétés comme des mélanges linéaires instantanés de multiples versions pondérées et décalées des processus $u_q(n)$ avec $q \in I$. En particulier, l'analyse montre que si Q est l'ordre de $\mathbf{H}(n)$ dans (2.4), et si l'inégalité

$$PL \geq \tilde{N}(Q + L) \quad (6.31)$$

est vérifiée, où $L = (2R + 1)$ est le nombre de décalages, P est le nombre de mélanges et \tilde{N} est le nombre de sources non-stationnaires, alors $\tilde{\mathbf{x}}(n)$ est un mélange linéaire instantané (sur-)déterminé de versions décalées et pondérées des processus $u_q(n)$ avec $q \in I$. La preuve de convergence obtenue dans la section 5.3 dans le cas instantané peut donc être appliquée à notre algorithme convolutif différentiel, ce qui garantit ici à nouveau la convergence globale cubique.

Si (6.31) n'est pas vérifiée, ce qui est en particulier le cas quand $P = \tilde{N}$, le mélange *reformulé de manière instantanée* est sous-déterminé. Cette sous-détermination est liée à l'ordre fini des filtres d'extraction équivalents de (2.5). Néanmoins, quand le rapport $PL/[\tilde{N}(Q + L)]$ associé à (6.31) tend vers 1 (ce qui est le cas quand $P = \tilde{N}$ et L est grand), notre configuration est presque déterminée c.-à-d. que la matrice de mélange instantanée équivalente est presque carrée. On peut donc encore obtenir quasiment une estimation d'un processus $u_q(n)$ en maximisant la valeur absolue du kurtosis différentiel non-normalisé sous la contrainte d'une puissance différentielle unitaire.

Evoquons maintenant l'approximation que nous avons mentionnée dans la section 6.2 concernant les points de la sphère unité accessibles en convolutif. Dans le cas linéaire instantané sous-déterminé, $\tilde{\mathbf{v}} = \tilde{\mathbf{M}}^t \mathbf{w}$, où $\tilde{\mathbf{M}}$ est une matrice orthogonale. Donc, faire varier \mathbf{w} permet d'atteindre n'importe quelle valeur de $\tilde{\mathbf{v}}$, et en particulier toute la sphère unité. Le même résultat s'applique aux mélanges convolutifs si (6.31) est vérifiée parce que, comme expliqué dans la section 2.5, ces mélanges peuvent être reformulés comme des mélanges linéaires instantanés. Au contraire, quand on considère des mélanges convolutifs tels que (6.31) n'est pas vérifiée, on ne peut pas garantir que tout point $\{v_{qd}, q \in I, d \in [D_{min}, \dots, D_{max}]\}$ de la sphère unité peut être atteint en faisant varier \mathbf{w} . Cependant, cet effet devient négligeable quand $PL/[\tilde{N}(K + L)]$ se rapproche de 1, comme expliqué ci-dessus. Ce phénomène est donc ignoré dans ce document.

6.5 Conclusion

Nous avons proposé dans ce chapitre un algorithme rapide de type point fixe opérant dans le domaine temporel et destiné à la séparation de mélanges convolutifs sous-déterminés. Notre approche, baptisée C-DFICA, est basée sur un processus de blanchiment différentiel spatio-temporel (ou convolutif) non-causal des observations qui permet d'appliquer des itérations d'optimisation à point fixe pour optimiser le critère du kurtosis différentiel. Nous avons prouvé sous certaines conditions la convergence et la stabilité de notre méthode à déflation vers les valeurs optimales du kurtosis différentiel. Nous proposons d'estimer les contributions des sources sur les capteurs grâce au calcul d'un filtre de Wiener différentiel non-causal qui permet de minimiser un critère quadratique différentiel défini comme la différence entre deux instants de la puissance de l'écart entre une observation et une version filtrée non-causale du processus d'innovation extrait.

Chapitre 7

Résultats expérimentaux

7.1 Introduction

Dans ce chapitre destiné à valider expérimentalement nos algorithmes différentiels rapides DFICA et C-DFICA, comme dans la partie précédente, nous commençons par illustrer le fonctionnement de nos méthodes en les testant sur des mélanges linéaires instantanés et convolutifs de sources de télécommunications et de sources audio. Pour chaque expérience réalisée, nous donnons aussi les performances de séparation obtenues avec les méthodes non-différentielles associées FastICA et C-FICA. Ensuite, nous effectuons une analyse statistique pour évaluer la robustesse de nos méthodes en fonction de la puissance du bruit stationnaire présent dans les mélanges. Nous comparons aussi les performances et la rapidité de notre algorithme C-DFICA avec une version différentielle de l'algorithme de Tugnait développée auparavant dans l'équipe.

7.2 Séparation de mélanges linéaires instantanés de signaux télécoms

Nous avons ici effectué une simulation de séparation d'un mélange linéaire instantané sous-déterminé de deux sources télécoms AM démodulées à 4 états et d'une source de bruit stationnaire. Nous avons choisi ici une source stationnaire uniformément répartie entre -2 et 2 (ce qui correspond à une source de puissance 1.3) et une matrice de mélange égale à

$$\mathbf{A} = \begin{bmatrix} 1 & 0.9 & 1 \\ 0.9 & 1 & 1 \end{bmatrix} \quad (7.1)$$

et dont la dernière colonne est associée à la source de bruit. Sur les deux domaines contenant chacun 100000 échantillons, les deux sources AM sont respectivement de puissances 1 et 0.01. Le SNR_{in} correspondant à cette configuration est de -14 dB en moyenne sur les deux domaines.

La figure 7.1 représente de haut en bas les deux sources AM sur une partie du domaine 1, les deux observations sur le même domaine et les deux sources séparées auxquelles ont été enlevées les composantes de bruit stationnaire. On constate que les sources séparées sont visuellement identiques aux sources réelles malgré la forte présence de bruit dans les observations.

Pour mesurer la performance de séparation, nous avons ici utilisé un critère défini à partir de la matrice $P \times N$ dite de performance $\mathbf{G} = [g_{ij}]$ qui est le produit de la matrice de mélange et de la matrice de séparation partielle estimée. Le critère de performances pour le $j^{\text{ème}}$ signal source s'écrit :

$$\forall j \in I, \text{ Perf}(j) = \max_i 10 \log_{10} \left(\frac{g_{ij}^2}{\sum_{k \in I, k \neq j} g_{ik}^2} \right). \quad (7.2)$$

Le critère $\text{Perf}(j)$ résultant est ensuite moyenné sur l'indice j pour obtenir un critère unique Perf pour une matrice de performances \mathbf{G} donnée. Le critère de performances obtenu par notre algorithme DFICA symétrique pour cette simulation est de 43.8 dB contre 0.1 dB pour l'algorithme FastICA symétrique standard.

Nous avons aussi testé les versions à déflation des algorithmes standard et différentiel. Les SIR_{out} obtenus pour les deux sources ont été respectivement de 33.2 dB et 32.5 dB pour notre algorithme DFICA (32.3 et 31.6 dB de $SIRI$) et de 3.0 dB et -21.5 dB pour l'algorithme FastICA standard (2.07 et -22.4 dB de $SIRI$)²³.

²³Les définitions des paramètres SIR_{out} et $SIRI$ utilisés dans le cas d'une séparation partielle sont données en annexe F.

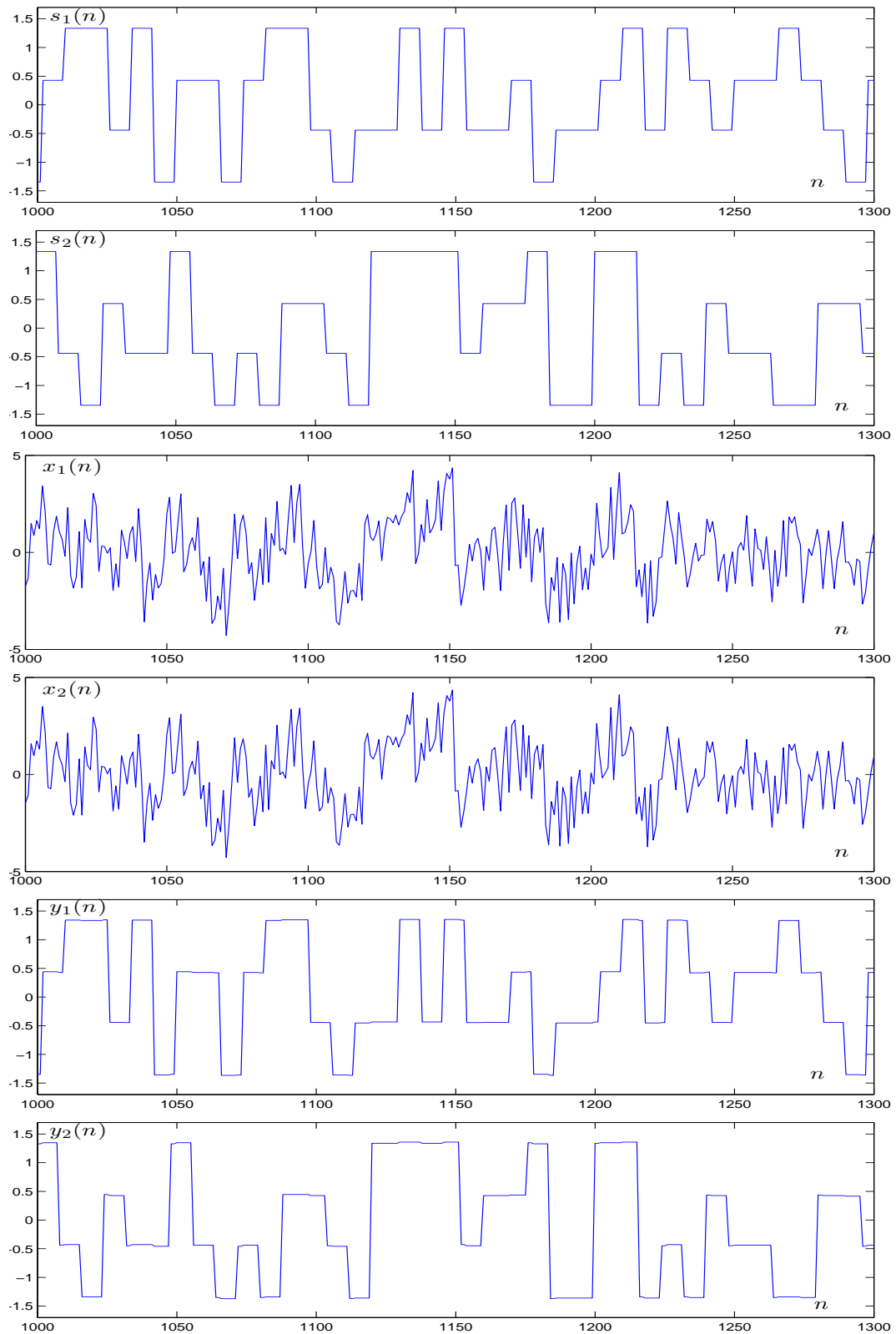


FIG. 7.1 – De haut en bas : sources 1 et 2, observations 1 et 2, sources estimées sans leur composante de bruit.

7.3 Séparation de mélanges linéaires instantanés de signaux audio

Nous avons effectué la même simulation que dans la section précédente en remplaçant les deux sources télécoms par deux sources audio correspondant à l'enregistrement de deux instruments (piano et basse) pendant 15 secondes à une fréquence d'échantillonnage de 44 kHz et dont nous avons sélectionné deux domaines de 100000 échantillons pour appliquer notre méthode différentielle (les deux domaines correspondent aux intervalles $[2.10^5, 3.10^5]$ et $[5.10^5, 6.10^5]$). Les deux sources sont de puissances 1.0 et 0.26 sur le premier domaine et de puissances 0.04 et 0.01 sur le second domaine. Nous avons choisi la même source de bruit et la même matrice de mélange que dans la section précédente. Le SNR_{in} correspondant à cette configuration est de -19.6 dB en moyenne sur les deux domaines de 100000 échantillons.

La figure 7.2 représente de haut en bas le signal source de basse, le signal de piano, les deux observations et les deux sources séparées sans leur composante de bruit. Comme dans la section précédente, malgré la forte présence de bruit, les sources estimées sont visuellement identiques aux sources réelles. L'indice de performances obtenu est de 37.3 dB pour notre algorithme différentiel DFICA symétrique contre 4.2 dB pour l'algorithme FastICA symétrique standard. Comme dans la section précédente, nous avons testé les versions à déflation des algorithmes DFICA et FastICA. Les $SIRI$ obtenus pour les deux sources ont été respectivement de 19.3 dB et 25.6 dB pour notre algorithme DFICA (24.2 et 18.9 dB de $SIRI$) et de 4.6 dB et -2.0 dB pour l'algorithme FastICA standard (7.3 et -6.6 dB de $SIRI$).

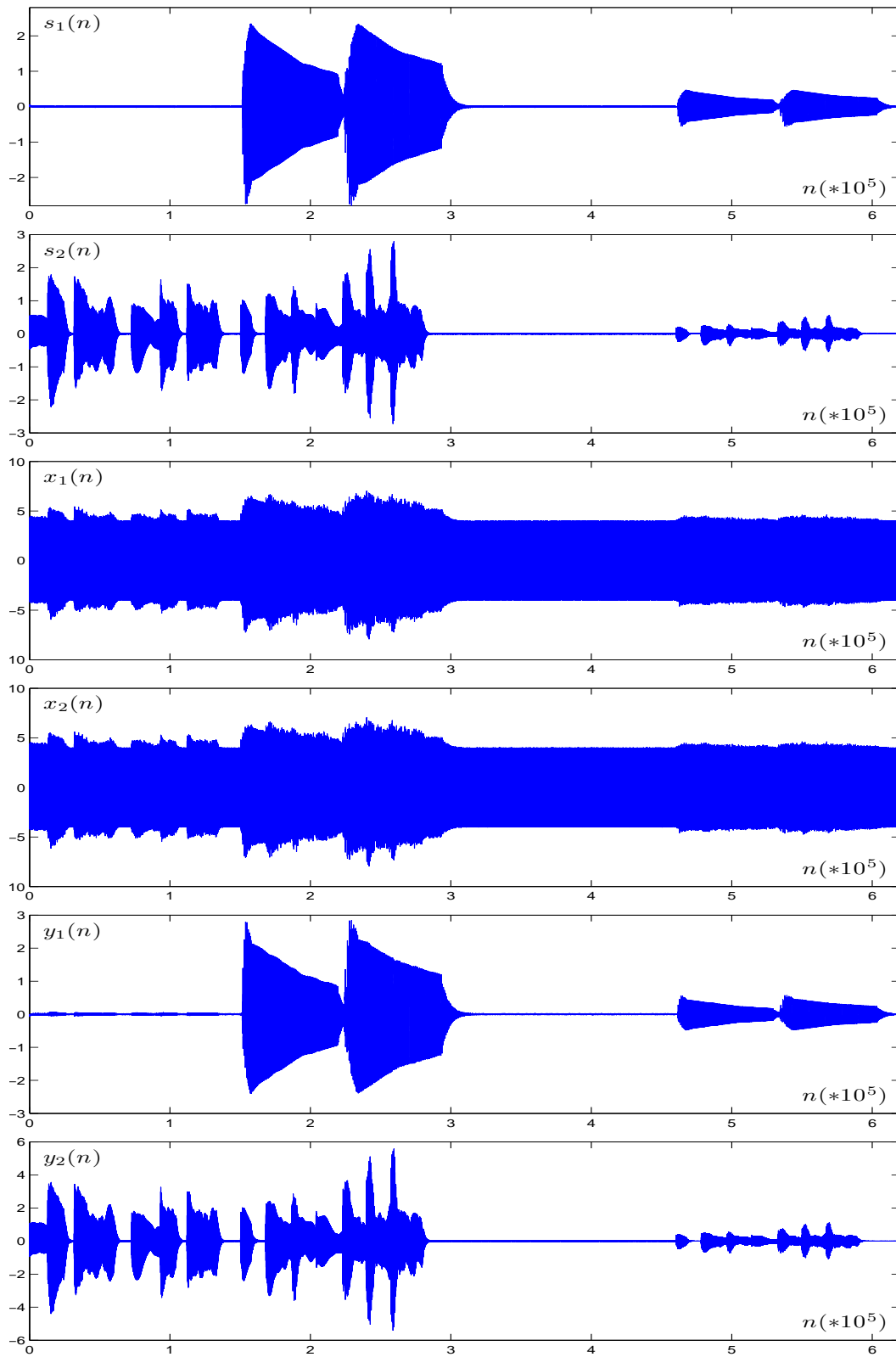


FIG. 7.2 – De haut en bas : sources 1 et 2, observations 1 et 2, sources estimées sans leur composante de bruit.

7.4 Séparation de mélanges convolutifs de signaux télécoms

Dans cette section, nous avons réalisé une simulation de séparation d'un mélange convolutif des deux sources télécoms AM à 4 états utilisées précédemment. Comme dans la section 7.1, les deux sources comportant 100000 échantillons sur chaque domaine sont de puissance 1 sur le premier domaine et de puissance 0.01 sur le deuxième. La source stationnaire est une source uniformément répartie entre -0.5 et +0.5 ce qui correspond à une puissance²⁴ de 0.082. Nous obtenons dans cette configuration un SNR_{in} moyen sur les deux domaines de 1.0 dB. Le jeu de filtres de 5 coefficients a été généré par une fonction uniformément répartie entre 0 et 1.

La figure 7.3 représente les deux sources utilisées sur une plage de 300 échantillons. La figure 7.4 représente de haut en bas les deux observations du mélange, les deux contributions réelles des sources 1 et 2 respectivement sur les capteurs 1 et 2 et les deux estimations de ces contributions. Les SIR_{out} des contributions estimées sont respectivement de 21.4 dB et 19.4 dB (les $SIRI$ sont respectivement de 19.1 dB et 20.1 dB) alors que l'on obtient dans la même configuration avec l'algorithme C-FICA non-différentiel des valeurs de SIR_{out} de 13.0 dB et 8.7 dB (pour des $SIRI$ de 10.7 dB et 9.4 dB).

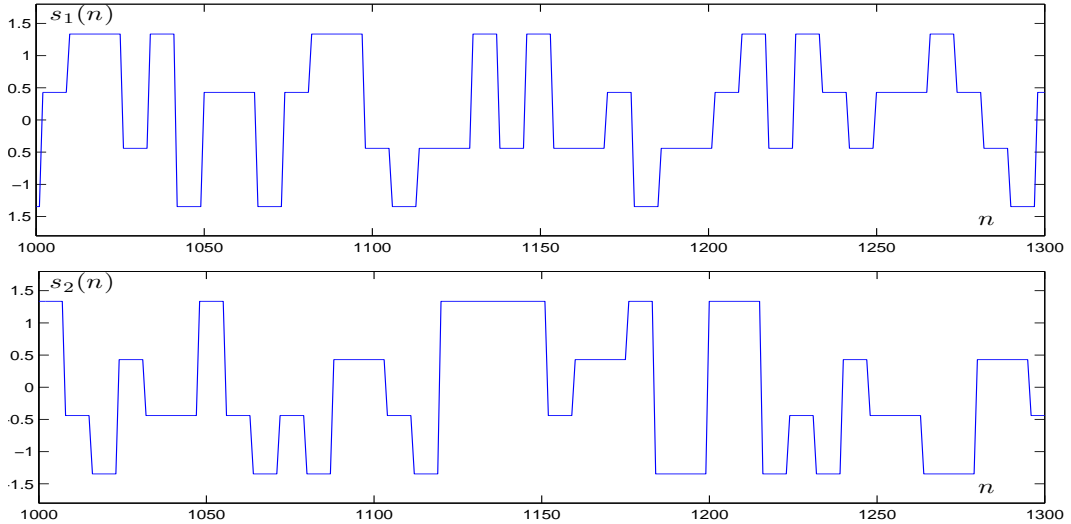


FIG. 7.3 – Sources utilisées 1 et 2 correspondant à des signaux télécoms AM démodulés à 4 états.

²⁴En convolutif, nous avons dû diminuer la puissance de bruit pour obtenir une séparation correcte du mélange.

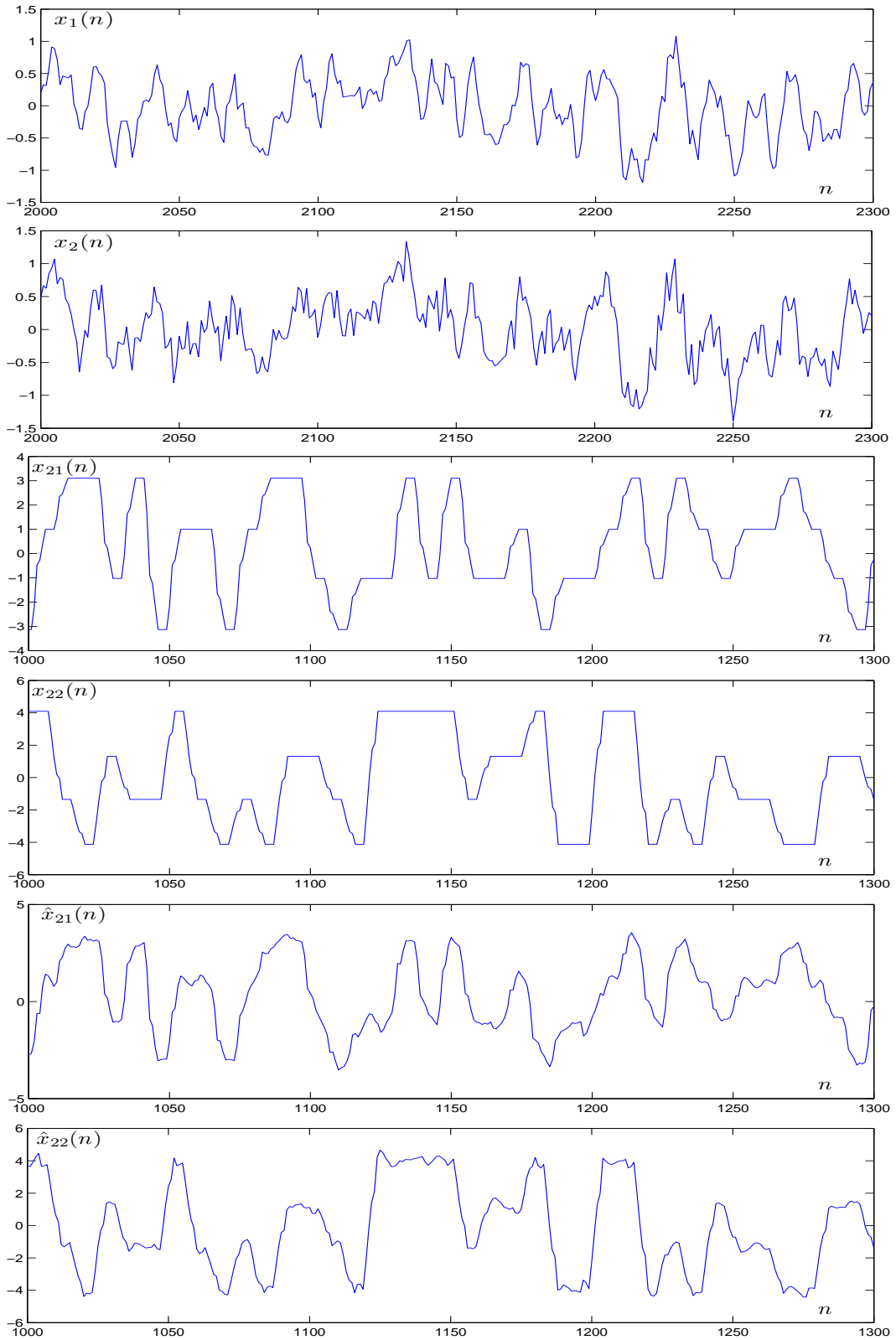


FIG. 7.4 – De haut en bas : observations 1 et 2, contributions réelles des sources 1 et 2 sur le capteur 2, contributions estimées sans leur composante de bruit.

7.5 Tests statistiques

Nous cherchons à comparer les performances de nos méthodes pour mélanges sous-déterminés avec leur versions non différentielles c.-à-d. les algorithmes FastICA et C-FICA. Nous avons d'abord testé notre méthode à déflation DFICA pour mélanges linéaires instantanés sous-déterminés dans le cas $N = 5$, $P = \check{N} = 2$, où les deux sources non-stationnaires correspondent à des signaux sonores de basse et de piano [63] avec 100000 échantillons sur chacun des deux domaines temporels. Les trois sources de bruit sont des signaux stationnaires de même puissance correspondant à des distributions uniforme, gaussienne et laplacienne. Leur facteur d'échelle et leur puissance résultante ont été variés dans nos tests pour mesurer l'évolution des performances en fonction du SNR_{in} des observations. Nous moyennons les deux valeurs de SNR_{in} associées aux deux domaines temporels.

Pour chaque facteur d'échelle appliqué aux sources de bruit normalisées, nous avons réalisé 100 simulations de Monte-Carlo en faisant varier les coefficients de la matrice de mélange avec une distribution uniforme dans $[-0.5, 0.5]$. Les performances de notre méthode de SAS sont mesurées par son Rapport Signal à Interférence de sortie (SIR_{out}), qui peut être comparé au SIR d'entrée (SIR_{in}) associé aux observations.

La figure 7.5 montre que quel que soit le SNR_{in} , le SIR_{out} est supérieur pour notre algorithme différentiel comparé à l'algorithme standard. De plus, les performances commencent à diminuer significativement pour un SNR_{in} inférieur à 10 dB pour notre algorithme différentiel, au lieu de 35 dB pour l'algorithme standard. En outre, pour notre version différentielle, le SIR_{out} est supérieur à 30 dB pour la première source extraite jusqu'à un SNR_{in} égal à 0 dB, au lieu de 20 dB pour l'algorithme standard. Ces résultats ont été obtenus pour un SIR_{in} moyen de 5 dB pour l'ensemble des 100 simulations.

Nous avons aussi comparé la version symétrique de notre algorithme instantané différentiel avec l'algorithme symétrique standard FastICA, avec la même configuration que ci-dessus. La figure 7.6 montre les mêmes avantages que ci-dessus pour notre algorithme différentiel comparé à l'algorithme standard. Cette fois, l'indice de performances $Perf$ défini en (7.2) commence à diminuer significativement pour SNR_{in} plus petit que 15 dB (resp. 35 dB) pour l'algorithme différentiel (resp. l'algorithme standard) et ce critère reste supérieur à 33 dB jusqu'à un SNR_{in} égal à 0 dB (resp. 16 dB). Sur cette figure, les SIR_{out} des contributions des sources, estimées au moyen de la corrélation différentielle (5.28) comme post-traitement, sont aussi représentées. Cela montre pour l'algorithme symétrique que le critère $Perf$ est plus élevé que le critère SIR_{out} qui requiert lui pour être mesuré l'estimation des contributions des sources par un post-traitement (le critère $Perf$ se base seulement sur la matrice de séparation estimée). Notons que ces SIR_{out} des contributions de sources sont légèrement supérieurs à ceux obtenus avec l'algorithme à déflation.

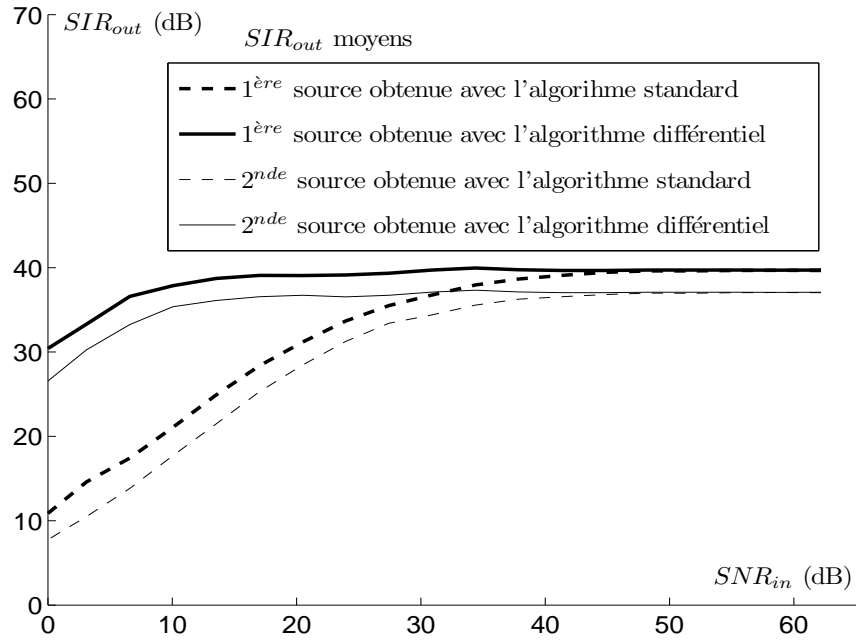


FIG. 7.5 – SIR_{out} obtenus avec les versions à déflation de FastICA et DFICA en fonction du SNR d'entrée.

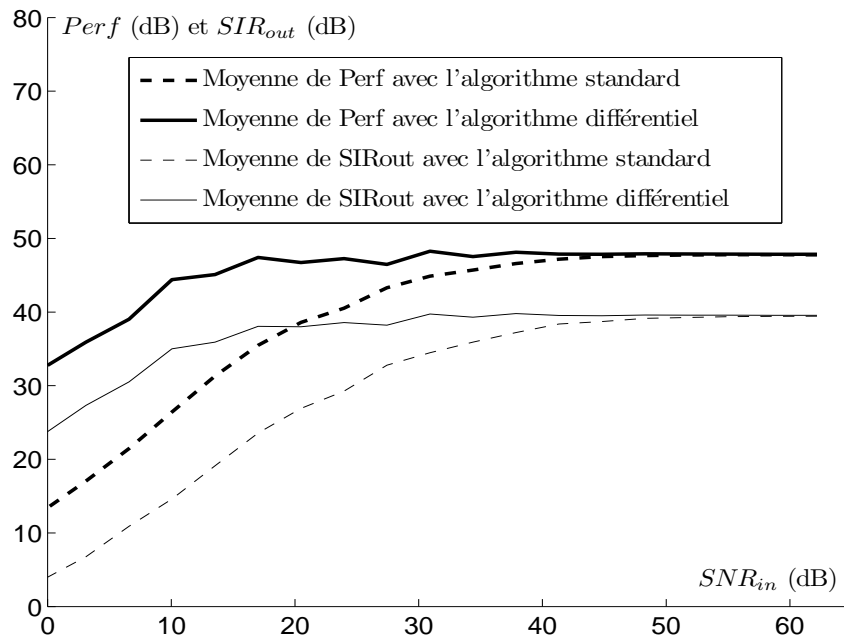


FIG. 7.6 – $Perf$ et SIR_{out} obtenus avec les versions symétriques de FastICA et DFICA en fonction du SNR d'entrée.

Dans le cas convolutif, nous avons testé nos algorithmes C-FICA et C-DFICA pour deux sources laplaciennes colorées par des filtres RIF d'ordre 10 et non-stationnaires à long terme. Elles sont stationnaires à court terme sur deux fenêtres temporelles de 100000 échantillons. Nous avons utilisé des filtres de mélange réels dont les réponses impulsionnelles d'ordre 64 ont été mesurées au niveau des oreilles d'une tête de mannequin [80]. Comme dans le cas linéaire instantané, les sources de bruit sont des sources pondérées à distributions uniforme, gaussienne et laplacienne dont les facteurs d'échelle varient pour modifier le SNR_{in} .

Pour chaque facteur d'échelle appliqué aux sources de bruit, nous avons considéré 153 jeux de filtres différents associés à différentes positions de sources. Nous avons seulement contraint les 2 sources utiles à n'être pas très proches les unes des autres (la différence angulaire a été prise supérieure ou égale à 20°) et nous avons placé les sources de bruit à des positions associées aux angles -90° , 30° et 150° . Comme avec les versions à déflation des algorithmes instantanés, nous exprimons les performances en termes de SIR_{out} des contributions des sources (mesurés comme expliqué en annexe F), cette fois obtenues en utilisant le filtrage de Wiener différentiel (6.30).

La figure 7.7 représente le SIR_{out} des algorithmes d'ACI à point fixe standard et différentiel en fonction du SNR_{in} . Comme dans le cas linéaire instantané, les performances de notre algorithme différentiel restent supérieures à celles de l'algorithme standard quel que soit le SNR_{in} . Plus précisément, elles commencent à diminuer significativement pour un SNR_{in} inférieur à 15 dB (resp. 35 dB) pour notre version différentielle (resp. la version standard). Pour la première source extraite, le SIR_{out} reste supérieur à 15 dB jusqu'à un SNR_{in} égal à 8 dB (resp. 22 dB).

Dans l'expérience suivante, nous avons cherché à évaluer la robustesse de notre algorithme différentiel pour les mélanges convolutifs en fonction du nombre d'échantillons de chaque domaine temporel. Pour cela, nous avons réglé le facteur d'échelle appliqué aux bruits stationnaires de façon à obtenir un SNR_{in} moyen de 11 dB pour toutes les positions de sources et pour chaque nombre d'échantillons n , et nous avons testé notre algorithme différentiel sur les 153 jeux de filtres utilisés dans l'expérience précédente. La figure 7.8 montre que le SIR_{out} moyen reste supérieur à 15 dB jusqu'à un nombre d'échantillons égal à 50000 échantillons.

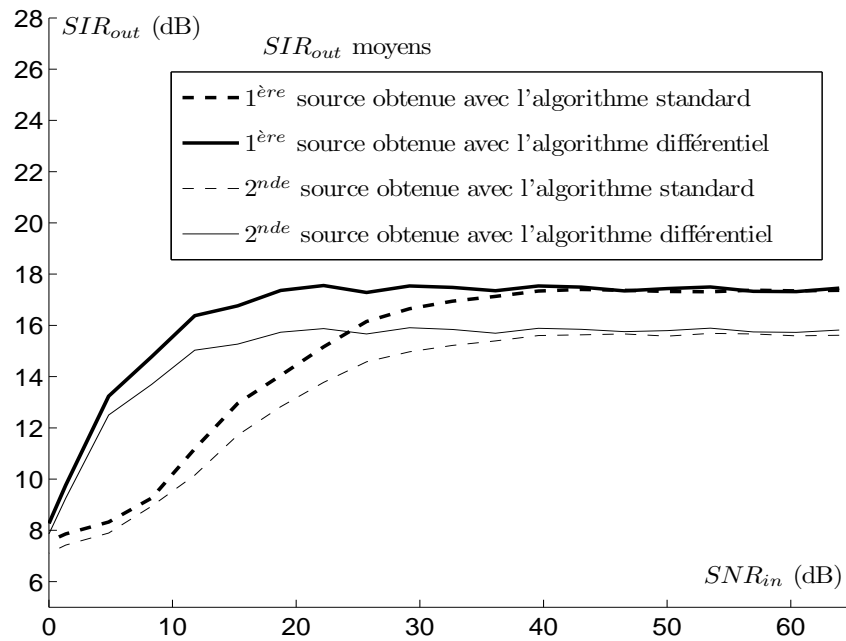


FIG. 7.7 – SIR_{out} obtenus avec C-FICA et C-DFICA en fonction du SNR d'entrée.

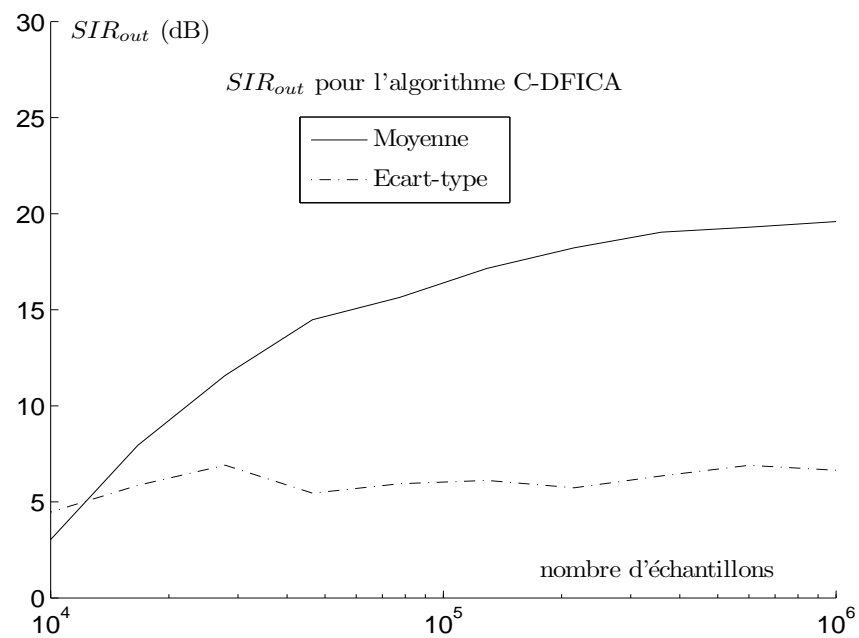


FIG. 7.8 – SIR_{out} obtenus avec C-DFICA en fonction du nombre d'échantillons.

Pour $\tilde{N} = P = 2$ et $N = 3$, nous avons comparé le temps de calcul de notre méthode convolutive différentielle rapide C-DFICA avec une version différentielle de l'algorithme de Tugnait développée dans [6], qui permet déjà une grande rapidité en utilisant un algorithme de Newton modifié. Nous avons choisi des processus d'innovation de sources de densité uniforme contenant $T = 10000$ échantillons sur chaque domaine temporel et colorés par des filtres générés aléatoirement par une loi uniforme. Les filtres de mélange ont aussi été générés par une loi uniforme. Nous avons fait varier l'ordre du modèle²⁵ Q entre 0 et 30 ($T > 10000$ ou $Q > 30$ entraînent des temps de calcul trop grands pour 100 tests de la méthode de Tugnait modifiée) et nous avons effectué 100 tests de Monte Carlo pour chaque configuration. Les résultats de la figure 7.9 représentent la durée de l'extraction du processus d'innovation, qui est de loin l'étape la plus gourmande en temps de calcul. Cela montre que notre méthode à point fixe est près de 500 fois plus rapide que la méthode de Tugnait différentielle. Nous avons aussi constaté des performances de séparation légèrement meilleures pour notre méthode (les SIR_{out} sont supérieurs de 1 dB environ).

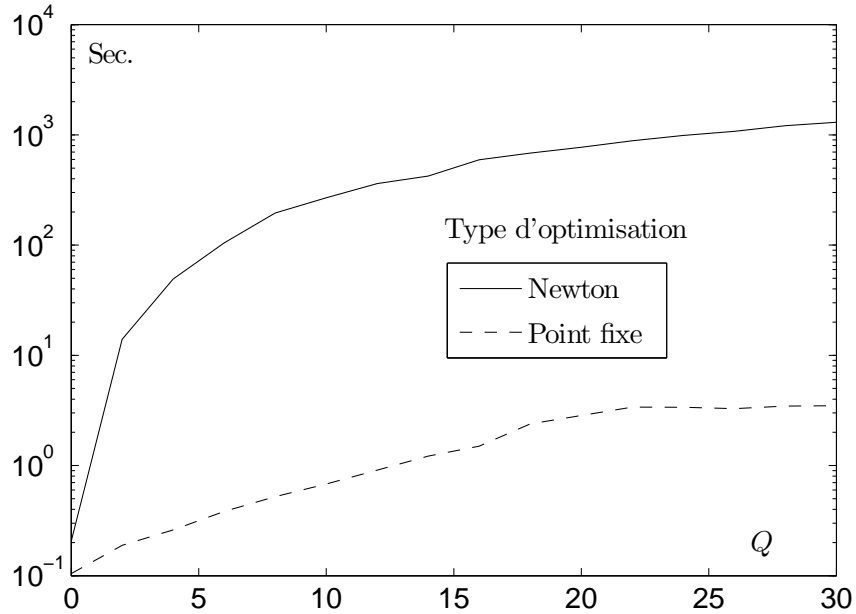


FIG. 7.9 – Temps d'extraction moyen du premier processus d'innovation en fonction de Q pour une optimisation de type Newton et pour notre optimisation de type point fixe (algorithme C-DFICA).

²⁵L'ordre du modèle est défini par la somme de l'ordre de coloration des sources et de l'ordre du mélange.

7.6 Conclusion

Dans ce chapitre, nous avons testé expérimentalement nos algorithmes différentiels DFICA et C-DFICA applicables aux mélanges sous-déterminés respectivement instantanés et convolutifs. Après avoir illustré le fonctionnement de nos méthodes sur des mélanges de sources de télécommunications et audio, nous avons présenté les résultats d'une analyse statistique donnant les performances de séparation en fonction de la puissance totale des sources de bruit stationnaires. Les méthodes différentielles se sont montrées plus résistantes que les approches standard en présence de sources de bruit stationnaires. Nous avons aussi comparé notre algorithme C-DFICA avec une version différentielle de l'algorithme de Tugnait développée auparavant dans l'équipe. Nous avons constaté que notre algorithme est près de 500 fois plus rapide tout en permettant une séparation légèrement meilleure en termes de *SIR*.

Troisième partie

Mélanges linéaires instantanés contenant un grand nombre d'échantillons

Chapitre 8

Algorithmes rapides optimisés pour les mélanges linéaires instantanés

8.1 Introduction

Ce chapitre présente une nouvelle approche pour optimiser les critères de séparation basés sur le kurtosis dans le cas de longs enregistrements de mélanges. Notre méthode est basée sur une étape d'identification polynômiale multivariable qui évite le calcul de statistiques à chaque mise à jour de l'algorithme d'optimisation à point fixe. Comparées à l'algorithme FastICA et à notre algorithme DFICA présenté dans la partie précédente, nos nouvelles méthodes sont très performantes pour les longs enregistrements d'un nombre modéré de sources mélangées. Elles sont donc particulièrement adaptées à la séparation aveugle d'images, en raison de l'augmentation actuelle du nombre de pixels des capteurs photo-sensibles. Nos algorithmes économisent aussi de l'espace mémoire en évitant le calcul et le stockage du vecteur d'observations blanchies.

8.2 Position du problème

Notons $\mathbf{s}(n) = [s_1(n), \dots, s_N(n)]^t$ le vecteur des sources réelles centrées et $\mathbf{x}(n) = [x_1(n), \dots, x_P(n)]^t$ le vecteur des observations. Nous étudions ici le cas des mélanges linéaires instantanés. En notant \mathbf{A} la matrice de mélange scalaire de taille $P \times N$ contenant des coefficients scalaires réels, la relation entre les vecteurs source et observation s'écrit sous forme matricielle

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) \tag{8.1}$$

Dans le cas (sur-)déterminé, nous faisons les hypothèses classiques suivantes concernant le modèle de mélange :

- $\mathbf{s}(n)$ est spatialement indépendant, c.-à-d. que ses composantes sont statistiquement indépendantes entre elles. Nous supposons aussi qu’au plus une de ses composantes est gaussienne.
- La matrice de mélange scalaire \mathbf{A} est de rang de colonne plein. Si $P > N$, nous pouvons utiliser l’Analyse en Composantes Principales (ACP) pour réduire le nombre d’observations à N .

Les hypothèses ci-dessus sont par exemple requises par l’algorithme FastICA à kurtosis développé par Hyvärinen et Oja [98] que nous nous proposons d’optimiser dans cette partie.

Dans le cas sous-déterminé ($P < N$), nous optimisons notre algorithme DFICA présenté dans la partie 2 et destiné à la séparation partielle de sources non-stationnaires en présence de sources de bruit stationnaires. Les hypothèses requises par cet algorithme sous-déterminé de type point fixe sont les suivantes :

- $\mathbf{s}(n)$ est composé de \tilde{N} sources non-stationnaires à long terme (avec $\tilde{N} \leq P$) dont les puissances varient dans le même sens entre deux instants particuliers n_1 et n_2 , et de $N - \tilde{N}$ sources stationnaires à long terme. Les N sources sont statistiquement indépendantes, et au plus une des \tilde{N} sources non-stationnaires est gaussienne aux deux instants n_1 et n_2 ²⁶.
- La matrice $\tilde{\mathbf{A}}$, composée des colonnes de \mathbf{A} associées aux sources non-stationnaires, est de rang de colonne plein. Si le nombre \tilde{N} de sources non-stationnaires est inférieur à P , nous pouvons effectuer une Analyse en Composantes Principales différentielle pour réduire le nombre d’observations à \tilde{N} .

Dans la suite, nous supposons que le nombre \tilde{N} de sources non-stationnaires est égal au nombre d’observations P , et nous notons T le nombre d’échantillons dans chacun des deux intervalles temporels.

Nous considérons ici le cas où les observations contiennent un grand nombre d’échantillons (typiquement plus de 100000 échantillons, ce qui est souvent le cas pour les sources d’images) ce qui implique que les algorithmes FastICA et DFICA demandent beaucoup de temps de calcul pour l’estimation des statistiques à chaque mise à jour des algorithmes d’optimisation. En raison de l’actuelle augmentation de taille des capteurs photo-sensibles de type CCD et CMOS, qui engendre un grand nombre d’échantillons, notre approche est par exemple une bonne alternative aux algorithmes classiques FastICA et DFICA dans le cadre du traitement d’images. Nous supposons que les sources à estimer ne sont pas trop nombreuses (typiquement jusqu’à 7 sources), ce qui est souvent le cas à l’exception des applications biomédicales où le nombre de

²⁶Nous excluons le cas très spécifique où le kurtosis normalisé et la puissance au carré des sources non-stationnaires varient de manière inversement proportionnelle.

mélanges enregistrés est souvent élevé.

8.3 Un algorithme optimisé rapide pour la SAS (sur-) déterminée

8.3.1 Principes et limitations de l'approche standard

Parmi les méthodes de SAS, nous considérons ici celles qui utilisent l'Analyse en Composantes Indépendantes (ACI) et en particulier celles qui cherchent à maximiser la non-gaussianité de signaux de sortie pour estimer les sources. Dans [61], il fut proposé d'effectuer cette séparation en deux étapes.

La première étape réalise un blanchiment en appliquant au vecteur d'observations une matrice dite de blanchiment obtenue grâce à la décomposition en vecteurs propres de la matrice d'autocorrélation des observations. Dans la seconde étape, un algorithme à gradient maximise la valeur absolue ou le carré du kurtosis non-normalisé d'une combinaison linéaire $y(n) = \mathbf{w}^t \mathbf{z}(n)$ des N signaux blanchis sous la contrainte $\|\mathbf{w}\| = 1$. Cela nous donne une estimation d'une des sources, les autres sources pouvant être estimées itérativement par une procédure à déflation [61].

Dans l'algorithme FastICA, le même processus de blanchiment des observations est appliqué mais il est cette fois suivi par un algorithme de type point fixe plus rapide que l'algorithme à gradient. L'itération de l'algorithme à point fixe met à jour le vecteur d'extraction avec le gradient du kurtosis et normalise ensuite ce vecteur.

Il existe deux versions de FastICA. La version parallèle, dont la convergence a été rigoureusement prouvée par Oja et Yuan dans [143], extrait toutes les sources en même temps en répétant alternativement les deux opérations suivantes : la mise à jour d'un jeu de vecteurs d'extraction par les gradients du kurtosis en ces points et l'orthonormalisation de ces vecteurs par l'opération symétrique $\mathbf{W} = \mathbf{W} (\mathbf{W}^t \mathbf{W})^{-\frac{1}{2}}$, où la matrice \mathbf{W} est composée des N vecteurs d'extraction arrangés par colonnes. La mise à jour de type point fixe des N vecteurs d'extraction est réalisée dans le logiciel FastICA [87] par l'opération

$$\mathbf{W} = (\mathbf{Z} * ((\mathbf{Z}' * \mathbf{W}) . ^3)) / \mathbf{T} - 3 * \mathbf{W} \quad (8.2)$$

où \mathbf{Z} est une matrice de taille $N \times T$ contenant les T échantillons des observations blanchies et \mathbf{W} implémente la matrice \mathbf{W} de taille $N \times N$. Comme la mise à jour (8.2) nécessite de stocker \mathbf{Z} et $(\mathbf{Z}' * \mathbf{W}) . ^3$, nous avons besoin de $2N \times T$ cases mémoires. De plus, on peut montrer que l'opération (8.2) demande de réaliser $(4N^2 + N)T + 2N^2$ opérations élémentaires. Dans [96] est évoquée la possibilité d'effectuer les moyennages dans l'opération de mise à jour en utilisant un nombre réduit d'échantillons, diminuant ainsi la charge de calcul. Néanmoins, ce nombre réduit d'échantillons a un effet sur la précision de l'estimation finale de \mathbf{W} . Comme nous cherchons à atteindre une bonne précision, nous utilisons ici toutes les données disponibles.

La version à déflation de FastICA extrait les sources les unes après les autres en appliquant une procédure d'orthogonalisation à déflation à la place de l'orthogonalisation symétrique mentionnée ci-dessus. Dans ce cas, à chaque étape de l'optimisation, un unique vecteur \mathbf{w} est mis à jour avec le gradient du kurtosis de $\mathbf{w}^t \mathbf{z}(n)$ (la matrice \mathbf{W} est remplacée par un unique vecteur \mathbf{w} de taille N) et le vecteur résultant est projeté sur l'espace orthogonal aux vecteurs d'extraction précédemment estimés. Ceci est réalisé par l'opération $\mathbf{w} = \mathbf{w} - \mathbf{W}\mathbf{W}^t \mathbf{w}$ où \mathbf{W} contient les vecteurs précédemment estimés arrangés par colonnes. Le vecteur résultant est ensuite normalisé par l'opération $\mathbf{w} = \mathbf{w} / \|\mathbf{w}\|$.

8.3.2 Une nouvelle approche pour optimiser le kurtosis

Dans cette section, nous proposons une méthode pour éviter l'important besoin en mémoire et le calcul des statistiques empiriques à chaque opération de mise à jour de l'optimisation de \mathbf{W} ou \mathbf{w} . Cette méthode est basée sur une identification polynômiale multivariable qui nous permet d'effectuer les calculs des algorithmes à point fixe ci-dessus dans un espace de coefficients polynômiaux. Elle évite aussi le calcul et le stockage du vecteur d'observations blanchies $\mathbf{z}(n)$. Cette nouvelle approche est attrayante dans beaucoup de configurations (avec $N \leq 7$) en termes de temps de calcul et d'espace mémoire requis, particulièrement dans le cas de longs enregistrements de mélange, alors qu'elle converge exactement vers les mêmes points que les algorithmes FastICA standard.

Notons $y(n) = \mathbf{w}^t \mathbf{z}(n)$ une combinaison linéaire des éléments du vecteur d'observations blanchies $\mathbf{z}(n)$ utilisé dans l'algorithme FastICA (le calcul de $\mathbf{z}(n)$ sera évité dans notre approche comme expliqué plus bas). Comme tous les signaux sont supposés être centrés, le kurtosis de $y(n)$ s'écrit :

$$\begin{aligned} kurt_y(n) &= kurt_{\mathbf{w}^t \mathbf{z}}(n) \\ &= E \{ (\mathbf{w}^t \mathbf{z}(n))^4 \} - 3E \{ (\mathbf{w}^t \mathbf{z}(n))^2 \}^2. \end{aligned} \quad (8.3)$$

Comme nous avons

$$\begin{aligned} E \{ (\mathbf{w}^t \mathbf{z}(n))^2 \} &= E \{ \mathbf{w}^t \mathbf{z}(n) \mathbf{z}^t(n) \mathbf{w} \} \\ &= \mathbf{w}^t E \{ \mathbf{z}(n) \mathbf{z}^t(n) \} \mathbf{w} \\ &= \mathbf{w}^t \mathbf{I} \mathbf{w} \\ &= \|\mathbf{w}\|^2, \end{aligned} \quad (8.4)$$

(8.3) devient

$$\begin{aligned}
 kurt_y(n) &= E \left\{ \sum_{i_1, i_2, i_3, i_4=1}^N \prod_{k=1}^4 w(i_k) z_{i_k}(n) \right\} - 3 \left(\sum_{i=1}^N w(i)^2 \right)^2 \\
 &= \sum_{i_1, i_2, i_3, i_4=1}^N \left(\prod_{k=1}^4 w(i_k) \right) E \left\{ \prod_{k=1}^4 z_{i_k}(n) \right\} - 3 \sum_{i_1, i_2=1}^N w(i_1)^2 w(i_2)^2.
 \end{aligned} \tag{8.5}$$

Comme (8.5) est un polynôme d'ordre 4 par rapport aux variables $w(1), \dots, w(N)$, nous pouvons définir un jeu de coefficients $(\alpha_{\mathbf{d}})_{\mathbf{d} \in D}$ avec

$$D = \left\{ \mathbf{d} \in \{0, \dots, 4\}^N \setminus \sum_{k=1}^N d(k) = 4 \right\}, \tag{8.6}$$

tel que

$$kurt_y(n) = \sum_{\mathbf{d} \in D} \alpha_{\mathbf{d}} \prod_{k=1}^N w(k)^{d(k)}. \tag{8.7}$$

L'éq. (8.7) correspond au développement canonique du kurtosis par rapport aux variables $w(1), \dots, w(N)$. Notons R le cardinal²⁷ de D , $\mathbf{d}_1, \dots, \mathbf{d}_R$ ses éléments et $\alpha_1, \dots, \alpha_R$ les valeurs de $(\alpha_{\mathbf{d}})_{\mathbf{d} \in D}$. Nous avons alors

$$kurt_y(n) = \sum_{r=1}^R \alpha_r \prod_{k=1}^N w(k)^{d_r(k)}. \tag{8.8}$$

Il serait possible de calculer directement les coefficients α_r qui dépendent de moments croisés d'ordre 4 des signaux $z_1(n), \dots, z_N(n)$. Nous proposons ici une méthode plus efficace en termes de temps de calcul. Soit $(\mathbf{v}_i)_{i=1..R}$ une famille de R vecteurs différents de taille N . En remplaçant \mathbf{w} par \mathbf{v}_i dans (8.8), nous avons

$$\forall i = 1 \dots R, \quad kurt_{\mathbf{v}_i \mathbf{z}}(n) = \sum_{j=1}^R \alpha_j \prod_{k=1}^N v_i(k)^{d_j(k)} \tag{8.9}$$

ce qui peut être exprimé par la relation matricielle

$$\mathbf{M} \boldsymbol{\alpha} = \mathbf{k} \tag{8.10}$$

où $\boldsymbol{\alpha}$ est le vecteur colonne des coefficients $(\alpha_r)_{r=1..R}$ et où la matrice $\mathbf{M} = [m_{ij}]_{i,j=1..R}$ et le vecteur colonne $\mathbf{k} = (k_i)_{i=1..R}$ sont définis respectivement par

$$\begin{cases} \forall i, j = 1..R, \quad m_{ij} = \prod_{k=1}^N v_i(k)^{d_j(k)} \\ \forall i = 1..R, \quad k_i = kurt_{\mathbf{v}_i \mathbf{z}}(n) \end{cases} \tag{8.11}$$

²⁷On peut montrer que $\text{card}(D) = R = N + \frac{3N(N-1)}{2} + \frac{N(N-1)(N-2)}{2} + \frac{N(N-1)(N-2)(N-3)}{24}$.

En choisissant R vecteurs d'extraction \mathbf{v}_i de taille N qui fournissent une matrice \mathbf{M} non-singulière, nous pouvons donc identifier les R coefficients $(\alpha_r)_{r=1..R}$ en calculant les kurtosis k_i des R signaux $\mathbf{v}_i^t \mathbf{z}(n)$ obtenus avec ces vecteurs d'extraction et en utilisant la relation inverse de (8.10), c.-à-d. $\boldsymbol{\alpha} = \mathbf{M}^{-1} \mathbf{k}$. Dans l'annexe G, nous proposons un choix particulier de vecteurs (\mathbf{v}_i) qui entraîne un bon conditionnement de la matrice \mathbf{M} . Notons que grâce au processus de blanchiment,

$$\forall i = 1, \dots, R, \quad E\{(\mathbf{v}_i^t \mathbf{z}(n))^2\} = \|\mathbf{v}_i\|^2, \quad (8.12)$$

comme le montre (8.4) et donc

$$kurt_{\mathbf{v}_i^t \mathbf{z}}(n) = E\{(\mathbf{v}_i^t \mathbf{z}(n))^4\} - 3\|\mathbf{v}_i\|^4, \quad (8.13)$$

si bien que nous n'avons à calculer qu'un moment simple d'ordre 4 pour chacun des R vecteurs d'extraction. De plus la relation

$$\begin{aligned} \forall i = 1 \dots R, \quad \mathbf{v}_i^t \mathbf{z}(n) &= \mathbf{v}_i^t \boldsymbol{\Delta}^{-1/2} \mathbf{E}^t \mathbf{x}(n) \\ &= \mathbf{u}_i^t \mathbf{x}(n) \end{aligned} \quad (8.14)$$

avec $\mathbf{u}_i^t = \mathbf{v}_i^t \boldsymbol{\Delta}^{-1/2} \mathbf{E}^t$ nous permet d'éviter le calcul du vecteur d'observations blanchies $\mathbf{z}(n)$. Dans la suite, \mathbf{V} et \mathbf{U} sont les matrices contenant respectivement les jeux de vecteurs $(\mathbf{u}_i)_{i=1..R}$ et $(\mathbf{v}_i)_{i=1..R}$ arrangés par colonnes.

Notons que la matrice \mathbf{M} est indépendante des sources et du processus de mélange. Il est donc possible de calculer son inverse une fois pour toutes pour le jeu de vecteurs $(\mathbf{v}_i)_{i=1..R}$ choisi pour un nombre de sources donné. Après avoir identifié le jeu de coefficients $(\alpha_r)_{r=1..R}$ en calculant les kurtosis k_i des R signaux associés aux vecteurs $(\mathbf{v}_i)_{i=1..R}$ et en calculant $\boldsymbol{\alpha} = \mathbf{M}^{-1} \mathbf{k}$, nous cherchons à maximiser la valeur absolue de $kurt_{\mathbf{w}^t \mathbf{z}}(n) = \sum_{r=1}^R \alpha_r \prod_{k=1}^N w(k)^{d_r(k)}$ par rapport à \mathbf{w} sous la contrainte $\|\mathbf{w}\| = 1$. Cela peut être réalisé de la même façon qu'avec l'algorithme FastICA standard mais ici dans l'espace des coefficients polynômiaux α_r de (8.8), au lieu de l'espace des signaux associé à (8.3). Comme dans la version à déflation de FastICA, nous répétons alternativement l'opération de mise à jour par le gradient du kurtosis et l'opération de normalisation (pour la version parallèle, N vecteurs sont mis à jour puis orthonormalisés). Nous utilisons donc le gradient de $kurt_{\mathbf{w}^t \mathbf{z}}(n)$ par rapport à \mathbf{w} , qui s'écrit

$$\frac{\partial kurt_{\mathbf{w}^t \mathbf{z}}(n)}{\partial \mathbf{w}} = \sum_{r=1}^R \alpha_r \frac{\partial \prod_{k=1}^N w(k)^{d_r(k)}}{\partial \mathbf{w}} \quad (8.15)$$

avec

$$\frac{\partial \prod_{k=1}^N w(k)^{d_r(k)}}{\partial w(j)} = \begin{cases} d_r(j) w(j)^{d_r(j)-1} \prod_{k \neq j} w(k)^{d_r(k)} & , \text{ si } d_r(j) \geq 1 \\ 0 & , \text{ si } d_r(j) = 0 \end{cases} \quad (8.16)$$

8.3.3 Résumé de la méthode proposée

Résumons maintenant la version parallèle de notre algorithme FastICA optimisé, que nous appelons O-FICA :

- Estimer la matrice d'auto-corrélation \mathbf{R}_x du vecteur d'observations $\mathbf{x}(n)$ et calculer sa décomposition en vecteurs propres qui donne une matrice orthogonale \mathbf{E} et une matrice diagonale $\mathbf{\Delta}$.
- Calculer $\mathbf{U} = \mathbf{E}\mathbf{\Delta}^{-1/2}\mathbf{V}$ où \mathbf{V} est constituée des vecteurs d'extraction proposés dans l'annexe G.
- Calculer les kurtosis $(k_i)_{i=1..R}$ des R signaux $\mathbf{u}_i^t \mathbf{x}(n)$, $i = 1..R$ et identifier le jeu de coefficients $\boldsymbol{\alpha} = (\alpha_r)_{r=1..R}$ au moyen de l'inverse de \mathbf{M} définie par (8.11) et calculée préalablement, avec la relation $\boldsymbol{\alpha} = \mathbf{M}^{-1}\mathbf{k}$.
- Initialiser \mathbf{W} avec N vecteurs de taille N différents arrangés par colonnes, c.-à-d. $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N] = [w_{ji}]$ et répéter jusqu'à la convergence :
 - 1) $\forall i, j = 1 \dots N, w_{ji} \leftarrow \sum_{r=1}^R \alpha_r d_r(j) w_{ji}^{max(d_r(j)-1,0)} \prod_{k \neq j} w_{ki}^{d_r(k)}$
 ce qui est équivalent à faire $\forall i = 1 \dots N, \mathbf{w}_i \leftarrow \frac{\partial kurt_{\mathbf{w}_i^t \mathbf{z}}(n)}{\partial \mathbf{w}_i}$.
 - 2) $\mathbf{W} \leftarrow \mathbf{W}(\mathbf{W}^t \mathbf{W})^{-\frac{1}{2}}$

Remarquons que ce nouvel algorithme ne demande pas d'estimer de statistiques à chaque étape de l'algorithme à point fixe, contrairement à l'algorithme FastICA standard. Il n'utilise que des coefficients polynômiaux α_r dont le nombre est indépendant de T , ce qui est plus efficace s'ils ne sont pas trop nombreux (c'est le cas typiquement pour $N \leq 7$). De plus, nous évitons de stocker les matrices \mathbf{Z} et $(\mathbf{Z}' * \mathbf{W})$.³ pendant l'algorithme à point fixe, ce qui représente $2NT$ cases mémoires. On peut montrer que nous ne stockons ici que $R^2 + R + RN + N^2$ mots mémoires pendant l'optimisation, valeur indépendante du nombre d'échantillons T .

Le passage à la version à déflation de notre algorithme O-FICA peut être déduit facilement de la relation entre les versions parallèle et à déflation de l'algorithme FastICA classique.

8.4 Un algorithme optimisé rapide pour la SAS sous-déterminée

8.4.1 Principe et limitations de notre approche DFICA

Comme l'algorithme FastICA standard, notre algorithme DFICA est composé de deux étapes : une étape de blanchiment différentiel suivie d'un al-

gorithme de type point fixe. Le blanchiment différentiel estime la matrice d'autocorrélation différentielle des observations définie par la différence entre les matrices d'autocorrélation estimées respectivement sur les deux intervalles temporels :

$$\mathbf{DR}_x(n_1, n_2) = \mathbf{R}_x(n_2) - \mathbf{R}_x(n_1). \quad (8.17)$$

Ensuite, sous la condition que $\mathbf{DR}_x(n_1, n_2)$ est définie positive²⁸, sa décomposition en vecteurs propres est calculée, ce qui donne une matrice orthogonale \mathbf{E} et une matrice diagonale $\mathbf{\Delta}$. La matrice de blanchiment différentielle est ensuite obtenue par $\mathbf{B} = \mathbf{\Delta}^{-1/2}\mathbf{E}^t$. Enfin, les observations sont blanchies différentiellement. Ce blanchiment différentiel implique une équivalence entre la normalisation d'un vecteur d'extraction \mathbf{w} et celle de la puissance différentielle du signal $y(n) = \mathbf{w}^t \mathbf{z}(n)$:

$$E\{y^2(n_2)\} - E\{y^2(n_1)\} = 1 \Leftrightarrow \|\mathbf{w}\|^2 = 1. \quad (8.18)$$

Nous avons prouvé dans la seconde partie de ce document que la maximisation de la valeur absolue du kurtosis différentiel de $\mathbf{w}^t \mathbf{z}(n)$, défini par

$$Dkurt_{\mathbf{w}^t \mathbf{z}}(n_1, n_2) = kurt_{\mathbf{w}^t \mathbf{z}}(n_2) - kurt_{\mathbf{w}^t \mathbf{z}}(n_1) \quad (8.19)$$

par rapport à \mathbf{w} sous la contrainte $\|\mathbf{w}\| = 1$ réalise la séparation partielle des sources non-stationnaires, c.-à-d. que le signal de sortie $y(n) = \mathbf{w}^t \mathbf{z}(n)$ est alors composé d'une unique source utile et de contributions des sources de bruit.

Comme pour l'algorithme FastICA standard, nous avons proposé deux versions de notre algorithme DFICA. La version parallèle extrait toutes les sources utiles en même temps en mettant à jour \tilde{N} vecteurs d'extraction avec les gradients du kurtosis différentiel en ces points, puis en les orthonormalisant au moyen de l'opération symétrique $\mathbf{W} = \mathbf{W}(\mathbf{W}^t \mathbf{W})^{-1/2}$ déjà mentionnée dans le chapitre précédent. La mise à jour de type point fixe des \tilde{N} vecteurs d'extraction peut être représentée en langage Matlab par les opérations

$$\begin{aligned} \mathbf{R1} &= (\mathbf{Z1} * \mathbf{Z1}') / T ; \\ \mathbf{D} &= \text{diag}(\mathbf{W}' * \mathbf{R1} * \mathbf{W})' ; \\ \mathbf{W} &= (\mathbf{Z2} * (\mathbf{Z2}' * \mathbf{W}) .^3) - \mathbf{Z1} * (\mathbf{Z1}' * \mathbf{W}) .^3) / T \\ &\quad - 3(\mathbf{I} + \mathbf{R1}) * \mathbf{W} + \mathbf{D}(\text{ones}(1, \text{length}(\mathbf{W})), :). * \mathbf{W}) ; \end{aligned} \quad (8.20)$$

où $\mathbf{Z1}$ et $\mathbf{Z2}$ sont deux matrices de taille $\tilde{N} \times T$ contenant les T échantillons des observations différentiellement blanchies sur les deux intervalles temporels, \mathbf{W} est une matrice de taille $\tilde{N} \times \tilde{N}$ implémentant les \tilde{N} vecteurs d'extraction arrangés par colonnes et $\mathbf{R1}$ est la matrice d'autocorrélation de $\mathbf{z}(n)$ sur le premier intervalle. La mise à jour (8.20) demande donc de stocker au moins \mathbf{Zi} et $(\mathbf{Zi}' * \mathbf{W}) .^3$ avec $i = 1$ ou 2 , ce qui représente $2\tilde{N}T$ cases mémoires.

²⁸Les conditions de la nature définie positive de $\mathbf{DR}_x(n_1, n_2)$ seront données dans la prochaine section.

De plus, on peut montrer que l'opération (8.20) nécessite de calculer $(10\tilde{N}^2 + 2\tilde{N})T + 6\tilde{N}^3 + \tilde{N}^2 + \tilde{N}$ opérations élémentaires.

Comme pour les algorithmes FastICA et O-FICA, une version à déflation peut être obtenue en remplaçant la matrice \mathbf{W} par un unique vecteur d'extraction \mathbf{w} et en appliquant une orthogonalisation à déflation au lieu de l'orthogonalisation symétrique.

Comme pour la version optimisée O-FICA de l'algorithme FastICA standard, nous proposons dans la prochaine section une approche pour éviter le calcul de statistiques à chaque étape de l'optimisation à point fixe.

8.4.2 Une nouvelle approche pour optimiser le kurtosis différentiel

Notons à nouveau $y(n) = \mathbf{w}^t \mathbf{z}(n)$ une combinaison linéaire des éléments du vecteur d'observations différentiellement blanchies $\mathbf{z}(n)$ obtenu dans l'algorithme DFICA.

Comme pour l'algorithme O-FICA, le calcul de $\mathbf{z}(n)$ sera évité dans notre nouvel algorithme comme expliqué plus bas. Dans la section précédente, nous avons montré que le kurtosis standard de $y(n)$ est un polynôme multivariable par rapport à $w(1), \dots, w(\tilde{N})$. Le kurtosis différentiel, défini comme une différence de deux kurtosis à deux instants, est donc aussi un polynôme multivariable de ces mêmes variables. On peut facilement l'exprimer en fonction des moments croisés d'ordre 4 des signaux $z_1(n), \dots, z_P(n)$:

$$\begin{aligned}
 Dkurt_y(n_1, n_2) &= kurt_y(n_2) - kurt_y(n_1) \\
 &= E\{(\mathbf{w}^t \mathbf{z}(n_2))^4\} - 3(E\{(\mathbf{w}^t \mathbf{z}(n_2))^2\})^2 - E\{(\mathbf{w}^t \mathbf{z}(n_1))^4\} + 3(E\{(\mathbf{w}^t \mathbf{z}(n_1))^2\})^2 \\
 &= E\left\{\sum_{i_1, i_2, i_3, i_4=1}^{\tilde{N}} \prod_{k=1}^4 w(i_k) z_{i_k}(n_2)\right\} - 3(E\left\{\sum_{i_1, i_2=1}^{\tilde{N}} w(i_1) w(i_2) z_{i_1}(n_2) z_{i_2}(n_2)\right\})^2 \\
 &\quad - E\left\{\sum_{i_1, i_2, i_3, i_4=1}^{\tilde{N}} \prod_{k=1}^4 w(i_k) z_{i_k}(n_1)\right\} + 3(E\left\{\sum_{i_1, i_2=1}^{\tilde{N}} w(i_1) w(i_2) z_{i_1}(n_1) z_{i_2}(n_1)\right\})^2 \\
 &= \sum_{i_1, i_2, i_3, i_4=1}^{\tilde{N}} \left(E\left\{\prod_{k=1}^4 z_{i_k}(n_2)\right\} - E\left\{\prod_{k=1}^4 z_{i_k}(n_1)\right\} \right) \prod_{k=1}^4 w(i_k) \\
 &\quad - 3 \sum_{i_1, i_2, i_3, i_4=1}^{\tilde{N}} \left(E\{z_{i_1}(n_2) z_{i_2}(n_2)\} E\{z_{i_3}(n_2) z_{i_4}(n_2)\} \right. \\
 &\quad \left. - E\{z_{i_1}(n_1) z_{i_2}(n_1)\} E\{z_{i_3}(n_1) z_{i_4}(n_1)\} \right) \prod_{k=1}^4 w(i_k). \tag{8.21}
 \end{aligned}$$

Comme le kurtosis standard, le kurtosis différentiel peut ainsi être exprimé de façon similaire à (8.7) en fonction d'un jeu de R coefficients $(\alpha_d)_{\mathbf{d} \in D}$. Au lieu de calculer les moments croisés identifiés dans (8.21), les coefficients polynômiaux peuvent être identifiés en calculant les kurtosis différentiels Dk_i de R signaux obtenus avec R vecteurs d'extraction $(\mathbf{v}_i)_{i=1..R}$ qui peuvent être choisis

comme indiqué dans l'annexe G²⁹. Comme dans le cas (sur-)déterminé, il n'est pas nécessaire de calculer le vecteur d'observations différentiellement blanchies $\mathbf{z}(n)$. Une fois que l'on a identifié le jeu de coefficients $(\alpha_r)_{r=1..R}$ en calculant les kurtosis différentiels de R signaux et en multipliant le vecteur résultant par l'inverse de \mathbf{M} , on peut utiliser le même type d'algorithme que dans l'algorithme O-FICA (sur-)déterminé pour maximiser la valeur absolue du kurtosis différentiel sous contrainte de puissance différentielle unitaire.

8.4.3 Résumé de la méthode proposée

Résumons maintenant notre algorithme DFICA optimisé, que nous appelons O-DFICA :

- Choisir deux intervalles temporels associés aux deux instants n_1 et n_2 de sorte que l'estimée de la matrice d'autocorrélation différentielle $\mathbf{DR}_\mathbf{x}(n_1, n_2) = \mathbf{R}_\mathbf{x}(n_2) - \mathbf{R}_\mathbf{x}(n_1)$ soit définie positive³⁰ et calculer sa décomposition en vecteurs propres qui donne une matrice orthogonale \mathbf{E} et une matrice diagonale $\mathbf{\Delta}$.
- Calculer $\mathbf{U} = \mathbf{E}\mathbf{\Delta}^{-1/2}\mathbf{V}$ où \mathbf{V} est constituée des vecteurs d'extraction proposés dans l'annexe G.
- Calculer les kurtosis différentiels $(Dk_i)_{i=1..R}$ des R signaux $\mathbf{u}_i^t \mathbf{x}(n)$, $i = 1..R$ et déterminer le jeu de coefficients $\boldsymbol{\alpha} = (\alpha_r)_{r=1..R}$ au moyen de la matrice inverse de \mathbf{M} (8.11) calculée préalablement, avec la relation $\boldsymbol{\alpha} = \mathbf{M}^{-1}\mathbf{D}\mathbf{k}$.
- Initialiser \mathbf{W} avec \tilde{N} vecteurs différents de taille \tilde{N} arrangés par colonnes, c.-à-d. $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{\tilde{N}}] = [w_{ji}]$ et répéter jusqu'à la convergence :

$$1) \forall i, j = 1 \dots \tilde{N}, \quad w_{ji} \leftarrow \sum_{r=1}^R \alpha_r d_r(j) w_{ji}^{\max(d_r(j)-1, 0)} \prod_{k \neq j} w_{ki}^{d_r(k)}$$

ce qui est équivalent à faire $\forall i = 1 \dots \tilde{N}, \quad \mathbf{w}_i \leftarrow \frac{\partial Dkurt_{\mathbf{w}_i^t \mathbf{z}}(n_1, n_2)}{\partial \mathbf{w}_i}$.

$$2) \mathbf{W} \leftarrow \mathbf{W}(\mathbf{W}^t \mathbf{W})^{-\frac{1}{2}}$$

Comme l'algorithme O-FICA, cet algorithme O-DFICA n'utilise que des coefficients polynômiaux dont le nombre est indépendant de T , pendant l'optimisation de type point fixe. Cela est plus efficace quand le nombre de sources est modéré (typiquement $\tilde{N} \leq 7$). Il évite aussi de stocker \mathbf{Z}_i et $(\mathbf{Z}_i^* \mathbf{W})$

²⁹La valeur de R dépend cette fois de \tilde{N} au lieu de N , c.-à-d. $R = \text{Card}(D) = \tilde{N} + \frac{3\tilde{N}(\tilde{N}-1)}{2} + \frac{\tilde{N}(\tilde{N}-1)(\tilde{N}-2)}{2} + \frac{\tilde{N}(\tilde{N}-1)(\tilde{N}-2)(\tilde{N}-3)}{24}$.

³⁰Nous avons prouvé dans la seconde partie que cette condition est équivalente à l'augmentation de puissance de toutes les sources non-stationnaires entre n_1 et n_2 . Si $\mathbf{DR}_\mathbf{x}(n_1, n_2)$ est définie négative, on peut permuter les deux intervalles temporels associés à n_1 et n_2 pour la rendre définie positive.

)³ avec $i = 1$ ou 2 , ce qui représente $2\tilde{N}T$ mots mémoires. Comme avec l'algorithme (sur-)déterminé, nous ne stockons ici que $R^2 + R + R\tilde{N} + \tilde{N}^2$ mots, valeur indépendante du nombre T d'échantillons.

Comme pour l'algorithme O-FICA, la version à déflation de notre algorithme sous-déterminé O-DFICA peut être déduite de la relation entre les algorithmes parallèles et à déflation de l'algorithme FastICA classique.

8.5 Conclusion

Dans ce chapitre, nous avons proposé des méthodes d'optimisation du kurtosis et du kurtosis différentiel basées sur l'identification du critère sous forme d'un polynôme relativement aux variables d'extraction. Nous avons montré que nos algorithmes pour mélanges (sur-) et sous-déterminés, baptisés respectivement O-FICA et O-DFICA, évitent l'estimation de statistiques à chaque itération de l'algorithme d'optimisation à point fixe. Ils évitent aussi de stocker le vecteur d'observations blanches qui est de grande taille lorsque le nombre d'échantillons est élevé. Notre approche est particulièrement intéressante en séparation aveugle d'images en raison de l'augmentation du nombre de pixels des capteurs photo-sensibles de types CCD et CMOS.

Chapitre 9

Résultats expérimentaux

9.1 Introduction

Dans ce chapitre, nous présentons des résultats expérimentaux qui comparent les vitesses de nos algorithmes O-FICA et O-DFICA avec celles des algorithmes standard FastICA et DFICA. Dans la première section, nous montrons l'intérêt de notre approche pour la séparation d'images astrophysiques recueillies par le télescope Hubble. Des tests de Monte-Carlo sont ensuite réalisés dans deux configurations : celle où le critère d'arrêt est fixé et celle où le temps de calcul est fixé. Dans le premier cas, nous mesurons le temps mis par les algorithmes à atteindre un critère d'arrêt permettant d'obtenir une performance de séparation moyenne donnée. Dans le deuxième cas, nous évaluons les performances moyennes qui peuvent être atteintes par les différents algorithmes en un temps fixé.

9.2 Application à l'astrophysique

Nous avons ici appliqué notre algorithme O-FICA à la séparation de deux types de rayonnement émis par un nuage interstellaire en présence du rayonnement d'une étoile voisine. Les nuages interstellaires sont à l'origine de la formation des étoiles, ce qui les rend particulièrement intéressants à étudier d'un point de vue physico-chimique.

Les télescopes qui scrutent ces nuages interstellaires, souvent en employant des filtres pour obtenir des images à différentes longueurs d'onde, permettent d'obtenir une distribution spatiale des rayonnements émis. Dans le cas des nébuleuses par réflexion, la répartition spatiale obtenue est en fait une répartition du mélange formé par deux types de rayonnements : l'émission rouge étendue (*ERE : Extended Red Emission* en anglais) dont le spectre s'étend de 540 à 900 nm avec un pic entre 600 et 800 nm, et la lumière diffusée (*scattered light* en anglais). En modifiant la longueur d'onde d'étude, on obtient des répartitions différentes du fait de l'émission plus ou moins importante de ces deux types de rayonnements aux longueurs d'onde considérées.

Admettons que le nombre de photons reçus dans la cellule d'un capteur CCD pour un type de rayonnement donné soit proportionnel à l'intensité de ce rayonnement dans la zone spatiale associée à ce pixel ainsi qu'à la densité spectrale du rayonnement à la longueur d'onde choisie :

$$\nu_j(x, y, \lambda) = C^{te} S_j(\lambda) r_j(x, y) \quad (9.1)$$

où $\nu_j(x, y, \lambda)$ est le nombre de photons du rayonnement j et de longueur d'onde λ reçus par le pixel de coordonnées x et y , $r_j(x, y)$ est l'intensité du rayonnement j dans la zone spatiale associée au pixel de coordonnées x et y , et $S_j(\lambda)$ est la densité spectrale d'émission du rayonnement j à la longueur d'onde λ .

Dans le cas de l'émission de N rayonnements différents, on obtient pour P longueurs d'onde différentes $\lambda_1, \dots, \lambda_P$ les équations

$$\forall i = 1 \dots P, \quad \nu(x, y, \lambda_i) = \sum_{j=1}^N \nu_j(x, y, \lambda_i) \quad (9.2)$$

où $\nu(x, y, \lambda_i)$ est le nombre total de photons de longueur d'onde λ_i reçus par le pixel (x, y) . En tenant compte de l'éq. (9.1), nous obtenons

$$\forall i = 1 \dots P, \quad \nu(x, y, \lambda_i) = \sum_{j=1}^N C^{te} S_j(\lambda_i) r_j(x, y), \quad (9.3)$$

ce qui s'écrit matriciellement

$$\boldsymbol{\nu}(x, y) = \mathbf{S} \mathbf{r}(x, y) \quad (9.4)$$

où les vecteurs $\mathbf{r}(x, y)$ et $\boldsymbol{\nu}(x, y)$ sont les vecteurs source et observation définis respectivement par $\mathbf{r}(x, y) = (r_j(x, y))_{j=1..N}$ et $\boldsymbol{\nu}(x, y) = (\nu(x, y, \lambda_i))_{i=1..P}$, et \mathbf{S} est la matrice de mélange définie par $\mathbf{S} = [C^{te} S_j(\lambda_i)]_{i=1..P, j=1..N}$.

L'équation (9.4) correspond à l'équation d'un mélange linéaire instantané d'images³¹ de coordonnées spatiales x et y .

Ce modèle simple effectue quelques approximations acceptables pour l'application que nous allons présenter. Il suppose en effet que les spectres des rayonnements ne varient pas spatialement, ce qui peut être irréaliste en cas de présence de forts transferts radiatifs, dans le cas de galaxies par exemple [140]. Néanmoins Witt et al. [223] ont montré que pour les nébuleuses par réflexion, ce phénomène n'intervient pas.

Lorsque le nombre de pixels du capteur photo-sensible est important, une contrainte calculatoire apparaît : l'analyse statistique des échantillons impose, si l'on veut tirer le meilleur parti des données, d'effectuer des moyennages spatiaux sur un grand nombre de valeurs. L'algorithme FastICA, comme nous

³¹Récemment des méthodes de séparation aveugle de sources ont été utilisées pour étudier la répartition d'espèces chimiques de nuages interstellaires ; le mélange y était interprété en termes de mélanges de spectres et non de répartitions spatiales de rayonnement [29, 138].

l'avons expliqué précédemment, effectue des calculs statistiques à chaque itération de l'algorithme de type point fixe, ce qui entraîne un temps de calcul important. Notre algorithme O-FICA effectue lui une analyse statistique dans un premier temps pour identifier la fonction polynômiale à optimiser, puis réalise l'optimisation du polynôme sans avoir à calculer de statistiques sur un grand nombre d'échantillons.

Nous comparons ici les temps de calcul des algorithmes O-FICA et FastICA pour l'analyse de la répartition spatiale de rayonnements à partir de deux images de la nébuleuse NGC 2023 prises par le télescope Hubble aux deux longueurs d'onde $\lambda_1=474$ nm et $\lambda_2=631$ nm. Le capteur CCD employé est de taille 4220×4298 , ce qui entraîne un nombre d'échantillons égal à 18137560. Cette valeur est très élevée et exige un temps de calcul important en utilisant l'algorithme FastICA.

Les figures 9.1 et 9.2 représentent les deux images obtenues par le télescope Hubble respectivement aux longueurs d'onde $\lambda_1=474$ nm et $\lambda_2=631$ nm.

Les figures 9.3 et 9.4 représentent les deux images obtenues en sortie des algorithmes FastICA ou O-FICA. Alors que l'algorithme FastICA a effectué la séparation en 115 sec., c.-à-d. près de deux minutes sur notre machine, l'algorithme O-FICA n'a mis que 4.5 secondes pour réaliser la même tâche.

Les deux répartitions spatiales de rayonnement des figures 9.3 et 9.4 sont intéressantes d'un point de vue astrochimique car elles nous renseignent sur les répartitions des espèces chimiques qui les émettent. Parmi les espèces chimiques présentes dans les nuages interstellaires, on peut citer les molécules polycycliques aromatiques hydrogénées (PAH : *Polycyclic Aromatic Hydrocarbons*) qui sont ionisées à divers degrés (PAH neutres, PAH^+ , PAH^{++} et PAH_2^+) et les très petits grains (VSG : *Very Small Grain* en anglais) formés par agrégation de ces molécules aromatiques. Des travaux menés en collaboration avec l'équipe d'Olivier Berné et de Christine Joblin du CESR [30] ont montré qu'un bon candidat à l'émission du rayonnement ERE est la molécule PAH_2^+ , présente en périphérie du nuage en raison d'un processus de destruction des VSG exposés aux photons UV. Etant le produit final d'une cascade de réactions de photodissociation, cette espèce est en effet relativement stable et abondante car elle se reforme efficacement à partir des molécules de PAH neutres et ionisés.

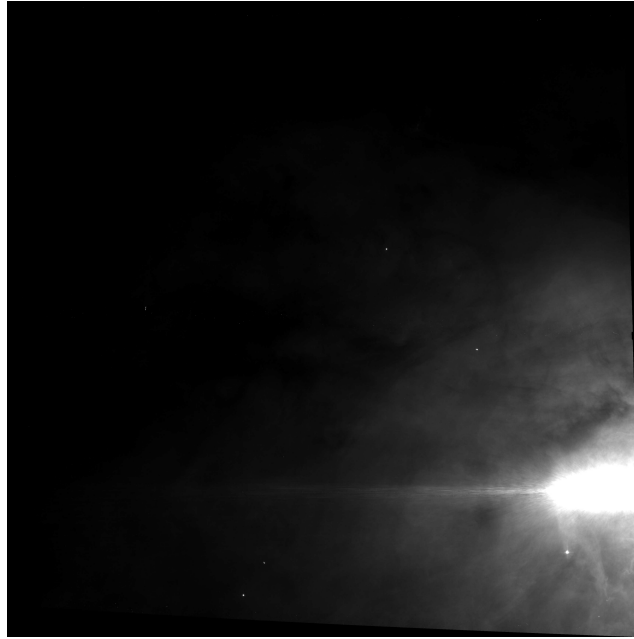


FIG. 9.1 – Image de la nébuleuse NGC 2023 prise par le télescope Hubble à la longueur d'onde $\lambda_1=474$ nm.

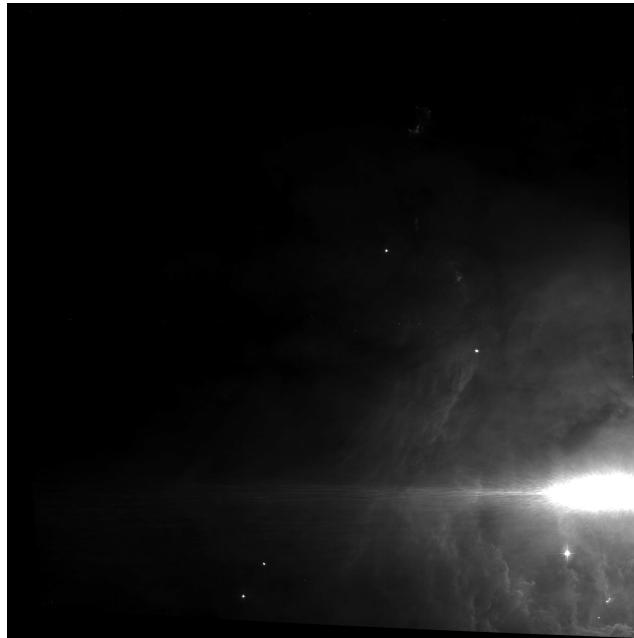


FIG. 9.2 – Image de la nébuleuse NGC 2023 prise par le télescope Hubble à la longueur d'onde $\lambda_2=631$ nm.



FIG. 9.3 – Image séparée 1 obtenue avec l'algorithme FastICA en 115 sec. et avec l'algorithme O-FICA en 4.5 sec.

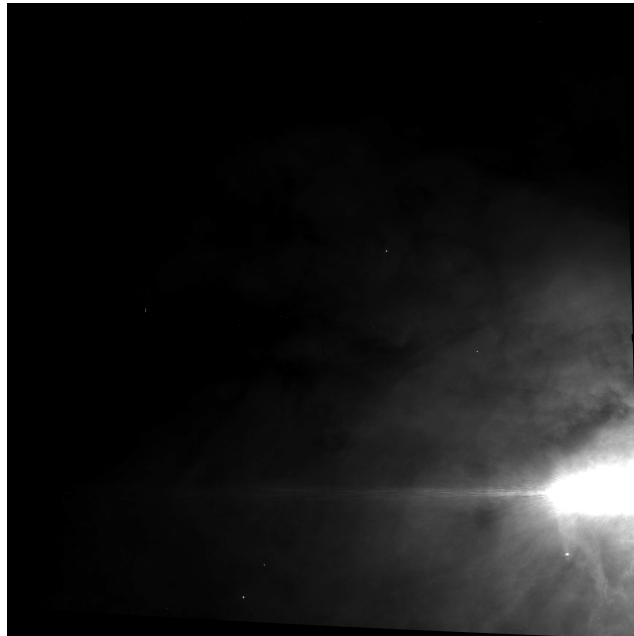


FIG. 9.4 – Image séparée 2 obtenue avec l'algorithme FastICA en 115 sec. et avec l'algorithme O-FICA en 4.5 sec.

9.3 Tests statistiques

Dans la première série de tests présentée ci-dessous, nous avons comparé les temps de calcul des algorithmes optimisés O-FICA et O-DFICA aux algorithmes standard associés FastICA et DFICA. Pour cela, nous avons mesuré le temps qu'ils prennent à atteindre le critère d'arrêt utilisé dans le logiciel Matlab FastICA. Brièvement, ce critère compare les directions des vecteurs d'extraction après les deux dernières mises à jour et l'optimisation est stoppée si ces directions varient moins qu'un certain seuil (nous avons réglé ce paramètre de seuil à 10^{-4} , qui est la valeur par défaut dans le logiciel). Nous avons testé les versions parallèles et à déflation des algorithmes pour séparer des sources uniformes et laplaciennes en nombres à peu près égaux.

Pour les algorithmes différentiels O-DFICA et DFICA, nous avons pris deux intervalles entre lesquels les puissances des sources varient dans le même sens. Sur ces intervalles, nous avons mélangé ces sources avec 3 sources de bruit stationnaires correspondant à des distributions uniforme, gaussienne et laplacienne.

La figure 9.5 représente les valeurs de t'/t en fonction du nombre d'échantillons T , où t et t' sont respectivement les temps de calcul des algorithmes standard et optimisé. Pour chaque configuration, nous avons réalisé 100 tests de Monte-Carlo en faisant varier les sources et les coefficients de mélange, et avons moyenné les valeurs de t'/t . On peut constater sur la figure 9.5 que pour un nombre de sources (un nombre de sources non-stationnaires dans le cas différentiel) inférieur ou égal à 7 pour les algorithmes (sur-)déterminés et inférieur ou égal à 6 pour les algorithmes différentiels, le rapport t'/t est nettement inférieur à 1 quand le nombre d'échantillons T est assez grand et décroît généralement quand T augmente, ce qui montre l'efficacité de nos méthodes optimisées.

Dans la seconde série de tests, nous avons représenté l'indice de performances en fonction du temps de calcul des algorithmes symétriques standard et optimisés dans la même configuration que ci-dessus. L'indice de performances mentionné ci-dessus est défini par

$$Perf = mean_j \left\{ max_i 10 \log_{10} \left(\frac{g_{ij}^2}{\sum_{j \neq i} g_{ij}^2} \right) \right\} \quad (9.5)$$

où les valeurs g_{ij} sont les coefficients de la matrice de performances \mathbf{G} égale au produit des matrices de mélange et de séparation. Dans le cas différentiel, nous ne prenons en compte que les colonnes de \mathbf{G} associées aux sources non-stationnaires et nous évaluons donc la qualité de la séparation partielle des sources utiles.

Comme précédemment, pour chaque configuration associée à un nombre de sources et à un temps de calcul donnés, nous avons réalisé 100 tests de Monte-Carlo et avons moyenné les valeurs de $Perf$. Nous avons choisi un nombre total

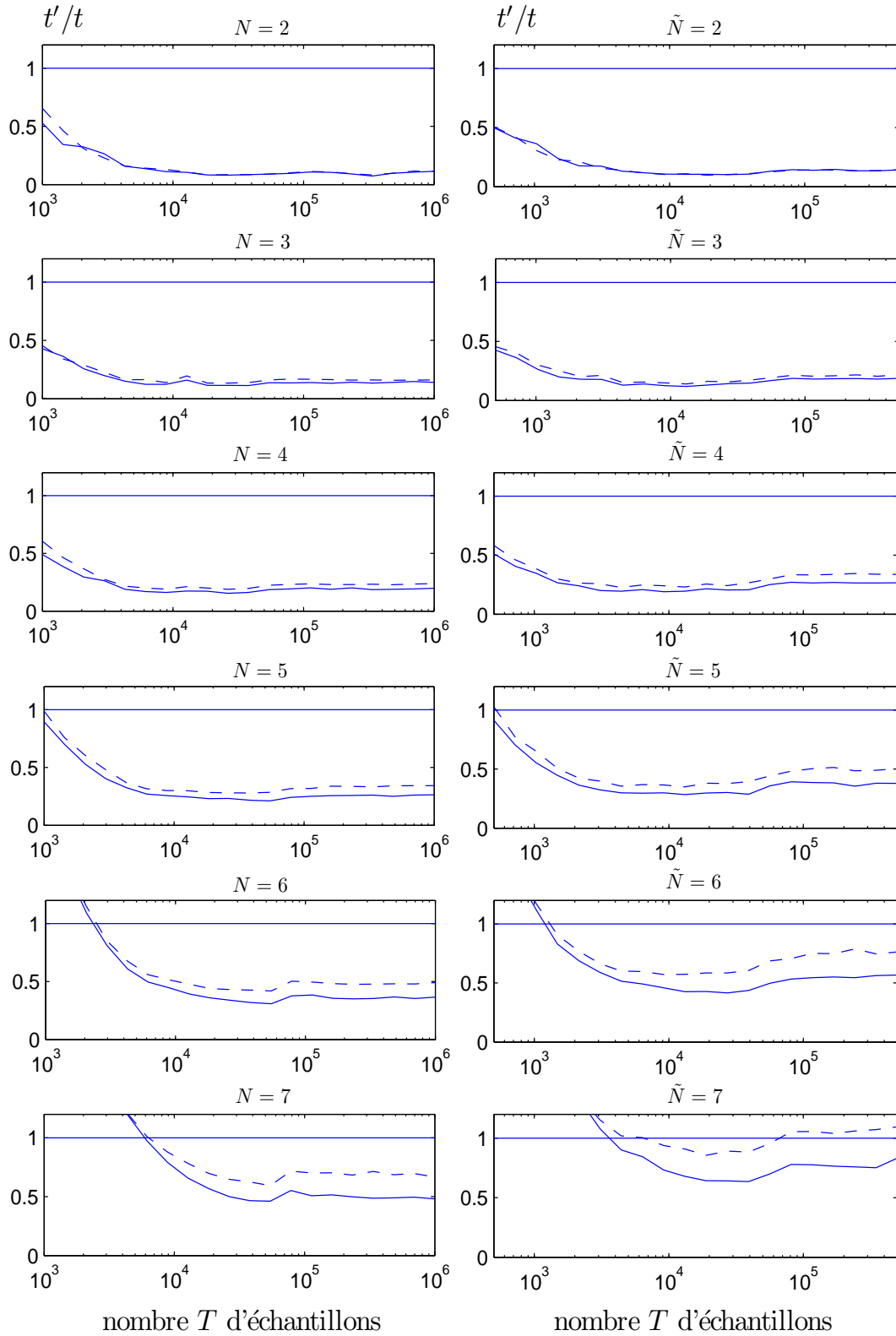


FIG. 9.5 – Rapport t'/t des temps de calcul en fonction du nombre T d'échantillons. En traits pleins : algorithmes symétriques, en pointillés : algorithmes à déflation. A gauche : algorithmes non-différentiels (sur-)déterminés FastICA et O-FICA. A droite : algorithmes différentiels sous-déterminés DFICA et O-DFICA.

d'échantillons égal à 100000 ($T = 50000$ dans le cas différentiel). La figure 9.6 montre que les algorithmes optimisés atteignent la plus grande valeur de $Perf$ en un temps de calcul plus court (excepté pour $\tilde{N} = 7$ dans le cas différentiel). Nous voyons aussi que pour les algorithmes FastICA et DFICA, l'indice de performances augmente progressivement contrairement à nos versions optimisées pour lesquelles l'indice augmente très rapidement au moment où les coefficients polynômiaux ont été estimés (le temps de l'optimisation dans le nouvel espace est négligeable comparé à l'identification polynômiale).

9.4 Conclusion

Dans ce chapitre, nous avons démontré l'intérêt de nos algorithmes optimisés O-FICA et O-DFICA par rapport aux méthodes standard associées, FastICA et DFICA. Une application en astrophysique a d'abord montré que dans le cas d'une séparation de deux images sources de dimension 4220×4298 , notre méthode O-FICA permet d'effectuer la séparation en un temps 25 fois plus court. Des tests statistiques ont ensuite montré que nos algorithmes permettent, lorsque le nombre d'observations donc de sources n'est pas trop élevé, d'obtenir la même performance de séparation en un temps plus court. Ces tests statistiques, pour des raisons de temps de calcul, ont été limitées à un nombre d'échantillons égal à 1 million. Nos méthodes se sont alors révélées 2 à 10 fois plus rapides que les méthodes standard. Pour des nombres d'échantillons très grands comme dans le cas des images astrophysiques de grande taille, le gain s'avère supérieur. Nos méthodes permettent aussi, en un temps de calcul donné, d'obtenir une meilleure séparation.

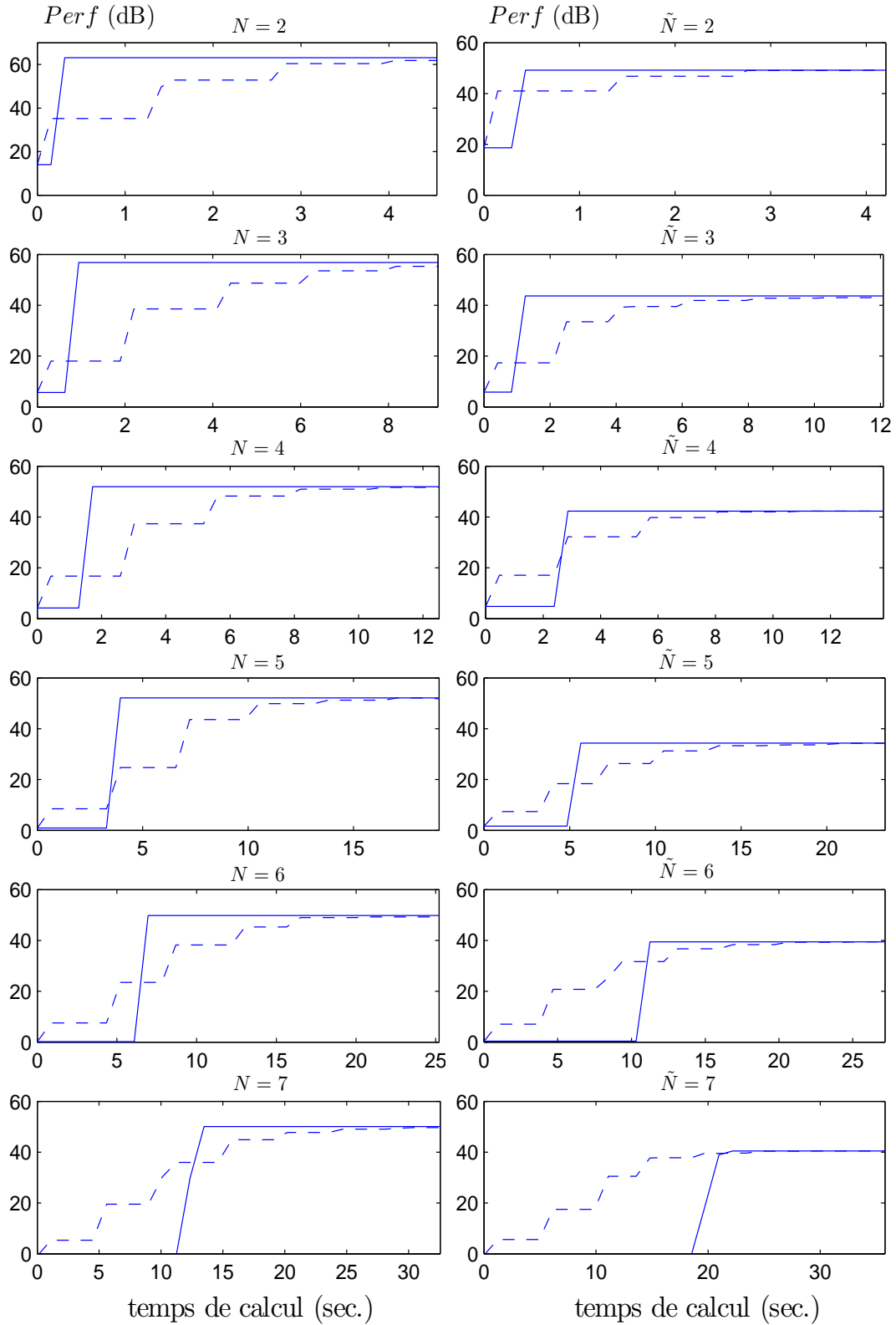


FIG. 9.6 – Indice de performances $Perf$ en fonction du temps de calcul.
 En traits pleins : algorithmes symétriques optimisés.
 En pointillés : algorithmes symétriques standard.
 A gauche : algorithmes non-différentiels (sur-)déterminés FastICA et O-FICA.
 A droite : algorithmes différentiels sous-déterminés DFICA et O-DFICA.

Conclusion générale

Ce travail de thèse a eu pour objectif de proposer des algorithmes temporels de type point fixe pour une classe de mélanges plus large que la classe des mélanges linéaires instantanés (sur-)déterminés, domaine d'application de l'algorithme FastICA. Nous nous sommes d'abord intéressés aux mélanges convolutifs (sur-)déterminés pour lesquels nous avons proposé un algorithme dénommé C-FICA, qui, grâce à une procédure de blanchiment convolutif non-causal et à une initialisation particulière des paramètres d'extraction, utilise des itérations d'optimisation de type point fixe pour estimer le processus d'innovation d'une source. Les contributions des sources sur les différents capteurs sont obtenues grâce à un critère quadratique correspondant à la puissance de la différence entre une observation et une version filtrée non-causale du processus d'innovation extrait. L'optimisation de ce critère conduit à l'adoption d'un processus de filtrage de Wiener non-causal. Des tests expérimentaux ont montré que notre approche permet de séparer des mélanges convolutifs artificiels de signaux réels audio pour des réponses impulsionnelles de filtres contenant 256 coefficients mesurés au niveau des oreilles d'une tête de mannequin. Notre méthode a aussi permis de séparer des mélanges de signaux de télécommunications pour des filtres artificiels d'ordre 10. Suivant le type de sources et la longueur des filtres, nous obtenons des *SIR* (rapport signal à interférence) de sortie compris entre 13 à 25 dB. Nous avons constaté que notre algorithme C-FICA permet d'effectuer la séparation lorsque deux sources se trouvent du même côté de la tête de mannequin, contrairement à d'autres algorithmes qui n'y parviennent pas dans ce cas de figure. Dans le cas des signaux audio, une procédure a été proposée, qui permet de rechercher les zones stationnaires des signaux de mélange. Des tests statistiques ont ensuite permis de valider notre algorithme en termes de performances de séparation et de rapidité. Nous avons notamment montré que notre algorithme est plus de 100 fois plus rapide et légèrement plus performant en termes de séparation que l'algorithme de Tugnait modifié (accéléré en utilisant une optimisation de type Newton), qui s'applique aux mêmes classes de mélanges et de sources. Cela permet d'utiliser des filtres de séparation d'ordre plus grand et d'obtenir ainsi de meilleures performances que cet algorithme de Tugnait modifié.

Dans la seconde partie de ce document, nous avons proposé des algorithmes de type point fixe pour les mélanges sous-déterminés instantanés ou convolutifs. Contrairement à beaucoup d'algorithmes destinés aux mélanges sous-déterminés, nous n'utilisons pas ici l'hypothèse de parcimonie des sources. Le concept de séparation différentielle de sources que nous employons nous permet d'effectuer la séparation partielle de sources utiles non-stationnaires en présence de sources de bruit stationnaires. L'algorithme que nous proposons pour les mélanges linéaires instantanés sous-déterminés, baptisé DFICA, utilise un blanchiment différentiel des observations de façon à employer un algorithme de type point fixe pour extraire les sources, soit itérativement par une approche à déflation, soit de façon symétrique en utilisant un critère de kurtosis dit différentiel. La convergence globale de ces algorithmes est prouvée rigoureu-

sement et l'on montre que cette convergence est cubique. Pour la séparation des mélanges convolutifs sous-déterminés, nous avons étendu le processus de blanchiment spatio-temporel non causal à la séparation différentielle de sources et avons proposé l'algorithme C-DFICA qui emploie un processus de blanchiment convolutif différentiel non-causal en préalable à l'optimisation du critère du kurtosis différentiel. Un critère quadratique différentiel est proposé pour l'estimation des contributions des sources sur les capteurs, et son optimisation est réalisée au moyen d'un processus de filtrage de Wiener non-causal dit différentiel. Nous avons ensuite réalisé des tests expérimentaux de séparation de mélanges sous-déterminés linéaires instantanés ou convolutifs de signaux réels audio ou de télécommunications. Dans le cas des mélanges convolutifs, comme dans la première partie, nous avons utilisé des filtres d'ordre 256 mesurés au niveau des oreilles d'une tête de mannequin. Ces tests ont confirmé l'utilité de nos méthodes, qui, contrairement aux méthodes standard, permettent d'effectuer la séparation partielle de sources utiles en présence de sources de bruit relativement puissantes. Des tests de Monte-Carlo ont permis ensuite de comparer les performances et la vitesse de notre algorithme C-DFICA avec une version différentielle de l'algorithme de Tugnait accéléré développée auparavant dans l'équipe. Ceux-ci ont montré que notre algorithme C-DFICA est à peu près 500 fois plus rapide, tout en donnant des performances de séparation légèrement meilleures.

Dans la dernière partie de ce document, nous avons présenté une nouvelle approche pour séparer rapidement les mélanges linéaires instantanés (sur-)déterminés ou sous-déterminés de sources contenant un grand nombre d'échantillons. Nos méthodes se fondent sur l'observation du fait que les algorithmes FastICA et DFICA calculent des moments empiriques à chaque itération de l'optimisation à point fixe du critère, ce qui implique le calcul de nombreuses opérations élémentaires. Pour y remédier, nous proposons d'identifier les critères du kurtosis et du kurtosis différentiel sous forme de polynômes multivariés relativement aux coefficients d'extraction appliqués au vecteur d'observation blanchi. Cette étape d'identification permet ensuite d'effectuer les itérations d'optimisation en un temps très court. Un test expérimental sur des données astrophysiques du télescope Hubble a montré que notre approche O-FICA pour mélanges (sur-)déterminés permet de diviser le temps de calcul par 25 par rapport à FastICA lorsque la taille de l'image avoisine les 16 millions de pixels. Des tests de Monte-Carlo ont ensuite montré que pour un nombre de sources inférieur ou égal à 7, nos algorithmes O-FICA et O-DFICA permettent d'atteindre un seuil de critère d'arrêt donné en un temps plus court, c.-à-d. jusqu'à 10 fois plus court pour un mélange de 2 sources contenant 1 million d'échantillons. Les itérations d'optimisation étant effectuées très rapidement grâce à notre approche, celle-ci permet en fait de réduire considérablement le seuil du critère d'arrêt sans affecter sensiblement le temps de calcul. Elle permet donc d'obtenir de meilleures performances de séparation en un temps donné.

Au sujet des méthodes proposées dans les deux premières parties de cette thèse, nous pouvons envisager à l’avenir de développer des procédures pour choisir au mieux les fenêtres d’extraction où les processus d’innovation sont estimés. Nos méthodes nécessitant que les sources à séparer soient stationnaires sur ces fenêtres, les performances de séparation pourraient se trouver améliorées. Dans le cas des mélanges (sur-)déterminés par exemple, la méthode proposée qui calcule les coefficients de corrélation de sorties obtenues sur plusieurs intervalles est relativement simple et pourrait être perfectionnée en y incluant des mesures de stationnarité. Dans le cas des méthodes différentielles qui demandent à la fois la stationnarité des sources sur chaque fenêtre d’analyse et des variations de puissances de sources de même signe d’une fenêtre à une autre, une procédure de choix automatique des deux fenêtres d’extraction tenant compte de ces deux contraintes pourrait être envisagée.

Nous pensons aussi étendre nos algorithmes C-FICA et C-DFICA aux mélanges convolutifs à filtres et à sources complexes. Cela serait intéressant pour pouvoir utiliser nos méthodes dans le cas d’applications en télécommunications. Nous souhaitons aussi proposer des versions symétriques de ces algorithmes de façon à éviter l’accumulation d’erreurs d’estimation durant le processus de déflation. Enfin, une version optimale de C-FICA, en se basant sur l’algorithme instantané EFICA qui atteint la borne de Cramer-Rao, pourrait aussi être envisagée.

Concernant la dernière partie de cette thèse, nous pensons à l’avenir étendre notre méthode O-FICA au critère de la néguentropie. En effet, il existe des approximations de la néguentropie qui utilisent des fonctions non-quadratiques $G(x)$ polynômiales. La néguentropie est alors encore un polynôme multi-variable relativement aux coefficients d’extraction. Une première étape d’identification de ce polynôme permettrait donc d’effectuer les itérations d’optimisation de façon plus rapide, ce qui diminuerait le temps de calcul par rapport à l’algorithme FastICA basé sur la néguentropie.

Dans les trois parties de cette thèse, nous avons proposé des méthodes appartenant à la classe de l’ACI. Ces méthodes supposent que les sources sont stationnaires et ergodiques. En effet, ces méthodes demandent l’estimation de statistiques par des moyennages empiriques, ce qui est réalisable seulement si les sources sont stationnaires et ergodiques. Récemment, nous avons mis en évidence qu’en choisissant des fenêtres d’extraction de taille réduite, du fait de la stationnarité à court terme des sources qui est mieux vérifiée en pratique, nous pouvons obtenir de meilleures performances de séparation pour des signaux non-stationnaires à long terme mais stationnaires à court terme. Nous développons actuellement, pour les mélanges linéaires instantanés (sur-)déterminés dans un premier temps, une méthode qui effectue des ACI successivement sur une série de zones temporelles ou spatiales de taille ré-

duite et éventuellement recouvrantes, puis classifie les vecteurs de séparation obtenus pour obtenir une série de vecteurs correspondant aux centroïdes des clusters détectés. Nous avons démontré que cette approche permet d'effectuer la séparation de sources empiriquement corrélées à long terme mais empiriquement décorréliées à court terme. Cela n'est pas le cas de la plupart des méthodes d'ACI car celles-ci commencent par blanchir empiriquement les observations et estiment ensuite une matrice unitaire réalisant la séparation à partir du vecteur d'observations blanchies : les sources estimées sont donc par construction empiriquement décorréliées à long terme. Nos premiers résultats expérimentaux montrent que notre approche, en utilisant l'algorithme de classification K-medians, donne une nette amélioration des performances de séparation dans le cas de sources audio et d'images non-stationnaires, et permet une séparation de sources corrélées à long terme, contrairement aux méthodes d'ACI standard qui échouent lorsqu'on effectue l'ACI sur l'intégralité des échantillons. Un article de conférence est en cours de soumission concernant ce travail. Cette approche pourra ensuite être étendue aux mélanges convolutifs (sur-)déterminés et sous-déterminés, ce qui donnera lieu prochainement à des travaux théoriques et expérimentaux.

Annexes

Annexe A : Blanchiment

L'étape de blanchiment d'un jeu d'observations centrées $x_1(n), \dots, x_P(n)$ consiste à décorréliser mutuellement ces observations entre elles. Le plus souvent, cette étape de décorrélation est suivie d'une étape de normalisation de sorte que les signaux obtenus sont de puissance unité. Soit \mathbf{R}_x la matrice d'autocorrélation des observations, dont chaque coefficient (i, j) est défini par le coefficient de corrélation entre la $i^{\text{ème}}$ observation et la $j^{\text{ème}}$ observation :

$$E \{x_i(n)x_j(n)\} \quad (1)$$

La matrice \mathbf{R}_x est alors définie matriciellement par

$$\mathbf{R}_x = E \{ \mathbf{x}(n)\mathbf{x}(n)^t \}. \quad (2)$$

Effectuons maintenant la diagonalisation de la matrice \mathbf{R}_x , qui donne une matrice orthogonale \mathbf{E} de vecteurs propres regroupés par colonnes, et une matrice diagonale Δ constituée des valeurs propres de \mathbf{R}_x . Soit $\mathbf{y}(n)$ le vecteur signal défini par

$$\mathbf{y}(n) = \mathbf{E}^t \mathbf{x}(n) \quad (3)$$

et calculons sa matrice d'autocorrélation \mathbf{R}_y :

$$\begin{aligned} \mathbf{R}_y &= E \{ \mathbf{y}(n)\mathbf{y}(n)^t \} \\ &= E \{ \mathbf{E}^t \mathbf{x}(n) (\mathbf{E}^t \mathbf{x}(n))^t \} \\ &= \mathbf{E}^t E \{ \mathbf{x}(n)\mathbf{x}(n)^t \} \mathbf{E} \\ &= \mathbf{E}^t \mathbf{R}_x \mathbf{E} \\ &= \mathbf{E}^t \mathbf{E} \Delta \mathbf{E}^t \mathbf{E} \\ &= \Delta \end{aligned} \quad (4)$$

La matrice d'autocorrélation du vecteur signal $\mathbf{y}(n)$ est donc diagonale, ce qui signifie que les signaux $y_1(n), \dots, y_P(n)$ sont mutuellement décorrélés (le vecteur d'observations $\mathbf{y}(n)$ est dit blanchi), et vérifient

$$\forall i, j = 1 \dots P, E \{ z_i(n)z_j(n) \} = \delta_{ij} \Delta_i \quad (5)$$

où δ_{ij} est le symbole de Kronecker défini par

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}, \quad (6)$$

et Δ_i est la $i^{\text{ème}}$ valeur propre de $\mathbf{R}_\mathbf{x}$. La matrice $\mathbf{R}_\mathbf{x}$ étant une matrice d'autocorrélation, on peut montrer qu'elle est définie positive, c.-à-d. que toutes ses valeurs propres (regroupées dans la matrice diagonale Δ) sont strictement positives.

Soit le vecteur signal $\mathbf{z}(n)$ défini par

$$\mathbf{z}(n) = \Delta^{-1/2} \mathbf{y}(n) = \Delta^{-1/2} \mathbf{E}^t \mathbf{x}(n), \quad (7)$$

où $\Delta^{-1/2}$ est la matrice diagonale constituée des racines carrées inverses des valeurs propres de $\mathbf{R}_\mathbf{x}$. Calculons la matrice d'autocorrélation de $\mathbf{z}(n)$:

$$\begin{aligned} \mathbf{R}_\mathbf{z} &= E \{ \mathbf{z}(n) \mathbf{z}(n)^t \} \\ &= E \{ \Delta^{-1/2} \mathbf{y}(n) (\Delta^{-1/2} \mathbf{y}(n))^t \} \\ &= \Delta^{-1/2} E \{ \mathbf{y}(n) \mathbf{y}(n)^t \} \Delta^{-1/2} \\ &= \Delta^{-1/2} \Delta \Delta^{-1/2} \\ &= \mathbf{I} \end{aligned} \quad (8)$$

La matrice d'autocorrélation de $\mathbf{z}(n)$ est donc égale à la matrice identité, ce qui implique que ses signaux sont mutuellement décorrélés et de puissance unité (le vecteur d'observations $\mathbf{z}(n)$ est dit blanchi et normalisé). La figure 1 illustre sous forme de schéma-bloc les deux étages successifs de décorrélation et de normalisation.

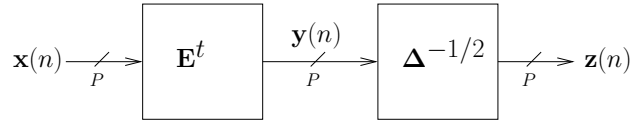


FIG. 1 – Schéma-bloc des étages de blanchiment et de normalisation.

On définit aussi parfois $\mathbf{z}(n)$ par

$$\mathbf{z}(n) = \mathbf{E} \Delta^{-1/2} \mathbf{E}^t \mathbf{x}(n), \quad (9)$$

définition qui fait intervenir la matrice $\mathbf{E} \Delta^{-1/2} \mathbf{E}^t$, dite matrice racine carrée inverse de la matrice d'autocorrélation $\mathbf{R}_\mathbf{x}$. On montre facilement que la matrice d'autocorrélation de ce nouveau vecteur signal $\mathbf{z}(n)$ est aussi égale à la matrice identité. La figure 2 ci-dessous illustre les deux possibilités de blanchiment.

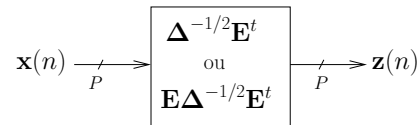


FIG. 2 – Schéma-bloc du processus de blanchiment et de normalisation.

Le processus de blanchiment (ou de décorrélation) et de normalisation décrit ci-dessus est appelé 'sphering' en anglais. Cette dénomination vient du fait que si l'on applique un vecteur d'extraction \mathbf{w} au vecteur d'observation blanchi normalisé $\mathbf{z}(n)$, le signal $\mathbf{w}^t \mathbf{z}(n)$ est de puissance unité si et seulement si le vecteur \mathbf{w} appartient à la sphère unité :

$$E \left\{ \left(\mathbf{w}^t \mathbf{z}(n) \right)^2 \right\} = 1 \Leftrightarrow \|\mathbf{w}\| = 1 \quad (10)$$

Annexe B : Estimation des contributions des sources

Considérons la différence $x'_k(n)$ entre l'observation $x_k(n)$ et une version filtrée du signal extrait précédent $e_l(n)$, c.-à-d.

$$x'_k(n) = x_k(n) - \sum_{r=-R'}^{R'} c(r)e_l(n-r) \quad (1)$$

où les $c(r)$ sont les coefficients d'un filtre RIF non causal. En raison des hypothèses de la section 2.1, $x_k(n)$ est un mélange RIF causal de tous les processus d'innovation, c.-à-d. que (2.4) nous donne

$$x_k(n) = \sum_{q=1}^N \sum_{r=0}^L h_{kq}(r)u_q(n-r). \quad (2)$$

De plus,

$$e_l(n) = \beta_l u_l(n - r_l) \quad (3)$$

Donc, si R' est assez grand pour que $r_l - R' \leq 0$ et $r_l + R' \geq L$, (1)-(3) nous donnent

$$x'_k(n) = \sum_{r=r_l-R'}^{r_l+R'} [h_{kl}(r) - \beta_l c(r - r_l)]u_l(n-r) + \sum_{q \neq l} \sum_{r=0}^L h_{kq}(r)u_q(n-r) \quad (4)$$

La puissance de $x'_k(n)$ vaut

$$\begin{aligned} E \{x_k'^2(n)\} &= \sum_{r=r_l-R'}^{r_l+R'} [h_{kl}(r) - \beta_l c(r - r_l)]^2 E \{u_l^2(n-r)\} \\ &\quad + \sum_{q \neq l} \sum_{r=0}^L h_{kq}^2(r) E \{u_q^2(n-r)\}. \end{aligned} \quad (5)$$

Comme nous avons supposé que toutes les puissances $E \{u_l^2(n-r)\}$ sont strictement positives, l'éq. (5) montre que la minimisation de $E \{x_k'^2(n)\}$ en fonction de $c(r)$ correspond exactement à

$$h_{kl}(r) - \beta_l c(r - r_l) = 0, \quad \forall r \in [r_l - R', \dots, r_l + R']. \quad (6)$$

En comparant ce résultat avec (4), nous en déduisons que pour obtenir une observation modifiée $x'_k(n)$ où la contribution de la $l^{\text{ème}}$ source a été enlevée consiste à minimiser $E\{x'^2_k(n)\}$. Nous utilisons la méthode de Newton pour optimiser ce critère, qui est noté J dans la suite et peut être exprimé par

$$J = E\left\{\left(x_k(n) - \sum_{r=-R'}^{R'} c(r)e_l(n-r)\right)^2\right\}. \quad (7)$$

L'équation de mise à jour du vecteur $\mathbf{c} = [c(-R'), \dots, c(R')]^t$ s'écrit

$$\mathbf{c} \leftarrow \mathbf{c} - \left(\frac{\partial^2 J}{\partial \mathbf{c}^2}\right)^{-1} \frac{\partial J}{\partial \mathbf{c}} \quad (8)$$

où $\frac{\partial J}{\partial \mathbf{c}}$ et $\frac{\partial^2 J}{\partial \mathbf{c}^2}$ sont respectivement le gradient et le Hessian du critère J . Calculons la $i^{\text{ème}}$ composante du gradient de J par rapport à \mathbf{c} :

$$\frac{\partial J}{\partial c_i} = 2E\left\{-e_l(n-i) \cdot \left(x_k(n) - \sum_{r=-R'}^{R'} c(r)e_l(n-r)\right)\right\} \quad (9)$$

En particulier

$$\left.\frac{\partial J}{\partial c_i}\right|_{\mathbf{c}=0} = 2E\{-e_l(n-i)x_k(n)\}. \quad (10)$$

Donc nous avons

$$\left.\frac{\partial J}{\partial \mathbf{c}}\right|_{\mathbf{c}=0} = -2\mathbf{r}_{e_l x_k}(n) \quad (11)$$

où $\mathbf{r}_{e_l x_k}(n_p)$ est le vecteur d'intercorrélation entre les signaux $e_l(n-i)$, avec $i \in [-R', R']$ et $x_k(n_p)$. Calculons maintenant la $(i, j)^{\text{ème}}$ composante du Hessian de J :

$$\frac{\partial^2 J}{\partial c_i \partial c_j} = 2E\{e_l(n-i)e_l(n-j)\}. \quad (12)$$

Donc, nous avons

$$\frac{\partial^2 J}{\partial \mathbf{c}^2} = 2\mathbf{R}_{e_l}(n) \quad (13)$$

où $\mathbf{R}_{e_l}(n_p)$ est la matrice d'autocorrélation associée aux valeurs de $e_l(n)$. En initialisant \mathbf{c} avec $\mathbf{c} = 0$, la valeur de \mathbf{c} après une itération de l'algorithme de Newton s'écrit

$$\begin{aligned} \mathbf{c}' &= -\left(\frac{\partial^2 J}{\partial \mathbf{c}^2}\right)^{-1} \left.\frac{\partial J}{\partial \mathbf{c}}\right|_{\mathbf{c}=0} \\ &= -\frac{1}{2}\mathbf{R}_{e_l}(n)^{-1}(-2)\mathbf{r}_{e_l x_k}(n) \\ &= \mathbf{R}_{e_l}(n)^{-1}\mathbf{r}_{e_l x_k}(n). \end{aligned} \quad (14)$$

Sachant que l'algorithme de Newton atteint un point stationnaire en une itération pour les critères quadratiques et que notre critère est quadratique positif, nous prouvons que le vecteur \mathbf{c}' correspond au maximum global de J . Notons que cette expression est celle d'un filtre de Wiener.

Annexe C : Définitions des critères SIR_{out} , SIR_{in} , $SIRI$ et SNR_{in}

Pour chaque source $s_j(n)$ et chaque contribution estimée de la source k dans l'observation i notée $\hat{x}_{ik}(n)$, nous définissons le $SIR_{out}(i, j, k)$ associé comme le rapport entre la puissance du "signal" et la puissance de l'"interférence" où

- le "Signal" est la valeur idéale $x_{ij}(n)$ de $\hat{x}_{ik}(n)$ qui est la contribution de $s_j(n)$ dans $x_i(n)$.
- l'"Interférence" est la déviation de $\hat{x}_{ik}(n)$ par rapport à sa valeur idéale, c.-à-d. $\hat{x}_{ik}(n) - x_{ij}(n)$.

Cela donne $\forall i \in \{1, \dots, P\}, \forall j \in \{1, \dots, N\}, \forall k \in \{1, \dots, P\}$,

$$SIR_{out}(i, j, k) = 10 \log_{10} \left(\frac{E \{x_{ij}^2(n)\}}{E \{(\hat{x}_{ik}(n) - x_{ij}(n))^2\}} \right). \quad (1)$$

Pour chaque source j et chaque sortie k , on considère seulement le seul $SIR_{out}(i, j, k)$ correspondant à l'observation i donnant le signal \hat{x}_{ik} qui a la puissance la plus élevée. Ensuite, pour chaque indice de source j , on considère seulement le maximum par rapport à l'indice de sortie k des valeurs $SIR_{out}(i, j, k)$. Cela nous donne un unique SIR pour chaque source utile.

Le SIR d'entrée (SIR_{in}) de la source j dans les observations est lui défini par

$$\forall j \in \{1, \dots, N\}, \quad SIR_{in}(j) = \max_i 10 \log_{10} \left(\frac{E \{x_{ij}^2(n)\}}{\sum_{k=1..N}^{k \neq j} E \{x_{ik}^2(n)\}} \right). \quad (2)$$

Pour une source donnée, on définit ensuite l'amélioration du SIR , $SIRI$ par

$$SIRI(j) = SIR_{out}(j) - SIR_{in}(j) \quad (3)$$

D'une façon similaire, nous pouvons définir le rapport signal à bruit des observations par

$$\forall i \in \{1, \dots, N\}, \quad SNR_{in}(i) = 10 \log_{10} \left(\frac{\sum_{k \in \mathcal{U}} E \{x_{ik}^2(n)\}}{\sum_{k \in \mathcal{B}} E \{x_{ik}^2(n)\}} \right) \quad (4)$$

où \mathcal{U} et \mathcal{B} sont respectivement les ensembles d'indices associées aux sources utiles et aux sources de bruit. Les $SNR_{in}(i)$ peuvent ensuite être moyennés pour obtenir un unique critère pour l'ensemble des observations.

Annexe D : Gradient du kurtosis différentiel

Le kurtosis non-normalisé d'un signal centré $y(n) = \mathbf{w}^t \mathbf{z}(n)$ est défini par

$$kurt_y(n) = E \{y^4(n)\} - 3 (E \{y^2(n)\})^2. \quad (1)$$

En calculant sa dérivée première par rapport à \mathbf{w} , nous obtenons

$$\begin{aligned} \frac{\partial kurt_y(n)}{\partial \mathbf{w}} &= \frac{\partial}{\partial \mathbf{w}} \left(E \{(\mathbf{w}^t \mathbf{z}(n))^4\} - 3 (E \{(\mathbf{w}^t \mathbf{z}(n))^2\})^2 \right) \\ &= E \{4\mathbf{z}(n)(\mathbf{w}^t \mathbf{z}(n))^3\} - 6E \{(\mathbf{w}^t \mathbf{z}(n))^2\} E \{2\mathbf{z}(n)\mathbf{w}^t \mathbf{z}(n)\} \\ &= 4E \{\mathbf{z}(n)y^3(n)\} - 12E \{y^2(n)\} E \{\mathbf{z}(n)\mathbf{z}(n)^t\} \mathbf{w} \\ &\propto E \{\mathbf{z}(n)y^3(n)\} - 3E \{y^2(n)\} \mathbf{R}_{\mathbf{z}}(n)\mathbf{w}. \end{aligned} \quad (2)$$

Notons que dans l'approche FastICA non différentielle, l'expression du gradient (2) peut être simplifiée en raison de l'étape de décorrélation et de normalisation qui entraîne $E \{y^2(n)\} = 1$ et $\mathbf{R}_{\mathbf{z}}(n) = \mathbf{I}$. L'expression du gradient est alors

$$\frac{\partial kurt_y(n)}{\partial \mathbf{w}} \propto E \{\mathbf{z}(n)y^3(n)\} - 3\mathbf{w}. \quad (3)$$

Dans notre cas, le gradient différentiel s'écrit

$$\begin{aligned} \frac{\partial Dkurt_y(n_1, n_2)}{\partial \mathbf{w}} &= \frac{\partial kurt_y(n_2)}{\partial \mathbf{w}} - \frac{\partial kurt_y(n_1)}{\partial \mathbf{w}} \\ &\propto E \{\mathbf{z}(n_2)y^3(n_2)\} - 3E \{y^2(n_2)\} \mathbf{R}_{\mathbf{z}}(n_2)\mathbf{w} \\ &\quad - E \{\mathbf{z}(n_1)y^3(n_1)\} + 3E \{y^2(n_1)\} \mathbf{R}_{\mathbf{z}}(n_1)\mathbf{w}. \end{aligned} \quad (4)$$

En utilisant les relations $\mathbf{R}_{\mathbf{z}}(n_2) - \mathbf{R}_{\mathbf{z}}(n_1) = \mathbf{I}$, $E \{y^2(n_i)\} = \mathbf{w}^t \mathbf{R}_{\mathbf{z}}(n_i) \mathbf{w} \ \forall i \in \{1, 2\}$ et $E \{y^2(n_2)\} - E \{y^2(n_1)\} = 1$, on peut simplifier l'expression (4) pour obtenir

$$\begin{aligned} \frac{\partial Dkurt_y(n_1, n_2)}{\partial \mathbf{w}} &\propto E \{\mathbf{z}(n_2)y^3(n_2)\} - E \{\mathbf{z}(n_1)y^3(n_1)\} \\ &\quad - 3 [\mathbf{R}_{\mathbf{z}}(n_1) + (1 + \mathbf{w}^t \mathbf{R}_{\mathbf{z}}(n_1) \mathbf{w}) \mathbf{I}] \mathbf{w}. \end{aligned} \quad (5)$$

Annexe E : La corrélation différentielle

Nous supposons ici que le signal extrait contient seulement une seule source utile ainsi que des sources de bruits de sorte que

$$y(n) = \varepsilon s_l(n) + \sum_{q \notin I} \alpha_q s_q(n) \quad (1)$$

avec $l \in I$. Le facteur d'échelle ε associé à $s_l(n)$ dans (1) peut être obtenu grâce à la puissance différentielle de $y(n)$ supposée égale à 1 et grâce à la puissance différentielle unitaire des sources non-stationnaires, c.-à-d.

$$Dpow_{s_l}(n_1, n_2) = 1, \quad \forall l \in I \quad (2)$$

qui implique que $\varepsilon = \pm 1$. Dans la suite, nous supposons sans perte de généralité que $\varepsilon = 1$ (autrement, nous pouvons redéfinir $s_l(n)$ par son opposée).

En outre, la $k^{ème}$ observation s'écrit

$$x_k(n) = \sum_{q=1}^N a_{kq} s_q(n). \quad (3)$$

Calculons maintenant la corrélation différentielle entre $y(n)$ et $x_k(n)$, définie par

$$Dcorr_{yx_k}(n_1, n_2) = E \{y(n_2)x_k(n_2)\} - E \{y(n_1)x_k(n_1)\}. \quad (4)$$

Donc

$$\begin{aligned} Dcorr_{yx_k}(n_1, n_2) &= a_{kl} Dcorr_{s_l s_l}(n_1, n_2) + \sum_{q \neq l} a_{kq} Dcorr_{s_l s_q}(n_1, n_2) \\ &\quad + \sum_{q_1 \notin I, q_2} \alpha_{q_1} a_{kq_2} Dcorr_{s_{q_1} s_{q_2}}(n_1, n_2) \\ &= a_{kl} Dpow_{s_l}(n_1, n_2) \\ &= a_{kl} \end{aligned} \quad (5)$$

comme $Dcorr_{s_{q_1} s_{q_2}}(n_1, n_2) = 0, \forall q_1 \neq q_2$ en raison de l'indépendance des sources centrées, $Dcorr_{s_q s_q}(n_1, n_2) = 0, \forall q \notin I$ en raison de la stationnarité

des sources de bruit et en prenant en compte (2).

$Dcorr_{yx_k}(n_1, n_2)$ donne donc le facteur d'échelle associé à la contribution de la $l^{ème}$ source dans la $k^{ème}$ observation.

Annexe F : Définitions des critères SIR_{out} , SIR_{in} et $SIRI$ pour les méthodes de séparation partielle

Nous définissons ici les critères utilisés dans le chapitre 3 pour évaluer les performances des algorithmes à déflation. Pour chaque test, nous appliquons d'abord les mélanges de toutes les sources au système de SAS et estimons les paramètres séparation. Comme nous cherchons à évaluer la qualité de la séparation partielle seulement entre les sources non-stationnaires, nous transférons ensuite seulement ces sources à travers les étages de mélange de séparation. Nous obtenons ainsi un jeu de contributions estimées $\hat{x}_{ik}(n)$ associées à un signal extrait $y_k(n)$ et à une observation $x_i(n)$. Pour chaque source $s_j(n)$ et chaque signal $\hat{x}_{ik}(n)$, nous définissons ensuite un $SIR_{out}(i, j, k)$ associé comme le rapport entre la puissance du "signal" et la puissance de l'"interférence" où

- le "Signal" est la valeur idéale $x_{ij}(n)$ de $\hat{x}_{ik}(n)$ qui est la contribution de $s_j(n)$ dans $x_i(n)$.
- l'"Interférence" est la déviation de $\hat{x}_{ik}(n)$ par rapport à sa valeur idéale, c.-à-d. $\hat{x}_{ik}(n) - x_{ij}(n)$.

Cela donne $\forall i \in \{1, \dots, P\}, \forall j \in I, \forall k \in \{1, \dots, P\}$,

$$SIR_{out}(i, j, k) = 10 \log_{10} \left(\frac{E \{x_{ij}^2(n)\}}{E \{(\hat{x}_{ik}(n) - x_{ij}(n))^2\}} \right). \quad (1)$$

Pour chaque source j et chaque sortie k , on considère seulement le seul $SIR_{out}(i, j, k)$ correspondant à l'observation i donnant le signal \hat{x}_{ik} qui a la puissance la plus élevée. Ensuite, pour chaque indice de source j , on considère seulement le maximum par rapport à l'indice de sortie k des valeurs $SIR_{out}(i, j, k)$. Nous calculons ensuite la moyenne de ces valeurs sur les deux intervalles temporels. Cela nous donne un unique SIR pour chaque source utile, qui définit la performance de sortie de notre système, noté $SIR_{out}(j)$ ci-après. La moyenne de ces $SIR_{out}(j)$ sur les sources peut ensuite être calculée. Nous faisons cette moyenne pour calculer le SIR_{out} global des méthodes symétriques.

Le SIR d'entrée des observations est définie de manière similaire. Nous définissons d'abord le SIR d'entrée associé à chaque source par

$$\forall j \in I, \quad SIR_{in}(j) = \max_i 10 \log_{10} \left(\frac{E \{x_{ij}^2(n)\}}{\sum_{k \in I, k \neq j} E \{x_{ik}^2(n)\}} \right). \quad (2)$$

Nous calculons ensuite le SIR_{in} global comme la moyenne sur les sources et sur les domaines temporels.

Annexe G : Choix de la famille de vecteurs d'extraction

Définissons les 5 ensembles³² :

$$\begin{aligned}
 E_1 &= \left\{ \mathbf{w} \in \{0, 1\}^N \mid \exists i_1, w(i_1) = 1, \forall i \neq i_1, w(i) = 0 \right\} \\
 E_2 &= \left\{ \mathbf{w} \in \{0, 1\}^N \mid \exists i_1, i_2, w(i_1) = w(i_2) = 1, \forall i \neq i_1, i_2, w(i) = 0 \right\} \\
 E_3 &= \left\{ \mathbf{w} \in \{0, 1, 2\}^N \mid \exists i_1, i_2, w(i_1) = 2w(i_2) = 2, \forall i \neq i_1, i_2, w(i) = 0 \right\} \\
 E_4 &= \left\{ \mathbf{w} \in \{0, 1, 2\}^N \mid \exists i_1, i_2, i_3, w(i_1) = 2w(i_2) = 2w(i_3) = 2, \forall i \neq i_1, i_2, i_3, w(i) = 0 \right\} \\
 E_5 &= \left\{ \mathbf{w} \in \{0, 1\}^N \mid \exists i_1, \dots, i_4, w(i_1) = \dots = w(i_4) = 1, \forall i \neq i_1, \dots, i_4, w(i) = 0 \right\}
 \end{aligned}$$

Les cardinaux de E_1, \dots, E_5 valent respectivement $N, \frac{N(N-1)}{2}, N(N-1), \frac{N(N-1)(N-2)}{2}, \frac{N(N-1)(N-2)(N-3)}{24}$. En notant $E = \cup_i E_i$, $\text{card}(E) = \sum_i \text{card}(E_i) = N + \frac{3N(N-1)}{2} + \frac{N(N-1)(N-2)}{2} + \frac{N(N-1)(N-2)(N-3)}{24} = \text{card}(D) = R$.

De plus, nous avons vérifié numériquement que cette famille donne une matrice \mathbf{M} définie par (8.11) non singulière pour un nombre de sources $N \leq 7$. Les conditionnements respectifs de \mathbf{M} (définis par le rapport de la plus grande et de la plus petite valeurs propres de \mathbf{M}) pour $N = 2, \dots, 7$ valent en effet 182, 414, 844, 1605, 2758, 4344, ce qui est assez faible pour les valeurs de R associées (respectivement 5, 15, 35, 70, 126, 210). Par exemple, le conditionnement moyen de matrices de dimension 210 avec des coefficients uniformément distribués entre 0 et 1 est supérieur à 50000. Nous avons donc une famille de vecteurs qui peut être utilisée pour identifier le jeu de coefficients $(\alpha_r)_{r=1..R}$ défini par (8.8).

³²Dans le cas sous-déterminé, N doit être remplacé par \tilde{N} dans ce qui suit.

Bibliographie

- [1] *Musical audio stream separation by non-negative matrix factorization*. Ds Actes *DMRN Summer Conference*, Glasgow, Royaume-Uni, juil. 2005. B. Wang and M. D. Plumbley.
- [2] *Investigating single-channel audio source separation methods based on non-negative matrix factorization*. Ds Actes *ICArn'06*, p. 18–19, Liverpool, Royaume-Uni, sept. 2006. B. Wang and M. D. Plumbley.
- [3] K. Abed-Meraim, J. F. Cardoso, A. Y. Gorokhov, P. Loubaton et E. Moulines: *On Subspace Methods for Blind Identification of Single-Input Multiple-Output FIR Systems*. IEEE Trans. on Signal Processing, 45(1) :42–55, janv. 1997.
- [4] K. Abed-Meraim, Y. Hua, P. Loubaton et E. Moulines: *Subspace Method for Blind Identification of Multichannel FIR Systems in Noise Field with Unknown Spatial Covariance*. IEEE Signal Processing Letters, 4(5) :135–137, mai 1997.
- [5] K. Abed-Meraim, P. Loubaton et E. Moulines: *A Subspace Algorithm for Certain Blind Identification Problems*. IEEE Trans. on Information Theory, 43(2) :499–511, mars 1997.
- [6] F. Abrard: *Méthodes de séparation aveugle de sources et applications*. Thèse de doctorat, *Université Paul Sabatier - Toulouse 3*, France, pages 29-41, mars 2003.
- [7] F. Abrard et Y. Deville: *A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources*. Signal Processing, 85(7) :1389–1403, juil. 2005.
- [8] F. Abrard, Y. Deville et J. Thomas: *Blind partial separation of underdetermined convolutive mixtures of complex sources based on differential normalized kurtosis*. Neurocomputing. A paraître.
- [9] F. Abrard, Y. Deville et P. White: *From blind source separation to blind source cancellation in the underdetermined case : a new approach based on time-frequency analysis*. Ds Actes *Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'01)*, San Diego, Californie, déc. 2001.
- [10] B. Albouy: *Méthodes de séparation aveugle de sources applicables à des signaux de parole*. Thèse de doctorat, *Université Paul Sabatier - Toulouse 3*, France, déc. 2004.

- [11] B. Albouy et Y. Deville: *A time-frequency blind source separation method based on segmented coherence function*. Ds Actes Intl. Work. on Artificial Neural Networks (IWANN'03), vol. 2, p. 289–296, Minorque, Espagne, juin 2003.
- [12] B. Albouy et Y. Deville: *Alternative structures and power spectrum criteria for blind segmentation and separation of convolutive speech mixtures*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'03), p. 361–366, Nara, Japon, avr. 2003.
- [13] S. Amari, A. Cichocki et H. H. Yang: *A New Learning Algorithm for Blind Signal Separation*. Ds Actes Conf. on Advances in Neural Information Processing Systems (NIPS'95), vol. 8, p. 757–763, Denver, Colorado, nov. 1995.
- [14] J. Anemüller et B. Kollmeier: *Adaptive separation of acoustic sources for anechoic conditions : A constrained frequency domain approach*. Speech Communications, 39(1-2) :79–95, janv. 2003.
- [15] P. Anttila, P. Paatero, U. Tapper et O. Järvinen: *Source identification of bulk wet deposition in Finlande by positive matrix factorization*. Atmospheric Environment, 29(14) :1705–1718, juil. 1995.
- [16] S. Araki, R. Mukai et S. Makino: *The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech*. IEEE Trans. on Speech and Audio Processing, 11(2) :109–116, mars 2003.
- [17] A. Aïssa-El-Bey, K. Abed-Meraim et Y. Grenier: *Blind Separation of Underdetermined Convolutive Mixtures Using Their Time-Frequency Representation*. IEEE Trans. on Audio, Speech, and Language Processing, 15(5) :1540–1550, juil. 2007.
- [18] A. Aïssa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani et Y. Grenier: *Underdetermined Blind Separation of Nondisjoint Sources in the Time-Frequency Domain*. IEEE Trans. on Signal Processing, 55(3) :897–907, mars 2007.
- [19] M. Babaie-Zadeh et C. Jutten: *A general approach for mutual information minimization and its application to blind source separation*. Signal Processing, 85(5) :975–995, mai 2005.
- [20] A. D. Back et A. C. Tsoi: *Blind deconvolution of signals using a complex recurrent network*. Ds Actes Intl. Work. on Neural Networks for Signal Processing (NNSP'94), p. 565–574, Ermioni, Grèce, sept. 1994.
- [21] R. Balan, J. Rosca et S. Rickard: *Non-square blind source separation under coherent noise by beamforming and time-frequency masking*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'03), p. 313–318, Nara, Japon, avr. 2003.
- [22] R. Balan, J. Rosca et S. Rickard: *Scalable non-square blind source separation in the presence of noise*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'03), vol. 5, p. 293–296, Hong Kong, Chine, avr. 2003.

-
- [23] A. J. Bell et T. J. Sejnowski: *An information-maximisation approach to blind separation and blind deconvolution*. Neural Computation, 7(6) :1129–1159, juil. 1995.
 - [24] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso et E. Moulines: *A Blind Source Separation Technique Using Second-Order Statistics*. IEEE Trans. on Signal Processing, 45(2) :434–444, fév. 1997.
 - [25] A. Belouchrani et M. G. Amin: *Blind Source Separation Based on Time-Frequency Signal Representations*. IEEE Trans. on Signal Processing, 46(11) :2888–2897, nov. 1998.
 - [26] A. Belouchrani et J. F. Cardoso: *Maximum likelihood source separation for discrete sources*. Ds Actes *European Signal Processing Conference (EUSIPCO'94)*, vol. 2, p. 768–771, Edimbourg, Royaume-Uni, sept. 1994.
 - [27] L. Benaroya, F. Bimbot et R. Gribonval: *Audio Source Separation With a Single Sensor*. IEEE Trans. on Audio, Speech, and Language Processing, 14(1), janv. 2006.
 - [28] L. Benaroya, L. M. Donagh, F. Bimbot et R. Gribonval: *Non negative sparse representation for Wiener based source separation with a single sensor*. Ds Actes *Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'03)*, vol. 6, p. 613–616, Hong Kong, Chine, avr. 2003.
 - [29] O. Berné, C. Joblin, Y. Deville, J. D. Smith, M. Rapaccioli, J. P. Bernard, J. Thomas, W. Reach et A. Abergel: *Analysis of the emission of very small dust particles from Spitzer spectro-imagery data using Blind Signal Separation methods*. Astronomy & Astrophysics, 469(2) :575–586, juil. 2007.
 - [30] O. Berné, C. Joblin, M. Rapacioli, J. Thomas, J. C. Cuillandre et Y. Deville: *Extended Red Emission and the evolution of carbonaceous nano-grains in NGC 7023*. Astronomy & Astrophysics, 479(3) :L41–44, mars 2008.
 - [31] F. Berthommier et S. Choi: *Several improvements of the Herault-Jutten model for speech segregation*. Ds Actes *Intl. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'03)*, p. 1089–1094, Nara, Japon, avr. 2003.
 - [32] A. Blin, S. Araki et S. Makino: *A sparseness-mixing matrix estimation (SMME) solving the underdetermined BSS for convolutive mixtures*. Ds Actes *Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'04)*, vol. 4, p. 85–88, Montréal, Canada, mai 2004.
 - [33] P. Bofill: *Underdetermined blind separation of delayed sound sources in the frequency domain*. Neurocomputing, 55(3-4) :627–641, oct. 2003.
 - [34] P. Bofill et E. Monte: *Underdetermined Convolved Source Reconstruction Using LP and SOCP, and a Neural Approximator of the Optimizer*. Ds Actes *Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'06)*, p. 569–576, Charleston, Caroline du Sud, mars 2006.

-
- [35] P. Bofill et M. Zibulevsky: *Blind separation of more sources than mixtures using sparsity of their short-time Fourier transform*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'00), p. 87–92, Helsinki, Finlande, juin 2000.
 - [36] P. Bofill et M. Zibulevsky: *Underdetermined blind source separation using sparse representations*. Signal Processing, 81(11) :2353–2362, nov. 2001.
 - [37] H. Bousbia-Salah, A. Belouchrani et K. Abed-Meraim: *Jacobi-like algorithm for blind signal separation of convolutive mixtures*. Electronic Letters, 37 :1049–1050, août 2001.
 - [38] H. Buchner, R. Aichner et W. Kellermann: *A Generalization of Blind Source Separation Algorithms for Convolutive Mixtures Based on Second-Order Statistics*. IEEE Trans. on Speech and Audio Processing, 13(1), janv. 2005.
 - [39] X. R. Cao et R. W. Liu: *General approach to blind source separation*. IEEE Trans. on Signal Processing, 44(3) :562–571, mars 1996.
 - [40] V. Capdevielle, C. Servière et J. L. Lacoume: *Blind separation of wide-band sources in the frequency domain*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'95), vol. 3, p. 2080–2083, Détroit, Michigan, mai 1995.
 - [41] J. F. Cardoso: *Source separation using higher order moments*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'89), vol. 4, p. 2109–2112, Glasgow, Royaume-Uni, mai 1989.
 - [42] J. F. Cardoso: *Infomax and Maximum Likelihood for Blind Source Separation*. IEEE Signal Processing Letters, 4(4) :112–114, avr. 1997.
 - [43] J. F. Cardoso: *Blind signal separation : statistical principles*. Proceedings of the IEEE, 9(10) :2009–2025, oct. 1998.
 - [44] J. F. Cardoso et A. Souloumiac: *Blind beamforming for non-Gaussian signals*. IEE Proceedings F Radar and Signal Processing, 140(6) :362–370, déc. 1993.
 - [45] N. Charkani: *Séparation auto-adaptative de sources pour des mélanges convolutifs. Application à la téléphonie mains libres dans les voitures*. Thèse de doctorat, Institut National Polytechnique de Grenoble, France, nov. 1996.
 - [46] N. Charkani et Y. Deville: *Self-adaptive separation of convolutively mixed signals with a recursive structure. Part I : Stability analysis and optimization of asymptotic behaviour*. Signal Processing, 73(3) :225–254, janv. 1999.
 - [47] N. Charkani et 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.

-
- [48] S. S. Chen, D. L. Donoho et M. A. Saunders: *Atomic decomposition by basis pursuit*. SIAM J Sci Comput, 20(1) :33–61, 1998.
 - [49] C. Choi: *Real-time binaural blind source separation*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'03), p. 567–572, Nara, Japon, avr. 2003.
 - [50] S. Choi et A. Cichocki: *Adaptive Blind Separation of Speech Signals : Cocktail Party Problem*. Ds Actes Intl. Conf. on Speech Processing (ICSP'97), p. 617–622, Séoul, Corée du Sud, août 1997.
 - [51] S. Choi et A. Cichocki: *A Hybrid Learning Approach to Blind Deconvolution of Linear MIMO Systems*. Electronic letters, 35(7) :1429–1430, août 1999.
 - [52] A. Cichocki et S. Amari: *Adaptive Blind Signal and Image Processing*. John Wiley & Sons, 2002.
 - [53] R. R. Coifman et M. V. Wickerhauser: *Entropy-Based Algorithms for Best Basis Selection*. IEEE Trans. on Information Theory, 38(2) :713–718, mars 1992.
 - [54] P. Comon: *Separation of sources using higher-order cumulants*. Ds Actes Advanced Algorithms and Architectures for Signal Processing IV, vol. 1152, p. 170–181, San Diego, Californie, août 1989.
 - [55] P. Comon: *Independent component analysis*. Ds Proc. Int. Sig. Proc. Workshop on Higher-Order Statistics (HOS'91), p. 111–120, Chamrousse, France, juil. 1991.
 - [56] P. Comon: *Blind Identification in Presence of Noise*. Ds Actes European Signal Processing Conference (EUSIPCO'92), p. 835–838, Bruxelles, Belgique, août 1992.
 - [57] P. Comon: *Independent component analysis, A new concept?* Signal Processing, 36(3) :287–314, avr. 1994.
 - [58] P. Comon: *Blind Identification and Source Separation in 2×3 Underdetermined Mixtures*. IEEE Trans. on Signal Processing, 52(1) :11–22, janv. 2004.
 - [59] P. Comon et E. Moreau: *Improved contrast dedicated to blind separation in communications*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'97), p. 3453–3456, Munich, Allemagne, avr. 1997.
 - [60] I. Daubechies: *Time-Frequency Localization Operators : A Geometric Phase Space Approach*. IEEE Trans. on Information Theory, 34(4) :605–612, juil. 1988.
 - [61] N. Delfosse et P. Loubaton: *Adaptive blind separation of independent sources : A deflation approach*. Signal Processing, 45(1) :59–83, juil. 1995.
 - [62] S. Deligne et R. Gopinath: *An EM algorithm for convolutive independent component analysis*. Neurocomputing, 49(1) :187–211, déc. 2002.

-
- [63] Y. Deville: *Sons de basse et de piano*. <http://www.ast.obs-mip.fr/deville>.
 - [64] Y. Deville: *Towards industrial applications of blind source separation and independent component analysis*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'99), p. 19–24, Aussois, France, janv. 1999.
 - [65] Y. Deville: *Panorama des applications biomédicales des méthodes de séparation aveugle de sources*. Ds Actes Groupe de Recherche et d'Etude du Traitement du Signal et des Images (GRETSI'03), vol. 1, p. 31–34, Paris, France, sept. 2003.
 - [66] Y. Deville: *Temporal and time-frequency correlation-based blind source separation methods*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'03), p. 1059–1064, Nara, Japon, avr. 2003.
 - [67] Y. Deville, M. Benali et F. Abrard: *Differential source separation for underdetermined instantaneous or convolutive mixtures : concept and algorithms*. Signal Processing, 84(10) :1759–1776, oct. 2004.
 - [68] Y. Deville, D. Bissessur, M. Puigt, S. Hosseini et H. Carfantan: *A time-scale correlation-based blind separation method applicable to correlated sources*. Ds Actes European Symp. on Artificial Neural Networks (ESANN'06), p. 337–344, Bruges, Belgique, avr. 2006.
 - [69] Y. Deville, J. Chappuis, S. Hosseini et J. Thomas: *Differential fast fixed-point BSS for underdetermined linear instantaneous mixtures*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'06), p. 48–56, Charleston, Caroline du Sud, mars 2006.
 - [70] Y. Deville et S. Hosseini: *Blind identification and separation methods for linear-quadratic mixtures and/or linearly independent non-stationary signals*. Ds Actes Intl. Symp. on Signal Processing and its Applications (ISSPA'07), Sharjah, Émirats arabes unis, fév. 2007.
 - [71] Y. Deville et M. Puigt: *Temporal and time-frequency correlation-based blind source separation methods. Part I : Determined and underdetermined linear instantaneous mixtures*. Signal Processing, 87(3) :374–407, mars 2007.
 - [72] D. Donoho: *On Minimum Entropy Deconvolution*. Ds Actes Applied Time series Analysis II, p. 556–608, Tulsa, Oklahoma, mars 1980.
 - [73] S. Douglas, H. Sawada et S. Makino: *A spatio-temporal FastICA algorithm for separating convolutive mixtures*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'05), vol. 5, p. 165–168, Philadelphie, Pennsylvanie, mars 2005.
 - [74] S. C. Douglas, M. Guptaand, H. Sawada et S. Makino: *Spatio-Temporal FastICA Algorithms for the Blind Separation of Convolutive Mixtures*. IEEE Trans. on Audio, Speech, and Language Processing, 15(5) :1511 – 1520, juil. 2007.

-
- [75] D. Erdogmus, K. E. Hild, Y. N. Rao et J. C. Principe: *Minimax Mutual Information Approach for Independent Component Analysis*. Neural Computation, 16(6) :1235–1252, juin 2004.
 - [76] J. Eriksson et V. Koivunen: *Blind Separation of a Class of Nonlinear ICA Models*. Ds Actes Intl. Symp. on Circuits and Systems (ISCAS'05), p. 5890–5893, Kobe, Japon, mai 2005.
 - [77] A. Ferréol, L. Albera et P. Chevalier: *Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM)*. IEEE Trans. on Signal Processing, 53(5) :1640–1653, mai 2005.
 - [78] C. Févotte: *Approche temps-fréquence pour la séparation aveugle de sources non-stationnaires*. Thèse de doctorat, Ecole Centrale de Nantes et Université de Nantes, France, pages 80-82, oct. 2003. <http://www.tsi.enst.fr/~cfevotte/>.
 - [79] M. Gaeta et J. L. Lacoume: *Source separation without prior knowledge : The maximum likelihood solution*. Ds Actes European Signal Processing Conference (EUSIPCO'90), p. 621–624, Barcelone, Espagne, sept. 1990.
 - [80] B. Gardner et K. Martin: *Head related transfer functions of a dummy head*, 1999. <http://sound.media.mit.edu/ica-bench>.
 - [81] B. Gardner et K. Martin: *HRTF Measurements of a KEMAR Dummy-Head Microphone*, 2000. <http://sound.media.mit.edu/KEMAR.html>.
 - [82] G. H. Golub et C. F. Van Loan: *Matrix Computations*. The John Hopkins University Press, 1996.
 - [83] A. Gorokhov et P. Loubaton: *Subspace-Based Techniques for Blind Separation of Convolutional Mixtures with Temporally Correlated Sources*. IEEE Trans. on Circuits and Systems I : Fundamental Theory and Applications, 44(9) :813–820, sept. 1997.
 - [84] R. Gribonval et S. Lesage: *A survey of Sparse Component Analysis for Blind Source Separation : principles, perspectives, and new challenges*. Ds Actes Europ. Symp. on Artificial Neural Networks (ESANN'06), p. 323–330, Bruges, Belgique, avr. 2006.
 - [85] R. Guidara, S. Hosseini et Y. Deville: *Blind separation of non-stationary images using Markov models*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'07), Londres, Royaume-Uni, sept. 2007.
 - [86] R. Guidara, S. Hosseini et Y. Deville: *Markovian blind separation of non-stationary temporally correlated sources*. Ds Actes Europ. Symp. on Artificial Neural Networks (ESANN'07), Bruges, Belgique, avr. 2007.
 - [87] Hugo Gävert, Jarmo Hurri, Jaakko Särelä et Aapo Hyvärinen: *FastICA for Matlab 7.x and 6.x, Version 2.5*, oct. 2005. <http://www.cis.hut.fi/projects/ica/fastica/code/dlcode.shtml>.

-
- [88] J. Herault et B. Ans: *Réseaux de neurones à synapses modifiables : décodage de messages sensoriels composites par un apprentissage non supervisé et permanent*. C. R. Académie des Sciences, Paris, série III, p. 525–528, 1984.
 - [89] J. Herault et C. Jutten: *Space or time adaptive signal processing by neural network models*. Ds Actes Intl. Conf. on Neural Networks for Computing, p. 206–211, Snowbird, Utah, avr. 1986.
 - [90] J. Herault, C. Jutten et B. Ans: *Détection de grandeurs primitives dans un message composite pour une architecture de calcul neuromimétique en apprentissage non supervisé*. Ds Actes Groupe de Recherche et d'Etude du Traitement du Signal et des Images (GRETSI'85), p. 1017–1022, Nice, France, mai 1985.
 - [91] K. E. Hild, D. Erdogmus et J. C. Principe: *Blind Source Separation Using Renyi's Mutual Information*. IEEE Signal Processing Letters, 8(6) :174–176, juin 2001.
 - [92] K. E. Hild, D. Erdogmus et J. C. Principe: *On-line minimum mutual information method for time-varying blind source separation*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'01), p. 126–131, San Diego, Californie, déc. 2001.
 - [93] K. E. Hild, D. Pinto, D. Erdogmus et J. C. Principe: *Convolutive Blind Source Separation by Minimizing Mutual Information Between Segments of Signals*. IEEE Trans. on Circuits and Systems - I, 52(10) :2188–2196, oct. 2005.
 - [94] M. V. Hulle: *Clustering approach to square and non-square blind source separation*. Ds Actes Intl. Work. on Neural Networks for Signal Processing (NNSP'99), p. 315–323, Madison, Wisconsin, août 1999.
 - [95] A. Hyvärinen: *New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit*. Ds Actes Conf. on Advances in Neural Information Processing Systems (NIPS'97), vol. 10, p. 273–279, Denver, Colorado, déc. 1997.
 - [96] A. Hyvärinen: *Fast and Robust Fixed-Point Algorithms for Independent Component Analysis*. IEEE Trans. on Neural Networks, 10(3) :626–634, mai 1999.
 - [97] A. Hyvärinen, J. Karhunen et E. Oja: *Independent Component Analysis*. John Wiley & Sons, 2001.
 - [98] A. Hyvärinen et E. Oja: *A Fast Fixed-Point Algorithm for Independent Component Analysis*. Neural Computation, 9(7) :1483–1492, oct. 1997.
 - [99] S. Ikeda et N. Murata: *A method of blind separation based on temporal structure of signals*. Ds Actes Intl. Conf. on Neural Information Processing (ICONIP'98), p. 737–742, Kitakyushu, Japon, oct. 1998.
 - [100] Y. Inouye et K. Tanebe: *Super-Exponential Algorithms for Multichannel Blind Deconvolution*. IEEE Trans. on Signal Processing, 48(3) :881–888, mars 2000.

-
- [101] A. Jourjine, S. Rickard et O. Yilmaz: *Blind separation of disjoint orthogonal signals : Demixing N sources from 2 mixtures*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'00), vol. 5, p. 2985–2988, Istanbul, Turquie, juin 2000.
 - [102] C. Jutten et M. Babaie-Zadeh: *Source separation : principles, current advances and applications*. Ds Actes Work. on Advanced Control and Diagnosis (ACD'06), Nancy, France, nov. 2006.
 - [103] C. Jutten, P. Comon et al.: *Séparation de Sources*, vol. 2, chapitre 13 (*Applications des méthodes de séparation aveugle de sources* par Y. Deville). Collection *Traité IC2, Information – Commande – Communication*, Hermès - Lavoisier, 2007. www.lavoisier.fr/fr/livres/index.asp?texte=2746215180&select=isbn&from=Hermes.
 - [104] C. Jutten et J. Herault: *Blind separation of sources, part I : An adaptive algorithm based on neuromimetic architecture*. Signal Processing, 24(1) :1–10, juil. 1991.
 - [105] C. Jutten, H. L. Nguyen-Thi, E. Dijkstra, E. Vittoz et J. Caelen: *Blind separation of sources : An algorithm for separation of convolutive mixtures*. Ds Actes Intl. Signal Processing Work. on Higher Order Statistics (HOS'91), p. 273–276, Chamrousse, France, juil. 1991.
 - [106] A. Kachenoura, L. Albera et L. Senhadji: *Séparation aveugle de sources en ingénierie biomédicale*. ITBM–RBM, 28 :20–34, 2007.
 - [107] M. Kawamoto, A. K. Barros, A. Mansour, K. Matsuoka et N. Ohnishi: *Blind Separation for Convolutive Mixtures of Non-stationary Signals*. Ds Actes Intl. Conf. on Neural Information Processing (ICONIP'98), p. 743–746, Kitakyushu, Japon, oct. 1998.
 - [108] M. Kawamoto, K. Matsuoka et N. Ohnishi: *A method of blind separation for convolved non-stationary signals*. Neurocomputing, 22(1-3) :157–171, nov. 1998.
 - [109] L. C. Khor, W. L. Woo et S. S. Dlay: *Non-sparse approach to underdetermined blind signal estimation*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'05), vol. 5, p. 309–312, Philadelphie, Pennsylvanie, mars 2005.
 - [110] M. Knaak, S. Araki et S. Makino: *Geometrically constraint ICA for convolutive mixtures of sound*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'03), vol. 2, p. 725–728, Hong Kong, Chine, avr. 2003.
 - [111] I. Kopriva, Z. Vevcic et H. Szu: *An adaptive short-time frequency domain algorithm for blind separation of nonstationary convolved mixtures*. Ds Actes Intl. Joint Conf. on Neural Networks (IJCNN'01), p. 424–429, Washington DC, juil. 2001.
 - [112] J. L. Lacoume et P. Ruiz: *Sources identification : A Solution Based on the Cumulants*. Ds Actes Work. on Spectrum, Estimation and Modelling, p. 199–203, août 1988.
 - [113] D. D. Lee et H. S. Seung: *Learning the parts of objects by non-negative matrix factorization*. Nature, 401 :788–791, oct. 1999.

-
- [114] D. D. Lee et H. S. Seung: *Algorithms for Non-negative Matrix Factorization*. Ds Actes *Conf. on Advances in Neural Information Processing Systems (NIPS'00)*, vol. 13, p. 556–562, Denver, Colorado, 2000.
 - [115] T. W. Lee, A. J. Bell et R. H. Lambert: *Blind separation of delayed and convolved sources*. Ds Actes *Conf. on Advances in Neural Information Processing Systems (NIPS'96)*, vol. 9, p. 758–764, Denver, Colorado, déc. 1996.
 - [116] S. Lesage, S. Krstulovic et R. Gribonval: *Séparation de sources dans le cas sous-déterminé : comparaison de deux approches basées sur des décompositions parcimonieuses*. Ds Actes *Groupe de Recherche et d'Etude du Traitement du Signal et des Images (GRETSI'05)*, Louvain-la-Neuve, Belgique, sept. 2005.
 - [117] S. Lesage, S. Krstulovic et R. Gribonval: *Under-Determined Source Separation : Comparison of Two Approaches Based on Sparse Decompositions*. Ds Actes *Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'06)*, p. 633–640, Charleston, Caroline du Sud, mars 2006.
 - [118] Y. Li, S. I. Amari, A. Cichocki, D. W. C. Ho et S. Xie: *Underdetermined Blind Source Separation Based on Sparse Representation*. *IEEE Trans. on Signal Processing*, 54(2), fév. 2006.
 - [119] Y. Li, A. Cichocki et S. Amari: *Sparse component analysis for blind source separation with less sensors than sources*. Ds Actes *Intl. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'03)*, p. 89–94, Nara, Japon, avr. 2003.
 - [120] Y. Li, A. Cichocki et S. Amari: *Analysis of sparse representation and blind source separation*. *Neural Computation*, 16(6) :1193–1234, 2004.
 - [121] J. K. Lin, D. G. Grier et J. D. Cowan: *Feature Extraction Approach to Blind Source Separation*. Ds Actes *Intl. Work. on Neural Networks for Signal Processing (NNSP'97)*, p. 398–405, Amelia Island, Floride, sept. 1997.
 - [122] P. Loubaton et P. Regalia: *Blind deconvolution of multivariate signals : a deflation approach*. Ds Actes *Intl. Conf. on Communications (ICC'93)*, vol. 2, p. 1160–1164, Genève, Suisse, mai 1993.
 - [123] D. G. Luenberger: *Optimization by Vector Space Methods*. John Wiley & Sons, New York, 1969.
 - [124] Z. Q. Luo et J. Lu: *On blind source separation using mutual information criterion*. *Mathematical programming*, 97(3) :587–603, août 2003.
 - [125] S. Makino, H. Sawada, R. Mukai et S. Araki: *Blind source separation of convolutive mixtures of speech in frequency domain*. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E88-A(7) :1640–1655, juil. 2005.
 - [126] S. Mallat et Z. Zhang: *Matching Pursuits With Time-Frequency Dictionaries*. *IEEE Trans. on Signal Processing*, 41(12) :3397–3415, 1993.
 - [127] A. Mansour, C. Jutten et P. Loubaton: *Adaptive Subspace Algorithm for Blind Separation of Independent Sources in Convolutional Mixture*. *IEEE Trans. on Signal Processing*, 48(2) :583–586, Febr. 2000.
 - [128] A. S. Master: *Sound source separation of N sources from stereo signals via fitting to N models each lacking one source*. rapport technique, Center

-
- for Computer Research in Music and Acoustics, Stanford University, 2003.
<http://www-ccrma.stanford.edu/asmaster/>.
- [129] A. S. Master: *Bayesian two source modeling for separation of N sources from stereo signals*. Ds Actes Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP'04), vol. 4, p. 281–284, Montréal, Canada, mai 2004.
 - [130] K. Matsuoka et M. Kawamoto: *Blind signal separation based on mutual information criterion*. Ds Actes Intl. Symp. on Nonlinear Theory and its Applications (NOLTA'95), p. 85–90, Las Vegas, Nevada, déc. 1995.
 - [131] K. Matsuoka, M. Ohya et M. Kawamoto: *A Neural Net for Blind Separation of Nonstationary Signals*. Neural Networks, 8 :411–419, 1995.
 - [132] J. McQueen: *Some methods for classification and analysis of multivariate observations*. Ds Actes Berkeley Symp. on Mathematical Statistics and Probability, vol. 1, p. 281–297, 1967.
 - [133] T. Melia et S. Rickard: *Extending the DUET Blind Source Separation technique*. Ds Actes Structure et Parcimonie pour la Représentation Adaptative de Signaux (SPARS'05), Rennes, France, nov. 2005.
 - [134] T. Melia et S. Rickard: *Underdetermined blind source separation in echoic environments using DESPRIT*. EURASIP Journal on Advances in Signal Processing, 2007 :Article ID 86484, 19 pages, 2007.
 - [135] N. Mitianoudis et M. E. Davies: *Audio Source Separation of Convolutional Mixtures*. IEEE Trans. on Speech and Audio Processing, 11(5) :489–497, sept. 2003.
 - [136] E. Moreau et E. Macchi: *New self-adaptive algorithms for source separation based on contrast functions*. Ds Actes Work. on Higher-Order Statistics (HOS'93), p. 215–219, Lake Tahoe, USA, juin 1993.
 - [137] E. Moulines, J. F. Cardoso et E. Gassiat: *Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'97), vol. 5, p. 3617–3620, Munich, Allemagne, avr. 1997.
 - [138] M. Rapacioli, C. Joblin et P. Boissel: *Spectroscopy of polycyclic aromatic hydrocarbons and very small grains in photodissociation regions*. Astronomy & Astrophysics, 429 :193–204, 2005.
 - [139] H. L. Nguyen-Thi et C. Jutten: *Blind source separation for convolutional mixtures*. Signal Processing, 45(2) :209–229, août 1995.
 - [140] D. Nuzillard et A. Bijaoui: *Blind source separation and analysis of multispectral astronomical images*. Astronomy & Astrophysics, 147 :129–138, nov. 2006.
 - [141] M. Ohata, K. Matsuoka et T. Mukai: *An adaptive blind separation method using para-Hermitian whitening filter for convolutionally mixed signals*. Signal Processing, 87(1) :33–50, janv. 2007.
 - [142] E. Oja: *Convergence of the symmetrical FastICA algorithm*. Ds Actes Intl. Conf. on Neural Information Processing (ICONIP'02), vol. 3, p. 1368–1372, Singapour, nov. 2002.
 - [143] E. Oja et Z. Yuan: *The FastICA Algorithm Revisited : Convergence Analysis*. IEEE Trans. on Neural Networks, 17(6) :1370–1381, nov. 2006.

-
- [144] R. K. Olsson et L. K. Hansen: *Blind separation of more sources than sensors in convolutive mixtures*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'06), vol. 5, p. 657–660, Toulouse, France, mai 2006.
 - [145] P. Paatero et U. Tapper: *Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values*. Environmetrics, 5 :111–126, 1994.
 - [146] L. Parra et C. Alvino: *Geometric Source Separation : Merging Convolutive Source Separation With Geometric Beamforming*. IEEE Trans. on Speech Audio Processing, 10(6) :352–362, sept. 2002.
 - [147] L. Parra et C. Spence: *Convolutive Blind Separation of Non-Stationary Sources*. IEEE Trans. on Speech Audio Processing, 8(3) :320–327, mai 2000.
 - [148] L. C. Parra, C. Spence et B. Vries: *Convolutive Source Separation and Signal Modeling with ML*. Ds Actes Intl. Symp. on Intelligent Systems (ISIS'97), Reggio de Calabre, Italie, sept. 1997.
 - [149] B. A. Pearlmutter et L. C. Parra: *Maximum Likelihood Blind Source Separation : A Context-Sensitive Generalization of ICA*. Ds Actes Conf. on Advances in Neural Information Processing Systems (NIPS'96), p. 613–619, Denver, Colorado, déc. 1996.
 - [150] M. S. Pedersen, J. Larsen, U. Kjems et L. C. Parra: *A survey of Convolutive Blind Source Separation Methods*. Eds. Jacob Benesty and Arden Huang, Springer, 2007. to appear as chapter in Multichannel Speech Processing Handbook.
 - [151] M. S. Pedersen et C. M. Nielsen: *Gradient flow convolutive blind source separation*. Ds Actes Intl. Work. on Machine Learning for Signal Processing (MLSP'04), p. 335–344, Saint Louis, Brésil, sept. 2004.
 - [152] J. M. Peterson et S. Kadambe: *A probabilistic approach for blind source separation of underdetermined convolutive mixtures*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'03), vol. 6, p. 581–584, Hong Kong, Chine, avr. 2003.
 - [153] D. T. Pham: *Mutual information approach to blind separation of stationary sources*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'99), p. 215–220, Aussois, France, janv. 1999.
 - [154] D. T. Pham: *Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion*. Signal Processing, 81(4) :855–870, avr. 2001.
 - [155] D. T. Pham: *Blind separation of non stationary non Gaussian sources*. Ds Actes European Signal Processing Conference (EUSIPCO'02), Toulouse, France, sept. 2002.
 - [156] D. T. Pham: *Exploiting source non stationarity and coloration in Blind Source Separation*. Ds Actes Intl. Conf. on Digital Signal Processing (DSP'02), vol. 1, p. 151–154, Santorin, Grèce, juil. 2002.
 - [157] D. T. Pham: *Mutual Information Approach to Blind Separation of Stationary Sources*. IEEE Trans. on Information Theory, 48(7) :1935–1946, juil. 2002.

-
- [158] D. T. Pham: *Mutual information approach to blind separation-deconvolution*. Ds Actes *European Signal Processing Conference (EUSIPCO'05)*, Antalya, Turquie, sept. 2005.
 - [159] D. T. Pham et J. F. Cardoso: *Blind Separation of Instantaneous Mixture of Non Stationary Sources*. Ds Actes *Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'00)*, p. 187–192, Helsinki, Finlande, juin 2000.
 - [160] D. T. Pham et J. F. Cardoso: *Blind Separation of Instantaneous Mixtures of Nonstationary Sources*. *IEEE Trans. on Signal Processing*, 49(9) :1837 – 1848, sept. 2001.
 - [161] D. T. Pham et P. Garat: *Blind Separation of Mixtures of Independent Sources Through a Quasi-Maximum Likelihood Approach*. *IEEE Trans. on Signal Processing*, 45(7) :1712–1725, juil. 1997.
 - [162] D. T. Pham, P. Garat et C. Jutten: *Separation of mixture of independent sources through a maximum likelihood approach*. Ds *Signal Processing VI, Proc. of the 6th European Signal Processing Conference (EUSIPCO'92)*, p. 771–774, Bruxelles, Belgique, août 1992.
 - [163] D. T. Pham, C. Servière et H. Boumaraf: *Blind separation of convolutive audio mixtures using nonstationarity*. Ds Actes *Intl. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'03)*, p. 975–980, Nara, Japon, avr. 2003.
 - [164] D. T. Pham, C. Servière et H. Boumaraf: *Blind separation of speech mixtures based on nonstationarity*. Ds Actes *Intl. Symp. on Signal Processing and its Applications (ISSPA'03)*, vol. 2, p. 73–76, Paris, France, juil. 2003.
 - [165] R. Prasad, H. Saruwatari, A. Lee et K. Shikano: *A fixed-point ICA algorithm for convoluted speech signal separation*. Ds Actes *Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'03)*, p. 579–584, Nara, Japon, avr. 2003.
 - [166] M. Puigt et Y. Deville: *Time-frequency ratio-based blind separation methods for attenuated and time-delayed sources*. *Mechanical Systems and Signal Processing*, 19(6) :1348–1379, nov. 2005.
 - [167] M. Puigt et Y. Deville: *A time-frequency correlation-based blind source separation method for time-delayed mixtures*. Ds Actes *Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'06)*, vol. 5, p. 853–856, Toulouse, France, mai 2006.
 - [168] M. Puigt et Y. Deville: *Two Time-Frequency Ratio-Based Blind Source Separation Methods for Time-Delayed Mixtures*. Ds Actes *Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'06)*, p. 682–690, Charleston, Caroline du Sud, mars 2006.
 - [169] M. Puigt et Y. Deville: *A new time-frequency correlation-based source separation method for attenuated and time shifted mixtures*. Ds Actes *Intl. Work. on Electronics, Control, Modelling, Measurement and Signals (ECMS'07)*, Liberec, République tchèque, mai 2007.

-
- [170] K. Rahbar et J. Reilly: *Blind source separation algorithm for MIMO convolutive mixtures*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'01), p. 242–247, San Diego, Californie, déc. 2001.
 - [171] K. Rahbar et J. P. Reilly: *A Frequency Domain Method for Blind Source Separation of Convolutional Audio Mixtures*. IEEE Trans. on Speech and Audio Processing, 13(5), sept. 2005.
 - [172] K. Rahbar, J. P. Reilly et 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), fév. 2004.
 - [173] J. P. Reilly et L. C. Mendoza: *Blind signal separation for convolutional mixing environments using spatial-temporal processing*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'99), vol. 3, p. 1437–1440, Phoenix, Arizona, mai 1999.
 - [174] M. E. Rhabi, G. Gelle, H. Fenniri et G. Delaunay: *A penalized mutual information criterion for blind separation of convolutional mixtures*. Signal Processing, 84(10) :1979–1984, oct. 2004.
 - [175] S. Rickard, R. Balan et J. Rosca: *Real-time time-frequency based blind source separation*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'01), p. 651–656, San Diego, Californie, déc. 2001.
 - [176] S. Rickard et M. Fallon: *The Gini index of speech*. Ds Actes Annual Conf. on Information Sciences and Systems (CISS'04), Princeton, New Jersey, March 2004.
 - [177] S. Rickard et O. Yilmaz: *On the approximate W-disjoint orthogonality of speech*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'02), vol. 1, p. 529–532, Orlando, Floride, mai 2002.
 - [178] J. Rosca, C. Borss et R. Balan: *Generalized sparse signal mixing model and application to noisy blind source separation*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'04), vol. 3, p. 877–880, Montréal, Canada, mai 2004.
 - [179] R. Saab, O. Yilmaz, M. J. McKeown et R. Abugharbieh: *Underdetermined sparse blind source separation with delays*. Ds Actes Structure et Parcimonie pour la Représentation Adaptative de Signaux (SPARS'05), Rennes, France, nov. 2005.
 - [180] R. Saab, O. Yilmaz, M. J. McKeown et R. Abugharbieh: *Underdetermined anechoic blind source separation via ℓ^q -basis-pursuit with $q < 1$* . IEEE Trans. on Signal Processing, to appear.
 - [181] S. Samadi, M. Babaie-Zadeh, C. Jutten et K. Nayebi: *Blind Source Separation by Adaptive Estimation of Score Function Difference*. Ds Actes Intl. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'04), Grenade, Espagne, sept. 2004.
 - [182] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee et K. Shikano: *Blind Source Separation Based on a Fast-Convergence Algorithm Combining ICA and Beamforming*. IEEE Trans. on Audio, Speech, and Language Processing, 14(2) :666–678, mars 2006.

-
- [183] H. Sawada, R. Mukai, S. Araki et S. Makino: *A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation*. IEEE Trans. on Speech and Audio Processing, 12(5) :530–538, sept. 2004.
 - [184] H. Saylani, Y. Deville, S. Hosseini et M. Habibi: *A multi-tag radio-frequency identification system using a new blind source separation method based on spectral decorrelation*. Ds Actes Intl. Symp. on Communications, Control and Signal Processing (ISCCSP'06), Marrakech, Maroc, mars 2006.
 - [185] C. Servière: *Blind source separation of convolutive mixtures*. Ds Actes Work. on Statistical Signal and Array Processing (SSAP'96), p. 316–319, Corfou, Grèce, juin 1996.
 - [186] C. Servière: *Separation of speech signals under reverberant conditions*. Ds Actes European Signal Processing Conference (EUSIPCO'04), p. 1693–1696, Vienne, Autriche, sept. 2004.
 - [187] O. Shalvi et E. Weinstein: *New Criteria for Blind Deconvolution of Non-minimum Phase Systems (Channels)*. IEEE Trans. on Information Theory, 36(2) :312–321, mars 1990.
 - [188] O. Shalvi et E. Weinstein: *Super-Exponential Methods for Blind Deconvolution*. IEEE Trans. on Information Theory, 39(2) :504–519, mars 1993.
 - [189] C. Simon: *Séparation aveugle de sources en mélanges convolutifs*. Thèse de doctorat, Université de Marne-la-Vallée, France, nov. 1999.
 - [190] C. Simon, P. Loubaton, C. Vignat, C. Jutten et G. d'Urso: *Blind Source Separation of convolutive mixtures by maximization of fourth-order cumulants : the non-i.i.d. case*. Ds Actes Asilomar Conf. on Signals, Systems, and Computers (ASILOMAR'98), p. 1584–1588, Monterey, Californie, nov. 1998.
 - [191] C. Simon, P. Loubaton, C. Vignat, C. Jutten et G. d'Urso: *Separation of a class of convolutive mixtures : a contrast function approach*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'99), vol. 3, p. 1429–1432, Phoenix, Arizona, mars 1999.
 - [192] P. Smaragdis: *Blind separation of convolved mixtures in the frequency domain*. Neurocomputing, 22(1-3) :21–34, nov. 1998.
 - [193] P. Smaragdis: *Non-negative matrix factor deconvolution ; extraction of multiple sound sources from monophonic inputs*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'04), p. 494–499, Grenade, Espagne, 2004.
 - [194] P. Smaragdis: *Convolutional Speech Bases and Their Application to Supervised Speech Separation*. IEEE Trans. on Audio, Speech, and Language Processing, 15(1) :1–12, janv. 2007.
 - [195] D. Smith, J. Lukasiak et I. Burnett: *A two channel, block-adaptive audio separation technique based upon time-frequency information*. Ds Actes European Signal Processing Conference (EUSIPCO'04), p. 393–396, Vienne, Autriche, sept. 2004.
 - [196] D. Smith, J. Lukasiak et I. Burnett: *Two Channel, Block Adaptive Audio Separation Using the Cross Correlation of Time Frequency Information*. Ds

- Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'04), p. 889–897, Grenade, Espagne, sept. 2004.
- [197] A. Taleb et C. Jutten: *On underdetermined source separation*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'99), vol. 3, p. 1445–1448, Phoenix, Arizona, mars 1999.
 - [198] A. R. Teixeira, A. M. Tomé, E. W. Lang et K. Stadlthanner: *Delayed AMUSE – A Tool for Blind Source Separation and Denoising*. Ds Actes Intl. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'04), Grenade, Espagne, sept. 2004.
 - [199] F. J. Theis, E. W. Lang et C. G. Puntonet: *A geometric algorithm for over-complete linear ICA*. Neurocomputing, 56 :381–398, janv. 2004.
 - [200] F. J. Theis, C. G. Puntonet et E. W. Lang: *Median-based clustering for under-determined blind signal processing*. IEEE Signal Processing Letters, 13(2) :96–99, fév. 2006.
 - [201] J. Thomas: *L'algorithme C-FICA*, 2007. <http://www.ast.obs-mip.fr/article569.html>.
 - [202] J. Thomas, Y. Deville et S. Hosseini: *Fixed-point algorithms for convolutive blind source separation based on non-gaussianity maximization*. Ds Actes Intl. Work. on Electronics, Control, Modelling, Measurement and Signals (ECMS'05), Toulouse, France, mai 2005.
 - [203] J. Thomas, Y. Deville et S. Hosseini: *Time-domain fast fixed point algorithms for convolutive ICA*. IEEE Signal Processing Letters, 13(4) :228–231, avr. 2006.
 - [204] J. Thomas, Y. Deville et S. Hosseini: *Differential Fast Fixed-point Algorithms for Underdetermined Instantaneous and Convolutive Partial Blind Source Separation*. IEEE Trans. on Signal Processing, 55(7) :3717–3729, juil. 2007.
 - [205] J. Thomas, Y. Deville et S. Hosseini: *Multivariate polynomial identification for blind image separation*. Ds Actes Intl. Work. on Electronics, Control, Modelling, Measurement and Signals (ECMS'07), p. 97–102, Liberec, République tchèque, mai 2007.
 - [206] J. Thomas, Y. Deville et S. Hosseini: *Un algorithme rapide à identification polynômiale multi-variable pour la séparation aveugle d'images*. Ds Actes Groupe de Recherche et d'Etude du Traitement du Signal et des Images (GRETSI'07), Troyes, France, sept. 2007.
 - [207] L. Tong, R. W. Liu, V. C. Soon et Y. F. Huang: *Indeterminacy and Identifiability of Blind Identification*. IEEE Trans. on Circuits and Systems, 38(5) :499–509, mai 1991.
 - [208] L. Tong, V. C. Soon, R. Liu et Y. Huang: *AMUSE : A New Blind Identification Algorithm*. Ds Actes Intl. Symp. on Circuits and Systems (ISCAS'90), vol. 3, p. 1784–1787, Nouvelle-Orléans, Louisiane, mai 1990.
 - [209] C. C. Took, S. Sanei, J. Chambers et S. Dunne: *Underdetermined Blind Source Separation of Temporomandibular Joint Sounds*. IEEE Trans. on Biomedical Engineering, 53(10) :2123–2126, oct. 2006.

-
- [210] K. Torkkola: *Blind separation of convolved sources based on information maximization*. Ds Actes Intl. Work. on Neural Networks for Signal Processing (NNSP'96), p. 423–432, Kyoto, Japon, sept. 1996.
 - [211] K. Torkkola: *Blind separation of delayed sources based on information maximization*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'96), vol. 6, p. 3509–3512, Atlanta, Géorgie, mai 1996.
 - [212] J. K. Tugnait: *Identification and Deconvolution of Multichannel Linear Non-Gaussian Processes Using Higher Order Statistics and Inverse Filter Criteria*. IEEE Trans. on Signal Processing, 45(3) :658–672, mars 1997.
 - [213] J. K. Tugnait: *On Blind Separation of Convolutional Mixtures of Independent Linear Signals in Unknown Additive Noise*. IEEE Trans. on Signal Processing, 46(11) :3117–3123, nov. 1998.
 - [214] J. K. Tugnait: *Adaptive blind separation of convolutional mixtures of independent linear signals*. Signal Processing, 73(11) :139–152, janv. 1999.
 - [215] S. V. Vaseghi: *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons, 2000.
 - [216] L. Vielva, D. Erdogmus, C. Pantaleon, I. Santamaria, J. Pereda et J. C. Principe: *Underdetermined blind source separation in a time-varying environment*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'02), vol. 3, p. 3049–3052, Orlando, Floride, mai 2002.
 - [217] L. Vielva, D. Erdogmus et J. C. Principe: *Underdetermined blind source separation using a probabilistic source sparsity model*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'00), p. 675–679, Helsinki, Finlande, 2000.
 - [218] E. Vincent et X. Rodet: *Underdetermined Source Separation with Structured Source Priors*. Ds Actes Intl. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'04), p. 327–332, Grenade, Espagne, sept. 2004.
 - [219] T. Virtanen: *Sound source separation using sparse coding with temporal continuity objective*. Ds Actes Intl. Computer Music Conf. (ICMC'03), p. 231–234, Singapour, oct. 2003.
 - [220] E. Weinstein, M. Feder et A. Oppenheim: *Multi-Channel Signal Separation by Decorrelation*. IEEE Trans. on Speech Audio Processing, 1(4) :405–413, oct. 1993.
 - [221] R. A. Wiggins: *On minimum entropy deconvolution*. Geop Exploration, 16 :21–35, janv. 1978.
 - [222] S. Winter, W. Kellermann, H. Sawada et S. Makino: *MAP-Based Underdetermined Blind Source Separation of Convolutional Mixtures by Hierarchical Clustering and ℓ^1 -Norm Minimization*. EURASIP Journal on Advances in Signal Processing, 2007 :Article ID 24717, 12 pages, 2007.
 - [223] A. N. Witt, K. D. Gordon, U. P. Vijh, P. H. Sell, T. L. Smith et R. H. Xie: *The Excitation of Extended Red Emission : New Constraints on Its Carrier from Hubble Space Telescope Observations of NGC 7023*. The Astrophysical Journal, 636 :303–315.

-
- [224] H. C. Wu et J. C. Principe: *Simultaneous Diagonalization in the Frequency Domain (SDIF) for Source Separation*. Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'99), p. 245–250, Aussois, France, janv. 1999.
 - [225] J. Xi et J. P. Reilly: *Blind separation and restoration of signals mixed in convolutive environment*. Ds Actes Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'97), vol. 2, p. 1327–1330, Munich, Allemagne, avr. 1997.
 - [226] M. Xiao, S. Xie et Y. Fu: *A Novel Approach for Underdetermined Blind Sources Separation in Frequency Domain*. Ds Actes Intl. Symp. on Neural Networks (ISNN'05), p. 484–489, Chongqing, Chine, mai 2005.
 - [227] M. Xiao, S. Xie et Y. Fu: *A statistically sparse decomposition principle for underdetermined blind source separation*. Ds Actes Intl. Symp. on Intelligent Signal Processing and Communication Systems (ISPACS'05), p. 165–168, Hong Kong, Chine, déc. 2005.
 - [228] R. Xu et D. Wunsch II: *Survey of Clustering Algorithms*. IEEE Transactions on Neural Networks, 16(3) :645–678, mai 2005.
 - [229] O. Yilmaz et S. Rickard: *Blind Separation of Speech Mixtures via Time-Frequency Masking*. IEEE Trans. on Signal Processing, 52(7) :1830–1847, juil. 2004.
 - [230] Y. Zhang, X. Shi et C. H. Chen: *A Gaussian mixture model for underdetermined independent component analysis*. Signal Processing, 86(7) :1538–1549, juil. 2006.
 - [231] M. Zibulevsky, P. Kisilev, Y. Y. Zeevi et B. A. Pearlmutter: *Blind source separation via multinode sparse representation*. Ds Actes Conf. on Advances in Neural Information Processing Systems (NIPS'01), vol. 14, p. 1049–1056, Vancouver, Canada, déc. 2001.
 - [232] M. Zibulevsky et Y. Y. Zeevi: *Extraction of a source from multichannel data using sparse decomposition*. Neurocomputing, 49(1-4) :163–173, déc. 2002.

Publications de l'auteur

Articles de revues

- **J. Thomas**, Y. Deville et S. Hosseini, *Time-domain fast fixed point algorithms for convolutive ICA*, IEEE Signal Processing Letters, vol. 13, no. 4, p. 228-231, avr. 2006.
- **J. Thomas**, Y. Deville et S. Hosseini, *Differential Fast Fixed-point Algorithms for Underdetermined Instantaneous and Convolutive Partial Blind Source Separation*, IEEE Transactions on Signal Processing, vol. 55, no. 7, p. 3717-3729, juil. 2007.
- O. Berné, C. Joblin, Y. Deville, J. D. Smith, M. Rapaccioli, J. P. Bernard, **J. Thomas**, W. Reach et A. Abergel, *Analysis of the emission of very small dust particles from Spitzer spectro-imagery data using Blind Signal Separation methods*, Astronomy & Astrophysics, vol. 469, no. 2, p. 575-586, juil. 2007.
- F. Abrard, Y. Deville et **J. Thomas**, *Blind partial separation of underdetermined convolutive mixtures of complex sources based on differential normalized kurtosis*, Neurocomputing, à paraître.
- O. Berné, C. Joblin, M. Rapacioli, **J. Thomas**, J.-C. Cuillandre, and Y. Deville, *Extended Red Emission and the evolution of carbonaceous nanograins in NGC 7023*, Astron. & Astrophysics, vol. 479, no. 3, pp. L41-44, Mar. 2008.

Actes de conférences

- **J. Thomas**, Y. Deville et S. Hosseini, *Fixed-point algorithms for convolutive blind source separation based on non-gaussianity maximization*, Ds Actes Intl. Work. on Electronics, Control, Modelling, Measurement and Signals (ECMS'05), Toulouse, France, mai 2005.
- Y. Deville, J. Chappuis, S. Hosseini et **J. Thomas**, *Differential fast fixed-point BSS for underdetermined linear instantaneous mixtures*, Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'06), p. 48-56, Charleston, Caroline du Sud, mars 2006.
- **J. Thomas**, Y. Deville et S. Hosseini, *Multivariate polynomial identification for blind image separation*, Ds Actes Intl. Work. on Electronics, Control, Modelling, Measurement and Signals (ECMS'07), p. 97-102, Liberec, République tchèque, mai 2007.
- **J. Thomas**, Y. Deville et S. Hosseini, *Un algorithme rapide à identification polynômiale multi-variable pour la séparation aveugle d'images*, Ds Actes Groupe de Recherche et d'Etude du Traitement du Signal et des Images (GRETSI'07), Troyes, France, sept. 2007.

AUTHOR : Johan Thomas

TITLE : *Time-domain fast fixed-point algorithms for the blind separation of convolutive and/or underdetermined mixtures*

THESIS SUPERVISORS : Yannick Deville and Shahram Hosseini

DEFENSE : the 12th of Decembre 2007 at the Observatoire Midi-Pyrénées of Toulouse, France.

ABSTRACT : The first part of this thesis deals with (over-)determined convolutive mixtures. We extend the FastICA algorithm to convolutive mixtures by developing a fixed-point algorithm operating in the time domain. We introduce a non-causal spatio-temporal whitening process, which, by initializing the extraction parameters in a particular way, allows us to use fixed-point optimization iterations. The estimation of the contributions in the observations is realized by means of a quadratic criterion optimized by a non-causal Wiener process.

In the second part concerning instantaneous and convolutive underdetermined mixtures, we extend the FastICA algorithm by using the differential separation concept. We propose differential observation whitening procedures which allow us to use fixed-point iterations to separate between them some sources of interest. In the convolutive case, the contributions are estimated thanks to a non-causal differential Wiener process.

The third part deals with instantaneous mixtures of sources containing a high number of samples. We first point out a limitation of the FastICA algorithm which, at each optimization iteration, computes some statistics on all available samples. As the kurtosis criterion is polynomial relatively to the extraction coefficients, we propose a procedure to identify this polynomial in order to perform the optimization in a more efficient computation space. This principle is then extended to underdetermined instantaneous mixtures.

KEYWORDS : Blind Source Separation (BSS), Independent Component Analysis (ICA), convolutive mixtures, underdetermined mixtures, long mixture recordings, FastICA.

AUTEUR : Johan Thomas

TITRE : *Algorithmes temporels rapides à point fixe pour la séparation aveugle de mélanges convolutifs et/ou sous-déterminés*

DIRECTEURS DE THÈSE : Yannick Deville et Shahram Hosseini

SOUTENANCE : le 12 décembre 2007 à l'Observatoire Midi - Pyrénées de Toulouse, France.

RÉSUMÉ : La première partie de cette thèse est consacrée aux mélanges convolutifs (sur-)déterminés. Nous cherchons à étendre l'algorithme FastICA aux mélanges convolutifs en proposant un algorithme à point fixe opérant dans le domaine temporel. Nous introduisons un processus de blanchiment spatio-temporel non-causal, qui, en initialisant les paramètres d'extraction d'une façon particulière, permet d'utiliser des itérations d'optimisation de type point fixe. L'estimation des contributions dans les observations est réalisée grâce à un critère quadratique optimisé par un filtre de Wiener non-causal.

Dans la deuxième partie, consacrée aux mélanges sous-déterminés instantanés et convolutifs, nous cherchons aussi à étendre l'algorithme FastICA en nous basant sur le concept de séparation différentielle. Nous proposons des procédures de blanchiment différentiel des observations qui permettent l'emploi d'itérations à point fixe pour séparer entre elles des sources dites utiles. En convolutif, les contributions sont estimées au moyen d'un filtre de Wiener différentiel non-causal.

La troisième partie, consacrée aux mélanges instantanés de sources contenant un grand nombre d'échantillons, met d'abord en évidence une limitation de l'algorithme FastICA, qui, à chaque itération d'optimisation, calcule des statistiques sur l'ensemble des échantillons disponibles. Le critère du kurtosis étant polynômial relativement aux coefficients d'extraction, nous proposons une procédure d'identification de ce polynôme qui permet d'effectuer l'optimisation dans un espace de calcul moins gourmand. Ce principe est ensuite appliqué aux mélanges instantanés sous-déterminés.

MOTS-CLÉS : Séparation Aveugle de Sources (SAS), Analyse en Composantes Indépendantes (ACI), mélanges convolutifs, mélanges sous-déterminés, longs enregistrements de mélanges, FastICA.

DISCIPLINE : Traitement du signal

Laboratoire d'Astrophysique de Toulouse - Tarbes
 Université de Toulouse – CNRS
 14, avenue Edouard Belin
 31400 Toulouse

