

Traitement des signaux parcimonieux et applications

Si Mohamed Aziz Sbai

▶ **To cite this version:**

Si Mohamed Aziz Sbai. Traitement des signaux parcimonieux et applications. Traitement du signal et de l'image [eess.SP]. Télécom Bretagne, Université de Bretagne Occidentale, 2012. Français. NNT : . tel-00789538

HAL Id: tel-00789538

https://theses.hal.science/tel-00789538

Submitted on 18 Feb 2013

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.

N° d'ordre : 2012telb0243

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'Université de Bretagne Occidentale

Ecole Doctorale – sicma

Traitement des signaux parcimonieux et applications

Thèse de Doctorat

Mention : Sciences et Technologies de l'Information et de la Communication

Présentée par **Si Mohamed Aziz Sbaï**

Département : Signal et Communications

Laboratoire : Lab-STIC Pôle : CACS/CID

Directeur de thèse : Dominique Pastor

Soutenance prévue le 20 Novembre 2012

Jury :

M. Yannick Berthoumieu, Professeur, IMS UMR 5218, Bordeaux (Rapporteur)
M. Yannick Deville, Professeur, Université Paul Sabatier Toulouse 3, Toulouse (Rapporteur)
M. Abdourramane Atto, Maître de conférences, Université de Savoie, Annecy (Examinateur)
M. Stéphane Azou, Professeur, ENIB, Brest (Examinateur)
M. Karim Abed-Meraim, Professeur, Université d'Orléans, Orléans (Examinateur)
M. Rémi Gribonval, Directeur de Recherche, INRIA, Rennes (Examinateur)
M. Dominique Pastor, Professeur, Télécom Bretagne, Brest (Directeur de thèse)
M. Abdeldjalil Aïssa El Bey, Maître de conférences, Télécom Bretagne, Brest (Encadrant)

Résumé

Quel que soit le domaine d'application, il est nécessaire de tirer profit de toute l'information a priori dans le but d'optimiser les résultats ou parfois même de manière à rendre un problème soluble. Dans ce contexte, la notion de parcimonie a émergé comme un a priori fondamental ces dernières années. On dit qu'un signal est parcimonieux dans une base s'il peut être décrit par un faible nombre de coefficients non nuls dans cette base.

L'objet de cette thèse est l'étude de nouveaux apports de l'hypothèse de parcimonie au traitement du signal. Deux domaines d'applications sont considérés. Outre l'utilisation de la parcimonie, ces deux domaines ont en commun la résolution de problèmes inverses sous-déterminés. Le premier concerne la séparation de sources. Dans ce domaine, la parcimonie a conduit au développement de différentes méthodes de séparation de sources. Les performances de ces méthodes sont sensibles au choix de certains paramètres, habituellement choisis de manière empirique. Dans cette thèse, on propose un formalisme statistique qui permet de réduire le nombre de ces paramètres, tout en préservant la qualité de la séparation. Le second domaine d'application étudié est l'acquisition compressée des signaux à alphabet fini. Une telle acquisition compressée permet de réduire la dimension des signaux à alphabet fini, tout en gardant l'information nécessaire à leur reconstruction. Une formalisation du problème permet de le relier à celui de la reconstruction des signaux parcimonieux à partir de mesures incomplètes.

Cette thèse est donc une exploration de nouvelles problématiques où l'intégration de la parcimonie conduit à de bonnes performances.

Table des matières

Introduction générale **1**

1 Parcimonie **7**

1.1 La parcimonie : Un a priori à succés 7

1.1.1 Bref historique 7

1.1.2 Parcimonie en traitement du signal 8

1.1.3 Mesures de parcimonie 9

1.1.4 Et la pratique dans tout ça ? 11

1.2 Intégration de la parcimonie 12

1.2.1 Optimisation global 12

1.2.2 Optimisation local et Algorithmes gloutons 16

1.3 Apports 16

1.3.1 Approximation et Compression 16

1.3.2 Débruitage 17

I Tests statistiques pour les méthodes de séparation de sources fondées sur la parcimonie des sources **21**

2 La séparation de sources **25**

2.1 Introduction 25

2.2 L'approche d'analyse en composantes indépendantes 26

2.3 L'approche de factorisation en matrices non-négatives 27

2.4 Méthodes SAS fondées sur la parcimonie 28
2.4.1 Modèle 28
2.4.2 Principe 28
2.4.3 Estimation de la matrice de mélange 29
2.4.4 Estimation des signaux sources 35

3 Tests statistiques pour les méthodes SAS basées sur la parcimonie des sources 39

3.1 Sélection multisource 40
3.1.1 Formulation statistique 40
3.1.2 Hypothèse de parcimonie du modèle 42
3.1.3 Test à seuil 43
3.1.4 Estimation de l'écart type 44
3.2 Sélection autosource 46
3.2.1 Le cas de signaux avec un faible taux de recouvrement 47
3.2.2 Le cas de signaux avec un fort taux de recouvrement 47

4 Résultats de simulations 49

4.1 Données de simulations 49
4.2 La méthode SUBSS 51
4.3 La minimisation ℓ_1 55
4.4 La méthode TIFROM 57
4.5 Masquage binaire (reconstruction DUET) 58
4.6 Mélange convolutif 59

5 Conclusions 63

II Acquisition compressée des signaux à alphabet fini 65

6 Introduction 69

6.1 Acquisition compressée 69

6.2 Formalisme d'acquisition compressée de signaux à alphabet fini 71
6.3 Cas particulier 72
6.4 Notations 74

7 Solvabilité et méthodes de reconstruction 75

7.1 Conditions de solvabilité 75
7.1.1 Hypothèse de solvabilité 75
7.1.2 Condition suffisante 76
7.2 Méthodes de reconstruction 77
7.2.1 Approche de reconstruction par régularisation : ℓ_1-additive . . . 78
7.2.2 Approche de reconstruction par transformation parcimonieuse : ℓ_1-indicateur 80
7.3 Discussions 85
7.4 Conclusion 87

8 Expérimentation numérique 89

8.1 Simulations 89
8.2 Résultats 90
8.2.1 Procédure de reconstruction par régularisation 90
8.2.2 Procédure de reconstruction par transformation parcimonieuse . 91
8.3 Séparation d'images Binaires 92
8.4 Conclusion 94

Conclusion générale 94

9 Conclusion générale et perspectives 97

9.1 Conclusion générale 97
9.2 Perspectives 98

A Séparation de sources 101

A.1 Estimation linéaire 101

B Acquisition compressée de signaux à alphabet fini **103**

B.1 Quelques généralités sur les lois binomiales 103

B.2 Preuve du Théorème 7.5 104

Bibliographie **107**

Liste des figures

1.1 Exemples de boules unités associées aux (quasi)-normes ℓ_q. ℓ_∞ est égale au max des valeurs absolues 10

1.2 Représentation temps-fréquence d'un signal audio. Plusieurs coefficients sont de faible amplitude, voire nulle. 12

1.3 (a) Exemple d'image régulière par morceaux (b) Histogramme des coefficients en ondelettes. La représentation de l'image en ondelette est parcimonieuse. 13

2.1 Organigramme des algorithmes de SAS procédants en deux étapes dans le domaine temps-fréquence. 30

2.2 Procédure itérative de traitement des sources cachées dans un mélange. (a) Exemple de recouvrement temps-fréquence dans un mélange de trois sources. Les sources 2 et 3 ne sont pas visibles. En supprimant la contribution de la source 1, nous obtenons un nouveau mélange. (b) Recouvrement temps-fréquence dans le nouveau mélange où toutes les sources sont visibles. 31

2.3 Diagramme de dispersion dans le domaine de parcimonie des sources. . 32

3.1 Organigramme des algorithmes de SAS proposés à deux étapes. 41

3.2 (a) Représentation temps-fréquence d'un mélange de signaux audio. Plusieurs coefficients sont de faible amplitude, voire nulle. (b) Représentation temps-fréquence d'un mélange de signaux audio bruité. Les petites valeurs sont masquées par le bruit. Les grandes valeurs restent visible. Ces valeurs ne sont pas énormément affectées par le bruit pour peu que le rapport signal à bruit ne soit pas trop élevé. La proportion de ces coefficients est inférieure ou égale à un demi. 42

4.1 (a)-(d) montrent les formes d'onde des signaux sources originaux dans le domaine temporel, (e)-(h) affichent les spectrogrammes de ces signaux sources dans le domaine temps-fréquence. 50

4.2 Spectrogramme d'un mélange des quatre sources de la figure 4.1, pour un RSB égal à 10dB). 51

4.3 Résultats de simulations : (a)-(d) montrent les formes d'onde des signaux sources reconstruits par l'algorithme SUBSS modifié dans le cas d'un RSB = 45dB, (e)-(h) montrent les formes d'onde des signaux sources reconstruits par l'algorithme SUBSS modifié dans le cas d'un RSB = 10dB. 52

4.4 EQMN en fonction du RSB : Comparaison des performances des algorithmes SUBSS standard, SUBSS modifié avec et sans débruitage, et de l'algorithme SUBSS modifié avec l'estimation MAD au lieu du DATE et sans débruitage. 53

4.5 EQMN en fonction du nombre de sources : Comparaison des performances de l'algorithme SUBSS standard et de l'algorithme SUBSS modifié sans débruitage. 53

4.6 EQMN en fonction du nombre de sources : Comparaison des performances de l'algorithme SUBSS standard et de l'algorithme SUBSS modifié sans débruitage. 54

4.7 Spectrogramme des quatre signaux sources à fréquence modulée. 55

4.8 Spectrogramme d'un mélange des quatre sources de la figure 4.7, pour un RSB égal à 10dB). 56

4.9 EQMN en fonction du RSB : Comparaison des performances de l'algorithme SUBSS standard et de l'algorithme SUBSS modifié sans débruitage pour les signaux à fréquence modulée. 56

4.10 EQMN en fonction du RSB : Comparaison des performances de l'algorithme fondée sur la minimisation (2.9) et les algorithmes cherchants une solution du problème (4.2) après sélection des points multisources, lorsque l'écart type du bruit est connue ou estimée via le DATE ou le MAD. 58

4.11 EQMN en fonction du RSB : Comparaison des performances de la méthode TIFROM standard , du SNT-TIFROM et du SNT-TIFROM modifié. 59

4.12 EQMN en fonction du RSB : Comparaison des performances de l'approche de reconstruction DUET et l'approche de reconstruction DUET modifiée. 60

4.13 EQMN en fonction du RSB : Comparaison des performances de la méthode SUBSS convolutive standard et la SUBSS convolutive modifiée. Chaque mélange est une somme de signaux sources filtrées dans lequel chaque filtre est un RIF généré aléatoirement d'ordre 4 61

8.1 Probabilité de reconstruction exacte de $(P_{\mathcal{F},1})$ pour $p = 2$ 90

8.2 Probabilité de reconstruction exacte de $(P_{\mathcal{F},1})$ pour $p = 4$ 91

8.3 Probabilité de reconstruction de l'approche par transformation parcimonieuse pour $p = 2$ 92

8.4 Probabilité de reconstruction de l'approche par transformation parcimonieuse pour $p = 4$ 93

8.5 Images originales. 94

8.6 Combinaison linéaires des images (a), (b), (c) et (d) de la figure 8.5. . . 95

8.7 Images reconstruites. 96

Introduction générale

Motivations

Ces dernières années, la notion de parcimonie a fait l'objet d'un grand intérêt de la part de la communauté du traitement du signal. La parcimonie a émergé comme un *a priori* fondamental. Aujourd'hui, on ne compte plus le nombre de publications citant le terme parcimonie, des sessions "parcimonie" ont vu le jour dans les conférences les plus importantes, et des conférences y sont même dédiées (SPARS). Incontestablement, la parcimonie est devenue une notion incontournable en traitement du signal.

Le but visé à travers cette thèse est d'explorer de nouveaux apports de la parcimonie en traitement du signal. Dans ce cadre, les travaux développés se décomposent principalement en deux parties :

La première partie concerne la contribution de la parcimonie statistique à la séparation de sources via des tests statistiques. Ces travaux s'inscrivent dans une suite logique des travaux réalisés par mes encadrants. En effet, d'une part, Aïssa-El-Bey et al. ont présenté dans [1] la méthode de projection en sous-espace. Cette méthode basée sur la parcimonie des sources dans le domaine temps-fréquence donne de bonnes performances de séparation de sources dans le cas sous-déterminé. Cependant, la méthode de projection en sous-espace est sensible au choix des différents seuils auxquels elle fait appel. D'autre part, dans [2], D. Pastor a introduit une notion de parcimonie de modèle. Cette notion formelle de parcimonie se rattache à un certain nombre de résultats théoriques énoncés en statistiques non-paramétriques et estimation robuste publiés depuis 2002. La nouveauté des résultats présentés dans cette partie est que cette notion contribue à la robustesse et/ou à l'amélioration des performances des méthodes de séparation de sources basées sur la parcimonie, notamment à la méthode de projection en sous-espace.

La seconde partie propose un nouveau domaine d'application de la parcimonie qui est l'acquisition compressée des signaux à alphabet fini. En d'autres termes, la recherche

de procédure (encodage et décodage) de réduction de dimension des signaux à alphabet fini, tout en gardant l'information nécessaire à leur reconstruction. L'étude des signaux à alphabet fini est motivée par le fait que ce type de signaux est très utilisé, entre autres, en communications numériques. Le département signal et communications, au sein du quel j'ai mené ma thèse, est spécialisé dans le traitement du signal, le codage de l'information et les communications numériques. Ainsi, comme on peut le constater le sujet traité dans cette partie est fortement corrélé aux différentes spécialisations de mon département d'accueil. Dans cette partie, l'idée novatrice consiste à voir les signaux à alphabet fini comme des signaux parcimonieux. En effet, les signaux parcimonieux sont des signaux à faible degré de liberté, qui sont plongés dans des espaces de grandes dimensions. C'est le cas des signaux à alphabet fini. Nous donnerons une formalisation mathématique qui relie ces deux types de signaux.

Ces deux parties ont en commun la résolution de problèmes inverses sous-déterminés en utilisant la parcimonie.

Cadre et organisation de l'étude

Le travail présenté dans ce manuscrit se décompose en deux parties. Chaque partie est associée à un domaine du traitement du signal et regroupe plusieurs chapitres. Ainsi, à la suite de cette introduction, le manuscrit s'organise comme suit :

- Un chapitre sur la parcimonie ;
- Une partie dédiée à la séparation aveugle de sources ;
- Une partie associée à l'acquisition compressée des signaux à alphabet fini ;
- Un chapitre de conclusions et perspectives.

La parcimonie : Un a priori à succès

Il est facile d'analyser les raisons pour lesquelles la parcimonie a vu un tel engouement de la part de la communauté du traitement du signal : tout d'abord, l'intégration de la parcimonie donne naissance à des techniques plus performantes dans différents domaines tels que l'estimation [3], le débruitage [4], la compression [5] ou même l'acqui-

sition [6]. Outre cet apport de la parcimonie, la motivation principale de la parcimonie vient de l'observation. La parcimonie est présente dans différents domaines : Les signaux sonores ont une décomposition parcimonieuse en temps-fréquences, la plupart des images ont une décomposition parcimonieuse sur une base d'ondelettes, on rencontre la notion de parcimonie statistique en traitement du signal sonar, radar ou en contre-mesure...

Afin de mieux comprendre l'importance de la parcimonie en théorie comme en pratique, nous commençons dans ce chapitre par définir les différentes notions de parcimonie, les critère de mesure de la parcimonie, puis nous exposerons l'intégration et l'apport de celles-ci dans différents domaines d'applications.

Tests statistiques pour la séparation aveugle de sources

La séparation aveugle de sources (SAS) est une discipline en plein essor qui trouve application dans une variété de domaines, comme le traitement audio [7, 8], les télécommunications [9, 10], l'ingénierie biomédicale [11, 12] ou encore le traitement radar [13]. Différentes méthodes permettent d'aborder ce problème. Cependant, depuis une dizaine d'années, les méthodes basées sur la parcimonie des sources ont fait l'objet d'un grand intérêt. Toutes les méthodes étudiées dans cette thèse appartiennent à cette famille. Ces méthodes procèdent en deux étapes. Dans un premier temps, ces méthodes trouvent des zones temps-fréquence où une seule source est présente afin d'estimer la matrice de mélange. Puis, dans un second temps, estiment les sources. A l'image de la méthode de projection en sous-espaces [1], les performances de séparation des méthodes basées sur la parcimonie des sources dépendent du choix des paramètres. L'utilisation de paramètres empiriques constitue la limitation majeure de ces techniques.

Le premier chapitre de cette partie concerne l'état de l'art des méthodes de séparation de sources. La séparation de sources est un domaine très vaste en traitement du signal qui a été traitée sous différents angles. Intentionnellement, nous limiterons l'étude bibliographique aux approches d'analyse en composantes indépendantes, de la factorisation en matrices non négatives, et enfin, de façon plus détaillée, les approches de séparation de sources basées sur la parcimonie.

Puis, dans un second chapitre, nous proposons deux tests statistiques afin d'améliorer les performances de séparation et de réduire le nombre de paramètres empiriques utilisés par les méthodes SAS basées sur la parcimonie.

Le troisième chapitre concerne les aspects expérimentaux d'évaluation des techniques de séparation de sources. Nous appliquons les tests statistiques décrits dans le

chapitre précédent à plusieurs méthodes SAS classiques, notamment à la méthode de projection en sous-espaces. Nous montrons que nos algorithmes statistiques réduisent le nombre de paramètres empiriques et permettent dans certain cas d'effectuer une meilleure séparation.

Enfin, dans un quatrième chapitre, nous terminerons cette partie par une conclusion.

Les travaux présentés dans cette partie ont donné lieu à une publication dans la revue *EURASIP journal on Advances in Signal processing* [14] et un article dans la conférence *ICASSP* [15].

Acquisition compressée de signaux à alphabet fini

L'acquisition compressée (CS pour compressed sensing) est une nouvelle technique d'acquisition et de compression qui exploite la parcimonie des signaux. Développée conjointement par Donoho [5] et Candès, Romberg et Tao [6], l'objectif principal du CS est de réduire la dimension d'un signal de façon non-adaptative, tout en gardant l'information nécessaire à sa reconstruction. Ceci est réalisé en projetant les signaux sur un petit nombre de vecteurs aléatoires choisis de façon non-adaptative. Nous nous intéressons dans cette partie à l'acquisition compressé des signaux à alphabet fini.

Après une brève analyse de l'état de l'art de l'acquisition compressée et une formalisation du problème d'acquisition compressée des signaux à alphabet fini, le premier chapitre positionne notre travail par rapport à l'état de l'art, notamment par rapport à l'approche proposée par Mangasarian [16] et introduit les différentes notations utilisées dans cette partie.

Dans le second chapitre, d'une part, nous établissons une condition suffisante qui permet de rendre le problème soluble. Cette condition étend celle de Mangasarian au cas d'alphabet fini quelconque. D'autre part, nous présentons deux approches de reconstruction. Ces deux méthodes ont en commun une relaxation convexe de la norme ℓ_0 en norme ℓ_1. La première méthode est fondée sur la régularisation, où l'on introduit une fonction objective qui évalue le critère basé sur l'alphabet fini. La seconde méthode est fondée sur la parcimonie. Nous dérivons dans ce cas une formulation analogue à celle de la reconstruction de signaux parcimonieux à partir de mesures incomplètes. Nous donnerons les conditions nécessaires et suffisantes à la reconstruction de signal à alphabet fini pour cette approche.

Le troisième chapitre traite de l'évaluation numérique des deux approches. Les résultats expérimentaux montrent alors qu'à partir d'un certain nombre de mesures la

reconstruction des signaux est parfaite pour les deux approches.

Une partie du contenu de cette seconde partie a fait l'objet d'une publication dans la conférence *ISSPA* 2012 [17]. Une première version de cet article a été présenté à *SPARS* 2011 [18]. Une version détaillée de l'article sera soumise à la revue *Elsevier Digital Signal Processing* [19].

Conclusions et perspectives

Ce chapitre est consacré à un récapitulatif des contributions et des résultats obtenus dans le cadre de ce travail, ainsi qu'une présentation des différentes pistes de recherche futurs qu'on peut envisager ou même, qu'on a partiellement envisagé.

CHAPITRE 1 Parcimonie

Comme nous l'avons signalé dans l'introduction générale, la parcimonie est devenue une notion incontournable en traitement du signal. Dans ce chapitre, nous analysons les raisons pour lesquelles la parcimonie a vu un tel engouement de la part de la communauté du traitement du signal. Après avoir défini les différentes notions liées à la parcimonie, nous exposerons l'intégration et l'apport de celle-ci dans différents domaines d'application.

1.1 La parcimonie : Un a priori à succés

Quels que soient le domaine d'application et la technique utilisée, l'optimisation des résultats nécessite de tirer profit de toute l'information a priori. Ces dernières années, la parcimonie a émergé comme un a priori fondamental.

1.1.1 Bref historique

Le dictionnaire Larousse définit le terme parcimonie comme une notion d'économie ou d'épargne. Ainsi, il est dur d'essayer de retracer historiquement la notion de parcimonie car la parcimonie ou plus précisément le principe de parcimonie est un principe de raisonnement philosophique avant tout. Ce raisonnement consiste à n'utiliser que le minimum d'hypothèses ou de causes élémentaires pour expliquer/décrire un phénomène observé. Ce principe est également connu sous le nom du rasoir d'Ockham (Wikipédia). Le rasoir d'Ockham est nommé d'après William (de) Ockham (un village près de East Horsley dans le Surrey). Il est généralement formulé par : "Les multiples ne doivent pas être utilisés sans nécessité" bien que cette formule soit connue avant d'Ockham. Il s'agit d'une métrique de simplicité pour l'élaboration de théories [20]. En sciences, la parcimonie intervient dans un grand nombre de domaines (mathématique, physique, sciences naturelles...). Par exemple en systématique [21], ce principe privilé-

gie la séquence avec le minimum de changements évolutifs pour expliquer des relations phylogéniques. En sismologie, dans le milieu des années 80, Santosa et Symes [22] ont étudié la reconstruction de train parcimonieux de Dirac. En statistique, la robustesse par rapport aux outliers peut s'interpréter par la présence d'un bruit parcimonieux dans le modèle. Ainsi, comme nous le verrons dans la première partie, la notion de parcimonie et les statistiques robustes sont étroitement liées.

1.1.2 Parcimonie en traitement du signal

La notion de parcimonie comme nous l'avons dit est liée à la notion d'économie. Ainsi, une représentation d'un vecteur réel est dite parcimonieuse si cette représentation est économique. En d'autres termes, il existe une façon plus économique de décrire le vecteur que de donner la valeur de tous ses éléments. En traitement du signal, cela signifie que la plupart des coefficients sont nuls, et seuls quelques coefficients ont des valeurs non-nuls. Ce type de parcimonie porte le nom de parcimonie stricte. La parcimonie au sens large traduit le fait que la majorité des coefficients a une valeur faible, tandis que le reste prend des valeurs assez grandes.

Contrairement à ces deux notions de parcimonie, large et stricte, qui caractérisent des signaux déterministes et qui ne constituent pas une définition absolue en soit, mais plutôt une propriété nécessitant une mesure d'évaluation, D. Pastor a introduit dans [2] une notion de parcimonie pour des vecteurs aléatoires. De plus, cette notion de parcimonie n'est pas liée à un processus donné mais plutôt à un modèle aléatoire. Le modèle aléatoire est le suivant : l'observation est une réalisation d'un vecteur aléatoire qui est soit du bruit, soit la somme d'un signal utile plus du bruit. Ainsi, dans la suite de ce manuscrit, nous dirons que ce modèle aléatoire est parcimonieux si le signal est plus souvent absent que présent et que, présent, il l'est avec une amplitude minimale garantie. Mathématiquement, étant données plusieurs observations, notons $\boldsymbol{Y}$, $\boldsymbol{X}$ et $\boldsymbol{W}$ les vecteurs aléatoires réels de dimension d représentant respectivement l'observation, le signal utile et le bruit présent dans le modèle. Le modèle est alors le suivant :

$$\begin{cases} \mathcal{H}_0 : & \boldsymbol{Y} = \boldsymbol{W} \\ \mathcal{H}_1 : & \boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{W} \end{cases} \tag{1.1}$$

Ce modèle est dit parcimonieux si :

1. La probabilité d'occurrence de l'hypothèse $\mathcal{H}_1$ est inférieure ou égal à un demi, c'est à dire $\mathbb{P}(\mathcal{H}_1) \leq 1/2$.

2. Il existe une certaine valeur α telle que $|\boldsymbol{X}| > \alpha$ presque sûrement.

Comme on peut le constater, la séparation entre la notion de signal utile et l'absence ou la présence de ce signal (valeur nulle ou non nulle) constitue la différence fondamentale, outre l'aspect aléatoire ou déterministe, entre la parcimonie du modèle [2] et la parcimonie d'un signal utile (strict ou large). Pour s'en rendre compte, considérons le cas où $d = 1$ et que l'on observe un vecteur de n réalisations de la variable aléatoire $\boldsymbol{Y}$, noté $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^T$, on a alors

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w}. \tag{1.2}$$

Quelque soit la distribution de la variable aléatoire $\boldsymbol{X}$, le vecteur $\boldsymbol{x}$ est strictement parcimonieux. Nous retrouvons alors le modèle de débruitage (d'estimation) des signaux parcimonieux [23]. Dans ce dernier, $\boldsymbol{x}$ représente le signal utile, alors que dans le modèle (1.1), le signal utile représente les composantes de $\boldsymbol{x}$ non nulles. Cette différence dans la définition du signal utile découle de la procédure d'estimation de $\boldsymbol{x}$. En effet, afin d'estimer $\boldsymbol{x}$, deux approches sont possibles : l'approche qui consiste à estimer directement $\boldsymbol{x}$, sous l'hypothèse que $\boldsymbol{x}$ est parcimonieux [23] et l'approche qui consiste à détecter les coefficients non nuls, sous l'hypothèse de parcimonie de présence (du modèle), puis à estimer ces coefficients via une procédure d'estimation [24].

1.1.3 Mesures de parcimonie

Etant donnée la définition de la parcimonie stricte, il est clair que la norme ℓ_0 est la mesure intuitive pour quantifier cette parcimonie. La norme ℓ_0 consiste à compter le nombre d'éléments non nuls d'un vecteur, c'est à dire, pour tout $\boldsymbol{x}$,

$$\|\boldsymbol{x}\|_0 = \#\left\{i : \boldsymbol{x}_i \neq 0\right\}.$$

Il est à noter que la fonction ℓ_0 ne constitue pas une norme. En effet, la propriété d'homogénéité n'est pas vérifiée : pour tout $\lambda \neq 0, \|\lambda x\|_0 = \|\boldsymbol{x}\|_0 \neq |\lambda| \|\boldsymbol{x}\|_0$. Cependant, nous conserverons dans la suite cet abus de langage, afin d'être cohérent avec la littérature.

Mis à part quelques exceptions, notamment la reconstruction des signaux à alphabet fini (thème abordé dans la seconde partie du manuscrit), les différents signaux rencontrés ne contiennent pas forcément plusieurs coefficients nuls. La compressibilité,

qui est une relaxation de la notion de parcimonie au sens stricte, est beaucoup plus appropriée.

Soit $\boldsymbol{x} \in \mathbb{R}^N$, classons les coefficients $\{|\boldsymbol{x}_i|\}_{1\leq i\leq N}$ par ordre décroissant. On note $|\boldsymbol{x}|_{(r)}$ le coefficient de rang r et $N_{\epsilon,\boldsymbol{x}}$ le nombre de coefficients dans $\boldsymbol{x}$ de valeur supérieure à ϵ, c'est à dire $N_{\epsilon,\boldsymbol{x}} = \#\{i : \boldsymbol{x}_i \geq \epsilon\}$. La notion de compressibilité est liée à la vitesse de décroissance des coefficients $\left\{|\boldsymbol{x}|_{(r)}\right\}_{1\leq r\leq N}$. Cette décroissance peut être évaluée à partir de la norme ℓ_1 et des quasi-normes ℓ_q où $0 < q < 1$:

$$\|\boldsymbol{x}\|_q = \left(\sum_{i=1}^{N} |x_i|^q\right)^{1/q}.$$

La figure 1.1 affiche différents exemples de boules unités associées aux (quasi)-normes ℓ_q.

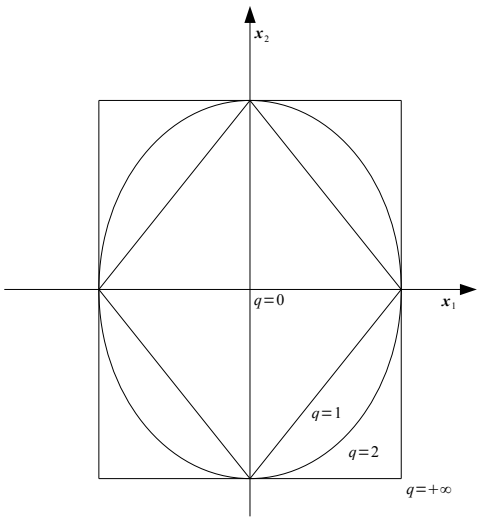


Figure 1.1 — Exemples de boules unités associées aux (quasi)-normes ℓ_q. ℓ_∞ est égale au max des valeurs absolues

Les quasi-normes $\omega\ell_q$ permettent aussi d'évaluer la décroissance des coefficients, elles sont définies par

$$\|\boldsymbol{x}\|_{\omega\ell_q} = \sup_{\epsilon>0} \epsilon\, N_{\epsilon,x}^{1/q}.$$

En effet, la définition de la quasi-norme $\omega\ell_q$ est équivalente à

$$\|\boldsymbol{x}\|_{\omega\ell_q} = \sup_{k\geq 1} k^{1/q} \, |\boldsymbol{x}|_{(k)} \, ,$$

ce qui implique que pour tout $k \in \{1, \cdots, N\}$,

$$|\boldsymbol{x}|_{(k)} \leq \|x\|_{\omega\ell_q} \, k^{-1/q} \leq \|\boldsymbol{x}\|_{\ell_q} \, k^{-1/q}.$$

On en déduit que $\left\{|\boldsymbol{x}|_{(r)}\right\}_{1\leq r\leq N}$ décroit au moins comme $k^{-1/q}$. Plus q est petit, plus rapide est la décroissance. Enfin, signalons que d'autres mesures de parcimonie ont été traitées dans la littérature. Dans [25], plusieurs mesures alternatives de parcimonie sont décrites. Ces mesures imitent le comportement de la norme ℓ_0 avec des propriétés différentes. Le kurtosis, les fonctions log et tanh sont des exemples de mesures de parcimonie. Dans [26], les auteurs proposent l'utilisation de l'index de Gini pour mesurer la parcimonie. Dans [27], Coifman et Wicherhauser ont proposé de mesurer la parcimonie en calculant l'entropie des coefficients. N'utilisant pas ces mesures de parcimonie dans la suite, nous ne détaillons pas celles-ci. Toutefois, signalons que le choix de la mesure de parcimonie n'est pas un détail mineur, car celle-ci peut induire des répercussions directes sur la structure du signal.

La norme ℓ_1 est la mesure de parcimonie la plus utilisée. Ceci est dû sans doute, d'une part, à l'essor qu'a connu le domaine d'optimisation convexe et plus précisément celui de la programmation linéaire, pour lequel il existe des algorithmes efficaces. D'autre part, à l'efficacité de la norme ℓ_1 à approcher la solution la plus parcimonieuse en terme de la norme ℓ_0. Nous détaillerons dans la suite ces points.

1.1.4 Et la pratique dans tout ça ?

La parcimonie comme tout a priori est utile lorsqu'on fait face à une indétermination ou une incertitude sur la solution d'un problème ou la cause d'un effet. Il est clair que l'on peut définir n'importe quel a priori et l'intégration de cet a priori nous permettra de réduire l'ensemble des solutions possibles. Cependant, au delà de la nécessité de réduire cet ensemble, il faut que cet a priori soit raisonnable et en adéquation avec la réalité. Si tel n'était pas le cas, on se contenterait toujours de rechercher la solution d'énergie minimale, car celle ci est à la fois très simple à calculer et à la fois unique. C'est là la motivation principale de l'utilisation de la parcimonie. La parcimonie est présente partout et constitue un a priori raisonnable pour différents types de signaux. Les signaux audio ont une décomposition parcimonieuse dans le domaine

temps-fréquence. La figure 1.2 montre une telle représentation pour un signal échantillonné à une fréquence de 8 KHz. Comme on peut le constater, plusieurs coefficients sont de faible amplitude, voire nulle. Ce résultat n'est pas spécifique au morceau choisi mais caractérise la grande majorité des signaux audio.

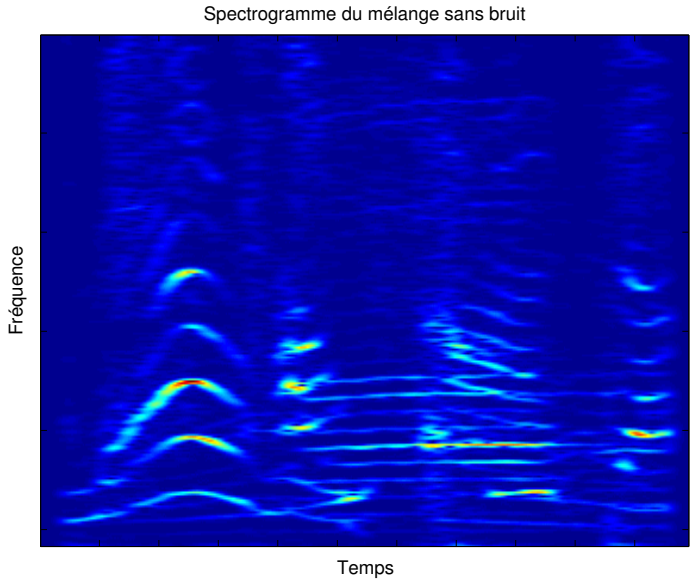


Figure 1.2 — Représentation temps-fréquence d'un signal audio. Plusieurs coefficients sont de faible amplitude, voire nulle.

Le même type de résultat est obtenu en décomposant une image courante dans une base d'ondelettes. Comme on peut le voir sur la figure 1.3, la transformation en ondelettes de l'image est parcimonieuse.

1.2 Intégration de la parcimonie

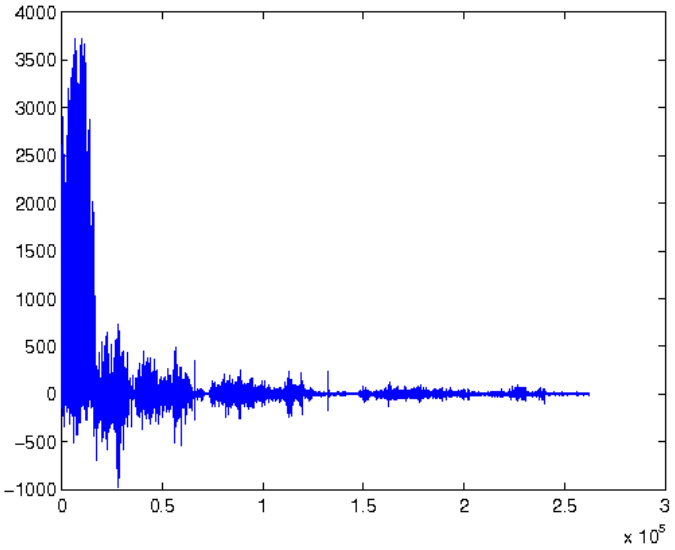
L'intégration de la parcimonie peut se faire de différentes façons.

1.2.1 Optimisation global

Soit $\boldsymbol{A}$ une matrice de dimension $n \times N$ à valeurs réelles, avec $n < N$, on définit le système d'équations linéaires sous-déterminé $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Ce système a moins d'équations que d'inconnues, et donc soit il a pas de solutions, dans le cas où $\boldsymbol{y}$ n'appartient pas à $Im\{\boldsymbol{A}\}$ l'ensemble image de la matrice $\boldsymbol{A}$, soit une infinité de solutions (l'ensemble des solutions forme alors un espace affine). Le cas où le système n'admet pas de solutions



(a)



(b)

Figure 1.3 — (a) Exemple d'image régulière par morceaux (b) Histogramme des coefficients en ondelettes. La représentation de l'image en ondelette est parcimonieuse.

n'est pas pertinent. Afin de contourner ce cas, nous supposerons tout le long de ce manuscrit que la matrice $\boldsymbol{A}$ est de rang plein, c'est-à-dire $Im\{\boldsymbol{A}\} = \mathbb{R}^n$.
Dans de nombreuses applications d'intérêt pratique, nous cherchons à résoudre un problème linéaire inverse où l'on tente de reconstruire un objet $\boldsymbol{f}$ (image, signal, etc.) à partir d'observations incomplètes de ses effets $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{f}$. Le fait qu'il existe une infinité

de solutions pour le système $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ se présente comme un problème majeur. Afin de réduire l'espace de recherche à éventuellement une unique solution $\boldsymbol{f}$, des informations a priori, telles que la parcimonie, sont nécessaires.
La régularisation est un moyen permettant l'intégration d'un a priori dans la recherche de solution. Le but est de privilégier une solution particulière, en introduisant une fonction de pénalité $g(\boldsymbol{x})$, qui favorise certaines propriétés désirables. Cette procédure limite la recherche de solutions aux seules solutions dotées de ces informations a priori, et donc permet de réduire l'espace de recherche.
Mathématiquement, on peut formuler le problème comme un problème d'optimisation générale (P)

$$(P) \ : \qquad \min_{\boldsymbol{x}} \ g(\boldsymbol{x}) \ \text{ sous la contrainte } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

Le choix de la fonction $g(\boldsymbol{x})$ n'est pas un détail mineur, car celle-ci a des répercussions directes sur le choix de la solution.
La norme euclidienne au carré $\|\boldsymbol{x}\|_2^2$ fut durant plusieurs années le choix prédominent de la fonction de pénalité $g(\boldsymbol{x})$. La résolution du problème (P_2) qui en résulte,

$$(P_2) \ : \qquad \min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_2^2 \ \text{ sous la contrainte } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

est connue sous différents noms selon les communautés : méthode des Frames [28], méthode des moindres carrés, rétroprojection filtrée... Cette méthode sélectionne parmi toutes les solutions de $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, celle ayant la norme ℓ_2 minimale. Cette solution unique est appelée la solution pseudo-inverse ou la solution d'énergie minimale, elle est calculée directement par :

$$\hat{\boldsymbol{x}} = \boldsymbol{A}^{+}\boldsymbol{y} = \boldsymbol{A}^{T}(\boldsymbol{A}\boldsymbol{A}^{T})^{-1}\boldsymbol{y}$$

où $\boldsymbol{A}^{+}$ est l'inverse généralisé de $\boldsymbol{A}$ [29]. L'unicité de la solution et la simplicité de son calcul ont sans doute contribué à la popularité du choix de la norme euclidienne au carré comme fonction de pénalité. Néanmoins, cette stratégie ne donne toujours pas de très bonnes performances, et peut être difficilement utilisé pour différentes applications (imagerie médical, séismologie, etc.).
Une fonction de pénalité alternative est la norme ℓ_0, cherchant ainsi les solutions (éventuellement une unique solution) les plus parcimonieuses. Ce problème peut être formulé comme une minimisation (P_0)

$$(P_0) \ : \qquad \min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_0 \ \text{ sous la contrainte } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

Ce problème se pose dans de nombreuses applications telle l'acquisition compressée [5], la tomographie [30], la classification des signaux [31]. Malheureusement, la recherche d'une solution du problème (P_0) est NP-difficile en générale [32] et ne peut être résolue en pratique pour des valeurs de N assez grandes. En effet, la résolution du problème (P_0) exige essentiellement une stratégie de recherche exhaustive, or celle-ci a une complexité algorithmique exponentielle en N. Une alternative est fournie par relaxation du problème. Il s'agit alors de "remplacer" la norme ℓ_0 par une norme favorisant la parcimonie et rendant le problème soluble (algorithme de complexité polynomiale). C'est là où intervient les mesures de parcimonie. Dans [33], Chen et al. ont introduit le *basis pursuit* qui consiste à résoudre le problème suivant :

$$(P_1) \ : \qquad \min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \ \text{ sous la contrainte } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

D'autres relaxations ont été étudiées dans la littérature, citons par exemple FOCUSS [34, 35], ℓ_1-pondéré [36, 37], ℓ_p avec $0 < p < 1$ [38, 39]. Toutefois, l'avantage du problème de minimisation (P_1), et la raison pour laquelle on s'intéressera à ce type de relaxation dans la suite du manuscrit, est qu'il est convexe d'une part, et d'autre part qu'il est équivalent au programme linéaire suivant :

$$\min_{(\boldsymbol{u},\boldsymbol{v})} \mathbf{1}^T\boldsymbol{v} \ \text{ sous les contraintes } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{u} \text{ et } \ -\boldsymbol{v} \leq \boldsymbol{u} \leq \boldsymbol{v}$$

La résolution de ce programme peut être effectuée efficacement en utilisant soit l'algorithme du simplexe [40], soit la méthode du point intérieur [41, 40]. On n'entrera pas ici dans le détail d'implémentations et de résolution numérique de cet algorithme. Le lecteur intéressé par ces aspects pourra se référer à [42, 40]. On utilisera en pratique une implémentation dans le paquet logiciel *cvx* développé sous MATLAB [43].
En présence de bruit dans le modèle, il ne s'agit plus de chercher la solution la plus parcimonieuse de $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ mais une approximation parcimonieuse. Une relaxation convexe du problème de minimisation de la norme ℓ_0 sachant que $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ conduit aux problèmes suivants :

- $\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \ \text{ s.c. } \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 < \epsilon$;
- *Dantzig selector* [44] : $\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \ \text{ s.c. } \ \left\|A^H(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})\right\|_\infty < (1 + t^{-1})\sqrt{2\log N}\sigma$;

– *Basis Pursuit De-Noising* [33] : $\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 + \lambda \|\boldsymbol{x}\|_1$

avec ϵ, t et λ des paramètres positifs.

1.2.2 Optimisation local et Algorithmes gloutons

Une autre approche d'intégration de la parcimonie consiste à régulariser le modèle en sélectionnant de manière itérative l'approximation courante optimale en fonction de certains critères, conduisant à des méthodes telles que l'*Iterative Hard Thresholding* [45], l'*orthogonal matching pursuit* [46] ou encore le *subspace pursuit* [47]. A partir d'une valeur initiale $\boldsymbol{x}^{(0)} = 0$, les différents algorithmes construisent de manière itérative, à chaque étape, une approximation $\boldsymbol{x}^{(k)}$ impliquant uniquement un faible nombre de coefficients dans l'ajustement aux données. A chaque itération, l'erreur résiduelle est évaluée. L'algorithme s'arrête si l'erreur résiduelle est inférieure à un seuil donné. Nous ne développons pas en détails ces algorithmes. Le lecteur intéressé par cette approche trouvera, par exemple, tout le matériel nécessaire dans [48, 49, 50, 51].

1.3 Apports

La notion de parcimonie intervient au coeur d'un grand nombre d'applications. La compression (l'approximation) est sans doute l'application qui vient en premier à l'esprit. Mais c'est, dans le début des années 90, grâce aux travaux de Donoho sur l'estimation statistique [23], que la notion de parcimonie s'est développée. La séparation de sources et l'acquisition compressée ont permis, par la suite, de poursuivre son essor. Ces deux domaines d'applications, plus précisément la contribution de la parcimonie à ces domaines, seront développés dans le premier chapitre (état de l'art) de chaque partie.

1.3.1 Approximation et Compression

De nos jours, le volume des données créées croît plus rapidement que les capacités de stockage, et beaucoup plus vite que la bande passante nécessaire à leur transmission. Il est, par conséquent, naturelle de chercher à compresser les données afin de les stocker ou les transmettre. Le but de la compression d'un signal est de réduire la dimension N de ce signal, tout en gardant l'information nécessaire à sa reconstruction. Les algorithmes de compression sont basés sur deux critères, qui sont le taux et la qualité de la compression.

Il existe plusieurs approches de compression. La compression par transformation en est un exemple. Cette approche procède en deux étapes, elle consiste à

1. décomposer un signal dans une base orthogonale ;

2. puis, quantifier les coefficients $\boldsymbol{x}$ de décomposition du signal dans cette base.

Etant donné un taux de compression fixé, l'approximation par les k-meilleurs termes consiste à retenir les k plus grands coefficients de décomposition. Il est à noter que ce processus de réduction de dimension est adaptatif dans le sens où, les indices des coefficients que l'on garde dépendent du signal à compresser. Ceci n'empêche pas son utilisation dans différentes applications (JPEG 2000).
La qualité de compression dépend du choix de la base orthogonale. Le théorème suivant montre une équivalence entre la parcimonie du signal et l'efficacité de la compression de celui-ci.

Théorème 1.1 (Donoho [52]) *Soit $q > 0$, on pose $m = 1/q - 1/2$. Alors, ils existent deux constantes positives α_1 et α_2 vérifiant :*

$$\alpha_1 \|\boldsymbol{x}\|_{c,m} \leq \|\boldsymbol{x}\|_{\omega\ell_q} \leq \alpha_2 \|\boldsymbol{x}\|_{c,m}$$

où $\|\boldsymbol{x}\|_{c,m} = \sup_{k\geq 1} k^m e_k(\boldsymbol{x})$ mesure la vitesse de décroissance de l'erreur $e_k(\boldsymbol{x})$ d'approximation par les k-meilleurs termes.

1.3.2 Débruitage

L'estimation des signaux dans un milieu bruité est un cadre qui a été largement étudié dans les domaines de traitement du signal et de statistiques. Le problème a été donc abordé de différentes façons et plusieurs méthodes d'estimation ont été développées. Par exemple, on peut citer les travaux de Wiener [53], d'Ephraïm-Malah [54] ou Stein [55], pour n'en citer que quelques uns. Mais c'est dans les années 90, que les travaux de Donoho et son équipe ont révolutionné la théorie de l'estimation statistique, grâce notamment aux estimateurs de seuillage dans les bases d'ondelettes et paquets d'ondelettes (*wavelet shrinkage* [23]). Ces travaux exploitaient les progrès récents en théorie des ondelettes pour concevoir un opérateur d'estimation non-linéaire simple et efficace. Plus généralement, Donoho a montré le rôle important que jouait la parcimonie dans l'efficacité des estimations.

Pour mieux voir l'apport de la parcimonie au débruitage, formalisons mathématiquement ceci. On considère le modèle suivant :

$$y_t = x_t + b_t, \quad t = 1, \ldots, n$$

où $\boldsymbol{y} = (y_t)_{1 \leq i \leq n}$ est l'observation bruitée du signal utile $\boldsymbol{x} = (x_t)_{1 \leq i \leq n}$ et $\boldsymbol{b} = (b_t)_{1 \leq i \leq n}$ est une réalisation d'un processus aléatoire $\boldsymbol{B}$. On suppose que $\boldsymbol{B}$ est indépendant du signal $\boldsymbol{x}$ et que $\boldsymbol{B} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2.\boldsymbol{\mathcal{I}}_n)$.

Le projecteur idéal $\hat{\boldsymbol{x}}^{\boldsymbol{\sigma}}$ est défini par :

$$\hat{x}_t^{\sigma} = \begin{cases} 0 & \text{si } |x_t| < \sigma \\ y_t & \text{si } |x_t| \geq \sigma \end{cases} \tag{1.3}$$

pour tout $t \in [\![1, n]\!]$. En pratique, ce projecteur n'est pas réalisable puisqu'il exige la connaissance des indices tels que $|x_t| \geq \sigma$, or cette information ne peut provenir que d'un oracle. Cependant, la notion d'estimateur oracle (estimateur nécessitant une information pertinente sur le signal que l'on cherche à estimer) est très importante dans l'évaluation des performances d'estimateurs pratiques. Le risque du projecteur idéal est donné par :

$$e(\hat{\boldsymbol{x}}^{\boldsymbol{\sigma}}) = \mathbb{E} \left\| \boldsymbol{x} - \hat{\boldsymbol{x}}^{\boldsymbol{\sigma}} \right\|^2 = \sum_{t=1}^{n} \min(x_t^2, \sigma^2).$$

Dans [23], Donoho et Johnstone ont montré que les seuillages dur et doux paramétrisés avec un seuil bien choisi permettent d'atteindre les performances du risque du projecteur idéale à un facteur logarithmique multiplicatif près. On rappelle que le seuillage dur est semblable au projecteur idéal à la différence que la condition porte sur l'observation au lieu du signal. Le seuillage doux contrairement au seuillage dur atténue les observations d'une valeur égale au seuil au lieu de garder la valeur intacte et forcer à zéro les autres (comme pour le seuillage dur)

Théorème 1.2 (Donoho et Johnstone [23]) *On suppose que $n \geq 2$ et que $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}, \sigma^2.\boldsymbol{\mathcal{I}}_n)$. Soit $\tilde{\boldsymbol{x}}$ une estimation de $\boldsymbol{x}$ via un seuillage dur ou doux de paramètre $\sigma\sqrt{2 \log n}$. Alors :*

$$e(\tilde{\boldsymbol{x}}) = \mathbb{E} \left\| \boldsymbol{x} - \tilde{\boldsymbol{x}} \right\|^2 \leq \Big(\sigma^2 + \sum_{t=1}^{n} \min(x_t^2, \sigma^2) \Big).$$

En outre, le théorème suivant montre une équivalence entre la parcimonie du signal et l'efficacité de l'estimation de celui-ci dans un milieu bruité.

Théorème 1.3 (Donoho [52]) *Soit $q > 0$, on pose $r = 1 - q/2$. Alors, ils existent deux constantes positives β_1 et β_2 vérifiant :*

$$\beta_1 \|\boldsymbol{x}\|_{e,r} \leq \|\boldsymbol{x}\|_{\omega\ell_q} \leq \beta_2 \|\boldsymbol{x}\|_{e,r}$$

Avec $\|\boldsymbol{x}\|_{e,r} = \sup_{\epsilon>0} (\epsilon^{-2r}\, e(\hat{\boldsymbol{x}}^{\boldsymbol{\epsilon}}))^{1/2}$ mesure la vitesse de décroissance du risque $e(\hat{\boldsymbol{x}}^{\boldsymbol{\epsilon}})$ d'estimation par un projecteur idéal.

Ainsi, la parcimonie joue un rôle important dans différents domaines puisqu'elle conduit à des bonnes performances. Nous nous intéressons dans la suite aux contributions de la parcimonie à la robustesse des méthodes de séparation de sources via les tests statistiques et à l'acquisition compressée des signaux à alphabet fini.

Première partie

Tests statistiques pour les méthodes de séparation de sources fondées sur la parcimonie des sources

Cette partie est consacrée à l'apport de tests statistiques aux méthodes de séparation de sources fondées sur la parcimonie des signaux sources. Nous nous plaçons dans le cadre de la séparation aveugle de sources (SAS). Les différentes méthodes SAS fondées sur la parcimonie des sources introduisent des paramètres empiriques. La performance de ces méthodes dépend du choix de ces paramètres. En général, la valeur de ces paramètres est relative aux expérimentations effectuées : niveau de bruit, données et/ou application traitées, etc. Cette sensibilité constitue une faiblesse de ces méthodes.

Le premier chapitre concerne les méthodes SAS classiques fondées sur l'hypothèse de parcimonie des sources. Ces méthodes procèdent en deux étapes : l'estimation de la matrice de mélange puis l'estimation des différentes sources. Nous rappelons ces étapes puis mettons en évidence les paramètres empiriques nécessaires pour celles ci.

Dans le second chapitre, deux tests statistiques sont proposés pour réduire le nombre de ces paramètres empiriques.

Le troisième chapitre concerne les aspects expérimentaux d'évaluation des techniques de séparation de sources. pour la séparation de sources. Nous appliquons les tests statistiques décrits dans le chapitre précédent à plusieurs méthodes SAS classiques, notamment à la méthode de projection en sous-espaces. Nous montrons que nos algorithmes statistiques réduisent le nombre de paramètres empiriques et permettent dans certain cas d'effectuer une meilleure séparation.

Enfin, dans un quatrième chapitre, nous terminerons cette partie par une conclusion.

Les travaux présentés dans cette partie ont donné lieu à une publication dans la revue *EURASIP journal on Advances in Signal processing* [14] et un article dans la conférence *ICASSP* [15].

CHAPITRE 2 La séparation de sources

2.1 Introduction

La séparation de sources est un problème qui vise à reconstruire/estimer les signaux sources à partir des mélanges obtenus par un réseau de capteurs. L'étude de ce problème n'est pas très récente, cependant, elle continue à susciter de l'intérêt. Ce sujet très actif continue encore à être d'actualité. Dans la littérature, on parle souvent de séparation *aveugle* de sources (SAS). Cette terminologie correspond à un manque de connaissance a priori sur le processus de mélange. La SAS est un important sujet de recherche qui trouve application dans de nombreux domaines tels le traitement audio [7, 8], les télécommunications [9, 10], l'ingénierie biomédicale [11, 12] ou encore le traitement radar [13]. Les problèmes de SAS peuvent être classés selon la nature du mélange et des sources. La nature du mélange dépend de trois paramètres. Le premier paramètre concerne les paramètres du mélange, c'est à dire le rapport entre le nombre de mélanges n et le nombre de sources présentes dans les différents mélanges N. Un mélange est dit sous-déterminé (resp., sur-déterminé) si $n/N < 1$ (resp., $n/N \geq 1$). Un second paramètre indique si le mélange est linéaire ou non linéaire. Dans le cadre de cette thèse, nous supposons les mélanges linéaires. Cette hypothèse rend le problème plus facile à traiter que le cas général où les mélanges sont non-linéaires (par exemple [56, 57, 58]). Néanmoins, l'hypothèse est acceptable dans de nombreuses applications. Enfin, un troisième paramètre qui qualifie la nature même du mélange. Lorsque les mélanges sont obtenus par combinaisons linéaires des différentes sources au même instant, le mélange est dit linéaire instantané. Le mélange est convolutif si au moins un des mélanges dépend de plusieurs versions retardées d'une même source. En ce qui concerne la nature des sources, plusieurs types de signaux ont été abordés tels que les signaux à alphabet fini [59, 60, 61], à cohérence temporelle [62], les signaux non stationnaires [63, 64], ou même cyclostationnaire [65]...

Le problème de SAS a été abordé de différentes façons. Suivant la nature du mélange

et des sources, il existe plusieurs types d'approches de SAS adaptées plus ou moins à ce modèle. Parmi elles, on peut citer les approches d'analyse en composantes indépendantes [66, 67, 68, 69], l'approche de factorisation en matrices non-négatives, ou encore les techniques de SAS fondées sur la représentation parcimonieuse des sources ...

2.2 L'approche d'analyse en composantes indépendantes

Les premières tentatives de séparation de sources sont apparues dans le cadre d'un problème de décodage du mouvement en neurobiologie. Dans [70, 71, 72], Jutten et Hérault ont développé un algorithme de séparation, en absence de bruit, de n sources à partir de n mélanges à base d'un réseau neuromimétrique. Ces travaux ont donné naissance au concept d'Analyse en Composantes Indépendantes (ACI). Formalisé plus précisément par Comon dans [69], l'ACI est devenue une approche incontournable en SAS. L'ACI consiste à résoudre le problème de SAS en estimant les sources par une transformation linéaire qui minimise la dépendance des sources. Historiquement, l'analyse en composantes indépendantes (ACI) peut être vue comme une extension de l'analyse en composantes principales (ACP), vu que l'ACP n'impose qu'une décorrélation des sources. La seule indépendance au second ordre ne permet pas de résoudre le problème de SAS. En effet, l'ACP introduit une indétermination à une matrice orthogonale près sur l'estimation de la matrice de mélange. L'ACI fondée sur une indépendance aux moments d'ordre supérieur à deux, donc une hypothèse plus forte, permet de lever cette indétermination introduite si l'on utilise uniquement l'ACP. Afin d'obtenir l'indépendance des sources, l'approche de base consiste à minimiser l'information mutuelle des sources estimées. L'information mutuelle est une notion liée à la théorie de l'information qui s'annule lorsque les sources sont indépendantes. Différentes techniques ont été proposées pour l'ACI chacune exploitant une propriété statistique des sources. Parmi ces propriétés, citons la non-stationarité [73, 63, 64] et la corrélation temporelle [62, 74]. Cependant, la propriété la plus exploitée dans l'ACI est celle de la non-gaussianité des sources. Dans ce contexte, il est donc naturel de chercher à maximiser la valeur absolue des cumulants d'ordre supérieur à deux tel que le kurtosis [66], puisque ceux ci sont nuls pour une Gaussienne. Tel que nous l'avons présenté, l'ACI se focalise uniquement sur l'estimation de la matrice de mélange qui maximise l'indépendance des sources. Pour cause, l'ACI a été développée principalement pour la SAS dans le cadre de mélange sur déterminé. Dans ce cadre, le problème de séparation de sources est équivalent à celui de l'estimation de la matrice de mélange. Les sources sont obtenues par inversion

généralisée [29]. Dans le cadre sous-déterminé, nous sommes en présence d'un manque d'informations. Il est donc assez naturel d'utiliser l'algorithme EM (*Expectation Maximization*, en anglais) [75] pour l'estimation de la matrice de mélange et des sources. Cependant, ce type d'approche nécessite des informations a priori sur la distribution des sources.

2.3 L'approche de factorisation en matrices non-négatives

La factorisation en matrices non négatives (NMF pour *Non-negative Matrix Factorization*) est une méthode de réduction dimensionnelle fréquemment utilisée pour extraire des formes caractéristiques sous-jacentes à des processus complexes. Initialement, la NMF a été proposée par Lee et Seung [76]. Ces travaux ont été inspirés par les travaux antérieurs de Paateero [77] sur la factorisation en matrices positives. Etant donnée une matrice non-négative $\boldsymbol{V}$ de dimension $n \times N$, le but de la NMF est d'approximer $\boldsymbol{V}$ par le produit de deux matrices non-négatives $\boldsymbol{W}$ et $\boldsymbol{H}$ respectivement de dimension $n \times r$ et $r \times N$, en minimisant une erreur de reconstruction $D(\boldsymbol{V}\|\boldsymbol{WH})$. L'entier naturel r est généralement choisi de telle sorte que $nr + rN \ll nN$, ce qui permet de réduire la dimension des données. Concernant l'erreur de reconstruction, différentes mesures $D(\cdot\|\cdot)$ ont été étudiées. Citons, par exemple, la distance Euclidienne [76], la divergence de Kullback Leibler (généralisée) [78], la divergence d'Itakura-Saito [79] ou encore les divergences Bregman [80] et Csiszar [81]. Lee et Seung proposent une règle de mise à jour multiplicative qui est équivalente à une descente de gradient avec un pas particulier. La simplicité de cette règle de mise à jour a sans doute contribué à la popularité de la NMF. Depuis, plusieurs algorithmes ont été proposés [82, 83]. D'abord introduite en traitement d'image par Lee et Seung [76] pour la reconnaissance de visages, la NMF est depuis largement utilisée pour divers problèmes et dans différents domaines d'applications [76, 84, 85]. Cependant, c'est en traitement audio que la NMF a suscité le plus grand intérêt, notamment en identification sonore [86, 87], en transcription musicale [88] ou encore en séparation de sources [89, 90, 91, 92] où la matrice $\boldsymbol{V}$ représente le spectre d'amplitude d'un signal audio. Ainsi, la NMF cherche à décrire le spectre comme une combinaison de composantes élémentaires récurrentes dans le mélange. Souvent, un assemblage de ces éléments simples donne une bonne estimation des différentes sources présentes. Cependant, l'arbitraire des composantes produites constitue un problème de l'approche de SAS fondée sur la NMF. Afin de pallier ce problème, plusieurs travaux ont émergé. Ces travaux sont basés sur l'ajout

de contraintes supplémentaires telle la mélodicité [93, 94], la régularité [95, 96], l'invariance par translation [97, 98] ou encore la parcimonie [91].

2.4 Méthodes SAS fondées sur la parcimonie

2.4.1 Modèle

Nous considérons le système de mélange instantané :

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{b}(t), \tag{2.1}$$

où t varie dans un certain ensemble fini de périodes d'échantillonnage tel que, pour tout t appartenant à cet ensemble, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_N(t)]^T$ est le vecteur des N signaux sources, $\boldsymbol{x}(t) = [x_1(t), x_2(t), \cdots, x_n(t)]^T$ représente le vecteur d'observations de dimension n, $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_N]$ est la matrice de mélange de dimension $n \times N$ et $\boldsymbol{b}(t) = [b_1(t), b_2(t), \cdots, b_n(t)]^T$ représente le bruit additive. On note, pour $i \in [\![1, n]\!]$ et $j \in [\![1, N]\!]$, (a_{ij}) les coefficients de la matrice $\boldsymbol{A}$ et on suppose que $(b_k(t))_{1 \leq k \leq n}$ est une suite de processus aléatoires Gaussiens, mutuellement décorrélés et indépendants des sources. Dans la suite, sauf mention contraire, nous aborderons le cas sous déterminé ($n < N$). Sans perte de généralité, nous supposons que les vecteurs colonnes de $\boldsymbol{A}$ sont de norme unité, c'est à dire $\|\boldsymbol{a}_i\| = 1$ pour tout $i \in [\![1, N]\!]$. En effet, dû à l'indétermination par rapport à l'échelle, les colonnes de la matrice $\boldsymbol{A}$ ne peuvent être estimées qu'à une constante multiplicative près.

2.4.2 Principe

La parcimonie ou plus précisément l'existence d'un domaine où les différents signaux sont parcimonieux constitue un principal atout permettant de résoudre le problème de SAS dans le cadre sous déterminé [99]. Dictée par la nature des signaux que l'on va traiter pour l'évaluation des algorithmes proposés, on s'intéressera au domaine temps-fréquence. Cependant, toute autre transformation parcimonieuse inversible peut être utilisée. Le traitement temps-fréquence des signaux [100] fournit un cadre efficace pour analyser les signaux non stationnaires, dont les caractéristiques fréquentielles évoluent dans le temps. Il nécessite la représentation du signal dans un espace à deux dimensions, qui est, le domaine temps-fréquence, fournissant ainsi une distribution de l'énergie du signal en fonction du temps et de la fréquence simultanément. L'exploitation des différences dans les distributions propre à chaque source constitue la base des méthodes

de séparation de sources temps-fréquence.
Dans la pratique, la transformée de Fourier à court terme (TFCT) est la représentation temps-fréquence la plus utilisée. Le processus de mélange peut être modélisé dans le domaine temps-fréquence par l'intermédiaire de la TFCT comme suit :

$$\mathcal{S}_{\boldsymbol{x}}(t,f) = \boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f) + \mathcal{S}_{\boldsymbol{b}}(t,f), \tag{2.2}$$

où $\mathcal{S}_{\boldsymbol{x}}(t,f)$, $\mathcal{S}_{\boldsymbol{s}}(t,f)$ et $\mathcal{S}_{\boldsymbol{b}}(t,f)$ sont respectivement les vecteurs des coefficients de la TFCT au point temps-fréquence (t,f) des observations, des sources et du bruit. Etant donné $\boldsymbol{x}(t)$, notre but est d'estimer $\boldsymbol{s}(t)$ ou de manière équivalente $\mathcal{S}_{\boldsymbol{s}}(t,f)$. Dans le cas sous déterminé, il est utile de considérer les deux problèmes suivants :

- L'estimation de la matrice de mélange : les différentes colonnes $(\boldsymbol{a}_i)_{1\leq i\leq N}$ sont estimées afin d'obtenir une estimation de la matrice de mélange $\boldsymbol{A}$.

- Puis, l'estimation des signaux sources connaissant une estimée de la matrice $\boldsymbol{A}$: une solution particulière de l'équation (2.2) est choisie en utilisant des contraintes appropriées.

Toutes les méthodes de SAS, où plus de sources sont présentes que de mélanges, ne décomposent pas forcément le problème en deux sous problèmes séparés. En effet, à titre d'exemple citons les articles [101, 102], où les deux étapes sont fusionnées en une maximisation de la vraisemblance des données. Un autre exemple est celui des méthodes de séparation de sources avec un seul capteur [103]. En générale, les méthodes de SAS basées sur la parcimonie procèdent en deux étapes. Ainsi, dans cette partie, nous suivons une approche de SAS en deux étapes. La figure 2.1 présente un organigramme d'une telle approche.

Nous détaillons maintenant quelques techniques d'estimation de la matrice de mélange et d'estimation des signaux sources exploitant l'hypothèse de parcimonie. L'objectif n'étant pas de donner une étude bibliographique exhaustive des techniques existantes, mais de détailler les méthodes qui nous concernent directement. Un éventail plus large des techniques de SAS dans le cas sous-déterminé peut être trouvé dans [104].

2.4.3 Estimation de la matrice de mélange

Les méthodes de SAS basées sur la parcimonie des signaux dans le domaine temps-fréquence partagent l'hypothèse principale suivante :

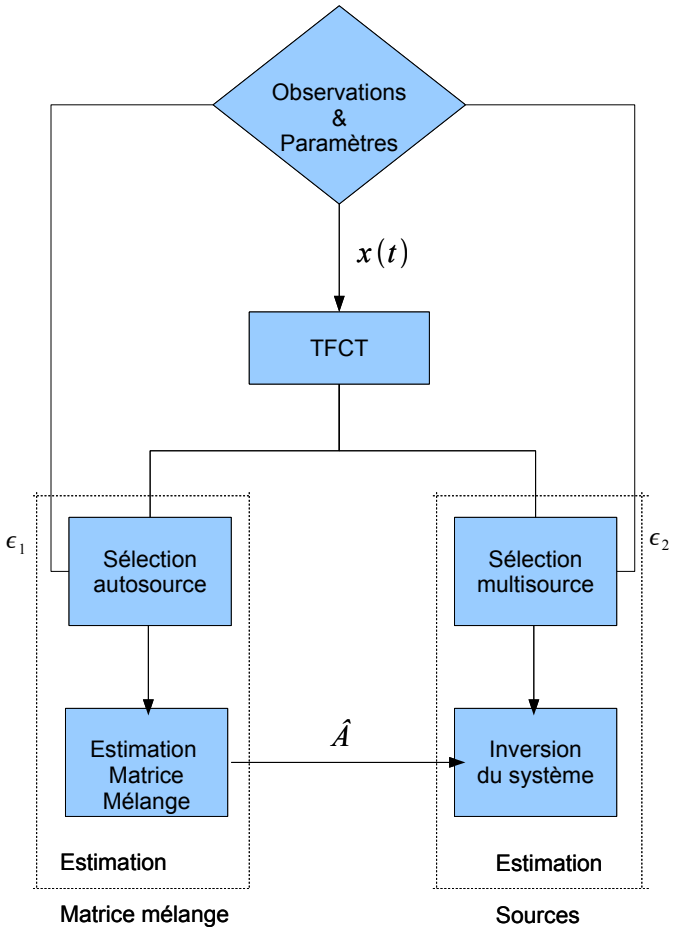


Figure 2.1 — Organigramme des algorithmes de SAS procédants en deux étapes dans le domaine temps-fréquence.

Hypothèse 2.1 *Pour chaque source, il existe un ensemble de zones temps-fréquence où seule cette source est présente.*

Une zone temps-fréquence peut désigner un point temps-fréquence [1, 105] ou une série de points temps-fréquence temporellement adjacents [106, 107]. L'hypothèse 2.1 est souvent raisonnable grâce à la parcimonie des représentations temps-fréquence des différentes sources, surtout quand le nombre de sources est modéré. En effet, plus les signaux sont décorrélés et parcimonieux en temps-fréquence plus la combinaison linéaire de ces sources fera apparaître ces zones temps-fréquence. Toutefois, notant que si cette hypothèse n'est pas satisfaite, c'est à dire qu'il existe une source qui n'est présente toute seule dans aucune zone temps-fréquence (source cachée), les auteurs dans

[106] proposent une technique itérative qui permet éventuellement de rendre *visible* ces sources. La technique consiste à supprimer la contribution des sources visibles (e.g. la source i) dans les différentes observations en estimant les paramètres du mélange propres à ces sources ($\boldsymbol{a}_i$) par n'importe quelle procédure d'estimation de paramètres de mélange décrite dans cette partie. Par cette réduction de nombre de sources, on espère rendre visible les sources qui étaient initialement cachées par d'autres sources. La figure 2.2-(a) présente une telle situation où seule la source 1 est visible. Les sources 2 et 3 ne sont pas visibles. Cependant, en annulant la contribution de la source visible 1, les sources 2 et 3 deviennent visibles (figure 2.2-(b)).

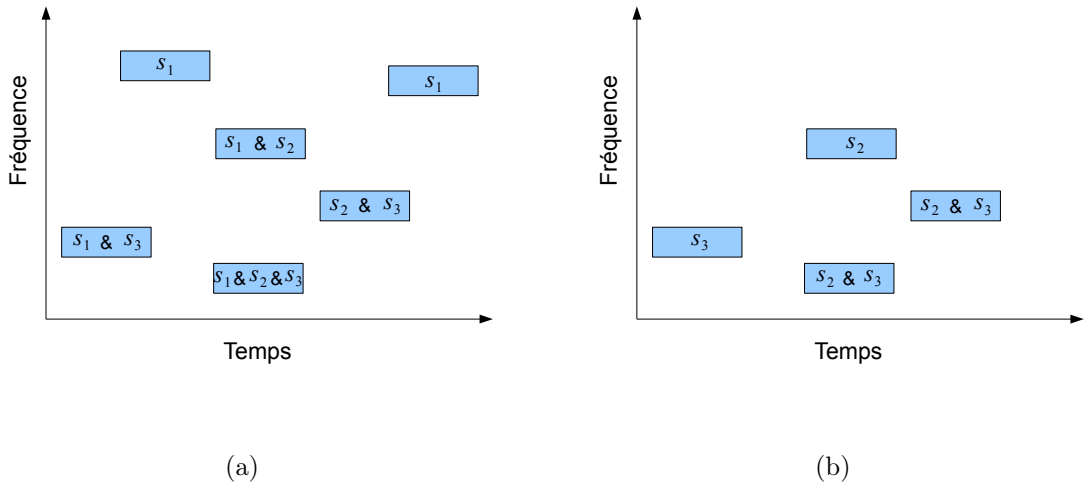


Figure 2.2 — Procédure itérative de traitement des sources cachées dans un mélange. (a) Exemple de recouvrement temps-fréquence dans un mélange de trois sources. Les sources 2 et 3 ne sont pas visibles. En supprimant la contribution de la source 1, nous obtenons un nouveau mélange. (b) Recouvrement temps-fréquence dans le nouveau mélange où toutes les sources sont visibles.

Comme nous l'avons déjà signalé, la première étape des méthodes de SAS dans le cas sous déterminé consiste à estimer la matrice de mélange $\boldsymbol{A}$ pour ensuite estimer les signaux sources. Dans la plupart des algorithmes de séparation de source en deux étapes [1, 108, 106, 109, 110, 111], une sélection autosource est effectuée. Par sélection autosource, on désigne la détection de régions temps-fréquence où une seule source est présente. L'intérêt de la sélection d'un point temps-fréquence est la suivante :

Soit (t, f) un point temps fréquence où une seule source est présente. Notons, par exemple, i l'indice de cette source et supposons que les observations ne sont pas bruitées. L'équation (2.2) devient alors

$$\mathcal{S}_{\boldsymbol{x}}(t,f) = \mathcal{S}_{s_i}(t,f) \cdot \boldsymbol{a}_i \tag{2.3}$$

Ainsi, étant donné un tel point, nous pouvons estimer le vecteur normalisé $\boldsymbol{a}_i$ à partir de $\mathcal{S}_{\boldsymbol{x}}(t,f)$. Dans le cas bruité, pour estimer $\boldsymbol{a}_i$, il suffit de moyenner sur tous les points temps-fréquence où la source i est présente. La figure 2.3 illustre bien ce résultat pour un mélange de 2 capteurs et 3 sources. Cette figure, dénommée diagramme de dispersion, représente la partie réelle des points $(\mathcal{S}_{x_1}(t,f), \mathcal{S}_{x_2}(t,f))$ lorsque (t,f) parcourt l'ensemble des points temps-fréquence. On en déduit que la connaissance de points temps-fréquence où une seule source est présente permet de rendre le problème d'estimation de la matrice de mélange trivial.

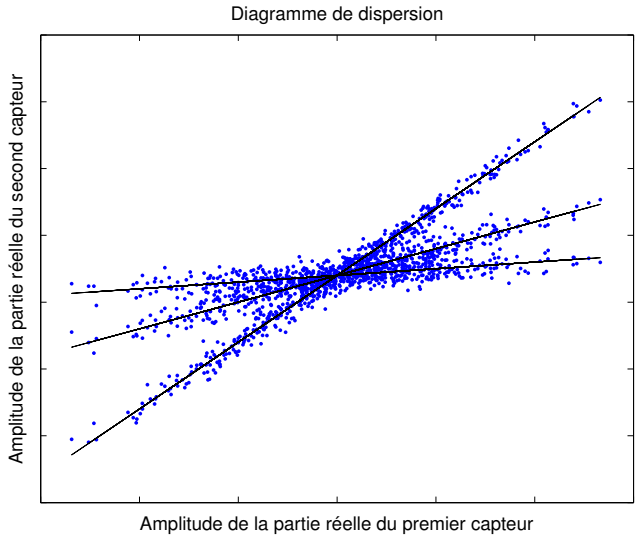


Figure 2.3 — Diagramme de dispersion dans le domaine de parcimonie des sources.

Nous pouvons décomposer les différentes techniques proposées pour estimer les paramètres de mélanges en deux classes. La première classe de méthodes, basée sur le diagramme de dispersion global, estime les paramètres de mélanges en utilisant la majorité des points temps-fréquence. Les approches présentées dans [108, 105, 112, 113, 114] appartiennent à cette classe. Les coefficients de la matrice de mélange sont par exemple estimés par un algorithme de classification, type *kmeans*, sur des paramètres du diagramme de dispersion 2.3, ou encore en étudiant les pics de l'histogramme des directions spatiales [105, 113, 112]. Une autre approche similaire a été proposée par Bofill et Zibulevski [108]. Cette approche consiste à estimer les paramètres de mélanges en maximisant une fonction potentielle Φ sur une grille discrète de directions. En calculant pour chaque point du diagramme de dispersion $z_1(t,f), z_2(t,f)$ les grandeurs

$l(t,f) = \sqrt{(z_1(t,f))^2 + (z_2(t,f))^2}$ et $\theta(t,f) = \tan^{-1}(z_1(t,f)/z_2(t,f))$, les directions des vecteurs colonnes de la matrice de mélange $\boldsymbol{A}$ sont données par les maxima locaux de la fonction potentielle

$$\Phi(\theta_k) = \sum_{(t,f)} l(t,f)\phi(\lambda(\theta_k - \theta(t,f)))$$

où $\theta_k = \pi/2K + k\pi/2K,\ k \in [\![1, K]\!]$ est une grille discrète de directions et $\phi = 1 - \frac{\alpha}{\pi/4}$ pour $|\alpha| < \pi/4$ et nulle sinon.
Les paramètres α, K et λ sont des paramètres de l'approche d'estimation. D'autre part, les auteurs proposent une étape de sélection des points temps-fréquence d'énergie significative, c'est à dire les points temps-fréquence tels que $l(t,f) > h$. La valeur du seuil h a été fixée à 0.3, environ un tiers de la tranche dynamique.

La seconde classe de méthodes est basée sur le diagramme de dispersion local. Cette classe relaxe l'hypothèse de présence d'une seule source à chaque points temps-fréquence, en supposant qu'il peut exister des points temps-fréquence où plusieurs sources se recouvrent à condition que l'hypothèse 2.1 soit satisfaite. Pour estimer la matrice de mélange, une technique de sélection de points autosources est alors proposée. La technique de SAS présentée par Aïssa-El-Bey et al. dans [1] et que l'on notera SUBSS, proposent de sélectionner les points autosources en appliquant une procédure de seuillage. Plus précisément, pour chaque point temps-fréquence (t,f) on applique le critère suivant :

$$\text{Si } \frac{\|\mathcal{S}_{\boldsymbol{x}}(t,f)\|}{\max_u\{\|\mathcal{S}_{\boldsymbol{x}}(t,u)\|\}} > \epsilon_2, \quad \text{garder } (t,f) \tag{2.4}$$

où ϵ_2 est un seuil empirique (typiquement, $\epsilon_2 = 0.05$). Les coefficients de la matrice $\boldsymbol{A}$ sont ensuite estimés via une procédure de classification de type *kmeans* des vecteurs :

$$\boldsymbol{v}(t,f) = \frac{\mathcal{S}_{\boldsymbol{x}}(t,f)}{\|\mathcal{S}_{\boldsymbol{x}}(t,f)\|}$$

où (t,f) parcourt l'ensemble des points autosources sélectionnés par la procédure (2.4). Les N classes obtenues, notées $\{\mathcal{C}_i | i \in [\![1, N]\!]\}$, permettent alors d'estimer les paramètres de mélanges et d'identifier Ω_i l'ensemble des points temps-fréquence correspondant à la présence de la même source i. Clairement, les vecteurs colonnes $\boldsymbol{a}_i$ de la matrice de mélange $\boldsymbol{A}$ sont estimés comme les centroïdes des différentes classes

$$\hat{\boldsymbol{a}}_i = \frac{1}{\#(\Omega_i)} \sum_{(t,f)\in\Omega_i} \boldsymbol{v}(t,f).$$

Une autre approche d'estimation des paramètres de mélanges a été proposée par Arbard et Deville dans [106]. Les auteurs proposent une méthode de SAS nommée TIFROM (*TIme Frequency Ratio Of Mixtures*, en anglais). L'estimation de la matrice de mélange dans TIFROM relaxe également l'hypothèse de présence d'une seule source à chaque points temps-fréquence. En définissant une zone temps-fréquence comme une série Γ_q de M fenêtres (points) temps-fréquence temporellement adjacents en demi-recouvrement, ils supposent que l'hypothèse 2.1 est satisfaite. Pour illustrer leur approche, considérons le cas de deux capteurs $n = 2$. L'extension à n capteurs a été présentée dans [115]. Les zones temps-fréquence autosources sont identifiées et traitées pour l'estimation des paramètres de la matrice de mélange en étudiant respectivement la variance et la moyenne des rapports d'observations suivants :

$$\alpha(t,f) = \frac{\mathcal{S}_{x_1}(t,f)}{\mathcal{S}_{x_2}(t,f)}$$

Sous l'hypothèse que la variation des sources est incohérente sur les M points temps-fréquence qui forment une zone temps-fréquence, les zones autosources correspondent aux zones (Γ_q, f_k) de variance nulle. Cette variance est calculée comme suit

$$\mathrm{var}[\alpha](\Gamma_q, f_k) = \frac{1}{M} \sum_{j=1}^{M} \left|\alpha(t_j, f_k) - \bar{\alpha}(\Gamma_q, f_k)\right|^2, \tag{2.5}$$

où $\bar{\alpha}(\Gamma_q, f_k)$ est la moyenne des rapports d'observations sur la zone temps-fréquence (Γ_q, f_k), et est définie par

$$\bar{\alpha}(\Gamma_q, f_k) = \frac{1}{M} \sum_{j=1}^{M} \alpha(t_j, f_k). \tag{2.6}$$

D'après l'équation (2.3), en notant i l'indice de la source présente dans la zone temps-fréquence autosource (Γ_q, f_k), cette moyenne est égale au rapport a_{1i}/a_{2i} et permet ainsi d'estimer les coefficients de la matrice de mélange (la colonne $\boldsymbol{a}_i$ est supposée de norme unité).

Pour résumer, la procédure de base pour estimer la matrice de mélange dans TIFROM consiste à classer les moyennes de l'expression (2.6) par ordre croissant des variances définies par (2.5) en fonction de toutes les zones temps-fréquence (Γ_q, f_k). Puis à sélectionner les N différentes premières valeurs.

Une amélioration de cette procédure d'estimation des paramètres de mélanges a été proposée par Deville et al. dans [116]. En posant

$$\begin{aligned}
\beta(t,f) &= \frac{\mathcal{S}_{x_2}(t,f)}{\mathcal{S}_{x_1}(t,f)} \\
\bar{\beta}(\Gamma_q, f_k) &= \frac{1}{M}\sum_{j=1}^{M}\beta(t_j, f_k) \\
\mathrm{var}[\beta](\Gamma_q, f_k) &= \frac{1}{M}\sum_{j=1}^{M}\left|\beta(t_j, f_k) - \bar{\beta}(\Gamma_q, f_k)\right|^2,
\end{aligned}$$

la modification consiste à rechercher les zones autosources à l'aide du paramètre :

$$\min\{\mathrm{var}[\alpha](\Gamma_q, f_k), \mathrm{var}[\beta](\Gamma_q, f_k)\} \tag{2.7}$$

La détection est réalisée en classant les moyennes de l'expression (2.6) par ordre croissant des paramètres définies par l'équation (2.7) en fonction de toutes les zones temps-fréquence (Γ_q, f_k), au lieu des seules variances $\mathrm{var}[\alpha](\Gamma_q, f_k)$. Ceci permet de rendre la procédure de détection symétrique par rapport aux sources.

Deville et al. proposent dans [107, 117, 118] plusieurs techniques de détection des zones temps-fréquence autosources. Ces techniques similaires à celle de TIFROM, utilisent des critères basés sur la cohérence et la corrélation au lieu de la variance définie par l'expression (2.5).

2.4.4 Estimation des signaux sources

Cette section présente un certain nombre de techniques utilisées dans la phase d'estimation des sources des algorithmes de SAS. Dans le cas sous déterminé, le système (2.2) a moins d'équations que d'inconnues, et donc il admet (en général) une infinité de solutions. Afin de reconstruire les signaux originaux, des hypothèses supplémentaires sont nécessaires, notamment la parcimonie des signaux sources.

En supposant que les sources soient quasi-disjointes dans le domaine temps-fréquence, les signaux originaux s'estiment à l'aide d'un masque temps-fréquence. C'est le principe de reconstruction dans la méthode de SAS DUET [105, 109] (pour *Degenerate Unmixing Estimation Technique* , en anglais). En effet, comme l'estimation des paramètres de mélange fournit une répartition des supports des sources dans le domaine temps-fréquence, la TFCT estimée de la source $\boldsymbol{s}_i$ est calculée par :

$$\hat{\mathcal{S}}_{s_i}(t,f) = \hat{\boldsymbol{a}}_i^H \mathcal{S}_{\boldsymbol{x}}(t,f) \cdot \mathbb{1}_{\Omega_i}(t,f) \tag{2.8}$$

où Ω_i est le support de la source $\boldsymbol{s}_i$ et $\mathbb{1}_{\boldsymbol{B}}$ est la fonction indicatrice de l'ensemble $\boldsymbol{B}$. En générale, cette procédure attribue une source à chaque point temps-fréquence, même si ce point est dû au bruit seul, ce qui dégrade les performances globales de la procédure. Afin de pallier ce problème, les points temps-fréquence d'énergie significative sont sélectionnés. Dans [1], par exemple, cette sélection est réalisée via un seuillage dont le seuil est fixé de façon empirique.

Bien que les méthodes TIFROM et TIFCORR n'exigent pas que les sources soient quasi-disjointes, elles souffrent cependant de la même limitation que les méthodes utilisant les masques temps-fréquence, dans le sens où la phase d'estimation des sources assigne aussi les points temps-fréquence où seul le bruit est présent aux différentes sources. Ceci est dû au fait que ces méthodes ont été développées dans le cas de mélange sans bruit. Les auteurs de [106, 107] ont étudié uniquement l'influence du bruit sur l'estimation de la matrice de mélange. Cependant, ce bruit a une influence aussi sur l'estimation des sources. Suivant les données du problème (détermination, nombre de sources cachées), les sources sont estimées dans TIFROM par inversion globale ou par atténuations successives. Pour illustrer cette limitation, supposons que $n = N$, c'est à dire le cas déterminé, et que l'hypothèse 2.1 est satisfaite. L'inversion globale consiste à inverser la matrice de mélange estimée $\hat{\boldsymbol{A}}$ fournie dans la première phase :

$$\hat{\mathcal{S}}_{\boldsymbol{s}}(t,f) = \hat{\boldsymbol{A}}^{-1} \mathcal{S}_{\boldsymbol{x}}(t,f)$$

Donc, en un point temps-fréquence comportant uniquement du bruit, le mélange sera réparti sur les différentes sources.

Connaissant une estimée de la matrice de mélange, une autre approche d'estimation des sources consiste à reconstruire les sources par minimisation sous contrainte de parcimonie. Comme nous l'avons vu dans le premier chapitre, la norme ℓ_1 est une bonne mesure de parcimonie. Ainsi, Bofill et Zibulevski, dans le cas sans bruit, propose de résoudre l'optimisation suivante :

$$\arg\min_{\boldsymbol{z}(t,f)} \|\boldsymbol{z}(t,f)\|_1 \quad \text{sous la contrainte} \quad \mathcal{S}_{\boldsymbol{x}}(t,f) = \boldsymbol{A}\boldsymbol{z}(t,f). \tag{2.9}$$

En présence de bruit, la contrainte doit être modifiée de manière à prendre en compte l'écart type du bruit. Dans la pratique, cet écart type du bruit est inconnu et doit être estimé.

Enfin, Aïssa-El-Bey et al [1] propose une technique d'estimation des sources basée sur la projection en sous-espace. L'approche fait appel aux deux hypothèses suivantes :

Hypothèse 2.2 *En tout point temps-fréquence, le nombre de sources actives est strictement inférieur au nombre n d'observations.*

Hypothèse 2.3 *Toute sous matrice carrée de $\boldsymbol{A}$ de taille n est de rang plein, c'est à dire, pour tout $\mathcal{J} \subset \{1, 2, \cdots, N\}$ tel que $\#(\mathcal{J}) \leq n$, les vecteurs colonnes $(\boldsymbol{a}_j)_{j\in\mathcal{J}}$ sont linéairement indépendants.*

L'approche de SAS par projection en sous-espaces identifie les sources présentes en un point temps-fréquence (t, f) en minimisant :

$$\mathcal{K} = \arg\min_{\mathcal{J}} \{\|\boldsymbol{\mathcal{Q}}_{\mathcal{J}} \mathcal{S}_{\boldsymbol{x}}(t, f)\| \mid \#(\mathcal{J}) \leq n\}$$

où $\boldsymbol{\mathcal{Q}}_{\mathcal{J}}$ est la matrice de projection orthogonale sur le sous espace bruit de la matrice $\boldsymbol{A}_{\mathcal{J}}$ formée par les colonnes $(\boldsymbol{a}_j)_{j\in\mathcal{J}}$. On a donc

$$\boldsymbol{\mathcal{Q}}_{\mathcal{J}} = \boldsymbol{\mathcal{I}}_n - \boldsymbol{A}_{\mathcal{J}} \boldsymbol{A}_{\mathcal{J}}^{\#}$$

où $\boldsymbol{A}_{\mathcal{J}}^{\#}$ est le pseudo-inverse de Moore-Penrose [29] de la matrice $\boldsymbol{A}_{\mathcal{J}}$.

Après avoir identifier les différentes sources présentes en un point temps-fréquence (t, f) et grâce à l'hypothèse 2.3, la méthode consiste à résoudre le problème linéaire localement surdéterminé résultant, c'est-à-dire,

$$\hat{\mathcal{S}}_{\boldsymbol{s}_{\mathcal{K}}}(t, f) = \boldsymbol{A}_{\mathcal{K}}^{\#} \mathcal{S}_{\boldsymbol{x}}(t, f)$$

Les points temps-fréquence (t, f) traités sont les points temps-fréquence correspondants à la présence de signaux sources. Une étape préliminaire de sélection des points d'énergie inférieure à un certain seuil empirique est donc réalisée.

CHAPITRE 3 Tests statistiques pour les méthodes SAS basées sur la parcimonie des sources

Ce chapitre est le noyau principal de cette partie. Il est dédié à une série d'améliorations qui seront apportées aux méthodes classiques de SAS présentés dans le chapitre précèdent. Ces améliorations concernent la sélection de points temps-fréquence nécessaires pour l'estimation de la matrice de mélange (sélection autosource) et la sélection de points temps-fréquence pour reconstruire les signaux originaux (sélection multisource). L'idée principale de l'approche décrite ci-dessous est de formuler ces problèmes de sélection de points temps-fréquence en problèmes statistiques consistants à rejeter ou à ne pas rejeter (dans notre cas accepter) la présence de sources dans le bruit, donnant lieu à deux tests d'hypothèses. Ces deux tests statistiques sont différents. En effet, l'estimation de la matrice de mélange nécessite la sélection des points où une unique source est présente, alors que cette contrainte est inutile pour le débruitage et l'estimation des sources.

La difficulté du problème soulevé par ces tests d'hypothèses binaires est double. D'une part, la distribution de probabilité des observations est inconnue, car celle ci dépend des distributions de probabilité des sources qui sont eux-mêmes inconnues. D'autre part, l'écart type du bruit est aussi inconnu. La décision statistique d'accepter ou de rejeter la présence de sources ne pourra donc se faire à l'aide de la théorie de vraisemblance usuelle ou une de ses extensions tel le rapport de vraisemblance généralisé [119] ou telles les approches basées sur l'invariance, ces approches requièrent la connaissance des distributions de probabilité mises en jeu.

Pour l'estimation des sources, le détecteur proposé est basé sur un modèle parcimonieux de sources en présence de bruit. Le détecteur commence par une estimation de l'écart type du bruit via l'algorithme DATE (pour d-Dimensional Amplitude Trim-

med Estimator), récemment introduit dans [120] et qui découle du MC-ESE. Puis, la vraie valeur inconnue de l'écart type du bruit est remplacée par la valeur estimée dans l'expression d'un test statistique, conçu également dans le cadre d'un modèle parcimonieux.

Pour l'estimation de la matrice de mélange, nous exploitons la nature physique des signaux pour détecter les points temps-fréquence où une seule source est présente. Pour les signaux avec un taux de recouvrement élevé, le SNT (pour Signal Norm Testing) est approprié pour la sélection de ces points temps-fréquence. Lorsque les signaux ont un faible taux de recouvrement, nous utilisons directement les points temps-fréquence fournis par la procédure d'estimation des sources.

La figure 3.1 présente un organigramme de l'approche proposée utilisant les algorithmes DATE et SNT.

3.1 Sélection multisource

Dans cette section, on s'intéresse à l'étape d'estimation des sources. Dans notre modèle, pour chaque capteur, les observations (coefficients TFCT) sont supposées résulter de signaux, aléatoirement présents ou absents, dans un bruit additif blanc gaussien indépendant (BABG). Il est clair que les points temps-fréquence dus au bruit seul sont inutiles pour estimer les sources. Par conséquent, une phase dans l'étape d'estimation des sources consiste à détecter les points temps-fréquence où les différentes sources sont présentes. Détecter ces points temps-fréquence équivaut à décider de la présence ou non d'un signal d'intérêt en un point temps-fréquence (t, f).

3.1.1 Formulation statistique

Etant donnés un point temps-fréquence (t, f) et $i \in [\![1, n]\!]$, il est donc naturel de formuler ce problème par un test d'hypothèse binaire tel que :

$$\begin{cases} \mathcal{H}_0 : & \mathcal{S}_{\boldsymbol{x}_i}(t, f) = \mathcal{S}_{\boldsymbol{b}_i}(t, f) \\ \mathcal{H}_1 : & \mathcal{S}_{\boldsymbol{x}_i}(t, f) = \Theta(t, f) + \mathcal{S}_{\boldsymbol{b}_i}(t, f), \end{cases} \tag{3.1}$$

où $\Theta(t, f)$ représente le mélange de signaux éventuellement présents au point temps-fréquence (t, f).

En supposant que les signaux $\mathcal{S}_{\boldsymbol{x}_i}(t, f)$, $\mathcal{S}_{\boldsymbol{b}_i}(t, f)$ et $\Theta(t, f)$ soient des réalisations de variables aléatoires complexes X, B et Θ respectivement, nous formalisons ce problème

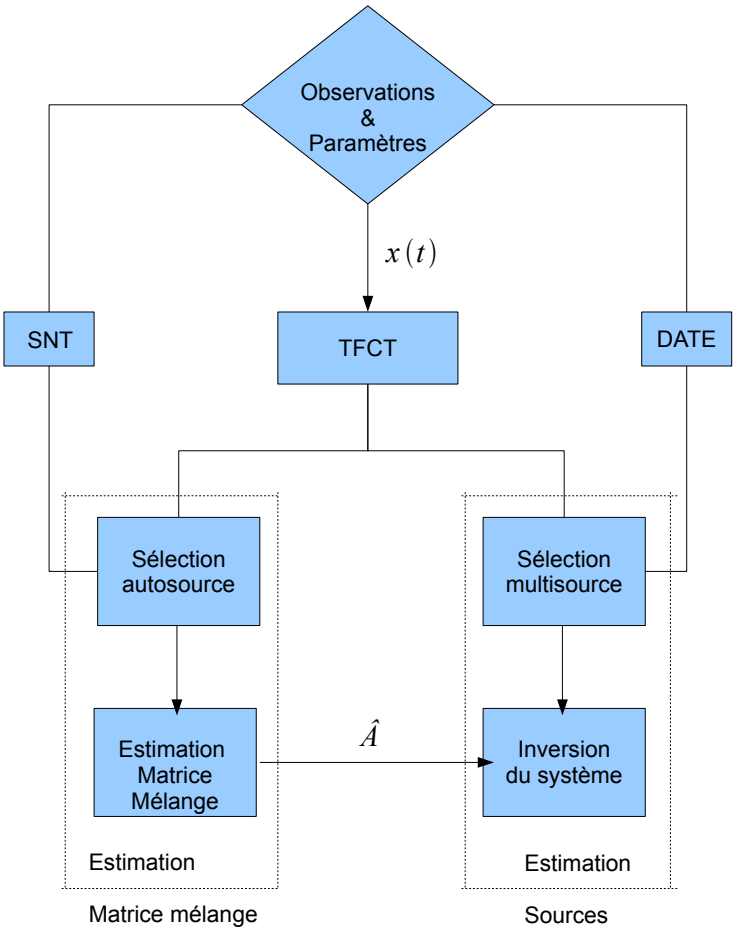


Figure 3.1 — Organigramme des algorithmes de SAS proposés à deux étapes.

comme suit :

$$\begin{cases} \mathcal{H}_0 : & X \sim \mathcal{N}_c(0, \sigma^2) \\ \mathcal{H}_1 : & X = \Theta + B, \end{cases} \tag{3.2}$$

où $B \sim \mathcal{N}_c(0, \sigma^2)$.

La difficulté est la suivante. Bien que $\mathcal{S}_{\boldsymbol{x}_i}(t, f)$ peut raisonnablement être modélisée par une variable aléatoire complexe X, la distribution de probabilité de X ne peut guère être connue. Les critères de Bayes, du minimax et de Neyman-Pearson, qui reposent sur l'emploi du rapport de vraisemblance, ne sont donc pas utilisables. Cependant, cette difficulté peut être surmontée par le recours au modèle parcimonieux.

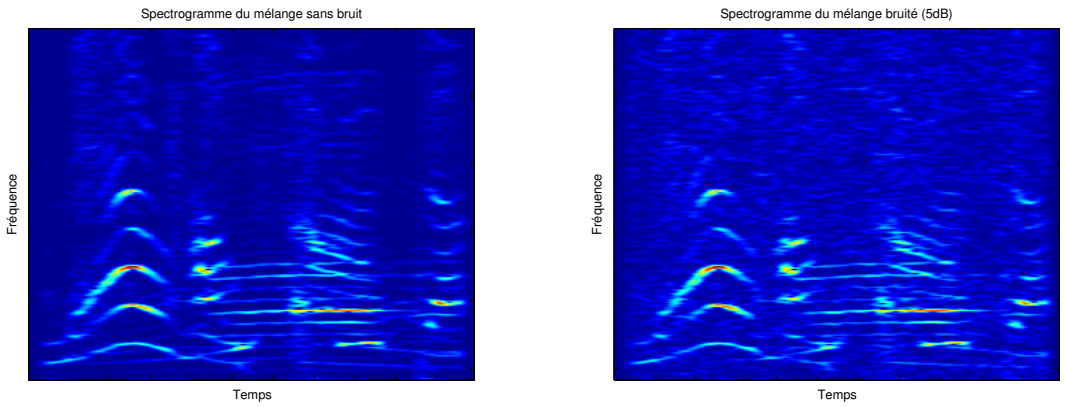
3.1.2 Hypothèse de parcimonie du modèle

Pour prendre la parcimonie du modèle, décrite dans l'état de l'art, en compte dans notre modèle (3.2), nous supposons que :

1. La probabilité d'occurrence de l'hypothèse $\mathcal{H}_1$ est inférieure ou égal à un demi, c'est à dire $\mathbb{P}(\mathcal{H}_1) \leq 1/2$.
2. Il existe une certaine valeur α telle que $|\Theta| > \alpha$ presque sûrement.

Ces deux hypothèses spécifient le modèle parcimonieux en bornant le manque de connaissances a priori sur la distribution du signal.

La figure 3.2-(a) affiche le spectrogramme obtenu par TFCT d'un mélange de signaux audio. Ce spectrogramme présente de nombreuses composantes temps-fréquence de faible amplitude, voire nulle. Lorsque ce mélange est corrompu par un bruit additif indépendant comme dans la figure 3.2-(b) , les petites composantes sont masquées alors que les grandes sont encore visibles. La proportion de ces grandes composantes reste inférieure ou égale à un demi. L'hypothèse de parcimonie du modèle est donc raisonnable.



(a) (b)

Figure 3.2 — (a) Représentation temps-fréquence d'un mélange de signaux audio. Plusieurs coefficients sont de faible amplitude, voire nulle. (b) Représentation temps-fréquence d'un mélange de signaux audio bruité. Les petites valeurs sont masquées par le bruit. Les grandes valeurs restent visible. Ces valeurs ne sont pas énormément affectées par le bruit pour peu que le rapport signal à bruit ne soit pas trop élevé. La proportion de ces coefficients est inférieure ou égale à un demi.

3.1.3 Test à seuil

Dans ce qui suit, nous modélisons la présence ou l'absence du signal Θ par une variable aléatoire de Bernoulli ε, définie sur le même espace probabilisé que X, B et Θ et indépendante de Θ et de B. Nous pouvons écrire le vecteur observé X sous la forme

$$X = \varepsilon\Theta + B$$

On a donc $\mathbb{P}(\mathcal{H}_1) = \mathbb{P}[\varepsilon = 1]$. Etant donnée une observation $\mathcal{S}_{\boldsymbol{x}_i}(t,f)$, le problème est alors de déterminer la valeur de ε pour cette réalisation, c'est à dire la valeur de l'indice de l'hypothèse qui est vraie. Etant donné un test $\mathcal{T}$, c'est à dire, une application de $\mathbb{C}^n$ vers $\{0,1\}$, on dit alors que $\mathcal{T}$ accepte (resp. rejette) l'hypothèse nulle $\mathcal{H}_0$ si $\mathcal{T}(\mathcal{S}_{\boldsymbol{x}_i}(t,f)) = 0$ (resp. $\mathcal{T}(\mathcal{S}_{\boldsymbol{x}_i}(t,f)) = 1$). En d'autres termes, la valeur de $\mathcal{T}(\mathcal{S}_{\boldsymbol{x}_i}(t,f))$ est la décision que l'on prend. La probabilité d'erreur du test est $\mathcal{P}_e = \mathbb{P}(\mathcal{T}(X) \neq \varepsilon)$. Rappelons que dans un cadre de représentation parcimonieuse, nous avons vu que les techniques de seuillage jouaient un rôle important en estimation de signaux parcimonieux. On s'intéressera donc aux tests à seuil. Pour tout réel $h \geq 0$, soit $\mathcal{T}_h$ le test à seuil défini pour tout $u \in \mathbb{C}$ par

$$\mathcal{T}_h(u) = \begin{cases} 1 & \text{si } |u| \geq h \\ 0 & \text{si } |u| < h. \end{cases} \tag{3.3}$$

La probabilité de fausse alarme de ce test est $\mathcal{P}_{FA} = \mathbb{P}(|B| \geq h)$. Cette probabilité ne dépend pas de la loi du signal. Dans de nombreuses applications, telles que le radar, nous voulons assurer une probabilité de fausse alarme constante fixée à l'avance. Sachant que $B \sim \mathcal{N}_c(0,\sigma^2)$, la probabilité de fausse alarme est $\mathcal{P}_{FA} = e^{-h^2/\sigma^2}$.
Ainsi, étant donnée une probabilité de fausse alarme $\mathcal{P}_{FA}$, le seuil qui permet de garantir cette probabilité pour le test à seuil (3.3) est $\lambda = \sigma\sqrt{-\ln \mathcal{P}_{FA}}$.

Pour prendre en compte l'hypothèse de parcimonie du modèle, les auteurs de [121] proposent le seuil de détection $\lambda_D(\alpha,\sigma) = (\sigma/\sqrt{2})\xi(\alpha\sqrt{2}/\sigma)$ comme seuil pour le test (3.3), où, pour tout réel positif ρ, $\xi(\rho) = \mathrm{I}_0^{-1}(e^{\rho^2/2})/\rho$ et I_0 est la fonction de Bessel modifiée de première espèce d'ordre 0. Notant $\mathcal{T}_{PEM}$ le test de vraisemblance. Ce test qui nécessite la connaissance de toutes les distributions de probabilité mises en jeu, est le test qui garantie la probabilité d'erreur minimale sur tous les tests possibles. Les auteurs de [121] montrent alors que la probabilité d'erreur du test $\mathcal{T}_{\lambda_D(\alpha,\sigma)}$ et celle du test $\mathcal{T}_{PEM}$ présentent un même maximum global.

Dans [122], les auteurs proposent de choisir $\alpha = \lambda_u = \sigma\sqrt{\ln 2nL}$ où L est le nombre total de points temps-fréquence par capteur. L'interprétation est la suivante. La figure

3.2 montre que les petites valeurs du signal sont masquées par le bruit. Ainsi, toute valeur ayant le même niveau que le bruit pourra être confondue avec du bruit, seules les grandes valeurs (par rapport au niveau du bruit) seront détectables. L'amplitude minimale du signal à détecter est donc la valeur maximale du bruit. Ceci justifie le choix $\alpha = \lambda_u$, puisque le seuil universel λ_u correspond à une bonne approximation de cette valeur maximale.

Il ne reste plus qu'à estimer l'écart type du bruit avec un algorithme d'estimation robuste. C'est ce que nous décrivons dans la sous section suivante.

3.1.4 Estimation de l'écart type

En ce qui concerne l'estimation de l'écart type du bruit, les solutions usuelles basées sur des estimateurs robustes standards tels le MAD (pour *Median Absolute Deviatione*) [123], le *Trimmed* ou le *Winsorized* [124] ne sont pas appropriés. En effet, en considérant le spectrogramme de la figure 3.2-(b), on peut pressentir que ces estimateurs ne seront pas précis à cause de la proportion élevée des points temps-fréquence où le signal est significativement présent. Par conséquent, la proportion de données aberrantes (*outliers*) par rapport à la distribution de probabilité du bruit est élevée. Dans les articles récents [120], [125], deux nouveaux estimateurs d'écart type du bruit ont été proposés, à savoir le MC-ESE et le DATE. Ces deux estimateurs reposent sur l'hypothèse de modèle parcimonieux. Nous nous contentons de décrire ces algorithmes dans le contexte de séparation de sources, une présentation exhaustive et théorique est fournie dans [120] et [125]. Néanmoins, signalons que le coût de calcul du DATE est très inférieur à celui du MC-ESE et que ces deux algorithmes dérivent du même théorème [3, Théorème 1], d'où leur ressemblance.

MC-ESE

Etant donné $i \in [\![1, n]\!]$, soient $Y^i_{(1)}, Y^i_{(2)}, \ldots, Y^i_{(L)}$ les L valeurs de $\mathcal{S}_{\boldsymbol{x}_j}(t, f)$ telles que $\left|Y^i_{(1)}\right| \leq \left|Y^i_{(2)}\right| \leq \cdots \leq \left|Y^i_{(L)}\right|$.

1. **[Première estimation]** :
   (a) On choisit une valeur réelle positive Q inférieure ou égale à $1 - \frac{L}{4(L/2-1)^2}$.
   (b) On pose $h = 1/\sqrt{4L(1-Q)}$
   (c) On calcule $k_{\min} = L/2 - hL$. D'après l'inégalité de Bienaymé-Chebyshev et l'hypothèse que la probabilité de présence du signal est inférieure à un

demi, la probabilité que le nombre d'observations dû au bruit uniquement soit supérieur à $k_{\min}$ est supérieure ou égale à Q. Dans la partie de simulations numériques, pour le calcul de $k_{\min}$, la valeur de Q a été fixée à 0.95.

(d) On calcule $\sigma_{\min} = \left|Y^i_{(k_{\min})}\right| / \sqrt{2}$ et $\sigma_{\max} = \left|Y^i_{(L)}\right| / \sqrt{2}$. $\sigma_{\min}$ et $\sigma_{\max}$ sont les bornes de l'intervalle de recherche.

(e) Puis, on résout la minimisation suivante sur $[\sigma_{\min}, \sigma_{\max}]$ par un algorithme standard telle que la routine *fminbnd.m* de MATLAB.

$$\widetilde{\sigma}_i = \arg\min_{s} \sup_{\ell \in [\![1,L]\!]} \left\{ \left| \frac{\sum_{r=1}^{L} \left|Y^i_{(r)}\right| \mathrm{I}(\left|Y^i_{(r)}\right| \leq \ell s/L)}{\sum_{r=1}^{L} \mathrm{I}(\left|Y^i_{(r)}\right| \leq \ell s/L)} - \frac{s}{\sqrt{2}} \frac{\Upsilon_1(\ell\sqrt{2}/L)}{\Upsilon_0(\ell\sqrt{2}/L)} \right| \right\}$$

où $\mathrm{I}(A)$ est la fonction indicatrice de de l'événement A, et pour tout $(q, x) \in [0, +\infty[^2$, $\Upsilon_q(x) = \int_0^x t^{q+1} e^{-t^2/2} dt$.

2. **[Seconde estimation]** :

On améliore la première estimation en posant

$$\widehat{\sigma}_i = \eta \, \frac{\sum_{r=1}^{L} \left|Y^i_{(r)}\right|^2 \mathrm{I}(\left|Y^i_{(r)}\right| \leq \widetilde{\sigma}_i)}{\sum_{r=1}^{L} \mathrm{I}(\left|Y^i_{(r)}\right| \leq \widetilde{\sigma}_i)} \tag{3.4}$$

où $\eta = \sqrt{2\Upsilon_0(\sqrt{2})/\Upsilon_2(\sqrt{2})} = 1.5467$.

La valeur finale estimée $\widehat{\sigma}$ de l'écart type du bruit est alors obtenue en moyennant les écarts types estimés $\widehat{\sigma}_i$ sur tous les capteurs, c'est à dire $\widehat{\sigma} = (1/n) \sum_{i=1}^{n} \widehat{\sigma}_i$.

DATE

Etant donné $i \in [\![1, n]\!]$, soient $Y^i_{(1)}, Y^i_{(2)}, \ldots, Y^i_{(L)}$ les L valeurs de $\mathcal{S}_{\boldsymbol{x}_j}(t, f)$ telles que $\left|Y^i_{(1)}\right| \leq \left|Y^i_{(2)}\right| \leq \cdots \leq \left|Y^i_{(L)}\right|$.

1. **[Intervalle de recherche]** :

(a) On choisit une valeur réelle positive Q inférieure ou égale à $1 - \frac{L}{4(L/2-1)^2}$.

(b) On pose $h = 1/\sqrt{4L(1-Q)}$

(c) On calcule $k_{\min} = L/2 - hL$. D'après l'inégalité de Bienaymé-Chebyshev et l'hypothèse que la probabilité de présence du signal est inférieure à un demi, la probabilité que le nombre d'observations dû au bruit uniquement soit supérieur à $k_{\min}$ est supérieure ou égale à Q. Dans la partie de simulations numériques, pour le calcul de $k_{\min}$, la valeur de Q a été fixée à 0.95.

2. **[Existence]** :

S'il existe une valeur k dans $\{k_{\min}, \ldots, L\}$ telle que

$$|Y^i_{(k)}| \leq \big(\mu_i(k)/\kappa\big)\xi\left(\sqrt{\log(2nL)}\right) < |Y^i_{(k+1)}| \tag{3.5}$$

où

$$\mu_i(k) = \begin{cases} \frac{1}{k}\sum_{r=1}^{k} |Y^i_{(r)}| & \text{si } k \neq 0 \\ 0 & \text{si } k = 0, \end{cases} \tag{3.6}$$

on pose $k^* = k$.

Sinon, on pose $k^* = k_{\min}$.

3. **[Valeur]** : L'écart type estimé $\widehat{\sigma}_i$ du $i^{\text{ème}}$ capteur est alors

$$\widehat{\sigma}_i = \mu_i(k^*)/\kappa, \tag{3.7}$$

où $\kappa = \sqrt{2}\Gamma(3/2)$ avec Γ la fonction de Gamma standard.

La valeur finale estimée $\widehat{\sigma}$ de l'écart type du bruit est alors obtenue en moyennant les écarts types estimés $\widehat{\sigma}_i$ sur tous les capteurs, c'est à dire $\widehat{\sigma} = (1/n)\sum_{i=1}^{n} \widehat{\sigma}_i$.

3.2 Sélection autosource

Dans cette section, on s'intéresse à l'étape d'estimation de la matrice de mélange. Nous proposons un test de sélection des points temps-fréquence correspondant à la présence d'une seule source. Pour effectuer cette sélection, nous faisons la distinction entre deux cas : les signaux à faible taux de recouvrement et les signaux à fort taux de recouvrement. Les signaux modulés en fréquence (resp. les signaux audio) sont des exemples typiques de signaux à faible (resp. fort) taux de chevauchement. Notant que, pour chaque classe, les procédures d'estimation proposées ci-dessous ont un coût de calcul raisonnable.

3.2.1 Le cas de signaux avec un faible taux de recouvrement

Etant donné que les sources ont un faible taux de recouvrement, nous supposons que les points temps-fréquence sélectionnés par le test à seuil de la section précédente correspondent, pour la plupart, à des points où une seule source est présente. En d'autres termes, nous négligeons l'effet des (rares) points où les sources se chevauchent sur l'estimation de la matrice de mélange, dans la mesure où l'impact de celles ci est réduit par l'effet de moyenne, essentielle à toute méthode d'estimation de la matrice de mélange.

3.2.2 Le cas de signaux avec un fort taux de recouvrement

Lorsque les signaux se chevauchent de manière significative dans le domaine temps-fréquence, nous ne pouvons plus utiliser la procédure décrite précédemment. En effet, la procédure statistique de la section précédente est destinée à détecter les points temps-fréquence où les sources sont présentes, alors que maintenant il est nécessaire de distinguer les points où une seule source est présente. Pour ce faire, on suppose que :

Hypothèse 3.1 *L'énergie des points temps-fréquence où plusieurs sources sont présentes est inférieure à l'énergie des points temps-fréquence où une seule source est présente.*

Cette hypothèse se justifie grâce à la parcimonie. En effet, pour une source donnée, la parcimonie dans le domaine temps-fréquence impose à peu de points TF d'être actifs, c-à-d d'avoir une grande énergie. En un tel point TF, il est peu probable qu'une autre source y soit très active. L'énergie des points temps-fréquence où une seule source présente est donc supérieure aux autres.

Notre but est donc de détecter les points temps-fréquence où l'énergie du signal est assez grande en présence de bruit. Fondamentalement, ce problème revient à comparer $|\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f)|$ à un certain seuil τ. La valeur de τ^2 représente donc le niveau d'énergie minimal à partir duquel on considère que l'énergie du signal est assez grande pour supposer qu'une seule source est présente au point (t,f). Pour tout réel strictement positif λ, il résulte du lemme [126, Lemme 4, résultat (iii)] que :

$$\mathbb{P}\big[\ |\mathcal{S}_{\boldsymbol{x}}(t,f)| > \lambda \ \big|\ |\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f)| < \tau\ \big] \leq 1 - F_{\chi^2_{2n}(\tau^2/\sigma^2)}\big(\lambda^2/\sigma^2\big)\,, \tag{3.8}$$

où $F_{\chi^2_d(\delta)}(\cdot)$ est la fonction de répartition d'une loi χ^2 non centrée de paramètre δ et à d degrés de liberté. Puisque chaque $\mathcal{S}_{\boldsymbol{x}}(t,f)$ est un vecteur aléatoire complexe de dimension n et, donc, un vecteur aléatoire réelle de dimension $2n$, le degré de liberté dans (3.8) est $2n$.

Etant donné un certain niveau $\gamma \in]0,1[$, il suffit alors de choisir

$$\lambda = \lambda(\tau,\gamma) = \sigma\sqrt{F^{-1}_{\chi^2_{2n}(\tau^2/\sigma^2)}(1-\gamma)}. \tag{3.9}$$

afin de garantir une probabilité de fausse alarme $\mathbb{P}\big[\ |\mathcal{S}_{\boldsymbol{x}}(t,f)| > \lambda \,\big|\, |\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f)| < \tau\ \big]$ inférieure ou égale au niveau γ.

Par conséquent, étant donné un point temps-fréquence (t,f) la règle de décision est :

$$\begin{cases} |\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f)| < \tau & \text{si } |\mathcal{S}_{\boldsymbol{x}}(t,f)| < \lambda(\tau,\gamma) \\ |\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f)| \geq \tau & \text{si } |\mathcal{S}_{\boldsymbol{x}}(t,f)| \geq \lambda(\tau,\gamma). \end{cases} \tag{3.10}$$

Pour estimer la matrice de mélange, nous gardons les points temps-fréquence (t,f) tels que $|\mathcal{S}_{\boldsymbol{x}}(t,f)| \geq \lambda(\tau,\gamma)$. Dans la pratique, sachant que la valeur réelle de σ est inconnue, nous remplaçons cette valeur par l'estimation $\widehat{\sigma}$ fournie par le DATE.
Bien que la détection proposée n'introduisent deux paramètres γ et τ, le choix de la valeur de ces paramètres ne dépend pas du rapport signal à bruit. Le paramètre τ, qui est indépendant du niveau du bruit, peut être fixé par une petite base de données sans bruit. De même, le niveau γ peut être déterminée par des tests préliminaires sur une base de données restreinte mais représentative.

CHAPITRE 4 Résultats de simulations

Afin d'évaluer les modifications apportées sur quelques algorithmes de séparation de sources, nous rapportons dans ce chapitre les résultats des simulations effectuées. On rappelle que le but recherché dans cette partie est de rendre les techniques de séparation moins sensibles aux différents paramètres dont elles dépendent. La robustesse des performances des techniques de séparation par rapport aux choix des seuils et donc par rapport aux données constitue le principal critère d'évaluation. Toutefois, nous montrons que dans certain cas, grâce notamment à l'estimation de l'écart type du bruit, qu'une amélioration des performances de séparation est obtenue.

4.1 Données de simulations

La matrice de mélange est choisie conformément à [127, Eq. (38)] de façon à modéliser les N sources arrivant à l'ensemble des capteurs à des angles différents $\theta_1, \theta_2, \ldots, \theta_M$. Sauf indication contraire, les signaux sources utilisés sont des signaux de parole aléatoirement choisis dans la base de données TI-digits [128]. Cette grande base de données recueillies dans un environnement calme est couramment utilisée dans le traitement de la parole. Les signaux de paroles sont échantillonnés à une fréquence de 8 KHz. La taille de ces signaux est de 8192 échantillons. Nous considérons le cas de 3 mélanges et 4 sources d'angles d'arrivés $\theta_1 = 15\,^\circ$, $\theta_2 = 30\,^\circ$, $\theta_3 = 45\,^\circ$ and $\theta_4 = 75\,^\circ$. Tout au long de ce chapitre, nous nous positionnerons dans un cadre bruité où le bruit est supposé blanc Gaussien. Dans la figure 4.1, la colonne de gauche (figures (a)-(d)) représente temporellement les 4 signaux sources choisis et la colonne de droite (figures (e)-(h)) donne leurs spectrogrammes respectifs.

La figure 4.2 affiche le spectrogramme d'un mélange de signaux sources pour un rapport signal à bruit (RSB) de 10dB. Comme aucun des coefficients de la matrice de mélange $\boldsymbol{A}$ générée n'est nul, le spectrogramme des autres mélanges remplit les mêmes points temps-fréquence avec différentes intensités. Ces spectrogrammes ne sont donc

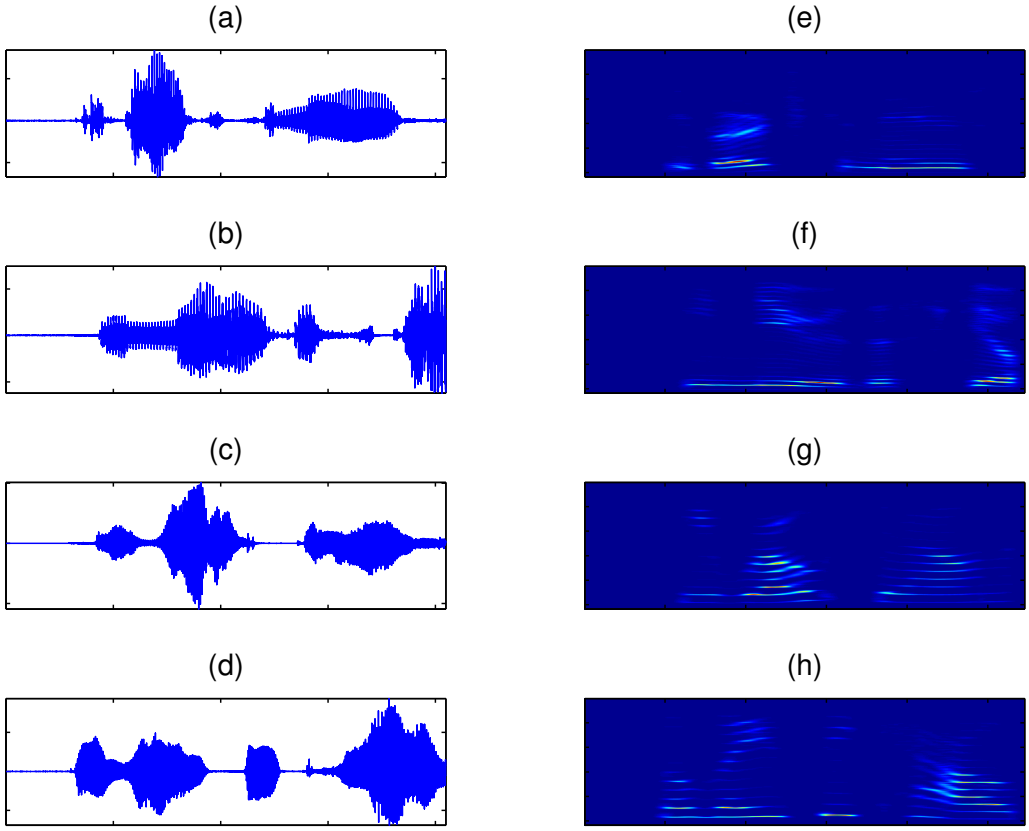


Figure 4.1 — (a)-(d) montrent les formes d'onde des signaux sources originaux dans le domaine temporel, (e)-(h) affichent les spectrogrammes de ces signaux sources dans le domaine temps-fréquence.

pas affichés.

Les deux paramètres requis pour l'estimation de la matrice de mélange sont fixés à $\tau = 4$ and $\gamma = 10^{-3}$ (Voir section discussion).

Enfin, la qualité de la séparation est mesurée par l'erreur quadratique moyenne normalisée (EQMN) :

$$\text{EQMN} = \frac{1}{N}\sum_{j=1}^{N} \min_i \left\{ 10\log_{10}\left(1 - \left(\frac{\langle \hat{s}_i, s_j\rangle}{\|\hat{s}_i\| \cdot \|s_j\|}\right)^2\right)\right\}. \tag{4.1}$$

Tout au long de ce chapitre, les EQMNs sont calculées sur 100 itérations Monté-Carlo.

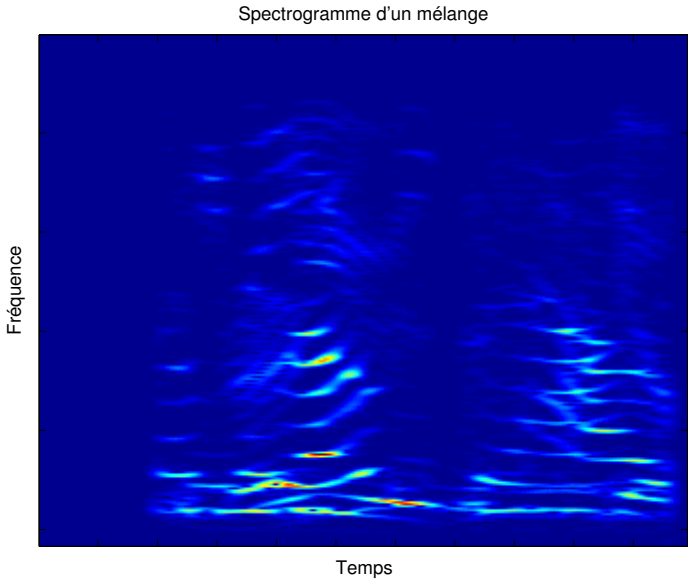


Figure 4.2 — Spectrogramme d'un mélange des quatre sources de la figure 4.1, pour un RSB égal à 10dB).

4.2 La méthode SUBSS

L'algorithme SUBSS modifié est obtenu en utilisant à la fois le DATE et le SNT pour la reconstruction des sources et l'estimation de la matrice de mélange respectivement, comme cela a été expliqué dans le chapitre précédent. Les formes d'ondes des signaux sources estimés par l'algorithme SUBSS modifié sont représentées dans la figure 4.3. Les quatre figures de la colonne de gauche affichent les représentations temps-fréquence des sources estimées dans le cas sans bruit (RSB = 45dB), et les quatre figures de la colonne de droite affichent les représentations temps-fréquence des sources estimées dans le cas bruité avec un RSB égale à 10dB.

Dans la figure 4.4, les performances de séparation de l'algorithme SUBSS modifié avec et sans débruitage sont comparées à celles obtenues par l'algorithme SUBSS standard [1]. Le débruitage mentionné ci-dessus est décrit en annexe comme une estimation linéaire standard.
Nous observons, d'une part, que l'algorithme SUBSS modifié avec débruitage produit de bien meilleurs résultats en terme d'EQMN que ceux obtenus par l'algorithme SUBSS original [1]. D'autre part, les algorithmes SUBSS original et SUBSS modifié sans débruitage ont des performances comparables. Cependant, l'algorithme SUBSS original [1] repose sur le choix de seuils empiriques qui sont manuellement choisis pour chaque

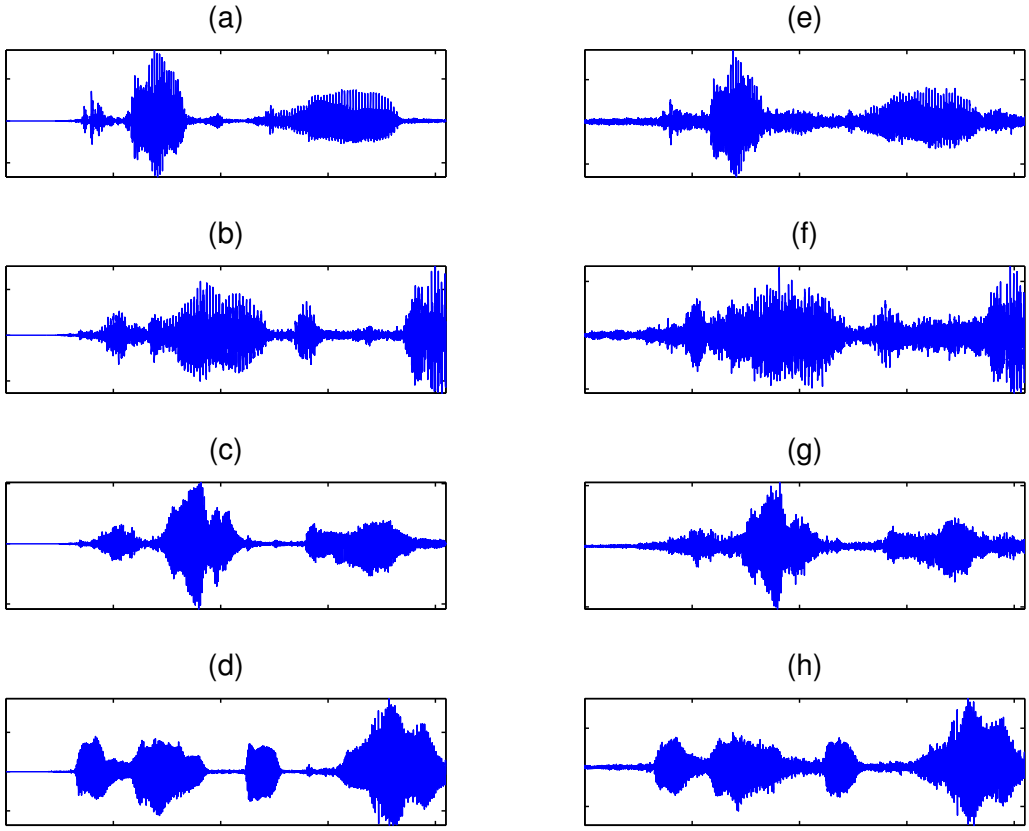


Figure 4.3 — Résultats de simulations : (a)-(d) montrent les formes d'onde des signaux sources reconstruits par l'algorithme SUBSS modifié dans le cas d'un RSB = 45dB, (e)-(h) montrent les formes d'onde des signaux sources reconstruits par l'algorithme SUBSS modifié dans le cas d'un RSB = 10dB.

RSB. En outre, la figure affiche les EQMNs obtenues en utilisant l'estimateur MAD au lieu du DATE dans le SUBSS modifié sans l'étape de débruitage. L'utilisation du MAD à la place du DATE induit une perte significative des performances, ce qui illustre la pertinence du DATE et du modèle parcimonieux.

Dans les figures 4.5 et 4.6, nous présentons les EQMNs obtenues par l'algorithme SUBSS modifié et l'algorithme SUBSS original lorsque le nombre de sources augmente pour un RSB = 10dB et RSB = 20dB. Dans les deux figures, on observe une dégradation de la qualité de séparation, car les hypothèses 2.1 et 2.2 sont plus fortes (une augmentation du nombre de sources implique une augmentation des probabilités d'interférence).

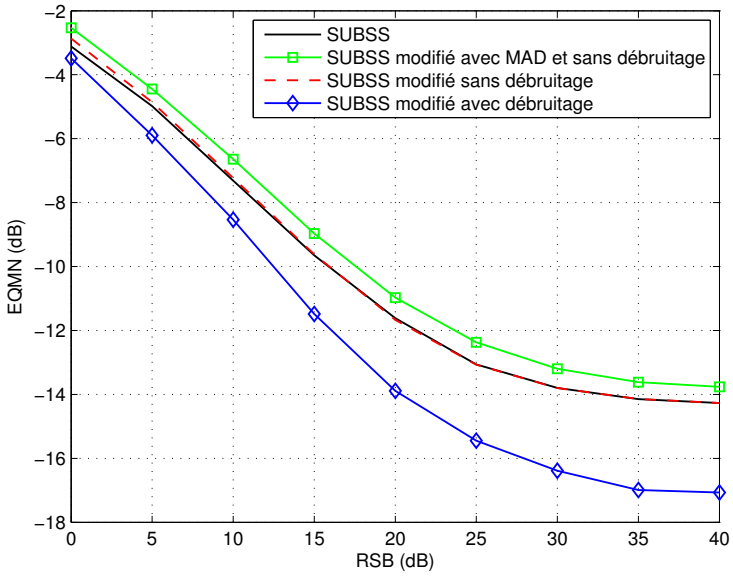


Figure 4.4 — EQMN en fonction du RSB : Comparaison des performances des algorithmes SUBSS standard, SUBSS modifié avec et sans débruitage, et de l'algorithme SUBSS modifié avec l'estimation MAD au lieu du DATE et sans débruitage.

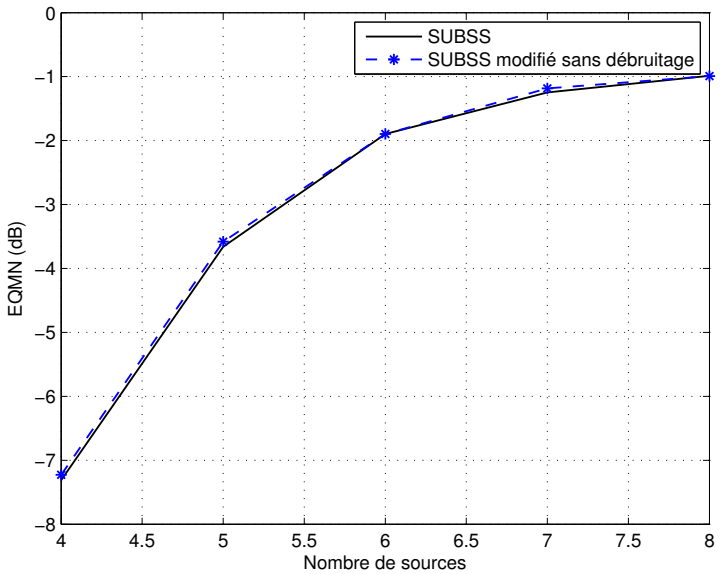


Figure 4.5 — EQMN en fonction du nombre de sources : Comparaison des performances de l'algorithme SUBSS standard et de l'algorithme SUBSS modifié sans débruitage.

Nous allons maintenant examiner le cas des signaux complexes modulés en fréquence. Ceux ci ont été générés en modifiant légèrement la fonction MATLAB *Make-*

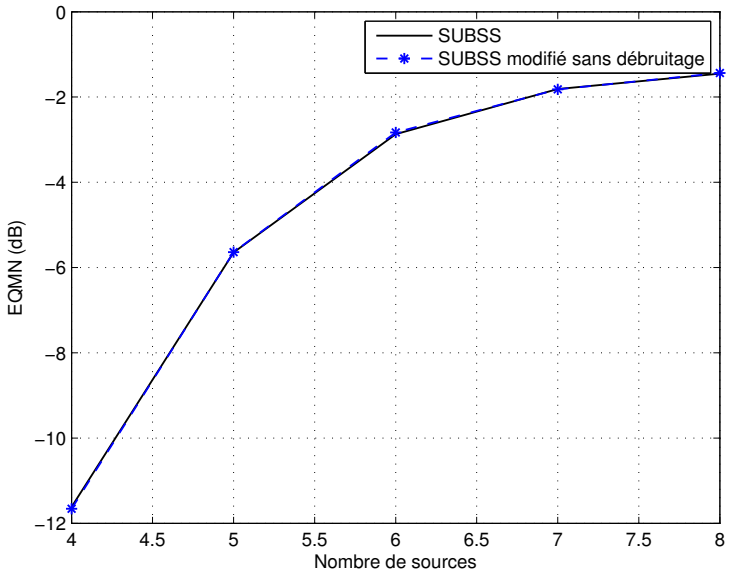


Figure 4.6 — EQMN en fonction du nombre de sources : Comparaison des performances de l'algorithme SUBSS standard et de l'algorithme SUBSS modifié sans débruitage.

Signal.m de la boite à outils *WAVELAB*, de manière à obtenir des signaux complexes modulés en fréquence. Les 4 signaux sources utilisés sont $s_1(t) = \sqrt{t(1-t)}e^{i\frac{\pi \mathrm{T}}{2}t^2}$, $s_2(t) = \sqrt{t(1-t)}e^{-i\frac{\pi \mathrm{T}}{4}t^2}$, $s_3(t) = e^{-i\pi \mathrm{T}t^2}$ et $s_4(t) = e^{i\frac{2}{3}\pi \mathrm{T}t}$, avec $t \in [0,1]$ et une taille d'observation $\mathrm{T} = 8192$ pour chaque signal. Trois de ces signaux modulés en fréquence sont des signaux à fréquence modulée linéaire (FML) et un est une pulsation fréquentielle pure. La figure 4.7 affiche alors les spectrogrammes des quatre signaux à fréquence modulée, tandis que la figure 4.8 représente le spectrogramme d'un mélange bruité (RSB = 10dB) de ces sources. Les spectrogrammes des autres mélanges ne sont pas affichés pour les mêmes raisons que celles indiquées précédemment pour les mélanges de signaux de parole.

Nous évaluons les performances de séparation de l'algorithme SUBSS modifié par rapport à l'algorithme SUBSS original. Comme spécifié dans la section, les seuils utilisés pour l'estimation de la matrice de mélange sont ceux de détection multisource. Par conséquent, aucun paramètre supplémentaire n'est nécessaire. Les résultats obtenus dans la figure 4.9 montre la pertinence de ce choix des seuils, qui s'explique par le fait que les signaux traités présentent très peu de chevauchement dans le domaine temps-fréquence.

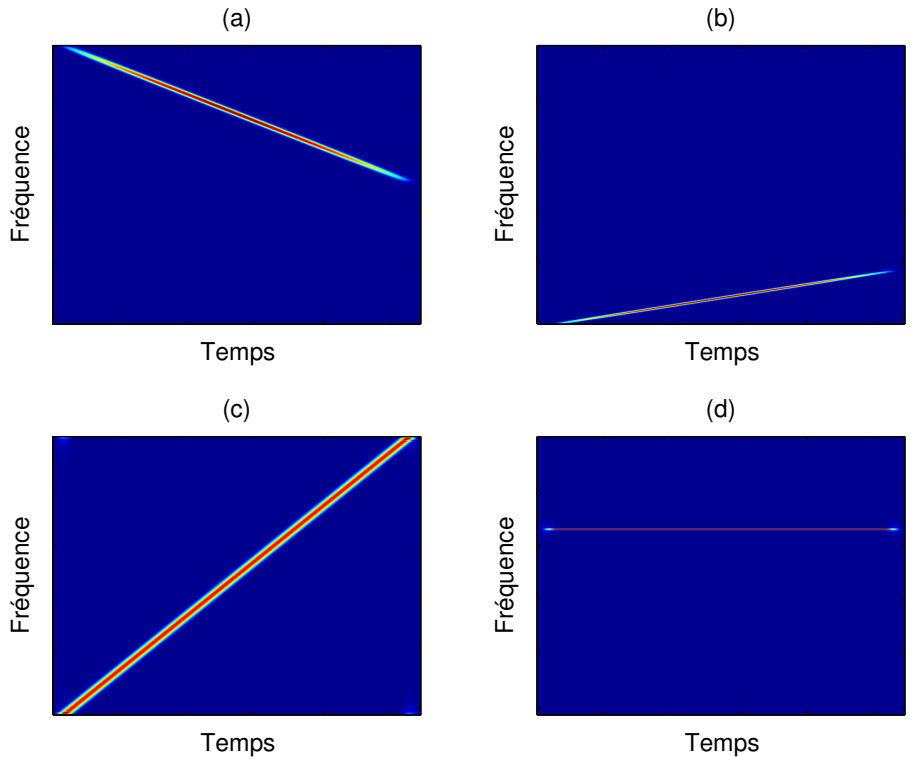


Figure 4.7 — Spectrogramme des quatre signaux sources à fréquence modulée.

4.3 La minimisation ℓ_1

L'estimation de l'écart type du bruit est une autre contribution de notre approche statistique aux méthodes de séparation de sources basées sur la parcimonie. En effet, certaines méthodes supposent connu l'écart type du bruit. En pratique cette hypothèse n'est jamais satisfaite et l'on doit estimer cette valeur. Par exemple, Afin de reconstruire les sources, Bofill et Zibulevsky, dans [108], utilisent la minimisation de la norme ℓ_1 (2.9). Dans le cas bruité, ils proposent de résoudre le problème d'optimisation :

$$\arg\min_{\boldsymbol{z}(t,f)} \frac{1}{2\sigma^2} \left\| \mathcal{S}_{\boldsymbol{x}}(t,f) - \boldsymbol{A}\boldsymbol{z}(t,f) \right\|_2^2 + \left\| \boldsymbol{z}(t,f) \right\|_1 .$$

Soit $\hat{\boldsymbol{z}}(t,f)$ une solution du problème. En posant $\mu = \left\| \mathcal{S}_{\boldsymbol{x}}(t,f) - \boldsymbol{A}\hat{\boldsymbol{z}}(t,f) \right\|_2^2$, alors $\hat{\boldsymbol{z}}(t,f)$ est aussi solution du problème

$$\arg\min_{\boldsymbol{z}(t,f)} \left\| \boldsymbol{z}(t,f) \right\|_1 \quad \text{sous la contrainte} \quad \left\| \mathcal{S}_{\boldsymbol{x}}(t,f) - \boldsymbol{A}\boldsymbol{z}(t,f) \right\|_2^2 \leq \mu.$$

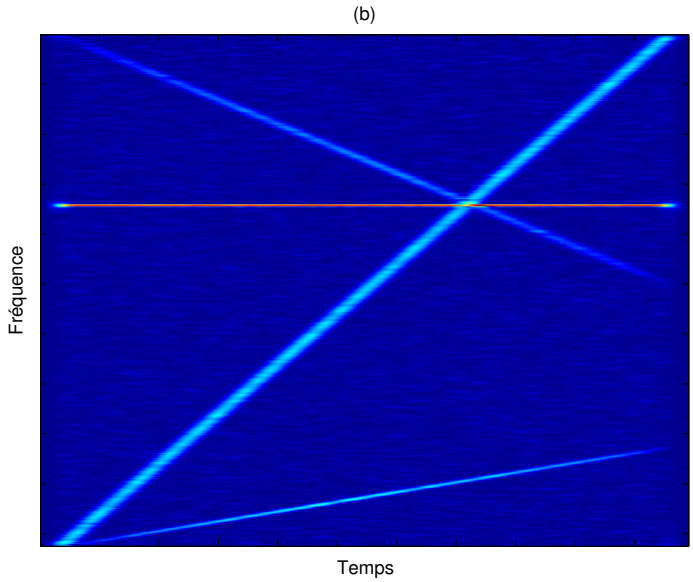


Figure 4.8 — Spectrogramme d'un mélange des quatre sources de la figure 4.7, pour un RSB égal à 10dB).

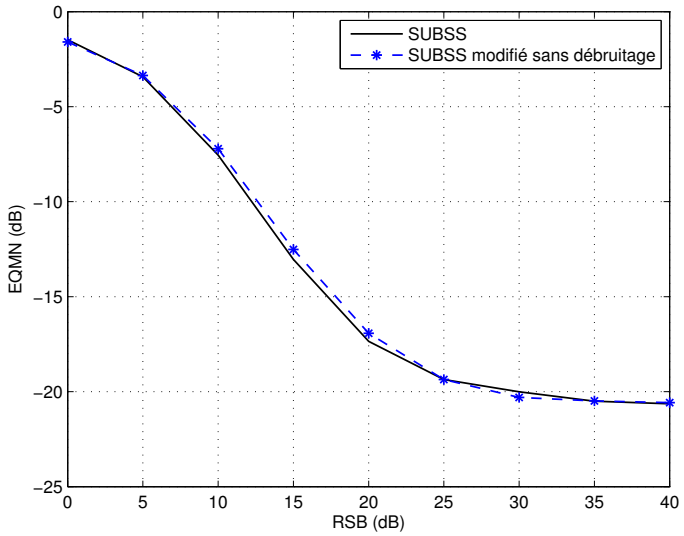


Figure 4.9 — EQMN en fonction du RSB : Comparaison des performances de l'algorithme SUBSS standard et de l'algorithme SUBSS modifié sans débruitage pour les signaux à fréquence modulée.

Suivant [129], nous choisissons de résoudre le problème d'optimisation suivant :

$$\arg\min_{\boldsymbol{z}(t,f)} \|\boldsymbol{z}(t,f)\|_1 \quad \text{sous la contrainte} \quad \|\mathcal{S}_{\boldsymbol{x}}(t,f) - \boldsymbol{A}\boldsymbol{z}(t,f)\|_2^2 \leq \sigma^2(n + 2\sqrt{n}). \quad (4.2)$$

Cette approche peut être améliorée de deux façons. Tout d'abord, par la résolution de ce problème d'optimisation uniquement sur les points temps-fréquence sélectionnés par la procédure multisource décrite dans la section 3.2. Puis, en remplaçant la vraie valeur inconnue de l'écart type du bruit par son estimation fournie par le DATE. Incluant ces deux modifications, nous appellerons cette technique par minimisation ℓ_1 modifiée. A cet égard, la figure 4.10 affiche les performances de séparation de sources par la minimisation (Minimisation ℓ_1) et celles de la minimisation ℓ_1 modifiée. Nous observons un gain significatif en terme d'EQMN, ceci est dû essentiellement à la prise en compte du bruit dans le modèle (problème 4.2). Il est également intéressant de remarquer que l'erreur d'estimation du DATE n'affecte pas de manière significative les performances de séparation par rapport au cas où l'écart type du bruit est parfaitement connu. En effet, la figure 4.10 affiche aussi les performances de séparation lorsque la sélection multisource et la contrainte de minimisation sont à la fois ajustées avec la valeur réelle de l'écart type (Minimisation ℓ_1 modifiée avec oracle). En revanche, nous obtenons une perte des performances lorsque l'on estime l'écart type du bruit à l'aide du MAD (Minimisation ℓ_1 modifiée avec MAD). Ce résultat est dû au fait que le DATE est plus robuste que le MAD.

4.4 La méthode TIFROM

Dans le cas sous déterminé, La méthode TIFROM garantie uniquement une séparation partielle des différentes sources. Par conséquent, afin de mieux évaluer la contribution de nos tests statistiques, nous considérons le cas déterminé où quatre signaux de paroles sont mixés ($n = N = 4$). Les coefficients de la matrice de mélange sont générés indépendamment à partir d'une loi normale. Dans la figure 4.11, nous présentons les EQMNs obtenues par le TIFROM, SNT-TIFROM et le SNT-TIFROM modifié, Plus précisément, SNT-TIFROM utilise le SNT pour une pré-sélection des zones temps-fréquence où une seul source est présente. SNT-TIFROM, comme TIFROM, n'effectue aucune sélection multisource pour l'estimation des sources. En revanche, le SNT-TIFROM effectue une sélection multisource et annule les différentes sources aux points temps-fréquence non sélectionnés. Ces résultats montrent que, sans perte de performance, le SNT-TIFROM permet une sélection des points temps-fréquence autosources sans recourir au seuil empirique requis par le TIFROM d'origine. En outre,

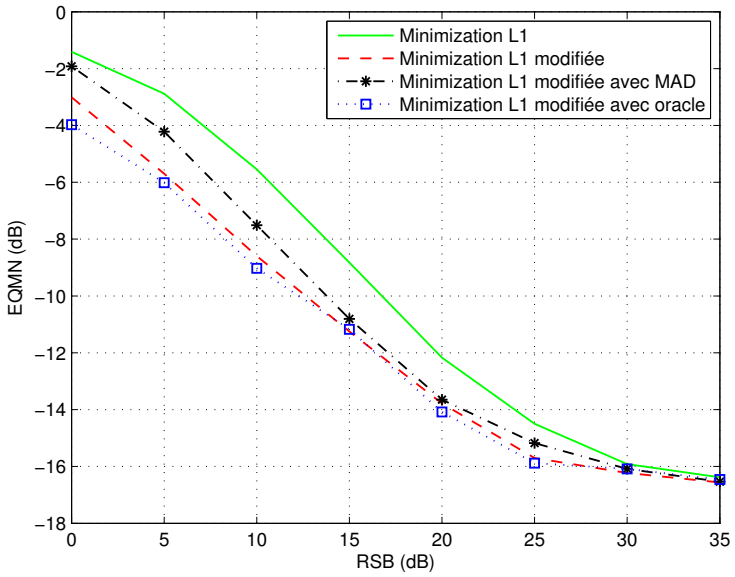


Figure 4.10 — EQMN en fonction du RSB : Comparaison des performances de l'algorithme fondée sur la minimisation (2.9) et les algorithmes cherchants une solution du problème (4.2) après sélection des points multisources, lorsque l'écart type du bruit est connue ou estimée via le DATE ou le MAD.

les performances du SNT-TIFROM modifié montre que le seuil de détection ajusté par le DATE est approprié pour la sélection des points multisources. Nous obtenons un gain pour des faibles RSB car la sélection multisources agit comme un débruitage non-linéaire. Ce gain diminue au fur et à mesure que le RSB augmente.

4.5 Masquage binaire (reconstruction DUET)

Dans cette section, on évalue l'apport de la sélection mutlisource basée sur le seuil de détection qui est ajusté par l'estimation fournie par le DATE à la technique de reconstruction par masquage binaire. Le masquage binaire, définie par l'équation (2.8), attribue une source à chaque point temps-fréquence, même si ce point est dû au bruit seul. Nous appellerons cette technique de reconstruction par "reconstruction DUET" car celle ci a été utilisée dans la méthode de SAS DUET. En supposant la matrice de mélange $\boldsymbol{A}$ connue, la figure 4.12 montre les performances de reconstruction. Dans cette simulation, les signaux traités sont les signaux modulés en fréquence définis ci dessus, de sorte que l'hypothèse d'orthogonalité des supports des sources soit quasiment satisfaite. D'une part, nous effectuons une reconstruction en utilisant l'ensemble des points du plan temps-fréquence (reconstruction DUET). D'autre part, nous considérons

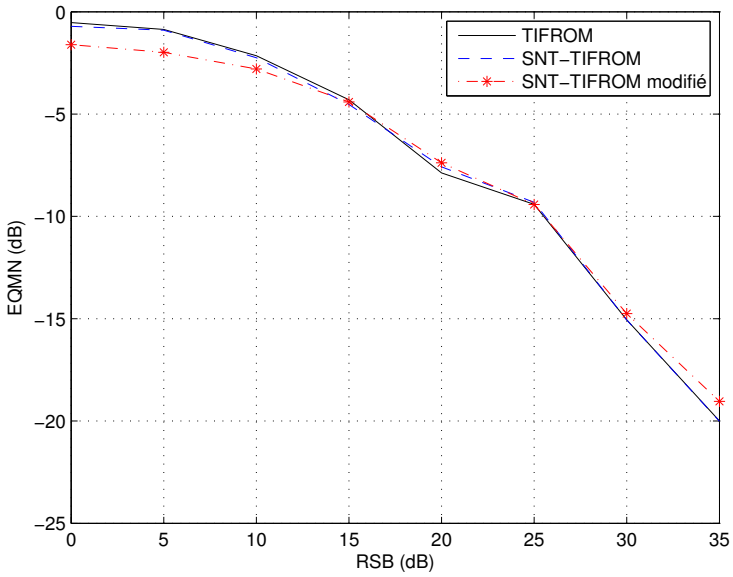


Figure 4.11 — EQMN en fonction du RSB : Comparaison des performances de la méthode TIFROM standard , du SNT-TIFROM et du SNT-TIFROM modifié.

la technique de reconstruction DUET modifiée, qui applique le masquage binaire (2.8) uniquement aux points multisources sélectionnés par notre test d'hypothèse décrit dans la section 3.2. Les résultats sont similaires à ceux précédemment obtenus dans le cas de la méthode TIFROM et ses versions modifiées. Ici, l'apport de la sélection multisource est plus élevé, car la représentation temps-fréquence des signaux modulés en fréquence est plus parcimonieuse que celle des signaux audio.

4.6 Mélange convolutif

Il existe une grande variété de stratégies possibles permettant de traiter le cadre convolutif. Ce cadre est plus réaliste que celui d'un mélange instantané dans différentes applications. Dans le cas de mélanges convolutifs, il est difficile de définir une famille de méthodes bien établie telle que celle évoquée dans cette partie. Cependant, malgré cette variété [130], on peut s'attendre à ce que le cadre statistique proposé dans le chapitre 3 puissent contribuer aux différentes méthodes de séparation de sources dans le cadre de mélanges convolutifs, notamment celles procédant dans le domaine où les sources sont parcimonieuses et pour lesquelles une identification des points temps-fréquence où seul le bruit est présent peut s'avérer utile. Par exemple, citons la méthode SUBSS convolutive présentée dans [131]. La procédure de détection des points multisources,

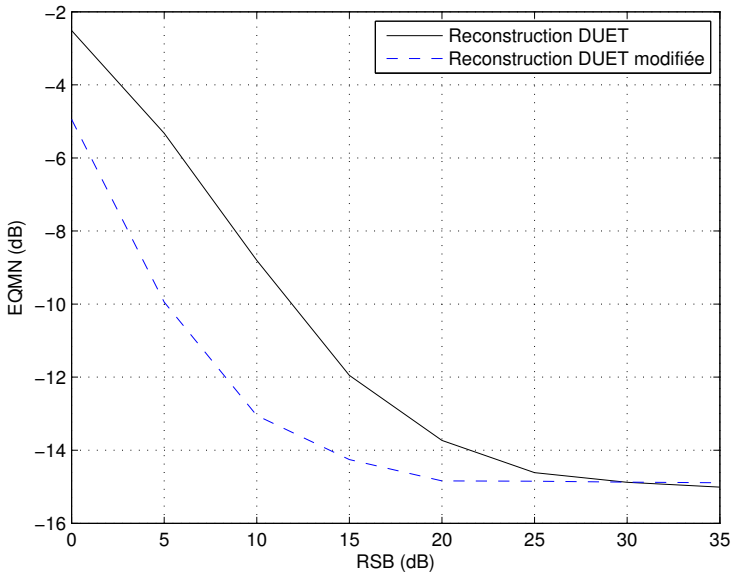


Figure 4.12 — EQMN en fonction du RSB : Comparaison des performances de l'approche de reconstruction DUET et l'approche de reconstruction DUET modifiée.

décrite dans la section 3.1, peut se greffer directement à cette méthode de SAS. La méthode SUBSS convolutive modifiée obtenue ne dépend plus de paramètre empirique dans la sélection multisource. Toutefois, la figure 4.13 montre que la suppression du paramètre empirique dont dépend la méthode SUBSS convolutive [131] ne dégrade pas les performances de séparation.

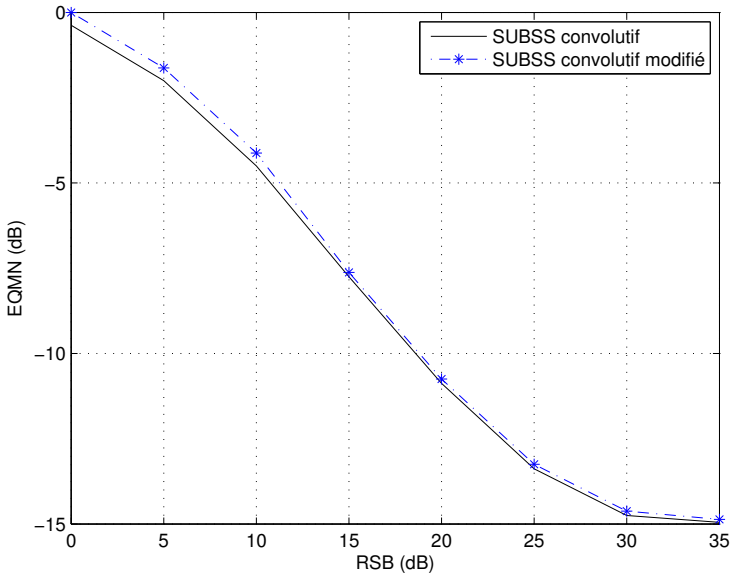


Figure 4.13 — EQMN en fonction du RSB : Comparaison des performances de la méthode SUBSS convolutive standard et la SUBSS convolutive modifiée. Chaque mélange est une somme de signaux sources filtrées dans lequel chaque filtre est un RIF généré aléatoirement d'ordre 4

CHAPITRE 5 Conclusions

Dans cette partie, les algorithmes présentés contribuent à la SAS dans le cas d'un mélange sous-déterminé, en évitant les choix empiriques de paramètres présents dans les méthodes usuelles de séparation basées sur la parcimonie. Notre premier algorithme visant à sélectionner les points temps-fréquence appropriés pour l'estimation des sources est entièrement automatique. Le second, consacré à l'estimation de la matrice de mélange nécessite deux paramètres seulement, quelque soit le RSB choisi.

Ces algorithmes sont très généraux. Ils ne sont pas dédiés à une méthode donnée de SAS basée sur l'hypothèse de parcimonie. Ils sont simples à appliquer, sans aucun ajustement. D'après les résultats du chapitre 4, nos procédures permettent d'améliorer, de simplifier ou d'apporter une robustesse par rapport aux données aux méthodes de SAS fondées sur la parcimonie des sources.

Plus précisément, la procédure de détection basée sur l'hypothèse de parcimonie du modèle, décrite dans la section 3.1, peut être utilisée comme un pré-traitement automatisé de la sélection multisource. Par exemple, la détection temps-fréquence dans [1] nécessite une valeur de seuil pour chaque valeurs du RSB. La procédure de détection de la section 3.1 permet alors d'éviter ce choix des paramètres empiriques, ce qui apporte une simplification significative et une certaine robustesse de la méthode. Utilisée comme un pré-traitement pour TIFROM [106], qui essentiellement ne comprend aucune phase de sélection de points temps-fréquence multisources, la sélection multisource que l'on propose améliore les performances de séparation.

Pour l'estimation de la matrice de mélange, notre approche, décrite dans la section 3.2, est basée sur l'hypothèse 3.1 et nécessite le choix de deux paramètres, qui sont, la tolérance et la probabilité de fausse alarme. Contrairement à [1] par exemple, ces paramètres ne varient pas en fonction du RSB. En outre, les hypothèses formulées par TIFROM peuvent être relaxer en utilisant la sélection autosource de la section 3.2. Il est également intéressant de noter que les paramètres dont nous avons besoin pour l'estimation de la matrice de mélange ont un sens physique, ce qui n'est pas le cas pour

certaines méthodes standards de SAS basées sur la parcimonie.

Deuxième partie

Acquisition compressée des signaux à alphabet fini

Dans le premier chapitre, nous avons vu le rôle important que jouait la parcimonie dans différents domaines. L'acquisition compressée contrairement aux autres domaines d'applications exploitant la parcimonie est un domaine qui a vu le jour grâce à la parcimonie. Dans cette partie, nous allons nous intéresser à l'acquisition compressée des signaux à alphabet fini. Etant donnés un alphabet fini $\mathcal{F} = \{a_1, \cdots, a_p\}$ et un signal $\boldsymbol{f} \in \mathcal{F}^N$, l'objectif est donc de reconstruire $\boldsymbol{f}$ à partir des observations $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{f}$ où $\boldsymbol{A}$ est une matrice connue (aléatoire) de dimension $n \times N$.

Après une brève analyse de l'état de l'art de l'acquisition compressée et une formalisation du problème d'acquisition compressée des signaux à alphabet fini, le premier chapitre positionne notre travail par rapport à l'état de l'art, notamment par rapport à l'approche proposée par Mangasarian [16] et introduit les différentes notations utilisées dans cette partie.

Dans le second chapitre, d'une part, nous établissons une condition suffisante qui permet de rendre le problème soluble. Cette condition étend celle de Mangasarian au cas d'alphabet fini quelconque. D'autre part, nous présentons deux approches de reconstruction. Ces deux méthodes ont en commun une relaxation convexe de la norme ℓ_0 en norme ℓ_1. La première méthode est fondée sur la régularisation, où l'on introduit une fonction objective qui évalue le critère basé sur l'alphabet fini. La seconde méthode est fondée sur la parcimonie. Nous dérivons dans ce cas une formulation analogue à celle de la reconstruction de signaux parcimonieux à partir de mesures incomplètes. Nous donnerons les conditions nécessaires et suffisantes à la reconstruction de signal à alphabet fini pour cette approche.

Le troisième chapitre traite de l'évaluation des performances numériques des deux approches.

Une partie du contenu de cette seconde partie a fait l'objet d'une publication dans la conférence *ISSPA* 2012 [17]. Une première version de cet article a été présenté à *SPARS* 2011 [18]. Une version détaillée de l'article sera soumise à la revue *Elsevier Digital Signal Processing* [19].

CHAPITRE 6 Introduction

6.1 Acquisition compressée

Dans la première partie, nous avons vu que la parcimonie d'un signal pouvait aider à réduire la dimension d'un signal, tout en gardant l'information nécessaire à sa reconstruction ou à son approximation. Ce processus de réduction de dimension dans l'approximation par les k-meilleurs termes est adaptatif dans le sens où les indices des coefficients que l'on garde dépendent de $\boldsymbol{x}$. L'objectif principal de l'acquisition compressée est double. Le premier est de réduire la dimension de $\boldsymbol{x}$ de façon non-adaptative. Cette propriété est connue sous le nom de l'universalité de l'acquisition. La seconde part du constat que dans la compression par transformation, afin de retenir les meilleurs termes, il est nécessaire de calculer tous les coefficients du signal dans la nouvelle base. L'acquisition compressée calcule directement les coefficients nécessaires à la reconstruction. Nul besoin de jeter des coefficients après leur calcul.
Supposons qu'il existe une représentation parcimonieuse $\boldsymbol{\theta}$ de $\boldsymbol{x}$ dans une base orthonormée $\mathcal{B}$ telle que $\|\boldsymbol{\theta}\|_0 = k$. Soit $\boldsymbol{A}$ une matrice de mesures de dimension $n \times N$ tel que $n < N$, il s'agit de reconstruire $\boldsymbol{x}$ à partir de $\boldsymbol{y} = \boldsymbol{Ax} = \boldsymbol{AB\theta}$ sachant que $\boldsymbol{\theta}$ est parcimonieux.

Quitte à se positionner dans le domaine $\mathcal{B}$ où le signal est parcimonieux, on peut supposer que $\boldsymbol{x}$ est parcimonieux d'ordre k, en d'autres termes $\boldsymbol{x} = \boldsymbol{\theta}$.
L'approximation par les k-meilleurs termes permet de réduire la dimension de $\boldsymbol{x}$ sans aucune perte d'informations. Cependant, la matrice de mesures $\boldsymbol{A}$ de dimension $k \times N$, qui consiste à sélectionner les k coefficients non nuls de $\boldsymbol{x}$ dépend du signal $\boldsymbol{x}$. La modification du signal implique une modification de la matrice de mélange.

Dans l'acquisition compressée, on cherche à construire une matrice de mesures universelle, ne dépendant pas du signal choisi. Il est évident, du à la non-adaptativité, que n est forcément plus grand que k.
L'acquisition compressée s'intéresse à la question suivante : Quelles sont les conditions

qui permettent d'obtenir des "performances identiques" entre l'acquisition compressée et l'approximation par les k-meilleurs termes, pour des nombres k et N fixés ? Ses conditions se traduisent par des propriétés sur la matrice $\boldsymbol{A}$ qui imposeront des conditions sur le nombre minimal de mesures.

Sachant que $\boldsymbol{x}$, vérifiant $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, est parcimonieux. La procédure de reconstruction dans l'acquisition compressée consiste à résoudre l'optimisation suivante :

$$(P_0) \ : \qquad \min_{\boldsymbol{c}} \ \|\boldsymbol{c}\|_0 \ \text{ sous la contrainte } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{c}$$

Comme nous l'avons noté, nous nous intéressons à la relaxation ℓ_1 du problème (P_0)

$$(P_1) \ : \qquad \min_{\boldsymbol{c}} \ \|\boldsymbol{c}\|_1 \ \text{ sous la contrainte } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{c}$$

Il s'agit alors dans un premier lieu de donner les conditions d'unicité du problème (P_0) puis les conditions d'équivalence des deux problèmes. Dans le but d'énoncer quelques résultats, nous introduisons les définitions suivantes :

Définition 6.1 *: **NullSpace Property (NSP) [132]***
On dit qu'une matrice $\boldsymbol{A}$ de dimension $n \times N$ satisfait la propriété NSP d'ordre k si

$$\forall\, \boldsymbol{h} \in \mathrm{Ker}\{\boldsymbol{A}\},\ \forall\, \boldsymbol{T} \subset \{1, \ldots, N\} \ tq\ \#(\boldsymbol{T}) = k,\ on\ a\ \ \|\boldsymbol{h}_T\|_1 < \|\boldsymbol{h}_{T^c}\|_1$$

Définition 6.2 *: **Restricted Isometry Property (RIP) [133]***
Soit $\boldsymbol{A}$ une matrice de dimension $n \times N$. Quelque soit l'entier $1 \leq k \leq N$, on définit la constante d'isométrie restreinte d'ordre k, notée δ_k, la plus petite quantité telle que $\boldsymbol{A}$ vérifie :

$$(1 - \delta_k)\|\boldsymbol{c}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{c}\|_2^2 \leq (1 + \delta_k)\|\boldsymbol{c}\|_2^2$$

pour tous les vecteurs parcimonieux $\boldsymbol{c}$ d'ordre k.

La constante δ_k mesure à quel point les vecteurs colonnes de $\boldsymbol{A}$, notés $(\boldsymbol{a}_j)_{1 \leq j \leq N}$ sont proches de former un système orthonormé lorsque l'on restreint notre attention aux combinaisons linéaires de pas plus de k vecteurs.

Les théorèmes suivants donnent des conditions d'unicité de la solution du problème (P_0) et d'équivalence des deux problèmes (P_0) et (P_1)

Théorème 6.1 *Soit $k \geq 1$, et soit $\boldsymbol{x}$ un vecteur parcimonieux d'ordre k tel que $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Notons δ_r la constante d'isométrie d'ordre r de $\boldsymbol{A}$. Si $\delta_{2k} < 1$, alors $\boldsymbol{x}$ est l'unique solution du problème (P_0).*

Théorème 6.2 *Soit $k \geq 1$, $\boldsymbol{A}$ satisfait la propriété NSP d'ordre k si et seulement si pour tout vecteur parcimonieux d'ordre k solution du problème (P_0) il est l'unique solution du problème (P_1).*

Théorème 6.3 *Soit $k \geq 1$, et soit $\boldsymbol{x}$ un vecteur parcimonieux d'ordre k tel que $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Notons δ_r la constante d'isométrie d'ordre r de $\boldsymbol{A}$. Si $\delta_{2k} < \sqrt{2} - 1$, alors le problème (P_1) admet une unique solution qui vaut $\boldsymbol{x}$.*

La condition RIP impose à la matrice de mesures un nombre minimal de mesures. Ce nombre dépend du choix des éléments de la matrice. On peut citer comme exemple classiques les matrices suivantes :

- Matrice Gaussienne [6], c'est-à-dire lorsque les éléments de $\boldsymbol{A}$ sont *i.i.d* selon une loi normal centrée de variance $1/n$: n est de l'ordre de $k \log(N/k)$;

- Matrice Bernoulli [6], c'est-à-dire lorsque les éléments de $\boldsymbol{A}$ sont *i.i.d* selon une loi de Bernoulli de paramètre $1/2$: n est de l'ordre de $k \log(N/k)$;

- Matrice Fourier sous déterminée [134], c'est-à-dire une sous matrice de la matrice de Fourier où les lignes sont choisies uniformément : n est de l'ordre de $k \log^4(N)$.

Signalons qu'il existe d'autres conditions, telle que la cohérence mutuelle, qui permettent de garantir l'unicité de la solution du problème (P_0) et les bonnes performances de la relaxation ℓ_1. Toutefois, le NSP et le RIP sont les outils privilégiés dans la littérature pour étudier l'équivalence des problèmes (P_0) et (P_1) la cohérence mutuelle.
Des résultats similaires permettent d'étendre cette étude au cas où le vecteur $\boldsymbol{x}$ n'est pas strictement parcimonieux [135] et/ou lorsque les observations sont bruitées [129, 44].

6.2 Formalisme d'acquisition compressée de signaux à alphabet fini

Nous nous plaçons dans le cadre d'acquisition compressée de signaux à alphabet fini. Soit $\mathcal{F} = \{a_1, a_2, \cdots, a_p\}$ un ensemble discret de cardinal p, on supposera, tout au long de ce chapitre, que les $(a_i)_{1 \leq i \leq p}$ sont différents et équiprobables, et soit $\boldsymbol{f} \in \mathcal{F}^N$. On se donne $n < N$ mesures linéaires prises sur $\boldsymbol{f}$

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{f}$$

où $\boldsymbol{A}$ est la matrice de mesures (d'observations) de dimension $n \times N$ choisie indépendamment de $\boldsymbol{f}$. L'objectif est de reconstruire $\boldsymbol{f}$ comme solution du système

$$\begin{cases} \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \\ \boldsymbol{x} \in \mathcal{F}^N \end{cases} \tag{6.1}$$

Dans ce but, on se doit de répondre aux deux questions suivantes :

- Sous quelles conditions la reconstruction est possible ? Autrement dit, sous quelles conditions $\boldsymbol{f}$ est l'unique solution de (6.1) ?

- Dans le cas où ses conditions sont satisfaites, quelle est la procédure qui nous permet de reconstruire $\boldsymbol{f}$?

La difficulté de la deuxième question réside dans le fait que l'on cherche une procédure de reconstruction à complexité polynomiale. En effet, une recherche exhaustive de vecteur dans $\mathcal{F}^N$ attaché aux données permet de reconstruire $\boldsymbol{f}$. Cependant, un tel algorithme a une complexité exponentielle en fonction de N. En fait, la simple vérification que $\boldsymbol{f}$ est l'unique solution de (6.1) est un problème NP-difficile [136].

6.3 Cas particulier

Dans les articles [137], [16], les auteurs étudient la reconstruction de $\boldsymbol{f}$ à partir du système (6.1) dans le cas particulier où $p = 2$ et $\mathcal{F} = \{-1, 1\}$. Ils proposent de reconstruire $\boldsymbol{f}$ en minimisant la norme ℓ_∞

$$\arg\min_x \|\boldsymbol{x}\|_\infty \quad \text{s.t} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{6.2}$$

Pour énoncer le résultat principal de cette étude, nous introduisons tout d'abord la définition suivante :

Définition 6.3 *: Matrice aléatoire générique [16]*
$\boldsymbol{A}$ est une matrice dite aléatoire générique de dimension $n \times N$ si tous les ensembles constitués de n colonnes de la matrice sont linéairement indépendants avec une probabilité égale à 1 et que toutes les colonnes sont distribuées symétriquement autour de l'origine.

Lorsque $\boldsymbol{A}$ est une matrice aléatoire générique, le théorème suivant calcule la probabilité que l'optimisation (6.2) admette une unique solution dans $\{-1,1\}^N$.

Théorème 6.4 (Mangasarian et al.) *Soient $\boldsymbol{A}$ une matrice aléatoire générique de dimension $n \times N$ et $\boldsymbol{f} \in \{-1,1\}^N$.*
Si $\boldsymbol{f}$ est une solution de (6.2), alors la probabilité que $\boldsymbol{f}$ soit l'unique solution de (6.2) est donnée par

$$p_{n,N} = 2^{-N+1} \sum_{i=0}^{n-1} \binom{N-1}{i}$$

Cela implique que, pour (n, N) proportionnellement large, la probabilité de reconstruction exacte de $\boldsymbol{f}$ via (6.2) tend vers un ou zéro selon que $\frac{n}{N} > \frac{1}{2}$ ou $\frac{n}{N} < \frac{1}{2}$ respectivement.

La preuve de ce théorème n'est pas reproduite, mais nous rappelons le principal résultat sur lequel elle est fondée et qui va nous servir dans la suite du manuscrit.

Lemme 6.1 (Wendel [138]) *Soient $\boldsymbol{A}$ une matrice aléatoire générique de dimension $n \times N$. La probabilité que toutes les colonnes de la matrice $\boldsymbol{A}$ soient dans le même hémisphère est égale à*

$$p_{n,N} = 2^{-N+1} \sum_{i=0}^{n-1} \binom{N-1}{i}$$

Finalement, notons que la probabilité $p_{n,N}$ est une fonction croissante qui passe rapidement de 0 à 1 au voisinage de $n \approx \frac{N}{2}$ (voir l'annexe B.1).

Remarques sur la procédure :

- La procédure de reconstruction proposée dans [16] consiste à résoudre le problème (6.2) et vérifier que la valeur minimale de la norme ℓ_∞ est égale à 1. Cette vérification constitue une faiblesse de cette procédure de reconstruction.

- La procédure de reconstruction est spécifique au cas $p = 2$ et $\mathcal{F} = \{-1, 1\}$.

Dans cette partie, nous cherchons à étendre cette étude pour des signaux à alphabet fini quelconque. Dans le cas où $\mathcal{F} = \{a_1, a_2\}$, un changement de variable permet de se ramener à l'alphabet $\{-1, 1\}$. Cependant, une généralisation de la procédure (6.2) dans le cas où $p > 2$ n'est pas triviale.

6.4 Notations

1. Soit $m \in \mathbb{N}^*$, on note $\mathbf{0}_m$ et $\mathbf{1}_m$ les vecteurs colonnes de dimension m dont tous les coefficients valent 0 et 1 respectivement.

2. Soit $\boldsymbol{x} \in \mathbb{R}^m$, on note :

$$\begin{aligned}
&\boldsymbol{x} \geqq \mathbf{0} \text{ si, pour tout } i \in [\![1, m]\!],\ \boldsymbol{x}_i \geq 0 \\
&\boldsymbol{x} \geq \mathbf{0} \text{ si } \boldsymbol{x} \geqq \mathbf{0} \text{ et } \boldsymbol{x} \neq \mathbf{0} \\
&\boldsymbol{x} > \mathbf{0} \text{ si, pour tout } i \in [\![1, m]\!],\ \boldsymbol{x}_i > 0
\end{aligned}$$

3. Enfin, on définit les deux matrices $\boldsymbol{B_a}$, $\boldsymbol{B_1}$ de taille $N \times Np$ par,

$$\boldsymbol{B_a} = \begin{pmatrix} \boldsymbol{a} & \mathbf{0}_p & \dots & \mathbf{0}_p \\ \mathbf{0}_p & \boldsymbol{a} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_p \\ \mathbf{0}_p & \dots & \mathbf{0}_p & \boldsymbol{a} \end{pmatrix}^T \qquad \boldsymbol{B_1} = \begin{pmatrix} \mathbf{1}_p & \mathbf{0}_p & \dots & \mathbf{0}_p \\ \mathbf{0}_p & \mathbf{1}_p & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_p \\ \mathbf{0}_p & \dots & \mathbf{0}_p & \mathbf{1}_p \end{pmatrix}^T$$

où $\boldsymbol{a} = (a_1, \cdots, a_p)^T$.

CHAPITRE 7 Solvabilité et méthodes de reconstruction

7.1 Conditions de solvabilité

Nous tentons dans cette section de répondre à la première question, c'est à dire sous quelles conditions $\boldsymbol{f}$ est l'unique vecteur de $\mathcal{F}^N$ solution de l'équation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{7.1}$$

7.1.1 Hypothèse de solvabilité

L'extraction ou la séparation de signaux à alphabet fini à partir d'un système linéaire sous déterminé sont des sujets qui s'apparentent à notre problème, c'est à dire l'acquisition compressée de signaux à alphabet fini. La plupart des articles sur ces sujets [60], [59], [139], supposent que :

($\mathcal{H}_1$) : Le modèle (7.1) est $\mathcal{F}$-bien posé,

dans le sens où la matrice $\boldsymbol{A}$ est injective sur $\mathcal{F}^N$: c'est à dire si $\boldsymbol{g}, \boldsymbol{g'} \in \mathcal{F}^N$ tel que $\boldsymbol{A}\boldsymbol{g} = \boldsymbol{A}\boldsymbol{g'}$, alors $\boldsymbol{g} = \boldsymbol{g'}$ [59].

Il est clair que, sans cette hypothèse, aucun algorithme n'est capable d'identifier le signal original. En effet, on sera en présence de plusieurs solutions du modèle (7.1) vérifiant les mêmes contraintes.

Cette hypothèse est équivalente à l'unicité de la solution $\boldsymbol{f}$ du système (6.1) que l'on se propose d'étudier.

7.1.2 Condition suffisante

Le théorème suivant donne une condition suffisante pour que l'hypothèse $(\mathcal{H}_1)$ soit satisfaite.

Théorème 7.1 *Soient $\boldsymbol{A}$ une matrice aléatoire générique de dimension $n \times N$ et $\boldsymbol{f} \in \mathcal{F}^N$.*
Pour (n, N) proportionnellement large, l'hypothèse $(\mathcal{H}_1)$ est satisfaite si

$$\frac{n}{N} > \frac{p-1}{p}.$$

La preuve de ce théorème est fondée sur un résultat démontré par Donoho dans [140]. Avant d'énoncer ce résultat, rappelons la définition suivante :

Définition 7.1 *: **Vecteur simple d'ordre k d'un hypercube [140]***
Soit $\boldsymbol{x} \in [\alpha, \beta]^N$, $\boldsymbol{x}$ est un vecteur simple d'ordre k si tous ses coefficients sont égal à α ou β, sauf au plus k éléments.

Le lemme suivant calcule la probabilité d'unicité d'une solution de $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ dans un hypercube lorsque l'équation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ admet comme solution un vecteur simple de l'hypercube. La preuve du lemme n'est pas reproduite, mais le principal résultat est résumé

Lemme 7.1 (Donoho et al.) *Soit $\boldsymbol{A}$ une matrice aléatoire générique de dimension $n \times N$ et soit $\hat{\boldsymbol{x}}$ un vecteur simple d'ordre k de $[\alpha, \beta]^N$ tel que $\boldsymbol{y} = \boldsymbol{A}\hat{\boldsymbol{x}}$.*
La probabilité que $\hat{\boldsymbol{x}}$ soit l'unique solution du système $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x} \in [\alpha, \beta]^N$ est

$$1 - p_{N-n,N-k}$$

où $p_{m,M}$ est la probabilité d'obtenir au plus $m-1$ fois "succès" à l'issue de $M-1$ épreuves de Bernoulli de paramètre $\frac{1}{2}$.

Preuve : Quitte à permuter $a_1, a_2, \cdots, a_p$, on peut supposer que $a_1 < a_2 < \cdots < a_p$.

Soit $\boldsymbol{f} \in \mathcal{F}^N$. Pour prouver que l'hypothèse $(\mathcal{H}_1)$ est satisfaite, c'est-à-dire que $\boldsymbol{f}$ est l'unique solution du système (6.1), il suffit de montrer que $\boldsymbol{f}$ est l'unique solution du système $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x} \in [a_1, a_p]^N$.

Supposons que $\boldsymbol{f}$ est une réalisation d'un vecteur aléatoire $\boldsymbol{F}$ dont la distribution de probabilité est uniforme sur $\mathcal{F}^N$. Comme les $(a_i)_{1 \leq i \leq p}$ sont équiprobables, la probabilité, pour tout $i \in [\![1, N]\!]$, que F_i soit différent de a_1 et de a_p est $q = \frac{p-2}{p}$.

Soit S_i la variable associée à F_i telle que S_i vaut 0 ou 1 selon que $F_i \in \{a_1, a_p\}$ ou pas. Alors $\boldsymbol{f}$ est un vecteur simple d'ordre $\sum_{i=1}^{N} s_i$.

La variable aléatoire S_i associée à $\boldsymbol{F}_i$ suit une loi de Bernoulli de paramètre q. Ainsi, $S = \sum_{i=1}^{N} S_i$ suit une loi binomiale de paramètres (N, q). Pour N assez grand, $\boldsymbol{f}$ est alors un vecteur simple d'ordre Nq (presque sûrement).

D'après le lemme 7.1, pour (n, N) proportionnellement large, la probabilité que $\boldsymbol{f}$ soit l'unique solution du système $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x} \in [a_1, a_p]^N$ est

$$1 - p_{N-n,N-Nq}$$

Le comportement asymptotique de la fonction $p_{N-n,N-Nq}$ (voir annexe B.1) prouve que

- Si $N - n < \frac{N-Nq}{2}$, c'est à dire $\frac{n}{N} > \frac{p-1}{p}$, alors $p_{N-n,N-Nq} = 0$
- Si $N - n > \frac{N-Nq}{2}$, c'est à dire $\frac{n}{N} < \frac{p-1}{p}$, alors $p_{N-n,N-Nq} = 1$

Ainsi, $\boldsymbol{f}$ est l'unique solution du système $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x} \in [a_1, a_p]^N$ si et seulement si $\frac{n}{N} > \frac{p-1}{p}$. Ce qui prouve que l'hypothèse $(\mathcal{H}_1)$ est satisfaite si $\frac{n}{N} > \frac{p-1}{p}$. ∎

7.2 Méthodes de reconstruction

Dans cette section, nous présentons deux approches de reconstruction : L'approche de reconstruction par régularisation et celle par transformation parcimonieuse. Ces deux méthodes ont en commun une relaxation convexe de la norme ℓ_0 en norme ℓ_1. Pour les deux approches, nous montrons tout d'abord que le problème de reconstruction est équivalent à une minimisation de la norme ℓ_0, puis nous relaxons ce problème en minimisation ℓ_1. Les performances des deux méthodes sont comparables pour $p = 2$, avec des complexités algorithmiques plus faibles dans le cas de la première méthode. Néanmoins, pour $p > 2$, les résultats expérimentaux décrits dans le chapitre suivant montrent que la seconde méthode est meilleure en terme de probabilité de reconstruction.

7.2.1 Approche de reconstruction par régularisation : ℓ_1-additive

Dans les problèmes inverses mal posés, tel que (7.1), la régularisation consiste à favoriser des solutions particulières ayants une propriété désirée. Pour ce faire, un terme de régularisation est introduit tel que la minimisation de ce terme promeut la propriété pour un signal donné. Dans cette section, nous étudions une approche de reconstruction de $\boldsymbol{f}$ basée sur cette technique. L'approche donc considère le critère d'alphabet fini comme un a priori et minimise la fonction objective associée.

Soit $\mathcal{F} = \{a_1, a_2, \cdots, a_p\}$ un ensemble discret de cardinal p. On pose

$$\mathcal{J}(\boldsymbol{x}) = \sum_{i=1}^{p} \|\boldsymbol{x} - a_i \mathbf{1}_N\|_0$$

Le théorème suivant montre que la minimisation de la fonction $\mathcal{J}(\boldsymbol{x})$ impose à la solution d'appartenir à $\mathcal{F}^N$.

Théorème 7.2 *Supposons $(\mathcal{H}_1)$ vérifiée, soit $\boldsymbol{f}$ l'unique solution du système (7.1) dont les coefficients $f_i \in \mathcal{F}$. La minimisation :*

$$(P_{\mathcal{F},0}): \quad \arg\min_x \mathcal{J}(\boldsymbol{x}) \;\; s.t \;\; \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

admet une unique solution dans $\mathbb{R}^N$, cette solution est égale à $\boldsymbol{f}$.

Preuve : Soit $\boldsymbol{x} \in \mathbb{R}^N$ une solution de l'équation (7.1), et soit $(N_i)_{1\leq i\leq p}$ la suite dans $\{0, 1, \cdots, N\}$ définie par

$$\forall i \in \{1, 2, \cdots, p\}, \qquad N_i = \#\Big\{ j \in \{1, 2, \cdots, N\} : \boldsymbol{x}_j = a_i \Big\}$$

Comme : $\forall i \in \{1, 2, \cdots, p\}, \quad \|\boldsymbol{x} - a_i \mathbf{1}_N\|_0 = N - N_i$
On déduit :

$$\sum_{i=1}^{p} \|\boldsymbol{x} - a_i \mathbf{1}\|_0 = Np - \sum_{i=1}^{p} N_i \geq N(p-1)$$

Cette valeur minimale est atteinte lorsque $\sum_{i=1}^{p} N_i = N$, c'est à dire pour $\boldsymbol{x} \in \mathcal{F}^N$. L'unicité donc de la solution $(P_{\mathcal{F},0})$ découle de l'hypothèse $(\mathcal{H}_1)$. ∎

L'interprétation de ce résultat est évidente, puisqu'on cherche à minimiser le nombre d'éléments différents des $(a_i)_{1\leq i\leq p}$. Cependant, la minimisation de la norme

ℓ_0 est un problème NP-difficile en général [32] et nécessite une stratégie de recherche exhaustive. D'après l'état de l'art sur l'acquisition compressée, une tentative de résolution du problème $(P_{\mathcal{F},0})$ consiste à "remplacer" la norme ℓ_0 par la norme ℓ_1. Nous obtenons alors une nouvelle procédure de reconstruction de $\boldsymbol{f}$.

Procédure ℓ_1-additive :

$$(P_{\mathcal{F},1}) : \quad \arg\min_x \sum_{i=1}^{p} \|\boldsymbol{x} - a_i \mathbf{1}\|_1 \ \text{ s.t } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

Contrairement à la norme ℓ_0, la norme ℓ_1 est convexe. La procédure de reconstruction $(P_{\mathcal{F},1})$ peut être résolue en un temps polynomial. Par contre, une solution du problème $(P_{\mathcal{F},1})$ n'est pas forcément une solution de $(P_{\mathcal{F},0})$.

Le théorème suivant donne les conditions nécessaires et suffisantes d'équivalence des deux problèmes $(P_{\mathcal{F},0})$ et $(P_{\mathcal{F},1})$ dans le cas binaire $(p = 2)$.

Théorème 7.3 (Cas binaire) *Soient $\boldsymbol{A}$ une matrice aléatoire générique de dimension $n \times N$, $\mathcal{F} = \{a_1, a_2\}$ un ensemble discret et $\boldsymbol{f} \in \mathcal{F}^N$ une solution de l'équation (7.1). La probabilité que $\boldsymbol{f}$ soit l'unique solution de la minimisation $(P_{\mathcal{F},1})$ est*

$$1 - p_{N-n,N}$$

Pour (n, N) proportionnellement large, la probabilité de reconstruction exacte de $\boldsymbol{f}$ via $(P_{\mathcal{F},1})$ tend vers un ou zéro selon que $\frac{n}{N} > \frac{1}{2}$ ou $\frac{n}{N} < \frac{1}{2}$ respectivement.

Preuve : Soit $\boldsymbol{x} \in \mathbb{R}^N$ une solution de l'équation (7.1). Quitte à permuter a_1, a_2, on peut supposer que $a_1 < a_2$.

En appliquant l'inégalité triangulaire sur la norme ℓ_1, on a :

$$\|\boldsymbol{x} - a_1 \mathbf{1}_N\|_1 + \|\boldsymbol{x} - a_2 \mathbf{1}_N\|_1 \geq N(a_2 - a_1)$$

On obtient une égalité si et seulement si

$$\forall i \in \{1, \cdots, N\}, \quad |x_i - a_1| + |x_i - a_2| = a_2 - a_1$$

Ce qui est équivalent à $x_i \in [a_1, a_2]$, pour tout $i \in \{1, \cdots, N\}$.

Donc $\boldsymbol{x}$ est solution de $(P_{\mathcal{F},1})$ si et seulement si $\boldsymbol{x} \in [a_1, a_2]^N$.

On en déduit que $\boldsymbol{f} \in \mathcal{F}^N$ est solution de $(P_{\mathcal{F},1})$ et que c'est l'unique solution si et seulement si $\boldsymbol{f}$ est l'unique solution du système

$$\begin{cases} \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \\ \boldsymbol{x} \in [a_1, a_2]^N \end{cases} \tag{7.2}$$

Or $\boldsymbol{f}$ est un vecteur simple d'ordre 0 de l'hypercube $[a_1, a_2]^N$.

D'après le lemme 7.1, la probabilité que $\boldsymbol{f}$ soit l'unique solution du système (7.2) est $1 - p_{N-n,N}$.

Pour (n, N) proportionnellement large, les généralités sur les lois binomiales (annexe B.1) montrent que

- Si $N - n < \frac{N}{2}$, c'est à dire $\frac{n}{N} > \frac{1}{2}$, alors la probabilité de reconstruction exacte de $\boldsymbol{f}$ via $(P_{\mathcal{F},1})$ tend vers un.

- Si $N - n > \frac{N}{2}$, c'est à dire $\frac{n}{N} < \frac{1}{2}$, alors la probabilité de reconstruction exacte de $\boldsymbol{f}$ via $(P_{\mathcal{F},1})$ tend vers zéros.

La probabilité de reconstruction exacte de $\boldsymbol{f}$ via $(P_{\mathcal{F},1})$ tend vers un ou zéro selon que $\frac{n}{N} > \frac{1}{2}$ ou $\frac{n}{N} < \frac{1}{2}$ respectivement. ∎

Contrairement aux approches proposées dans [137] et [16], la procédure de reconstruction par régularisation donne un cadre qui permet d'étendre l'acquisition compressée dans le cas où $p > 2$.

7.2.2 Approche de reconstruction par transformation parcimonieuse : ℓ_1-indicateur

Dans cette section, une seconde approche est proposée. L'idée consiste à voir la propriété d'alphabet fini comme une contrainte de parcimonie. Ceci ce justifie par le fait que les signaux parcimonieux sont des signaux à faible degré de liberté, qui sont plongés dans des espaces de grandes dimensions. C'est le cas des signaux à alphabet fini. L'approche proposée formalise cette idée et reformule le problème de reconstruction de signaux à alphabet fini en un problème de reconstruction de signaux parcimonieux à partir de mesures largement incomplètes.

Soit $\boldsymbol{f} \in \mathcal{F}^N$. On note $\boldsymbol{\epsilon_i} = [I(f_i = a_1), I(f_i = a_2), \cdots, I(f_i = a_p)]^T$ le vecteur indicateur associé à f_i, où $I(f_i = a_j)$ est la fonction indicatrice qui vaut 1 si $f_i = a_j$ et 0 sinon. On note $\hat{\boldsymbol{s}}$ le vecteur de dimension Np tel que :

$$\hat{\boldsymbol{s}} = \begin{pmatrix} \boldsymbol{\epsilon_1} \\ \boldsymbol{\epsilon_2} \\ \vdots \\ \boldsymbol{\epsilon_N} \end{pmatrix}$$

Nous appellerons $\hat{\boldsymbol{s}}$ le vecteur indicateur associé à $\boldsymbol{f}$. Comme les a_i sont différents, un vecteur de $\mathcal{F}^N$ admet un unique vecteur indicateur.
Sachant que $\boldsymbol{f} = \boldsymbol{B_a}\hat{\boldsymbol{s}}$, il suffit alors de reconstruire $\hat{\boldsymbol{s}}$ pour pouvoir reconstruire $\boldsymbol{f}$. L'intérêt majeure d'avoir introduit le vecteur $\hat{\boldsymbol{s}}$ est son aspect parcimonieux. En effet, $\hat{\boldsymbol{s}}$ est un vecteur de $\mathbb{R}^{Np}$ tel que $\|\hat{\boldsymbol{s}}\|_0 = N$.
Supposons maintenant que $\boldsymbol{f}$ soit une solution de l'équation (7.1). Il est clair que $\hat{\boldsymbol{s}}$ est solution de l'équation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{B_a}\boldsymbol{x} \tag{7.3}$$

Signalons que la démarche peut paraître saugrenue au premier abord : pour un même nombre d'observations, l'équation (7.3) admet beaucoup plus d'inconnues que l'équation (7.1) (Np contre N), donc plus d'incertitude sur la solution. De plus, la reconstruction de $\hat{\boldsymbol{s}}$ n'est pas nécessaire pour reconstruire $\boldsymbol{f}$. Cependant, l'idée principale consiste à "transformer" l'a priori d'appartenance à un alphabet fini en un a priori de parcimonie, puis d'exploiter l'état de l'art sur la recherche de solution parcimonieuse.
Tenant compte de la première remarque, nous réduisons l'espace des solutions possibles en ajoutant au modèle des contraintes.
Par construction, le vecteur $\hat{\boldsymbol{s}}$ vérifie :

$$\mathbf{1}_N = \boldsymbol{B_1}\hat{\boldsymbol{s}}$$

On a alors

$$\boldsymbol{\Phi}\hat{\boldsymbol{s}} = \begin{pmatrix} \boldsymbol{A}\boldsymbol{B_a} \\ \boldsymbol{B_1} \end{pmatrix} \hat{\boldsymbol{s}} = \begin{pmatrix} \boldsymbol{y} \\ \mathbf{1}_N \end{pmatrix} = \boldsymbol{b}.$$

Nous pouvons maintenant reformuler le problème de reconstruction de signaux à alphabet fini en un problème de recherche de solution parcimonieuse d'un système sous déterminé. Le théorème suivant relie ces deux problèmes.

Théorème 7.4 *Supposons $(\mathcal{H}_1)$ vérifiée. La minimisation :*

$$(TP_{\mathcal{F},0}) : \quad \arg\min_{\boldsymbol{s}} \|\boldsymbol{s}\|_0 \;\; s.t \;\; \boldsymbol{\Phi}\boldsymbol{s} = \boldsymbol{b}.$$

admet une unique solution dans $\mathbb{R}^{Np}$*, cette solution est égale à* $\hat{\boldsymbol{s}}$*.*

Preuve : Soit $\boldsymbol{s} \in \mathbb{R}^{Np}$ tel que $\boldsymbol{B_1 s} = \mathbf{1}_N$.
En notant, pour tout $i \in [\![1, N]\!]$, $T_i = [\![(i-1)p+1, ip]\!]$ et $\boldsymbol{s}_{T_i}$ la restriction de $\boldsymbol{s}$ sur T_i, on a :

$$\boldsymbol{B_1 s} = \mathbf{1}_N \iff \forall i \in [\![1, N]\!],\ \mathbf{1}_p^T \boldsymbol{s}_{T_i} = 1$$

d'où $\|\boldsymbol{s}_{T_i}\|_0 \geq 1$, et donc $\|\boldsymbol{s}\|_0 \geq N$.
D'autre part, $\|\hat{\boldsymbol{s}}\|_0 = N$ et $\boldsymbol{\Phi}\hat{\boldsymbol{s}} = \boldsymbol{b}$. Ce qui prouve que $\hat{\boldsymbol{s}}$ est une solution de $(TP_{\mathcal{F},0})$.
Pour prouver l'unicité, il suffit de remarquer que si $\boldsymbol{B_1 s} = \mathbf{1}_N$ et $\|\boldsymbol{s}\|_0 = N$ alors $\|\boldsymbol{s}_{T_i}\|_0 = 1$ et donc $\boldsymbol{s}$ est un vecteur indicateur associé à un vecteur de $\mathcal{F}^N$.
Or d'après l'hypothèse $(\mathcal{H}_1)$, $\boldsymbol{f}$ est l'unique solution de (7.1) dans $\mathcal{F}^N$. Donc $\boldsymbol{s} = \hat{\boldsymbol{s}}$ ∎

Afin de reconstruire $\boldsymbol{f}$, il suffit donc de minimiser $(TP_{\mathcal{F},0})$ et de poser $\boldsymbol{f} = \boldsymbol{B_a}\hat{\boldsymbol{s}}$. Malheureusement, comme nous l'avons vu précédemment, la recherche de solutions parcimonieuses d'un système linéaire sous déterminé est un problème NP-difficile en général [32]. Naturellement, de la même manière que dans l'approche ℓ_1-additive, on remplace la norme ℓ_0 par la norme ℓ_1 pour obtenir la nouvelle procédure de reconstruction de $\boldsymbol{f}$.

Procédure ℓ_1-indicateur :

1. Résoudre la minimisation

$$(TP_{\mathcal{F},1}) : \quad \arg\min_s \|\boldsymbol{s}\|_1 \ \text{ s.t } \ \boldsymbol{\Phi s} = \boldsymbol{b}$$

2. Poser $\tilde{\boldsymbol{f}} = \boldsymbol{B_a}\tilde{\boldsymbol{s}}$ où $\tilde{\boldsymbol{s}}$ est une solution de $(TP_{\mathcal{F},1})$.

Au début, en procédant ainsi, on a cherché à reconstruire $\hat{\boldsymbol{s}}$ la solution parcimonieuse de l'équation $\boldsymbol{\Phi s} = \boldsymbol{b}$. En général, la minimisation $(TP_{\mathcal{F},1})$ échoue à reconstruire $\hat{\boldsymbol{s}}$. Toutefois, ce qui nous intéresse ici, n'est pas tant la reconstruction du vecteur indicateur de $\boldsymbol{f}$, mais le vecteur $\boldsymbol{f}$. Le théorème suivant donne des conditions nécessaires et suffisantes à la réussite de la procédure ℓ_1-indicateur.

Théorème 7.5 *Supposons* $(\mathcal{H}_1)$ *vérifiée. Pour tout* $i \in [\![1, N]\!]$*, posons* $T_i = [\![(i-1)p+1, ip]\!]$*. On note*

$$\mathcal{C}^{Np} = \{\boldsymbol{h} \in \mathbb{R}^{Np} |\ \forall i \in [\![1, N]\!], \textit{il existe au maximum un } n_i \in T_i,\ h_{n_i} < 0\}$$

Alors, pour tout $\boldsymbol{f} \in \mathcal{F}^N$, $\boldsymbol{f}$ *est l'unique sortie de la procédure de reconstruction par transformation parcimonieuse si et seulement si*

$$\boldsymbol{\Phi h} = \mathbf{0}, \;\; \boldsymbol{h} \in \mathcal{C}^{Np} \;\; \textit{n'admet pas de solution } \boldsymbol{h} \textit{ tel que } \boldsymbol{B_a h} \neq \mathbf{0} \tag{7.4}$$

Afin de ne pas alourdir davantage le chapitre par un flot d'équations mathématiques, la démonstration du théorème 7.5 sera donnée dans l'annexe B.2.

Remarques :

- Comme pour la propriété RIP, ce résultat est déterministe dans le sens où il n'implique aucune probabilité.

- Pour un $\boldsymbol{f} \in \mathcal{F}^N$ fixé, les indices $(n_i)_{1 \leq i \leq N}$ de l'ensemble $\mathcal{C}^{Np}$ sont choisis en en fonction de $\boldsymbol{f}$. L'indice n_i est l'indice de la composante non nulle du vecteur $[\mathbf{0}^T_{(i-1)p}, \boldsymbol{\epsilon}^T_{\boldsymbol{i}}, \mathbf{0}^T_{(N-i)p}]^T$.

Il suffit donc de trouver une matrice $\boldsymbol{\Phi}$ et à moindre mesure un alphabet $\mathcal{F}$ tel que la condition du théorème 7.5 soit vérifiée. Néanmoins, cette condition n'est pas facile à vérifier. Le théorème suivant donne un exemple de matrice vérifiant la condition dans le cas $p = 2$.

Théorème 7.6 (Cas binaire) *Soit* $\boldsymbol{A}$ *une matrice aléatoire générique de dimension* $n \times N$.
La matrice $\boldsymbol{A}$ *vérifie la condition (7.4) avec une probabilité*

$$p_{n,N} = 2^{-N+1} \sum_{i=0}^{n-1} \binom{N-1}{i}$$

Cela implique que, pour (n, N) *proportionnellement large,* $\boldsymbol{A}$ *satisfait la condition (7.4) si* $\frac{n}{N} > \frac{1}{2}$ *et ne la satisfait pas si* $\frac{n}{N} < \frac{1}{2}$.

Preuve : Cette démonstration est fortement inspirée de la démonstration du théorème 6.4 dans [16].
Soit $\boldsymbol{h} \in \mathbb{R}^{2N}$, on note, pour tout $i \in [\![1, N]\!]$, $\alpha_i = \text{sgn}(h_{(i-1)p+2})$ où sgn représente la fonction signe. On pose

$$\boldsymbol{D} = \text{diag}(\alpha_1, \alpha_2, \cdots, \alpha_N)$$

Montrons que la condition (7.4) est équivalente à la condition que toutes les colonnes de la matrice $\boldsymbol{AD}$ soient dans le même hémisphère.

On remarque tout d'abord que la condition $\boldsymbol{B_1 h} = \mathbf{0}$ implique que $\boldsymbol{h} \in \mathcal{C}^{2N}$. En présence de $\boldsymbol{\Phi h} = \mathbf{0}$, la condition $\boldsymbol{h} \in \mathcal{C}^{2N}$ est donc obsolete.
Supposons que $\boldsymbol{\Phi h} = \mathbf{0}$, on a $\boldsymbol{B_1 h} = \mathbf{0}$. On note $\hat{\boldsymbol{h}}$ le vecteur de $\mathbb{R}_+^N$ tel que

$$\hat{h}_i = \left|h_{(i-1)p+1}\right| = \left|h_{(i-1)p+2}\right|$$

On a alors :

$$\boldsymbol{h} = \begin{pmatrix} -\alpha_1 \hat{h}_1 \\ \alpha_1 \hat{h}_1 \\ -\alpha_2 \hat{h}_2 \\ \alpha_2 \hat{h}_2 \\ \vdots \\ -\alpha_N \hat{h}_N \\ \alpha_N \hat{h}_N \end{pmatrix}$$

d'où

$$\boldsymbol{B_a h} = (a_2 - a_1) \begin{pmatrix} \alpha_1 \hat{h}_1 \\ \alpha_2 \hat{h}_2 \\ \vdots \\ \alpha_N \hat{h}_N \end{pmatrix} = (a_2 - a_1) \boldsymbol{D} \hat{\boldsymbol{h}}$$

La condition "$\boldsymbol{\Phi h} = \mathbf{0},\ (\boldsymbol{h} \in \mathcal{C}^{Np})$ n'admet pas de solution $\boldsymbol{h}$ tel que $\boldsymbol{B_a h} \neq \mathbf{0}$" est équivalente à "$\boldsymbol{AD\hat{h}} = \mathbf{0},\ \hat{\boldsymbol{h}} \geqq \mathbf{0}$ n'admet pas de solution $\hat{\boldsymbol{h}}$ tel que $\hat{\boldsymbol{h}} \neq \mathbf{0}$".

Le théorème de l'alternative de Gordan, appliqué à la matrice $\boldsymbol{AD}$, montre que cette dernière condition est équivalente à la condition que toutes les colonnes de la matrice $\boldsymbol{AD}$ soient dans le même hémisphère, c'est à dire qu'il existe $\boldsymbol{r} \in \mathbb{R}^n$ tel que $\boldsymbol{DA}^T \boldsymbol{r} > \mathbf{0}$.

Théorème 7.7 (Gordan) *Soit $\boldsymbol{G}$ une matrice réelle de dimension (n, k), alors un et un seul des systèmes suivants a une solution :*

- $\boldsymbol{G}^T \boldsymbol{r} > \mathbf{0}$ *pour* $\boldsymbol{r} \in \mathbb{R}^n$
- $\boldsymbol{Gx} = \mathbf{0}$ *pour* $\boldsymbol{x} \geq \mathbf{0}$

Comme les a_i sont équiprobables, α_i est une variable de Bernoulli de paramètre $\frac{1}{2}$, d'où $\boldsymbol{AD}$ est une matrice aléatoire générique. Le lemme 6.1 et les généralités sur les lois binomiales achèvent la preuve. ∎

Compte tenu des résultats numériques présentés dans le chapitre suivant, nous généralisons ce résultat dans le cas $p > 2$.

Conjecture 7.1 (Cas géneral) *Soit $\boldsymbol{A}$ une matrice aléatoire générique de dimension $n \times N$.*
Pour (n, N) proportionnellement large, la matrice $\boldsymbol{A}$ satisfait la condition (7.4) si $\frac{n}{N} > \frac{p-1}{p}$ et ne la satisfait pas si $\frac{n}{N} < \frac{p-1}{p}$.

7.3 Discussions

– Il est clair que cette étude à pour intérêt le cas où N est assez grand. En effet, si N est faible (par exemple de l'ordre de l'unité pour $p = 2$), la recherche exhaustive de solution $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ dans $\mathcal{F}^N$ est plus appropriée, tant d'un point de vue complexité que du point de vue performance de reconstruction, que les deux procédures de reconstruction présentées dans le manuscrit.

– En analyse harmonique appliquée, la relaxation de la norme ℓ_0 en norme ℓ_1 a pour but de trouver la solution donnée par la minimisation de la norme ℓ_0. Dans cette perspective, différents articles ([132], [141], [142] [143] pour n'en citer que quelques-uns) ont étudié les conditions qui garantissent l'équivalence entre la norme ℓ_0 et la norme ℓ_1. Ces conditions sont suffisantes dans notre cas mais loin d'être nécessaires. Pour cause, l'objectif n'est pas de reconstruire le vecteur indicateur $\hat{\boldsymbol{s}}$ mais plutôt $\boldsymbol{f}$. Dans le cas $p > 2$, on observe que la solution de $(TP_{\mathcal{F},1})$, fournie par la boite à outils Matlab *cvx*, un programme de résolution d'optimisation convexe [43], n'est pas la solution la plus parcimonieuse de $\boldsymbol{\Phi}\boldsymbol{s} = \boldsymbol{b}$. Toutefois, cette solution conduit à $\boldsymbol{f}$.

– La condition $\frac{n}{N} > \frac{p-1}{p}$ apparait à la fois comme condition suffisante d'unicité de la solution et à la fois dans la procédure de reconstruction par transformation parcimonieuse lorsque l'hypothèse $(\mathcal{H}_1)$ est vérifiée. On peut penser alors que la procédure de reconstruction échoue à cause du fait qu'il existe plusieurs solutions du problème et que dans le cas où on a unicité de la solution, la procédure réussie toujours à reconstruire le signal. En d'autres termes, que la norme ℓ_0 et la norme ℓ_1 sont $\mathcal{F}$-équivalentes, c'est à dire conduisent à la même solution par la transformation $\boldsymbol{B}_{\boldsymbol{a}}$, si et seulement si l'hypothèse $(\mathcal{H}_1)$ est vérifiée.
L'exemple suivant montre que ceci n'est pas le cas : $\mathcal{F} = \{0, 1\}$, $\boldsymbol{f} = \mathbf{1}_4$ et $\boldsymbol{A}$ la réalisation suivante d'une matrice aléatoire de dimension 3×4.

$$\boldsymbol{A} = \begin{pmatrix} 0.4889 & -0.3034 & 0.8884 & -0.8095 \\ 1.0347 & 0.2939 & -1.1471 & -2.9443 \\ 0.7269 & -0.7873 & -1.0689 & 1.4384 \end{pmatrix}$$

La procédure de reconstruction échoue à reconstruire $\boldsymbol{f}$ même si l'hypothèse $(\mathcal{H}_1)$ est vérifiée (la solution donnée par l'algorithme est $\tilde{\boldsymbol{f}} = (0.4695, 0.3121, 0.8705, 0.7954)^T$).

– L'acquisition compressée de signaux à alphabet fini peut être liée au problème d'acquisition compressée de signaux parcimonieux dans un dictionnaire cohérent et redondant $\boldsymbol{D}$. Les auteurs de [144] proposent une méthode de reconstruction via une minimisation d'analyse ℓ_1

$$\arg\min_{\boldsymbol{g}} \left\|\boldsymbol{D}^T\boldsymbol{g}\right\|_1 \text{ s.t } \boldsymbol{A}\boldsymbol{g} = \boldsymbol{y}.$$

Dans le cas où $\boldsymbol{D}$ est une concaténation de base orthogonale, ce qui est notre cas ($\boldsymbol{D} = \boldsymbol{B_a}$), la minimisation d'analyse ℓ_1 n'est pas appropriée. La seconde approche que nous avons proposée est liée à la méthode alternative décrite brièvement dans [144] qui consiste à résoudre la minimisation de synthèse ℓ_1

$$\arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ s.t } \boldsymbol{A}\boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}.$$

et poser $\hat{\boldsymbol{f}} = \boldsymbol{D}\boldsymbol{x}$. La différence entre les deux approches est l'ajout d'observations (contraintes) à nos mesures ($\boldsymbol{B_1 s} = \boldsymbol{1}_N$). Sans cette étape, la méthode échoue à reconstruire $\boldsymbol{f}$.

– La condition $\frac{n}{N} > \frac{p-1}{p}$ peut être réécrite sous la forme $Mp < N$ où M est le nombre de mesures manquantes. Sachant que des conditions de reconstruction similaires peuvent être trouvées dans [145], on peut se demander si cette condition n'est pas basée sur un principe d'incertitude vu que M représente l'information manquante, $p = |\mathcal{F}|$ quantifie l'information connue et N représente la dimension du vecteur qu'on cherche à reconstruire.

7.4 Conclusion

Dans un contexte d'acquisition compressée de signaux à alphabet fini, nous avons établi une condition suffisante qui permet de rendre le problème soluble. Cette condition étend celle de Mangasarian au cas d'alphabet fini quelconque.
Nous avons présenté deux procédures de reconstruction de signaux à alphabet fini à partir d'observations incomplètes. Ces procédures sont basées sur une relaxation convexe de la norme ℓ_0 en norme ℓ_1.

La première procédure considère le critère d'alphabet fini comme un a priori et introduit un terme de régularisation, afin de favoriser les signaux dont les éléments appartiennent à l'alphabet. Dans le cas binaire, c'est à dire alphabet à deux éléments, le théorème 7.3 calcule la probabilité de réussite et d'échec de la procédure de reconstruction, pour un type de matrices de mélange aléatoire.

La seconde procédure reformule le problème de reconstruction comme un problème de recherche de solution parcimonieuse d'un système sous déterminé. Le théorème 7.5 donne une condition nécessaire et suffisante de reconstruction. Cette condition est déterministe. Dans le cas binaire, nous avons calculé la probabilité qu'un type de matrices de mélange aléatoire de dimension $n \times N$ vérifie cette condition. En grande dimension, cette condition est satisfaite si $\frac{n}{N} > \frac{1}{2}$ et non satisfaite si $\frac{n}{N} < \frac{1}{2}$. La conjecture généralise ce résultat lorsque l'alphabet est de cardinal supérieur à deux.

CHAPITRE 8 Expérimentation numérique

Dans ce chapitre, nous présentons les performances de reconstruction des deux approches. Le but est d'évaluer de façon empirique la borne sur le rapport $\frac{n}{N}$ qui permet de reconstruire exactement le signal original.

8.1 Simulations

Dans ce cadre, nous effectuons une série d'expériences conçues comme suit :

1. Etant donnés deux entiers naturels n et N, on génère la matrice $\boldsymbol{A}$ telle que les coefficients de la matrice soient des variables aléatoires indépendantes et identiquement distribuées ($i.i.d$) selon une loi normale centrée réduite ;
2. Pour un entier $p \geq 2$ fixé, on génère aléatoirement un alphabet $\mathcal{F}$, et on génère un vecteur $\boldsymbol{f}$ dont les composantes sont uniformément choisies dans $\mathcal{F}$;
3. On calcule $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{f}$.
4. On résout le système $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x},\ \boldsymbol{x} \in \mathcal{F}^N$ avec l'une des deux approches de reconstruction. Cette solution sera notée $\boldsymbol{f}^*$;
5. On compare $\boldsymbol{f}$ et $\boldsymbol{f}^*$. La reconstruction sera considérée comme exacte si l'erreur relative $\|\boldsymbol{f}^* - \boldsymbol{f}\|_2 / \|\boldsymbol{f}\|_2$ est inférieure à 10^{-8} ;
6. On répète 100 fois l'expérience pour différentes valeurs de $\boldsymbol{A}$, $\mathcal{F}$ et $\boldsymbol{f}$;
7. On calcule la probabilité de reconstruction exacte pour un (n, N, p) fixé.
8. On répète pour différentes valeurs de n, N et p.

Nous utilisons la boite à outils Matlab *cvx*, un programme de résolution d'optimisation convexe [43], pour résoudre les minimisations ℓ_1 introduites par les deux procédures.

8.2 Résultats

8.2.1 Procédure de reconstruction par régularisation

Les figures 8.1, 8.2 montrent les résultats de reconstruction pour $p = 2$ et pour $p = 4$ respectivement avec différentes valeurs de N.

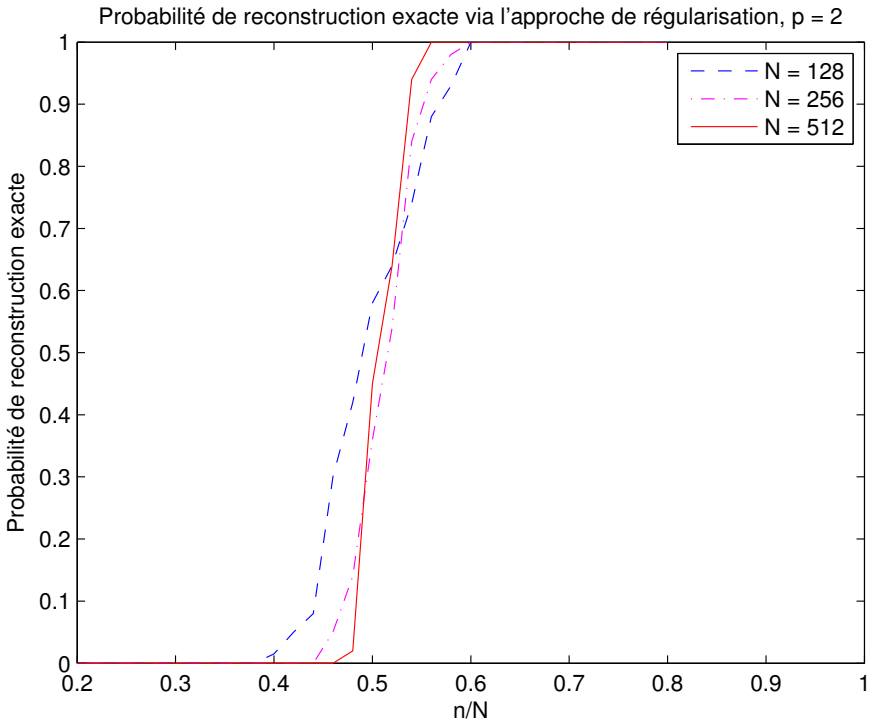


Figure 8.1 — Probabilité de reconstruction exacte de $(P_{\mathcal{F},1})$ pour $p = 2$

Dans la figure 8.1, les résultats expérimentaux correspondent bien à nos prédictions théoriques (Théorème 7.3) et mettent en évidence les deux points suivants :

- La probabilité de reconstruction par minimisation $(P_{\mathcal{F},1})$ tend vers 1 pour $\frac{n}{N} > \frac{1}{2}$ et tend vers 0 pour $\frac{n}{N} < \frac{1}{2}$. En acquisition compressée, ce résultat est connu sous le nom de "phase de transition".

- Le caractère asymptotique : Plus (n, N) est proportionnellement large, plus précis sont les résultats (c'est à dire plus la convergence est rapide).

Dans le cas $p > 2$, nous pouvons noter sur les résultats illustrés par la figure 8.2 que

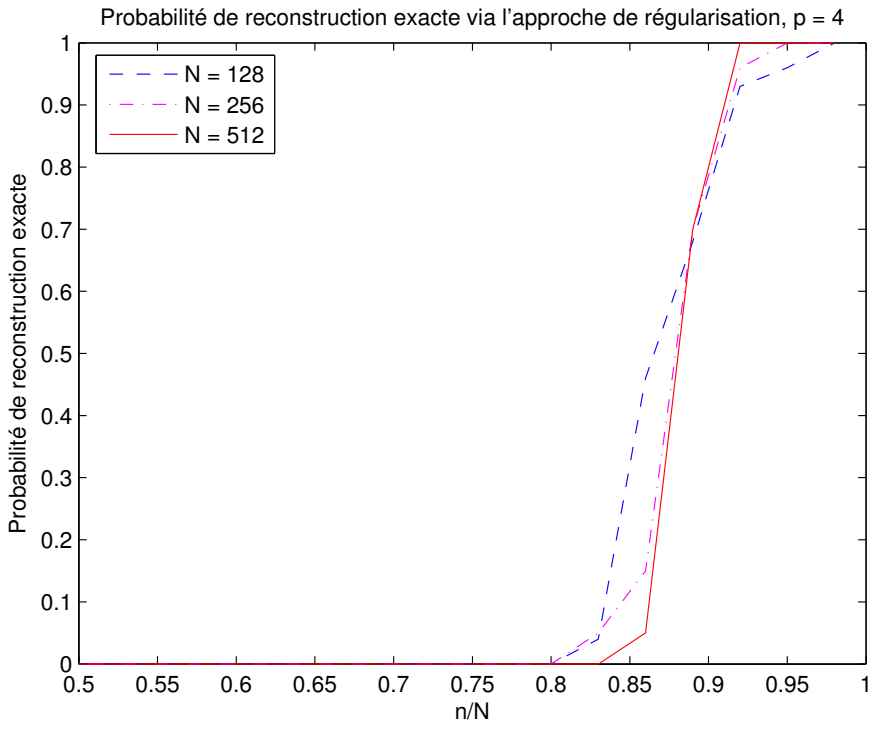


Figure 8.2 — Probabilité de reconstruction exacte de $(P_{\mathcal{F},1})$ pour $p = 4$

l'augmentation de la taille d'alphabet engendre une translation de la borne sur $\frac{n}{N}$ dans la phase de transition, tout en gardant le même caractère asymptotique des résultats. Pour $p = 4$, cette borne est à peu près égale à 0.88.

8.2.2 Procédure de reconstruction par transformation parcimonieuse

Les figures 8.3, 8.4 montrent les résultats de reconstruction pour $p = 2$ et pour $p = 4$ respectivement avec différentes valeurs de N.

Dans la figure 8.3, les résultats expérimentaux correspondent bien à nos prédictions théoriques (Théorèmes 7.5 et 7.6) et mettent en évidence les deux points suivants :

- La phase de transition : La probabilité de reconstruction par transformation parcimonieuse tend vers 1 pour $\frac{n}{N} > \frac{1}{2}$ et tend vers 0 pour $\frac{n}{N} < \frac{1}{2}$.

- Le caractère asymptotique : Plus (n, N) est proportionnellement large, plus précis sont les résultats (c'est à dire plus la convergence est rapide).

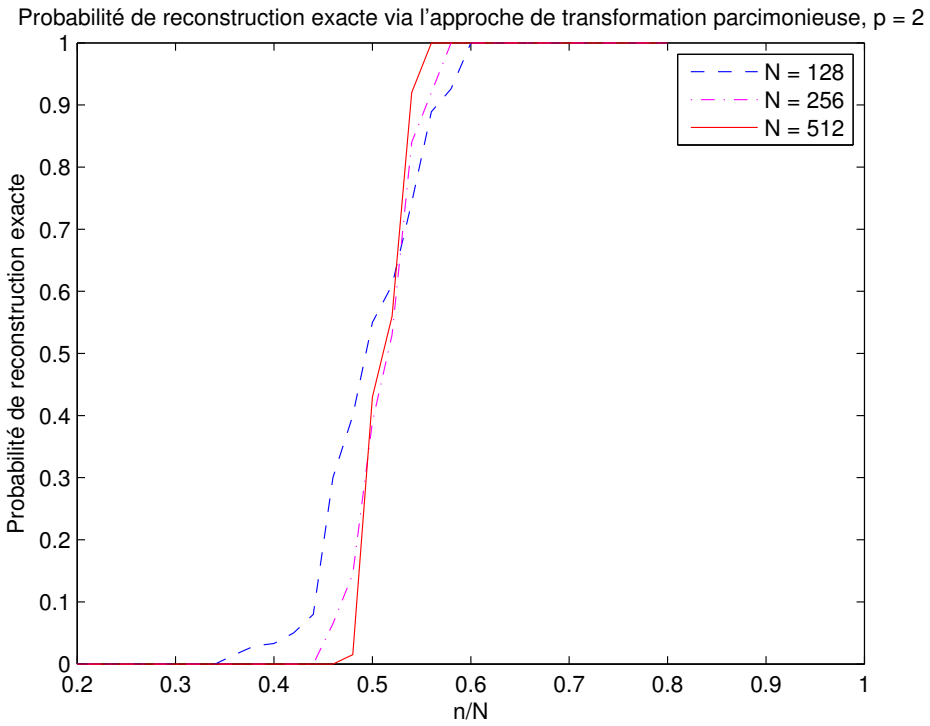


Figure 8.3 — Probabilité de reconstruction de l'approche par transformation parcimonieuse pour $p = 2$

Le cas $p > 2$, illustrée par la figure 8.4 montre que l'augmentation de la taille d'alphabet engendre une translation de la borne sur $\frac{n}{N}$ dans la phase de transition, tout en gardant le même caractère asymptotique des résultats. Pour $p = 4$, cette borne est à peu près égale à $\frac{2}{3} = 0,75$. Cette valeur correspond à la borne donnée dans la conjecture 7.1.

8.3 Séparation d'images Binaires

A titre purement illustratif, nous présentons ici une seconde application qui concerne la séparation d'images. Pour cela, nous utilisons quatre images binaires de taille 256×256 (voir figure 8.5).

Afin de satisfaire la condition $\frac{n}{N} > \frac{p-1}{p}$, nous considérons trois combinaisons linéaires des images originales. Les coefficients de la matrice de mélange sont générées indépendamment suivant une loi normale centrée réduite. La figure 8.6 illustre un exemple d'images issues d'un tel mélange.

Pour chaque pixel, on applique la procédure de reconstruction ℓ_1-indicateur sur les vecteurs formées par concaténation vectorielle des pixels des quatre sources.

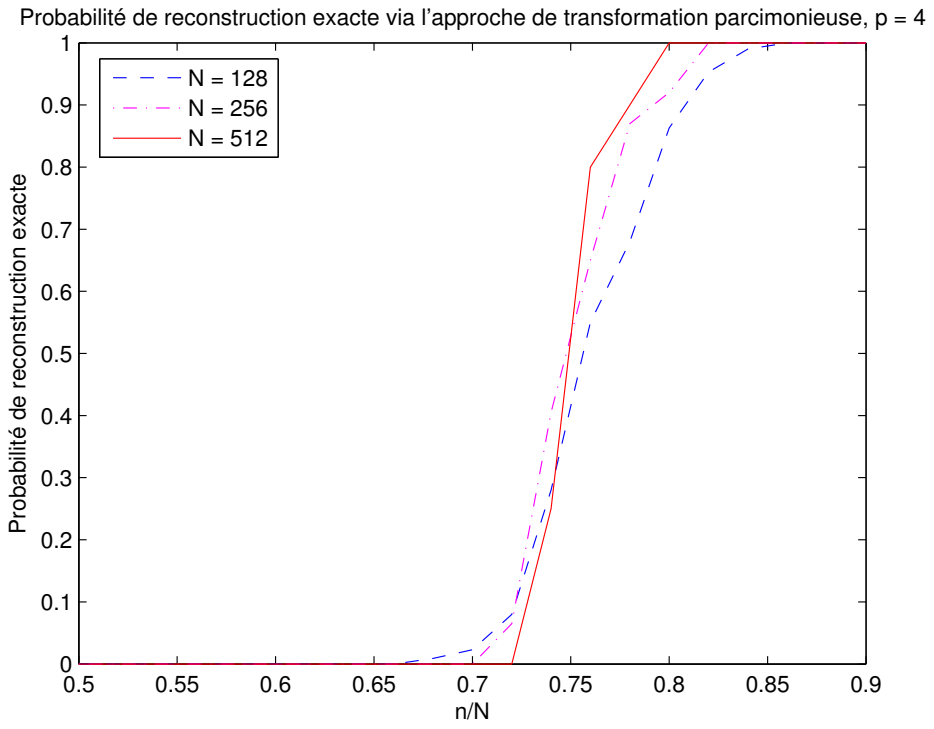


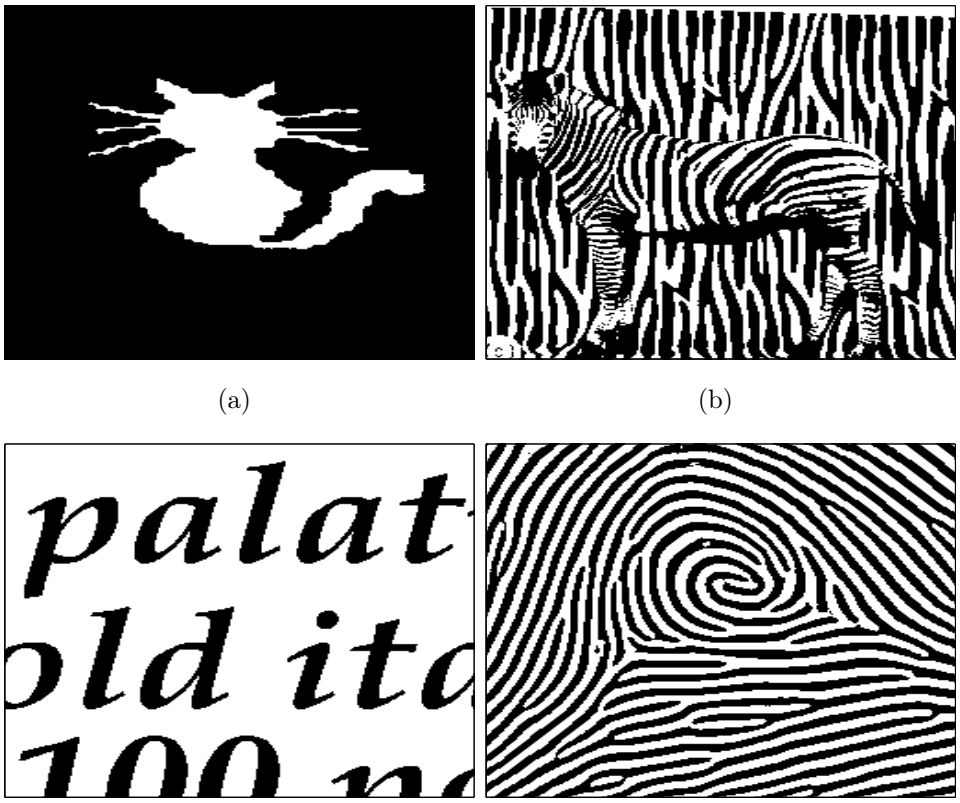
Figure 8.4 — Probabilité de reconstruction de l'approche par transformation parcimonieuse pour $p = 4$

La figure 8.7 montre le résultat de séparation. La qualité de la séparation est mesurée par le rapport signal à bruit RSB $= 20 \log_{10}(\|X\| / \|X - \hat{X}\|)$, où X, $\hat{X}$ sont respectivement l'image originale et l'image reconstruite. Le tableau suivant donne les différentes valeurs du RSB.

Images	Chat	Texte	Zèbre	Empreinte
RSB (dB)	17.15	34.73	22.49	20.23

Tableau 8.1 — RSB des différentes images

Notons que ces résultats n'ont pas un réel intérêt pratique compte tenu de la première remarque dans la section 7.3 du chapitre précèdent, puisqu'ils correspondent au cas où $N = 4$, alors que les résultats prennent un réel intérêt lorsque N est grand et qu'une recherche exhaustive devient trop coûteuse.
La figure 8.7 montre que même avec N petit (N $= 4$), et malgré les erreurs de reconstruction (tableau 8.1), les algorithmes proposés fournissent déjà un bon résultat.



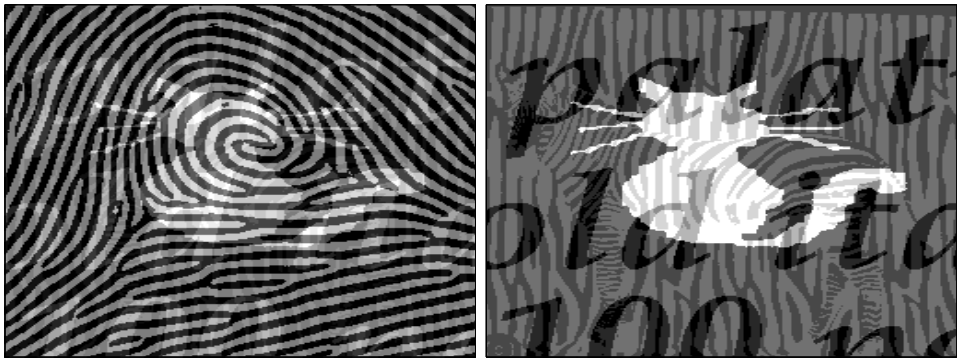
(a) (b)

(c) (d)

Figure 8.5 — Images originales.

8.4 Conclusion

Lorsque les composantes de la matrice $\boldsymbol{A}$ sont des variables aléatoires indépendantes et identiquement distribuées ($i.i.d$) selon une loi normale centrée, les simulations montrent qu'il existe un seuil ρ tel que : pour $n/N < \rho$ la probabilité de reconstruction de $\boldsymbol{f}$ est nulle et pour $n/N > \rho$ la probabilité de reconstruction est égale à 1. Dans cette dernière situation, on obtient donc une reconstruction exacte. Les performances des deux méthodes sont comparables pour $p = 2$, avec des complexités algorithmiques plus faibles dans le cas de la première méthode. En effet, la première approche résout une minimisation dans $\mathbb{R}^N$, alors que la seconde méthode effectue différentes multiplications matricielles en plus de la résolution d'une minimisation dans $\mathbb{R}^{2N}$. Néanmoins, pour $p > 2$, la seconde méthode est meilleure en terme de probabilité de reconstruction. Le seuil de transition pour la méthode de reconstruction par transformation parcimonieuse est inférieur au seuil de transition pour la méthode de reconstruction par régularisation.

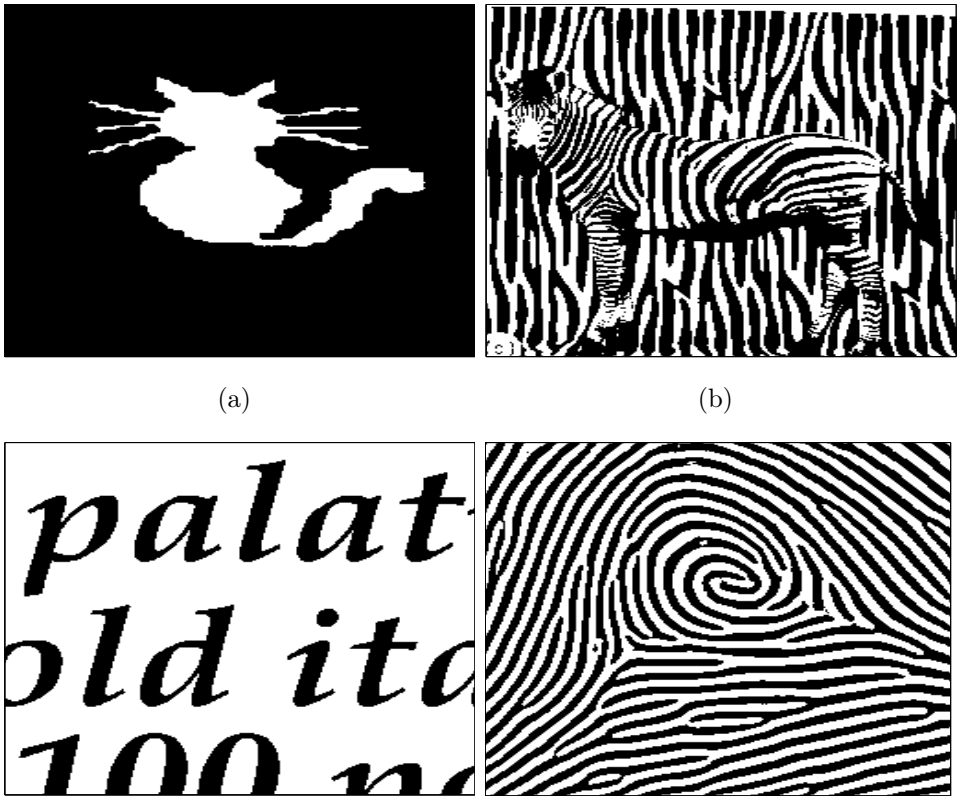


(a) (b)



(c)

Figure 8.6 — Combinaison linéaires des images (a), (b), (c) et (d) de la figure 8.5.



(a) (b)

(c) (d)

Figure 8.7 — Images reconstruites.

CHAPITRE 9 Conclusion générale et perspectives

9.1 Conclusion générale

La parcimonie et, plus généralement, la compressibilité, ont joué et continuent de jouer un rôle important dans de nombreux domaines de traitement du signal (estimation, détection, compression, traitement radar). Dans ce manuscrit, nous avons exploré l'apport de la parcimonie en traitement du signal. Deux problématiques ont été traitées. Le premier apport de la parcimonie concerne la séparation de sources. Le second concerne l'acquisition compressée.

La contribution de la parcimonie à la séparation de sources n'est pas nouvelle. Durant ces dernières années, l'utilisation de la parcimonie dans la séparation aveugle de sources, notamment dans le cadre sous-déterminé a connu un intérêt grandissant de la part de la communauté du traitement de signal. Cela a conduit au développement de différentes techniques de séparation de sources basées sur la parcimonie. Ces méthodes introduisent en général des paramètres empiriques. Les performances de ces méthodes, bien quelles soient meilleures que les autres approches de séparation de sources, dépendent du choix de ces paramètres. Nous avons proposé des tests statistiques permettants de rendre plus automatique le choix de ces paramètres. Ces algorithmes sont très généraux et peuvent être insérés facilement aux différentes techniques de SAS fondées sur la parcimonie. Les résultats du chapitre 4 montrent que nos procédures permettent d'améliorer, de simplifier ou d'apporter une robustesse par rapport aux données aux techniques usuelles de SAS.

L'acquisition compressée est sans aucun doute l'application la plus populaire et la plus connue de la parcimonie. La parcimonie conduit à un protocole efficace d'acquisition des signaux parcimonieux. Une contribution majeure de notre travail concerne l'acquisition compressée des signaux à alphabet fini. D'une part, nous avons

établi une condition suffisante qui permet de rendre le problème soluble. D'autre part, nous avons présenté deux approches de reconstruction. Ces deux méthodes ont en commun une relaxation convexe de la norme ℓ_0 en norme ℓ_1. La première méthode est fondée sur la régularisation, où l'on introduit une fonction objective qui évalue le critère basé sur l'alphabet fini. La seconde méthode est fondée sur la parcimonie. Bien que la seconde méthode conduisent à de meilleures résultats, il nous a paru intéressant d'exposer la première approche compte rendu de sa faible complexité algorithmique par rapport à la seconde méthode.

9.2 Perspectives

Nous donnons ici quelques pistes à explorer pour étendre l'utilité des algorithmes développés lors de nos travaux de thèse. Les perspectives possibles de ces travaux sont nombreuses.

En ce qui concerne l'acquisition compressée des signaux à alphabet fini, les performances de l'algorithme ℓ_1 indicateur sont garanties par la conjecture 7.1. Les résultats des simulations, présentés dans le chapitre 8, appuient cette conjecture. Dans le cas $p = 2$, une démonstration théorique des résultats énoncés par la conjecture 7.1 (théorème 7.6) a été donnée, il s'agit alors de fournir une preuve théorique de la conjecture 7.1 dans le cas général.

Dans le cas où une telle démonstration est fournie, peut-on pour autant dire que le problème d'acquisition compressée des signaux à alphabet fini est résolu ? Bien évidemment, la réponse est non. Nos travaux explorent un nouveau domaine d'application de la parcimonie et s'inscrivent ainsi comme les premières briques de ce domaine. Plusieurs perspectives sont envisageables. L'étude du comportement des deux algorithmes (ℓ_1 additif et le ℓ_1 indicateur) vis à vis du bruit nous semble la plus incontournable. En effet, le modèle étudié dans cette partie ne comporte pas de bruit, il s'agit d'étudier la stabilité des algorithmes en présence de bruit dans les mesures. Cette extension au modèle bruité permettra d'augmenter le champs applicatif de ces travaux, notamment en communications numériques. Des résultats préliminaires encourageants ont été obtenus.

Par ailleurs, une piste de recherche intéressante concerne la conception des matrices de mesures à défaut d'une meilleure caractérisation de la condition (7.4).

D'autre part, dans la séparation de sources, il serait nécessaire d'évaluer ces algorithmes de manière beaucoup plus large et plus précise sur une base de données plus grande et par l'intermédiaire de la boite à outils *BSSEval*, téléchargeable à partir de [146].

D'un point de vue théorique, la question que l'on peut se poser est dans quelle mesure les tests statistiques utilisés ci dessus dans le cas d'un mélange instantané peuvent être éventuellement exploitées dans le cas convolutif, en particulier en complément des résultats discutés dans la section 4.6. Un autre aspect concerne l'extension du modèle au cas de bruit coloré.

La plupart des techniques de séparation aveugle de sources fondées sur la parcimonie procèdent en deux étapes, qui sont l'estimation de la matrice de mélange et l'estimation des sources. Etant donné que les techniques de résolution de ces deux sous-problèmes de séparation de sources peuvent être considérées comme des composantes élémentaires interchangeables, il serait souhaitable, d'un point de vue pratique, de construire une boîte à outils impliquant l'ensemble de ces composantes élémentaires pour des futurs études et développements.

ANNEXE A Séparation de sources

A.1 Estimation linéaire

La technique SUBSS présentée par Aïssa-El-Bey dans [1] identifie l'ensemble d'indices des sources présentes en un point temps-fréquence (t, f). Notons $\mathcal{K}$ cet ensemble d'indices. L'équation (2.2) devient alors :

$$\mathcal{S}_{\boldsymbol{x}}(t,f) = \boldsymbol{A}_{\mathcal{K}}\mathcal{S}_{\boldsymbol{s}_{\mathcal{K}}}(t,f) + \mathcal{S}_{\boldsymbol{b}}(t,f) \tag{A.1}$$

Afin de simplifier les écritures ci-après, nous omettons parfois les indices (t, f). Comme nous l'avons vu précédemment dans la sous section 2.4.4, les coefficients de la TFCT sont estimées via le pseudo inverse, c'est-à-dire,

$$\hat{\mathcal{S}}_{\boldsymbol{s}_{\mathcal{K}}} = \boldsymbol{A}_{\mathcal{K}}^{\#}\mathcal{S}_{\boldsymbol{x}} \tag{A.2}$$

L'étape de sélection de points multisources permet de rejeter les points temps-fréquences contenant uniquement du bruit. Afin "d'enlever" la contribution du bruit des points mutisources, nous proposons d'utiliser l'estimation de l'écart type du bruit fournie par le DATE. Ainsi, au lieu d'estimer les coefficients de la TFCT des sources présentes en un point temps-fréquence par l'équation (A.2), la séparation de sources est maintenant effectuée comme suit :

$$\widehat{\mathcal{S}}_{\boldsymbol{s}_{\mathcal{K}}} = \boldsymbol{R}_{\boldsymbol{s}_{\mathcal{K}}}\boldsymbol{A}_{\mathcal{K}}^{H}(\boldsymbol{A}_{\mathcal{K}}\boldsymbol{R}_{\boldsymbol{s}_{\mathcal{K}}}\boldsymbol{A}_{\mathcal{K}}^{H} + \widehat{\sigma}^2\boldsymbol{I}_n)^{-1}\mathcal{S}_{\boldsymbol{x}} \tag{A.3}$$

où $\widehat{\sigma}$ est l'écart type du bruit estimé par le DATE et $\boldsymbol{R}_{\boldsymbol{s}_{\mathcal{K}}} = \mathbb{E}[\mathcal{S}_{\boldsymbol{s}_{\mathcal{K}}}\mathcal{S}_{\boldsymbol{s}_{\mathcal{K}}}^{H}]$. Le calcul de l'estimateur linéaire optimal (A.3) est classique. Il s'agit de minimiser le risque $\mathbb{E}[\|\mathcal{S}_{\boldsymbol{s}_{\mathcal{K}}} - \boldsymbol{D}\mathcal{S}_{\boldsymbol{x}}\|^2]$ lorsque $\boldsymbol{D}$ parcourt l'ensemble des matrices de dimension $\#(\mathcal{K}) \times n$. Dans la pratique, la matrice $\boldsymbol{R}_{\boldsymbol{s}_{\mathcal{K}}}$ est inconnue et doit être estimée. D'après l'équation (A.1), nous avons

$$\boldsymbol{R_x} = \boldsymbol{A_K} \boldsymbol{R_{s_K}} \boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_n$$

Etant donnée que la matrice $\boldsymbol{A}$ et l'écart type du bruit σ ont été estimés auparavant, la matrice de covariance empirique du mélange $\widehat{\boldsymbol{R}}_{\boldsymbol{x}} = \frac{1}{\#t} \sum_t \mathcal{S}_{\boldsymbol{x}}(t,f) \mathcal{S}_{\boldsymbol{x}}(t,f)^H$, où $\#t$ représente le nombre de fenêtres temporelles sur le quel l'estimation est calculée, fournie une estimation de $\boldsymbol{R_{s_K}}$ (compte tenue de l'hypothèse (2.2))

ANNEXE B Acquisition compressée de signaux à alphabet fini

B.1 Quelques généralités sur les lois binomiales

La probabilité $p_{m,M}$ introduite dans le théorème de Wendel est la probabilité d'obtenir au plus $m-1$ fois "succès" à l'issue de $M-1$ épreuves de Bernoulli avec un paramètre $\frac{1}{2}$.
Soit une suite $(X_i)_{1\leq i\leq M}$ de variables aléatoires de Bernoulli avec un paramètre $\frac{1}{2}$, on pose $S=\sum_{i=1}^{M} X_i$.
Alors,

$$p_{m+1,M+1} = Pr(S \leq m-1)$$

D'après l'inégalité de Hoeffding [147],

$$\begin{cases} Pr(S \leq m-1) \leq e^{-2M(\frac{m}{M}-\frac{1}{2})^2} & \text{Si } \frac{m}{M} < \frac{1}{2} \\ Pr(S \leq m-1) \geq 1 - e^{-2M(\frac{m}{M}-\frac{1}{2})^2} & \text{Si } \frac{m}{M} > \frac{1}{2} \end{cases}$$

Pour (m, M) proportionnellement large, on a

$$\begin{cases} p_{m+1,M+1} \approx 0 & \text{Si } \frac{m}{M} < \frac{1}{2} \\ p_{m+1,M+1} \approx 1 & \text{Si } \frac{m}{M} > \frac{1}{2} \end{cases}$$

B.2 Preuve du Théorème 7.5

Montrons tout d'abord que le vecteur indicateur $\hat{\boldsymbol{s}}$ de $\boldsymbol{f}$ est une solution de $(TP_{\mathcal{F},1})$. Pour tout $i \in [\![1, N]\!]$, on pose

$$T_i = [\![(i-1)p + 1, ip]\!]$$

Soit $\boldsymbol{s} \in \mathbb{R}^{Np}$ tel que $\boldsymbol{\Phi s} = \boldsymbol{b}$, on a

$$\begin{aligned}
\|\boldsymbol{s}\|_1 = \sum_{i=1}^{Np} |s_i| &= \sum_{i=1}^{N} \sum_{j \in T_i} |s_j| \\
&\geq \sum_{i=1}^{N} \left| \sum_{j \in T_i} s_j \right| \\
&= \sum_{i=1}^{N} |(\boldsymbol{B_1 s})_i| \\
&= \|\boldsymbol{B_1 s}\|_1 \\
&= \|\mathbf{1}_N\|_1 \\
&= N
\end{aligned}$$

Comme $\|\hat{\boldsymbol{s}}\|_1 = \|\hat{\boldsymbol{s}}\|_0 = N$, le vecteur indicateur $\hat{\boldsymbol{s}}$ de $\boldsymbol{f}$ est une solution de $(TP_{\mathcal{F},1})$. On déduit que l'ensemble des solutions de $(TP_{\mathcal{F},1})$ est égale à

$$\mathcal{S} = \{\hat{\boldsymbol{s}} + \boldsymbol{h} \mid \boldsymbol{\Phi h} = \mathbf{0} \; et \; \|\hat{\boldsymbol{s}} + \boldsymbol{h}\|_1 = \|\hat{\boldsymbol{s}}\|_1\}$$

Soit $\boldsymbol{h} \in \mathbb{R}^{Np}$ tel que $\boldsymbol{B_1 h} = \mathbf{0}_N$ et $\|\hat{\boldsymbol{s}} + \boldsymbol{h}\|_1 = \|\hat{\boldsymbol{s}}\|_1$. Par construction de $\hat{\boldsymbol{s}}$ on a

$$\|\hat{\boldsymbol{s}} + \boldsymbol{h}\|_1 = \|\hat{\boldsymbol{s}}\|_1 = \|\boldsymbol{B_1}\hat{\boldsymbol{s}}\|_1 = \|\boldsymbol{B_1}(\hat{\boldsymbol{s}} + \boldsymbol{h})\|_1$$

Ainsi,

$$\sum_{i=1}^{N} \sum_{j \in T_i} |(\hat{\boldsymbol{s}} + \boldsymbol{h})_j| = \sum_{i=1}^{N} \left| \sum_{j \in T_i} (\hat{\boldsymbol{s}} + \boldsymbol{h})_j \right|$$

Or, pour tout $i \in [\![1, N]\!]$,

$$\left| \sum_{j \in T_i} (\hat{\boldsymbol{s}} + \boldsymbol{h})_j \right| \leq \sum_{j \in T_i} |(\hat{\boldsymbol{s}} + \boldsymbol{h})_j|$$

Alors,

$$\left| \sum_{j \in T_i} (\hat{\boldsymbol{s}} + \boldsymbol{h})_j \right| = \sum_{j \in T_i} |(\hat{\boldsymbol{s}} + \boldsymbol{h})_j|$$

Soit $i \in [\![1, N]\!]$. Comme $\sum_{j \in T_i} (\hat{\boldsymbol{s}} + \boldsymbol{h})_j = (\boldsymbol{B_1}\hat{\boldsymbol{s}} + \boldsymbol{B_1}\boldsymbol{h})_i = 1$, on a

$$\sum_{j \in T_i} |(\hat{\boldsymbol{s}} + \boldsymbol{h})_j| = 1$$

D'autre part, il existe $n_i \in T_i$ tel que $\hat{s}_{n_i} = 1$ et $\forall j \in T_i \setminus \{n_i\}, \hat{s}_j = 0$.
d'où,

$$\sum_{j \in T_i} |(\hat{\boldsymbol{s}} + \boldsymbol{h})_j| = |1 + h_{n_i}| + \sum_{\substack{j \in T_i \\ j \neq n_i}} |h_j|$$

Sachant que $\boldsymbol{B_1}\boldsymbol{h} = \boldsymbol{0}$, alors

$$h_{n_i} = -\sum_{\substack{j \in T_i \\ j \neq n_i}} h_j$$

On déduit que

$$\left| 1 - \sum_{\substack{j=1 \\ j \neq n_i}} h_j \right| = 1 - \sum_{\substack{j=1 \\ j \neq n_i}} |h_j|$$

Or,

$$\sum_{\substack{j=1 \\ j \neq n_i}}^{p} h_j \leq \sum_{\substack{j=1 \\ j \neq n_i}}^{p} |h_j|$$

Il en découle que

$$\sum_{\substack{j=1 \\ j \neq n_i}} h_j = \sum_{\substack{j=1 \\ j \neq n_i}} |h_j| \leq 1$$

d'où

$$\forall i \in [\![1, N]\!], \forall j \in T_i \setminus \{n_i\}, h_j \geq 0 \text{ et } h_{n_i} = -\sum_{\substack{j \in T_i \\ j \neq n_i}} h_j \geq -1$$

On pose

$$\mathcal{K}_{\hat{\boldsymbol{s}}}^{Np} = \{\boldsymbol{x} \in \mathbb{R}^{Np} |\ \forall i \in [\![1, N]\!], \forall j \in T_i \setminus \{n_i\}, x_j \geq 0 \text{ et } -1 \leq x_{n_i} \leq 0\}$$

Alors $\boldsymbol{h} \in \mathcal{K}_{\hat{\boldsymbol{s}}}^{Np}$.

Réciproquement, il est clair que si $\boldsymbol{h} \in \mathcal{K}_{\hat{\boldsymbol{s}}}^{Np}$ et $\boldsymbol{B_1}\boldsymbol{h} = \boldsymbol{0}$ alors $\|\hat{\boldsymbol{s}} + \boldsymbol{h}\|_1 = \|\hat{\boldsymbol{s}}\|_1$.

Etant donné un vecteur indicateur $\hat{\boldsymbol{s}}$ de $\boldsymbol{f}$, on a ainsi caractérisé l'ensemble des solutions possibles de $(TP_{\mathcal{F},1})$. Il en découle que la procédure de reconstruction par transformation parcimonieuse réussie à reconstruire $\boldsymbol{f}$ si et seulement si

$$\boldsymbol{\Phi h} = \mathbf{0},\ \ \boldsymbol{h} \in \mathcal{K}_{\hat{\boldsymbol{s}}}^{Np} \text{ n'admet pas de solution } \boldsymbol{h} \text{ tel que } \boldsymbol{B_a h} \neq \mathbf{0}$$

Ce qui est équivalent à

$$\boldsymbol{\Phi h} = \mathbf{0},\ \ \boldsymbol{h} \in \mathcal{C}_{\hat{\boldsymbol{s}}}^{Np} \text{ n'admet pas de solution } \boldsymbol{h} \text{ tel que } \boldsymbol{B_a h} \neq \mathbf{0}$$

où

$$\mathcal{C}_{\hat{\boldsymbol{s}}}^{Np} = \{\boldsymbol{x} \in \mathbb{R}^{Np} |\ \forall i \in [\![1, N]\!], \forall j \in T_i \setminus \{n_i\}, x_j \geq 0 \text{ et } x_{n_i} \leq 0\}$$

L'équivalence est due au fait que si $\boldsymbol{h} \in \mathcal{C}_{\hat{\boldsymbol{s}}}^{Np}$ alors $\boldsymbol{h} / \|\boldsymbol{h}\|_\infty \in \mathcal{K}_{\hat{\boldsymbol{s}}}^{Np}$.

Sachant que $\mathcal{C}^{Np}$ est l'union des $\mathcal{C}_{\hat{\boldsymbol{s}}}^{Np}$ où $\hat{\boldsymbol{s}}$ parcourt l'ensemble des vecteurs indicateurs, on a reconstruction parfaite de la procédure si et seulement si

$$\boldsymbol{\Phi h} = \mathbf{0},\ \ \boldsymbol{h} \in \mathcal{C}^{Np} \text{ n'admet pas de solution } \boldsymbol{h} \text{ tel que } \boldsymbol{B_a h} \neq \mathbf{0}$$

D'où le résultat. ∎

Bibliographie

[1] A. Aïssa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier. Underdetermined blind separation of nondisjoint sources in the time-frequency domain. *IEEE Transactions on Signal Processing*, 55(3) :897–907, Mars 2007.

[2] D. Pastor and A. M. Atto. *Wavelet shrinkage : from sparsity and robust testing to smooth adaptation*, chapter Fractals and Related Fields. Birkhauser, 2010.

[3] Dominique Pastor. A theoretical result for processing signals that have unknown distributions and priors in white Gaussian noise. *Computational statistics and data analysis*, 52(6) :3167 – 3186, february 2008.

[4] D. Donoho and I. Johnstone. deal spatial adaptation by wavelet shrinkage. *Biometrikay*, 81(3) :425–455, 1994.

[5] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4) :1289–1306, 2006.

[6] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2) :489–509, 2006.

[7] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier. Underdetermined blind audio source separation using modal decomposition. *EURASIP Journal on Audio, Speech, and Music Processing*, 2007(85438) :1–15, 2007.

[8] S. Arberet. *Estimation robuste et apprentissage aveugle de modèles pour la séparation de sources sonores.* PhD thesis, Université de Rennes 1, 2008.

[9] K. Abed-Meraim, S. Attallah, T.J. Lim, and M.O. Damen. A blind interference canceller in DS-CDMA. In *IEEE International Symposium on Spread Spectrum Techniques and Applications*, pages 358–362, Parsippany, Septembre 2000.

[10] I. Durán-Díaz and S. A. Cruces-Alvarez. A joint optimization criterion for blind DS-CDMA detection. *EURASIP Journal on Advances in Signal Processing*, 2007(79248) :1–11, 2007.

[11] A. Kachenoura, L. Albera, and L. Senhadji. Séparation aveugle de sources en ingénierie biomédicale. *Ingénierie et Recherche en BioMédicale (IRBM)*, 28(1) :20–34, Mars 2007.

[12] A. Rouxel, D. Le Guennec, and O. Macchi. Unsupervised adaptive separation of impulse signals applied to EEG analysis. In *IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP)*, volume 1, pages 420–423, Istanbul, Turquie, Juin 2000.

[13] V. Varajarajan and J. Krolik. Multichannel system identification methods for sensor array calibration in uncertain multipath environments. In *IEEE Signal Processing Workshop on Statistical Signal Processing (SSP)*, pages 297–300, Singapour, Octobre 2001.

[14] S. M. Aziz Sbaï, A. Aïssa-El-Bey, and D. Pastor. Contribution of Statistical Tests to Sparseness-Based Blind Source Separation. *EURASIP journal on applied signal processing*, Juillet 2012.

[15] S. M. Aziz Sbaï, A. Aïssa-El-Bey, and D. Pastor. Robust underdetermined blind audio source separation of sparse signals in the time-frequency domain. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3716–3719, Prague, République Tchèque, Mai 2011.

[16] O. L. Mangasarian and B. Recht. Probability of unique integer solution to a system of linear equations. *European Journal of Operational Research*, 214(1) :27–30, April 2011.

[17] S. M. Aziz Sbaï, A. Aïssa-El-Bey, and D. Pastor. Underdetermined Source Separation of Finite Alphabet Signals Via L1 Minimization. In *ISSPA 2012 : 11th IEEE International Conference on Information Sciences, Signal Processing and their Applications*, pages 625 – 628, Montreal, Quebec, Canada, Juillet 2012. SC - Dépt. Signal et Communications (Institut Mines-Télécom-Télécom Bretagne-UEB), Lab-sticc - Laboratoire en sciences et technologies de l'information, de la communication et de la connaissance (UMR CNRS 6285 - Télécom Bretagne - Université de Bretagne Occidentale - Université de Bretagne Sud).

[18] S. M. Aziz Sbaï, A. Aïssa-El-Bey, and D. Pastor. Recovery of finite alphabet signals from incomplete measurements. In *SPARS*, Edimbourg, Ecosse, Juin 2011. SC - Dépt. Signal et Communications (Institut Mines-Télécom-Télécom

Bretagne-UEB), Lab-sticc - Laboratoire en sciences et technologies de l'information, de la communication et de la connaissance (UMR CNRS 6285 - Télécom Bretagne - Université de Bretagne Occidentale - Université de Bretagne Sud).

[19] S. M. Aziz Sbaï, A. Aïssa-El-Bey, and D. Pastor. Recovery of Finite Alphabet Solutions of Underdetermined Linear System. *To be submitted to Elsevier Digital Signal Processing*, Octobre 2012.

[20] L. Bauer, editor. *The linguistics student's handbook*. Edinburgh : Edinburgh University Press. page 155, 2007.

[21] G. Lecointre. Le statut de la parcimonie. *Biosystema*, 22 :7–14, 2004.

[22] F. Santosa and W.W.Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7 :1307–1330, 1986.

[23] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3) :425–455, 1994.

[24] A. M. Atto. *Analyse en ondelettes et par paquets d'ondelettes de processus aléatoires stationnaires, et application à l'estimation non-paramètrique*. PhD thesis, Université Européenne de Bretagne, 2008.

[25] J. Karvanen and A. Cichocki. Measuring sparseness of noisy signals. In *4TH INT. SYMP. ON INDEPENDENT COMPONENT ANALYSIS AND BLIND SIGNAL SEPARATION (ICA2003*, pages 125–130, 2003.

[26] D. Zonoobi, , and Y.V. Venkatesh. Gini index as sparsity measure for signal reconstruction from compressive samples. *IEEE Journal of Selected Topics in Signal Processing*, 5 :927–932, 2011.

[27] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38 :713–718, 1992.

[28] I. Daubechies, editor. *Ten Lectures on Wavelets*. Siam, 1992.

[29] R. Penrose. A generalized inverse for matrices. *Cambridge Philosophical Society*, pages 406–413, 1954.

[30] H. Y. Liao and G. Sapiro. Sparse representations for limited data tomography. In *ISBI*, pages 1375–1378, Paris, France, Mai 2008.

[31] K. Huang and S. Aviyente. Sparse representation for signal classification. In *In Adv. NIPS*, 2006.

[32] B. K. Natarajan. Sparse approximate solutions to linear systems. *ISIAM J. Comput.*, 24 :227–234, 1995.

[33] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal Sci. Comput.*, 20(1) :33–61, 1999.

[34] I. F. Gorodnitsky, J. George, and B. D. Rao. Neuromagnetic source imaging with FOCUSS : A recursive weighted minimum norm algorithm. *IElectrocephalography and Clinical Neurophysiology*, 95 :231–251, 1995.

[35] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS : A re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3) :600–395, Mars 1997.

[36] E. J. Candes, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.*, 14 :877–905, 2008.

[37] Y.B.Zhao and D. Li. Reweighted ℓ_1 minimization for sparse solutions to underdetermined linear systems. *SIAM Journal on Optimization*, page To appear, 2012.

[38] R. Chartrand. Exact reconstructions of sparse signals via nonconvex minimization. *IEEE Signal Processing Letter*, 14(10) :707–710, 2007.

[39] M. Lai and J. Wang. An unconstrained ℓ_q minimization with $0 < q \leq 1$ for sparse solution of under-determined linear systems, 2009.

[40] J. Nocedal and S. J. Wright, editors. *Numerical Optimization.* Springer, second edition edition, 1999.

[41] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4) :373–395, 1984.

[42] S. P. Boyd and L. Vandenberghe, editors. *Convex Optimization.* Combridge University Press, 2004.

[43] M. Grant and S. Boyd. Cvx : Matlab software for disciplined convex programming (web page and software). http ://stanford.edu/ boyd/cvx, 2008.

[44] E. Candès and T.Tao. The dantzig selector : Statistical estimation when p is much larger than n. *Ann. Statist.*, 35(6) :2313–2351, 2007.

[45] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal*, 27(3) :265–274, 2009.

[46] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit : recursive function approximation with applications to wavelet decomposition. In *Dans Asilomar Conf. on Signals, Systems and Comput.*, volume 1, pages 40–44, Novembre 1993.

[47] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5) :2230–2249, 2009.

[48] J. Tropp, A. Gilbert, and M. Strauss. Algorithms for simultaneous sparse approximations ; part i : Greedy pursuit. *Signal Proc. – Sparse approximations in signal and image processing*, 86 :572–588, 2006.

[49] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12) :3397–3415, Décembre 1993.

[50] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Stark. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 58(2) :1094–1121, Février 2012.

[51] D. Needell and J.A. Tropp. Cosamp : Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(12) :301–321, 2009.

[52] D. L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Appl. Comput. Harmon. Anal.*, 1 :100–115, 1993.

[53] N. Wiener, editor. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series : With Engineering Applications*. The technology Press of the MIT, 2004.

[54] Y. Ephraïm and D. Malah. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-33(2) :443–445, 1985.

[55] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9 :1135–1151, 1981.

[56] Y. Deville and S. Hosseini. Blind identification and separation methods for linear-quadratic mixtures and/or linearly independent non-stationary signals. In *Actes dans 9th International Symposium on Signal Processing and its Applications (ISSPA)*, Sharjah, Emirates arabes unis, Février 2007.

[57] A. Taleb and C. Jutten. Source separation in post-nonlinear mixtures. *IEEE Transaction on Signal Processing*, 47(10) :2807–2820, 1999.

[58] C. Jutten, B. Babaie-ZadehHosseini, and S. Hosseini. Three easy ways for separating nonlinear mixtures ? *Signal Processing*, 84(2) :217–229, Février 2004.

[59] Y. Li, A. Cichoki, and L. Zhang. Blind separation and extraction of binary sources. *IEICE Transactions on Fundamentals*, E86-A(3) :580–589, March 2003.

[60] M. Rostami, M. Babaie-Zadeh, S. Samadi, and C. Jutten. Blind source separation of discrete finite alphabet sources using a single mixture. In *IEEE Statistical Signal Processing Workshop (SSP'11)*, pages 713–716, Nice, France, 2011.

[61] B. C. Ihm, D. J. Park, and Y. H. Kwon. Overcomplete blind source separation of finite alphabet sources. *IEICE Trans. on INF. et SYST.*, E84-D(1) :209–212, Janvier 2001.

[62] K. Abed-Meraim, A. Belouchrani, J. F. Cardoso, and E. Moulines. A blind source separation technique using second order statistics. *IEEE Transactions on Signal Processing*, 45(2) :434–444, Février 1997.

[63] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8 :411–419, 1995.

[64] A. Belouchrani and M. G. Amin. Blind source separation based on time-frequency signal representations. *IEEE Transactions on Signal Processing*, 46(11) :2888–2897, Novembre 1998.

[65] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua. Blind source separation using second order cyclostationary statistics. *IEEE Transactions on Signal Processing*, 49(4) :694–701, Avril 2001.

[66] J. F. Cardoso. Blind signal separation : Statistical principles. *Actes dans IEEE*, 86(10) :2009–2025, Octobre 1998.

[67] A. Hyvärinen, J. Karhunen, and E. Oja, editors. *Independent Component Analysis*. Wiley and Sons, 2001.

[68] P. Comon and C. Jutten, editors. *Handbook of Blind Source Separation : Independent Component Analysis and Blind Deconvolution*. Academic Press, Février 2010.

[69] P. Comon. Independent component analysis : a new concept ? *Signal Processing, Elsevier*, 36(3) :287–314, Avril 1994.

[70] J. Hérault, C. Jutten, and B. Ans. Détection de grandeur primitives dans un message composite par une architecture de calcul neuromimétrique en apprentissage non supervisé. In *Colloque GRETSI sur le traitement du signal et des images*, pages 1017–1022, Nice, France, Mai 1985.

[71] C. Jutten. *Calcul neuromimétrique et traitement du signal : analyse en composantes indépendantes*. PhD thesis, Université Joseph Fourier-Institut National Polytechnique de Grenoble, 1987.

[72] C. Jutten and J. Hérault. Blind separation of sources, part i : an adaptative algorithm based on neuromimetic architecture. *SignalProcessing*, 24 :1–10, Janvier 1991.

[73] L. Parra and C. Spence. Convolutive blind source separation of non-stationary sources. *IEEE transactions on Speech and Audio Processing*, 8(3) :320–327, Mai 2000.

[74] D. T. Pham and P. Garat. Blind separation of mixture of independent sources through a quasi-maximum likeelihood approach. *IEEE transactions on Signal Processing*, 45(7) :1712–1725, Juillet 1997.

[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, Series B*, 39(1) :1–38, 1977.

[76] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Dans Nature*, 401 :788–791, 1999.

[77] P. Pateero. Least squares formulation of robust non-negative factor analysis. *Dans Chemometrics and Intelligent Laboratory Systems*, 37 :23–35, 1997.

[78] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural and Information Processing Systems (NIPS)*, pages 556–562. MIT Press, 2000.

[79] C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the itakura-saito divergence. with application to music analysis. *Neural Computation*, 21(3) :793–830, Mars 2009.

[80] I. S. Dhillon and S. Sra. Generalized nonnegative matrix approximations with bregman divergences. *Advances in Neural Information Processing Systems (NIPS)*, 19, 2005.

[81] A. Cichocki, R. Zdunek, and S. Amari. Csiszar's divergences for non-negative matrix factorization : Family of new algorithms. In *Independent Component Analysis and Blind Signal Separation, ICA 2006*, 2006.

[82] C.-J. Lin. Projected gradient methods for non-negative matrix factorization. Technical report, National Taiwan university, 2005.

[83] R. Zdunek and A. Cichocki. Fast nonnegative matrix factorization algorithm using projected gradient approach for large scale problems. *Computational intelligence and neuroscience*, 2008.

[84] K. Drakakis, S. Rickard, R. de Fréin, and A. Cichocki. Analysis of financial data using non-negative matrix factorization. *International Mathematical Forum*, 3 :1853–1870, 2008.

[85] S. Young, P. Fogel, and D. Hawkins. Clustering scotch whiskies using non-negative matrix factorization. *Joint Newsletter for the Section on Physical and Engineering Sciences and the Quality and Productivity Section of the American Statistical Association*, 14 :11–13, 2006.

[86] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2003.

[87] P. Smaragdis. Non-negative matrix factor deconvolution ; extraction of multiple sound sources from monophonic inputs. In *5th International Conference on Independent Component Analysis and Blind Signal Separation*, 2004.

[88] E. Vincent, N. Bertin, and R. Badeau. Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 109–112, 2008.

[89] B. Wang and M. D. Plumbley. Investigating single-channel audio source separation methods based on non-negative matrix factorization. In *the ICA Research Network International Workshop*, 2006.

[90] A. Cichocki, R. Zdunek, and S. Amari. New algorithms for non-negative matrix factorization in applications to blind source separation. In *Acoustics, Speech and Signal Processing*, 2006.

[91] T. Virtanen. Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3) :1066–1074, Mars 2007.

[92] P. Smaragdis. Convolutive speech bases and their application to speech separation. *IEEE Transactions on Audio, Speech and Language Processing*, 15(1) :1–12, Janvier 2007.

[93] J. Durrieu, B. David, and G. Richard. A musically motivated mid-level representation for pitch estimation and musical audio source separation. *Selected Topics in Signal Processing, IEEE Journal of*, 5(6) :1180 –1191, oct. 2011.

[94] J.L. Durrieu and J.P. Thiran. Musical audio source separation based on user-selected F0 track. In *Proc. of International Conference on Latent Variable Analysis and Signal Separation*, Tel-Aviv, Israel, March 12-15 2012.

[95] T. Virtanen. Algorithm for the separation of harmonic sounds with time-frequency smoothness constraint. In *Proc. of the 6th Conf. on Digital Audio Effects (DAFx-03)*, pages 35–40, Londres, UK, September 2003.

[96] E. Vincent, N. Bertin, and R. Badeau. Adaptive harmonic spectral decomposition for multiple pitch estimation. *IEEE trans. on Audio, Speech and Language Proc. (TASLP)*, 18(3) :528–537, March 2010.

[97] P. Smaragdis. Relative-pitch tracking of multiple arbitrary sounds. *Journal of the Ac. Soc. of America*, 125 :3406–3413, 2009.

[98] G.J. Mysore and P. Smaragdis. Relative pitch estimation of multiple instruments. In *Proc. IEEE Intl. Conf. Acoust. Speech Signal Processing (ICASSP'09)*, pages 313–316, 2009.

[99] R. Gribonval and S. Lesage. A survey of blind component analysis for blind source separation : principles, perspectives and new challenges. In *Dans Europ. Sympo. on Artificial Neural Networks (ESANN)*, pages 323–330, Bruges, Belgique, 2006.

[100] L. Cohen, editor. *Time-frequency analysis*. Prentice Hall PTR, 1995.

[101] M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 1998.

[102] O. Bermond, E. Moulines, and J. F . Cardoso. Séparation et déconvolution aveugle de signaux bruités : modélisation par mélange de gaussiennes. In *GRETSI*, pages 1427–1430, Grenoble, France, Septembre 1997.

[103] L. Benaroya, F. Bimbot, and R. Gribonval. Audio sources separation with a single sensor. *IEEE transactions on Audio, Speech and Language Processing*, 14(1) :191–199, Janvier 2006.

[104] 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*, 15(1) :18–33, Juillet 2005.

[105] A. Jourjine, S. Rickard, and O. Yilmaz. Blind separation of disjoint orthogonal signals : Demixing N sources from 2 mixtures. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, pages 2985–2988, Istanbul, Turquie, Juin 2000.

[106] F. Abrard and Y. Deville. A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. *Signal Processing*, 85(7) :1389–1403, Juillet 2005.

[107] Y. Deville and M. Puigt. Temporal and time-frequency correlation-based blind source separation methods. part I : Determined and underdetermined linear instantaneous mixtures. *Signal Processing*, 87(3) :374–407, Mars 2007.

[108] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81(11) :2353–2362, Novembre 2001.

[109] O. Yilmaz and S. Rickard. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, 52(7) :1830–1847, Juillet 2004.

[110] T. Melia and S. Rickard. Underdetermined blind source separation in echoic environments using DESPRIT. *EURASIP Journal on Advances in Signal Processing*, 2007(86484) :1–19, 2007.

[111] S. Arberet, R. Gribonval, and F. Bimbot. A robust method to count and locate audio sources in a multichannel underdetermined mixture. *IEEE Transactions on Signal Processing*, 58(1) :121–133, Janvier 2010.

[112] F. J. Theis, C. G. Puntonet, and E. W. Lang. A histogram-based overcomplete ICA algorithm. In *Actes dans Independent Component Analysis and Blind Source Separation (ICA)*, pages 1071–1076, Nara, Japon, Avril 2003.

[113] A. Prieto, B. Prieto, C. G. Puntonet, A. Cañas, and P. Martin-Smith. Geometric separation of linear mixtures of sources : application to speech signals. In *Actes dans Independent Component Analysis and Blind Source Separation (ICA)*, pages 295–300, Janvier 1999.

[114] J. K. Lin and G. Grier. Faithful representation of separable distributions. *Neural Comp.*, 9 :1305–1320, 1997.

[115] F. Abrard and Y. Deville. Extension of the "time-frequency ratio of mixtures" blind source separation to more than 2 channels. In *Actes dans International Joint Conference on Neural Networks (ECMS)*, pages 1–6, Liberec, République Tchèque, Juin 2003.

[116] Y. Deville, M. Puigt, and B. Albouy. Time-frequency blind signal separation : extended methods, performance evaluation for speech sources. In *Actes dans International Joint Conference on Neural Networks (IJCNN)*, pages 255–260, Budapest, Hongrie, Juillet 2004.

[117] Y. Deville. Temporal and time-frequency correlation-based blind source separation methods. In *Independent Component Analysis and Blind Source Separation (ICA)*, pages 1059–1064, Nara, Japon, Avril 2003.

[118] B. Albouy and Y. Deville. A time-frequency blind source separation method based on segmented correlation function. In *International Work conference on*

Artificial and Natural Neural Networks(IWANN), volume 2, pages 289–296, Mao, Menorca, Espagne, Juin 2003.

[119] S. M. Kay, editor. *Fundamentals of Statistical Signal Processing, Volume II : Detection Theory*. Prentice Hall, 1998.

[120] D. Pastor and F.-X. Socheleau. Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences. *submitted to IEEE Transactions on Signal Processing*, 2011.

[121] D. Pastor, R. Gay, and A. Gronenboom. A sharp upper bound for the probability of error of likelihood ratio test for detecting signals in white gaussian noise. *IEEE Transactions on Information Theory*, 48(1) :228–238, January 2002.

[122] A. M. Atto, D. Pastor, and G. Mercier. Detection thresholds for non-parametric estimation. *Signal, Image and Video processing*, 2(3) :207–223, February 2008.

[123] F.R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346) :383–393, June 1974.

[124] P.J. Huber and E.M. Ronchetti. *Robust Statistics, second edition*. John Wiley and Sons, 2009.

[125] François-Xavier Socheleau, Dominique Pastor, and Abdeldjalil Aissa El Bey. Robust statistics based noise variance estimation : application to wideband interception of non-cooperative communications. *IEEE transactions on aerospace and electronic systems*, 47(1) :746 – 755, january 2011.

[126] Dominique Pastor. Signal norm testing in additive and independant standard Gaussian noise. Technical report, Institut Mines-Télécom ; Télécom Bretagne, UEB, Lab-STICC UMR CNRS 3192, 2011.

[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*, 2005(17) :2828–2847, 2005.

[128] R. Leonard. A database for speaker-independent digit recognition. In *IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP)*, volume 9, pages 328–331, San Diego, California, USA, Mars 1984.

[129] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8) :1207–1223, Août 2006.

[130] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra. A survey of convolutive blind source separation methods. In *Springer Handbook of Speech Processing*. Springer Press, Novembre 2007.

[131] 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*, 15(5) :1540–1550, Juillet 2007.

[132] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Journal of American Mathematical Society*, 22 :211–231, 2009.

[133] E. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory,*, 51(12) :4203–4215, Décembre 2005.

[134] M. Rudelson and R. Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8) :1025–1045, 2008.

[135] E. J. Candès and T. Tao. Near optimal signal recovery from random projections : Universal encoding strategies ? *IEEE Transactions on Information Theory*, 52(12) :5406–5425, Décembre 2006.

[136] C. H. Papadimitriou. On the complexity of unique solution. *Journal of the ACM*, 31(2) :392–400, 1984.

[137] O. L. Mangasarian and M. C. Ferris. Uniqueness of integer solution of linear equations. Technical report, Technical report, University of Wisconsin, Madison, Wisconsin, July 2009.

[138] J.G. Wendel. A problem in geometric probability. *Mathematica Scandinavica*, 11 :109–111, 1962.

[139] K. I. Diamantaras. Blind separation of two multi-level sources from a single linear mixture. In *IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP'08)*, pages 93–98, Cancun, Mexico, October 2008.

[140] D. L. Donoho and J. Tanner. Counting the faces of randomly-projected hypercubes and orthants, with applications. *Discrete Comput. Geom.*, 43(3) :522–541, April 2010.

[141] R. Gribonval and M. Nielson. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12) :3320–3325, December 2003.

[142] S. Foucart. A note on guaranted sparse recovery via ℓ_1 minimization. *App. Comput. Harmon. Anal.*, 29(1) :97–103, 2010.

[143] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proc. Nat. Acad. Sci. USA*, 100(5) :2197–2202, 2002.

[144] D. Needell E. J. Candes, Y. C. Eldar and P. Randall. Compressed sensing with coherent and redundant dictionaries. *In Press, Applied and Computational Harmonic Analysis*, 2011.

[145] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math*, 49 :906–931, 1989.

[146] C. Févotte, R. Gribonval, and E. Vincent. A toolbox for performance measurement in (blind) source separation. available at `http://bass-db.gforge.inria.fr/bss\_eval/`.

[147] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58 :13–30, Mar. 1963.