



**HAL**  
open science

# Séparation aveugle de sources audio dans le contexte automobile

Abdeldjalil Aissa El Bey

► **To cite this version:**

Abdeldjalil Aissa El Bey. Séparation aveugle de sources audio dans le contexte automobile. Traitement du signal et de l'image [eess.SP]. Télécom ParisTech, 2007. Français. NNT : . tel-00189685

**HAL Id: tel-00189685**

**<https://pastel.hal.science/tel-00189685>**

Submitted on 21 Nov 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Thèse

présentée pour obtenir le grade de docteur

de l'École nationale supérieure  
des télécommunications

Spécialité : Traitement du Signal et des Images

**Abdeldjalil AISSA-EL-BEY**

**Séparation aveugle de sources audio**

soutenue le 06 juillet 2007 devant le jury composé de

Pierre Comon

Rapporteurs

Christian Jutten

Yannick Deville

Examineurs

Gaël Richard

Karim Abed-Meraim

Directeurs de thèse

Yves Grenier

École nationale supérieure des télécommunications



# Table des matières

<b>Abréviations</b>	<b>xv</b>
<b>Notations</b>	<b>xvii</b>
<b>Introduction générale</b>	<b>1</b>
Présentation du problème . . . . .	1
Motivations . . . . .	2
Plan du document et contributions . . . . .	3
<b>1 La séparation de sources audio</b>	<b>5</b>
1.1 Bref historique . . . . .	5
1.2 Les principes de la séparation aveugle de sources . . . . .	7
1.2.1 Problème de la séparation de sources . . . . .	7
1.2.2 Analyse en composantes indépendantes . . . . .	9
1.2.3 Maximum de vraisemblance . . . . .	11
1.3 La séparation de sources audio . . . . .	13
1.3.1 Mélanges instantanés . . . . .	14
1.3.2 Mélanges convolutifs . . . . .	22
1.4 Conclusion . . . . .	24
<b>2 La séparation de sources audio sous-déterminée</b>	<b>27</b>
2.1 Introduction . . . . .	27
2.2 Mélanges instantanés . . . . .	28
2.2.1 Méthodes basées sur un modèle probabiliste . . . . .	28
2.2.2 Méthodes basées sur une représentation parcimonieuse . . . . .	30
2.3 Mélanges convolutifs . . . . .	34

2.4	Conclusion . . . . .	35
<b>I</b>	<b>La SAS audio utilisant la décomposition modale</b>	<b>37</b>
<b>3</b>	<b>La SAS audio utilisant la décomposition modale : mélanges instantanés</b>	<b>39</b>
3.1	Introduction . . . . .	39
3.2	Formalisation du problème et hypothèses . . . . .	40
3.3	L'algorithme MD-UBSS . . . . .	42
3.3.1	Estimation des composantes modales . . . . .	43
3.3.2	Classification et estimation des signaux sources . . . . .	45
3.3.3	Cas des composantes modales communes . . . . .	47
3.4	L'algorithme MD-UBSS modifié . . . . .	48
3.5	Discussion . . . . .	50
3.6	Simulations . . . . .	51
3.7	Conclusion . . . . .	57
<b>4</b>	<b>La SAS audio utilisant la décomposition modale : mélanges convolutifs</b>	<b>59</b>
4.1	Introduction . . . . .	59
4.2	Formalisation du problème et hypothèses . . . . .	60
4.3	L'algorithme MD-UBSS convolutif . . . . .	61
4.3.1	Estimation du canal de mélange . . . . .	62
4.3.2	Association des composantes modales et estimation des sources . . . . .	66
4.4	L'algorithme MD-BSS convolutif sur-déterminé . . . . .	67
4.4.1	Synthèse des signaux . . . . .	67
4.4.2	Critère de parcimonie . . . . .	69
4.5	Discussion . . . . .	71
4.6	Simulations . . . . .	71
4.7	Conclusion . . . . .	74
<b>II</b>	<b>La SAS audio utilisant la représentation temps-fréquence</b>	<b>77</b>
<b>5</b>	<b>La SAS audio utilisant la représentation temps-fréquence : mélanges instantanés</b>	<b>79</b>
5.1	Introduction . . . . .	79
5.2	Formalisation du problème et hypothèses . . . . .	80
5.2.1	Distributions temps-fréquence . . . . .	80

5.2.2	Conditions TF sur les sources . . . . .	82
5.3	Approche TF-UBSS basée sur la classification pour des sources TF- disjointes . . . . .	85
5.3.1	Algorithme TF-UBSS quadratique basé sur la classification . . . . .	85
5.3.2	Algorithme TF-UBSS linéaire basé sur la classification . . . . .	88
5.4	Approche TF-UBSS basée sur la projection en sous-espace pour des sources TF-non-disjointes . . . . .	92
5.4.1	Algorithme TF-UBSS quadratique basé sur la projection en sous-espace . . . . .	92
5.4.2	Algorithme TF-UBSS linéaire basé sur la projection en sous- espace . . . . .	94
5.5	Discussion . . . . .	96
5.6	Simulations . . . . .	98
5.6.1	Simulations de l'algorithme TF-UBSS linéaire . . . . .	98
5.6.2	Simulations de l'algorithme TF-UBSS quadratique . . . . .	102
5.7	Conclusion . . . . .	104
<b>6</b>	<b>La SAS audio utilisant la représentation temps-fréquence : mélanges convo- lutifs</b>	<b>107</b>
6.1	Introduction . . . . .	107
6.2	Formalisation du problème et hypothèses . . . . .	108
6.3	L'algorithme TF-CUBSS . . . . .	109
6.3.1	Identification du canal . . . . .	109
6.3.2	Algorithme TF-CUBSS basé sur la classification . . . . .	109
6.3.3	Algorithme TF-CUBSS basé sur la projection en sous-espace . . . . .	111
6.4	Discussion . . . . .	113
6.5	Simulations . . . . .	114
6.6	Conclusion . . . . .	121
<b>III</b>	<b>La SAS audio utilisant les techniques itératives</b>	<b>123</b>
<b>7</b>	<b>La SAS audio utilisant la parcimonie temporelle</b>	<b>125</b>
7.1	Introduction . . . . .	125
7.2	L'algorithme ISBS pour les mélanges instantanés . . . . .	126
7.3	L'algorithme ISBS pour les mélanges convolutifs . . . . .	128
7.3.1	Résolution des ambiguïtés de permutation et d'échelle . . . . .	129

7.4	Simulations . . . . .	130
7.5	Conclusion . . . . .	133
<b>8</b>	<b>La SAS utilisant les statistiques d'ordre deux</b>	<b>137</b>
8.1	Introduction . . . . .	137
8.2	Algorithme de séparation utilisant le gradient relatif . . . . .	138
8.2.1	Fonction de contraste basée sur les statistiques d'ordre deux	138
8.2.2	Algorithme de décorrélation utilisant le gradient relatif . . .	138
8.2.3	Implémentation adaptative . . . . .	141
8.3	Analyse de performances . . . . .	141
8.3.1	Analyse théorique des performances . . . . .	141
8.3.2	Validations et simulations . . . . .	144
8.4	Conclusion . . . . .	147
	<b>Conclusion générale</b>	<b>149</b>
<b>A</b>	<b>Identifiabilité au second ordre</b>	<b>153</b>
A.1	Conditions nécessaires et suffisantes d'identifiabilité . . . . .	153
A.2	Identifiabilité partielle . . . . .	155
A.3	Test d'identifiabilité . . . . .	156
A.3.1	Résultat théorique . . . . .	156
A.3.2	Test d'identifiabilité utilisant des techniques de rééchantillonnage . . . . .	157
A.3.3	Techniques de rééchantillonnage : jackknife . . . . .	158
A.4	Discussion . . . . .	160
A.5	Simulations et résultats . . . . .	161
A.6	Conclusion . . . . .	165
<b>B</b>	<b>Démonstrations</b>	<b>167</b>
B.1	Démonstration du théorème A.3 . . . . .	167
B.2	Démonstration du théorème A.4 . . . . .	168
	<b>Bibliographie</b>	<b>170</b>

# Table des figures

2.1	Amplitude du canal gauche en fonction de l'amplitude du canal droit pour un mélange stéréo de 3 sources parcimonieuses. . . . .	31
2.2	Représentation temps-fréquence d'un signal de parole : la représentation dans le domaine temps-fréquence permet d'illustrer la propriété de parcimonie des signaux audio. . . . .	32
3.1	Représentation temps-fréquence de trois composantes modales (nous avons utilisé la transformée de Fourier à court terme). . . . .	40
3.2	Illustration des données à classifier, ou nous représentons les différents vecteurs estimés $\hat{a}_i^j$ et leurs centroïdes. . . . .	46
3.3	Exemple de séparation de 4 signaux audio pour 3 capteurs : signaux originaux (ligne du haut); signaux estimés par pseudo-inversion quand la matrice de mélange est connue (ligne du milieu); signaux estimés par l'algorithme MD-UBSS utilisant la méthode de décomposition EMD (ligne du bas). . . . .	52
3.4	Performances de l'algorithme de séparation MD-UBSS dans le cas de 4 sources audio pour 3 capteurs : les courbes représentent la valeur moyenne de l'erreur quadratique de l'estimation des signaux sources (EQMN) en fonction du rapport signal à bruit (RSB). . . . .	53
3.5	EQMN en fonction de $L$ pour 4 sources audio et 3 capteurs : comparaison des performances de l'algorithme MD-UBSS (ESPRIT) pour $L$ variant dans la plage $[10 \dots 40]$ avec $RSB=10\text{dB}$ et $RSB=30\text{dB}$ . . . . .	54
3.6	Performances de l'algorithme de séparation MD-UBSS pour 3 capteurs : la courbe représente la valeur moyenne de l'erreur quadratique de l'estimation des signaux sources en fonction de leur nombre. . . . .	54



3.7 Performances des algorithmes de séparation MD-UBSS pour 4 sources audio et 3 capteurs : les courbes représentent la valeur moyenne de l'erreur quadratique de l'estimation des signaux sources en fonction du rapport signal à bruit pour l'algorithme ESPRIT classique ainsi que celle pour l'algorithme ESPRIT avec projection en sous-espace. . . . .	55
3.8 Performances des l'algorithmes de séparation MD-UBSS pour 4 sources audio et 3 capteurs : les courbes représentent la valeur moyenne de l'erreur quadratique de l'estimation des signaux sources en fonction du rapport signal à bruit pour l'algorithme MD-UBSS avec sélection optimale ainsi que celle pour l'algorithme MD-UBSS modifié. . . . .	56
3.9 Estimation de la matrice de mélange : la courbe représente la valeur moyenne de l'erreur quadratique de l'estimation de la matrice de mélange en fonction du rapport signal à bruit pour l'algorithme MD-UBSS pour 4 signaux audio et 3 capteurs. . . . .	57
3.10 EQMN en fonction du RSB pour $N = 4$ sources audio et $M = 3$ capteurs avec une taille d'observation $T = 10000$ : comparaison des performances de séparation de l'algorithme MD-UBSS (EMD et paramétrique) et l'algorithme DUET modifié. . . . .	58
4.1 Représentation temporelle de 4 signaux audio : cette représentation illustre la propriété de parcimonie des signaux audio (il existe des intervalles de temps où une seule source est présente). . . . .	61
4.2 Histogramme représentant le nombre des intervalles de temps pour chaque nombre estimé de sources pour 4 sources audio et 3 capteurs dans le cas du mélange convolutif. . . . .	64
4.3 Exemple de séparation aveugle de sources pour 3 sources audio et 4 capteurs dans le cas du mélange convolutif : la première ligne représente les signaux originaux, les deux lignes suivantes représentent les $M$ mélanges et la dernière représente les signaux estimés par notre algorithme. . . . .	70

4.4 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 10000$  échantillons dans le cas du mélange convolutif : comparaison des performances de la méthode d'identification de canal basée sur le modèle SIMO et celle utilisant les modèles SIMO et MIMO. . . . . 72

4.5 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 10000$  échantillons dans le cas du mélange convolutif : comparaison des performances de l'algorithme MD-UBSS dans le cas où la réponse impulsionnelle du canal  $\mathbf{H}$  est connue et lorsque elle est perturbée par un bruit Gaussien pour différentes valeurs du EQMNC. . . . . 73

4.6 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 10000$  dans le cas du mélange convolutif : comparaison des performances de l'algorithme MD-UBSS dans le cas où la réponse impulsionnelle du canal  $\mathbf{H}$  est connue et lorsque elle est estimée. . . . . 74

4.7 RISF en fonction du RSB pour  $N = 2$  sources audio et  $M = 2$  capteurs avec une taille d'observation  $T = 10000$  : comparaison des performances de séparation de l'algorithme MD-UBSS, l'algorithme de Y. Deville et l'algorithme de L. Parra. . . . . 75

5.1 La condition de sources TF-disjointes :  $\Omega_1 \cap \Omega_2 = \emptyset$  (où  $\Omega_1 \cap \Omega_2 \approx \emptyset$ ). 83

5.2 La condition de sources TF-non-disjointes :  $\Omega_1 \cap \Omega_2 \neq \emptyset$ . . . . . 84

5.3 Exemple de simulation de l'algorithme TF-UBSS linéaire basé sur la projection en sous-espace dans le cas de  $N = 4$  sources audio et  $M = 3$  capteurs. Les quatres premières figures représentent les signaux sources, les trois figures du milieu représentent les trois mélanges, et les quatres dernières figures représentent les signaux sources estimés par notre algorithme. . . . . 99

5.4 Exemple de simulation (dans le domaine temporel) de l'algorithme TF-UBSS linéaire basé sur la projection en sous-espace dans le cas de  $N = 4$  sources audio et  $M = 3$  capteurs. (a) les signaux sources originaux, (b) les signaux estimés par notre algorithme, (c) les signaux capteurs (mélanges). . . . . 100

5.5 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 25000$  échantillons : comparaison des performances des algorithmes TF-UBSS linéaires basés sur la classification et sur la projection en sous-espace. . . . . 101

5.6 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 25000$  échantillons : performances de la méthode d'estimation de la matrice de mélange  $\mathbf{A}$ . 101

5.7 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 25000$  échantillons : comparaison des performances de séparation de l'algorithme TF-UBSS basé sur la projection en sous-espace dans le cas où la matrice de mélange  $\mathbf{A}$  est exactement connue, et ceux obtenues en utilisant notre méthode d'estimation de la matrice  $\mathbf{A}$ . . . . . 102

5.8 EQMN en fonction du nombre de sources  $N$  pour  $M = 3$  capteurs avec une taille d'observation  $T = 25000$  : comparaison des performances des algorithmes TF-UBSS basés sur la classification et celui basé sur la projection en sous-espace. . . . . 103

5.9 EQMN en fonction du RSB pour  $N = 5$  sources audio et  $M = 4$  capteurs avec une taille d'observation  $T = 25000$  échantillons : comparaison des performances de l'algorithme TF-UBSS basé sur la classification et l'algorithme TF-UBSS basé sur la projection en sous-espace pour différentes valeurs de  $\mathcal{J}$ . . . . . 104

5.10 Exemple de simulation pour l'algorithme TF-UBSS quadratique dans le cas de  $N = 4$  sources FML et  $M = 3$  capteurs. De gauche à droite, les figures représentent respectivement les signatures TF des sources originales, des sources estimées par la méthode basée sur la classification, et celles des sources estimées par la méthode basée sur la projection en sous-espace. . . . . 105

5.11 EQMN en fonction du RSB pour  $N = 4$  sources LFM et  $M = 3$  capteurs avec une taille d'observation  $T = 256$  échantillons : comparaison des performances des algorithmes TF-UBSS quadratiques basés sur la classification et sur la projection en sous-espace. . . . . 106

6.1 Exemple de simulation de l’algorithme TF-CUBSS basé sur la projection en sous-espace dans le cas de  $N = 4$  sources audio et  $M = 3$  capteurs. Les quatre premières figures représentent les signaux sources, les trois figures du milieu représentent les trois mélanges, et les quatre dernières figures représentent les signaux sources estimés par notre algorithme. . . . . 115

6.2 Exemple de simulation (dans le domaine temporel) de l’algorithme TF-CUBSS basé sur la projection en sous-espace dans le cas de  $N = 4$  sources audio et  $M = 3$  capteurs. (a) les signaux sources originaux, (b) les signaux estimés par notre algorithme, (c) les signaux capteurs (mélanges). . . . . 116

6.3 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances des algorithmes TF-CUBSS basés sur la classification et sur la projection en sous-espace. . . . . 117

6.4 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances de séparation de l’algorithme TF-CUBSS basé sur la classification dans le cas où le canal de mélange  $\mathbf{H}$  est exactement connu, et celles obtenues en utilisant notre méthode d’estimation du canal  $\mathbf{H}$ . . . . . 117

6.5 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances de séparation de l’algorithme TF-CUBSS basé sur la projection en sous-espace dans le cas où le canal de mélange  $\mathbf{H}$  est exactement connu, et ceux obtenues en utilisant notre méthode d’estimation du canal  $\mathbf{H}$ . . . . . 118

6.6 EQMN en fonction du RSB pour  $N = 5$  sources audio et  $M = 4$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances de l’algorithme TF-CUBSS basé sur la classification et l’algorithme TF-CUBSS basé sur la projection en sous-espace pour différentes valeurs de  $\mathcal{J}$ . . . . . 119

6.7 EQMN en fonction du nombre de sources  $N$  pour  $M = 3$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances des algorithmes TF-CUBSS basés sur la classification et celui basé sur la projection en sous-espace. . . . . 119

6.8 EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 50000$  : comparaison des performances de séparation de l'algorithme TF-CUBSS basé sur la projection en sous-espace pour différentes valeurs de la taille du canal de mélange  $\mathcal{K}$ . . . . . 120

6.9 RISF en fonction du RSB pour  $N = 2$  sources audio et  $M = 2$  capteurs avec une taille d'observation  $T = 50000$  : comparaison des performances de séparation de l'algorithme TF-CUBSS basé sur la projection en sous-espace, l'algorithme de Y. Deville et l'algorithme de L. Parra. . . . . 121

7.1 Exemple de simulation de l'algorithme ISBS dans le cas de  $N = 2$  sources audio et  $M = 2$  capteurs. De haut en bas on a les signaux sources originaux, les signaux capteurs (mélanges) et les signaux estimés par notre algorithme. . . . . 131

7.2 RIS en fonction du RSB pour  $N = 2$  sources audio et  $M = 5$  capteurs avec une taille d'observation  $T = 20000$  dans le cas du mélange instantané : comparaison des performances de séparation entre notre algorithme ISBS, l'algorithme SOBI et l'algorithme du Newton relatif. 132

7.3 EQMN en fonction du RSB pour  $N = 2$  sources audio et  $M = 5$  capteurs avec une taille d'observation  $T = 20000$  dans le cas du mélange instantané : comparaison des performances de séparation entre notre algorithme ISBS, l'algorithme SOBI et l'algorithme du Newton relatif. . . . . 133

7.4 RIS en fonction du nombre d'itérations pour  $N = 2$  sources audio et  $M = 5$  capteurs avec une taille d'observation  $T = 20000$  dans le cas du mélange instantané : convergence de l'algorithme ISBS pour différentes valeurs du RSB. . . . . 134

7.5 RIS en fonction du nombre de capteurs  $M$  pour  $N = 2$  sources audio une taille d'observation  $T = 20000$  dans le cas du mélange instantané : comparaison des performances de séparation entre notre algorithme ISBS et l'algorithme SOBI. . . . . 134

7.6	RISF en fonction du RSB pour $N = 2$ sources audio et $M = 2$ capteurs avec une taille d'observation $T = 20000$ dans le cas du mélange convolutif : comparaison des performances de séparation entre notre algorithme ISBS, l'algorithme de Y. Deville ; l'algorithme de L. Parra et l'algorithme du Newton relatif. . . . .	135
8.1	RIS en fonction du RSB pour $N = 2$ sources auto-régressives et $M = 4$ capteurs avec une taille d'observation $T = 1000$ : comparaison des performances entre notre algorithme de décorrélation IDA et l'algorithme SOBI. . . . .	144
8.2	RIS en fonction de la taille des observations pour $N = 2$ sources auto-régressives et $M = 4$ capteurs avec un RSB = 40dB : comparaison des performances empiriques et les performances théoriques de notre algorithme IDA. . . . .	145
8.3	RIS en fonction du RSB pour $N = 2$ sources auto-régressives et $M = 4$ capteurs avec une taille d'observation $T = 1000$ : comparaison des performances empiriques et des performances théoriques de notre algorithme IDA. . . . .	146
8.4	RIS en fonction du nombre de capteurs $M$ pour $N = 2$ sources auto-régressives et une taille d'observation $T = 1000$ : représentation des performances théoriques de l'algorithme IDA pour différentes valeurs du RSB. . . . .	146
8.5	RIS en fonction du décalage spectral $\delta\theta$ pour $N = 2$ sources auto-régressives et $M = 4$ capteurs avec une taille d'observation $T = 1000$ : représentation des performances théoriques de l'algorithme IDA. . . . .	147
8.6	RIS en fonction du nombre de retards $K$ pour $N = 2$ sources auto-régressives et $M = 4$ capteurs avec une taille d'observation $T = 1000$ : représentation des performances théoriques de l'algorithme IDA. . . . .	148
A.1	Taux de succès en fonction du RSB pour $N = 2$ sources auto-régressives et $M = 2$ capteurs avec $\beta = 99.7\%$ : comparaison des performances de notre procédure de test pour différentes tailles d'observations $T$ . . . . .	162
A.2	Taux de succès en fonction de la taille des observations $T$ pour $N = 2$ sources auto-régressives et $M = 2$ capteurs avec un RSB = 25dB : comparaison des performances de notre procédure de test pour différentes valeurs de l'intervalle de confiance $\beta$ . . . . .	163

---

A.3	Taux de succès en fonction de la valeur de l'intervalle de confiance $\beta$ pour $N = 2$ sources auto-régressives et $M = 2$ capteurs avec un RSB = 25dB : comparaison des performances de notre procédure de test pour différentes tailles d'observations $T$ . . . . .	164
A.4	Taux de succès en fonction du décalage spectral $\delta\theta$ pour $N = 2$ sources auto-régressives et $M = 5$ capteurs avec un RSB = 30dB. . . . .	164
A.5	Valeurs moyennes de $ \mathcal{E}(i, j) $ et les seuils $1 - \alpha_{(i, j)}$ en fonction du RSB pour $N = 3$ sources et $M = 5$ capteurs : les 2 deux premières sources sont des processus aléatoires complexes suivant une loi Gaussienne et temporellement blancs et la troisième est un signal auto-régressif du premier ordre. . . . .	165

# Liste des tableaux

3.1	Algorithme MD-UBSS dans le cas du mélange instantané . . . . .	43
4.1	Algorithme MD-UBSS dans le cas du mélange convolutif . . . . .	62
5.1	Algorithme TF-UBSS quadratique basé sur la classification . . . . .	86
5.2	Algorithme TF-UBSS linéaire basé sur la classification . . . . .	89
5.3	Algorithme TF-UBSS quadratique basé sur la projection en sous- espace . . . . .	94
5.4	Algorithme TF-UBSS linéaire basé sur la projection en sous-espace .	96
6.1	Algorithme TF-CUBSS basé sur la classification . . . . .	110
6.2	Algorithme TF-CUBSS basé sur la projection en sous-espace . . . .	113
7.1	Algorithme ISBS pour <i>Iterative Sparse Blind Separation</i> . . . . .	128
A.1	Test d'identifiabilité utilisant les techniques de rééchantillonnage .	158
A.2	Test d'identifiabilité utilisant la technique jackknife . . . . .	160





# Abréviations

---

Pour des raisons de lisibilité, la signification d'une abréviation ou d'un acronyme n'est souvent rappelée qu'à sa première apparition dans le texte d'un chapitre. Par ailleurs, puisque nous utilisons toujours l'abréviation la plus usuelle, il est fréquent que ce soit le terme anglais qui soit employé.

<b>ACI</b>	<i>Analyse en Composantes Indépendantes</i>
<b>ACP</b>	<i>Analyse en Composantes Principales</i>
<b>AIC</b>	<i>Akaike Information Criterion</i>
<b>BSS</b>	<i>Blind Source Separation</i>
<b>dB</b>	<i>decibels</i>
<b>EM</b>	<i>Expectation Maximization</i>
<b>EMD</b>	<i>Empirical Mode Decomposition</i>
<b>EQMN</b>	<i>Erreur Quadratique Moyenne Normalisée</i>
<b>ESPRIT</b>	<i>Estimation of Signal Parameters via Rotational Invariance Techniques</i>
<b>ICA</b>	<i>Independent Component Analysis</i>
<b>i.i.d</b>	<i>indépendant et identiquement distribué</i>
<b>MIMO</b>	<i>Multiple-Input Multiple-Output</i>
<b>RIS</b>	<i>Rapport Interférences à Signal</i>
<b>RSB</b>	<i>Rapport Signal à Bruit</i>
<b>TFCT</b>	<i>Transformée de Fourier à Court Terme</i>
<b>SAS</b>	<i>Séparation Aveugle de Sources</i>
<b>SIMO</b>	<i>Single-Input Multiple-Output</i>
<b>STFT</b>	<i>Short Time Fourier Transform</i>
<b>UBSS</b>	<i>Underdetermined Blind Source Separation</i>



# Notations

---

Nous avons regroupé ci-dessous les principales notations employées dans les différents chapitres du document. Dans la mesure du possible, nous avons tenté de conserver les mêmes notations d'un chapitre à l'autre.

$\boldsymbol{v}$	vecteur
$\boldsymbol{M}$	matrice
$(\cdot)^T$	transposé
$(\cdot)^*$	conjugué
$(\cdot)^H$	conjugué transposé
$\Re$	partie réelle
$\Im$	partie imaginaire
$ \cdot $	module
$\mathbb{E}[\cdot]$	espérance mathématique
$\widehat{(\cdot)}$	estimateur d'un paramètre
$\det(\cdot)$	déterminant d'une matrice
$\text{diag}(\boldsymbol{v})$	matrice diagonale construite à partir du vecteur $\boldsymbol{v}$
$\text{diag}(\boldsymbol{M})$	matrice diagonale construite à partir des éléments diagonaux de la matrice $\boldsymbol{M}$
$\text{trace}(\cdot)$	trace d'une matrice
$\boldsymbol{M}^\#$	pseudo-inverse au sens de Moore-Penrose $\boldsymbol{M}^\# = (\boldsymbol{M}^H \boldsymbol{M})^{-1} \boldsymbol{M}^H$

# Introduction générale

---

## Présentation du problème

Parce qu'elles s'appliquent à toutes les étapes d'une chaîne d'acquisition, d'analyse, de transfert et de restitution des données, les techniques du traitement du signal trouvent des applications dans pratiquement tous les domaines de la technologie, particulièrement en audio où on cherche à améliorer les techniques d'enregistrement et de compression pour obtenir la plus grande qualité sonore possible. Les techniques de correction d'écho permettent de réduire les effets de réflexions acoustiques dans la pièce. Le traitement du son s'est largement amélioré grâce aux ordinateurs. La synthèse sonore permet en outre de créer des sons artificiels ou de recréer les sons d'instruments naturels. Elle a été à l'origine de nombreux bouleversements en musique. La majorité des signaux audio sont des mélanges auxquels contribuent plusieurs sources. Ainsi un dialogue de film réalisé à la table de mixage est un mélange synthétique entre les paroles successives de plusieurs locuteurs. Et un fond sonore de rue enregistré au micro est un mélange réel entre les bruits simultanés de plusieurs voitures. L'étude des mélanges audio a de nombreuses applications dont la plus populaire est le problème du cocktail (*cocktail party problem*). Ce nom provient de la situation rencontrée lors d'une fête. Comment comprendre ce que dit votre voisin parmi les autres voix, la musique, le bruit des verres et les autres bruits de fond ? L'audition humaine effectue dans ce cas une écoute sélective basée sur les caractéristiques spatiales et spectro-temporelles des sources sonores en présence. Elle prend également en compte des informations a priori telles que la position spatiale des sources fournie par la vision et les particularités apprises de la voix et du langage. Ce type de comportement peut être

copié par un système artificiel visant la séparation de sources pour l'écoute ou bien la transcription du texte prononcé.

Dans ce travail, nous considérons le problème de la séparation de sources audio issue d'un mélange enregistré à partir de plusieurs microphones. Ces enregistrements correspondent à des types de mélanges assez variés, selon le rapport entre le nombre de sources et de capteurs, l'environnement d'enregistrement, le matériel utilisé...etc.

## Motivations

Cette thèse se déroule dans le cadre d'une étude sur la séparation de sources audio dans un milieu réverbérant. Plus précisément, l'application visée est la séparation de sources pour le débruitage de la parole dans les véhicules. Cette étude est menée par l'ENST d'une part et FTR&D (Lannion) d'autre part.

Dans la suite de ce document, nous exposerons plusieurs techniques de séparation aveugle de sources dont de nombreuses méthodes basées sur les statistiques d'ordre supérieur. Cette voie ne sera pas considérée vu le coût de calcul, la complexité et la faible vitesse de convergence de ces méthodes.

Par contre, sachant que nous dédions notre étude à la séparation de sources audio, nous exploiterons les propriétés des signaux audio afin de mettre en oeuvre des méthodes de séparation plus efficaces. En particulier, nous tirerons parti des propriétés de non-stationnarité et de parcimonie dans différents domaines ainsi que du caractère harmonique des signaux, pour améliorer les performances de séparation, réduire le coût de calcul, améliorer la vitesse de convergence et traiter le cas plus délicat de la séparation de sources sous-déterminée (c'est à dire le cas où l'on a moins de capteurs que de sources).

Effectivement, nous avons consacré une bonne partie de notre travail au cas sous-déterminé. Ce choix se motive par le fait qu'une telle configuration est plus plausible dans le cas d'un enregistrement dans un environnement réel, étant donné que le choix du nombre de sources n'est pas un paramètre contrôlable. En plus, en partant du principe de "celui qui peut le plus peut le moins", on peut très facilement généraliser ces méthodes pour le cas des mélanges sur-déterminés, ce qui nous offrira un large choix d'applications.

## Plan du document et contributions

Nous présenterons ci-après un aperçu des différents chapitres de ce mémoire.

Nous commencerons dans les chapitres 1 et 2 par présenter le problème de séparation de sources en expliquant son principe, en donnant son modèle mathématique et en citant quelques exemples d'application de la séparation de sources. Dans une deuxième phase, nous donnons une présentation structurée de quelques méthodes de séparation de sources avec un intérêt particulier pour les méthodes dédiées aux signaux audio.

Dans les chapitres 3 et 4 nous montrerons comment effectuer la séparation de sources audio en utilisant une méthode basée sur des algorithmes de décomposition modale (EMD ou ESPRIT) [1–4]. Les avantages de cette approche résident dans le fait qu'elle permet de traiter le cas des mélanges instantanés et convolutifs, et elle nous permet en particulier, de traiter le cas sous-déterminé. Cette approche se base sur le fait que les signaux audio (et particulièrement les signaux musicaux) peuvent être bien modélisés localement par une somme de signaux périodiques. Ces signaux seront donc décomposés en utilisant les algorithmes de décomposition modale et recombinaison par classification suivant leurs directions spatiales regroupant ainsi les composantes de chacune des sources. Cette méthode peut être utilisée dans le cas du mélange convolutif sur-déterminé [5], en lui conjuguant une décomposition en valeurs et vecteurs propres et un critère de parcimonie approprié.

Notons que le terme de “parcimonie” sera utilisé tout au long de ce document pour définir le caractère creux ou clairsemé d'un signal ou d'une représentation.

Nous présenterons dans le chapitre 5 une méthode qui permet d'effectuer la séparation de mélange instantané de sources audio dans le cas sous-déterminé en utilisant la propriété de parcimonie des signaux audio dans le domaine temps-fréquence [6–9]. Nous proposons deux méthodes utilisant différentes transformées dans le domaine temps-fréquence. La première utilise les distributions temps-fréquence quadratiques DTFQ (ou STFD pour *Spatial Time Frequency Distribution*), la deuxième utilise la transformée de Fourier à court terme TFCT (ou STFT pour *Short Time Fourier Transform*). Ces deux méthodes supposent que les sources sont disjointes dans le domaine temps-fréquence ; c'est à dire qu'une seule source est présente dans chaque point temps-fréquence. Nous proposons ensuite de relâcher cette contrainte en supposant que les sources ne sont pas forcément disjointes dans le domaine temps-fréquence. En particulier, le nombre de sources

présentes en un point temps–fréquence doit être strictement inférieur au nombre de capteurs. Pour effectuer la séparation dans ce cas de figure, on utilise la projection en sous-espace qui permet d’identifier les sources présentes dans chaque point temps–fréquence et de déterminer la valeur correspondante à chaque source. Une contribution sous-jacente est celle d’une nouvelle méthode d’estimation de la matrice de mélange dans le cas sous-déterminé.

Dans le chapitre 6, nous introduirons une approche basée sur la transformation dans le domaine temps–fréquence pour résoudre le problème de séparation de mélange convolutif de sources audio dans le cas sous-déterminé [10, 11]. Cette approche nécessite une pré estimation du canal de mélange qui sera réalisée en exploitant la propriété de parcimonie temporelle des signaux audio et la structure des canaux acoustiques.

Dans le chapitre 7, nous avons exploité la propriété de parcimonie des signaux audio dans le domaine temporel. Nous avons proposé une méthode itérative utilisant une technique du gradient relatif qui minimise une fonction de contraste basée sur la norme  $\ell_p$ . Cette norme pour  $p < 2$  (ou même  $p \leq 1$ ) est considérée comme une bonne mesure de parcimonie [12–14]. Les simulations prouvent que la méthode proposée surpasse d’autres méthodes basées sur l’indépendance des sources.

Finalement, nous nous sommes intéressés dans la chapitre 8 à une méthode itérative de séparation de sources utilisant les statistiques d’ordre deux (SOS pour Second Order Statistics) [15, 16]. Les statistiques d’ordre deux des données observées s’avèrent suffisantes pour séparer des sources mutuellement non corrélées à condition que les vecteurs de corrélation temporelle des sources considérées soient linéairement indépendants par paires. En appliquant une technique de gradient relatif, on déduit un algorithme itératif qui possède un certain nombre de propriétés attrayantes, comme sa simplicité et sa généralisation aux cas adaptatifs ou convolutifs. Nous proposons aussi une analyse de performances asymptotiques de cet algorithme qui sera validée par plusieurs simulations.

Nous terminons notre exposé par une conclusion qui résume les apports de nos travaux et ouvre un ensemble de perspectives en séparation de sources audio en particulier et en séparation de sources en général.



# La séparation de sources audio

---

## 1.1 Bref historique

La séparation aveugle de sources (SAS) consiste à estimer  $N$  signaux inconnus (les sources) d'après la seule connaissance de  $M$  mélanges de ces signaux (les observations). Le terme aveugle signifie que les sources ne sont pas observées et qu'on ne connaît pas les paramètres de mélange. La séparation aveugle de sources possède de nombreuses applications [17, 18], notamment dans les domaines du traitement du signal audio et du traitement du signal biomédical. Dans le domaine de l'audio, la séparation aveugle de sources concerne par exemple le séparation d'extraits musicaux stéréophoniques ou le rehaussement de la parole pour la téléphonie mobile ou pour l'élaboration de prothèses auditives [19]. Dans le domaine biomédical, de nombreux travaux existent sur la séparation de signaux électroencéphalogrammes (EEG) et électrocardiogrammes (ECG). La séparation aveugle de sources possède plusieurs degrés de difficulté, selon les caractéristiques des sources et surtout du mélange. Plusieurs paramètres sont à prendre en considération. Le premier paramètre est le nombre  $M$  d'observations par rapport au nombre  $N$  de sources. Intuitivement, on conçoit aisément que le cas sur-déterminé ( $M \geq N$ ) est plus simple à résoudre que le cas sous-déterminé ( $M < N$ ), ce dernier cas ne pouvant généralement être résolu qu'au prix d'une importante information a priori sur les sources. Un deuxième paramètre est la nature du mélange. Le mélange le plus simple est le mélange linéaire instantané : à chaque instant, les obser-

vations sont des combinaisons linéaires des sources au même instant. Dans la réalité, les mélanges sont souvent convolutifs : à chaque instant les observations dépendent aussi des valeurs des signaux sources aux instants précédents. En audio par exemple, lorsque deux musiciens (jouant d'un instrument acoustique) sont enregistrés par différents micros dans une salle, l'effet de réverbération de la salle produit au niveau des micros un mélange convolutif. Dans la réalité, les mélanges sont aussi souvent variant dans le temps. Par exemple lorsque des musiciens se déplacent, les fonctions de transfert de la salle entre les émetteurs (les musiciens) et les récepteurs (les micros) changent suivant les positions des émetteurs et récepteurs, produisant donc sur les micros un mélange variant dans le temps. Enfin, les mélanges peuvent être non-linéaires, par exemple certains micros ont un comportement non-linéaire, bien souvent négligeable, par rapport à l'intensité des sources enregistrées (les micros de studio). Enfin, un troisième paramètre est la nature des sources. La plupart des méthodes de séparation aveugle de sources reposent sur l'hypothèse que les sources sont mutuellement indépendantes, au moins à l'ordre deux. Ceci est une contrainte forte qui ne peut pas toujours être vérifiée en pratique.

Les recherches dans le domaine de la séparation de sources ont été initiées en 1985 par C. Jutten et J. Héroult [20, 21]. Ces derniers ont développé, à partir d'un problème de décodage du mouvement en neurobiologie, un algorithme de séparation à base d'un *réseau neuromimétique*. Cet algorithme cherche à résoudre le problème du mélange instantané sans bruit avec une matrice de mélange carrée et des signaux à valeurs réelles. En 1988 J. L. Lacoume et P. Ruiz [22] ont proposé un algorithme minimisant une fonction objective basée sur les cumulants croisés des sorties d'un filtre inverse, qui sont les outils naturels de mesure d'indépendance. En 1989, J. F. Cardoso [23] propose de considérer les propriétés algébriques des cumulants d'ordre quatre en utilisant l'information d'ordre deux pour rendre unitaire la matrice de mélange puis de l'identifier par des techniques de diagonalisation. En revanche, il montre dans [24] et [25] que les cumulants d'ordre quatre sont suffisants pour permettre, à eux seuls, d'estimer la matrice de mélange.

L'approche développée par P. Comon [26], en 1990, élargit la théorie de l'Analyse en Composantes Principales (ACP) à l'Analyse en Composantes Indépendantes (ACI) et introduit la notion de fonction de contraste. P. Comon propose comme contraste à maximiser la somme des cumulants d'ordre quatre. Indépendamment des travaux de P. Comon, M. Gaeta et J. L. Lacoume [27] proposent de résoudre

le problème de séparation par une technique du maximum de vraisemblance approchée. En 1993, N. Delfosse et Ph. Loubaton [28] proposent une approche de type “déflation” qui consiste à adapter, après blanchiment, un filtre ligne pour extraire une des sources, simultanément un complément orthogonal est obtenu, la sortie de ce dernier contient un mélange de l’ensemble des autres sources auxquelles on applique un filtre semblable de dimension inférieure. B. Laheld et J. F. Cardoso [29] ont développé en 1993 une famille d’algorithmes adaptatifs basés sur la notion de gradient matriciel relatif. En absence de bruit, ces algorithmes ont l’avantage de fournir des performances indépendantes de la matrice de mélange et donc de son conditionnement. En 1992, D. T. Pham, P. Garat et C. Jutten [30] résolvent le problème de séparation par une méthode fondée sur le principe du maximum de vraisemblance en considérant des observations non bruitées. Depuis ces travaux pionniers en séparation de sources, ce domaine de recherche a pris une importance grandissante au point où un workshop annuel lui a été dédié (*ICA Workshop*). Ainsi, une multitude de méthodes et d’algorithmes ont été proposés depuis, incluant celle basée sur les statistiques d’ordre deux proposée par A. Belouchrani et al. en 1997 dans [31] et celle pour la séparation de sources non-stationnaire proposée par D. T. Pham et J. F. Cardoso dans [32, 33] en 2001.

## 1.2 Les principes de la séparation aveugle de sources

### 1.2.1 Problème de la séparation de sources

Le modèle du mélange instantané suppose, à l’instant  $t$ , l’existence de  $N$  signaux sources notés  $s_i(t)$ , pour  $i = 1, \dots, N$  dont on observe  $M$  mélanges linéaires instantanés, appelés signaux capteurs. Le modèle s’écrit alors sous la forme :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t) \quad t = 0, \dots, T - 1, \quad (1.1)$$

où les  $M$  composantes du vecteur  $\mathbf{x}(t) \triangleq [x_1(t), \dots, x_M(t)]^T$  représentent les signaux capteurs, les  $N$  signaux sources  $s_1(t), \dots, s_N(t)$  supposés statistiquement indépendants sont contenus dans le vecteur  $\mathbf{s}(t) \triangleq [s_1(t), \dots, s_N(t)]^T$ , tandis que les coefficients des mélanges forment une matrice notée  $\mathbf{A} \triangleq [\mathbf{a}_1, \dots, \mathbf{a}_N]$  de dimension  $M \times N$ , appelée matrice de mélange où  $\mathbf{a}_i = [a_{1i}, \dots, a_{Mi}]^T$  contient les coefficients du mélange et on note par  $\mathbf{w}(t)$  un possible bruit additif.

Dans un contexte aveugle où l’on n’a aucune information a priori sur le mélange, une identification complète de la matrice de mélange  $\mathbf{A}$  est impossible. En effet,

l'échange d'un facteur complexe quelconque entre un signal source et le vecteur colonne de la matrice de mélange  $\mathbf{A}$  correspondant n'affecte en rien les observations. Ceci est montré par la relation suivante :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t) = \sum_{i=1}^N \frac{\mathbf{a}_i}{\alpha_i} \alpha_i s_i(t) + \mathbf{w}(t) . \quad (1.2)$$

Par ailleurs, l'indice  $i$  associé à une certaine source est arbitraire. Les sources seront donc connues à une permutation près. Finalement, nous appellerons **séparation de source** l'opération qui, à partir des signaux capteurs, nous permet dans le cas idéal (sans bruit) d'obtenir un ensemble de signaux proportionnels aux signaux émis par chacune des sources. De même, la matrice de mélange ne peut être estimée qu'à une constante et une permutation près de ses colonnes.

Le principe général de la séparation de sources consiste à appliquer des transformations aux observations pour obtenir des signaux *statistiquement indépendants* qui seront proportionnels aux signaux sources que l'on veut estimer. Ceci se formalise comme suit :

$$\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t) , \quad (1.3)$$

où  $\mathbf{B}$  est une matrice séparante dès que  $\mathbf{y}(t)$  est une estimation des sources aux indéterminations d'échelle et de permutation près. Pour la recherche des composantes indépendantes une première solution consiste à rendre les observations indépendantes à l'ordre deux en exploitant la matrice d'autocovariance de ces dernières, mais on se rend compte rapidement que celle-ci est insuffisante pour identifier la matrice de mélange. En effet, la matrice d'autocovariance des observations est donnée par :

$$\mathbf{R}_x = \mathbb{E}[\mathbf{x}(t)\mathbf{x}(t)^H] = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mathbf{R}_w , \quad (1.4)$$

$\mathbf{R}_s$  est une matrice diagonale que l'on supposera souvent égale à la matrice identité (il n'y a aucune perte de généralité du fait de l'indétermination d'amplitude des sources) et  $\mathbf{R}_w$  est la matrice d'autocovariance du bruit. Sans bruit, l'autocovariance devient  $\mathbf{R}_x = \mathbf{A}\mathbf{A}^H$ , On voit ainsi que la matrice de mélange  $\mathbf{A}$  est une racine carrée de  $\mathbf{R}_x$ . Toutefois la racine carrée d'une matrice Hermitienne positive n'est pas unique et n'est identifiée qu'à une matrice unitaire près. Ceci montre que l'on a besoin d'information supplémentaire (autre que la matrice de covariance) pour l'identification du mélange  $\mathbf{A}$ . Souvent (du moins pour des signaux sources *i.i.d*), cette information est donnée par les statistiques d'ordre supérieur de l'observation. Ceci nécessite en particulier des signaux sources non-Gaussiens

car dans le cas Gaussien les informations à l'ordre supérieur à deux dépendent uniquement des moments d'ordre un et deux.

Malheureusement, le modèle instantané de mélanges est plutôt insuffisant dans le cas des sources enregistrées dans un vrai environnement acoustique (cas d'une source sonore et d'un microphone dans une chambre). Les recherches précédentes ont prouvé que le signal capturé par le microphone peut être bien représenté par une convolution du signal de source avec un filtre RIF (réponse impulsionnelle finie) d'ordre élevé, modélisant l'acoustique de la pièce entre la source et le capteur [34]. Pour augmenter la précision du modèle, nous pourrions utiliser des filtres RII (réponse impulsionnelle infinie) pour modéliser l'acoustique de la pièce. Cependant, les filtres RII sont moins stables et ont besoin de condition de phase minimale pour être inversés. Pour contourner ce problème nous modéliserons le canal à l'aide des filtres RIF. Alors on peut écrire :

$$\mathbf{x}(t) = \sum_{k=0}^{\mathcal{K}} \mathbf{H}(k)\mathbf{s}(t-k) + \mathbf{w}(t), \quad (1.5)$$

où  $\forall k \in [0, \mathcal{K}]$ ,  $\mathbf{H}(k)$  est une matrice de taille  $M \times N$ .

Les premiers travaux modélisant cette situation de mélanges ont été initiés au début des années 90 ; selon ce que l'on trouve dans la littérature [35, 36] le premier travail qui traite le mélange convolutif est celui de C. Jutten et al. [37, 38]. Le mélange convolutif, relativement au mélange instantané, a été moins étudié mais il existe néanmoins beaucoup d'algorithmes pour ce type de mélange qui convient à une variété d'applications physiques.

## 1.2.2 Analyse en composantes indépendantes

Formalisée par P. Comon [39], l'analyse en composantes indépendantes (ACI) est une technique équivalente à la résolution du problème de séparation de sources (dans son expression la plus simple). Il s'agit de la généralisation de l'analyse en composantes principales (ACP). Le principe est de transformer linéairement des données vectorielles en données dont les composantes sont statistiquement indépendantes, alors que l'ACP ne les transforme qu'en composantes décorrélées. Il s'avère que, dans un contexte non bruité, cela est équivalent à l'estimation des paramètres du modèle instantané. Plus précisément, P. Comon définit l'analyse en composantes indépendantes d'un vecteur aléatoire  $\mathbf{x}(t)$  comme la donnée d'une matrice de séparation  $\mathbf{B}$  telle que  $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$  soit à composantes indépendantes.

Il montre que dans le cas  $M \geq N$ , et pour un vecteur  $\mathbf{x}$  suivant le modèle (1.1) non bruité, une matrice  $\mathbf{B}$  est séparante si et seulement si :

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

où  $\mathbf{\Lambda}$  est une matrice diagonale non-singulière (représente l'indétermination d'échelle) et  $\mathbf{P}$  est une matrice de permutation. On appelle une telle matrice  $\mathbf{\Lambda}\mathbf{P}$  une matrice de permutation généralisée. La recherche de la matrice  $\mathbf{B}$  est donc le but de l'ACI, elle s'effectue en général par minimisation d'un critère de séparation, c'est à dire un critère mesurant l'indépendance statistique de  $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$  :

$$\mathbf{B} = \arg \min \mathcal{C}(\mathbf{y}(t), \mathbf{B}) . \quad (1.7)$$

Pour mesurer l'indépendance, deux approches principales sont utilisées : soit on minimise l'information mutuelle entre les composantes du vecteur  $\mathbf{y}$ . Clairement, l'absence d'information mutuelle traduit l'indépendance des signaux. Soit aussi par maximisation de la "non-gaussianité" des composantes de  $\mathbf{y}$ . En effet, par le théorème central limite, le mélange des signaux *i.i.d* donne un signal dont la loi est proche (asymptotiquement) de la Gaussienne. Ainsi, séparer les signaux consiste à s'éloigner de la gaussianité. Une méthode simple pour le faire revient à maximiser (en valeur absolue) les cumulants d'ordre supérieur, car ceux de la Gaussienne sont nuls.

Basés sur ces principes généraux, plusieurs types de fonctions de contrastes pour l'ACI ont été étudiées ces dix dernières années. Citons par exemple les contributions de P. Comon [39–41], de J. F. Cardoso [42–44] dont les fonctions de contrastes sont basées sur l'utilisation des statistiques d'ordre supérieur et de S. I. Amari [45, 46] dont l'approche utilise des notions de théorie de l'information (principe de *l'infomax*). La plupart de ces fonctions de contrastes découlent du principe de l'estimation au maximum de vraisemblance. En effet, pour le modèle de l'ACI, maximum de vraisemblance, minimum d'entropie, minimum d'information mutuelle et infomax coïncident, si l'on impose des contraintes supplémentaires de blancheur sur les sources [47, 48]. On applique souvent un *pré-blanchiment* des observations avant d'adapter une matrice de séparation. Le principe est celui de l'ACP, qui permet de diminuer la dimension du problème et de rendre les observations indépendantes à l'ordre deux, c'est à dire décorréélées. On appelle matrice de blanchiment  $\mathbf{W}$  une matrice telle que :

$$\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t) ,$$

soit un vecteur spatialement blanc, c'est à dire (dans le cas sans bruit) :

$$\mathbb{E}[z(t)z(t)^H] = \mathbf{W}\mathbf{A}\mathbf{A}^H\mathbf{W}^H = \mathbf{I}_N .$$

Si  $\mathbf{A}$  est de taille  $M \times N$ , une matrice de blanchiment  $\mathbf{W}$  (par exemple une racine carrée de la matrice symétrique  $\mathbf{R}_x^{-1}$  projetée sur son espace image) est de taille  $N \times M$ . On est donc ramené à un modèle de mélange carré :

$$z(t) = \mathbf{W}\mathbf{A}s(t) = \mathbf{U}s(t) , \tag{1.8}$$

où la matrice de mélange  $\mathbf{U}$  est une matrice unitaire. Un tel pré-blanchiment permet de réduire la dimension de l'espace de paramètres à estimer, puisque l'espace des matrices unitaires est de dimension  $N(N - 1)/2$  au lieu de  $MN$  pour l'espace initial. Blanchir les observations, puis estimer la rotation manquante est un problème équivalent à la minimisation d'un contraste sous la contrainte de blancheur :

$$\mathbb{E}[\mathbf{y}(t)\mathbf{y}(t)^H] = \mathbf{I}_N .$$

On parle dans ce cas de *contraste orthogonal*, et sous cette contrainte, maximiser la vraisemblance est équivalent à minimiser l'information mutuelle [48].

### 1.2.3 Maximum de vraisemblance

L'estimation au sens du maximum de vraisemblance a été largement étudiée dans le cas d'un mélange inversible et dans le contexte d'un bruit d'observation négligeable. M. Gaeta et J. L. Lacoume [27] sont les premiers à considérer cette solution, en approchant la vraisemblance par un développement de Gram-Charlier. Leur approche rejoint celle de P. Comon [39], basée sur un développement d'Edgeworth. Les fonctions de contraste obtenues sont des approximations de la log-vraisemblance, et font intervenir les cumulants d'ordre supérieur à deux. Une approche basée sur la vraisemblance elle-même dans le cas carré non-bruité a été présentée par D. T. Pham, P. Garat et C. Jutten dans [30]. De nombreuses autres contributions considérant le maximum de vraisemblance pour la séparation de sources ont été présentées, citons par exemple S. I. Amari [45] et A. Belouchrani [49, 50].

Le principe du maximum de vraisemblance, connaissant la distribution des sources, consiste à maximiser

$$p(\mathbf{x}; \mathbf{A}) = |\det(\mathbf{A})|^{-1}r(\mathbf{A}^{-1}\mathbf{x}) , \tag{1.9}$$

où  $|\det(\mathbf{A})|$  désigne la valeur absolue du déterminant de  $\mathbf{A}$  et  $r(\mathbf{s})$  est la densité de la distribution des sources par rapport à la mesure de Lebesgue. Dans [51] O. Bermond définit la fonction **score** associée au paramètre matriciel  $\mathbf{A}$  comme la dérivée logarithmique de la vraisemblance par rapport à cette matrice :

$$S(\mathbf{x}, \mathbf{A}) = \frac{\partial}{\partial \mathbf{A}} \ln p(\mathbf{x}; \mathbf{A}) \quad (1.10)$$

$$= -\mathbf{A}^{-T} + \mathbf{A}^{-T} \psi(\mathbf{A}^{-1} \mathbf{x}) \mathbf{x}^T \mathbf{A}^{-T}, \quad (1.11)$$

où

$$\psi(\mathbf{u}) \triangleq -\frac{\partial}{\partial \mathbf{u}} \ln r(\mathbf{u}), \quad (1.12)$$

est l'opposé de la dérivée logarithmique de la densité des sources. Il s'agit en fait du **score** associé au paramètre de translation pour la densité  $r$ . On notant  $\mathbf{y}(t) = \mathbf{A}^{-1} \mathbf{x}(t)$ , l'estimée des sources à la sortie du séparateur, l'expression du **score** associé à la matrice de mélange devient simplement :

$$S(\mathbf{y}, \mathbf{A}) = \mathbf{A}^{-T} [\psi(\mathbf{y}) \mathbf{y}^T - \mathbf{I}_N]. \quad (1.13)$$

L'estimateur du maximum de vraisemblance  $\widehat{\mathbf{A}}_{mv}$  de la matrice  $\mathbf{A}$  pour une série d'observations indépendantes, est la solution de l'équation (où  $\mathbf{y}(t) = \widehat{\mathbf{A}}_{mv}^{-1} \mathbf{x}(t)$ ) :

$$\frac{1}{T} \sum_{t=0}^{T-1} \psi(\mathbf{y}(t)) \mathbf{y}(t)^T = \mathbf{I}_N. \quad (1.14)$$

J. F. Cardoso [44] et S. I. Amari [45] ont introduit respectivement les notions de gradient *relatif* et *naturel*. Le gradient relatif utilise la structure de groupe de l'espace des matrices inversibles, tandis que le gradient naturel considère cet espace comme étant Riemannien en le munissant d'une métrique locale donnée par la matrice d'information de Fisher (après approximation, le gradient naturel fournit la même définition que le gradient relatif dans le cas de la séparation de sources [52]). Il s'agit en fait de la dérivée de la log-vraisemblance en terme de système global ; le gradient relatif  $S_r(\mathbf{y})$  est défini par :

$$\ln p(\mathbf{x}; \mathbf{A}(\mathbf{I}_N + \mathcal{E})) = \ln p(\mathbf{x}; \mathbf{A}) + \langle S_r(\mathbf{y}), \mathcal{E} \rangle + o(\|\mathcal{E}\|), \quad (1.15)$$

et l'expression du **score** relatif est :

$$S_r(\mathbf{y}) = \psi(\mathbf{y}) \mathbf{y}^T - \mathbf{I}_N. \quad (1.16)$$

L'intérêt d'une telle expression réside dans le fait que ce **score** relatif ne fait intervenir le mélange qu'à travers la sortie du séparateur  $\mathbf{y}(t) = \mathbf{A}^{-1} \mathbf{x}(t)$ . Ce point a été



souligné par J. F. Cardoso et B. Laheld [44], comme résultant de l'équivariance du problème de séparation de sources. En d'autres termes, l'identification, ou de manière équivalente la séparation, n'est formellement pas plus difficile pour une matrice  $\mathbf{A}$  quelconque (et même mal conditionnée) que pour la matrice identité. Excepté pour certaines applications (comme les communications numériques par exemple), on ne dispose pas d'information a priori sur la distribution des sources, en conséquence, on ne connaît pas l'expression du **score** de translation  $\psi$ . Plusieurs approches ont été proposées pour réaliser l'estimation par maximum de vraisemblance dans ce cas. On peut avoir recours à des approximations de la vraisemblance utilisant les statistiques d'ordre supérieur des sources (cumulants), obtenues à l'aide du développement d'Edgeworth [39] ou de Gram-Charlier [46]. Une approche plus directe a été proposée par D. T. Pham et P. Garat [53], qui proposent d'approximer le **score**  $\psi$  par une combinaison linéaire :

$$\phi_{\alpha}(\mathbf{s}) \triangleq \sum_{i=1}^L \alpha_i f_i(\mathbf{s}) , \quad (1.17)$$

où  $f_1, \dots, f_L$  est une famille de fonctions fixée. L'intérêt de cette approche est que la combinaison linéaire qui minimise l'erreur quadratique moyenne  $\mathbb{E}[(\phi_{\alpha}(\mathbf{s}) - \psi(\mathbf{s}))^2]$  entre le vrai **score** et son approximation ne dépend pas explicitement de  $\psi$ . Cette approximation optimale est donnée par :

$$\phi_{opt}(\mathbf{s}) = \mathbb{E}[\mathbf{f}'(\mathbf{s})]^T \mathbb{E}[\mathbf{f}(\mathbf{s})\mathbf{f}(\mathbf{s})^T]^{-1} \mathbf{f}(\mathbf{s}) , \quad (1.18)$$

où  $\mathbf{f}(\mathbf{s}) \triangleq (f_1(\mathbf{s}), \dots, f_L(\mathbf{s}))^T$  est le vecteur colonne des fonctions de base, et  $\mathbf{f}'(\mathbf{s})$  est le vecteur colonne de leurs dérivées. L'expression (1.18) peut en outre être facilement estimée en remplaçant l'opérateur d'espérance mathématique par celui d'espérance empirique. Il existe d'autres méthodes dont le principe est de remplacer le **score** non-linéaire  $\psi$  par des fonctions non-linéaires choisies de manière ad hoc.

### 1.3 La séparation de sources audio

La séparation de sources audio est depuis quelques années déjà l'objet de nombreux travaux. Le grand intérêt qu'elle suscite est dû non seulement à ses multiples applications, mais aussi aux problèmes scientifiques originaux qu'elle soulève. Parmi les applications de la séparation de sources audio, on peut citer notamment la séparation de sources musicales visant une restitution haute qualité, la

séparation du signal utile du bruit environnant en téléphonie mobile, mais aussi la séparation à des fins d'indexation dans le domaine du multimedia, ainsi que les applications de reconnaissance de la parole en "cocktail party" ou la localisation de sources pour l'analyse de scènes auditives. D'une application à l'autre, la difficulté de séparation varie en fonction du nombre de sources par rapport au nombre de capteurs, du niveau de bruit, de la dépendance entre les sources, de la nature du mélange, etc. Par exemple, il est a priori plus difficile de séparer deux instruments à cordes enregistrés en mono et jouant en harmonie et en synchronie que deux locuteurs ayant des conversations indépendantes et enregistrés à l'aide d'un réseau de microphones. Ainsi, le cadre audio offre une grande palette de problèmes et une large variété de niveaux de difficulté au sein d'un même problème. Les méthodes de résolution sont donc diverses, chacune étant souvent spécifique à un type de problème et généralement évaluée avec des signaux tests et des critères choisis selon l'application. On conçoit donc qu'il est difficile de savoir à quelle catégorie de problèmes et à quel niveau de difficulté chaque méthode peut s'appliquer, ce qui rend difficile la comparaison de plusieurs méthodes.

Dans ce chapitre nous allons présenter quelques méthodes de séparation de sources suivant les caractéristiques des mélanges. On commencera par le cas des mélanges instantanés dans la section 1.3.1 et dans la section 1.3.2 on verra quelques méthodes pour le cas des mélanges convolutifs.

### 1.3.1 Mélanges instantanés

On suppose dans cette section que le mélange est sur-déterminé, c'est à dire que le nombre de capteurs est supérieur ou égal au nombre de sources  $M \geq N$ . Ce cas a été largement étudié et on peut décomposer les méthodes existant en trois classes :

- méthodes basées sur les statistiques d'ordre supérieur (sources *i.i.d*),
- méthodes basées sur la cohérence temporelle des signaux sources,
- méthodes basées sur la non-stationnarité des signaux sources.

Pour les méthodes basées sur les statistiques d'ordre supérieur, on utilise l'ACI décrite dans la section 1.2.2, et suivant différentes fonctions de contrastes (maximum de vraisemblance, minimum d'entropie, minimum d'information mutuelle, infomax etc).

## Méthodes basées sur la cohérence temporelle

Dans le cas où les signaux sources possèdent une cohérence temporelle, L. Féty [54] et L. Tong [55] montrent qu'il est possible de séparer les différents signaux en utilisant les matrices d'intercovariance des signaux capteurs. Ces matrices possèdent une structure simple qui permet de résoudre le problème de l'identification aveugle par la procédure de décomposition propre. Dans cette section, on va décrire une technique d'identification aveugle basée sur une diagonalisation conjointe d'un ensemble de matrices d'intercovariance [56]. La robustesse est significativement accrue en utilisant un ensemble de matrices plutôt qu'une seule matrice comme dans [54, 55].

Les signaux sources sont supposés stationnaires au second ordre, de moyennes nulles, de puissances finies et mutuellement décorrélés. On note  $\rho_i(\tau)$  la suite des coefficients de covariance du processus  $s_i(t)$  :

$$\mathbb{E}[s_i(t)s_j(t')^*] = \rho_i(t-t')\delta_{ij} \quad \forall t, t', \quad (1.19)$$

où  $\delta_{ij}$  désigne le symbole de Kronecker.

Le bruit additif  $w(t)$  est modélisé par un processus aléatoire complexe stationnaire, temporellement blanc, de moyenne nulle et de matrice de covariance  $\mathbf{R}_w$  :

$$\mathbb{E}[w(t+\tau)w(t)^H] = \delta(\tau)\mathbf{R}_w. \quad (1.20)$$

Ce bruit additif est supposé décorrélé des signaux sources.

Sous ces hypothèses, les matrices d'intercovariance ont la structure simple suivante :

$$\mathbf{R}_x(0) = \mathbb{E}[\mathbf{x}(t)\mathbf{x}(t)^H] = \mathbf{A}\mathbf{R}_s(0)\mathbf{A}^H + \mathbf{R}_w \quad (1.21)$$

$$\mathbf{R}_x(\tau) = \mathbb{E}[\mathbf{x}(t+\tau)\mathbf{x}(t)^H] = \mathbf{A}\mathbf{R}_s(\tau)\mathbf{A}^H \quad \tau \neq 0, \quad (1.22)$$

où  $\mathbf{R}_s(\tau) \triangleq \mathbb{E}[\mathbf{s}(t+\tau)\mathbf{s}(t)^H]$  est la matrice d'intercovariance des signaux sources, diagonale grâce à l'équation (1.19), ayant pour éléments diagonaux les  $\rho_i(\tau)$  pour  $1 \leq i \leq N$ . Dans ce qui suit, on se propose de résoudre le problème de l'identification aveugle en utilisant uniquement les matrices d'intercovariance  $\mathbf{R}_x(\tau)$  des observations.

Soit  $\mathbf{W}$  une matrice de blanchiment complexe de dimension  $N \times M$  qui, appliquée aux signaux capteurs  $\mathbf{x}$ , permet de se ramener à un mélange unitaire des signaux sources :

$$\mathbf{W}\mathbf{A} = \mathbf{U}. \quad (1.23)$$

Comme on l'a vu dans la section 1.2.2, la matrice de blanchiment peut être déterminée à partir de la matrice de covariance des observations. On note qu'il est ici possible de blanchir avec une combinaison linéaire définie positive des matrices d'intercovariance prises à des retards différents de zéro. Cette façon de procéder permet de s'affranchir de toute modélisation du bruit.

Soient maintenant les matrices d'intercovariance blanchies  $\underline{\mathbf{R}}_x(\tau)$  définies par :

$$\underline{\mathbf{R}}_x(\tau) = \mathbf{U} \mathbf{R}_s(\tau) \mathbf{U}^H \quad \forall \tau \neq 0, \quad (1.24)$$

Comme  $\mathbf{U}$  est unitaire et  $\mathbf{R}_s(\tau)$  diagonale, la relation (1.24) signifie que les matrices d'intercovariance blanchies se diagonalisent sous la même transformation unitaire  $\mathbf{U}$ . Un principe simple d'identification aveugle consiste alors en une diagonalisation conjointe des matrices d'intercovariance blanchies. On trouve cette idée sous différentes formes dans [54, 55].

Rappelons que si les valeurs propres d'une matrice sont uniques, il n'en est pas de même pour les vecteurs propres. Pour les valeurs propres distinctes, les vecteurs propres normés sont déterminés à une phase et une permutation près. Comme les vecteurs propres des matrices d'intercovariance blanchies sont les colonnes de la matrice  $\mathbf{U}$ , on trouve les mêmes indéterminations que dans le problème de la séparation de sources.

La diagonalisation d'une seule matrice  $\underline{\mathbf{R}}_x(\tau)$  résout le problème de la séparation de sources si ses valeurs propres sont distinctes : les vraies indéterminations apparaissent dans le cas de valeurs propres dégénérées. A priori, il n'est pas possible de déterminer un retard  $\tau$  pour lequel les valeurs propres de  $\underline{\mathbf{R}}_x(\tau)$  soient distinctes. Si les sources ont des spectres de formes différentes, la dégénérescence des valeurs propres est peu probable. Mais ce problème n'est pas purement académique car il est évident que quand les valeurs propres de  $\underline{\mathbf{R}}_x(\tau)$  sont proches de la dégénérescence, la robustesse de la détermination de  $\mathbf{U}$  à partir d'une décomposition propre est sérieusement affectée.

La situation est plus favorable si l'on considère la diagonalisation simultanée d'un ensemble  $\{\underline{\mathbf{R}}_x(\tau_i) | i = 1, \dots, K\}$  de  $K$  matrices d'intercovariance blanchies. Celles-ci sont simultanément diagonalisées par  $\mathbf{U}$  comme il est montré par la relation (1.24). Ainsi, au lieu de diagonaliser une seule matrice ou une combinaison linéaire de matrices comme le font respectivement L. Féty [54] et L. Tong [55], on détermine la rotation manquante en diagonalisant conjointement un ensemble de matrices d'intercovariance blanchies dans une même base  $\mathbf{U}$ . La condition sur la détermination de la matrice  $\mathbf{U}$  est donnée par le lemme suivant [56] :

**Lemme 1.1** *Supposons qu'il existe une matrice unitaire  $U = [\mathbf{u}_1, \dots, \mathbf{u}_N]$  qui diagonalise simultanément  $K$  matrices  $M_k$ ,  $k = 1, \dots, K$  de dimension  $N \times N$ ;  $U^H M_k U = \text{diag}(d_1(k), \dots, d_N(k))$ , pour tout  $k$ . La matrice unitaire  $U$  sera unique à une permutation et une phase près de ses colonnes si et seulement si*

$$\forall i \neq j \quad \exists k \quad d_i(k) \neq d_j(k) . \quad (1.25)$$

Selon ce lemme, la matrice unitaire  $U$  qui diagonalise simultanément les matrices d'intercovariance blanchies est unique à une permutation et une phase près de ses colonnes si et seulement si pour toute paire  $(i, j)$  de sources, il existe au moins un retard  $\tau_l$  dans  $(\tau_1, \dots, \tau_K)$  tel que  $\rho_i(\tau_l) \neq \rho_j(\tau_l)$ . Il est clair que des sources ayant des spectres normalisés identiques, ne peuvent pas être séparées.

Nous considérons par la suite la diagonalisation conjointe de plusieurs matrices d'intercovariance. Cette approche permet de réduire la probabilité des cas d'indétermination et d'améliorer a priori l'efficacité statistique grâce à l'utilisation d'un ensemble plus large de statistiques.

En pratique, on ne dispose que de statistiques empiriques aussi bien pour l'estimation du blanchisseur que pour l'estimation des matrices d'intercovariance. Une diagonalisation conjointe exacte n'est pas possible à cause des erreurs d'estimation. Il faut donc utiliser une diagonalisation conjointe approchée. Pour cela, remarquons que diagonaliser une matrice  $M$  de taille  $N \times N$  est équivalent à trouver une base orthonormée  $\mathbf{V} = \{\mathbf{v}_i | i = 1, \dots, N\}$  telle que, dans cette base, la somme des modules carrés des termes non-diagonaux soit minimale. La norme d'une matrice étant constante sous transformation unitaire, ceci revient à trouver la base  $\mathbf{V}$  minimisant le critère :

$$\mathcal{C}(M, \mathbf{V}) \triangleq - \sum_i |\mathbf{v}_i^H M \mathbf{v}_i|^2 , \quad (1.26)$$

sur l'ensemble des matrices unitaires  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$ .

Pour un ensemble de matrices, on étend naturellement ce critère en définissant la diagonalisation conjointe approchée d'une famille  $\widetilde{\mathbf{M}} = \{M_k | i = 1, \dots, K\}$  de  $K$  matrices comme la minimisation du critère :

$$\mathcal{C}(\mathbf{V}) \triangleq \mathcal{C}(M_k, \mathbf{V}) = - \sum_{k,i} |\mathbf{v}_i^H M_k \mathbf{v}_i|^2 , \quad (1.27)$$

sous la même contrainte unitaire. L'algorithme de diagonalisation conjointe utilisé dans [56] est une généralisation de la technique de Jacobi [57] de la diagonalisation d'une seule matrice Hermitienne . L'idée originale de la diagonalisation

conjointe d'un ensemble de matrices Hermitiennes a été introduite par J. F. Cardoso et A. Souloumiac dans [58, 59]. Le schéma général de l'algorithme décrit ci-dessus qui est nommé **SOBI**, est le suivant [56] :

1. estimation du blanchisseur  $\widehat{\mathbf{W}}$ ,
2. blanchiment des données :  $\widehat{\mathbf{x}}(t) = \widehat{\mathbf{W}}\mathbf{x}(t)$ ,
3. estimation de  $K$  matrices d'intercovariance  $\widehat{\mathbf{R}}_x(\tau_k)$  de  $\widehat{\mathbf{x}}(t)$  à différents retards  $\tau_k$ ,  $k = 1, \dots, K$ ,
4. diagonalisation conjointe approchée des matrices d'intercovariance dans une base  $\widehat{\mathbf{U}} = [\widehat{\mathbf{u}}_1, \dots, \widehat{\mathbf{u}}_N]$ ,
5. estimation de la matrice de mélange par :  $\widehat{\mathbf{A}} = \widehat{\mathbf{W}}\# \widehat{\mathbf{U}}$ ,
6. estimation des signaux sources par :  $\widehat{\mathbf{s}}(t) = \widehat{\mathbf{A}}^H \widehat{\mathbf{R}}_x(0)^{-1} \widehat{\mathbf{x}}(t)$ .

### Méthodes basées sur la non-stationnarité des signaux sources

Nous nous intéressons dans cette section à la séparation de sources non-stationnaires, c'est-à-dire dont les statistiques varient au cours du temps. Nous verrons dans cette section que la méthode de séparation présentée dans la section précédente et donc construite sur l'hypothèse de stationnarité des sources, peut donner des résultats acceptables sur des signaux non-stationnaires. Cependant, nous allons voir que les performances de séparation peuvent être accrues en utilisant le caractère non-stationnaire des sources.

Supposons que les sources soient modélisées par des séquences non-stationnaires blanches mutuellement décorrélées. On note :

$$\mathbf{R}_s(t, \tau) \triangleq \mathbb{E}[\mathbf{s}(t + \tau)\mathbf{s}(t)^H]. \quad (1.28)$$

Le fait que les sources soient blanches implique que  $\mathbf{R}_s(t, \tau) = \mathbf{0}$  pour  $\tau \neq 0$ , et on note plus simplement :

$$\mathbf{R}_s(t, 0) = \mathbf{R}_s(t) = \text{diag}(\sigma_1^2(t), \dots, \sigma_N^2(t)), \quad (1.29)$$

où  $\sigma_i^2(t)$  est la variance de la  $i^{\text{e}}$  source à l'instant  $t$ . Dans le cas non-bruité, on a :

$$\mathbf{R}_x(t) = \mathbf{A}\mathbf{R}_s(t)\mathbf{A}^H. \quad (1.30)$$

Disposant d'une matrice d'orthonormalisation  $\mathbf{W}$ , une méthode simple de séparation, inspirée de SOBI (section 1.3.1), consisterait à diagonaliser simultanément un ensemble de matrices  $\underline{\mathbf{R}}_x(t) = \mathbf{W}\mathbf{R}_x(t)\mathbf{W}^H$  pour différents instants

$t_1, \dots, t_K$ . En supposant les observations  $\mathbf{x}(t)$  stationnaires par morceaux, une estimation de  $\mathbf{R}_x(t)$  pourrait être réalisée par une moyenne temporelle locale, approche évoquée dans [60].

D. T. Pham et J. F. Cardoso [33] ont montré dans le cas non-bruité, réel et déterminé ( $M = N$ ) que cette approche peut être connectée à l'estimateur du maximum de vraisemblance de  $\mathbf{A}^{-1}$  en modélisant les sources par des séquences Gaussiennes stationnaires par morceaux. Les auteurs montrent en outre que dans le cas  $M = N$  un algorithme de diagonalisation de matrices Hermitiennes positives permet d'estimer  $\mathbf{A}$  sans avoir recours à une orthonormalisation préalable.

Supposons donc que les sources et la matrice de mélange sont à valeurs réelles avec  $M = N$ . En outre, la densité de chaque échantillon  $s(t)$  des sources est supposée Gaussienne, c'est à dire :

$$\forall \mathbf{u}, \quad p_{s(t)}(\mathbf{u}) = \frac{1}{\sqrt{\det(2\pi \mathbf{R}_s(t))}} \exp \left\{ -\frac{1}{2} \mathbf{u}^T \mathbf{R}_s(t)^{-1} \mathbf{u} \right\}. \quad (1.31)$$

Sous l'hypothèse de gaussianité, les sources supposées blanches et mutuellement décorrélées deviennent indépendamment distribuées et mutuellement indépendantes. Ceci va nous permettre d'écrire simplement le critère du maximum de vraisemblance associé à toutes ces hypothèses. D'après l'équation (1.31), on a :

$$\begin{aligned} \ln p_{s(t)}(\mathbf{u}) &= -\frac{1}{2} \ln \det(2\pi \mathbf{R}_s(t)) - \frac{1}{2} \mathbf{u}^T \mathbf{R}_s(t)^{-1} \mathbf{u} \\ &= -\frac{1}{2} \ln \det(2\pi \mathbf{R}_s(t)) - \frac{1}{2} \text{trace}(\mathbf{R}_s(t)^{-1} \mathbf{u} \mathbf{u}^T), \end{aligned}$$

avec  $\mathbf{x}(t) = \mathbf{A} \mathbf{s}(t)$ , la densité de probabilité de  $\mathbf{x}(t)$  est :

$$p_{\mathbf{x}(t)}(\mathbf{u}) = |\det(\mathbf{A})|^{-1} p_{s(t)}(\mathbf{A}^{-1} \mathbf{u}), \quad (1.32)$$

d'où :

$$\begin{aligned} \ln p_{\mathbf{x}(t)}(\mathbf{u}) &= -\ln \det(\mathbf{A}) + \ln p_{s(t)}(\mathbf{A}^{-1} \mathbf{u}) \\ &= -\ln \det(\mathbf{A}) - \frac{1}{2} \ln \det(2\pi \mathbf{R}_s(t)) - \frac{1}{2} \text{trace}(\mathbf{R}_s(t)^{-1} \mathbf{A}^{-1} \mathbf{u} \mathbf{u}^T \mathbf{A}^{-T}). \end{aligned}$$

Les sources étant supposées indépendamment distribuées et mutuellement indépendantes, les observations sont également indépendamment distribuées, d'où :

$$p_{\mathbf{x}}(\mathbf{x}) = \prod_{t=0}^{T-1} p_{\mathbf{x}(t)}(\mathbf{x}(t)), \quad (1.33)$$

et

$$\frac{1}{T} \ln p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{T} \sum_{t=0}^{T-1} \ln p_{\mathbf{x}(t)}(\mathbf{x}(t)) . \quad (1.34)$$

En posant  $\mathbf{B} = \mathbf{A}^{-1}$ , la log-vraisemblance de  $\mathbf{B}$  (et  $\mathbf{R}_s$ ) par rapport à  $\mathbf{x}$  s'écrit donc :

$$\mathcal{L}(\mathbf{B}, \mathbf{R}_s) = \ln \det(\mathbf{B}) - \frac{1}{2T} \sum_{t=0}^{T-1} \ln \det(2\pi \mathbf{R}_s(t)) + \text{trace}(\mathbf{R}_s(t)^{-1} \mathbf{B} \mathbf{x}(t) \mathbf{x}(t)^T \mathbf{B}^T) .$$

En notant  $\mathbf{y}(t) = \mathbf{B} \mathbf{x}(t)$ , le gradient relatif de  $\mathcal{L}(\mathbf{B}, \mathbf{R}_s)$  par rapport à  $\mathbf{B}$  s'écrit [33] :

$$\nabla_{\mathbf{B}} \mathcal{L}(\mathbf{B}, \mathbf{R}_s) = \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{R}_s(t)^{-1} \mathbf{y}(t) \mathbf{y}(t)^T - \mathbf{I}_N . \quad (1.35)$$

Optimiser la vraisemblance par annulation du gradient relatif revient à écrire :

$$\forall (i, j) \quad \frac{1}{T} \sum_{t=0}^{T-1} \frac{y_i(t) y_j(t)}{\sigma_i^2(t)} = \delta_{ij} . \quad (1.36)$$

Cette expression possède une analogie évidente avec l'expression du gradient relatif de la vraisemblance de  $\mathbf{B}$  dans le cas de sources *i.i.d.* Les termes non-diagonaux ( $i \neq j$ ) de cette équation matricielle expriment une forme de décorrélation moyenne entre les composantes de  $\mathbf{y}$ . Cependant nous n'avons plus ici besoin de réaliser la décorrélation non-linéaire des observations : ici les statistiques d'ordre deux suffisent, c'est la diversité temporelle qui est utilisée.

La résolution des équations données par (1.36) nécessite la connaissance a priori du profil de variance des sources  $\mathbf{R}_s$  qui n'est généralement évidemment pas disponible en pratique. Comme pour l'écriture du critère du maximum de vraisemblance dans le cas de sources *i.i.d.*, nous allons voir que le critère  $\mathcal{L}(\mathbf{B}, \mathbf{R}_s)$  peut être maximisé suivant  $\mathbf{B}$  et  $\mathbf{R}_s$  et, en supposant les sources stationnaires par morceaux, ramené à la diagonalisation simultanée d'un ensemble de covariances empiriques. Supposons que l'intervalle  $[0, T - 1]$  puisse être divisé en  $K$  intervalles  $I_1, \dots, I_K$  sur lesquels les sources sont stationnaires, c'est à dire, tels que :

$$\forall k \in [1, K], \quad \forall t \in I_k, \quad \mathbf{R}_s(t) = \mathbf{R}_s^{(k)} , \quad (1.37)$$

où  $\mathbf{R}_s^{(k)}$  est une matrice constante. En outre, on pose :

$$\widehat{\mathbf{R}}_x^{(k)} = \frac{1}{\#I_k} \sum_{t \in I_k} \mathbf{x}(t) \mathbf{x}(t)^T , \quad (1.38)$$

où  $\#I_k$  désigne le nombre d'éléments de  $I_k$ . On notera  $d_k = \#I_k/T$  la proportion d'éléments dans  $I_k$ . En remaniant l'expression de  $\mathcal{L}(\mathbf{B}, \mathbf{R}_s)$ , il a été montré dans



[61] que le critère du maximum de vraisemblance peut s'écrire comme une somme de divergences de Kullback entre des Gaussiennes de covariances  $\mathbf{R}_x^{(k)}$  et  $\widehat{\mathbf{R}}_x^{(k)}$ . En effet, on a :

$$\begin{aligned} \ln p_{\mathbf{x}(t)}(\mathbf{u}) &= -\ln \det(\mathbf{A}) - \frac{1}{2} \ln \det(2\pi \mathbf{R}_s(t)) - \frac{1}{2} \text{trace}(\mathbf{R}_s(t)^{-1} \mathbf{A}^{-1} \mathbf{u} \mathbf{u}^T \mathbf{A}^{-T}) \\ &= \frac{1}{2} [\ln \det(\mathbf{A}^{-T} \mathbf{R}_s(t)^{-1} \mathbf{A}^{-1}) - N \ln 2\pi - \text{trace}(\mathbf{R}_s(t)^{-1} \mathbf{A}^{-1} \mathbf{u} \mathbf{u}^T \mathbf{A}^{-T})] . \end{aligned}$$

On a donc d'après (1.33) :

$$\begin{aligned} \mathcal{L}(\mathbf{B}, \mathbf{R}_s) &= \\ &= -\frac{1}{2T} \sum_{k=1}^K \sum_{t \in I_k} \text{trace}(\mathbf{R}_s(t)^{-1} \mathbf{B} \mathbf{x}(t) \mathbf{x}(t)^T \mathbf{B}^T) - \ln \det(\mathbf{B}^T \mathbf{R}_s(t)^{-1} \mathbf{B}) + N \ln 2\pi \\ &= -\frac{1}{2} \sum_{k=1}^K d_k [\text{trace}(\mathbf{R}_s^{(k)-1} \mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T) - \ln \det(\mathbf{B}^T \mathbf{R}_s^{(k)-1} \mathbf{B}) + N \ln 2\pi] \end{aligned}$$

La divergence de Kullback  $\mathcal{K}_g(\mathbf{R}_1 \parallel \mathbf{R}_2)$  entre deux vecteurs aléatoires centrés Gaussiens de dimension  $N$  et de covariances  $\mathbf{R}_1$  et  $\mathbf{R}_2$  est donnée par [33] :

$$\mathcal{K}_g(\mathbf{R}_1 \parallel \mathbf{R}_2) = \frac{1}{2} [\text{trace}(\mathbf{R}_2^{-1} \mathbf{R}_1) - \ln \det(\mathbf{R}_2^{-1} \mathbf{R}_1) - N] . \quad (1.39)$$

Le critère du maximum de vraisemblance s'écrit alors :

$$\mathcal{L}(\mathbf{B}, \mathbf{R}_s) = - \sum_{k=1}^K d_k \mathcal{K}_g(\mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T \parallel \mathbf{R}_s^{(k)}) . \quad (1.40)$$

Si  $\mathbf{R}$  est une matrice positive et  $\mathbf{D}$  est une matrice positive diagonale alors :

$$\mathcal{K}_g(\mathbf{R} \parallel \mathbf{D}) = \mathcal{K}_g(\mathbf{R} \parallel \text{diag}(\mathbf{R})) + \mathcal{K}_g(\text{diag}(\mathbf{R}) \parallel \mathbf{D}) , \quad (1.41)$$

d'où

$$\mathcal{L}(\mathbf{B}, \mathbf{R}_s) = - \sum_{k=1}^K d_k \{ \mathcal{K}_g(\mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T \parallel \text{diag}(\mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T)) + \mathcal{K}_g(\text{diag}(\mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T) \parallel \mathbf{R}_s^{(k)}) \} .$$

Il est alors facile de maximiser  $\mathcal{L}(\mathbf{B}, \mathbf{R}_s)$  par rapport aux covariances des sources. Pour une valeur fixée de  $\mathbf{B}$ , il suffit de choisir  $\forall k \in [1, K]$ ,  $\mathbf{R}_s^{(k)} = \text{diag}(\mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T)$ . Le critère de vraisemblance ainsi maximisé par rapport aux covariances des sources devient :

$$\mathcal{L}(\mathbf{B}) = - \sum_{k=1}^K d_k \mathcal{K}_g(\mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T \parallel \text{diag}(\mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T)) , \quad (1.42)$$

en notant :

$$\text{off}_2(\mathbf{M}) \triangleq \mathcal{K}_g(\mathbf{M} \parallel \text{diag}(\mathbf{M})) ,$$

on a :

$$\mathcal{L}(\mathbf{B}) = - \sum_{k=1}^K d_k \{ \text{off}_2(\mathbf{B} \widehat{\mathbf{R}}_x^{(k)} \mathbf{B}^T) \} . \quad (1.43)$$

La quantité  $\text{off}_2$  est une mesure de diagonalité. En effet, d'après les propriétés de la divergence de Kullback,  $\text{off}_2(\mathbf{M})$  est positif, nul si et seulement si  $\mathbf{M} = \text{diag}(\mathbf{M})$ , c'est à dire si  $\mathbf{M}$  est diagonale. La maximisation du critère  $\mathcal{L}(\mathbf{B})$  est donc ramenée à la diagonalisation simultanée de l'ensemble de matrices positives Hermitiennes  $\widehat{\mathbf{R}}_x^{(k)}$  au sens de la mesure de diagonalité  $\text{off}_2$ . La méthode exposée ci-dessus est nommée **BGML** dans [33] pour "*Block-Gaussian Maximum Likelihood*" et son schéma général est le suivant :

1. division de l'intervalle  $[0, T - 1]$  en  $K$  intervalles adjacents,
2. calcul des  $K$  covariances locales  $\{\widehat{\mathbf{R}}_x^{(k)}, k \in [1, K]\}$ ,
3. diagonalisation simultanée, sans contrainte orthonormale, de l'ensemble  $\{\widehat{\mathbf{R}}_x^{(k)}, k \in [1, K]\}$ ,
4. la diagonalisation simultanée fournit directement une estimation  $\widehat{\mathbf{B}}$  de  $\mathbf{A}^{-1}$ .

### 1.3.2 Mélanges convolutifs

Les premiers efforts sur la séparation de sources de mélanges convolutifs ont été fait dans le domaine temporel [62]. Inspiré par les méthodes de déconvolution aveugle, K. Torkkola [63] a modélisé le procédé de séparation comme un filtre RIF

$$\mathbf{y}(t) = \sum_{k=0}^{\kappa} \mathbf{G}(k) \mathbf{x}(t - k) , \quad (1.44)$$

dont il faut estimer les coefficients  $g_{ij}$ . Pour résoudre ce problème, il opta pour une méthode de maximisation de l'information mutuelle [64]. T. W. Lee et al. [65] ont modélisé la procédure de séparation comme un filtre IIR en supposant que l'environnement de l'enregistrement est à phase minimale, ce qui n'est pas toujours valable.

Mukai et al. [66–68] ont étudié le problème de séparation de mélanges convolutifs en tenant compte des propriétés acoustiques des signaux. Ils traitent le problème en fréquence en appliquant une transformée de Fourier à court terme (TFCT) :

$$\mathcal{S}_x(t, f) = \mathbf{H}(f) \mathcal{S}_s(t, f) + \mathcal{S}_w(t, f) , \quad (1.45)$$

où  $\mathcal{S}_x(t, f)$  est le vecteur temps-fréquence (TF) du mélange,  $\mathcal{S}_s(t, f)$  est le vecteur TF des signaux sources et  $\mathbf{H}(f)$  est la transformée de Fourier du filtre de convo-

lution. En appliquant une matrice de séparation  $\mathbf{G}(f)$  pour chaque fréquence

$$\mathcal{S}_y(t, f) = \mathbf{G}(f)\mathcal{S}_x(t, f) , \quad (1.46)$$

tel que  $\mathbf{G}(f)$  est calculée d'une manière itérative en utilisant

$$\mathbf{G}(f)^{(p+1)} = \mathbf{G}(f)^{(p)} + \mu[\mathbf{I}_N - \mathbf{f}(\mathcal{S}_y)\mathcal{S}_y^H]\mathbf{G}(f)^{(p)} , \quad (1.47)$$

où  $\mathbf{f}(\cdot)$  est une fonction non-linéaire.

P. Smaragdis [69, 70] a aussi proposé de travailler uniquement dans le domaine temps-fréquence, pour transformer le produit de convolution en multiplication, donc le problème est défini en tant qu'estimation de la matrice de séparation pour chaque trame de fréquence qui peut avoir une solution au sens du maximum de vraisemblance (l'algorithme du gradient naturel peut être utilisé pour résoudre ce problème, bien qu'il soit dérivé d'un principe différent). Malheureusement, la résolution du problème indépendamment pour chaque trame de fréquence produit des problème de permutation, puisqu'il y'a l'ambiguïté inhérente de permutation des lignes de la matrices de séparation. Pour résoudre ce problème P. Smaragdis a proposé un arrangement adaptatif pour appliquer un certain accouplement des fréquences pour les trames voisines.

L. Parra et al. [60] ont également travaillé dans le domaine temps-fréquences en utilisant la non-stationnarité pour effectuer la séparation, leur solution au problème était d'imposer une contrainte à la longueur du filtre de séparation. Ceci est réalisé en appliquant un opérateur de projection aux évaluation du filtre à chaque itération. Vu l'hypothèse de non-stationnarité des sources, la matrice de covariance est exprimée en fonction de l'indice  $\tau$  :

$$\mathbf{R}_x(\tau, f) = \mathbf{H}(f)\mathbf{R}_s(\tau, f)\mathbf{H}^H(f) + \mathbf{R}_w(\tau, f) , \quad (1.48)$$

en exploitant l'hypothèse d'indépendance qui implique la décorrélation, les matrices  $\mathbf{R}_s$  et  $\mathbf{R}_w$  sont diagonales. Ainsi le problème se ramène à une diagonalisation conjointe des matrices de même fréquence :

$$\mathbf{R}_s(\tau, f) = \mathbf{G}(f)(\mathbf{R}_x(\tau, f) - \mathbf{R}_w(\tau, f))\mathbf{G}^H(f) , \quad (1.49)$$

le critère de diagonalisation utilisé est adaptatif et utilise la méthode des moindres carrés. Enfin, pour résoudre le problème de permutation, les auteurs imposent la contrainte de continuité ("lissage") dans le domaine des fréquences. Une autre solution est d'employer le *Beamforming* (formation de voie) [71] pour aligner les données le long de l'axe de fréquence en utilisant les directions d'arrivée du signal.

H. Bousbia et al. [72] ont utilisé la non-stationnarité des signaux comme suit. Soit  $\tilde{\mathbf{x}}(t) = [\mathbf{x}(t)^T \dots \mathbf{x}(t-d)^T]^T$  tel que :

$$\tilde{\mathbf{x}}(t) = \mathbf{A}_d \tilde{\mathbf{s}}(t) + \tilde{\mathbf{w}}(t), \quad (1.50)$$

où  $\tilde{\mathbf{s}}(t) = [\mathbf{s}(t)^T \dots \mathbf{s}(t-d-L)^T]^T$ . Il est montré dans [72] que les matrices d'auto-covariance (sans bruit) du signal  $\tilde{\mathbf{x}}(t)$  s'écrivent sous la forme

$$\mathbf{R}_{\tilde{\mathbf{x}}}(t) = \mathbf{A}_d \mathbf{R}_{\tilde{\mathbf{s}}}(t) \mathbf{A}_d^H, \quad (1.51)$$

où  $\mathbf{R}_{\tilde{\mathbf{s}}}(t)$  est bloc diagonal. Ainsi, trouver une matrice de séparation qui inverserait  $\mathbf{A}_d$  revient à bloc-diagonaliser les matrices  $\mathbf{R}_{\tilde{\mathbf{x}}}(t)$  pour  $t = t_1, \dots, t_K$ . Ceci est une généralisation de la méthode de séparation de signaux non-stationnaires par diagonalisation conjointe proposée par D. T. Pham et J. F. Cardoso dans [33].

## 1.4 Conclusion

Comme on l'a vu dans ce chapitre, le problème de séparation de sources est arrivé à une certaine maturité du moins pour le cas de mélanges instantanés. De nombreuses méthodes ont été proposées pour résoudre ce dernier se basant sur les statistiques d'ordre supérieur. Pour la suite de cette thèse cette piste ne sera pas envisagée en priorité vu le coût de calcul, la complexité et la faible vitesse de convergence de ces méthodes.

Comme cité précédemment cette étude se déroule dans le cadre de la séparation de sources audio. En regard des caractéristiques du signal audio et plus spécialement ceux du signal de parole, nous proposons d'exploiter celles-ci en se concentrant sur des méthodes qui se basent sur les propriétés suivantes :

- la non-stationnarité des signaux sources,
- la cohérence temporelle des signaux sources,
- le caractère parcimonieux des sources,
- le caractère localement harmonique des signaux sources.

En effet, de telles méthodes permettent de remédier aux points faibles des méthodes utilisant les statistiques d'ordre supérieur, notamment, elle permettent :

- de réduire sensiblement pour certaines de ces méthodes le coût de calcul de la séparation de sources,
- d'améliorer les performances de la séparation en utilisant des informations spécifiques supplémentaires, relatives aux signaux sources considérés,

- améliorer les vitesses de convergence (réaliser la séparation avec des tailles plus réduites des observations),
- de permettre la séparation de plus de sources que de capteurs.



# La séparation de sources audio sous-déterminée

---

## 2.1 Introduction

Ce chapitre est consacré à l'étude d'un problème plus délicat, celui de la séparation aveugle de sources où l'on dispose de moins de capteurs que de sources. Ce cas a été moins étudié que le cas carré ou le cas sur-déterminé, mais ces dernières années ont vu apparaître un intérêt croissant dans ce domaine.

Le problème de séparation de sources n'est pas équivalent à celui de l'estimation de la matrice de mélange dans le cas sous-déterminé. En effet, même si l'on connaît la matrice  $A$ , il existe un espace infini de solutions à l'équation  $x(t) = As(t)$  (la matrice  $A$  n'est pas inversible). Afin de contourner ce problème on a recours à une information a priori forte sur la nature des sources que l'on veut séparer. La plupart des méthodes proposées pour le cas sous-déterminé utilisent comme information a priori le caractère parcimonieux des sources. Ceci revient à supposer qu'il existe une base de fonctions (un domaine transformé) où la représentation des sources est creuse.

Le but principal de ce chapitre est d'introduire la problématique de la SAS sous-déterminée et ses motivations, puis de présenter certaines solutions et techniques qui ont été proposées dans la littérature afin de mieux introduire nos contributions dans la suite du document.

## 2.2 Mélanges instantanés

Afin de résoudre le problème de la séparation aveugle de sources dans le cas sous-déterminé, plusieurs méthodes ont été proposées. Ces approches qui traitent en générale le cas instantané peuvent être décomposées en deux classes :

- méthodes basées sur un modèle probabiliste des signaux sources,
- méthodes basées sur une représentation parcimonieuse des signaux sources dans un domaine transformé.

### 2.2.1 Méthodes basées sur un modèle probabiliste

Parmi les travaux pionniers dans la SAS sous-déterminée, ceux de P. Comon et O. Grellier [73] qui ont proposé une méthode d'extraction de sources dans le cas de sources complexes et discrètes, en exploitant leurs distributions particulières. Cependant d'autres méthodes ont été proposées, utilisant des approximations de la densité de probabilité des signaux sources [74].

Dans le cas de mélanges sous-déterminés, le mélange n'est plus séparable par inversion de la matrice de mélange. Il faut estimer conjointement la matrice de mélange et les sources qui permettent de reconstruire le mélange, éventuellement avec une faible erreur résiduelle. O. Bermond a proposé d'utiliser un algorithme de type *Expectation-Maximization* (EM) [51, 75] pour l'estimation de la matrice de mélange ainsi que des sources, au sens du maximum de vraisemblance. En effet, si l'on considère que la matrice de mélange et la covariance du bruit sont les paramètres à estimer, alors on a affaire à des observations incomplètes car les sources sont inconnues. La vraisemblance des données observées connaissant les paramètres  $\mathbf{A}$  et  $\mathbf{R}_w$  est :

$$p(\mathbf{x}|\mathbf{A}, \mathbf{R}_w) = \frac{1}{(2\pi)^{N/2} \sqrt{|\det \mathbf{R}_w|}} \int_{\mathbf{s}} r(\mathbf{s}) \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{A}\mathbf{s})^T \mathbf{R}_w^{-1} (\mathbf{x} - \mathbf{A}\mathbf{s}) \right\} d\mathbf{s} ,$$

où  $r(\mathbf{s})$  est la densité a priori des sources. Dans le cas de sources indépendantes, cette densité est factorisable. On a utilisé ici la loi des observations, sachant la matrice de mélange et la covariance du bruit. Cette loi est marginale par rapport aux sources, d'où l'intégrale dans l'expression précédente. En effet, la densité des données complètes  $(\mathbf{x}, \mathbf{s})$ , s'écrit :

$$p(\mathbf{x}, \mathbf{s}|\mathbf{A}, \mathbf{R}_w) = \frac{1}{(2\pi)^{N/2} \sqrt{|\det \mathbf{R}_w|}} r(\mathbf{s}) \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{A}\mathbf{s})^T \mathbf{R}_w^{-1} (\mathbf{x} - \mathbf{A}\mathbf{s}) \right\} .$$



L'algorithme EM permet de trouver un maximum local pour la densité marginale, en itérant deux étapes :

- une étape E (*Expectation*) : on calcule l'espérance de fonctions (statistiques exhaustives) des données non observées,
- une étape M (*Maximization*) : on estime les nouvelles valeurs des paramètres comme si les statistiques sur les données complètes étaient observées.

En langage mathématique, cela s'écrit :

- étape E : calculer  $Q(\theta, \theta^{(k)}) = \mathbb{E}[\ln p(\mathbf{x}, \mathbf{s}|\theta)|\mathbf{x}, \theta^{(k)}]$ ,
- étape M : maximiser  $\theta^{(k+1)} = \arg \max_{\theta} Q(\theta, \theta^{(k)})$ .

Dans le cas de la séparation de sources bruitées, l'étape M est simple. Après calculs, on obtient [51] :

$$\begin{aligned} \mathbf{A}^{(k+1)} &= \mathbf{R}_{xs}^{(k)} (\mathbf{R}_s^{(k)})^{-1} \\ \mathbf{R}_w^{(k+1)} &= \mathbf{R}_x - \mathbf{R}_{xs}^{(k)} (\mathbf{R}_s^{(k)})^{-1} (\mathbf{R}_{xs}^{(k)})^T \end{aligned}$$

où l'on a dénoté les moments empiriques :

$$\begin{aligned} \mathbf{R}_{xs}^{(k)} &= \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{x}(t) \mathbb{E}[\mathbf{s}(t)^T | \mathbf{x}, \mathbf{A}^{(k)}, \mathbf{R}_w^{(k)}] \\ \mathbf{R}_s^{(k)} &= \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\mathbf{s}(t) \mathbf{s}(t)^T | \mathbf{x}, \mathbf{A}^{(k)}, \mathbf{R}_w^{(k)}] . \end{aligned}$$

Toute la difficulté réside dans le calcul de ces moments (l'étape E). Une option serait de calculer les espérances par une méthode de Monte-Carlo. Cependant, cela devrait être fait à chaque étape, pour chaque index de temps  $t$ . En pratique, c'est beaucoup trop coûteux algorithmiquement. Au contraire, une option minimale consiste à remplacer les espérances sur  $\mathbf{s}(t)$  par la valeur au maximum de vraisemblance  $\hat{\mathbf{s}}(t)$  :

$$\hat{\mathbf{s}}(t)^{(k)} = \arg \max_{\mathbf{s}} \left\{ \ln r(\mathbf{s}) - \frac{1}{2} (\mathbf{x} - \mathbf{A}^{(k)} \mathbf{s})^T (\mathbf{R}_w^{(k)})^{-1} (\mathbf{x} - \mathbf{A}^{(k)} \mathbf{s}) \right\} . \quad (2.1)$$

Pour aller un peu plus loin, on peut utiliser, au premier ordre, une approximation de Laplace. On écrit pour cela le Hessien de la densité complète au point  $\hat{\mathbf{s}}(t)$  :

$$\begin{aligned} \mathcal{H}(\mathbf{s}) &= -\frac{\partial^2}{\partial \mathbf{s}^2} \ln p(\mathbf{x}, \mathbf{s} | \mathbf{A}, \mathbf{R}_w) \\ &= \mathbf{A}^T \mathbf{R}_w^{-1} \mathbf{A} - \frac{\partial^2}{\partial \mathbf{s}^2} r(\mathbf{s}) . \end{aligned}$$

On en déduit alors l'approximation Gaussienne autour du maximum  $\hat{\mathbf{s}}$ , suivante :

$$p(\mathbf{s} | \mathbf{A}, \mathbf{R}_w, \mathbf{x}) \approx \frac{1}{(2\pi)^{M/2} \sqrt{|\det(\mathcal{H}(\hat{\mathbf{s}}))|}} \exp \left\{ -\frac{1}{2} (\mathbf{s} - \hat{\mathbf{s}})^T \mathcal{H}(\hat{\mathbf{s}}) (\mathbf{s} - \hat{\mathbf{s}}) \right\} ,$$

d'où finalement le résultat :

$$\begin{aligned}\mathbb{E}[\mathbf{s}(t)^T | \mathbf{x}, \mathbf{A}^k, \mathbf{R}_w^k] &= \hat{\mathbf{s}} \\ \mathbb{E}[\mathbf{s}(t)\mathbf{s}(t)^T | \mathbf{x}, \mathbf{A}^k, \mathbf{R}_w^k] &= \hat{\mathbf{s}}\hat{\mathbf{s}}^T + \mathcal{H}(\hat{\mathbf{s}})^{-1} .\end{aligned}$$

Une autre option pour estimer les moments empiriques consiste à modéliser la densité a priori sur les sources  $r(\mathbf{s})$  par un modèle de mélange de Gaussiennes (modèle paramétrique). En effet, dans le cas Gaussien, tel qu'on l'a vu précédemment, les calculs sont simples et s'étendent naturellement aux mélanges de Gaussiennes dans un formalisme de données incomplètes [76].

## 2.2.2 Méthodes basées sur une représentation parcimonieuse

L'autre classe de méthodes est celles qui se basent sur la représentation parcimonieuse des signaux sources. En d'autres termes, elles supposent que si les formes d'onde des sources sont modélisées par des décompositions parcimonieuses dans une base donnée, l'estimation consiste à réaliser un compromis entre la quantité d'erreur résiduelle et la parcimonie des coefficients de décomposition supposés indépendants. Les calculs sont effectués par des méthodes de type gradient naturel ou de Newton [77], par apprentissage bayésien variationnel [78, 79] ou par des méthodes de types Monte Carlo utilisant une représentation en chaînes de Markov [80].

La grande majorité des algorithmes existants ne considèrent que le cas du mélange instantané. T. W. Lee et al. [77] et F. J. Theis et al. [81] séparent des mélanges de parole en modélisant la parcimonie des formes d'onde temporelles. M. E. Davies [82] et C. Févotte [80] utilisent la base de Transformée en Cosinus Discrète Modifiée (TCDM) pour séparer des sources musicales.

Dans le cas particulier de mélanges stéréo, certains algorithmes proposent d'effectuer la séparation en deux étapes : estimation de la matrice de mélange, puis estimation des sources sachant la matrice de mélange [81, 83, 84]. La première étape ne repose pas sur le modèle probabiliste des sources. Elle consiste à décomposer le mélange dans la base où les sources sont parcimonieuses,

$$s_i(t) = \sum_{k=1}^{\mathcal{L}} C_{ik} \varphi_k(t) , \quad (2.2)$$

où  $C_{ik}$  est le coefficient associé à la fonction  $\varphi_k(t)$  qui appelée atome ou élément du dictionnaire, puis à définir une direction spatiale pour chaque atome. Les directions spatiales des sources sont alors calculées en étudiant les pics [83] ou les

champs réceptifs, de l'histogramme des directions spatiales [81], ce qui fournit la matrice de mélange (à un gain près sur chaque colonne). La deuxième étape

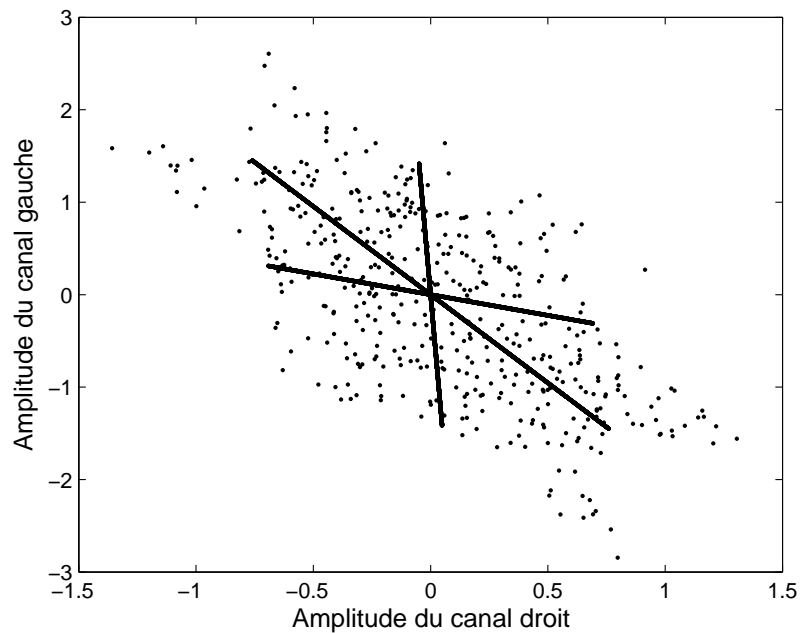


Fig. 2.1 – Amplitude du canal gauche en fonction de l'amplitude du canal droit pour un mélange stéréo de 3 sources parcimonieuses.

reste basée sur le modèle probabiliste, mais dans certains cas elle peut aussi se décrire à partir des informations spatiales. Par exemple, lorsque les coefficients de décomposition des sources sont supposés Laplaciens de même variance, les coefficients estimés sur un atome donné sont non nuls pour deux sources et nuls pour les autres. Ces deux sources sont celles dont les directions spatiales sont les plus proches de la direction spatiale observée sur cet atome [81]. La performance de ces algorithmes dépend du degré de parcimonie effectif des sources.

Prenons l'exemple d'un mélange stéréo de sources de distribution Laplacienne de même variance dans une base d'atomes temps-fréquence. Deux types d'erreurs existent lorsque les sources ne sont pas parfaitement disjointes. Premièrement les sources sont mal séparées aux points temps-fréquence contenant trois sources ou plus car seules deux sources en sont extraites. Deuxièmement, la séparation est mauvaise aussi aux points contenant deux sources périphériques (provenant de la gauche et de la droite) car ils sont attribués aux sources centrales (provenant du milieu) dont la direction spatiale est plus proche de celle observée [85].

D'autres distributions parcimonieuses diminuent ces problèmes sans les suppri-

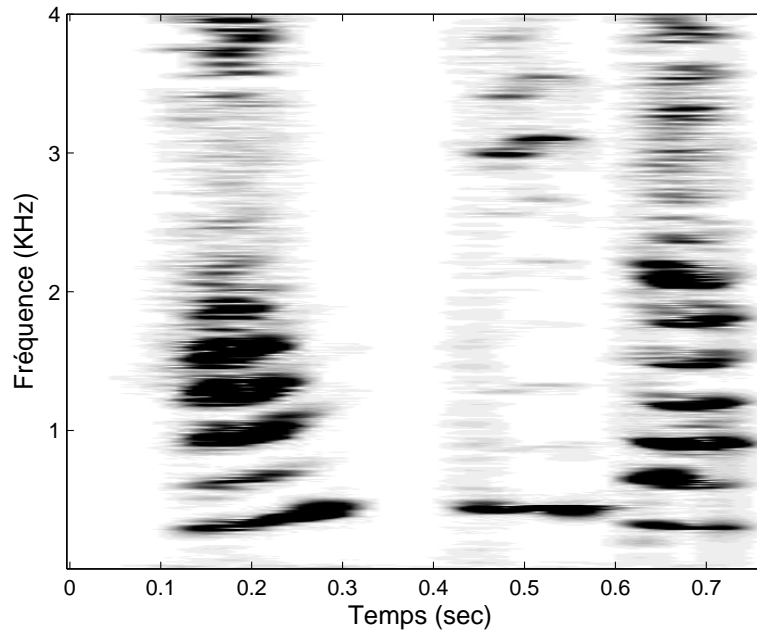


Fig. 2.2 – Représentation temps-fréquence d'un signal de parole : la représentation dans le domaine temps-fréquence permet d'illustrer la propriété de parcimonie des signaux audio.

mer [86]. En pratique la parcimonie limitée des sources musicales en temps-fréquence se manifeste par des artefacts sur les sources estimées [80] et par une dégradation de performance lorsque le nombre de sources augmente [83]. Plusieurs approches sont possibles pour améliorer les résultats. M. Zibulevsky et al. [83] augmentent la parcimonie des sources en utilisant un dictionnaire redondant au lieu d'une base. Vielva, Erdogan et Principe [86] modélisent chaque source par une distribution de variance appropriée connue a priori.

Une des méthodes exploitant la parcimonie des signaux audio dans le domaine temps-fréquence (TF), celle proposée par O. Yilmaz et al. [87] pour le cas des mélanges stéréo. Cette méthode dénommée DUET, utilise la transformée de Fourier à court terme (TFCT) et suppose qu'au plus une seule source est présente dans chaque point TF (les sources sont disjointes dans le domaine TF). Après la transformation dans le domaine TF utilisant la TFCT l'équation (1.1) dans le cas sans bruit, devient

$$\mathcal{S}_x(t, f) = \mathbf{A}\mathcal{S}_s(t, f), \quad (2.3)$$

où  $\mathcal{S}_x(t, f)$  est le vecteur TF du mélange de dimension  $M \times 1$ , et  $\mathcal{S}_s(t, f)$  est le vecteur TF des signaux sources de dimension  $N \times 1$ . Sous l'hypothèse que les sources sont

disjointes dans le domaine TF, l'équation (2.3) devient

$$\mathcal{S}_x(t, f) = \mathbf{a}_i \mathcal{S}_{s_i}(t, f), \quad \forall (t, f) \in \Omega_i, \forall i. \quad (2.4)$$

où  $\Omega_i$  représente le support TF de la source  $s_i$ . La méthode proposée dans [87] est adaptée pour le cas particulier d'une matrice de mélange de 2 lignes dont la première contient que des uns, alors on peut réécrire l'équation (2.3)

$$\begin{bmatrix} \mathcal{S}_{x_1}(t, f) \\ \mathcal{S}_{x_2}(t, f) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ a_{21} & \dots & a_{2N} \end{bmatrix} \begin{bmatrix} \mathcal{S}_{s_1}(t, f) \\ \vdots \\ \mathcal{S}_{s_N}(t, f) \end{bmatrix}, \quad (2.5)$$

par conséquent, l'équation (2.4) devient

$$\begin{bmatrix} \mathcal{S}_{x_1}(t, f) \\ \mathcal{S}_{x_2}(t, f) \end{bmatrix} = \begin{bmatrix} 1 \\ a_{2i} \end{bmatrix} \mathcal{S}_{s_i}(t, f),$$

il en résulte que

$$a_{2,i} = \frac{\mathcal{S}_{x_2}(t, f)}{\mathcal{S}_{x_1}(t, f)}. \quad (2.6)$$

Ainsi, tous les points dont le rapport donné par (2.6) donne la même valeur, appartiennent au même support TF  $\Omega_i$  de la source  $s_i(t)$ . Donc, l'estimée de la TFCT de  $s_i(t)$  peut être calculée par

$$\hat{\mathcal{S}}_{s_i}(t, f) = \begin{cases} \mathcal{S}_{x_1}(t, f), & \forall (t, f) \in \Omega_i, \\ 0, & \text{sinon.} \end{cases}$$

Finalement l'estimée de la source  $\hat{s}_i(t)$  est obtenue par transformation de  $\hat{\mathcal{S}}_{s_i}(t, f)$  au domaine temporel en utilisant la TFCT inverse.

Il existe cependant un cas particulier, qui est la séparation de sources utilisant un seul capteur [88]. Ce domaine a commencé à être actif depuis les années 2000. Une proportion importante de chercheurs à avoir abordé ce problème provient du domaine du traitement de la parole, notamment dans le cadre de la reconnaissance de la parole en milieu bruité. Par ailleurs, depuis que la séparation de sources est devenue un domaine à part entière en traitement du signal, en particulier grâce aux succès de la communauté ACI, quelques chercheurs se sont attaqués au problème spécifique de la séparation avec un seul capteur. Leurs approches sont différentes : adaptation de méthode de type ACI, utilisation de modèles prédictifs ou encore modélisation par mélange de lois sur des données spectrales.

L'approche en sous-espaces indépendants dans le domaine spectral rentre dans le cadre de ces méthodes. Cette approche *Independent Subspace Analysis* ou ISA se fonde sur une décomposition du spectrogramme du signal observé en sous-espaces indépendants. Le spectrogramme est décomposé en une somme d'atomes temps-fréquence séparables :

$$\mathcal{S}_{\mathbf{x}(t)}(t, f) = \mathbf{u}_1(f)\alpha_1(t) + \cdots + \mathbf{u}_N(f)\alpha_N(t) . \quad (2.7)$$

Cela revient en fait à décomposer le spectre en atomes fréquentiels de base  $\mathbf{u}_i(f)$  modulés dans le temps par une fonction d'activation  $\alpha_i(t)$ . M. A. Casey et W. Westner [89] utilisent une décomposition en valeurs singulières pour obtenir cette factorisation : dans ce cas, les atomes fréquentiels sont orthogonaux. Les auteurs regroupent ensuite les divers atomes pour former les sources et la reconstruction se fait par simple inversion du spectrogramme ou par projection.

## 2.3 Mélanges convolutifs

Dans le cas plus réaliste du mélange convolutif, le problème de SAS sous-déterminé est beaucoup plus délicat à résoudre [62]. De nombreuses applications de séparation de sources sont limitées par le nombre de microphones disponibles. Il n'est pas toujours garanti que le nombre de sources soient supérieure ou égal au nombre de capteurs. Avec les filtres linéaires, il n'est pas possible en général de séparer plus de  $(M - 1)$  sources du signal. En revanche, en employant des techniques non-linéaires, il est possible d'extraire un plus grand nombre de signaux sources. La grande majorité des techniques utilisées pour séparer plus de sources que de capteurs dans le cas convolutif est basée sur la parcimonie. Si les signaux sources ne se recouvrent pas (disjoints) dans le domaine temps-fréquence (TF), il est alors possible de les séparer. Un masque peut être appliqué dans le domaine TF pour atténuer l'énergie du signal d'interférence tout en préservant les points TF où le signal d'intérêt est dominant. Souvent un masque binaire est employé donnant des résultats perceptuels satisfaisants même pour des sources à faible taux de recouvrement [90, 91]. Ces méthodes fonctionnent bien sous le modèle convolutif simplifié, dit modèle anéchoïque qui est défini comme suit :

$$x_i(t) = \sum_{j=1}^N a_{ij}s_j(t - \delta_{ij}) + w_i(t) \quad \text{pour } i = 1, \dots, M , \quad (2.8)$$

où  $\delta_{ij}$  représente le délai nécessaire au signal issu de la  $j^{\text{e}}$  source pour atteindre le  $i^{\text{e}}$  capteur avec une atténuation  $a_{ij}$ . Cependant, sous la condition de réverbération,

la représentation des signaux en TF est moins parcimonieuse. Dans un environnement modérément réverbérant, la séparation de sources sous-déterminée peut être réalisée en combinant l'analyse en composantes indépendantes (ACI) et le masquage binaire en temps-fréquence [92]. Les  $(N - M)$  premiers signaux sont extraits du mélange en appliquant un masquage TF estimé à partir des directions d'arrivées des signaux sources. Les  $M$  sources restantes seront séparées par des techniques de SAS conventionnelles. Quand le masquage binaire est appliqué au signal, des artefacts (bruit musical) sont souvent introduits. Dans le but de réduire le bruit musical, un masquage dit doux a été introduit dans [92, 93].

La plupart des méthodes de masquage temps-fréquence n'utilisent pas efficacement l'information issue de plus de deux capteurs. Cependant, quelques méthodes proposent d'utiliser cette information supplémentaire [91, 94].

La classification a été également employée pour la séparation de sources parcimonieuses [94–99]. Si les sources sont projetées dans un espace où les atomes de chaque source sont regroupés ensemble, le problème de séparation de source peut être résolu en utilisant des algorithmes de classification. Dans [92] le masque temps-fréquence est déterminé en appliquant une classification sur les amplitudes et les retards.

Le cas très particulier de la SAS convolutive mono-canal peut être résolu à condition d'avoir d'importantes informations additionnelles. Dans [100], des informations a priori sur les signaux sources sont utilisées pour construire des fonctions de base pour chaque signal source, afin d'isoler des signaux musicaux à partir de leurs mélanges. De même, dans [101] la parcimonie des signaux audio a été exploitée dans le but d'extraire les différents instruments de musique présents dans le mélange.

## 2.4 Conclusion

Nous avons exposé dans ce chapitre le problème de séparation aveugle de sources dans le cas sous-déterminé. Nous avons présenté plusieurs méthodes proposées dans la littérature, destinées à ce cas de figure, utilisant les statistiques d'ordre supérieur ou la parcimonie des signaux audio.

Dans la suite de nos travaux, nous nous intéresserons uniquement aux méthodes exploitant la parcimonie des signaux audio et nous écarterons les méthodes basées sur les statistiques d'ordre supérieur pour les mêmes raisons que celles évoquées dans la section 1.4. Nous proposerons des techniques exploitant la

propriété de parcimonie dans différents domaines transformés (temps-fréquence, temps-échelle), et nous relâcherons l'hypothèse de non-recouvrement (sources disjointes) utilisée dans la plupart des méthodes exposées précédemment.



# Première partie

## La SAS audio utilisant la décomposition modale

---

Dans cette partie, nous présenterons des techniques de séparation aveugle de sources utilisant la décomposition modale. L'utilisation de la décomposition modale a été inspirée par le fait que les signaux audio et plus particulièrement les signaux musicaux peuvent être modélisés par une somme de sinusoides amorties. C'est cette propriété que nous proposons d'exploiter afin d'effectuer la séparation de sources dans le cas sous-déterminé. Par ailleurs, les travaux présentés dans cette partie ont donné lieu à une publication dans la revue EURASIP Journal on Audio, Speech & Music Processing.



# La SAS audio utilisant la décomposition modale : mélanges instantanés

---

## 3.1 Introduction

Dans ce chapitre, nous proposons une approche nommée MD-UBSS (*Modal Decomposition - Underdetermined Blind Source Separation*) pour la résolution du problème de la SAS sous-déterminée. Cette approche se base sur la décomposition modale des signaux observés [102, 103]. Plus précisément, on propose d'appliquer des techniques de décomposition modale à des signaux supposés *localement périodiques*, sans qu'ils soient nécessairement harmoniques au sens de Fourier. Les signaux audio et plus particulièrement les signaux musicaux peuvent être modélisés par une somme de sinusoides amorties [104, 105]. C'est cette propriété que nous proposons d'exploiter ici pour la séparation de sources audio par le biais de la décomposition modale.

Bien que nous considérons ici une application audio, la méthode proposée peut être employée pour n'importe quelle autre application où les signaux sources peuvent être représentés par une somme de composantes modales. Ceci inclut, en particulier la séparation de signaux NMR (*Nuclear Magnetic Resonance*) dans [106, 107] et les signaux de machine rotative dans [108].

Notons que cette représentation modale des sources est un cas particulier de la représentation parcimonieuse des sources, qui est souvent exploitée pour effectuer la séparation de sources dans le cas sous-déterminé [109]. En effet, un signal composé par une somme de sinusoïdes (ou sinusoïdes amorties) occupe seulement une petite région dans le plan temps-fréquence ; sa représentation TF est parcimonieuse. Ceci est illustré par la figure 3.1 où nous représentons la distribution TF d'un signal modal de trois composantes.

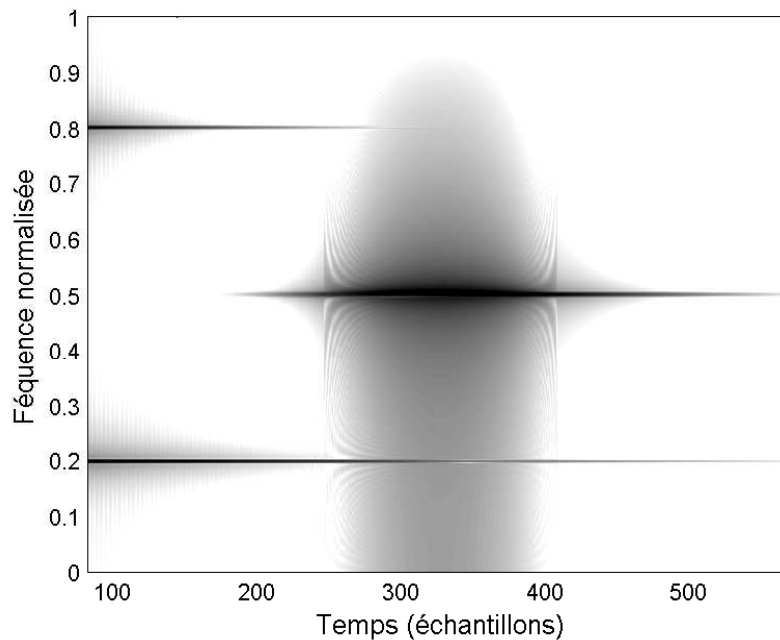


Fig. 3.1 – Représentation temps-fréquence de trois composantes modales (nous avons utilisé la transformée de Fourier à court terme).

## 3.2 Formalisation du problème et hypothèses

Pour commencer, nous considérons d'abord le cas du mélange instantané, et nous traiterons le problème du mélange convolutif sur et sous-déterminé dans le prochain chapitre. Nous utiliserons l'équation (1.1) afin de modéliser le cas du mélange instantané avec  $N > M$ , et nous supposons, sans perte de généralité, que les vecteurs colonnes de la matrice de mélange  $\mathbf{A}$  sont de norme unité ;  $\|\mathbf{a}_i\| = 1$  pour  $i \in \mathcal{N}$ , où  $\mathcal{N} = \{1, \dots, N\}$ .

Dans la suite nous allons établir les hypothèses nécessaires à notre méthode MD-UBSS.

**Hypothèse 3.1** *Les vecteurs colonnes de  $\mathbf{A}$  sont deux à deux linéairement indépendants.*

Cette hypothèse implique, que pour tout couple d'indice  $i \neq j \in \mathcal{N}$ , les vecteurs  $\mathbf{a}_i$  et  $\mathbf{a}_j$  sont linéairement indépendants. Cette hypothèse est d'autant plus nécessaire que si elle n'est pas vérifiée, le cas où  $\mathbf{a}_2 = \alpha \mathbf{a}_1$  peut se présenter, et donc l'équation (1.1) pourra se réécrire comme :

$$\mathbf{x}(t) = [\mathbf{a}_1, \mathbf{a}_3, \dots, \mathbf{a}_N][s_1(t) + \alpha s_2(t), s_3(t), \dots, s_N(t)]^T,$$

et par conséquent la séparation de  $s_1(t)$  et  $s_2(t)$  devient impossible. Cette hypothèse est utilisée dans la suite (dans l'étape de classification) pour séparer les composantes modales des sources en utilisant leurs directions spatiales qui sont données par les vecteurs colonnes de  $\mathbf{A}$ .

Comme mentionné précédemment, la résolution du problème de SAS sous-déterminé exige de fortes hypothèses a priori sur les signaux sources. Pour notre méthode, la parcimonie des signaux est considérée en termes de représentation modale des signaux comme l'indique l'hypothèse fondamentale suivante :

**Hypothèse 3.2** *Les signaux sources sont modélisés comme une somme de composantes modales.*

En effet, nous supposons ici que chaque signal source  $s_i(t)$  est la somme de  $l_i$  composantes modales  $c_i^j(t)$ ,

$$s_i(t) = \sum_{j=1}^{l_i} c_i^j(t), \quad (3.1)$$

où les  $c_i^j(t)$  représentent des sinusoides amorties ou des signaux (pseudo) harmoniques.

Les techniques standards de SAS sont basées sur l'hypothèse d'indépendance statistique des signaux sources. Dans le cas de la SAS sous-déterminée, l'indépendance statistique des signaux sources est souvent remplacée par l'hypothèse d'avoir des signaux à supports disjoints. Ceci signifie qu'il existe un domaine transformé où la représentation des signaux sources est à support disjoint ou quasi-disjoint. Cette hypothèse se traduit dans notre cas par l'hypothèse de quasi-orthogonalité des composantes modales.

**Hypothèse 3.3** *Les composantes modales des sources sont quasi-orthogonales.*

$$\frac{\langle c_i^j | c_{i'}^{j'} \rangle}{\|c_i^j\| \|c_{i'}^{j'}\|} \approx 0 \quad \text{pour} \quad (i, j) \neq (i', j'), \quad (3.2)$$

où

$$\langle c_i^j | c_{i'}^{j'} \rangle \triangleq \sum_{t=0}^{T-1} c_i^j(t) c_{i'}^{j'}(t), \quad (3.3)$$

et

$$\|c_i^j\|^2 = \langle c_i^j | c_i^j \rangle. \quad (3.4)$$

Dans le cas des signaux sinusoïdaux, la quasi-orthogonalité des composantes modales n'est rien d'autre que la quasi-orthogonalité de Fourier de deux composantes sinusoïdales avec des fréquences distinctes. Ceci peut être observé dans le domaine fréquentiel par le fait que leurs supports respectifs soient disjoints.

### 3.3 L'algorithme MD-UBSS

En se basant sur le modèle précédent, nous proposons une approche de séparation en deux étapes :

- **étape d'analyse** : dans cette étape, nous appliquons une décomposition modale sur chaque sortie de capteur afin d'en extraire toutes les composantes harmoniques ou pseudo-harmoniques. Nous utiliserons deux algorithmes de décomposition : soit l'algorithme EMD (*Empirical Mode Decomposition*) introduit en [102, 103] soit un algorithme de décomposition paramétrique ESPRIT (*Estimation of Signal Parameters via Rotational Invariance Techniques*) qui estime les paramètres des composantes modales modélisées par des sinusoides amorties.
- **étape de synthèse** : dans cette étape, nous regroupons ensemble les composantes modales correspondant au même signal source afin de reconstituer le signal d'origine. Ceci est fait par une méthode de classification utilisant les directions spatiales des composantes que l'on estime par corrélation avec le signal observé.

Notons, qu'en appliquant cette approche sur toutes les sorties d'antenne, nous obtenons  $M$  estimées de chacune des sources. Nous avons observé que la qualité d'estimation des sources varie d'un capteur à un autre. Ceci dépend fortement des coefficients du mélange, en particulier, du rapport interférences à signal (RIS) de la source désirée. Par conséquent, nous proposons une méthode aveugle de sélection pour choisir la meilleure des  $M$  estimées d'une source donnée. Le tableau 3.1 donne un récapitulatif de cet algorithme.

Tab. 3.1 – Algorithme MD-UBSS dans le cas du mélange instantané

1. Extraire toutes les composantes harmoniques de chaque capteur en appliquant la décomposition modale.
2. Estimer les directions spatiales par l'équation (3.10) et leurs appliquer une classification vectorielle en utilisant l'algorithme  $k$ -means [110].
3. Estimer les sources en regroupant ensemble les composantes modales correspondantes à la même direction spatiale.
4. Sélectionner les sources par (3.12).

### 3.3.1 Estimation des composantes modales

#### Analyse des signaux utilisant l'EMD

Une nouvelle technique non-linéaire, appelée EMD a été récemment introduite par N. E. Huang et al. pour représenter les signaux non-stationnaires [102]. Dans l'algorithme EMD, le signal non stationnaire est considéré à l'échelle de ses oscillations locales. Plus précisément, la décomposition d'un signal  $z(t)$  par l'EMD se résume comme suit :

1. identification de tous les extremums de  $z(t)$  ; ceci est réalisé par l'algorithme décrit dans [111],
2. interpolation entre les minima (respectivement maxima) du signal, ce qui permettra de former une enveloppe  $e_{min}(t)$  (respectivement  $e_{max}(t)$ ) ; plusieurs techniques d'interpolation peuvent être utilisées ; dans nos simulations, nous avons utilisé l'interpolation spline [111],
3. calcul de la moyenne  $m(t) = (e_{min}(t) + e_{max}(t))/2$ ,
4. extraction du détail  $d(t) = z(t) - m(t)$ ,
5. itération des étapes précédentes sur le résidu<sup>1</sup>  $m(t)$  jusqu'à l'obtention d'un résidu final d'énergie quasi-nulle (en pratique la condition d'arrêt de l'algorithme est  $\|m(t)\| \leq \epsilon$ , où  $\epsilon$  est une valeur seuil donnée).

<sup>1</sup>En effet, le signal moyen  $m(t)$  est également le signal résiduel après extraction du détail  $d(t)$  ;  $m(t) = z(t) - d(t)$ .

Par conséquent, en appliquant l'algorithme EMD sur les  $x_i$  qui s'écrivent  $x_i(t) = \sum_{j=1}^N a_{ij} s_j(t) = \sum_{j=1}^N \sum_{k=1}^{l_j} a_{ij} c_j^k(t)$ , on obtient des estimées  $\hat{c}_j^k(t)$  des composantes  $c_j^k(t)$  (à une constante scalaire  $a_{ij}$  près).

### Analyse paramétrique de signaux

Dans cette section, nous présentons une solution alternative pour l'étape d'analyse des signaux. Pour ce faire, nous adopterons une modélisation en somme de sinusoides amorties des signaux sources :

$$s_i(t) = \Re \left\{ \sum_{j=1}^{l_i} \alpha_i^j (z_i^j)^t \right\}, \quad (3.5)$$

correspondant à :

$$c_i^j(t) = \Re \left\{ \alpha_i^j (z_i^j)^t \right\}, \quad (3.6)$$

où  $\alpha_i^j = \beta_i^j e^{j\theta_i^j}$  représente l'amplitude complexe et  $z_i^j = e^{d_i^j + j2\pi f_i^j}$  est le  $j^{\text{e}}$  pôle de la source  $s_i$  où  $d_i^j$  est le facteur négatif d'amortissement et  $f_i^j$  est la fréquence. On note par  $L_{tot}$  le nombre total de composantes modales ;

$$L_{tot} = \sum_{i=1}^N l_i.$$

Pour l'extraction des composantes modales, nous proposons d'employer l'algorithme ESPRIT (Estimation of Signal Parameters via Rotational Invariance Technique) qui estime les pôles des signaux en exploitant la propriété d'invariance par décalage de la matrice Hankel de données  $[\mathcal{H}(x_k)]_{n_1 n_2} \triangleq x_k(n_1 + n_2)$  de taille  $D \times (T - D)$ ,  $D$  étant un paramètre de fenêtrage choisi dans la plage  $T/3 \leq D \leq 2T/3$ . Plus précisément, nous employons l'algorithme de Kung introduit dans [112] qui peut être résumé par les étapes suivantes :

1. former la matrice Hankel des données  $\mathcal{H}(x_k)$ ,
2. estimer le sous-espace signal  $\mathbf{U}^{(L_{tot})} = [\mathbf{u}_1, \dots, \mathbf{u}_{2L_{tot}}]$  de dimension  $2L_{tot}$  de la matrice  $\mathcal{H}(x_k)$  par la décomposition en valeurs et vecteurs singuliers de cette dernière ( $\mathbf{u}_1, \dots, \mathbf{u}_{2L_{tot}}$  sont les vecteurs singuliers principaux gauche de  $\mathcal{H}(x_k)$ ),
3. résoudre (au sens des moindres carrés) l'équation d'invariance par décalage

$$\mathbf{U}_{\downarrow}^{(L_{tot})} \Psi = \mathbf{U}_{\uparrow}^{(L_{tot})} \Leftrightarrow \Psi = \mathbf{U}_{\downarrow}^{(L_{tot})\#} \mathbf{U}_{\uparrow}^{(L_{tot})}, \quad (3.7)$$



où  $\Psi = \Phi \Delta \Phi^{-1}$ , et  $\Phi$  est une matrice non-singulière de dimension  $2L_{tot} \times 2L_{tot}$  et  $\Delta = \text{diag}(z_1^1, z_1^{1*}, \dots, z_1^{l_1}, z_1^{l_1*}, \dots, z_N^{l_N}, z_N^{l_N*})$ . Les flèches  $\downarrow$  et  $\uparrow$  dénotent, les opérateurs de decimation de la dernière et la première lignes respectivement,

4. estimer les pôles comme étant les valeurs singulières de la matrice  $\Psi$ ,
5. estimer les amplitudes complexes en résolvant le critère des moindres carrés suivant :

$$\min_{\alpha_k} \|\mathbf{x}_k - \mathbf{Z} \alpha_k\|^2 \Leftrightarrow \alpha_k = \mathbf{Z}^\# \mathbf{x}_k, \quad (3.8)$$

où  $\mathbf{x}_k = [x_k(0), \dots, x_k(T-1)]^T$  est le vecteur des observations et  $\mathbf{Z}$  est la matrice Vandermonde construite à partir des pôles estimés :

$$\mathbf{Z} = [z_1^1, z_1^{1*}, \dots, z_1^{l_1}, z_1^{l_1*}, \dots, z_N^{l_N}, z_N^{l_N*}],$$

avec  $z_i^j = [1, z_i^j, (z_i^j)^2, \dots, (z_i^j)^{T-1}]^T$  et  $\alpha_k$  est le vecteur des amplitudes complexes :

$$\alpha_k = \frac{1}{2} [a_{k1} \alpha_1^1, a_{k1} \alpha_1^{1*}, \dots, a_{k1} \alpha_1^{l_1}, a_{k1} \alpha_1^{l_1*}, \dots, a_{kN} \alpha_N^{l_N}, a_{kN} \alpha_N^{l_N*}]^T.$$

### 3.3.2 Classification et estimation des signaux sources

#### Synthèse des signaux utilisant la classification vectorielle

Pour la synthèse des signaux sources, on observe que l'hypothèse de quasi-orthogonalité nous permet d'avoir :

$$\frac{\langle \mathbf{x} | \mathbf{c}_i^j \rangle}{\|\mathbf{c}_i^j\|^2} \triangleq \frac{1}{\|\mathbf{c}_i^j\|^2} \begin{bmatrix} \langle x_1 | \mathbf{c}_i^j \rangle \\ \vdots \\ \langle x_M | \mathbf{c}_i^j \rangle \end{bmatrix} \approx \mathbf{a}_i, \quad (3.9)$$

où  $\mathbf{a}_i$  représente la  $i^e$  colonne de  $\mathbf{A}$ . nous pouvons alors associer chaque composante  $\hat{\mathbf{c}}_j^k$  à une direction spatiale (vecteur colonne de  $\mathbf{A}$ ) que l'on estime par

$$\hat{\mathbf{a}}_j^k = \frac{\langle \mathbf{x} | \hat{\mathbf{c}}_j^k \rangle}{\|\hat{\mathbf{c}}_j^k\|^2}. \quad (3.10)$$

Le vecteur  $\hat{\mathbf{a}}_j^k$  sera égal approximativement au vecteur  $\mathbf{a}_i$  (à une constante scalaire près) si  $\hat{\mathbf{c}}_j^k$  est une estimée d'une composante modale de la source  $i$ . Par conséquent, deux composantes du même signal source étant associées au même vecteur colonne de  $\mathbf{A}$ , nous proposons de regrouper ces composantes par classification sur les vecteurs  $\hat{\mathbf{a}}_j^k$  selon  $N$  classes (voir figure 3.2). Pour cela, nous calculerons

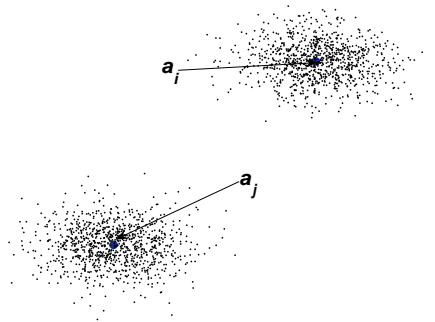


Fig. 3.2 – Illustration des données à classifier, ou nous représentons les différents vecteurs estimés  $\hat{\mathbf{a}}_i^j$  et leurs centroïdes.

d'abord les vecteurs normalisés

$$\bar{\mathbf{a}}_j^k = \frac{\hat{\mathbf{a}}_j^k e^{-j\psi_j^k}}{\|\hat{\mathbf{a}}_j^k\|}$$

où  $\psi_j^k$  est la phase du premier élément du vecteur  $\hat{\mathbf{a}}_j^k$  (le but est de forcer le premier élément à être un réel positif). Puis, ces vecteurs sont classés en utilisant l'algorithme  $k$ -means [110], qui peut se résumer comme suit :

1. initialement, sélectionner  $N$  centroïdes (vecteurs) arbitrairement parmi l'ensemble des vecteurs  $\bar{\mathbf{a}}_j^k$  à classifier,
2. assigner chaque vecteur  $\bar{\mathbf{a}}_j^k$  à la classe dont le centroïde est le plus proche au sens de la distance Euclidienne; si  $\mathbf{y}_1, \dots, \mathbf{y}_N$  sont les centroïdes des  $N$  classes, on assigne le vecteur  $\bar{\mathbf{a}}_j^k$  à la classe  $i_0$  qui satisfait :

$$i_0 = \arg \min_i \|\bar{\mathbf{a}}_j^k - \mathbf{y}_i\| ,$$

3. quand tous les vecteurs ont été assignés, recalculer les positions des  $N$  centroïdes de la façon suivante : le centroïde de chaque classe est calculé comme la valeur moyenne des vecteurs appartenant à cette classe,
4. répéter l'étape 2 et 3 jusqu'à ce qu'il n'y ait aucun changement de vecteurs entre les classes.

Finalement, on pourra reconstruire les sources initiales à une constante près en additionnant les différentes composantes d'une même classe,

$$\hat{s}_i(t) = \sum_{c_i} \hat{c}_i^j(t), \quad (3.11)$$

où  $\mathcal{C}_i$  représente la  $i^e$  classe.

### Association et sélection

Nous remarquons, qu'en appliquant l'approche décrite précédemment (analyse plus synthèse) sur toutes les sorties d'antenne  $x_1(t), \dots, x_M(t)$ , nous obtenons  $M$  estimées de chacune des sources. Nous avons observé que la qualité d'estimation des sources varie d'un capteur à un autre, et ceci dépend fortement des coefficients du mélange, en particulier, du RIS de la source désirée. Par conséquent, nous proposons une méthode aveugle de sélection pour choisir la meilleure des  $M$  estimées d'une source donnée. Pour cela, nous avons besoin tout d'abord de regrouper les estimées d'une même source ensemble. Ceci est réalisé par corrélation ; un signal est associé aux  $(M - 1)$  signaux (les signaux issus des  $(M - 1)$  autres capteurs) qui lui sont les plus corrélés. Le facteur de corrélation de deux signaux  $s_1$  et  $s_2$  est calculé par  $\frac{\langle s_1 | s_2 \rangle}{\|s_1\| \|s_2\|}$ .

Une fois l'association des sources effectuée, on propose de sélectionner la source estimée qui a l'énergie maximale ;

$$\hat{s}_i(t) = \arg \max_{\tilde{s}_i^j(t)} \left\{ E_i^j = \sum_{t=0}^{T-1} |\tilde{s}_i^j(t)|^2, \quad j = 1, \dots, M \right\}, \quad (3.12)$$

où  $E_i^j$  représente l'énergie de la  $i^e$  source obtenue à partir du  $j^e$  capteur. On peut envisager d'autres méthodes de sélection basées sur la dispersion autour du centroïde de chaque classe, le nombre de composantes pour chaque estimée, etc.

### 3.3.3 Cas des composantes modales communes

Dans cette section, nous supposons qu'une composante  $c_j^k(t)$ , associée au pôle  $z_j^k$ , peut être présente dans plusieurs sources. Ceci est le cas pour certains signaux musicaux tels que ceux traités dans [113]. Pour simplifier, nous supposons ici qu'une composante appartient au plus à deux sources. Supposons donc que la composante sinusoidale  $(z_j^k)^t$  est présente dans les sources  $s_{j_1}(t)$  et  $s_{j_2}(t)$  avec les amplitudes  $\alpha_{j_1}$  et  $\alpha_{j_2}$  respectivement. Il s'en suit que la direction spatiale associée à cette composante est une combinaison linéaire des vecteurs colonnes  $\mathbf{a}_{j_1}$  et  $\mathbf{a}_{j_2}$ . Plus précisément, nous avons :

$$\hat{\mathbf{a}}_j^k = \frac{1}{\|z_j^k\|^2} \begin{bmatrix} \mathbf{x}_1^T z_j^k \\ \vdots \\ \mathbf{x}_M^T z_j^k \end{bmatrix} \approx \alpha_{j_1} \mathbf{a}_{j_1} + \alpha_{j_2} \mathbf{a}_{j_2}. \quad (3.13)$$

Il s'agit maintenant de trouver les indices  $j_1$  et  $j_2$  des deux sources associées à cette composante, ainsi que les amplitudes  $\alpha_{j_1}$  et  $\alpha_{j_2}$ . Pour ce faire, on propose une approche basée sur la projection en sous-espace. Supposons que la matrice de mélange  $\mathbf{A}$  est connue et vérifie les conditions  $M > 2$  et tous les triplets de vecteurs colonnes de  $\mathbf{A}$  sont linéairement indépendants. Observons alors :

$$\mathbf{P}_{\tilde{\mathbf{A}}}^{\perp} \hat{\mathbf{a}}_j^k = \mathbf{0} , \quad (3.14)$$

si et seulement si  $\tilde{\mathbf{A}} = [\mathbf{a}_{j_1} \ \mathbf{a}_{j_2}]$ , où  $\tilde{\mathbf{A}}$  est une matrice formée à partir d'une paire de vecteurs colonnes de  $\mathbf{A}$  et où  $\mathbf{P}_{\tilde{\mathbf{A}}}^{\perp}$  représente la matrice de projection orthogonale sur le sous-espace bruit de  $\tilde{\mathbf{A}}$ ;

$$\mathbf{P}_{\tilde{\mathbf{A}}}^{\perp} = \mathbf{I}_M - \tilde{\mathbf{A}} (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^H . \quad (3.15)$$

En pratique, en tenant compte du terme de bruit, on détecte les colonnes  $j_1$  et  $j_2$  en minimisant :

$$(j_1, j_2) = \arg \min_{(l, m)} \left\{ \|\mathbf{P}_{\tilde{\mathbf{A}}}^{\perp} \hat{\mathbf{a}}_j^k\| \mid \tilde{\mathbf{A}} = [\mathbf{a}_l \ \mathbf{a}_m] \right\} . \quad (3.16)$$

Une fois  $\tilde{\mathbf{A}}$  trouvée, on estime les pondérations  $\alpha_{j_1}$  et  $\alpha_{j_2}$  par :

$$\begin{bmatrix} \alpha_{j_1} \\ \alpha_{j_2} \end{bmatrix} = \tilde{\mathbf{A}}^{\#} \hat{\mathbf{a}}_j^k , \quad (3.17)$$

Dans cette section, nous avons traité toutes les composantes comme étant associées à deux signaux sources. Si jamais une composante n'est présente que dans une seule source, un des deux coefficients estimés par l'équation (3.15) devrait être nul ou proche de zéro.

Dans ce qui précède, la matrice de mélange  $\mathbf{A}$  est supposée connue, il faut donc l'estimer avant d'appliquer la projection en sous-espace. Nous proposons d'estimer les colonnes de la matrice  $\mathbf{A}$  comme étant les centroïdes moyennés (comme chaque source est estimée  $M$  fois, et que chaque sources correspond à une classe et à un centroïde, alors, on moyenne les centroïdes des  $M$  classes d'une source donnée) des  $N$  classes obtenues dans l'étape de synthèse de notre algorithme. Cette approche suppose implicitement que la majorité des composantes n'appartiennent qu'à une seule source, et par conséquent, que les vecteurs  $\hat{\mathbf{a}}_j^k$  représentent dans leurs majorité une des colonnes de la matrice  $\mathbf{A}$ .

### 3.4 L'algorithme MD-UBSS modifié

Nous proposons dans cette section d'améliorer l'algorithme précédant en terme de coût de calcul et en qualité d'estimation dans le cas où l'hypothèse 3.3 est

mal satisfaite<sup>2</sup>. D'abord, afin d'éviter de réestimer les composantes modales pour chaque capteur, nous utiliserons toutes les données observées pour estimer (une seule fois) les pôles des signaux sources. C'est pourquoi, nous appliquerons la technique ESPRIT sur la matrice de covariance moyennée des données  $\mathbb{H}(\mathbf{x})$  définie par :

$$\mathbb{H}(\mathbf{x}) = \sum_{i=1}^M \mathcal{H}(x_i) \mathcal{H}(x_i)^H, \quad (3.18)$$

et nous appliquerons les étapes 1 à 4 de l'algorithme de Kung décrit dans la section 3.3.1 afin d'obtenir tous les pôles  $z_i^j$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, l_i$ . De cette façon, nous réduisons de manière significative le coût de calcul et évitons le problème du choix de la "meilleure source estimée" de l'algorithme précédent.

Dans ce qui suit, nous présenterons les étapes d'estimation des pôles et des directions spatiales respectivement, ainsi que la recombinaison des signaux sources. Afin de relâcher l'hypothèse 3.3, on peut réécrire le modèle de donnée comme suit :

$$\Gamma \mathbf{z}(t) = \mathbf{x}(t), \quad (3.19)$$

où  $\Gamma \triangleq [\gamma_1^1, \bar{\gamma}_1^1, \dots, \gamma_N^{l_N}, \bar{\gamma}_N^{l_N}]$ ,  $\gamma_i^j = \beta_i^j e^{j\phi_i^j} \mathbf{b}_i^j$  et  $\bar{\gamma}_i^j = \beta_i^j e^{-j\phi_i^j} \mathbf{b}_i^j$ , où  $\mathbf{b}_i^j$  est un vecteur de norme unité, représentant la direction spatiale de la  $i^e$  composante ( $\mathbf{b}_i^j = \mathbf{a}_k$  si la composante  $(z_i^j)^t$  appartient au  $k^e$  signal source) et

$$\mathbf{z}(t) \triangleq [(z_1^1)^t, (z_1^{1*})^t, \dots, (z_N^{l_N})^t, (z_N^{l_N*})^t]^T.$$

L'estimation de  $\Gamma$  est obtenue au sens des moindres carrés comme suit :

$$\min_{\Gamma} \|\mathbf{X} - \Gamma \mathbf{Z}\|^2 \Leftrightarrow \Gamma = \mathbf{X} \mathbf{Z}^\#, \quad (3.20)$$

où  $\mathbf{X} = [\mathbf{x}(0), \dots, \mathbf{x}(T-1)]$  et  $\mathbf{Z} = [\mathbf{z}(0), \dots, \mathbf{z}(T-1)]$ . Après avoir estimé  $\Gamma$ , on estime la phase de chaque pôle comme suit :

$$\phi_i^j = \frac{\arg(\bar{\gamma}_i^{jH} \gamma_i^j)}{2}. \quad (3.21)$$

La direction spatiale de chaque composante modale est estimée par :

$$\hat{\mathbf{a}}_i^j = \gamma_i^j e^{-j\phi_i^j} + \bar{\gamma}_i^j e^{j\phi_i^j} = 2\beta_i^j \mathbf{b}_i^j. \quad (3.22)$$

Enfin, on regroupe les composantes modales par classification des vecteurs  $\hat{\mathbf{a}}_i^j$  en  $N$  classes. Après l'étape de classification, on obtient  $N$  classes avec  $N$  centroïdes

<sup>2</sup>C'est le cas où les composantes modales sont proches ou pour des composantes modales à fort facteur d'amortissement.

de norme unitaire  $\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_N$  correspondant aux estimés des vecteurs colonnes de la matrice de mélange  $\mathbf{A}$ . Si le pôle  $z_i^j$  appartient à la  $k^e$  classe, alors selon (3.20), son amplitude peut être estimée par :

$$\beta_i^j = \frac{\hat{\mathbf{a}}_k^T \hat{\mathbf{a}}_i^j}{2}. \quad (3.23)$$

On pourra reconstruire les sources originales à une constante près en additionnant toutes les composantes modales appartenant à la même classe  $\mathcal{C}_k$  comme suit :

$$\hat{s}_k(t) = \Re e \left\{ \sum_{\mathcal{C}_k} \beta_i^j e^{j\phi_i^j} (z_i^j)^t \right\}. \quad (3.24)$$

Nous noterons qu'il est possible d'assigner chaque composante modale à deux (ou plus) sources comme dans la section 3.3.3 en utilisant les équations (3.14)-(3.17).

## 3.5 Discussion

Dans cette section, nous produisons quelques commentaires afin obtenir plus de lisibilité sur la méthode de séparation proposée :

### Cas sur-déterminé

Dans le cas sur-déterminé, on peut effectuer la séparation par une inversion de la matrice de mélange  $\mathbf{A}$ . Cette dernière est estimée à partir des centroïdes des  $N$  classes (le centroïde de la  $i^e$  classe représente l'estimée de la  $i^e$  colonne de la matrice de mélange  $\mathbf{A}$ ).

### Estimation du nombre de sources

C'est une tâche difficile dans le cas sous-déterminé. Il existe certaines solutions utilisant des approches tensorielles [114] ou des techniques de classifications avec estimation conjointe du nombre de classes [110]. Cependant, ces méthodes sont très sensibles au bruit, à la dynamique des signaux sources et au conditionnement de la matrice de mélange. Dans ce manuscrit, on suppose que le nombre de sources est connu (ou correctement estimé).

### Le nombre des composantes modales

Dans l'approche paramétrique, nous avons à choisir le nombre des composantes modales  $L_{tot}$  dont on a besoin pour une bonne approximation du signal audio. En effet, une petite valeur de  $L_{tot}$  mène à une mauvaise représentation du signal et une grande valeur de  $L_{tot}$  augmente le coût de calcul. En fait,  $L_{tot}$  dépend de la complexité du signal.

### Approche hybride de séparation

Il est très probable que la qualité de séparation peut être améliorée en combinant l'approche proposée avec des techniques de filtrage spatial appropriées. En effet, on a observé que la qualité de séparation dépend des coefficients de la matrice de mélange  $A$ . Le filtrage spatial peut être employé pour améliorer le RIS du signal source désiré et par conséquent améliorer sa qualité d'extraction.

### Cas sans bruit

Dans le cas sans bruit (avec une modélisation parfaite en somme de sinusoides amorties) l'estimation des composantes modales utilisant l'algorithme ESPRIT sera parfaite. Ceci impliquera, une estimation exacte des colonnes de la matrice de mélange en utilisant une estimation au sens des moindres carrés et donc une restauration des sources quasi-parfaite.

## 3.6 Simulations

Nous présentons ici quelques résultats de simulation pour illustrer l'efficacité de notre algorithme de séparation. Pour cela, nous considérons une antenne composée de  $M = 3$  capteurs recevant  $N = 4$  signaux sources audio arrivant suivant les angles d'arrivés  $\theta_1 = 15$ ,  $\theta_2 = 30$ ,  $\theta_3 = 45$  et  $\theta_4 = 75$  degrés respectivement. La taille des observations  $T = 10000$  échantillons (les signaux sources sont échantillonnés à une fréquence de 8 KHz). Les signaux observés sont corrompus par un bruit blanc additif de covariance  $\sigma^2 \mathbf{I}_M$  ( $\sigma^2$  étant la puissance du bruit). La qualité de la séparation est mesurée par l'erreur quadratique moyenne normalisée (EQMN)

des sources estimées pour  $N_r = 100$  réalisations aléatoires du bruit.

$$\text{EQMN}_i \triangleq \frac{1}{N_r} \sum_{r=1}^{N_r} \min_{\alpha} \left( \frac{\|\alpha \hat{\mathbf{s}}_{i,r} - \mathbf{s}_i\|^2}{\|\mathbf{s}_i\|^2} \right) \quad (3.25)$$

$$\text{EQMN}_i = \frac{1}{N_r} \sum_{r=1}^{N_r} 1 - \left( \frac{\hat{\mathbf{s}}_{i,r}^T \mathbf{s}_i}{\|\hat{\mathbf{s}}_{i,r}\| \|\mathbf{s}_i\|} \right)^2 \quad (3.26)$$

$$\text{EQMN} = \frac{1}{N} \sum_{i=1}^N \text{EQMN}_i. \quad (3.27)$$

où  $\mathbf{s}_i \triangleq [s_i(0), \dots, s_i(T-1)]$ ,  $\hat{\mathbf{s}}_{i,r}$  (défini de manière similaire) est la  $r^{\text{e}}$  estimée de la source  $\mathbf{s}_i$  et  $\alpha$  est un facteur scalaire qui compense l'indétermination d'échelle de la SAS.

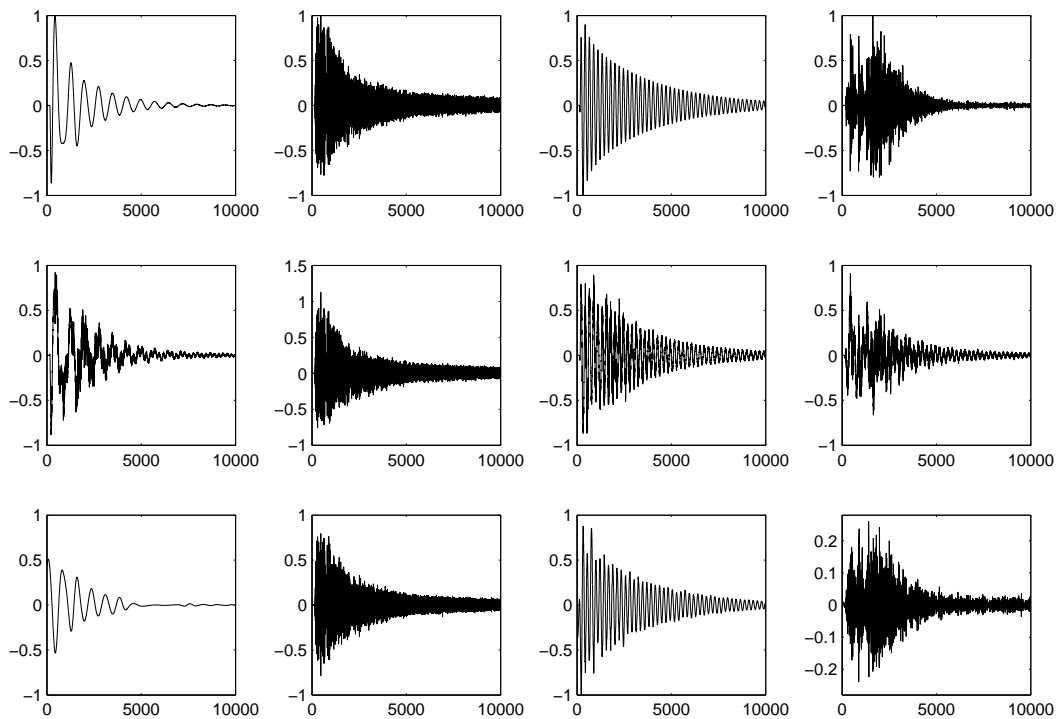


Fig. 3.3 – Exemple de séparation de 4 signaux audio pour 3 capteurs : signaux originaux (ligne du haut) ; signaux estimés par pseudo-inversion quand la matrice de mélange est connue (ligne du milieu) ; signaux estimés par l'algorithme MD-UBSS utilisant la méthode de décomposition EMD (ligne du bas).

La figure 3.3 représente les résultats obtenus en utilisant notre algorithme MD-UBSS avec la méthode de décomposition EMD pour les 4 sources audio



représentées par la première ligne de la figure. La deuxième ligne montre le résultat obtenu en utilisant la pseudo-inverse de la matrice de mélange  $\mathbf{A}$  pour estimer les sources originales en connaissant  $\mathbf{A}$  avec exactitude, et la dernière donne les estimées des sources en utilisant notre algorithme.

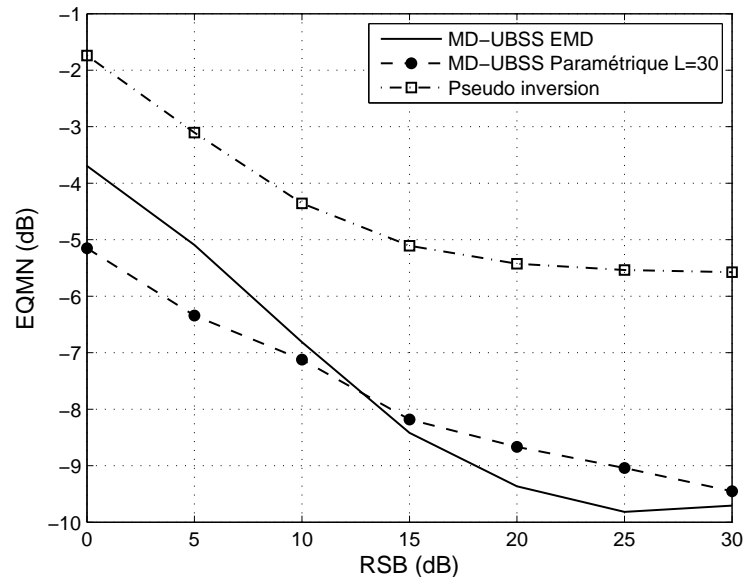


Fig. 3.4 – Performances de l'algorithme de séparation MD-UBSS dans le cas de 4 sources audio pour 3 capteurs : les courbes représentent la valeur moyenne de l'erreur quadratique de l'estimation des signaux sources (EQMN) en fonction du rapport signal à bruit (RSB).

Dans la figure 3.4, on représente la variation de l'erreur quadratique moyenne normalisée (EQMN) des estimées des signaux sources en fonction du rapport signal à bruit (RSB). Sur cette figure nous comparons les résultats obtenus par l'algorithme MD-UBSS avec les techniques de décomposition EMD et ESPRIT. Le nombre de composantes modales  $L = 30$  composantes par signal source ( $L_{tot} = NL$ ). Comme référence, on représente les performances obtenues par application de la pseudo-inverse de la matrice  $\mathbf{A}$  (supposée exactement connue) [115]. On observe que nos algorithmes (basés sur les décompositions EMD et ESPRIT) produisent de meilleurs résultats que l'application de la pseudo-inverse de la matrice exacte de mélange.

La figure 3.5 illustre l'effet du choix du nombre de composantes modales  $L$  modélisant le signal audio. Des valeurs trop petites ou trop grandes de  $L$  dégradent les performances de notre méthode. En d'autres termes, il existe un

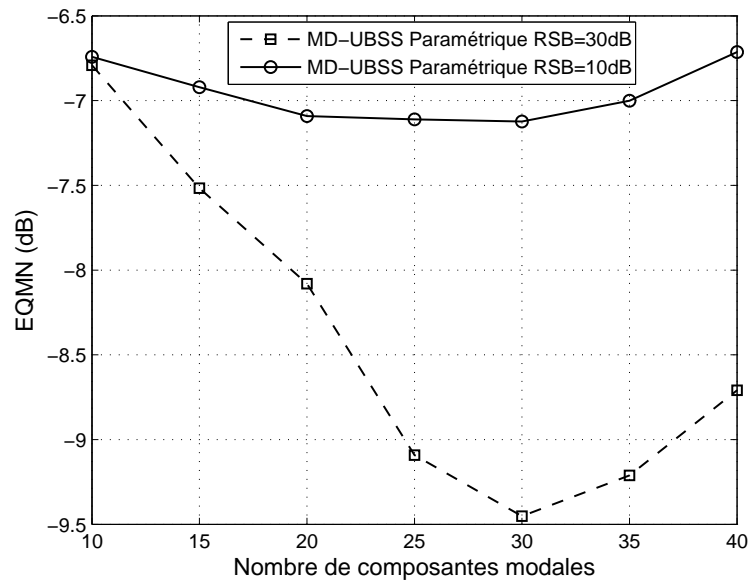


Fig. 3.5 – EQMN en fonction de  $L$  pour 4 sources audio et 3 capteurs : comparaison des performances de l'algorithme MD-UBSS (ESPRIT) pour  $L$  variant dans la plage  $[10 \dots 40]$  avec  $RSB=10\text{dB}$  et  $RSB=30\text{dB}$ .

choix optimal de  $L$  qui dépend du type de signal à modéliser.

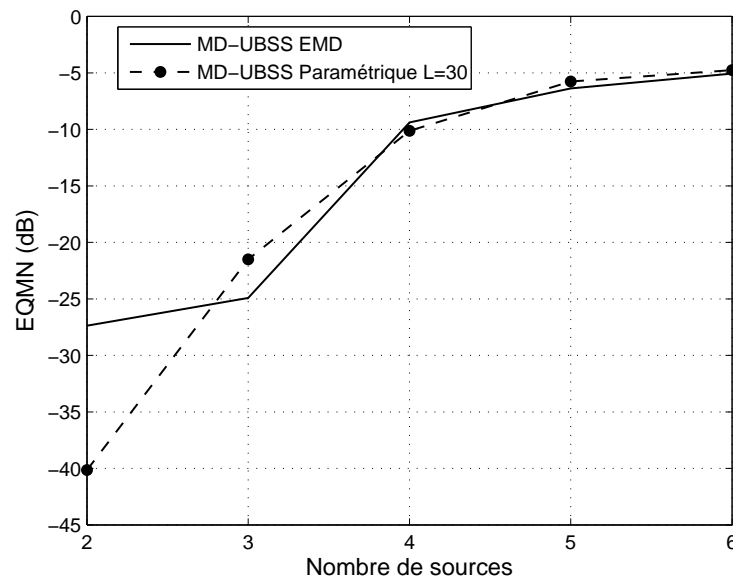


Fig. 3.6 – Performances de l'algorithme de séparation MD-UBSS pour 3 capteurs : la courbe représente la valeur moyenne de l'erreur quadratique de l'estimation des signaux sources en fonction de leur nombre.

La figure 3.6 représente la variation de l'EQMN des estimées des signaux sources en fonction du nombre de sources qui varie de 2 à 6. Dans les cas où  $N = 2$  et  $N = 3$  (cas sur-déterminé) la séparation est effectuée par pseudo-inversion de l'estimée de la matrice de mélange  $\mathbf{A}$ . Dans le cas sous-déterminé, l'utilisation des techniques de décomposition EMD et ESPRIT présente des performances similaires. Cependant, la technique ESPRIT produit de meilleurs résultats dans le cas sur-déterminé que la technique EMD.

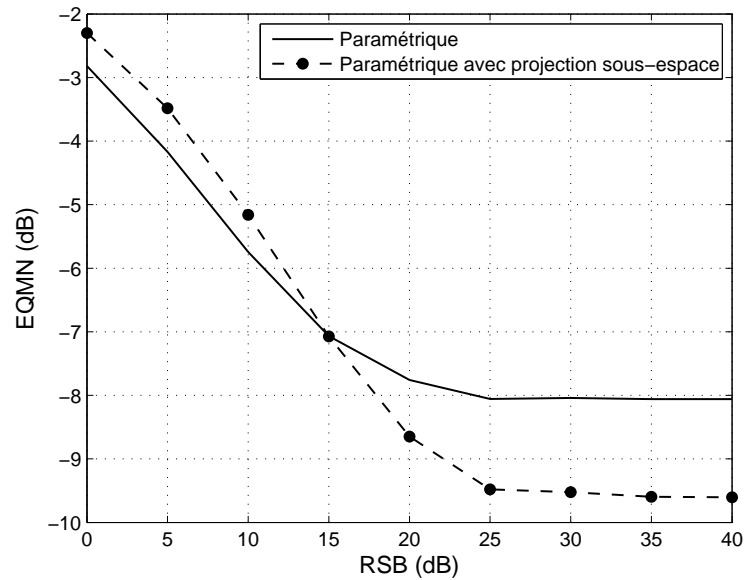


Fig. 3.7 – Performances des algorithmes de séparation MD-UBSS pour 4 sources audio et 3 capteurs : les courbes représentent la valeur moyenne de l'erreur quadratique de l'estimation des signaux sources en fonction du rapport signal à bruit pour l'algorithme ESPRIT classique ainsi que celle pour l'algorithme ESPRIT avec projection en sous-espace.

La figure 3.7 représente la variation de l'EQMN des estimées des signaux sources en fonction du RSB. Sur cette figure nous comparons les résultats obtenus par l'utilisation de la décomposition ESPRIT uniquement et celles obtenues avec l'utilisation de la décomposition ESPRIT conjuguée à la projection en sous-espace. On peut observer un gain à moyen et fort RSB mais une légère dégradation à faible RSB pour la méthode ESPRIT avec projection en sous-espace. Cette dégradation est due au fait que l'on associe systématiquement toute composante à deux sources est que l'estimation des amplitudes par (3.17) est fortement bruitée à faible RSB. Ainsi, si une composante n'appartient effectivement qu'à une seule

source  $s_{j_1}$  le coefficient de la deuxième source  $\alpha_{j_2}$  devrait être nul. En présence de bruit, l'équation (3.17) donne une valeur non nulle de ce coefficient qui dégrade l'estimation des sources.

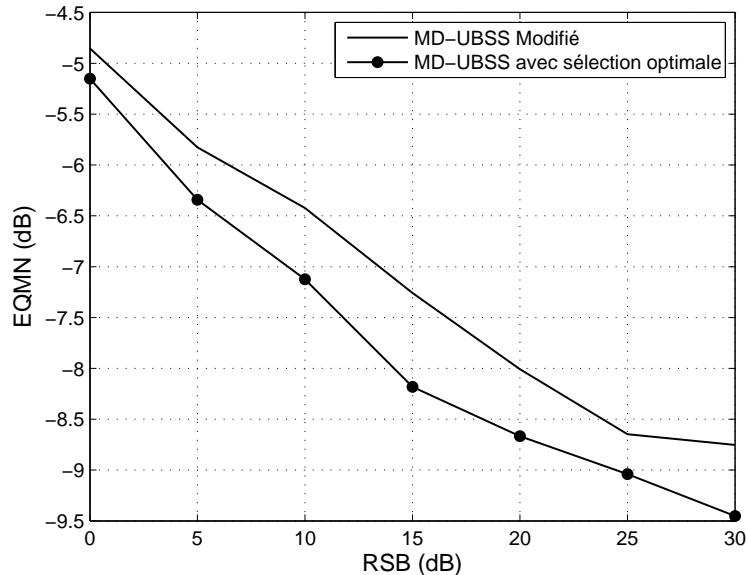


Fig. 3.8 – Performances des l’algorithmes de séparation MD-UBSS pour 4 sources audio et 3 capteurs : les courbes représentent la valeur moyenne de l’erreur quadratique de l’estimation des signaux sources en fonction du rapport signal à bruit pour l’algorithme MD-UBSS avec sélection optimale ainsi que celle pour l’algorithme MD-UBSS modifié.

Dans la figure 3.8, on compare les performances de séparation obtenues par l’algorithme MD-UBSS utilisant la technique de décomposition ESPRIT et l’algorithme MD-UBSS modifié décrit dans la section 3.4.

La figure 3.9, illustre les performances d’estimation de la matrice de mélange  $\mathbf{A}$  en utilisant la méthode de classification décrite dans la section 3.5. On observe que l’on obtient une bonne qualité d’estimation de la matrice de mélange, ce qui confirme le fait que la majorité des composantes modales appartiennent effectivement à une seule source.

Dans la figure 3.10, on compare les performances de séparation de notre algorithme MD-UBSS utilisant les techniques de décomposition EMD et ESPRIT et l’algorithme DUET modifié [87] dans le cas sous-déterminé avec  $N = 4$  signaux sources et  $M = 3$  capteurs. Les signaux sources utilisés sont deux voix d’hommes, une voix de femme et un instrument de musique. On observe un gain de perfor-

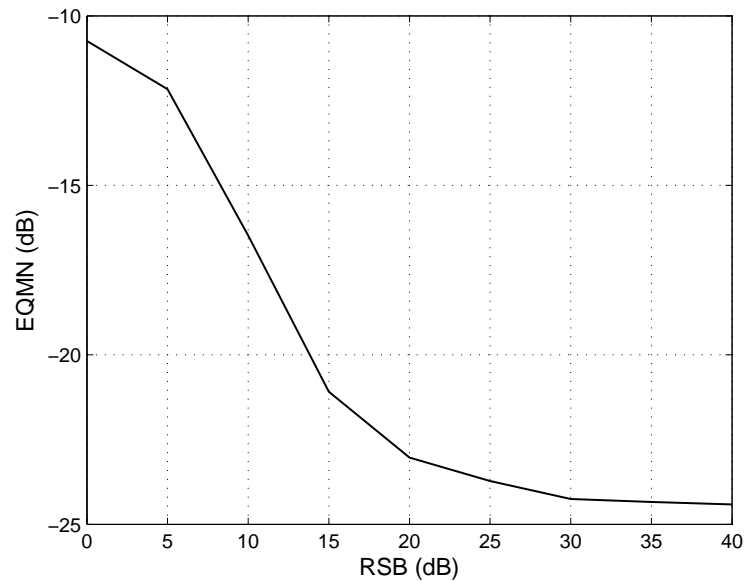


Fig. 3.9 – Estimation de la matrice de mélange : la courbe représente la valeur moyenne de l'erreur quadratique de l'estimation de la matrice de mélange en fonction du rapport signal à bruit pour l'algorithme MD-UBSS pour 4 signaux audio et 3 capteurs.

mances en faveur de l'algorithme DUET modifié. Ceci s'explique par le fait que les signaux utilisés ont été mal décomposés en raison de leur structure (l'hypothèse de quasi-orthogonalité n'est pas ou mal satisfaite). Une mauvaise décomposition implique forcément une mauvaise synthèse qui explique la dégradation de la qualité de la séparation.

### 3.7 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode de séparation aveugle de sources audio MD-UBSS, utilisant des techniques de décomposition modale (EMD et ESPRIT). L'avantage majeur de la méthode proposée est sa capacité à séparer plus de sources que de capteurs et dans ce cas elle produit une meilleure qualité de séparation que ce que l'on obtient par pseudo-inversion de la matrice de mélange (même si elle est connue). La méthode proposée procède en deux étapes : une étape d'analyse où toutes les composantes modales sont estimées, suivie par l'étape de synthèse pour regrouper (classifier) ensemble les composantes modales et reconstruire les signaux sources. Pour l'étape d'ana-

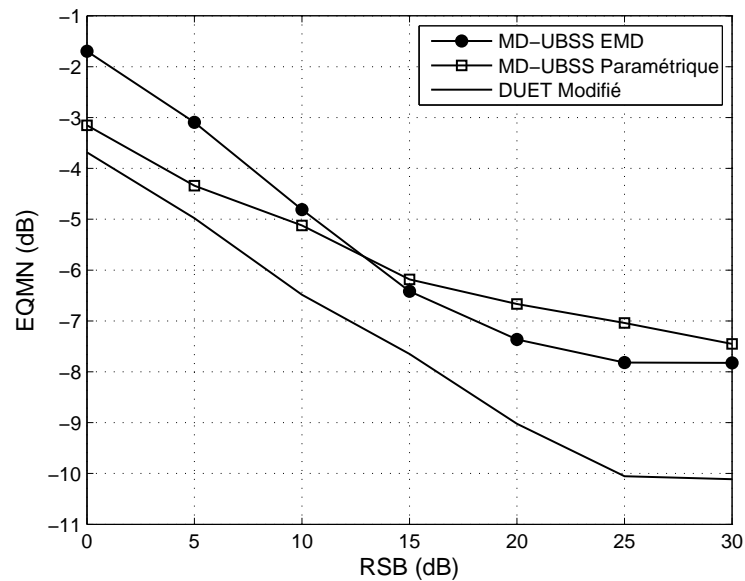


Fig. 3.10 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 10000$  : comparaison des performances de séparation de l'algorithme MD-UBSS (EMD et paramétrique) et l'algorithme DUET modifié.

lyse, deux techniques sont utilisées et comparées basées respectivement sur les algorithmes EMD et ESPRIT. Une version modifiée de l'algorithme MD-UBSS ainsi qu'une approche basée sur la projection en sous-espace sont proposées, afin de relâcher l'hypothèse de quasi-orthogonalité et permettre aux signaux sources de partager les composantes modales communes. Ces approches permettent d'améliorer la qualité de séparation et de diminuer le coût de calcul de notre méthode.

# La SAS audio utilisant la décomposition modale : mélanges convolutifs

---

## 4.1 Introduction

Dans ce chapitre, nous proposons une approche inspirée de la méthode MD-UBSS présentée précédemment, pour la résolution du problème de la SAS dans le cas du mélange convolutif. Cette approche se base sur la décomposition modale des signaux observés en utilisant la technique ESPRIT, en supposant que les signaux sources peuvent être modélisés par une somme de sinusoides amorties et que le produit de convolution préserve les différents modes des signaux sources.

Nous présenterons une première version, qui permet de traiter le cas du mélange convolutif sous-déterminé. Dans cette version un pré-traitement s'impose, en faisant précéder l'algorithme de séparation par une étape d'identification du canal de mélange. Quant à la deuxième version, elle conjugue la décomposition modale et un critère de parcimonie basé sur la norme  $\ell_p$  pour traiter le cas convolutif sur-déterminé.

## 4.2 Formalisation du problème et hypothèses

Dans ce chapitre nous adopterons la modélisation du mélange convolutif exposée par l'équation (1.5) :

$$\mathbf{x}(t) = \sum_{k=0}^{\mathcal{K}} \mathbf{H}(k) \mathbf{s}(t-k) + \mathbf{w}(t), \quad (4.1)$$

où  $\forall k \in [0, \mathcal{K}]$ ,  $\mathbf{H}(k)$  est une matrice de taille  $M \times N$ . Les signaux sources sont supposés, comme dans le cas instantané décomposable en somme de sinusoides amorties (hypothèse 3.2) satisfaisant approximativement l'hypothèse de quasi-orthogonalité (hypothèse 3.3). Le canal satisfait l'hypothèse de diversité suivante.

**Hypothèse 4.1** *Le canal est tel que chaque vecteur colonne de*

$$\mathbf{H}(z) \triangleq \sum_{k=0}^{\mathcal{K}} \mathbf{H}(k) z^{-k} \triangleq [\mathbf{h}_1(z), \dots, \mathbf{h}_N(z)]$$

*est irréductible, c.-à-d. les entrées de  $\mathbf{h}_i(z)$  notées  $h_{ij}(z)$ ,  $j = 1, \dots, M$ , n'ont pas de zéros en commun  $\forall i$ . D'ailleurs, toute paire de vecteurs colonnes de  $\mathbf{H}(z)$  forme une matrice polynômiale  $\tilde{\mathbf{H}}(z)$  irréductible ;  $\text{Range}(\tilde{\mathbf{H}}(z)) = 2 \quad \forall z$ .*

Sachant que le produit de convolution préserve les différents modes du signal, nous exploiterons cette propriété pour extraire les différentes composantes modales des signaux sources en utilisant l'algorithme ESPRIT présenté précédemment dans le cas du mélange instantané. Cependant, en utilisant l'hypothèse de quasi-orthogonalité, la corrélation d'une composante modale donnée correspondant à un pôle  $z_i^j$  issue de la source  $s_i$  avec le signal observé  $\mathbf{x}(t)$  mène à une estimation du vecteur  $\mathbf{h}_i(z_i^j)$ . Par conséquent, deux composantes des pôles respectifs  $z_i^j$  et  $z_i^k$  issues du même signal source  $s_i$  produiront les directions spatiales  $\mathbf{h}_i(z_i^j)$  et  $\mathbf{h}_i(z_i^k)$  qui ne sont pas forcément colinéaires. C'est pourquoi, la méthode de classification des composantes modales employée dans le cas des mélanges instantanés ne pourra pas être appliquée dans le contexte convolutif.

Dans la perspective de résoudre ce problème, nous proposons deux techniques différentes, la première nécessite une identification de la réponse impulsionnelle du canal de mélange et permet de traiter le cas du mélange convolutif sous-déterminé. En ce qui concerne la deuxième, elle utilise un critère de parcimonie approprié afin de résoudre le problème du mélange convolutif sur-déterminé.



### 4.3 L'algorithme MD-UBSS convolutif

On considère dans cette section le cas du mélange convolutif sous-déterminé et afin de résoudre ce problème, il est nécessaire d'identifier le canal de mélange. Le problème d'identification du canal dans le cas sur-déterminé est un problème très difficile et devient presque impossible dans le cas sous-déterminé sans informations supplémentaire sur les sources considérées. Dans ce travail et de manière similaire à [116], on exploitera la propriété de parcimonie des signaux audio en supposant que de temps en temps seulement une seule source est présente dans le mélange. En d'autres termes, on considère l'hypothèse suivante :

**Hypothèse 4.2** *Il existe, périodiquement, des intervalles de temps où seulement une seule source est présente dans le mélange. Ceci est vérifié pour tous les signaux sources du mélange considéré (voir figure 4.1).*

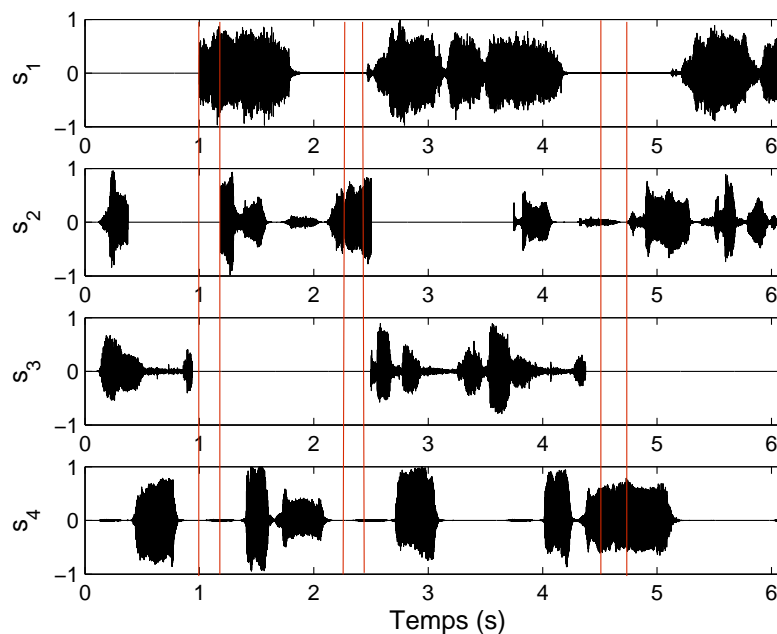


Fig. 4.1 – Représentation temporelle de 4 signaux audio : cette représentation illustre la propriété de parcimonie des signaux audio (il existe des intervalles de temps où une seule source est présente).

Afin de détecter ces intervalles de temps, on propose d'utiliser un critère de test d'information afin d'estimer le nombre de sources présentes dans le mélange. Une solution alternative utilise la technique de sélection de trame dans [117, 118] qui

Tab. 4.1 – Algorithme MD-UBSS dans le cas du mélange convolutif

1. Estimer le canal ; critère AIC [119] pour détecter le nombre de sources et appliquer un algorithme d'identification aveugle du canal [120, 121] pour estimer la réponse impulsionnelle, suivi d'une classification vectorielle des réponses impulsionnelles estimées.
2. Extraire toutes les composants harmoniques de chaque capteur en appliquant la décomposition modale paramétrique (technique ESPRIT).
3. Estimer les directions spatiales par l'équation (4.13).
4. Estimer les sources en regroupant ensemble les composantes modales correspondantes à la même source (canal).
5. Sélectionner les sources par l'équation (3.12).

exploite la structure de la fonction de densité spectrale des observations. L'algorithme MD-UBSS dans le cas du mélange convolutif est résumé par le tableau 4.1.

### 4.3.1 Estimation du canal de mélange

En nous basant sur l'hypothèse 4.2, nous proposons dans cette section d'appliquer des techniques d'identification SIMO (*Single Input Multiple Output*) pour l'estimation aveugle de la réponse impulsionnelle du canal. Il en résulte que pour résoudre le problème principal d'identification, il est nécessaire de résoudre trois sous-problèmes : Premièrement, on sélectionnera les intervalles de temps où une seule source est effectivement présente dans le mélange. Puis, pour chaque intervalle de temps sélectionné, on appliquera une technique d'identification aveugle SIMO pour estimer les paramètres du canal. Enfin, en appliquant l'approche précédente (détection des intervalles et identification SIMO), un même canal peut être estimé plusieurs fois. C'est pourquoi, on doit regrouper ensemble (classifier) les canaux estimés en  $N$  classes correspondant aux  $N$  canaux sources.

### Estimation du nombre de sources

On définit le vecteur spatio-temporel suivant :

$$\mathbf{x}_d(t) = [\mathbf{x}^T(t), \dots, \mathbf{x}^T(t-d+1)]^T = \sum_{k=1}^N \mathbf{H}_k \mathbf{s}_k(t) + \mathbf{w}_d(t), \quad (4.2)$$

où  $\mathbf{H}_k$  sont des matrices de Sylvester par bloc de taille  $dM \times (d + \mathcal{K})$  et  $\mathbf{s}_k(t) \triangleq [s_k(t), \dots, s_k(t - K - d + 1)]^T$ .  $d$  est la taille de la fenêtre de traitement choisie. Sous l'hypothèse que les canaux n'ont pas de zéros en commun et pour une taille de fenêtre assez grande (voir [119] pour plus de détails), les matrices  $\mathbf{H}_k$  sont de rang colonne plein. Par conséquent, dans le cas sans bruit, le rang de la matrice de covariance des observations  $\mathbf{R}_d \triangleq \mathbb{E}[\mathbf{x}_d(t)\mathbf{x}_d(t)^H]$  est égal au  $\min\{q(d + \mathcal{K}), dM\}$  où  $q$  représente le nombre de sources présentes dans l'intervalle de temps sur le quel a été estimée la matrice de covariance. En particulier, pour  $q = 1$ , on aura la valeur minimale du rang égale à  $(d + \mathcal{K})$ .

Donc, notre approche consiste à estimer le rang de la matrice de covariance  $\mathbf{R}_d$ , estimée sur plusieurs intervalles de temps et sélectionner ceux qui correspondent à la plus petite valeur du rang  $r = d + \mathcal{K}$ . Dans le cas où  $p$  sources sont actives (présentes) dans l'intervalle de temps considéré, le rang serait  $r = q(d + \mathcal{K})$  et par conséquent  $q$  peut être estimé comme l'entier le plus proche de  $\frac{r}{d + \mathcal{K}}$ .

L'estimation de la valeur du rang est réalisée par le critère d'information d'Akaike (AIC) [119] selon :

$$r = \arg \min_k \left[ -2 \log \left( \frac{\prod_{i=k+1}^{Md} \lambda_i^{1/(Md-k)}}{\frac{1}{Md-k} \sum_{i=k+1}^{Md} \lambda_i} \right)^{(Md-k)T_s} + 2k(2Md - k) \right], \quad (4.3)$$

ou bien par le critère d'information MDL [119] selon :

$$r = \arg \min_k \left[ -\log \left( \frac{\prod_{i=k+1}^{Md} \lambda_i^{1/(Md-k)}}{\frac{1}{Md-k} \sum_{i=k+1}^{Md} \lambda_i} \right)^{(Md-k)T_s} + \frac{1}{2}k(2Md - k) \log N \right], \quad (4.4)$$

où  $\lambda_1 \geq \dots \geq \lambda_{Md}$  représentent les valeurs propres de  $\mathbf{R}_d$  et  $T_s$  est la taille des intervalles de temps. Notons qu'à cette étape il n'est pas nécessaire de connaître exactement la taille du canal  $\mathcal{K}$  tant que  $d > \mathcal{K}$  (une sur-estimation de la taille du canal est suffisante). Dans ce cas la présence d'une seule source est caractérisée par :

$$d < r < 2d.$$

La figure 4.2 illustre l'efficacité de la méthode proposée sur un enregistrement de 6 secondes de  $M = 3$  mélanges convolutifs de  $N = 4$  sources. La fréquence d'échantillonnage est à 8 KHz et la taille des intervalles de temps  $T_s = 200$  échantillons. Les coefficients du filtre sont choisis aléatoirement et l'ordre du canal est  $\mathcal{K} = 6$ . On peut observer que le cas  $q = 1$  (une seule source) représente approximativement 10% des intervalles de temps considérées dans ce contexte.

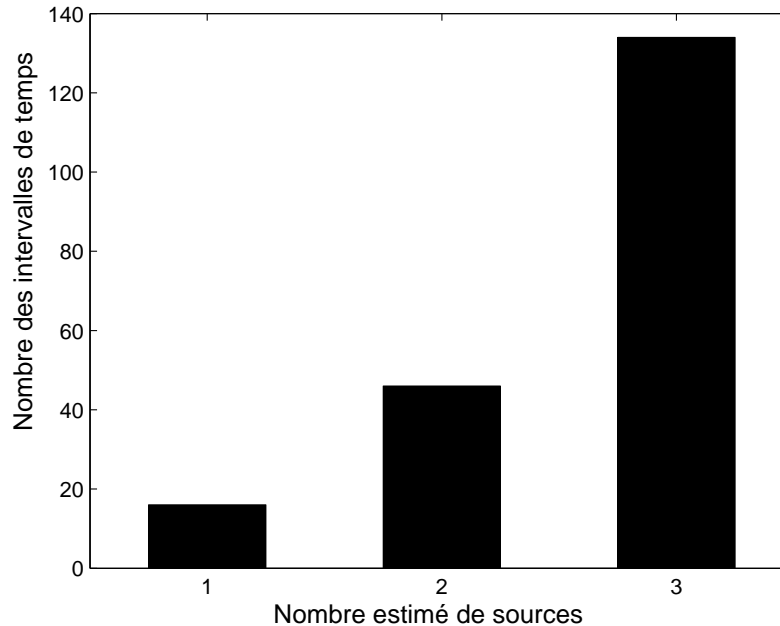


Fig. 4.2 – Histogramme représentant le nombre des intervalles de temps pour chaque nombre estimé de sources pour 4 sources audio et 3 capteurs dans le cas du mélange convolutif.

### Identification aveugle de canal

Pour réaliser l'identification aveugle du canal, nous avons utilisé dans ce chapitre l'algorithme des relations croisées (*Cross-Relation* (CR)) décrit dans [120, 121]. Considérons un intervalle de temps où une seule source  $s_i$  est présente dans le mélange. Dans ce cas, on peut considérer le système SIMO de  $M$  sorties donné par :

$$\mathbf{x}(t) = \sum_{k=0}^{\mathcal{K}} \mathbf{h}_i(k) s_i(t - k) + \mathbf{w}(t) , \quad (4.5)$$

où  $\mathbf{h}_i(k) = [h_{i1}(k) \dots h_{iM}(k)]^T$ ,  $k = 0, \dots, \mathcal{K}$ . A partir de (4.5), les sorties non bruitées  $x_j(k)$ ,  $1 \leq j \leq M$  sont données par :

$$x_j(k) = h_{ij}(k) * s_i(k), \quad 1 \leq j \leq M, \quad (4.6)$$

où "\*" dénote le produit de convolution. En utilisant la propriété de commutativité de la convolution, il en résulte que :

$$h_{il}(k) * x_j(k) = h_{ij}(k) * x_l(k), \quad 1 \leq j < l \leq M. \quad (4.7)$$

Cette équation linéaire est satisfaite par chaque paire de canaux. Il a été montré que réciproquement, les  $M(M-1)/2$  relations croisées précédentes, caractérisent de manière unique les paramètres du canal. Par conséquent, nous pouvons établir le théorème suivant [120] :

**Théorème 4.1** *Sous l'hypothèse que les canaux n'ont pas de zéros en commun, l'ensemble des relations croisées (dans le cas sans bruit) :*

$$x_l(k) * h'_j(k) - x_j(k) * h'_l(k) = 0, \quad 1 \leq l < j \leq M, \quad (4.8)$$

où  $\mathbf{h}'(z) = [h'_1(z) \dots h'_M(z)]^T$  est un vecteur polynomial de taille  $M \times 1$  et de degré  $\mathcal{K}$ , est satisfait ssi  $\mathbf{h}'(z) = \alpha \mathbf{h}_i(z)$  pour une constante scalaire donnée  $\alpha$ .

En collectant toutes les paires possibles des  $M$  canaux, on peut établir un ensemble d'équations linéaires. Sous forme matricielle, cet ensemble d'équations s'écrit comme :

$$\mathcal{X}_M \mathbf{h}_i = \mathbf{0}, \quad (4.9)$$

où  $\mathbf{h}_i \triangleq [h_{i1}(0) \dots h_{i1}(\mathcal{K}), \dots, h_{iM}(0) \dots h_{iM}(\mathcal{K})]^T$  et  $\mathcal{X}_M$  est définie par :

$$\mathcal{X}_n = \left[ \begin{array}{cc|c} \mathbf{X}_2 & -\mathbf{X}_1 & \\ \hline \mathbf{X}_{n-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}_n & \mathbf{0} & -\mathbf{X}_1 \\ & \ddots & \vdots \\ \mathbf{0} & \mathbf{X}_n & -\mathbf{X}_{n-1} \end{array} \right], \quad (4.10)$$

avec  $n = 3, \dots, M$  et :

$$\mathbf{X}_n = \begin{bmatrix} x_n(\mathcal{K}) & \dots & x_n(0) \\ \vdots & & \vdots \\ x_n(T_s - 1) & \dots & x_n(T_s - \mathcal{K} - 1) \end{bmatrix}. \quad (4.11)$$

En présence de bruit, l'équation (4.9) peut être résolue au sens des moindres carrés selon :

$$\hat{\mathbf{h}}_i = \arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^H \mathbf{X}_M^H \mathbf{X}_M \mathbf{h} , \quad (4.12)$$

dont la solution correspond au vecteur propre associé à la plus petite valeur propre de la matrice  $\mathbf{X}_M^H \mathbf{X}_M$ .

**Remarque :** Nous avons présenté dans ce chapitre la version basique de la méthode CR. Dans [122] une version améliorée de cette méthode est proposée (introduite dans un schéma adaptatif), exploitant la nature quasi-parcimonieuse des réponses impulsionnelles acoustiques.

### Classification des canaux estimés

La première étape de notre méthode d'identification de canal consiste à détecter les intervalles de temps où une seule source est effectivement présente dans le mélange. Cependant, le même signal source  $s_i$  peut être présent seul dans plusieurs intervalles de temps (voir figure 4.1 et figure 4.2), ceci impliquera qu'on estimera plusieurs version du même vecteur canal  $\mathbf{h}_i$ . Nous nous retrouverons finalement, avec plusieurs estimées de chaque canal source que nous devons regrouper ensemble en  $N$  classes. Ceci est réalisé par classification des vecteurs estimés en utilisant l'algorithme  $k$ -means. L'estimée du  $i^e$  canal est évaluée comme le centroïde de la  $i^e$  classe.

### 4.3.2 Association des composantes modales et estimation des sources

Pour la synthèse des signaux sources, on observe que l'hypothèse de quasi-orthogonalité conduit à :

$$\hat{\mathbf{h}}_i^j = \frac{\langle \mathbf{x} | \tilde{\mathcal{C}}_i^j \rangle}{\|\tilde{\mathcal{C}}_i^j\|^2} \propto \mathbf{h}_i(z_i^j) , \quad (4.13)$$

où  $z_i^j = e^{d_i^j + j2\pi f_i^j}$  est le pôle associé à la composante  $\tilde{\mathcal{C}}_i^j$ ;  $\tilde{\mathcal{C}}_i^j(t) = \Re \{ \alpha_i^j (z_i^j)^t \}$ . Par conséquent, nous proposons de regrouper ces composantes en minimisant le

critère<sup>1</sup> :

$$\hat{c}_i^j \in \mathcal{C}_i \Leftrightarrow i = \arg \min_l \left\{ \min_{\alpha} \|\hat{\mathbf{h}}_i^j - \alpha \mathbf{h}_l(z_i^j)\|^2 \right\} \quad (4.14)$$

$$i = \arg \min_l \left\{ \|\hat{\mathbf{h}}_i^j\|^2 - \frac{|\mathbf{h}_l^H(z_i^j) \hat{\mathbf{h}}_i^j|^2}{\|\mathbf{h}_l(z_i^j)\|^2} \right\} \quad (4.15)$$

où  $\mathbf{h}_l$  est la  $l^{\text{e}}$  colonne de la matrice  $\mathbf{H}$  estimée dans la section 4.3.1 et  $\mathbf{h}_l(z_i^k)$  est calculé par :

$$\mathbf{h}_l(z_i^j) = \sum_{k=0}^{\mathcal{K}} \mathbf{h}_l(k) (z_i^j)^{-k}. \quad (4.16)$$

On pourra reconstruire les sources initiales à une constante près en additionnant toutes les composantes appartenant à une même classe en utilisant (3.11).

De manière similaire au cas du mélange instantané, une composante modale peut être assignée à deux signaux sources ou plus, ce qui permettra de relâcher l'hypothèse de quasi-orthogonalité et améliorera la qualité des estimées à moyen et fort RSB (voir section 3.3.3).

## 4.4 L'algorithme MD-BSS convolutif sur-déterminé

Dans cette section, nous présenterons une autre technique de SAS basée sur la décomposition modale, destinée au cas du mélange convolutif sur-déterminé. En remarquant comme dans la section précédente que le produit de convolution préserve les différents modes du signal, nous pourrions donc utiliser la technique de décomposition ESPRIT pour extraire les composantes modales à partir des différents mélanges. En se basant sur l'hypothèse que les signaux audio sont parcimonieux dans le domaine modal (hypothèse 3.3), nous pourrions donc estimer nos signaux sources en exploitant cette propriété grâce à un critère de mesure de parcimonie adéquat.

### 4.4.1 Synthèse des signaux

En considérant le modèle du mélange convolutif (4.1) et pour une fenêtre d'observation  $g$  assez large, on a :

$$\mathbf{x}_g(t) \triangleq [\mathbf{x}^T(t) \dots \mathbf{x}^T(t+g-1)]^T = \mathcal{H} \mathbf{s}_g(t), \quad (4.17)$$

<sup>1</sup>Nous minimisons par rapport un scalaire  $\alpha$  à cause de l'indétermination d'échelle inhérente à l'identification aveugle de canal ;  $\mathbf{h}_i(z)$  est estimé à une constante scalaire près comme le montre le théorème 4.1

où  $\mathcal{H}$  est une matrice Sylvester par bloc de rang colonne plein (voir [123] pour plus de détails) et  $\mathbf{s}_g(t) = [\mathbf{s}^T(t - \mathcal{K}) \dots \mathbf{s}^T(t + g - 1)]^T$ . Donc, la matrice de données  $\mathbf{X}_g = [\mathbf{x}_g(0) \dots \mathbf{x}_g(T - g)]$  est donnée par :

$$\mathbf{X}_g = \mathcal{H} \mathbf{S}_g . \quad (4.18)$$

Cette structure suggère d'exploiter la méthode du sous-espace signal [124] pour estimer les signaux sources. Plus précisément, en utilisant une décomposition en valeurs singulières de la matrice  $\mathbf{X}_g$ , on peut écrire (dans le cas sans bruit) :

$$\mathbf{X}_g = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix} , \quad (4.19)$$

où  $\mathbf{V}_n$  est la base du sous-espace orthogonal à l'image de  $\mathbf{S}_g^H$ . Ainsi,  $\mathbf{V}_n$  satisfait :

$$\mathbf{S}_g \mathbf{V}_n = \mathbf{0} . \quad (4.20)$$

En utilisant la structure Hankel par bloc de  $\mathbf{S}_g$ , on peut transformer l'égalité ci-dessus en :

$$\mathbf{V}_n \mathbf{S}_T = \mathbf{0} , \quad (4.21)$$

où  $\mathbf{S}_T = [\mathbf{s}(-\mathcal{K}) \dots \mathbf{s}(T - 1)]^T$  et  $\mathbf{V}_n$  est la matrice de projection sur le sous-espace bruit construite à partir de  $\mathbf{V}_n$  comme il est montré dans [124]. En utilisant la décomposition modale des signaux sources (3.1), on peut écrire que :

$$\mathbf{S}_T = \mathbf{C} \boldsymbol{\Lambda} , \quad (4.22)$$

où  $\mathbf{C}$  est la matrice dont les vecteurs colonnes correspondent aux composantes modales :

$$\mathbf{c}_i^j = [c_i^j(-\mathcal{K}) \dots c_i^j(T - 1)]^T .$$

Par conséquent, l'équation (4.21) devient :

$$\mathbf{V}_n \mathbf{C} \boldsymbol{\Lambda} = \mathbf{0} . \quad (4.23)$$

En d'autres termes  $\boldsymbol{\Lambda}$  appartient au noyau de :

$$\mathbf{Q} \triangleq \mathbf{C}^H \mathbf{V}_n^H \mathbf{V}_n \mathbf{C} .$$

Soit  $\mathbf{P}$  la matrice dont les  $N$  vecteurs colonnes forment une base du noyau de  $\mathbf{Q}$ . Donc,  $\boldsymbol{\Lambda}$  satisfait :

$$\boldsymbol{\Lambda} = \mathbf{P} \tilde{\boldsymbol{\Lambda}} , \quad (4.24)$$

où  $\tilde{\boldsymbol{\Lambda}}$  est une matrice inconnue de taille  $N \times N$ . Ceci signifie que la solution par sous-espace fournit un mélange instantané des signaux sources. Afin d'estimer la matrice  $\tilde{\boldsymbol{\Lambda}}$ , nous proposons d'utiliser un critère de parcimonie que nous allons détailler dans la section suivante.



## 4.4.2 Critère de parcimonie

Nous remarquons, que sous l'hypothèse 3.2, chaque signal source  $s_i$  est une combinaison linéaire d'un nombre réduit à  $l_i$  de composantes modales. En d'autres termes, les vecteurs colonnes de  $\Lambda$  sont parcimonieux au sens que leurs éléments sont nuls sauf pour un nombre réduit d'éléments.

Maintenant, pour trouver la matrice  $\Lambda$  appropriée, on exploitera cette propriété en cherchant une matrice non-singulière  $\tilde{\Lambda}$  de taille  $N \times N$  telle que  $\Lambda$  soit parcimonieuse. Pour garantir que tous les signaux sources soient extraits (la non-singularité de la matrice  $\tilde{\Lambda}$ ), on procède comme suit :

Sans perte de généralité, on suppose que les signaux sources sont décorrélés et de norme unité vérifiant :

$$\frac{1}{T} \mathbf{S}_T^H \mathbf{S}_T \approx \mathbf{I}, \quad (4.25)$$

par conséquent,

$$\frac{1}{T} \tilde{\Lambda}^H \mathbf{P}^H \mathbf{C}^H \mathbf{C} \mathbf{P} \tilde{\Lambda} \approx \mathbf{I}. \quad (4.26)$$

Donc  $\tilde{\Lambda}$  peut être estimée à une matrice unitaire près  $\mathbf{U}$ , comme l'inverse de la racine carrée de :

$$\mathbf{Y} = \frac{1}{T} \mathbf{P}^H \mathbf{C}^H \mathbf{C} \mathbf{P}, \quad (4.27)$$

$$\tilde{\Lambda} = \mathbf{Y}^{-\frac{1}{2}} \mathbf{U}. \quad (4.28)$$

Pour obtenir la matrice unitaire restante  $\mathbf{U}$ , on utilise la propriété de parcimonie de la matrice  $\Lambda$ . On estime  $\mathbf{U}$  d'une manière telle qu'on minimise :

$$\|\Lambda\|_p = \|\mathbf{P} \mathbf{Y}^{-\frac{1}{2}} \mathbf{U}\|_p, \quad (4.29)$$

où  $\|\cdot\|_p$  représente la norme  $\ell_p$  pour  $p < 2$  qui est connue pour être une bonne mesure de parcimonie [125]. Pour minimiser l'équation (4.29) sous la contrainte unitaire, on décompose  $\mathbf{U}$  comme un produit de rotation de Givens de sorte que nous minimiserons le critère itérativement de la même manière que dans [59] où à chaque itération une seule rotation scalaire est optimisée en utilisant une technique de recherche en ligne.

Dans la figure 4.3, nous présentons un exemple d'application de notre algorithme de séparation sur un mélange convolutif. Pour cela on considère un système de  $M = 4$  capteurs recevant un mélange de  $N = 3$  signaux audio sans bruit additif. Les coefficients du filtre de mélange sont choisis de manière aléatoire et son ordre  $\mathcal{K} = 4$ . La première ligne représente les signaux originaux, les deux lignes suivantes représentent les  $M$  mélanges et la dernière représente les signaux estimés par notre algorithme.

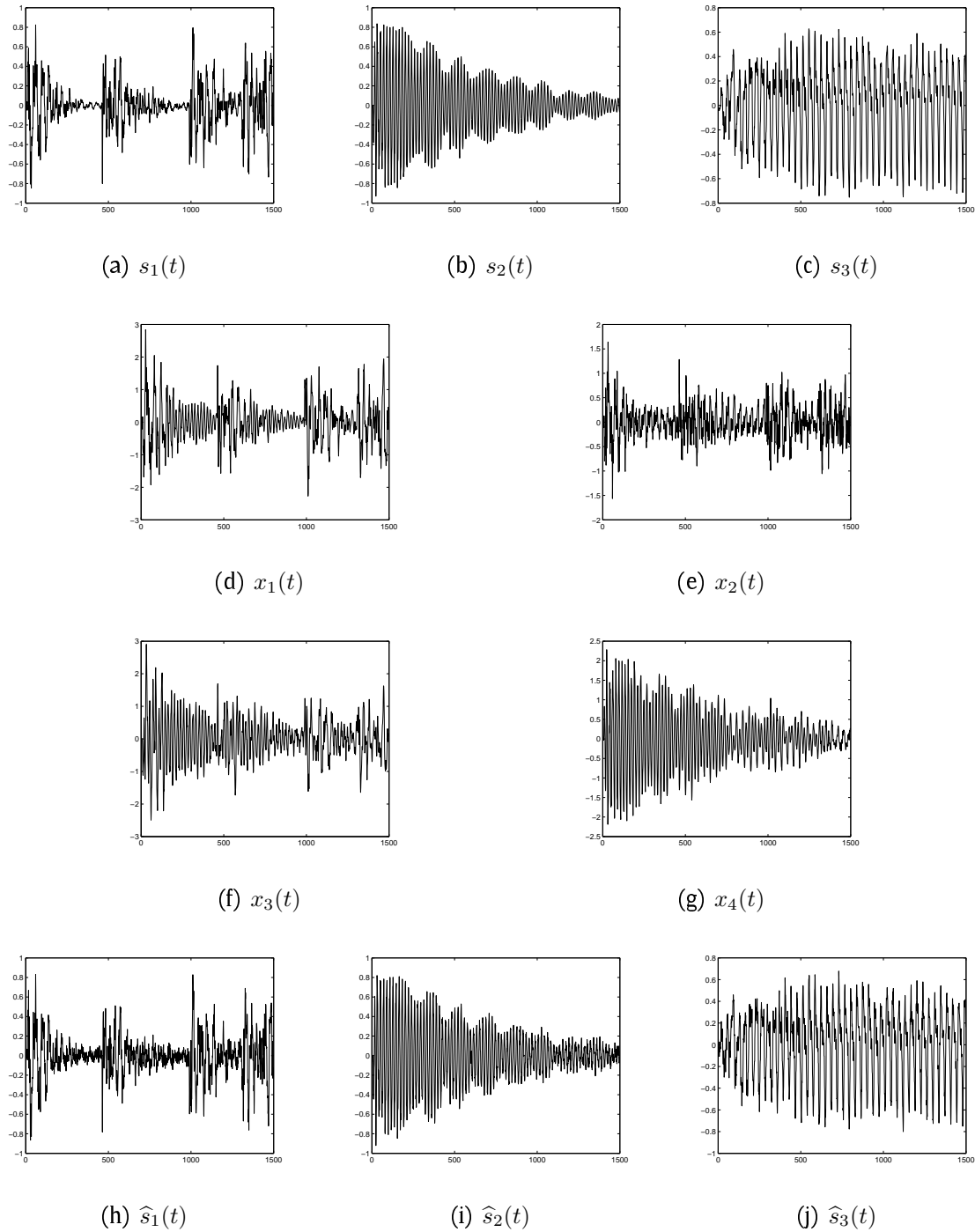


Fig. 4.3 – Exemple de séparation aveugle de sources pour 3 sources audio et 4 capteurs dans le cas du mélange convolutif : la première ligne représente les signaux originaux, les deux lignes suivantes représentent les  $M$  mélanges et la dernière représente les signaux estimés par notre algorithme.

## 4.5 Discussion

Nous discutons dans cette section certains points relatifs à l'algorithme MD-UBSS convolutif. Nous noterons que plusieurs points de discussion étudiés dans la section 3.5 peuvent être repris pour cet algorithme, notamment les discussions sur l'estimation du nombre de sources, le nombre des composantes modales et le cas sans bruit.

### Identification de canal SIMO et MIMO

Nous avons opté dans notre travail pour l'estimation du canal de mélange en utilisant les techniques SIMO. Cependant, il est aussi possible d'estimer les canaux en utilisant les techniques sur-déterminées MIMO en considérant les intervalles de temps où le nombre de sources actives est inférieur au nombre de capteurs, au lieu de considérer uniquement ceux où une seule source est effectivement présente. L'avantage de cette procédure est que le nombre d'intervalles de temps pris en considération est plus important (voir figure 4.1). L'inconvénient réside dans le fait que l'identification aveugle des systèmes MIMO est plus difficile que dans le cas SIMO, et introduit donc une plus grande erreur d'estimation (voir figure 4.4).

## 4.6 Simulations

Nous présentons dans cette section, quelques résultats de simulation pour illustrer l'efficacité de notre algorithme de séparation MD-UBSS convolutif. Pour cela, nous considérons une antenne composée de  $M = 3$  capteurs recevant  $N = 4$  signaux sources audio. La taille des observations  $T = 10000$  échantillons (les signaux sources sont échantillonnés à une fréquence de 8 KHz). Les signaux observés sont corrompus par un bruit blanc additif de covariance  $\sigma^2 \mathbf{I}_M$  ( $\sigma^2$  étant la puissance du bruit). La qualité de la séparation est mesurée par l'erreur quadratique moyenne normalisée (EQMN) des sources estimées pour  $N_r = 100$  réalisations aléatoires du bruit. Le canal de mélange  $\mathbf{H}$  est choisi tel que  $\mathcal{K} = 6$ . Dans la figure 4.4, on compare les performances de la méthode d'identification de canal basée sur le modèle SIMO (dans ce cas on choisit uniquement les intervalles de temps où une seule source est présente en utilisant le critère AIC) et celle utilisant les modèles SIMO et MIMO (dans ce cas on choisit les intervalles de temps où le système est sur-déterminé  $p < M$ ). On observe que l'identification basée sur le modèle

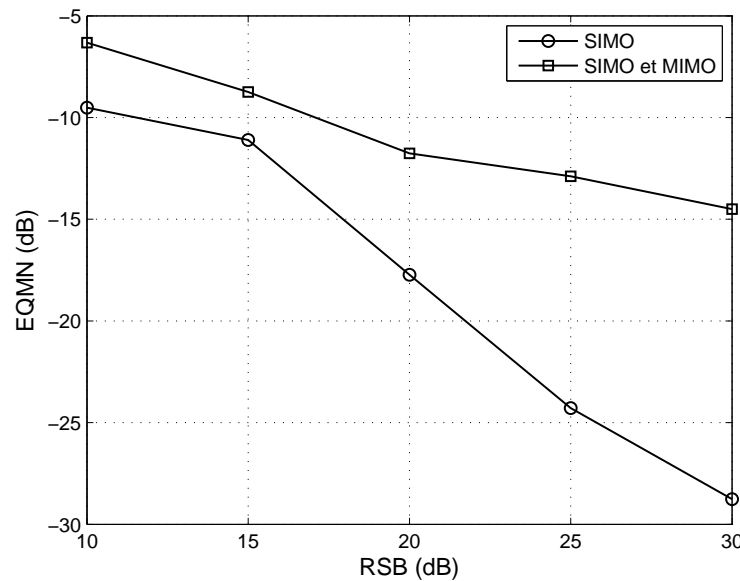


Fig. 4.4 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 10000$  échantillons dans le cas du mélange convolutif : comparaison des performances de la méthode d'identification de canal basée sur le modèle SIMO et celle utilisant les modèles SIMO et MIMO.

SIMO produit de meilleurs résultats que ceux obtenus par la méthode utilisant les modèles SIMO et MIMO, ce qui confirme l'analyse établie dans la section 4.5. La figure 4.5 présente les performances de séparation dans le cas du mélange convolutif quand la réponse impulsionnelle du canal  $\mathbf{H}$  est connue, comparées à celles obtenues avec un canal approximatif  $\overline{\mathbf{H}} = \mathbf{H} + \delta\mathbf{H}$ , où les éléments de  $\delta\mathbf{H}$  sont *i.i.d* suivant une distribution Gaussienne. Ceci est réalisé pour différentes valeurs de l'erreur quadratique moyenne normalisée du canal (EQMNC) définie par :

$$\text{EQMNC} = 10 \log \frac{\|\mathbf{H} - \overline{\mathbf{H}}\|^2}{\|\mathbf{H}\|^2}. \quad (4.30)$$

Il est clair que la qualité de séparation dépend fortement de la qualité d'estimation du canal.

La figure 4.6 représente les performances de séparation de notre algorithme MD-UBSS quand on utilise la réponse impulsionnelle exacte du canal  $\mathbf{H}$ , comparées à celles obtenues avec notre méthode d'estimation du canal en utilisant un modèle SIMO. Pour des valeurs de RSB supérieures à 20dB, l'estimation du canal est suffisamment bonne pour obtenir presque la même qualité de séparation qu'avec le canal exact. Curieusement, à un RSB = 20dB, l'EQMN du canal estimé est approxi-

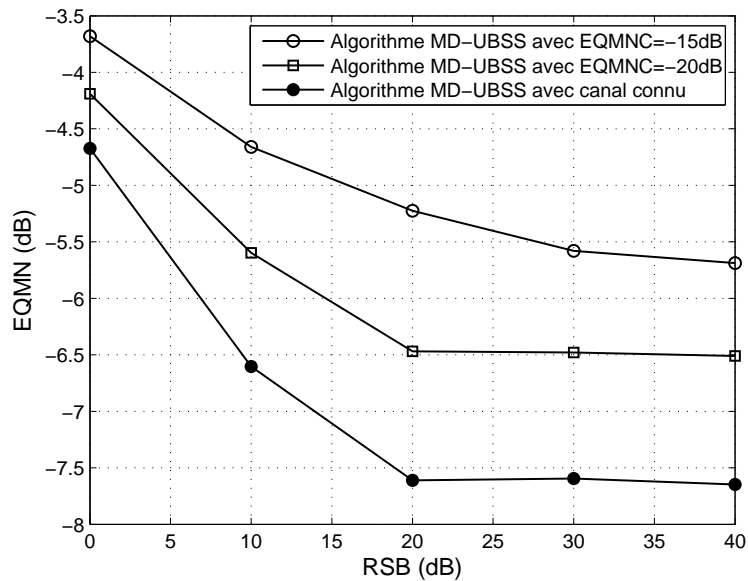


Fig. 4.5 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 10000$  échantillons dans le cas du mélange convolutif : comparaison des performances de l'algorithmes MD-UBSS dans le cas où la réponse impulsionnelle du canal  $\mathbf{H}$  est connue et lorsque elle est perturbée par un bruit Gaussien pour différentes valeurs du EQMNC.

mativement égale à  $-18\text{dB}$  (voir figure 4.4) un niveau d'erreur correspondant à une dégradation non-négligeable montrée par la figure 4.5. Cette contradiction vient du fait que dans l'expérience de la figure 4.5, l'erreur d'estimation du canal a été supposée spatialement blanche, suivant une distribution Gaussienne. Or, l'erreur d'estimation réelle est spatialement colorée (voir par exemple [126] où une expression explicite de la covariance asymptotique de l'erreur du canal est donnée) qui semble être favorable à notre méthode de séparation.

Dans la figure 4.7, on compare les performances de séparation de notre algorithme MD-UBSS, de l'algorithme de Y. Deville dans [118] et l'algorithme de L. Parra dans [60] dans le cas sur-déterminé avec  $N = 2$  signaux sources et  $M = 2$  capteurs. Les signaux sources utilisés sont deux signaux de parole (un homme et une femme). Les algorithmes cités dans [60, 118] séparent les signaux sources à un filtre près et par conséquent, on utilise dans cette expérience le critère de rapport interférences à signal fréquentiel (RISF) défini dans [60] à défaut d'utiliser l'EQMN. On observe un gain de performances en faveur des algorithmes de Y. Deville et L. Parra et particulièrement pour les hauts RSB. Ceci s'explique par le fait que les signaux

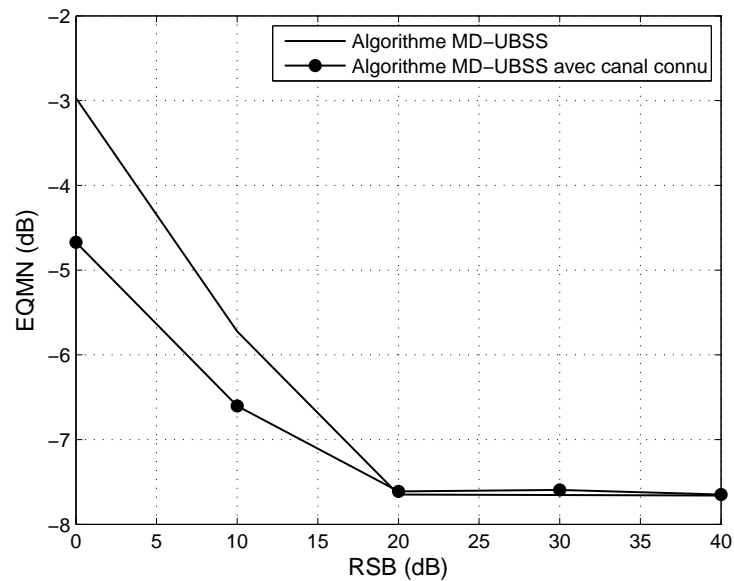


Fig. 4.6 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 10000$  dans le cas du mélange convolutif : comparaison des performances de l'algorithmes MD-UBSS dans le cas où la réponse impulsionnelle du canal  $\mathbf{H}$  est connue et lorsque elle est estimée.

utilisés ont été mal décomposés en raison de leur structure (l'hypothèse de quasi-orthogonalité n'est pas ou mal satisfaite). Une mauvaise décomposition implique forcément une mauvaise synthèse qui explique la dégradation de la qualité de la séparation.

## 4.7 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode de SAS audio dénommée MD-UBSS convolutif, utilisant la décomposition modale (ESPRIT). L'avantage majeur de la méthode introduite est sa capacité à séparer plus de sources que de capteurs dans le cas du mélange convolutif. Nous avons présenté aussi une nouvelle technique d'identification aveugle du canal de mélange, étape nécessaire à l'algorithme de séparation MD-UBSS. Pour le cas sur-déterminé nous avons introduit une version de l'algorithme MD-UBSS convolutif, utilisant un critère de parcimonie approprié conjugué à une décomposition en valeurs et vecteurs propres.

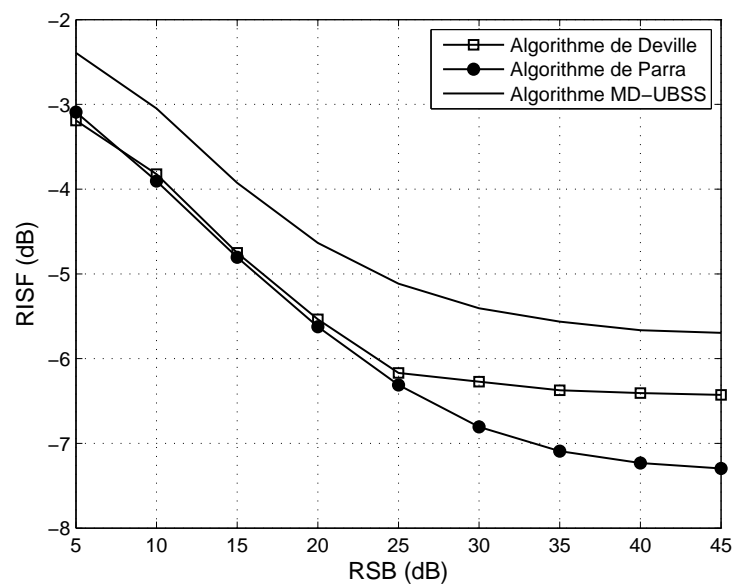


Fig. 4.7 – RISF en fonction du RSB pour  $N = 2$  sources audio et  $M = 2$  capteurs avec une taille d'observation  $T = 10000$  : comparaison des performances de séparation de l'algorithme MD-UBSS, l'algorithme de Y. Deville et l'algorithme de L. Parra.





## **Deuxième partie**

# **La SAS audio utilisant la représentation temps-fréquence**

---

Dans cette partie, nous présenterons des techniques de séparation aveugle de sources utilisant la propriété de parcimonie des signaux audio dans le domaine temps-fréquence. Ces techniques ont l'avantage de pouvoir traiter le cas sous-déterminé. Notre choix s'est porté sur cette classe de techniques en raison de leur efficacité par rapport aux techniques utilisant la décomposition modale. Par ailleurs, les travaux présentés dans cette section ont donné lieu à deux publications dans les revues IEEE Transactions on Signal Processing et IEEE Transactions on Audio, Speech & Language Processing.



# La SAS audio utilisant la représentation temps-fréquence : mélanges instantanés

---

## 5.1 Introduction

Récemment, plusieurs méthodes de SAS sous-déterminée pour des sources non-stationnaires ont été proposées, exploitant leurs propriété de parcimonie dans le domaine temps-fréquence (TF) [87, 127, 128]. La première méthode utilise les distributions temps-fréquence quadratiques DTFQ, quant à la deuxième elle utilise la transformée de Fourier à court terme TFCT. L'hypothèse principale utilisée par ces méthodes est que les signaux sources sont disjoints dans le plan TF ; en d'autres termes, il y a tout au plus une source présente à chaque point du domaine TF. Cette hypothèse est plutôt restrictive. Cependant ces méthodes ont montré qu'elles obtenaient de bons résultats sous la condition de parcimonie des signaux sources dans le domaine temps-fréquence.

Dans ce chapitre, on présentera une nouvelle méthode dénommée TF-UBSS pour *Time-Frequency Underdetermined Blind Source Separation*, qui permettra de relâcher la condition qui suppose que les signaux sources sont disjoints dans le plan TF, en supposant qu'ils sont non-disjoint ; qu'il est possible que plusieurs sources soit présentes dans le même point du plan TF. En particulier, on se limi-

tera au scénario où le nombre de sources présentes en chaque point est inférieur au nombre de capteurs. Sous cette hypothèse, la séparation des sources non-disjointes est réalisée en utilisant une projection en sous-espace. La projection en sous-espace permet d'identifier les sources qui contribuent en chaque point, et par conséquent, estimer les valeurs correspondant à chacune de ces sources.

## 5.2 Formalisation du problème et hypothèses

Dans ce chapitre, nous traiterons uniquement le problème de SAS dans le cas du mélange instantané. Pour ce faire, nous admettrons que les signaux observés sont régis par le modèle instantané défini par l'équation (1.1) :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t) ,$$

et nous supposerons, sans perte de généralité, que les vecteurs colonnes de la matrice de mélange  $\mathbf{A}$  sont de norme unité ;  $\|\mathbf{a}_i\| = 1$  pour  $i \in \mathcal{N}$ , où  $\mathcal{N} = \{1, \dots, N\}$ . Afin de réaliser la séparation de source en utilisant l'algorithme TF-UBSS, il est nécessaire que les vecteurs colonnes de la matrice de mélange  $\mathbf{A}$  soient deux à deux linéairement indépendants (hypothèse 3.1).

Dans la suite nous allons établir et définir les hypothèses et les outils nécessaires pour effectuer la séparation de sources en utilisant l'algorithme TF-UBSS.

### 5.2.1 Distributions temps-fréquence

Le traitement du signal en TF fournit des outils efficaces pour l'analyse des signaux non-stationnaires, dont la fréquence varie dans le temps. Ce concept est l'extension naturelle du traitement dans les domaines temporel et fréquentiel qui implique une représentation en deux dimensions des signaux. Ce qui permet de produire une distribution de l'énergie des signaux en fonction du temps et de la fréquence simultanément. Pour cette raison, la représentation TF est généralement désignée comme la distribution temps-fréquence (DTF).

La classe générale des DTF quadratiques d'un signal analytique  $z(t)$  est définie comme dans [129] par :

$$\rho_{zz}(t, f) \triangleq \iiint_{-\infty}^{\infty} e^{j2\pi\nu(u-t)} \Gamma(\nu, \tau) \times z(u + \frac{\tau}{2}) z^*(u - \frac{\tau}{2}) e^{-j2\pi f\tau} d\nu du d\tau , \quad (5.1)$$

où  $\Gamma(\nu, \tau)$  est une fonction de paramétrisation à deux dimensions choisie dans le plan Doppler-retard. On peut désigner une DTF avec certaines propriétés désirées en contraignant correctement  $\Gamma$ .

Quand  $\Gamma(\nu, \tau) = 1$  nous obtenons la distribution de Wigner-Ville (DWV) :

$$\rho_{zz}^{\text{dwv}}(t, f) \triangleq \int_{-\infty}^{+\infty} z\left(t + \frac{\tau}{2}\right) z^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau . \quad (5.2)$$

La DWV est la DTF la plus largement étudiée dans la littérature. Elle permet de maximiser la concentration d'énergie dans le plan TF autour de la fréquence instantanée pour les signaux à fréquence modulée linéaire (FML). Cependant, elle introduit ce qu'on appelle les "termes-croisés". Ces termes sont dus à la multiplication quadratique des versions décalées du signal.

Une autre DTF bien connue et très utilisée en pratique, est la transformée de Fourier à court terme (TFCT) :

$$\mathcal{S}_z(t, f) \triangleq \int_{-\infty}^{+\infty} z(\tau) \hbar(\tau - t) e^{-j2\pi f\tau} d\tau , \quad (5.3)$$

où  $\hbar(t)$  est la fonction fenêtre. Notons que la TFCT est une DTF linéaire<sup>1</sup>, et sa version quadratique est appelée spectrogramme (SPEC), définie par :

$$\rho_{zz}^{\text{spec}}(t, f) \triangleq |\mathcal{S}_z(t, f)|^2 . \quad (5.4)$$

Notons que dans nos méthodes de séparation, nous avons utilisé la version discrète de cette transformée, qui s'écrit comme suit :

$$\mathcal{S}_z(t, f) \triangleq \sum_{\tau=-\infty}^{+\infty} z(\tau) \hbar(\tau - t) e^{-j2\pi f\tau} , \quad (5.5)$$

Il est clair que d'après cette définition, l'effet des termes-croisés est absent dans la TFCT, et par conséquent dans le SPEC. Cependant, cette DTF à une faible résolution TF par rapport à la DWV. Mais le faible coût d'implémentation de la TFCT et donc du SPEC, en comparaison à celui de la DWV, ainsi que l'absence de l'effet des termes-croisés, justifient le fait que la TFCT est plus utilisée en pratique, et spécialement pour les signaux audio. Mais pour des signaux de type FML, il est préférable d'utiliser la DWV.

Afin de combiner la haute résolution de la DWV et la propriété d'absence de termes-croisés du SPEC, on définit la distribution Wigner-Ville masquée (DWVM) par :

$$\rho_{zz}^{\text{mwvd}}(t, f) \triangleq \rho_{zz}^{\text{wvd}}(t, f) \cdot \rho_{zz}^{\text{spec}}(t, f) . \quad (5.6)$$

<sup>1</sup>En effet, la TFCT ne représente pas la distribution d'énergie dans le plan TF. Cependant, dans un soucis de simplification, nous la considérerons comme une DTF.

Il existe plusieurs DTF dans la littérature, notamment celles qui offrent une haute résolution TF avec un minimum de termes-croisés, par exemple la  $B$ -distribution [130].

L'inconvénient majeur des DTF citées précédemment est leur résolution temporelle et fréquentielle fixe. Afin d'obtenir une résolution temps-fréquence variable, on propose d'utiliser la transformée en ondelettes (TO) [131] qui est définie comme suit :

$$\mathcal{W}_z(t, \lambda) = \frac{1}{\sqrt{\lambda}} \int_{-\infty}^{+\infty} z(\tau) \psi^* \left( \frac{\tau - t}{\lambda} \right) d\tau, \quad (5.7)$$

où  $\mathcal{W}_z(t, \lambda)$  est la représentation temps-échelle (TE) du signal  $z(t)$  à l'instant  $t$  et à l'échelle  $\lambda$  et  $\psi$  est une fonction dite famille d'ondelettes de moyenne nulle :

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0,$$

normalisée à  $\|\psi\| = 1$ , et centrée au voisinage de  $t = 0$ .

### 5.2.2 Conditions TF sur les sources

Après avoir introduit le concept du traitement du signal TF comme un outil utilisé pour l'analyse des signaux non-stationnaires, certaines conditions TF doivent être vérifiées par les signaux sources. Notons que la méthode TF introduite dans [132] ne permet pas de traiter le cas de la SAS sous-déterminé, du fait que la matrice de mélange n'est pas inversible. Afin de traiter le problème de la SAS sous-déterminé, on recherche une représentation parcimonieuse des signaux sources [133]. En d'autres termes, si les sources peuvent avoir une représentation parcimonieuse dans un certain domaine, alors la séparation doit être effectuée dans ce domaine afin d'exploiter cette propriété de parcimonie.

#### sources TF-disjointes

Récemment, plusieurs méthodes de SAS sous-déterminée ont vu le jour, notamment celles dans [127] et [87], dans lesquelles le domaine TF a été choisi comme domaine transformé afin d'accroître la propriété de parcimonie des signaux. Ces deux articles reposent sur l'hypothèse que les signaux sources sont disjointes dans le domaine TF. Mathématiquement parlant, si  $\Omega_1$  et  $\Omega_2$  sont les supports TF des deux signaux sources  $s_1(t)$  et  $s_2(t)$  respectivement, alors  $\Omega_1 \cap \Omega_2 = \emptyset$ . Cette condition peut être illustrée par la figure 5.1. Néanmoins, c'est une hypothèse plutôt stricte. Une hypothèse plus pratique, est de supposer que les sources sont

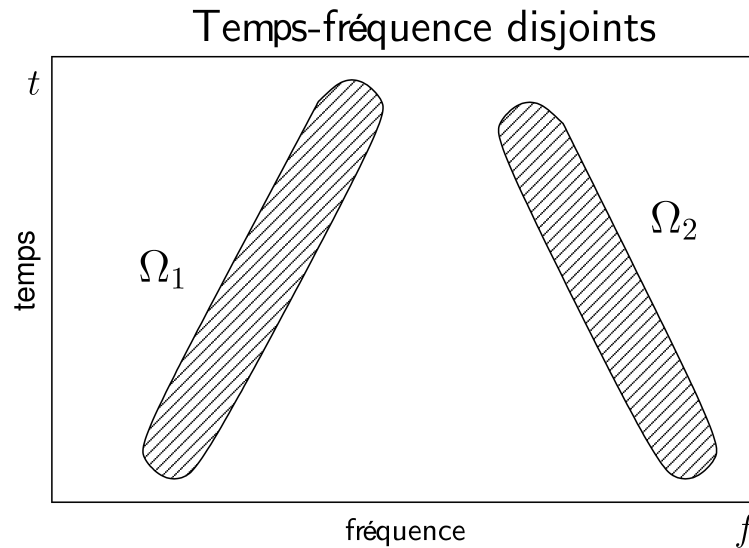


Fig. 5.1 – La condition de sources TF-disjointes :  $\Omega_1 \cap \Omega_2 = \emptyset$  (où  $\Omega_1 \cap \Omega_2 \approx \emptyset$ ).

presque-disjointes dans le domaine TF [127], en ayant quelques points de recouvrement. Pour cette hypothèse, la SAS sous-déterminée peut être effectuée en utilisant les deux méthodes citées précédemment.

### Sources TF-non-disjointes

Dans ce chapitre, nous souhaitons relâcher la condition de TF-disjointes en supposant que les sources sont non-disjointes dans le domaine TF ; comme illustré dans la figure 5.2. Ceci est motivé par les inconvénients de la méthode dans [127]. Bien que cette méthode produise de bons résultats sous la condition que les sources sont presque-disjointes dans le plan TF, elle ne traite pas explicitement les régions TF où les sources sont supposées avoir des points de recouvrement. Le point de recouvrement de deux sources est assigné de manière aléatoire à une des deux sources. Il en résulte, que la source à laquelle le point a été assigné contiendra une information supplémentaire issue de la deuxième source, tandis que cette dernière perdra cette information. Cette perte d'information peut être compensée dans une certaine mesure, par une interpolation au point de recouvrement en utilisant la synthèse TF. Cependant, l'information perdue pour la source précédente, se retrouve comme interférence dans l'autre source, ce qui amène vers une dégradation des performances de séparation si aucun traitement n'est appliqué. Si le nombre des points d'intersection augmente (la condition de TF-presque-disjointes n'est plus vérifiée), les performances de séparation sont amenées à se dégrader à moins

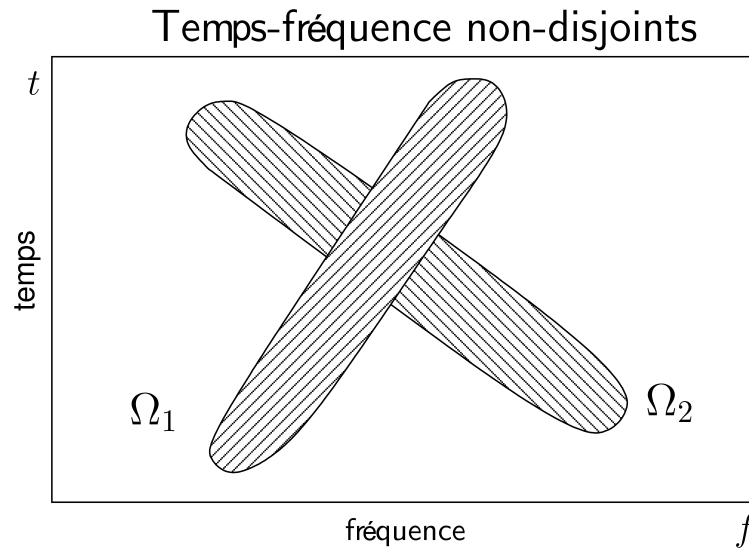


Fig. 5.2 – La condition de sources TF-non-disjointes :  $\Omega_1 \cap \Omega_2 \neq \emptyset$ .

que l'on fournisse un traitement approprié à ces points d'intersection.

Dans la suite nous proposerons une solution à ce problème en utilisant la projection en sous-espace. C'est pourquoi, nous admettrons que les sources sont non-disjointes dans le domaine TF; c'est-à-dire, que l'on autorise plusieurs sources à être présentes à chaque point du domaine TF. Toutefois, nous posons les limitations suivantes :

**Hypothèse 5.1** *Le nombre de sources dont l'énergie contribue à chaque point TF est strictement inférieure au nombre de capteurs.*

En d'autres termes, pour une configuration à  $M$  capteurs, il existe au plus  $(M - 1)$  sources à chaque point du domaine TF. Pour le cas particulier où  $M = 2$ , l'hypothèse 5.1 est réduite à la condition TF-disjointes. Nous posons aussi une autre hypothèse sur la disposition TF des sources.

**Hypothèse 5.2** *Pour chaque source, il existe des régions dans le domaine TF, où elle est présente seule.*

Notons que cette hypothèse est facilement vérifiable et non-restrictive dans le cas de signaux audio ou de signaux de type FML. D'autre part, il est nécessaire de signaler que l'hypothèse 5.2 n'est pas une restriction due à l'utilisation de la projection en sous-espace, mais elle sera utilisée dans la suite, afin d'estimer la matrice de mélange. Dans le cas peu vraisemblable où cette hypothèse n'est pas



vérifiable, la matrice de mélange peut être obtenue en utilisant d'autres méthodes, par exemple celle décrite dans [114], dans ce cas l'hypothèse 5.2 peut être omise.

## 5.3 Approche TF-UBSS basée sur la classification pour des sources TF-disjointes

### 5.3.1 Algorithme TF-UBSS quadratique basé sur la classification

Dans cette section, nous présentons la méthode proposée dans [127] basée sur l'idée de classification vectorielle où elle sera désignée par le nom *d'algorithme TF-UBSS quadratique basé sur la classification*.

Pour un signal  $\mathbf{z}(t) = [z_1(t), \dots, z_N(t)]^T$ , la matrice de distribution temps-fréquence quadratique (DTFQ) est donnée par [132] :

$$\mathbf{D}_{\mathbf{z}\mathbf{z}}(t, f) \triangleq \begin{bmatrix} \rho_{z_1 z_1}(t, f) & \cdots & \rho_{z_1 z_N}(t, f) \\ \vdots & \ddots & \vdots \\ \rho_{z_N z_1}(t, f) & \cdots & \rho_{z_N z_N}(t, f) \end{bmatrix}, \quad (5.8)$$

où, pour  $i, j \in \mathcal{N}$ ,  $\rho_{z_i z_j}(t, f)$  est la DTFQ-croisée entre  $z_i(t)$  et  $z_j(t)$  obtenue en utilisant l'équation (5.2), mais en remplaçant le premier  $z$  par  $z_i$  et le deuxième par  $z_j$ . Par définition, la DTFQ prend en compte la diversité spatiale. En appliquant la DTFQ définie par l'équation (5.8) au modèle du mélange instantané défini par l'équation (1.1) (dans le cas sans bruit), on obtient la structure TF suivante :

$$\mathbf{D}_{\mathbf{x}\mathbf{x}}(t, f) = \mathbf{A} \mathbf{D}_{\mathbf{s}\mathbf{s}}(t, f) \mathbf{A}^H, \quad (5.9)$$

où  $\mathbf{D}_{\mathbf{s}\mathbf{s}}(t, f)$  et  $\mathbf{D}_{\mathbf{x}\mathbf{x}}(t, f)$  sont respectivement, la matrice de DTFQ des sources et la matrice de DTFQ des mélanges.

Appelons *points TF auto-source* les points où se concentre la véritable énergie d'une source dans le domaine TF, et *points TF cross-source* les points où contribue une fausse énergie (due à l'effet des termes-croisés de la DTFQ). Notons que, les autres points étant sans contribution d'énergie, la valeur de leurs DTF est idéalement nulle. Sous l'hypothèse que toutes les sources sont disjointes dans le domaine TF, il y a seulement une source présente à chaque point auto-source. Donc, la structure de  $\mathbf{D}_{\mathbf{x}\mathbf{x}}(t, f)$  est réduite à :

$$\mathbf{D}_{\mathbf{x}\mathbf{x}}(t_a, f_a) = \rho_{s_i s_i}(t_a, f_a) \mathbf{a}_i \mathbf{a}_i^H, \quad \forall (t_a, f_a) \in \Omega_i, \quad (5.10)$$

Tab. 5.1 – Algorithme TF-UBSS quadratique basé sur la classification

1. Calculer la matrice de DTFQ des mélanges par (5.11).
2. Effectuer un seuillage du bruit et sélectionner les points auto-source par l'équation (5.12).
3. Classifier les directions spatiales en utilisant l'équation (5.13) et l'algorithme  $k$ -means ; Estimer les DTF des sources par l'équation (5.14).
4. Synthétiser les DTF des sources estimées [134].

où  $\Omega_i$  dénote, le support TF de la source  $s_i(t)$ .

L'observation fournit par l'équation (5.10) suggère que tous les points  $(t_a, f_a) \in \Omega_i$  et l'ensemble des matrices DTFQ  $\{\mathbf{D}_{xx}(t_a, f_a)\}$  qui leur correspondent ont le même vecteur propre principal  $\mathbf{a}_i$ . Cette observation est l'idée générale qui régit la méthode de séparation basée sur la DTFQ décrite dans [127]. En effet, l'article [127] propose plusieurs algorithmes et précise que le choix de DTF doit être fait avec précaution afin d'obtenir une représentation de la DTF des mélanges propre (sans termes-croisés), et choisit la DWVM comme meilleure candidate. Cette algorithme est résumé dans le tableau 5.1.

### Calcul de la matrice de DTFQ des mélanges

La matrice de DTFQ des mélanges basée sur la DWVM est calculée comme suit :

$$[\mathbf{D}_{xx}^{\text{dvw}}(t, f)]_{k,l} = \rho_{x_k x_l}^{\text{dvw}}(t, f) \quad (5.11a)$$

$$[\mathbf{D}_{xx}^{\text{tfct}}(t, f)]_{k,l} = \begin{cases} \mathcal{S}_{x_k}(t, f), & \text{pour } k = l, \\ 0, & \text{sinon,} \end{cases} \quad (5.11b)$$

$$\mathbf{D}_{xx}^{\text{dwvm}}(t, f) = \mathbf{D}_{xx}^{\text{dvw}}(t, f) \odot |\mathbf{D}_{xx}^{\text{tfct}}(t, f)|^2 \quad (5.11c)$$

Dans l'équation (5.11),  $k, l \in \mathcal{N}$ , et  $\odot$  dénote le produit de Hadamard.

### Seuillage du bruit et selection des points auto-source

La procédure de "seuillage du bruit" est utilisée afin de garder les points ayant suffisamment d'énergie ; les points auto-source. Une façon de réaliser cette opération

est : pour chaque trame de temps  $(t_p, f)$  de la représentation TF, appliquer le critère suivant pour tous les points de fréquence  $f_q$  appartenant à cette trame temporelle :

$$\text{Si } \frac{\|D_{xx}^{\text{dwvm}}(t_p, f_q)\|}{\max_f \{\|D_{xx}^{\text{dwvm}}(t_p, f)\|\}} > \epsilon_1, \quad \text{garder } (t_p, f_q), \quad (5.12)$$

où  $\epsilon_1$  est un seuil de faible valeur (typiquement  $\epsilon_1 = 0.05$ ). L'ensemble de tous les points auto-source est noté par  $\Omega$ . Tant que les sources sont TF-disjointes, on aura  $\Omega = \bigcup_{i=1}^N \Omega_i$ .

### Classification vectorielle et estimation des DTF des sources

Pour chaque point  $(t_a, f_a) \in \Omega$ , calculer la direction spatiale qui lui est associée  $\mathbf{a}(t_a, f_a)$  comme suit :

$$\mathbf{a}(t_a, f_a) = \frac{\text{diag}(D_{xx}^{\text{tft}}(t_a, f_a))}{\|\text{diag}(D_{xx}^{\text{tft}}(t_a, f_a))\|}, \quad (5.13)$$

et on forcera son premier élément à être réel et positif, et ceci sans perte de généralité. Une fois l'ensemble des directions spatiales  $\{\mathbf{a}(t_a, f_a) | (t_a, f_a) \in \Omega\}$  obtenu, on pourra les classifier en  $N$  classes en utilisant n'importe quel algorithme de classification non-supervisé (voir [110] pour les différentes méthodes de classifications). L'algorithme de classification utilisé dans [127] est plutôt sensible en raison de l'utilisation de différents seuils. Si le nombre de sources est supposé connu ou éventuellement bien estimé, on pourra utiliser l'algorithme de classification  $k$ -means [110] afin d'avoir de bonnes performances de classification. La sortie de l'algorithme de classification est un ensemble de  $N$  classes  $\{\mathcal{C}_i | i \in \mathcal{N}\}$ . En plus, la collecte de tous les points correspondant à tous les vecteurs de la classe  $\mathcal{C}_i$ , formera le support TF  $\Omega_i$  de la source  $s_i(t)$ . Donc, on estime la DTF de la source  $s_i(t)$  (à un constante scalaire près) comme :

$$\hat{\rho}_{s_i}^{\text{dwv}}(t, f) = \begin{cases} \text{trace}\{D_{xx}^{\text{dwv}}(t, f)\}, & (t, f) \in \Omega_i, \\ 0, & \text{sinon.} \end{cases} \quad (5.14)$$

### Synthèse TF des sources

Une fois la DTF des sources  $\hat{\rho}_{s_i}^{\text{wvd}}(t, f)$  obtenue, l'estimation des signaux sources  $s_i(t)$  peut être réalisée à travers un algorithme de synthèse TF. La méthode présentée dans [134] est utilisée pour la synthèse TF d'une estimée de la DWV, en se basant

sur la propriété d'inversion de la DWV suivante [129] :

$$x(t) = \frac{1}{x^*(0)} \int_{-\infty}^{\infty} \rho_x^{\text{dwv}}\left(\frac{t}{2}, f\right) e^{j2\pi ft} df,$$

Ce qui implique que le signal peut être reconstruit à une constante exponentielle complexe près  $e^{j\alpha} = x^*(0)/|x(0)|$  avec  $|x(0)| \neq 0$ .

On peut observer dans cette version de l'algorithme TF-UBSS quadratique, que ce ne sont pas tous les éléments des matrices de DTFQ qui sont utilisés, mais uniquement les éléments diagonaux. Ceci peut être pris en compte afin de réduire le coût de calcul.

### 5.3.2 Algorithme TF-UBSS linéaire basé sur la classification

Comme expliqué précédemment, la TFCT est souvent utilisée pour l'analyse des signaux de paroles ou audio en raison de son faible coût de calcul. Par conséquent, nous présenterons brièvement dans cette section la méthode basée sur la TFCT proposée dans [87], et nous proposerons simultanément *l'algorithme TF-UBSS linéaire basé sur la classification* utilisant la TFCT qui permet de pallier quelques faiblesses de la méthode proposée dans [87].

Premièrement, par la transformation dans le domaine TF en utilisant la TFCT, le modèle du mélange décrit par l'équation (1.1) devient (dans le cas sans bruit) :

$$\mathcal{S}_x(t, f) = \mathbf{A}\mathcal{S}_s(t, f), \quad (5.15)$$

où  $\mathcal{S}_x(t, f)$  est le vecteur TFCT des mélanges et  $\mathcal{S}_s(t, f)$  est le vecteur TFCT des sources. Sous l'hypothèse que toutes les sources sont disjointes dans le domaine TF, l'équation (5.15) est réduite à

$$\mathcal{S}_x(t_a, f_a) = \mathbf{a}_i \mathcal{S}_{s_i}(t_a, f_a), \quad \forall (t_a, f_a) \in \Omega_i, \quad \forall i \in \mathcal{N}. \quad (5.16)$$

Maintenant, dans [87], la structure de la matrice de mélange est particulière car elle a seulement deux lignes (la méthode est valable pour le cas où on dispose de 2 capteurs) et la première ligne de la matrice de mélange contient que des uns. Alors, l'équation (5.16) peut être réécrite comme :

$$\begin{bmatrix} \mathcal{S}_{x_1}(t_a, f_a) \\ \mathcal{S}_{x_2}(t_a, f_a) \end{bmatrix} = \begin{bmatrix} 1 \\ a_{2,i} \end{bmatrix} \mathcal{S}_{s_i}(t_a, f_a),$$

il en résulte que

$$a_{2,i} = \frac{\mathcal{S}_{x_2}(t_a, f_a)}{\mathcal{S}_{x_1}(t_a, f_a)}. \quad (5.17)$$

Tab. 5.2 – Algorithme TF-UBSS linéaire basé sur la classification

1. Calculer la TFCT des mélanges en utilisant l'équation (5.18); seuillage du bruit par (5.19).
2. Classifier les directions spatiales estimées par les équations (5.20) et (5.21).
3. Estimer les DTF des sources par (5.22).
4. Synthétiser les DTF des signaux estimés en utilisant la TFCT inverse [135].

Ainsi, tous les points dont les ratios calculés par l'équation (5.17) ont la même valeur, forment le support TF  $\Omega_i$  de la source  $s_i(t)$ . Donc, la TFCT estimée de la source  $s_i(t)$  est calculée par :

$$\hat{\mathcal{S}}_{s_i}(t, f) = \begin{cases} \mathcal{S}_{x_1}(t, f), & \forall (t, f) \in \Omega_i, \\ 0, & \text{sinon.} \end{cases}$$

La source estimée  $\hat{s}_i(t)$  est alors obtenue en convertissant  $\hat{\mathcal{S}}_{s_i}(t, f)$  au domaine temporel en utilisant la TFCT inverse [135]. Notons que, l'extension de la méthode de SAS sous-déterminée dans [87] à plus de deux capteurs est difficile dans ces conditions. Deuxièmement, le calcul du rapport entre les deux éléments de l'équation (5.17) est sujet à l'introduction d'erreurs si le dénominateur est proche de zéro. Afin de remédier aux problèmes mentionnés ci-dessus, nous proposons une version modifiée de la précédente méthode désignée par *algorithme TF-UBSS linéaire basé sur la classification*. A partir des observations (5.16), on peut en déduire l'algorithme de séparation présenté de ce qui suit qui est résumé dans le tableau 5.2.

### Calcul de la TFCT des mélanges et seuillage du bruit

On calcule la TFCT du mélange,  $\mathcal{S}_x(t, f)$ , en appliquant l'équation (5.3) pour chaque mélanges de  $x(t)$ , comme suit :

$$\mathcal{S}_{x_i}(t, f) = \sum_{\tau=-\infty}^{+\infty} x_i(\tau) \hat{h}(\tau - t) e^{-j2\pi f\tau}, \quad i = 1, \dots, M, \quad (5.18a)$$

$$\mathcal{S}_x(t, f) = [\mathcal{S}_{x_1}(t, f), \dots, \mathcal{S}_{x_M}(t, f)]^T. \quad (5.18b)$$

Puisque la TFCT est complètement dépourvue de termes-croisés, alors un point avec une valeur de DTF non-nulle est idéalement un point auto-source. En pratique, on sélectionne tous les points auto-source en appliquant uniquement la procédure de seuillage du bruit comme dans l'algorithme TF-UBSS quadratique basé sur la classification. Plus précisément, pour chaque trame de temps  $(t_p, f)$  de la représentation TF, on applique le critère suivant pour tous les points de fréquence  $f_q$  appartenant à cette trame temporelle :

$$\text{Si } \frac{\|\mathcal{S}_{\mathbf{x}}(t_p, f_q)\|}{\max_f \{\|\mathcal{S}_{\mathbf{x}}(t_p, f)\|\}} > \epsilon_2, \quad \text{garder } (t_p, f_q), \quad (5.19)$$

où  $\epsilon_2$  est un seuil de faible valeur (typiquement  $\epsilon_2 = 0.05$ ). Alors, l'ensemble des points sélectionnés  $\Omega$ , est exprimé par  $\Omega = \cup_{i=1}^N \Omega_i$ , où  $\Omega_i$  est le support TF de la source  $s_i(t)$ . Notons que, l'effet d'étalement de l'énergie du bruit et la localisation de celle des sources dans le domaine TF permet d'augmenter la robustesse de la méthode proposée vis à vis du bruit (voir la Partie IV de [129]). Ainsi, en utilisant l'équation (5.19), nous souhaitons garder uniquement les points TF où l'énergie du signal est significative. De plus, et toujours en raison de l'étalement de l'énergie du bruit dans le plan TF, la contribution du bruit dans les points TF des sources est relativement négligeable du moins pour des valeurs modérées et hautes du rapport signal à bruit. D'autre part, on note que le seuillage du bruit tout comme le masquage temps-fréquence induisent une distortion non-linéaire dans le signal reconstruit. L'élimination ou la réduction de cette distortion est un problème ouvert qui soulève beaucoup d'interrogations et motive un bon nombre de travaux de recherche y compris ceux qui tentent d'atténuer cette distortion dans le domaine temps-fréquence [136].

### **Classification vectorielle et estimation des DTF des sources**

La procédure de classification peut être effectuée de manière similaire au cas de l'algorithme TF-UBSS quadratique. Premièrement, on obtient les vecteurs de direction spatiale par :

$$\mathbf{v}(t_a, f_a) = \frac{\mathcal{S}_{\mathbf{x}}(t_a, f_a)}{\|\mathcal{S}_{\mathbf{x}}(t_a, f_a)\|}, \quad (t_a, f_a) \in \Omega, \quad (5.20)$$

et on forcera leur premier élément à être réel et positif, et ceci sans perte de généralité.

Ensuite, on classifera ces vecteurs en  $N$  classes  $\{\mathcal{C}_i | i \in \mathcal{N}\}$ , en utilisant l'algorithme de classification  $k$ -means. La collecte de tous les points correspondant à tous les vecteurs de la classe  $\mathcal{C}_i$ , formera le support TF  $\Omega_i$  de la source  $s_i(t)$ . Alors, le vecteur colonne  $\mathbf{a}_i$  de la matrice de mélange  $\mathbf{A}$  est estimé comme le centroïde de cet ensemble de vecteurs :

$$\hat{\mathbf{a}}_i = \frac{1}{\#\mathcal{C}_i} \sum_{(t,f) \in \Omega_i} \mathbf{v}(t, f), \quad (5.21)$$

où  $\#\mathcal{C}_i$  est le nombre de vecteurs dans la classe  $\mathcal{C}_i$ .

Donc, on peut estimer la TFCT de chaque source  $s_i(t)$  par :

$$\hat{\mathcal{S}}_{s_i}(t, f) = \begin{cases} \hat{\mathbf{a}}_i^H \mathcal{S}_x(t, f), & \forall (t, f) \in \Omega_i, \\ 0, & \text{sinon .} \end{cases} \quad (5.22)$$

à partir de l'équation (5.16), on a :

$$\hat{\mathbf{a}}_i^H \mathcal{S}_x(t, f) = \hat{\mathbf{a}}_i^H \mathbf{a}_i \mathcal{S}_{s_i}(t, f) \approx \mathcal{S}_{s_i}(t, f), \quad \forall (t, f) \in \Omega_i .$$

### Utilisation de la transformée en ondelettes

Il est intéressant de noter, que l'utilisation de la transformée en ondelettes comme transformation linéaire, afin d'exploiter la parcimonie des signaux audio dans le domaine TE et un choix judicieux, car elle permet de pallier l'inconvénient majeur de la TFCT qui est la résolution temps-fréquence fixe et de garder son principal avantage qui est le coût de calcul relativement faible.

En appliquant l'équation (5.7) au modèle des observations dans (1.1) (dans le cas sans bruit), on obtient l'expression suivante :

$$\mathcal{W}_x(t, \lambda) = \mathbf{A} \mathcal{W}_s(t, \lambda), \quad (5.23)$$

où  $\mathcal{W}_x(t, \lambda) = [\mathcal{W}_{x_1}(t, \lambda), \dots, \mathcal{W}_{x_M}(t, \lambda)]^T$  est le vecteur temps-échelle (TE) des mélanges,  $\mathcal{W}_s(t, \lambda) = [\mathcal{W}_{s_1}(t, \lambda), \dots, \mathcal{W}_{s_N}(t, \lambda)]^T$  est le vecteur TE des sources. Comme pour la TFCT la TO est une transformation linéaire, complètement dépourvue de termes-croisés. Par conséquent, on pourra utiliser les mêmes étapes de seuillage du bruit et de classification vectorielle pour sélectionner les points auto-source et estimer la matrice de mélange. Une fois la matrice de mélange estimée, on pourra reconstruire la distribution temps-échelle de chaque sources  $s_i$  comme suit :

$$\widehat{\mathcal{W}}_{s_i}(t, \lambda) = \begin{cases} \hat{\mathbf{a}}_i^H \mathcal{W}_x(t, \lambda), & \forall (t, \lambda) \in \Omega_i, \\ 0, & \text{sinon .} \end{cases}$$

## 5.4 Approche TF-UBSS basée sur la projection en sous-espace pour des sources TF-non-disjointes

Nous avons proposé dans la section précédente, des méthodes TF-UBSS basées sur la classification, utilisant des DTF quadratiques comme la DWVM, ou bien des DTF linéaires comme la TFCT, qui sont résumés dans les Tableaux 5.1 et 5.2 respectivement. Ces méthodes sont fondées sur l'hypothèse que les sources sont TF-disjointes, ce qui a mené aux structures TF données par les équations (5.10) et (5.16). Quand les sources sont non-disjointes dans le domaine TF, alors ces équations ne sont plus valides.

Sous la condition de TF-non-disjointes, exposée dans l'hypothèse 5.1, on propose dans cette section deux méthodes alternatives, une pour les DTF quadratiques et une autre pour les DTF linéaires, afin de résoudre le problème de la SAS sous-déterminée en utilisant la projection en sous-espace.

### 5.4.1 Algorithme TF-UBSS quadratique basé sur la projection en sous-espace

Rappelons que les deux premières étapes de l'algorithme TF-UBSS quadratique basé sur la classification ne sont pas fondées sur l'hypothèse que les sources soit TF-disjointes (voir le tableau 5.1). C'est pourquoi, on peut réutiliser ces étapes pour obtenir l'ensemble des points auto-source  $\Omega$ . Maintenant, sous la condition de TF-non-disjointes, on considère le point auto-source  $(t_b, f_b) \in \Omega$  tel qu'il y a  $\mathcal{J}$  sources,  $\mathcal{J} < M$ , présentes en ce point. Notre but est d'identifier les sources présentes au point  $(t_b, f_b)$  et estimer la contribution d'énergie de chacune.

Notons  $\alpha_1, \dots, \alpha_{\mathcal{J}} \in \mathcal{N}$  les indices des sources présentes au point  $(t_b, f_b)$ , et définissons les termes suivants :

$$\tilde{\mathbf{s}}(t) = [s_{\alpha_1}(t), \dots, s_{\alpha_{\mathcal{J}}}(t)]^T, \quad (5.24a)$$

$$\tilde{\mathbf{A}} = [\mathbf{a}_{\alpha_1}, \dots, \mathbf{a}_{\alpha_{\mathcal{J}}}] . \quad (5.24b)$$

Alors, sous l'hypothèse 5.1, l'équation (5.9) est réduite à :

$$\mathbf{D}_{xx}^{\text{dvw}}(t_b, f_b) = \tilde{\mathbf{A}} \mathbf{D}_{ss}^{\sim}(t_b, f_b) \tilde{\mathbf{A}}^H, \quad (5.25)$$

ainsi, sachant que  $\mathbf{D}_{ss}^{\sim}$  est une matrice de rang plein, on a :

$$\text{Range}\{\mathbf{D}_{xx}(t_b, f_b)\} = \text{Range}\{\tilde{\mathbf{A}}\} . \quad (5.26)$$



Soit  $\mathcal{P}$  la matrice de projection orthogonale sur le sous-espace bruit de  $\mathbf{D}_{xx}^{\text{dvw}}(t_b, f_b)$ . Alors, à partir de l'équation (5.26), on obtient :

$$\mathcal{P} = \mathbf{I}_M - \mathbf{V}\mathbf{V}^H, \quad (5.27)$$

et

$$\begin{cases} \mathcal{P}\mathbf{a}_i = \mathbf{0}, & \forall i \in \{\alpha_1, \dots, \alpha_{\mathcal{J}}\}, \\ \mathcal{P}\mathbf{a}_i \neq \mathbf{0}, & \text{sinon.} \end{cases} \quad (5.28)$$

Dans l'équation (5.27),  $\mathbf{V}$  est la matrice formée par les  $\mathcal{J}$  principaux vecteurs propres de la matrice  $\mathbf{D}_{xx}(t_b, f_b)$ .

En supposant que la matrice  $\mathbf{A}$  est connue ou correctement estimée, l'exploitation de l'équation (5.28) nous permet d'identifier les indices  $\alpha_1, \dots, \alpha_{\mathcal{J}}$ , et donc, les sources présentes au point  $(t_b, f_b)$ . En pratique, en tenant compte de l'effet du bruit, on peut estimer ces indices en détectant les  $\mathcal{J}$  plus petites valeurs de l'ensemble  $\{\|\mathcal{P}\mathbf{a}_i\| \mid i \in \mathcal{N}\}$ , que l'on peut exprimer par :

$$\{\alpha_1, \dots, \alpha_{\mathcal{J}}\} = \arg \min^{\mathcal{J}} \{\|\mathcal{P}\mathbf{a}_i\| \mid i \in \mathcal{N}\}, \quad (5.29)$$

où  $\min^{\mathcal{J}}$  dénote la minimisation pour obtenir les  $\mathcal{J}$  plus petites valeurs. Les contributions TF des  $\mathcal{J}$  sources au point  $(t_b, f_b)$  sont estimées comme les éléments diagonaux de la matrice suivante :

$$\widehat{\mathbf{D}}_{ss}(t_b, f_b) \approx \widetilde{\mathbf{A}}^{\#} \mathbf{D}_{xx}(t_b, f_b) (\widetilde{\mathbf{A}}^{\#})^H, \quad (5.30)$$

Dans ce qui suit, nous allons présenter une méthode d'estimation de la matrice de mélange  $\mathbf{A}$  en utilisant l'hypothèse 5.2. Cette hypothèse indique que, pour chaque source  $s_i(t)$ , il existe une région dans le domaine TF  $\Theta_i$  où elle est présente seule. En d'autres termes,  $\Theta_i$  contient uniquement les points auto-source de  $s_i(t)$ . Donc, on peut réutiliser les propriétés déduites de l'équation (5.10) dans le cas TF-disjoints, mais pour certaines regions TF, comme décrit ci-dessus :

$$\mathbf{D}_{xx}(t, f) = \rho_{s_i s_i}(t, f) \mathbf{a}_i \mathbf{a}_i^H, \quad \forall (t, f) \in \Theta_i, \forall i \in \mathcal{N}.$$

L'union de ces régions,  $\Theta = \bigcup_{i=1}^N \Theta_i$ , est détectée par l'expression suivante :

$$\text{Si } \left| \frac{\lambda_{\max}\{\mathbf{D}_{xx}^{\text{dvw}}(t, f)\}}{\text{trace}\{\mathbf{D}_{xx}^{\text{dvw}}(t, f)\}} - 1 \right| < \epsilon_3, \text{ alors } (t, f) \in \Theta, \quad (5.31)$$

où  $\epsilon_3$  est un seuil de faible valeur (typiquement  $\epsilon_3 \leq 0.1$ ) et  $\lambda_{\max}\{\mathbf{D}_{xx}^{\text{dvw}}(t, f)\}$  dénote la plus grande valeur propre de la matrice  $\mathbf{D}_{xx}^{\text{dvw}}(t, f)$ . Sous ces conditions, on peut

Tab. 5.3 – Algorithme TF-UBSS quadratique basé sur la projection en sous-espace

1. Calculer la matrice de DTFQ des mélanges par (5.11).
2. Effectuer un seuillage du bruit et sélectionner les points auto-source par l'équation (5.12).
3. Sélectionner les vecteurs par l'équation (5.31); estimation de la matrice de mélange par (5.32) et (5.33).
4. Pour tous les points auto-source, appliquer la projection en sous-espace pour estimer les DTF des sources en utilisant les équations (5.27), (5.29) et (5.30)
5. Synthétiser les DTF des sources estimées [134].

appliquer la même procédure de classification vectorielle que dans la section 5.3.1 pour estimer la matrice de mélange  $\mathbf{A}$ . Plus précisément, nous obtenons d'abord tous les vecteurs de direction spatiale :

$$\mathbf{v}(t, f) = \frac{\text{diag}(\mathbf{D}_{xx}^{\text{tftct}}(t, f))}{\|\text{diag}(\mathbf{D}_{xx}^{\text{tftct}}(t, f))\|}, \quad \forall (t, f) \in \Theta. \quad (5.32)$$

Ensuite, nous regroupons ces vecteurs en  $N$  classes  $\{\mathcal{C}_i | i \in \mathcal{N}\}$  en utilisant l'algorithme de classification  $k$ -means. La collecte de tous les points correspondant à tous les vecteurs de la classe  $\mathcal{C}_i$ , formera la région TF  $\Theta_i$  de la source  $s_i(t)$ . Enfin, les vecteurs colonnes de matrice  $\mathbf{A}$  seront estimés comme les centroïdes de ces classes par :

$$\hat{\mathbf{a}}_i = \frac{1}{\#\mathcal{C}_i} \sum_{(t,f) \in \Theta_i} \mathbf{v}(t, f), \quad \forall i \in \mathcal{N}, \quad (5.33)$$

où  $\#\mathcal{C}_i$  est le nombre de points présent dans la région  $\Theta_i$ . Le tableau 5.3 donne un résumé de l'algorithme TF-UBSS quadratique basé sur la projection en sous-espace.

### 5.4.2 Algorithme TF-UBSS linéaire basé sur la projection en sous-espace

De manière similaire, on présente dans cette section l'algorithme TF-UBSS linéaire basé sur la projection en sous-espace pour des sources TF-non-disjointes en uti-

lisant la TFCT. Nous utiliserons également la première étape de l'algorithme TF-UBSS linéaire basé sur la classification (voir tableau 5.2) pour obtenir l'ensemble de tous les points auto-source  $\Omega$ . Sous la condition de TF-non-disjoints, on considère le point auto-source  $(t_b, f_b) \in \Omega$  tel qu'il y a  $\mathcal{J}$  sources,  $\mathcal{J} < M$ , présentes en ce point. Alors, l'équation (5.15) est réduite comme suit :

$$\mathcal{S}_x(t_b, f_b) = \widetilde{\mathbf{A}} \widetilde{\mathcal{S}}_{\tilde{\mathbf{s}}}(t_b, f_b), \quad \forall (t_b, f_b) \in \Omega, \quad (5.34)$$

où  $\widetilde{\mathbf{A}}$  et  $\tilde{\mathbf{s}}$  sont définis par l'équation (5.24).

Soit  $\mathcal{Q}$  la matrice de projection orthogonale sur le sous-espace bruit de la matrice  $\widetilde{\mathbf{A}}$ . Alors,  $\mathcal{Q}$  peut être calculée par :

$$\mathcal{Q} = \mathbf{I}_M - \widetilde{\mathbf{A}} (\widetilde{\mathbf{A}}^H \widetilde{\mathbf{A}})^{-1} \widetilde{\mathbf{A}}^H. \quad (5.35)$$

Nous avons par conséquent l'observation suivante :

$$\begin{cases} \mathcal{Q} \mathbf{a}_i = \mathbf{0}, & i \in \{\alpha_1, \dots, \alpha_{\mathcal{J}}\} \\ \mathcal{Q} \mathbf{a}_i \neq \mathbf{0}, & \text{sinon} \end{cases}. \quad (5.36)$$

En supposant que la matrice  $\mathbf{A}$  est connue ou correctement estimée, l'exploitation de l'équation (5.36) nous permet d'identifier les indices  $\alpha_1, \dots, \alpha_{\mathcal{J}}$ , et donc, les sources présentes au point  $(t_b, f_b)$ . En pratique, en tenant compte de l'effet du bruit, on peut détecter les vecteurs colonnes de  $\widetilde{\mathbf{A}}$  en minimisant :

$$\{\alpha_1, \dots, \alpha_{\mathcal{J}}\} = \arg \min_{\beta_1, \dots, \beta_{\mathcal{J}}} \left\{ \|\mathcal{Q} \mathcal{S}_x(t, f)\| \mid \widetilde{\mathbf{A}}_{\beta} = [\mathbf{a}_{\beta_1}, \dots, \mathbf{a}_{\beta_{\mathcal{J}}}] \right\}. \quad (5.37)$$

Puis, les valeurs des DTF des  $\mathcal{J}$  sources au point TF  $(t_b, f_b)$  sont estimées par :

$$\widehat{\mathcal{S}}_{\tilde{\mathbf{s}}}(t, f) \approx \widetilde{\mathbf{A}}^{\#} \mathcal{S}_x(t, f). \quad (5.38)$$

La matrice de mélange  $\mathbf{A}$  est estimée en utilisant la technique de classification vectorielle des directions spatiales, de manière similaire à celle décrite dans l'algorithme TF-UBSS précédent. Enfin, les vecteurs colonnes de la matrice de mélange sont estimés comme les centroïdes des  $N$  classes issues de l'algorithme de classification. Le tableau 5.4 donne un résumé de l'algorithme TF-UBSS linéaire basé sur la projection en sous-espace.

### Utilisation de la transformée en ondelettes

Comme pour le cas de l'algorithme TF-UBSS pour des sources TF-disjointes, l'utilisation de la TO pour le cas des sources TF-non-disjointes est possible. En supposant que pour chaque point TE il y a  $\mathcal{J}$  sources,  $\mathcal{J} < M$  présentes au plus. Alors

Tab. 5.4 – Algorithme TF-UBSS linéaire basé sur la projection en sous-espace

1. Calculer la TFCT des mélanges en utilisant l'équation (5.18); seuillage du bruit par (5.19).
2. Estimer la matrice de mélange  $\mathbf{A}$  par l'équation (5.21) et l'algorithme  $k$ -means.
3. Pour tous les points auto-source, appliquer la projection en sous-espace pour estimer les DTF des sources en utilisant les équations (5.35), (5.37) et (5.38).
4. Synthétiser les DTF des signaux estimés en utilisant la TFCT inverse [135].

le modèle du mélange peut être réécrit comme :

$$\mathcal{W}_{\mathbf{x}}(t, \lambda) = \tilde{\mathbf{A}}\tilde{\mathcal{W}}_{\tilde{\mathbf{s}}}(t, \lambda), \quad \forall (t, \lambda) \in \Omega,$$

où  $\tilde{\mathbf{A}}$  et  $\tilde{\mathbf{s}}$  sont définis de la même manière que dans la section précédente. Par conséquent on pourra utiliser les mêmes étapes de seuillage du bruit, estimation de la matrice de mélange et projection en sous-espace, pour déterminer les points auto-source et les sources présentes en chaque point TE. Finalement on estimera les distributions temps-échelle des  $\mathcal{J}$  sources par :

$$\widehat{\mathcal{W}}_{\tilde{\mathbf{s}}}(t, \lambda) \approx \tilde{\mathbf{A}}^{\#}\mathcal{W}_{\mathbf{x}}(t, f).$$

## 5.5 Discussion

Nous discutons dans cette section certains points relatifs aux algorithmes TF-UBSS proposés et leurs applications.

### Le nombre de sources

Le nombre de sources  $N$  est supposé connu pour la méthode de classification vectorielle  $k$ -means que nous avons utilisé. Cependant, il existe des méthodes de classification [110] qui effectuent l'estimation des classes aussi bien que l'estimation du nombre de classes  $N$ . Dans nos simulations, nous avons observé que la

plupart du temps le nombre de classes est sur-estimé, ce qui conduit à une mauvaise qualité de séparation. C'est pourquoi, une estimation robuste du nombre de sources dans le cas de la SAS sous-déterminée demeure un problème ouvert et difficile qui mérite une attention particulière pour nos travaux futurs.

### Le recouvrement des sources

Dans l'approche basée sur la projection en sous-espace, nous devons évaluer le nombre de sources  $\mathcal{J}$  présentes à un point TF donné. Ceci peut être réalisé en estimant le nombre de valeurs propres non-nulles de la matrice  $D_{xx}^{dwv}(t, f)$  en utilisant des critères comme MDL ou AIC [119]. Il est possible aussi de considérer une valeur fixe (maximum) de  $\mathcal{J}$  qui sera utilisée pour tous les points auto-source. En effet, si le nombre de sources réellement présentes au point TF est inférieur à  $\mathcal{J}$ , nous estimerons des valeurs de DTF des sources proches de zéro. Par exemple, si on suppose que  $\mathcal{J} = 2$  sources sont présentes à un point TF donné, où une seule source contribue effectivement, alors on estimera une des deux valeurs de la DTF proche de zéro. Cette approche augmente légèrement l'erreur d'estimation des signaux sources (en particulier pour des RSB faibles), mais a comme avantage la simplicité de mise en oeuvre par rapport à l'utilisation de critères basés sur la théorie de l'information. Dans nos simulations, nous avons choisi cette solution avec le paramètre  $\mathcal{J} = 2$  ou  $\mathcal{J} = 3$ .

### DTF quadratique contre linéaire

Nous avons proposé deux algorithmes utilisant les DTF quadratique et linéaire. L'utilisation de la DTF quadratique est préférable pour des signaux de type FML et pour des tailles d'observation faibles ou modérées. Pour la séparation de signaux audio, la taille des observations est généralement grande, et par conséquent, afin de réduire le coût de calcul, on préférera l'utilisation de l'algorithme TF-UBSS linéaire (TFCT). En général, la version quadratique offre des performances légèrement meilleures que la version linéaire mais un coût de calcul nettement plus important.

### La qualité de séparation par rapport au nombre de sources

Bien que nous soyons dans le cas du mélange sous-déterminé, le nombre de sources  $N$  ne doit pas être très grand par rapport au nombre de capteurs. En effet,

quand  $N$  augmente, le niveau d'interférence des sources augmente à son tour, et par conséquent, l'hypothèse d'avoir des sources TF-disjointes sera mal satisfaite. Par ailleurs, pour un très grand nombre de sources, la probabilité d'avoir deux sources spatialement proches est très importante ; les directions spatiales  $\mathbf{a}_i$  et  $\mathbf{a}_j$  sont presque linéairement dépendantes. Dans ce cas, les performances de l'algorithme de classification vectorielle se dégraderont significativement. En bref, les propriétés de parcimonie et de séparation spatiale sont deux facteurs de limitation à l'augmentation du nombre de sources. La figure 5.8 illustre la dégradation des performances de séparation par rapport au nombre de sources.

## 5.6 Simulations

### 5.6.1 Simulations de l'algorithme TF-UBSS linéaire

Dans cette section, nous présentons quelques résultats de simulation pour illustrer l'efficacité de notre algorithme de séparation TF-UBSS linéaire. Pour cela, nous considérons une antenne composée de  $M = 3$  capteurs recevant  $N = 4$  signaux sources audio arrivant suivant les angles d'arrivées  $\theta_1 = 15$ ,  $\theta_2 = 30$ ,  $\theta_3 = 45$  et  $\theta_4 = 75$  degrés respectivement. La taille des observations  $T = 25000$  échantillons (les signaux sources sont échantillonnés à une fréquence de 8 KHz). Les signaux observés sont corrompus par un bruit blanc additif de covariance  $\sigma^2 \mathbf{I}_M$  ( $\sigma^2$  étant la puissance du bruit). La qualité de la séparation est mesurée par l'erreur quadratique moyenne normalisée (EQMN) des sources estimées pour  $N_r = 100$  réalisations aléatoires du bruit.

Dans la figure 5.3, les quatre figures de la première ligne, illustrent les représentations TF des signaux sources originaux. Les trois figures de la ligne du milieu, illustrent les représentations TF des  $M$  mélanges et les quatre dernières illustrent les représentation TF des sources estimées par l'algorithme TF-UBSS basé sur la projection en sous-espace (tableau 5.4). La figure 5.4 représente les mêmes signaux dans le domaine temporel.

Dans la figure 5.5, on compare les performances de séparation obtenues par l'algorithme TF-UBSS basé sur la projection en sous-espace avec  $\mathcal{J} = 2$  et l'algorithme TF-UBSS basé sur la classification (tableau 5.2). On observe que l'algorithme basé sur la projection en sous-espace produit de meilleurs résultats que ceux obtenus par l'algorithme basé sur la classification, ce qui conforte l'hypothèse 5.1 qui affirme que les sources ne sont pas nécessairement disjointes dans le domaine TF.

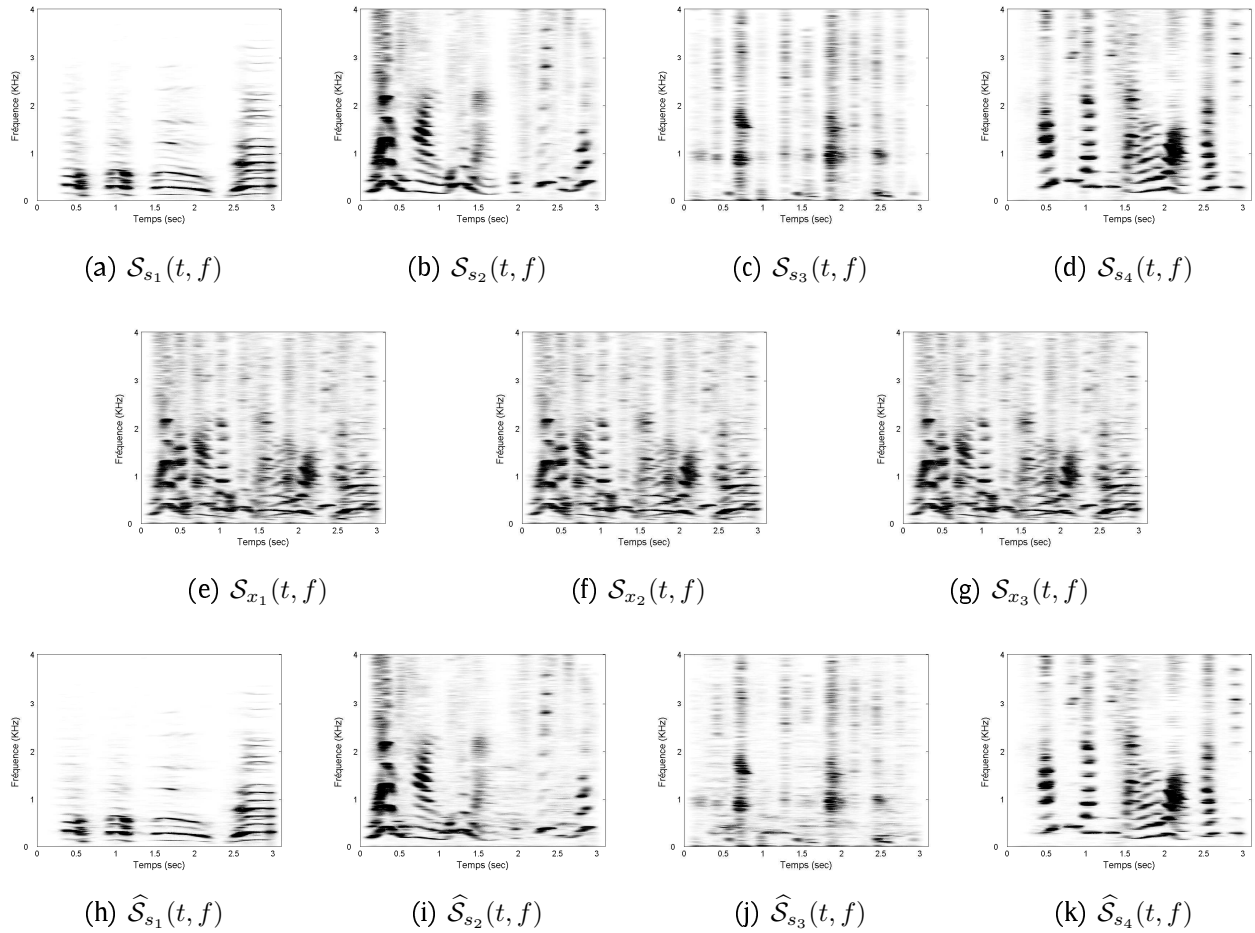
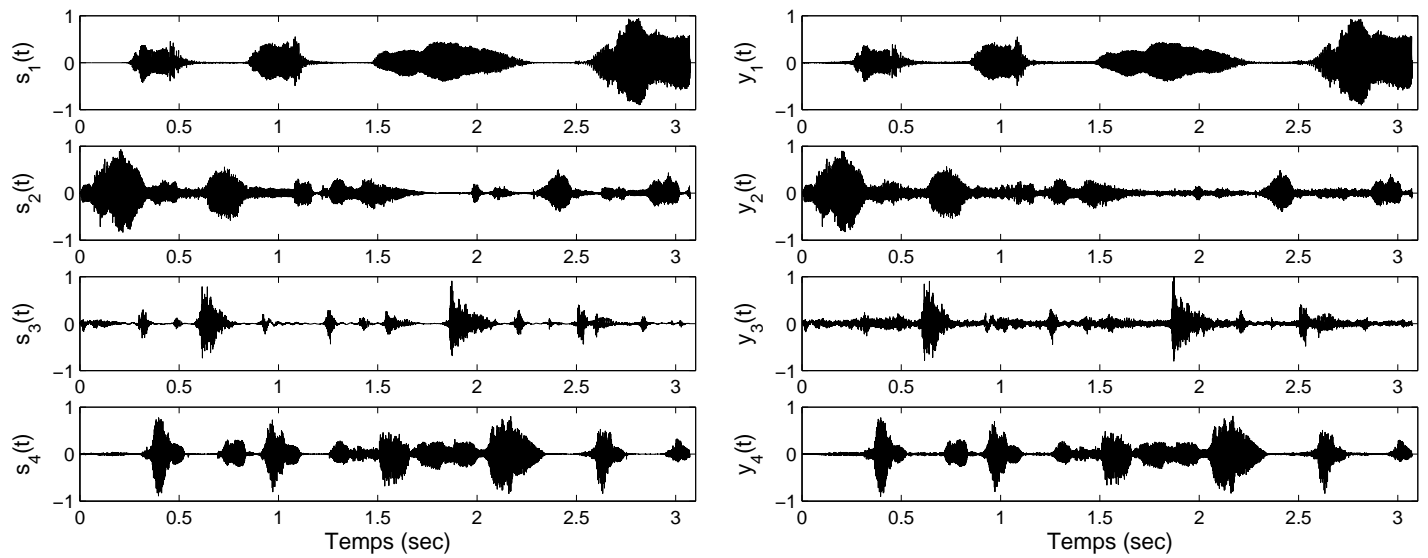


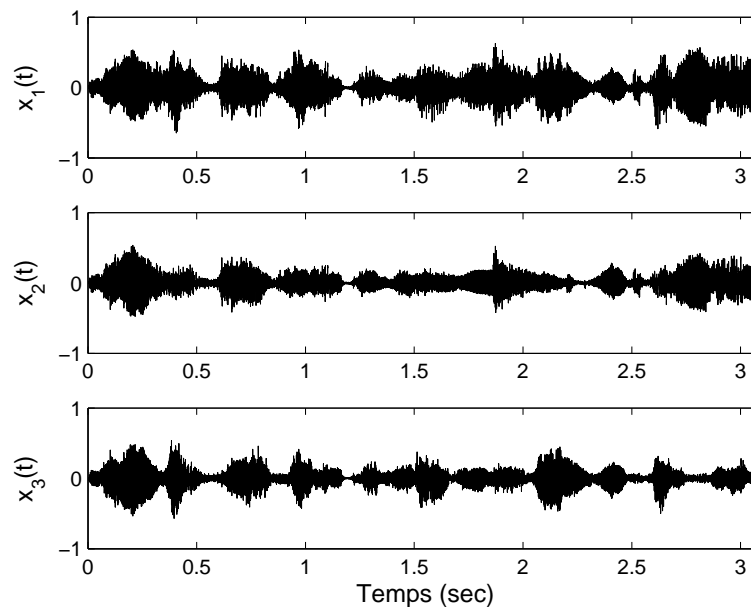
Fig. 5.3 – Exemple de simulation de l’algorithme TF-UBSS linéaire basé sur la projection en sous-espace dans le cas de  $N = 4$  sources audio et  $M = 3$  capteurs. Les quatre premières figures représentent les signaux sources, les trois figures du milieu représentent les trois mélanges, et les quatre dernières figures représentent les signaux sources estimés par notre algorithme.

Dans la méthode utilisant la projection en sous-espace, on a besoin d’estimer la matrice de mélange  $\mathbf{A}$ . Ceci est réalisé par la méthode décrite précédemment. La figure 5.6 représente l’EQMN calculée sur l’estimée de la matrice de mélange  $\mathbf{A}$  en fonction du RSB. Il est clair que la méthode proposée pour l’estimation de la matrice de mélange produit des performances satisfaisantes. Tandis que la figure 5.7 représente les performances de séparation dans le cas où on suppose la matrice  $\mathbf{A}$  connue, comparées à ceux obtenues en utilisant notre méthode d’estimation de la matrice  $\mathbf{A}$ . On peut en conclure que l’erreur d’estimation sur la matrice de mélange  $\widehat{\mathbf{A}}$ , n’influe pas sur la qualité de séparation de l’algorithme TF-UBSS. La figure 5.8 illustre la rapide dégradation de la qualité de séparation lors de



(a) Signaux sources

(b) Signaux estimés



(c) Signaux mélanges

Fig. 5.4 – Exemple de simulation (dans le domaine temporel) de l’algorithme TF-UBSS linéaire basé sur la projection en sous-espace dans le cas de  $N = 4$  sources audio et  $M = 3$  capteurs. (a) les signaux sources originaux, (b) les signaux estimés par notre algorithme, (c) les signaux capteurs (mélanges).

l’augmentation du nombre de sources qui varie entre  $N = 4$  et  $N = 7$ . Ceci confirme



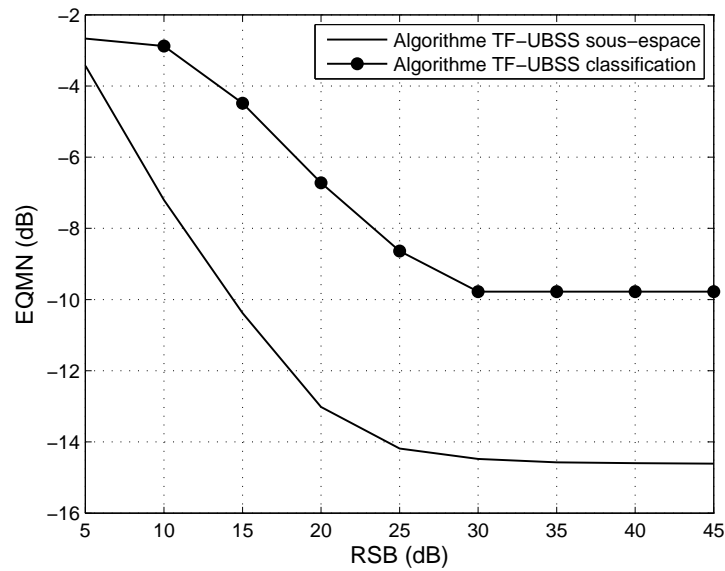


Fig. 5.5 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 25000$  échantillons : comparaison des performances des algorithmes TF-UBSS linéaires basés sur la classification et sur la projection en sous-espace.

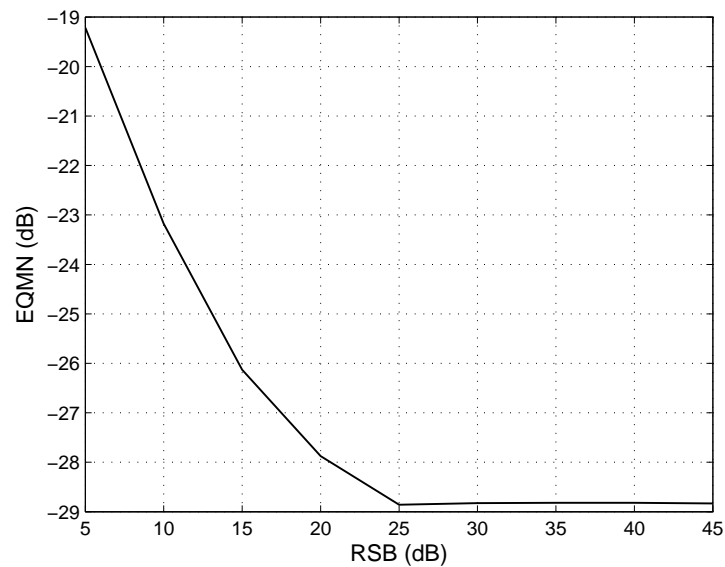


Fig. 5.6 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 25000$  échantillons : performances de la méthode d'estimation de la matrice de mélange  $\mathbf{A}$ .

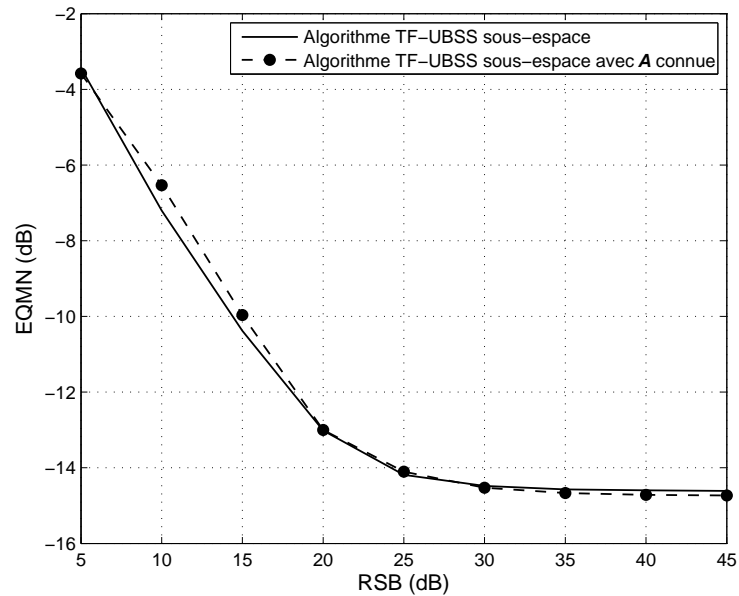


Fig. 5.7 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 25000$  échantillons : comparaison des performances de séparation de l'algorithme TF-UBSS basé sur la projection en sous-espace dans le cas où la matrice de mélange  $\mathbf{A}$  est exactement connue, et ceux obtenues en utilisant notre méthode d'estimation de la matrice  $\mathbf{A}$ .

les remarques faites dans la section 5.5.

Dans la figure 5.9, on compare les performances obtenues par l'algorithme TF-UBSS basé sur la projection en sous-espace pour  $\mathcal{J} = 2$  et  $\mathcal{J} = 3$  ainsi que l'algorithme TF-UBSS basé sur la classification. Dans cette expérience, nous avons utilisé  $M = 4$  capteurs et  $N = 5$  signaux sources. On peut observer que pour les hauts RSB, le cas où  $\mathcal{J} = 3$  produit de meilleures performances de séparation que ceux obtenues dans le cas où  $\mathcal{J} = 2$ . Cependant, pour les bas RSB, une grande valeur de  $\mathcal{J}$  augmente le bruit d'estimation (comme mentionné dans la section 5.5) et par conséquent dégrade la qualité de séparation.

### 5.6.2 Simulations de l'algorithme TF-UBSS quadratique

Dans ces simulations, on utilisera la même configuration pour l'algorithme TF-UBSS quadratique ; le même nombre de capteurs et de sources ( $M = 3$ ,  $N = 4$ ). Les sources utilisées sont des signaux FML, avec des tailles d'observation  $T = 256$  échantillons.

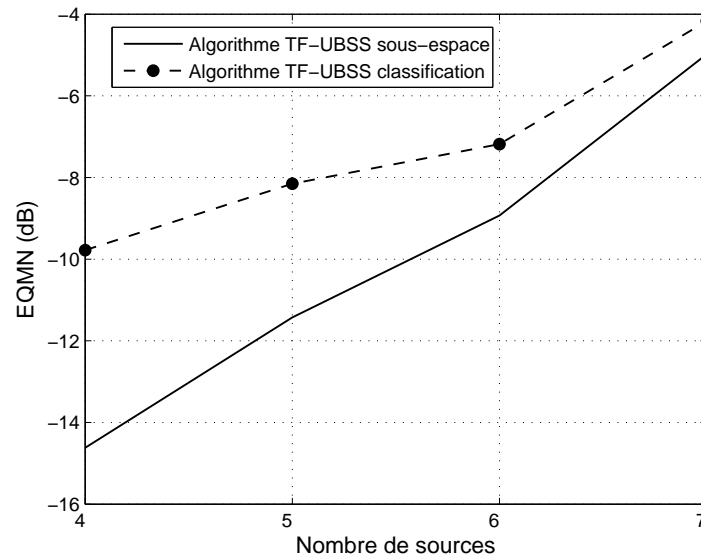


Fig. 5.8 – EQMN en fonction du nombre de sources  $N$  pour  $M = 3$  capteurs avec une taille d’observation  $T = 25000$  : comparaison des performances des algorithmes TF-UBSS basés sur la classification et celui basé sur la projection en sous-espace.

On compare l’algorithme TF-UBSS quadratique basé sur la classification tableau 5.1 et l’algorithme TF-UBSS quadratique basé sur la projection en sous-espace tableau 5.3. Les figures 5.10-(a,d,g,j) représentent les DTF (utilisant la DWV) des quatre sources. Les figures 5.10-(b,e,h,k) montrent les DTF des sources estimées en utilisant l’algorithme basé sur la classification, tandis que les figures 5.10-(c,f,i,l) sont ceux obtenues par l’algorithme basé sur la projection en sous-espace.

A partir des figures 5.10-(b,e) on peut observer que les points de recouvrement entre les sources  $s_1$  et  $s_2$  ont été pris par la source  $s_2$  avec l’algorithme basé sur la classification. En d’autres termes, en utilisant l’algorithme basé sur la projection en sous-espace, les points d’intersection seront redistribués aux deux sources (figures 5.10-(c,f)). En général, les points d’intersection dans le cas non-disjoint ont été explicitement traités. Cela fournit une comparaison visuelle des performances. Dans la figure 5.11, on compare les performances de séparation statistiques entre l’algorithme basé sur la classification et celui basé sur la projection en sous-espace utilisant les DTFQ, évaluées pour 100 réalisations aléatoires du bruit.

On peut noter que le gain dans ce cas en faveur de la méthode sous-espace est moins important que celui obtenu précédemment pour des signaux audio. Ceci

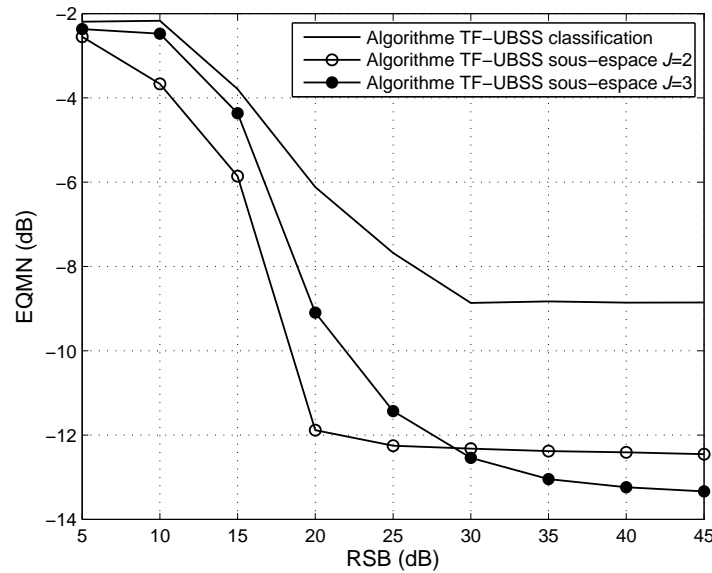


Fig. 5.9 – EQMN en fonction du RSB pour  $N = 5$  sources audio et  $M = 4$  capteurs avec une taille d’observation  $T = 25000$  échantillons : comparaison des performances de l’algorithme TF-UBSS basé sur la classification et l’algorithme TF-UBSS basé sur la projection en sous-espace pour différentes valeurs de  $\mathcal{J}$ .

est dû au fait que la region de recouvrement dans ce cas est moins importante que dans le cas des signaux audio. Néanmoins, ce résultat confirme l’observation visuelle précédente en ce qui concerne le gain de performances en faveur de la méthode sous-espace.

## 5.7 Conclusion

Dans ce chapitre nous avons introduit de nouvelles méthodes pour la séparation aveugle de sources dans le cas sous-déterminé pour des sources TF-non-disjointes et non-stationnaires en utilisant leurs représentations temps-fréquence. Les principaux avantages des algorithmes de séparation proposés sont : premièrement, une hypothèse plus réaliste sur la parcimonie des sources, c’est à dire que les sources ne sont pas nécessairement disjointes dans le domaine temps-fréquence, et deuxièmement, un traitement explicite des points de recouvrement en utilisant une projection en sous-espace, ce qui a conduit à une amélioration de la qualité de séparation. Des résultats de simulations illustrent l’efficacité de nos algorithmes selon différents scénarios comparés à ceux existant dans la littérature.

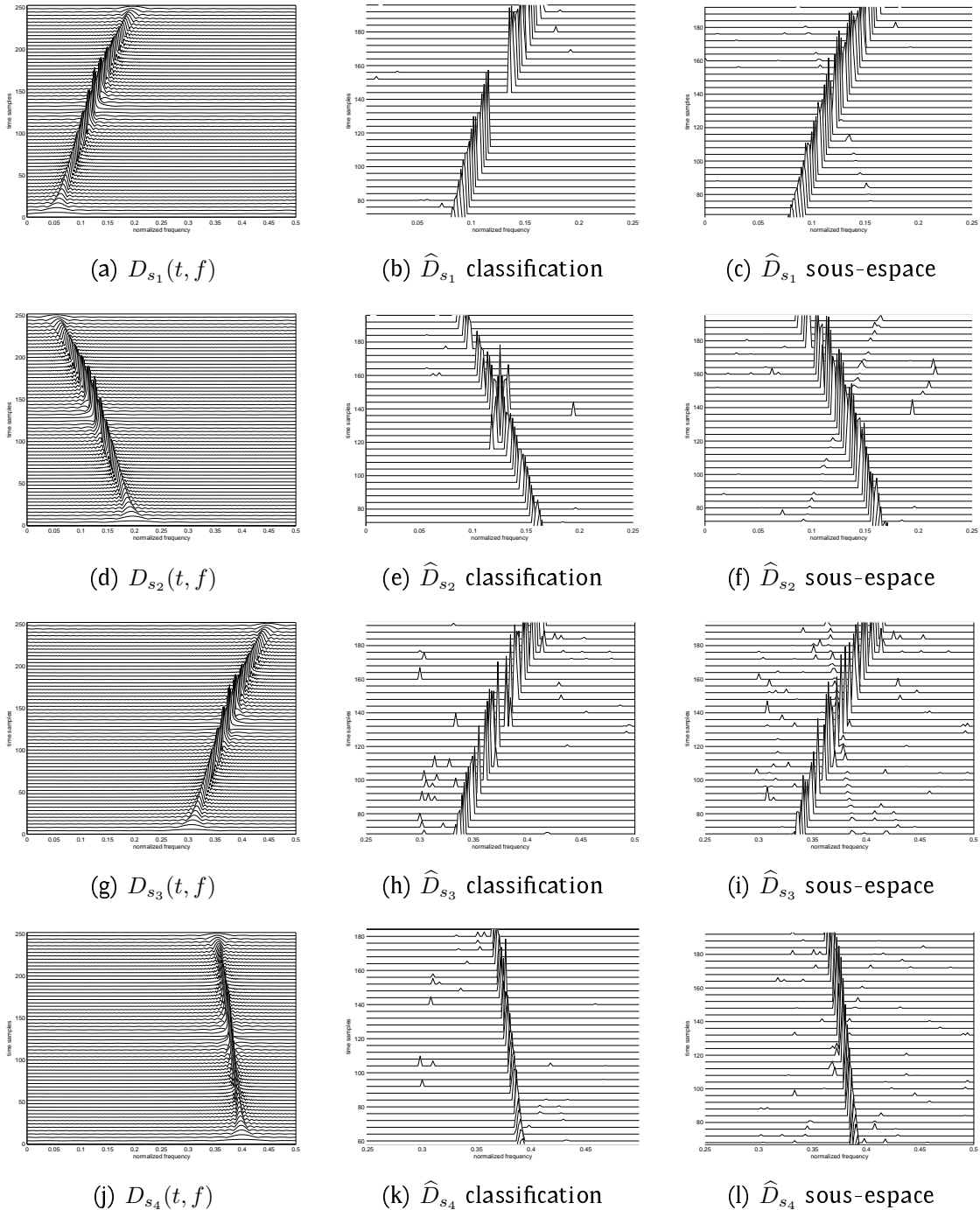


Fig. 5.10 – Exemple de simulation pour l’algorithme TF-UBSS quadratique dans le cas de  $N = 4$  sources FML et  $M = 3$  capteurs. De gauche à droite, les figures représentent respectivement les signatures TF des sources originales, des sources estimées par la méthode basée sur la classification, et celles des sources estimées par la méthode basée sur la projection en sous-espace.

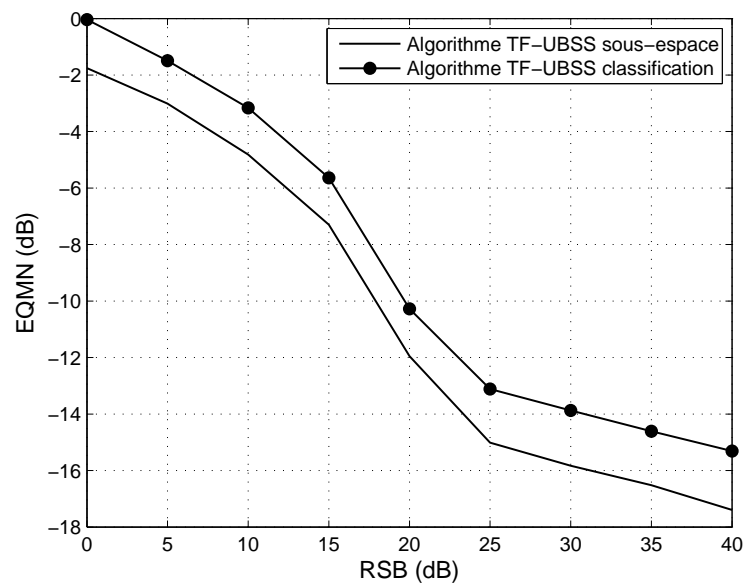


Fig. 5.11 – EQMN en fonction du RSB pour  $N = 4$  sources LFM et  $M = 3$  capteurs avec une taille d'observation  $T = 256$  échantillons : comparaison des performances des algorithmes TF-UBSS quadratiques basés sur la classification et sur la projection en sous-espace.

# La SAS audio utilisant la représentation temps-fréquence : mélanges convolutifs

---

## 6.1 Introduction

La séparation aveugle de sources dans le cas convolutif sous-déterminé reste un problème relativement peu traité dans la littérature comparé au cas instantané, et sa résolution est un des challenges de la séparation aveugle de sources. Dans le cas du mélange instantané sous-déterminé, plusieurs méthodes exploitent la parcimonie des sources dans différents domaines transformés ont été proposées [87,127,133,137,138]. Toutes ces méthodes procèdent grossièrement comme suit : transformer les signaux observés dans un domaine approprié dans le but d'accroître leurs parcimonie, utiliser des techniques de masquage pour séparer les sources en supposant qu'elles soient disjointes dans le domaine transformé et pour finir, synthétiser les sources estimées dans le domaine temporel [109]. La contrainte majeure de ces méthodes est le fait de supposer que les sources soient disjointes dans le domaine transformé. Cette hypothèse est assez restrictive et limitative, bien que ces méthodes aient prouvé leurs efficacité dans les conditions de faible densité (faible recouvrement) des sources.

Ce chapitre est consacré à la séparation aveugle de sources audio dans le cas du

mélange convolutif sous-déterminé, en ayant comme but, le relâchement de l'hypothèse qui impose aux sources d'être disjointes dans le domaine transformé. Par conséquent, dans ce qui suit, nous présenterons deux versions d'un algorithme dénommé TF-CUBSS pour *Time-Frequency Convolutive Underdetermined Blind Source Separation*. La première version est basée sur la classification vectorielle et sur le fait que les sources soient disjointes dans le domaine temps-fréquence. Quant à la deuxième, elle est basée sur la projection en sous-espace et tolère que plusieurs sources se recouvrent dans un même point du domaine temps-fréquence.

## 6.2 Formalisation du problème et hypothèses

Dans ce chapitre, nous admettrons que les signaux observés sont modélisés par l'équation (1.5) qui s'écrit :

$$\mathbf{x}(t) = \sum_{k=0}^{\mathcal{K}} \mathbf{H}(k) \mathbf{s}(t - k) + \mathbf{w}(t), \quad (6.1)$$

où  $\forall k \in [0, \mathcal{K}]$ ,  $\mathbf{H}(k)$  est une matrice de taille  $M \times N$  ( $N > M$ ) et en supposant que le canal vérifie l'hypothèse 4.1 qui implique que les entrées de  $\mathbf{h}_i(z)$  notées  $h_{ij}(z)$ ,  $j = 1, \dots, M$ , n'ont pas de zéros en commun  $\forall i$ .

Afin d'identifier le canal de mélange  $\mathbf{H}(z)$ , il est nécessaire que les signaux sources vérifient l'hypothèse 4.2 qui suppose qu'il existe, périodiquement, des intervalles de temps où seulement une seule source est présente dans le mélange, et ceci est vérifié pour tous les signaux sources du mélange considéré.

Dans le même esprit que la méthode TF-UBSS linéaire présentée dans le chapitre 5, nous proposons une solution au problème de la SAS convolutive sous-déterminée, basée sur la représentation parcimonieuse des signaux audio dans le domaine temps-fréquence. C'est pourquoi, nous admettrons que les signaux sources vérifient l'hypothèse 5.1 qui prétend que le nombre de sources dont l'énergie contribue à chaque point TF est strictement inférieure au nombre de capteurs.



## 6.3 L'algorithme TF-CUBSS

### 6.3.1 Identification du canal

Comme mentionné précédemment, la méthode de SAS présentée ci-dessous nécessite la connaissance a priori ou une estimation du canal. Par conséquent, nous proposons d'utiliser la même procédure d'estimation du canal que celle décrite dans la section 4.3.1 qui se base sur l'hypothèse 4.2.

Premièrement, on sélectionne les intervalles de temps où une seule source est effectivement présente dans le mélange. Puis, pour chaque intervalle de temps sélectionné, on applique une technique d'identification aveugle SIMO pour estimer les paramètres du canal. Enfin, en appliquant l'approche précédente (détection des intervalles et identification SIMO), un même canal sera estimé plusieurs fois. Donc, on doit regrouper ensemble (classifier) les canaux estimés en  $N$  classes correspondant aux  $N$  canaux sources en utilisant l'algorithme de classification  $k$ -means.

### 6.3.2 Algorithme TF-CUBSS basé sur la classification

Comme nous l'avons déjà expliqué dans le chapitre précédent, la TFCT est fréquemment utilisée pour le traitement des signaux audio en raison de son faible coût de calcul et sa simplicité de mise en oeuvre. De ce fait, nous proposons une méthode de SAS utilisant la TFCT et permettant de traiter le cas convolutif sous-déterminé. Par analogie aux méthodes proposées dans le chapitre précédent, nous ferons référence à cette méthode par *Algorithme TF-CUBSS basé sur la classification*. En appliquant la TFCT au modèle du mélange convolutif défini par l'équation (1.5), on obtient (dans le cas sans bruit) :

$$\mathcal{S}_x(t, f) = \mathbf{H}(f)\mathcal{S}_s(t, f), \quad (6.2)$$

où  $\mathcal{S}_x(t, f)$  est le vecteur TFCT des mélanges,  $\mathcal{S}_s(t, f)$  est le vecteur TFCT des sources et  $\mathbf{H}(f) = [\mathbf{h}_1(f) \dots \mathbf{h}_N(f)]$  est la transformée de Fourier du canal de mélange. Sous l'hypothèse que toutes les sources sont disjointes dans le domaine TF, l'équation (6.2) est réduite à :

$$\mathcal{S}_x(t, f) = \mathbf{h}_i(f)\mathcal{S}_{s_i}(t, f), \quad \forall (t, f) \in \Omega_i, \forall i \in \mathcal{N}. \quad (6.3)$$

où  $\Omega_i$  est le support TF de la source  $s_i(t)$ . Ainsi, deux points TF  $(t_1, f_1)$  et  $(t_2, f_2)$  appartenant à la même région  $\Omega_i$  (correspondant au signal source  $s_i(t)$ ) seront associés au même canal  $\mathbf{h}_i$ . C'est à partir de cette observation que nous avons déduit

Tab. 6.1 – Algorithme TF-CUBSS basé sur la classification

1. Estimer le canal ; critère AIC [119] pour détecter le nombre de sources et appliquer un algorithme d'identification aveugle du canal [120, 121] pour estimer la réponse impulsionnelle, suivi d'une classification vectorielle.
2. Calculer la TFCT des mélanges en utilisant l'équation (5.18); seuillage du bruit par (6.4).
3. Classifier les directions spatiales estimées par les équations (6.5) et (6.6).
4. Estimer les DTF des sources par (6.7).
5. Synthétiser les DTF des signaux estimées en utilisant la TFCT inverse [135].

l'algorithme de séparation résumé dans le tableau 6.1 dont le détail est donné dans ce qui suit.

Premièrement, on calcule la TFCT des mélanges  $\mathcal{S}_x(t, f)$ , en appliquant l'équation (5.3) à chacun des mélanges de  $x(t)$ . Une fois les transformées temps-fréquence des mélanges obtenues, on appliquera un seuillage, qui atténuera l'effet du bruit et réduira le coût de calcul, puisque seule les points TF sélectionnés seront traités par notre algorithme. Plus précisément, pour chaque trame de fréquence  $(t, f_p)$  de la représentation TF, on applique le critère suivant pour tous les points temporels  $t_q$  appartenant à cette trame :

$$\text{Si } \frac{\|\mathcal{S}_x(t_q, f_p)\|}{\max_t \{\|\mathcal{S}_x(t, f_p)\|\}} > \epsilon_1, \quad \text{garder } (t_q, f_p), \quad (6.4)$$

où  $\epsilon_1$  est un seuil de faible valeur (typiquement  $\epsilon_1 = 0.01$ ). Alors, l'ensemble des points sélectionnés  $\Omega$ , est exprimé par  $\Omega = \bigcup_{i=1}^N \Omega_i$ . Notons que, l'effet d'étalement de l'énergie du bruit et la localisation de celle des sources dans le domaine TF permet d'augmenter la robustesse de la méthode proposée vis à vis du bruit (voir la Partie IV de [129]). Ainsi, en utilisant l'équation (6.4), nous souhaitons garder uniquement les points TF où l'énergie du signal est significative. De plus, en raison de l'étalement de l'énergie du bruit dans le plan TF, la contribution du bruit dans les points TF des sources est relativement, négligeable du moins pour des valeurs

modérées et hautes du RSB.

Après le seuillage du bruit, la procédure de classification peut être réalisée comme suit : pour chaque point TF, on calcule son vecteur de direction spatiale par

$$\mathbf{v}(t, f) = \frac{\mathcal{S}_{\mathbf{x}}(t, f)}{\|\mathcal{S}_{\mathbf{x}}(t, f)\|}, \quad (t, f) \in \Omega, \quad (6.5)$$

et on forcera son premier élément à être réel et positif, et ceci sans perte de généralité. Ensuite, nous regroupons ces vecteurs en  $N$  classes  $\{\mathcal{C}_i | i \in \mathcal{N}\}$  en minimisant le critère

$$\mathbf{v}(t, f) \in \mathcal{C}_i \iff i = \arg \min_k \left\| \mathbf{v}(t, f) - \frac{\hat{\mathbf{h}}_k(f) e^{-j\theta_k}}{\|\hat{\mathbf{h}}_k(f)\|} \right\|^2 \quad (6.6)$$

où  $\hat{\mathbf{h}}_k(f)$  est la transformée de Fourier du  $k^{\text{e}}$  vecteur canal estimé (voir section 6.3.1) et  $\theta_k$  est l'argument de  $\hat{h}_{k1}(f)$  (afin de forcer la première élément à être réel positif). La collecte de tous les points correspondant à tous les vecteurs de la classe  $\mathcal{C}_i$ , formera le support TF  $\Omega_i$  de la source  $s_i(t)$ .

Alors, on peut estimer la TFCT de chaque source  $s_i(t)$  par :

$$\hat{\mathcal{S}}_{s_i}(t, f) = \begin{cases} \frac{\hat{\mathbf{h}}_i^H(f)}{\|\hat{\mathbf{h}}_i(f)\|^2} \mathcal{S}_{\mathbf{x}}(t, f), & \forall (t, f) \in \Omega_i, \\ 0, & \text{sinon,} \end{cases} \quad (6.7)$$

et à partir de l'équation (6.3), on a :

$$\frac{\hat{\mathbf{h}}_i^H(f)}{\|\hat{\mathbf{h}}_i(f)\|^2} \mathcal{S}_{\mathbf{x}}(t, f) = \frac{\hat{\mathbf{h}}_i^H(f) \mathbf{h}_i(f)}{\|\hat{\mathbf{h}}_i(f)\|^2} \mathcal{S}_{s_i}(t, f) \approx \mathcal{S}_{s_i}(t, f), \quad \forall (t, f) \in \Omega_i.$$

### 6.3.3 Algorithme TF-CUBSS basé sur la projection en sous-espace

Nous avons exposé ci-dessus l'algorithme TF-CUBSS basé sur la classification. Cette méthode se base sur l'hypothèse que les sources sont disjointes dans le domaine TF, ce qui a mené à la structure TF décrite par l'équation (6.3). Cette dernière n'est plus valable si les sources sont non-disjointes dans le domaine temps-fréquence.

Sous la condition TF-non-disjoints, établie dans l'hypothèse 5.1, nous proposons dans cette section une méthode alternative utilisant la projection en sous-espace. On rappelle que les deux premières étapes de l'algorithme TF-CUBSS basé sur la classification ne sont pas fondées sur l'hypothèse que les sources soit TF-disjointes (voir le tableau 6.1). Par conséquent, on peut réutiliser ces étapes pour obtenir le canal estimé et tous les points TF de  $\Omega$ . Sous la condition TF-non-disjoints,

on considère le point  $(t, f) \in \Omega$  tel qu'il y a  $\mathcal{J}$  sources  $s_{\alpha_1}(t), \dots, s_{\alpha_{\mathcal{J}}}(t)$  présentes en ce point, avec  $\mathcal{J} < M$  où  $\alpha_1, \dots, \alpha_{\mathcal{J}} \in \mathcal{N}$  dénotent les indices des sources actives. Notre but est d'identifier les sources présentes au point  $(t, f)$  et estimer la contribution d'énergie de chacune.

Définissons les termes suivants :

$$\tilde{\mathbf{s}}(t) = [s_{\alpha_1}(t), \dots, s_{\alpha_{\mathcal{J}}}(t)]^T, \quad (6.8a)$$

$$\tilde{\mathbf{H}}_{\alpha}(f) = [\mathbf{h}_{\alpha_1}(f), \dots, \mathbf{h}_{\alpha_{\mathcal{J}}}(f)]. \quad (6.8b)$$

Ainsi, l'équation (6.2) peut être réduite à :

$$\mathcal{S}_{\mathbf{x}}(t, f) = \tilde{\mathbf{H}}_{\alpha}(f) \mathcal{S}_{\tilde{\mathbf{s}}}(t, f). \quad (6.9)$$

Soit  $\tilde{\mathbf{H}}_{\beta}(f) = [\mathbf{h}_{\beta_1}(f), \dots, \mathbf{h}_{\beta_{\mathcal{J}}}(f)]$  et  $\mathcal{Q}_{\beta}(f)$  la matrice de projection orthogonale sur le sous-espace bruit de  $\tilde{\mathbf{H}}_{\beta}(f)$  calculée par :

$$\mathcal{Q}_{\beta}(f) = \mathbf{I}_M - \tilde{\mathbf{H}}_{\beta}(f) \left( \tilde{\mathbf{H}}_{\beta}^H(f) \tilde{\mathbf{H}}_{\beta}(f) \right)^{-1} \tilde{\mathbf{H}}_{\beta}^H(f). \quad (6.10)$$

Alors, on a l'observation suivante :

$$\begin{cases} \mathcal{Q}_{\beta}(f) \mathbf{h}_i(f) = \mathbf{0}, & i \in \{\beta_1, \dots, \beta_{\mathcal{J}}\} \\ \mathcal{Q}_{\beta}(f) \mathbf{h}_i(f) \neq \mathbf{0}, & \text{sinon} \end{cases}. \quad (6.11)$$

Par conséquent, comme  $\mathcal{S}_{\mathbf{x}}(t, f) \in \text{Range}\{\tilde{\mathbf{H}}_{\alpha}(f)\}$ , on a :

$$\begin{cases} \mathcal{Q}_{\beta}(f) \mathcal{S}_{\mathbf{x}}(t, f) = \mathbf{0}, & \text{si } \{\beta_1, \dots, \beta_{\mathcal{J}}\} = \{\alpha_1, \dots, \alpha_{\mathcal{J}}\} \\ \mathcal{Q}_{\beta}(f) \mathcal{S}_{\mathbf{x}}(t, f) \neq \mathbf{0}, & \text{sinon} \end{cases}. \quad (6.12)$$

Si  $\mathbf{H}(f)$  a déjà été estimée par la méthode présentée dans la section 6.3.1, alors à partir des observations précédentes, on peut déduire un critère pour détecter les indices  $\alpha_1, \dots, \alpha_{\mathcal{J}}$ , et donc la contribution de chaque sources en ce point TF. En pratique, en tenant compte de l'effet du bruit, on détecte les vecteurs colonnes de  $\tilde{\mathbf{H}}_{\alpha}(f)$  en minimisant :

$$\{\alpha_1, \dots, \alpha_{\mathcal{J}}\} = \arg \min_{\beta_1, \dots, \beta_{\mathcal{J}}} \{ \|\mathcal{Q}_{\beta}(f) \mathcal{S}_{\mathbf{x}}(t, f)\| \}. \quad (6.13)$$

Ensuite, les valeurs des DTF des  $\mathcal{J}$  sources au point TF  $(t, f)$  sont estimées par :

$$\hat{\mathcal{S}}_{\tilde{\mathbf{s}}}(t, f) \approx \tilde{\mathbf{H}}_{\alpha}^{\#}(f) \mathcal{S}_{\mathbf{x}}(t, f), \quad (6.14)$$

Dans nos simulations, le problème d'optimisation de l'équation (6.13) est résolu en utilisant une recherche exhaustive. Cette solution reste raisonnable du point de vu coût de calcul pour un nombre de capteurs assez faible, mais devient prohibitive si  $M$  est très grand. Le tableau 6.2 donne un résumé de l'algorithme TF-CUBSS basé sur la projection en sous-espace.

Tab. 6.2 – Algorithme TF-CUBSS basé sur la projection en sous-espace

1. Estimer le canal ; critère AIC [119] pour détecter le nombre de sources et appliquer un algorithme d'identification aveugle du canal [120, 121] pour estimer la réponse impulsionnelle, suivi d'une classification vectorielle.
2. Calculer la TFCT des mélanges en utilisant l'équation (5.18); seuillage du bruit par (6.4).
3. Pour tous les points TF, détecter les sources actives en utilisant (6.10) et(6.13).
4. Estimer les DTF des sources par (6.14).
5. Synthétiser les DTF des signaux estimées en utilisant la TFCT inverse [135].

## 6.4 Discussion

Nous discuterons dans cette section certains points relatifs à l'algorithme TF-CUBSS proposé. Nous noterons que plusieurs points de discussion étudiés dans la section 5.5 peuvent être repris pour cet algorithme, notamment les discussions sur le nombre de sources, le recouvrement des sources et la qualité de séparation par rapport au nombre de sources.

### Le cas sur-déterminé

Notre algorithme peut être simplifié dans le cas sur-déterminé où  $M \geq N$ . Dans ce contexte, l'algorithme peut être réduit aux étapes d'estimation du canal, calcul de la TFCT et seuillage du bruit et enfin estimation des TFCT des sources en utilisant la pseudo-inverse de la matrice de mélange pour chaque fréquence :

$$\hat{\mathcal{S}}_s(t, f) = \hat{\mathbf{H}}^\#(f) \mathcal{S}_x(t, f) .$$

### Utilisation de DTF quadratique

L'algorithme présenté dans ce chapitre utilise la TFCT afin d'exploiter la parcimonie des signaux sources. En se basant sur les résultats obtenus par l'algorithme

TF-UBSS présenté dans le chapitre précédent, on peut envisager une version de l'algorithme TF-CUBSS utilisant la DTF quadratique en gardant les mêmes principes que ceux de l'algorithme TF-CUBSS décrit dans les sections précédentes. Plus précisément pour le cas des sources TF-disjointes, une fois que le vecteur propre principal de  $\mathbf{D}_{xx}^{\text{dvw}}(t, f)$  est extrait et comparé aux vecteurs colonnes de  $\mathbf{H}(f)$ , on trouve la direction spatiale<sup>1</sup>  $\mathbf{h}_i(f)$  associée à  $\mathbf{D}_{xx}^{\text{dvw}}(t, f)$  et par conséquent :

$$\hat{\rho}_{s_i}^{\text{dvw}}(t, f) = \mathbf{h}_i(f)^H \mathbf{D}_{xx}^{\text{dvw}}(t, f) \mathbf{h}_i(f) .$$

De la même manière, on peut étendre ce résultat au cas des sources TF-non-disjointes, en supposant que la détection des sources contribuant au même point TF est effectuée par le biais de la projection en sous-espace, et donc les estimées TF des sources<sup>2</sup>  $\widetilde{\mathbf{D}}_{ss}^{\text{dvw}}(t, f)$  sont calculées par :

$$\widetilde{\mathbf{D}}_{ss}^{\text{dvw}}(t, f) \approx \widetilde{\mathbf{H}}(f)^\# \mathbf{D}_{xx}^{\text{dvw}}(t, f) (\widetilde{\mathbf{H}}(f)^\#)^H .$$

## 6.5 Simulations

Dans cette section, nous présentons quelques résultats de simulation pour illustrer l'efficacité de notre algorithme de séparation TF-CUBSS dans le cas du mélange convolutif. Pour cela, nous considérons une antenne composée de  $M = 3$  capteurs recevant  $N = 4$  signaux sources audio. Les canaux de mélanges sont choisis de manière aléatoire avec  $\mathcal{K} = 8$ . La taille des observations  $T = 50000$  échantillons (les signaux sources sont échantillonnés à une fréquence de 8 KHz). Les signaux observés sont corrompus par un bruit blanc additif de covariance  $\sigma^2 \mathbf{I}_M$  ( $\sigma^2$  étant la puissance du bruit). La qualité de la séparation est mesurée par l'erreur quadratique moyenne normalisée (EQMN) des sources estimées pour  $N_r = 100$  réalisations aléatoires du bruit.

Dans la figure 6.1, les quatre figures de la première ligne, illustrent les représentations TF des signaux sources originaux. Les trois figures de la ligne du milieu, illustrent les représentations TF des  $M$  mélanges et les quatre dernières illustrent les représentation TF des sources estimées par l'algorithme TF-CUBSS basé sur la projection en sous-espace (tableau 6.2). La figure 6.2 représente les mêmes signaux dans le domaine temporel.

<sup>1</sup>Nous supposons sans perte de généralité que les vecteurs  $\mathbf{h}_i(f)$  sont normalisés  $\|\mathbf{h}_i(f)\| = 1$ .

<sup>2</sup>Le vecteur  $\widetilde{\mathbf{s}}$  et la matrice  $\widetilde{\mathbf{H}}$  sont définis de la même façon que dans la section précédente.

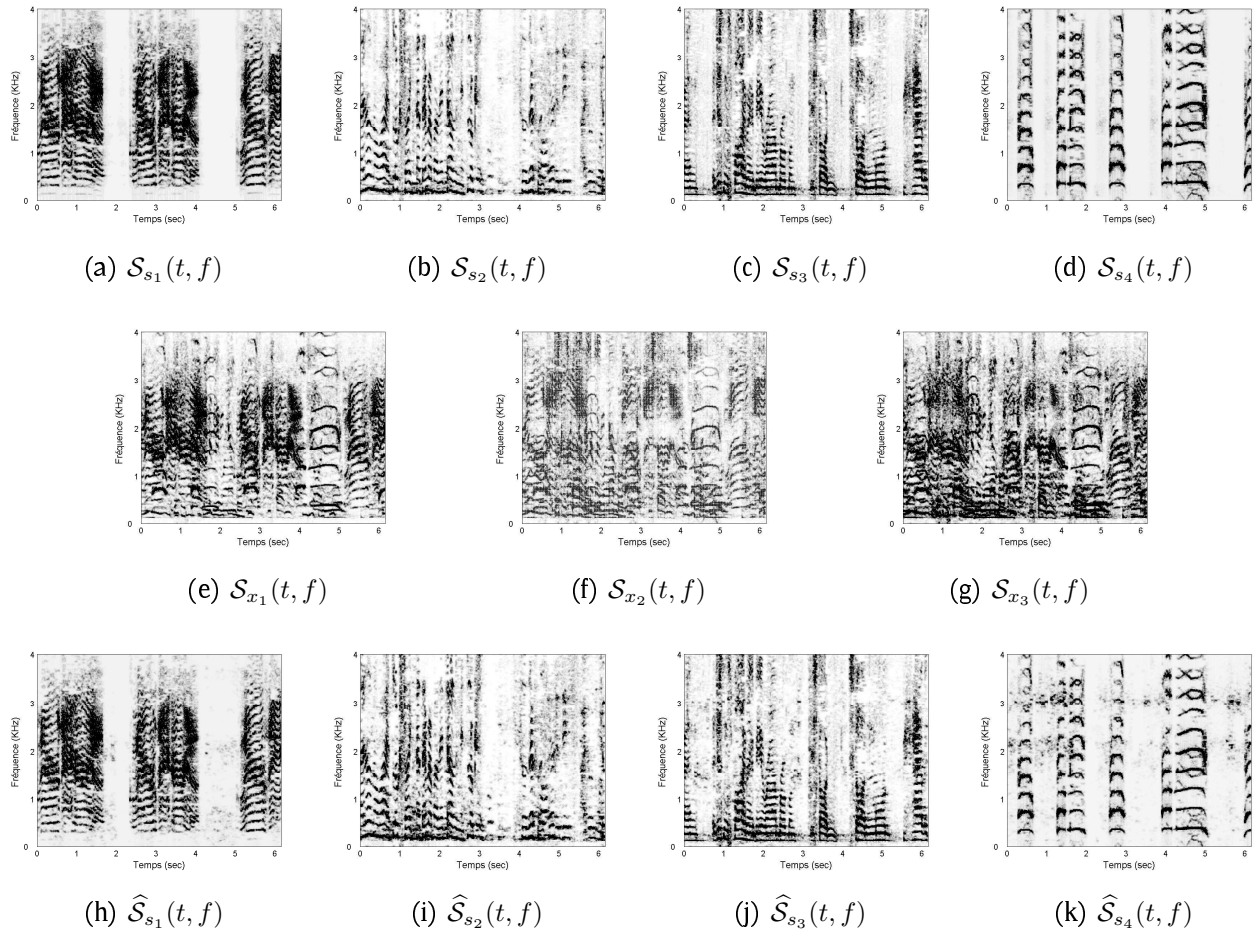
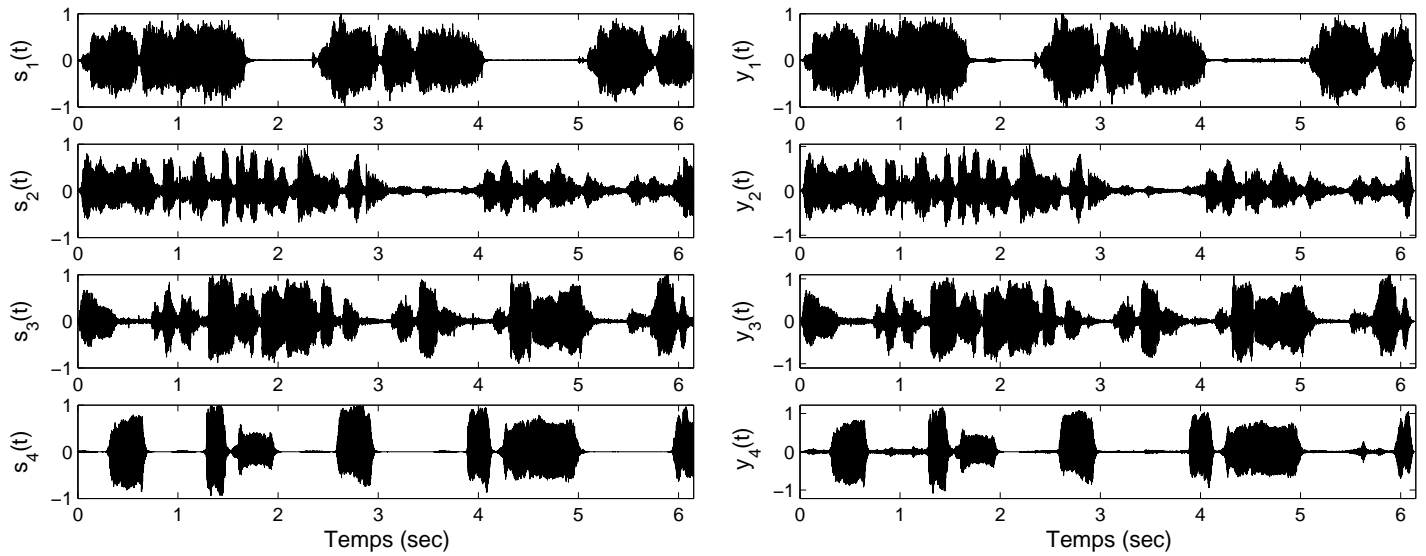


Fig. 6.1 – Exemple de simulation de l’algorithme TF-CUBSS basé sur la projection en sous-espace dans le cas de  $N = 4$  sources audio et  $M = 3$  capteurs. Les quatre premières figures représentent les signaux sources, les trois figures du milieu représentent les trois mélanges, et les quatre dernières figures représentent les signaux sources estimés par notre algorithme.

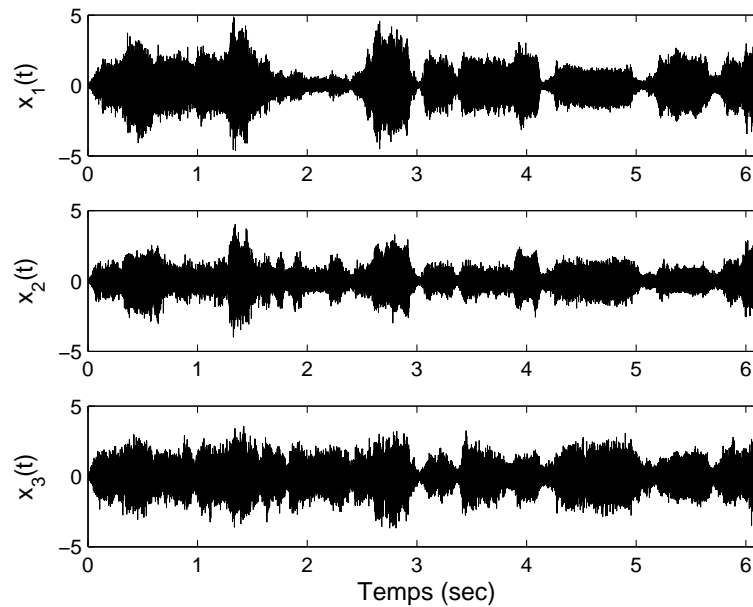
Dans la figure 6.3, on compare les performances de séparation obtenues par l’algorithme TF-CUBSS basé sur la projection en sous-espace avec  $\mathcal{J} = 2$  et l’algorithme TF-CUBSS basé sur la classification. On observe que la version de l’algorithme TF-CUBSS basé sur la projection en sous-espace produit de meilleurs résultats que celle utilisant la classification. C’est dû principalement au fait que le nombre de points de recouvrement dans le domaine TF est élevé, à tel point que l’hypothèse de TF-disjoints utilisée dans l’algorithme TF-CUBSS basé sur la classification est mal satisfaite.

La figure 6.4 (respectivement la figure 6.5) illustrent les performances de séparation de l’algorithme TF-CUBSS basé sur la classification (respectivement



(a) Signaux sources

(b) Signaux estimés



(c) Signaux mélanges

Fig. 6.2 – Exemple de simulation (dans le domaine temporel) de l'algorithme TF-CUBSS basé sur la projection en sous-espace dans le cas de  $N = 4$  sources audio et  $M = 3$  capteurs. (a) les signaux sources originaux, (b) les signaux estimés par notre algorithme, (c) les signaux capteurs (mélanges).

l'algorithme TF-CUBSS basé sur la projection en sous-espace) dans le cas où le



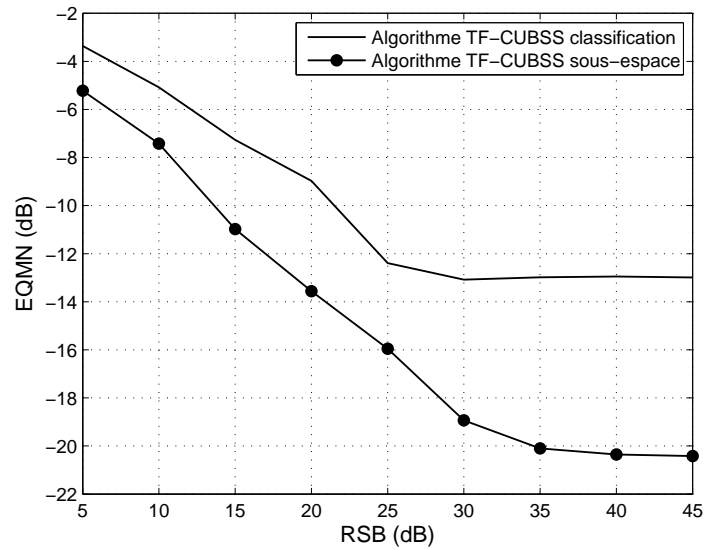


Fig. 6.3 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances des algorithmes TF-CUBSS basés sur la classification et sur la projection en sous-espace.

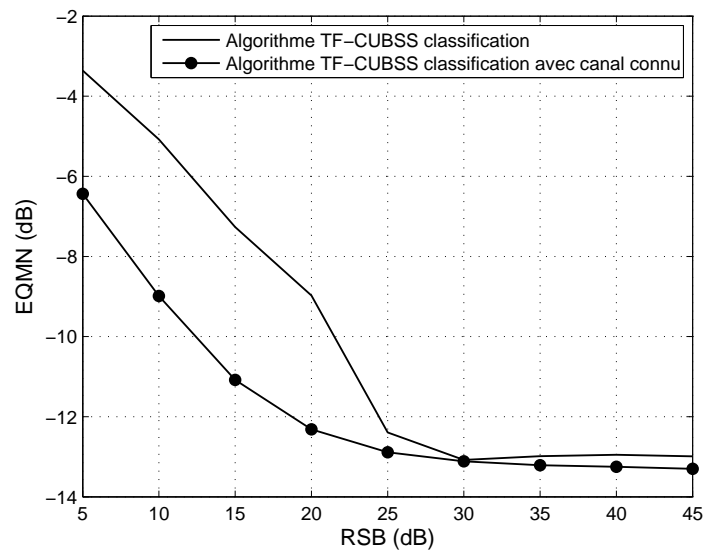


Fig. 6.4 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances de séparation de l’algorithme TF-CUBSS basé sur la classification dans le cas où le canal de mélange  $\mathbf{H}$  est exactement connu, et celles obtenues en utilisant notre méthode d’estimation du canal  $\mathbf{H}$ .

canal de mélange  $\mathbf{H}$  est a priori connu, comparées à celles obtenues en utilisant notre méthode d'estimation du canal. On observe une dégradation de la qualité de séparation pour les faibles RSB en raison de l'importance de l'erreur d'estimation du canal. Mais cette dégradation devient négligeable pour les hauts RSB.

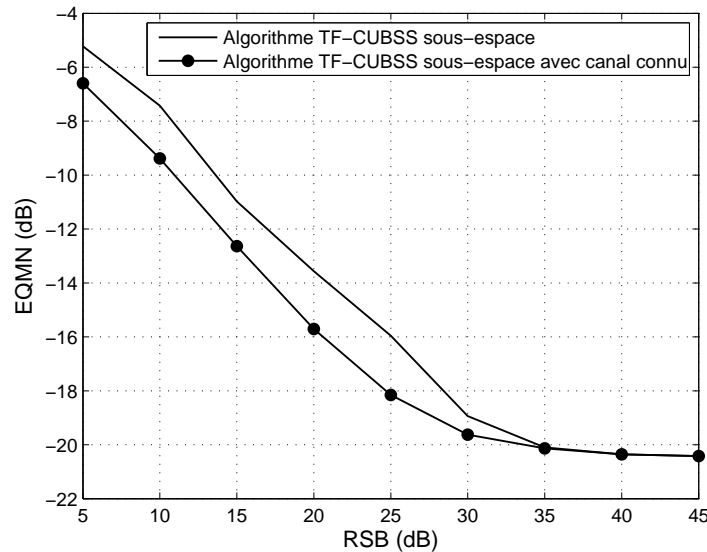


Fig. 6.5 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 50000$  : comparaison des performances de séparation de l'algorithme TF-CUBSS basé sur la projection en sous-espace dans le cas où le canal de mélange  $\mathbf{H}$  est exactement connu, et ceux obtenues en utilisant notre méthode d'estimation du canal  $\mathbf{H}$ .

Dans la figure 6.6, on compare les performances obtenues par l'algorithme TF-CUBSS basé sur la projection en sous-espace pour  $\mathcal{J} = 2$  et  $\mathcal{J} = 3$  ainsi que l'algorithme TF-CUBSS basé sur la classification. Dans cette expérience, nous avons utilisé  $M = 4$  capteurs et  $N = 5$  signaux sources. On peut observer que pour les hauts RSB, le cas où  $\mathcal{J} = 3$  produit de meilleures performances de séparation que ceux obtenues dans le cas où  $\mathcal{J} = 2$ . Cependant, pour les bas RSB, une grande valeur de  $\mathcal{J}$  augmente le bruit d'estimation (comme mentionné dans la section 6.4) et par conséquent dégrade la qualité de la séparation.

La figure 6.7 illustre la rapide dégradation de la qualité de la séparation lors de l'augmentation du nombre de sources qui varie entre  $N = 4$  et  $N = 7$ . Ceci confirme les remarques faites dans la section 6.4.

La figure 6.8 représente les performances de l'algorithme TF-CUBSS basé sur la projection en sous-espace en considérant des canaux à réponses impulsion-

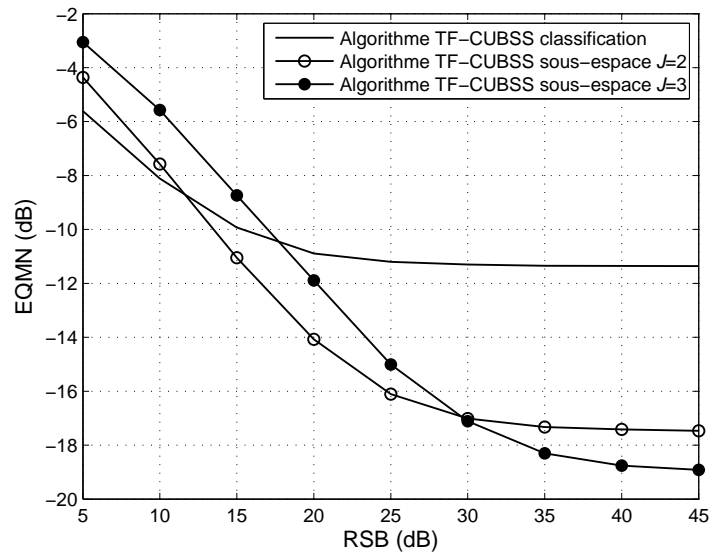


Fig. 6.6 – EQMN en fonction du RSB pour  $N = 5$  sources audio et  $M = 4$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances de l’algorithme TF-CUBSS basé sur la classification et l’algorithme TF-CUBSS basé sur la projection en sous-espace pour différentes valeurs de  $\mathcal{J}$ .

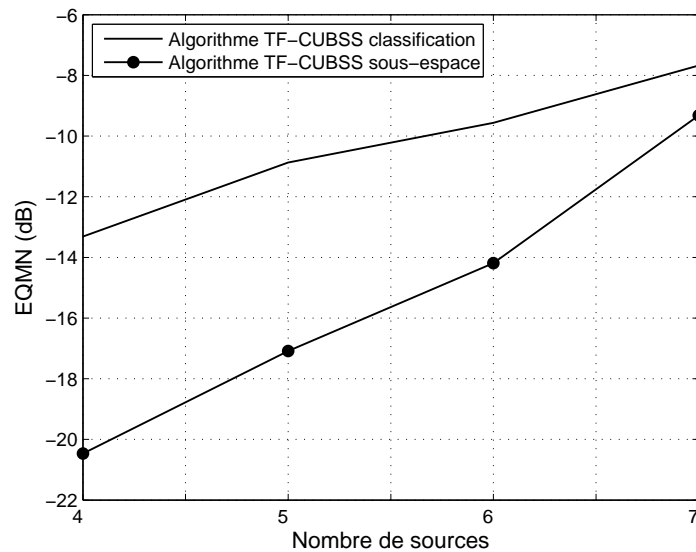


Fig. 6.7 – EQMN en fonction du nombre de sources  $N$  pour  $M = 3$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances des algorithmes TF-CUBSS basés sur la classification et celui basé sur la projection en sous-espace.

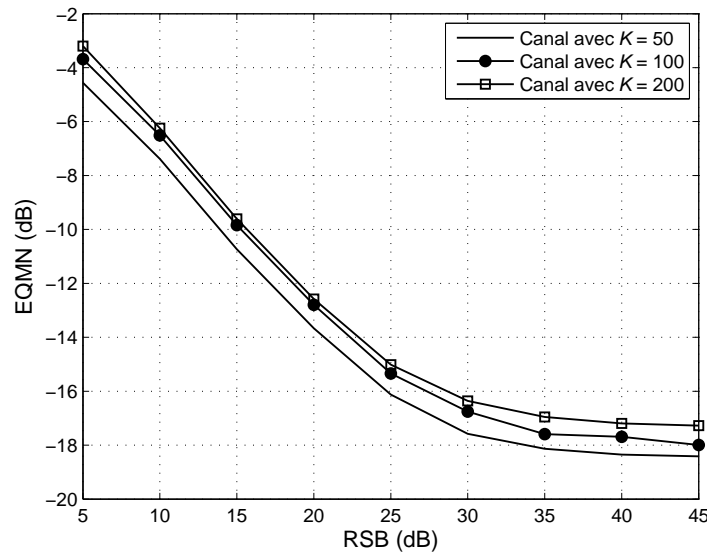


Fig. 6.8 – EQMN en fonction du RSB pour  $N = 4$  sources audio et  $M = 3$  capteurs avec une taille d'observation  $T = 50000$  : comparaison des performances de séparation de l'algorithme TF-CUBSS basé sur la projection en sous-espace pour différentes valeurs de la taille du canal de mélange  $\mathcal{K}$ .

nelles longues. Plus précisément, les courbes représentent les performances de séparation pour des canaux de taille  $\mathcal{K} = 50, 100$  et  $200$  respectivement. On observe une légère dégradation des performances quand la taille des canaux augmente, mais la qualité de séparation reste tout à fait acceptable.

Dans la figure 6.9, on compare les performances de séparation de notre algorithme TF-CUBSS, de l'algorithme de Y. Deville dans [118] et l'algorithme de L. Parra dans [60] dans le cas sur-déterminé avec  $N = 2$  signaux sources et  $M = 2$  capteurs. Les algorithmes cités dans [60, 118] séparent les signaux sources à un filtre près et par conséquent, on utilise dans cette expérience le critère de rapport interférences à signal fréquentiel (RISF) défini dans [60] à défaut d'utiliser l'EQMN. On observe un gain de performances significatif en faveur de notre algorithme et particulièrement pour les hauts RSB. De plus, notre algorithme possède les avantages suivants : (i) il peut traiter le cas sous-déterminé, (ii) il estime les sources à une constante scalaire près et non à un filtre près comme dans [60, 118], (iii) la procédure de sélection des intervalles de temps que nous proposons ne nécessite aucun seuillage contrairement à la procédure utilisée dans [118].

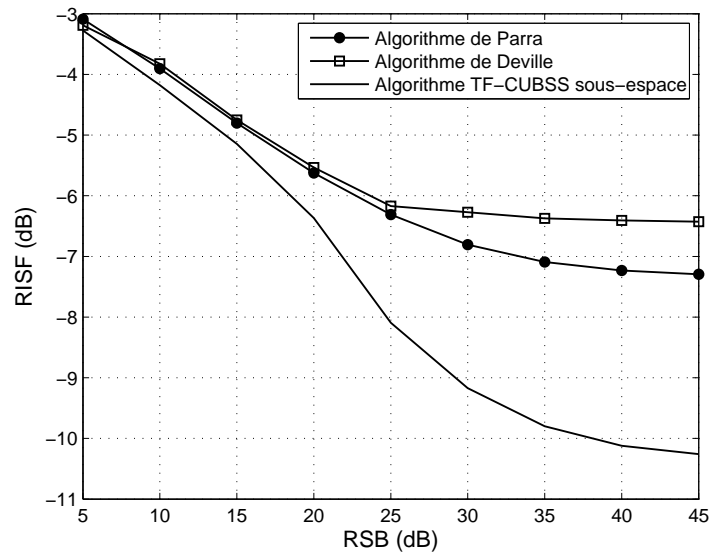


Fig. 6.9 – RISF en fonction du RSB pour  $N = 2$  sources audio et  $M = 2$  capteurs avec une taille d’observation  $T = 50000$  : comparaison des performances de séparation de l’algorithme TF-CUBSS basé sur la projection en sous-espace, l’algorithme de Y. Deville et l’algorithme de L. Parra.

## 6.6 Conclusion

Dans ce chapitre, nous avons introduit deux nouvelles méthodes pour la séparation aveugle de sources dans le cas sous-déterminé convolutif. Ces méthodes utilisent la représentation temps-fréquence des signaux sources en les supposant TF-disjointes pour la première méthode, et TF-non-disjointes pour la deuxième. La première méthode proposée a l’avantage de la simplicité tandis que la seconde utilise une hypothèse moins contraignante sur la parcimonie des sources, en supposant que les sources ne sont pas nécessairement disjointes dans le domaine temps-fréquence, et en appliquant un traitement explicite au point de recouvrement par le biais de la projection en sous-espace, ce qui permet d’améliorer la qualité de séparation. Des résultats de simulations illustrent les performances de nos méthodes dans différents cas de figure.



# Troisième partie

## La SAS audio utilisant les techniques itératives

---

Dans cette dernière partie, nous présenterons deux classes de techniques qui réalisent la séparation de sources en minimisant des fonctions de contraste de manière itérative. La première classe présentée utilise une fonction de contraste basée sur la norme  $\ell_p$  qui est considérée comme une bonne mesure de la parcimonie. En effet, nous exploiterons la propriété de parcimonie des signaux audio dans le domaine temporel à travers cette fonction de contraste. La deuxième classe introduite utilise un principe de séparation de sources plus classique qui repose sur la décorrélation des observations. Cette décorrélation est effectuée en exploitant des fonctions de contraste basées sur les statistiques d'ordre deux uniquement. Notons que ces techniques sont dédiées uniquement au cas sur-déterminé, du fait qu'elles reposent sur l'hypothèse que la matrice de mélange soit inversible.





# La SAS audio utilisant la parcimonie temporelle

---

## 7.1 Introduction

Nous avons présenté précédemment, plusieurs méthodes de SAS dédiées aux signaux non-stationnaires, et en particulier aux signaux audio. Certaines de ces méthodes utilisent l'indépendance statistique des sources en supposant qu'elles sont localement stationnaires afin de résoudre le problème de la séparation [33]. C'est une hypothèse forte qui n'est pas toujours vérifiée [139]. Pour remédier à ce problème, nous proposerons dans ce chapitre une nouvelle méthode de séparation dans le cas sur-déterminé qui manie le modèle de mélange instantané en utilisant l'hypothèse de parcimonie des sources dans le domaine temporel. Puis, on généralisera cet algorithme au cas du mélange convolutif, en transformant le problème du mélange convolutif en plusieurs mélanges instantanés dans le domaine fréquentiel, et en séparant ces mélanges pour chaque composante fréquentielle.

L'utilisation de la parcimonie a fait son apparition dans le domaine de la séparation de sources à la fin des années 90 [18, 77, 140, 141], utilisant la décomposition des signaux dans une base de dictionnaire ou des modèles probabilistes des signaux sources. Dans ce chapitre nous proposons une méthode dédiée au cas du mélanges instantané sur-déterminé ( $M \geq N$ ) où les observations sont

régies par le modèle (1.1) avec une matrice de mélange  $\mathbf{A}$  de rang plein. Cette méthode est basée sur la minimisation d'une fonction de contraste utilisant la norme  $\ell_p$ , par le biais d'une technique du gradient relatif.

## 7.2 L'algorithme ISBS pour les mélanges instantanés

Avant de commencer, nous proposons d'appliquer une étape optionnelle de blanchiment afin de remettre les mélanges au même niveau d'énergie et de réduire le nombre de paramètres à estimer. Plus précisément, l'étape de blanchiment est appliquée aux mélanges avant l'utilisation de notre algorithme. Le blanchiment est effectué en appliquant aux mélanges une matrice  $\mathbf{W}$  de taille  $N \times M$  tel que  $\mathbb{E}[\mathbf{W}\mathbf{x}(t)\mathbf{x}(t)^H\mathbf{W}^H] = \mathbf{I}_N$ . Comme montré dans la section 1.2.2, la matrice  $\mathbf{W}$  peut être calculée comme l'inverse de la racine carrée de la matrice de covariance des mélanges  $\mathbf{R}_x = \mathbb{E}[\mathbf{x}(t)\mathbf{x}(t)^H]$  dans le cas sans bruit (pour plus de détails voir [31]). Dans ce qui suit, nous appliquerons notre algorithme de séparation sur les données blanchies :

$$\mathbf{x}_w(t) = \mathbf{W}\mathbf{x}(t) .$$

Nous proposons dans ce qui suit un algorithme itératif pour la séparation de signaux audio parcimonieux, dénommé ISBS pour *Iterative Sparse Blind Separation*. Il est bien connu que les signaux audio sont caractérisés par leurs propriété de parcimonie dans le domaine temporel [18, 141], qui est mesurée par leurs norme  $\ell_p$  où  $0 \leq p < 2$ . Plus précisément, nous définissons la fonction de contraste basée sur la parcimonie comme suit :

$$\mathcal{F}_p(\mathbf{s}) = \sum_{i=1}^N [\mathfrak{J}_p(s_i)]^{\frac{1}{p}} , \quad (7.1)$$

où

$$\mathfrak{J}_p(s_i) = \frac{1}{T} \sum_{t=0}^{T-1} |s_i(t)|^p . \quad (7.2)$$

L'algorithme permet d'estimer une matrice de séparation  $\mathbf{B}$  tel que,

$$\mathbf{B} = \arg \min_{\mathbf{B}} \{ \mathcal{G}_p(\mathbf{B}) \} , \quad (7.3)$$

où

$$\mathcal{G}_p(\mathbf{B}) \triangleq \mathcal{F}_p(\mathbf{z}) , \quad (7.4)$$

et  $\mathbf{z}(t) \triangleq \mathbf{B}\mathbf{x}_w(t)$  représente les estimées des sources. L'approche choisie pour résoudre (7.3) est inspirée de [53]. C'est une technique en blocs qui consiste à

rechercher itérativement le minimum de l'équation (7.3) sous la forme :

$$\mathbf{B}^{(k+1)} = (\mathbf{I}_N + \boldsymbol{\epsilon}^{(k)})\mathbf{B}^{(k)} \quad (7.5)$$

$$\mathbf{z}^{(k+1)}(t) = (\mathbf{I}_N + \boldsymbol{\epsilon}^{(k)})\mathbf{z}^{(k)}(t) . \quad (7.6)$$

A l'itération  $k$ , la matrice  $\boldsymbol{\epsilon}^{(k)}$  est obtenue par linéarisation locale de  $\mathcal{F}_p(\mathbf{B}^{(k+1)}\mathbf{x}_w)$ . Cette approche a l'avantage de la simplicité de calcul de  $\boldsymbol{\epsilon}^{(k)}$  sous l'hypothèse que la matrice  $\mathbf{B}^{(k)}$  est proche de la matrice de séparation. Cette procédure est décrite dans ce qui suit.

A l'itération  $(k + 1)$ , le fonction décrite par l'équation (7.3) peut être développée comme suit :

$$\begin{aligned} \mathfrak{J}_p(z_i^{(k+1)}) &= \frac{1}{T} \sum_{t=0}^{T-1} \left| z_i^{(k)}(t) + \sum_{j=1}^N \epsilon_{ij}^{(k)} z_j^{(k)}(t) \right|^p \\ &= \frac{1}{T} \sum_{t=0}^{T-1} |z_i^{(k)}(t)|^p \left| 1 + \sum_{j=1}^N \epsilon_{ij}^{(k)} \frac{z_j^{(k)}(t)}{z_i^{(k)}(t)} \right|^p . \end{aligned}$$

Sous l'hypothèse que  $\mathbf{B}^{(k)}$  est proche de la matrice de séparation, on a

$$|\epsilon_{ij}^{(k)}| \ll 1$$

et ainsi, l'approximation au premier ordre de  $\mathfrak{J}_p(z_i^{(k+1)})$  est donnée par :

$$\begin{aligned} \mathfrak{J}_p(z_i^{(k+1)}) &\approx \frac{1}{T} \sum_{t=0}^{T-1} |z_i^{(k)}(t)|^p + p \sum_{j=1}^N \Re(\epsilon_{ij}^{(k)}) \Re \left( |z_i^{(k)}(t)|^{p-1} e^{-j\phi_i^{(k)}(t)} z_j^{(k)}(t) \right) \\ &\quad - \Im(\epsilon_{ij}^{(k)}) \Im \left( |z_i^{(k)}(t)|^{p-1} e^{-j\phi_i^{(k)}(t)} z_j^{(k)}(t) \right) \end{aligned} \quad (7.7)$$

où  $\phi_i^{(k)}(t)$  est l'argument du nombre complexe  $z_i^{(k)}(t)$ .

En utilisant l'équation (7.7), l'équation (7.1) peut être réécrite sous une forme plus compacte comme suit :

$$\mathcal{G}_p(\mathbf{B}^{(k+1)}) = \mathcal{G}_p(\mathbf{B}^{(k)}) + \Re \left\{ \text{trace} \left( \boldsymbol{\epsilon}^{(k)*} \mathcal{R}^{(k)H} \mathbf{D}^{(k)H} \right) \right\} \quad (7.8)$$

où l'élément  $ij$  de la matrice  $\mathcal{R}^{(k)}$  est donné par :

$$\mathcal{R}_{ij}^{(k)} = \frac{1}{T} \sum_{t=0}^{T-1} |z_i^{(k)}(t)|^{p-1} e^{-j\phi_i^{(k)}(t)} z_j^{(k)}(t) \quad (7.9)$$

et

$$\mathbf{D}^{(k)} = \left[ \text{diag} \left( \mathcal{R}_{11}^{(k)}, \dots, \mathcal{R}_{NN}^{(k)} \right) \right]^{\frac{1}{p}-1} . \quad (7.10)$$

Tab. 7.1 – Algorithme ISBS pour *Iterative Sparse Blind Separation*

1. Initialiser  $\mathbf{B}^{(1)} = \mathbf{I}_N$  ( $\mathbf{z}^{(1)}(t) = \mathbf{B}^{(1)}\mathbf{x}_w(t)$ ).
2. Pour  $k = 1, \dots, K$ , Calculer  $\mathcal{R}^{(k)}$  et  $\mathbf{D}^{(k)}$  par (7.9) et (7.10).
3. Mettre à jour la matrice de séparation  $\mathbf{B}^{(k+1)}$  par (7.5).
4. Mettre à jour les sources estimées par (7.6).

En utilisant une technique du gradient,  $\epsilon^{(k)}$  peut être choisi comme :

$$\epsilon^{(k)} = -\mu \mathbf{D}^{(k)} \mathcal{R}^{(k)*} \quad (7.11)$$

où  $\mu > 0$  est le pas du gradient. En remplaçant l'équation (7.11) dans (7.8) on obtient,

$$\mathcal{G}_p(\mathbf{B}^{(k+1)}) = \mathcal{G}_p(\mathbf{B}^{(k)}) - \mu \|\mathbf{D}^{(k)} \mathcal{R}^{(k)}\|^2. \quad (7.12)$$

Ainsi  $\mu$  contrôle la décroissance du critère. Maintenant, afin d'éviter à l'algorithme de converger vers la solution triviale  $\mathbf{B} = \mathbf{0}$ , on normalise la puissance des sources résultantes de l'application de la matrice de séparation;  $\rho_{z_i}^{(k+1)} \triangleq \frac{1}{T} \sum_{t=0}^{T-1} |z_i^{(k+1)}(t)|^2 = 1, \forall i$ . En utilisant une approximation au premier ordre, cette normalisation nous mène à :

$$\epsilon_{ii}^{(k)} = \frac{1 - \rho_{z_i}^{(k)}}{2\rho_{z_i}^{(k)}}. \quad (7.13)$$

Après convergence de l'algorithme, la matrice de séparation  $\mathbf{B} = \mathbf{B}^{(\mathfrak{F})}$  est appliquée au signal de mélange blanchi  $\mathbf{x}_w$  afin d'obtenir une estimation des signaux sources originaux.  $\mathfrak{F}$  dénote ici le nombre d'itérations qui peut être ou bien choisi a priori ou donné par un critère d'arrêt de la forme  $\|\mathbf{B}^{(k+1)} - \mathbf{B}^{(k)}\| < \delta$  où  $\delta$  est une valeur seuil. L'algorithme ISBS peut être résumé par le tableau 7.1.

### 7.3 L'algorithme ISBS pour les mélanges convolutifs

Malheureusement, et comme nous l'avons expliqué dans les chapitres précédents, le cas du mélange instantané est peu vraisemblable dans le cas d'un enregistrement réel où le modèle convolutif est plus adéquat.

Par conséquent, et dans le but d'exploiter les qualités de l'algorithme ISBS décrit dans la section précédente, nous appliquerons une TFCT aux données observées

$\mathbf{x}(t)$  du modèle convolutif décrit par l'équation (1.5), qui devient dans le cas sans bruit :

$$\mathcal{S}_x(t, f) \approx \mathbf{H}(f)\mathcal{S}_s(t, f), \quad (7.14)$$

où  $\mathcal{S}_x(t, f)$  est le vecteur TFCT des mélanges,  $\mathcal{S}_s(t, f)$  est le vecteur TFCT des sources et  $\mathbf{H}(f) = [\mathbf{h}_1(f) \dots \mathbf{h}_N(f)]$  est la transformée de Fourier du canal de mélange. On remarque que, pour chaque trame de fréquence, le mélange convolutif est réduit à un simple mélange instantané. Donc, on peut envisager d'appliquer notre algorithme ISBS pour chaque fréquence afin de séparer nos signaux. Comme résultat, on obtient pour chaque fréquence, les estimées TF de nos sources

$$\hat{\mathcal{S}}_s(t, f) = \mathbf{B}(f)\mathcal{S}_x(t, f). \quad (7.15)$$

Il semble naturel qu'afin de reconstruire les signaux séparés, il soit nécessaire d'aligner les vecteurs  $\hat{\mathcal{S}}_s(t, f)$  obtenus pour chaque fréquence et leurs appliquer une TFCT inverse. Pour cela, nous devons d'abord résoudre les ambiguïtés de permutation et d'échelle comme montré ci-dessous.

### 7.3.1 Résolution des ambiguïtés de permutation et d'échelle

Dans cette section, nous proposons de résoudre le problème des ambiguïtés de permutation et d'échelle, en appliquant une matrice de permutation  $\mathbf{\Pi}(f)$  et une matrice d'échelle  $\mathcal{C}(f)$ .

$$\mathbf{G}(f) = \mathbf{\Pi}(f)\mathcal{C}(f)\mathbf{B}(f). \quad (7.16)$$

La matrice d'échelle  $\mathcal{C}(f)$  est une matrice diagonale de taille  $N \times N$  que l'on peut calculer comme dans [142] par  $\mathcal{C}(f) = \text{diag}(\mathbf{B}(f)^\#)$ . Pour la matrice de permutation  $\mathbf{\Pi}(f)$ , on exploite la propriété de continuité des filtres acoustique dans le domaine fréquentiel [143]. Afin d'aligner les sources estimées de deux trames de fréquences successives, on teste le degré de rapprochement entre  $\mathbf{G}(f_n)\mathbf{G}(f_{n-1})^\#$  et une matrice diagonale. En effet, en utilisant la représentation donnée par l'équation (7.16), on peut estimer la matrice de permutation en minimisant :

$$\mathbf{\Pi}(f_n) = \arg \min_{\tilde{\mathbf{\Pi}}} \left\{ \sum_{i \neq j} \left( \tilde{\mathbf{\Pi}}\mathcal{C}(f_n)\mathbf{B}(f_n)\mathbf{G}(f_{n-1})^\# \right)_{ij}^2 \right\}. \quad (7.17)$$

Dans nos simulations, nous avons utilisé une recherche exhaustive pour résoudre l'équation (7.17). Néanmoins, quand le nombre de sources est grand, la recherche exhaustive devient prohibitive. Dans ce cas, on pourra estimer  $\mathbf{\Pi}(f_n)$  comme la

matrice avec des uns dans les éléments  $ij$  satisfaisant  $|\mathcal{M}(f_n)|_{ij} = \max_k |\mathcal{M}(f_n)|_{ik}$  et des zéros ailleurs avec  $\mathcal{M}(f_n) = \mathcal{C}(f_n)\mathbf{B}(f_n)\mathbf{G}(f_{n-1})^\#$ . L'avantage de cette solution est sa simplicité. Mais malencontreusement, elle peut mener à une solution erronée dans un contexte difficile. Une solution alternative consiste à décomposer  $\mathbf{\Pi}(f_n)$  comme un produit de permutations élémentaires<sup>1</sup>  $\mathbf{\Pi}_{(pq)}$ . Cette dernière est considérée à une itération donnée, seulement si elle fait décroître le critère donné par l'équation (7.17), et donc si

$$|\mathcal{M}(f_n)|_{pq}^2 + |\mathcal{M}(f_n)|_{qp}^2 > |\mathcal{M}(f_n)|_{pp}^2 + |\mathcal{M}(f_n)|_{qq}^2$$

Finalement, on obtient :

$$\mathbf{\Pi}(f_n) = \prod_{\text{nb d'itérations}} \prod_{1 \leq p < q \leq N} \widetilde{\mathbf{\Pi}}_{(pq)}, \quad (7.18)$$

$\widetilde{\mathbf{\Pi}}_{(pq)}$  étant ou bien une matrice identité ou bien la matrice de permutation précédente  $\mathbf{\Pi}_{(pq)}$  en fonction de la règle de décision binaire définie ci-dessus. Le processus itératif est arrêté, quand toutes les matrices  $\widetilde{\mathbf{\Pi}}_{(pq)}$  sont égales à des matrices identités. Nous avons observé que une ou deux itérations sont suffisantes pour obtenir notre matrice de permutation.

Enfin, on applique la matrice de séparation mise à jour  $\mathbf{G}(f)$  au mélange dans le domaine fréquentiel :

$$\widehat{\mathcal{S}}_s(t, f) = \mathbf{G}(f)\mathcal{S}_x(t, f). \quad (7.19)$$

## 7.4 Simulations

Nous présentons ici quelques résultats de simulation pour illustrer l'efficacité de notre algorithme de séparation ISBS. Pour cela, nous considérons une antenne composée de  $M = 5$  capteurs recevant  $N = 2$  signaux sources audio arrivant suivant les angles d'arrivés  $\theta_1 = 15$  et  $\theta_2 = 45$  degrés respectivement. La taille des observations  $T = 20000$  échantillons (les signaux sources sont échantillonnés à une fréquence de 8 KHz). Les signaux observés sont corrompus par un bruit blanc additif de covariance  $\sigma^2 \mathbf{I}_M$  ( $\sigma^2$  étant la puissance du bruit). La qualité de la séparation est mesurée par le rapport interférences à signal (RIS) des sources estimées pour  $N_r = 100$  réalisations aléatoires du bruit.

$$\text{RIS} \triangleq \sum_{p \neq q} \frac{\mathbb{E} [ |(\mathbf{B}\mathbf{A})_{pq}|^2 ] \rho_q}{\mathbb{E} [ |(\mathbf{B}\mathbf{A})_{pp}|^2 ] \rho_p} \quad (7.20)$$

<sup>1</sup> $\mathbf{\Pi}_{(pq)}$  est définie tel que pour un vecteur donné  $\mathbf{y}$ ,  $\widetilde{\mathbf{y}} = \mathbf{\Pi}_{(pq)}\mathbf{y}$  si et seulement si  $\widetilde{y}(k) = y(k)$ , pour  $k \notin \{p, q\}$ ,  $\widetilde{y}(p) = y(q)$  et  $\widetilde{y}(q) = y(p)$ .

où  $\rho_i = \mathbb{E}[|s_i(t)|^2]$  est la puissance de la  $i^{\text{e}}$  source, évaluée par  $\frac{1}{T} \sum_{t=0}^{T-1} |s_i(t)|^2$ .

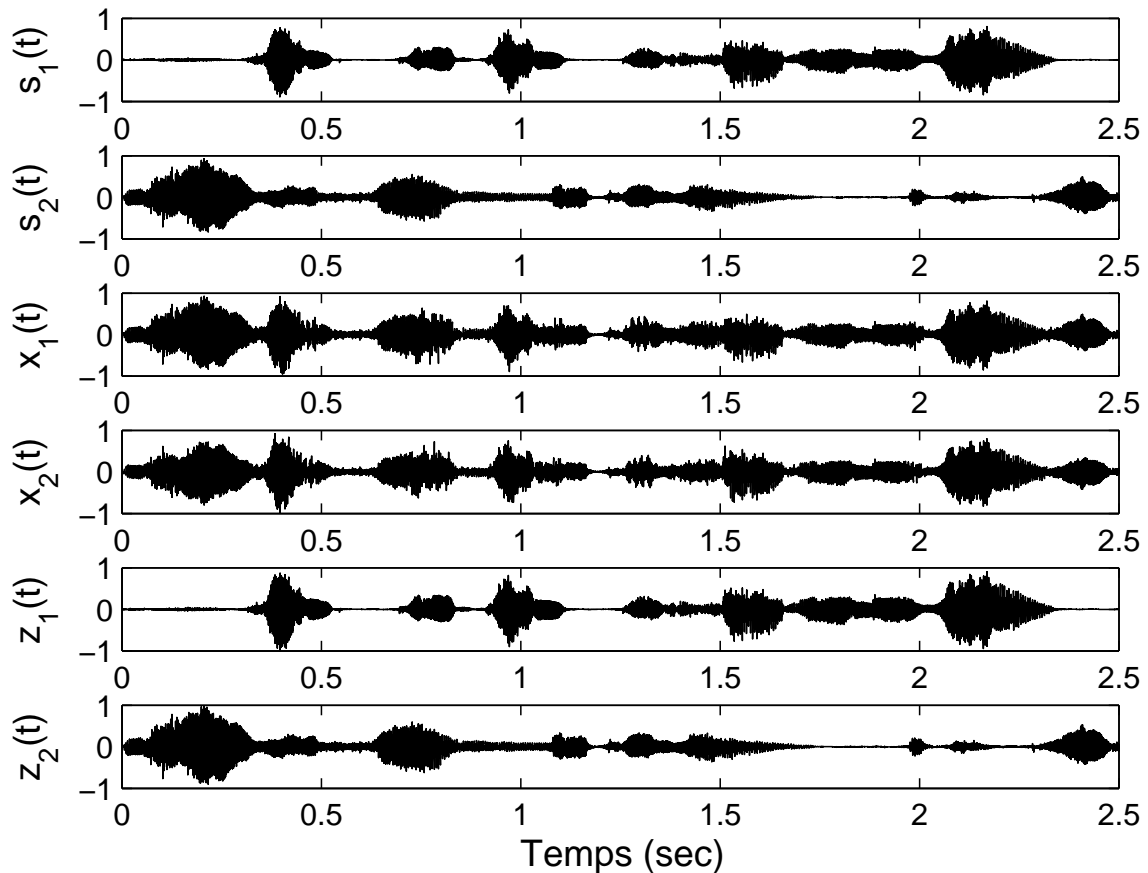


Fig. 7.1 – Exemple de simulation de l’algorithme ISBS dans le cas de  $N = 2$  sources audio et  $M = 2$  capteurs. De haut en bas on a les signaux sources originaux, les signaux capteurs (mélanges) et les signaux estimés par notre algorithme.

Dans la figure 7.1, les deux premières figures représentent les signaux sources originaux. Les deux suivantes représentent les  $M = 2$  mélanges instantanés et les deux dernières représentent les signaux sources estimés par l’algorithme ISBS.

Dans la figure 7.2, on compare les performances de séparation obtenues par notre algorithme ISBS dans le cas du mélange instantané à ceux de l’algorithme du Newton relatif développé par M. Zibulevsky dans [141] où le cas des signaux parcimonieux est considéré, ainsi que les performances obtenues par l’algorithme SOBI [31]. On représente le RIS en fonction du RSB. Il est clair que notre algorithme produit de meilleurs résultats que les deux autres méthodes. On peut établir la même conclusion pour la figure 7.3 où on représente l’EQMN des mêmes

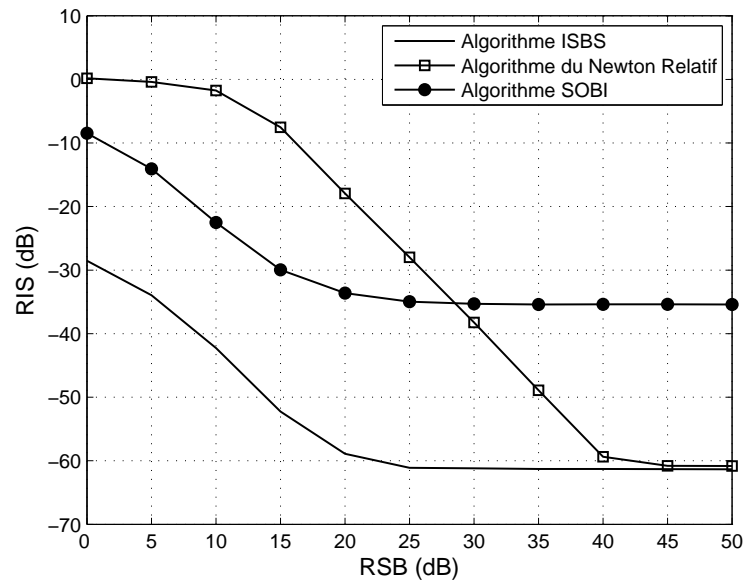


Fig. 7.2 – RIS en fonction du RSB pour  $N = 2$  sources audio et  $M = 5$  capteurs avec une taille d'observation  $T = 20000$  dans le cas du mélange instantané : comparaison des performances de séparation entre notre algorithme ISBS, l'algorithme SOBI et l'algorithme du Newton relatif.

algorithmes en fonction du RSB.

La figure 7.4 représente l'évolution du RIS en fonction du nombre d'itérations pour l'algorithme ISBS dans le cas du mélange instantané. On observe que l'algorithme ISBS nécessite une dizaine d'itérations pour converger, ce qui peut être considéré comme une bonne vitesse de convergence.

Dans la figure 7.5, on compare les performances de notre algorithme ISBS et ceux de l'algorithme SOBI en fonction du nombre de capteurs et pour différentes valeurs du RSB. On observe un net avantage pour notre algorithme pour toute la plage de variation du nombre de capteurs.

Dans la figure 7.6, on représente les performances dans le cas du mélange convolutif  $2 \times 2$  ( $N = 2$  signaux sources et  $M = 2$  capteurs), de notre algorithme ISBS, de l'algorithme de Y. Deville dans [118], de l'algorithme de L. Parra dans [60] et de l'extension de l'algorithme de M. Zibulevsky au cas du mélange convolutif. Les canaux sont choisis de telle sorte que  $\mathcal{K} = 256$ . On compare ces algorithmes en utilisant le rapport interférences à signal fréquentiel (RISF) introduit dans [60] tenant compte du fait que les sources estimées sont obtenues à un filtre près. On observe un gain significatif en faveur de notre algorithme ISBS et particulièrement pour les



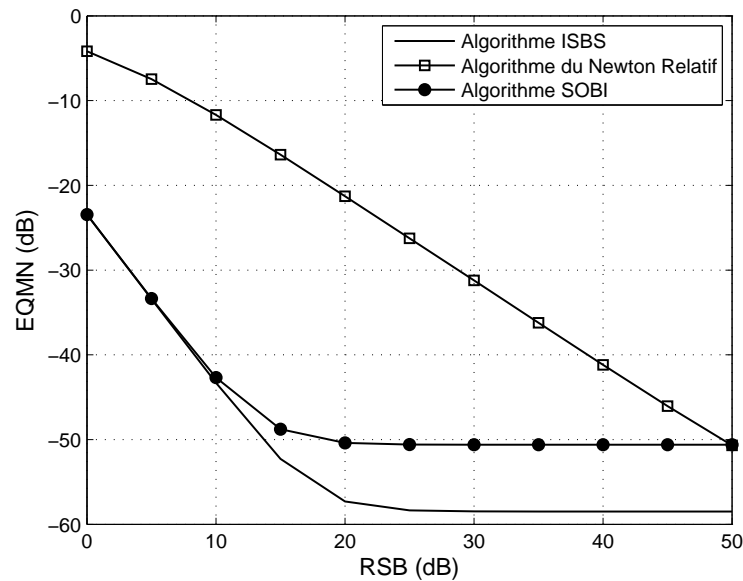


Fig. 7.3 – EQMN en fonction du RSB pour  $N = 2$  sources audio et  $M = 5$  capteurs avec une taille d’observation  $T = 20000$  dans le cas du mélange instantané : comparaison des performances de séparation entre notre algorithme ISBS, l’algorithme SOBI et l’algorithme du Newton relatif.

bas RSB. De plus, la complexité de l’algorithme ISBS est de l’ordre de  $2N^2T + \mathcal{O}(N^2)$  multiplications par itération tandis que la complexité de l’algorithme du Newton relatif dans [141] est de l’ordre de  $2N^4 + N^3T + N^6/6$  multiplications par itération.

## 7.5 Conclusion

Dans ce chapitre, nous avons présenté une méthode de SAS pour des sources parcimonieuses dans le cas du mélange instantané ainsi que son extension au cas du mélange convolutif. Nous avons aussi introduit une fonction de contraste basée sur la norme  $\ell_p$  pour la mesure de la parcimonie ainsi qu’un algorithme itératif basé sur la technique du gradient naturel pour la minimisation de cette fonction. Les résultats des simulations mettent en évidence l’utilité de la méthode. La méthode proposée devance les méthodes existantes en termes de qualité de séparation et de coût de calcul dans les deux cas instantané et convolutif.

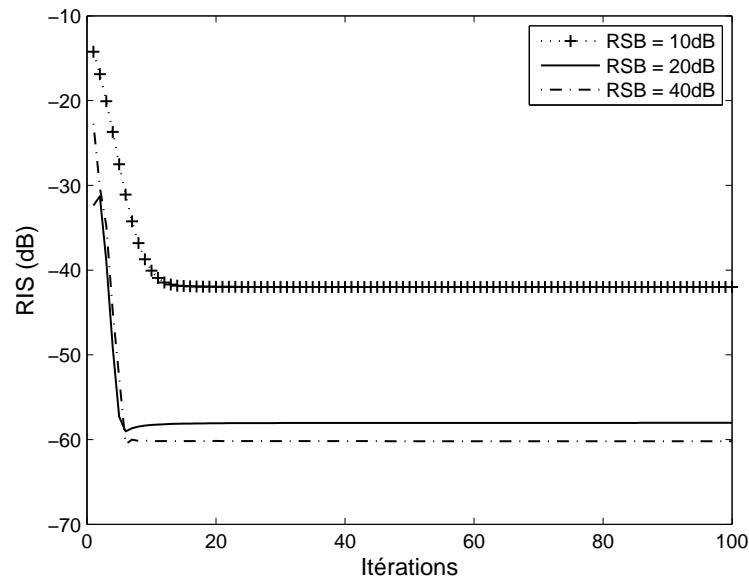


Fig. 7.4 – RIS en fonction du nombre d'itérations pour  $N = 2$  sources audio et  $M = 5$  capteurs avec une taille d'observation  $T = 20000$  dans le cas du mélange instantané : convergence de l'algorithme ISBS pour différentes valeurs du RSB.

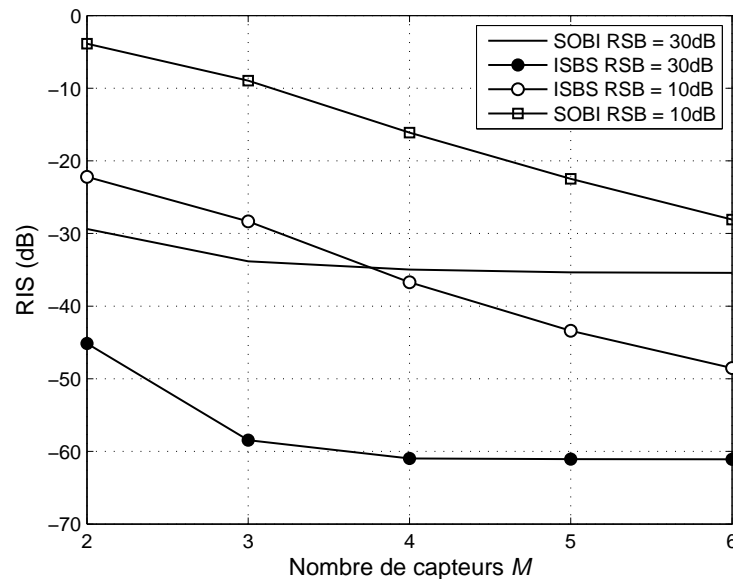


Fig. 7.5 – RIS en fonction du nombre de capteurs  $M$  pour  $N = 2$  sources audio une taille d'observation  $T = 20000$  dans le cas du mélange instantané : comparaison des performances de séparation entre notre algorithme ISBS et l'algorithme SOBI.

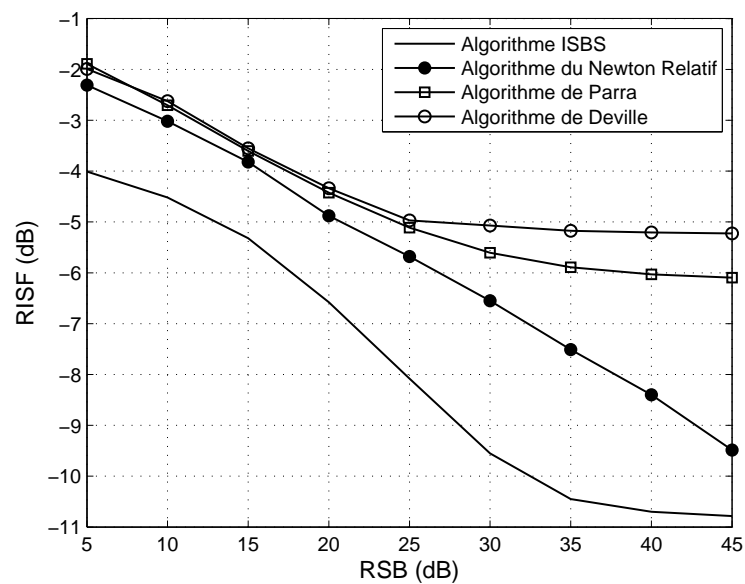


Fig. 7.6 – RISF en fonction du RSB pour  $N = 2$  sources audio et  $M = 2$  capteurs avec une taille d'observation  $T = 20000$  dans le cas du mélange convolutif : comparaison des performances de séparation entre notre algorithme ISBS, l'algorithme de Y. Deville ; l'algorithme de L. Parra et l'algorithme du Newton relatif.



# La SAS utilisant les statistiques d'ordre deux

---

## 8.1 Introduction

La séparation aveugle de sources exploitant les statistiques d'ordres deux a été largement étudiée dans la littérature. Les méthodes proposées [31, 33, 54, 55] utilisent des informations a priori supplémentaires. Dans ces algorithmes, la seule information disponible dans les observations est interprétée soit par le calcul des matrices de corrélation, soit par des matrices de covariance ou soit par les matrices spectrales.

Dans ce chapitre, Nous présenterons une méthode de séparation basée sur la décorrélation de sources stationnaires, ainsi qu'une analyse théorique des performances asymptotiques de la méthode proposée.

Nous supposons que les observations sont régies par le modèle du mélange instantané (1.1), où la matrice de mélange  $\mathbf{A}$  est supposée de rang plein ( $M \geq N$ ). Les signaux sources  $s_i(t)$ ,  $1 \leq i \leq N$  sont supposés ayant une cohérence temporelle et mutuellement décorrélés, avec une moyenne nulle et des moments du second ordre décrit par :

$$\mathbf{R}_s(\tau) \triangleq \mathbb{E} [\mathbf{s}(t + \tau)\mathbf{s}(t)^H] = \text{diag}(\rho_1(\tau), \dots, \rho_N(\tau)) \quad , \quad (8.1)$$

où  $\rho_i(\tau) \triangleq \mathbb{E}[s_i(t + \tau)s_i^*(t)]$ . Le bruit additif  $\mathbf{w}(t)$  est modélisé comme un process aléatoire complexe blanc stationnaire avec une moyenne nulle. Dans ce cas,

la séparation des sources peut être effectuée par la décorrélation des signaux à différents retards temporels.

## 8.2 Algorithme de séparation utilisant le gradient relatif

Dans cette section nous allons présenter un algorithme de SAS nommé IDA pour *Iterative Decorrelation Algorithm*, destiné aux sources possédant une cohérence temporelle. Cet algorithme est basé sur la minimisation d'une fonction de contraste basée à son tour sur les statistiques d'ordre deux en utilisant la technique du gradient relatif.

### 8.2.1 Fonction de contraste basée sur les statistiques d'ordre deux

En se basant sur le théorème A.2 (voir annexe A), on pourra définir plusieurs fonctions de contraste pour la décorrélation des signaux. Soit  $\tau_1, \tau_2, \dots, \tau_K$  avec  $K \geq 1$  les retards temporels et  $\mathbf{z}(t) = [z_1(t), \dots, z_N(t)]^T$  un vecteur de dimension  $N \times 1$  donné par  $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$  et définissons  $r_{ij}(\tau_k) \triangleq \mathbb{E}[z_i(t + \tau_k)z_j(t)^*]$ . Dans [144] le critère suivant<sup>1</sup> est utilisé :

$$G(\mathbf{z}) = \sum_{k=1}^K \log |\text{diag}(\mathbf{R}_z(\tau_k))| - \log |\mathbf{R}_z(\tau_k)| \quad (8.2)$$

où  $\mathbf{R}_z(\tau_k) = \mathbb{E}[\mathbf{z}(t + \tau_k)\mathbf{z}(t)^H]$ . Un autre critère utilisé dans [145] est :

$$G(\mathbf{z}) = \sum_{k=1}^K \sum_{i,j}^N \left[ |r_{ij}(\tau_k) + r_{ji}(\tau_k)|^2 + |r_{ij}(\tau_k) - r_{ji}(\tau_k)|^2 \right] + \sum_{i=1}^N \left| \sum_{k=1}^K |r_{ii}(\tau_k)| - 1 \right|^2 \quad (8.3)$$

Les équations (8.2) et (8.3) sont des fonctions non-négatives qui seront nulles si et seulement si la matrice  $\mathbf{R}_z(\tau_k) = \mathbb{E}[\mathbf{z}(t + \tau_k)\mathbf{z}(t)^H]$  est diagonale pour  $k = 1, \dots, K$  ou de manière équivalente si la condition (A.2) (voir annexe A) est vérifiée.

### 8.2.2 Algorithme de décorrélation utilisant le gradient relatif

Le critère de séparation que nous avons présenté prend la forme suivante :

$$\mathbf{B} \text{ est une matrice de séparation} \iff G(\mathbf{z}) = 0, \quad (8.4)$$

où  $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$  et  $G$  est la fonction de contraste décrite par l'équation (8.3). L'approche que nous avons choisi pour résoudre (8.4) est inspirée de [53]. C'est

<sup>1</sup>Dans cet article, seulement le cas où  $\tau_1 = 0$  est considéré.

une technique en bloc basée sur le traitement des  $T$  échantillons observés et qui consiste à trouver où le critère de l'équation (8.4) s'annule. La solution est obtenue itérativement sous la forme :

$$\mathbf{B}^{(p+1)} = (\mathbf{I}_N + \boldsymbol{\epsilon}^{(p)})\mathbf{B}^{(p)} \quad (8.5)$$

$$\mathbf{z}^{(p+1)}(t) = (\mathbf{I}_N + \boldsymbol{\epsilon}^{(p)})\mathbf{z}^{(p)}(t) \quad (8.6)$$

à l'itération  $p$ , la matrice  $\boldsymbol{\epsilon}^{(p)}$  est déterminée à partir de la linéarisation de  $G(\mathbf{B}\mathbf{x})$ . C'est une approximation de la technique du Newton avec l'avantage que  $\boldsymbol{\epsilon}^{(p)}$  peut être calculée d'une manière très simple (sans inversion de Hessien) sous l'hypothèse additionnelle que  $\mathbf{B}^{(p)}$  est proche d'une matrice de séparation. Cette procédure est illustrée dans ce qui suit :

Tous d'abord, on considère le cas sans bruit et éventuellement le cas du retard non-nul ( $\tau_i \neq 0$  pour  $i = 1, \dots, K$ ). En utilisant l'équation (8.6), on aura :

$$r_{ij}^{(p+1)}(\tau_k) = r_{ij}^{(p)}(\tau_k) + \sum_{q=1}^N \epsilon_{jq}^{*(p)} r_{iq}^{(p)}(\tau_k) + \sum_{l=1}^N \epsilon_{il}^{(p)} r_{lj}^{(p)}(\tau_k) + \sum_{l,q=1}^N \epsilon_{il}^{(p)} \epsilon_{jq}^{*(p)} r_{lq}^{(p)}(\tau_k), \quad (8.7)$$

où

$$r_{ij}^{(p)}(\tau_k) \triangleq \mathbb{E}[z_i^{(p)}(t + \tau_k) z_j^{(p)}(t)^*] \quad (8.8)$$

$$\approx \frac{1}{T - \tau_k} \sum_{t=1}^{T-\tau_k} z_i^{(p)}(t + \tau_k) z_j^{(p)}(t)^* . \quad (8.9)$$

Sous l'hypothèse que  $\mathbf{B}^{(p)}$  est proche d'une matrice de séparation, il s'en suit que

$$|\epsilon_{ij}^{(p)}| \ll 1$$

et

$$|r_{ij}^{(p)}(\tau_k)| \ll 1 \text{ pour } i \neq j$$

et ainsi, l'approximation au premier ordre de  $r_{ij}^{(p+1)}(\tau_k)$  est donnée par :

$$r_{ij}^{(p+1)}(\tau_k) \approx r_{ij}^{(p)}(\tau_k) + \epsilon_{ji}^{*(p)} r_{ii}^{(p)}(\tau_k) + \epsilon_{ij}^{(p)} r_{jj}^{(p)}(\tau_k), \quad (8.10)$$

de manière similaire, on a :

$$r_{ji}^{(p+1)}(\tau_k) \approx r_{ji}^{(p)}(\tau_k) + \epsilon_{ij}^{*(p)} r_{jj}^{(p)}(\tau_k) + \epsilon_{ji}^{(p)} r_{ii}^{(p)}(\tau_k). \quad (8.11)$$

A partir des équations (8.10) et (8.11), on peut écrire :

$$\begin{aligned} r_{ij}^{(p+1)}(\tau_k) + r_{ji}^{(p+1)}(\tau_k) &\approx 2r_{jj}^{(p)}(\tau_k)\Re(\epsilon_{ij}^{(p)}) + 2r_{ii}^{(p)}(\tau_k)\Re(\epsilon_{ji}^{(p)}) + (r_{ij}^{(p)}(\tau_k) + r_{ji}^{(p)}(\tau_k)) \\ r_{ij}^{(p+1)}(\tau_k) - r_{ji}^{(p+1)}(\tau_k) &\approx 2jr_{jj}^{(p)}(\tau_k)\Im(\epsilon_{ij}^{(p)}) - 2jr_{ii}^{(p)}(\tau_k)\Im(\epsilon_{ji}^{(p)}) + (r_{ij}^{(p)}(\tau_k) - r_{ji}^{(p)}(\tau_k)) \end{aligned}$$

où  $j = \sqrt{-1}$ . En remplaçant les équations précédentes dans l'équation (8.3), on obtient le problème de minimisation au sens des moindres carrés (MC) suivant :

$$\min \left\| \begin{bmatrix} \mathbf{r}_{jj}^{(p)} & \mathbf{r}_{ii}^{(p)} \end{bmatrix} \mathbf{E}_{ij}^{(p)} + \begin{bmatrix} \frac{1}{2}(\mathbf{r}_{ij}^{(p)} + \mathbf{r}_{ji}^{(p)}) & \frac{1}{2j}(\mathbf{r}_{ij}^{(p)} - \mathbf{r}_{ji}^{(p)}) \end{bmatrix} \right\| ,$$

où

$$\mathbf{E}_{ij}^{(p)} \triangleq \begin{bmatrix} \Re e(\epsilon_{ij}^{(p)}) & \Im m(\epsilon_{ij}^{(p)}) \\ \Re e(\epsilon_{ji}^{(p)}) & -\Im m(\epsilon_{ji}^{(p)}) \end{bmatrix} , \quad (8.12)$$

$$\mathbf{r}_{ij}^{(p)} = [r_{ij}^{(p)}(\tau_1), \dots, r_{ij}^{(p)}(\tau_K)]^T . \quad (8.13)$$

La solution de ce problème de minimisation au sens des moindres carrés est donnée par :

$$\mathbf{E}_{ij}^{(p)} = - \begin{bmatrix} \mathbf{r}_{jj}^{(p)} & \mathbf{r}_{ii}^{(p)} \end{bmatrix}^\# \begin{bmatrix} \frac{1}{2}(\mathbf{r}_{ij}^{(p)} + \mathbf{r}_{ji}^{(p)}) & \frac{1}{2j}(\mathbf{r}_{ij}^{(p)} - \mathbf{r}_{ji}^{(p)}) \end{bmatrix} . \quad (8.14)$$

Les équations (8.12) à (8.14) fournissent une expression explicite de  $\epsilon_{ij}^{(p)}$  pour  $i \neq j$ . Dans le cas où  $i = j$ , la minimisation de l'équation (8.3) en utilisant l'approximation au premier ordre conduit à :

$$\left| \sum_{k=1}^K r_{ii}^{(p)}(\tau_k) \left( 1 + 2\Re e(\epsilon_{ii}^{(p)}) \right) \right| - 1 = 0 . \quad (8.15)$$

Sans perte de généralité, on tire avantage de l'indétermination inhérente d'échelle de la SAS pour supposer que  $\epsilon_{ii}$  est réel et donc  $\Re e(\epsilon_{ii}) = \epsilon_{ii}$ . Par conséquent, on obtient :

$$\epsilon_{ii}^{(p)} = \frac{1 - \sum_{k=1}^K |r_{ii}^{(p)}(\tau_k)|}{2 \sum_{k=1}^K |r_{ii}^{(p)}(\tau_k)|} . \quad (8.16)$$

Dans le cas d'un signal réel, la minimisation au sens des moindres carrés devient :

$$\min \left\| \mathbf{H}_{ij}^{(p)} \mathbf{e}_{ij}^{(p)} + \boldsymbol{\psi}_{ij}^{(p)} \right\| ,$$

où

$$\mathbf{H}_{ij}^{(p)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{r}_{jj}^{(p)} & \mathbf{r}_{ii}^{(p)} \end{bmatrix} , \quad (8.17)$$

$$\mathbf{e}_{ij}^{(p)} = [\epsilon_{ij}^{(p)}, \epsilon_{ji}^{(p)}]^T , \quad (8.18)$$

$$\boldsymbol{\psi}_{ij}^{(p)} = \begin{bmatrix} \mathbf{r}_{ij}^{(p)} \\ \mathbf{r}_{ji}^{(p)} \end{bmatrix} , \quad (8.19)$$



et  $\otimes$  dénote le produit de Kronecker. La solution de ce problème de minimisation est donnée par :

$$e_{ij}^{(p)} = -\mathbf{H}_{ij}^{(p)\#} \psi_{ij}^{(p)}. \quad (8.20)$$

**Remarque :** L' avantage principal de l' algorithme précédent est sa flexibilité et sa facilité d' implémentation dans le cas adaptatif. En outre, cet algorithme peut être modifié afin de traiter le cas du mélange convolutif comme montré dans [16].

### 8.2.3 Implémentation adaptative

Afin de déduire une version adaptative de l' algorithme présenté dans le section précédente, on remplace dans les formules précédentes l' indice des itérations  $p$  par l' indice de temps  $t$  et on estime adaptativement les coefficients de corrélation  $r_{ij}^{(t)}(\tau_k)$ . L' algorithme adaptatif peut être décrit comme suit : à l' instant  $(t + 1)$

- mettre à jour les matrices de corrélation  $\mathbf{R}_z(\tau_k)$ ,  $k = 1, \dots, K$ , en utilisant la technique de moyennage suivante :

$$\begin{aligned} \mathbf{z}(t + 1) &= \mathbf{B}^{(t)} \mathbf{x}(t + 1) \\ \mathbf{R}_z^{(t+1)}(\tau_k) &= (1 - \lambda_{t+1}) \mathbf{R}_z^{(t)}(\tau_k) + \lambda_{t+1} \mathbf{z}(t + 1) \mathbf{z}(t + 1 - \tau_k)^H \end{aligned}$$

où  $\lambda_t$  est une séquence décroissante et positive. Notons que  $r_{ij}^{(t)}(\tau_k)$  est l' élément  $ij$  de la matrice  $\mathbf{R}_z^{(t)}(\tau_k)$ ,

- estimer  $\epsilon^{(t+1)}$  en utilisant les équations (8.14) et (8.16) et la mise à jour des coefficients de corrélation  $r_{ij}^{(t+1)}(\tau_k)$ ,
- mettre à jour la matrice de séparation, les matrices de corrélation  $\mathbf{R}_z(\tau_k)$ ,  $k = 1, \dots, K$ , et les sources estimées  $\mathbf{z}(t + 2 - \tau_k)$ ,  $k = 1, \dots, K$  :

$$\begin{aligned} \mathbf{B}^{(t+1)} &= (\mathbf{I}_N + \epsilon^{(t+1)}) \mathbf{B}^{(t)} \\ \mathbf{R}_z^{(t+1)}(\tau_k) &= (\mathbf{I}_N + \epsilon^{(t+1)}) \mathbf{R}_z^{(t+1)}(\tau_k) (\mathbf{I}_N + \epsilon^{(t+1)})^H \\ \mathbf{z}(t + 2 - \tau_k) &= (\mathbf{I}_N + \epsilon^{(t+1)}) \mathbf{z}(t + 2 - \tau_k) \end{aligned}$$

## 8.3 Analyse de performances

### 8.3.1 Analyse théorique des performances

Dans cette section, nous présenterons une analyse de performances asymptotiques (pour un grand nombre d' échantillons) de la méthode de SAS présentée dans la section précédente. On considère le cas d' un mélange instantané avec des sources

complexes i.i.d satisfaisant en plus la condition d'identifiabilité  $\sum_{k \in \mathbb{Z}} |\rho_i(\tau_k)| < +\infty$  pour  $i = 1, \dots, N$ . Le bruit est supposé Gaussien avec une variance  $\sigma^2 \mathbf{I}_M$  et nous supposons que l'indétermination de permutation est  $\mathbf{P} = \mathbf{I}_N$ . Dans ce cas on peut écrire que :

$$\mathbf{BA} = \mathbf{I}_N + \boldsymbol{\delta}, \quad (8.21)$$

donc, la qualité de la séparation est mesurée en utilisant le rapport interférences à signal (RIS) [31] défini comme :

$$\text{RIS} \triangleq \sum_{p \neq q} \frac{\mathbb{E} [ |(BA)_{pq}|^2 ] \rho_q(0)}{\mathbb{E} [ |(BA)_{pp}|^2 ] \rho_p(0)} \quad (8.22)$$

$$= \sum_{p \neq q} \mathbb{E} [ |\delta_{pq}|^2 ] \frac{\rho_q(0)}{\rho_p(0)}. \quad (8.23)$$

Notre analyse de performances est dérivée de l'expression de la variance asymptotique :

$$\lim_{T \rightarrow +\infty} T E \left( |\delta_{pq}|^2 \right) \quad (8.24)$$

En utilisant une approche similaire à [53] basée sur les théorèmes centrale-limite et de continuité, on obtient le résultat suivant :

**Théorème 8.1** *Le vecteur  $\boldsymbol{\delta}_{ij} \triangleq [\Re(\delta_{ij}) \Re(\delta_{ji}) \Im(\delta_{ij}) - \Im(\delta_{ji})]^T$  suit asymptotiquement une distribution Gaussienne avec une matrice de covariance asymptotique*

$$\boldsymbol{\Delta}_{ij} \triangleq \lim_{T \rightarrow +\infty} T E \left( \boldsymbol{\delta}_{ij} \boldsymbol{\delta}_{ij}^T \right) \quad (8.25)$$

$$= \frac{1}{4} \boldsymbol{\mathcal{H}}_{ij}^{\#} \boldsymbol{\Psi}_{ij} \boldsymbol{\mathcal{H}}_{ij}^{\#\star} \quad (8.26)$$

où

$$\boldsymbol{\mathcal{H}}_{ij} = \mathbf{I}_2 \otimes \begin{bmatrix} \boldsymbol{\rho}_i^T & \boldsymbol{\rho}_j^T \end{bmatrix} \quad (8.27)$$

$$\boldsymbol{\Psi}_{ij} = \begin{bmatrix} \Xi_{11}^{(ij)} & \Xi_{12}^{(ij)} \\ \Xi_{21}^{(ij)} & \Xi_{22}^{(ij)} \end{bmatrix} \quad (8.28)$$

avec  $\boldsymbol{\rho}_i = [\rho_i(\tau_1), \rho_i(\tau_2), \dots, \rho_i(\tau_K)]$  et

$$\begin{aligned} \Xi_{11}^{(ij)} &= \Phi_{11}^{(ij)} + \Phi_{12}^{(ij)} + \Phi_{21}^{(ij)} + \Phi_{22}^{(ij)} \\ \Xi_{12}^{(ij)} &= j \left[ \Phi_{11}^{(ij)} - \Phi_{12}^{(ij)} + \Phi_{21}^{(ij)} + \Phi_{22}^{(ij)} \right] \\ \Xi_{21}^{(ij)} &= -j \left[ \Phi_{11}^{(ij)} + \Phi_{12}^{(ij)} - \Phi_{21}^{(ij)} + \Phi_{22}^{(ij)} \right] \\ \Xi_{22}^{(ij)} &= \Phi_{11}^{(ij)} - \Phi_{12}^{(ij)} - \Phi_{21}^{(ij)} + \Phi_{22}^{(ij)} \end{aligned}$$

et

$$\begin{aligned}\Phi_{11}^{(ij)}(k, k') &= \sum_{\tau \in \mathbb{Z}} r_{ii}(k + \tau) r_{jj}(k' + \tau) \\ \Phi_{22}^{(ij)}(k, k') &= \sum_{\tau \in \mathbb{Z}} r_{ii}(k' + \tau) r_{jj}(k + \tau) \\ \Phi_{12}^{(ij)}(k, k') &= \sum_{\tau \in \mathbb{Z}} r_{ii}(k + \tau) r_{jj}(k' - \tau) \\ r_{ii}(k) &= \rho_i(k) + \delta(k) \sigma^2 \mathbf{b}_i \mathbf{b}_i^T\end{aligned}$$

et  $\Phi_{21}^{(ij)} = \Phi_{12}^{(ij)T}$  et  $\mathbf{b}_i$  représente la  $i^e$  ligne de  $\mathbf{B} = \mathbf{A}^\#$ .

Dans le cas de signaux réels, le résultat précédent devient :

$$\Delta_{ij} = \mathcal{H}_{ij}^\# \Psi_{ij} \mathcal{H}_{ij}^{\#T} \quad (8.29)$$

où

$$\mathcal{H}_{ij} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes [\boldsymbol{\rho}_i^T \ \boldsymbol{\rho}_j^T] \quad (8.30)$$

$$\Psi_{ij} = \begin{bmatrix} \Phi_{11}^{(ij)} & \Phi_{12}^{(ij)} \\ \Phi_{21}^{(ij)} & \Phi_{22}^{(ij)} \end{bmatrix} \quad (8.31)$$

Ce résultat peut être réécrit en utilisant l'hypothèse de normalisation présentée dans le théorème A.1 de l'annexe A ( $\|\boldsymbol{\rho}_i\| = 1 \ \forall i$ ) comme :

$$\Delta_{ij} = \frac{1}{(\rho_{ij}^2 - 1)^2} \begin{bmatrix} 1 & -\rho_{ij} \\ -\rho_{ij} & 1 \end{bmatrix} \mathcal{D} \begin{bmatrix} 1 & -\rho_{ij} \\ -\rho_{ij} & 1 \end{bmatrix} \quad (8.32)$$

où

$$\rho_{ij} = \boldsymbol{\rho}_i \boldsymbol{\rho}_j^T \quad (8.33)$$

et

$$\mathcal{D} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma_i(f) \Gamma_j(f) \boldsymbol{\nu}_{ij} \boldsymbol{\nu}_{ij}^T df \quad (8.34)$$

où

$$\boldsymbol{\nu}_{ij} = [\Re e(\varrho_i(f)) \ \Re e(\varrho_j(f))]^T \quad (8.35)$$

avec

$$\varrho_l(f) = \sum_{k=1}^K \rho_l(\tau_k) \exp(-j2\pi\tau_k f) \quad \text{pour } l \in \{i, j\} \quad (8.36)$$

et

$$\Gamma_l(f) = \Gamma_{s_l}(f) + \sigma^2 \|\mathbf{b}_l\|^2, \quad \text{pour } l \in \{i, j\} \quad (8.37)$$

où  $\Gamma_{s_l}(f)$  est la fonction de densité spectrale de puissance de la  $l^e$  source. En remplaçant (8.37) dans (8.34), on pourra réécrire :

$$\begin{aligned} \mathcal{D} = & \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma_{s_i}(f) \Gamma_{s_j}(f) \mathbf{v}_{ij} \mathbf{v}_{ij}^T df + \sigma^4 \|\mathbf{b}_i\|^2 \|\mathbf{b}_j\|^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{v}_{ij} \mathbf{v}_{ij}^T df \\ & + \sigma^2 \|\mathbf{b}_i\|^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma_{s_j}(f) \mathbf{v}_{ij} \mathbf{v}_{ij}^T df + \sigma^2 \|\mathbf{b}_j\|^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma_{s_i}(f) \mathbf{v}_{ij} \mathbf{v}_{ij}^T df \end{aligned}$$

### 8.3.2 Validations et simulations

Dans cette partie, nous présenterons quelques simulations pour l'évaluation des performances de notre algorithme de séparation IDA. On considérera dans nos simulations une antenne composée de  $M = 4$  capteurs recevant  $N = 2$  signaux sources, en présence d'un bruit blanc stationnaire. Les deux signaux sources suivent un modèle auto-régressif d'ordre un, avec les coefficients  $a_1 = 0.95e^{j0.5}$  et  $a_2 = 0.5e^{j0.7}$ . On utilisera un nombre de retards  $K = 5$  et on estimera le RIS pour  $N_r = 100$  réalisations aléatoires.

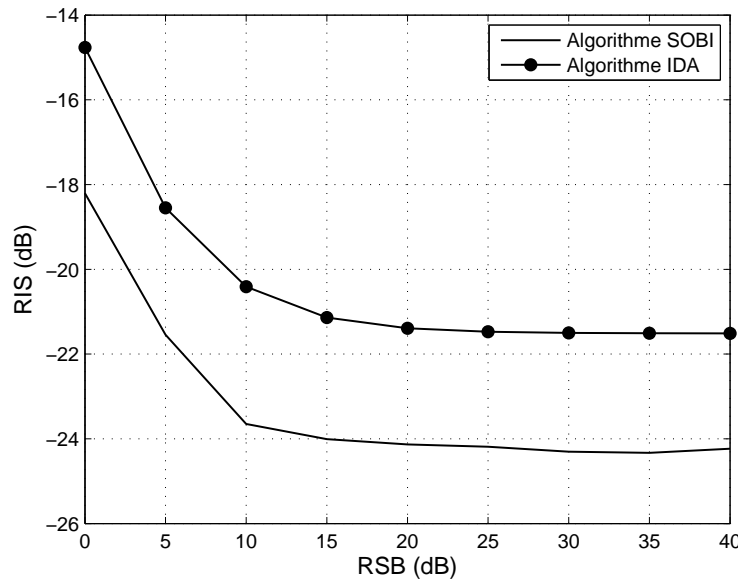


Fig. 8.1 – RIS en fonction du RSB pour  $N = 2$  sources auto-régressives et  $M = 4$  capteurs avec une taille d'observation  $T = 1000$  : comparaison des performances entre notre algorithme de décorrélation IDA et l'algorithme SOBI.

Dans la figure 8.1, on compare les performances de séparation obtenues par notre algorithme de décorrélation IDA, et celles de l'algorithme SOBI en fonction du RSB. On observe en général que l'algorithme SOBI fournit de meilleurs résultats

que notre algorithme IDA, mais cela reste raisonnable si on prend en considération le faible coût de calcul et la simplicité de l'algorithme IDA ainsi que l'absence d'étape de blanchiment contrairement à l'algorithme SOBI.

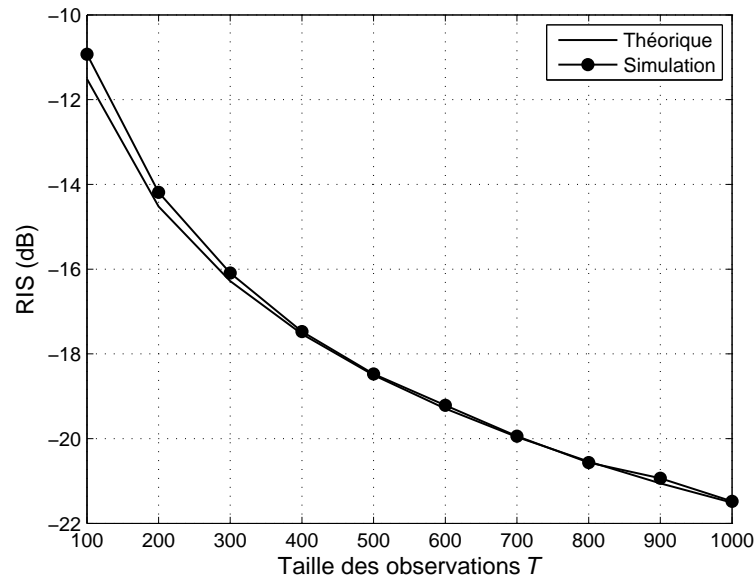


Fig. 8.2 – RIS en fonction de la taille des observations pour  $N = 2$  sources auto-régressives et  $M = 4$  capteurs avec un RSB = 40dB : comparaison des performances empiriques et les performances théoriques de notre algorithme IDA.

Dans la simulation de la figure 8.2, le RIS est tracé en fonction de la taille des observations pour un RSB de 40dB. Cette figure montre que l'expression asymptotique du RIS est pertinente dès que la taille des observations est supérieure à 100 échantillons. Dans les courbes précédentes, les termes  $\mathbb{E}[|\delta_{ij}|^2]$  et  $\mathbb{E}[|\delta_{ji}|^2]$  intervenants dans le calcul du RIS, sont remplacés par  $\frac{\Delta_{ij(1,1)} + \Delta_{ij(3,3)}}{T}$  et  $\frac{\Delta_{ij(2,2)} + \Delta_{ij(4,4)}}{T}$  respectivement. Cela signifie que les conditions asymptotiques sont atteintes pour des tailles d'observations relativement faibles.

La figure 8.3 représente la variation du RIS en fonction du RSB. On compare les performances empiriques avec les performances théoriques pour une taille d'observations  $T = 1000$ . On observe que les deux courbes sont confondues, ce qui implique le bien-fondé de l'expression théorique de l'erreur d'estimation dans les conditions asymptotiques.

La figure 8.4 représente la variation du RIS en fonction du nombre de capteurs  $M$  en utilisant l'expression théorique pour une taille d'observations  $T = 1000$ . On observe que plus le nombre de capteurs est élevé, plus le niveau du RIS est

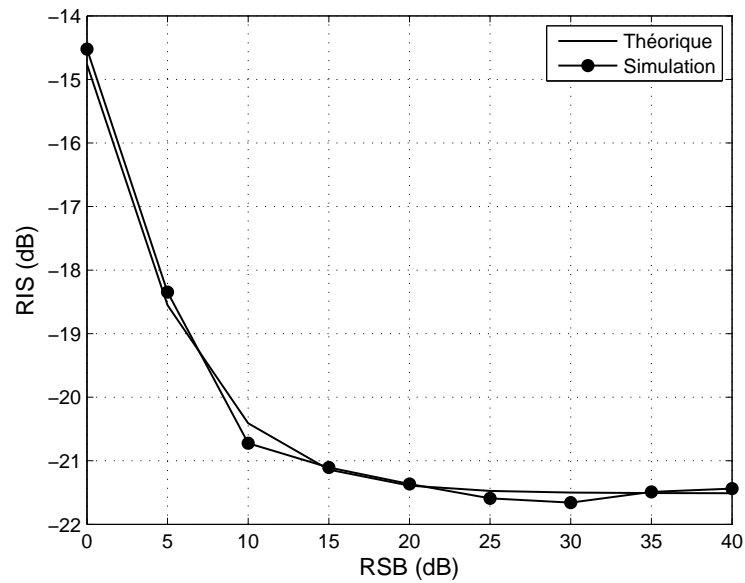


Fig. 8.3 – RIS en fonction du RSB pour  $N = 2$  sources auto-régressives et  $M = 4$  capteurs avec une taille d'observation  $T = 1000$  : comparaison des performances empiriques et des performances théoriques de notre algorithme IDA.

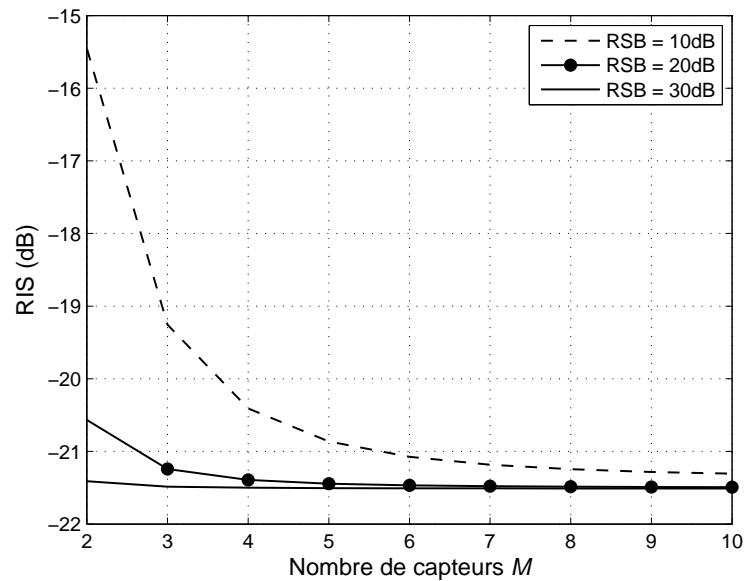


Fig. 8.4 – RIS en fonction du nombre de capteurs  $M$  pour  $N = 2$  sources auto-régressives et une taille d'observation  $T = 1000$  : représentation des performances théoriques de l'algorithme IDA pour différentes valeurs du RSB.

bas pour des valeurs du RSB faibles. Pour des valeurs élevées du RSB, l'effet du nombre de capteurs est négligeable sur les performances de la séparation.

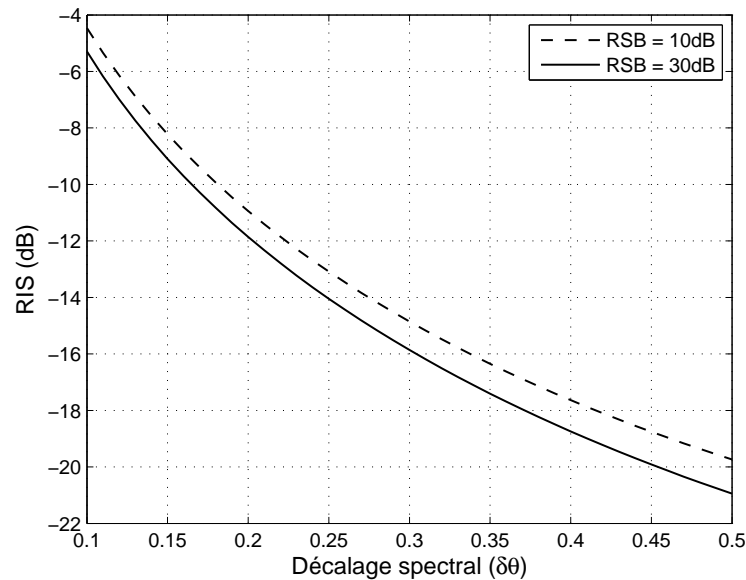


Fig. 8.5 – RIS en fonction du décalage spectral  $\delta\theta$  pour  $N = 2$  sources auto-régressives et  $M = 4$  capteurs avec une taille d'observation  $T = 1000$  : représentation des performances théoriques de l'algorithme IDA.

La figure 8.5 représente la variation du RIS en fonction du décalage spectral  $\delta\theta$  pour une taille d'observations  $T = 1000$ . Le décalage spectral  $\delta\theta$  représente le recouvrement spectral des deux sources. En d'autres termes, les coefficients des deux processus auto-régressifs sont  $a_1 = 0.5e^{j0.5}$  et  $a_2 = 0.5e^{j(0.5+\delta\theta)}$ . Dans cette figure, le bruit est supposé spatialement blanc et son niveau est gardé constant à 10dB et 30dB. Les courbes mettent en évidence le fait que les performances de séparation s'améliorent avec l'augmentation du décalage spectral  $\delta\theta$ .

Dans la simulation de la figure 8.6, on représente la variation du RIS en fonction du nombre de retards  $K$  utilisés dans l'algorithme de décorrélation.

## 8.4 Conclusion

Ce chapitre a introduit une méthode de séparation basée sur la décorrélation de sources stationnaires. Nous avons introduit par la même occasion une fonction de contraste basée sur les statistiques d'ordre deux minimisée à l'aide d'une technique du gradient relatif. Une analyse théorique des performances asymptotiques

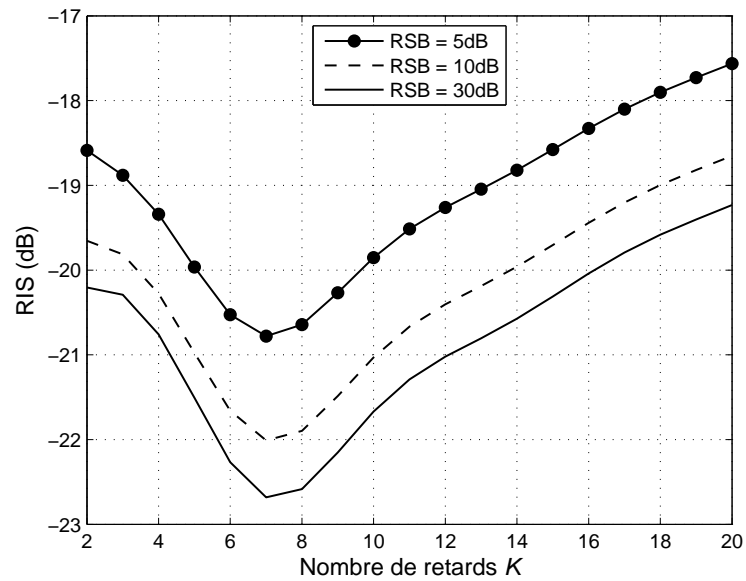


Fig. 8.6 – RIS en fonction du nombre de retards  $K$  pour  $N = 2$  sources auto-régressives et  $M = 4$  capteurs avec une taille d'observation  $T = 1000$  : représentation des performances théoriques de l'algorithme IDA.

de la méthode proposée a été introduite, et validée à travers différents scénarios de simulations.



# Conclusion générale

---

Cette thèse a abordé le problème de la séparation aveugle de sources audio. Plusieurs méthodes de séparation de sources ont été développées et étudiées en fonction de la configuration du mélange.

En premier lieu, nous avons présenté une nouvelle méthode de séparation dans le cas du mélange instantané, utilisant des techniques de décomposition modale. L'avantage majeur de la méthode proposée est sa capacité à séparer plus de sources que de capteurs. La méthode proposée procède en deux étapes : une étape d'analyse où toutes les composantes modales sont estimées, suivie par l'étape de synthèse pour regrouper (classifier) ensemble les composantes modales et reconstruire les signaux sources. Pour l'étape d'analyse, deux techniques de décomposition sont utilisées et comparées. Une version modifiée de l'algorithme ainsi qu'une approche basée sur la projection en sous-espace sont proposées, afin de relâcher l'hypothèse de quasi-orthogonalité et permettre aux signaux sources de partager les composantes modales communes respectivement. Ces approches permettent d'améliorer la qualité de la séparation et de diminuer le coût de calcul de l'algorithme initial. Nous avons étendu cet algorithme au cas du mélange convolutif sous-déterminé et nous avons introduit aussi une nouvelle technique d'identification aveugle du canal de mélange, étape nécessaire à l'algorithme de séparation. Pour le cas sur-déterminé convolutif, nous avons présenté une autre version, utilisant un critère de parcimonie approprié conjugué à une décomposition en valeurs et vecteurs propres.

En second lieu, nous avons introduit de nouvelles méthodes de séparation dans le cas du mélange instantané sous-déterminé pour des sources TF-non-disjointes et non-stationnaires en utilisant les représentations temps-fréquence. Les prin-

cipaux avantages des algorithmes de séparation proposés sont : premièrement, une hypothèse plus réaliste sur la parcimonie des sources, c'est à dire que les sources ne sont pas nécessairement disjointes dans le domaine temps-fréquence, et deuxièmement, un traitement explicite des points de recouvrement en utilisant une projection en sous-espace, ce qui a conduit à une amélioration de la qualité de séparation. Nous avons présenté aussi, une extension des ces algorithmes au cas du mélange convolutif sous-déterminé en lui combinant une étape de pré-estimation du canal de mélange.

Par la suite, nous avons présenté une méthode de séparation dédiée aux sources parcimonieuses. Cette méthode a été développée initialement pour le cas du mélange instantané et étendue par la suite au cas du mélange convolutif. Nous avons aussi introduit une fonction de contraste basée sur la norme  $\ell_p$  pour la mesure de la parcimonie ainsi qu'un algorithme itératif basé sur la technique du gradient relatif pour la minimisation de cette fonction. Les résultats des simulations mettent en évidence l'utilité de la méthode.

Pour finir, nous avons présenté une méthode de séparation basée sur la décorrélation de sources stationnaires. Nous avons introduit par la même occasion une fonction de contraste basée sur les statistiques d'ordre deux minimisée à l'aide d'une technique du gradient relatif. Une analyse théorique des performances asymptotiques de la méthode proposée a été introduite, et validée à travers différents scénarios de simulations.

## Perspectives

A l'issue de cette étude, nous avons constaté qu'il rest beaucoup de chemin à parcourir avant d'envisager une solution complète pour la séparation de sources. Plus précisément, nous envisageons les pistes suivantes pour nos éventuels travaux futurs dans la perspective de combler les lacunes des méthodes présentées précédemment :

- développer des techniques d'identification adaptées aux canaux long : ces techniques nous permettraient de résoudre le problème de la taille des canaux à estimer soulevé par la méthode CR utilisée dans les chapitres 4 et 6 et ainsi obtenir une solution plus adaptée aux canaux acoustiques,
- utiliser des techniques d'interpolation dans le domaine temps-fréquence : l'utilisation de techniques d'interpolation dans le domaine temps-fréquence pourra améliorer la qualité des signaux estimés en comblant les "vides" dus au mas-

quage temps-fréquence utilisé par les méthodes présentées dans la partie II de ce rapport,

- méthode d'estimation de la matrice de mélange dans le cas sous-déterminé : cette méthode nous offrira la possibilité de nous affranchir de quelques hypothèses contraignantes ainsi que l'utilisation de techniques de seuillage pour les méthodes temps-fréquence,
- validation des algorithmes développés dans le cadre de cette étude en utilisant des signaux issus d'enregistrements effectués dans un environnement réel.

Cette liste est bien entendu non exhaustive et nous permet d'avoir une visibilité à moyen et court terme sur nos travaux futurs.



## Identifiabilité au second ordre

---

Dans cette annexe, nous proposerons une nouvelle technique de test de la condition d'identifiabilité au second ordre pour la séparation aveugle de sources. En d'autres termes, ce test nous permettra de vérifier de manière aveugle, à partir des observations, si les sources ont été correctement séparées ou pas.

### A.1 Conditions nécessaires et suffisantes d'identifiabilité

La condition d'identifiabilité au second ordre a été proposée par Tong et al. dans [146], qui ont montré que les sources sont séparables en utilisant les statistiques d'ordre deux seulement (cas de grand nombre de matrices de corrélation) si elles ont des fonctions de densité spectrale différentes. En pratique, la SAS est effectuée sur un ensemble réduit de matrices de corrélation. Par conséquent, les résultats d'identifiabilité précédents ont été généralisés à ce cas dans [31, 147] ce qui a mené aux conditions nécessaires et suffisantes d'identifiabilité établies par le théorème suivant :

**Théorème A.1** *Soit  $\tau_1 < \tau_2 < \dots < \tau_K$  avec  $K \geq 1$  les retards temporels, et définissons  $\boldsymbol{\rho}_i = [\rho_i(\tau_1), \rho_i(\tau_2), \dots, \rho_i(\tau_K)]$  et  $\tilde{\boldsymbol{\rho}}_i = [\Re(\boldsymbol{\rho}_i), \Im(\boldsymbol{\rho}_i)]$ . En tirant profit de l'indétermination d'échelle de la SAS, on suppose sans perte de généralité que les sources sont normalisées tel que  $\|\boldsymbol{\rho}_i\| = \|\tilde{\boldsymbol{\rho}}_i\| = 1, \forall i^1$ . Alors, la SAS peut*

<sup>1</sup>On suppose ici implicitement que  $\boldsymbol{\rho}_i \neq 0$ , sinon le signal source ne sera pas détecté (et a fortiori

être effectuée en utilisant les matrices de corrélation des données observées aux retards  $\tau_1, \tau_2, \dots, \tau_K$  si et seulement si pour tous  $1 \leq i \neq j \leq N$  :

$$\tilde{\rho}_i \text{ et } \tilde{\rho}_j \text{ sont linéairement indépendants} \quad (\text{A.1})$$

**Démonstration** : La démonstration de la suffisance du théorème est donnée dans l'annexe B. Nous démontrons ici uniquement que la condition (A.1) est nécessaire à l'accomplissement de la SAS en utilisant les matrices de corrélation

$$\mathbf{R}_x(\tau_k) \triangleq \mathbb{E}[\mathbf{x}(t + \tau_k)\mathbf{x}(t)^H], \quad k = 1, \dots, K.$$

En effet, si deux signaux sources  $s_1$  et  $s_2$  ont des coefficients de corrélation tel que  $\tilde{\rho}_1 = \epsilon \tilde{\rho}_2$  où  $\epsilon = \pm 1$ , alors tout signal "virtuel" de la forme  $\tilde{\mathbf{x}}(t) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(t) + \mathbf{w}(t)$  où  $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \mathbf{a}_3, \dots, \mathbf{a}_N]$  et  $\tilde{\mathbf{s}}(t) = [\tilde{s}_1(t), \tilde{s}_2(t), s_3(t), \dots, s_N(t)]^T$  avec

$$[\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2] = [\mathbf{a}_1, \mathbf{a}_2]\mathbf{T}$$

et

$$\begin{bmatrix} \tilde{s}_1(t) \\ \tilde{s}_2(t) \end{bmatrix} = \mathbf{T}^{-1} \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix}$$

avec

$$\mathbf{T} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}, \quad \text{si } \epsilon = +1$$

$$\mathbf{T} = \begin{bmatrix} \cosh \theta & \sinh \theta \\ \sinh \theta & \cosh \theta \end{bmatrix}, \quad \text{si } \epsilon = -1$$

vérifie  $\mathbf{R}_{\tilde{\mathbf{x}}}(\tau_k) = \mathbf{R}_x(\tau_k)$  et  $\mathbf{R}_{\tilde{\mathbf{s}}}(\tau_k) = \mathbf{R}_s(\tau_k)$  pour  $k = 1, \dots, K$ .  $\square$

On remarque à partir de la condition (A.1) que la SAS peut être réalisée en utilisant une seule matrice de corrélation  $\mathbf{R}_x(\tau_k)$  à condition que les vecteurs  $[\Re(\rho_i(\tau_k)), \Im(\rho_i(\tau_k))]$  et  $[\Re(\rho_j(\tau_k)), \Im(\rho_j(\tau_k))]$  soient linéairement indépendants pour tous  $i \neq j$ .

Notons aussi, qu'à partir de la condition (A.1), la SAS peut être effectuée si au plus une des sources est temporellement blanche. Réciproquement, on rappelle qu'en utilisant les statistiques d'ordre supérieur, la SAS ne peut être effectuée que si au plus une des sources est Gaussienne.

Sous la condition du Théorème A.1, la SAS peut être réalisée par décorrélation selon le résultat suivant :

ne sera pas estimé) à partir de l'ensemble considéré de matrices de corrélation. Cette hypothèse sera maintenue dans la suite du chapitre.

**Théorème A.2** Soit  $\tau_1, \tau_2, \dots, \tau_K$  avec  $K \geq 1$  les retards temporels et  $z(t) = [z_1(t), \dots, z_N(t)]^T$  un vecteur de dimension  $N \times 1$  donné par  $z(t) = \mathbf{B}\mathbf{x}(t)$ . Définissons  $r_{ij}(\tau_k) \triangleq \mathbb{E}[z_i(t + \tau_k)z_j(t)^*]$ . Si la condition d'identifiabilité est maintenue, alors  $\mathbf{B}$  est une matrice de séparation ( $\mathbf{B}\mathbf{x}(t) = \mathbf{P}\Lambda s(t)$  pour une matrice de permutation  $\mathbf{P}$  et une matrice diagonale non-singulière  $\Lambda$  dans le cas sans bruit) si et seulement si

$$r_{ij}(\tau_k) = 0 \quad \text{et} \quad \sum_{k=1}^K |r_{ii}(\tau_k)| > 0 \quad (\text{A.2})$$

pour tous  $1 \leq i \neq j \leq N$  et  $k = 1, 2, \dots, K$ .

**Démonstration** : voir la démonstration du Théorème A.3 dans l'annexe B.

Notons que, si un des retards est nul, le résultat du Théorème A.2 ne sera maintenu que sous l'hypothèse du mélange sans bruit. Dans ce cas, on peut remplacer la condition  $\sum_{k=1}^K |r_{ii}(\tau_k)| > 0$  par  $r_{ii}(0) > 0$ , pour  $i = 1, \dots, N$ . En d'autres termes, si tous les retards sont non nuls et si le bruit est temporellement blanc (mais peut être spatialement coloré avec une matrice de covariance spatiale inconnue) alors le résultat précédent restera valable sans l'hypothèse d'avoir un mélange sans bruit.

## A.2 Identifiabilité partielle

On croit généralement que quand les conditions d'identifiabilité ne sont pas réunies, la SAS ne peut pas être effectuée. Ceci n'est pas entièrement vrai car il est possible de séparer partiellement les sources dans le sens où nous pouvons extraire ceux qui satisfont les conditions d'identifiabilité. Plus précisément, les sources peuvent être séparées en blocs dont chacun contient un mélange des sources qui ne sont pas séparables en utilisant l'ensemble considéré de statistiques. Par exemple, considérons le mélange de trois sources tel que  $\tilde{\rho}_1 = \tilde{\rho}_2$  tandis que  $\tilde{\rho}_1$  et  $\tilde{\rho}_3$  sont linéairement indépendants. Dans ce cas, la source  $s_3$  peut être extraite alors que les sources  $s_1$  et  $s_2$  ne le seront pas. En d'autres termes, en décorrélant les signaux observés aux retards considérés, on obtient trois signaux, où un d'entre eux étant  $s_3$  (à une constante près) et les deux autres sont des mélanges linéaires de  $s_1$  et  $s_2$ .

Ce résultat peut être formulé mathématiquement comme suit : supposons qu'il y ait  $d$  groupes distincts de sources où chacun d'entre eux contient  $N_i$  signaux sources avec les mêmes (à un signe près) vecteurs de corrélation  $\tilde{\rho}_i$ ,  $i = 1, \dots, d$  (il est clair que  $N = N_1 + \dots + N_d$ ). Les vecteurs de corrélation  $\tilde{\rho}_1, \dots, \tilde{\rho}_d$  sont deux

à deux linéairement indépendants. on écrit  $\mathbf{s}(t) = [\mathbf{s}_1(t)^T, \dots, \mathbf{s}_d(t)^T]^T$  où chaque sous-vecteur  $\mathbf{s}_i(t)$  contient les  $N_i$  signaux sources ayant pour vecteur de corrélation  $\tilde{\boldsymbol{\rho}}_i$ .

**Théorème A.3** Soit  $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$  un vecteur de dimension  $N \times 1$  satisfaisant l'équation (A.2) pour tout  $1 \leq i \neq j \leq N$  et  $k = 1, \dots, K$ . Alors, il existe une matrice de permutation  $\mathbf{P}$  tel que

$$\bar{\mathbf{z}}(t) \triangleq \mathbf{P}\mathbf{z}(t) = [\bar{\mathbf{z}}_1(t)^T, \dots, \bar{\mathbf{z}}_d(t)^T]^T$$

où  $\bar{\mathbf{z}}_i(t) = \mathbf{U}_i\mathbf{s}_i(t)$ ,  $\mathbf{U}_i$  est une matrice non-singulière de dimension  $N_i \times N_i$ . Par ailleurs, les sources appartenant au même groupe, c'est à dire ayant le même vecteur de corrélation (à un signe près)  $\tilde{\boldsymbol{\rho}}_i$ , ne pourront pas être séparées en utilisant uniquement les matrices de corrélation  $\mathbf{R}_x(\tau_k)$ ,  $k = 1, \dots, K$ .

Ce résultat montre que quand certaines sources ont le même vecteur de corrélation (à un signe près) alors le mieux que l'on puisse faire en utilisant les statistiques d'ordre deux, est de les séparer par blocs. Cependant, ce résultat serait inutile si nous ne pouvions pas vérifier la dépendance linéaire des vecteurs de corrélation  $\tilde{\boldsymbol{\rho}}_i$  et diviser les signaux par groupes (comme cela a été montré ci-dessus) selon leurs vecteurs de corrélation. Ceci nous conduit à l'important problème de test de la condition d'identifiabilité qui sera discuté dans ce qui suit.

## A.3 Test d'identifiabilité

### A.3.1 Résultat théorique

La condition nécessaire et suffisante d'identifiabilité (A.1) dépend des coefficients de corrélation des signaux sources. Par conséquent, il est impossible a priori de vérifier si les sources sont "séparables" ou pas à partir de l'ensemble des matrices de corrélation des signaux observés. Cependant, il est possible de vérifier a posteriori si les sources ont été "séparées" ou pas. Nous avons le résultat suivant :

**Théorème A.4** Soit  $\tau_1 < \tau_2 < \dots < \tau_K$  avec  $K \geq 1$  les retards temporels et  $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$ . On suppose que  $\mathbf{B}$  est une matrice tel que  $\mathbf{z}(t)$  satisfait<sup>2</sup> l'équation (A.2)

<sup>2</sup>En raison de l'indétermination inhérente du problème de SAS, on suppose sans perte de généralité que les signaux originaux et estimés ont la même pondération, c'est à dire  $\|\tilde{\boldsymbol{\rho}}_i\| = 1$ .



pour tous  $1 \leq i \neq j \leq N$  et  $k = 1, \dots, K$ . Alors il existe une matrice de permutation  $P$  tel que pour  $k = 1, \dots, K$ .

$$\mathbb{E}[\mathbf{z}(t + \tau_k)\mathbf{z}(t)^H] = \mathbf{P}\mathbf{R}_s(\tau_k)\mathbf{P}^T$$

En d'autres termes, les entrées de  $\bar{\mathbf{z}}(t) \triangleq \mathbf{P}\mathbf{z}(t)$  ont les mêmes coefficients de corrélation que ceux de  $s(t)$  aux retards  $\tau_1, \dots, \tau_K$ , c'est à dire que  $\mathbb{E}[\bar{z}_i(t + \tau_k)\bar{z}_i(t)^*] = \rho_i(\tau_k)$  pour  $k = 1, \dots, K$  et  $i = 1, \dots, N$ .

**Démonstration** : voir la démonstration dans l'annexe B.

En se servant du théorème A.4, l'existence de la condition (A.1) peut être vérifiée en utilisant une approximation des coefficients de corrélation  $r_{ii}(\tau_k) \triangleq \mathbb{E}[z_i(t + \tau_k)z_i(t)^*]$ . Il est important de préciser que même si l'équation (A.2) est valide, cela ne signifie pas que les signaux sources ont été séparés. Trois situations peuvent se présenter :

1. pour toutes les paires  $(i, j)$ , les vecteurs de corrélation  $\tilde{\rho}_i$  et  $\tilde{\rho}_j$  calculés à partir de  $\mathbf{z}(t)$ , sont linéairement indépendants ; alors on est sûr que les sources ont été séparées et que le signal  $\mathbf{z}(t) = \mathbf{s}(t)$  à une indétermination près. En fait, le test de la condition d'identifiabilité est équivalent au test d'indépendance par mesure des angles entre les vecteurs  $\tilde{\rho}_i$  et  $\tilde{\rho}_j$  pour tout  $1 \leq i \neq j \leq N$ . Plus l'angle entre  $\tilde{\rho}_i$  et  $\tilde{\rho}_j$  est grand, meilleure est la qualité de séparation (voir l'analyse de performance faite par Belouchrani et al. dans [31]),
2. pour toutes les paires  $(i, j)$ , les vecteurs de corrélation  $\tilde{\rho}_i$  et  $\tilde{\rho}_j$  sont linéairement dépendants. Ainsi les sources n'ont pas été séparées et  $\mathbf{z}(t)$  est toujours une combinaison linéaire des signaux  $\mathbf{s}(t)$ ,
3. quelques paires  $(i, j)$  parmi l'ensemble de toutes les paires, sont telles que les vecteurs  $\tilde{\rho}_i$  et  $\tilde{\rho}_j$  sont linéairement dépendants. Alors les sources seront séparables par blocs.

Maintenant, ayant seulement une seule réalisation du signal sous la main, nous proposons d'utiliser une technique de rééchantillonnage pour évaluer la statistique nécessaire pour la mise en oeuvre de la procédure de test.

### A.3.2 Test d'identifiabilité utilisant des techniques de rééchantillonnage

Notons, qu'en pratique les coefficients de corrélation des signaux sources sont calculés à partir d'un ensemble de données finies et bruitées. En raison des effets

Tab. A.1 – Test d’identifiabilité utilisant les techniques de rééchantillonnage

1. Estimer la matrice de séparation  $\mathbf{B}$  et  $\mathbf{z}(t) \triangleq \mathbf{B}\mathbf{x}(t)$  en utilisant un algorithme existant de SAS exploitant les statistiques d’ordre deux (par exemple SOBI [31]).
2. Pour chaque composante  $z_i(t)$ , estimer le vecteur normalisé correspondant  $\tilde{\boldsymbol{\rho}}_i$ .
3. Calculer le produit scalaire  $\hat{\mathcal{E}}_{ij} = |\tilde{\boldsymbol{\rho}}_i \tilde{\boldsymbol{\rho}}_j^T|$  pour chaque paire  $(i, j)$  pour tous  $1 \leq i \neq j \leq N$ .
4. Estimer  $\hat{\sigma}_{(i,j)}$  la déviation standard de  $\hat{\mathcal{E}}_{ij}$  en utilisant une technique de rééchantillonnage.
5. Choisir  $\alpha_{(i,j)}$  en fonction de l’intervalle de confiance. Par exemple pour avoir un intervalle de confiance égal à 99.7% on choisit  $\alpha_{(i,j)} = 3\hat{\sigma}_{(i,j)}$ , et on compare  $|\hat{\mathcal{E}}_{ij} - 1|$  à  $\alpha_{(i,j)}$  afin de tester si les sources  $i$  et  $j$  ont été séparées ou pas.

communs du bruit et de la taille des observations, il est impossible d’obtenir les coefficients de corrélation exacts des sources afin évaluer la condition d’identifiabilité. La condition d’identifiabilité devrait être testée en utilisant un certain seuil  $\alpha$ , c’est à dire que nous déciderons que les vecteurs de corrélation  $\tilde{\boldsymbol{\rho}}_i$  et  $\tilde{\boldsymbol{\rho}}_j$  sont linéairement indépendants si  $||\tilde{\boldsymbol{\rho}}_i \tilde{\boldsymbol{\rho}}_j^T| - 1| > \alpha$ .

Afin de trouver  $\alpha$  on exploite le fait que l’erreur d’estimation de  $\tilde{\boldsymbol{\rho}}_i \tilde{\boldsymbol{\rho}}_j^T$  est asymptotiquement Gaussienne<sup>3</sup> et par conséquent, on peut établir un intervalle de confiance d’une telle variable en fonction de sa variance. Cet algorithme peut être résumé par le tableau A.1.

### A.3.3 Techniques de rééchantillonnage : jackknife

Dans plusieurs applications en traitement du signal, on est intéressé par l’estimation d’un certain nombre de paramètres inconnus d’un processus aléatoire, en

<sup>3</sup>Plus précisément, on peut prouver que l’erreur estimée  $\sqrt{T}\delta(\tilde{\boldsymbol{\rho}}_i \tilde{\boldsymbol{\rho}}_j^T)$  est asymptotiquement (pour une taille d’observation  $T$  assez grande) Gaussienne avec une moyenne nulle et une variance finie.

utilisant un ensemble d'observations fini. On est aussi intéressé par la distribution échantillonnée des estimateurs, de sorte que les moyennes, les variances, et les cumulants respectifs puissent être calculés.

Le bootstrap [148] a été introduit par B. Efron [149] comme une approche pour calculer les intervalles de confiance de paramètres dans des circonstances où les méthodes standard ne peuvent pas être appliquées. Le bootstrap a par la suite été utilisé pour résoudre plusieurs autres problèmes qui seraient trop compliqués pour l'analyse statistique traditionnelle. Avec des mots simples, le bootstrap fait avec un ordinateur ce que l'expérimentateur ferait en pratique, c'est à dire si c'était possible, il ou elle répéterait l'expérience. Avec le bootstrap, les observations sont aléatoirement réassignées, et les estimées recalculées. Ces assignations et calculs sont faits des centaines ou des milliers de fois et traités comme des expériences répétées.

Le jackknife [150] est une autre technique de rééchantillonnage pour l'estimation de la déviation standard. Comme alternative au bootstrap, la méthode jackknife peut être pensée comme le tirage de  $n$  réalisations de taille  $(n - 1)$  sans remplacement de l'échantillon original de taille  $n$  [150].

Supposons que nous avons l'échantillon  $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$  et  $\hat{\vartheta}$  est une estimée calculée à partir de l'échantillon  $\mathcal{X}$ . La méthode jackknife est basée sur des échantillons amputés d'une observation pour chaque réalisation,

$$\mathcal{X}^{(i)} = \{X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}$$

avec  $i = 1, 2, \dots, n$ , appelés échantillons jackknife. Cet  $i^{\text{e}}$  échantillon jackknife consiste en l'ensemble des données avec la  $i^{\text{e}}$  observation supprimée. Pour chaque  $i^{\text{e}}$  échantillon jackknife, on calcule la  $i^{\text{e}}$  estimée jackknife  $\hat{\vartheta}^{(i)}$  de  $\vartheta$ ,  $i = 1, 2, \dots, n$ . L'estimation jackknife de la déviation standard de  $\hat{\vartheta}$  est définie par :

$$\hat{\sigma} = \sqrt{\frac{n-1}{n} \sum_{i=1}^n \left( \hat{\vartheta}^{(i)} - \frac{1}{n} \sum_{j=1}^n \hat{\vartheta}^{(j)} \right)^2} \quad (\text{A.3})$$

Le jackknife est moins coûteux que le bootstrap, si  $n$  est inférieur au nombre d'itérations utilisées par le bootstrap pour l'estimation de la déviation standard puisque il requiert le calcul de  $\hat{\vartheta}$  seulement pour les  $n$  échantillons jackknife. Par exemple, si 25 rééchantillonnages sont nécessaires pour l'estimation de la déviation standard avec le bootstrap, avec un échantillon de taille  $n = 10$  il n'en faut que 10 pour le jackknife, alors il est clair que le jackknife sera moins coûteux que le bootstrap.

Tab. A.2 – Test d'identifiabilité utilisant la technique jackknife

1. Pour chaque signal  $\mathbf{z}_i = [z_i(0), \dots, z_i(T-1)]^T$ , engendrer  $T$  vecteurs jackknife tel que  $\mathbf{z}_i^{(j)} = [z_i(0), \dots, z_i(j-1), z_i(j+1), \dots, z_i(T-1)]^T$  et  $j = 0, 1, \dots, T-1$ .

2. Pour chaque vecteur jackknife  $\mathbf{z}_i^{(j)}$ , estimer le vecteur correspondant  $\tilde{\boldsymbol{\rho}}_i^{(j)}$ .

3. Estimer la matrice  $\hat{\mathcal{E}}$  telle que son élément  $ij$  est défini par

$$\hat{\mathcal{E}}_{ij} = \frac{1}{T} \sum_{k=0}^{T-1} \frac{\langle \tilde{\boldsymbol{\rho}}_i^{(k)}, \tilde{\boldsymbol{\rho}}_j^{(k)} \rangle}{\|\tilde{\boldsymbol{\rho}}_i^{(k)}\| \|\tilde{\boldsymbol{\rho}}_j^{(k)}\|}$$

où  $\langle \cdot, \cdot \rangle$  dénote le produit scalaire.

4. Estimer la déviation standard de  $\hat{\mathcal{E}}_{ij}$  par :

$$\hat{\sigma}_{(i,j)} = \sqrt{\frac{T-1}{T} \sum_{k=0}^{T-1} \left( \frac{\langle \tilde{\boldsymbol{\rho}}_i^{(k)}, \tilde{\boldsymbol{\rho}}_j^{(k)} \rangle}{\|\tilde{\boldsymbol{\rho}}_i^{(k)}\| \|\tilde{\boldsymbol{\rho}}_j^{(k)}\|} - \frac{1}{T} \sum_{l=0}^{T-1} \frac{\langle \tilde{\boldsymbol{\rho}}_i^{(l)}, \tilde{\boldsymbol{\rho}}_j^{(l)} \rangle}{\|\tilde{\boldsymbol{\rho}}_i^{(l)}\| \|\tilde{\boldsymbol{\rho}}_j^{(l)}\|} \right)^2}$$

Afin de tester la séparabilité des signaux estimés, nous avons opté pour la méthode jackknife afin d'estimer la déviation standard du produit scalaire  $\mathcal{E}_{ij}$  pour  $i, j = 1, 2, \dots, N$ . Ceci est fait selon les étapes données dans le tableau A.2.

## A.4 Discussion

Nous fournissons ici quelques commentaires importants afin d'obtenir plus de lisibilité sur la méthode de test considérée ainsi que ses applications et extensions potentielles.

### Qualité de la séparation

L'analyse de performance asymptotique de SOBI tirée de [31], montre que les performances de la séparation de deux sources  $s_i$  et  $s_j$  dépendent de l'angle entre les vecteurs de corrélation respectives  $\tilde{\boldsymbol{\rho}}_i$  et  $\tilde{\boldsymbol{\rho}}_j$ . Donc, la mesure de cet angle donne une idée sur le niveau d'interférence des deux sources considérées. Par conséquent,

on peut utiliser la mesure de cet angle non seulement pour évaluer la séparabilité des deux sources, mais aussi pour garantir une qualité de séparation minimale. Le choix du seuil  $\alpha_{(i,j)}$  est par conséquent une question importante actuellement à l'étude.

### Procédure de séparation à deux étapes

La méthode de test peut être incorporée dans une procédure de séparation à deux étapes où la première étape consiste en une méthode de décorrélation au second ordre (SOBI). La deuxième étape serait une méthode de séparation basée sur les statistiques d'ordre supérieur appliquée seulement quand le test indique un échec de la séparation à la première étape.

### Séparation partielle

Dans beaucoup de situations pratiques, on pourrait être intéressé par seulement un ou quelques signaux sources. C'est le cas par exemple dans le problème de réduction d'interférence dans [151] ou dans des applications de surveillance de centrales électriques [152]. Dans ces situations, le résultat d'identifiabilité partielle est d'un grand intérêt car il montre que le signal source désiré peut être extrait même si une séparation complète des sources ne peut pas être réalisée.

### Statistiques d'ordre supérieur

Nous pensons que des méthodes de test semblable peuvent être employées pour des méthodes de SAS basée sur les statistiques d'ordre supérieur, du moins pour des méthodes de type JADE [43], qui sont basées sur la décorrélation d'ordre quatre.

## **A.5 Simulations et résultats**

Nous présentons dans cette section quelques résultats de simulations pour illustrer les performances de notre méthode de test. Pour notre environnement de simulation, nous considérerons une antenne composée de  $M = 2$  capteurs, recevant  $N = 2$  signaux auto-régressifs du premier ordre (avec des coefficients  $a_1 = 0.95e^{j0.5}$  et  $a_2 = 0.5e^{j0.7}$ ) de puissance unité, en présence d'un bruit complexe blanc et stationnaire. Les sources considérées sont séparables selon le résultat d'identifi-

bilité. En d'autres termes, leurs vecteurs de corrélation respectifs  $\tilde{\rho}_1$  et  $\tilde{\rho}_2$  sont linéairement indépendants. Nous utiliserons l'algorithme SOBI [31] avec  $K = 10$  pour obtenir les sources décorréliées. Les statistiques sont évaluées pour  $N_r = 100$  tirages aléatoires.

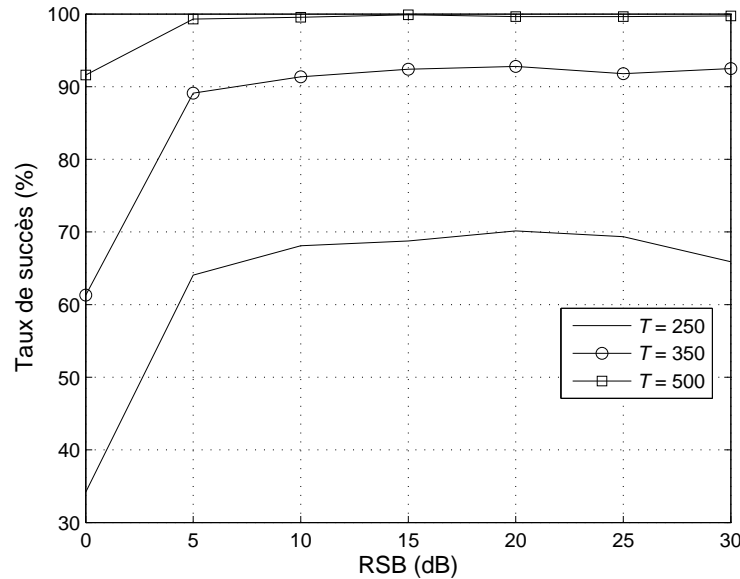


Fig. A.1 – Taux de succès en fonction du RSB pour  $N = 2$  sources auto-régressives et  $M = 2$  capteurs avec  $\beta = 99.7\%$  : comparaison des performances de notre procédure de test pour différentes tailles d'observations  $T$ .

On présente premièrement dans la figure A.1 un exemple de simulation où on compare le taux de succès de notre procédure de test (le succès signifie que nous décidons que les deux sources ont été séparées) pour détecter la séparabilité de sources pour différentes tailles d'observations en fonction du RSB. L'intervalle de confiance est fixé à  $\beta = 99.7\%$ . On peut observer à partir de cette figure que les performances de notre procédure de test se dégradent de manière significative pour de petites tailles d'observations, du fait de l'amplification de l'erreur d'estimation, et du fait que nous utilisons la normalité asymptotique des statistiques considérées.

Dans la figure A.2, on présente un exemple de simulation où on compare le taux de succès en fonction de la taille des observations pour différentes valeurs de l'intervalle de confiance. Le RSB est fixé à 25dB. Il est clair que plus l'intervalle de confiance est faible, plus le taux de succès de notre procédure de test est élevé. De plus, comme observé dans la figure A.1, le taux de succès augmente rapidement

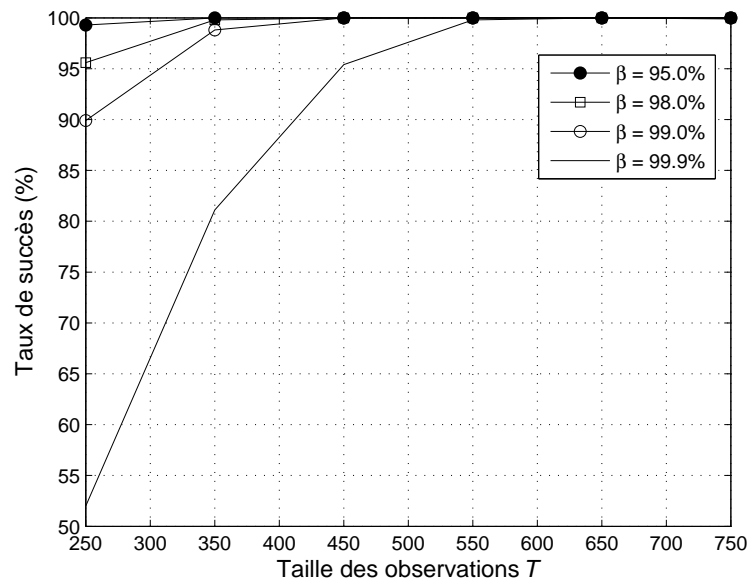


Fig. A.2 – Taux de succès en fonction de la taille des observations  $T$  pour  $N = 2$  sources auto-régressives et  $M = 2$  capteurs avec un RSB = 25dB : comparaison des performances de notre procédure de test pour différentes valeurs de l'intervalle de confiance  $\beta$ .

quand la taille des observations croît.

Dans la figure A.3, on introduit un exemple de simulation où on représente le taux de succès de notre procédure de test, en fonction de la valeur de l'intervalle de confiance  $\beta$  pour différentes tailles d'observations et un RSB de 25dB. Cette figure montre d'une façon ou d'une autre l'évolution du taux de succès (ou le taux de fausses alarmes) et confirme les résultats des deux figures précédentes.

Dans la simulation présentée dans la figure A.4, on utilise des signaux sources avec les paramètres  $a_1 = 0.5e^{j0.5}$  et  $a_2 = 0.5e^{j(0.5+\delta\theta)}$ , où  $\delta\theta$  représente le décalage spectral entre les deux sources. Le nombre de capteurs est  $M = 5$ , la taille des observations est  $T = 1000$  et le RSB = 30dB. La figure A.4 expose le taux de succès en fonction du décalage spectral  $\delta\theta$ . Comme on peut le voir, l'utilisation de notre procédure de test avec de faibles valeurs de  $\delta\theta$ , conduit à un taux important de décision de "non-séparabilité". En effet, quand  $\delta\theta$  est proche de zéro, les deux vecteurs de corrélation  $\tilde{\rho}_1$  et  $\tilde{\rho}_2$  sont proches de la dépendance linéaire, ce qui implique une dégradation de la qualité de la séparation des deux sources, et par conséquent les mauvais résultats obtenus par notre procédure de test.

Dans la dernière figure, on suppose qu'il existe trois sources. Les deux premières

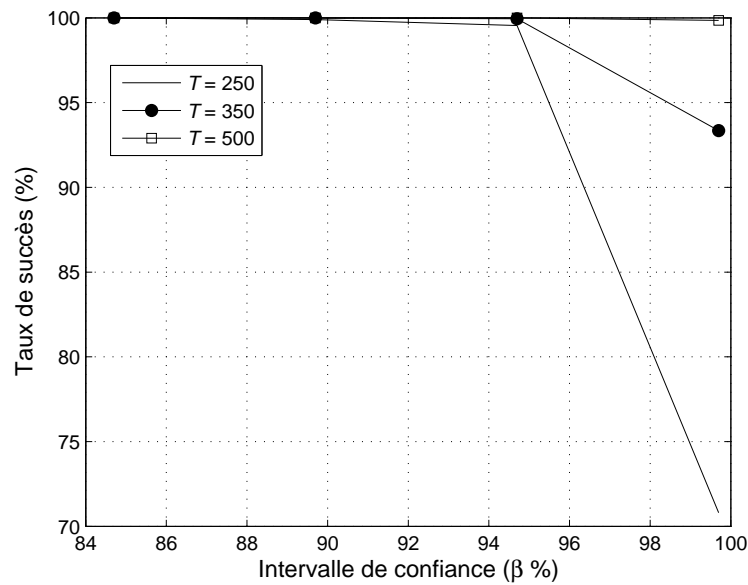


Fig. A.3 – Taux de succès en fonction de la valeur de l'intervalle de confiance  $\beta$  pour  $N = 2$  sources auto-régressives et  $M = 2$  capteurs avec un RSB = 25dB : comparaison des performances de notre procédure de test pour différentes tailles d'observations  $T$ .

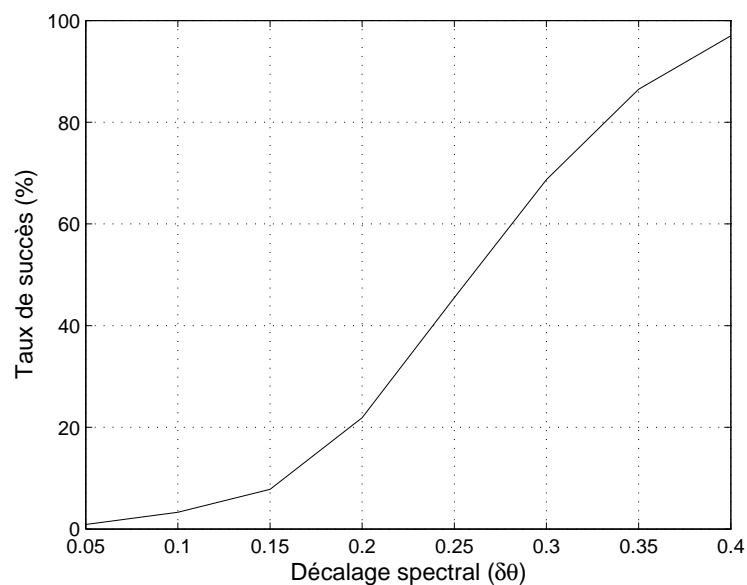


Fig. A.4 – Taux de succès en fonction du décalage spectral  $\delta\theta$  pour  $N = 2$  sources auto-régressives et  $M = 5$  capteurs avec un RSB = 30dB.



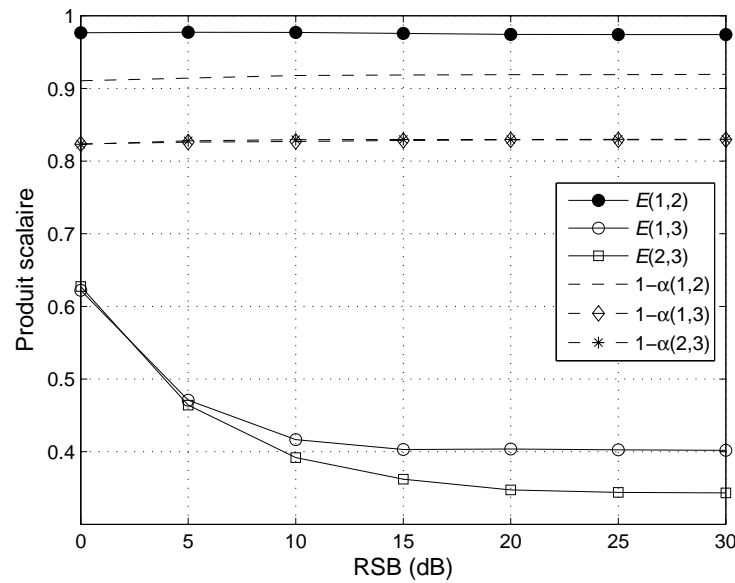


Fig. A.5 – Valeurs moyennes de  $|\mathcal{E}(i, j)|$  et les seuils  $1 - \alpha_{(i,j)}$  en fonction du RSB pour  $N = 3$  sources et  $M = 5$  capteurs : les 2 deux premières sources sont des processus aléatoires complexes suivant une loi Gaussienne et temporellement blancs et la troisième est un signal auto-régressif du premier ordre.

sources sont des processus aléatoires complexes suivant une loi Gaussienne et temporellement blancs (par conséquent  $\tilde{\rho}_1 = \tilde{\rho}_2$ ), et la troisième est un signal auto-régressif du premier ordre avec un coefficient  $a_3 = 0.95e^{j0.5}$ . Dans la figure A.5, on compare la valeur moyenne du produit scalaire entre  $\tilde{\rho}_i$  et  $\tilde{\rho}_j$  ( $i, j = 1, 2, 3$ ) avec leurs valeurs seuil  $1 - \alpha_{(i,j)}$  en fonction du RSB. La taille des observations est fixée à  $T = 500$  et le nombre de capteurs est  $M = 5$ . Cet exemple illustre la situation où deux sources (dans cette exemple les sources 1 et 2) ne sont pas séparables (ceci est confirmé par les résultats du test) tandis que la troisième est correctement extraite (la figure montre clairement que  $\mathcal{E}(1, 3) < 1 - \alpha_{(1,3)}$  et  $\mathcal{E}(2, 3) < 1 - \alpha_{(2,3)}$ ).

## A.6 Conclusion

Cette annexe a introduit une nouvelle méthode de test de la condition d'identifiabilité au second ordre pour la séparation aveugle de sources. En d'autres termes, ce test nous permet de vérifier de manière aveugle, à partir des observations, si les sources ont été correctement séparées ou pas. Afin d'évaluer les statistiques dont on a besoin pour la procédure de test on a utilisé la technique de rééchantillonnage

jackknife. Les résultats des simulations illustrent et évaluent l'efficacité de cette méthode de test du moins pour des tailles d'observation modérées et grandes.

## Démonstrations

---

### B.1 Démonstration du théorème A.3

L'équation (A.1) est équivalente à<sup>1</sup>

$C\mathbf{R}_s(\tau_k)C^H$  diagonale pour  $k = 1, \dots, K$

$$C \left( \sum_{k=1}^K \mathbf{R}_s(\tau_k)^H C^H C \mathbf{R}_s(\tau_k) \right) C^H > 0$$

où  $C = BA$  et l'inégalité  $M > 0$  indique que la matrice  $M$  est définie positive. L'inégalité précédente implique en particulier que  $C$  est de rang plein. Notons que  $C\mathbf{R}_s(\tau_k)C^H$  est diagonale si et seulement si les deux matrices  $C\Re(\mathbf{R}_s(\tau_k))C^H$  et  $C\Im(\mathbf{R}_s(\tau_k))C^H$  sont diagonales.

Alors nous montrerons qu'il existe des combinaisons linéaires de  $\Re(\mathbf{R}_s(\tau_k))$  et  $\Im(\mathbf{R}_s(\tau_k))$ ,  $\mathcal{R}_1 = \sum_{k=1}^K \alpha_k \Re(\mathbf{R}_s(\tau_k)) + \alpha'_k \Im(\mathbf{R}_s(\tau_k))$  et  $\mathcal{R}_2 = \sum_{k=1}^K \beta_k \Re(\mathbf{R}_s(\tau_k)) + \beta'_k \Im(\mathbf{R}_s(\tau_k))$ , telles que **(i)**  $\mathcal{R}_2$  est non-singulière et **(ii)** les éléments diagonaux de  $\mathcal{R}_1 \mathcal{R}_2^{-1}$  prennent  $d$  valeurs distinctes de multiplicité  $N_1, \dots, N_d$  respectivement. Une façon simple de le prouver est de considérer des combinaisons linéaires de la forme  $\alpha_k = x^k, \alpha'_k = x^{k+K}$  et  $\beta_k = y^k, \beta'_k = y^{k+K}$  ( $x$  et  $y$  étant des scalaires réels). Les éléments diagonaux de  $\mathcal{R}_1$  et  $\mathcal{R}_2$  sont  $P_i(x) = \sum_{k=1}^K \Re(\rho_i(\tau_k))x^k + x^K \sum_{k=1}^K \Im(\rho_i(\tau_k))x^k$  et  $P_i(y)$ ,  $i = 1, \dots, N$  respectivement. Sous l'hypothèse du

---

<sup>1</sup>Le terme de bruit est négligé. Cependant, ce terme est nul si tous les retards sont non-nuls et si le bruit est temporellement blanc.

Théorème A.2, on a<sup>2</sup>

$$\begin{aligned} P_i(y) &\neq 0 \text{ pour tout } i \\ P_i(x)P_j(y) &\neq P_i(y)P_j(x) \text{ si } i \text{ et } j \text{ appartiennent à deux groupes distincts} \\ P_i(x)P_j(y) &\equiv P_i(y)P_j(x) \text{ si } i \text{ et } j \text{ appartiennent au même groupe} \end{aligned}$$

Par conséquent,

$$P(x, y) \triangleq \prod_{i=1}^d P_i(y) \prod_{1 \leq i < j \leq d} (P_i(x)P_j(y) - P_j(x)P_i(y)) \neq 0$$

(ici  $i$  et  $j$  dénotent les indices des groupes et non ceux des sources) et ainsi il existe une infinité de valeurs de  $(x, y)$  telles que  $P(x, y) \neq 0$ . Soit  $(x_0, y_0)$  des valeurs telles que  $\mathcal{R}_1 = \sum_{k=1}^K x_0^k \Re e(\mathbf{R}_s(\tau_k)) + x_0^{k+K} \Im m(\mathbf{R}_s(\tau_k))$  et  $\mathcal{R}_2 = \sum_{k=1}^K y_0^k \Re e(\mathbf{R}_s(\tau_k)) + y_0^{k+K} \Im m(\mathbf{R}_s(\tau_k))$ .  $\mathcal{R}_1$  et  $\mathcal{R}_2$  satisfont les conditions (i) et (ii).

Afin de compléter la démonstration, notons que

$$\mathbf{M}_1 \triangleq \sum_{k=1}^K x_0^k \mathbf{C} \Re e(\mathbf{R}_s(\tau_k)) \mathbf{C}^H + x_0^{k+K} \mathbf{C} \Im m(\mathbf{R}_s(\tau_k)) \mathbf{C}^H = \mathbf{C} \mathcal{R}_1 \mathbf{C}^H$$

est diagonale et

$$\mathbf{M}_2 \triangleq \sum_{k=1}^K y_0^k \mathbf{C} \Re e(\mathbf{R}_s(\tau_k)) \mathbf{C}^H + y_0^{k+K} \mathbf{C} \Im m(\mathbf{R}_s(\tau_k)) \mathbf{C}^H = \mathbf{C} \mathcal{R}_2 \mathbf{C}^H$$

est diagonale et non-singulière. Il s'en suit que,  $\mathbf{M}_1 \mathbf{M}_2^{-1} = \mathbf{C} \mathcal{R}_1 \mathcal{R}_2^{-1} \mathbf{C}^{-1}$  est une matrice diagonale avec  $d$  valeurs propres distinctes (condition (ii)).

En utilisant la théorie spectrale standard [153], on conclut que  $\mathbf{C} = \mathbf{P} \mathbf{\Lambda}$  pour une matrice de permutation  $\mathbf{P}$  donnée et une matrice bloc-diagonale non-singulière  $\mathbf{\Lambda}$  donnée,  $\mathbf{\Lambda} = \text{diag}(\mathbf{U}_1, \dots, \mathbf{U}_d)$  où  $\mathbf{U}_i$  est une matrice non-singulière de taille  $N_i \times N_i$ . Enfin, le fait que des sources appartiennent au même groupe ne peuvent pas être séparées en utilisant l'ensemble des statistiques considérées est simplement la conséquence directe de la nécessité de la condition (A.1).  $\square$

## B.2 Démonstration du théorème A.4

Selon le développement ci-dessus, on a  $\mathbf{C} = \mathbf{P} \text{diag}(\mathbf{U}_1, \dots, \mathbf{U}_d)$ . D'autre part le produit  $\mathbf{C} \mathbf{R}_s(\tau_k) \mathbf{C}^H$  est une matrice diagonale pour  $k = 1, \dots, K$ . Cela induit que

<sup>2</sup>Les égalités ci-dessous sont interprétées en termes d'égalités polynômiales.

$U_i \Sigma_i U_i^H$  est diagonale pour  $i = 1, \dots, d$ , où les  $\Sigma_i$  sont des matrices diagonales dont les éléments diagonaux sont égaux à  $\pm 1$ . De plus, comme  $\Sigma_i$  est une matrice réelle diagonale et comme on admet que les vecteurs de corrélation sont normalisés, alors  $U_i \Sigma_i U_i^H$  est une matrice diagonale dont les éléments sont égaux à  $\pm 1$ . Finalement, comme la signature (le nombre des valeurs propres positives et le nombre des valeurs propres négatives) de la forme quadratique  $Q$  est invariante par multiplication à droite et à gauche par une matrice non-singulière et sa transposée conjuguée respectivement, alors  $U_i \Sigma_i U_i^H$  à la même signature que  $\Sigma_i$ . En d'autres termes, il existe une matrice de permutation  $P_i$  de taille  $N_i \times N_i$  tel que  $U_i \Sigma_i U_i^H = P_i \Sigma_i P_i^T$ . nous en déduisons que pour  $k = 1, \dots, K$

$$\mathbb{E}[\mathbf{z}(t + \tau_k) \mathbf{z}(t)^H] = \mathbf{C} \mathbf{R}_s(\tau_k) \mathbf{C}^H = \mathbf{P} \mathbf{R}_s(\tau_k) \mathbf{P}^T, \quad (\text{B.1})$$

pour une matrice de permutation  $\mathbf{P}$  donnée. L'équation (B.1) signifie que les éléments  $z_i(t)$  de  $\mathbf{z}(t)$  ont les mêmes coefficients de corrélation que  $s_{P(i)}(t)$  for  $i = 1, \dots, N$ , où  $P(1), \dots, P(N)$  sont les images de  $1, \dots, N$  par la permutation  $\mathbf{P}$ . Ceci conclut la vérification du théorème A.4.  $\square$



# Bibliographie

- [1] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Underdetermined blind audio source separation using modal decomposition”, *EURASIP Journal on Audio, Speech & Music Processing*, vol. 2007, pp. 1–15, March 2007.
- [2] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Séparation aveugle sous-déterminée de sources audio par la méthode EMD (Empirical Mode Decomposition)”, in *Actes 20<sup>e</sup> colloque GRETSI sur le Traitement du Signal et des Images*, Louvain-La-Neuve, Belgique, Septembre 2005.
- [3] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Blind separation of audio sources using modal decomposition”, in *Proc. 8<sup>th</sup> International Symposium on Signal Processing and Its Applications ISSPA*, Sydney, Australia, August 2005, vol. 2, pp. 451–454.
- [4] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Underdetermined audio source separation using fast parametric decomposition”, in *Proc. 9<sup>th</sup> International Symposium on Signal Processing and its Applications ISSPA*, Sharjah, U.A.E., February 2007.
- [5] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Blind separation of audio sources convolutive mixtures using parametric decomposition”, in *Proc. 9<sup>th</sup> International Workshop on Acoustic Echo and Noise Control IWAENC*, Eindhoven, Pays-bas, September 2005, vol. 1, pp. 161–164.
- [6] A. Aïssa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier, “Underdetermined blind separation of non-disjoint sources using time-frequency representation”, *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 897–907, March 2007.
- [7] N. Linh-Trung, A. Aïssa-El-Bey, K. Abed-Meraim, and A. Belouchrani, “Underdetermined blind source separation of non-disjoint nonstationary

- sources in time–frequency domain”, in *Proc. 8<sup>th</sup> International Symposium on Signal Processing and Its Applications ISSPA*, Sydney, Australia, August 2005, vol. 1, pp. 46–49.
- [8] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Underdetermined blind source separation of audio sources in time–frequency domain”, in *Proc. Signal Processing with Adaptive Sparse Structured Representations SPARS*, Rennes, France, November 2005, vol. 1, pp. 115–118.
- [9] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Séparation aveugle sous-déterminée de sources en utilisant la décomposition en paquet d’ondelettes”, in *Actes 21<sup>e</sup> colloque GRETSI sur le Traitement du Signal et des Images*, Troyes, France, Septembre 2007.
- [10] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Blind separation of underdetermined convolutive mixtures using their time–frequency representation”, *IEEE Transactions on Audio, Speech & Language Processing*, vol. 15, no. 5, July 2007.
- [11] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Underdetermined blind separation of audio sources from the time–frequency representation of their convolutive mixtures”, in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Hawaii, USA, April 2007, vol. 1, pp. 153–156.
- [12] A. Aïssa-El-Bey, H. Bousbia-Salah, K. Abed-Meraim, and Y. Grenier, “Audio source separation using sparsity”, in *Proc. 10<sup>th</sup> International Workshop on Acoustic Echo and Noise Control IWAENC*, Paris, France, September 2006.
- [13] W. Soudène, A. Aïssa-El-Bey, K. Abed-Meraim, and A. Beghdadi, “Blind image separation using sparse representation”, in *Proc. 14<sup>th</sup> International Conference on Image Processing ICIP*, San Antonio, Texas, USA, September 2007.
- [14] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Blind audio source separation using sparsity based criterion for convolutive mixture case”, in *Proc. 7<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, London, UK, September 2007.
- [15] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “On the identifiability testing in blind source separation using resampling technique”, in *Proc. 6<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Charleston, SC, USA, March 2006, vol. LCNS 3889, pp. 755–764.



- 
- [16] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, "Iterative blind source separation by decorrelation : algorithm and performance analysis", in *Proc. 14<sup>th</sup> European Signal Processing Conference EUSIPCO*, Florence, Italy, September 2006.
- [17] A. K. Nandi, *Blind estimation using higher-order statistics*, Kluwer Academic Publishers, Boston, January 1999.
- [18] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing : Learning Algorithms and Applications*, John Wiley and Sons, March 2003.
- [19] E. Vincent, C. Févotte, and R. Gribonval, "A tentative typology of audio source separation tasks", in *Proc. 4<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Nara, Japan, April 2003, pp. 715–720.
- [20] J. Héroult, C. Jutten, and B. Ans, "Détection de grandeur primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé", in *Actes 10<sup>e</sup> colloque GRETSI sur le Traitement du Signal et des Images*, Nice, France, Mai 1985, pp. 1017–1022.
- [21] C. Jutten, *Calcul neuromimétique et traitement du signal : analyse en composantes indépendantes*, PhD thesis, Université Joseph Fourier-Institut National Polytechnique de Grenoble, Grenoble, France, 1987.
- [22] J. L. Lacoume and P. Ruiz, "Sources identification : a solution based on cumulants", in *Proc. 4<sup>th</sup> Annual ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, MN, USA, August 1988.
- [23] J. F. Cardoso, "Source separation using higher order moments", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Glasgow, May 1989, vol. 4, pp. 2109–2112.
- [24] J. F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. blind identification of more sources than sensors", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Toronto, Canada, April 1991, vol. 5, pp. 3109–3112.
- [25] J. F. Cardoso, "Iterative techniques for blind source separation using only fourth order cumulants", in *Proc. 6<sup>th</sup> European Signal Processing Conference EUSIPCO*, Brussels, Belgium, 1992, pp. 739–742.
- [26] P. Comon, "Analyse en composantes indépendantes et identification aveugle", *Traitement du signal*, vol. 7, no. 5, pp. 435–450, 1990.

- [27] M. Gaeta and J. L. Lacoume, “Estimateurs du maximum de vraisemblance étendus à la séparation de sources non gaussiennes”, *Traitement du signal*, vol. 7, no. 5, pp. 419–434, 1990.
- [28] N. Delfosse and P. Loubaton, “Séparation adaptative de sources indépendantes par une approche de déflation”, in *Actes 14<sup>e</sup> colloque GRETSI sur le Traitement du Signal et des Images*, Juan Les Pins, France, Septembre 1993, pp. 313–316.
- [29] B. Laheld and J. F. Cardoso, “Séparation adaptative de sources en aveugle, implémentation complexe sans contraintes”, in *Actes 14<sup>e</sup> colloque GRETSI sur le Traitement du Signal et des Images*, Juan Les Pins, France, Septembre 1993, pp. 329–332.
- [30] D. T. Pham, P. Garat, and C. Jutten, “Separation of a mixture of independent sources through a maximum likelihood approach”, in *Proc. 6<sup>th</sup> European Signal Processing Conference EUSIPCO*, Brussels, Belgium, 1992, pp. 771–774.
- [31] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, “A blind source separation technique using second-order statistics”, *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, February 1997.
- [32] D. T. Pham and J. F. Cardoso, “Blind separation of instantaneous mixtures of non stationary sources”, in *Proc. 2<sup>nd</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Helsinki, Finlande, June 2000, pp. 187–192.
- [33] D. T. Pham and J. F. Cardoso, “Blind separation of instantaneous mixtures of non stationary sources”, *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1837–1848, September 2001.
- [34] K. B. Christensen, “The application of digital signal processing to largescale simulation of room acoustics : frequency response modeling and optimization software for a multichannel dsp engine”, *Journal Audion Engineering Society*, vol. 40, no. 4, pp. 260–276, April 1992.
- [35] V. Capdevielle, *Séparation de sources large bande à l’aide des moments d’ordre supérieur*, PhD thesis, Institut National Polytechnique de Grenoble, France, 1995.
- [36] A. Mansour, *Contribution à la séparation aveugle de sources*, PhD thesis, Institut National Polytechnique de Grenoble, France, 1997.

- 
- [37] C. Jutten, L. Nguyen-Thi, E. Dijkstra, E. Vittoz, and J. Caelen, "Blind separation of sources : an algorithm for separation of convolutive mixtures", in *Proc. International Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 1991, pp. 275–278.
- [38] H. L. Nguyen-Thai and C. Jutten, "Blind source separation for convolutive mixtures", *Signal Processing*, vol. 45, no. 2, pp. 209–229, 1995.
- [39] P. Comon, "Independent component analysis, a new concept?", *Signal Processing, Elsevier, Special issue on Higher-Order Statistics*, vol. 36, pp. 287–314, April 1994.
- [40] P. Comon, P. Chevalier, and V. Capdevielle, "Performance of contrast-based blind source separation", in *Proc. 1<sup>st</sup> Signal Processing Advances in Wireless Communication SPAWC*, April 1997, pp. 345–348.
- [41] P. Comon and E. Moureau, "Improved contrast dedicated to blind separation in communications", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Munich, Germany, April 1997, pp. 3453–3456.
- [42] J. F. Cardoso and P. Comon, "Independent component analysis, a survey of some algebraic methods", in *Proc. International Symposium on Circuits and Systems ISCAS*, 1996, vol. 2, pp. 93–96.
- [43] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals", *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, December 1993.
- [44] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation", *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, December 1996.
- [45] S. I. Amari, "Neural learning in structured parameter spaces–natural riemannian gradient", in *Proc. Neural Information Processing System–Natural and Synthetic*, Colorado, USA, December 1996, pp. 127–133.
- [46] S. I. Amari, A. Cichocki, and H. Yang, "A new learning algorithm for blind signal separation", in *Advances in Neural Information Processing Systems*, Denver, USA, 1996, vol. 8, pp. 757–763.
- [47] J. F. Cardoso, "Infomax and maximum likelihood for source separation", *IEEE Signal Processing Letters*, vol. 4, no. 4, pp. 112–114, April 1997.

- 
- [48] J. F. Cardoso, “Blind signal separation : statistcal principles”, *Proceeding of the IEEE special issue on blind identification and estimation*, vol. 86, no. 10, pp. 2009–2025, October 1998.
- [49] A. Belouchrani and J. F. Cardoso, “Maximum likelihood source separation for discrete sources”, in *Proc. 7<sup>th</sup> European Signal Processing Conference EUSIPCO*, Edinburgh, September 1994, pp. 768–771.
- [50] A. Belouchrani and J. F. Cardoso, “Maximum likelihood source separation by the expectation-maximization technique : deterministic and stochastic implementation”, in *Proc. International Symposium on Non-Linear Theory and Applications NOLTA*, Las Vegas, NV, USA, 1995, pp. 49–53.
- [51] O. Bermond, *Méthodes statistiques pour la séparation de sources*, PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France, 2000.
- [52] J. F. Cardoso, “Learninig in manifolds : the case of source separation”, in *Proc. 9<sup>th</sup> IEEE SP Workshop on Statistical Signal and Array Processing SSAP*, Portland, Oregon, September 1998, pp. 136–139.
- [53] D. T. Pham and P. Garat, “Blind separation of mixture of independent sources through a quasi-maximum likelihood approach”, *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1712–1725, July 1997.
- [54] L. Féty, *Méthodes de traitement d’antenne adaptées aux radio-communications*, PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France, 1988.
- [55] L. Tong and R. Liu, “Blind estimation of correlated source signals”, in *Proc. Asilomar Conference on Signals, Systems and Computers*, November 1990, vol. 1, pp. 258–262.
- [56] A. Belouchrani, *Séparation autodidacte de sources : algorithmes, performances et applications à des signaux expérimentaux*, PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France, Juillet 1995.
- [57] G. H. Golub and C. F. V. Loan, *Matrix computations*, The Johns Hopkins University Press, October 1996.
- [58] J. F. Cardoso and A. Souloumiac, “An efficient technique for blind separation of complex sources”, in *Proc. IEEE Signal Processing Workshop on Higher Order Statistics*, Lake Tahoe, USA, June 1993, pp. 275–279.

- 
- [59] J. F. Cardoso and A. Souloumiac, “Jacobi angles for simultaneous diagonalization”, *SIAM Journal of Matrix Analysis and Applications*, vol. 17, no. 1, pp. 161–164, January 1996.
- [60] L. Parra and C. Spence, “Convulsive blind separation of non-stationary sources”, *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.
- [61] C. Févotte, *Approche temps-fréquence pour la séparation aveugle de sources non-stationnaires*, PhD thesis, Ecole Centrale de Nantes et de l’Université de Nantes, Nantes, France, Octobre 2003.
- [62] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, “A survey of convolutive blind source separation methods”, in *Springer Handbook of Speech (to appear)*. Springer Press, September 2007.
- [63] K. Torkkola, “Blind separation of convolved sources based on information maximization”, in *Proc. IEEE Workshop on Neural Networks and Signal Processing*, Kyoto, Japan, September 1996, pp. 423–432.
- [64] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution”, *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [65] T. W. Lee, A. J. Bell, and R. Lambert, “Blind separation of delayed and convolved sources”, in *Proc. Advances in Neural Information Processing Systems 9*. 1997, pp. 758–764, Cambridge, MA : MIT Press.
- [66] R. Mukai, S. Araki, and S. Makino, “Separation and dereverberation performance of frequency domain blind source separation”, in *Proc. 3<sup>rd</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, San Diego, California, USA, December 2001, pp. 230–235.
- [67] R. Mukai, S. Araki, and S. Makino, “Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment”, in *Proc. of European Conference on Speech Communication and Technology EUROSPEECH*, September 2001, pp. 2599–2603.
- [68] R. Mukai, S. Araki, H. Sawada, and S. Makino, “Evaluation of separation and dereverberation performance in frequency domain blind source separation”, *Journal of Acoustic Science and Technology*, vol. 25, no. 2, pp. 119–126, March 2004.

- [69] P. Smaragdis, “Information theoretic approaches to source separation”, Master’s thesis, MIT Media Lab, June 1997.
- [70] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain”, *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, November 1998.
- [71] N. Mitianoudis and M. E. Davies, “Audio source separation of convolutive mixtures”, *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 489–497, September 2003.
- [72] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meraim, “Jacobi-like algorithm for blind signal separation of convolutive mixtures”, *IEE Electronics Letters*, vol. 37, no. 16, pp. 1049–1050, August 2001.
- [73] P. Comon and O. Grellier, “Nonlinear inversion of underdetermined mixtures”, in *Proc. 1<sup>st</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Aussois, France, January 1999, pp. 461–465.
- [74] M. Lewicki and T. J. Sejnowski, “Learning nonlinear overcomplete representations for efficient coding”, in *Proc. Advances in neural Information Processing Systems*, 1998, vol. 10, pp. 815–821.
- [75] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.
- [76] O. Bermond, E. Moulines, and J. F. Cardoso, “Séparation et déconvolution aveugle de signaux bruités : modélisation par mélange de gaussiennes”, in *Actes 16<sup>e</sup> colloque GRETSI sur le Traitement du Signal et des Images*, Grenoble, France, Septembre 1997, pp. 1427–1430.
- [77] T. W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, “Blind source separation of more sources than mixtures using overcomplete representations”, *IEEE Signal Processing Letters*, vol. 6, no. 4, pp. 87–90, April 1999.
- [78] H. Attias, “Independent factor analysis”, *Neural Computation*, vol. 11, no. 4, pp. 803–851, 1999.
- [79] J. Miskin and D. MacKay, “Ensemble learning for blind source separation”, *Independent Component Analysis : Principles and Practice*, pp. 209–233, 2001.
- [80] C. Févotte, “Bayesian approach for blind separation of undetermined mixtures of sparse sources”, in *Proc. 5<sup>th</sup> Symposium on Independent Com-*

- ponent Analysis and Blind Source Separation ICA*, Granada, Spain, September 2004, pp. 398–405.
- [81] F. J. Theis, C. Puntonet, and E. W. Lang, “A histogram-based overcomplete ica algorithm”, in *Proc. 4<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Nara, Japan, April 2003, pp. 1071–1076.
- [82] M. E. Davies, “A sparse mixture of gaussians model for blind separation of more sources than sensors”, in *Proc. N2SP Workshop*, 2002.
- [83] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev, “Blind source separation by sparse decomposition in a signal dictionary”, *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.
- [84] R. Gribonval, “Piecewise linear source separation”, in *Proc. SPIE Conference “Wavelets : Applications in Signal and Image Processing”*, San Diego, California, USA, August 2003, vol. 5207, pp. 297–310.
- [85] E. Vincent and X. Rodet, “Underdetermined source separation with structured source priors”, in *Proc. 5<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Granada, Spain, September 2004, pp. 327–334.
- [86] L. Vielva, D. Erdogmus, and J. C. Príncipe, “Underdetermined blind source separation using a probabilistic source sparsity model”, in *Proc. 3<sup>rd</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, San Diego, California, USA, December 2001, pp. 675–679.
- [87] O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking”, *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July 2004.
- [88] E. L. Benaroya, *Séparation de plusieurs sources sonores avec un seul microphone*, PhD thesis, Université de Rennes 1, Rennes, France, Juin 2003.
- [89] M. A. Casey and W. Westner, “Separation of mixed audio sources by independent subspace analysis”, in *Proc. of the International Computer Music Conference*, Berlin, 2000.
- [90] S. Rickard and O. Yilmaz, “On the approximate w-disjoint orthogonality of speech”, in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Orlando, Florida, USA, May 2002, vol. 1, pp. 529–532.

- [91] S. Rickard, T. Melia, and C. Fearon, "DESPRIT - histogram based blind source separation of more sources than sensors using subspace methods", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA*, New York, USA, October 2005, pp. 5–8.
- [92] A. Blin, S. Araki, and S. Makino, "Underdetermined blind separation of convolutive mixtures of speech using time-frequency mask and mixing matrix estimation", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 7, pp. 1693–1700, July 2005.
- [93] S. Araki, S. Makino, H. Sawada, and R. Mukai, "Reducing musical noise by a fine-shift overlap-add method applied to source separation using a time-frequency mask", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Philadelphia, USA, March 2005, vol. 3, pp. 81–84.
- [94] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind extraction of a dominant source signal from mixtures of many sources", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Philadelphia, USA, March 2005, vol. 3, pp. 61–64.
- [95] J.M. Peterson and S. Kadambe, "A probabilistic approach for blind source separation of underdetermined convolutive mixtures", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, April 1998, vol. 6, pp. 581–584.
- [96] D. Luengo, I. Santamaria, L. Vielva, and C. Pantaleon, "Underdetermined blind separation of sparse sources with instantaneous and convolutive mixtures", in *Proc. 13<sup>th</sup> IEEE Workshop on Neural Networks for Signal Processing NNSP*, September 2003, pp. 279–288.
- [97] S. Winter, H. Sawada, S. Araki, and S. Makino, "Overcomplete BSS for convolutive mixtures based on hierarchical clustering", in *Proc. 5<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Granada, Spain, September 2004, pp. 652–660.
- [98] S. Winter, H. Sawada, S. Araki, and S. Makino, "Hierarchical clustering applied to overcomplete BSS for convolutive mixtures", in *Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing SAPA*, October 2004.



- 
- [99] S. Winter, W. Kellermann, H. Sawada, and S. Makino, "Map-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and  $\ell_1$ -norm minimization", *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–12, January 2007.
- [100] P. Smaragdis, "Non negative matrix factor deconvolution, extraction of multiple sound sources from monophonic inputs", in *Proc. 5<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Granada, Spain, September 2004, pp. 494–499.
- [101] M. S. Pederson, T. Lehn-Schiøler, and J. Larsen, "BLUES from music : Blind underdetermined extraction of sources from music", in *Proc. 6<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Charleston, SC, USA, March 2006, pp. 392–399.
- [102] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Trung, and H. H. Liu, "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis", *Royal Society London A*, vol. 454, no. 1971, pp. 903–995, March 1998.
- [103] P. Flandrin, G. Rilling, and P. Goncalvès, "Empirical mode decomposition as a filter bank", *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, February 2004.
- [104] R. Boyer and K. Abed-Meraim, "Audio modeling based on delayed sinusoids", *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 2, pp. 110–120, March 2004.
- [105] J. Nieuwenhuijse, R. Heusens, and Ed. F. Deprettere, "Robust exponential modeling of audio signals", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, May 1998, vol. 6, pp. 3581–3584.
- [106] D. Nuzillard and J.-M. Nuzillard, "Application of blind source separation to 1-d and 2-d nuclear magnetic resonance spectroscopy", *IEEE Signal Processing Letters*, vol. 5, no. 8, pp. 209–211, August 1998.
- [107] H. Park, S. Van Huffel, and L. Elden, "Fast algorithms for exponential data modeling", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Adelaide, Australia, April 1994, vol. 4, pp. 25–28.
- [108] C. Serviere, V. Capdevielle, and J.-L. Lacoume, "Separation of sinusoidal sources", in *Proc. IEEE SP Workshop on Higher-Order Stat.*, July 1997, pp. 344–348.

- 
- [109] P. O'Grady, B. Pearlmutter, and S. Rickard, "Survey of sparse and non-sparse methods in source separation", *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 18–33, July 2005.
- [110] I. E. Frank and R. Todeschini, *The data analysis handbook*, Elsevier Science, Amsterdam, September 1994.
- [111] G. Rilling, P. Flandrin, and P. Gonçalvès, "Empirical mode decomposition", <http://perso.ens-lyon.fr/patrick.flandrin/emd.html>, 2007.
- [112] S. Y. Kung, K. S. Arun, and D. V. Bhaskan Rao, "Space-time and singular-value decomposition based approximation methods for the harmonic retrieval problem", *Journal of Optical Society of America*, vol. 73, no. 12, pp. 1799–1811, 1983.
- [113] J. Rosier and Y. Grenier, "Unsupervised classification techniques for multipitch estimation", in *116<sup>th</sup> Convention of the Audio Engineering Society*, Berlin, Germany, May 2004.
- [114] L. De Lathauwer, B. Moor, and J. Vandewalle, "ICA techniques for more sources than sensors", in *HOS Proc. of IEEE Signal Processing Workshop*, June 1999, pp. 121–124.
- [115] M.Z. Ikram, "Blind separation of delayed instantaneous mixtures : A cross-correlation based approach", in *Proc. 2<sup>nd</sup> Symposium on Signal Processing and Information Technology ISSPIT*, Marrakech, Morocco, December 2002.
- [116] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment", *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 882–895, September 2005.
- [117] B. Albouy and Y. Deville, "Méthode temps-fréquence de séparation aveugle de sources basée sur la fonction de cohérence segmentée", in *Actes 19<sup>e</sup> colloque GRETSI sur le Traitement du Signal et des Images*, Paris, France, Septembre 2003.
- [118] B. Albouy and Y. Deville, "Alternative structures and power spectrum criteria for blind segmentation and separation of convolutive speech mixtures", in *Proc. 4<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Nara, Japan, April 2003, pp. 361–366.
- [119] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria", *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. ASSP-33, no. 2, pp. 387–392, April 1985.

- [120] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification", *IEEE Transactions on Signal Processing*, vol. 43, pp. 2982–2993, December 1995.
- [121] A. Aïssa-El-Bey, M. Grebici, K. Abed-Meraim, and A. Belouchrani, "Blind system identification using cross-relation methods : further results and developments", in *Proc. 7<sup>th</sup> International Symposium on Signal Processing and Its Applications ISSPA*, Paris, France, July 2003, vol. 1, pp. 649–652.
- [122] R. Ahmad, A. W. H. Khong, and P. Naylor, "Proportionate frequency domain adaptive algorithms for blind channel identification", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, Toulouse, France, May 2006, vol. 5, pp. 29–32.
- [123] K. Abed-Meraim, Ph. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems", *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 499–511, March 1997.
- [124] A. J. Van Der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems", *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 173–190, January 1997.
- [125] M. S. O'Brien, A. N. Sinclair, and S. M. Kramer, "Recovery of a sparse spike time series by  $\ell_1$  norm deconvolution", *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3353–3365, December 1994.
- [126] W. Qiu and Y. Hua, "Performance comparison of subspace and cross-relation methods for blind channel identification", *Signal Processing - Special Issue on Subspace Methods*, vol. 50, pp. 71–81, April 1996.
- [127] N. Linh-Trung, A. Belouchrani, K. Abed-Meraim, and B. Boashash, "Separating more sources than sensors using time-frequency distributions", *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 17, pp. 2828–2847, 2005.
- [128] B. Barkat and K. Abed-Meraim, "Algorithms for blind components separation and extraction from the time-frequency distribution of their mixture", *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 13, pp. 2025–2033, 2004.
- [129] B. Boashash, Ed., *Time Frequency Signal Analysis and Processing : Method and Applications*, Elsevier, Oxford, 2003.

- 
- [130] B. Barkat and B. Boashash, "A high-resolution quadratic time-frequency distribution for multicomponent signal analysis", *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2232–2239, October 2001.
- [131] S. Mallat, Ed., *Une exploration des signaux en ondelettes*, Les éditions de l'école polytechnique, Palaiseau, France, 2000.
- [132] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations", *IEEE Transactions on Signal Processing*, vol. 46, no. 11, pp. 2888–2897, November 1998.
- [133] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations", *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, November 2001.
- [134] G. F. Boudreaux-Bartels and T. W. Parks, "Time-varying filtering and signal estimation using Wigner distributions", *IEEE Transactions on Acoustic, Speech, Signal Processing*, vol. ASSP-34, no. 3, pp. 442–451, March 1986.
- [135] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time fourier transform", *IEEE Transactions on Acoustic, Speech, Signal Processing*, vol. ASSP-32, no. 2, pp. 236–243, April 1984.
- [136] J. Rosca, T. Gerkmann, and D. C. Balcan, "Statistical inference of missing speech data in the ICA domain", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, May 2006, vol. 5, pp. 617–620.
- [137] F. Abrard and Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources", *Signal Processing*, vol. 85, no. 7, pp. 1389–1403, July 2005.
- [138] J. Rosca, C. Borss, and R. Balan, "Generalized sparse signal mixing model and application to noisy blind source separation", in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, May 2004, vol. 3, pp. 877–880.
- [139] D. Smith, J. Lukasiak, and I. S. Burnett, "An analysis of the limitations of blind signal separation application with speech", *Signal Processing*, vol. 86, no. 2, pp. 353–359, February 2006.
- [140] R. Gribonval, "Sparse decomposition of stereo signals with matching pursuit and application to blind separation of more than two sources from a stereo mixture", in *Proc. International Conference Acoustics, Speech, and*

- Signal Processing ICASSP*, Orlando, Florida, USA, May 2002, vol. 3, pp. 3057–3060.
- [141] M. Zibulevsky, “Sparse source separation with relative Newton method”, in *Proc. 4<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Nara, Japan, April 2003, pp. 897–902.
- [142] N. Murata, S. Ikeda, and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals”, *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, Oct. 2001.
- [143] D. T. Pham, C. Servière, and H. Boumaraf, “Blind separation of convolutive audio mixtures using nonstationarity”, in *Proc. 4<sup>th</sup> Symposium on Independent Component Analysis and Blind Source Separation ICA*, Nara, Japan, April 2003, pp. 981–986.
- [144] M. Kawamoto, K. Matsuoka, and M. Oya, “Blind separation of sources using temporal correlation of the observed signal”, *IEICE Transactions on Fundamentals of Electronics Communications and Computing*, vol. E80A, no. 4, pp. 695–704, April 1997.
- [145] K. Abed-Meraim, Y. Hua, and A. Belouchrani, “A general framework for blind source separation using second order statistics”, in *Proc. 8<sup>th</sup> IEEE Digital Signal Processing Workshop*, Utah, USA, August 1998.
- [146] L. Tong, R. Liu, V.C. Soon, and Y.H. Huang, “Indeterminacy and identifiability of blind identification”, *IEEE Transactions on Circuits and Systems*, vol. 38, no. 5, pp. 499–509, May 1991.
- [147] K. Abed-Meraim, Y. Xiang, and Y. Hua, “Generalized second order identifiability condition and relevant testing technique”, in *Proc. International Conference Acoustics, Speech, and Signal Processing ICASSP*, June 2000, vol. 5, pp. 2989–2992.
- [148] A. M. Zoubir and B. Boashash, “The bootstrap and its application in signal processing”, *IEEE Signal Processing Magazine*, vol. 15, no. 1, pp. 56–76, January 1998.
- [149] B. Efron, “The jackknife, the bootstrap and other resampling plans”, in *CBMS Monograph 38, Society for Industrial and Applied Mathematics (SIAM)*, Philadelphia, USA, 1982.
- [150] R. G. Miller, “The jackknife - a review”, *Biometrika*, vol. 61, no. 1, pp. 1–15, 1974.

- [151] A. Belouchrani, M. G. Amin, and C. Wang, "Interference mitigation in spread spectrum communications using blind source separation", in *Proc. Asilomar Conference on Signals, Systems and Computers*, November 1996, vol. 1, pp. 718–722.
- [152] G. D'Urso, P. Prieur, and C. Vincent, "Blind identification methods applied to electricite de france's civil works and power plants monitoring", in *Proc. of the IEEE Signal Processing Workshop on Higher-Order Statistics*, July 1997, pp. 82–86.
- [153] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.