

Faculté des  
sciences  
d'Orsay



UNIVERSITÉ  
PARIS-SUD 11

UNIVERSITÉ PARIS-SUD 11  
LABORATOIRE CEA-IRFU/SEDI-SAP

# THÈSE DE DOCTORAT

Spécialité PHYSIQUE

Ecole Doctorale « Sciences et Technologies de l'Information des  
Télécommunications et des Systèmes »

Présentée par Jérôme Bobin

## DIVERSITÉ MORPHOLOGIQUE ET ANALYSE DE DONNÉES MULTIVALUÉES

Soutenue le 1<sup>er</sup> Octobre 2008 devant les membres du jury :

M. BIJAOUI Albert	(Rapporteur)
M. FADILI Jalal	(Co-directeur)
M. FROMENT Jacques	
M. PUGET Jean-Loup	(Président)
M. STARCK Jean-Luc	(Directeur de thèse)
M. VANDERGHEYNST Pierre	(Rapporteur)

*A Shanti . . .*

*« Le hasard ne profite qu'aux esprits préparés. »  
Louis Pasteur*

*« Qui fait confiance au non-savoir entre dans des ténèbres, aveugle  
Qui fait confiance au savoir entre dans des ténèbres plus épaisses encore. »  
Isha Upanishad - (VI ème siècle av. J-C)*

---

# Remerciements

En tout premier lieu, je tiens à remercier Jean-Luc Starck pour la très large liberté qu'il m'a laissée . . . entre deux sessions de Challenge WG2/Planck ! Plus sérieusement, j'ai tout particulièrement apprécié sa disponibilité, ses conseils avisés et sa clairvoyance. Au-delà des connaissances acquises durant ces trois années, j'ai pu ainsi approcher la recherche et le travail du chercheur sous un regard différent fort éloigné des sempiternels stéréotypes.

Je remercie mon co-directeur, Jalal Fadili, pour ses conseils avisés et ses encouragements.

Mes remerciements vont également à Yassir Moudden, qui a très largement contribué à cette thèse par de très nombreuses discussions et son constant soutien.

Messieurs Albert Bijaoui et Pierre Vandergheynst ont accepté d'être les rapporteurs de cette thèse, et je les en remercie, de même que pour leur participation au Jury.

Messieurs Jean-Loup Puget et Jacques Froment m'ont fait l'honneur de participer au Jury de soutenance ; je les en remercie profondément.

Je tiens aussi à mentionner le plaisir que j'ai eu à travailler au sein de l'IRFU/SEDI et Sap, et j'en remercie ici tous les membres.

Je remercie enfin toute ma famille et mes amis pour leur soutien sans faille durant ces trois années.

---





**Titre** Diversité morphologique et analyse de données multivaluées

**Résumé** Ce mémoire de thèse introduit une nouvelle méthodologie pour l'analyse de données multivaluées et plus particulièrement pour la résolution de problèmes de séparation de sources. Les approches classiques de séparation de sources cherchent à tirer partie d'une certaine diversité entre les composantes à estimer. En parallèle, la dernière décennie a vu apparaître une nouvelle communauté scientifique dont les travaux ont essentiellement porté sur la recherche et l'utilisation de représentations dites parcimonieuses des signaux. Ce travail de thèse a été dédié à la conception d'une nouvelle approche des problèmes de séparation de source aveugle fondée sur l'utilisation de telles représentations parcimonieuses des signaux. Nous introduisons en particulier une nouvelle méthode de séparation de source aveugle intitulée Analyse en Composantes Morphologiques Généralisées que nous appliquerons dans divers domaines : l'analyse de données multispectrales, l'extraction de composantes astrophysiques dans cadre de la mission spatiale Planck de l'Agence Spatiale Européenne et la restauration adaptative de données multivaluées.

**Mots-clés** Diversité morphologique, parcimonie, représentations parcimonieuses, ondelettes, séparation de source aveugle, données multivaluées

**Title** Morphological diversity and multivalued data analysis

**Abstract** This report describes a new methodology for multivalued data analysis. More precisely, we will focus on source separation problems. Classical source separation techniques aim at devising quantitative measures of diversity between the sources to be estimated. During the last decade, an active field of research has focused on using sparse representations for a wide range of signal processing problems. In this report, we introduce a new way of solving blind source separation problems in the so-called sparse representation paradigm. We devise a new blind source separation technique coined Generalized Morphological Component Analysis. This source separation method has been applied to various multivalued data analysis issues : hyperspectral data analysis, extraction of astrophysical components in the scope of the Planck mission of the European Space Agency and adaptive restoration of multivalued data.

**Keywords** Morphological diversity, sparsity, sparse representations, wavelets, blind source separation, multivalued data

# Table des matières

<b>Table des figures</b>	<b>xiii</b>
<b>Liste des publications</b>	<b>xiv</b>
<b>Sigles et abréviations</b>	<b>xvii</b>
<b>Notations et définitions</b>	<b>xviii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Prologue</b>	<b>3</b>
1.1 De la représentation d'un signal . . . . .	4
1.1.1 Analogies en neurosciences . . . . .	4
1.1.2 Structures, parcimonie et représentations . . . . .	5
1.2 Diversité morphologique . . . . .	8
1.2.1 Analyse en composantes morphologiques . . . . .	8
1.2.2 Le bon grain et l'ivraie . . . . .	9
<b>2 Dictionnaires multivalués et décomposition parcimonieuse</b>	<b>11</b>
2.1 Décomposition d'un signal dans un dictionnaire redondant . . . . .	12
2.1.1 De la décomposition parcimonieuse de signaux . . . . .	12
2.1.2 Convexification . . . . .	13
2.1.3 Solution approchée . . . . .	14
2.1.4 Algorithmes de décomposition dans des dictionnaires redondants . . . . .	15
2.2 L'analyse en composantes morphologiques multicanale . . . . .	17
2.2.1 Analyse en composantes morphologiques monocanale . . . . .	17
2.2.2 MCA, un algorithme de décomposition parcimonieuse ? . . . . .	22
2.3 Vers une représentation parcimonieuses des données multivaluées . . . . .	28
2.3.1 Dictionnaires multicanaux . . . . .	29
2.3.2 Extension . . . . .	30
2.4 Analyse en composantes morphologiques multivaluées . . . . .	31
2.4.1 Principe . . . . .	31
2.4.2 Algorithme . . . . .	32

---

<b>3</b>	<b>Données multivaluées et séparation de sources</b>	<b>35</b>
3.1	Données multivaluées : du modèle de mélange à la séparation de sources . .	36
3.1.1	Données multivaluées et modèle de mélange . . . . .	36
3.1.2	Séparation par analyse en composantes décorrélées . . . . .	37
3.1.3	Séparation et analyse en composantes indépendantes (ACI) . . . . .	38
3.2	Séparation et analyse en composantes mutuellement parcimonieuses . . . . .	43
3.2.1	Parcimonie et séparation de sources . . . . .	43
3.2.2	Analyse en composantes mutuellement parcimonieuses . . . . .	50
<b>4</b>	<b>Diversité morphologique pour la séparation de sources en aveugle</b>	<b>59</b>
4.1	Parcimonie et diversité morphologique pour la séparation de sources en aveugle . . . . .	60
4.1.1	Extension du cadre mMCA à la séparation de sources en aveugle . .	60
4.1.2	Analyse en composantes morphologiques généralisées . . . . .	61
4.1.3	Le rôle du seuillage dans l'algorithme GMCA . . . . .	64
4.1.4	Premiers résultats expérimentaux . . . . .	64
4.2	Une version rapide de GMCA . . . . .	65
4.2.1	Principe . . . . .	65
4.2.2	Algorithme . . . . .	68
4.3	Résultats en SSA . . . . .	74
4.4	Estimation du nombre de sources . . . . .	82
<b>5</b>	<b>Extension à l'analyse de données hyperspectrales</b>	<b>89</b>
5.1	Morphologies spatiales et spectrales . . . . .	90
5.1.1	Modélisation de l'information spectrale . . . . .	90
5.1.2	L'algorithme HypGMCA . . . . .	92
5.1.3	Comparaison avec GMCA . . . . .	94
5.2	Application à l'extraction de composantes physiques à partir de données hyperspectrales réelles . . . . .	96
5.2.1	Données Mars Express . . . . .	96
5.2.2	Résultats préliminaires . . . . .	97
<b>6</b>	<b>Application aux données Planck</b>	<b>103</b>
6.1	Les données CMB : contexte et enjeux . . . . .	104
6.1.1	Fond diffus cosmologique . . . . .	104
6.1.2	Mission spatiale Planck . . . . .	104
6.1.3	Objectifs . . . . .	108
6.2	Hypothèse du modèle linéaire . . . . .	109
6.2.1	Modélisation linéaire sur cartes carrées . . . . .	109
6.3	Application aux données Planck/WG2 . . . . .	112
6.3.1	Des données réalistes . . . . .	112
6.3.2	Un modèle physique pour GMCA . . . . .	113
6.3.3	Résultats . . . . .	116

<b>7</b>	<b>Nouvelles perspectives pour la restauration de données multivaluées</b>	<b>132</b>
7.1	Restauration de données multicanales . . . . .	133
7.2	Application à la restauration adaptative de données multicanales . . . . .	138
<b>8</b>	<b>Epilogue</b>	<b>146</b>
<b>A</b>	<b>Conditions de convergence de l'Analyse en Composantes Morphologiques (MCA)</b>	<b>148</b>
<b>B</b>	<b><i>Compressed Sensing en Astronomie</i></b>	<b>152</b>
B.1	A brief introduction to compressed sensing . . . . .	154
B.1.1	The gist of compressed sensing . . . . .	154
B.1.2	Signal recovery . . . . .	155
B.2	Compressed Sensing in Astronomy . . . . .	156
B.2.1	Compressed Sensing as a way of designing sensors . . . . .	156
B.2.2	A flexible compression/decompression scheme . . . . .	157
B.2.3	Example : joint recovery of multiple observations . . . . .	159
B.3	A real-world application : the Herschel project . . . . .	160
B.3.1	The Herschel dilemma . . . . .	160
B.3.2	Compressed sensing for the Herschel data . . . . .	162
B.3.3	A toy-example . . . . .	163
B.3.4	Realistic data . . . . .	165
B.4	Conclusion . . . . .	166
B.5	Acknowledgment . . . . .	167
	<b>Bibliographie</b>	<b>172</b>

# Table des figures

1.1	Deux images "naturelles". . . . .	4
1.2	Une image synthétique complexe. . . . .	9
1.3	Composantes morphologiques. . . . .	9
2.1	Stratégie de sélection du seuil. . . . .	22
2.2	RSB en Decibels de la première composante morphologique extraite MCA/MOM ( $-20 \log_{10} (\ \varphi_1 - \tilde{\varphi}_1\ _{\ell_2})$ ). <b>Abscisse :</b> $p_1$ . <b>Ordonnées :</b> $p_2$ . . . . .	24
2.3	<b>Haut à droite :</b> courbe $\ell_1/\ell_2$ de MCA/MOM ( <i>trait continu</i> ) et de Basis Pursuit ( <i>trait discontinu</i> ). <b>Haut à droite :</b> courbe $\ell_0/\ell_1$ de MCA/MOM ( <i>trait continu</i> ) et de Basis Pursuit ( <i>trait discontinu</i> ). <b>Bas à gauche :</b> courbe d'approximation non linéaire ( $\ell_0/\ell_2$ ) de MCA/MOM ( <i>trait continu</i> ) et de Basis Pursuit ( <i>trait discontinu</i> ). <b>Bas à droite :</b> comparaison du temps d'exécution (en secondes) : MCA/MOM ( <i>trait continu</i> ) et Basis Pursuit ( <i>trait discontinu</i> ). . . . .	25
2.4	Exemples de données multivaluées. . . . .	29
3.1	<b>Un point de vue intuitif :</b> représentation en nuage de points des échantillons de la source $\mathbf{s}_1$ en fonction de la source $\mathbf{s}_2$ . . . . .	45
3.2	<b>Un point de vue intuitif :</b> courbes équiprobabilité de la densité de probabilité jointe $f_S(\mathbf{s}_1, \mathbf{s}_2)$ . . . . .	46
3.3	<b>Un point de vue intuitif - Gauche :</b> représentation en nuage de points des échantillons de la source $\mathbf{s}_1$ en fonction de la source $\mathbf{s}_2$ . <b>Droite :</b> Représentation en nuage de points des échantillons de l'observation $\mathbf{x}_1$ en fonction de $\mathbf{x}_2$ . . . . .	46
3.4	<b>Sources non indépendantes - Gauche :</b> représentation en nuage de points des échantillons de la source $\mathbf{s}_1$ en fonction de la source $\mathbf{s}_2$ . <b>Droite :</b> Représentation en nuage de points des échantillons de l'observation $\mathbf{x}_1$ en fonction de $\mathbf{x}_2$ . . . . .	57
4.1	Evolution du critère sur la matrice de mélange $C_A$ lorsque la variance du bruit varie : GMCA ( <i>trait continu</i> ), EFICA, ( $\star$ ) :, RNA (+). <b>Abscisses :</b> rapport signal à bruit en dB. <b>Ordonnées :</b> valeur du critère sur la matrice de mélange. . . . .	65

---

4.2	Courbes équiprobables de la <i>pdf</i> jointe de 2 sources indépendantes sources générées selon une distribution gaussienne généralisée de paramètre $\theta = 0, 5$ . <b>Gauche :</b> <i>pdf</i> jointe de sources indépendantes. <b>Droite :</b> <i>pdf</i> jointe de 2 mélanges. . . . .	71
4.3	Courbes équiprobables de la <i>pdf</i> jointe de 2 sources indépendantes sources générées selon une distribution gaussienne généralisée de paramètre $\theta = 0, 5$ après seuillage. <b>Gauche :</b> <i>pdf</i> jointe de sources indépendantes. <b>Droite :</b> <i>pdf</i> jointe de 2 mélanges après seuillage. . . . .	71
4.4	<b>Le rôle de la parcimonie - première colonne :</b> sources originales. <b>Seconde colonne :</b> mélanges avec bruit gaussien additif (RSB = 19dB). <b>Troisième colonne :</b> sources estimées avec GMCA en DWT. <b>Quatrième colonne :</b> sources estimées avec GMCA dans le dictionnaire redondant $\Phi$ formé par l'union d'une DWT et d'une DCT. . . . .	75
4.5	<b>Rôle de la parcimonie :</b> évolution du critère sur la matrice de mélange pour DWT-GMCA (trait discontinu) et (DWT+DCT)-GMCA (trait continu). . .	76
4.6	Sources originales de taille $256 \times 256$ . . . . .	77
4.7	Mélanges sans bruit. . . . .	78
4.8	Sources estimées par GMCA. . . . .	79
4.9	<b>GMCA conduit une solution plus parcimonieuse - Abscisses :</b> nombre d'itérations. <b>Ordonnées :</b> divergence de parcimonie $\ \tilde{\mathbf{S}}\ _1 - \ \mathbf{S}\ _1$ . . . . .	80
4.10	<b>Haut :</b> sources originales de taille $256 \times 256$ . <b>Bas :</b> deux mélanges. Le bruit gaussien additif est tel que le RSB est égal à 10dB. . . . .	81
4.11	Evolution du coefficient de corrélation entre sources estimées et sources originales pour différentes valeurs de variance du bruit. Trait continu : GMCA, trait pointillé : JADE, symbole ( $\star$ ) : EFICA, symbole (+) : RNA. <b>Abscisses :</b> RSB en dB. <b>Ordonnées :</b> coefficient de corrélation entre les sources estimées et les sources originales. . . . .	82
4.12	Evolution du critère sur la matrice de mélange $C_A$ pour différentes valeurs de variance du bruit. Trait continu : GMCA, trait pointillé : JADE, symbole ( $\star$ ) : EFICA, symbole (+) : RNA. <b>Abscisses :</b> RSB en dB. <b>Ordonnées :</b> critère sur la matrice de mélange $C_A$ . . . . .	83
4.13	Ensemble des 15 sources utilisées pour l'évaluation du comportement de GMCA en fonction du nombre de sources. . . . .	83
4.14	<b>Gauche :</b> évolution du critère normalisé sur la matrice de mélange lorsque le nombre d'itérations $P_{\max}$ augmente. <b>Abscisses :</b> nombre d'itérations. <b>Ordonnées :</b> critère normalisé sur la matrice de mélange. Le nombre de sources varie comme suit : trait continu : $n = 2$ , trait discontinu : $n = 5$ , symbole ( $\square$ ) : $n = 10$ , symbole ( $\circ$ ) : $n = 15$ . <b>Droite :</b> temps de calcul lorsque le nombre de sources croît. <b>Abscisses :</b> nombre de sources. <b>Ordonnées :</b> temps de calcul en secondes. Le nombre d'itérations varie comme suit : trait continu : $P_{\max} = 10$ , trait discontinu : $P_{\max} = 100$ , symbole ( $\circ$ ) : $P_{\max} = 1000$ . . . . .	84

4.15	<b>Estimation du nombre de sources avec GMCA- Abscisses :</b> nombre de sources réel. <b>Ordonnées :</b> Nombre de sources estimée avec GMCA. Chaque point est calculé comme la moyenne de 25 réalisations. . . . .	87
4.16	<b>GMCA (points) versus PCA (trait continu) - Abscisses :</b> nombres de sources. <b>Ordonnées - Haut :</b> RSB en dB <b>Bas :</b> critère de parcimonie $C_1$ . Chaque point est calculé à partir de 25 réalisations. . . . .	88
5.1	Ensemble d'images utilisées lors des expérimentations. . . . .	95
5.2	Six mélanges de taille $128 \times 128$ parmi les 128 canaux. Le RSB est de 20dB. . . . .	96
5.3	<b>figures de gauche :</b> sources estimées avec la version originale ( <i>i.e.</i> sans contrainte spectrale) de GMCA. <b>figures de droite :</b> sources estimées avec HypGMCA. . . . .	99
5.4	<b>Evolution du critère sur la matrice de mélange <math>C_A</math> comme une fonction du RSB en dB.</b> <i>Ligne continue :</i> résultats avec GMCA. Symbole $\bullet$ : résultats avec HypGMCA. . . . .	100
5.5	<b>Abscisse :</b> nombre de sources. <b>Ordonnées- gauche :</b> RSB de la solution. <b>Droite :</b> critère de parcimonie $C_{\ell_1}$ . <i>Ligne continue :</i> résultats de GMCA. Symbole $\bullet$ : résultats de HypGMCA. . . . .	100
5.6	<b>De gauche à droite :</b> observations Mars Express pour les fréquences 1,38 - 1,75 - 1,94 et 2,41 $\mu m$ . . . . .	101
5.7	<b>Image de gauche :</b> source de glace de $CO_2$ estimée. <b>Graphique de droite - ligne continue :</b> spectre de référence de la glace de $CO_2$ . Symbole $\bullet$ : spectre estimé. . . . .	102
5.8	<b>Image de gauche :</b> source de glace de $H_2O$ estimée. <b>Graphique de droite - ligne continue :</b> spectre de référence de la glace de $H_2O$ . Symbole $\bullet$ : spectre estimé. . . . .	102
6.1	Spectre de puissance théorique du CMB et mesures issues des dernières missions d'observation (issu de (Hin06)). . . . .	105
6.2	<b>Sources simulées - Gauche :</b> CMB. <b>Centre :</b> poussière galactique. <b>Droite :</b> SZ. . . . .	110
6.3	<b>Données simulées Planck/HFI - RSB global (sur les 6 canaux) = 2.7dB</b> . . . . .	110
6.4	<b>Colonne de gauche :</b> coefficients de corrélation entre les sources originales et leurs estimées - <b>Colonne de droite :</b> écart-type des coefficients de corrélation. <b>Première ligne :</b> CMB. <b>Seconde ligne :</b> poussière galactique. <b>Troisième ligne :</b> composante SZ. <b>Quatrième ligne :</b> critère sur la matrice de mélange $C_A$ . <b>Légende :</b> JADE : <i>ligne pointillée avec <math>\square</math></i> - SMICA : <i>tirets avec <math>\circ</math></i> - GMCA : <i>ligne continue</i> . . . . .	120
6.5	<b>Colonne de gauche :</b> coefficients de corrélation entre les sources originales et leurs estimées - <b>Colonne de droite :</b> écart-type des coefficients de corrélation. <b>Première ligne :</b> CMB. <b>Seconde ligne :</b> poussière galactique. <b>Troisième ligne :</b> SZ. <b>Légende :</b> GMCA avec <i>a priori</i> : <i>ligne en pointillé</i> - GMCA : <i>ligne continue</i> . <b>Abscissa :</b> RSB in dB. . . . .	121

6.6	Données WG2/CH2. . . . .	122
6.7	<b>Haut</b> : composante CMB originale. <b>Bas</b> : composante CMB estimée par GMCA. <b>Unité</b> : $10^{-3}$ K. . . . .	123
6.8	<b>Haut</b> : résidu GMCA convolué à 20 arcminutes. <b>Bas</b> : résidu ADAMIS convolué à 20 arcminutes. <b>Unité</b> : $10^{-3}$ K. . . . .	124
6.9	<b>Haut</b> : résidu MEM convolué à 20 arcminutes. <b>Bas</b> : résidu CCA convolué à 20 arcminutes. <b>Unité</b> : $10^{-3}$ K. . . . .	125
6.10	<b>Haut</b> : résidu GMCA convolué à 60 arcminutes. <b>Bas</b> : résidu ADAMIS convolué à 60 arcminutes. <b>Unité</b> : $10^{-3}$ K. . . . .	126
6.11	<b>Haut</b> : résidu MEM convolué à 60 arcminutes. <b>Bas</b> : résidu CCA convolué à 60 arcminutes. <b>Unité</b> : $10^{-3}$ K. . . . .	127
6.12	<b>Haut</b> : écart-type du résidu convolué à 20 arcminutes par bande de latitudes de largeur $20^\circ$ . <b>Bas</b> : écart-type du résidu convolué à 60 arcminutes par bande de latitudes de largeur $20^\circ$ . <b>Abscisses</b> : latitude centrale. La bande centrée en $90^\circ$ n'est pas représentée. <b>Ordonnées</b> : écart-type du résidu en mK. . . . .	128
6.13	<b>Kurtosis des cartes estimées par échelle d'ondelette</b> - <b>Abscisses</b> : échelle d'ondelette ; l'équivalence en harmonique sphérique est donnée en haut de la figure. <b>Ordonnées</b> : kurtosis. <b>Trait continu</b> : kurtosis calculé à partir des données masquées puis après IDM. <b>Trait discontinu</b> : kurtosis calculé à partir des données masquées. . . . .	129
6.14	<b>Spectre de puissance <math>C(\ell)</math></b> - <b>Abscisses</b> : échelle angulaire en échelle logarithmique. <b>Ordonnées</b> : température d'antenne au carré ( $\text{mK}^2$ ) normalisée en $\ell(\ell + 1)/(2\pi)$ . . . . .	130
6.15	<b>Erreur d'estimation du spectre de puissance <math>e(\ell)</math></b> - <b>Abscisses</b> : échelle angulaire en échelle logarithmique. <b>Ordonnées</b> : température d'antenne au carré ( $\text{mK}^2$ ) normalisée en $\ell(\ell + 1)/(2\pi)$ . . . . .	131
7.1	<b>Haut à gauche</b> : image originale de taille $256 \times 256$ avec un bruit additif blanc gaussien. Le RSB est égal à 15 dB. <b>Haut à droite</b> : débruitage par ondelettes de chaque couche couleur indépendamment. <b>Bas</b> : débruitage dans le dictionnaire $\Pi \otimes \Phi^{(2)}$ ; $\Pi$ a été estimé par GMCA. . . . .	139
7.2	Versions agrandies. <b>Haut à gauche</b> : image originale avec un bruit additif blanc gaussien. Le RSB est égal à 15 dB. <b>Haut à droite</b> : débruitage par ondelettes de chaque couche couleur indépendamment. <b>Bas</b> : débruitage dans le dictionnaire $\Pi \otimes \Phi^{(2)}$ ; $\Pi$ a été estimé par GMCA. . . . .	140
7.3	Débruitage adaptatif d'images couleurs. <b>Abscisses</b> : RSB en dB. <b>Ordonnées</b> : gain de RSB en dB par rapport à un débruitage sur les couches RVB. Trait continu : débruitage adaptatif via GMCA, Trait discontinu : JADE, Symbole '●' YUV, Symbole '+' YCC. . . . .	141



7.4	Restauration d'images couleurs. (a) image originale <i>Barbara</i> . (b) image masquée - 90% des pixels couleurs sont manquants. (c) image restaurée en utilisant MCA sur chaque canal séparément. (d) image restaurée à partir de l'approche multicanale adaptative fondée sur mMCA. . . . .	144
7.5	Versions agrandies. (a) image originale <i>Barbara</i> . (b) image masquée - 90% des pixels couleurs sont manquants. (c) image restaurée en utilisant MCA sur chaque canal séparément. (d) image restaurée à partir de l'approche multicanale adaptative fondée sur mMCA. . . . .	145
B.1	<b>Left :</b> Input image $x$ of size $512 \times 512$ . <b>Middle :</b> Estimate from the average of 100 images compressed by JPEG2000 with a compression rate $\rho = 0.1$ . <b>Right :</b> Estimate from 100 pictures compressed by CS with a compression rate $\rho = 0.1$ . . . . .	161
B.2	<b>Zoomed versions - Left :</b> Input image $x$ of size $512 \times 512$ . <b>Middle :</b> Estimate from the average of 100 images compressed by JPEG2000 with a compression rate $\rho = 0.1$ . <b>Right :</b> Estimate from 100 pictures compressed by CS with a compression rate $\rho = 0.1$ . . . . .	168
B.3	<b>Top left :</b> Original image of size $128 \times 128$ the total intensity of which is $f = 3500$ . <b>Top right :</b> First input noisy map (out of 6). White Gaussian with variance $\sigma_n^2 = 1$ was added. <b>Bottom left :</b> Mean of the 6 input images. <b>Bottom right :</b> Reconstruction from noiselet-based CS projections. The iterative algorithm described in Section B.2 has been used with 100 iterations.	169
B.4	<b>Detection rate when the intensity of the input data varies :</b> <i>Solid line</i> Resolution defined by the Rayleigh criterion of the CS-based reconstruction. <i>o</i> : Resolution of the solution provided by the mean of 6 images. . . .	169
B.5	<b>Top left :</b> Original image of size $128 \times 128$ the total intensity of which is $f = 1000$ . <b>Top right :</b> First input noisy map (out of 6). White Gaussian with variance $\sigma_n^2 = 1$ was added. <b>Bottom left :</b> Mean of the 6 input images. <b>Bottom right :</b> Reconstruction from noiselet-based CS projections. The iterative algorithm has been used with 100 iterations. . . . .	170
B.6	<b>The Rayleigh criterion - Left :</b> The point sources are not resolved. <b>Middle :</b> Resolution limit. <b>Right :</b> Fully resolved point sources. . . . .	170
B.7	<b>Top left :</b> Original image of size $32 \times 64$ with a total intensity of $f = 4500$ . <b>Top right :</b> First input noisy map (out of 6). The PACS data already contains approximately Gaussian noise. <b>Bottom left :</b> Mean of the 6 input images. <b>Bottom right :</b> Reconstruction from noiselet-based CS projections. The iterative algorithm has been used with 100 iterations. . . . .	171

# Liste des publications

## Chapitre d'ouvrage

- ◆ J.Bobin, J.-L. Starck, Y.Moudden, J. Fadili, *Blind Source Separation : the Sparsity Revolution*, Advances in Imaging and Electron Physics, Vol. 152, p. 221-298 - Peter W. Hawkes Ed. - 2008.

## Articles de revue avec comité de lecture

- ◆ S.M.Leach, J.-F.Cardoso, C.Baccigalupi, R.B.Barreiro, M.Betoule, J.Bobin, A.Bonaldi, G.de Zotti, J.Delabrouille, C.Dickinson, H.K.Eriksen, J.González-Nuevo, F.K.Hansen, D.Herranz, M.LeJeune, M.López-Caniego, E.Martinez-González, M.Massardi, J.-B.Melin, M.-A.Miville-Deschênes, G.Patanchon, S.Prunet, S.Ricciardi, E.Salerno, J.L.Sanz, J.-L.Starck, F.Stivoli, V.Stolyarov, R.Stompor and P.Vielva , *Component separation methods for the Planck mission*, Astronomy and Astrophysics, 2008 - en cours de publication.
  - ◆ J.Bobin, J.-L. Starck and R. Ottensamer, *Compressed Sensing in Astronomy*, IEEE Journal of Selected Topics in Signal Processing, 2008 - en cours de publication.
  - ◆ J.Bobin, Y.Moudden, J. Fadili, J.-L. Starck, *Morphological Diversity and Sparsity for Multivariate Data Restoration*, Journal of Mathematical Imaging and Vision, 2008 - en cours de publication.
  - ◆ J.Bobin, Y.Moudden, J.-L. Starck, J. Fadili, N. Aghanim, *SZ and CMB reconstruction using GMCA*, Statistical Methodology, 2008 - Vol. 4, p.307-317.
  - ◆ J.Bobin, J.-L. Starck, J. Fadili, Y.Moudden, *Sparsity and Morphological Diversity in Blind Source Separation*, IEEE Transactions on Image Processing, Vol.16, N°11, p. 2662-2674, November 2007.
  - ◆ J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, and D.L. Donoho, *Morphological Component Analysis : An Adaptive Thresholding Strategy*, IEEE Transactions on Image Processing, Vol.16, N°11, p. 2675-2681, November 2007.
  - ◆ P. Abrial, Y. Moudden, J.L. Starck, J. Bobin, M.J. Fadili, B. Afeyan and M. Nguyen, *Morphological Component Analysis and Inpainting on the Sphere : Ap-*
-

*plication in Physics and Astrophysics*, Journal of Fourier Analysis and Applications (JFAA), special issue on « Analysis on the Sphere », Vol.13, N°6, pp 729-748, 2007.

◆ J. Bobin, Y. Moudden, J.-L. Starck and M. Elad, *Morphological Diversity and Source Separation*, IEEE Signal Processing Letters, Vol.13, N°7, p. 409-412, July 2006.

## Articles de conférence

◆ J. Bobin, J.-L. Starck, *Compressed Sensing in Astronomy*, Astronomical Data Analysis (ADA V), Marseille, Grèce, Mai 2008.

◆ Y. Moudden, J. Bobin, J.-L. Starck, *An iterative thresholding algorithm for joint deconvolution and source separation of multichannel data*, Astronomical Data Analysis (ADA V), Marseille, Grèce, Mai 2008.

◆ J. Bobin, Y. Moudden, J. Fadili, J.-L. Starck, *Morphological diversity and sparsity in blind sources separation*, Int. Conference on Independent Component Analysis and Signal Separation (ICA 2007), London, UK, September 2007.

◆ J. Bobin, J. Fadili, Y. Moudden, J.-L. Starck, *Morphological diversity and sparsity : New insights into multivariate data analysis*, Proceedings of the SPIE conference wavelets (SPIE 2007), San Diego, USA, August 2007.

◆ J. Bobin, Y. Moudden, J.-L. Starck, J. Fadili, N. Aghanim, *SZ and CMB reconstruction using GMCA*, Astronomical Data Analysis (ADA IV), Marseille, France, September 2006.

◆ J. Bobin, Y. Moudden, J.-L. Starck, *Enhanced Source Separation by Morphological Component Analysis*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006), Toulouse, France, May 2006.

◆ J. Bobin, Y. Moudden, J.-L. Starck and M. Elad, *Multichannel Morphological Component Analysis*, Proceedings of SPARS'05, Rennes, France, November 2005.

◆ J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, and D.L. Donoho, *Morphological Component Analysis*, Proceedings of the SPIE conference wavelets, Vol. 5914, San Diego, USA, July 2005.

## Rapport technique

◆ J. Bobin, J.-L. Starck, *Compressed Sensing for Hershel*, Rapport technique - CEA Saclay, Septembre 2006.

## Communications orales

◆ GDR-ISIS, journées problèmes inverses - Paris - Juin 2008.

◆ ADA V - Heraklion, Grèce - Mai 2008.

◆ Communication au CEA/SEDI sur l'analyse de données multivaluées - Mars 2008.

◆ Communication à l'université de Grenade, Espagne - Février 2006.

◆ Communication à l'université de Caen - Avril 2006.

# Sigles et abbréviations

ACI	Analyse en Composantes Indépendantes
ACMP	Analyse en Composantes Mutuellement Parcimonieuses
ACP	Analyse en Composantes Principales
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
dB	décibels
DCT	Discrete Cosine Transform Transformée en Cosinus Discrète
EFICA	Efficient FastICA
<i>e.g.</i>	exempli gratia
GMCA	Generalized Morphological Component Analysis Analyse en Composantes Morphologiques Généralisées
HypGMCA	Generalized Morphological Component Analysis for hyperspectral data Analyse en Composantes Morphologiques Généralisées pour des données hyperspectrales
IDM	Interpolation de Données Manquantes
<i>i.e.</i>	id est
<i>i.i.d.</i>	indépendamment et identiquement distribué
JADE	Joint Approximate Diagonalization of Eigen-matrices
LDCT	Local Discrete Cosine Transform Transformée en Cosinus Discrète Locale
MAP	Maximum A Posteriori
MCA	Morphological Component Analysis Analyse en Composantes Morphologiques
MDL	Minimum Description Length
mMCA	multichannel Morphological Component Analysis Analyse en Composantes Morphologiques multicanale
MOM	Mean Of Max ou moyenne des maxima
MP	Matching Pursuit
MV	Maximum de Vraisemblance
OMP	Orthogonal Matching Pursuit
<i>pdf</i>	probability density function densité de probabilité
<i>resp.</i>	respectivement
RNA	Relative Newton Algorithm
RSB	Rapport Signal sur Bruit
RVB	Rouge Vert Bleu
<i>s.c.</i>	sous la contrainte
SSA	Séparation de Source Aveugle

---

# Notations et définitions

## Notations

Les vecteurs sont notés en minuscule et en gras. La  $i$ -ème entrée du vecteur  $\mathbf{x}$  est notée  $x[i]$ . Sauf indication, un vecteur sera noté sous forme d'une ligne :  $\mathbf{x} = [x[1], \dots, x[t]]$ . Les matrices sont notées en majuscule et en gras. L'élément figurant à la  $i$ -ème ligne et  $j$ -ème colonne de  $\mathbf{X}$  est noté  $X[i, j]$ . La  $i$ -ème ligne sera notée  $\mathbf{x}_i$  et la  $j$ -ème colonne  $\mathbf{x}^j$ .

## Définitions

### Normes

Pour un vecteur  $\mathbf{x}$  de taille  $t$ , on définit la norme  $\ell_p$  avec  $p > 0$  comme suit :

$$\|\mathbf{x}\|_{\ell_p} = \left( \sum_{k=1}^t |x[k]|^p \right)^{1/p}$$

Notons deux cas limites :

- **Cas**  $p \rightarrow 0$  : la (pseudo)-norme  $\ell_0$  d'un vecteur  $\mathbf{x}$  est égale au nombre de ses entrées non nulles.
- **Cas**  $p \rightarrow +\infty$  : la norme  $\ell_\infty$  est définie comme suit :  $\|\mathbf{x}\|_{\ell_\infty} = \max_k |x[k]|$ .

Pour les matrices, certaines normes particulières seront utilisées. L'extension de la norme  $\ell_p$  aux matrices sera la suivante :

$$\|\mathbf{X}\|_p = \left( \sum_{i,j} |X[i, j]|^p \right)^{1/p}$$

Nous utiliserons également la norme de Frobenius sous sa forme pondérée. Soit  $\Sigma$  une matrice symétrique, définie positive :

$$\|\mathbf{X}\|_{F,\Sigma} = \left( \text{Trace}(\mathbf{X}\Sigma^{-1}\mathbf{X}^T) \right)^{1/2}$$


---

## Décomposition parcimonieuse

Considérons le problème d'optimisation suivant :

$$\min_{\alpha} \|\alpha\|_{\ell_p} \text{ s.c. } \mathbf{x} = \alpha \Phi$$

où  $\mathbf{x}$  est un vecteur de taille  $t$ ,  $\alpha$  est un vecteur de taille  $T$  et  $\Phi$  une matrice de taille  $T \times t$ . Sous réserve d'existence, l'ensemble des solutions de ce problème sera noté  $\mathcal{V}_{\ell_p}^{\Phi}(\mathbf{x})$ . Si la solution à ce problème existe et est unique, nous noterons  $\Delta_{\Phi}(\mathbf{x})$  cette solution.

## Miscellanées

L'espérance d'une variable aléatoire  $\mathbf{x}$  est notée  $\mathbb{E}\{\mathbf{x}\}$ .

Le support d'un vecteur  $\mathbf{x}$  est défini comme suit :

$$\Lambda(\mathbf{x}) = \left\{ i \mid |x[i]| > 0 \right\}$$

Pour une matrice  $\mathbf{X}$  de rang colonne plein, on définit le pseudo-inverse (à gauche)  $\mathbf{X}^{\dagger}$  comme suit :

$$\mathbf{X}^{\dagger} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

où  $\mathbf{X}^T$  est la transposée de  $\mathbf{X}$ .

Soit  $\mathbf{X}$  une matrice pseudo-inversible à gauche et  $\tilde{\mathbf{X}}$  une estimée de  $\mathbf{X}$  définie à une permutation et un changement d'échelle près à droite (noté  $\mathbf{P}$ ), il nous sera utile de définir un critère de divergence entre  $\mathbf{X}$  et  $\tilde{\mathbf{X}}$  :

$$C_{\mathbf{X}} = \|\mathbf{I} - \tilde{\mathbf{X}}^{\dagger} \mathbf{P} \mathbf{X}\|_1$$


---

# Introduction

Ce mémoire de thèse a pour objet l'étude de nouvelles méthodologies pour l'analyse de données multivaluées. Ce type particulier de données est formé de l'observation simultanée du même processus physique « sous différents angles ». La prise de vue d'une même scène dans des bandes de fréquences différentes en imagerie multispectrale fournit un exemple typique de telles données multivaluées. L'analyse de ces signaux est également spécifique : la présence de structures (spectres d'émission ou d'absorption dans le cas des données multispectrales) requiert le traitement cohérent des différentes observations.

Dans ce mémoire de thèse, nous proposons le développement d'une nouvelle méthodologie d'analyse de données multivaluées fondée sur l'utilisation de représentations dites parcimonieuses des signaux. A cet effet, le Chapitre 1 introduit les aspects fondamentaux relatifs à l'utilisation de telles représentations, en particulier dans le cadre de la séparation de signaux. La dernière décennie a vu l'essor tant théorique que pratique de telles méthodes d'analyses de signaux ; le Chapitre 2 propose un tour d'horizon des principaux résultats d'analyse parcimonieuse de signaux. Nous y introduirons également une extension au cas du traitement des données multivaluées.

Nous nous concentrerons ensuite sur une modélisation particulière de ces données : le cas du modèle de mélange. Il s'agira alors de modéliser chaque observation comme un mélange linéaire de signaux élémentaires appelés plus généralement « sources ». L'analyse de ces données ainsi modélisées conduit à la résolution du problème de séparation de sources en aveugle. Le Chapitre 3 présente une revue des principales approches proposées pour la résolution de ces problèmes. Nous introduirons également les aspects conceptuels d'une nouvelle méthodologie de séparation de sources en aveugle fondée sur l'utilisation de représentations parcimonieuses : l'Analyse en Composantes Mutuellement Parcimonieuses (ACMP). Nous développons au Chapitre 4, un nouvel algorithme d'analyse de données multivaluées intitulé Analyse en Composantes Morphologiques Généralisées (GMCA) pour la résolution des problèmes de séparation de sources en aveugle.

Au vu de l'expansion des observations hyperspectrales, en particulier en géosciences, nous proposons au Chapitre 5 une extension de l'algorithme GMCA à la séparation de sources aveugles à partir de données hyperspectrales. Nous présenterons des résultats préliminaires

---



d'extraction de composantes à partir des données *Mars Express*.

La mission spatiale Planck de l'Agence Spatiale Européenne (ESA) sera lancée au cours du premier semestre de l'année 2009. Ce satellite fournira des données multispectrales dans le domaine des micro-ondes en particulier pour l'analyse du fond diffus cosmologique (plus célèbre sous le nom de rayonnement fossile). L'analyse de ces données étant cruciale pour la communauté astrophysique, le développement de méthodes de séparation de sources dédiées à ces données est d'importance. Au Chapitre 6, nous présentons une application de l'algorithme GMCA à l'analyse de données réalistes proches de celles que Planck devrait fournir.

Enfin, au Chapitre 7, nous présenterons une extension de l'ACMP et de l'algorithme GMCA à la résolution adaptative de problèmes inverses classiques (débruitage et interpolation de données manquantes).

Au cours de la dernière décennie, une nouvelle théorie de l'échantillonnage a vu le jour : le *compressed sensing* (échantillonnage compressé). Nous proposons en Annexe B un article décrivant la première application de cette théorie à la compression des données de la mission spatiale Herschel de l'ESA.

---

# Chapitre 1

## Prologue

**A** l'heure du développement des technologies numériques, le *vulgum pecus* est submergé de données tout aussi diverses que complexes. Il est bien souvent difficile de traiter des données dont la quantité va en s'accroissant. Le développement de techniques capables d'extraire rapidement l'information essentielle à partir de telles données est ainsi crucial. A cette fin, de récents développements en mathématiques appliquées ont porté sur l'élaboration de concepts et de techniques conduisant à représenter des données complexes de manière économe permettant ainsi l'extraction des structures essentielles de ces données. Tout au long de cette thèse, la question de la distinction et de la séparation de signaux sera étudiée. La séparation de signaux est un problème ancien aux multiples applications : électro-cardiogramme et l'exemple emblématique de la séparation du pouls d'un fœtus et de la mère qui le porte, électro-encéphalogramme pour l'extraction de réponses élémentaires du cortex cérébral, imagerie multispectrale et hyperspectrale tant en astrophysique, géophysique ou pour l'observation de la terre. La liste est loin d'être exhaustive. Nous proposons d'aborder le problème général de la séparation de signaux sous l'angle de la caractérisation de ces derniers via une représentation économe ou parcimonieuse. Ce premier chapitre introduit les concepts très généraux auxquels il sera fait très largement référence par la suite.

---

## 1.1 De la représentation d'un signal

Tout au long de cette thèse, la question de la séparation de signaux est posée. Considérons ainsi les deux images dites "naturelles"<sup>1</sup> présentées en figure 1.1. Toutes deux représentent la même scène d'un point de vue descriptif (bateau, lac, bâtiment, etc ···). Néanmoins, l'oeil du lecteur n'a aucune difficulté à percevoir les différences qui distinguent ces deux illustrations.



FIG. 1.1: Deux images "naturelles".

### 1.1.1 Analogies en neurosciences

Le fonctionnement du système visuel humain est une machinerie d'une efficacité surprenante dont l'étude peut nous apporter des enseignements sur les modélisations possibles en traitement du signal. Au cours de ce paragraphe nous allons faire un tour d'horizon des résultats significatifs que les neurosciences ont pu mettre à jour sur le fonctionnement du système visuel humain.

Au-delà de l'essor récent de cette discipline, les premiers résultats concernant le fonctionnement des premières aires du cortex visuel mammalien sont à mettre à l'actif de pionniers tels que Attneave ([Att54](#)) ou Barlow ([Bar61](#)) au cours de la seconde moitié du 20<sup>ème</sup> siècle. Leurs travaux ont en particulier fait apparaître que la première aire visuelle (nommée V1) est le siège d'un premier traitement de l'information visuelle. Ce traitement primordial s'apparente à un codage de l'information visuelle. Dès lors, de nombreuses études ont été menées pour caractériser plus finement les spécificités de ce codage. Citons en particulier les travaux d'intérêt sur le sujet suivant : ([SO01](#); [OF06](#); [BS97](#); [Fie99](#)). Il résulte de ces différents travaux des observations des plus fascinantes :

- **Sensibilité à des stimuli bien précis** : il apparaît que le codage effectué au niveau de l'aire V1 est tel que certaines *formes* ou *structures* des images sont privilégiées.

---

<sup>1</sup>Au sens où elles peuvent faire partie de notre environnement visuel.

En particulier (voir (OF06; Fie99)), il semble que le traitement effectué au niveau de V1 soit équivalent à un codage de la scène visuelle à partir de *stimuli élémentaires* proches d'ondelettes de Gabor. En d'autres termes, ce codage visuel est efficace sur les structures orientées de type *contour*. Cette propriété inattendue a été fort justement nommée "ubiquity of edges" par (DF01).

- **Réponse parcimonieuse** : des études identiques (en particulier (OF04)) ont montré que cette *efficacité* du codage de stimuli particuliers (contours) se traduisait par une réponse *parcimonieuse* de l'activité neuronale. Plus précisément, en présence de stimuli de type "contour", la plupart des neurones sont inactifs et peu ont une activité significative. Cette propriété a été initiatrice de la théorie du *codage parcimonieux* (sparse coding). Comme le suggèrent Olshausen, « *Sparse coding provides an efficient means of representing data found in the natural world. Moreover, it provides a means of efficiently forming associations and storing memory, and it achieves all of this using relatively small amounts of energy* ».
- **Décohérence des réponses neuronales** : Olshausen (SO01) suggère également que les réponses neuronales doivent être décohérentes (au sens où elles doivent être différenciées) : « *the code cannot be efficient if the effort of encoding any particular piece of information is duplicated in more than one neuron* ». Cette hypothèse de décohérence des réponses neuronales est donc motivée par le même souci de "compresser" l'information du signal visuel initial ; un codage efficace ne devant pas coder plusieurs fois des structures identiques <sup>2</sup>.

Ce bref voyage au pays des neurosciences nous apportent plusieurs enseignements. Le plus important réside sûrement dans cette notion de *code adapté* aux images à traiter. De plus, le code utilisé au niveau de l'aire V1 est loin d'être quelconque : il s'agit d'un code élaboré pour "capturer" les structures particulières des scènes visuelles qui nous entourent : les contours. La notion de parcimonie est donc intimement liée à la structure même des signaux et à la manière dont celles-ci sont codées.

## 1.1.2 Structures, parcimonie et représentations

### Parcimonie : définitions

Par analogie au *codage parcimonie* introduit dans le domaine des neurosciences, nous allons définir un peu plus formellement le modèle de représentation parcimonieuse que l'on retrouvera tout au long du présent rapport.

---

<sup>2</sup>(Fie99) avance l'hypothèse que les forts liens entre les scènes naturelles et le codage parcimonieux (forte prédominance des contours) peut s'interpréter à la lumière de l'évolution naturelle. En d'autres termes, le codage opéré au niveau du système visuel mammalien serait l'aboutissement de l'adaptation de ce même code à son environnement.

D'ores et déjà, le décor à planter est le suivant : un signal  $\mathbf{x}$  est un vecteur de dimension  $t$  dont les entrées sont réelles :

$$\mathbf{x} = [x_1, \dots, x_t] \quad (1.1)$$

On dispose d'un code  $\mathcal{D}$  dont les éléments sont des vecteurs de  $\mathbb{R}^t$ . L'ensemble des éléments de  $\mathcal{D}$  génère  $\mathbb{R}^t$  de sorte que tout signal  $\mathbf{x}$  peut être synthétisé à partir du code  $\mathcal{D}$ . Plus précisément, nous noterons  $\phi_i$  le  $i$ -ème élément du code  $\mathcal{D}$ ; ce dernier contenant  $T > t$  mots. Il existe un ensemble de scalaires  $\{\alpha_i\}_{i=1, \dots, T}$  tel que  $\mathbf{x}$  s'écrit comme suit :

$$\mathbf{x} = \sum_{i=1}^T \alpha_i \phi_i \quad (1.2)$$

Puisque nous manipulons des vecteurs, ils sera par la suite plus judicieux de définir  $\Phi^3$  comme la matrice de taille  $T \times t$  dont les lignes sont les éléments de  $\mathcal{D}$ . Puisque les éléments de  $\Phi$  sont les "mots" du code  $\mathcal{D}$ , il est d'usage de parler de *dictionnaire* pour  $\Phi$ . En définissant  $\alpha = [\alpha_1, \dots, \alpha_T]$ , on obtient la formulation matricielle suivante :

$$\mathbf{x} = \alpha \Phi \quad (1.3)$$

Par analogie avec le codage parcimonieux, le vecteur  $\alpha$  peut s'interpréter comme l'ensemble des activités neuronales liées au stimulus  $\mathbf{x}$ . Les éléments de  $\mathcal{D}$  (appelés également *atomes* (CDS98)) représentent ainsi les stimuli élémentaires qui, du point de vue de  $\mathbf{x}$ , sont ses structures les plus significatives. Un codage efficace pour  $\mathbf{x}$  devra également conduire au caractère parcimonieux des scalaires  $\{\alpha_i\}_{i=1, \dots, T}$ .

D'un point de vue général, il est commun de rencontrer des définitions de la parcimonie de  $\alpha$  peu explicites : *dire que  $\alpha$  est parcimonieux signifie que la plupart de ses entrées sont proches de zéro et peuvent prendre des valeurs significatives*. Plus formellement, nous utiliserons deux types de modélisations pour définir le caractère parcimonieux d'un vecteur  $\alpha$ .

### Modèles déterministes :

Il s'agit ici de caractériser le comportement des entrées de  $\alpha$ . Le vecteur  $\alpha$  est dit *K-sparse* si seules  $K$  de ses entrées sont non nulles. Dans ce cas, on définit le support (DH01) de  $\alpha$  comme l'ensemble des indices des entrées non-nulles de  $\alpha$  :

$$\Lambda(\alpha) = \{i | \alpha[i] \neq 0\} \quad (1.4)$$

Un tel vecteur est également caractérisé par sa (pseudo) norme  $\ell_0$  puisque :

$$\|\alpha\|_{\ell_0} = \text{Card}(\text{Support}(\alpha)) = K \quad (1.5)$$

Ce modèle présente un intérêt essentiellement théorique comme nous le verrons au Chapitre 2.

---

<sup>3</sup>Par la suite, nous confondrons abusivement  $\mathcal{D}$  et  $\Phi$  bien que constituant deux objets bien différents.

En pratique des modèles plus réalistes consistent à borner le taux de décroissance des entrées de  $\alpha$  en valeur absolue. Supposons que les entrées de  $\alpha$  sont ordonnées de sorte que  $|\alpha[1]| \geq |\alpha[2]| \cdots \geq |\alpha[T]|$ ,  $C$  une constante strictement positive et  $p$  appartenant à l'intervalle  $]0, 1]$ , une définition de parcimonie pour  $\alpha$  est la suivante :

$$|\alpha[i]| \leq C i^{-1/p} \text{ pour } i = 1, \dots, p \quad (1.6)$$

### Un modèle probabiliste :

Nous verrons par la suite qu'il peut être nécessaire de définir une distribution pour  $\alpha$ . Par analogie avec les neurosciences, une telle distribution se caractérise comme suit :

- **Indépendance** : les entrées du vecteurs  $\alpha$  sont indépendamment et identiquement distribuées (*i.i.d.*) suivant une distribution de densité  $f_\alpha$ . Ces entrées étant indépendantes dans leur ensemble, elles vérifieront :  $\forall i \neq j \in \{1, \dots, t\}; \quad f_\alpha(x[i], x[j]) = f_\alpha(x[i]) f_\alpha(x[j])$ .
- **Parcimonie** : la densité de probabilité  $f_\alpha$  sera choisie de sorte que la probabilité pour  $\alpha$  d'avoir des entrées de fortes amplitudes soit faible. Dans cette thèse, nous choisirons le plus souvent un modèle de signaux générés suivant une densité de probabilité leptokurtique (de kurtosis positif) telles que les gaussiennes généralisées<sup>4</sup> avec  $\theta < 1$  :

$$\forall i; \quad f_\alpha(\alpha[i]) \propto \exp(-\kappa |\alpha[i]|^\theta) \quad (1.7)$$

### Structures et représentations

Le plus souvent, les signaux à traiter possèdent des structures spécifiques (caractère périodique ou transitoire pour des signaux unidimensionnels, textures ou contours pour des images).

Le caractère parcimonieux du codage décrit précédemment est lié aux structures du signal  $\mathbf{x}$  et du choix du dictionnaire  $\Phi$ . Les développements récents en analyse harmonique ont permis la construction d'un très large éventail de "représentations" (bases, trames ajustées) adaptées à des structures particulières ; le caractère adapté signifiant leur capacité à apporter une représentation parcimonieuse de signaux spécifiques. En ce sens, le choix d'une représentation (*i.e.* le "code"  $\mathcal{D}$ ) s'effectue en fonction d'une connaissance *a priori* du contenu des données à traiter. Ainsi, à une classe de structures correspond un ensemble de *formes d'ondes* adaptées (*i.e.* les atomes du dictionnaire  $\Phi$ ). Parmi la multiplicité des

---

<sup>4</sup>Notons que les deux catégories de modèles ne sont pas si différentes. En effet, dans le cas du modèle probabiliste que nous venons d'introduire, la propriété suivante est vérifiée pour  $\theta = p$  :

$$\forall \epsilon, \exists C_1 \in \mathbb{R}^+ \text{ tel que } \mathbb{E} \left\{ \|\alpha\|_{\ell_p}^p \right\} \leq C_1 \text{ avec une probabilité supérieure à } 1 - \epsilon$$

Revenons à la définition déterministe précédente caractérisant le degré de décroissance  $p$  des entrées de  $\alpha$ . Ainsi, la norme  $\ell_p$  de  $\alpha$  est également bornée avec cette définition puisqu'en définissant  $C_2 = C \sum_i i^{-1} : \|\alpha\|_{\ell_p}^p \leq C_2$

---

formes d'ondes particulières<sup>5</sup>, nous utiliserons dans le cas du traitement d'image :

- **Transformation en cosinus (local)** : bien adaptée aux images dont le contenu est globalement (ou localement) oscillant.
- **Transformation en ondelettes** : adapté (*i.e.* optimal dans le cas d'une approximation m-termes - voir (Mal98)) aux singularités ponctuelles.
- **Trame ajustée de Curvelets** : adapté aux images  $C^2$  par morceau contenant des contours  $C^2$ .

Le plus souvent, ces ensembles d'atomes sont associés à des opérateurs implicites de transformation rapide qui en rendent l'utilisation facilitée.

## 1.2 Diversité morphologique

Introduite dans le contexte très particulier des problématiques de séparation contour/texture dans une image (ESDQ05), la notion de diversité morphologique est un outil puissant de caractérisation du contenu d'un signal. La diversité morphologique sera le concept fondamental des développements des prochains chapitres.

### 1.2.1 Analyse en composantes morphologiques

Dans le cas d'une image complexe telle que celle présentée en figure 1.2, aucune base n'est adaptée à la fois à la représentation parcimonieuse d'un contenu globalement oscillant et continu par morceau. Une idée simple consiste alors à décomposer cette image au contenu complexe en une combinaison linéaire de *composantes morphologiques* contenant chacune des structures élémentaires différentes (voir figure 1.3) :

$$\mathbf{x} = \varphi_1 + \varphi_2 \quad (1.8)$$

où  $\varphi_1$  (respectivement  $\varphi_2$ ) est supposée admettre une représentation parcimonieuse dans la base  $\Phi_1$  (respectivement  $\Phi_2$ ). Le concept de diversité morphologique intervient : les deux composantes  $\varphi_1$  et  $\varphi_2$  présentent des structures ou de manière équivalente des **morphologies** différentes. Cette différence ou **diversité** de morphologies se caractérise formellement comme suit :

$$\|\varphi_i \Phi_i^T\|_{\ell_p} < \|\varphi_i \Phi_{j \neq i}^T\|_{\ell_p} \text{ où } 0 < p \leq 1 \quad (1.9)$$

En résumé, deux composantes aux morphologies différentes sont certes parcimonieuses dans une représentation qui leur est adaptée mais "moins parcimonieuse" dans des celles qui ne le sont pas. En s'appuyant sur ce principe de diversité morphologique, il est alors

---

<sup>5</sup> Consulter tout particulièrement : <http://lcd.siva.free.fr/siva-wits-where-is-the-starlet.html>.

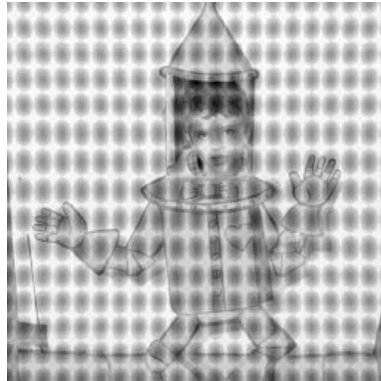


FIG. 1.2: Une image synthétique complexe.

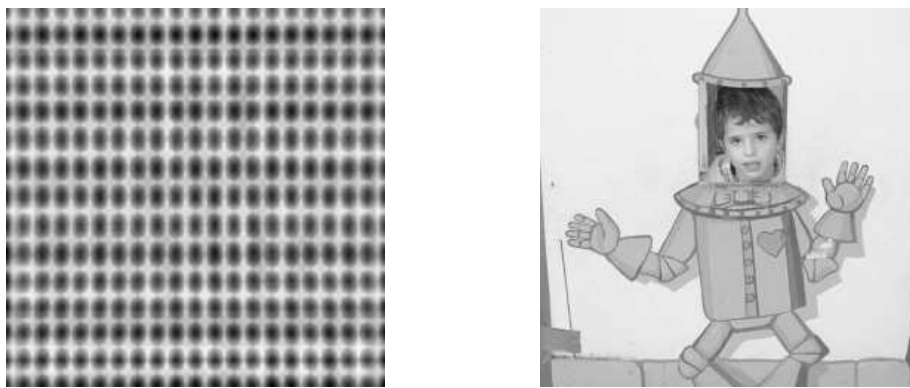


FIG. 1.3: Composantes morphologiques.

possible de séparer les deux composantes  $\varphi_1$  et  $\varphi_2$  à partir de leurs seules différences de structures/morphologies.

D'un point de vue général, nous verrons au Chapitre 2 qu'une telle modélisation peut être équivalente à une représentation du signal  $\mathbf{x}$  dans le dictionnaire **redondant** ( $T > t$ ) union de  $\Phi_1$  et  $\Phi_2$ . L'une des clefs des méthodes que nous développerons par la suite est le fait que la redondance du dictionnaire de représentation permet d'obtenir une représentation plus parcimonieuse.

### 1.2.2 Le bon grain et l'ivraie

L'un des atouts des représentations dites parcimonieuses est la caractérisation des structures inhérentes à chaque signal individuel. Ainsi, l'étape suivante consiste à s'appuyer sur de telles représentations pour les **distinguer** voire les séparer.

A l'origine, la notion de diversité morphologique a permis la distinction de signaux possédant des morphologies distinctes. Par la suite, il sera également nécessaire de pouvoir



distinguer des signaux ayant des morphologies proches. Considérons deux signaux  $\mathbf{x}_1$  et  $\mathbf{x}_2$  ; nous introduisons deux catégories de diversité morphologique permettant de les différencier :

- **Diversité morphologique au sens strict** : il s'agit de celle que nous venons d'introduire. Dans ce cas  $\mathbf{x}_1$  et  $\mathbf{x}_2$  présentent des structures ou morphologies différentes.
- **Diversité morphologique au sens large** : dans ce cas  $\mathbf{x}_1$  et  $\mathbf{x}_2$  présentent des morphologies proches (telles que les images de la figure 1.1). *Comment est-il alors possible de les distinguer ?* Ces deux images contiennent toutes deux des contours et des parties texturées (localement oscillantes). Si les structures de chacune des images sont identiques de part leur type (contour, texture, ...etc.), elles demeurent néanmoins différentes (structures qui diffèrent de part leurs positions, amplitudes, échelles ... etc.). Supposons que les deux signaux  $\mathbf{x}_1$  et  $\mathbf{x}_2$  admettent une représentation parcimonieuse dans une représentation  $\Phi$ . Si  $\mathbf{x}_1$  et  $\mathbf{x}_2$  se distinguent par leurs structures les plus fortes, alors, d'un point de vue empirique, leurs supports dans  $\Phi$  doivent être différenciés. En particulier, les entrées de  $\alpha_1$  et  $\alpha_2$  les plus significatives doivent être différentes. A cette fin, nous introduisons la notion de  $\delta$ -support d'un vecteur  $\alpha$  :

$$\Lambda_\delta(\alpha) = \{k; |\alpha[k]| > \delta \|\alpha\|_\infty\} \quad (1.10)$$

où  $\|\alpha\|_\infty = \max_k |\alpha[k]|$ . En particulier, il est possible de comparer deux signaux en terme de  $\delta$ -supports disjoints. En effet, deux signaux  $\mathbf{x}_1 = \alpha_1 \Phi$  et  $\mathbf{x}_2 = \alpha_2 \Phi$  seront dits à support  $\delta$ -disjoints dans  $\Phi$  s'il existe  $0 \leq \delta < 1$  tel que :

$$\Lambda_\delta(\alpha_1) \cap \Lambda_\delta(\alpha_2) = \emptyset \quad (1.11)$$

Nous noterons en particulier la plus petite valeur de  $\delta$  telle que deux signaux sont à support  $\delta$ -disjoints :

$$\delta^* = \min\{\delta; \Lambda_\delta(\alpha_1) \cap \Lambda_\delta(\alpha_2) = \emptyset\} \quad (1.12)$$

Notons le cas particulier  $\delta^* = 0$  pour lequel les supports des deux signaux  $\mathbf{x}_1$  et  $\mathbf{x}_2$  dans  $\Phi$  sont strictement disjoints. La valeur de  $\delta^*$  mesure, en quelque sorte, un degré de diversité entre  $\mathbf{x}_1$  et  $\mathbf{x}_2$ . Le cas où  $\delta^* = 1$  caractérise celui où les deux signaux n'admettent pas de supports  $\delta$ -disjoints dans  $\Phi$ .

Cette définition de la diversité morphologique sera le fondement de la méthode de séparation de signaux introduite au Chapitre 3.

## Chapitre 2

# Dictionnaires multivalués et décomposition parcimonieuse

**L**E Chapitre 1 a introduit les concepts élémentaires sur lesquels seront fondés les développements des Chapitres 3 et 4 pour la séparation de signaux. Notre approche met plus précisément en exergue la recherche de représentations dites *parcimonieuses* capables de caractériser la morphologie des signaux à séparer. Afin de représenter de manière creuse ou économe (synonymes de parcimonieuse ici) de très larges classes de signaux, l'utilisation de dictionnaires redondants (*i.e.* le code  $\mathcal{D}$  introduit au Chapitre 1) a apporté une avancée considérable au cours de la dernière décennie. La décomposition de signaux dans des dictionnaires redondants a soulevé de nombreux problèmes tant théoriques que pratiques. Ce chapitre présente un rapide tour d'horizon des problèmes de décomposition parcimonieuse de signaux qui ont agité la communauté scientifique. La seconde partie est consacrée à l'introduction d'un outil d'analyse (plus précisément de décomposition parcimonieuse) de données multivaluées. Ainsi, l'Analyse en Composantes Morphologiques Multivaluées (mMCA) est introduit. L'algorithme qui lui est associé constitue le fondement des algorithmes de séparation que nous développons par la suite.

---

## 2.1 Décomposition d'un signal dans un dictionnaire redondant

### 2.1.1 De la décomposition parcimonieuse de signaux

#### La résolution d'un problème sous-déterminé

Pour l'analyse de signaux, les représentations redondantes apportent une flexibilité dont la matérialisation directe est la possibilité d'obtenir des représentations très creuses ou parcimonieuses pour de larges classes de signaux. Rappelons qu'au sein du paradigme des signaux parcimonieux, un signal unidimensionnel  $\mathbf{x}$ <sup>1</sup> est modélisé comme la combinaison linéaire de signaux élémentaires ou atomes appartenant à un dictionnaire redondant d'atomes  $\Phi = [\phi_1, \dots, \phi_T]^T$  comme suit :

$$\mathbf{x} = \sum_{i=1}^T \alpha_i \phi_i \quad (2.1)$$

Supposons que le signal  $\mathbf{x}$  soit  $K$ -sparse dans le dictionnaire  $\Phi$  ; *i.e* exactement  $K < T$  entrées du vecteur des coefficients  $\alpha = [\alpha_1, \dots, \alpha_T]$  sont non nulles. Evidemment, nous parlerons de représentation creuse lorsque le nombre de ces coefficients non nuls sera faible :  $K \ll T$ . Soit  $\Lambda$  l'ensemble des index des entrées non nulles de  $\alpha$  (*i.e* le support de  $\alpha$ ),  $\mathbf{x}$  est synthétisé parfaitement à partir de ces  $K$  atomes :

$$\mathbf{x} = \sum_{i \in \Lambda} \alpha[i] \phi_i \text{ avec } \text{Card}(\Lambda) = K \quad (2.2)$$

Du fait de la redondance de  $\Phi$ , l'équation régissant la synthèse du vecteur  $\mathbf{x}$  à partir d'éléments de  $\Phi$  :

$$\mathbf{x} = \alpha \Phi \quad (2.3)$$

est un système sous-déterminé (rappelons que  $t < T$ ) ; de fait, en supposant  $\Phi$  de rang colonne plein, il existe une **infinité** de solutions  $\alpha$  vérifiant la relation (2.3).

Néanmoins, la solution vérifiant (2.2) est particulière puisqu'elle est creuse. Les auteurs de (DH01) ont fort logiquement proposé la recherche de cette solution particulière par la résolution du problème d'optimisation suivant :

$$\min_{\alpha} \|\alpha\|_{\ell_0} \text{ s.c. } \mathbf{x} = \alpha \Phi \quad (2.4)$$

consistant en la recherche de la solution qui contient le moins d'entrées actives ou non nulles.

#### Propriétés en terme de cohérence mutuelle

De nombreuses études ont été consacrées à la caractérisation de ce problème et plus particulièrement aux propriétés vérifiées par sa ou ses solutions. Nous supposons par la suite que chaque atome de  $\Phi$  (chaque ligne) est de norme  $\ell_2$  unité :  $\forall i; \|\phi_i\|_{\ell_2} = 1$ .

---

<sup>1</sup>monovalué par opposition aux données multivaluées que nous introduirons par la suite.

L'un des paramètres privilégiés de caractérisation du problème de l'équation (2.4) est la *cohérence mutuelle*  $\mu_{\Phi}$  du dictionnaire  $\Phi$ . Introduite en premier lieu par Donoho (DH01) (mais également présent dans (MZ93)), la cohérence mutuelle de  $\Phi$  est définie comme suit :

$$\mu_{\Phi} = \max_{i \neq j} |\phi_i \phi_j^T| \quad (2.5)$$

La cohérence mutuelle d'une matrice  $\Phi$  de taille  $T \times t$  vérifie les inégalités suivantes (SH04) :

$$\sqrt{\frac{T-t}{t(T-1)}} \leq \mu_{\Phi} \leq 1 \quad (2.6)$$

avec deux cas particuliers :

- **Dictionnaire incohérent** : il s'agit du cas où  $\mu_{\Phi}$  est faible ; le cas limite étant celui des matrices Grassmanniennes (SH04) pour lesquelles la borne inférieure est atteinte. Dans le cas où  $\Phi = [\Phi_1^T, \Phi_2^T]^T$  est formé de deux bases orthonormales, la borne inférieure est en particulier atteinte pour les trames dites *équiangulaire* pour lesquelles le produit scalaire entre deux atomes quelconques est soit 0 soit  $\mu_{\Phi}$ . Un exemple simple est celui formé par la base de sinus et la base canonique (liée à la matrice identité). D'un point de vue empirique, les matrices redondantes incohérentes contiennent des atomes relativement différents (au sens de leurs produits scalaires).
- **Dictionnaire cohérent** : le cas où  $\mu_{\Phi}$  est élevé correspond à celui de matrices dites cohérentes pour lesquelles certains atomes sont peu dissemblables ; le cas limite  $\mu_{\Phi} = 1$  est celui où deux atomes au moins sont strictement colinéaires.

En somme, la cohérence mutuelle caractérise un degré de "ressemblance" des atomes d'un dictionnaire  $\Phi$ . Ce paramètre est d'autant plus intéressant qu'il peut être déterminé aisément.

En particulier, Donoho *et al.* (DH01) ont montré que la condition suivante :

$$\|\alpha\|_{\ell_0} < \frac{1}{2} \left( 1 + \frac{1}{\mu_{\Phi}} \right) \quad (2.7)$$

est suffisante pour garantir l'unicité de la solution du problème de l'équation (2.4). La borne induite par cette condition d'unicité peut être raffinée dans certains cas ; en particulier celui où  $\Phi$  est formé d'une union de bases orthonormales (EB02; GN03; FN03a; BDE07).

Notons également que des résultats comparables ont été obtenus à partir d'une caractérisation différente de la matrice redondante  $\Phi$  : la propriété d'isométrie restreinte (RIP) introduite par Candès (CT06; Can08).

### 2.1.2 Convexification

Le problème de l'équation (2.4) s'avère, par sa nature combinatoire, NP-difficile (voir (Nat95)). Il ne peut donc être question de résoudre ce problème directement. Le caractère

---

combinatoire du problème de l'équation (2.4) provient en particulier de la *non-convexité* de la norme  $\ell_0$ . L'astuce (CDS98) consiste à approximer la norme  $\ell_0$  par une norme  $\ell_p$  convexe. Celle-ci devant être apte à mesurer la parcimonie d'un élément de  $\mathbb{R}^t$ , le candidat idéal n'est alors autre que la norme  $\ell_1$ . Donoho *et al.* (CDS98) ont proposé la substitution du problème de l'équation (2.4) par le suivant :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s.c. } \mathbf{x} = \alpha\Phi \quad (2.8)$$

Cette approche est plus connue sous le nom de *Basis Pursuit* (ou "poursuite de base"). Le problème (2.8) n'étant qu'une approximation du problème de décomposition parcimonieuse initial, il est important de poser la question de l'*équivalence* de ces deux problèmes. De façon tout à fait intéressante, il a été montré (voir (DE03; GN03)) que la condition suivante :

$$\|\alpha\|_{\ell_0} < \frac{1}{2} \left( 1 + \frac{1}{\mu_{\Phi}} \right) \quad (2.9)$$

est suffisante pour garantir l'équivalence entre les problèmes (2.8) et (2.4). Comme pour le cas de l'unicité de la solution du problème (2.4) cette borne est peu réaliste : BP est ainsi capable de résoudre des problèmes de décomposition en norme  $\ell_0$  au-delà de la borne (voir l'étude d'Elad *et al.* (EZ04)). A titre indicatif, notons que Tropp (Tro04; Tro06b) a raffiné ces résultats par une généralisation de la cohérence mutuelle d'un dictionnaire  $\Phi$ .

### 2.1.3 Solution approchée

La contrainte d'égalité  $\mathbf{x} = \alpha\Phi$  intervenant dans les problèmes (2.4) et (2.8) est souvent trop forte et de fait relaxée. Cette relaxation s'effectue en ne recherchant qu'une solution *approchée* à ces problèmes. Elle consiste à remplacer la contrainte d'égalité par une mesure de qualité d'approximation de  $\mathbf{x}$  par la "solution" obtenue. Comme proposé dans (CDS98), une manière classique d'une distance entre les données à décomposer  $\mathbf{x}$  et la solution approchée  $\alpha\Phi$  s'effectue par une mesure quadratique. La solution approchée s'obtient alors à une erreur de reconstruction  $\epsilon$  près. Le problème (2.4) devient alors le problème approché suivant :

$$\min_{\alpha} \|\alpha\|_{\ell_0} \text{ s.c. } \|\mathbf{x} - \alpha\Phi\|_{\ell_2} \leq \epsilon \quad (2.10)$$

Le problème formulé en norme  $\ell_1$  peut être écrit sous forme approchée de façon identique comme suit :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s.c. } \|\mathbf{x} - \alpha\Phi\|_{\ell_2} \leq \epsilon \quad (2.11)$$

Ce dernier est plus connu sous le nom de *Basis Pursuit Denoising* (CDS98).

Comme souligné dans (BDE07), la résolution du problème approché permet de définir une quasi-solution même lorsque la solution du problème (2.8) n'existe pas. Le problème

---

approché précédent peut être écrit (par une reparamétrisation univoque entre  $\epsilon$  et  $\gamma$ ) sous la forme d'un lagrangien augmenté comme suit :

$$\min_{\alpha} \gamma \|\alpha\|_{\ell_1} + \frac{1}{2} \|\mathbf{x} - \alpha \Phi\|_{\ell_2}^2 \quad (2.12)$$

Rappelons que  $\mathbf{x}$  est  $K$ -sparse ; Tropp (Tro06b) montre que si la condition d'unicité et d'équivalence (2.7) est vérifiée, le support de la solution  $\alpha_\gamma$  de (2.12) est inclus dans celui de la solution exacte au problème (2.4).

Le problème de l'équation (2.12) présente l'intérêt pratique de pouvoir introduire un terme additif de bruit borné (en norme  $\ell_2$ ) :

$$\mathbf{x}_n = \mathbf{x} + \mathbf{n} \quad (2.13)$$

avec  $\|\mathbf{n}\|_{\ell_2} < \epsilon$ . Dans ce cas, les développements de Tropp (Tro06b) sont valides, fournissant en particulier une borne sur l'erreur d'approximation de  $\alpha$  par la solution  $\alpha_\gamma$  du problème de l'équation (2.11). Les problèmes (2.11) et (2.12) sont donc stables sous hypothèse de bruit borné. De tels résultats de stabilité ont également été prouvés par Fuchs (Fuc06) et Donoho (DET06).

Les développements de cette dernière décennie ont permis de dresser une description assez détaillée du problème classique en norme  $\ell_0$  et en particulier du comportement de sa version convexe en norme  $\ell_1$ . Dès lors, la difficulté principale réside dans la recherche de la solution du problème de l'équation (2.8).

### 2.1.4 Algorithmes de décomposition dans des dictionnaires redondants

Au cours de ce paragraphe, nous proposons un rapide tour d'horizon des algorithmes aptes à recouvrer la solution parcimonieuse tant convoitée. Nous avons déjà évoqué que le problème de poursuite de base (2.8) (voir (CDS98)) peut être reformulé sous la forme d'un problème de programmation linéaire. Malgré le perfectionnement récent des algorithmes de programmation linéaire (KKB07; CWB08; Can07), leur coup calculatoire est le plus souvent élevé. En parallèle, il a été prouvé que d'autres classes d'algorithmes (algorithmes "gloutons", homotopie ou plus récemment par seuillage itéré) convergent (sous certaines conditions de parcimonie (Tro04; DT06)) vers la solution unique des problèmes (2.4) et (2.8).

#### Algorithmes "gloutons"

Les algorithmes dits "gloutons" (*greedy* en anglais) ont été originellement construits pour l'obtention d'approximations  $m$ -termes<sup>2</sup>. Dans cette famille d'algorithmes, le plus connu est sans nul doute (Orthogonal) *Matching Pursuit* (MZ93; PRK93). Ce type d'algorithme itératif se caractérise par la sélection séquentielle d'un unique atome entrant dans

---

<sup>2</sup>Approximation de  $\mathbf{x}$  à partir des  $m$  entrées les plus significatives de  $\alpha$

l'approximation de  $\mathbf{x}$ .

A titre indicatif, un proche cousin du problème de Basis Pursuit a été développé dans le domaine des statistiques sous le nom de LASSO (HTF01; Tib96). Une méthode empirique nommée *Least Angle Regression* (LARS) a ensuite été proposée pour résoudre le problème du LASSO (EHJT04). Parallèlement, Osborne *et al.* (OPT00), en étudiant le problème initial du LASSO et plus particulièrement son dual convexe, ont proposé la recherche de sa solution par une méthode de continuation de l'homotopie (fondée sur la géométrie particulière du problème BP (DET06; Plu07; MCW05)).

Fort heureusement, une étude de Donoho (DT06) a mis au jour la similitude (voire l'équivalence) entre ces différents algorithmes pour la résolution de problèmes de décomposition parcimonieuse dans un dictionnaire redondant. Cette même étude a également prouvé que ces différents algorithmes dits de régression "stepwise" (sélectionnant un seul atome au maximum, à chaque pas ou itération) résolvent le problème de décomposition de l'équation (2.4) sous l'hypothèse d'unicité bien connue de l'équation (2.7).

Néanmoins, les algorithmes gloutons ou plus généralement "stepwise" ne sélectionnant qu'un et un seul atome par itération sont coûteux en temps de calcul. Récemment, une approche "stagewise" de OMP pour laquelle plusieurs atomes sont sélectionnés par itération a été proposé par Donoho *et al.* (DTDS06). Nous y reviendrons au cours de la Section 2.2.

### Algorithmes de seuillage itératif : des alternatives plus rapides

Récemment, des approches alternatives fondées sur des méthodes de seuillage itératif ont été proposées pour la résolution du problème *Basis Pursuit Denoising* de l'équation (2.12). D'un point de vue général, l'utilisation du seuillage itératif en traitement du signal sous contrainte de parcimonie (en particulier en restauration d'images) émerge depuis une décennie (voir en particulier (FN03b; DDM04; CW05; CCPW07)).

Considérons le cas simple où le dictionnaire  $\Phi$  est une base orthonormale :  $\Phi\Phi^T = I$ . Dans ce cas, le problème de Basis Pursuit relaxé sous forme de Lagrangien de l'équation (2.12) peut s'écrire comme suit :

$$\min_{\alpha} \gamma \|\alpha\|_{\ell_1} + \frac{1}{2} \|\mathbf{x}\Phi^T - \alpha\|_{\ell_2}^2 \quad (2.14)$$

En notant  $\alpha_{\mathbf{x}} = \mathbf{x}\Phi^T$  comme la représentation des données  $\mathbf{x}$  dans la base  $\Phi$ , le problème précédent devient :

$$\min_{\alpha} \gamma \|\alpha\|_{\ell_1} + \frac{1}{2} \|\alpha_{\mathbf{x}} - \alpha\|_{\ell_2}^2 \quad (2.15)$$

Ce dernier problème admet une solution exacte bien connue sous l'appellation de seuillage doux de seuil  $\gamma$  :

$$\alpha = S_{\gamma}(\alpha_{\mathbf{x}}) \quad (2.16)$$

où l'opérateur de seuillage  $S_\gamma$  est défini comme suit :

$$\forall i = 1, \dots, t; \quad S_\gamma(\alpha_{\mathbf{x}})[i] = \begin{cases} 0, & \text{si } \alpha_{\mathbf{x}}[i] < \gamma \\ \alpha_{\mathbf{x}}[i] - \gamma \text{ Signe}(\alpha_{\mathbf{x}}[i]), & \text{sinon} \end{cases} \quad (2.17)$$

où  $S_\gamma(\alpha_{\mathbf{x}})[i]$  (resp.  $\alpha_{\mathbf{x}}[i]$ ) définit le  $i$ -ème élément du vecteur  $S_\gamma(\alpha_{\mathbf{x}})$  (resp.  $\alpha_{\mathbf{x}}$ ). La solution au problème de l'équation (2.14), dans le domaine direct (*i.e.* celui des échantillons) s'obtient donc comme suit :

$$\mathbf{x}^\star = S_\lambda(\mathbf{x}\Phi^T)\Phi \quad (2.18)$$

Cette solution s'obtient donc par la simple séquence : i) Transformation directe, ii) Seuillage puis iii) Transformation inverse ... sous la condition d'orthogonalité de  $\Phi$ .

Dans le cas où  $\Phi$  représente un dictionnaire redondant d'atomes ou signaux élémentaires, la solution n'est plus obtenue par l'application d'un opérateur de seuillage comme dans l'équation (2.18). Néanmoins, l'excellente étude menée par Elad dans (Ela06) montre que, dans le cas où  $\Phi$  est redondant, l'opération de seuillage définie dans l'équation (2.18) constitue un premier pas vers la solution du problème de l'équation (2.14). De plus, Elad a illustré empiriquement qu'un processus de seuillage itératif permet de se rapprocher de la solution en question. Dès lors, ces travaux motivent en grande partie l'utilisation de processus de seuillage itératif pour résoudre le problème de *Basis Pursuit Denoising*. Néanmoins, l'explosion de l'utilisation des algorithmes fondées sur le seuillage itératif apparaît relativement tard dans le cas où le dictionnaire  $\Phi$  est redondant. Ainsi, les premiers algorithmes de seuillages itératifs dédiés à la résolution du problème de l'équation (2.14) n'apparaissent qu'en 2007 par des approches assez diverses (voir (FNW07; FR07; MZ07) pour n'en citer que quelques uns). Notons également que certaines de ces approches tendent à promouvoir le développement d'algorithmes de seuillage itératif dur dans l'objectif d'approcher la solution (sous hypothèse d'unicité) de la version en norme  $\ell_0$  du problème de l'équation (2.14) (voir (BL07; BD07)). Néanmoins, l'extension au cas  $\ell_0$  du problème de l'équation (2.14) n'étant pas convexe, les approches par seuillage itératifs ne peuvent viser qu'une convergence vers un minimum local. Quel que soit l'approche, la forme générale de l'équation de mis à jour à l'itération ( $h$ ) de ces algorithmes itératifs est la suivante :

$$\alpha^{(h+1)} = S_\lambda(\alpha^{(h)} + \mu(\mathbf{x} - \alpha^{(h)}\Phi)\Phi^T) \quad (2.19)$$

où  $\mu$  est un scalaire de sorte que le rayon spectral de  $\mu\Phi\Phi^T$  est inférieur à 1.

## 2.2 L'analyse en composantes morphologiques multicanale

### 2.2.1 Analyse en composantes morphologiques monocanale

#### Principe et modèle

Dans les exemples du Chapitre 1, l'utilisation de dictionnaires de signaux élémentaires redondants est fondée sur le souhait de "mieux" représenter (au sens parcimonieusement)



des signaux aux morphologies diverses. En dehors du cadre de la décomposition de signaux dans un dictionnaire redondant, Starck *et al.* (SED05; ESDQ05) ont introduit en 2004 le concept de diversité morphologique dans le cadre d'un problème de décomposition d'images en composantes de contour et de texture.

L'idée de départ d'une décomposition fondée sur la diversité morphologique est le constat qu'il existe des ensembles de signaux élémentaires qui représentent parcimonieusement l'une et une seule des deux composantes (contour ou texture). En particulier, une trame de curvelets (CD99; CDDY06; SCD02) est bien adaptée à la représentation parcimonieuse des contours. Par ailleurs, une texture modélisée comme une composante globalement oscillante sera particulièrement bien représentée dans une base de cosinus discrets.

Plus formellement, nous supposons que  $\mathbf{x}$  peut s'écrire comme une combinaison linéaire de  $D$  composantes dites morphologiques (dans notre exemple  $D = 2$ ) :

$$\mathbf{x} = \sum_{i=1}^D \boldsymbol{\varphi}_i + \mathbf{n} \quad (2.20)$$

où les composantes morphologiques  $\{\boldsymbol{\varphi}_i\}_{i=1,\dots,D}$  admettent respectivement une représentation parcimonieuse dans un ensemble de bases orthonormales<sup>3</sup>  $\{\boldsymbol{\Phi}_i\}_{i=1,\dots,D}$ . Le vecteur  $\mathbf{n}$  modélise soit les imperfections du modèle soit une contamination des données par du bruit. Le principe fondateur de la diversité morphologique est celui de la décoherence (dans leur ensemble) de l'ensemble de bases  $\{\boldsymbol{\Phi}_i\}_{i=1,\dots,D}$ . De façon équivalente, la diversité morphologique suppose que chaque composante est particulièrement parcimonieuse dans une et une seule des  $D$  bases de l'ensemble  $\{\boldsymbol{\Phi}_i\}_{i=1,\dots,D}$  (ici au sens d'une norme  $\ell_1$ ) :

$$\forall i, j \in \{1, \dots, D\}; \quad \|\boldsymbol{\varphi}_i \boldsymbol{\Phi}_j^T\|_{\ell_1} < \|\boldsymbol{\varphi}_i \boldsymbol{\Phi}_{j \neq i}^T\|_{\ell_1} \quad (2.21)$$

De fait, cette diversité *inter*-composantes doit permettre la restauration des différentes composantes à partir du signal original  $\mathbf{x}$ . La propriété de diversité, formalisée dans l'équation (2.21), sous-tend l'utilisation de la parcimonie respective des composantes morphologiques dans des bases décoherentes (au sens de la cohérence mutuelle d'une matrice) :

$$\forall k \neq k' \in \{1, \dots, D\}; \quad \mu_{[\boldsymbol{\Phi}_k, \boldsymbol{\Phi}_{k'}]} < 1 \quad (2.22)$$

où  $\mu_{[\boldsymbol{\Phi}_k, \boldsymbol{\Phi}_{k'}]}$  définit la cohérence mutuelle des bases  $\boldsymbol{\Phi}_k$  et  $\boldsymbol{\Phi}_{k'}$ .

Starck *et al.* (SED05) ont ainsi proposé de rechercher les composantes morphologiques les plus parcimonieuses dans leurs bases respectives. Une première approche consiste alors à rechercher la solution (sous réserve d'existence et d'unicité) du problème en norme  $\ell_0$  suivant :

$$\min_{\{\boldsymbol{\varphi}_j\}} \sum_{j=1}^D \|\boldsymbol{\varphi}_j \boldsymbol{\Phi}_j^T\|_{\ell_0} \quad \text{s.c} \quad \left\| \mathbf{x} - \sum_{j=1}^D \boldsymbol{\varphi}_j \right\|_{\ell_2} \leq \epsilon \quad (2.23)$$

sous hypothèse de bruit borné :  $\|\mathbf{n}\|_{\ell_2} < \epsilon$

---

<sup>3</sup>Nous verrons ultérieurement le cas où les trames telles que les curvelets peuvent être adoptées à la place des bases orthonormales.

**Algorithme d'Analyse en Composantes Morphologiques (MCA)**

Nous proposons d'approcher le problème de l'équation (2.23) sous une forme relaxée :

$$\{\varphi_j\} = \underset{\{\varphi_j\}}{\operatorname{Argmin}} \gamma^2 \sum_{j=1}^D \|\varphi_j \Phi_j^T\|_{\ell_0} + \left\| \mathbf{x} - \sum_{j=1}^D \varphi_j \right\|_{\ell_2}^2 \quad (2.24)$$

Ce problème nécessite l'estimation de  $D$  "paramètres" (les composantes morphologiques) couplés pour lesquels une estimation par *Block-coordinate relaxation* (BCR - voir (SAP00)) semble appropriée.

**Block-coordinate relaxation :** Inspiré du *Block-coordinate relaxation* (BCR - voir (SAP00)), l'algorithme MCA estime itérativement chacune des composantes en supposant les autres fixées. Ainsi, à l'itération  $(h)$ , la  $i$ -ème composante morphologique est estimée comme suit :

$$\varphi_i^{(h)} = \underset{\varphi_i}{\operatorname{Argmin}} \gamma^2 \|\varphi_i \Phi_i^T\|_{\ell_0} + \left\| \mathbf{r}_i^{(h)} - \varphi_i \right\|_{\ell_2}^2 \quad (2.25)$$

où  $\mathbf{r}_i^{(h)} = \mathbf{x} - \sum_{j \neq i} \varphi_j^{(h-1)}$  est le résidu calculé à partir des composantes morphologiques courantes différentes de  $\varphi_i$ . Définissons maintenant les coefficients de  $\varphi_i$  dans sa base associée  $\Phi_i$  :

$$\alpha_i = \varphi_i \Phi_i \quad (2.26)$$

La matrice  $\Phi_i$  associée à la base  $\Phi_i^4$  étant orthonormale, le vecteur  $\alpha_i$  est unique. De plus, cette propriété nous permet d'écrire plus simplement le problème de l'équation (2.25) :

$$\alpha_i^{(h)} = \underset{\alpha_i}{\operatorname{Argmin}} \gamma^2 \|\alpha_i\|_{\ell_0} + \left\| \mathbf{r}_i^{(h)} \Phi_i^T - \alpha_i \right\|_{\ell_2}^2 \quad (2.27)$$

La solution à ce problème est unique et obtenue par seuillage dur de la manière suivante :

$$\alpha_i^{(h)} = \mathcal{H}_\gamma \left( \mathbf{r}_i^{(h)} \Phi_i^T \right) \quad (2.28)$$

où l'opérateur de seuillage dur est défini comme suit :

$$\forall i = 1, \dots, t; \quad \mathcal{H}_\gamma(\mathbf{u})[i] = \begin{cases} 0, & \text{si } u[i] < \gamma \\ u[i], & \text{sinon} \end{cases} \quad (2.29)$$

**Seuillage décroissant :** Dans l'algorithme MCA, le rôle joué par le seuil  $\gamma$  est déterminant. En effet, l'approche par BCR et seuillage alterné dans les différentes bases  $\{\Phi_i\}_{i=1, \dots, D}$  est équivalente à la sélection d'atomes de manière alternée dans ces mêmes bases. Considérons le cas simple  $D = 2$ . A une itération donnée  $(h)$ , la première composante morphologique  $\varphi_1$  est estimée de la sorte :

$$\varphi_1^{(h)} = \mathcal{H}_\gamma \left( \mathbf{r}_1^{(h)} \Phi_1^T \right) \Phi_1 \quad (2.30)$$

---

<sup>4</sup>Rappelons que nous nous placerons toujours dans des espaces de dimension finie ; de fait nous confondrons sans ambiguïtés bases, matrices associées et opérateur d'analyse/synthèse.

Supposons, pour simplifier, que le modèle en composantes morphologiques est exact et qu'aucun bruit ne perturbe les données. Le résidu  $\mathbf{r}_1^{(h)}$  peut alors s'écrire comme suit :

$$\mathbf{r}_1^{(h)} = \boldsymbol{\varphi}_1 + (\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2^{(h-1)}) \quad (2.31)$$

L'opérateur de seuillage va sélectionner des coefficients du résidu  $\mathbf{r}_1^{(h)}$  projeté dans la base  $\Phi_1$  :

$$\mathbf{r}_1^{(h)} \Phi_1^T = \alpha_1 + \delta_1^{(h-1)} \quad (2.32)$$

où  $\delta_1^{(h-1)} = (\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2^{(h-1)}) \Phi_1^T$ . Intuitivement, la diversité morphologique implique que  $\boldsymbol{\varphi}_1$  est plus parcimonieux que  $(\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2^{(h-1)})$  dans  $\Phi_1$ . De fait, les coefficients *les plus significatifs* du résidu projeté  $\mathbf{r}_1^{(h)} \Phi_1^T$  devraient essentiellement être affectés à  $\alpha_1$  et donc  $\boldsymbol{\varphi}_1$ .

En conséquence, le choix d'un "bon seuil"  $\gamma$  doit conduire à la sélection des coefficients les plus forts (en valeur absolue). C'est ainsi que Starck *et al.* (voir (SED05; ESDQ05)) ont proposé l'utilisation empirique d'un seuil décroissant et de fait fonction de  $(h)$ . L'algorithme MCA est alors le suivant :

1. Choix d'un nombre d'itérations  $P_{\max}$  et du seuil initial  $\gamma^{(0)}$
2. Tant que  $\gamma^{(h)}$  est plus élevé qu'une valeur minimale fixée  $\gamma_{\min}$  (e.g. dépendant en particulier de la variance du bruit  $\sigma$ )
  - Les composantes morphologiques  $\{\boldsymbol{\varphi}_i\}_{i=1,\dots,D}$  sont calculées à l'itération  $(h)$  de la façon suivante :
    - Pour tout  $i = 1, \dots, D$ 
      - Calcul du résidu  $\mathbf{r}_i^{(h)}$  en supposant les composantes morphologiques  $\boldsymbol{\varphi}_{j \neq i}^{(h-1)}$  fixées :
 
$$\mathbf{r}_i^{(h)} = \mathbf{x} - \sum_{j \neq i} \boldsymbol{\varphi}_j^{(h-1)}$$
      - Estimer les coefficients courants  $\alpha_i^{(h)}$  par seuillage dur de seuil  $\gamma^{(h)}$  :
 
$$\alpha_i^{(h)} = \mathcal{H}_{\gamma^{(h)}}(\mathbf{r}_i^{(h)} \Phi_i^T)$$
      - L'estimée de  $\boldsymbol{\varphi}_i$  est obtenue par reconstruction :
 
$$\boldsymbol{\varphi}_i^{(h)} = \alpha_i^{(h)} \Phi_i$$
    - Décroissance du seuil  $\gamma^{(h)}$  selon une stratégie donnée

### Stratégie de décroissance du seuil

**Choix des seuils  $\gamma^{(0)}$  et  $\gamma_{\min}$  :** une analyse fine du fonctionnement de l'algorithme MCA fait rapidement apparaître que le choix du seuil est un des points fondamentaux de l'algorithme. Dans l'article original (voir (SED05)), le seuil décroissait linéairement entre deux valeurs extrêmes :

- **Valeur initiale  $\gamma^{(0)}$  :** la valeur initiale est donnée par :

$$\max_i \|\mathbf{x} \Phi_i^T\|_{\infty} \quad (2.33)$$

C'est à dire par la valeur absolue du coefficient le plus fort obtenu par projection de  $\mathbf{x}$  dans l'ensemble de bases  $\{\Phi_i\}_{i=1,\dots,D}$ . En pratique, il s'agit du premier coefficient que MCA peut sélectionner.

- **Valeur finale  $\gamma_{\min}$**  : supposons que les données  $\mathbf{x}$  soient entachées de bruit  $\mathbf{n}$ . Dans le cas  $D = 2$ , l'équation (2.31), définissant le résidu à l'itération  $(h)$ , devient alors :

$$\mathbf{r}_1^{(h)} = \varphi_1 + (\varphi_2 - \varphi_2^{(h-1)}) + \mathbf{n} \quad (2.34)$$

Supposons, par souci de simplicité, qu'à la convergence ( $h = P_{\max}$ ), la seconde composante morphologique soit parfaitement estimée ; la relation précédente est alors la suivante :

$$\mathbf{r}_1^{(P_{\max})} = \varphi_1 + \mathbf{n} \quad (2.35)$$

L'estimation de  $\varphi_1$  est alors équivalente à un problème de débruitage sous contrainte de parcimonie dans la base orthonormale  $\Phi_1$ . Dans le cas où  $\mathbf{n}$  est blanc gaussien de variance  $\sigma^2$ , le choix  $\gamma_{\min} = \rho\sigma$  (typiquement  $\rho = 3$ ) semble raisonnable ... tout en restant un critère d'arrêt empirique.

**Stratégie de décroissance** : en pratique, une décroissance linéaire du seuil peut s'avérer peu efficace. Une "bonne" décroissance de seuil devrait en effet être dépendante du signal à décomposer ainsi que de l'ensemble de bases  $\{\Phi_i\}_{i=1,\dots,D}$ . Dans (BSF<sup>+</sup>07), nous avons mené une étude sur la stratégie de décroissance du seuil  $\gamma$ . Dans le cas  $D = 2$ , en accord avec l'équation de mise à jour (2.31), un bon choix de seuil  $\gamma^{(h)}$  doit alors conduire à la sélection de coefficients de  $\alpha_1$  tout en évitant ceux du résidu  $\alpha_2 - \alpha_2^{(h-1)}$ . Néanmoins, la quantité  $\delta_1^{(h-1)} = (\alpha_2 - \alpha_2^{(h-1)}) \Phi_2 \Phi_1^T$  n'est pas accessible. Une quantité calculable à l'itération  $(h)$  est le résidu *total* :

$$\mathbf{r}^{(h)} = \mathbf{x} - \varphi_1^{(h-1)} - \varphi_2^{(h-1)} \quad (2.36)$$

$$= (\alpha_1 - \alpha_1^{(h)}) \Phi_1 + (\alpha_2 - \alpha_2^{(h)}) \Phi_2 \quad (2.37)$$

Définissons les constantes suivantes :

$$m_1 = \|\mathbf{r}^{(h)} \Phi_1^T\|_{\infty} \quad (2.38)$$

$$m_2 = \|\mathbf{r}^{(h)} \Phi_2^T\|_{\infty} \quad (2.39)$$

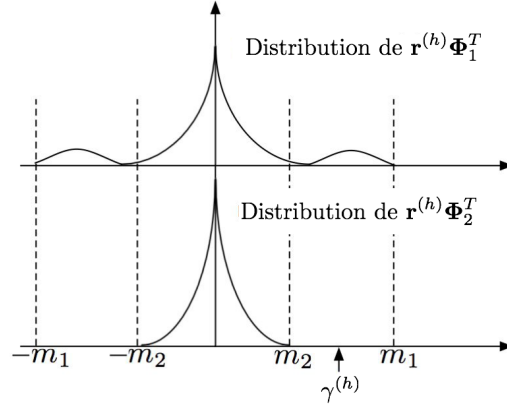
Supposons de plus que  $m_1 > m_2$  (le cas inverse conduirait à la même analyse). La Figure 2.2.1 illustre ce que pourrait être la distribution des coefficients de  $\mathbf{r}^{(h)}$  dans  $\Phi_1$  et  $\Phi_2$ . Dans (BSF<sup>+</sup>07), nous définissons une stratégie de choix de seuil appelée MOM (pour "Mean-Of-Max") définie comme suit :

$$m_2 < \gamma^{(h)} < m_1 \quad (2.40)$$

Dans le même article, nous donnons des conditions garantissant une bonne sélection de coefficients à l'itération  $(h)$  vérifiant :

$$\gamma^{(h)} > \left\| (\alpha_2 - \alpha_2^{(h-1)}) \Phi_2 \Phi_1^T \right\|_\infty \quad (2.41)$$

Les avantages de la stratégie MOM sont les suivants :



**Fig. 2.1:** Stratégie de sélection du seuil.

- **Adaptivité** : dans (BSF<sup>+</sup>07), nous montrons que, en comparaison avec la stratégie de décroissance linéaire, la stratégie MOM conduit à une décroissance bien mieux adaptée au signal à décomposer et en conséquence plus rapide.
- **Aucun paramètre à régler** : la stratégie MOM définit, sans paramètre, une suite de seuils dépendants directement du problème (signal  $x$  et ensemble de bases). La stratégie MOM ne nécessite donc pas, *a priori*, la manipulation de paramètres.

### 2.2.2 MCA, un algorithme de décomposition parcimonieuse ?

#### Des liens étroits avec les problématiques de décomposition parcimonieuse

Dans cette partie, nous considérons que les données ne sont pas contaminées par du bruit :  $\mathbf{n} = \mathbf{0}$ .

**Des liens formels** : rappelons qu'à l'origine, l'estimation des composantes morphologiques par principe de moindre parcimonie s'effectue comme suit :

$$\min_{\{\varphi_j\}} \sum_{j=1}^D \|\varphi_j \Phi_j^T\|_{\ell_0} \text{ s.c } \mathbf{x} = \sum_{j=1}^D \varphi_j \quad (2.42)$$

L'ensemble  $\{\Phi_j\}_{j=1,\dots,D}$  étant constitué de  $D$  bases orthonormales, il est aisé de définir les coefficients  $\alpha_j = \varphi_j \Phi_j^T$  et d'écrire de nouveau le problème précédent :

$$\min_{\{\alpha_j\}} \sum_{j=1}^D \|\alpha_j\|_{\ell_0} \text{ s.c } \mathbf{x} = \sum_{j=1}^D \alpha_j \Phi_j \quad (2.43)$$

Définir le dictionnaire redondant  $\Phi$  comme la concaténation des  $D$  bases orthogonales  $\{\Phi_j\}_{j=1,\dots,D}$  ainsi que le vecteur  $\alpha$  comme la concaténation des  $D$  vecteurs  $\{\alpha_j\}_{j=1,\dots,D}$  conduit sans difficulté à l'équivalence entre le problème précédent et le problème classique de décomposition parcimonieuse d'un signal  $\mathbf{x}$  dans un dictionnaire redondant  $\Phi$  :

$$\min_{\alpha} \|\alpha\|_{\ell_0} \text{ s.c } \mathbf{x} = \alpha \Phi \quad (2.44)$$

**Evidences empiriques :** de fait, MCA cherche à résoudre un problème de décomposition parcimonieuse dans le cas particulier où  $\Phi$  est formé d'une union de bases orthogonales. La question posée précédemment devient alors une interrogation sur la capacité de MCA à résoudre le problème de l'équation (2.44). Dans (BSF<sup>+</sup>07), nous avons fourni quelques expérimentations soutenant une réponse positive.

Dans cette partie, nous considérons le cas simple  $D = 2$ .

Cas parfait - Diagramme de phase Succès/Echec : afin d'évaluer les succès et les échecs de MCA dans un problème de décomposition parcimonieuse, nous définissons la classe de signaux aléatoires  $\mathcal{BG}(p_1, p_2)$  telle que si  $\mathbf{x}$  appartient à  $\mathcal{BG}(p_1, p_2)$ , alors il peut être défini comme une combinaison linéaire de deux composantes morphologiques

$$\mathbf{x} = \alpha_1 \Phi_1 + \alpha_2 \Phi_2 \quad (2.45)$$

où  $\alpha_1$  et  $\alpha_2$  sont deux vecteurs de tailles  $t$  dont les entrées sont tirées suivant une loi Bernoulli-gaussienne. Une telle loi fait intervenir une variable aléatoire supplémentaire régissant l'état d'activité d'un coefficient  $\tau$  tel que

$$\forall i \in \{1, 2\}; \quad \tau_i = \begin{cases} 1, & \text{avec une probabilité } p_i \\ 0, & \text{avec une probabilité } 1 - p_i \end{cases} \quad (2.46)$$

Les coefficients  $\{\alpha_i[k]\}_{i=1,2;k=1,\dots,t}$  sont alors définis conditionnellement à  $\tau_{i=1,2}$  comme suit :

$$\forall i \in \{1, 2\} \text{ et } k \in \{1, \dots, t\}; \quad f(\alpha_i[k]) = \begin{cases} \frac{1}{2} \exp\left(-\frac{\alpha_i[k]^2}{2}\right), & \text{si } \tau_i = 1 \\ \delta(\alpha_i[k]), & \text{si } \tau_i = 0 \end{cases} \quad (2.47)$$

Sur la Figure 2.2.2 est représenté le rapport signal-sur-bruit (RSB) après estimation par MCA/MOM de la première composante morphologique  $\varphi_i$  dont on note l'estimée  $\tilde{\varphi}_1^5$ . Le

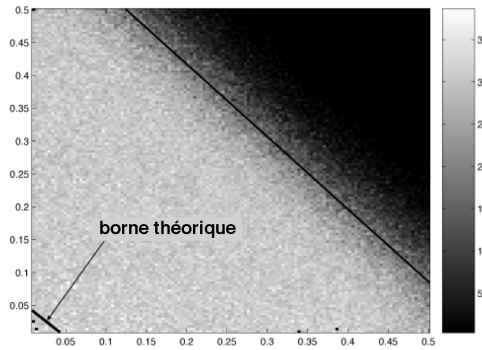
---

<sup>5</sup>Notons que de l'orthogonalité de  $\Phi_1$ , on a l'égalité  $\|\varphi_1 - \tilde{\varphi}_1\|_{\ell_2} = \|\alpha_1 - \tilde{\alpha}_1\|_{\ell_2}$  ce qui constitue un bon critère de convergence en erreur quadratique mais également en support sur  $\alpha_1$ .

dictionnaire  $\Phi$  est formé de l'union de la base canonique (ou identité) et de la base en cosinus discrets. Les mesures de la Figure 2.2.2 ont été obtenues pour des valeurs différentes des probabilités d'activité des coefficients  $p_1$  et  $p_2$  allant de 0,01 à 0,5. Chaque point est évalué à partir de la moyenne de 25 décompositions. Le nombre d'échantillons par composante morphologique est  $t = 512$ .

Si l'on définit une borne de réussite lorsque  $RSB > 20\text{dB}$ , la partie grisée représente les couples  $(p_1, p_2)$  pour lesquels MCA est un succès. Nous avons reporté sur la figure la borne théorique d'unicité/équivalence définie à partir de la cohérence mutuelle du dictionnaire  $\Phi$ . Dans ce cas, cette borne est égale à  $p_1 + p_2 = \frac{1}{t\mu_\Phi}$  dans laquelle  $\mu_\Phi = 1/\sqrt{t}$  (voir (GN03)). Cette expérience simple apporte deux enseignements :

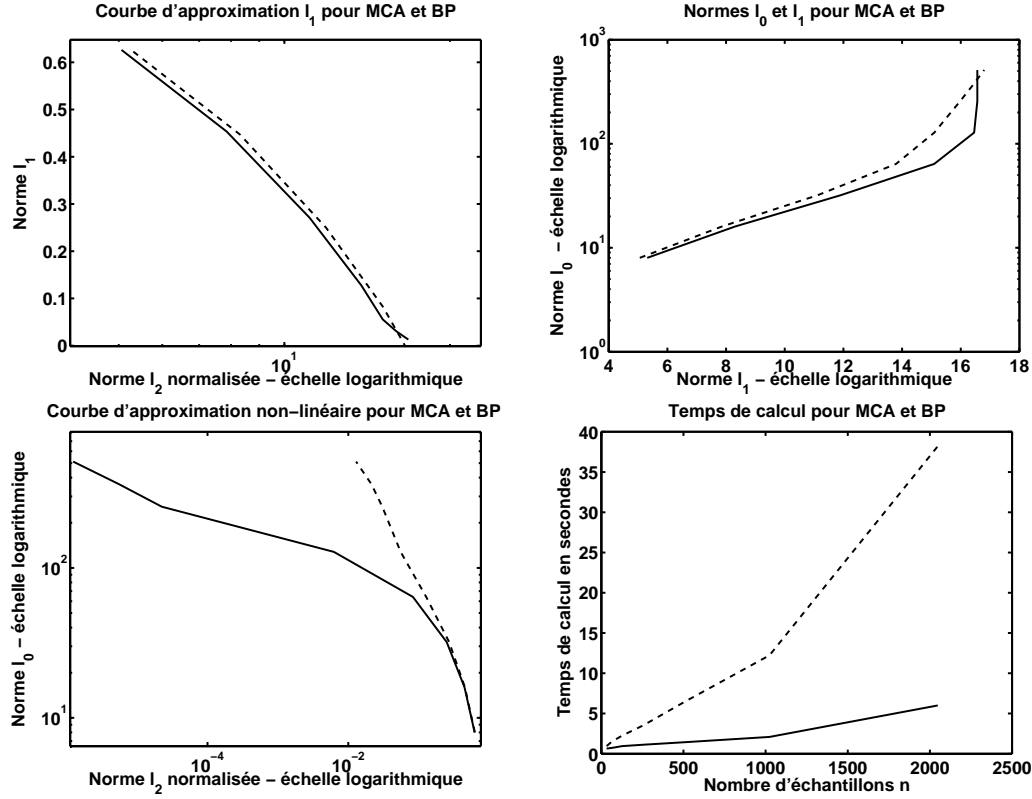
- **Une borne peu réaliste** : comme nous l'avons déjà souligné, la borne d'unicité/équivalence obtenue à partir de la cohérence mutuelle de  $\Phi$  est de loin très pessimiste.
- **MCA/MOM apporte une solution approchée** : MCA/MOM est capable d'apporter des solutions approchées au problème de nature combinatoire de l'équation (2.44).



**Fig. 2.2:** RSB en Decibels de la première composante morphologique extraite MCA/MOM ( $-20 \log_{10} (\|\varphi_1 - \tilde{\varphi}_1\|_{\ell_2})$ ). **Abscisse :**  $p_1$ . **Ordonnées :**  $p_2$ .

#### Cas réel - Comparaison entre MCA et *Basis Pursuit* :

Rappelons que l'analyse en composantes morphologiques a été développée dans le cadre de la séparation contour/texture en imagerie. Toutefois la similitude formelle entre le problème résolu par la MCA et celui du problème de décomposition parcimonieuse dans une union de bases suggère d'effectuer une comparaison empirique de ces deux problèmes. Il a été montré que l'algorithme *Basis Pursuit* converge vers la solution de décomposition en norme  $\ell_0$  lorsque les conditions classiques sur la cohérence mutuelle du dictionnaire  $\Phi$  sont vérifiées (voir entre autres (DE03; GN03)). Les expériences qui vont suivre proposent une comparaison empirique de MCA et de BP dans le cadre d'un problème de décomposition parcimonieuse dans un dictionnaire formé de l'union de deux bases orthonormales  $\Phi = [\Phi_1, \Phi_2]$ . Les résultats présentés en Figure 2.3 sont tels que chaque observation a



**FIG. 2.3:** Haut à droite : courbe  $\ell_1/\ell_2$  de MCA/MOM (*trait continu*) et de Basis Pursuit (*trait discontinu*). Haut à droite : courbe  $\ell_0/\ell_1$  de MCA/MOM (*trait continu*) et de Basis Pursuit (*trait discontinu*). Bas à gauche : courbe d'approximation non linéaire ( $\ell_0/\ell_2$ ) de MCA/MOM (*trait continu*) et de Basis Pursuit (*trait discontinu*). Bas à droite : comparaison du temps d'exécution (en secondes) : MCA/MOM (*trait continu*) et Basis Pursuit (*trait discontinu*).

été calculée comme la moyenne de 50 décompositions de signaux appartenant au modèle Bernoulli-gaussien  $\mathcal{BG}(p_1, p_2)$  avec  $p_1$  et  $p_2$  aléatoires. Le dictionnaire  $\Phi$ , quant à lui, est formé de l'union d'une base orthogonale d'ondelettes et de la base liée à la transformation en cosinus discrète globale (DCT globale). Le graphique en haut et à gauche de la Figure 2.3 montre la courbe  $\ell_1/\ell_2$  après décomposition pour BP et MCA/MOM. La courbe  $\ell_0/\ell_1$  est représentée en haut et à droite de la même figure. La différence entre MCA/MOM et BP conduit aux observations suivantes : i) pour le jeu de signaux utilisés lors de ces simulations, certains ne vérifient pas la condition d'équivalence entre les problèmes  $\ell_0$  et  $\ell_1$  ; ii) d'autre part, la décomposition obtenue par MCA/MOM tend à être plus parcimonieuse (distribution plus piquée) que celle qu'apporte BP. On en déduit de fait, en accord avec le diagramme de phase du paragraphe précédent, que MCA/MOM tend à résoudre un problème bien plus proche de celui en norme  $\ell_0$  que  $\ell_1$  lorsque l'équivalence n'est pas garantie. La différence entre les deux problèmes  $\ell_0$  et  $\ell_1$  est particulièrement mise en évidence sur le



graphique représentant la courbe d'approximation linéaire (courbe  $\ell_2/\ell_0$ ) en bas et à gauche de la Figure 2.3. Cette courbe met en avant la capacité de MCA à apporter une solution bien plus parcimonieuse que celle apportée par une décomposition en norme  $\ell_1$ . Au-delà de ces considérations liées au degré de parcimonie, la courbe, en bas et à droite de la Figure 2.3, illustre la différence de rapidité d'exécution des deux algorithmes. MCA utilisant pour  $\Phi_1$  (resp.  $\Phi_2$ ) les transformées rapides implicites en ondelettes orthogonales (resp. en DCT globale), son temps d'exécution est particulièrement faible.

En conclusion, il semble que MCA soit capable de résoudre (au moins de manière approchée) le problème de décomposition parcimonieuse en norme  $\ell_0$ . Les propositions suivantes montrent, sous certaines hypothèses, que MCA converge vers la solution du problème de l'équation (2.4).

### Conditions de convergence

**Proposition 1 (Propriété de sélection exacte)** *Supposons que  $\mathbf{x}$  soit  $K$ -sparse tel que :*

$$\mathbf{x} = \sum_{j=1}^D \sum_{i \in \Lambda_j} \alpha_j[i] \phi_j[i]$$

où  $K = \sum_j \text{Card}(\Lambda_j)$ . A l'itération  $h$ , supposons que le résidu  $\mathbf{r}^{(h)}$  est  $K$ -sparse de sorte que :

$$\mathbf{r}^{(h)} = \sum_{j=1}^D \sum_{i \in \Lambda_j} \beta_j[i] \phi_j[i]$$

On note :

$$\begin{aligned} (\tau^*, i^*) &= \underset{\tau, i \in \Lambda_\tau}{\text{Argmax}} |\beta_\tau[i]| \\ (\tau^\dagger, i^\dagger) &= \underset{\tau \neq \tau^*, i \in \Lambda_\tau}{\text{Argmax}} |\beta_\tau[i]| \\ \beta^* &= |\beta_{\tau^*}[i^*]| \\ \beta^\dagger &= |\beta_{\tau^\dagger}[i^\dagger]| \end{aligned}$$

On suppose qu'il existe, à l'itération  $(h)$ , un scalaire  $0 \leq \rho < 1$  tel que  $\beta^\dagger = \rho \beta^*$ . Si la condition suivante est vérifiée :

$$K < \mu_\Phi^{-1} \frac{1-\rho}{1+\rho}$$

Alors MCA/MOM sélectionne, à l'itération  $h$ , des coefficients appartenant au support de  $\mathbf{x}$ .

**Proposition 2 (Convergence)** *Supposons que  $\mathbf{x}$  soit  $K$ -sparse tel que :*

$$\mathbf{x} = \sum_{j=1}^D \sum_{i \in \Lambda_j} \alpha_j[i] \phi_j[i]$$


---

où  $K = \sum_j \text{Card}(\Lambda_j)$ .

Supposons que les hypothèses de la Propriété de sélection exacte sont vérifiées. Si  $K < \mu_{\Phi}^{-1} \frac{1-\rho}{1+\rho}$  alors MCA/MOM converge exponentiellement vers  $\mathbf{x}$  et sa décomposition la plus parcimonieuse au sens  $\ell_0$  dans  $\Phi$ . Plus précisément, le résidu converge vers 0 avec une vitesse exponentielle.

### Un algorithme de régression *stagewise*

Revenons à la propriété de "bonne sélection atomique". Il apparaît alors que, muni de la stratégie de sélection MOM, l'algorithme MCA sélectionne, pour une itération ( $h$ ), un ensemble d'atomes appartenant à une même composante morphologique  $\varphi_{i^*}$ . De plus, le choix de  $i^*$  s'effectue dans le sens de l'atome de  $\Phi$  le plus cohérent avec le résidu total  $\mathbf{r}^{(h)}$ . Finalement, au sens de la stratégie MOM, le choix de  $i^*$  s'effectue comme suit :

$$i^* = \underset{i}{\text{Argmax}} \left\| \mathbf{r}^{(h)} \Phi_i^T \right\|_{\ell_\infty} \quad (2.48)$$

Considérons le cas extrême où, au lieu de sélectionner un **ensemble d'atomes** de  $\Phi_{i^*}$ , MCA ne sélectionne que **l'atome** de  $\Phi_{i^*}$  le plus cohérent avec le résidu total  $\mathbf{r}^{(h)}$ . Cette version *stepwise* (i.e. ne sélectionnant qu'un **unique atome** à chaque itération) serait totalement équivalente au très classique algorithme de *Matching Pursuit* (MZ93; GV06). De fait, des développements de l'article (BMFS07b), il découle une interprétation de l'algorithme MCA/MOM comme une version *stagewise* (i.e. sélectionnant **plusieurs atomes** à chaque itération) de l'algorithme *Matching Pursuit*. Du point de vue du temps de calcul de ces deux algorithmes, MCA/MOM présente deux avantages capitaux dans le cas de problèmes de grande dimension :

- **Aspect *stagewise*** : contrairement aux algorithmes *stepwise* dont *Matching Pursuit* est un exemple, MCA/MOM sélectionne un ensemble d'atomes. Une telle approche *stagewise* permet de pallier, en partie, les limitations des algorithmes *stepwise* dans le cas de problèmes de grande dimension.
- **Utilisation explicite d'opérateurs rapides** : à l'origine, l'algorithme MCA propose une méthode de décomposition d'une image complexe en composantes morphologiques. La restauration de ces dernières se fonde sur la décoherence des représentations dans lesquelles les-dites composantes morphologiques sont liées. Dans un cadre expérimental, l'un des attraits majeurs de MCA est alors de faire intervenir des représentations admettant des opérateurs implicites **rapides** telles que les ondelettes, curvelets<sup>6</sup>, transformée en cosinus discrets (locaux ou globaux),... etc. Les opérations

---

<sup>6</sup>Notons que les curvelets forment non pas une base orthonormale mais une trame ajustée. Dans ce cas, leur utilisation sort entièrement du cadre MCA que nous venons de présenter. Néanmoins, en pratique, leur utilisation au sein de MCA conduit à des résultats satisfaisants. Une des interprétations possibles est que, tel que noté par Elad dans (Ela06), l'opération de seuillage effectuée à chaque itération de MCA constitue le

de l'algorithme MCA faisant intervenir les bases  $\{\Phi_i\}_{i=1,\dots,D}$  (*resp.* leur transposée) sont, en pratique, effectuées par leur opérateurs implicites de synthèse (*resp.* d'analyse).

Au final, ces deux points permettent d'obtenir un algorithme de décomposition parcimonieuse rapide tout en présentant des performances théoriques semblables à celle présentées par un algorithme *stepwise* tel que *Matching Pursuit*. Par la suite, l'utilisation de MCA sera justifiée pour résoudre des problèmes de décomposition parcimonieuse dans un dictionnaire redondant formé d'une union de bases orthonormales.

Remarquons que, dans le cadre des problématiques de décomposition dans un dictionnaire redondant, les premières approches *stagewise* ont été proposées par Donoho *et al.* dans (DTDS06). Ces auteurs proposent en particulier une version *stagewise* de l'algorithme OMP (*StOMP - Stagewise Orthogonal Matching Pursuit*).

L'interprétation de MCA comme un algorithme *stagewise* de décomposition parcimonieuse (BSF<sup>+</sup>07; BMFS07b) a été développée parallèlement à StOMP et ceci dans le souhait d'obtenir des algorithmes de décomposition rapides et efficaces applicables dans des problèmes de grande dimension. Nous avons voulu montrer ici qu'en dehors du cadre initial de décomposition en composantes morphologiques, MCA se place entièrement dans le cadre des problématiques de décomposition parcimonieuse dans des dictionnaires redondants.

## 2.3 Vers une représentation parcimonieuses des données multivaluées

Dans le cadre de l'analyse de données multivaluées  $\mathbf{X}$  (écrites sous la forme d'une matrice  $m \times t$ ), il est tout à fait clair que les structures présentes dans ces données sont autant *spatiales* que *spectrales*. Classiquement, les structures spatiales induisent l'existence d'un dictionnaire  $\Phi$  dans lequel chacune des  $m$  observations admet un certain niveau de parcimonie. Prenons l'exemple des deux images de la Figure 2.4 : celle de gauche présente une observation *Mars Express*<sup>7</sup> dont chaque pixel peut être modélisé comme un mélange de spectres d'absorption. L'image de gauche représente *Barbara* dans sa version couleur présentant donc 3 observations liées aux canaux rouge, vert et bleu. De notre point de vue, ce type de données multivaluées (image hyperspectrale pour la première et couleur pour la seconde) présentent un point commun : le fait qu'en général, ces données sont également structurées suivant leur direction spectrale.

Dans le domaine du traitement parcimonieux des données multivaluées, deux approches aux philosophies différentes s'opposent :

---

"premier pas" vers la solution du problème en présence d'une représentation redondante telle qu'une trame de curvelets.

<sup>7</sup><http://www.esa.int/marsexpress>

- **Recherche d'une parcimonie jointe** : ce type d'approche consiste soit en l'utilisation d'une mesure de parcimonie inter-canaux favorisant l'apparition de supports de décomposition joints entre les canaux (voir (TGS06; Tro06a; CH05)) ou directement en promouvant les supports joints entre les canaux (voir (CREKD05; FR06)). Si de telles conceptions en matière de données multivaluées peuvent être relativement bien adaptées dans le cas des données couleur (les structures telles que les contours ou les textures sont effectivement jointes entre les canaux), elles ne sont pas nécessairement "optimales" dans le cas où la taille du problème (ici le nombre d'observations  $m$ ) est important. Nous reviendrons sur ce point au Chapitre 7.
- **Recherche d'une décomposition parcimonieuseuse multicanale** : dans une approche plus classique, les structures (tant *spatiales* que *spectrales*) induisent l'existence d'une représentation, ici multicanale, dans laquelle l'ensemble des données  $\mathbf{X}$  admettent une représentation parcimonieuse.



FIG. 2.4: Exemples de données multivaluées.

Dans nombre de problèmes inverses, la recherche d'un degré de parcimonie important est la clef d'une bonne estimation. C'est de fait la seconde approche que nous adoptons par la suite.

### 2.3.1 Dictionnaires multicanaux

Le concept de dictionnaire redondant multicanal a tout d'abord été introduit par Grignonval (RM05) dans le cadre des problèmes de séparation de source sous-déterminée. Dans ce mémoire, les dictionnaires multicanaux seront le fondement de la plupart des développements que nous introduirons par la suite (séparation de source, restauration,  $\dots$  etc.). Les données multicanales  $\mathbf{X}$  s'écrivent sous la forme d'une matrice de taille  $m \times t$  contenant

$m$  observées de  $t$  échantillons. Dire que  $\mathbf{X}$  admet une représentation parcimonieuse dans le dictionnaire **multicanal**  $\Psi$  signifie qu'elle peut s'écrire comme une combinaison linéaire d'un "faible" nombre d'atomes multicanaux  $\{\psi_\gamma\}_{\gamma \in \Gamma}$ <sup>8</sup> :

$$\mathbf{X} = \sum_{\gamma \in \Gamma} \alpha_\gamma \psi_\gamma \quad (2.49)$$

où, finalement, les atomes multicanaux  $\{\psi_\gamma\}_{\gamma \in \Gamma}$  sont des matrices de taille  $m \times t$ . Jusqu'ici, sans hypothèses supplémentaires concernant la structure des atomes multicanaux, le problème que nous considérons est totalement identique au cas monocal.

Notre objectif est de construire un outil d'analyse pour des données ayant des structures à la fois *spatiales* et *spectrales*. Par exemple, dans le cas de l'imagerie hyperspectrale, un outil d'analyse adapté est celui qui, spatialement pourrait représenter des contours (*i.e.* atomes de curvelets) tout en ayant, spectralement, la possibilité d'analyser des signaux ayant une signature spectrale particulière (par exemple du béton pour l'analyse de données hyperspectrales de scènes urbaines). De fait, un tel atome **multicanal** peut être représenté sous la forme du produit tensoriel d'un atome *spatial*  $\phi_j$  et d'un atome *spectral*  $\xi_i$ . En définissant  $\Gamma = \{[i, j]\}_{i=1, \dots, m; j=1, \dots, t}$  l'ensemble des couples d'indices, chaque atome multicanal s'écrit alors comme suit :

$$\forall [i, j] \in \Gamma; \quad \psi_{\gamma=[i, j]} = \pi_i \otimes \phi_j \quad (2.50)$$

L'ensemble du dictionnaire multicanal  $\Psi$  s'écrit alors comme le produit tensoriel des deux dictionnaires *spectraux* et *spatiaux* :  $\Psi = \Pi \otimes \Phi$ . Les données multicanales  $\mathbf{X}$  sont alors vues sous l'angle particulier d'une combinaison linéaire d'éléments de  $\mathbb{R}^m \otimes \mathbb{R}^t$ .

### 2.3.2 Extension

Une fois ces définitions particulières posées, il est possible d'étendre les résultats classiques du cas monocal au cas multicanal (voir (BMFS07b)). Tout d'abord, comme dans le cas monocal, la décomposition dans un dictionnaire redondant requiert la notion de cohérence entre atomes. Assez naturellement, le produit scalaire entre deux atomes multicanaux  $\psi_{ij}$  et  $\psi_{pq}$  sera défini comme suit :

$$\forall i, j, p, q; \quad \langle \psi_{ij}, \psi_{pq} \rangle = \langle \pi_i, \pi_p \rangle \langle \phi_j, \phi_q \rangle \quad (2.51)$$

La norme induite sera la norme de Frobenius de la matrice  $\psi_{ij}$ . La cohérence mutuelle du dictionnaire multicanal  $\Psi$  est par conséquent définie comme suit :

$$\mu_\Psi = \max \{\mu_\Pi, \mu_\Phi\} \quad (2.52)$$

Dans ce cas, la condition d'unicité et d'équivalence de l'équation (2.7) des problèmes de décomposition  $\ell_0$  et  $\ell_1$  peut être étendue au cas multicanal. Définissons  $\alpha = [\alpha_{ij}]_{i=1, \dots, m; j=1, \dots, t}$ ,

---

<sup>8</sup> $\Gamma$  représente un ensemble d' "indices" qui sera précisé par la suite.

le problème de décomposition parcimonieuse dans le dictionnaire  $\Psi$  s'écrit alors comme suit :

$$\min_{\alpha} \|\alpha\|_{\ell_0} \text{ s.c. } \mathbf{X} = \sum_{i,j} \alpha_{ij} \Psi_{ij} \quad (2.53)$$

De façon analogue, si la condition  $\|\alpha\|_{\ell_0} < 1/2(1 + 1/\mu_{\Psi})$  est vérifiée, le problème précédent possède une unique solution et sa variante en norme  $\ell_1$  lui est équivalente. Néanmoins, ceci mérite quelques commentaires. En effet, intuitivement, disposer de  $m$  observations peut, dans certains cas, rendre le problème de décomposition plus "facile" (cohérence mutuelle multicanale plus faible). De la définition précédente, la cohérence mutuelle du dictionnaire multicanal  $\mu_{\Psi} = \max\{\mu_{\Pi}, \mu_{\Phi}\}$  peut prendre des valeurs plus faibles (si  $\mu_{\Pi} < \mu_{\Phi}$ ) que dans le cas monocal (cohérence  $\mu_{\Phi}$ ). Ce phénomène est essentiellement la conséquence du faible nombre d'hypothèses sur la construction des atomes multicanaux. En effet, jusqu'ici, rien interdit les atomes  $\{\Psi_{ij}\}_{i,j}$  à partager des spectres ou des morphologies spatiales identiques. Si de plus, nous faisons l'hypothèse que le dictionnaire  $\Psi$  contient des atomes dont la morphologie spatiale et le spectre sont différents, alors la cohérence mutuelle devient :

$$\mu_{\Psi} = \mu_{\Pi} \mu_{\Phi} \quad (2.54)$$

Cette cohérence est de fait toujours inférieure ou égale à celle du problème monocal.

## 2.4 Analyse en composantes morphologiques multivaluées

### 2.4.1 Principe

Dans ce paragraphe, nous proposons une extension du concept de diversité morphologique au cas multicanal. Par la suite, nous supposerons que les données multicanales  $\mathbf{X}$  sont structurées de sorte que chaque canal/observation  $\{\mathbf{x}_i\}_{i=1,\dots,m}$  est modélisé comme la combinaison linéaire de  $D$  composantes morphologiques :

$$\forall i \in \{1, \dots, m\}; \quad \mathbf{x}_i = \sum_{j=1}^D \varphi_{ij} \quad (2.55)$$

où chaque composante morphologique  $\varphi_{ij}$  est supposée parcimonieuse dans la base orthonormale  $\Phi_j$ . De façon équivalente, chaque canal  $\{\mathbf{x}_i\}_{i=1,\dots,m}$  est de fait parcimonieux dans le dictionnaire redondant  $\Phi$  formé de l'union des  $D$  bases orthonormales  $\{\Phi_i\}_{i=1,\dots,D}$ .

Chaque colonne de la matrice  $\mathbf{X}$  admet une décomposition parcimonieuse dans le dictionnaire redondant  $\Pi$  formé de l'union des  $D'$  bases orthonormales  $\{\Pi_i\}_{i=1,\dots,D'}$ . Cette formulation est celle que nous choisissons pour modéliser les structures inter-canaux présentes dans bon nombre de données multicanales telles les données hyperspectrales. D'un point de vue global, les données multicanales  $\mathbf{X}$  admettent alors une représentation parcimonieuse dans le dictionnaire redondant **multicanal**  $\Psi = [\Pi_1 \cdots \Pi_{D'}] \otimes [\Phi_1 \cdots \Phi_D]$ . Par analogie au

---

cas monocal, les données  $\mathbf{X}$  sont modélisées comme la combinaison linéaire de  $D \times D'$  composantes morphologiques :

$$\mathbf{X} = \sum_{j=1}^D \sum_{k=1}^{D'} \varpi_{jk} \quad (2.56)$$

où  $\varpi_{jk}$  admet une représentation parcimonieuse dans  $\Pi_k \otimes \Phi_j$ . Dans ce nouveau cadre multicanal, séparer deux composantes morphologiques  $\varpi_{ip}$  and  $\varpi_{jq \neq ip}$  s'effectue en accord avec le principe de diversité morphologique multicanale :

- **Morphologies spatiales ou temporelles (resp. spectrales)** : dans le cas où  $i \neq j$  et  $p = q$  (resp.  $i = j$  et  $p \neq q$ ), les composantes morphologiques en jeu présentent des structures spectrales (resp. spatiales) comparables. La distinction entre celles-ci s'effectue alors à partir de leur diversité spatiale (resp. spectrale). La cohérence mutuelle des bases multicanales en jeu  $\Pi_p \otimes \Phi_i$  et  $\Pi_p \otimes \Phi_j$  (resp.  $\Pi_p \otimes \Phi_i$  et  $\Pi_q \otimes \Phi_i$ ) est bornée supérieurement par  $\mu_\Phi$  (resp.  $\mu_\Pi$ ). Dans ce cas, la séparation de telles composantes morphologiques est équivalente à la résolution de multiples problèmes de décompositions monovalués.
- **Morphologies spatiales et spectrales** : lorsque  $i \neq j$  et  $p \neq q$ , les composantes morphologiques se distinguent par leurs diversités spatiale et spectrale. Notons qu'ici, la cohérence entre  $\Pi_p \otimes \Phi_i$  et  $\Pi_q \otimes \Phi_j$  est plus faible :  $\mu_\Pi \mu_\Phi \leq \max\{\mu_\Pi, \mu_\Phi\}$ . Ce cas est celui où la notion de diversité multicanale apporte un gain important.

Dans le cadre du traitement des données multivaluées telles que nous les avons définies, leur analyse efficace devra prendre en compte leurs diversités tant *spectrales* que *spatiales*. Dans le cas monocal, la MCA offre un outil d'analyse privilégié. Nous proposons d'étendre la MCA au cas de l'analyse de données multivaluées. L'Analyse en Composantes Morphologiques Multivaluées (mMCA) cherche la solution au problème suivant :

$$\min_{\{\varpi_{jk}\}} \gamma^2 \sum_{j=1}^D \sum_{k=1}^{D'} \|\Pi_k^T \varpi_{jk} \Phi_j^T\|_0 + \left\| \mathbf{X} - \sum_{j=1}^D \sum_{k=1}^{D'} \varpi_{jk} \right\|_F^2 \quad (2.57)$$

Nous avons mis en lumière les liens forts entre le problème de l'analyse en composantes morphologiques et celle de la recherche de décompositions parcimonieuses dans une union de bases. Ces résultats restent valables dans le cas multicanal.

### 2.4.2 Algorithme

Dans le cadre défini par mMCA, les données  $\mathbf{X}$  sont modélisées comme la combinaison linéaire de  $D \times D'$  composantes morphologiques  $\{\varpi_{jk}\}_{j=1, \dots, D; k=1, \dots, D'}$ . Définissons  $\Lambda_{jk}$  le *support* de la composante morphologique  $\varpi_{jk}$  dans le sous-dictionnaire  $\Psi_{jk} = \Pi_k \otimes \Phi_j$ .  $\mathbf{X}$  est  $K$ -sparse dans le dictionnaire pris dans son ensemble avec  $K = \sum_{j,k} \text{Card}(\Lambda_{jk})$ . Les

---

données se décomposent alors comme suit :

$$\mathbf{X} = \sum_{j=1}^D \sum_{k=1}^{D'} \boldsymbol{\varpi}_{jk} = \sum_{j=1}^D \sum_{k=1}^{D'} \sum_{i \in \Lambda_{jk}} \alpha_{jk}[i] \boldsymbol{\psi}_{jk}[i] \quad (2.58)$$

De façon similaire à l'algorithme MCA, mMCA estime itérativement chacune des composantes morphologiques  $\boldsymbol{\varpi}_{jk}$  alternativement (approche BCR, voir (SAP00)). Chaque matrice de coefficients  $\alpha_{jk}$  s'obtient alors comme suit :

$$\alpha_{jk} = \underset{\alpha_{jk}}{\operatorname{Argmin}} \gamma^2 \|\alpha_{jk}\|_0 + \|\mathbf{R}_{jk} - \mathbf{\Pi}_k \alpha_{jk} \mathbf{\Phi}_j\|_F^2 \quad (2.59)$$

où  $\mathbf{R}_{jk} = \mathbf{X} - \sum_{p,q \neq j,k} \mathbf{\Pi}_q \alpha_{pq} \mathbf{\Phi}_p$  est le terme de résidu.

De l'orthonormalité des bases  $\{\mathbf{\Phi}_j\}_j$  et  $\{\mathbf{\Pi}_k\}_k$ , l'équation (2.59) se formule également comme suit :

$$\alpha_{jk} = \underset{\alpha_{jk}}{\operatorname{Argmin}} \gamma^2 \|\alpha_{jk}\|_0 + \|\mathbf{\Pi}_k^T \mathbf{R}_{jk} \mathbf{\Phi}_j^T - \alpha_{jk}\|_F^2 \quad (2.60)$$

qui admet comme solution unique  $\alpha_{jk} = \mathcal{H}_\gamma(\mathbf{\Pi}_k^T \mathbf{R}_{jk} \mathbf{\Phi}_j^T)$ .

L'algorithme mMCA est décrit ci-dessous :

1. Fixer  $P_{\max}$  et le seuil  $\gamma^{(0)}$ .
2. Tant que  $\gamma^{(h)}$  est plus élevé qu'une valeur minimale fixée  $\gamma_{\min}$  (e.g. dépendant en particulier de la variance du bruit  $\sigma$ )
  - Les composantes morphologiques  $\{\boldsymbol{\varpi}_i\}_{i=1,\dots,D; i=k,\dots,D'}$  sont calculées à l'itération  $(h)$  de la façon suivante :
 

Pout tout  $i = 1, \dots, D$  et  $k = 1, \dots, D'$

    - Calcul du résidu  $\mathbf{R}_{jk}^{(h)}$  en supposant les composantes morphologiques  $\boldsymbol{\varpi}_{pq \neq jk}$ ,  $\tilde{\boldsymbol{\varpi}}_{pq \neq jk}^{(h-1)}$  fixées :
 
$$\mathbf{R}_{jk}^{(h)} = \mathbf{X} - \sum_{pq \neq jk} \tilde{\boldsymbol{\varpi}}_{pq \neq jk}^{(h-1)}$$
    - Estimer les coefficients courants de  $\tilde{\boldsymbol{\varpi}}_{jk}^{(h)}$  par seuillage dur de seuil  $\gamma^{(h)}$  :
 
$$\tilde{\alpha}_{jk}^{(h)} = \mathcal{H}_{\gamma^{(h)}}(\mathbf{\Pi}_k^T \mathbf{R}_{jk}^{(h)} \mathbf{\Phi}_j^T)$$
    - L'estimée de  $\boldsymbol{\varpi}_{jk}$  est obtenue par reconstruction :
 
$$\tilde{\boldsymbol{\varpi}}_{jk}^{(h)} = \mathbf{\Pi}_k \tilde{\alpha}_{jk}^{(h)} \mathbf{\Phi}_j$$
  - Décroissance du seuil  $\gamma$  selon une stratégie donnée.

#### Stratégie de seuillage :

Nous avons montré que, dans le cas monocanal, MCA/MOM permet, sous certaines conditions relatives à la cohérence du dictionnaire, d'atteindre la décomposition en norme



$\ell_0$ . De façon analogue, la stratégie MOM s'étend facilement au cas multicanal. A chaque itération  $h$  le résidu est projeté sur chaque sous-dictionnaire et nous définissons :

$$m_{j_0 k_0}^{(h-1)} = \max_{i,j} \left\| \Pi_k^T \left( \mathbf{X} - \sum_{p,q} \Pi_q \tilde{\alpha}_{pq}^{(h-1)} \Phi_p \right) \Phi_j^T \right\|_{\ell_\infty} \quad (2.61)$$

$$m_{j_1 k_1}^{(h-1)} = \max_{\{i,j\} \neq \{i_0, j_0\}} \left\| \Pi_k^T \left( \mathbf{X} - \sum_{p,q} \Pi_q \tilde{\alpha}_{pq}^{(h-1)} \Phi_p \right) \Phi_j^T \right\|_{\ell_\infty} \quad (2.62)$$

Le seuil MOM multicanal (mMOM) est alors défini comme suit :

$$\gamma^{(h)} = \frac{1}{2} \{ m_{j_0 k_0}^{(h-1)} + m_{j_1 k_1}^{(h-1)} \} \quad (2.63)$$

Muni d'une telle stratégie de choix de seuil, mMCA/mMOM vérifie des conditions de convergence en tout point similaires à celles vérifiées par MCA/MOM. Des expériences de simulations sont présentées dans ([BMFS07b](#)).

---

## Chapitre 3

# Données multivaluées et séparation de sources

L'ANALYSE en composantes morphologiques multicanales introduite au Chapitre 2 fournit un outil de décomposition parcimonieuse de données multivaluées. Elle permet en particulier de séparer des composantes morphologiques multicanales admettant une stricte diversité morphologique spectrale et/ou spatiale. Dans ce chapitre, nous considérons le cas où les signaux multivalués à séparer admettent une diversité morphologique au sens large. Dans ce cas, il convient de s'intéresser aux problèmes de Séparation de Source Aveugle (SSA).

La première partie de ce chapitre propose une introduction aux problèmes de SSA ainsi qu'aux méthodes les plus classiques développées pour les résoudre telle que l'Analyse en Composantes Indépendantes (ACI). Nous introduisons ensuite une nouvelle approche de la SSA fondée sur la recherche de signaux ou composantes admettant une représentation la plus creuse possible dans un dictionnaire  $\Phi$ . Cette nouvelle conception de la SSA conduit à définir l'Analyse en Composantes Mutuellement Parcimonieuses (ACMP).

---

### 3.1 Données multivaluées : du modèle de mélange à la séparation de sources

L'appellation "données multivaluées" est un terme générique qu'il est nécessaire de préciser. Une grande variété de modèles formels peut être utilisée pour la caractérisation de ce type de données : nous nous intéresserons en particulier au cas du modèle de mélange linéaire.

#### 3.1.1 Données multivaluées et modèle de mélange

D'un point de vue formel, des données multivaluées sont constituées de différents "échantillons" regroupés dans un tableau ou plus précisément une matrice de taille  $m \times n$  :  $\mathbf{X}$ . Conformément à la notation choisie jusqu'à présent, chacune des  $m$  lignes de  $\mathbf{X}$  contient les échantillons d'une observation : physiquement, la cohérence (*i.e.* équivalent de structure) des données suivant une ligne de  $\mathbf{X}$  est la cohérence que présente une observation prise comme un signal individuel. Par exemple, ces échantillons sont ceux d'une image, d'une série temporelle ou tout autre type de signal.

Il est intéressant de constater que la particularité des données multivaluées réside dans le fait que la matrice  $\mathbf{X}$  n'est pas qu'un empilement de  $m$  signaux sans lien : nous supposons qu'il existe une structure **inter-canal** ou **inter-observation**.

Prenons l'exemple de l'imagerie multispectrale : chaque observation est une image de la même scène mais observée dans des bandes de fréquences différentes  $\{\nu_i\}_{i=1,\dots,m}$ . L'élément  $X[i, j]$  (à la  $i$ -ème ligne et  $j$ -ème colonne) est la réponse, au pixel  $j$ , d'un ensemble de phénomènes physiques à la fréquence  $\nu_i$ .

Une modélisation correcte des données multivaluées  $\mathbf{X}$  doit donc prendre en compte cette double structure **spatiale/temporelle** et **inter-canal**. Suivant l'objectif que l'on se fixe, plusieurs approches sont possibles. Nous évoquerons au Chapitre 7 certaines approches récentes mises en oeuvre pour la résolution de problèmes de restauration de données multivaluées.

Nous nous intéressons ici à un problème plus particulier : celui de l'extraction de signaux. Dans ce type de problématique, chaque canal  $\{\mathbf{x}_i\}_{i=1,\dots,m}$  (les lignes de  $\mathbf{X}$ ) est supposé être défini comme une fonction d'un ensemble de composantes ou sources élémentaires  $\{\mathbf{s}_k\}_{k=1,\dots,n}$  :

$$\mathbf{x}_i = \mathcal{F}_i(\mathbf{s}_1, \dots, \mathbf{s}_n) \quad (3.1)$$

où  $\mathcal{F}_i$  permet la modélisation de la dépendance des données aux composantes  $\{\mathbf{s}_k\}_{k=1,\dots,n}$ . La dépendance en  $i$  de  $\mathcal{F}_i$  entraîne que les composantes peuvent intervenir de façon différente d'une observation à l'autre. Le problème de l'extraction de signaux consiste alors à estimer les composantes  $\{\mathbf{s}_k\}_{k=1,\dots,n}$  à partir de la connaissance de  $\mathbf{X}$  et de  $\{\mathcal{F}_1, \dots, \mathcal{F}_m\}$ . Dans le cas général, la séparation de signaux, plus généralement appelée **séparation de sources** est

---

équivalente à un problème inverse dont la complexité dépendra essentiellement de l'expression des fonctions  $\{\mathcal{F}_1, \dots, \mathcal{F}_m\}$ .

Cette formalisation fait appel à des fonctions de  $\{\mathcal{F}_1, \dots, \mathcal{F}_m\}$  qui peuvent potentiellement apporter une grande souplesse de modélisation au détriment d'une certaine complexité. L'exemple le plus célèbre d'application d'un tel type de modèle est celui de la "cocktail party" (Car) : lors d'une soirée plus ou moins arrosée,  $n$  convives prennent part à des discussions. On dispose  $m$  microphones dans la salle. Dans le modèle précédent, la fonction  $\mathcal{F}_i$  modélise la manière dont les  $n$  convives contribuent au signal enregistré par le microphone numéroté  $i$ . Le défi de l'extraction de sources consiste alors à retrouver la contribution de chacun des convives à partir de leurs différents "mélanges" captés par les  $m$  microphones. Dans un cadre très général, les fonctions  $\{\mathcal{F}_1, \dots, \mathcal{F}_m\}$  peuvent modéliser : i) la contribution de chaque convive (en terme de puissance sonore au niveau des microphones), ii) des effets physiques divers tels que des effets d'écho ... etc.

Le modèle le plus simple que l'on peut concevoir est celui qui consiste à négliger tous les effets physiques secondaires (retards, échos, réverbération, convolution ... etc) et à ne prendre en compte que la contribution en terme de puissance sonore de chacun des convives. En conséquence, le modèle devient extrêmement simple puisque le signal capté au niveau du microphone numéroté  $i$  est, en première approximation, la combinaison linéaire du signal sonore émis par chaque convive :

$$\forall i = 1, \dots, m; \quad \mathbf{x}_i = \sum_{j=1}^n a_{ij} \mathbf{s}_j \quad (3.2)$$

où les fonctions  $\{\mathcal{F}_1, \dots, \mathcal{F}_m\}$  de l'équation (3.1) deviennent de simples opérateurs linéaires. Le modèle de l'équation (3.2) est classiquement nommé **modèle de mélange**. Les données  $\mathbf{X}$  sont alors modélisées comme de "simples" combinaisons linéaires de sources  $\mathbf{S}$  de sorte que :

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (3.3)$$

L'objet de la séparation de sources sera d'estimer des sources  $\mathbf{S}$  à partir de la connaissance des données  $\mathbf{X}$ .

Sous l'angle de la modélisation des structures des données  $\mathbf{X}$ , la matrice de mélange capture, dans ce modèle linéaire, les cohérences inter-canales présentes dans les données  $\mathbf{X}$ . De fait, en tant que signal, les sources doivent se distinguer par leur absence de structure mutuelle ; nous reviendrons sur ce point par la suite.

Dans le cadre de la séparation de sources aveugle, la matrice de mélange  $\mathbf{A}$  est une inconnue du problème et doit être estimée au même titre que les sources  $\mathbf{S}$  à partir de la connaissance des données  $\mathbf{X}$ .

### 3.1.2 Séparation par analyse en composantes décorréées

Dans un cadre probabiliste, supposons que les sources  $\{\mathbf{s}_i\}_{i=1, \dots, n}$  sont des signaux aléatoires (réalisations de variables aléatoires). Par souci de simplicité, nous les supposons

de moyenne nulle et de variance 1. En accord avec le raisonnement développé au paragraphe précédent, les sources  $\mathbf{S}$  ne doivent pas présenter de cohérence "mutuelle". Autrement dit elles doivent être "différentes". Une "mesure" simple de cohérence entre signaux est leur corrélation. De fait, des sources décohérentes (*i.e.* sans structures les unes par rapport aux autres) seront décorréliées. Résoudre le problème  $\mathbf{X} = \mathbf{A}\mathbf{S}$  s'effectuera de telle sorte que :

$$\mathbb{E}\{\mathbf{S}\mathbf{S}^T\} = \mathbf{I} \quad (3.4)$$

Au-delà de sa simplicité, cette solution est très facile à obtenir via une analyse en composantes principales (ACP). Néanmoins, même sous contrainte de décorrélation, le problème reste mal posé. En effet, la matrice de covariance des sources  $\mathbb{E}\{\mathbf{S}\mathbf{S}^T\}$  est invariante par toute transformation orthogonale  $\mathbf{S} \leftarrow \mathbf{P}\mathbf{S}$  conduisant à des solutions différentes. De fait, la décorrélation n'est pas une mesure de décohérence suffisamment forte pour conduire à une résolution correcte du problème de séparation de sources. En d'autres termes, les sources ne sont pas identifiables sous la seule hypothèse de décorrélation. Comme l'a souligné Cardoso dans (Car03), la séparation de sources nécessite d'aller au-delà de la décorrélation.

### 3.1.3 Séparation et analyse en composantes indépendantes (ACI)

Dans cette partie, nous continuons à considérer les sources comme des signaux aléatoires.

#### Generalités sur l'ACI

Le célèbre article de Comon (Com94) a ouvert la voie à l'*Analyse en Composantes Indépendantes* (ACI). Dans le cadre de l'ACI, la décohérence ou diversité entre les sources  $\mathbf{S}$  se "mesure" en terme d'indépendance statistique.

Plus formellement, supposons que les sources  $\{\mathbf{s}_i\}_{i=1,\dots,n}$  soient distribuées suivant la même densité de probabilité  $f_{\mathbf{S}}(\cdot)$ . Dans le cadre de l'ACI, la "mesure" théorique de décohérence mutuelle des sources n'est autre que leur indépendance statistique :

$$f_{\mathbf{S}}(s_1, \dots, s_n) = \prod_{i=1}^n f_{\mathbf{S}}(s_i) \quad (3.5)$$

Dans le cadre de la séparation de sources, une Analyse en Composantes Indépendantes (ACI) est équivalente à la recherche de sources indépendantes. De manière équivalente, l'ACI cherche à estimer une matrice de dé-mélange  $\mathbf{B}$  (inverse de la matrice de mélange  $\mathbf{A}$  dans le cas  $m = n$ ) telle que les sources estimées  $\tilde{\mathbf{S}} = \mathbf{B}\mathbf{A}\mathbf{S}$  soient statistiquement indépendantes. Par souci de simplicité et de clarté, nous supposons dans cette partie que la matrice de mélange  $\mathbf{A}$  est carrée et inversible ( $m = n$  et  $\det(\mathbf{A}) > 0$ ).

L'un des éléments qui ont fait de l'ACI un outil privilégié de la résolution de problèmes de SSA est que, si les sources à estimer sont effectivement indépendantes, rechercher les sources indépendantes conduit effectivement à la solution du problème : celui-ci devient

alors bien-posé. En effet, le théorème de Darmois-Linnik (Dar53; Com94) montre que si au plus une source est distribuée suivant une distribution gaussienne, et si les estimées  $\tilde{\mathbf{S}} = \mathbf{BAS}$  sont statistiquement indépendantes alors  $\mathbf{B}$  est une matrice de séparation. En d'autres termes, les sources estimées  $\tilde{\mathbf{S}}$  sont égales aux vraies sources  $\mathbf{S}$  à un facteur d'échelle (la multiplication de  $\mathbf{S}$  par une matrice diagonale, et la multiplication de  $\mathbf{A}$  par l'inverse de cette même matrice laisse le produit  $\mathbf{AS}$  invariant) et à une permutation près.

Par conséquent, si au plus une des sources est gaussienne, rétablir l'indépendance des sources conduit à l'estimation de  $\mathbf{S}$  et de la matrice de mélange  $\mathbf{A} = \mathbf{B}^{-1}$ . Le théorème de Darmois-Linnik confère de fait aux partisans de l'ACI un argument puissant : la séparation est atteignable via la "seule" indépendance des sources.

Cependant, la mise en oeuvre pratique d'une contrainte d'indépendance statistique n'est pas aussi simple qu'une décorrélation. En effet, dans le cas général, il n'existe pas de mesure exacte de l'indépendance de signaux aléatoires. C'est à partir de ce point que débute ce que nous appellerons l'ACI "pratique" ou comment essayer de mettre au point des algorithmes capables d'approcher les solutions de l'ACI. Par la suite, nous allons faire un rapide tour d'horizon des techniques d'ACI "pratiques" les plus classiques. Celles-ci n'auront de cesse de chercher : 1) à correctement modéliser les structures internes de chacune des sources, 2) à mettre au point des "mesures approchées" d'indépendance statistique entre les sources.

**Indépendance et non-gaussianité :** dans ce paragraphe, nous allons mettre l'accent sur un point particulier de l'ACI qui a été le point de départ d'un grand nombre de techniques d'ACI "pratique". Rappelons tout d'abord que dans le cadre de l'ACI et en accord avec les hypothèses d'application du théorème de Darmois-Linnik, au maximum une seule des  $n$  sources peut être distribuée suivant une distribution gaussienne. Par souci de simplicité, nous supposons qu'aucune des sources  $\{\mathbf{s}_i\}_{i=1,\dots,n}$  n'est générée suivant une distribution gaussienne. On fera également l'hypothèse que les sources ont toutes la même distribution marginale  $f_{\mathbf{S}}$ .

Considérons maintenant l'une des  $m$  observations :  $\mathbf{x}_i = \sum_{j=1}^n a_{ij}\mathbf{s}_j$

Les sources étant indépendantes, la donnée  $\mathbf{x}_i$  est une combinaison linéaire de  $n$  signaux indépendants deux-à-deux. Si on suppose que  $n$  tend en limite vers  $+\infty$ , le théorème central limite s'applique et nous indique que la distribution de  $\mathbf{x}_i$  tend vers une loi normale. Ce point motive l'adage couramment énoncé en ACI : "mélanger tend à rendre un processus gaussien". Evidemment, en général,  $n$  ne prend pas des valeurs suffisamment grandes pour qu'un tel processus soit clairement perceptible. Néanmoins, il semble raisonnable d'un point de vue strictement empirique, qu'un processus de démixage ou de séparation devrait tendre à l'effet inverse : celui de chercher à rendre les sources les moins "gaussienne" possible.

Plus formellement, une "mesure" classique de l'indépendance entre signaux  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  est leur information mutuelle  $\mathcal{I}(\mathbf{s}_1, \dots, \mathbf{s}_n)$ . Elle est définie comme la divergence de Kullback-

Leibler entre la densité jointe  $f_{\mathbf{S}}(\mathbf{s}_1, \dots, \mathbf{s}_n)$  et le produit des densités marginales :

$$\mathcal{I}(\mathbf{s}_1, \dots, \mathbf{s}_n) = \mathcal{K} \left[ f_{\mathbf{S}}(\mathbf{s}_1, \dots, \mathbf{s}_n), \prod_{i=1}^n f_{\mathbf{S}}(\mathbf{s}_i) \right] \quad (3.6)$$

$$= \int_{\mathbf{S}} f_{\mathbf{S}}(\mathbf{s}_1, \dots, \mathbf{s}_n) \log \left( \frac{f_{\mathbf{S}}(\mathbf{s}_1, \dots, \mathbf{s}_n)}{\prod_{i=1}^n f_{\mathbf{S}}(\mathbf{s}_i)} \right) d\mathbf{S} \quad (3.7)$$

De façon évidente, l'information mutuelle  $\mathcal{I}(\mathbf{s}_1, \dots, \mathbf{s}_n)$  s'annule lorsque la densité jointe  $f_{\mathbf{S}}(\mathbf{s}_1, \dots, \mathbf{s}_n)$  est factorielle ; en d'autres termes lorsque les signaux  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  sont statistiquement indépendants. L'article de Lee *et al.* (LGBS98) propose un cadre unificateur de la plupart des approches d'Analyse en Composantes Indépendantes sous l'angle de la théorie de l'information et plus particulièrement de la minimisation de l'information mutuelle.

De façon intéressante, l'expression de l'information mutuelle de l'équation (3.6) peut se décomposer sous la forme suivante (voir (Car03)) :

$$\mathcal{I}(\mathbf{s}_1, \dots, \mathbf{s}_n) = C(\mathbf{s}_1, \dots, \mathbf{s}_n) - \sum_{i=1}^n \mathcal{G}(\mathbf{s}_i) + K \quad (3.8)$$

où le terme  $C(\mathbf{S}) = \mathcal{K}[\mathcal{N}(\mathbb{E}\{\mathbf{S}\}, \mathbf{R}_{\mathbf{S}}) \mathcal{N}(\mathbb{E}\{\mathbf{S}\}, \text{diag}(\mathbf{R}_{\mathbf{S}}))]$  est la divergence de Kullback entre une approximation gaussienne des sources (de même moyenne et matrice de covariance) et une approximation gaussienne décorrélée. Ce terme quantifie donc un degré de corrélation entre les sources. Le second terme fait intervenir individuellement les divergences de Kullback entre la densité de chacune des sources avec leur approximation gaussienne  $\mathcal{G}(\mathbf{s}_i) = \mathcal{K}[f(\mathbf{s}_i), \mathcal{N}(\mathbb{E}\{\mathbf{s}_i\}, \sigma_{\mathbf{s}_i}^2)]$  (où  $\sigma_{\mathbf{s}_i}^2$  est la variance de  $\mathbf{s}_i$ ). Ce terme quantifie en quelque sorte le degré de non-Gaussianité de chacune des sources. Le terme  $K$  est une constante. Minimiser l'information mutuelle des sources est équivalent à : i) minimiser la corrélation entre les sources et ii) maximiser leur non-Gaussianité. Ce dernier point met particulièrement en lumière l'intuition que nous venons de développer.

**Point de vue algorithmique :** en pratique, toute méthode d'ACI propose l'estimation d'une matrice de "démélange" ou de "séparation"  $\mathbf{B}$  telle que les sources estimées  $\tilde{\mathbf{S}} = \mathbf{B}\mathbf{X}$  sont indépendantes. Le paragraphe précédent a introduit l'approche la plus classique en ACI : l'indépendance des sources est obtenue en annulant leur information mutuelle.

Dans l'article de Bell *et al.* (BS95), les auteurs proposent de reformuler l'information mutuelle des signaux  $\{\tilde{\mathbf{s}}_1, \dots, \tilde{\mathbf{s}}_n\}$  en introduisant leur entropie :

$$\mathcal{I}(\tilde{\mathbf{s}}_1, \dots, \tilde{\mathbf{s}}_n) = \sum_{i=1}^n \mathcal{H}(\tilde{\mathbf{s}}_i) - \mathcal{H}(\tilde{\mathbf{s}}_1, \dots, \tilde{\mathbf{s}}_n) \quad (3.9)$$

Ils montrent en particulier que maximiser l'entropie  $\mathcal{H}(\mathbf{y}_1, \dots, \mathbf{y}_n)$  de composantes telles que  $\mathbf{y}_i = g(\tilde{\mathbf{s}}_i)$  (où  $g$  est une fonction scalaire non-linéaire, monotone et inversible) permet de minimiser approximativement leur information mutuelle. En conséquence, maximiser

l'entropie des signaux  $\mathbf{y}_1, \dots, \mathbf{y}_n$  permet d'approcher la minimisation de l'information mutuelle des sources estimées  $\{\tilde{\mathbf{s}}_1, \dots, \tilde{\mathbf{s}}_n\}$ . Notons que, dans cet article, l'introduction de cette fonction non-linéaire  $g$  est motivée par l'utilisation d'un traitement par réseau de neurones pour lequel  $g$  est la fonction de transfert.

En somme, selon l'approche de Bell *et al.*, maximiser l'entropie de  $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$  :

$$\mathcal{H}(\mathbf{y}_1, \dots, \mathbf{y}_n) = \sum_{i=1}^n \mathcal{H}(\mathbf{y}_i) - \mathcal{I}(\mathbf{y}_1, \dots, \mathbf{y}_n) \quad (3.10)$$

est équivalente à : 1) minimiser l'information mutuelle des  $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$  (et de fait des sources estimées) et 2) maximiser les entropies marginales de chacun. Dans le cas où les variables sont d'amplitudes bornées (de *d.d.p.* à support compact), ce second terme est maximal lorsque les variables  $\{\mathbf{y}_i\}_{i=1, \dots, n}$  admettent des densités uniformes. En somme, la fonction non-linéaire tend à "uniformiser" la densité de probabilité des sources estimées. Plus formellement,  $g$  doit vérifier :

$$\left| \frac{\partial g(\tilde{\mathbf{s}}_i)}{\partial \tilde{\mathbf{s}}_i} \right| = p(\tilde{\mathbf{s}}_i) \quad (3.11)$$

En d'autres termes,  $g$  est la fonction de répartition de  $\mathbf{u}_i$ .

Intéressons-nous maintenant au gradient de  $\mathcal{H}(\mathbf{y}_1, \dots, \mathbf{y}_n)$  par rapport à  $\mathbf{B}$  (voir (LGBS98)) :

$$\frac{\partial \mathcal{H}(\mathbf{y}_1, \dots, \mathbf{y}_n)}{\partial \mathbf{B}} = \frac{\partial}{\partial \mathbf{B}} \log |\det(\mathbf{B})| + \frac{\partial}{\partial \mathbf{B}} \log \prod_{i=1}^n \left| \frac{\partial \mathbf{y}_i}{\partial \tilde{\mathbf{s}}_i} \right| \quad (3.12)$$

En définissant la fonction "score"  $h$  :

$$h(\tilde{\mathbf{s}}_i) = -\frac{\partial \log p(\tilde{\mathbf{s}}_i)}{\partial \tilde{\mathbf{s}}_i} \quad (3.13)$$

le gradient de  $\mathcal{H}(\mathbf{y}_1, \dots, \mathbf{y}_n)$  par rapport à  $\mathbf{B}$  s'écrit alors simplement :

$$\frac{\partial \mathcal{H}(\mathbf{y}_1, \dots, \mathbf{y}_n)}{\partial \mathbf{B}} = \mathbf{B}^{-T} - h(\tilde{\mathbf{S}}) \mathbf{X}^T \quad (3.14)$$

Rappelons que  $\mathbf{X} = \mathbf{B}^{-1} \tilde{\mathbf{S}}$ , d'où :

$$\frac{\partial \mathcal{H}(\mathbf{y}_1, \dots, \mathbf{y}_n)}{\partial \mathbf{B}} = (\mathbf{I} - h(\tilde{\mathbf{S}}) \tilde{\mathbf{S}}^T) \mathbf{B}^{-T} \quad (3.15)$$

Cette expression fait apparaître que l'annulation du gradient s'obtient en moyenne sous la condition suivante :

$$\mathbb{E} \{ h(\tilde{\mathbf{S}}) \tilde{\mathbf{S}}^T \} = \mathbf{I} \quad (3.16)$$

Cette condition de convergence est équivalente à une condition de décorrélation subtile entre les sources estimées. Le caractère non-linéaire de  $g$  fait intervenir les moments d'ordre supérieur de la densité de probabilité des sources. En conséquence, intuitivement, cette

---



condition de "décorrélation" non-linéaire fait intervenir une décohérence entre sources à des ordres supérieurs approchant ainsi l'indépendance des sources. Nous reviendrons ultérieurement sur cette condition de convergence.

L'approche fondée sur la théorie de l'information que nous venons de présenter n'est pas la seule qui a pu conduire au développement d'algorithmes d'ACI. Nous donnons ci-dessous quelques références à différentes approches dont les liens forts ont été mis en évidence par Lee *et al.* dans (LGBS98) :

- **Approche par Maximum de Vraisemblance (MV)** : l'ACI consistant à rechercher une matrice de séparation (inverse de la matrice de mélange  $\mathbf{A}$ ), elle peut être considérée plus classiquement sous l'angle de l'estimation du paramètre  $\mathbf{B}$ . De ce point de vue l'estimation de  $\mathbf{B}$  au sens du Maximum de Vraisemblance (MV) a été considérée pour résoudre les problèmes de SSA. Nous citerons en particulier les articles suivants fondés sur une approche par MV : (Car97; BL97; PGJ92).
- **Statistiques d'ordre supérieur** : en séparation de sources, comme nous l'avons souligné, la simple décorrélation des sources (décorrélation linéaire correspondant au cas  $g = \text{Id}$ ) n'est pas suffisante pour résoudre les problèmes de SSA. L'intervention de "décorrélations d'ordre supérieur" faisant intervenir les statistiques d'ordre supérieur des sources à estimer permet en particulier de dépasser la simple décorrélation pour approcher l'indépendance. De plus, en ACI, la non-gaussianité des sources joue un rôle prépondérant. Il semble de fait naturel que les statistiques d'ordre supérieur soient adaptées pour évaluer le caractère non-gaussien des sources à estimer. Nous citerons en particulier les documents suivants : (HKO01; BMCM97; Car99; JH91). Historiquement, Comon (Com94) proposa un algorithme d'ACI fondé sur une approximation de la divergence de Kullback-Leibler (à partir de son développement de Edgeworth). Cette approximation faisait d'ores et déjà intervenir les statistiques d'ordre supérieur des signaux à séparer.

A partir de ces éléments d'ordre général caractérisant les méthodes d'Analyse en Composantes Indépendantes, de nombreux développements algorithmiques ont été proposés pour l'amélioration des performances des algorithmes d'ACI. Par souci de concision, nous ne détaillerons pas l'ensemble des algorithmes qui ont pu être mis en oeuvre.

**Limitations du cadre de l'ACI** : l'élégance théorique de l'Analyse en Composantes Indépendantes est l'un de ses plus grands attraits. Néanmoins, d'un point de vue plus large, le cadre de l'ACI rencontre quelques limites :

- **Importance de la fonction de contraste** : même si cela n'apparaît qu'implicitement, l'ACI nécessite une information forte sur les sources : celle de la connaissance de leur
-

densité de probabilité. Lors des développements précédents, cette information est nécessaire pour déterminer avec précision la fonction de contraste adaptée aux sources à séparer. La connaissance de la densité de probabilité intervient ainsi implicitement dans la condition de convergence de l'équation (3.16).

Si une déviation par rapport à la fonction de contraste "optimale" est acceptable pour la séparation (voir (AC96)), le choix de cette fonction a une influence sur la qualité de convergence des algorithmes d'ACI.

Ainsi, certaines approches ont mis l'accent sur l'estimation de conjointe de la matrice de séparation  $\mathbf{B}$  et de la densité de probabilité (via une modélisation par mélange de gaussiennes (McG97) ou parmi une famille de densité paramétrées (ZK06)). Néanmoins, dans un cadre général de séparation en composantes, l'estimation ou l'adaptation de la densité de probabilité peut s'avérer plus complexe.

- **Un cadre peu adapté à la modélisation du bruit** : seuls quelques articles se sont intéressés à l'influence du bruit dans le cadre de l'ACI (voir (Dav04; ZK06)). Le cadre original de l'ACI néglige toute perturbation ou bruit dont la modélisation peut s'avérer fondamentale en pratique.

D'un point de vue théorique, Davies (Dav04) a montré que la présence de bruit additif pouvait engendrer une dégénérescence du modèle d'ACI : il n'est plus identifiable. Dans le cas où le bruit est additif et gaussien, l'utilisation de statistiques d'ordre supérieur permet l'estimation efficace de la matrice de mélange  $\mathbf{A} = \mathbf{B}^{-1}$ . Rappelons en effet que les statistiques d'ordre supérieur sont aveugles à un bruit gaussien additif. Néanmoins, cette propriété n'est plus valable lorsque le bruit n'est plus gaussien. Là encore, le cadre de l'ACI n'est pas adapté puisque appliquer la matrice de séparation  $\mathbf{B}$  aux données  $\mathbf{X}$  n'apporte pas une estimation satisfaisante des sources  $\mathbf{S}$ .

Un dernier détail d'ordre pratique doit également être ajouté : l'ACI suppose explicitement l'égalité du nombre de canaux et du nombre de sources (la matrice  $\mathbf{A}$  est carrée inversible). En conséquence, lorsque  $m > n$ , une réduction de la dimensionnalité des données (souvent opérée par une ACP) est effectuée. Une prise en compte efficace du bruit dans un contexte de séparation de sources devrait nécessairement mettre en oeuvre toute l'information mise à disposition par les données. En d'autres termes, dans le cas  $m > n$ , tous les canaux devraient être mis à profit pour estimer la matrice de mélange  $\mathbf{A}$  et les sources  $\mathbf{S}$ .

## 3.2 Séparation et analyse en composantes mutuellement parcimonieuses

### 3.2.1 Parcimonie et séparation de sources

#### Diversité pour la séparation de sources

---

Par la suite, nous proposons d'introduire une approche différente de la séparation de sources en aveugle fondée sur le concept de parcimonie. Supposons que les sources  $\{s_1, \dots, s_n\}$  (dans les exemples qui suivront nous choisirons  $n = 2$  par souci de clarté) sont telles que chacun de leurs échantillons est indépendamment et identiquement distribué selon une distribution  $f_S$  symétrique et fortement piquée à l'origine de sorte que son kurtosis est supérieur à 3. Ce type de distribution est dite leptokurtique.

Nous définissons la représentation en nuage de points de deux signaux  $s_1$  et  $s_2$  comme la représentation graphique du couple  $(s_1[k], s_2[k])$  pour  $k = 1, \dots, t$ . De l'indépendance des sources  $s_1$  et  $s_2$  et de leur distribution piquée en 0 (parcimonieuse), cette représentation fait apparaître une morphologie étoilée particulièrement caractéristique. La figure 3.1 montre la représentation en nuage de points de deux sources dont les échantillons sont *i.i.d.* suivant une distribution gaussienne généralisée de paramètre  $\theta = 0.5$  et de variance 1. Cette figure met particulièrement en lumière la forme étoilée propre à des signaux parcimonieux. L'alignement des branches de l'étoile de la représentation en nuages de points est révélatrice de l'indépendance des sources  $s_1$  et  $s_2$ . Par la suite, cette configuration particulière de la représentation en nuages de points de signaux parcimonieux et indépendants sera mise à profit pour interpréter certaines conditions de convergence des algorithmes de séparation de sources parcimonieuses.

Plus précisément, l'apport de la parcimonie pour la séparation de sources est multiple. Considérons la figure 3.1 :

- **Zone (1) :** la zone rosée (1) de cette figure illustre le principe de parcimonie énoncé précédemment pour un signal unique : une faible probabilité pour des échantillons de forte amplitude et *vice versa*.
- **Zone (2) :** sous hypothèse d'indépendance, la probabilité pour que deux échantillons  $s_1[k]$  et  $s_2[k]$  aient simultanément des amplitudes fortes est faible. De sorte qu'avec une probabilité fixée  $1 - \epsilon$  (avec  $\epsilon \ll 1$ ), si un échantillon  $s_1[k]$  est de forte amplitude alors l'échantillon  $s_2[k]$  a une amplitude bornée faible fonction de  $\epsilon$ . En effet, dans le cas simple où les échantillons de chacune des sources ont été tirés suivant la même distribution gaussienne généralisée de paramètre  $\theta$  et de variance 1 (choisi pour simplifier l'analyse), les courbes d'équiprobabilité<sup>1</sup>  $1 - \epsilon$  de la densité jointe  $f_S(s_1, s_2)$  sont les frontières de boules  $\ell_\theta$  de rayon  $r(\epsilon)$  (fonction de epsilon). Cette propriété est illustrée par la figure 3.2. Pour une valeur élevée de l'amplitude de  $s_1[k]$ , l'amplitude de  $s_2[k]$  est bornée avec une probabilité  $1 - \epsilon$  à une amplitude potentiellement faible (fonction d'epsilon) :

$$|s_1[k]|^\theta + |s_2[k]|^\theta = \tau_1 \ln\left(\frac{1}{1 - \epsilon}\right) + \tau_2 \quad (3.17)$$

où  $\tau_1$  et  $\tau_2$  sont deux constantes indépendantes de  $s_1[k]$ ,  $s_2[k]$  et  $\epsilon$ .

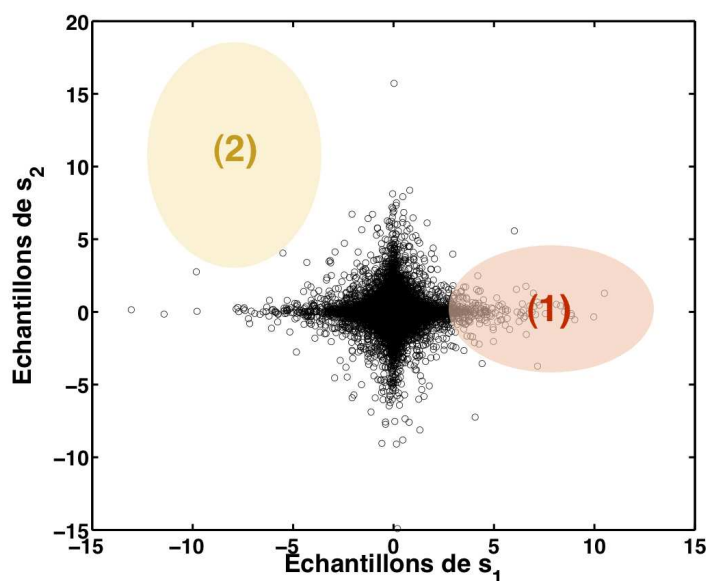
---

<sup>1</sup>On définit la courbe équiprobabilité de probabilité  $p$  pour la densité  $f$  comme la frontière du volume  $\mathcal{V}_f(p) = \{(x_1, \dots, x_n) \mid \int_{-x_1}^{x_1} \dots \int_{-x_n}^{x_n} f(x_1, \dots, x_n) dx_1 \dots dx_n \leq p\}$ .

En somme, en séparation de sources, la parcimonie renforce la diversité entre les sources. Cette diversité se manifeste par le fait que deux sources ne peuvent pas être "fortement" actives simultanément (*i.e.* pour le même échantillon  $k$ ).

Supposons maintenant que nous observons des mélanges de ces sources. La matrice de mélange est tirée aléatoirement. La figure 3.3 montre la représentation en nuage de points des sources initiales (figure de gauche) et des mélanges (figure de droite). Dans le cas du mélange, les branches de la forme étoilée que nous obtenons sont alignées suivant des droites dont les vecteurs directeurs sont définis par les colonnes de la matrice  $\mathbf{A}$  (flèches rouges). De fait, dans le cas de mélanges de sources indépendantes et parcimonieuses, les échantillons de forte amplitude sont partagés par les deux observations simultanément.

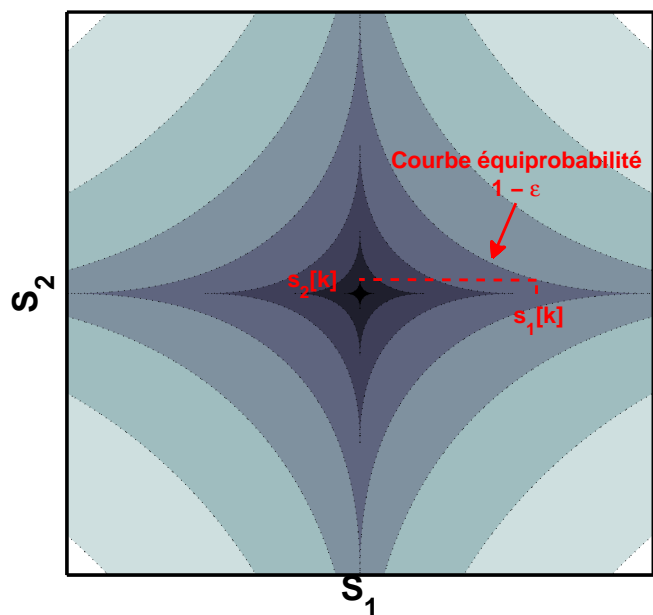
En somme, d'un point de vue intuitif, une séparation de sources parcimonieuse efficace consiste à aligner sur les axes les branches d'une étoile ... les prochains paragraphes nous éclaireront sur les moyens de formaliser ces considérations intuitives.



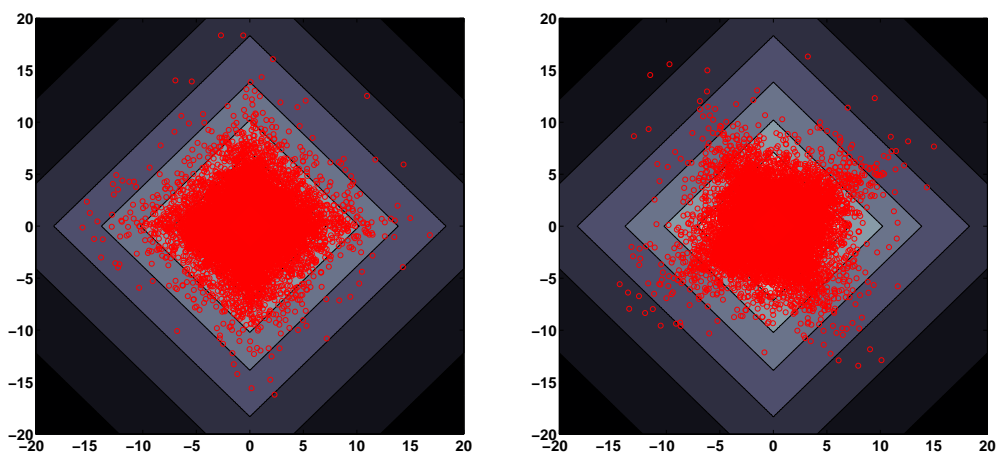
**FIG. 3.1: Un point de vue intuitif :** représentation en nuage de points des échantillons de la source  $s_1$  en fonction de la source  $s_2$ .

### Approches classiques pour la séparation de sources parcimonieuse

Le cas extrême - disjonction des supports : supposons à l'extrême que les sources sont suffisamment parcimonieuses pour qu'elles ne puissent pas être actives simultanément. Rap-



**FIG. 3.2:** Un point de vue intuitif : courbes équiprobabilité de la densité de probabilité jointe  $f_S(s_1, s_2)$ .



**FIG. 3.3:** Un point de vue intuitif - **Gauche** : représentation en nuage de points des échantillons de la source  $s_1$  en fonction de la source  $s_2$ . **Droite** : Représentation en nuage de points des échantillons de l'observation  $x_1$  en fonction de  $x_2$ .

pelons le support  $\Lambda(\mathbf{x})$  du signal  $x$  comme suit :

$$\Lambda(\mathbf{x}) = \left\{ k \mid |x[k]| > 0 \right\} \quad (3.18)$$

En somme, le support de  $\mathbf{x}$  indexe les échantillons de  $\mathbf{x}$  dont l'amplitude est non nulle (échantillons actifs). Dans le cas présent, les sources ne pouvant être actives simultanément, leurs supports sont *disjoints* :

$$\Lambda(\mathbf{s}_1) \cap \Lambda(\mathbf{s}_2) = \emptyset \quad (3.19)$$

Cette caractérisation des sources pose deux problèmes :

- **Aspect fondamental** : dans le cas de la séparation de sources à supports disjoints, la densité jointe des sources est singulière puisque son domaine de définition est restreint aux axes :

$$f_{\mathbf{S}}(\mathbf{s}_1, \mathbf{s}_2) = 0 \text{ si } s_1[k] \neq 0 \text{ et } s_2[k] \neq 0 \quad (3.20)$$

Peut-on alors parler d'indépendance des sources ? La réponse est évidemment non. En effet, considérons la probabilité conditionnelle suivante :

$$P(s_1[k] = \tau_1 \mid s_2[k] = \tau_2) = 0 \quad (3.21)$$

avec  $\tau_1, \tau_2 > 0$ . Evidemment, dans l'hypothèse de sources indépendantes, on a :

$$P(s_1[k] = \tau_1 \mid s_2[k] = \tau_2) = f_{\mathbf{S}}(s_1[k])$$

Elle est indépendante de la valeur prise par  $s_2[k]$ . Ce n'est plus le cas lorsque les sources sont supposées indépendantes.

- **Limites pratiques** : l'hypothèse de supports disjoints est de loin trop forte. En pratique, des signaux naturels n'admettent pas de représentation les rendant exactement disjoints. De plus, en présence de bruit, la notion de support devient totalement instable.

Ce type d'approche (voir (GTC05; BBZZ05; Vin07)) a essentiellement été mis en oeuvre pour la séparation de sources en contexte sous-déterminé (le nombre de sources est supérieur aux nombres d'observations). En effet, dans ce cas, sans hypothèse de disjonction, le problème inverse à résoudre est mal posé. Avec l'hypothèse de disjonction des supports, le problème revient essentiellement à segmenter l'ensemble des échantillons en sous-ensembles liés à des sources distinctes.

Il semble de fait plus réaliste de relaxer la disjonction des supports en introduisant la notion de support  $\delta$ -disjoint  $\Lambda_{\delta}(\mathbf{x})$  (BSFM07; BMFS07a) :

$$\Lambda_{\delta}(x) = \left\{ k \mid |x[k]| > \delta \|\mathbf{x}\|_{\infty} \right\} \quad (3.22)$$


---

Dans ce cas, on peut montrer que des sources indépendantes sont  $\delta$ -disjointes avec une forte probabilité (fonction de  $\delta$ ). A cet effet, supposons que les sources  $s_1$  and  $s_2$  sont *i.i.id.* suivant la même distribution laplacienne de précision  $\mu$  (gaussienne généralisée de paramètre  $\theta = 1$ ). Chaque échantillon est donc tiré suivant la loi :  $f_S(s_i[k]) = \frac{\mu}{2} \exp(-\mu|s_i[k]|)$ .

Définissons la proposition logique  $H_\delta = “|s_1[k^*]| = \|s_1\|_\infty > \delta, |s_2[k^*]| = \|s_2\|_\infty > \delta \text{ et } \forall k \neq k^*, |s_{i=1,2}[k]| \leq \delta”$ .  $H_\tau$  décrit donc des sources  $s_1$  et  $s_2$  admettant des supports strictement  $\delta$ -jointes ; dans le cas contraire, si  $H_\tau$  est sûrement fausse alors  $s_1$  et  $s_2$  ont des supports au moins  $\delta$ -disjointes. Nous définissons  $P_{|s|>\delta} = P(|s_i[k]| > \delta)$  and  $P_{|s|\leq\delta} = P(|s_i[k]| \leq \delta)$ . De l'indépendance des entrées des sources  $s_{i=1,2}$  suivant la même loi  $f_S$ , alors  $P(H_\delta)$  est telle que :

$$P(H_\delta) = t P_{|s|>\delta}^2 P_{|s|\leq\delta}^{2(t-1)} \quad (3.23)$$

Notons  $P_{|s|>\delta} = \exp(-\mu\delta)$  et  $P_{|s|\leq\delta} = 1 - \exp(-\mu\delta)$  de sorte que nous en déduisons :

$$P(H_\delta) = t \exp(-2\mu\delta) (1 - \exp(-\mu\delta))^{2(t-1)}. \quad (3.24)$$

Puisqu'en pratique  $t \gg 1$ , la proposition  $H_\delta$  est presque sûrement fausse lorsque  $t \rightarrow \infty$ . En d'autres termes, des sources indépendantes et laplaciennes sont asymptotiquement  $\delta$ -disjointes.

Inférence bayésienne - une approche universelle : la séparation de sources en aveugle est un problème inverse présentant deux "paramètres" à estimer : les sources  $\mathbf{A}$  et la matrice de mélange  $\mathbf{S}$ . Rappelons que le modèle de mélange avec terme de bruit est le suivant :

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \quad (3.25)$$

Rappelons également que l'analyse en composantes indépendantes (ACI) peut s'interpréter sous l'angle d'une estimation des paramètre  $\mathbf{A}$  et  $\mathbf{S}$  au sens du Maximum de Vraisemblance (MV - voir (Car97)) :

$$\text{Au sens de l'ACI : } \{\mathbf{A}, \mathbf{S}\} = \underset{\{\mathbf{A}, \mathbf{S}\}}{\text{Argmax}} P(\mathbf{X}|\mathbf{A}, \mathbf{S}) \quad (3.26)$$

La parcimonie des sources  $\mathbf{S}$  intervient alors comme une information *a priori* supplémentaire. L'inférence bayésien offre un cadre particulièrement bien adapté pour l'estimation de paramètres pour lesquels des informations *a priori* sont disponibles.

Dans le contexte de la séparation de sources parcimonieuses, il est fait l'hypothèse que ces dernières admettent une distribution *a priori*  $f_S$  fortement piquée à l'origine. Sans information supplémentaire sur la dépendance des sources, nous supposons de plus qu'elles sont statistiquement indépendantes :  $f_S(s_1, \dots, s_n) = \prod_{i=1}^n f_{s_i}(s_i)$ . Nous ferons également l'hypothèse que les échantillons d'une même source sont indépendamment et identiquement distribués suivant  $f_{s_i}$ . Le cadre inférentiel bayésien permet en particulier de relaxer ces différentes hypothèses simplificatrices. Il permet en particulier de pouvoir introduire une modélisation plus subtile des sources (voir (CR03; IMD06; FG06)).

Classiquement, l'estimation des paramètres  $\mathbf{A}$  et  $\mathbf{S}$  peut s'effectuer au sens d'un Maximum A Posteriori (MAP) comme suit :

$$\text{Au sens d'un MAP : } \{\mathbf{A}, \mathbf{S}\} = \underset{\{\mathbf{A}, \mathbf{S}\}}{\text{Argmax}} P(\mathbf{A}, \mathbf{S} | \mathbf{X}) \quad (3.27)$$

De la règle de Bayes (voir l'excellent ouvrage de Jaynes ([Jay03](#))) :

$$P(\mathbf{A}, \mathbf{S} | \mathbf{X}) = \frac{P(\mathbf{X} | \mathbf{A}, \mathbf{S}) P(\mathbf{A}, \mathbf{S})}{P(\mathbf{X})} \quad (3.28)$$

On déduit, sous hypothèse d'indépendance entre la matrice de mélange et les sources :

$$P(\mathbf{A}, \mathbf{S} | \mathbf{X}) \propto P(\mathbf{X} | \mathbf{A}, \mathbf{S}) P(\mathbf{A}) \prod_{i=1}^n f_{\mathbf{S}}(\mathbf{s}_i) \quad (3.29)$$

A ce point, notons qu'introduire une distribution *a priori* pour la matrice de mélange  $\mathbf{A}$  est un "jeu d'enfant", du moins formellement. Cette potentialité du cadre inférentiel bayésien sera en particulier mis à profit pour l'analyse de données hyperspectrales au Chapitre 5. Ici nous supposons qu'aucune information *a priori* ne permet de préciser  $P(\mathbf{A})$ . Nous choisirons de ce fait une distribution peu informative pour  $P(\mathbf{A})$  telle qu'une loi uniforme  $\mathcal{U}_{\mathbf{A}}$  qu'il n'est pas nécessaire d'explicitier. Ainsi, nous déduisons l'estimation de  $\mathbf{A}$  et  $\mathbf{S}$  au sens d'un MAP comme suit :

$$\{\mathbf{A}, \mathbf{S}\} = \underset{\{\mathbf{A}, \mathbf{S}\}}{\text{Argmax}} P(\mathbf{X} | \mathbf{A}, \mathbf{S}) \prod_{i=1}^n f_{\mathbf{S}}(s_i) \quad (3.30)$$

Notons que des approches alternatives sont possibles : ainsi Zhang *et al.* ([ZCA99](#)) proposent d'estimer la matrice de mélange via un MAP marginalisé (sous-entendu par rapport aux sources). Leur approche nécessite néanmoins le calcul explicite de  $\int_{\mathbf{S}} P(\mathbf{A}, \mathbf{S} | \mathbf{X}) d\mathbf{S}$  que les auteurs simplifient en recourant à une classique approximation de laplace (voir ([Rob01](#))).

Le travail pionnier de Zibulevsky ([ZP01](#)) a proposé l'utilisation de loi *a priori* parcimonieuse dans le cadre d'inférence bayésienne que nous venons d'introduire. Dans cet article, les auteurs font l'hypothèse que chaque source  $\{\mathbf{s}_i\}_{i=1, \dots, n}$  admet une distribution *a priori* parcimonieuse de sorte que :

$$f_{\mathbf{S}}(s_i[k]) \propto \exp(-\mu_i g_{\gamma}(s_i[k])) \quad (3.31)$$

où  $g_{\gamma}(s_i[k]) = |s_i[k]|^{\gamma}$  avec  $\gamma \leq 1^2$ . Dans ce cadre inférentiel, Zibulevsky propose l'estimation des paramètres  $\mathbf{A}$  et  $\mathbf{S}$  au sens du Maximum A Posteriori (MAP). De façon équivalente,

---

<sup>2</sup>Notons que l'application par échantillon de  $g_{\gamma}(\cdot)$  au vecteur  $s_i$  est équivalente au calcul de sa norme  $\ell_{\gamma}$ .



l'estimation des paramètres s'effectue par minimisation de l'opposé du logarithme de la loi *a posteriori* :

$$\{\mathbf{A}, \mathbf{S}\} = \underset{\{\mathbf{A}, \mathbf{S}\}}{\text{Argmin}} -\mathcal{L}(\mathbf{X}|\mathbf{A}, \mathbf{S}) + \sum_{i,k} \mu_i g_\gamma(s_i[k]) \quad (3.32)$$

où  $\mathcal{L}(\mathbf{X}|\mathbf{A}, \mathbf{S}) = \log(P(\mathbf{X}|\mathbf{A}, \mathbf{S}))$  n'est autre que la log-vraisemblance. En supposant que la matrice de bruit  $\mathbf{N}$  est composée d'échantillons *i.i.d.* suivant une distribution gaussienne centrée de variance 1, le problème à résoudre est le suivant :

$$\{\mathbf{A}, \mathbf{S}\} = \underset{\{\mathbf{A}, \mathbf{S}\}}{\text{Argmin}} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \sum_{i,k} \mu_i g_\gamma(s_i[k]) \quad (3.33)$$

où  $\|\mathbf{Y}\|_F^2 = \text{trace}(\mathbf{Y}\mathbf{Y}^T)$  est la norme de Frobenius de la matrice  $\mathbf{Y}$ . D'un point de vue pratique, la résolution du problème d'optimisation s'effectue au moyen d'un algorithme inspiré de la méthode de Newton : le Relative Newton Algorithm (RNA - voir (Zib03) pour de plus amples détails).

Remarque : revenons sur l'expression de l'estimateur MAP de l'équation (3.30). En supposant que la variance du terme de bruit  $\mathbf{N}$  et telle que  $\mathbb{E}\{\|\mathbf{N}\|_F^2\} \rightarrow 0$ , la vraisemblance tend vers une distribution de Dirac comme le souligne judicieusement Knuth (Knu99). En conséquence, l'estimateur du couple de paramètres  $\mathbf{A}, \mathbf{S}$  au sens du Maximum A Posteriori s'exprime comme suit :

$$\{\mathbf{A}, \mathbf{S}\} = \underset{\{\mathbf{A}, \mathbf{S}\}}{\text{Argmax}} \delta(\mathbf{X} - \mathbf{AS}) \prod_{i=1}^n f_S(s_i) \quad (3.34)$$

Au final, l'ACI classique peut-être interprétée comme un cas particulier d'une inférence au sens du MAP en limite de bruit "faible".

### 3.2.2 Analyse en composantes mutuellement parcimonieuses

#### Principe

Revenons au problème auquel nous nous sommes intéressés initialement : la décomposition de signaux multivalués dans un dictionnaire multicanal (potentiellement redondant). Rappelons que  $\mathbf{X}$  est une matrice de taille  $m \times t$  comportant  $m$  canaux ou observations contenant chacun  $t$  échantillons/entrées. Le problème de la décomposition de  $\mathbf{X}$  dans un dictionnaire multicanal  $\mathbf{\Psi} = \mathbf{\Pi} \otimes \mathbf{\Phi}$  consiste alors à résoudre le problème suivant :

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad \text{s.c. } \mathbf{X} = \mathbf{\Pi} \boldsymbol{\alpha} \mathbf{\Phi} \quad (3.35)$$

Transposons ce problème à celui de la séparation source non aveugle : la matrice de mélange  $\mathbf{A}$  est supposée connue. La restauration des sources  $\mathbf{S} = \boldsymbol{\alpha} \mathbf{\Phi}$  sous hypothèse de parcimonie

---

des sources dans  $\Phi$  s'obtient en résolvant le problème d'optimisation suivant :

$$\min_{\alpha} \|\alpha\|_1 \text{ s.c. } \mathbf{X} = \mathbf{A}\alpha\Phi \quad (3.36)$$

où la composante "spectrale" du dictionnaire multicanal  $\Psi$  est la matrice de mélange  $\mathbf{A}$ .

Plaçons-nous maintenant dans le cadre de la séparation de sources en aveugle (SSA) ; la composante spectrale du dictionnaire multicanal  $\Psi$  devient au même titre que  $\mathbf{S}$  une inconnue du problème. Dans ce contexte, nous définissons l'Analyse en Composantes Mutuellement Parcimonieuses (ACMP) comme la recherche du couple formé par la matrice de mélange  $\mathbf{A}$  et les sources  $\mathbf{S}$  telles que ces dernières sont les plus parcimonieuses possible au sens de la norme  $\|\cdot\|_1$  définie plus haut. En somme, le problème linéaire de l'équation (3.36) est remplacé par le problème bilinéaire suivant :

$$\min_{\mathbf{A}, \alpha} \|\alpha\|_1 \text{ s.c. } \mathbf{X} = \mathbf{A}\alpha\Phi \quad (3.37)$$

En l'état, plusieurs questions sont soulevées par l'introduction d'une telle analyse de données multivaluées. En particulier quel sens peut-on donner à l'Analyse en Composantes Mutuellement Parcimonieuses ? L'ACMP permet-elle (en supposant le modèle de mélange exact) de résoudre le problème de la SSA ? Existe-t-il des liens entre des approches a priori éloignées : ACMP et ACI ?

C'est à ces interrogations que les paragraphes à venir vont tenter d'apporter des éléments de réponse.

### Interprétation en terme de "meilleure base"

Jusqu'à présent, nous avons essentiellement considéré les données  $\mathbf{X}$  sous l'angle d'une collection de  $m$  observations/canaux dont chacun était formé de  $t$  entrées. Nous proposons de changer de point de vue en "transposant" la matrice  $\mathbf{X}$  : elle est formée de la concaténation de  $t$  signaux ayant  $m$  entrées (*i.e.* les colonnes de la matrice  $\mathbf{X}$ ).

Pour simplifier l'analyse et l'interprétation de l'ACMP, nous supposons que le nombre de sources est identique au nombre de canaux :  $m = n$  (*i.e.* la matrice  $\mathbf{A}$  est carrée). La matrice de mélange  $\mathbf{A}$  est supposée inversible :  $\det(\mathbf{A}) \neq 0$ . De plus, la composante temporelle ou spatiale  $\Phi$  du dictionnaire multicanal  $\Psi$  sera supposée orthogonale. Nous définissons alors les coefficients de projection de chacune des lignes de  $\mathbf{X}$  dans  $\Phi$  comme suit :

$$\mathbf{X} = \alpha_{\mathbf{X}}\Phi \quad (3.38)$$

où  $\alpha_{\mathbf{X}}$  a la même dimension que  $\mathbf{X}$ . Afin d'assurer la cohérence des notations, les coefficients de projection de lignes de  $\mathbf{S}$  dans  $\Phi$  seront notés  $\alpha_{\mathbf{S}}$ . La matrice  $\alpha_{\mathbf{S}}$  peut alors être vue comme un ensemble de  $t$  vecteurs colonne de taille  $n \times 1$  notés  $\{\alpha_{\mathbf{S}}^k\}_{k=1, \dots, t}$  tels que :

$$\forall k = 1, \dots, t; i = 1, \dots, n; \quad \alpha_{\mathbf{S}}^k[i] = \alpha_{\mathbf{S}}[i, k] \quad (3.39)$$

Le problème de l'équation (3.37) donnant corps à l'ACMP peut alors être reformulé de la sorte :

$$\min_{\mathbf{A}, \{\alpha_{\mathbf{S}}^k\}} \sum_{k=1}^t \|\alpha_{\mathbf{S}}^k\|_{\ell_1} \quad \text{s.c. } \forall k = 1, \dots, t; \alpha_{\mathbf{X}}^k = \mathbf{A} \alpha_{\mathbf{S}}^k \quad (3.40)$$

Considérons le problème individuel suivant :

$$\min_{\mathbf{A}, \alpha_{\mathbf{S}}^k} \|\alpha_{\mathbf{S}}^k\|_{\ell_1} \quad \text{s.c. } \alpha_{\mathbf{X}}^k = \mathbf{A} \alpha_{\mathbf{S}}^k \quad (3.41)$$

Le problème de l'équation (3.41) est équivalent à la recherche de la "base"  $\mathbf{A}$  dans laquelle le signal unidimensionnel  $\alpha_{\mathbf{S}}^k$  est le plus parcimonieux possible. Ce problème fait donc entièrement écho à la problématique de recherche de la meilleure base (au sens parcimonieuse - *Best Sparsifying Basis* ou BSB) que l'on retrouve en particulier chez Saito *et al.* (SB03; Sai04).

En somme, le problème de l'équation (3.40) peut s'interpréter comme la recherche de la base la plus parcimonieuse pour un ensemble de signaux  $\{\alpha_{\mathbf{X}}^k\}_{k=1, \dots, t}$  :

Résoudre un problème d'Analyse en Composantes Mutuellement Parcimonieuses c'est chercher la base la plus parcimonieuse (BSB) pour un **ensemble** de vecteurs.

Reformuler cette phrase consisterait à souligner que l'estimation de  $\mathbf{A}$  au sens de l'ACMP c'est également rechercher la "base"  $\mathbf{A}$  dans laquelle les vecteurs  $\{\alpha_{\mathbf{X}}^k\}_{k=1, \dots, t}$  sont **conjointement** les plus parcimonieux possible (au sens de la norme  $\ell_1$ ).

### ACMP et BSB

Dans ce paragraphe, nous proposons d'aller un peu plus loin dans la formalisation des liens entre ACMP et BSB pour un ensemble de signaux. A cet effet, plaçons-nous dans un cadre probabiliste plus confortable. Supposons que chaque signal  $\{\alpha_{\mathbf{S}}^k\}_{k=1, \dots, t}$  est constitué de  $n$  échantillons tirés *indépendamment* et *indentiquement* suivant une distribution laplacienne  $f_{\mathbf{S}}$  de précision  $\mu$  :

$$k = 1, \dots, t; \forall i = 1, \dots, n; \quad f_{\mathbf{S}}(\alpha_{\mathbf{S}}^k[i]) \propto \exp(-\mu |\alpha_{\mathbf{S}}^k[i]|) . \quad (3.42)$$

Clairement, les signaux  $\{\alpha_{\mathbf{S}}^k\}_{k=1, \dots, t}$  sont mutuellement indépendants et leur densité de probabilité marginale est définie comme suit :

$$f_{\mathbf{S}}(\alpha_{\mathbf{S}}^k) = \prod_{i=1}^n f_{\mathbf{S}}(\alpha_{\mathbf{S}}^k[i]) \propto \exp(-\mu \|\alpha_{\mathbf{S}}^k\|_1) . \quad (3.43)$$

De façon tout à fait intéressante, le concept de "meilleure base" au sens d'un critère de parcimonie a été formalisé par Donoho dans (Don93) dans un contexte différent. Il s'agissait pour lui de caractériser la notion de "meilleure base" au sens d'une norme  $\ell_p$  minimale pour des espaces fonctionnels particuliers. Notre problème quant à lui est restreint aux signaux

de dimension finie (plus exactement de taille  $n$ ). Néanmoins, nous allons montrer que s'inspirer du travail de Donoho sur les BSB pour des espaces fonctionnels permet d'obtenir des résultats instructifs sur l'ACMP.

Dans un cadre probabiliste (en particulier via l'équation (3.43)), l'ensemble des vecteurs d'intérêt  $\{\alpha_S^k\}_{k=1,\dots,t}$  appartient avec une forte probabilité (si  $C \gg 1/\mu$ ) à l'ensemble défini comme suit :

$$\Omega = \{\alpha \in \mathbb{R}^n \mid \|\alpha\|_1 \leq C\} \quad (3.44)$$

En d'autres termes, il s'agit des signaux de taille  $n$  contenus dans la boule  $\ell_1$  de rayon  $C$ . La probabilité avec laquelle un vecteur défini par l'équation (3.43) appartient à  $\Omega$  dépend évidemment directement de  $C$  et  $\mu$ . Cette définition est à rapprocher du concept déterministe en dimension infinie d'hyperrectangle défini dans (Don93).

Chaque vecteur  $\alpha_X^k$  observé est, par définition, la version transformée du vecteur  $\alpha_S^k$  :

$$\forall k = 1, \dots, t; \quad \alpha_X^k = \mathbf{A} \alpha_S^k \quad (3.45)$$

Par souci de simplicité, nous supposons que la matrice  $\mathbf{A}$  est orthonormale. Il s'ensuit que chacun des vecteurs  $\{\alpha_X^k\}_{k=1,\dots,t}$  appartient, avec une forte probabilité, à  $\mathbf{A}\Omega$  défini comme l'image de  $\Omega$  by  $\mathbf{A}$ .

Sur ce point le lien avec les travaux de Donoho (Don93) est évident. Il est possible de montrer que résoudre le problème de l'ACMP permet effectivement de résoudre le problème de la SSA comme le souligne la proposition suivante :

**Proposition 1** *En supposant que les entrées des sources (dans le domaine transformé  $\Phi$ )  $\alpha_S$  sont indépendamment et identiquement distribuées. Supposons que la matrice  $\mathbf{A}$  est orthonormale. Alors minimiser la norme  $\|\cdot\|_1$  permet de restaurer les sources  $\alpha_S$  :*

$$\mathbb{E} \{\|\alpha_S\|_1\} \leq \mathbb{E} \{\|\mathbf{A} \alpha_S\|_1\} . \quad (3.46)$$

La preuve est inspirée du Lemme 4 de (Don93).

Ce résultat important suscite plusieurs remarques :

- **La parcimonie d'un mélange** : de la proposition précédente, nous déduisons qu'un mélange de sources parcimonieuses statistiquement indépendantes est moins parcimonieux asymptotiquement (en terme de norme  $\|\cdot\|_1$ ). En résumé, cette propriété fait écho à la non-Gaussianité présentée comme un point de vue intuitif du fonctionnement des méthodes d'ACI.
- **L'ACMP résout les problèmes de SSA parcimonieuse** : rappelons que l'Analyse en Composantes Mutuellement Parcimonieuses estime une matrice de mélange  $\mathbf{A}$  telle que les sources aient une norme  $\|\cdot\|_1$  la plus faible possible. Cette proposition suggère que dans le cas où le modèle de sources est exact et que ces sources sont indépendantes, l'ACMP permet de restaurer les sources. L'ACMP permet donc de résoudre

les problèmes de SSA dans le cas où les signaux en jeu admettent des représentations parcimonieuses.

### Liens avec l'ACI

Remarquons que dans le cas du travail de Donoho (Don93), la recherche d'une BSB<sup>3</sup> est motivée par le souhait d'obtenir pour  $\mathbf{S}$  des processus dits "diagonaux". Dans un cadre probabiliste, cette "diagonalité" serait simplement une propriété d'indépendance mutuelle des entrées de chacune des sources  $\{\alpha_s^k\}_{k=1,\dots,t}$ . On comprend alors que cette propriété soit fortement utile tant pour la compression que l'estimation statistique puisqu'elle permet en particulier de recourir à des estimateur par seuillage individuel sur chaque entrée (Don93; DVDD98; Can05). Cette remarque met particulièrement en lumière les liens entre deux domaines apparemment fort éloignés que sont la recherche de bases optimales au sens d'un critère de parcimonie pour des signaux définis comme des éléments d'un espace fonctionnel en analyse harmonique et l'analyse en composantes indépendantes. Notons que Meyer avait donné quelques pistes intuitives de ces liens dans (Mey03).

Plus formellement, du point de vue de l'ACI, supposons que les sources  $\mathbf{S}$  sont générées à partir d'une distribution laplacienne comme cela est décrit dans l'équation (3.43). une approche très classique en ACI consiste à estimer la matrice de démélange  $\mathbf{B} = \mathbf{A}^{-1}$  au sens du Maximum de Vraisemblance. Dans ce cadre, l'estimation des sources  $\alpha_s$  (dans le domaine transformé  $\Phi$ ), s'effectue comme suit<sup>4</sup> :

$$\max_{\mathbf{B}} P(\alpha_{\mathbf{X}} | \mathbf{B}, \alpha_s) . \quad (3.47)$$

Puisque  $\alpha_{\mathbf{X}} = \mathbf{A}\alpha_s$  et les entrées de  $\alpha_s$  sont *i.i.d.* suivant une distribution laplacienne de précision, on déduit :

$$\max_{\mathbf{A}, \alpha_s} \|\alpha_s\|_1 + \tau \log |\det(\mathbf{A})| \quad \text{s.c. } \alpha_{\mathbf{X}} = \mathbf{A}\alpha_s \quad (3.48)$$

où  $\tau$  est une constante. En supposant  $\mathbf{A}$  orthogonale (hypothèse classique en ACI après blanchiment des données  $\mathbf{X}$  - (HKO01)), on obtient :

$$\max_{\mathbf{A}, \alpha_s} \|\alpha_s\|_1 \quad \text{s.c. } \alpha_{\mathbf{X}} = \mathbf{A}\alpha_s \quad (3.49)$$

Cette équation est en tout point similaire à l'équation (3.37). En somme, dans le cadre de l'ACI, l'ACMP est équivalente à une estimation de la matrice  $\mathbf{A}$  et des sources  $\mathbf{S}$  au sens

<sup>3</sup>Donoho parle de bases inconditionnelles pour des espaces fonctionnels particuliers.

<sup>4</sup>Nous passons les difficultés techniques liées à la non-dérivabilité de la norme  $\|\cdot\|_1$  en 0. Une approche simple consiste à régulariser cette norme en la définissant par  $\|\mathbf{Y}\|_1 = \lim_{\epsilon \rightarrow 0} g_{\epsilon}(\mathbf{Y})$ . La fonction  $g_{\epsilon}$  s'applique individuellement sur chacune des entrées de la matrice  $\mathbf{Y}$  de sorte que pour un scalaire  $y$  :  $g_{\epsilon}(y) = \sqrt{y^2 + \epsilon^2}$ .

du Maximum de Vraisemblance (MV). Ceci confirme les liens manifestes entre l'ACMP et l'ACI.

**Sous hypothèse d'indépendance statistique et de sources laplaciennes, en supposant  $\mathbf{A}$  orthogonale, l'Analyse en Composantes Mutuellement Parcimonieuses est équivalente à l'Analyse en Composantes Indépendantes.**

Notons que dans ce cas, la norme  $\|\cdot\|_1$  est une fonction de contraste pour l'ACI parcimonieuse (sous hypothèse de sources laplaciennes). En d'autres termes, la Proposition 1 suggère que la norme  $\|\cdot\|_1$  permet de discriminer entre les sources effectivement indépendantes (correspondant dans ce cas aux plus parcimonieuses) et leurs mélanges. En particulier l'intérêt pratique est évident. En effet, en ACI classique, la mesure d'indépendance de choix est l'information mutuelle  $\mathcal{I}$  des sources estimées. Il est souvent difficile de manipuler cette quantité dont la mise en oeuvre pratique nécessite bien souvent l'utilisation d'approximations. Dans le cadre de l'ACMP, minimiser la norme  $\|\cdot\|_1$  des sources estimées conduit à la solution de l'ACI. La norme  $\|\cdot\|_1$  étant une grandeur aisée à manipuler, l'accès aux solutions de l'ACI est donc facilitée sans recours à des "mesures" approchées d'indépendance.

La figure 3.3 illustre dans le cas  $m = n = 2$  l'intuition sous-jacente à l'utilisation de l'ACMP pour résoudre les problèmes d'ACI. La figure de droite représente la superposition de courbes d'équiprobabilité de la densité idéale gaussienne généralisée et de la réalisation de deux mélanges aléatoires de sources gaussiennes généralisées en nuage de points (cercles rouges). La projection du nuage sur l'un ou l'autre axe conduit intuitivement à une distribution moins piquée que la projection sur les axes de mélanges (colonnes de la matrice  $\mathbf{A}$ ) représentée sur la figure de gauche.

Un point de vue différent consiste à mesurer le rayon de la boule  $\ell_1$  dans laquelle l'ensemble des vecteurs  $\{\alpha_s^k\}_{k=1,\dots,t}$  s'inscrit. Dans le cas du mélange (figure de droite), le rayon de boule  $\ell_1$  nécessaire à inscrire le nuage de point du mélange est supérieur au rayon de la boule nécessaire à inscrire le nuage de point des sources. Ceci explique en partie pourquoi l'ACMP (qui, rappelons-le, recherche la solution la plus parcimonieuse) permet de résoudre le problème de l'ACI.

Ce point de vue justifie, du moins empiriquement, l'équivalence des solutions de l'ACMP et de l'ACI sous hypothèses de sources parcimonieuses (du moins laplaciennes) et indépendantes.

### Relaxation de la contrainte d'égalité

L'ajout d'une perturbation linéaire sur le modèle de source permet en particulier de modéliser : 1) un bruit instrumental dans un contexte de traitement de données réaliste et/ou 2) relaxer le modèle de source en ajoutant un terme de perturbation linéaire. Le modèle en jeu est alors le suivant :

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \quad (3.50)$$


---

En supposant de nouveau que le dictionnaire temporel/spatial  $\Phi$  est orthogonal, l'expression générale précédente s'écrit simplement dans le domaine transformé :

$$\alpha_X = A\alpha_S + \alpha_N \quad (3.51)$$

Si de plus on suppose que la perturbation est bornée :

$$\|\alpha_N\|_F < \epsilon \quad (3.52)$$

Le problème d'optimisation définissant l'ACMP à l'équation (3.37) s'écrit alors sous forme relaxée comme suit :

$$\min_{A, \alpha} \|\alpha\|_1 \text{ s.c. } \|X - A\alpha\Phi\|_F < \epsilon \quad (3.53)$$

Au-delà du cadre défini par l'ACI, l'ACMP permet de prendre en compte facilement (au moins formellement) une perturbation additive du modèle ou une composante de bruit additif dont l'utilité est cruciale dans un cadre applicatif.

Le problème sous contrainte de l'équation (3.53) est équivalent au problème sous forme de Lagrangien suivant (via une reparamétrisation adéquate  $\epsilon \rightarrow \gamma$ ) :

$$\min_{A, \alpha} \gamma \|\alpha\|_1 + \frac{1}{2} \|X - A\alpha\Phi\|_F^2 \quad (3.54)$$

où  $\gamma$  est défini de manière univoque en fonction de  $\epsilon$ . Notons que ce problème est en tout point équivalent à l'estimation des paramètres  $A$  et  $\alpha_S$  au sens d'un Maximum A Posteriori dans un cadre d'inférence bayésienne sous les hypothèses suivantes :

- **Vraisemblance** : les entrées de la perturbation  $N$  sont *i.i.d.* suivant une distribution gaussienne centrée et de variance. Notons que pour un bruit gaussien de variance  $\sigma_N$  quelconque, la condition de perturbation bornée, utilisée pour la relaxation de la contrainte d'égalité du problème de l'ACMP, est vérifiée avec une forte probabilité si l'on définit :  $\epsilon = \sqrt{t + 2\sqrt{2t}\sigma_N}$  (constituant une borne confortable pour une distribution du  $\chi_t^2$  à  $t$  degrés de liberté suivie par  $\|N\|_F^2$ ). Notons au passage que dans le cas où la matrice de covariance du bruit n'est pas proportionnelle à l'identité (*i.e.* les variances de bruit sont différentes d'un canal à l'autre avec potentiellement un certain degré de corrélation), l'adaption des équations (3.53) (3.54) s'effectue simplement en changeant la métrique  $\|\cdot\|_F$  par une version pondérée par la matrice de covariance du bruit  $\Sigma_N$  :  $\|Y\|_{F, \Sigma_N}^2 = \text{trace}(Y\Sigma_N^{-1}Y^T)$ .
  - **Lois *a priori*** : la distribution *a priori* caractérisant les sources n'est autre que la loi laplacienne déjà utilisée dans le modèle de l'équation (3.43).
-

### En dehors de l'ACI ?

*Qu'en est-il si l'indépendance des sources n'est plus vraie ?* Lorsque les sources ne sont pas indépendantes, la Proposition 1 n'est pas aussi facile à démontrer (d'ailleurs rien ne garantit qu'elle reste vraie). Deux remarques néanmoins :

- **Définition de l'ACMP** : la dépendance potentielle des sources à estimer ne prive pas l'Analyse en Composantes Mutuellement Parcimonies d'intérêt. Il est en effet toujours intéressant (du moins conceptuellement) de rechercher des composantes les plus parcimonieuses possibles (sans garantir l'existence de telles composantes). Comme nous l'avons souligné au cours des Chapitre 1 et 2, le choix du dictionnaire  $\Phi$  est lié à une information *a priori* connue sur les sources à estimer. Sous hypothèse de diversité morphologique des sources, les structures les plus fortes sont "différentes" (au sens peu cohérentes) d'une source à l'autre. De fait, même si l'indépendance statistique des sources n'est pas vraie ou n'a pas de sens, l'hypothèse de diversité morphologique permet l'utilisation de l'ACMP au moins pour la recherche des structures les plus fortes des sources en jeu.

Illustrons ce point par un exemple simple : soit  $\tau$  une variable aléatoire distribuée suivant une distribution de Bernoulli de sorte que  $\tau = 0$  ou  $1$  avec la même probabilité  $1/2$  et  $\{\tau_k\}_{k=1,\dots,t}$  des réalisations de  $\tau$ . Considérons deux sources  $s_1$  et  $s_2$  générées comme suit :

$$\forall k = 1, \dots, t; \quad s_1[k] \quad \text{est tiré suivant} \quad \begin{cases} \mathcal{N}(0, \sigma_1^2), & \text{si } \tau_k = 1 \\ \mathcal{N}(0, \sigma_2^2), & \text{sinon} \end{cases} \quad (3.55)$$

$$\forall k = 1, \dots, t; \quad s_2[k] \quad \text{est tiré suivant} \quad \begin{cases} \mathcal{N}(0, \sigma_2^2), & \text{si } \tau_k = 1 \\ \mathcal{N}(0, \sigma_1^2), & \text{sinon} \end{cases} \quad (3.56)$$

avec  $\sigma_1^2 \gg \sigma_2^2$  (ici  $\sigma_1 = 1$  et  $\sigma_2 = 10$ ). Les deux sources dépendent essentiellement de la réalisation  $\tau$ , elles ne sont donc pas indépendantes. Néanmoins comme le suggère la Figure 3.4, les sources (à gauche) tendent à être plus parcimonieuses que les mélanges (à droite). Bien que les sources soient dépendantes, l'ACMP reste valable pour ce cas particulier.

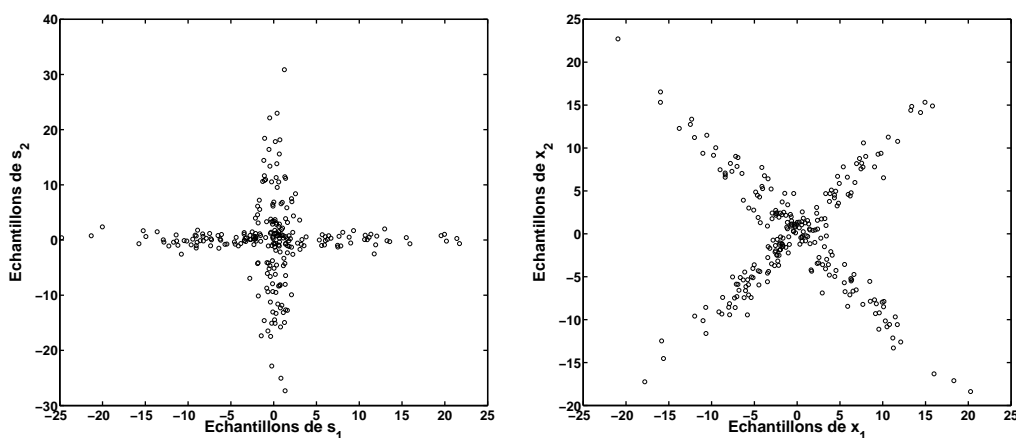
- **Modélisation de la dépendance** : nous avons vu que l'ACMP peut également être interprétée du point de vue de l'inférence bayésienne. Dans ce paradigme inférentiel, la question de la propriété d'indépendance des sources revêt une importance particulière.

En effet, si **aucune information** concernant une quelconque dépendance des sources n'est disponible, faire l'hypothèse de leur indépendance est justifiée par le désir de choisir un modèle peu informatif. A l'inverse, si la dépendance des sources est avérée, elle doit entrer dans la modélisation du problème. De fait, si un modèle de la dépendance des sources est disponible ou possible, celle-ci peut (et doit) être prise en compte (via le choix de la distribution *a priori* des sources) dans le cadre de l'ACMP.

---



*Qu'en est-il si le modèle de mélange n'est pas vérifié ?* Dans un tel cas, quelle est l'utilité de l'ACMP ? Ce type d'approche est très lié à la recherche de "meilleure base" (au sens de celle conduisant à maximiser le kurtosis des composantes estimées (Sai04), les plus parcimonieuses (Don93) ou les plus "indépendantes" (BS97; SO03)) à partir d'un ensemble significatifs de signaux que l'on peut qualifier d'apprentissage. Dès lors, la mise en oeuvre d'une Analyse en Composantes Mutuellement Parcimonieuses peut-être justifiée par le souhait d'obtenir des composantes parcimonieuses. La parcimonie pourrait permettre d'obtenir des composantes plus "interprétables". D'un point de vue plus pragmatique, la parcimonie est une propriété précieuse pour la résolution de problèmes inverses (en particulier de restauration). Ce dernier point fera l'objet du développement du Chapitre 7.



**FIG. 3.4: Sources non indépendantes - Gauche :** représentation en nuage de points des échantillons de la source  $s_1$  en fonction de la source  $s_2$ . **Droite :** Représentation en nuage de points des échantillons de l'observation  $x_1$  en fonction de  $x_2$ .

## Chapitre 4

# Diversité morphologique pour la séparation de sources en aveugle

**L**E fondement de l'Analyse en Composantes Mutuellement Parcimonieuses est l'estimation des sources ou composantes les plus parcimonieuses possibles au sens de la norme  $\|\cdot\|_1$  dans un dictionnaire  $\Phi$ . A cette fin, suivant le principe de diversité morphologique, il semble souhaitable de représenter au mieux (de façon la plus parcimonieuse) chacune des sources à séparer. Nous l'avons vu, l'utilisation d'un dictionnaire de signaux élémentaires (ou atomes) redondant offre la possibilité de représenter de façon particulièrement parcimonieuse de larges classes de signaux. Fonder une technique d'ACMP en prenant appui sur les méthodes de décomposition parcimonieuse décrites dans le Chapitre 2 semble de fait une idée à prospecter.

Plus particulièrement, l'algorithme mMCA introduit au cours du Chapitre 2 pour l'analyse de données multivaluées dans un dictionnaire multicanal redondant offre une base particulièrement appréciable à partir de laquelle bâtir une méthode pratique d'ACMP. Ce chapitre étend le cadre mMCA à la résolution du problème de séparation de sources en aveugle (SSA) et plus précisément d'ACMP.

---

## 4.1 Parcimonie et diversité morphologique pour la séparation de sources en aveugle

Rappelons que dans le cadre de la SSA, les données multicanales  $\mathbf{X}$  sont modélisées comme une combinaison linéaire de sources  $\mathbf{S}$ . Pour généraliser, les données seront agrégées d'une matrice  $\mathbf{N}$  modélisant un bruit instrumental ou des imperfections du modèle :

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \quad (4.1)$$

### 4.1.1 Extension du cadre mMCA à la séparation de sources en aveugle

En accord avec le cadre mMCA introduit précédemment, nous supposons que les sources admettent individuellement une représentation parcimonieuse dans le dictionnaire  $\Phi$ . Ce dernier est la concaténation de  $D$  bases orthonormales  $\{\Phi_i\}_{i=1,\dots,D} : \Phi = [\Phi_1^T, \dots, \Phi_D^T]^T$ . De fait, chacune des sources  $\{\mathbf{s}_i\}_{i=1,\dots,n}$  est la combinaison linéaire de  $D$  composantes morphologiques comme suit :

$$\forall i \in \{1, \dots, n\}; \quad \mathbf{s}_i = \sum_{k=1}^D \varphi_{ik} = \sum_{k=1}^D \alpha_{ik} \Phi_k \quad (4.2)$$

Dans le cas où la matrice de mélange  $\mathbf{A}$  est connue, l'estimation des sources consiste à résoudre le problème suivant :

$$\min_{\{\alpha_{ik}\}} \sum_{i=1}^n \sum_{k=1}^D \|\alpha_{ik}\|_{\ell_1} \text{ s.c. } \left\| \mathbf{X} - \sum_{i=1}^n \sum_{k=1}^D \mathbf{a}^i \alpha_{ik} \Phi_k \right\|_F < \epsilon \quad (4.3)$$

où  $\mathbf{a}^i$  est la  $i$ -ème colonne de  $\mathbf{A}$ . Le vecteur  $\alpha_{ik}$  contient alors les coefficients de la  $j$ -ème composante morphologique de la  $i$ -ème source. Ce type de problème s'inscrit dans le cadre mMCA (ici par une minimisation en norme  $\ell_1$ ) dans le cas où le dictionnaire *spectral*  $\Pi$  n'est pas une union de bases orthogonales mais l'union de "spectres" ou plus généralement de vecteurs  $\{\mathbf{a}^i\}_{i=1,\dots,n}$ . Notons que ce problème d'inversion peut être traité par l'algorithme mMCA même dans le cas sous-déterminé où  $m < n$ .

Dans le cadre de la séparation de sources en aveugle, la matrice de mélange  $\mathbf{A}$  est une inconnue du problème et doit être estimée. Nous proposons alors d'adapter le problème de l'équation (4.3) comme suit :

$$\min_{\{\mathbf{a}^i\}, \{\alpha_{ik}\}} \sum_{i=1}^n \sum_{k=1}^D \|\alpha_{ik}\|_{\ell_1} \text{ s.c. } \left\| \mathbf{X} - \sum_{i=1}^n \sum_{k=1}^D \mathbf{a}^i \alpha_{ik} \Phi_k \right\|_F < \epsilon \quad (4.4)$$

Résoudre ce problème revient à chercher une matrice de mélange  $\mathbf{A}$  telle que les sources soient mutuellement les plus parcimonieuses possibles au sens de l'ACMP. Le paragraphe suivant présente une adaptation de l'algorithme mMCA au cas de la séparation de sources en aveugle pour la résolution du problème de l'équation (4.4).

### 4.1.2 Analyse en composantes morphologiques généralisées

L'algorithme mMCA a été conçu pour résoudre le problème de l'équation (4.3) dans lequel le dictionnaire multicanal (ici  $\Psi = \mathbf{A} \otimes \Phi$ ) est supposé connu. Dans le cas de la séparation de sources en aveugle la matrice  $\mathbf{A}$  est également une inconnue du problème. Dans ce paragraphe, nous introduisons un nouvel outil d'analyse intitulé Analyse en Composantes Morphologiques Généralisées (GMCA) dont l'objectif est : i) d'estimer un ensemble de composantes morphologiques de façon similaire à mMCA, ii) d'estimer une matrice de mélange  $\mathbf{A}$  telle que ces composantes morphologiques sont les plus parcimonieuses possibles.

De façon similaire aux développements menant à l'algorithme mMCA, l'équation (4.4) peut s'écrire sous la forme d'un lagrangien :

$$\min_{\{\mathbf{a}^i\}, \{\alpha_{ik}\}} \gamma \sum_{i=1}^n \sum_{k=1}^D \|\alpha_{ik}\|_{\ell_1} + \frac{1}{2} \left\| \mathbf{X} - \sum_{i=1}^n \sum_{k=1}^D \mathbf{a}^i \alpha_{ik} \Phi_k \right\|_F^2 \quad (4.5)$$

L'équation (4.6) peut être reformulée afin de faire intervenir la matrice des sources  $\mathbf{S}$  ainsi que la matrice de mélange :

$$\{\tilde{\mathbf{A}}, \tilde{\mathbf{S}}\} = \underset{\mathbf{A}, \mathbf{S}}{\text{Argmin}} \gamma \sum_{i=1}^n \sum_{k=1}^D \|\varphi_{ik} \Phi_k^T\|_{\ell_1} + \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 \quad (4.6)$$

où chaque ligne de  $\mathbf{S}$  est une combinaison linéaire de  $D$  composantes morphologiques monocanales  $\mathbf{s}_i = \sum_{k=1}^D \varphi_{ik}$ .

La difficulté majeure soulevée par le problème d'optimisation de l'équation (4.6) est qu'elle fait intervenir deux paramètres  $\mathbf{A}$  et  $\mathbf{S}$  couplés. Il n'existe pas de solution analytique à ce problème. Par découpage du produit  $\mathbf{AS}$  en  $n \times D$  composantes morphologiques :  $\mathbf{AS} = \sum_{i,k} \mathbf{a}^i \varphi_{ik}$ , le problème à résoudre fait intervenir un ensemble de matrices de rang 1 :  $\{\mathbf{a}^i \varphi_{ik}\}_{i,k}$ . En s'inspirant des méthodes de "Block coordinate relaxation" (BCR - voir (BST98)), le problème de l'équation (4.6) peut être résolu itérativement par une suite alternée de sous-problèmes faisant intervenir séparément les couples  $\{\mathbf{a}^i, \varphi_{ik}\}$ .

On définit le  $\{i, k\}$ -ème résidu multicanal par  $\mathbf{R}_{i,k} = \mathbf{X} - \sum_{\{p,q\} \neq \{i,k\}} \mathbf{a}^p \varphi_{pq}$ . Ce terme contient la partie des données  $\mathbf{X}$  auxquelles ont été retirées les contributions des composantes morphologiques multicanales autres que  $\mathbf{a}^i \varphi_{ik}$ . L'estimation de la composante  $\varphi_{ik} = \alpha_{ik} \Phi_k$  s'effectue en supposant que  $\mathbf{A}$  et  $\varphi_{\{pq\} \neq \{ik\}}$  sont fixés. Ceci conduit à la résolution du problème suivant pour chaque composante :

$$\tilde{\varphi}_{ik} = \underset{\varphi_{ik}}{\text{Argmin}} \gamma \|\varphi_{ik} \Phi_k^T\|_{\ell_1} + \frac{1}{2} \|\mathbf{R}_{i,k} - \mathbf{a}^i \varphi_{ik}\|_F^2 \quad (4.7)$$

ou de façon équivalente,

$$\tilde{\alpha}_{ik} = \underset{\alpha_{ik}}{\text{Argmin}} \gamma \|\alpha_{ik}\|_{\ell_1} + \frac{1}{2} \|\mathbf{R}_{i,k} \Phi_k^T - \mathbf{a}^i \alpha_{ik}\|_F^2 \quad (4.8)$$

puisque  $\Phi_k$  est une matrice orthogonale. En analyse convexe, une condition nécessaire classique pour que  $\tilde{\alpha}_{ik}$  soit un minimiseur de la fonctionnelle ci-dessus est que le vecteur nul soit un élément de sa subdifférentielle à  $\tilde{\alpha}_{ik}$  (Roc70), en d'autres termes :

$$0 \in -\frac{1}{\|\mathbf{a}^i\|_2^2} \mathbf{a}^{iT} \mathbf{R}_{i,k} \Phi_k^T + \alpha_{ik} + \frac{\gamma}{\|\mathbf{a}^i\|_2^2} \partial \|\alpha_{ik}\|_{\ell_1} \quad (4.9)$$

où  $\partial \|\alpha_{ik}\|_{\ell_1}$  est le subgradient défini comme suit (en partie grâce à la séparabilité de la norme  $\ell_1$ ) :

$$\partial \|\alpha\|_1 = \left\{ \mathbf{u} \in \mathbb{R}^t \left| \begin{array}{ll} u[l] = \text{sign}(\alpha[l]), & l \in \Lambda(\alpha) \\ u[l] \in [-1, 1], & \text{sinon.} \end{array} \right. \right\}.$$

L'équation (4.9) peut être réécrite pour donner la solution suivante :

$$\tilde{\alpha}_{ik}[l] = \begin{cases} 0, & \text{if } \left| (\mathbf{a}^{iT} \mathbf{R}_{i,k} \Phi_k^T)[l] \right| \leq \gamma \\ \alpha'[l], & \text{sinon} \end{cases} \quad (4.10)$$

où  $\alpha' = \frac{1}{\|\mathbf{a}^i\|_2^2} \mathbf{a}^{iT} \mathbf{X}_{i,k} \Phi_k^T - \frac{\gamma}{\|\mathbf{a}^i\|_2^2} \text{sign}(\mathbf{a}^{iT} \mathbf{X}_{i,k} \Phi_k^T)$ . Cette solution exacte n'est rien d'autre qu'un seuillage doux. De fait, la solution pour la composante morphologique  $\varphi_{ik}$  est :

$$\tilde{\varphi}_{ik} = \mathcal{S}_{\gamma'} \left( \frac{1}{\|\mathbf{a}^i\|_2^2} \mathbf{a}^{iT} \mathbf{X}_{i,k} \Phi_k^T \right) \Phi_k \text{ avec } \gamma' = \frac{\gamma}{\|\mathbf{a}^i\|_2^2} \quad (4.11)$$

En supposant que  $\{\mathbf{a}^p\}_{p \neq i}$  et  $\mathbf{S}$  sont fixés, l'estimée de  $\mathbf{a}^i$  s'obtient comme la solution d'un problème quadratique :

$$\tilde{\mathbf{a}}^i = \frac{1}{\|\mathbf{s}_i\|_2^2} \left( \mathbf{X} - \sum_{p \neq i} \mathbf{a}^p \mathbf{s}_p \right) \mathbf{s}_i^T \quad (4.12)$$

où  $\mathbf{s}_k = \sum_{k=1}^D \varphi_{ik}$ . On peut noter que dans le cas où  $\gamma$  est fixe (Tse01, Théorème 4.1) nous permet de prouver que les solutions des problèmes alternés (4.11) et (4.12) sont définies et bornées et que chaque point d'accumulation est un point stationnaire de la fonctionnelle (4.6).

Notons qu'en pratique, chaque colonne de  $\mathbf{A}$  est de norme  $\ell_2$  unité à chaque itération. Ce point technique permet en particulier de pallier la difficulté classique liée à l'invariance d'échelle du produit  $\mathbf{AS}$  dans l'équation (4.1). L'algorithme GMCA est alors le suivant :

1. Fixer le nombre d'itérations  $P_{\max}$  et le seuil  $\gamma^{(0)}$
2. Tant que  $\gamma^{(h)}$  est supérieur à  $\gamma_{\min}$  (e.g. fonction du niveau de bruit),  
 Pour  $i = 1, \dots, n$ 
  - Pour  $k = 1, \dots, D$ 
    - Calcul du résidu  $\mathbf{r}_{ik}^{(h)}$  en supposant que  $\varphi_{\{pq\} \neq \{ik\}}, \tilde{\varphi}_{\{pq\} \neq \{ik\}}^{(h-1)}$  sont fixés :
 
$$\mathbf{r}_{ik}^{(h)} = \tilde{\mathbf{a}}^{(h-1)T} \left( \mathbf{X} - \sum_{\{p,q\} \neq \{i,k\}} \tilde{\mathbf{a}}^{p(h-1)} \tilde{\varphi}_{\{pq\}}^{(h-1)} \right)$$
    - Estimation de  $\tilde{\varphi}_{ik}^{(h)}$  par seuillage de seuil  $\gamma^{(h)}$  :
 
$$\tilde{\alpha}_{ik}^{(h)} = \mathcal{S}_{\gamma^{(h)}} \left( \mathbf{r}_{ik}^{(h)} \Phi_k^T \right)$$
    - Estimation de  $\varphi_{ik}^{(h)}$  par reconstruction à partir des coefficients sélectionnés  $\tilde{\alpha}_{ik}^{(h)}$  :
 
$$\tilde{\varphi}_{ik}^{(h)} = \tilde{\alpha}_{ik}^{(h)} \Phi_k$$
    - Mise-à-jour de  $\mathbf{a}^i$  en supposant que  $\mathbf{a}^{p \neq k^{(h)}}$  et les composantes morphologiques  $\tilde{\varphi}_{pq}^{(h)}$  sont fixées :
 
$$\tilde{\mathbf{a}}^{(h)} = \frac{1}{\|\tilde{\mathbf{s}}_i^{(h)}\|_2^2} \left( \mathbf{X} - \sum_{p \neq i} \tilde{\mathbf{a}}^{p(h-1)} \tilde{\mathbf{s}}_p^{(h)} \right) \tilde{\mathbf{s}}_i^{(h)T}$$
  - Décroissance du seuil  $\gamma^{(h)}$ .

### Le choix du dictionnaire $\Phi$

Chaque itération de l'algorithme précédent peut être divisée en deux étapes : i) décomposition parcimonieuse dans  $\mathbf{A} \otimes \Phi$  (via l'estimation des différentes composantes morphologiques multicanales) et ii) Estimation des colonnes de la matrices  $\mathbf{A}$ . En somme, il s'agit de l'algorithme mMCA introduit au Chapitre 2 auquel a été ajouté une étape d'estimation de  $\mathbf{A}$ . Dès lors, le choix du dictionnaire  $\Phi$  est fondamental. En particulier, plus la distribution des coefficients dans  $\Phi$  sera piquée (forte parcimonie), meilleure devrait être la qualité de séparation. Le choix de  $\Phi$  se fera donc d'abord par rapport aux informations *a priori* dont on dispose sur les sources. De plus, l'algorithme GMCA nécessite un ensemble de multiplications par les matrices  $\Phi_k^T$  and  $\Phi_k$ . Pour des raisons de complexité de calcul, il est préférable de faire appel à des bases ou trames ajustées pour lesquelles il existe des opérateurs implicites rapides (transformation en cosinus discrets locale ou globale, ondelettes, curvelets, etc  $\dots$ ). Avec de tels dictionnaires, les matrices  $\Phi_k^T$  et  $\Phi_k$  ne sont jamais explicitement construites mais remplacées par des opérateurs rapides implicites de synthèse et d'analyse.

### Complexité calculatoire

Notons tout d'abord que l'essentiel du temps de calcul est lié aux multiplications par les matrices  $\Phi_k^T$  et  $\Phi_k$  à chaque itération et pour chaque composante. Comme nous l'avons déjà souligné, le recours aux opérateurs implicites rapides liés à  $\Phi_k$  ou son adjoint est capital pour l'application de l'algorithme à des problèmes de grande dimension. Au cours de

l'analyse suivante, nous définissons  $V_k$  comme le coût numérique de l'opérateur  $\Phi_k$  et de son adjoint.

A chaque itération, le calcul du résidu multicanal pour chaque  $\{i, k\}$  engendre un coût de  $O(nDmt)$  opérations. Chaque étape de la boucle 'Pour' de l'algorithme GMCA calcule la corrélation du résidu avec  $\mathbf{a}^{i^T}$  nécessitant ainsi  $O(mt)$  opérations. Ensuite, la matrice  $\Phi_k^T$  est appliquée. L'opération de seuillage précède la reconstruction de la composante morphologique  $\varphi_{ik}$ . Ces opérations nécessitent  $O(2V_k + T)$  opérations. Les sources sont ensuite reconstruites avec un coût  $O(nDt)$ . La mise à jour des colonnes de la matrice  $\mathbf{A}$  est facturée  $O(mt)$  opérations.

En supposant, sans perte de généralité, que  $n \sim m \ll t$ , et  $V_k = O(t)$  ou  $O(t \log t)$  pour la plupart des transformations rapides, l'algorithme GMCA dans son ensemble nécessite  $O(P_{\max} n^2 Dt) + O(2P_{\max} n \sum_{k=1}^D V_k + nDT)$  opérations. Son coût calculatoire peut être important pour des problèmes de grande dimension. Dans la Section 4.2.1, nous présenterons une version rapide de l'algorithme GMCA permettant son application à des problèmes de grandes dimensions en économisant du temps de calcul.

### 4.1.3 Le rôle du seuillage dans l'algorithme GMCA

#### La nécessité d'un seuil décroissant

Dans l'algorithme mMCA, l'introduction du seuil décroissant est évidente : elle est liée à une sélection de coefficients dans l'espace transformé par analogie aux techniques traditionnelles de décomposition parcimonieuse gloutonnes (MP, OMP, ... etc).

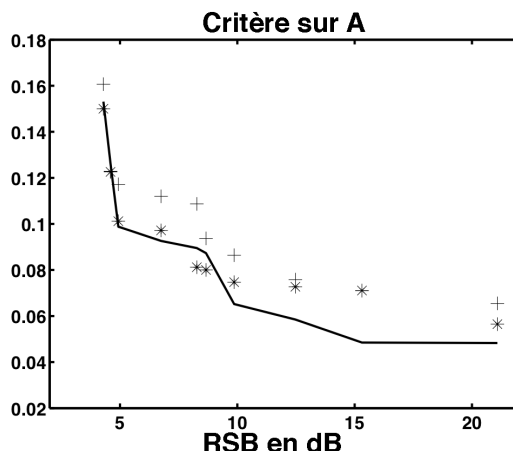
L'algorithme GMCA est, quant à lui, la combinaison d'un algorithme de décomposition parcimonieuse dans un dictionnaire multicanal  $\mathbf{A} \otimes \Phi$  et d'une procédure d'"apprentissage" d'une partie de ce dictionnaire :  $\mathbf{A}$ . Si l'utilisation du seuil décroissant semble justifié pour la phase de décomposition, qu'en est-il pour l'estimation de  $\mathbf{A}$  ? D'un point de vue intuitif, l'utilisation du seuil décroissant a un impact important sur l'estimation de la matrice  $\mathbf{A}$ . En effet, au début de l'algorithme, le seuil étant de forte valeur, la matrice  $\mathbf{A}$  est estimée à partir des coefficients sélectionnés les plus forts. Or, ces coefficients de forte amplitude jouent un rôle prépondérant : au nom du principe de diversité morphologique, un coefficient fort est, avec une forte probabilité, lié à une seule source. De fait, la matrice  $\mathbf{A}$  est, dès les premières itérations de l'algorithme, estimée à partir d'éléments de sources déjà différenciés. Alors que le seuil décroît au fur et à mesure des itérations, le nombre croissant de coefficients entrant dans l'estimation de  $\mathbf{A}$  affine son estimation. L'utilisation du seuil décroissant est donc cruciale dans l'algorithme GMCA puisqu'elle lui apporte une certaine robustesse en particulier par rapport au bruit, comme nous le verrons par la suite.

### 4.1.4 Premiers résultats expérimentaux

Nous donnons ici une première illustration de principe de l'algorithme GMCA. Considérons deux sources  $\mathbf{s}_1$  et  $\mathbf{s}_2$  admettant une représentation parcimonieuse dans le dictionnaire formé par l'union d'une base de cosinus et d'une base orthogonal d'ondelettes. Plus

---

précisément, les deux sources sont telles que leurs coefficients dans le dictionnaire  $\Phi$  sont générés suivant une distribution bernoulli-gaussienne (définie au Chapitre 2) : la probabilité pour qu'un coefficient  $\{\alpha_{1,2}[k]\}_{k=1,\dots,T}$  soit non nul est  $p = 0.01$  et l'amplitude des coefficients non nuls est distribuée suivant une distribution gaussienne centrée et de variance 1. Les deux sources sont de taille  $t = 1024$ . Un bruit additif blanc gaussien est ajouté au mélange. L'expérience suivante compare la robustesse de GMCA et de méthodes de référence vis-à-vis du bruit. La figure 4.1 montre l'évolution du critère  $C_A$  lorsque la variance



**Fig. 4.1:** Evolution du critère sur la matrice de mélange  $C_A$  lorsque la variance du bruit varie : GMCA (trait continu), EFICA, (★) ; RNA (+). **Abscisses :** rapport signal à bruit en dB. **Ordonnées :** valeur du critère sur la matrice de mélange.

du bruit sur les données augmente. GMCA est comparée à la méthode "Relative Newton Algorithm" (RNA) (Zib03) et à EFICA (ZKO06) dont l'intérêt est de modéliser la parcimonie des sources dans le cadre de l'ICA. En particulier, EFICA est une version adaptée du célèbre algorithme FastICA (Hyv99) qui a été adapté pour modéliser des sources fortement leptokurtotiques (en l'occurrence des gaussiennes généralisées de paramètre  $\theta < 2$ ). RNA et EFICA ont été appliquées sur les données transformées dans une base orthogonale d'ondelettes. La figure 4.1 montre à ce niveau que GMCA se comporte au moins aussi bien que ces deux méthodes classiques.

Notons que, dans le cas où les sources ont des morphologies différentes, chacune est formée d'une et unique composante morphologique. Dans ce cas, nos premiers résultats de séparation de sources en aveugle ont été présentés dans (BMSE06).

## 4.2 Une version rapide de GMCA

### 4.2.1 Principe

Dans le cadre de la séparation de sources en aveugle, l'analyse en composantes morphologiques généralisée (GMCA) est une extension d'un algorithme de décomposition par-



monieuse dans un dictionnaire redondant multicanal (mMCA) dans le cas où le dictionnaire multicanal (la matrice de mélange  $\mathbf{A}$ ) lui-même doit être estimé. L'algorithme que nous avons proposé nécessite un coût de calcul qui peut être prohibitif dans le cas où les données à manipuler sont de grande taille. Ce coût calculatoire provient essentiellement de l'utilisation répétée à chaque itération des opérations d'analyse et de synthèse des transformées formant le dictionnaire spatial/temporel  $\Phi$ . Dans cette partie, nous allons voir comment, à l'aide de quelques hypothèses simplificatrices, une version rapide et efficace de GMCA peut être conçue.

### Introduction : le cas où $\Phi$ est orthogonal :

Supposons que le dictionnaire jadis redondant  $\Phi$  se réduit à une simple base orthogonale. Le problème de l'équation (4.6) se simplifie alors comme suit :

$$\{\tilde{\mathbf{A}}, \tilde{\mathbf{S}}\} = \underset{\mathbf{A}, \mathbf{S}}{\text{Argmin}} \gamma \sum_{i=1}^n \|\alpha_i\|_{\ell_1} + \frac{1}{2} \|\alpha_{\mathbf{X}} - \mathbf{A}\alpha\|_F^2 \text{ avec } \mathbf{S} = \alpha_{\mathbf{S}}\Phi \quad (4.13)$$

où chaque ligne de  $\alpha_{\mathbf{X}} = \mathbf{X}\Phi^T$  contient les coefficients de projection de  $\mathbf{X}$  dans  $\Phi$ .

Dans ce cas, l'algorithme GMCA s'avère extrêmement simple puisque l'algorithme peut être entièrement développé dans le domaine des coefficients : les données sont transformées une seule fois.

### Un cas plus intéressant : le cas où $\Phi$ est redondant

Dans le paragraphe précédent, l'orthogonalité de  $\Phi$  permet le développement de l'algorithme GMCA dans le domaine des coefficients. Cette simplification permet de faire l'économie d'une phase analyse/synthèse à chaque itération. Néanmoins, nous avons souligné à plusieurs reprises que le recours à des dictionnaires redondants plutôt qu'à de simples bases orthogonales doit apporter des résultats de séparation supérieurs. Dans le cas général, la simplification valable dans le cas où  $\Phi$  est orthogonale n'est plus valable dans le cas redondant.

Dans ce paragraphe, nous allons montrer que dans le cas où les sources sont suffisamment parcimonieuses dans  $\Phi$ , il est possible d'avoir accès à des simplifications identiques à celles qui sont valables dans le cas orthogonal.

Nous considérerons par la suite que les données  $\{\mathbf{x}_i\}_{i=1,\dots,m}$  admettent une solution unique au problème de décomposition  $\ell_0$  (i.e.  $\mathcal{V}_{\ell_0}^{\Phi}(\mathbf{x}_i)$  est un singleton pour tout  $i \in \{1, \dots, m\}$ ). Nous supposons également que les sources admettent une solution unique au problème de décomposition  $\ell_0$  (i.e.  $\mathcal{V}_{\ell_0}^{\Phi}(\mathbf{s}_i)$  est un singleton pour tout  $i \in \{1, \dots, n\}$ ). Nous définissons  $\alpha_{\mathbf{X}} = [\Delta_{\Phi}(\mathbf{x}_1)^T, \dots, \Delta_{\Phi}(\mathbf{x}_m)^T]^T$  et  $\alpha_{\mathbf{S}} = [\Delta_{\Phi}(\mathbf{s}_1)^T, \dots, \Delta_{\Phi}(\mathbf{s}_n)^T]^T$ .

Quel que soit l'algorithme de décomposition, décomposer un signal dans un dictionnaire redondant est un processus **non linéaire**. La simplification valable dans le cas où  $\Phi$

est orthogonal reste vraie dans le cas redondant sous la condition suivante :

$$\forall \mathbf{A} \text{ telle que } \det(\mathbf{A}) \neq 0; \quad \mathbf{X} = \mathbf{A}\mathbf{S} \Rightarrow \boldsymbol{\alpha}_{\mathbf{X}} = \mathbf{A}\boldsymbol{\alpha}_{\mathbf{S}} \quad (4.14)$$

En d'autres termes, il s'agit du cas où la décomposition préserve tout mélange linéaire des sources.

Simplifions l'analyse en ne considérant qu'un unique mélange de signaux : soit  $\mathbf{y}$  la combinaison linéaire de  $m$  signaux ( $y$  est finalement un canal individuel dans le modèle de la SSA) :

$$\mathbf{y} = \sum_{i=1}^m v_i \mathbf{y}_i \quad (4.15)$$

En supposant que les signaux  $\{\mathbf{y}_i\}_{i=1,\dots,m}$  admettent une solution unique au problème de décomposition  $\ell_0$  (sous-entendu dans  $\Phi$ ), nous définissons  $\boldsymbol{\alpha}_i = \Delta_{\Phi}(\mathbf{y}_i)$  pour tout  $i \in \{1, \dots, m\}$ . Comme nous l'avons défini précédemment,  $\mathcal{V}_{\ell_0}^{\Phi}(\mathbf{y})$  est l'ensemble de toutes les solutions au problème  $\ell_0$  synthétisant parfaitement  $\mathbf{y}$  : pour tout  $\boldsymbol{\alpha} \in \mathcal{V}_{\ell_0}^{\Phi}(\mathbf{y})$ ;  $\mathbf{y} = \boldsymbol{\alpha}\Phi$ . Notons  $\boldsymbol{\alpha}^*$  le vecteur suivant :

$$\boldsymbol{\alpha}^* = \sum_{i=1}^m v_i \boldsymbol{\alpha}_i \quad (4.16)$$

Si  $\boldsymbol{\alpha}^*$  appartient à  $\mathcal{V}_{\ell_0}^{\Phi}(\mathbf{y})$ , une solution suffisante pour que la décomposition parcimonieuse en norme  $\ell_0$  préserve la linéarité n'est autre que l'unicité de la solution. Ainsi, dans le cas général, en accord avec les résultats décrits dans (DH01), si les conditions suivantes sont vérifiées :

$$\|\boldsymbol{\alpha}\|_0 < (\mu_{\Phi}^{-1} + 1)/2 \quad (4.17)$$

$$\|\boldsymbol{\alpha}^*\|_0 < (\mu_{\Phi}^{-1} + 1)/2 \quad (4.18)$$

alors il existe une solution unique au problème de décomposition parcimonieuse en norme  $\ell_0$ , et cette solution est  $\boldsymbol{\alpha}^*$ . Ainsi, si toutes les sources et leurs mélanges admettent une décomposition parcimonieuse unique dans  $\Phi$  au sens de la condition (4.17), alors l'opérateur de décomposition  $\Delta_{\Phi}(\cdot)$  préserve la linéarité de tout mélange.

Par la suite, nous supposons que l'opération de décomposition parcimonieuse dans  $\Phi$ ,  $\Delta_{\Phi}(\mathbf{y})$  est effectuée en utilisant l'algorithme MCA.

**Dans le cadre de la SSA :** dans le contexte de la séparation de sources en aveugle chaque observation  $\{\mathbf{x}_i\}_{i=1,\dots,m}$  est une combinaison linéaire de  $n$  sources :

$$\mathbf{x}_i = \sum_{j=1}^n a_{ij} \mathbf{s}_j \quad (4.19)$$

En accord avec le paragraphe précédent, si les sources et les observations sont suffisamment parcimonieuses dans  $\Phi$  (en particulier si elles vérifient la condition de l'équation (4.17))

---

alors le modèle linéaire est préservé après décomposition :

$$\Delta_{\Phi}(\mathbf{x}_i) = \sum_{j=1}^n a_{ij} \Delta_{\Phi}(\mathbf{s}_j) \quad (4.20)$$

Ainsi, de façon analogue au cas où  $\Phi$  est orthogonal, l'estimation de  $\mathbf{A}$  et  $\mathbf{S}$  peut être effectuée en résolvant le problème de l'équation (4.13).

### 4.2.2 Algorithme

De manière similaire à l'algorithme GMCA initial, le problème de l'équation (4.13) peut être résolu via une estimation itérative et alternée des coefficients des sources  $\alpha_{\mathbf{S}}$  et de la matrice de mélange  $\mathbf{A}$ . A cette fin, le schéma itératif suivant est utilisé pour l'estimation de  $\mathbf{A}$  and  $\alpha_{\mathbf{S}}$  :

- **Mise à jour des sources** : la matrice  $\mathbf{A}$  fixée, le problème de l'équation (4.13) possède une solution unique. Celle-ci peut être obtenue par un algorithme itératif de Landweber projeté (CW05) :

$$\tilde{\alpha}_{\mathbf{S}} = \mathcal{S}_{\gamma}(\tilde{\alpha}_{\mathbf{S}} + \mathbf{M}(\alpha_{\mathbf{X}} - \tilde{\mathbf{A}}\tilde{\alpha}_{\mathbf{S}})) \quad (4.21)$$

où  $\mathbf{M}$  une matrice de relaxation. Celle-ci influence la direction de la descente du gradient. Pour assurer la convergence de cet algorithme, la matrice de relaxation doit être telle que le rayon spectral de  $\mathbf{I} - \mathbf{M}\mathbf{A}$  est borné par 1. Le choix  $\mathbf{M} = \tilde{\mathbf{A}}^{\dagger}$  (pseudo-inverse de  $\mathbf{A}$  de rang-colonne plein), donne  $\tilde{\alpha}_{\mathbf{S}} = \mathcal{S}_{\gamma}(\tilde{\mathbf{A}}^{\dagger}\alpha_{\mathbf{X}})$ .  $\mathcal{S}_{\gamma}$  est l'opérateur de seuillage doux.

- **Mise à jour de la matrice de mélange  $\mathbf{A}$  au sens des moindres carrés** :

$$\tilde{\mathbf{A}} = \alpha_{\mathbf{X}}\tilde{\alpha}_{\mathbf{S}}^T (\tilde{\alpha}_{\mathbf{S}}\tilde{\alpha}_{\mathbf{S}}^T)^{-1} \quad (4.22)$$

Notons que le processus alterné précédent est proche de celui du *Codage parcimonieux/Apprentissage de dictionnaire* introduit dans un cadre de la restauration adaptative d'images (AEB06). Ce processus alterné conduit à l'algorithme GMCA "rapide" suivant :

1. Appliquer l'algorithme MCA à chaque ligne de  $\alpha_{\mathbf{X}}$  :  
 $\alpha_{\mathbf{X}} = [\Delta_{\Phi}(\mathbf{x}_i)]^T$
2. Fixer le nombre d'itérations  $P_{\max}$  et du seuil  $\gamma^{(0)}$
3. Tant que le seuil  $\gamma^{(h)}$  sont supérieurs à  $\gamma_{\min}$  (e.g. fonction de la variance du bruit),
  - Estimation des sources  $\alpha_{\mathbf{S}}$  à l'itération  $h$  en supposant la matrice  $\mathbf{A}$  fixée :  
 $\alpha_{\mathbf{S}}^{(h+1)} = \mathcal{S}_{\gamma^{(h)}}(\tilde{\mathbf{A}}^{\dagger(h)}\alpha_{\mathbf{X}})$  :
  - Mise-à-jour de  $\mathbf{A}$  en supposant  $\alpha_{\mathbf{S}}$  fixée :  
 $\tilde{\mathbf{A}}^{(h+1)} = \alpha_{\mathbf{X}}\tilde{\alpha}_{\mathbf{S}}^{(h)T} (\tilde{\alpha}_{\mathbf{S}}^{(h)}\tilde{\alpha}_{\mathbf{S}}^{(h)T})^{-1}$
  - Décroissance du seuil  $\gamma^{(h)}$ .

**Complexité calculatoire** La version "rapide" de GMCA nécessite l'application de l'algorithme MCA sur chaque canal. Ceci est plus rapide que la version initiale (voir Section 4.1.2). En effet, dès lors que MCA est appliqué sur chaque canal, chaque itération nécessite  $O(P_{\max} n^2 D t)$  opérations élémentaires.

### Un algorithme du point fixe

L'algorithme GMCA "rapide" se décompose en deux étapes : (i) estimer  $\mathbf{S}$  en supposant  $\mathbf{A}$  fixée, (ii) inférer la matrice  $\mathbf{A}$  en supposant les sources  $\mathbf{S}$  fixées. Dans cette version de l'algorithme GMCA, les sources sont estimées via un version seuillée de l'estimateur au sens des moindres :

$$\tilde{\alpha}_{\mathbf{S}} = \mathcal{S}_{\gamma}(\tilde{\mathbf{A}}^{\dagger} \alpha_{\mathbf{X}}) \quad (4.23)$$

où  $\tilde{\mathbf{A}}^{\dagger}$  est le pseudo-inverse de l'estimation courante de la matrice de mélange  $\tilde{\mathbf{A}}$ . L'étape suivante a pour objet l'estimation de la matrice de mélange  $\mathbf{A}$  au sens des moindres carrés :

$$\tilde{\mathbf{A}} = \alpha_{\mathbf{X}} \tilde{\alpha}_{\mathbf{S}}^T (\tilde{\alpha}_{\mathbf{S}} \tilde{\alpha}_{\mathbf{S}}^T)^{-1} \quad (4.24)$$

Définissons  $\hat{\alpha}_{\mathbf{S}} = \tilde{\mathbf{A}}^{\dagger} \alpha_{\mathbf{X}}$  telle que  $\tilde{\alpha}_{\mathbf{S}} = \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})$  et écrivons de nouveau l'équation précédente comme suit :

$$\tilde{\mathbf{A}} = \tilde{\mathbf{A}} \hat{\alpha}_{\mathbf{S}} \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T (\mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}}) \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T)^{-1} \quad (4.25)$$

De façon intéressante, cet algorithme est un algorithme du point fixe dont la condition de stationnarité est la suivante :

$$\hat{\alpha}_{\mathbf{S}} \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T = \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}}) \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T \quad (4.26)$$

### Etude de la convergence :

**Cas des sources à supports  $\gamma$ -disjoints :** Supposons dans ce paragraphe que les sources sont à supports  $\gamma$ -disjoints. Un point fixe de l'algorithme GMCA est atteint sous la condition suivante :

$$\hat{\alpha}_{\mathbf{S}} \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T = \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}}) \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T \quad (4.27)$$

De la non-linéarité de  $\mathcal{S}_{\gamma}(\cdot)$  le premier terme n'est pas symétrique en général alors que le second l'est. La condition de stationnarité est de faite une condition de *symétrisation* de la matrice  $\hat{\alpha}_{\mathbf{S}} \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T$ . Examinons en détail le cas où  $n = 2$ . Nous considérons deux sources  $\mathbf{s}_1$  et  $\mathbf{s}_2$ . Les éléments diagonaux de la condition de stationnarité sont tels que :

$$\begin{aligned} [\hat{\alpha}_{\mathbf{S}} \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T][1, 1] &= \sum_{k=1}^T \hat{\alpha}_1[k] \mathcal{S}_{\gamma}(\hat{\alpha}_1[k]) \\ &= [\mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}}) \mathcal{S}_{\gamma}(\hat{\alpha}_{\mathbf{S}})^T][1, 1] \end{aligned} \quad (4.28)$$

La condition de convergence est donc toujours vérifiée pour les éléments diagonaux. Les éléments hors-diagonaux de (4.27) sont comme suit :

$$\begin{aligned} \sum_{k \in \Lambda_\gamma(\hat{\alpha}_2)} \hat{\alpha}_1[k] \hat{\alpha}_2[k] &= \sum_{k \in \Lambda_\gamma(\hat{\alpha}_1) \cap \Lambda_\gamma(\hat{\alpha}_2)} \hat{\alpha}_1[k] \hat{\alpha}_2[k] \\ \sum_{k \in \Lambda_\gamma(\hat{\alpha}_1)} \hat{\alpha}_1[k] \hat{\alpha}_2[k] &= \sum_{k \in \Lambda_\gamma(\hat{\alpha}_1) \cap \Lambda_\gamma(\hat{\alpha}_2)} \hat{\alpha}_1[k] \hat{\alpha}_2[k] \end{aligned} \quad (4.29)$$

Définissons  $\gamma^\star$  la plus petite valeur de  $\gamma$  telle que  $\mathbf{s}_1$  et  $\mathbf{s}_2$  sont à support  $\gamma$ -disjoint. Supposons également que  $\hat{\mathbf{s}}_1$  et  $\hat{\mathbf{s}}_2$  sont  $\gamma$ -disjoints et  $\gamma^\dagger$  est le plus petit scalaire  $\gamma$  tel que ces sources sont à support  $\gamma$ -disjoint. De fait si  $\gamma > \gamma^\dagger$  alors :

$$\sum_{k \in \Lambda_\gamma(\hat{\alpha}_p) \cap \Lambda_\gamma(\hat{\alpha}_q)} \hat{\alpha}_1[k] \hat{\alpha}_2[k] = 0 \quad (4.30)$$

Intuitivement, lorsque les sources sont suffisamment parcimonieuses, les mélanges ont des  $\gamma$ -supports tels que :  $\gamma^\star \leq \gamma^\dagger$  avec égalité lorsque les sources sont parfaitement estimées. De fait, pour tout  $\gamma^\star \leq \gamma < \gamma^\dagger$  la condition de convergence n'est pas vérifiée pour les éléments hors-diagonaux (4.27) puisque :

$$\sum_{k \in \Lambda_\gamma(\hat{\alpha}_q)} \hat{\alpha}_1[k] \hat{\alpha}_2[k] \neq 0 \text{ et } \sum_{k \in \Lambda_\gamma(\hat{\alpha}_p)} \hat{\alpha}_1[k] \hat{\alpha}_2[k] \neq 0 \quad (4.31)$$

La condition de convergence est vérifiée lorsque  $\gamma^\star = \gamma^\dagger$  ; *i.e.* les sources sont correctement estimées à une erreur  $\gamma^\star$  près. Lorsque les sources sont à supports strictement disjoints ( $\gamma^\star = 0$ ), la condition de convergence est vérifiée lorsque les sources sont exactement estimées. En somme les solutions stationnaires de l'algorithme GMCA sont des solutions du problèmes de SSA.

**Point de vue statistique - cas des sources parcimonieuses indépendantes :** Dans ce paragraphe nous faisons l'hypothèse que les sources  $\mathbf{s}_1$  et  $\mathbf{s}_2$  sont des processus aléatoires. Nous supposons également que  $\alpha_1[k]$  et  $\alpha_2[k]$  sont *i.i.d.* selon une distribution symétrique et leptokurtique, unimodale en 0 et monotone croissante sur  $\mathbb{R}^-$ . Par exemple toute distribution gaussienne généralisée de paramètre  $\theta < 2$  vérifie ces conditions. La figure 4.2 représente la *pdf* jointe de deux sources parcimonieuses (à gauche) ainsi que la *pdf* jointe de deux de leurs mélanges (à droite). En prenant l'espérance des deux termes intervenant dans la condition du point fixe équation (4.26) :

$$\sum_{k \in \Lambda_\gamma(\hat{\alpha}_2)} \mathbb{E}\{\hat{\alpha}_1[k] \hat{\alpha}_2[k]\} = \sum_{k \in \Lambda_\gamma(\hat{\alpha}_1) \cap \Lambda_\gamma(\hat{\alpha}_2)} \mathbb{E}\{\hat{\alpha}_1[k] \hat{\alpha}_2[k]\} \quad (4.32)$$

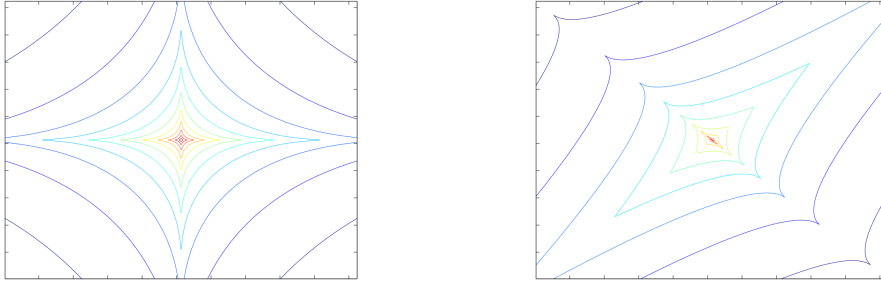
De façon symétrique,

$$\sum_{k \in \Lambda_\gamma(\hat{\alpha}_1)} \mathbb{E}\{\hat{\alpha}_1[k] \hat{\alpha}_2[k]\} = \sum_{k \in \Lambda_\gamma(\hat{\alpha}_1) \cap \Lambda_\gamma(\hat{\alpha}_2)} \mathbb{E}\{\hat{\alpha}_1[k] \hat{\alpha}_2[k]\} \quad (4.33)$$


---

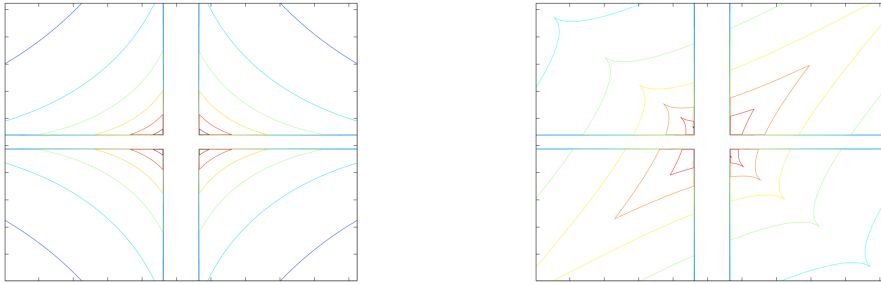
Intuitivement les sources sont correctement séparées lorsque les branches des courbes d'équiprobabilité (voir figure 4.2 sur la gauche) de la *pdf* jointe des sources sont colinéaires aux axes.

Les conditions (4.32) et (4.33) sont-elle équivalentes à la condition de séparation évoquées ci-dessus ? Notons que si les sources sont parfaitement estimées alors  $\mathbb{E}\{\mathcal{S}_\gamma(\alpha_S) \mathcal{S}_\gamma(\alpha_S)^T\}$



**Fig. 4.2:** Courbes équiprobables de la *pdf* jointe de 2 sources indépendantes sources générées selon une distribution gaussienne généralisée de paramètre  $\theta = 0,5$ . **Gauche :** *pdf* jointe de sources indépendantes. **Droite :** *pdf* jointe de 2 mélanges.

est diagonale et  $\mathbb{E}\{\alpha_S \mathcal{S}_\gamma(\alpha_S)\} = \mathbb{E}\{\mathcal{S}_\gamma(\alpha_S) \mathcal{S}_\gamma(\alpha_S)\}$ . Comme prévu, l'ensemble des solutions acceptables (les sources initiales à une permutation et un facteur d'échelle près) vérifient la condition de convergence. Supposons que  $\hat{\alpha}_1$  et  $\hat{\alpha}_2$  sont des mélanges décorré-



**Fig. 4.3:** Courbes équiprobables de la *pdf* jointe de 2 sources indépendantes sources générées selon une distribution gaussienne généralisée de paramètre  $\theta = 0,5$  après seuillage. **Gauche :** *pdf* jointe de sources indépendantes. **Droite :** *pdf* jointe de 2 mélanges après seuillage.

lés des coefficients des sources originales  $\alpha_1$  and  $\alpha_2$  ; l'application d'un seuillage sur  $\hat{\alpha}_1$  et  $\mathcal{S}_\gamma(\hat{\alpha}_2)$  (respectivement  $\hat{\alpha}_1$  et  $\mathcal{S}_\gamma(\hat{\alpha}_2)$ ) les corrèle à moins que la *pdf* jointe des sources estimées  $\alpha_1$  et  $\hat{\alpha}_2$  aient les mêmes symétries que l'opérateur de seuillage (cette propriété a été remarquée également par Field (Fie99)). La figure 4.3 apporte un point de vue intuitif sur ce point. Sur la gauche de la figure 4.3 montre la *pdf* jointe de deux sources parfaitement séparées après avoir subi l'opération de seuillage dur. Quelle que soit la valeur du seuil, les sources seuillées sont toujours décorréélées puisque leur *pdf* jointe vérifie les symétries de

l'opérateur de seuillage. A l'opposé, sur la droite de la figure 4.3, l'application de l'opérateur de seuillage entraîne une corrélation des sources non séparées après seuillage.

Pour une valeur fixe de  $\gamma$ , certains points fixes conduisent à des sources seuillées  $\hat{\alpha}_1$  and  $\hat{\alpha}_2$  décorréliées. La figure 4.3 fournit une bonne intuition à cette remarque : pour une valeur de  $\gamma$  donnée, il existe deux types de points fixes : (i) ceux qui dépendent de la valeur de  $\gamma$  (graphique de droite) (ii) ceux qui restent invariants quelle que soit la valeur de  $\gamma$  (graphique à gauche de la figure 4.3). Ce dernier type de points fixes est le seul à caractériser les solutions acceptables (*i.e.* les sources originales à une permutation et un facteur d'échelle près).

Plus formellement, les termes hors-diagonaux de la condition du point fixe des équations (4.26) font apparaître des termes de "corrélation seuillée" : la corrélation d'une source avec la version seuillée d'une autre source. Définissons la corrélation seuillée de deux signaux avec deux seuils différents  $\gamma$  et  $\gamma'$  comme suit :

$$C_{\gamma,\gamma'}(\alpha_1, \alpha_2) = \sum_{k \in \Lambda_\gamma(\alpha_1) \cap \Lambda_{\gamma'}(\alpha_2)} \mathbb{E}\{\alpha_1[k]\alpha_2[k]\} \quad (4.34)$$

$$= \mathbb{E}\{\mathcal{S}_\gamma(\alpha_1) \mathcal{S}_{\gamma'}(\alpha_2)^T\} \quad (4.35)$$

Les conditions du point fixe se s'écrivent ainsi pour les termes hors-diagonale :

$$C_{0,\gamma'}(\hat{\alpha}_1, \hat{\alpha}_2) = C_{\gamma,\gamma'}(\hat{\alpha}_1, \hat{\alpha}_2) \quad (4.36)$$

$$C_{\gamma,0}(\hat{\alpha}_1, \hat{\alpha}_2) = C_{\gamma,\gamma'}(\hat{\alpha}_1, \hat{\alpha}_2) \quad (4.37)$$

Définissons les quantités suivantes :

$$A_1(\gamma, \gamma') = C_{0,\gamma'}(\hat{\alpha}_1, \hat{\alpha}_2) - C_{\gamma,\gamma'}(\hat{\alpha}_1, \hat{\alpha}_2) \quad (4.38)$$

$$A_2(\gamma, \gamma') = C_{\gamma,0}(\hat{\alpha}_1, \hat{\alpha}_2) - C_{\gamma,\gamma'}(\hat{\alpha}_1, \hat{\alpha}_2) \quad (4.39)$$

Les conditions du point fixe (4.36) deviennent alors :

$$A_1(\gamma, \gamma') = 0 \quad (4.40)$$

$$A_2(\gamma, \gamma') = 0 \quad (4.41)$$

On remarque que :

$$A_1(\gamma, \gamma') = C_{0,\gamma'}(\hat{\alpha}_1 - \mathcal{S}_\gamma(\hat{\alpha}_1), \hat{\alpha}_2) \quad (4.42)$$

$$A_2(\gamma, \gamma') = C_{\gamma,0}(\hat{\alpha}_1, \hat{\alpha}_2 - \mathcal{S}'_\gamma(\hat{\alpha}_2)) \quad (4.43)$$

Des symétries axiales de  $(\alpha_1, \alpha_2) \rightarrow C_{0,\gamma'}(\hat{\alpha}_1 - \mathcal{S}_\gamma(\hat{\alpha}_1), \hat{\alpha}_2)$  et  $(\alpha_1, \alpha_2) \rightarrow C_{\gamma,0}(\hat{\alpha}_1, \hat{\alpha}_2 - \mathcal{S}'_\gamma(\hat{\alpha}_2))$ , on déduit que  $A_1(\gamma, \gamma')$  et  $A_2(\gamma, \gamma')$  s'annulent pour deux cas distincts :

- **La densité jointe de  $\hat{\alpha}_1$  et  $\hat{\alpha}_2$  n'est pas factorielle** :  $A_1(\gamma, \gamma')$  et  $A_2(\gamma, \gamma')$  s'annulent pour les valeurs de  $\gamma$  et  $\gamma'$  dénombrables fonctions de la *pdf* marginale des sources et de la matrice de mélange  $\mathbf{A}$ . De tels points fixes ne sont pas des solutions acceptables.
- **La densité jointe de  $\hat{\alpha}_1$  et  $\hat{\alpha}_2$  est factorielle** :  $A_1(\gamma, \gamma')$  et  $A_2(\gamma, \gamma')$  s'annulent pour toute valeur de  $\gamma$  et  $\gamma'$ . Le caractère factoriel de la densité jointe des sources estimées est équivalent à leur indépendance. Les solutions ainsi mises en valeur sont des solutions acceptables du problème de séparation.

En résumé, seuls les points fixes vérifiant les conditions (4.36) pour **toutes valeurs de seuils**  $\gamma$  et  $\gamma'$  conduisent à des solutions du problème de séparation. Ainsi, l'utilisation de seuils décroissants dans l'algorithme GMCA permet (en partie) d'éviter les points stationnaires non acceptables dépendants de  $\gamma$  et  $\gamma'$ . De fait, le seuillage décroissant permet de rendre l'algorithme GMCA plus robuste aux points fixes indésirables. Notons que dans l'algorithme GMCA présenté auparavant, seul un seuil  $\gamma$  est utilisé ; ces derniers développements tendent à justifier l'usage de seuils différents.

### Liens avec les algorithmes d'ACI

Nous avons vu au cours du Chapitre 2 que l'Analyse en Composantes Mutuellement Parcimonieuses est très proche de l'ACI en limite de bruit faible. De façon intéressante, la condition du point fixe (4.27) est en très semblable à des conditions de convergence classiques des algorithmes d'ACI. En effet, dans (Cic05), Cichocki *et al.* proposent une revue des algorithmes d'ACI les plus classiques. Ainsi, l'équation (4.27) est très semblable aux conditions de convergence des algorithmes d'ACI : la matrice de démixage  $\mathbf{B}$  est estimée sous la condition de convergence telle que la matrice  $\mathbb{E}\{h(\mathbf{B}\mathbf{X})\mathbf{B}\mathbf{X}^T\}$  est symétrique (dans cette expression  $h(\cdot)$  est la fonction *score* de l'ACI). La fonction de seuillage  $\mathcal{S}_\gamma(\cdot)$  joue un rôle similaire à celui joué par la fonction *score*  $h(\cdot)$  de l'ACI.

La condition de convergence (4.36) est atteinte lorsque la *pdf* jointe des sources estimées possède les symétries que la fonction *score*. De fait, par analogie avec l'ACI, toute fonction  $h(\cdot)$  possédant des symétries identiques pourrait être candidate au statut de fonction *score*. Ceci explique en particulier que la convergence peut être garantie pour les algorithmes d'ACI lorsque la fonction *score* n'est pas optimale (voir (LGBS98)). Pour l'ACMP, l'utilisation de l'opérateur de seuillage est double :

- **Robustesse au bruit** : l'utilisation de l'opérateur de seuillage est également lié à la volonté de sélectionner des coefficients peu perturbés par le bruit. Il garantit de fait une certaine robustesse au bruit lors du processus de séparation.
- **Robustesse aux minima locaux** : du point de vue de l'optimisation, le recours à un seuil décroissant permet d'apporter une certaine robustesse vis à vis des minima locaux.



En retour, il serait intéressant d'étudier l'ACI sous l'angle d'un problème de minimisation de la log-vraisemblance sous contrainte comme suit :

$$\min_{\mathbf{A}, \mathbf{S}} -\log P(\mathbf{X}|\mathbf{A}, \mathbf{S}) \text{ s.c. } \mathbf{X} = \mathbf{A}\mathbf{S} \quad (4.44)$$

Problème dont la résolution pourrait s'effectuer sous la forme d'un lagrangien :

$$\text{Argmin}_{\mathbf{A}, \mathbf{S}} -\gamma \log P(\mathbf{X}|\mathbf{A}, \mathbf{S}) + \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 \quad (4.45)$$

Suivant la forme de  $P$ , la solution du problème précédent en supposant  $\mathbf{A}$  fixée pourrait s'exprimer sous la forme d'un "seuillage" pour lequel  $\gamma$  jouerait le rôle du seuil. Un algorithme du type GMCA pourrait alors être développé pour résoudre ce problème.

### Prise en compte du bruit

Rappelons que cette version rapide de GMCA nécessite l'utilisation de MCA comme pré-traitement en supposant que les sources sont suffisamment parcimonieuses pour que la décomposition préserve la linéarité du mélange. En présence de bruit, de telles conditions sont plus difficiles à déterminer (voir (DET06) et (Fuc06) pour la prise en compte du bruit dans les décompositions parcimonieuses). En pratique, l'algorithme MCA est stoppé à une valeur de seuil choisie en fonction de la variance du bruit (typiquement  $3\sigma_N$  où  $\sigma_N$  est l'écart-type du bruit). MCA sélectionne alors les coefficients des données  $\mathbf{X}$  dans  $\Phi$  les plus significatifs donc peu perturbés par le bruit. Intuitivement, sous réserve de parcimonie suffisante des sources, la linéarité du mélange est, en première approximation, préservée pour les coefficients les plus forts.

## 4.3 Résultats en SSA

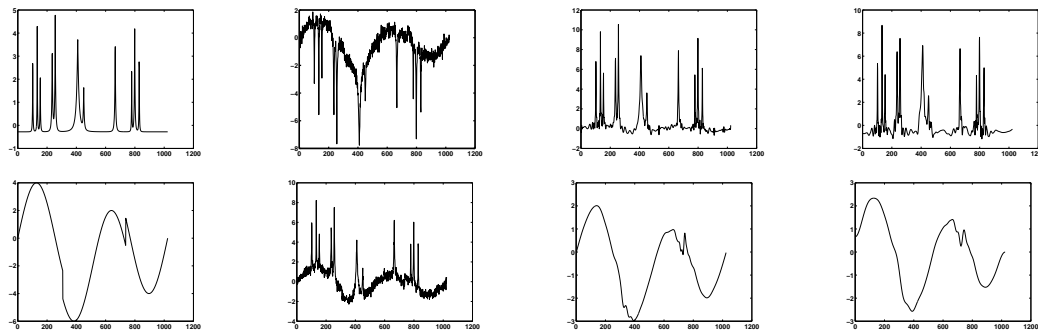
### Plus de parcimonie pour plus de robustesse

En séparation de sources en aveugle, le rôle de la diversité morphologique est double :

- **Séparabilité** : quel que soit le point de vue adopté pour la modélisation des sources, plus les sources sont parcimonieuses, plus leurs supports dans  $\Phi$  sont  $\gamma$ -disjoints avec une valeur de  $\gamma$  faible (en probabilité ou strictement). Ainsi, une plus grande parcimonie des sources devrait faciliter le processus de séparation.
  - **Robustesse au bruit et aux imperfections du modèle** : le "degré" de parcimonie des sources permet de limiter la perturbation du bruit sur les coefficients les plus significatifs. Une plus grande parcimonie des sources devrait ainsi apporter plus de robustesse au bruit.
-

Considérons le cas  $n = 2$  sources unidimensionnelles contenant  $t = 1024$  échantillons. Ces sources sont les signaux *Bump* et *HeaviSine* fournis par la boîte à outils WaveLab - voir (Wav05). La première colonne de la figure 4.4 montre les deux sources. Les mélanges sont aléatoires, et un bruit blanc gaussien centré est ajouté de sorte que  $RSB = 19\text{dB}$ . Les  $m = 2$  observations sont visibles sur la second colonne de la figure 4.4. Nous supposons que MCA préserve la linéarité du mélange (voir le choix du dictionnaire  $\Phi$  ci-dessous). Les troisième et quatrième colonnes de la figure 4.4 représentent les estimées des sources obtenues avec GMCA calculé respectivement dans les domaines suivants : (i) une simple transformée en ondelettes orthogonale (DWT) et (ii) l'union de la même transformée en ondelettes (DWT) et la tranformée en cosinus discrète (DCT). A première vue, l'algorithme GMCA conduit à des solution correctes dans les deux cas.

La figure 4.5 représente l'évolution du critère sur la matrice de mélange  $C_A$  alors que le

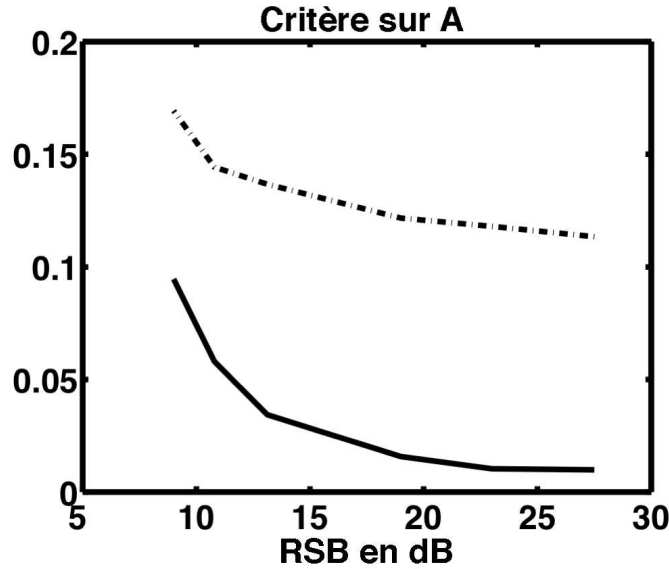


**FIG. 4.4: Le rôle de la parcimonie - première colonne :** sources originales. **Seconde colonne :** mélanges avec bruit gaussien additif ( $RSB = 19\text{dB}$ ). **Troisième colonne :** sources estimées avec GMCA en DWT. **Quatrième colonne :** sources estimées avec GMCA dans le dictionnaire redondant  $\Phi$  formé par l'union d'une DWT et d'une DCT.

$RSB$  du bruit augmente. Dans la figure 4.5, la courbe en pointillés montre le comportement de GMCA calculé avec la DWT ; la ligne continue correspond aux résultats de GMCA lorsque  $\Phi$  est l'union de la DWT et de la DCT. Globalement, quel que soit le domaine transformé, l'algorithme GMCA fournit des résultats acceptables puisque  $C_A$  est relativement faible dans les deux cas. Plus précisément, les valeurs de  $C_A$  obtenues dans le domaine MCA sont près de 5 fois plus faibles. Cette expérience met tout d'abord en lumière l'efficacité de la parcimonie et de la diversité morphologique pour la séparation aveugle de sources. Si la redondance de  $\Phi$  tend à complexifier le problème, le plus grand degré de parcimonie de sources à estimer facilite leur estimation. Ceci est ce que Donoho nomme le "blessing of dimensionality" (Don00).

### GMCA et l'ACMP

Nous proposons d'évaluer empiriquement la capacité de GMCA à obtenir la solution de l'ACMP ; *i.e.* la plus parcimonieuse en terme de norme  $\|\cdot\|_1$ . A cet effet, nous proposons l'expérience de principe suivante ; les données  $\mathbf{X}$  sont composées de 4 mélanges (figure 4.7)



**FIG. 4.5: Rôle de la parcimonie** : évolution du critère sur la matrice de mélange pour DWT-GMCA (trait discontinu) et (DWT+DCT)-GMCA (trait continu).

dont chacun est la combinaison linéaire de 4 sources (figure 4.6). La matrice de mélange a été générée aléatoirement. L'algorithme GMCA a été utilisé après application d'une transformation en ondelettes biorthogonales sur les données  $\mathbf{X}$  ; voir (Mal98). Les sources estimées sont représentées sur la figure 4.8. Notons que ces résultats ont été obtenus avec la boîte à outils GMCALab (Bob07).

Sur la figure 4.9, nous avons relevé l'évolution des valeurs de la divergence sur le degré de parcimonie des sources  $\|\tilde{\mathbf{S}}\|_1 - \|\mathbf{S}\|_1$  au cours des 500 itérations de GMCA. Cette divergence sur le degré de parcimonie caractérise la différence de normes  $\|\cdot\|_1$  entre les sources estimées et les vraies sources. Clairement, l'algorithme GMCA tend, au fur-et-à-mesure des itérations, à estimer des sources de plus en plus parcimonieuses au sens d'une plus faible norme  $\|\cdot\|_1$  des sources estimées  $\tilde{\mathbf{S}}$ . De façon intéressante, la solution vers laquelle converge GMCA a un degré de parcimonie en norme  $\|\cdot\|_1$  semblable à celui des sources originales.

### Robustesse au bruit

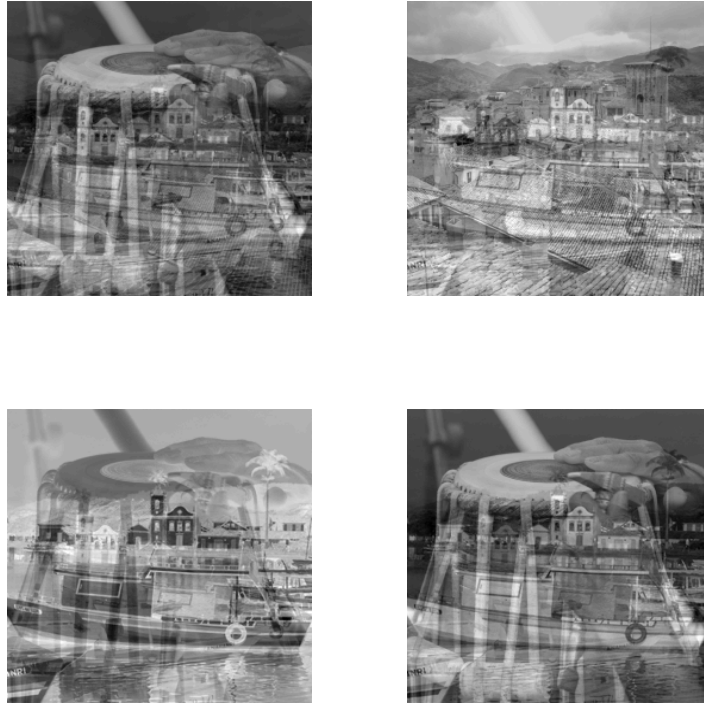
Dans ce paragraphe, nous proposons d'évaluer la robustesse de l'algorithme GMCA lorsqu'un bruit additif perturbe les données. A cet effet, nous comparons GMCA à trois méthodes de séparation de sources en aveugle :

- **JADE** : est un célèbre algorithme d'ACI fondé sur les cumulants d'ordre 4 (voir (Car99)).



**Fig. 4.6:** Sources originales de taille  $256 \times 256$ .

- **Relative Newton Algorithm** : il s'agit d'un algorithme de SSA mis au point par Zibulevsky ([Zib03](#)). Cet algorithme présente la particularité de mettre à profit la parcimonie des sources ; ces dernières sont estimées au sens d'un Maximum A Posteriori dans un cadre Bayésien. Au cours des différentes expérimentations, l'algorithme RNA est appliqué sur les données après application d'une transformation en ondelettes biorthogonales (2D-DWT).
- **EFICA** : cette méthode d'ACI (une version de FastICA - voir ([Hyv99](#))) propose l'estimation de sources gaussiennes généralisées parcimonieuses ( $\theta < 2$ ) dont le paramètre  $\theta$  est estimée de manière courante. Ceci permet d'adapter la fonction *score* de l'ACI aux sources à séparer. L'algorithme EFICA est appliqué sur les données après application d'une transformation en ondelettes biorthogonales (2D-DWT).



**FIG. 4.7:** Mélanges sans bruit.

La figure 4.10 montre les sources originales (en haut) ainsi que les 2 mélanges (en bas). La matrice de mélange  $\mathbf{A}$  est telle que  $\mathbf{x}_1 = 0.25\mathbf{s}_1 + 0.5\mathbf{s}_2 + \mathbf{n}_1$  and  $\mathbf{x}_2 = -0.75\mathbf{s}_1 + 0.5\mathbf{s}_2 + \mathbf{n}_2$  où  $\mathbf{n}_1$  et  $\mathbf{n}_2$  sont des vecteurs de bruit gaussiens blancs, centrés tel que le RSB=10dB. La matrice de covariance du bruit  $\Sigma_{\mathbf{N}}$  est donc de la forme  $\sigma_{\mathbf{N}}^2 \mathbf{I}$ .

Ces différentes méthodes sont comparées suivant les critères suivants : (i) nous étudions l'évolution de la corrélation des sources estimées avec les sources originales pour différentes valeurs de RSB ; (ii) nous relevons l'évolution du critère sur la matrice de mélange  $C_{\mathbf{A}}$ . L'algorithme GMCA a été utilisé après application d'une MCA sur chaque observation. Le dictionnaire  $\Phi$  formé par l'union d'une trame de curvelets (voir (CDDY06; Cur06)) et d'une transformation en cosinus discrète locale (LDCT). Ce choix de dictionnaire est particulièrement adapté aux images dites "naturelles" pouvant contenir des structures du type contour et texture.

Sur la figure 4.11, nous avons relevé l'évolution de la corrélation de la source 1 avec son estimée (figure de gauche) et de la source 2 avec son estimée (figure de droite) en fonction



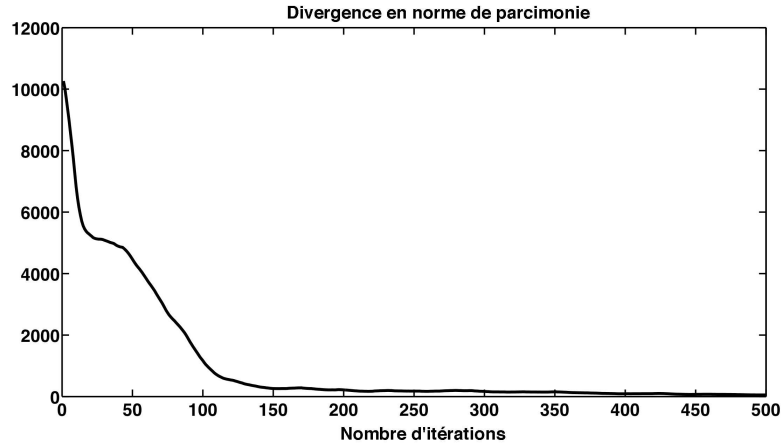
**FIG. 4.8:** Sources estimées par GMCA.

d'une variance du bruit. GMCA, RNA et EFICA conduisent à des solutions similaires en terme de corrélation ; la valeur de cette dernière est proche de 1 pour ces trois méthodes. Pour ces sources particulières, JADE fournit des résultats peu satisfaisants. Ceci peut être la conséquence d'une faible corrélation entre les deux sources à estimer.

La corrélation étant relativement sensible au bruit, le critère sur la matrice de mélange permet une meilleure discrimination des différentes méthodes. La figure 4.12 montre l'évolution du critère sur la matrice de mélange pour différentes valeurs du niveau de bruit (RSB). Alors que la corrélation sur les sources ne permettait pas de discriminer RNA et GMCA, le critère sur la matrice de mélange permet de les distinguer. Tout d'abord, il confirme l'inadéquation de JADE pour ce type de sources. Enfin, GMCA semble conduire à de meilleurs résultats puisqu'elle apporte une solution dont le critère sur la matrice de mélange est plus faible d'un facteur 10 par rapport à JADE et d'un facteur 2 par rapport à RNA et EFICA.

En résumé, nous pouvons conclure sur les points suivants :

---



**Fig. 4.9: GMCA conduit une solution plus parcimonieuse** - Abscisses : nombre d'itérations. Ordonnées : divergence de parcimonie  $\|\tilde{\mathbf{S}}\|_1 - \|\mathbf{S}\|_1$ .

- **Rôle de la parcimonie** : parmi ces méthodes, seule JADE a été appliquée sur les sources sans transformation préalable (pas de parcimonie des sources). La parcimonie des sources (et donc leur diversité morphologique) facilite effectivement la séparation.
- **Le degré de parcimonie** : grâce à la parcimonie mutuelle des sources dans le domaine transformé, les trois méthodes EFICA, RNA et GMCA fournissent des solutions satisfaisantes. Néanmoins, grâce à l'utilisation de représentations redondantes conduisant à de plus grands degrés de parcimonie, GMCA est capable de fournir de meilleurs résultats.

### Temps de calcul et grandes dimensions

Dans ce paragraphe, nous proposons d'étudier le comportement de GMCA en grandes dimensions : lorsque le nombre de sources devient important. Nous supposons que les données ne sont pas perturbées par du bruit. En effet, pour un même nombre d'échantillons  $t$ , il semble plus complexe de séparer un nombre de sources  $n$  important. Au cours des expérimentations suivantes, GMCA est appliqué sur des mélanges aléatoires de  $n = 2$  à 15 sources. Le nombre de mélanges  $m$  est égal au nombre de sources :  $m = n$ . les sources sont choisies aléatoirement parmi un ensemble de 15 images (de taille  $t = 128 \times 128$  pixels). Ces sources sont représentées sur la figure 4.13. Par souci de simplicité, GMCA est calculé après avoir appliqué une transformation en curvelets rapide (CDDY06) sur les données  $\mathbf{X}$ . Par la suite, nous analysons la convergence de GMCA en terme de critère sur la matrice de



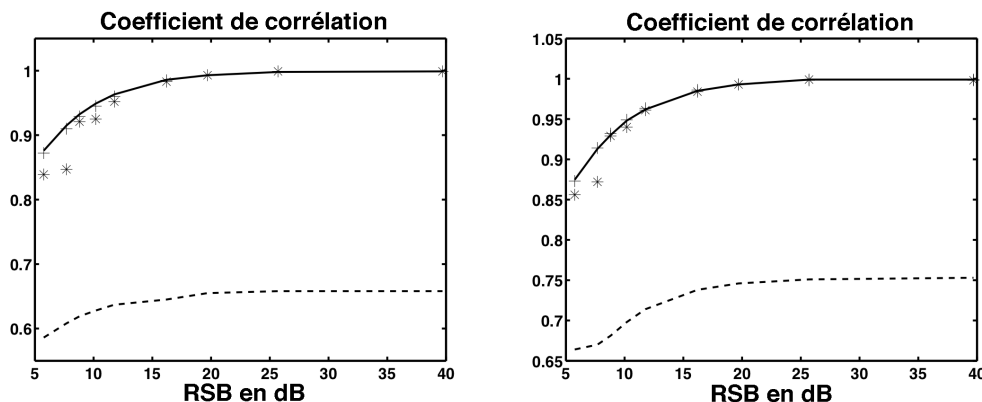
**Fig. 4.10:** **Haut :** sources originales de taille  $256 \times 256$ . **Bas :** deux mélanges. Le bruit gaussien additif est tel que le RSB est égal à 10dB.

mélange  $C_A$ . Le critère sur la matrice de mélange est normalisé comme suit :

$$\bar{C}_A = \frac{C_A}{n^2} \quad (4.46)$$

afin de ne pas dépendre du nombre de sources  $n$ . Le graphique de gauche de la figure 4.14 montre l'évolution du critère sur la matrice de mélange à mesure que le nombre d'itérations  $P_{\max}$  varie de 2 à 1000. Quel que soit le nombre de sources, le critère sur la matrice de mélange normalisé chute après 50 itérations. Lorsque  $P_{\max} > 100$ , l'algorithme GMCA tend à se stabiliser. Dès lors, augmenter le nombre d'itérations ne conduit pas à une amélioration substantielle de la qualité de la séparation. Lorsque le nombre de sources augmente, le critère sur la matrice de mélange augmente à la convergence ( pour  $P_{\max} > 100$ ). De fait, comme prévu, pour un nombre fixé d'échantillons par observation  $t$ , la complexité de la





**FIG. 4.11:** Evolution du coefficient de corrélation entre sources estimées et sources originales pour différentes valeurs de variance du bruit. Trait continu : GMCA, trait pointillé : JADE, symbole ( $\star$ ) : EFICA, symbole (+) : RNA. **Abscisses :** RSB en dB. **Ordonnées :** coefficient de corrélation entre les sources estimées et les sources originales.

séparation de sources en aveugle se dégrade avec le nombre de sources  $n$ . Fort heureusement, GMCA produit des solutions admettant un critère sur la matrice de mélange faible (inférieur à 0.025) caractérisant une séparation satisfaisante jusqu'à  $n = 15$  au moins.

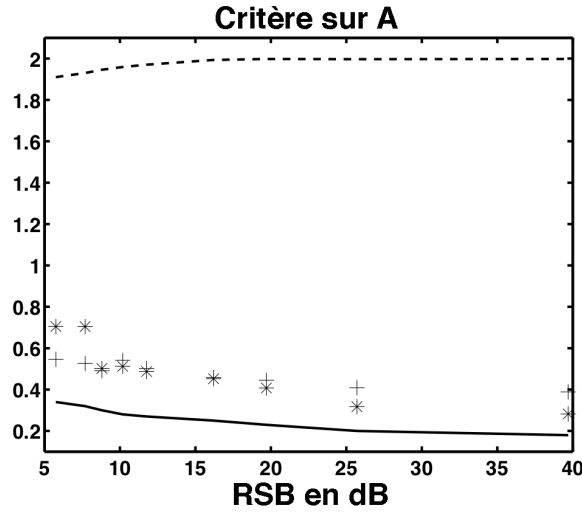
Le graphique de droite sur la figure 4.14 montre le temps de calcul <sup>1</sup> de GMCA lorsque le nombre de sources  $n$  varie. Rappelons ici que l'algorithme "rapide" GMCA se décompose en deux étapes : i) calcul de  $\alpha_{\mathbf{X}}$  par application d'une transformation, ii) estimation de la matrice  $\mathbf{A}$  et des sources  $\alpha_{\mathbf{S}}$ . Evidemment, ce graphique tend à montrer que le coût de calcul augmente avec le nombre de sources. Notons que pour  $m = n$ , le coût calculatoire de l'étape i) est proportionnel au nombre d'observations (ici  $m = n$ ) et indépendant du nombre d'itérations  $P_{\max}$  et devient donc négligeable par rapport à ii) pour de fortes valeurs de  $P_{\max}$ . Le temps de calcul de GMCA tend alors à être proportionnel à  $P_{\max}$ .

## 4.4 Estimation du nombre de sources

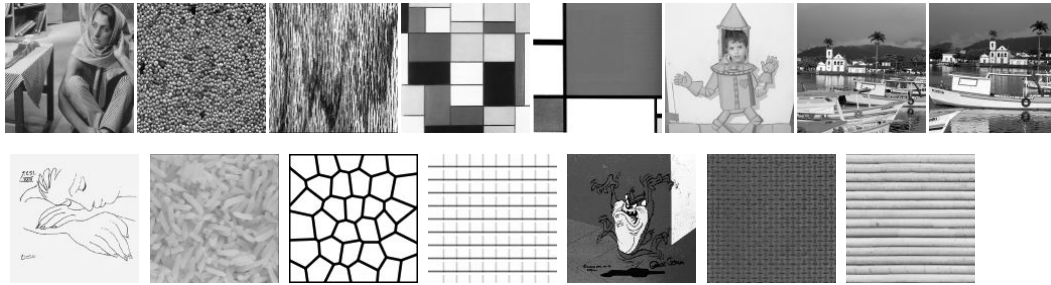
Jusqu'à présent et d'une manière générale en séparation de sources en aveugle, le nombre de sources  $n$  est supposé connu. En pratique, il est rare que  $n$  soit effectivement un paramètre connu par avance. En supposant le modèle de mélange vérifié, le nombre de sources est la dimension du sous-espace de  $\mathbb{R}^m$  généré par les colonnes de  $\mathbf{X}$ . En effet, le rang (ligne) de la matrice  $\mathbf{A}$  est exactement égal à  $n$ . Une mauvaise estimation du nombre de sources  $n$  peut apporter un certain nombre de difficultés :

- **Sous-estimation :** au sein de l'algorithme GMCA, sous-estimer le nombre de sources conduira à l'estimation de solutions qui seront des combinaisons linéaires des sources originales. La solution risque donc d'être sous-optimale en terme de parcimonie des

<sup>1</sup>Les expérimentations ont été menées sur un Apple PowerMac G5 - 2Ghz sous IDL.



**FIG. 4.12:** Evolution du critère sur la matrice de mélange  $C_A$  pour différentes valeurs de variance du bruit. Trait continu : GMCA, trait pointillé : JADE, symbole (★) : EFICA, symbole (+) : RNA. **Abscisses :** RSB en dB. **Ordonnées :** critère sur la matrice de mélange  $C_A$ .



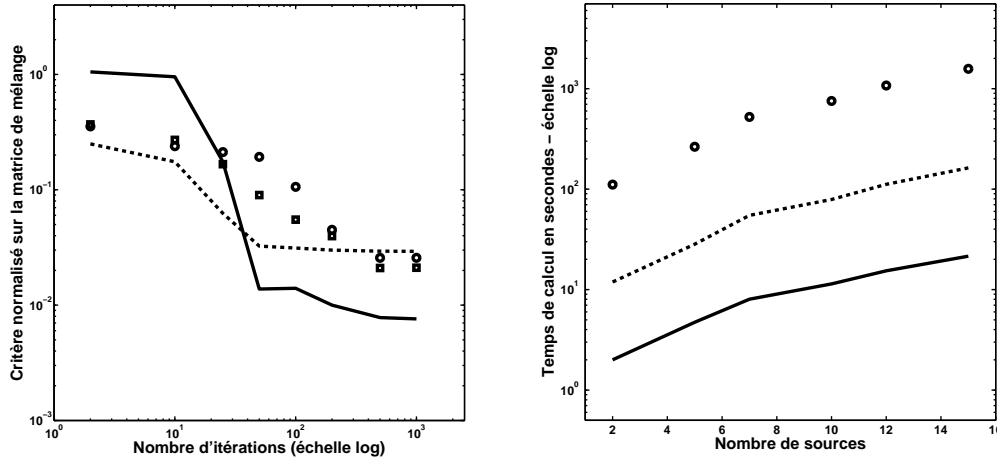
**FIG. 4.13:** Ensemble des 15 sources utilisées pour l'évaluation du comportement de GMCA en fonction du nombre de sources.

sources estimées.

- **Sur-estimation :** en cas de sur-estimation, l'algorithme GMCA aura à faire face à des matrices de mélanges estimées qui ne seront plus de rang plein. Le problème d'optimisation impliquera des matrices mal-conditionnées.

L'estimation des sources est un point crucial. Néanmoins, ajouter  $n$  au cortège des inconnues à estimer complexifie le problème de séparation.

A notre connaissance, rares sont les travaux qui ont proposé des solutions à l'estimation du nombre de sources  $n$ . Récemment, les auteurs de (Bal07) ont proposé d'aborder ce problème sous l'angle du "minimum description length" (MDL). Une approche classique consisterait à recourir à des critères usuels de sélection de modèle tel que AIC (Aka70) ou BIC (Sch78). De tels critères proposent un compromis entre la complexité du modèle (ici le nombre de



**FIG. 4.14:** *Gauche* : évolution du critère normalisé sur la matrice de mélange lorsque le nombre d'itérations  $P_{\max}$  augmente. **Abscisses** : nombre d'itérations. **Ordonnées** : critère normalisé sur la matrice de mélange. Le nombre de sources varie comme suit : trait continu :  $n = 2$ , trait discontinu :  $n = 5$ , symbole ( $\square$ ) :  $n = 10$ , symbole ( $\circ$ ) :  $n = 15$ . *Droite* : temps de calcul lorsque le nombre de sources croît. **Abscisses** : nombre de sources. **Ordonnées** : temps de calcul en secondes. Le nombre d'itérations varie comme suit : trait continu :  $P_{\max} = 10$ , trait discontinu :  $P_{\max} = 100$ , symbole ( $\circ$ ) :  $P_{\max} = 1000$ .

sources) et la capacité du modèle à représenter les données.

Dans ce paragraphe, nous proposons une étude préliminaire de l'estimation du nombre de sources  $n$  au sein du cadre GMCA. Nous supposons que les données  $\mathbf{X}$  ne sont pas perturbées par du bruit.

Pour un nombre de sources fixé  $p < n$  (où  $n$  est le "vrai" nombre de sources), le problème soulevé par l'ACMP consiste à résoudre le problème suivant :

$$\min_{\mathbf{A}, \alpha} \|\alpha\|_1 \quad \text{s.c.} \quad \|\mathbf{X} - \mathbf{A}\alpha\Phi\|_F < \epsilon. \quad (4.47)$$

$\mathbf{A}, \alpha \mid \text{ColDim}(\mathbf{A})=p$

où  $\text{ColDim}(\mathbf{A})$  est le nombre de colonnes de la matrice  $\mathbf{A}$ . Estimer le nombre de sources conjointement à la matrice de mélange et aux sources consisterait alors à chercher la solution du problème suivant :

$$\min_p \left\{ \min_{\mathbf{A}, \alpha} \|\alpha\|_1 \quad \text{s.c.} \quad \|\mathbf{X} - \mathbf{A}\alpha\Phi\|_F < \epsilon \right\}. \quad (4.48)$$

$\mathbf{A}, \alpha \mid \text{ColDim}(\mathbf{A})=p$

Notons  $\mathcal{P}_{p,\epsilon}$  le problème de l'équation (4.47). De façon intéressante, si  $p < n$ , il existe une valeur minimale  $\epsilon^*(p)$  pour laquelle si  $\epsilon < \epsilon^*(p)$ , le problème  $\mathcal{P}_{p,\epsilon}$  n'a aucune solution. Pour une valeur fixée  $p < n$ , la valeur minimale  $\epsilon^*(p)$  est atteinte en particulier en approximant les données  $\mathbf{X}$  avec ses  $p$  premiers vecteurs singuliers.

Ainsi, pour  $p < n$ ,  $\epsilon^*(p)$  est toujours non nul puisque les données sont confinées dans un sous-espace de dimension  $n$ . En conséquence, lorsque  $p = n$ , le problème  $\mathcal{P}_{n,\epsilon}$  a au moins une solution pour  $\epsilon = \epsilon^*(n) = 0$ .

Au final, il est envisageable d'estimer conjointement la matrice de mélange  $\mathbf{A}$ , les sources  $\mathbf{S}$  et le nombre de sources  $n$  par le biais d'une approche constructive. Nous proposons en effet de rechercher la solution de chacun des problème  $\mathcal{P}_{p,\epsilon}$  pour des valeurs croissantes de  $p \geq 1$  pour  $\epsilon = \epsilon^*(p)$ . Puisque l'algorithme GMCA peut fournir une solution approchée à  $\mathcal{P}_{p,\epsilon^*(p)}$  pour une valeur de  $p$  fixée, nous proposons d'adapter l'algorithme GMCA pour l'estimation du nombre de sources  $n$  comme suit :

Tant que  $\|\mathbf{X} - \mathbf{A}\alpha\Phi\|_F > 0$  et  $p \leq m$  :

1- Incrémenter la valeur de  $p$  en ajoutant une nouvelle colonne à  $\mathbf{A}$  - cette étape est décrite ci-dessous.

2- Résoudre  $\mathcal{P}_{p,\epsilon^*(p)}$  en utilisant l'algorithme GMCA pour la valeur  $p$  fixée :

$$\min_{\mathbf{A}, \alpha} |\text{ColDim}(\mathbf{A})=p| \|\alpha\|_1 \quad \text{s.c.} \quad \|\mathbf{X} - \mathbf{A}\alpha\Phi\|_F < \epsilon^*(p) .$$

**Le rôle de l'algorithme GMCA :** L'algorithme décrit précédemment recherche une suite de solutions aux problèmes  $\mathcal{P}_{p_i, \epsilon_i}$  pour une suite de valeurs  $\{p_i\}_i$ . Idéalement une estimation optimale devrait fournir les valeurs  $\{i, \epsilon^*(i)\}_{i=1, \dots, n}$  des solutions de  $\mathcal{P}_{i, \epsilon^*(i)}$  conduisant à la valeur optimale  $\epsilon^*(n) = 0$  pour  $i = n$ . Malheureusement, cette suite optimale est difficile à obtenir en pratique : la suite optimale  $\{i, \epsilon^*(i)\}_{i=1, \dots, n}$  est inconnue *a priori*. Heureusement, pour une valeur de  $p$  fixée, le processus de seuillage décroissant dans l'algorithme GMCA (Etape 2 de l'algorithme précédent) pourrait permettre d'accéder aux solutions proches de celles de  $\mathcal{P}_{p, \epsilon^*(p)}$ . En effet, dans le cadre de l'algorithme GMCA, il existe une application bijective entre les valeurs de  $\epsilon$  de l'équation (4.47) et le seuil  $\gamma$  utilisé dans GMCA de sorte que ces deux approches ont des solutions équivalentes. Evidemment, lorsque  $\epsilon \rightarrow 0$  alors  $\gamma \rightarrow 0$ . Ainsi, en pratique, pour une valeur fixée de  $p$ , faire tendre le seuil vers 0 dans l'algorithme GMCA devrait conduire à une solution approchée du problème  $\mathcal{P}_{p, \epsilon^*(p)}$ .

Ainsi l'étape 2 de l'algorithme précédent est effectuée en calculant les estimées de  $\mathbf{A}$  et  $\mathbf{S} = \alpha\Phi$  pour  $p$  fixé avec GMCA et avec pour seuil final  $\gamma_{\min} = 0$ .

Notons au passage que la suite optimale  $\{i, \epsilon^*(i)\}_{i=1, \dots, n}$  peut être calculée par avance. En effet, pour un nombre de sources fixé  $p = i$ , l'erreur d'approximation optimale est obtenue en projetant les données sur le sous-espace de dimension  $p$  généré par les  $p$  vecteurs singuliers liés aux  $p$  valeurs singulières les plus significatives de  $\mathbf{X}$ . Ainsi, une décomposition en valeurs singulières de  $\mathbf{X}$  pourrait être effectuée au préalable pour déterminer la suite des valeurs optimales  $\{i, \epsilon^*(i)\}_{i=1, \dots, n}$ . Ce raisonnement ne serait plus valable en présence de bruit.

**Augmentation du nombre de composantes :** Dans l'algorithme décrit précédemment, l'étape 1 conduit à l'ajout d'une colonne supplémentaire à la matrice de mélange courante  $\mathbf{A}$ . Le choix le plus simple consiste à choisir cette nouvelle colonne aléatoirement. Néanmoins, des choix plus astucieux sont envisageables :

- **Orthogonalité :** si la matrice de mélange est supposée orthogonale, la nouvelle colonne peut être choisie orthogonale au sous-espace généré par les colonnes de  $\mathbf{A}$  avec  $\text{ColDim}(\mathbf{A}) = p - 1$ .
- **Spectres connus :** si un ensemble de spectres est connu *a priori*, la nouvelle colonne peut être choisie parmi les spectres encore non utilisés. Ce spectre peut en particulier être choisi en fonction de sa cohérence avec le résidu. Définissons  $\mathcal{A}$  comme l'ensemble de spectres  $\{\eta_l \in \mathcal{A}\}_{l=1, \dots, \text{Card}(\mathcal{A})}$  et  $\mathcal{A}_c$  l'ensemble des spectres non encore utilisés (*i.e.* les spectres qui n'ont pas été choisis précédemment), alors la  $p$ -ème colonne de  $\mathbf{A}$  peut être choisie comme suit :

$$\eta_{l^*} = \underset{\eta_l \in \mathcal{A}_c}{\text{Argmax}} \left| \sum_{k=1}^t \frac{1}{\|\eta_l\|_{\ell_2}^2} \eta_l^T [\mathbf{X} - \mathbf{A}\mathbf{S}]^k \right|, \quad (4.49)$$

où  $[\mathbf{X} - \mathbf{A}\mathbf{S}]^k$  est la  $k$ -ième colonne de  $\mathbf{X} - \mathbf{A}\mathbf{S}$ .

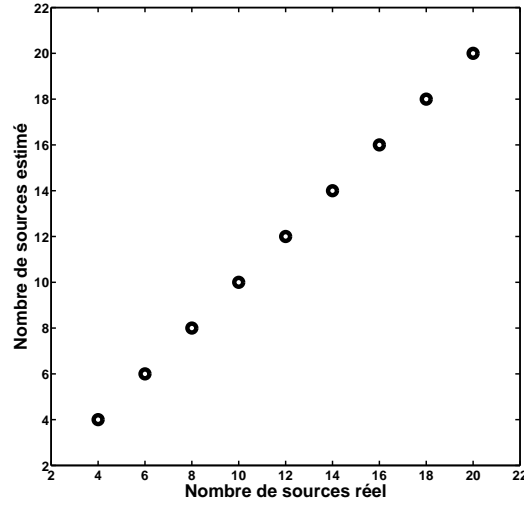
Tout autre information *a priori* peut être mise à profit pour une composition judicieuse de  $\mathbf{A}$ .

### Une expérience préliminaire

Dans cette expérience, les données sont le mélange de  $n$  sources. Les entrées de  $\mathbf{S}$  ont été générées aléatoirement (*i.i.d.*) suivant une distribution Laplacienne de précision  $\mu = 1$  ( $\Phi = \mathbf{I}$ ). Les éléments de la matrice de mélange sont tirés aléatoirement *i.i.d.* suivant une distribution gaussienne centrée de variance 1.

L'expérience préliminaire suivante compare l'ACP vu comme une méthode de sélection de sous-espace classique et l'algorithme GMCA adapté au cas où le nombre de sources est inconnu. En l'absence de bruit, seules les  $n$  valeurs propres fournies par l'ACP sont non nulles. L'ACP fournit ainsi le nombre correct de sources. Dans cette expérience, le nombre d'observations est  $m = 64$ . Chaque observation possède  $t = 256$  entrées. Le nombre de sources  $n$  varie de 2 à 20. La figure 4.15 montre le nombre de sources estimées par GMCA en fonction du nombre exact de sources. Chaque point a été calculé à partir de 25 réalisations différentes des données. Pour chacune de ces réalisations, GMCA estime correctement le nombre exact de sources.

La figure 4.16 propose une comparaison des performances de l'ACP et de GMCA en terme de qualité de séparation. Dans cette expérience, le nombre d'observations est  $m = 128$ . Chaque canal possède  $t = 2048$  entrées. Le graphique en haut de la figure 4.16 montre le RSB moyen en dB des sources estimées. Ainsi les sources estimées par GMCA sont bien



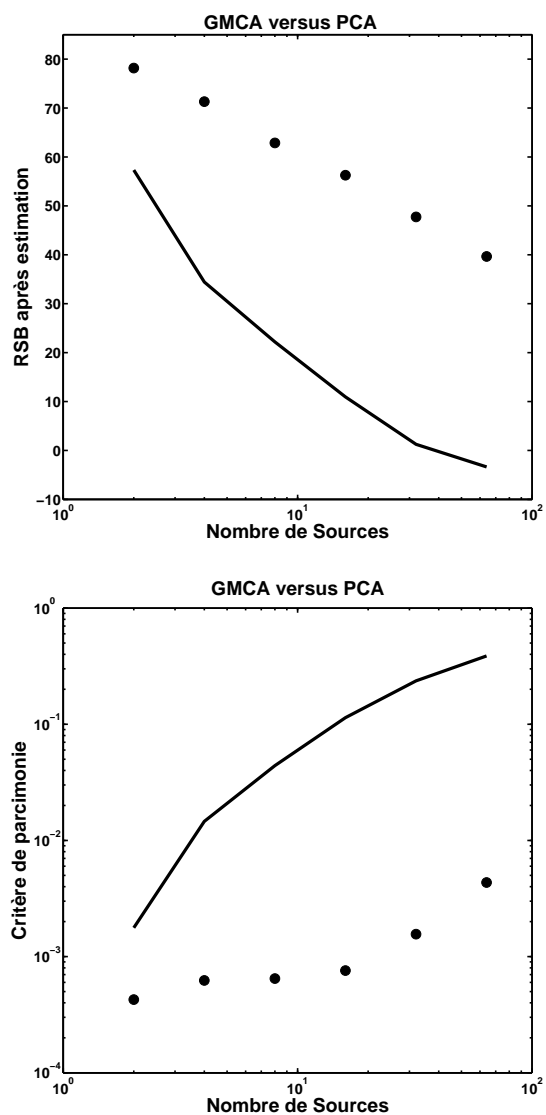
**FIG. 4.15: Estimation du nombre de sources avec GMCA-** **Abscisses :** nombre de sources réel.  
**Ordonnées :** Nombre de sources estimée avec GMCA. Chaque point est calculé comme la moyenne de 25 réalisations.

plus proches des sources exactes que celles fournies par une ACP. Si l'on définit le critère en norme  $\|\cdot\|_1$  suivant :

$$C_1 = \frac{\sum_{i=1}^n \|\mathbf{a}^i \mathbf{s}_i - \tilde{\mathbf{a}}^i \tilde{\mathbf{s}}_i\|_1}{\sum_{i=1}^n \|\mathbf{a}^i \mathbf{s}_i\|_1}, \quad (4.50)$$

où  $\sim$  caractérise les paramètres estimés.  $C_1$  quantifie une divergence en terme de parcimonie entre les sources et les sources estimées. Le graphique au bas de la figure 4.16 représente l'évolution du critère  $C_1$  lorsque le nombre de sources varie. Comme prévu l'algorithme GMCA fournit des sources bien plus parcimonieuses que celles obtenues par ACP.

Ainsi, l'adaptation proposée de l'algorithme GMCA pour l'estimation conjointe de la matrice de mélange  $\mathbf{A}$ , des sources  $\mathbf{S}$  et du nombre de sources  $n$  permet d'obtenir le nombre correct de source au même titre que l'ACP. De façon intéressante, au-delà de l'estimation du nombre de sources, GMCA apporte des solutions au problème de SSA satisfaisantes car elle met à profit les informations *a priori* disponibles pour les sources (ici leur parcimonie). Par la suite il sera intéressant d'étudier le comportement de GMCA pour l'estimation du nombre de composantes en particulier dans le cas où les données sont bruitées.



**FIG. 4.16:** GMCA (points) versus PCA (trait continu) - Abscisses : nombres de sources. Ordonnées - Haut : RSB en dB Bas : critère de parcimonie  $C_1$ . Chaque point est calculé à partir de 25 réalisations.

## Chapitre 5

# Extension à l'analyse de données hyperspectrales

Au cours des chapitres précédents, les problèmes de séparation de sources en aveugle ont été abordés dans leur généralité. Plus précisément, les techniques de SSA gagnent à être adaptées au traitement de données auxquelles elles sont appliquées. Un contexte, certes encore vaste, où la séparation aveugle de composantes est d'importance est celui de l'analyse de données hyperspectrales. Dans ce cadre particulier, les spécificités des données sont nombreuses. Dans ce chapitre, nous exposons une extension du cadre de l'Analyse en Composantes Morphologique Généralisée (GMCA) au cas de la séparation de sources hyperspectrales. Cette extension s'effectuera via l'introduction d'informations *a priori* adéquates sur la matrice de mélange.

Afin d'éviter toute ambiguïté, nous supposerons que les données hyperspectrales sont caractérisées par deux aspects importants :

1. **Grande dimension** : Le nombre d'observations  $m$  en imagerie hyperspectrale est important (typiquement plusieurs dizaines à plusieurs centaines). En conséquence, les données hyperspectrales sont des données de grandes dimensions.
  2. **Interprétation physique de la matrice de mélange** : le nombre potentiellement important de canaux nous fournit un échantillonnage d'une grandeur spectrale physique (longueur d'onde ... etc). Ces spectres physiques sont le plus souvent des "mélanges" de spectres (par exemple des spectres d'absorption) dont la régularité entraîne une structure forte. C'est cette structure ou morphologie spectrale que nous proposons de mettre à profit pour la séparation de sources en aveugle.
-



## 5.1 Morphologies spatiales et spectrales

### 5.1.1 Modélisation de l'information spectrale

Intuitivement, l'application de méthodes de séparation de composantes à des données hyperspectrales (par exemple en imagerie hyperspectrale) est motivée par le souhait d'extraire des composantes interprétables physiquement. Prenons l'exemple de l'imagerie hyperspectrale d'une zone urbaine ; le cube de données obtenu est composé de plusieurs clichés (sur des bandes de longueur d'onde différentes) d'un champ d'observation pouvant contenir des bâtiments dont le matériau (béton, brique ...) n'aura pas la même **signature spectrale** que le gazon, le bitume ou l'eau du fleuve situé non loin. En conséquence, l'objectif d'une analyse en composantes pour cette catégorie de données est l'extraction de composantes admettant des **signatures spectrales** propres.

Dans le modèle de mélange, chacune des sources  $\{\mathbf{s}_k\}_{k=1,\dots,n}$  est liée à une signature spectrale distincte  $\mathbf{a}^k$ .

Rappelons que l'élément essentiel de la séparation de sources en aveugle est la mise en évidence de la diversité présentée par les sources. La particularité des données hyperspectrales est qu'elles sont le siège d'une double source de diversité :

1. **Diversité spatiale** : de façon analogue aux développements déjà présentés pour la séparation de sources en aveugle, les sources sont *mutuellement parcimonieuses*<sup>1</sup> dans un dictionnaire **spatial**  $\Phi$ .
2. **Diversité spectrale** : les sources hyperspectrales admettent non seulement des morphologies spatiales différentes mais également des morphologies spectrales distinctes (diversité morphologique au sens large définie au Chapitre 1). Nous supposons en particulier que les signatures spectrales des différentes composantes présentent une représentation *mutuellement parcimonieuse* dans un **dictionnaire spectral**  $\Pi$ .

Ainsi, l'Analyse en Composantes Mutuellement Parcimonieuses de données hyperspectrales est équivalente à une décomposition du cube de données  $\mathbf{X}$  en une combinaison linéaire de  $n$  matrices de rang 1  $\{\mathbf{X}_k = \mathbf{a}^k \mathbf{s}_k\}_{k=1,\dots,n}$  avec l'information *a priori* que chacune des matrices admet une représentation parcimonieuse dans le dictionnaire **multicanal**  $\Pi \otimes \Phi$ .

Dans le cas non bruité, la SSA peut être abordé sous l'angle d'une décomposition Matrice de mélange - Sources impliquant un **découplage** fort entre la matrice de mélange et les sources. La tâche de séparation consiste alors en l'estimation de deux paramètres  $\mathbf{A}$  et  $\mathbf{S}$  dont le seul lien est l'adéquation aux données  $\mathbf{X} = \mathbf{AS}$ . Néanmoins, un tel découplage ne peut être qu'apparent : preuve en est l'irréductible indétermination du produit  $\mathbf{AS} = \mathbf{APP}^{-1}\mathbf{S}$  où  $\mathbf{P}$  est une matrice inversible quelconque. Une autre difficulté est celle de

---

<sup>1</sup> Voir Chapitre 1 et 2.

l'asymétrie du traitement de  $\mathbf{A}$  et  $\mathbf{S}$ . Alors même que ces deux paramètres jouent des rôles en tous points symétriques :

$$\mathbf{X} = \mathbf{AS} \longleftrightarrow \mathbf{X}^T = \mathbf{S}^T \mathbf{A}^T \quad (5.1)$$

la séparation de sources en aveugle classique provoque une rupture de symétrie en modélisant uniquement les sources  $\mathbf{S}$ . Le traitement des données hyperspectrales permet de rétablir cette symétrie entre les paramètres  $\mathbf{S}$  et  $\mathbf{A}$ . En effet, dans ce contexte précis, les données peuvent être interprétées comme une combinaison linéaire de  $n$  matrices de rang 1. Cette différence qui pourrait paraître formelle voire superficielle a son importance. Ainsi, le couple  $\mathbf{a}^k \mathbf{s}_k$  est intimement lié : i) d'une part, la composante  $\mathbf{s}_k$  n'a pas nécessairement de sens physique au sens où elle n'est pas liée à une grandeur physique ; ii) d'autre part, la grandeur interprétable physiquement est bien le produit  $\mathbf{a}^k \mathbf{s}_k$ . Dès lors, il n'y a donc plus lieu de traiter de manière asymétrique les paramètres spectraux et spatiaux.

Notons que cette symétrie est rarement mise en valeur. Citons néanmoins les travaux d'Hyvarinen (HK02), Stone (Sto77) et Hoyer (Hoy04).

Par la suite, nous ferons les hypothèses suivantes :

- **Le bruit** : le terme de bruit additif  $\mathbf{N}$  sera tel que chacune des entrées sera générée indépendamment selon une distribution gaussienne centrée de matrice de covariance  $\Sigma_{\mathbf{N}}$ .
- **Le dictionnaire multicanal** : du fait de la grande taille potentielle des données, il sera difficile de manipuler des dictionnaires multicanaux redondants. Nous supposons que  $\Phi$  et  $\Pi$  sont des matrices orthogonales liées à des bases orthogonales de  $\mathbb{R}^t$  et  $\mathbb{R}^m$  respectivement.
- **Les composantes** : comme indiqué auparavant, les composantes hyperspectrales seront définies comme la matrice  $\mathbf{X}_k = \mathbf{a}^k \mathbf{s}_k$ . Nous ferons l'hypothèse qu'elles admettent une représentation parcimonieuse dans la base multicanale  $\Pi \otimes \Phi$ . Par symétrie du traitement des paramètres spectraux et spatiaux, nous choisissons une loi *a priori* jointe sur le produit de ces deux termes  $\mathbf{X}_k = \mathbf{a}^k \mathbf{s}_k$ . En définissant :

$$\mathbf{v}_k = \mathbf{s}_k \Phi^T \quad (5.2)$$

$$\boldsymbol{\eta}^k = \Pi^T \mathbf{a}^k \quad (5.3)$$

On obtient :

$$\mathbf{X}_k = \Pi \alpha_k \Phi = \Pi \boldsymbol{\eta}^k \mathbf{v}_k \Phi = \mathbf{a}^k \mathbf{s}_k \quad (5.4)$$

où  $\alpha_k$  est une matrice de rang 1. L'information *a priori* de parcimonie des paramètres  $\mathbf{v}_k$  et  $\boldsymbol{\eta}^k$  est modélisée sous la forme du choix d'une distribution jointe définie en fonction de leur produit :

$$p(\boldsymbol{\eta}^k, \mathbf{v}_k) = p(\boldsymbol{\eta}^k \mathbf{v}_k) \propto \exp(-\gamma \|\boldsymbol{\eta}^k \mathbf{v}_k\|_1) \propto \exp(-\gamma \sum_{i,j} |\eta^k[i] v^j[k]|) \quad (5.5)$$

Cette distribution *a priori* garantit la symétrie du modèle de sources et l'invariance par changement d'échelle.

L'estimation des paramètres  $\{\boldsymbol{\eta}^k, \mathbf{v}_k\}$  au sens d'un maximum *a posteriori* (MAP) s'obtient alors comme suit :

$$\min_{\{\boldsymbol{\eta}^k, \mathbf{v}_k\}} \gamma \sum_{k=1}^n \|\boldsymbol{\eta}^k \mathbf{v}_k\|_1 + \frac{1}{2} \left\| \mathbf{X} - \sum_{k=1}^n \boldsymbol{\Pi} \boldsymbol{\eta}^k \mathbf{v}_k \boldsymbol{\Phi} \right\|_{F, \boldsymbol{\Sigma}_N}^2 \quad (5.6)$$

qui peut également s'écrire sous la forme suivante :

$$\min_{\{\alpha_k\}} \gamma \sum_{k=1}^n \|\alpha_k\|_1 + \frac{1}{2} \left\| \mathbf{X} - \sum_{k=1}^n \mathbf{X}_k \right\|_{F, \boldsymbol{\Sigma}_N}^2 \quad \text{avec } \mathbf{X}_k = \boldsymbol{\Pi} \alpha_k \boldsymbol{\Phi} \text{ et } \forall k, \text{rang}(\mathbf{X}_k) = 1 \quad (5.7)$$

Ces deux problèmes étendent donc l'ACMP (ici sous forme relaxée) à la séparation de composantes hyperspectrales.

### 5.1.2 L'algorithme HypGMCA

De l'orthogonalité de  $\boldsymbol{\Phi}$  et  $\boldsymbol{\Pi}$ , le problème de l'équation (5.6) s'écrit dans le domaine transformé comme suit :

$$\min_{\{\boldsymbol{\eta}^k, \mathbf{v}_k\}} \sum_{k=1}^n \gamma \|\boldsymbol{\eta}^k \mathbf{v}_k\|_1 + \frac{1}{2} \left\| \mathbf{X} - \sum_{k=1}^n \boldsymbol{\Pi} \boldsymbol{\eta}^k \mathbf{v}_k \boldsymbol{\Phi} \right\|_{F, \boldsymbol{\Sigma}'_N}^2 \quad (5.8)$$

où  $\alpha_{\mathbf{X}} = \boldsymbol{\Pi}^T \mathbf{X} \boldsymbol{\Phi}^T$  est composé de la projection des données dans le dictionnaire multicanal  $\boldsymbol{\Psi}$  et  $\boldsymbol{\Sigma}'_N = \boldsymbol{\Pi}^T \boldsymbol{\Sigma}_N \boldsymbol{\Pi}$  est la matrice de covariance du bruit  $\mathbf{N}$  dans la base multicanale  $\boldsymbol{\Psi}$ . Par souci de simplicité, nous choisissons  $\boldsymbol{\Sigma}_N = \sigma_N^2 \mathbf{I}$  et nous effectuons le changement de paramétrisation  $\gamma \leftarrow \gamma \sigma_N^2$ . Ainsi, si l'on définit  $\boldsymbol{\eta}$  la matrice de taille  $m \times n$  comme la concaténation en colonne des signatures spectrales  $\{\boldsymbol{\eta}^k\}_{k=1, \dots, n}$  et  $\mathbf{v}$  la matrice de taille  $n \times t$  formée par la concaténation en ligne des sources  $\{\mathbf{v}_k\}_{k=1, \dots, n}$ , le problème précédent peut être reformulé comme suit :

$$\min_{\{\boldsymbol{\eta}^k, \mathbf{v}_k\}} \gamma \sum_{k=1}^K \|\boldsymbol{\eta}^k \mathbf{v}_k\|_1 + \frac{1}{2} \|\alpha_{\mathbf{X}} - \boldsymbol{\eta} \mathbf{v}\|_F^2 \quad (5.9)$$

De façon analogue aux développements conduisant à l'algorithme "rapide" GMCA, nous proposons de résoudre le problème précédent par le biais d'un algorithme de Landweber projeté (voir Chapitre 4).

#### Estimation des signatures spectrales

En supposant les sources  $\mathbf{v}$  fixées, le problème à résoudre est le suivant :

$$\min_{\{\boldsymbol{\eta}^k\}} \gamma \sum_{k=1}^K \|\boldsymbol{\eta}^k \mathbf{v}_k\|_1 + \frac{1}{2} \|\alpha_{\mathbf{X}} - \boldsymbol{\eta} \mathbf{v}\|_F^2 \quad (5.10)$$


---

La forme générale de l'itération de mise à jour de  $\boldsymbol{\eta}$  est la suivante :

$$\boldsymbol{\eta}^{(+)} = \mathcal{S}_\gamma \left( \boldsymbol{\eta}^{(-)} - (\boldsymbol{\alpha}_X - \boldsymbol{\eta}^{(-)} \boldsymbol{v}) \mathbf{R}_\eta \right) \quad (5.11)$$

où  $\mathbf{R}_\eta$  est une matrice de relaxation qui tend à orienter la direction du gradient. Cette matrice de relaxation doit être telle que le rayon spectral de la matrice  $\mathbf{I} - \mathbf{R}_\eta$  soit borné au-dessus par 1. Nous adoptons de fait le choix suivant dont l'intérêt est de faire apparaître le hessien du terme quadratique de l'équation 5.10 :

$$\mathbf{R}_\eta = \boldsymbol{v}^T (\boldsymbol{v} \boldsymbol{v}^T)^{-1} \quad (5.12)$$

conduit ainsi à l'équation de mise à jour suivante :

$$\boldsymbol{\eta}^{(+)} = \mathcal{S}_\gamma \left( \boldsymbol{\eta}^{(-)} - (\boldsymbol{\alpha}_X - \boldsymbol{\eta}^{(-)} \boldsymbol{v}) \boldsymbol{v}^T (\boldsymbol{v} \boldsymbol{v}^T)^{-1} \right) \quad (5.13)$$

Cette dernière équation se simplifie comme suit :

$$\boldsymbol{\eta}^{(+)} = \mathcal{S}_\gamma \left( \boldsymbol{\alpha}_X \boldsymbol{v}^T (\boldsymbol{v} \boldsymbol{v}^T)^{-1} \right) \quad (5.14)$$

qui s'interprète aisément comme une version seuillée (subissant l'opérateur de seuillage  $\mathcal{S}_\gamma$ ) de l'estimateur de  $\boldsymbol{\eta}$  au sens des moindres carrés. Notons également que cet opérateur de seuillage  $\mathcal{S}_\gamma$  agit sur chacune des colonnes de  $\boldsymbol{\eta}$  avec un seuil  $\gamma_k$  défini comme suit :

$$\forall k = 1, \dots, n; \quad \gamma_k = \gamma \frac{\|\boldsymbol{v}_k\|_{\ell_1}}{\|\boldsymbol{v}_k\|_{\ell_2}^2} \quad (5.15)$$

### Estimation des sources

En supposant les sources  $\boldsymbol{\eta}$  fixées, le problème à résoudre est le problème symétrique suivant :

$$\min_{\{\boldsymbol{v}^k\}} \sum_{k=1}^K \gamma \|\boldsymbol{\eta}^k \boldsymbol{v}_k\|_1 + \frac{1}{2} \|\boldsymbol{\alpha}_X - \boldsymbol{\eta} \boldsymbol{v}\|_F^2 \quad (5.16)$$

La forme générale de l'itération de mise à jour de  $\boldsymbol{\eta}$  est la suivante :

$$\boldsymbol{v}^{(+)} = \mathcal{S}_{\gamma'} \left( \boldsymbol{v}^{(-)} - \mathbf{R}_v (\boldsymbol{\alpha}_X - \boldsymbol{\eta} \boldsymbol{v}^{(-)}) \right) \quad (5.17)$$

La matrice de relaxation  $\mathbf{R}_v$  doit être telle que le rayon spectral de la matrice  $\mathbf{I} - \mathbf{R}_v$  doit être borné au-dessus par 1. Nous adoptons le choix suivant :

$$\mathbf{R}_v = (\boldsymbol{\eta}^T \boldsymbol{\eta})^{-1} \boldsymbol{\eta}^T \quad (5.18)$$

formule qui conduit ainsi à l'équation de mise à jour suivante :

$$\boldsymbol{v}^{(+)} = \mathcal{S}_{\gamma'} \left( (\boldsymbol{\eta}^T \boldsymbol{\eta})^{-1} \boldsymbol{\eta}^T \boldsymbol{\alpha}_X \right) \quad (5.19)$$


---

qui s'interprète aisément comme une version seuillée (subissant l'opérateur de seuillage  $S_{\gamma'}$ ) de l'estimateur de  $\boldsymbol{\eta}$  au sens des moindres carrés. Notons également que cet opérateur de seuillage  $S_{\gamma'}$  agit sur chacune des lignes de  $\boldsymbol{\nu}$  avec un seuil  $\gamma'_k$  défini comme suit :

$$\forall k = 1, \dots, n; \quad \gamma'_k = \gamma \frac{\|\boldsymbol{\eta}_k\|_{\ell_1}}{\|\boldsymbol{\eta}_k\|_{\ell_2}^2} \quad (5.20)$$

En conclusion, l'algorithme d'extension de GMCA au traitement des données hyperspectrales intitulé HypGMCA est le suivant :

1. Fixer le nombre d'itérations  $P_{\max}$  ainsi que le seuil  $\gamma^{(0)}$
2. Transformation des données  $\mathbf{X}$  en  $\boldsymbol{\alpha}_\mathbf{X}$
3. Tant que  $\gamma^{(h)}$  est supérieur au seuil final  $\gamma_{\min}$ ,
  - Mise à jour de  $\boldsymbol{\nu}$  en supposant  $\boldsymbol{\eta}$  fixé :  
 $\boldsymbol{\nu}^{(h+1)} = S_{\gamma_k^{(h)}} \left( (\boldsymbol{\eta}^t \boldsymbol{\eta})^{-1} \boldsymbol{\eta}^t \boldsymbol{\alpha}_\mathbf{X} \right)$
  - Mise à jour de  $\boldsymbol{\eta}$  en supposant  $\boldsymbol{\nu}$  fixé :  
 $\boldsymbol{\eta}^{(h+1)} = S_{\gamma_k^{(h)}} \left( \boldsymbol{\alpha}_\mathbf{X} \boldsymbol{\nu}^t (\boldsymbol{\nu} \boldsymbol{\nu}^t)^{-1} \right)$
  - Décroissance du seuil  $\gamma^{(h)}$ .
4. Arrêt lorsque  $\gamma^{(h)} < \gamma_{\min}$ .
5. Transformer les estimées  $\boldsymbol{\eta}$  et  $\boldsymbol{\nu}$  en  $\mathbf{A}$  et  $\mathbf{S}$

### 5.1.3 Comparaison avec GMCA

#### Données hyperspectrales : comparaison avec GMCA

En accord avec le modèle de mélange, les données  $\mathbf{X}$  sont modélisées comme une combinaison linéaire de  $n$  sources. Dans cette expérience de principe, les sources sont tirées au hasard parmi un ensemble de 32 images de taille  $128 \times 128$  apparaissant sur la figure 5.1. Le nombre de sources est  $n = 5$ . Les spectres sont générés aléatoirement de sorte que chacun de leurs échantillons est *i.i.d.* suivant une Laplacienne de paramètre  $\mu = 1$  dans une base orthogonale d'ondelettes. Le nombre de canaux est  $m = 128$ . De plus, un bruit gaussien additif centré et blanc de matrice de covariance  $\boldsymbol{\Sigma}_\mathbf{N} = \sigma_\mathbf{N}^2 \mathbf{I}$  est ajouté aux données.

Nous allons tout d'abord comparer l'algorithme GMCA original avec son extension aux données hyperspectrales. Nous proposons d'évaluer la robustesse de chaque méthode au bruit pour des RSB variant de 0 à 40 dB. La figure 5.2 montre 6 des 128 canaux bruités pour lesquels RSB = 20dB. Les deux algorithmes ont été appliqués aux données avec 100 itérations. Les représentations sont les suivantes : curvelets pour  $\boldsymbol{\Phi}$  et DWT pour  $\boldsymbol{\Pi}$ . Sur la figure 5.3 sont représentées les sources estimées par GMCA (images de gauche) et HypGMCA (images de droite). Visuellement, l'algorithme HypGMCA semble apporter de meilleurs résultats. Plus quantitativement, la figure 5.4 montre l'évolution du critère  $C_\mathbf{A}$  lorsque le RSB varie de 0 à 40dB. Il est donc clair que la prise en compte de la double contrainte de parcimonie : spatiale et spectrale conduit à une estimation de  $\mathbf{A}$  et  $\mathbf{S}$  bien plus précise et robuste vis-à-vis du bruit.

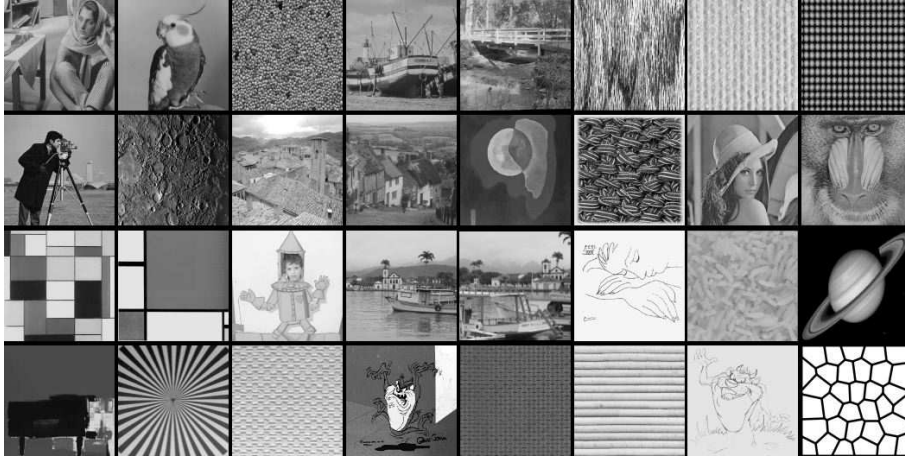


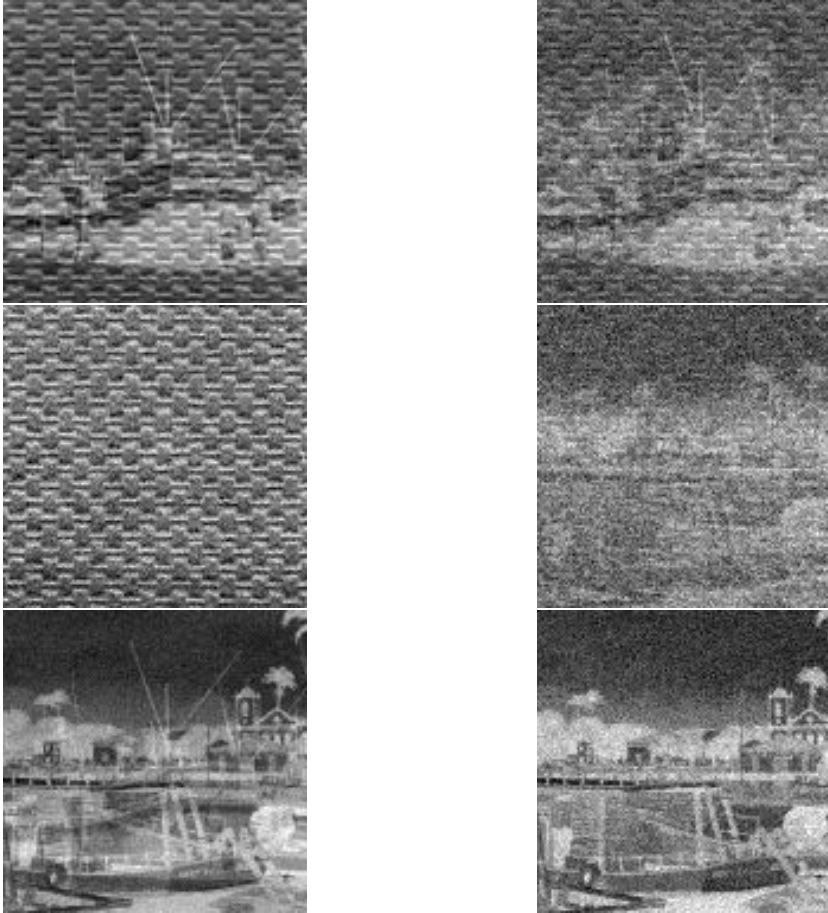
Fig. 5.1: Ensemble d'images utilisées lors des expérimentations.

### Comportement en grandes dimensions

Intuitivement, à nombre d'échantillons  $t$  et d'observations  $m$  fixés, l'augmentation du nombre de sources devrait rendre le problème de séparation de sources plus difficile ; le nombre de "paramètres" à estimer grandissant avec  $n$ . Une modélisation plus fine des données telle qu'elle est faite dans HypGMCA devrait apporter plus de robustesse en grande dimension (*i.e.* ici grande dimension signifiant nombre de sources élevé).

A cette fin, nous proposons une expérience de principe unidimensionnelle. Les entrées de  $\mathbf{S}$  sont *i.i.d.* suivant une loi Laplacienne de paramètre  $\mu = 1$  ( $\Phi$  est la base canonique). Les entrées de la matrice de mélange sont également *i.i.d.* suivant une loi Laplacienne de paramètre  $\mu = 1$  ( $\Pi$  est la base canonique). Les données sont contaminées par un bruit additif gaussien centré, blanc et de variance 1. Le nombre d'échantillons est  $t = 2048$  ; le nombre de canaux est  $m = 128$ . La figure 5.5 montre une comparaison entre GMCA et son extension au cadre des données hyperspectrales. Chaque point est la moyenne de 100 tests indépendants. Le graphique de gauche de la figure 5.5 illustre l'évolution du RSB lorsque le nombre de sources varie de 2 à 64. Lorsque le nombre de sources est faible, l'apport de la contrainte de parcimonie spectrale est certes faible mais positif. A mesure que le nombre de sources croît, les solutions de HypGMCA sont plus robustes que celle de GMCA. Par exemple, pour  $n = 64$ , l'algorithme HypGMCA dépasse (en terme de RSB) l'algorithme GMCA de 12dB. Le graphique de droite de la figure 5.5 montre le comportement des deux algorithmes en terme de critère de parcimonie  $C_{\ell_1}$ . Comme prévu, la contrainte de parcimonie spectrale conduit effectivement à une solution plus parcimonieuse et ceci d'autant plus que le nombre de sources devient important.

Au vu de ces expériences, l'ajout d'une contrainte spectrale est profitable à plus d'un titre : i) une meilleure stabilité de l'algorithme au bruit, ii) une plus grande robustesse lorsque la



**Fig. 5.2:** Six mélanges de taille  $128 \times 128$  parmi les 128 canaux. Le RSB est de 20dB.

dimension du problème croît (*i.e.* le nombre de sources).

## 5.2 Application à l'extraction de composantes physiques à partir de données hyperspectrales réelles

Dans cette section, nous proposons les résultats préliminaires de l'application de l'algorithme HypGMCA à l'extraction de composantes dans des données hyperspectrales réelles.

### 5.2.1 Données Mars Express

Nous proposons d'appliquer l'algorithme HypGMCA à l'extraction des composantes de glace d'eau et de dioxyde de carbone dans des données hyperspectrales de la calotte

---

polaire martienne. Ces données<sup>2</sup> ont été obtenues par la sonde *Mars Express*<sup>3</sup> et plus particulièrement à partir de l'instrument OMEGA (Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité). OMEGA possède une résolution spatiale allant de 4 à 300 km. Les données utilisées ont été capturées dans le proche infrarouge. Les données hyperspectral que nous allons analyser sont constituées de 128 canaux dont les fréquences varient dans l'intervalle  $[0,93\mu\text{m} - 2,73\mu\text{m}]$  avec une résolution de  $0,013\mu\text{m}$ . Ces données ont été calibrées de sorte que chaque pixel quantifie une valeur de *réflectance* (définie comme le rapport entre l'irradiance en chacun des pixels et l'irradiance solaire au sol). L'une des principales difficultés associées au traitement des données hyperspectrales réside dans le peu de validité du modèle de mélange. Les composantes d'intérêt font l'objet de mélanges non-linéaires (faible résolution spatiale, effets atmosphériques, ... etc). Néanmoins, il peut être intéressant (voir (MHS<sup>+</sup>08)) de considérer un modèle de sources mettant à profit de multiples informations *a priori* : parcimonie spatiale des sources, parcimonie de leurs spectres, positivité des sources, etc.

Dans l'expérience qui va suivre, certains canaux extrêmement bruités ont été mis de côté de sorte que  $\mathbf{X}$  est formé de 117 canaux ; chacun étant une image de taille  $872 \times 128$  pixels. La figure 5.6 présente 4 de ces 117 canaux pour des fréquences variant de  $1,38\mu\text{m}$  à  $2,4\mu\text{m}$ .

**Hypothèses** Nous avons effectué la recherche de 10 sources. L'algorithme HypGMCA a été utilisé avec les contraintes suivantes :

- **Parcimonie spatiale des sources** : on peut tout à fait supposer que les composantes  $H_2O$  et  $CO_2$  (celles que nous serons aptes à caractériser) possèdent une morphologie spatiale parcimonieuse dans une base d'ondelettes orthogonale.
- **Parcimonie des spectres** : les spectres d'intérêt sont structurés (spectres d'absorption/émission) de sorte qu'on les supposera parcimonieux dans une base d'ondelettes orthogonale.
- **Positivité des sources** : on fera l'hypothèse que les sources à estimer sont positives. Dans l'algorithme HypGMCA, cette contrainte de positivité est obtenue par projection à chaque itération de  $\mathbf{S}$  sur le cône des vecteurs à entrées positives.

Le nombre d'itérations est fixé à  $P_{max} = 250$ .

### 5.2.2 Résultats préliminaires

Parmi les 10 sources estimées, les plus corrélées avec les spectres de référence des glaces de  $H_2O$  et  $CO_2$  ont été identifiées. La figure (5.7) présente les résultats obtenus pour

---

<sup>2</sup>A ce titre, je tiens à remercier O. Forni pour nous avoir gracieusement fourni ces données ainsi que les spectres de référence.

<sup>3</sup>Voir [www.esa.int/marsexpress](http://www.esa.int/marsexpress).

---

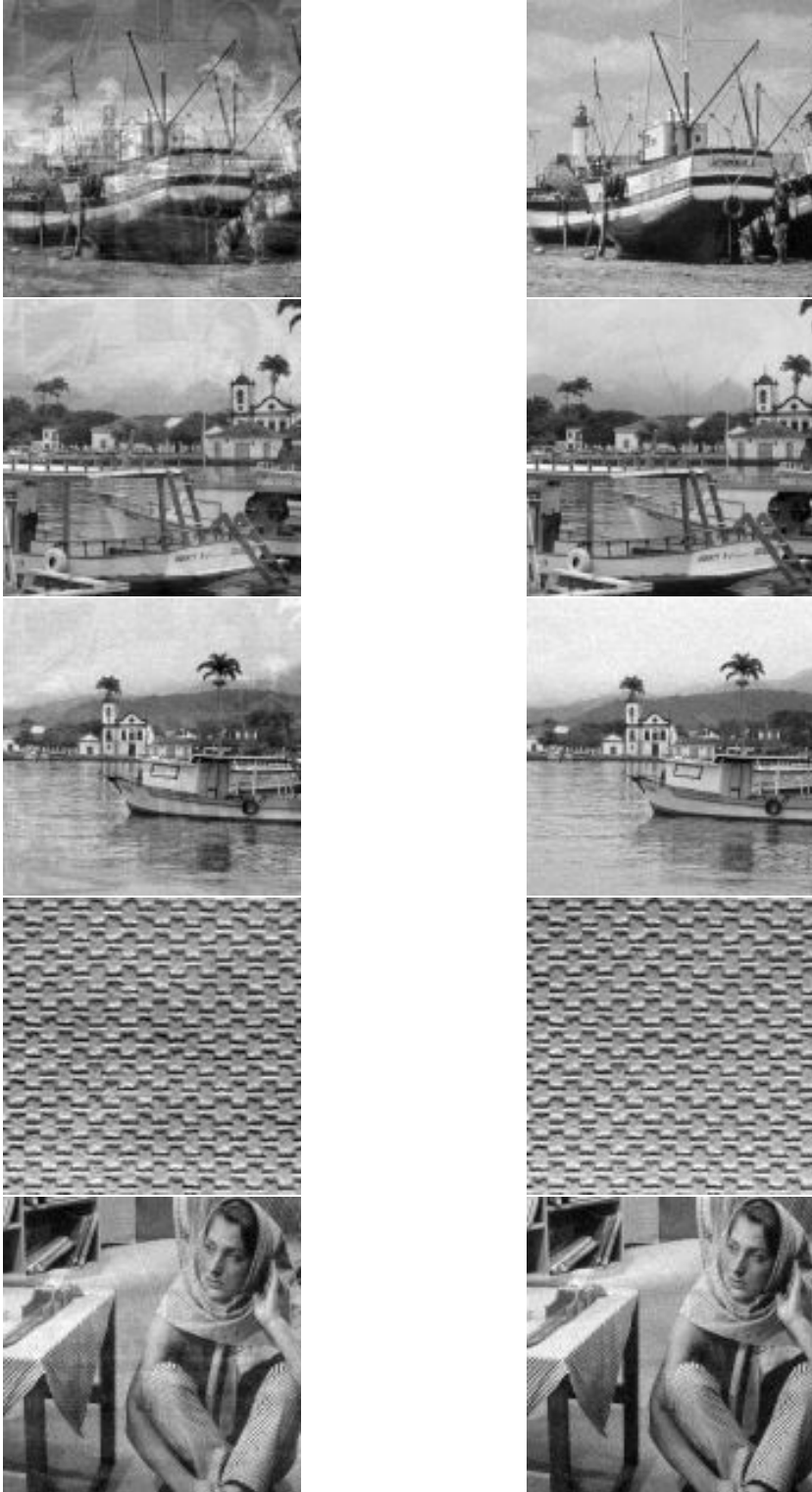


la glace de  $CO_2$ . L'image de gauche est la source  $CO_2$  estimée (image extraite de taille  $256 \times 128$ ). Visuellement, la glace de  $CO_2$  est concentrée spatialement autour du pôle en accord avec les résultats de (JMS07; MHS<sup>+</sup>08). Le graphique de droite présente le spectre de glace de  $CO_2$  de référence (ligne continue) ainsi que le spectre estimé. Compte tenu des non linéarités non modélisées des mélanges réels, l'adéquation du spectre estimé au spectre de référence est assez remarquable.

La figure 5.8 présente des résultats identiques pour la glace d'eau. L'interprétation du comportement spatial de la source est plus difficile à caractériser. La glace d'eau semble présente entre différents interstices présents à la surface. Du point de vue du spectre, le spectre estimé est remarquablement similaire au spectre de référence pour des fréquences supérieures à  $1\mu m$ .

Il serait intéressant de continuer l'étude sur différents points :

- **Plus large spectre** : de façon comparable à l'étude présentée dans (MHS<sup>+</sup>08), la même portion de la calotte polaire martienne a été observée sur la bande  $2,55 - 5,11\mu m$  en 128 canaux de résolution de  $0,020\mu m$ . L'utilisation jointe de ces deux jeux de données serait effectivement intéressante.
  - **Approche semi-aveugle** : ces résultats ont été obtenus en aveugle. Il peut également être extrêmement intéressant d'adopter une approche semi-aveugle en contraignant en partie les spectres  $H_2O$  et  $CO_2$ .
  - **Interprétation des spectres** : nous n'avons présenté que 2 des 10 composantes estimées par HypGMCA. A titre indicatif, aucune des 8 autres sources ne présente des caractéristiques spectrales proches des glaces de  $CO_2$  et  $H_2O$ . Il serait donc passionnant d'approfondir l'interprétation de ces composantes ; ceci n'a pas pu être fait jusqu'à présent et fera l'objet d'une prochaine étude.
-



**FIG. 5.3:** figures de gauche : sources estimées avec la version originale (*i.e.* sans contrainte spectrale) de GMCA. figures de droite : sources estimées avec HypGMCA.

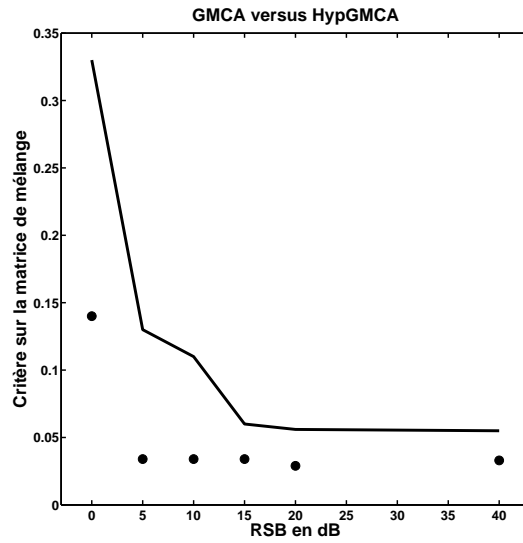


FIG. 5.4: Evolution du critère sur la matrice de mélange  $C_A$  comme une fonction du RSB en dB. Ligne continue : résultats avec GMCA. Symbole • : résultats avec HypGMCA.

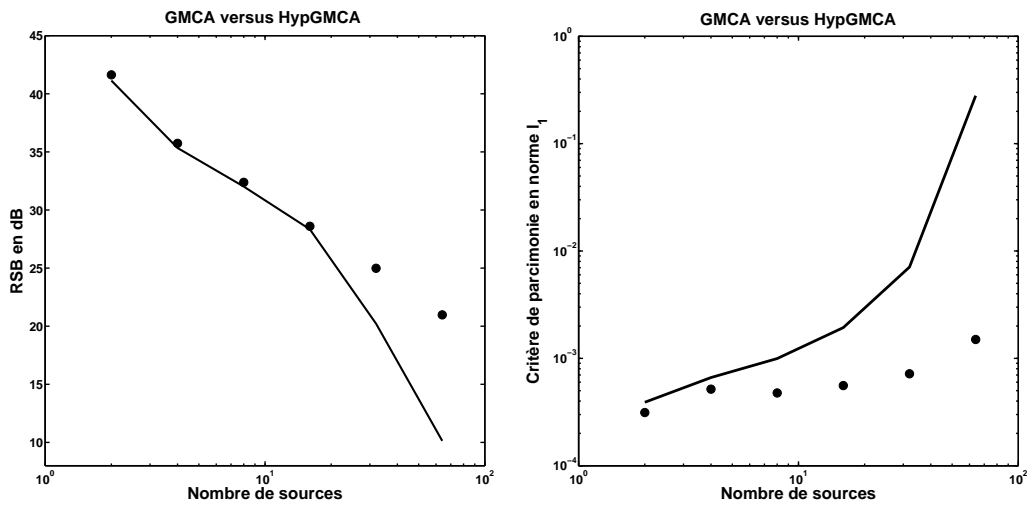
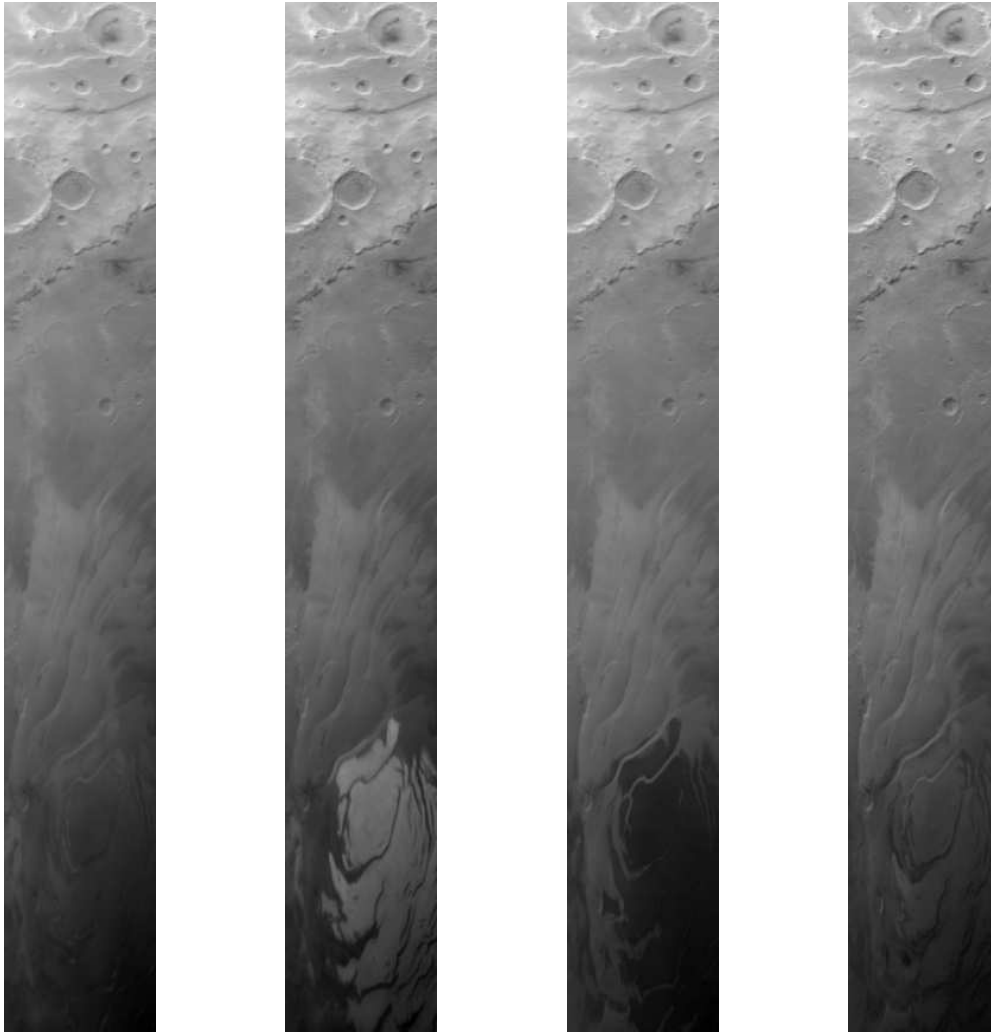
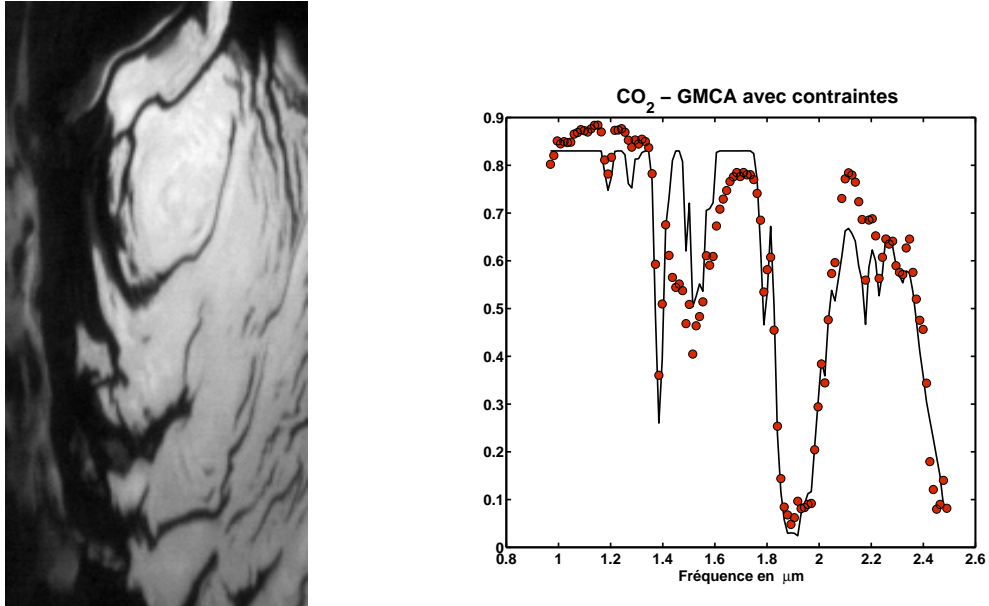


FIG. 5.5: Abscisse : nombre de sources. Ordonnées- gauche : RSB de la solution. Droite : critère de parcimonie  $C_{\ell_1}$ . Ligne continue : résultats de GMCA. Symbole • : résultats de HypGMCA.

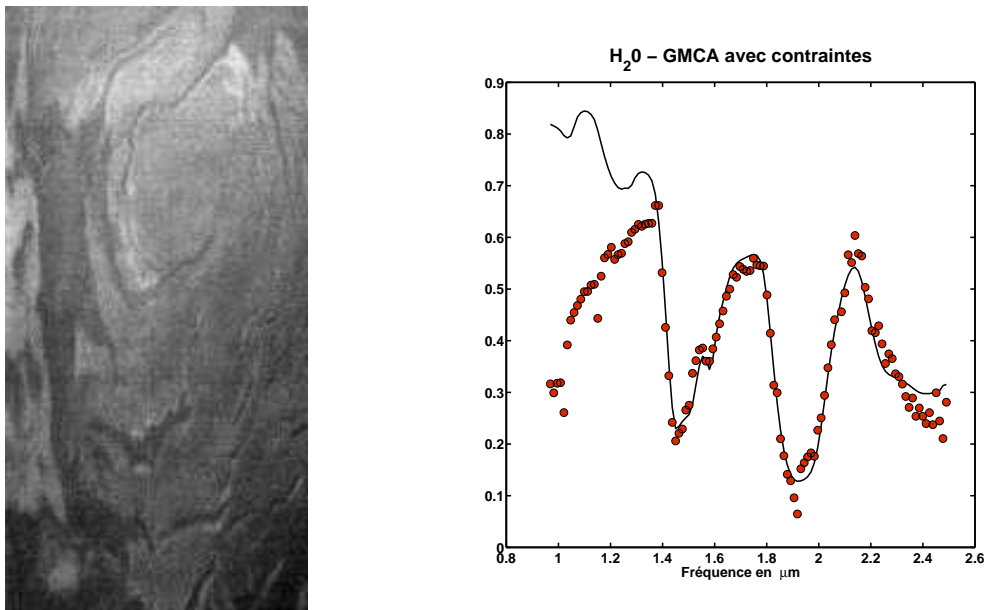


**FIG. 5.6:** De gauche à droite : observations Mars Express pour les fréquences 1,38 - 1,75 - 1,94 et 2,41 $\mu m$ .

---



**FIG. 5.7:** Image de gauche : source de glace de  $CO_2$  estimée. Graphique de droite - ligne continue : spectre de référence de la glace de  $CO_2$ . Symbole • : spectre estimé.



**FIG. 5.8:** Image de gauche : source de glace de  $H_2O$  estimée. Graphique de droite - ligne continue : spectre de référence de la glace de  $H_2O$ . Symbole • : spectre estimé.

## Chapitre 6

# Application aux données Planck

**E**N vue du lancement durant le premier semestre de l'année 2009 de la mission spatiale Planck de l'Agence Spatiale Européenne (ESA), cette partie fait l'objet de l'application de la séparation de sources en aveugle à la séparation de composantes astrophysiques. Ce projet d'envergure aura en particulier pour objectif l'étude du fond diffus cosmologique. Celui-ci constitue l'une des clefs de voûte des théories actuelles d'expansion de l'Univers ; il est donc l'un des éléments auxquels les cosmologistes portent une grande attention. Après une présentation du contexte scientifique du projet Planck, nous verrons comment l'Analyse en Composantes Morphologiques Généralisées peut être étendue pour l'extraction de composantes astrophysiques. Les résultats que nous présentons seront essentiellement liés à l'estimation du fond diffus cosmologique.

---

## 6.1 Les données CMB : contexte et enjeux

### 6.1.1 Fond diffus cosmologique

Le fond diffus cosmologique (Cosmological Microwave Background - CMB) est considéré par les cosmologistes comme l'une des clefs de voûte des théories actuelles de l'expansion de l'Univers (théorie du Big Bang). Le CMB<sup>1</sup> est le reliquat du rayonnement le plus ancien (d'où son appellation traditionnelle de "rayonnement fossile"). Selon la théorie du Big Bang, le fond diffus cosmologique est la trace du passage rapide de l'état opaque (lié à une forte ionisation de l'Univers) à transparent de l'Univers. Cet événement dont le déroulement serait daté de  $3.10^5 - 4.10^5$  années après le Big Bang constitue donc la "lumière" la plus ancienne observable actuellement. Elle revêt une importance capitale pour les cosmologistes puisque le comportement très particulier du CMB est riche d'enseignements.

La théorie du Big Bang prévoit que le fond diffus cosmologique soit un champ d'émission isotrope gaussien caractérisé par son spectre de puissance. Spectralement<sup>2</sup>, le CMB se comporte comme un corps noir dont la température est de l'ordre de  $2,735K$ .

Cette prévision fait du CMB un signal apparemment simple mais riche d'informations pour les cosmologistes. Le rayonnement CMB observé par Planck sera mesuré sur une sphère dont l'observateur est le centre. Dans ce cas, le CMB est un champ gaussien caractérisé par son spectre de puissance en harmoniques sphériques<sup>3</sup>. Mathématiquement, la matrice de covariance du CMB dans l'espace des harmoniques sphériques est diagonale ; celle-ci est appelée spectre de puissance. Ce dernier est une fonction de l'échelle angulaire  $l$  (équivalent de la fréquence sur la sphère) :  $C(l)$ . La figure 6.1 montre le spectre théorique du CMB normalisé ainsi que différentes mesures effectuées par les dernières missions de mesure du CMB. On peut observer sur la figure 6.1 que le spectre de puissance du CMB présente une série d'oscillations (appelées "modes acoustiques"). Les caractéristiques de ces différents modes (position et amplitude) permettent en particulier d'apporter des contraintes sur les paramètres cosmologiques intervenant dans les modèles d'expansion (voir (LR05)). L'estimation du CMB et en particulier de son spectre de puissance est ainsi cruciale pour contraindre les modèles de Big Bang qui décrivent l'évolution de l'Univers.

### 6.1.2 Mission spatiale Planck

Au cours du premier semestre de l'année 2009, une nouvelle mission d'observation du CMB sera lancée : la mission Planck de l'Agence Spatiale Européenne (ESA)<sup>4</sup>. Cette mission fait suite à une série de missions spatiales (COBE, WMAP) qui ont déjà apporté de nombreuses informations sur le CMB (Jun96). La dernière de ces missions (WMAP) a

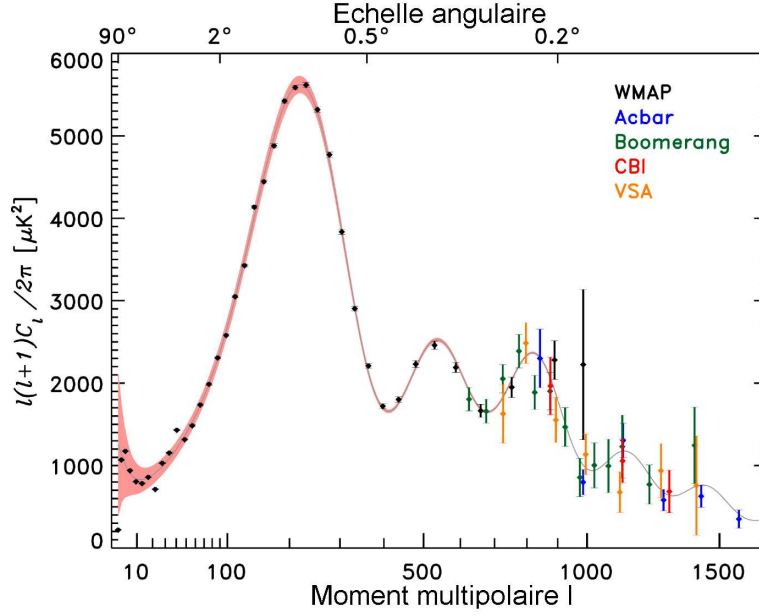
---

<sup>1</sup>Observé la première fois par Penzias et Wilson en 1963. Cette découverte leur a valu le prix Nobel de physique en 1978.

<sup>2</sup>Comportement par bande de fréquences.

<sup>3</sup>Equivalent de l'espace de Fourier pour la sphère.

<sup>4</sup>Nous invitons le lecteur à se reporter au site <http://planck.esa.int>.



**Fig. 6.1:** Spectre de puissance théorique du CMB et mesures issues des dernières missions d'observation (issu de (Hin06)).

ainsi confirmé la théorie sur les deux premiers modes du spectre de puissance du CMB. WMAP apporte des observations du ciel entier (sphère entière) jusqu'à une résolution de 15 arcminutes sur 5 canaux dont les fréquences varient de 20 à 90 GHz. Le défi porté par Planck est l'observation du ciel à une résolution plus fine (allant jusqu'à 5 arcminutes) sur 9 canaux de 30 à 857 GHz.

Outre le fond diffus cosmologique, chaque canal Planck capturera différentes composantes astrophysiques (BG99) que nous allons décrire par la suite.

### Fond diffus cosmologique

Les connaissances actuelles du fond diffus cosmologique sont les suivantes :

- **Loi d'émission** : sous l'angle de la modélisation de données multicanales, l'émission de fond diffus cosmologique à la fréquence  $\nu$ ,  $\mathbf{x}_\nu^{cmb}$  est le produit de la carte des fluctuations du CMB  $\Delta T^{cmb}$  (indépendante de  $\nu$ ) et d'une loi d'émission définie comme la dérivée de celle d'un corps noir à la température du CMB<sup>5</sup> ( $\bar{T} \simeq 2,735K$ ) :

$$\mathbf{x}_\nu^{cmb} = \Delta T^{cmb} \left[ \frac{\partial B_\nu(T)}{\partial T} \right]_{T=\bar{T}} \quad (6.1)$$

De sorte que la composante CMB  $\mathbf{X}^{cmb}$  contenue dans les données  $\mathbf{X}$  peut s'écrire sous la forme d'une composante de rang 1, dans un modèle de fluctuations linéaires

<sup>5</sup>Par la suite, les données seront manipulées en température d'antenne.



autour de  $\bar{T}$ . En d'autres termes, le CMB est présent dans les données  $\mathbf{X}$  sous la forme d'un mélange parfait dont la colonne de mélange sera supposée connue :

$$\mathbf{a}^{cmb} = \left[ \left[ \frac{\partial B_\nu(T)}{\partial T} \right]_{T=\bar{T}} \right]_{\nu=\{30, \dots, 857\text{GHz}\}} \quad (6.2)$$

La source correspondante,  $\mathbf{s}^{cmb} = \Delta T^{cmb}$  sera définie comme la carte des fluctuations de température du fond diffus cosmologique. Notons que cette loi d'émission prévue par la théorie est validée par les observations CMB antérieures : spatiales telles que COBE et WMAP ou des sondes atmosphériques telles que Boomerang et Archeops.

- **Stationnarité et caractère gaussien** : la carte de fluctuation de température que l'on notera  $\mathbf{s}^{cmb}$  est, en théorie, un champ gaussien stationnaire entièrement défini par son spectre de puissance  $C(\ell)$  dans l'espace des harmoniques sphériques. En réalité, la recherche de non-gaussianités dans les données du CMB est également réalisée pour tester certains modèles d'inflation de l'Univers. En l'état actuel des connaissances sur le CMB, le caractère gaussien de  $\mathbf{s}^{cmb}$  peut être considéré comme vérifié. Les observations Planck, de part leur résolution, pourraient permettre d'apporter de nouvelles données pour tester le caractère gaussien du champ de fluctuation de température du fond diffus cosmologique.
- **Spectre de puissance** : comme nous l'avons souligné, sous hypothèse du caractère gaussien du fond diffus cosmologique, le champ  $\mathbf{s}^{cmb}$  est statistiquement déterminé par son spectre de puissance dans l'espace des harmoniques sphériques  $C(\ell)$ . Ce spectre de puissance (angulaire puisque les données sont définies sur la sphère), est connu avec une faible erreur pour des valeurs de  $\ell$  pouvant atteindre  $\ell = 800$  notamment grâce aux données WMAP. Le spectre  $C(\ell)$  sera donc le "paramètre" sur lequel se portera plus particulièrement notre attention par la suite.

## Composantes galactiques

Ces composantes galactiques sont directement liées aux émissions de la voie lactée :

- **Émissions Synchrotron** : les émissions synchrotron sont provoquées par le mouvement en spirale de particules chargées dans un champ magnétique. Dans notre galaxie, de tels effets sont particulièrement intenses au niveau du centre galactique et peuvent s'étendre sur de plus hautes latitudes (par convention, l'équateur de la sphère d'observation). Le cube de données synchrotron est modélisé comme le produit d'un champ avec une loi d'émission suivant une loi de puissance  $\nu^{-\beta}$  pour chaque pixel  $i$  de la sphère :

$$\forall \nu; \quad x_\nu^{sync}[i] = s^{sync}[i] \nu^{-\beta^{sync}[i]} \quad (6.3)$$


---

Le cube de données synchrotron  $\mathbf{X}^{sync}$  est donc entièrement déterminé par les champs<sup>6</sup>  $\mathbf{s}^{sync}$  et  $\beta^{sync}$ . Ce dernier champ d'indice spectral fluctue typiquement entre 2.5 et 3.1. Malheureusement, les émissions synchrotron ne peuvent être modélisées comme une composante de rang 1 : il n'intervient donc pas comme un mélange linéaire.

- **Emission Free-Free** : l'émission "Free-Free" provient de l'interaction d'électrons libres avec des ions dans le milieu galactique ionisé. Au cours de ces interactions, l'énergie perdue par l'électron engendre l'émission d'un photon. En première approximation, le cube de données "Free-Free" peut être modélisé comme une composante de rang 1 issue du produit d'un champ de température et d'une loi d'émission en loi de puissance d'indice spectral  $\beta^{FF} \simeq 2,15$ . De sorte qu'à la fréquence  $\nu$ , la composante liée au "Free-Free" intervient comme suit :

$$\forall \nu; \quad \mathbf{x}_\nu^{FF} = \mathbf{s}^{FF} \nu^{-\beta^{FF}} \quad (6.4)$$

- **Emissions thermiques de poussières** : différents mécanismes peuvent être à l'origine des émissions de poussières galactiques. La description du rayonnement lié aux poussières est, à notre connaissance, uniquement effectuée qu'à l'aide d'une loi empirique. La contribution liée aux émissions de poussières est modélisée pour chaque pixel  $i$  de la sphère comme suit :

$$\forall \nu; \quad x_\nu^P[i] \propto \frac{\nu^{\beta^P[i]}}{\exp(h\nu/k s^P[i]) - 1} \quad (6.5)$$

où  $\beta^P$  est le champ d'indices spectraux de la loi d'émission supposée varier aux alentours de 2-3. Le champ  $\mathbf{s}^P$  est le champ de température dont les fluctuations sont comprises entre 16 et 18K. Tout comme la composante synchrotron, la composante de poussière n'intervient pas sous la forme d'un mélange linéaire.

Cette liste de composantes est non exhaustive. En particulier, certains astrophysiciens supposent l'existence d'une composante d'émissions de poussières dite "anormale" (non liée à des émission thermiques). Controversée, nous n'évoquons pas cette composante par la suite.

### Effet "Sunyaev Zel'Dovich"

L'effet Sunyaev Zel'Dovich (effet SZ) est la diffusion compton inverse des photons du fond diffus cosmologique avec les électrons libres d'un milieu ionisé. Il y a plus précisément deux types d'effet SZ :

- **Effet SZ thermique** : l'effet SZ thermique est lié à la diffusion des photons dans un gaz d'électrons à haute température tels que ceux des amas de galaxies. En première

---

<sup>6</sup>Au sens d'une carte sur la sphère.

approximation, les données de SZ thermique peuvent être modélisées comme une composante de rang 1 (mélange linéaire) :

$$\mathbf{X}^{SZ} = \mathbf{a}^{SZ} \mathbf{s}^{SZ} \quad (6.6)$$

- **Effet SZ cinétique** : l’effet SZ cinétique est lié à la diffusion des photons dans un gaz d’électrons en mouvement par rapport au CMB. Le SZ cinétique suit une loi d’émission similaire à celle du CMB. De faible amplitude, il est de fait difficile à distinguer des fluctuations du fond diffus cosmologique. Etant fortement corrélé aux amas de galaxies, l’estimation du SZ cinétique peut s’effectuer en post-traitement sur la carte de CMB.

### 6.1.3 Objectifs

En première approximation, les différentes composantes astrophysiques s’ajoutent linéairement au niveau des données :

$$\mathbf{X} = \mathbf{X}^{cmb} + \mathbf{X}^{SZ} + \mathbf{X}^{FF} + \mathbf{X}^{sync} + \mathbf{X}^P + \mathbf{X}^{PS} + \mathbf{N} \quad (6.7)$$

Nous nous intéresserons principalement à la modélisation des 5 premières composantes. Hormis ces composantes astrophysiques, les données  $\mathbf{X}$  sont entachées de sources ponctuelles (galaxies infra-rouge, radio, quasars, etc)  $\mathbf{X}^{PS}$  ayant chacune leur propre loi d’émission. Autant dire que la matrice  $\mathbf{X}^{PS}$  n’est évidemment pas de rang 1. Le terme  $\mathbf{N}$  modélise le bruit instrumental.

Chacune des composantes que nous venons d’énumérer revêt un intérêt pour différentes communautés d’astrophysiciens. Nous avons choisi de nous intéresser plus particulièrement à l’extraction du fond diffus cosmologique. Celui-ci est une composante de premier ordre : selon la théorie du Big Bang, son spectre de puissance est une source d’information cruciale sur les paramètres intervenant dans les équations de Friedmann régissant l’évolution de notre Univers. Par ailleurs, tester la gaussianité ou la non-gaussianité du champ de fluctuation de température du CMB permet de tester certains modèles d’inflations. Les objectifs que nous nous fixons seront donc les suivants :

- **Estimation de la carte des fluctuations de température du CMB** : nous proposons d’appliquer l’algorithme de séparation de sources GMCA pour l’extraction de la carte de fluctuations de température du CMB  $\mathbf{s}^{cmb}$ . Une estimation correcte du CMB requiert une décomposition efficace du cube de données : c’est-à-dire l’estimation jointe d’un maximum de composantes galactiques.
- **Estimation du spectre de puissance du CMB** : la qualité du spectre de puissance calculé à partir du champ  $\mathbf{s}^{cmb}$  estimé sera notre critère de qualité principal.

## 6.2 Hypothèse du modèle linéaire

Au niveau du processus de séparation des données, la difficulté majeure provient des composantes de "Poussières" et "Synchrotron" dont le comportement peut être fortement éloigné de celui d'un mélange linéaire. Une première approche simple consiste à nier l'évidence et à supposer naïvement que ces deux composantes astrophysiques se comportent comme des mélanges linéaires :

$$\mathbf{X}^{sync} = \mathbf{a}^{sync} \mathbf{s}^{sync} \quad (6.8)$$

$$\mathbf{X}^P = \mathbf{a}^P \mathbf{s}^P \quad (6.9)$$

Outre sa simplicité, cette approximation peut être considérée comme valable sur certaines zones de la sphère des données (en particulier les hautes latitudes éloignées des fortes émissions galactiques). En effet, le caractère non linéaire (au sens du mélange) des comportements des composantes de "Poussières" et de "Synchrotron" est en partie lié aux variations de  $\beta^{sync}$ ,  $\beta^P$  et  $\mathbf{s}^P$ . Il s'avère que ces champs varient peu pour des portions du ciel éloignées du centre galactique. Sur de telles zones, le modèle de mélange linéaire est valable en première approximation. Notons de plus que l'estimation du CMB (ou d'autres composantes) sur les portions du ciel reste intéressante en particulier pour effectuer des tests de non-gaussianité.

Par la suite, nous proposons d'appliquer GMCA aux données Planck synthétiques sur des cartes carrées (modélisant des portions de ciel sur lesquelles l'approximation linéaire est justifiée) dans la Section 6.2.1.

Le consortium ESA/Planck a mis en place différents groupes de travail (Working Group) dont l'un (WG2) est dédié à la séparation de composantes. Au sein de WG2, des expérimentations de séparation sont proposées sur des données synthétiques réalistes proches de celles que Planck devraient fournir. Nous proposons dans la Section 6.3 d'appliquer GMCA sur de telles données synthétiques.

### 6.2.1 Modélisation linéaire sur cartes carrées

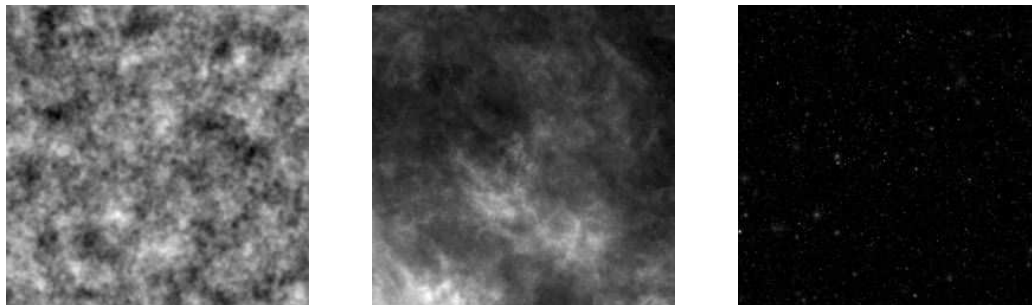
En première approximation, nous supposons que les différentes composantes astrophysiques se mélangent selon un modèle linéaire. Ainsi, l'observation au niveau du détecteur  $i$  est une combinaison linéaire de  $n$  composantes  $\{\mathbf{s}_j\}_{j=1,\dots,n}$  :  $\mathbf{x}_i = \sum_{j=1}^n A[i, j] \mathbf{s}_j + \mathbf{n}_i$ . Physiquement, les coefficients  $A[i, j]$  quantifient la loi d'émission de la source  $\mathbf{s}_j$  dans la  $i$ -ème bande de fréquence de l'instrument. Le terme  $\mathbf{n}_i$  modélise le bruit instrumental. Nous supposons de plus que la matrice de bruit  $\mathbf{N}$  (concaténation en colonne des vecteurs lignes  $\{\mathbf{n}_i\}_{i=1,\dots,n}$ ) est telle que ses éléments sont *i.i.d.* suivant une distribution gaussien centrée et de matrice de covariance  $\Sigma_{\mathbf{N}}$  diagonale connue.

Les premiers résultats présentés dans l'article (BMS<sup>+</sup>08) font intervenir un modèle de mélange particulièrement simple : seuls les 6 canaux Planck (HFI) de 100, 143, 217, 353, 545 et 857 GHz sont utilisés. Chaque canal est modélisé comme une combinaison linéaire

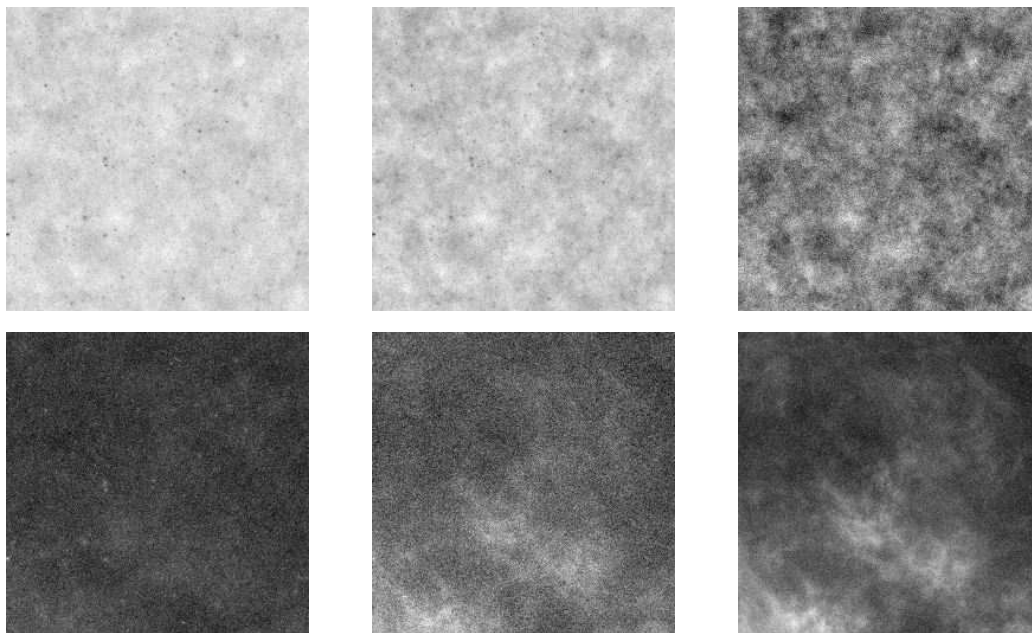
---

"réaliste" de 3 composantes : CMB, SZ et Poussière. Ces données ont également fait l'objet de l'étude présentée dans (MCSD05). Le seul effet instrumental est l'ajout de bruit dont la matrice de covariance  $\Sigma_N$  est diagonale. Chaque entrée de sa diagonale  $\Sigma_N[i, i] = \sigma_{i,N}^2$  est égale à la variance moyenne du bruit instrumental attendu au niveau du  $i$ -ème canal Planck. Les données  $\mathbf{X}$  sont des cartes carrées de taille  $\sqrt{t} = 300$  pixels correspondant à une zone de ciel de  $12,5^\circ$ . Les données sont représentées sur la figure 6.3. Les composantes initiales issues de simulations de CMB, SZ et Poussières sont visibles sur la figure 6.2.

## Résultats



**FIG. 6.2: Sources simulées - Gauche : CMB. Centre : poussière galactique. Droite : SZ.**



**FIG. 6.3: Données simulées Planck/HFI - RSB global (sur les 6 canaux) = 2.7dB**

Nous comparons GMCA avec d'autres méthodes déjà utilisées sur des données Planck

---

telles que JADE (voir (CS96)) et SMICA (voir (DCP03; MCSD05)). Cette étude propose en particulier d'évaluer la robustesse des ces différentes méthodes de séparation de sources en aveugle face au bruit instrumental. Les expériences ont été menées pour 7 valeurs différentes de RSB *global* variant de 1.7 à 16.7dB de sorte que la matrice de covariance du bruit  $\Sigma_N$  est proportionnelle à la valeur nominale attendue pour Planck. Le RSB global pour Planck sera de l'ordre de 10dB. Chaque point expérimental a été calculé à partir de 30 réalisations aléatoires de bruit avec des réalisations différentes de CMB, SZ et Poussières galactiques.

L'algorithme GMCA a été utilisé dans sa version "rapide" après application d'une transformation en ondelettes orthogonale bidimensionnelle sur les données. La valeur de  $P_{max}$  est fixée à 100 itérations. La figure 6.4 montre la moyenne des coefficients de corrélation des sources originales et de leurs estimées sur les 30 expériences pour différentes valeurs de RSB global. SMICA a été conçue pour l'extraction de champs gaussiens isotropes ; la diversité nécessaire à la séparation de sources est fondée sur la diagonalisation jointe par bande de fréquences (dans le domaine de Fourier) de la matrice de covariance des sources estimées. Il n'est donc pas étonnant que la carte de CMB soit bien estimée. GMCA tend à avoir un comportement proche de celui de SMICA pour le CMB. La composante de poussières galactiques est également bien estimée par SMICA et GMCA. La composante SZ est celle qui s'éloigne le plus des hypothèses du cadre de SMICA ; le SZ est en effet lié à des amas de galaxies spatialement très concentrés. Là encore, le comportement défaillant de SMICA est prévisible. En terme de coefficient de corrélation, GMCA présente donc des performances similaires à SMICA pour le CMB et les poussières ; GMCA se comportant également correctement pour l'extraction de la composante SZ.

Comme nous l'avons souligné à maintes reprises, le critère sur la matrice de mélange  $C_A$  est plus sensible aux erreurs de séparation dans un contexte de données bruitées. La dernière ligne de la figure 6.4 montre l'évolution du critère  $C_A$  lorsque la variance globale du bruit instrumental varie. Ce critère tend à discriminer plus finement ces trois méthodes en faveur de GMCA.

### Contrainte sur la matrice de mélange

Dans une configuration plus réaliste, la séparation n'est que partiellement aveugle. En effet, la loi d'émission du CMB est celle d'un corps noir à 2,375 °K. La colonne de la matrice  $A$  correspondant au CMB est de fait extrêmement bien connue. Dans ce paragraphe, nous montrons qu'il est facile d'ajouter cette contrainte forte dans GMCA. Au cours de l'expérience proposée, CMB-GMCA est mis au point en contraignant la colonne de  $A$  liée au CMB. La figure 6.5 montre l'évolution du coefficient de corrélation entre les sources originales et les sources estimées avec et sans information *a priori* sur la matrice  $A$ . Comme prévu, l'estimation de la carte de CMB est meilleure lorsque l'information *a priori* est ajoutée. Cette amélioration est également perceptible sur les autres composantes ; cela n'a rien d'étonnant : lorsque la colonne de  $A$  liée au CMB est fixée, seuls 2/3 des coefficients de  $A$  restent à estimer ; diminuant d'autant la complexité du problème.

## 6.3 Application aux données Planck/WG2

### 6.3.1 Des données réalistes

Les expériences de cette partie ont été réalisées dans le cadre du groupe de travail Planck dédié à la séparation de composantes (Planck-WG2). Au sein de ces groupes de travail, des "défis" de séparation de composantes ont cherché à évaluer différentes méthodes de séparation de composantes sur des données simulées réalistes. Nous présentons ici les 4 meilleurs résultats parmi 10 laboratoires ayant pris part à la seconde série de tests de séparation entre 2007 et 2008 (Challenge-2).

**Données et effets instrumentaux** Les données simulées fournies pour ce test de séparation sont les 9 canaux Planck dans les bandes de fréquence : 30, 44, 70, 100, 143, 217, 353, 545 et 857 GHz. Ces observations sont celles qui devraient être fournies par le consortium Planck à l'issue d'une année d'observation. Chaque observation est une carte échantillonnées sur la sphère utilisant la pixellisation HEALPix<sup>7</sup>. Le nombre total de pixels par canal est  $t = 50331648$ . Les 9 canaux Planck fournissent un jeu de données colossal contenant près de 450 millions de pixels. Les effets instrumentaux modélisés au niveau des données WG2/CH2 sont les suivants :

- **Bruit instrumental non-corrélé et non-stationnaire** : le bruit instrumental au niveau de la carte du ciel reconstruite à partir d'une année d'observation est lié à la manière dont le ciel est balayé par Planck. Plus précisément, pour le pixel  $k$  du canal  $i$ , la contribution  $\mathbf{N}[i, k]$  aura une variance inversement proportionnelle au nombre de fois où la zone du ciel correspondante aura été balayée. En première approximation, le bruit sera considéré gaussien, de moyenne nulle et tel que :

$$\forall i \text{ et } k \neq k'; \quad \mathbb{E}\{\mathbf{N}[i, k]\mathbf{N}[i, k']\} = 0 \quad (6.10)$$

$$\forall k \text{ et } i \neq i'; \quad \mathbb{E}\{\mathbf{N}[i, k]\mathbf{N}[i', k]\} = 0 \quad (6.11)$$

$$\forall k \text{ et } i; \quad \mathbb{E}\{\mathbf{N}[i, k]^2\} = \sigma_{\mathbf{N}i,k}^2 \quad (6.12)$$

où  $\sigma_{\mathbf{N}i,k}^2$  est la variance effective du bruit au pixel  $k$  du canal  $i$ . Etant lié à l'instrument et au balayage du ciel par le satellite, cette variance effective est connue.

- **Résolution et convolution** : le ciel est observé par Planck dans différentes bandes de fréquence. L'ouverture de l'instrument étant fixe, chaque canal possède donc une résolution spatiale différente. Le tableau suivant présente la résolution de chaque canal (en arcminute). L'ouverture circulaire de l'instrument  $D$  étant finie, il est équivalent de considérer que les données sont convoluées avec un noyau gaussien dont la largeur à mi-hauteur est la résolution de l'instrument.

---

<sup>7</sup>Voir <http://healpix.jpl.nasa.gov/>.

Dans les simulations proposées, ce noyau de convolution est isotrope et parfaitement connu.

Chaque observation effective  $\mathbf{y}_i$  peut être modélisée comme suit :

$$\forall i = 1, \dots, m; \quad \mathbf{y}_i = (\mathbf{h}_i \star \mathbf{x}_i) + \mathbf{n}_i \quad (6.13)$$

où  $\mathbf{x}_i$  est l'observation au canal  $i$  idéale : non-convoluée et sans bruit. Le terme  $\mathbf{h}_i$  correspond au noyau de convolution pour le canal  $i$ . Les données  $\mathbf{Y}$  de WG2/CH2 sont visibles sur la Figure 6.6.

### 6.3.2 Un modèle physique pour GMCA

#### Modélisation des composantes astrophysiques

Les composantes astrophysiques en jeu dans les données simulées WG2/CH2 sont les suivantes :

- **Fond diffus cosmologique (CMB)** : nous ne ferons aucune hypothèse sur le spectre de puissance du CMB qui sera l'un de nos paramètres d'intérêt. L'information *a priori* d'importance que nous utiliserons est le comportement spectral du CMB (*i.e.* spectre d'un corps noir à 2,375 °K). La colonne de la matrice de mélange  $\mathbf{A}$  sera donc fixée.
- **Effet Sunyaev Zel'Dovich (SZ)** : de façon analogue aux hypothèses faites pour le CMB, le modèle de mélange linéaire du SZ dans les bandes de fréquences observées par Planck est bien vérifié. De plus, le comportement spectral du SZ est également bien connu. La colonne de  $\mathbf{A}$  liée à la composante de SZ sera également supposée connue.
- **Emission "Free-Free"** : en première approximation, le modèle de mélange linéaire est bien vérifié pour la composante "Free-Free". Son comportement spectral est celui d'une loi de puissance de paramètre (indice spectral)  $\beta^{FF} \simeq 2,15$  (DC05).
- **Émissions "Synchrotron"** : les modèles physiques prévoient que les émissions synchrotron possèdent un comportement spectral en loi de puissance d'indice spectral variant spatialement. Le modèle de mélange linéaire n'est plus fixé. Bonaldi *et al.* (BBS<sup>+</sup>06) ont proposé d'approximer le comportement du synchrotron comme celui

Canal (en GHz)	30	44	70	100	143	217	353	545	857
Résolution (en arcminute)	33	24	14	10	7	5	5	5	5

Tab. 6.1: Résolution spatiale par canal



d'un mélange linéaire dont l'indice spectral  $\beta^{sync}$  est à estimer. Cette hypothèse devrait être vérifiée pour les grandes structures du synchrotron (grandes échelles). De plus, Haslam *et al.* (HSSW82) ont proposé une carte de synchrotron à une résolution proche du quart de celle qui sera atteinte par Planck. Nous proposons d'ajouter une étape d'estimation dans l'algorithme GMCA durant laquelle l'indice spectral  $\beta^{sync}$  est estimé à partir d'un cube de résidu  $\mathbf{R} = \mathbf{X} - \sum_{i \neq sync} \mathbf{a}^i \mathbf{s}_i$  et de  $\mathbf{s}_{Haslam}$ , la carte de synchrotron à basse résolution. L'estimation de  $\beta^{sync}$  s'effectue à partir de signaux à résolution identique (ramenée à celle de la carte d'Haslam) comme suit :

$$\beta^{sync} = \underset{\beta}{\text{Argmin}} \left\| \mathbf{R} - \mathbf{a}^{sync}(\beta) \mathbf{s}_{Haslam} \right\|_F \quad (6.14)$$

- **Composantes de "Poussières"** : la composante d'émissions thermiques de "Poussières" se comportent comme un mélange non-linéaire. Cette composante est plus complexe à modéliser puisqu'elle fait intervenir deux jeux de paramètres (température et indice spectral) variant spatialement. Fort heureusement, les comportements spectraux du CMB et de la poussière sont opposés : le CMB domine dans les canaux basse fréquence Planck alors que la poussière domine dans les canaux haute fréquence. De plus, la composante de poussière est une composante très structurée admettant une représentation parcimonieuse en ondelettes. GMCA ayant été conçu pour être sensible à ce type de composantes, il semble raisonnable de capturer la composante (ou plutôt devrions-nous parler des composantes) de poussière en aveugle dans un premier temps. Une composante anormale de poussière (le "spinnig dust") est également incluse dans les données WG2/CH2. Le niveau de cette composante étant plus faible nous l'ignorons dans ces tests.
- **Sources ponctuelles** : les sources ponctuelles radio ou infrarouges sont des phénomènes ponctuels. Chaque source présentant un spectre propre, les sources ponctuelles sont plutôt des composantes de nuisance dans l'objectif d'extraire les composantes diffuses telles que le CMB. Leur détection est l'un des problèmes soulevés lors du traitement des données (voir (SJFC<sup>+</sup>08)). A l'issue de leur détection, elles sont le plus généralement masquées.

### Algorithme GMCA pour Planck

La linéarité du modèle de SSA présente un avantage majeur dès lors que l'estimation de grandeurs physiques est nécessaire : par exemple, connaissant les propriétés statistiques du bruit, la linéarité du modèle permet très simplement de déduire les propriétés statistiques du bruit au niveau des sources. La version de GMCA que nous allons décrire traite donc les données sous l'angle d'un mélange linéaire de composantes. Au-delà de l'algorithme déjà décrit au Chapitre 4, la méthode GMCA est adaptée pour contraindre certaines composantes à vérifier les propriétés physiques décrites lors du paragraphe précédent.

Dans le cadre du modèle linéaire de mélange, chaque observation  $\{\mathbf{x}_i\}_{i=1, \dots, m}$  s'écrit comme

la combinaison linéaire de  $n$  composantes ou sources. Les données réelles seront dans ce cas des versions convoluées de ces observations idéales de sorte que les données réellement observées seront définies comme suit :

$$\forall i = 1, \dots, m; \quad \mathbf{y}_i = (\mathbf{h}_i \star \mathbf{x}_i) + \mathbf{n}_i \quad (6.15)$$

$$= \sum_{j=1}^n A[i, j](\mathbf{h}_i \star \mathbf{s}_j) + \mathbf{n}_i \quad (6.16)$$

Le modèle de mélange n'est donc plus linéaire au niveau des données observées  $\mathbf{Y}$ <sup>8</sup>. Deux approches peuvent être envisagées : une approche rigoureuse consisterait à déconvoluer chaque observation avant d'appliquer GMCA.

- **Déconvolution des données** : chaque observation est déconvoluée séparément. Les données possédant une structure intercanale, une déconvolution séparée de chaque canal n'est pas appropriée. Au-delà de ces considérations, certains noyaux de convolution (en particulier pour les premiers canaux basse fréquence) conduisent à une déconvolution mal conditionnée.
- **Adapter GMCA à la déconvolution jointe** : l'algorithme GMCA est constitué de l'estimation successive des sources et de la matrice de mélange. L'étape d'estimation des sources pour  $\mathbf{A}$  fixée pourrait intégrer la déconvolution.

De part la taille colossale des données, nous avons tout d'abord préféré appliquer GMCA sur les données  $\mathbf{Y}$ . La convolution est ensuite compensée au niveau de la carte de CMB estimée. En première approximation, les structures diffuses les plus significatives contenues dans les données sont fortement dominées par leurs fréquences les plus faibles pour lesquelles l'opérateur de convolution a le moins d'impact. Ce dernier point permet donc de justifier l'estimation de la matrice de mélange  $\mathbf{A}$  à partir des données  $\mathbf{Y}$ .

L'algorithme GMCA adapté au traitement des données Planck est le suivant :

---

<sup>8</sup>Ici, il s'agit d'une non-linéarité au sens du modèle de mélange ; le modèle reste linéaire au sens usuel puisqu'il est défini comme une composition d'opérateurs linéaires.

1. Transformation en ondelettes sur la sphère des données  $\alpha_Y$  :
2. Fixer le nombre d'itérations  $P_{\max}$  et du seuil  $\gamma^{(0)}$   

$$\alpha_Y = Y\Phi^T$$
3. Tant que les seuils  $\gamma^{(0)}$  sont supérieurs à  $\gamma_{\min}$  (e.g. fonction de la variance du bruit),
  - Estimation des sources  $\alpha_S$  à l'itération  $h$  en supposant la matrice  $\mathbf{A}$  fixée :  

$$\alpha_S^{(h+1)} = \mathcal{H}_{\gamma^{(h)}} \left( \mathbf{A}^{\dagger(h)} \alpha_Y \right) :$$
  - Mise-à-jour des colonnes de  $\mathbf{A}$  libres notées  $\mathbf{A}^\circ$  en supposant  $\alpha_S$  fixée :  

$$\tilde{\mathbf{A}}^\circ{}^{(h+1)} = \alpha_Y \tilde{\alpha}_S^{(h)T} \left( \tilde{\alpha}_S^{(h)} \tilde{\alpha}_S^{(h)T} \right)^{-1}$$
  - Estimation de l'indice spectral  $\beta^{\text{sync}}$  du synchrotron à partir du résidu  $\alpha_R = \alpha_Y - \sum_{i \neq \text{sync}} \mathbf{a}^i \alpha_{S_i}$  :  

$$\beta^{\text{sync}} = \text{Argmin}_\beta \|\alpha_R - \mathbf{a}^{\text{sync}} \alpha_{\text{Haslam}}\|_2$$
  - Décroissance du seuil  $\gamma^{(h)}$ .

### 6.3.3 Résultats

Nous nous concentrerons essentiellement sur l'estimation de la carte de CMB et de son spectre de puissance.

#### Cartes et résidus

La figure 6.7 représente la carte CMB simulée originale (en haut) et la composante CMB estimée à l'aide de GMCA (en bas). Notons que l'algorithme GMCA estime une matrice de mélange  $\mathbf{A}$  utilisée ensuite pour estimer une version bruitée de la carte de CMB obtenue par application du pseudo-inverse de  $\mathbf{A}$  aux données  $\mathbf{Y}$ . Dans l'hypothèse d'une séparation parfaite des composantes, la carte CMB estimée est isotrope et donc entièrement caractérisée par son spectre de puissance. La convolution des données engendre une perte de puissance qu'il est possible de compenser par le calcul d'un filtre effectif fonction des noyaux de convolution  $\{\mathbf{h}_i\}_{i=1, \dots, m}$  et de la matrice de mélange  $\mathbf{A}$ . Le débruitage de la carte de CMB est ensuite effectué via un filtrage de Wiener. La carte GMCA estimée présentée à la figure 6.7 est donc la carte finale après filtrage de Wiener. Le filtrage s'effectue dans l'espace des harmoniques sphériques (équivalent du domaine de Fourier sur la sphère) ; il est bien établi que le filtrage de Wiener induit un biais d'autant plus important que le rapport-signal-à-bruit (RSB) est faible. En première approximation, le spectre du bruit est plat (stationnaire dans le domaine des harmoniques sphériques). Le spectre spatial du CMB ayant une décroissance de l'ordre de  $1/\ell$ , le biais de la composante filtrée sera important pour les valeurs de  $\ell$  grandes. Ce point explique donc la différence visuelle entre la carte originale et la carte de CMB estimée qui fait apparaître une déficience de structures haute fréquence (grandes valeurs de  $\ell$ ) sur la carte estimée.

Afin d'analyser plus finement la carte estimée, l'observation du résidu  $\mathbf{r}_{CMB} = \mathbf{s}_{CMB} - \tilde{\mathbf{s}}_{CMB}$  (entre la carte d'entrée et la carte estimée) convolué par un noyau gaussien permet une évaluation plus adéquate de l'estimation des grandes structures (petites valeurs de  $\ell$ ). Nous proposons ici de comparer la carte estimée par GMCA avec les groupes ADAMIS, CCA et MEM dont les méthodes sont décrites dans (SJFC<sup>+</sup>08). Les figures 6.8 et 6.9 présentent les résidus convolués à l'aide d'un noyau gaussien de 20 arcminutes (correspondant à une largeur à mi-hauteur à  $\ell = 500$ ). A première vue, le résidu aux grandes latitudes (en dehors du centre galactique) semble relativement faible comparé aux autres résultats proposés. Le centre galactique fait essentiellement apparaître un résidu de poussières galactiques. L'ajout d'une contrainte supplémentaire dédiée à l'estimation d'une composante de "poussières" devrait améliorer ce résultat.

Les figures 6.10 et 6.11 représentent les mêmes résidus convolués par un noyau gaussien de 60 arcminutes dont la largeur à mi-hauteur est de l'ordre de  $\ell = 180$ . L'observation de ce résidu convolué fait essentiellement apparaître les très grandes structures. Il soutient les observations faites à partir du résidu convolué à 20 arcminutes : GMCA avec contraintes physiques permet d'obtenir une composante de CMB dont les grandes échelles sont préservées en dehors du centre galactique.

La figure 6.12 montre l'écart-type des résidus convolués à 20 (en haut) et 60 arcminutes (en bas) par bandes de latitude de largeur  $20^\circ$  ; la valeur  $90^\circ$  correspondant au centre galactique. Cette dernière étant fortement contaminée par les composantes galactiques, elle n'a pas été représentée. Le graphique du bas correspondant au résidu convolué à 60 arcminutes tend à confirmer le bon comportement de GMCA en dehors du centre galactique. De façon étonnante, le résultat est légèrement différent pour le résidu convolué à 20 arcminutes puisqu'il fait apparaître un résidu plus faible pour ADAMIS. Après analyse plus fine des cartes estimées, il semble que cet effet soit dû à la compensation directe de la déconvolution. Cette dernière tend à faire apparaître des effets d'anneaux (*ringing*) autour des structures résiduelles de sources ponctuelles ou de composante galactique. Une gestion plus rigoureuse de la déconvolution devrait permettre d'atténuer cet effet. Ce travail est en cours.

### Non-gaussianités

L'étude de la non-gaussianité de la carte de CMB estimée est fondamentale (AFB01). Au-delà du test du caractère gaussien de la carte, la détection de non-gaussianités dans la carte de CMB est un enjeu pour tester certains modèles d'inflation de l'Univers (tels que les

modèles de  $f_{NL}$  (YW08)). Les données CH2/WG2 n'incluant pas de tels modèles de non-gaussianité du CMB, il peut être intéressant d'évaluer l'influence des résidus de séparation sur le caractère gaussien de la carte estimée. La champ purement gaussien se caractérisant par des cumulants d'ordre supérieur nuls, Starck *et al.* (SAF04; JSD<sup>+</sup>05) ont proposé l'utilisation de méthodes d'analyse multi-échelles pour la détection de non-gaussianité. L'analyse du kurtosis par bande d'ondelettes de la carte estimée est un détecteur robuste de non-gaussianités.

La figure 6.13 représente le kurtosis par échelle d'ondelettes des cartes GMCA, ADAMIS, CCA et GIBBS. Notons que la méthode GIBBS a été conçue pour l'estimation des très grandes structures du CMB (*i.e.*  $\ell < 200$ ). Les sources de non-gaussianité classiques sont : i) les résidus de composantes galactiques après séparation et ii) les sources ponctuelles. Pour pallier en partie l'impact des sources ponctuelles sur le calcul du kurtosis par échelle d'ondelettes, les sources connues sont masquées. Elle sont ensuite éventuellement compensées par interpolation des données manquantes (IDM - voir Chapitre VII). L'IDM des données masquées (voir (AMS<sup>+</sup>07)) permet de combler les données manquantes par des structures statistiquement proches de celles qui les entourent. Pour certaines applications, cette opération d'interpolation permet d'éviter une gestion complexe de données manquantes. La figure 6.13 présente le kurtosis par échelle d'ondelettes pour les cartes masquées (trait discontinu) et après IDM (trait continu). Quelle que soit la méthode de séparation utilisée, le kurtosis augmente pour les échelles les plus fines ( $\ell > 1000$ ) caractérisant *a priori* des sources ponctuelles résiduelles. Une augmentation plus contenue est également visible pour les grandes échelles ( $\ell < 150$ ); ce phénomène peut être la conséquence de résidus galactiques en particulier sur l'équateur. Notons que l'impact de GMCA et CCA sur la non-gaussianité semble plus limité. Bien que prévue pour l'estimation des grandes échelles du CMB, la méthode GIBBS semble peu efficace pour limiter la contamination galactique. Une étude complémentaire d'intérêt consisterait à inclure un modèle de non-gaussianité lors des simulations WG2 afin de caractériser l'impact de ces différentes méthodes pour la détection de non-gaussianité dans les cartes de CMB estimées.

### Spectre de puissance

La mission Planck devrait permettre d'estimer le spectre de puissance du CMB avec une erreur plus faible que son prédécesseur WMAP jusqu'à  $\ell \sim 3000$ . Ceci permettra à l'évidence d'estimer plus finement les paramètres cosmologiques du modèle d'inflation  $\Lambda$ -CDM<sup>9</sup>. La figure 6.14 présente les spectres de puissance estimés à partir des cartes GMCA, ADAMIS, CCA et MEM ainsi que le spectre de la carte originale (en trait noir). Afin de discriminer ces différentes méthodes, la figure 6.15 présente l'erreur d'estimation du spectre de puissance :

$$e(\ell) = |C(\ell) - \tilde{C}(\ell)| \quad (6.17)$$

---

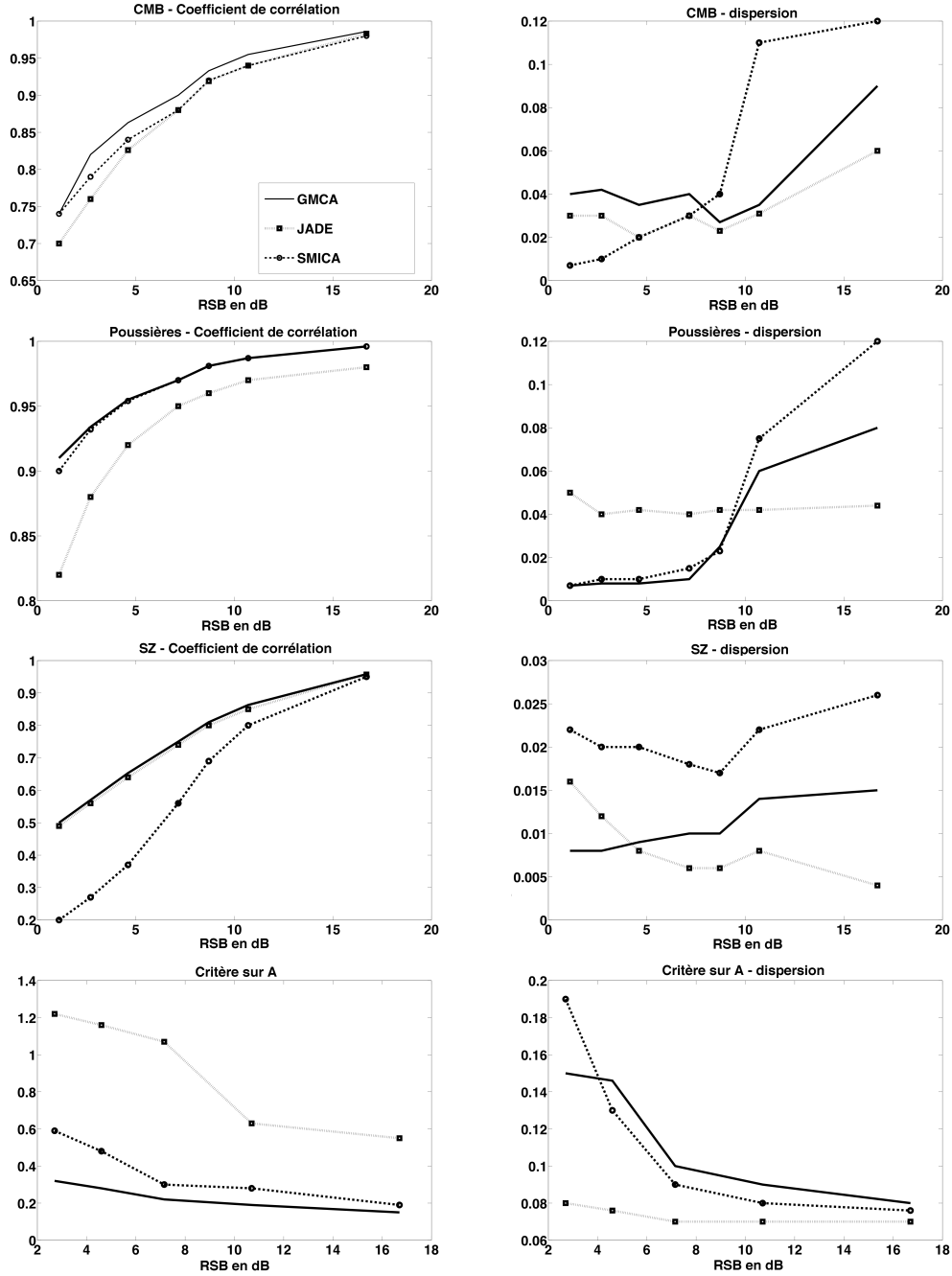
<sup>9</sup> $\Lambda$ -Cold Dark Matter : il s'agit d'un modèle de Big Bang intégrant une contribution d'énergie noire (via la constante cosmologique  $\Omega_\Lambda$ ) et de matière noire froide - voir <http://space.mit.edu/home/tegmark/varenna.ps>.

De façon intéressante, hormis la méthode MEM qui présente un biais important à  $\ell = 200$  et  $1300 < \ell < 1800$ , les différentes méthodes présentent des résultats d'estimation du spectre de puissance tout à fait comparables.

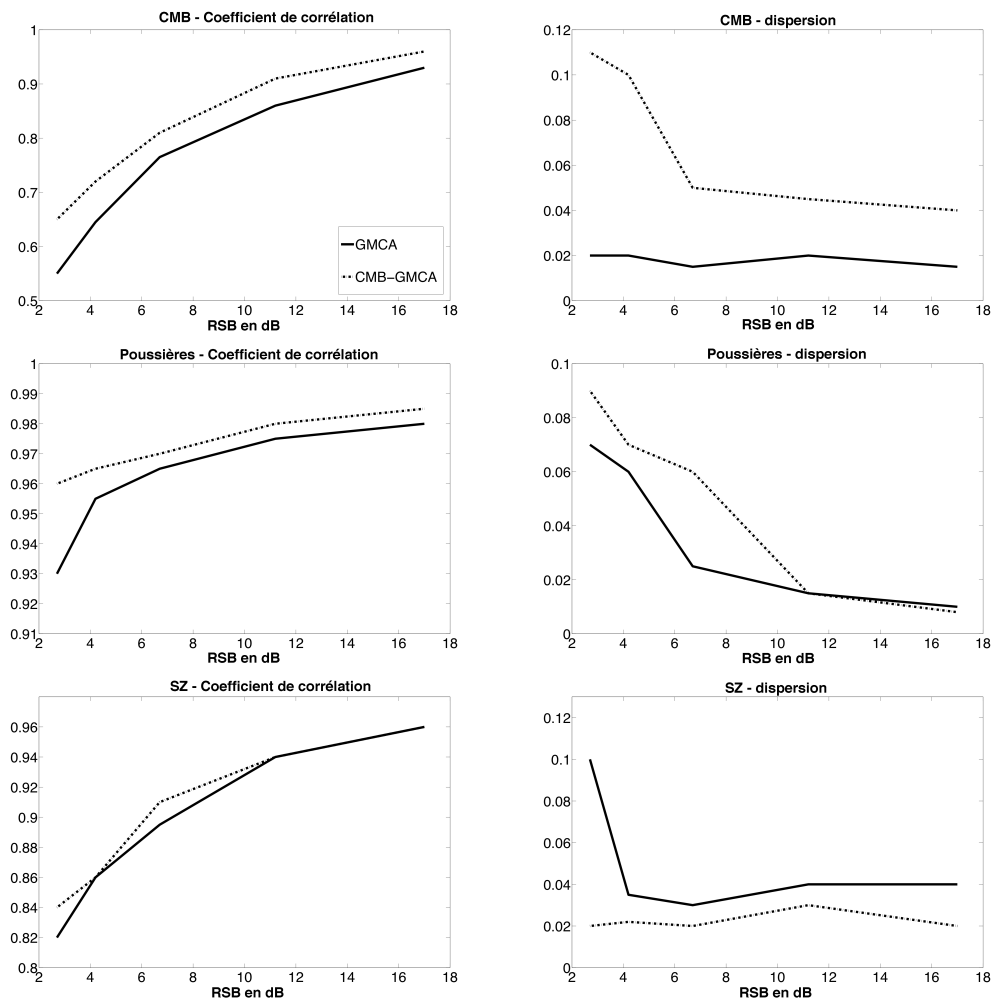
### GMCA et Planck : quelle suite ?

Ces différents résultats ont montré le potentiel de GMCA pour l'estimation d'une carte de CMB relativement peu contaminée en dehors du centre galactique. Différents points restent à analyser :

- **Déconvolution** : comme nous l'avons vu, une gestion plus fine de la déconvolution pourrait permettre une amélioration de l'estimée de la carte de CMB. Une étude en cours est celle de la déconvolution multicanale jointe à l'estimation de la matrice au sein de l'algorithme GMCA.
  - **Débruitages de la carte** : par souci de simplicité et de rapidité (au regard de la taille importante des données), le filtrage de Wiener est effectué intégralement dans le domaine des harmoniques sphériques sous hypothèse de stationnarité du bruit. Les paramètres statistiques du bruit étant connus, il semble possible de construire un estimateur de la carte de CMB tenant compte du caractère non-stationnaire du bruit.
  - **Modélisation de l'émission thermique de poussière** : au même titre que l'émission de synchrotron, il semble tout à fait intéressant d'estimer une composante de poussière paramétrée par un indice spectral  $\beta^P$  et une température  $T_P$ .
-

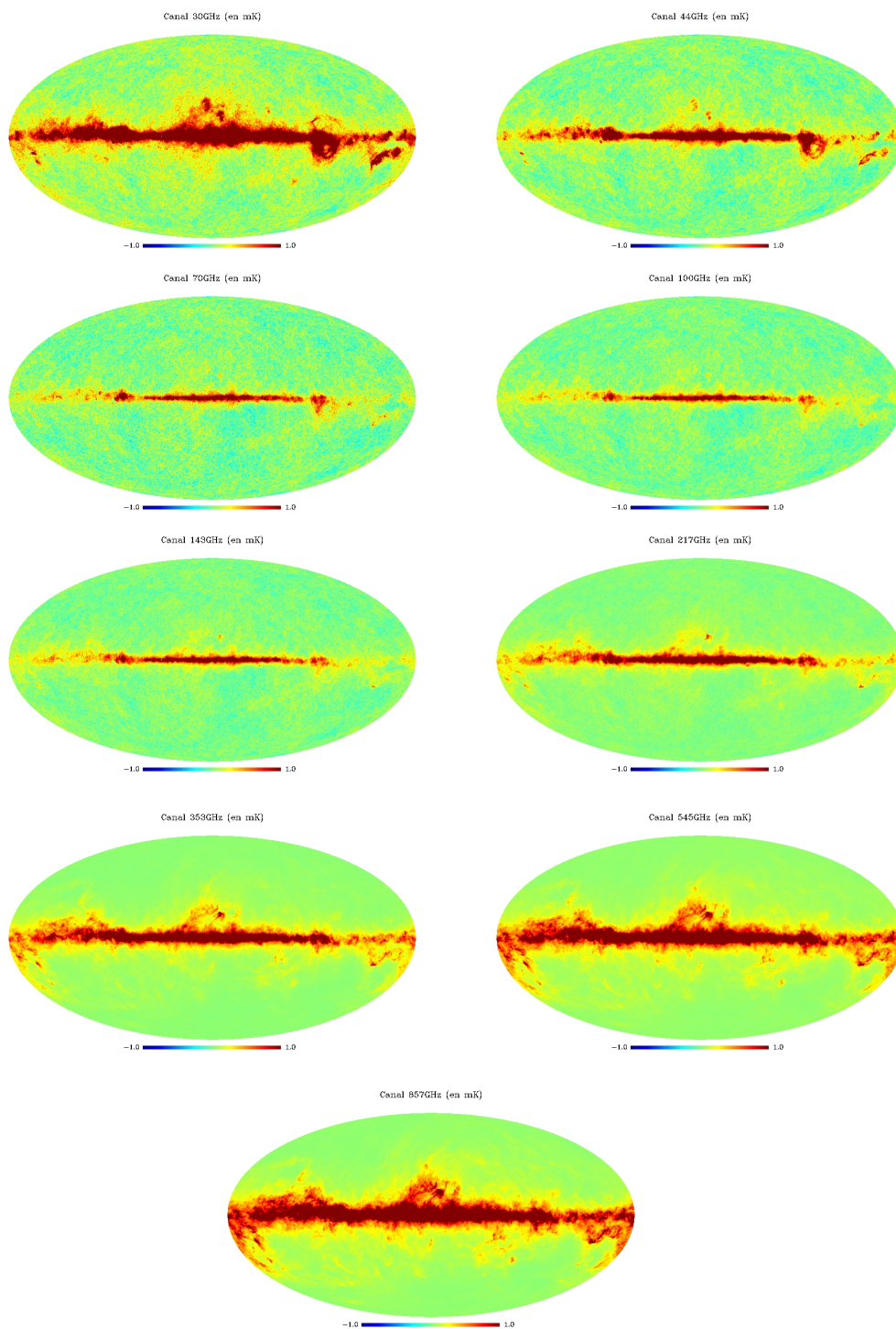


**FIG. 6.4:** Colonne de gauche : coefficients de corrélation entre les sources originales et leurs estimées - Colonne de droite : écart-type des coefficients de corrélation. Première ligne : CMB. Seconde ligne : poussière galactique. Troisième ligne : composante SZ. Quatrième ligne : critère sur la matrice de mélange  $C_A$ . Légende : JADE : ligne pointillée avec  $\square$  - SMICA : tirets avec  $\circ$  - GMCA : ligne continue.

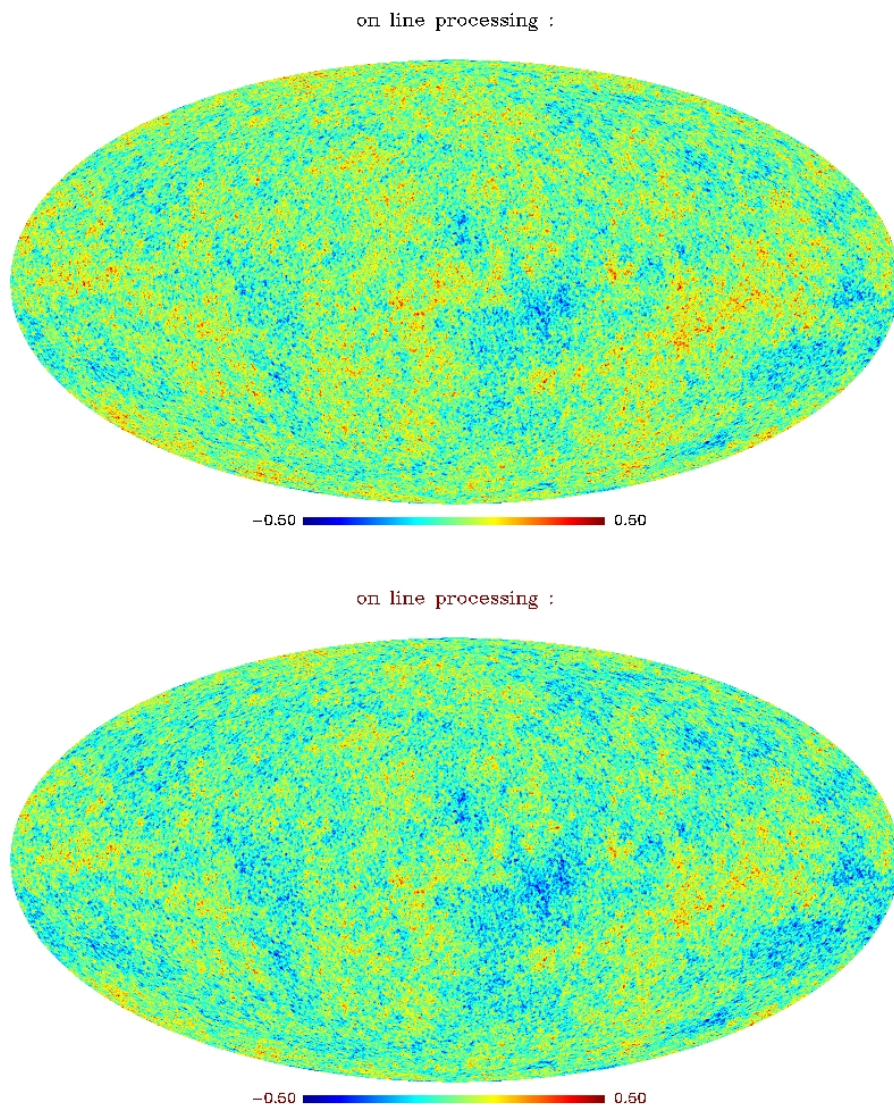


**FIG. 6.5:** Colonne de gauche : coefficients de corrélation entre les sources originales et leurs estimées - Colonne de droite : écart-type des coefficients de corrélation. Première ligne : CMB. Seconde ligne : poussière galactique. Troisième ligne : SZ. Légende : GMCA avec *a priori* : ligne en pointillé - GMCA : ligne continue. Abscissa : RSB in dB.





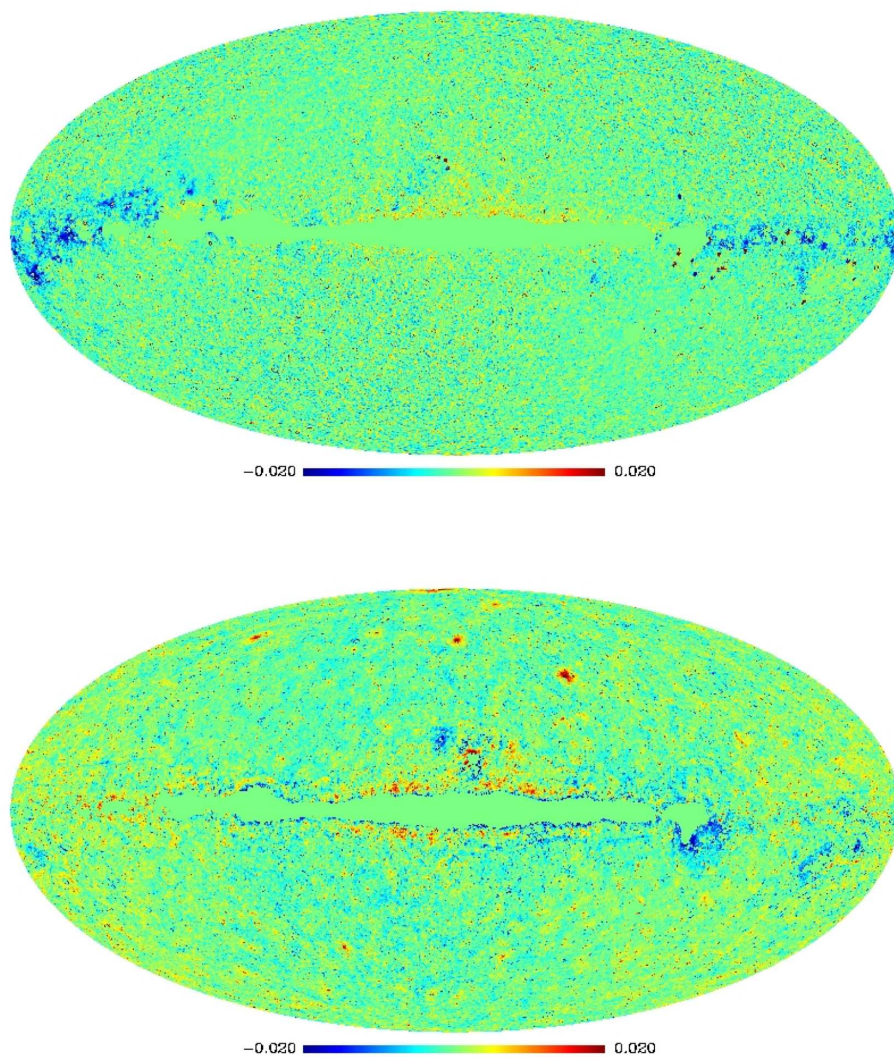
**FIG. 6.6:** Données WG2/CH2.



**FIG. 6.7:** Haut : composante CMB originale. Bas : composante CMB estimée par GMCA. Unité :  $10^{-3}\text{K}$ .

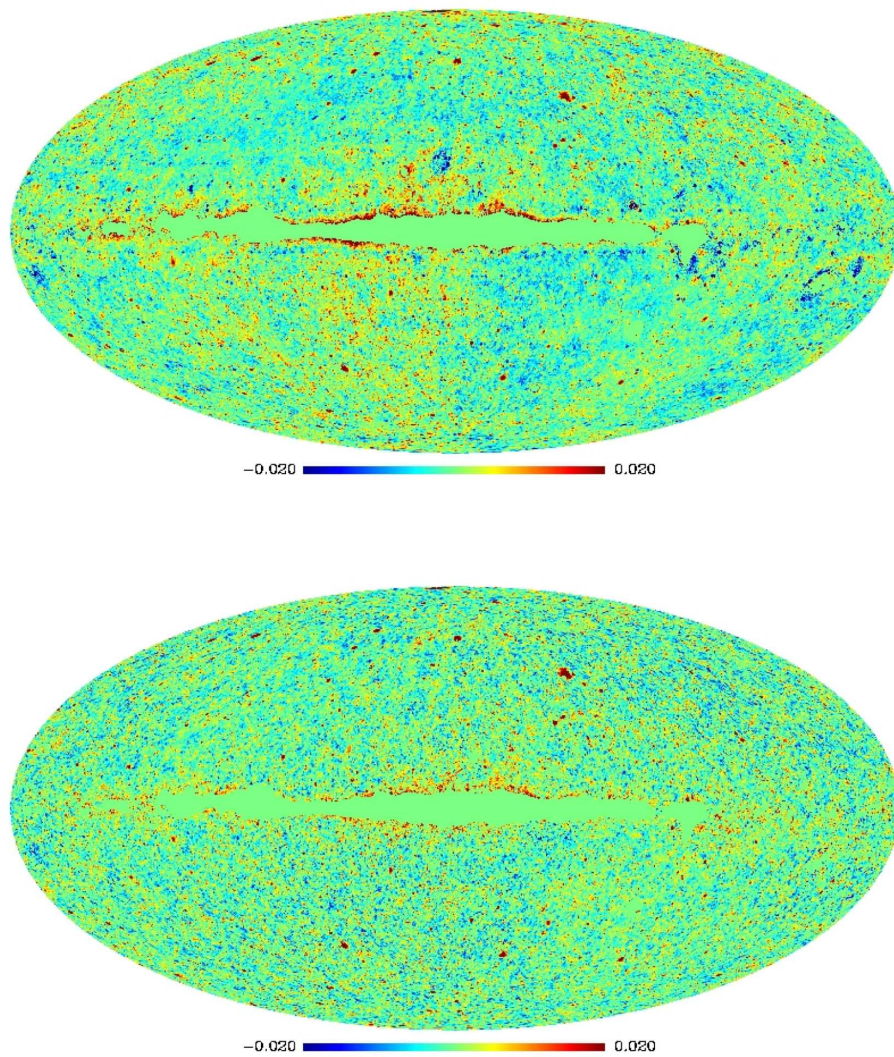
---



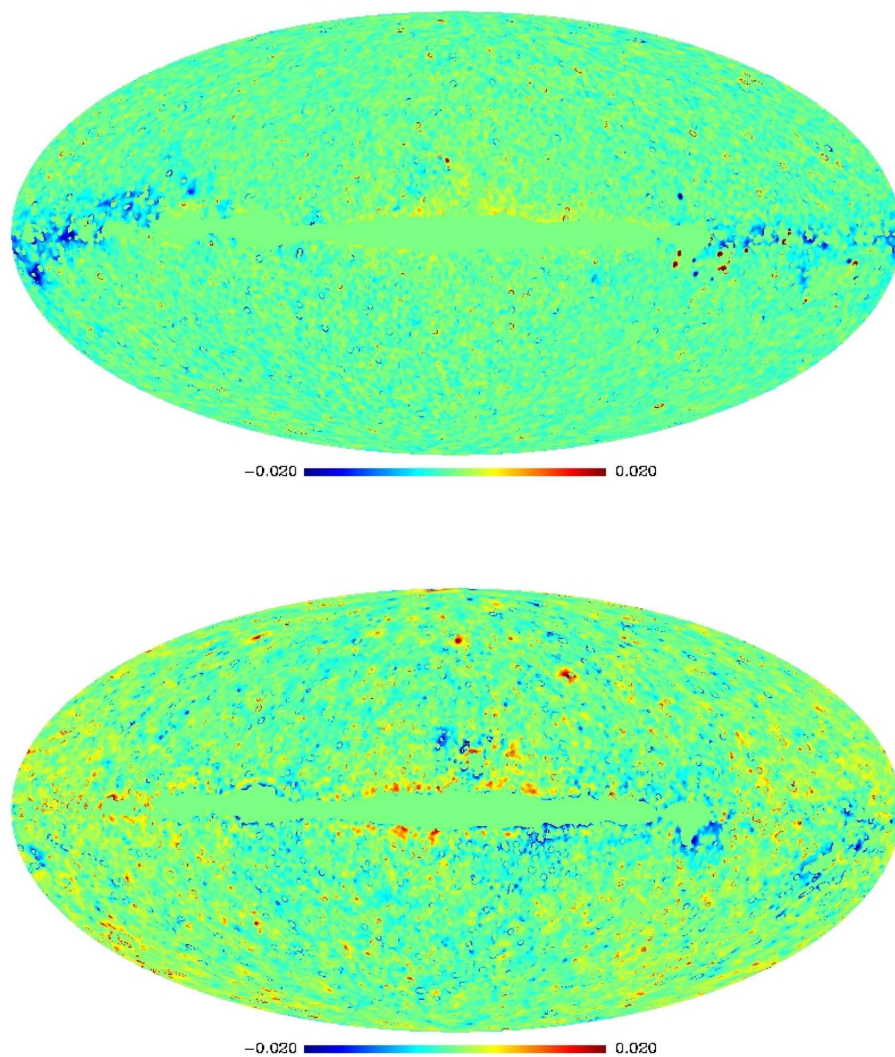


**FIG. 6.8:** **Haut** : résidu GMCA convolué à 20 arcminutes. **Bas** : résidu ADAMIS convolué à 20 arcminutes. **Unité** :  $10^{-3}\text{K}$ .

---



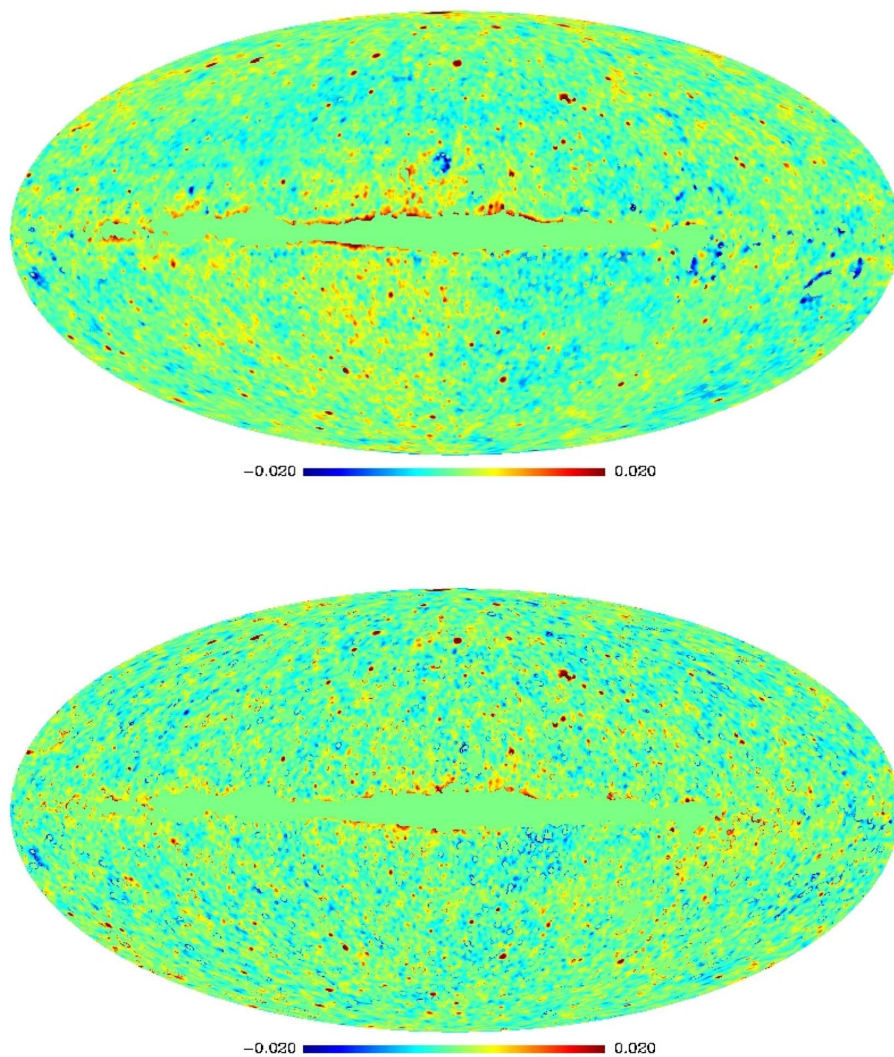
**FIG. 6.9:** Haut : résidu MEM convolué à 20 arcminutes. Bas : résidu CCA convolué à 20 arcminutes.  
Unité :  $10^{-3}\text{K}$ .



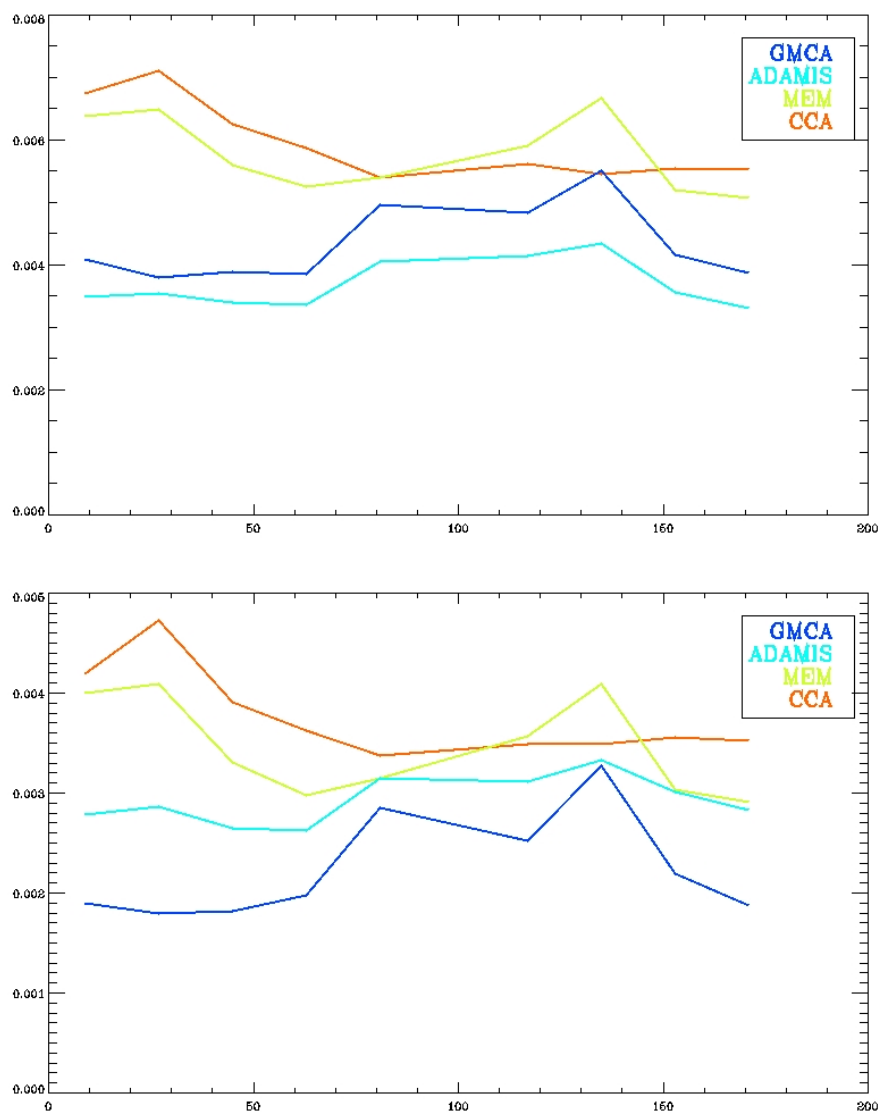
**FIG. 6.10:** **Haut :** résidu GMCA convolué à 60 arcminutes. **Bas :** résidu ADAMIS convolué à 60 arcminutes. **Unité :**  $10^{-3}\text{K}$ .

---

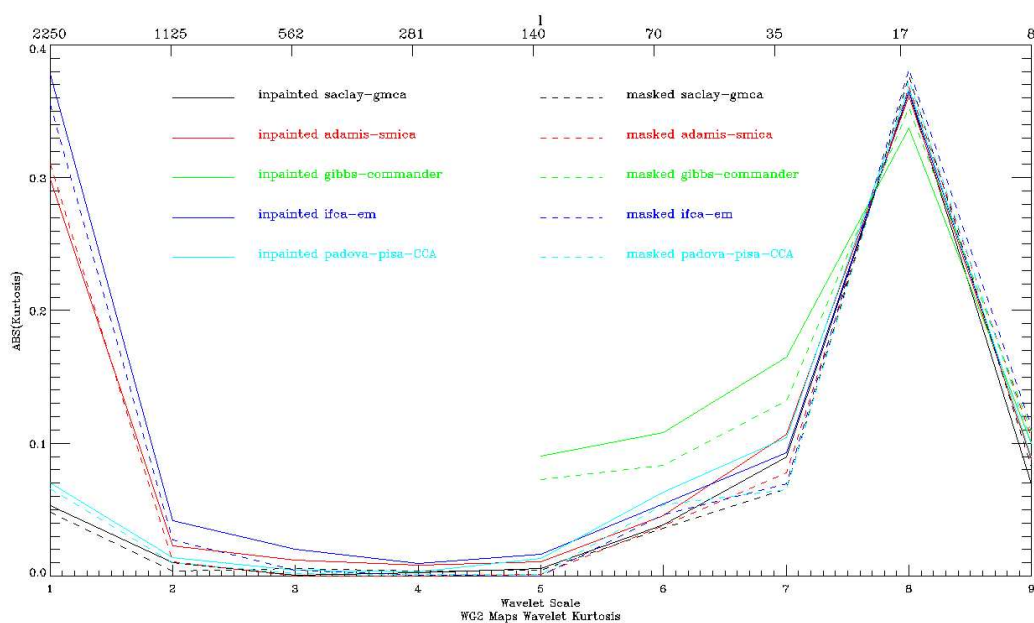




**FIG. 6.11:** Haut : résidu MEM convolué à 60 arcminutes. Bas : résidu CCA convolué à 60 arcminutes. Unité :  $10^{-3}\text{K}$ .

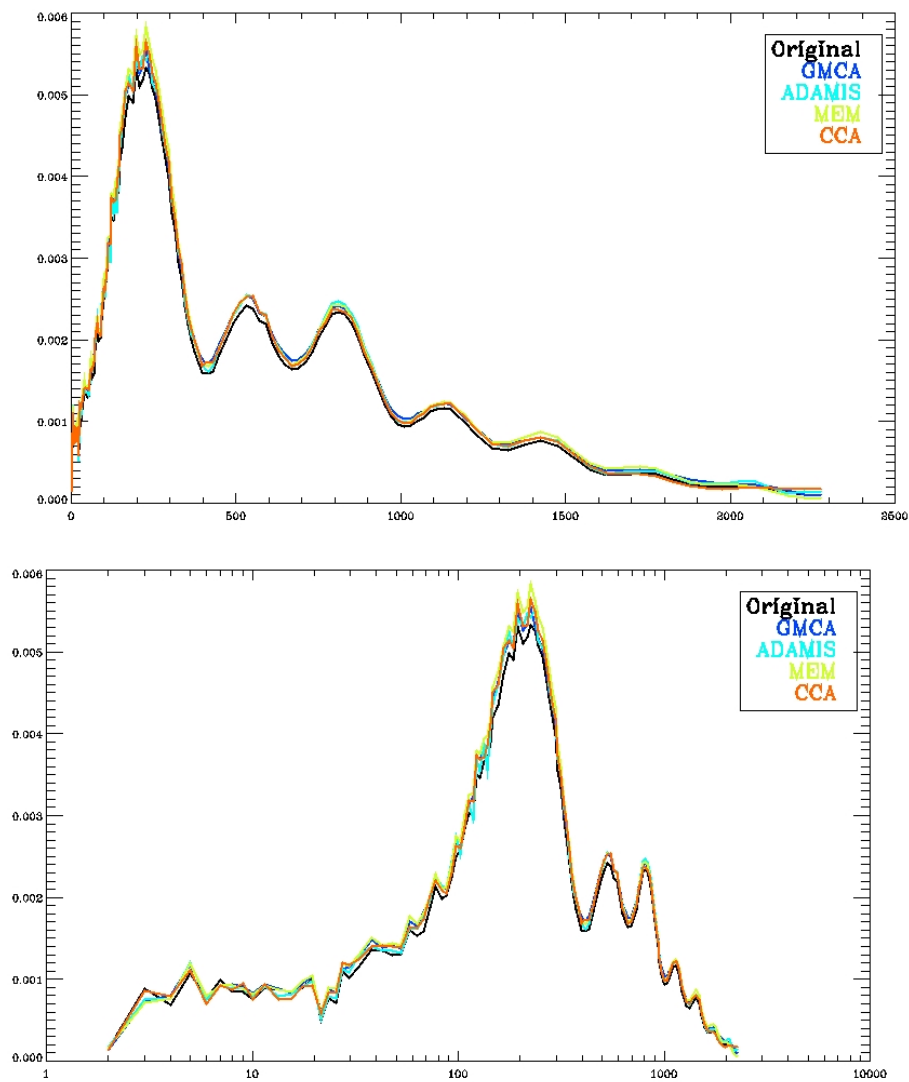


**FIG. 6.12:** Haut : écart-type du résidu convolué à 20 arcminutes par bande de latitudes de largeur 20°. Bas : écart-type du résidu convolué à 60 arcminutes par bande de latitudes de largeur 20°. Abscisses : latitude centrale. La bande centrée en 90° n'est pas représentée. Ordonnées : écart-type du résidu en mK.

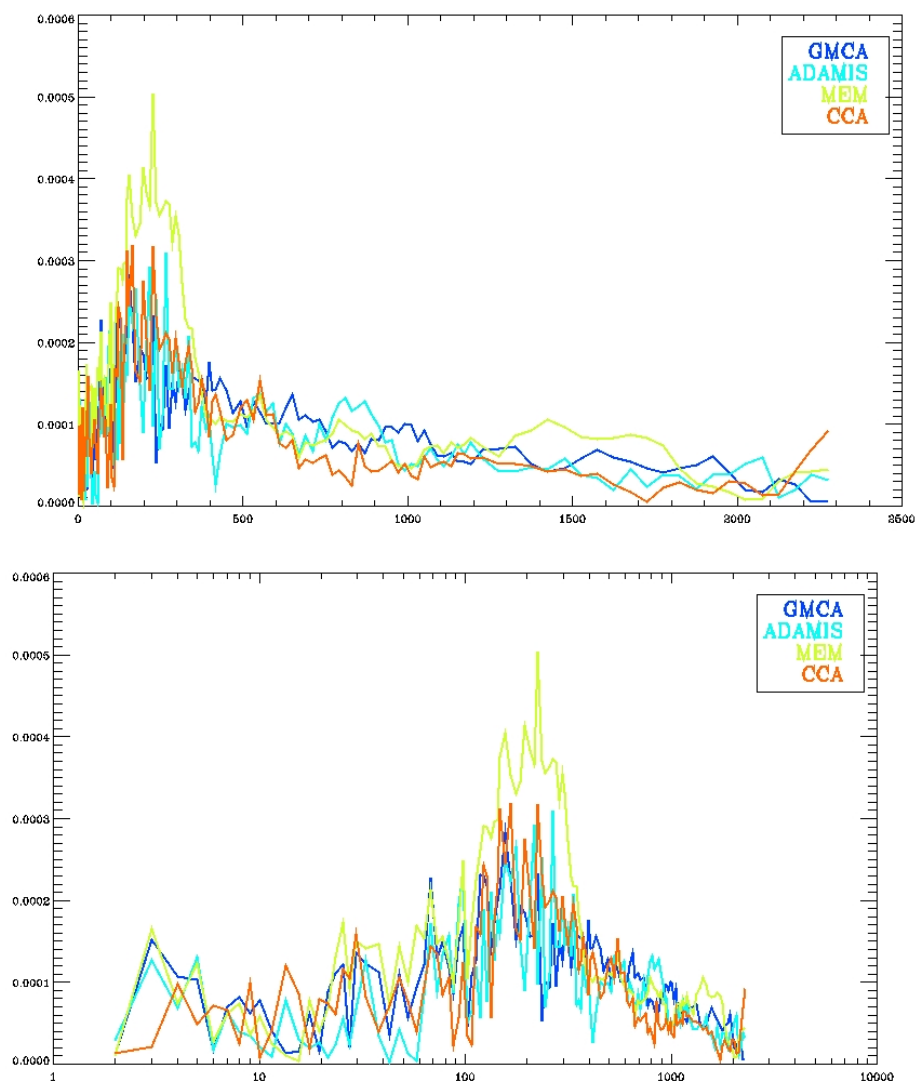


**FIG. 6.13: Kurtosis des cartes estimées par échelle d'ondelette - Abscisses : échelle d'ondelette ; l'équivalence en harmonique sphérique est donnée en haut de la figure. Ordonnées : kurtosis. Trait continu : kurtosis calculé à partir des données masquées puis après IDM. Trait discontinu : kurtosis calculé à partir des données masquées.**





**FIG. 6.14:** Spectre de puissance  $C(\ell)$  - Abscisses : échelle angulaire en échelle logarithmique. Ordonnées : température d'antenne au carré ( $\text{mK}^2$ ) normalisée en  $\ell(\ell + 1)/(2\pi)$



**FIG. 6.15: Erreur d'estimation du spectre de puissance  $e(\ell)$  - Abscisses : échelle angulaire en échelle logarithmique. Ordonnées : température d'antenne au carré ( $\text{mK}^2$ ) normalisée en  $\ell(\ell+1)/(2\pi)$**

## Chapitre 7

# Nouvelles perspectives pour la restauration de données multivaluées

Au cours de la dernière décennie, de nombreuses techniques promouvant la recherche d'une solution parcimonieuse dans une représentation fixée ont été proposées pour la résolution de nombreux problèmes inverses classiques (débruitage, déconvolution, super-résolution, interpolation de données manquantes). Compte tenu du rôle essentiel du choix de la représentation des données à restaurer, divers approches ont consisté à adapter la représentation aux données à restaurer. Cette dernière approche est équivalente à un "apprentissage" d'une "meilleure" représentation ; le plus souvent dans le sens d'une plus grande parcimonie des signaux à restaurer.

La fin du Chapitre 3 a mis en avant l'interprétation de l'Analyse en Composantes Mutuellement Parcimonieuses (ACMP) comme une méthode de recherche de meilleure base dans un cadre plus général. En effet, l'ACMP tend à promouvoir l'estimation d'une matrice de mélange (d'une manière générale, une représentation spectrale) dans laquelle les sources sont parcimonieuses. Dans le cas où le modèle de mélange n'est pas valide, l'obtention d'une représentation plus parcimonieuse devrait permettre d'apporter de meilleurs résultats de restauration.

Dans ce chapitre, nous proposons d'étendre le cadre de l'ACMP à celui de la résolution **adaptative** de problèmes de restauration de données multivaluées. La restauration adaptative de données multivaluées consiste alors à estimer conjointement les données à restaurer et la représentation dans laquelle ces données sont parcimonieuses.

---

## 7.1 Restauration de données multicanales

### Principe général

Le plus souvent, les données multicanales originales  $\mathbf{X}$  ne sont pas *directement* observables. Les données auxquelles l'observateur a accès sont une version *dégradée*  $\mathbf{Y}$  des données initiales  $\mathbf{X}$  :

$$\mathbf{Y} = \mathcal{F}(\mathbf{X}) + \mathbf{N}. \quad (7.1)$$

où  $\mathcal{F}$  est une application de l'espace de définition de  $\mathbf{X}$  à celui de  $\mathbf{Y}$ . Cette application  $\mathcal{F}$  peut modéliser une variété importante de fonctionnelles de *dégradation* intervenant dans des problèmes inverses classiques en traitement du signal : débruitage, déconvolution, mélange et séparation de source, super-résolution ou encore interpolation de données manquantes (IDM). Dans l'équation (7.1), la matrice  $\mathbf{N}$  modélise une perturbation additive indépendante du signal d'intérêt  $\mathbf{X}$  pouvant être une déviation par rapport au modèle de perturbation  $\mathcal{F}(\mathbf{X})$  ou tout simplement une composante de bruit instrumental additif.

### Quelques problèmes inverses classiques

Dans ce paragraphe, nous allons énumérer quelques problèmes inverses de restauration classique pouvant être décrits par l'équation (7.1) :

- **Débruitage** : le cas du problème de débruitage est un problème d'estimation classique. Dans ce cas, les données observées  $\mathbf{Y}$  sont une version bruitée des données originales  $\mathbf{X}$  :

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}$$

où  $\mathbf{N}$  modélise une perturbation additive de bruit. Dans ce cas l'application de perturbation  $\mathcal{F}$  est simplement l'identité :  $\mathcal{F}(\mathbf{X}) = \mathbf{X}$ .

- **Déconvolution** : le plus souvent, les instruments multispectraux (*i.e.* fournissant des observation multi-longueur d'onde) sont constitués de système de "détecteurs"<sup>1</sup> placés au foyer d'un unique instrument ayant une ouverture fixe. La largeur typique (largeur à mi-hauteur) de la *PSF* (*i.e.* point spread function ou réponse impulsionnelle) de l'instrument variant selon le facteur  $\lambda/D$  (où  $D$  est l'ouverture de l'instrument et  $\lambda$  la longueur d'onde d'observation), chacune des observations  $\mathbf{y}_i$  est une version convoluée de la donnée initiale  $\mathbf{x}_i$  avec un noyau de convolution  $\mathbf{h}_i$ . Ainsi chaque canal observé  $\mathbf{y}_i$  peut s'écrire sous la forme suivante :

$$\mathbf{y}_i = \mathbf{h}_i \star \mathbf{x}_i + \mathbf{n}_i \quad (7.2)$$

---

<sup>1</sup>Au sens large, il peut s'agir de photomètres, spectromètres, cellule CCD ...etc.

Dans ce cas, l'application  $\mathcal{F}$  est définie comme suit :

$$[\mathcal{F}(\mathbf{X})]_i = \mathbf{x}_i \star \mathbf{h}_i \quad (7.3)$$

- **Super-résolution** : Dans le cas de la super-résolution, l'application  $\mathcal{F}$  provoque une dégradation de la résolution des données initiales  $\mathbf{X}$ . Nous ne traiterons pas de ce problème dans cette thèse.
- **Interpolation de données manquantes** : Le problème d'interpolation de données manquantes (IDM) intervient lorsque les données observées  $\mathbf{Y}$  sont incomplètes : certains d'échantillons manquants. D'un point de vue formel, l'application  $\mathcal{F}$  est équivalente à un opérateur de masquage binaire. Soit  $\Omega$  l'ensemble des indices des échantillons de  $\mathbf{X}$  effectivement observés (non manquants). L'application  $\mathcal{F}$  est alors définie comme suit :

$$\mathcal{F}(\mathbf{X})[i, j] = \begin{cases} X[i, j] & \text{si } \{i, j\} \in \Omega \\ 0 & \text{sinon} \end{cases} \quad (7.4)$$

## Restauration adaptative par recherche d'une représentation parcimonieuse

### La parcimonie comme modélisation des structures d'un signal

Lors de ces dernières années, la plupart des problèmes inverses décrits précédemment ont été résolus en promouvant la parcimonie de la solution dans une représentation  $\Phi$  ; voir par exemple (FS05; Gul06a; Gul06b; FNW07; FN03b; NCB04; SNM03). Dans le cas de la restauration de données multicanales, le modèle est le suivant :

$$\mathbf{Y} = \mathcal{F}(\mathbf{X}) + \mathbf{N} . \quad (7.5)$$

Où le matrice  $\mathbf{N}$  modélise un bruit instrumental supposé borné en norme  $\ell_2$ . Estimer le signal multivaluée  $\mathbf{X}$  en minimisant sa norme  $\ell_1$  dans  $\Phi$  s'effectue en résolvant le problème d'optimisation suivant :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s.c. } \|\mathbf{Y} - \mathcal{F}(\mathbf{X})\|_F < \epsilon \text{ avec } \mathbf{X} = \alpha\Phi \quad (7.6)$$

Suivant l'expression de  $\mathcal{F}$ , cette équation peut être écrite sous la forme d'un lagrangien comme suit :

$$\min_{\alpha} \gamma \|\alpha\|_{\ell_1} + \frac{1}{2} \|\mathbf{Y} - \mathcal{F}(\mathbf{X})\|_F^2 \text{ avec } \mathbf{X} = \alpha\Phi \quad (7.7)$$

L'équation (7.7) définit un estimateur de  $\mathbf{X}$  obtenu par une approche uniquement monocanale. Une approche multicanale à ce problème requiert la modélisation de la double cohérence (intra/inter-canal) inhérente aux données multivaluées. La contrainte de parcimonie de  $\mathbf{X}$  dans  $\Phi$  permet de modéliser les seules structures temporelles ou spatiales du signal  $\mathbf{X}$ . La modélisation des structures spectrales des données nécessite le recours à une représentation multicanale adéquate  $\Pi \otimes \Phi$ . Le problème de l'équation (7.7) devient alors le suivant :

$$\min_{\alpha} \gamma \|\alpha\|_{\ell_1} + \frac{1}{2} \|\mathbf{Y} - \mathcal{F}(\mathbf{X})\|_F^2 \text{ avec } \mathbf{X} = \Pi\alpha\Phi \quad (7.8)$$


---

Afin de saisir la nuisance entre ces deux approches, nous allons considérer un cas d'école dans un cadre de débruitage.

**Illustration simple :** Prenons le cas simple où le signal  $\mathbf{X}$  n'est constitué que d'un unique élément de la base orthonormale multicanale  $\mathbf{\Pi} \otimes \mathbf{\Phi}$  ( $\mathbf{\Pi}$  et  $\mathbf{\Phi}$  sont supposées orthonormales). Considérons les données bruitées suivantes :

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} \quad (7.9)$$

Dans le cas où  $\mathbf{X} = \tau \pi_i \phi_j$  ( $\pi_i$  est le  $i$ -ème élément de  $\mathbf{\Pi}$  et  $\phi_j$  est le  $j$ -ème élément de  $\mathbf{\Phi}$ ;  $\tau$  est un scalaire). Supposons que  $\mathbf{N}$  est une matrice dont les entrées sont *i.i.d.* gaussiennes, centrées et de variance  $\sigma_N^2$ . Afin d'estimer  $\mathbf{X}$ , deux approches sont alors possibles :

- **Approche monocanale :** Résoudre le problème de l'équation (7.7) consiste, dans ce cas, à appliquer un seuillage doux à chaque échantillon de  $\mathbf{Y}\mathbf{\Phi}^T$  de façon indépendante. Classiquement, le biais et la variance de l'estimateur de  $\mathbf{X}$  ainsi mis en oeuvre seront liés au rapport signal-sur-bruit (RSB)  $\rho_{mono}$  au niveau de chacun des coefficients de  $\mathbf{Y}\mathbf{\Phi}^T$ . Dans le cas présent,  $\rho$  est borné de la sorte :

$$\rho_{mono} \leq \max_k |\pi_i[k]|^2 \left( \frac{\tau}{\sigma_N} \right)^2 \quad (7.10)$$

Dans le cas où  $\mathbf{\Pi} \neq \mathbf{I}^2$  on obtient l'inégalité **stricte** suivante :  $\rho_{mono} < \left( \frac{\tau}{\sigma_N} \right)^2$ .

- **Approche multicanale :** Dans le cas de l'approche multicanale, la résolution de l'équation (7.8) conduit à un seuillage doux de chacun des coefficients de  $\mathbf{\Pi}^T \mathbf{Y} \mathbf{\Phi}^T$  indépendamment. Dans ce cas, le RSB  $\rho_{multi}$  au niveau du coefficient  $\pi_i^T \mathbf{Y} \phi_j^T$  est alors exactement égal à :  $\rho_{multi} = \left( \frac{\tau}{\sigma_N} \right)^2$ .

En résumé, une approche multicanale/multivaluée (dans le cas où le signal à estimer présente effectivement des structures/cohérences multicanales) conduit à une meilleure extraction des structures du signal.

### Représentation adaptée

Dans un cadre général, un dictionnaire fixe  $\mathbf{\Pi} \otimes \mathbf{\Phi}$  n'est pas toujours adéquat pour la représentation parcimonieuse de large classes de signaux multivalués. Adapter la représentation dans laquelle les signaux sont modélisés revêt alors un intérêt certain. Les problématiques de recherche de bases adaptées ou de "meilleures" bases peuvent être classées en deux grandes catégories :

---

<sup>2</sup>Le cas égalité serait équivalent au cas de l'absence de structure/cohérence inter-canal.

- **Recherche d'une meilleure base dans une classe de bases paramétrées** : développé dans le cas monovalué, ce type d'approche considère la recherche d'une "meilleure" base pour un signal  $\mathbf{x}$  parmi un ensemble de bases paramétrées  $\{\Phi_\omega\}$  par  $\omega$ . Les ensembles de bases par arbre de décomposition (paquets d'ondelettes et de cosinus (Mal98) ou encore les bandelettes (LM05)) sont des exemples de tels ensembles de bases paramétrées. Le choix d'une meilleure base pour  $\mathbf{x}$  parmi l'ensemble  $\{\Phi\}_\omega$  consiste à sélectionner celle qui minimise un critère  $C(\mathbf{x}, \omega)$  fonction de  $\omega$ . Les critères classiques pour  $C(\mathbf{x}, \omega)$  sont par exemple la norme  $\ell_p$  (avec  $p < 2$ ) ou l'entropie (voir (CW92)) des coefficients de projection de  $\mathbf{x}$  dans  $\Phi_\omega$ . La recherche d'une meilleure base au sens de la norme  $\ell_1$  a en particulier été explorée par Peyré dans (G.P06). Nous n'entrerons pas plus dans les détails de ces techniques de recherche de meilleure base.
- **Recherche d'une base adaptée pour un ensemble de signaux** : contrairement au cas précédent pour lequel une meilleure base est recherchée à partir d'un seul signal  $\mathbf{x}$ , ici il est fait l'hypothèse qu'un ensemble de *réalisations* de signaux  $\{\mathbf{x}_i\}_{i=1,\dots,t}$  appartenant à une même classe de signaux notée  $\mathcal{B}$  est disponible. L'ensemble des signaux  $\{\mathbf{x}_i\}_{i=1,\dots,t}$  est alors considéré comme un ensemble d'apprentissage à partir duquel est "appris" (en anglais *learning*) une meilleure base. La recherche d'une base adaptée à partir d'un ensemble d'apprentissage a été mise en oeuvre dans divers domaines. Citons en particulier, les travaux de Olshausen (SO03) relatifs à la recherche d'une "meilleure" bases pour un ensemble d'images naturelles. Chez Olshausen, la "meilleure" base est celle dont les éléments sont indépendants (au sens de la maximisation du kurtosis - voir Chapitre 3). Notons encore les travaux d'Elad (AEB06; MES07) et de Peyré (Pey07) pour lesquels l'ensemble d'apprentissage est un jeu de patches obtenus par découpage spatial ou temporel d'un unique signal  $\mathbf{x}$ . Ici encore, une base adaptée  $\Phi$  (de patch en l'occurrence) est *apprise* au sens d'une maximisation de la parcimonie des patches de  $\mathbf{x}$  dans  $\Phi$ .

Au Chapitre 3, nous avons mis en avant une des propriétés de l'analyse en composantes mutuellement parcimonieuse : dans le cas où le modèle de sources est vérifié, elle permet d'obtenir la meilleure base (en terme d'une norme  $\|\cdot\|_1$  minimale) pour un ensemble de signaux correspondant aux colonnes de la matrice  $\mathbf{X}$ . L'ACMP peut ainsi être interprétée comme une technique de recherche de meilleure base pour une ensemble de signaux. Remarquons qu'Olshausen avait déjà mis en oeuvre un algorithme d'ACI pour la recherche d'une base adaptée à un ensemble de signaux.

Si le modèle de mélange n'est pas vérifié, rechercher une représentation spectrale  $\Pi$  telle que les coefficients  $\alpha = \Pi^T \mathbf{X} \Phi^T$  sont les plus parcimonieux possibles fait encore sens. Considérons ainsi le problème suivant :

$$\min_{\alpha, \Pi} \gamma \|\alpha\|_1 + \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\|_F^2 \text{ avec } \mathbf{X} = \Pi \alpha \Phi \quad (7.11)$$

Lorsque le modèle de mélange n'est pas vérifié, ce problème n'est pas convexe en général. Néanmoins, il peut être intéressant d'adapter la représentation  $\mathbf{\Pi}$  afin d'obtenir une matrice  $\alpha$  plus parcimonieuse facilitant ainsi la restauration de  $\mathbf{X}$ . L'utilisation de l'ACMP fait donc également sens lorsque le modèle de mélange n'est pas nécessairement vérifié. Le caractère adaptatif provient de la recherche **jointe** de  $\mathbf{\Pi}$  et de fait une partie du dictionnaire multicanal est "appris" à partir des données. Formellement, l'estimation du signal  $\mathbf{X}$  et de la représentation adaptée sera effectué selon le problème général suivant :

$$\min_{\alpha, \mathbf{\Pi}} \gamma \|\alpha\|_1 + \frac{1}{2} \|\mathbf{Y} - \mathcal{F}(\mathbf{X})\|_F^2 \text{ avec } \mathbf{X} = \mathbf{\Pi} \alpha \mathbf{\Phi} \quad (7.12)$$

Nous proposerons d'étendre l'algorithme GMCA pour la résolution du problème de l'équation (7.12). Durant cette thèse, deux approches ont été mises en oeuvre pour la restauration adaptative de données multivaluées :

- **Schéma "Offline"** : Dans le cas du débruitage, tel que l'a souligné Elad dans (EY08), la décomposition **la plus parcimonieuse** n'est pas nécessairement la meilleure ; par exemple, la solution peut présenter des artefacts visuels inacceptables (*e.g.* effets d'anneaux ou *ringing*). Une approche ad hoc différente de celle décrite par l'équation (7.12) consiste à appliquer l'algorithme GMCA sur  $\mathbf{Y} = \mathbf{X} + \mathbf{N}$  à l'aide d'une représentation spatiale  $\mathbf{\Phi}^{(1)}$  pour l'estimation de  $\mathbf{\Pi}$  (correspondant formellement à une matrice de mélange en SSA). Enfin  $\mathbf{X}$  est estimée dans le dictionnaire  $\mathbf{\Pi} \otimes \mathbf{\Phi}^{(2)}$  où  $\mathbf{\Phi}^{(2)}$  est mieux adaptée au débruitage d'image sans pour autant conduire à une décomposition très parcimonieuse. Cette méthode empirique permet d'établir un compromis entre la qualité de restauration des structures spatiales de  $\mathbf{X}$  dans  $\mathbf{\Phi}^{(2)}$  et une estimation adéquate d'une représentation spectrale  $\mathbf{\Pi}$  dont on sait qu'elle sera d'autant mieux estimée que  $\mathbf{\Phi}^{(1)}$  conduira à une représentation parcimonieuse.
- **Schéma "Online"** : Le schéma précédent ne résout pas le problème de l'équation (7.12). De plus, il est sans intérêt lorsque l'application de perturbation  $\mathcal{F}$  n'est pas linéaire. Cette approche conduit à l'estimation jointe de  $\mathbf{X}$  et de la matrice  $\mathbf{\Pi}$  telle qu'elle est décrite par l'équation (7.12).

#### Remarque - Qu'advient-il de $\mathbf{\Phi}$ ?

Dans le cas multicanal rien n'interdit la recherche d'une représentation multicanale adaptée. Une approche jointe consiste à traiter le problème suivant :

$$\min_{\alpha, \mathbf{\Pi}, \mathbf{\Phi}} \gamma \|\alpha\|_1 + \frac{1}{2} \|\mathbf{Y} - \mathcal{F}(\mathbf{X})\|_F^2 \text{ avec } \mathbf{X} = \mathbf{\Pi} \alpha \mathbf{\Phi} \quad (7.13)$$

La dimension spatiale des données multivaluées est le plus souvent bien supérieure à leur dimension spectrale :  $t \gg m$  ; la complexité du problème devient alors importante. Appliquer GMCA à ce problème peut ainsi être numériquement instable.

Dans le cas où la dimension spatiale  $t$  n'est pas trop élevée (quelques centaines), il peut-être envisageable de traiter le problème de l'équation 7.13. D'une manière plus pratique,



contraindre  $\Phi$  à appartenir à une classe de bases définies via un arbre de décomposition (par exemple : paquets d'ondelettes ou de cosinus, bandelettes) peut également être une extension numériquement viable des développements que nous venons décrire.

## 7.2 Application à la restauration adaptative de données multicanales

### Schéma "offline" - application au débruitage d'images couleur

Nous avons mis en exergue l'importance d'une modélisation des données dans une représentation *adaptée*. Dans le cas où le signal à restaurer  $\mathbf{X}$  est multivaluée il semble important de chercher à modéliser les structures de chaque canal ainsi que celles apparaissant entre eux. Le cas de l'imagerie couleur est particulièrement difficile puisque la dimensionnalité spectrale des données est faible :  $m = 3$ . La faible dimension de  $\Pi$  ne permet pas de capturer toute la richesse des structures couleurs que peuvent présenter ces images ; nous verrons néanmoins qu'une adaptation de  $\Pi$  permet un gain de restauration.

Dans ce paragraphe, le signal à restaurer  $\mathbf{X}$  est composé de 3 canaux (*i.e.* les couches couleurs - Rouge, Vert et Bleu). Chaque couche est formée d'une image de taille  $\sqrt{t} \times \sqrt{t}$ . Dans le cas du débruitage, la fonction de *perturbation*  $\mathcal{F}$  est l'identité de sorte que :

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} \quad (7.14)$$

Ce paragraphe présente les résultats préliminaires de restauration d'images multivaluées présentés dans (BSFM07). Dans cette étude, il n'était pas question de résoudre le problème de l'équation (7.12). Il s'agissait d'utiliser l'algorithme GMCA comme un moyen de recherche d'une "meilleure" base  $\Pi$  (*i.e.* conduisant à une représentation plus parcimonieuse des données) puis de résoudre le problème de l'équation (7.8) dans la représentation  $\Pi \otimes \Phi$ . Notons qu'une telle approche a été proposée par (SQ01) où la base  $\Pi$  était la base de Karhunen-Loeve des données (obtenue par analyse en composantes principales des données).

Par la suite, le débruitage *adaptatif* proposé se déroule en deux étapes :

- **Apprentissage de  $\Pi$**  : cette étape conduit à utiliser GMCA comme une technique de recherche de la matrice  $\Pi$  (pas nécessairement orthogonale) conduisant à une représentation plus parcimonieuses des données dans  $\Pi \otimes \Phi^{(1)}$ .
- **Restauration de  $\mathbf{X}$**  : restauration des données dans le dictionnaire  $\Pi \otimes \Phi^{(2)}$ .

L'image en haut à gauche de la Figure 7.1 montre la donnée bruitée  $\mathbf{Y}$ . Le bruit  $\mathbf{N}$  est une réalisation gaussienne, centrée et blanche de sorte que RSB=15dB. L'image en haut à droite est obtenue par débruitage de  $\mathbf{Y}$  suivant le problème de l'équation (7.7) où  $\Phi$  est une trame d'ondelettes non-décimées (SFM07). Ce résultat est celui de l'approche dite monocanale où la restauration est faite sur chaque couche couleur indépendamment. En d'autres termes,

il n'y a aucune prise en compte de possibles structures inter-canales.

L'image en bas de la Figure 7.1 représente le résultat du débruitage adaptatif à 2 étapes décrit précédemment avec : une trame de curvelets pour  $\Phi^{(1)}$  et une trame d'ondelettes non-décimées pour  $\Phi^{(2)}$ . Visuellement, la prise en compte de structures globales couleurs (rappelons que dans ce cas la dimensionnalité spectrale est faible :  $m = 3$ ) permet d'obtenir un résultat de restauration de meilleure qualité. On note en particulier que les contours sont mieux restaurés. La Figure 7.2 propose des versions agrandies d'une partie des images de la Figure 7.1. Dans le cas du traitement des images couleurs, il peut être intéressant de



**Fig. 7.1:** Haut à gauche : image originale de taille  $256 \times 256$  avec un bruit additif blanc gaussien. Le RSB est égal à 15 dB. Haut à droite : débruitage par ondelettes de chaque couche couleur indépendamment. Bas : débruitage dans le dictionnaire  $\Pi \otimes \Phi^{(2)}$  ;  $\Pi$  a été estimé par GMCA.

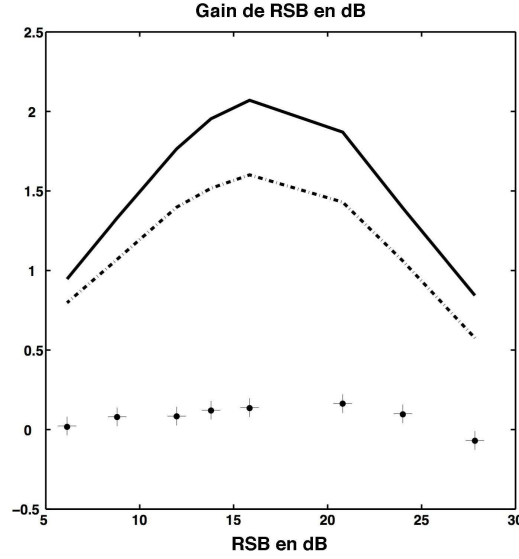
travailler non pas dans l'espace initial RVB mais dans celui formé par une luminance et deux chrominances (YUV et YCC). Nous proposons d'appliquer le même mécanisme de

restauration dans ces représentations couleurs *non-adaptatives*. Nous mettons également en oeuvre ce schéma de restauration dans un espace couleur adaptatif obtenu par l'application de l'algorithme d'ACI Jade sur les données  $\mathbf{Y}$ .

La Figure 7.3 montre le gain de RSB (en dB) des différentes méthodes (YUV, YCC, JADE et GMCA) par rapport à l'approche monocanale dans l'espace RVB. De façon intéressante, l'algorithme Jade permet d'obtenir un gain pouvant atteindre 1 dB. Au final, la recherche d'une "meilleure base" promouvant la parcimonie permet d'atteindre un gain de RSB de 2dB. Cette série de tests conforte l'impression visuelle de la Figure 7.1.



**Fig. 7.2:** Versions agrandies. **Haut à gauche :** image originale avec un bruit additif blanc gaussien. Le RSB est égal à 15 dB. **Haut à droite :** débruitage par ondelettes de chaque couche couleur indépendamment. **Bas :** débruitage dans le dictionnaire  $\Pi \otimes \Phi^{(2)}$  ;  $\Pi$  a été estimé par GMCA.



**Fig. 7.3:** Débruitage adaptatif d'images couleurs. **Abscisses :** RSB en dB. **Ordonnées :** gain de RSB en dB par rapport à un débruitage sur les couches RVB. Trait continu : débruitage adaptatif via GMCA, Trait discontinu : JADE, Symbole '•' YUV, Symbole '+' : YCC.

### Application à l'interpolation de données manquantes dans des images couleurs

Nous proposons ici une approche adaptative fondée sur l'ACMP et GMCA dans le cadre du problème de l'interpolation de données manquantes (IDM) dans des images couleurs. De tels problèmes de restauration interviennent lorsque certains échantillons des données  $\mathbf{Y}$  sont manquants. Notons  $\Omega$  l'ensemble des échantillons de  $\mathbf{Y}$  effectivement observés (*i.e.* non manquants). L'application de la dégradation  $\mathcal{F}$  est équivalente à un opérateur de masquage de sorte que :

$$\mathbf{Y} = \mathcal{M}_{\Omega} \odot \mathbf{X} \quad (7.15)$$

où  $\mathcal{M}_{\Omega}$  est un masque binaire multicanal qui s'applique à  $\mathbf{X}$  échantillon par échantillon :

$$\forall j \in \{1, \dots, t\}, i \in \{1, \dots, m\}; \quad \mathcal{M}_{\Omega}[i, j] = \begin{cases} 1 & \text{si } \{i, j\} \in \Omega \\ 0 & \text{sinon} \end{cases} \quad (7.16)$$

Le problème de restauration est formellement proche de celui de l'estimation d'un signal  $\mathbf{X}$  à partir de mesures incomplètes. La résolution de problèmes d'IDM sous hypothèse de parcimonie de  $\mathbf{X}$  dans une représentation appropriée peut être considérée comme un cas particulier du *Compressed Sensing* (Voir en Annexe).

Une extension de MCA a été proposée pour résoudre les problèmes d'IDM dans le cas monocal (ESDQ05; FS05; FSM06). L'algorithme mMCA (voir Chapitre 2) est adapté selon (FS05). A cet effet, nous supposons que  $\Phi$  est formé d'une union de  $D$  bases orthogonales

$\{\Phi_j\}_{j=1,\dots,D}$ . La résolution du problème de l'IDM s'effectue alors comme suit :

$$\{\alpha_{j=1,\dots,D}\} = \underset{\{\alpha_{j=1,\dots,D}, \mathbf{\Pi}\}}{\text{Argmin}} \gamma \sum_{j=1}^D \|\alpha_j\|_1 + \frac{1}{2} \left\| \mathbf{Y} - \mathcal{M}_\Omega \odot \left( \mathbf{\Pi} \sum_{j=1}^D \alpha_j \Phi_j \right) \right\|_F^2. \quad (7.17)$$

Dans le cadre de l'optimisation par BCR (BST98), ce problème peut être résolu itérativement (voir (BMFS07b)). A chaque itération ( $h$ ), les coefficients  $\{\alpha_j\}$  sont estimés alternativement à partir des données partiellement reconstruites  $\mathbf{Y}^{(h)}$  :

$$\mathbf{Y}^{(h)} = \mathbf{Y} + \mathcal{M}_\Omega^c \odot \tilde{\mathbf{X}}^{(h-1)} \quad (7.18)$$

où  $\mathcal{M}_\Omega^c$  est le complément binaire du masque  $\mathcal{M}_\Omega$  et  $\tilde{\mathbf{X}}^{(h-1)}$  est l'estimée à l'itération ( $h-1$ ) du signal  $\mathbf{X}$  à restaurer. A partir de ce schéma non-adaptatif, la matrice  $\mathbf{\Pi}$  est estimée au sens des moindres carrés comme suit :

$$\mathbf{\Pi}^{(h)} = \arg \min_{\mathbf{\Pi}} \left\| \mathbf{Y}^{(h)} - \mathbf{\Pi} \mathbf{X}^{(h)} \right\|_F^2. \quad (7.19)$$

Si l'on définit les composantes morphologiques  $\forall i = 1, \dots, D$ ;  $\varpi_j = \alpha_j \Phi_j$  l'algorithme proposé est alors le suivant :

1. Fixer le nombre d'itérations  $P_{\max}$  et le seuil  $\gamma^{(0)}$ .
2. Tant que  $\gamma^{(h)}$  que le seuil final  $\gamma_{\min}$ ,
  - a. Calculer  $\mathbf{Y}^{(h)} = \mathbf{Y} + \mathcal{M}^c \odot \tilde{\mathbf{X}}^{(h-1)}$ .
  - b. Initialiser chaque composante morphologique  $\{\tilde{\varpi}_j\}^{(h-1)}$ .  
 Pour  $j = 1, \dots, D$ 
    - Calculer le résidu  $\mathbf{R}_j^{(h)}$  en supposant  $\varpi_{\{p\} \neq \{j\}}$ ,  $\tilde{\varpi}_{p \neq j}^{(h-1)}$  fixés :  

$$\mathbf{R}_j^{(h)} = \mathbf{Y}^{(h)} - \sum_{p \neq j} \tilde{\varpi}_p^{(h-1)}.$$
    - Estimer les coefficients des composantes  $\tilde{\varpi}_j^{(h)}$  par seuillage de seuil  $\gamma^{(h)}$  :  

$$\tilde{\alpha}_j^{(h)} = \mathcal{S}_{\gamma^{(h)}} \left( \mathbf{\Pi}^{(h)T} \mathbf{R}_j^{(h)} \Phi_j^T \right).$$
    - Estimer  $\varpi_j$  par reconstruction à partir de  $\tilde{\alpha}_j^{(h)}$  :  

$$\tilde{\varpi}_j^{(h)} = \mathbf{\Pi}^{(h)} \tilde{\alpha}_j^{(h)} \Phi_j.$$
  - c. Mise à jour de  $\tilde{\mathbf{X}}^{(h)} = \sum_{j=1}^D \tilde{\varpi}_j^{(h)}$ .
  - d. Mise à jour de  $\mathbf{\Pi}$  :  

$$\mathbf{\Pi}^{(h+1)} = \mathbf{Y}^{(h)} \left[ \sum_{j=1}^D \varpi_j^{(h)} \Phi_j \right]^\dagger.$$
3. Décroissance du seuil  $\gamma^{(h)}$ .

L'image en haut à gauche de la Figure 7.4 est l'image couleur originale non bruitée *Barbara*. Celle en haut à droite représente la donnée  $\mathbf{Y}$  dont 90% des échantillons ont été masqués aléatoirement. Le résultat de l'approche monocanale (consistant à appliquer l'algorithme de (FS05) sur chaque canal de  $\mathbf{Y}$  séparément) est montré en bas à gauche. La

dernière image, en bas à droite, représente le résultat de l'algorithme adaptatif proposé. Visuellement, les deux approches apportent des résultats convaincants. La Figure 7.5 propose la version agrandie des images précédentes. Elle met en relief l'apport de la méthode adaptative : certaines aberrations chromatiques provoquant un effet de "moiré" disparaissent dans le cas où une "meilleure" représentation couleur est estimée. Cette impression visuelle est confirmée par un RSB en hausse de 1dB.

**Remarque :**

Au cours de ces expériences simples, le choix du traitement d'images couleurs n'est pas anodin. Il s'agit en quelque sorte d'une situation délicate pour la restauration de données multicanales. En effet, la méthode d'analyse de signaux parcimonieux mMCA que nous avons développée sous diverses formes dans ce rapport de thèse n'a recours qu'à des représentations séparables de la forme  $\mathbf{\Pi} \otimes \mathbf{\Phi}$ . Dans le cas des images couleurs, la dimension spectrale  $m = 3$  laisse très peu de degrés de liberté pour *capturer* la richesse des structures couleurs que présentent les images couleurs *naturelles*. En un certain sens, la méthode proposée n'est pas "optimale" pour la restauration des images couleurs. L'approche par patch d'Elad (MES07) permet une meilleure adaptation à la richesse des structures couleurs d'une image. Néanmoins, l'utilisation de représentations par patch apporte également un certain nombre de difficultés.

Dans le cas où la dimension spectrale des données augmente (*a fortiori* dans le cas des données hyperspectrales), la dimension de  $\mathbf{\Pi}$  augmentant, l'apport de cette méthode devrait être bien plus conséquent. Néanmoins, le traitement de données hyperspectrales par exemple posera une autre difficulté majeure : celui de la dimension des données  $\mathbf{X}$ . En effet, dans le cas où  $m \simeq \sqrt{t}$  (c'est le cas des données hyperspectrales, au moins à un ordre de grandeur près), la dimension des données  $\mathbf{X}$  (au sens du rang de la matrice  $\mathbf{X}$ ) peut être plus faible que la dimension spectrale des données  $\mathbf{Y}$  :  $n < m$ . Dans ce cas, l'estimation jointe de la dimension des données (*i.e.* le nombre de sources) est importante lorsque les données en jeu sont des données à grande dimension spectrale  $m$ . En résumé, une méthode de restauration adéquate consisterait à estimer conjointement  $\mathbf{X}$ ,  $\mathbf{\Pi}$  et  $n$ .

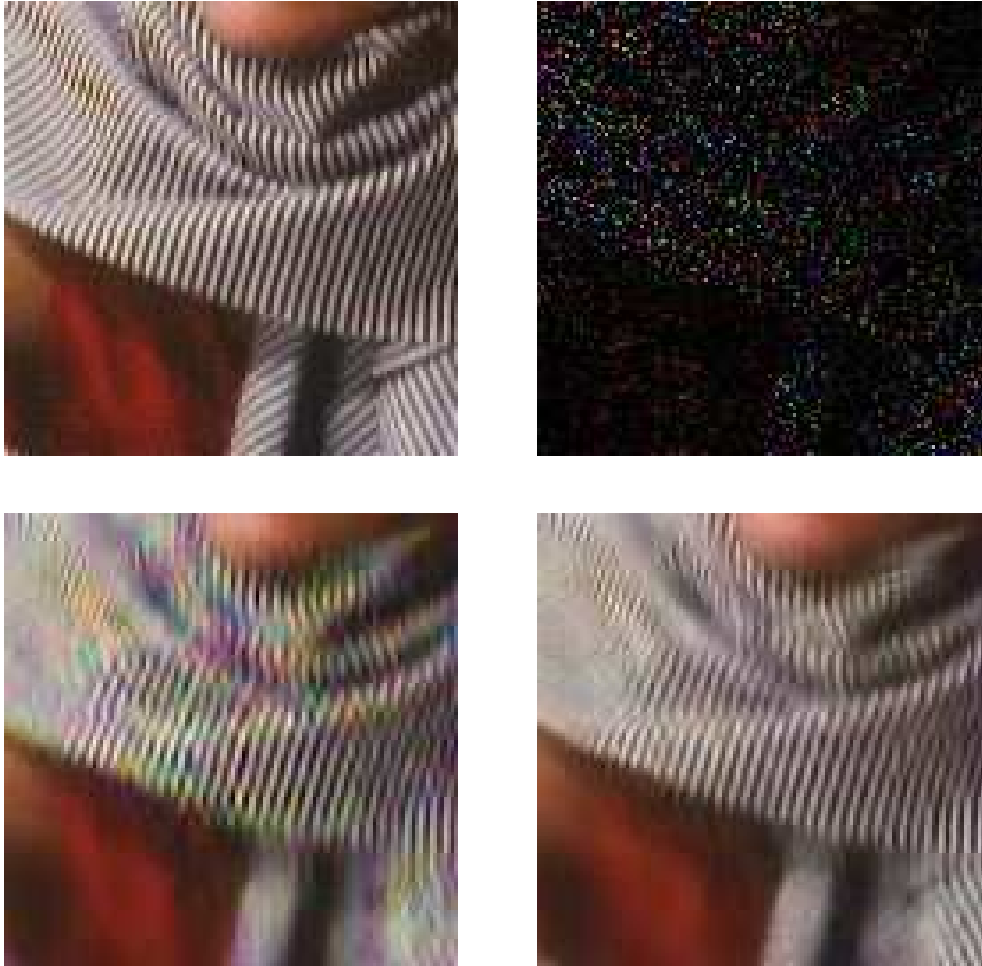
---





**Fig. 7.4:** Restauration d'images couleurs. (a) image originale *Barbara*. (b) image masquée - 90% des pixels couleurs sont manquants. (c) image restaurée en utilisant MCA sur chaque canal séparément. (d) image restaurée à partir de l'approche multicanale adaptative fondée sur mMCA.

---



**Fig. 7.5:** Versions agrandies. (a) image originale *Barbara*. (b) image masquée - 90% des pixels couleurs sont manquants. (c) image restaurée en utilisant MCA sur chaque canal séparément. (d) image restaurée à partir de l'approche multicanale adaptative fondée sur mMCA.

---



## Chapitre 8

# Epilogue

Ces trois années de thèse avaient pour objet l'étude de la séparation de sources en aveugle sous l'angle de la recherche de sources parcimonieuses. Nous avons préalablement proposé dans (BMSE06) une extension directe du cadre de l'analyse en composantes morphologiques (MCA). Celle-ci a mis au jour la nécessité d'une extension des problèmes de décomposition parcimonieuse aux données multicanales dans (BSF<sup>+</sup>07; BMFS07b). A partir de ces travaux initiaux, nous espérons que cette thèse aura permis d'apporter un regard différent des problèmes de séparation de sources en aveugle. Ainsi, parallèlement aux approches classiques telles que l'Analyse en Composantes Indépendantes (ACI), l'Analyse en Composantes Mutuellement Parcimonieuse (ACMP) offre un cadre hybride entre SSA classique et recherche de meilleure base (au sens plus parcimonieuse).

Au-delà de ces considérations "abstraites", l'ACMP trouve également sa place dans le vaste paradigme de l'inférence bayésienne. Elle offre ainsi un cadre flexible de mise en oeuvre dans différents contextes applicatifs. Ce point a d'ailleurs été souligné par l'extension de GMCA pour l'analyse de données hyperspectrales au Chapitre 5 ainsi que pour l'analyse des données Planck au Chapitre 6. Enfin, fort de l'interprétation de l'ACMP comme méthode de recherche de meilleure base, le Chapitre 7 a proposé une extension de l'algorithme GMCA à la résolution de différents problèmes inverses classiques impliquant des données multicanales tels que l'estimation de données bruitées ou l'interpolation de données manquantes (IDM).

Au-delà des multiples développements ayant fait l'objet de cette thèse, il est possible de dresser une liste des différents points qu'il sera intéressant d'explorer par la suite :

- **Séparation de source** : une extension de GMCA aux problèmes de séparation de sources en aveugle sous-déterminés (le nombre de sources est supérieur au nombre d'observations) est à l'étude. Elle est d'autant plus justifiée que les récentes approches proposés pour la résolution de problèmes de SSA dans le cas sous-déterminé ont mis en avant l'utilisation de représentation parcimonieuse. En théorie, rien n'empêche l'application directe de GMCA à ce type de données ; ceci doit être validé expérimentalement.
-

Dans le cadre de la mission Planck, le développement d'un outil d'analyse de données polarisées est nécessaire ; une extension de l'algorithme GMCA au traitement de telles données est à l'étude.

- **Restauration de données :** nous avons montré que GMCA s'étend assez simplement à la résolution adaptative de certains problèmes inverses classiques. L'exemple de l'analyse des données Planck a particulièrement mis en lumière la nécessité de la résolution jointe des problèmes de séparation en aveugle de données convoluées.
  - **Analyse de données multivaluées :** un certain nombre de problèmes de traitement de données, dont la SSA est un exemple particulier, nécessitent la décomposition d'une matrice en une combinaison linéaire de matrice de rang 1. On peut citer pour exemple l'analyse exploratoire de données multivaluées, la réduction de dimension, certains problèmes de classification ... etc. Dans le cas général, l'analyse en composantes principales (ACP) est l'une des méthodes privilégiées proposées pour la résolution de ce genre de problèmes. Dans certains contextes applicatifs, il peut être intéressant de contraindre les composantes recherchées. A cette fin, GMCA pourrait être une sorte d'ACP sous contraintes. Nous envisageons d'appliquer GMCA de la sorte pour le traitement de multiples spectres de masse dans le cadre d'une collaboration avec la School of Medicine de l'université de Stanford.
-

## Annexe A

# Conditions de convergence de l'Analyse en Composantes Morphologiques (MCA)

**Proposition (Propriété de sélection exacte)** Supposons que  $\mathbf{x}$  soit  $K$ -sparse tel que :

$$\mathbf{x} = \sum_{j=1}^D \sum_{i \in \Lambda_j} \alpha_j[i] \phi_j[i]$$

où  $K = \sum_j \text{Card}(\Lambda_j)$ . A l'itération  $h$ , supposons que le résidu  $\mathbf{r}^{(h)}$  est  $K$ -sparse de sorte que :

$$\mathbf{r}^{(h)} = \sum_{j=1}^D \sum_{i \in \Lambda_j} \beta_j[i] \phi_j[i]$$

On note :

$$(\tau^\star, i^\star) = \underset{\tau, i \in \Lambda_\tau}{\text{Argmax}} |\beta_\tau[i]|$$

$$(\tau^\dagger, i^\dagger) = \underset{\tau \neq \tau^\star, i \in \Lambda_\tau}{\text{Argmax}} |\beta_\tau[i]|$$

$$\beta^\star = |\beta_{\tau^\star}[i^\star]|$$

$$\beta^\dagger = |\beta_{\tau^\dagger}[i^\dagger]|$$

On suppose qu'il existe, à l'itération  $(h)$ , un scalaire  $0 \leq \rho < 1$  tel que  $\beta^\dagger = \rho \beta^\star$ . Si la condition suivante est vérifiée :

$$K < \mu_\Phi^{-1} \frac{1-\rho}{1+\rho}$$

Alors MCA/MOM sélectionne, à l'itération  $h$ , des coefficients appartenant au support de  $\mathbf{x}$ .

---

**Preuve (Propriété de sélection exacte)** Définissons les notations suivantes :

$$\begin{aligned}
 (\tau_1, i_1) &= \underset{\tau, i \in \Lambda_\tau}{\operatorname{Argmax}} \left| \langle \mathbf{r}^{(h)}, \boldsymbol{\phi}_\tau[i] \rangle \right| \\
 (\tau_2, i_2) &= \underset{\tau \neq \tau_1, i \in \Lambda_\tau}{\operatorname{Argmax}} \left| \langle \mathbf{r}^{(h)}, \boldsymbol{\phi}_\tau[i] \rangle \right| \\
 m_1 &= \max_{\tau, i \in \Lambda_\tau} \left| \langle \mathbf{r}^{(h)}, \boldsymbol{\phi}_\tau[i] \rangle \right| \\
 m_2 &= \max_{\tau \neq \tau_1, i \in \Lambda_\tau} \left| \langle \mathbf{r}^{(h)}, \boldsymbol{\phi}_\tau[i] \rangle \right| \\
 (\tau^\star, i^\star) &= \underset{\tau, i \in \Lambda_\tau}{\operatorname{Argmax}} |\beta_\tau[i]| \\
 (\tau^\dagger, i^\dagger) &= \underset{\tau \neq \tau^\star, i \in \Lambda_\tau}{\operatorname{Argmax}} |\beta_\tau[i]| \\
 \beta^\star &= |\beta_{\tau^\star}[i^\star]| \\
 \beta^\dagger &= |\beta_{\tau^\dagger}[i^\dagger]| \\
 \omega_\tau &= \max_i \left| \langle \mathbf{r}^{(h)}, \boldsymbol{\phi}_\tau[i] \rangle \right|
 \end{aligned}$$

Par définition, la règle de seuillage MOM sélectionne de nouveaux coefficients dans  $\Phi_{\tau^\star}$  comme suit tels que :

$$\forall i, \tau; \quad \left| \langle \mathbf{r}^{(h)}, \boldsymbol{\phi}_{\tau^\star}[i] \rangle \right| > \gamma^{(h)}$$

où  $\gamma^{(h)} = \frac{1}{2} (m_1 + m_2)$ .

Afin de montrer qu'une telle stratégie de sélection de coefficients vérifie la propriété de sélection exacte, il suffit de vérifier les inégalités suivantes :

$$\left| \langle \mathbf{r}^{(h)}, \boldsymbol{\phi}_{\tau^\star}[i] \rangle \right| < \gamma^{(h)}, \quad \forall i \in \Lambda_{\tau^\star}^c \quad (\text{A.1})$$

où  $\Lambda_{\tau^\star}^c$  est le complémentaire du support  $\Lambda_{\tau^\star}$  restreint au sous-dictionnaire  $\Phi_{\tau^\star}$ .

**a) Etablissons une condition pour que  $\tau^\star = \tau_1$  :**

D'une part, l'inégalité suivante est vérifiée :

$$\omega_{\tau^\star} \geq \beta^\star - \mu_\Phi K \beta^\dagger \quad (\text{A.2})$$

$$\geq \beta^\star (1 - \rho \mu_\Phi K) \quad (\text{A.3})$$

Nous avons utilisé ici le fait que  $\left| \langle \boldsymbol{\phi}_{\tau^\star}[i], \boldsymbol{\phi}_p[j] \rangle \right| \leq \mu_\Phi$  et  $\beta^\dagger = \rho \beta^\star$ .

D'autre part, la propriété suivante est vérifiée :

$$\omega_{\tau \neq \tau^\star} \leq \beta^\dagger + \mu_\Phi K \beta^\star \quad (\text{A.4})$$

$$\leq \beta^\star (\rho + \mu_\Phi K) \quad (\text{A.5})$$

En résumé, si  $\beta^*(\rho + \mu_\Phi K) < \beta^*(1 - \rho\mu_\Phi K)$  alors  $\forall \tau; \omega_{\tau^*} > \omega_{\tau \neq \tau^*}$ . En d'autres termes, si  $K < \mu_\Phi^{-1} \frac{1-\rho}{1+\rho}$  alors  $\forall \tau; \omega_{\tau^*} > \omega_{\tau \neq \tau^*}$ . On en déduit que  $m_1 = \omega_{\tau^*}$ .

**b) Sélection exacte :**

Rappelons que  $\gamma^{(h)} = \frac{1}{2}(m_1 + m_2)$ . Afin de montrer que MCA/MOM sélection des atomes de  $\Phi_{\tau^*}$  appartenant à  $\Lambda_{\tau^*}$ , il doit être montré que :

$$\forall i \in \Lambda_{\tau^*}^c; \gamma^{(h)} > \left| \langle \mathbf{r}^{(h)}, \phi_{\tau^*}[i] \rangle \right|$$

D'une part, une borne inférieure au seuil  $\gamma^{(h)}$  peut être obtenue en bornant inférieurement  $m_1$  et  $m_2$ . Sans hypothèses supplémentaires, il est difficile de dire plus que  $m_2 > 0$ . De sorte que :

$$\begin{aligned} \gamma^{(h)} &> \frac{1}{2}m_1 \\ &> \frac{1}{2}\beta^*(1 - \rho K \mu_\Phi) \end{aligned} \quad (\text{A.6})$$

D'autre part, on prouve la propriété suivante :

$$\left| \langle \mathbf{r}^{(h)}, \phi_{\tau^*}[i] \rangle \right| \leq K\beta^* \mu_\Phi \quad (\text{A.7})$$

$$\leq \rho K \mu_\Phi \beta^* \quad (\text{A.8})$$

où tous les coefficients de la composante  $\tau^*$  s'annulent puisque  $\Phi_{\tau^*}$  est orthonormale.

Des équations (A.6) et (A.7), nous déduisons que l'inégalité suivante doit être vérifiée pour que la condition (A.1) soit valide :

$$\frac{1}{2}(1 - \rho K \mu_\Phi) > \rho K \mu_\Phi$$

Soit :  $K < \frac{\mu_\Phi^{-1}}{3\rho}$ . Puisque pour  $0 < \rho < 1$ ,  $\frac{1}{3\rho} > \frac{1-\rho}{1+\rho}$  alors la condition  $K < \mu_\Phi^{-1} \frac{1-\rho}{1+\rho}$  conduit à la validité de (A.1) et de fait à celle de la propriété de sélection exacte.

Remarquons que  $\rho$  définit un facteur de contraste entre la composante morphologique  $\varphi_{\tau^*}$  et les autres. Si  $\rho$  tend vers 0,  $\varphi_{\tau^*}$  est largement prédominante (voire uniquement présente dans le résidu  $\mathbf{r}^{(h)}$ ), la sélection d'atomes actifs liés à  $\varphi_{\tau^*}$  est donc facilitée. Lorsque  $\rho$  tend vers 1, au moins deux composantes morphologiques sont semblablement actives ; un plus grand degré de parcimonie est de fait nécessaire pour sélectionner des atomes de  $\varphi_{\tau^*}$ .

**Proposition (Convergence)** Supposons que  $\mathbf{x}$  soit  $K$ -sparse tel que :

$$\mathbf{x} = \sum_{j=1}^D \sum_{i \in \Lambda_j} \alpha_j[i] \phi_j[i]$$


---

où  $K = \sum_j \text{Card}(\Lambda_j)$ .

Supposons que les hypothèses de la Propriété de sélection exacte sont vérifiées. Si  $K < \mu_{\Phi}^{-1} \frac{1-\rho}{1+\rho}$  alors MCA/MOM converge exponentiellement vers  $\mathbf{x}$  et sa décomposition la plus parcimonieuse au sens  $\ell_0$  dans  $\Phi$ . Plus précisément, le résidu converge vers 0 avec une vitesse exponentielle.

**Preuve (Convergence)** La démonstration est en tout point semblable à celle de la convergence de MP ou OMP (Tro04). En effet, l'algorithme MCA/MOM peut être interprété comme une version stagewise de l'algorithme MP. MCA/MOM est initialisé de sorte que  $\mathbf{r}^{(0)} = \mathbf{x}$ . Par hypothèse, à chaque itération  $h \geq 0$  et pour la composante  $\tau^*$ , MCA/MOM sélectionne des atomes effectivement activés par  $\Lambda_{\tau^*}$ . La nouvelle valeur du résidu est alors calculée comme suit :

$$\mathbf{r}^{(h)} = \sum_{i \in I_{\tau^*}} \langle \mathbf{r}^{(h)}, \phi_{\tau^*}[i] \rangle \phi_{\tau^*}[i] + \mathbf{r}^{(h+1)}$$

où  $I_{\tau^*} \subseteq \Lambda_{\tau^*}$ , et  $\Phi_{\Lambda_{\tau^*}}$  est la restriction de  $\Phi$  aux atomes indexés par  $\Lambda_{\tau^*}$ . De l'orthogonalité des atomes de  $\Phi_{\Lambda_{\tau^*}}$ , on déduit :

$$\begin{aligned} \|\mathbf{r}^{(h+1)}\|^2 &= \|\mathbf{r}^{(h)}\|^2 - \left\| \sum_{i \in I_{\tau^*}} \langle \mathbf{r}^{(h)}, \phi_{\tau^*}[i] \rangle \phi_{\tau^*}[i] \right\|^2 \\ &\leq \|\mathbf{r}^{(h)}\|^2 - \|\mathbf{r}^{(h)} \Phi_{\Lambda_{\tau^*}}\|_{\ell_\infty}^2 \end{aligned} \quad (\text{A.9})$$

La suite de la preuve est semblable à celle de (GV06, Théorème 2), du lemme (Tro04, Lemme 2.3) fondé sur le théorème du disque de Gershgorin, et de la proposition (Tro04, Proposition 2.1). On déduit :

$$\begin{aligned} \|\mathbf{r}^{(h+1)}\|^2 &\leq \kappa \|\mathbf{r}^{(h)}\|^2 \\ &\leq \dots \\ &\leq \kappa^{-(h+1)} \|\mathbf{x}\|^2 \end{aligned} \quad (\text{A.10})$$

où  $0 \leq \kappa = \frac{(K-1)(1+\mu_{\Phi})}{K}$ . Puisque  $K\mu_{\Phi} < \frac{1-\rho}{1+\rho}$  où  $0 \leq \rho < 1$  on déduit que  $\kappa \leq 1 - 1/K^2 \leq 1$ . Ceci prouve ainsi la convergence de MCA/MOM.

## Annexe B

# *Compressed Sensing* en Astronomie

La mission spatiale Herschel (ESA) sera lancée durant le premier semestre de l'année 2009. L'ESA est confrontée à un important problème de compression de données : le système de compression avec pertes actuellement mis en place au sein du satellite ne répond pas aux attentes de performances prévues : un facteur de compression de 6 reste nécessaire. La solution proposée par l'ESA consiste à n'envoyer que la moyenne de 6 images consécutives. Cependant, le satellite Herschel pouvant effectuer des sessions de prises de vue rapides, la moyenne de 6 images consécutives entraîne une dégradation irréversible importante de la résolution des images envoyées. Récemment, de nouvelles avancées en traitement de l'information ont conduit à l'édification d'une nouvelle théorie de l'échantillonnage appelée *compressed sensing*. Cette théorie permet l'échantillonnage compressé de données supposées parcimonieuses. Dans (BSO08), nous avons proposé une application du *compressed sensing* (CS) pour la résolution du problème de compression de la mission Herschel. Nous montrons en particulier que le CS permet une compression non-adaptative et peu coûteuse en temps de calcul applicable dans le cadre d'une mission spatiale. De plus, nous montrons que l'étape de décodage du (CS) permet la décompression jointe d'un ensemble de données cohérentes (par exemple 6 images consécutives pour Herschel) améliorant ainsi la qualité de restauration des données compressées. Au sein du projet Herschel, la mise en oeuvre du CS permet une amélioration de 30% de la résolution des images reconstruites pour une sensibilité identique par rapport à la solution actuelle de l'ESA. Nous donnons ici en annexe le texte original de l'article (BSO08).

## Introduction

From year to year, the quantity of astronomical data increases at an ever growing rate. In part this is due to very large digital sky surveys in the optical and near infrared, which

---

in turn has been made possible by the development of digital imaging arrays such as CCDs (charge-coupled devices). The size of digital arrays is continually growing, pushed by the demands of astronomical research for ever larger quantities of data in ever shorter time periods. As a result, the astronomical community is also confronted with a rather desperate need for data compression techniques. Several techniques have in fact been used, or even developed, for astronomical data compression (to quote a few : (WPL92; SMPA96)). For some projects, we need to achieve huge compression ratios, which cannot be obtained by current methods without introducing unacceptable distortions. Furthermore, for most astronomical data compression problems, three main properties must be under control : resolution (point spread function), sensitivity (ability to detect low level signals) and photometry. The Herschel satellite<sup>1</sup>, which will be launched in late 2008/early 2009, is faced with a similar problem. Indeed the photometer data need to be compressed by a factor of approximately 6 to be transferred. As the CPU load has to be extremely small, conventional compression methods cannot be used.

Recently, an alternative sampling theory has emerged which shows that signals can be recovered from far fewer samples (measurements) than what the Nyquist/Shannon sampling theory states. This new theory coined *compressed sensing* or (compressive sensing) introduced in the seminal papers (CRT06; Don06) relies on the compressibility of signals or more precisely on the property for some signals to be sparsely represented. From the compressed sensing viewpoint, sparse signals could be acquired “economically” (from a few samples) *without loss of information*. It introduces new conceptions in data acquisition and sampling.

*Scope of the paper :* We propose a new alternative approach for the transmission of astronomical images, based on CS. Section B.1 reviews the principle of the CS theory. We will see that CS can be used in different ways for data compression purposes : i) sensor design or ii) as a classical “compression-decompression” two-stage scheme. In this paper, we will particularly focus on the latter. Indeed, in practical situations (more particularly for onboard applications), CS provides a simple coding or compression stage that only requires a low computational burden. Most of the computational complexity is in the decoding step. In Section B.2 we show how CS offers us a flexible data compression framework as : i) compression et decompression are decoupled and ii) CS is able to account for the redundancy of the data due to some particular observational strategies to enhance the decoding process. It is particularly profitable when multiple observations of the same sky area are available. This happens very often in astronomical imaging when we need to build a large map from a micro-scan or a raster-scan. Section B.3 highlights the effectiveness of the proposed CS-based compression for solving the Herschel data compression problem. Indeed, we show the advantage of CS over the averaging approach (proposed by the European Spatial Agency - ESA) which has been considered so far.

---

<sup>1</sup>See <http://www.esa.int/science/herschel>



## B.1 A brief introduction to compressed sensing

In this section, we give a brief and non-exhaustive review of compressed sensing and show how this new sampling theory will probably lead to a “revolution” in signal processing and communication theory. For more exhaustive tutorials in this field, we refer the reader to the review papers (Can06; Bar07). In this paper, we will assume that the signal  $x$  belongs to  $\mathbb{R}^t$  (written as a column vector with  $t$  entries or samples). We will also assume that  $x$  is **compressible**.

### B.1.1 The gist of compressed sensing

**Compressibility** The content of most astronomical images is often well structured : diffuse gas clouds, point sources ... to name only a few. Recent advances in harmonic analysis have provided tools that efficiently represent such structures (wavelets, curvelets, ... to name a few). In this context, efficient representations mean sparse representations. Let us consider a signal  $x$  of size  $t$ . Let  $\Phi$  be an orthonormal basis (e.g. classically, an orthogonal wavelet basis for astronomical data processing) and let us consider the projection of  $x$  in  $\Phi$  :

$$\alpha = \Phi^T x \quad (\text{B.1})$$

The signal  $x$  is said to be **sparse** in  $\Phi$  if most entries of the so-called coefficient vector  $\alpha$  are zero or close to zero and thus only a few have significant amplitudes. In other words, the signal  $x$  can be efficiently approximated (with low  $\ell_2$  approximation error or distortion) from only a few significant coefficients. Then such sparse signal is said to be *compressible*. Note that, in the last decade, sparsity has emerged as one of the leading concepts in a wide range of signal processing applications. More formally, we will distinguish two categories of compressible signals :

- **Strict sparsity** : only  $K \ll t$  entries of  $\alpha$  are different from zero.  $x$  is said to be  $K$ -sparse in  $\Phi$ .
- **Wide sense compressibility** : A more realistic definition of compressibility consists in describing how the entries of  $\alpha$  behave. Let us consider that,  $\alpha[1] \geq \dots \geq \alpha[t]$ . Then  $x$  is said to be compressible in  $\Phi$ , if there exists  $p \in ]0, 1]$  such that :  $|\alpha[i]| \leq C i^{-1/p}$ . Here  $p$  defines a kind of sparsity or compressibility degree. Real-world data are more akin to be wide sense compressible.

**Acquiring incoherent measurements** Assuming that  $x$  is compressible (i.e.  $x$  has a sparse representation in a particular basis  $\Phi$ ),  $x$  can be efficiently approximated (with low distortion) from only a few entries of  $\alpha = \Phi^T x$ . In the Compressed Sensing framework, the signal  $x$  is not acquired directly ; it is *observed* or *measured* from  $M < t$  measurements  $\{y_k\}_{k=1, \dots, M}$ . These measurements are obtained by projecting the signal  $x$  on a set of so-called measurement vectors  $\{\theta_k\}_{k=1, \dots, M}$  as follows :

$$y_k = \langle x, \theta_k \rangle \quad (\text{B.2})$$


---

Each sample  $y_k$  is then the scalar product of  $x$  with a specific *measurement* vector  $\theta_k$ . The gist of compressed sensing relies on two major concepts : i) the data to be compressed are indeed compressible ; more precisely the data  $x$  have a “structured” content so that they can be sparsely represented in some basis  $\Phi$  ; ii) the measurement vectors  $\{\theta_k\}_{k=1,\dots,M}$  are non adaptive (they should not depend on  $x$ ) and incoherent with the basis in which  $x$  is assumed to be sparse.

As stated in (Can06), two main categories of measurement ensembles can be used for CS coding :

- **Random measurements** :  $\Theta$  is not explicitly used ; the measurements  $\{y_k\}_{k=1,\dots,M}$  are random linear combinations of the entries of  $x$ . Fourier, Binary or Gaussian measurements are widely used. In the CS framework, incoherent measurements can be obtained by using random ensembles (see (DY06; CRT06) and references therein). Randomness is likely to provide *incoherent* projections.
- **Incoherent measurements** : In that case,  $\Theta$  is a deterministic basis which is assumed to be *incoherent* with  $\Phi$ . More quantitatively, incoherence between  $\Theta$  and  $\Phi$  is measured by their mutual coherence :  $\mu = \max_{i,k} |\langle \phi_i, \theta_k \rangle|$ . The lower  $\mu$  is, the more incoherent  $\Theta$  and  $\Phi$  are. In practical situations, typical astronomical data are compressible in a wavelet basis  $\Phi$  ; a good choice for  $\Theta$  is the noiselet basis (CGM01).

In this paper, measurement vectors are designed by selecting at random a set (indexed by  $\Lambda$ ) of vectors from a deterministic orthonormal basis  $\Theta$  as suggested in (CR06) :  $y = \Theta_\Lambda x$ .

An empirical interpretation : In the CS framework, the signal to be transferred is  $y = \Theta\Phi\alpha$ . Let recall that the backbone of CS is twofold :

- **the data are compressible** : it means that only a few entries of  $\alpha$  have a significant amplitude ;  $x$  is then almost entirely determined from only a few entries  $\alpha$ .
- **the measurements are incoherent** : the measurement matrix  $\Theta$  and  $\Phi$  are incoherent. From a empirical point of view, the incoherence of  $\Theta$  and  $\Phi$  means that the information carried by a few entries of  $\alpha$  is spread all over the  $M$  entries of  $y = \Theta\Phi\alpha$ . Each sample  $y_k$  is likely to contain a piece of information of each significant entry of  $x$ . As  $M < t$ , the ratio  $\rho = M/t$  is equivalent to a compression ratio.

### B.1.2 Signal recovery

**Exact solutions** The previous paragraph emphasized on the way the compression step should be devised. The decompression step amounts to recovering the original signal  $x$  out of the compressed signal  $y = \Theta_\Lambda x$ . Furthermore,  $x$  is known *a priori* to be compressible in  $\Phi$ . Then the recovery problem boils down to emphasizing on the sparsity of the vector  $\alpha = x\Phi$ . As proposed in (CRT06; Don06), the decompression or decoding step is equivalent to solving the following optimization problem :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s.t. } y = \Theta_\Lambda \Phi \alpha \quad (\text{B.3})$$

Several strong recovery results in the particular CS framework have been proved based on specific assumptions with random measurement ensembles (see (Don06; CRT06)). For instance, in the extreme strict sparsity case where only  $K$  entries of  $\alpha$  are non-zero, conditions

---

are given in (CRT06) showing that the problem in Equation (B.3) provides the *exact* solution  $\alpha$ . Nevertheless, the data are often corrupted by noise. A more realistic compression model would be the following :

$$y = \Theta_{\Lambda}(x + n) \quad (\text{B.4})$$

where  $n$  is a white Gaussian noise with variance  $\sigma_n^2$ . As the measurement matrix  $\Theta_{\Lambda}$  is a sub-matrix of the orthonormal matrix  $\Theta$ , the projected noise  $n_{\Lambda} = \Theta_{\Lambda}n$  is still white and Gaussian with the same variance  $\sigma_n^2$ . The projected data are then recast as follows :  $y = \Theta_{\Lambda}x + n_{\Lambda}$ . The recovery step then boils down to solving the next optimization problem :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s.t. } \|y - \Theta_{\Lambda}\Phi\alpha\|_{\ell_2} \leq \epsilon \quad (\text{B.5})$$

where  $\epsilon$  is an upper bound of  $\|n\|_{\ell_2}$ . Defining  $\epsilon = \sqrt{M + 2\sqrt{2M}}\sigma_n$  provides a reasonable upper bound on the noise  $\ell_2$  norm, with overwhelming probability. In (Can08) conditions are given that guarantee some optimality results from the problem in Equation (B.5).

The convex program (second-order cone program) in Equation (B.5) then provides an efficient and robust mechanism to provide an approximate to the signal  $x$ .

## B.2 Compressed Sensing in Astronomy

In the next sections, we focus on applying the compressed sensing framework to astronomical remote sensing. In Section B.1, we show that compressed sensing and more precisely its way of coding information provides alternatives to astronomical instrument design. Section B.1.2 gives emphasis on the ability of CS decoding to easily account for the redundancy of the data (for instance, provided by specific observational strategies) thus improving the whole compression performances.

### B.2.1 Compressed Sensing as a way of designing sensors

The philosophy of compressed sensing (*i.e.* projecting onto incoherent measurement ensembles) should be directly applied in the design of the detector. Devising an optical system that directly “measures” incoherent projections of the input image would provide a compression system that encodes in the analog domain.

Interestingly, this kind of measurement paradigm is far from being science-fiction. Indeed, in the field of  $\gamma$ -ray imaging, the so-called coded-masks<sup>2</sup> (see (Ski02) and references therein) are used since the sixties and are currently operating in the ESA/Integral space mission<sup>3</sup>. In  $\gamma$ -ray (high energy) imaging, coded masks are used as aperture masks scattering the incoming  $\gamma$  photons. Empirically,  $\gamma$ -ray data are overwhelmingly composed of point sources (*i.e.* already rather sparse). The coded mask scatters the photons coming from each

---

<sup>2</sup>We invite the interested readers to visit the following site that is devoted to coded aperture imaging : <http://astrophysics.gsfc.nasa.gov/cai/>.

<sup>3</sup>See <http://sci.esa.int/science-e/www/area/index.cfm?fareaid=21>.

point sources over almost all the detectors. Each detector then provide incoherent projection. The last step to compressed sensing would amount to selecting only a few detector's signal (potentially at random) to be transmitted thus completing the compression stage. The first application of compressed sensing then dates back to the sixties ! In the compressed sensing community, the coded mask concept has inspired the design of the celebrated "compressed sensing camera" (TLW<sup>+</sup>06) that provide effective image compression with a single pixel. Simulations involving coded mask for compressive imaging have been made by R. Willet in (MW08).

Similarly, (radio-) interferometric data are acquired in the Fourier domain (*i.e.* visibilities), and the observed signal is known to be sparse in the direct domain for unresolved objects or in the wavelet domain for extended sources. A posteriori, it is not surprising that the best methods which have been proposed in the past for reconstructing such images are based on sparsity. Indeed, the famous CLEAN algorithm (H74) and its multiscale version (WS88; SM06) can be seen as matching pursuit algorithms (LAM97), and therefore enforce the sparsity of the solution, in the direct space for CLEAN and in the wavelet space for Multiresolution CLEAN. Recent  $l_0$  and  $l_1$  CS reconstruction methods (see for instance (DTDS06), (CWB08)) would certainly improve the computation time to derive these dirac-sparse and wavelet-sparse solutions.

### B.2.2 A flexible compression/decompression scheme

In this Section, we will particularly focus on using Compressed Sensing as a classical "compression/decompression" two stages technique. We will emphasize on showing that Compressed Sensing have several advantages over standard compression techniques such as the celebrated JPEG2000<sup>4</sup> compression standard.

#### Computational complexity

The onboard computational complexity of CS compression depends on the choice of  $\Theta_\Lambda$  (for instance, noiselets in the forthcoming experiments). The only computational cost required by a CS-based compression is the computation of  $\Theta_\Lambda x$ . When noiselets are used, their computational cost evolves as  $O(t)$  thus involving a low CPU load which is lower than the computational burden required by JPEG2000 (typically  $O(t \log(t))$ ). Furthermore, the CS projections (noiselets in the forthcoming examples), require no further encoding in contrast to classical compression methods such as JPEG2000.

#### Decoupling

In contrast to classical compression techniques, there is a complete decoupling between the compression and the decompression in the CS framework. Therefore the decompression step can be changed while keeping the same compressed data. This is a very nice property.

---

<sup>4</sup>See <http://www.jpeg.org/jpeg2000/>.

Indeed, we have seen that the quality of the decompressed data is related to the sparsity of the data in a given basis  $\Phi$ . If we discover in a few years a new basis which leads to a better sparsity of the data, then we can still improve the quality of the decompressed data.

### A data fusion perspective : accounting for the redundancy of the data

In astronomy, remote sensing data involving specific scanning strategies (raster scans) often provide redundant data which cannot be accounted for by standard compression techniques. In a general framework, let us consider that  $N$  observations of the same sky area are available :  $\{y_i\}_{i=1,\dots,N}$  such that :

$$\forall i = 1, \dots, N; \quad y_i = \Theta_{\Lambda_i} x + n_i \quad (\text{B.6})$$

where  $\{\Theta_{\Lambda_i}\}_{i=1,\dots,N}$  are  $N$  independent random sub-matrices of  $\Theta$  with  $\text{Card}(\Lambda_i) = M$ . Clearly, it would be worth recovering  $x$  from the  $N$  compressed observations  $\{y_i\}_{i=1,\dots,N}$ . We then propose to substitute the decompression problem in Equation (B.5) with the following :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s.t. } \sum_{i=1}^N \|y_i - \Theta_{\Lambda_i} \Phi \alpha\|_{\ell_2} \leq \sqrt{N} \epsilon \quad (\text{B.7})$$

Owing to the convexity of this problem, it can be recast in the following Lagrangian form :

$$\min_{\alpha} \gamma \|\alpha\|_{\ell_1} + \frac{1}{2} \sum_{i=1}^N \|y_i - \Theta_{\Lambda_i} \Phi \alpha\|_{\ell_2}^2 \quad (\text{B.8})$$

Let define  $\Pi_i$  as the diagonal matrix of size  $t \times t$  the entries of which are defined as follows :

$$\forall k \in \{1, \dots, t\}; \quad \Pi_i[k, k] = \begin{cases} 1 & \text{if } k \in \Lambda_i \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.9})$$

where  $\Pi_i[k, k]$  is the  $k$ -th diagonal element of  $\Pi_i$ . Let define the signal  $y_i^\#$  of size  $t$  as follows :

$$y_{i\Lambda_i}^\# = y_i \text{ and } y_{i\Lambda_i^c}^\# = 0 \quad (\text{B.10})$$

where  $\Lambda_i^c$  is complement of  $\Lambda_i$  in  $\{1, \dots, t\}$ . Inspired by recent iterative thresholding methods ([ESDQ05](#); [CW05](#)), we propose solving the problem in Equation (B.8) by means of a projected Landweber iterative algorithm. At iteration ( $h$ ), the coefficients  $\alpha$  would be updated as follows :

$$\alpha^{(h)} = S_\gamma \left\{ \alpha^{(h-1)} + \frac{1}{N} \sum_{i=1}^N \Phi^T \Theta^T \Pi_i (y_i^\# - \Pi_i \Theta \Phi \alpha^{(h-1)}) \right\} \quad (\text{B.11})$$

where  $S_\gamma$  is the soft-thresholding operator with threshold  $\gamma$ . Further calculation entails appreciable simplifications :

$$\alpha^{(h)} = S_\gamma \left\{ \frac{1}{N} \Phi^T \Theta^T \sum_{i=1}^N [y_i^\# + (\mathbf{I} - \Pi_i) \Theta \Phi \alpha^{(h-1)}] \right\} \quad (\text{B.12})$$

where  $\mathbf{I}$  is the identity matrix of size  $t \times t$ .

---

**Choosing the regularization parameter  $\gamma$**  The problem in Equation (B.12) is a general sparse decomposition problem in a redundant dictionary (see (BDE07; Tro06b) and references therein). The choice of the regularization parameter  $\gamma$  is crucial as it balances between the sparsity constraint and the how the solution fits the data. Classical approaches would advocate the use of cross-validation to estimate an optimal value of  $\gamma$ . Nevertheless, cross-validation is computationally expensive and thus not appropriate for large scale problems.

In a different framework, the same kind of optimization problem has been solved using a specific iterative hard-thresholding algorithm coined Morphological Component Analysis (MCA - see (SED04)) for which the threshold decreases. Inspired by MCA, hard-thresholding is used and the threshold  $\gamma$  is decreased at each iteration. It starts from  $\gamma^{(0)} = \|\Phi^T \Theta^T \sum_i y_i^\# \|_\infty$  and decreases towards  $\gamma_{\min}$ . The value of  $\gamma_{\min}$  is 0 in the noiseless case. When noise corrupts the data  $y^\#$ ,  $\gamma_{\min}$  may depend on the noise level. In Section B.3, numerical results are given. In these experiments, noise contamination is assumed to be white Gaussian with zero mean and variance  $\sigma_n^2$ . In this case, the final threshold is chosen as  $\gamma_{\min} = 3\sigma_n$  which gives an upper bound for noise coefficients with overwhelming probability.

### B.2.3 Example : joint recovery of multiple observations

For instance, let us consider that the data are made of  $N = 100$  images  $\{x_i\}_{i=1,\dots,N}$  such that each image  $x_i$  is a noiseless observation of the same sky area  $x$ . We propose to decompress the set of observations  $\{x_i\}_{i=1,\dots,N}$  in a joint recovery scheme. Each observation is compressed using CS such that  $\rho = M/t = 0.1$ . Compression is made by solving the problem in Equation (B.8) using the iterative thresholding algorithm described previously. Figure B.1 provides the compression/decompression results of  $x$  using CS and JPEG2000. At first sight, both techniques behave similarly. Figure B.2 depicts the zoomed versions of the previous pictures. Clearly, CS provides better visual results. Recall that in astronomy, the main properties to be under control are : i) the resolution, ii) the sensitivity and iii) the photometry. Let us have a look at photometry by defining the intensity of  $x$  as :  $f = \sum_{j=1}^t x[j]$ . Both compression techniques can be compared in terms of their relative intensity error defined as the ratio :  $(f - \tilde{f})/f$  where  $\tilde{f}$  is the intensity of the recovered image. Quantitative results are presented in Table I.

	Recovery SNR	Relative intensity error
JPEG2000	9.77dB	$2.610^{-3}$
CS	46.8dB	$1.8210^{-6}$

**TAB. B.1: CS versus JPEG2000 : Recovery SNR in dB and Relative intensity error.**

Where the  $\text{SNR} = -20 \log \frac{x - \tilde{x}}{x}$  where  $\tilde{x}$  is the recovered signal. This huge difference between both compression strategies is the consequence of a fundamental property of CS : **the linearity of the compression**. In contrast to standard compression techniques (such as JPEG2000), the CS-based compression is linear. The data to transmit are indeed simple linear projections :  $y_i = \Theta_{\Lambda_i}(x_i + n_i)$  where  $n_i$  models instrumental noise. Whatever the compression rate (*i.e.*  $\rho = M/t$ ), the incoherence between the measurement vectors  $\Theta_{\Lambda_i}$  and the data is likely to guarantee that  $x$  does not belong to the null space of  $\Theta_{\Lambda_i}$ . As a consequence, the compressed data always contain a piece of information belonging to  $x$ . Standard compression methods (which are non-linear) do not verify this crucial property. For low level details, a standard compression method will kill its high frequencies and they will never be recovered whatever the number of times this source is observed.

**The need for joint decompression** Assume that  $x$  is compressible. More precisely, for a fixed distortion  $\epsilon$ , there exists  $K(\epsilon)$  such that for  $K \geq K(\epsilon) \ll t$ ,  $\|x - x_K\|_{\ell_2} \leq \epsilon$  where  $x_K$  is the  $K$ -term approximation of  $x$  (*i.e.* synthesized from the  $K$  entries of  $\alpha$  that have the most significant amplitudes). The CS theory tells us that the same error  $\epsilon$  (*i.e.* distortion) can be obtained by solving the problem in Equation B.7 from  $M \geq M(\epsilon)$  incoherent measurements. Let us assume that, the set  $\{\Lambda_i\}_{i=1, \dots, N}$  have been generated at random such that, with high probability,  $\forall i \neq i'; \Lambda_i \cap \Lambda_{i'} = \emptyset$  then, the set of measurement vector  $\{\Theta_{\Lambda_i}\}$  is equivalent to a global measurement matrix  $\Theta_{\Lambda}$  where  $\text{Card}(\Lambda) = NM$ . To summarize, recovering  $x$  from the set of measurements  $\{y_i\}_{i=1, \dots, N}$  is likely to provide the same reconstruction that would be able to get from  $NM$  measurements thus leading to better recovery performances.

### B.3 A real-world application : the Herschel project

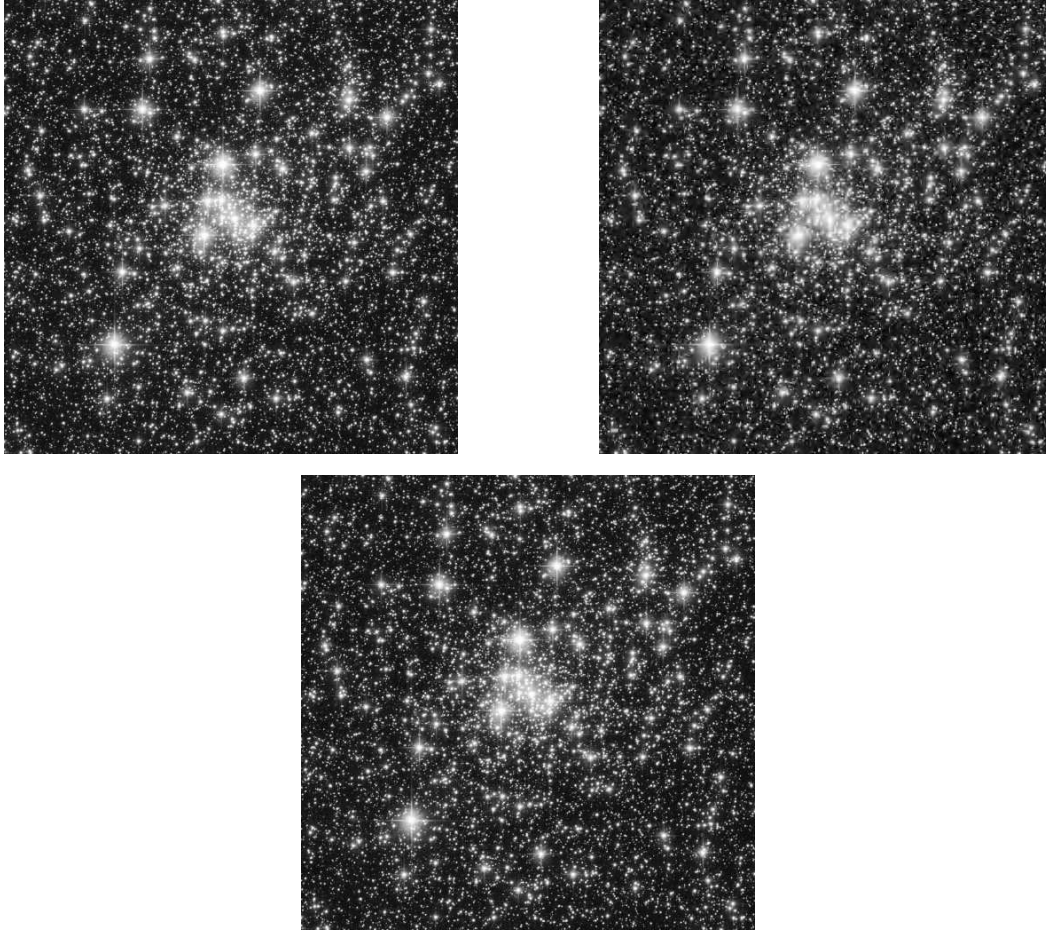
Herschel is one of the cornerstone missions of the European Space Agency (ESA). This space telescope has been designed to observe in the far-infrared and sub-millimeter wavelength range. Its launch is scheduled for the fall of 2008/spring of 2009. The shortest wavelength band, 57-210  $\mu\text{m}$ , is covered by PACS (Photodetector Array Camera and Spectrometer) (PWB<sup>+</sup>06), which provides low to medium resolution spectroscopy and dual-band photometry. When PACS is used as a photometer, it will simultaneously image with its two bolometer arrays, a  $64 \times 32$  and a  $32 \times 16$  matrix, both read out at 40 Hz. The ESA is faced with a challenging problem : conventional low-cost compression techniques cannot achieve a satisfactory compression rate. In this Section, we propose a new CS-based compression scheme for the Herschel/PACS data that yield an elegant and effective way to overcome the Herschel compression dilemma.

#### B.3.1 The Herschel dilemma

The Herschel space telescope is partially hampered by the narrowness of the transmission band compared to the large amount of data to be transferred. This handicap stems

---





**Fig. B.1:** **Left :** Input image  $x$  of size  $512 \times 512$ . **Middle :** Estimate from the average of 100 images compressed by JPEG2000 with a compression rate  $\rho = 0.1$ . **Right :** Estimate from 100 pictures compressed by CS with a compression rate  $\rho = 0.1$ .

from the limitation of conventional compression techniques to provide adequate compression rate with low computational cost, given the high readout noise. More quantitatively, the data have to be compressed in real time by a factor of 6 with very low CPU power.

**Problem statement** The Herschel spacecraft is about to be launched ; ESA is faced with a compression problem as the data need to be compressed by a factor of  $N = 6$ . Up to now, the only acceptable solution for the ESA (with respect to computational cost and quality) to overcome this need for a higher compression rate is the average of  $N = 6$  consecutive images ([BBO<sup>+</sup>05](#)). Indeed, the compression code has no information about the scan speed or the scan direction and a shift-and-add averaging solution is not possible. Other compression techniques such as JPEG or JPEG2000 are also not acceptable because of computation

---



time constraints.

Herschel will observe wide sky areas thus requiring fast scanning strategies. Herschel/PACS will provide sets of consecutive  $64 \times 32$  images that will be shifted with a typical shift value  $\lambda = 1$  pixel in fast scanning mode. Unfortunately, the shift value  $\lambda$  is comparable to the FWHM (full width at half maximum) of the instrumental PSF (point spread function) is  $\delta \approx 3$  pixels. As a consequence, averaging  $N$  consecutive images will entail a catastrophic loss of spatial resolution. This can be catastrophic for some scientific programs. Furthermore, averaging is far less optimal for noise reduction as averaging shifted signals does not yield a  $\sqrt{N}$  noise variance reduction.

An effective compression scheme would have to balance between the following performance criteria :

- **Spatial resolution** : averaging fast scanned data entails a lower spatial resolution. An effective compression scheme should provide a lower resolution loss.
- **Sensitivity** : averaging will reduce noise but will also blur the data thus entailing a loss of sensitivity. Sensitivity (*i.e.* ability to detect low level details or sources) after compression/decompression must be under control.

### B.3.2 Compressed sensing for the Herschel data

The Herschel/PACS mission needs a compression rate equal to  $\rho = 1/N$  with  $N = 6$ . A first approach would amount to compressing independently each image. As stated earlier, accounting for the redundancy of the data is profitable to enhance the global compression/decompression performances. Then, compressing/decompressing  $N = 6$  consecutive images jointly would be more relevant. If we consider a stack of  $N = 6$  consecutive images  $\{x_i\}_{i=1, \dots, N}$ , the simplest generative model is the following :

$$\forall i \in \{1, \dots, N\}; \quad x_i = \mathcal{T}_{\lambda_i}(x) + n_i \quad (\text{B.13})$$

where  $\mathcal{T}_{\lambda_i}$  is an operator that shifts the original image  $x$  with a shift  $\lambda_i$ . In practice,  $x = x_1$  and  $\lambda_1 = 0$ . The term  $n_i$  models instrumental noise or model imperfections. According to the compressed sensing framework, each signal is projected onto the subspace ranged by a subset of columns of  $\Theta$ . Each compressed observation is then obtained as follows :

$$\forall i \in \{1, \dots, N\}; \quad y_i = \Theta_{\Lambda_i} x_i \quad (\text{B.14})$$

where the sets  $\{\Lambda_i\}$  are such that :

$$\sum_i \Pi_i = \mathbf{I} \quad (\text{B.15})$$

$$\forall i \neq i'; \quad \Lambda_i \cap \Lambda_{i'} = \emptyset \quad (\text{B.16})$$

where the cardinality of each subset is  $M = \lfloor t/N \rfloor$ . When there is no shift between consecutive images, these conditions guarantee that the signal  $x$  can be reconstructed univocally

---

from  $\{y_i\}_{i=1,\dots,N}$ , up to noise. Furthermore,  $x$  is assumed to be positive. The decoding step amounts to seeking the signal  $x$  as follows :

$$\min_x \gamma \|\Phi^T \alpha\|_{\ell_1} + \frac{1}{2} \sum_{i=1}^N \|y_i - \Theta_{\Lambda_i} \mathcal{T}_{\lambda_i}(x)\|_{\ell_2}^2 \quad (\text{B.17})$$

We propose solving this problem by using an adapted version of the iterative algorithm we introduced in Section B.1.2. Furthermore, the content of astronomical data is often positive. Constraining the solution to be positive would help solving the recovery problem. Assuming that the shifting operator  $\mathcal{T}_{\lambda_i}$  is invertible<sup>5</sup>, we substitute Equation (B.12) by the following Equation :

$$x^{(h)} = \frac{1}{N} \Phi \mathcal{S}_\gamma \left\{ \Phi^T \sum_{i=1}^N \mathcal{T}_{-\lambda_i} \left( \Theta^T [y_i^\# + (\mathbf{I} - \Pi_i) \mathcal{T}_{\lambda_i}(x^{(h-1)})] \right) \right\} \quad (\text{B.18})$$

The positivity constraint is accounted for by projecting at each iteration the solution of the previous update equation on the cone generated by the vectors having positive entries :  $x^{(h)} \leftarrow P_C(x^{(h)})$  where the projector  $P_C$  is defined as follows :

$$\forall i = 1, \dots, t; \quad P_C(x)[i] = \begin{cases} x[i] & \text{if } x[i] \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.19})$$

where  $P_C(x)[i]$  is the  $i$ -th entry of  $P_C(x)$ . Convergence is guaranteed as long as shifting the image does not deteriorate the original signal. In practice, this condition is not valid, only a portion of the image can be recovered with precision. Similarly to the discussion in Section II, joint decompression should be profitable as sets of consecutive shifted images provide redundant data. In the next section, we illustrate the good performances of the proposed decoding scheme.

## Notations

In the next experiments, the data will be made of pointwise sources ; it is worth defining some useful notations. Recall that we assume the telescope's PSF to have a FWHM equal to  $\delta$ . The shift between the original datum  $x$  and the  $i$ -th datum  $x_i$  is  $\lambda_i$ . The intensity  $f$  of the datum  $x$  (defined in Section B.2). We also assume the  $x$  has positive entries.

### B.3.3 A toy-example

In the following experiments, the datum  $x$  is a  $128 \times 128$  image. The instrument is assumed to have a FWHM  $\delta = 3$  pixels. For the sake of simplicity, each shift  $\lambda_i = i$  pixels. White Gaussian noise is added to account for the instrumental noise. As we stated earlier, three main properties must be under control : i) spatial resolution, ii) sensitivity and iii)

---

<sup>5</sup>This assumption is true when shifting the image does not deteriorate the original signal.

photometry. Concerning the last point, it was shown in section B.2.3, that CS was very efficient compared to JPEG2000. In this section, we will particularly focus on sensitivity and spatial resolution.

### Sensitivity

In this experiment, the datum contains 49 point sources that have been uniformly scattered. The amplitude of each point source is generated at random with a Gaussian distribution. The top-left picture of Figure B.3 shows the input data  $x$ . The additive Gaussian noise has a fixed unit variance. The top-right panel of Figure B.3 features the data  $x$  contaminated with noise. Comparisons between the MO6 (“Mean of 6 images”) and CS methods are made by evaluating for varying intensity value (from 700 to 140000; it is equivalent to a SNR varying from  $-13.2$  to  $33\text{dB}$ ) the rate of detected point sources. To avoid false detection, the same pre-processing step is performed : i) “à trous” bspline wavelet transform (see (SMB06)), ii)  $5\sigma_M$  hard-thresholding<sup>6</sup> where  $\sigma_M$  is the residual standard deviation estimated by a Median Absolute Deviation (MAD) at each wavelet scale, iii) reconstruction. The bottom-left panel of Figure B.3 features such filtered decoded image using the MO6 strategy. The bottom-right picture in Figure B.3 shows the filtered CS-based solution. In this experiment the total intensity of the point sources is set to 3500. At first sight, both methods provide similar detection performances. As expected, the CS-based solution has a better spatial resolution.

Figure B.4 shows the detection rate (with no false detection) of each method for intensities varying from  $f = 700$  to  $f = 140000$ . At high intensity (higher than  $f = 10^4$ ), both MO6 and CS provide rather similar detection performances. Interestingly, at low intensity, CS provides slightly better results. This unexpected phenomenon is partly due to the spread that results from the average of shifted images.

MO6 is theoretically (for low shifts) near-optimal for point source detection. In contrast, this experiment shows that CS can provide similar or better detection performances than MO6.

### Resolution

Spatial resolution is a crucial instrumental feature. Averaging shifted images clearly deteriorates the final spatial resolution of Hershel/PACS. In this experiment, the original datum  $x$  is made of a couple of point sources. In the worst case, these point sources are aligned along the scan direction. The top-left picture of Figure B.5 features the original signal  $x$ . In the top-right panel of Figure B.5, the intensity of the point sources is set to  $f = 1000$  while the noise variance is  $\sigma_n^2 = 1$ . The SNR of the data to compress is equal to  $2.7\text{dB}$ .

---

<sup>6</sup>Such  $5\sigma_M$  is likely to avoid false detection as it defines a rather conservative threshold.

The MO6 solution (*resp.* the CS-based solution) is shown on the left (*resp.* right) at the bottom of Figure B.5. As expected, the spatial resolution of the MO6 is clearly worse than the resolution of the input datum  $x$ . Visually, the CS-based solution mitigate the resolution loss.

For different intensity of the datum  $x$  (from 100 to 2000), the spatial resolution is evaluated according to the Rayleigh criterion. The Rayleigh criterion is the generally accepted criterion for the minimum resolvable detail : two point sources are resolved when the first minimum is lower than the amplitude at half maximum of a single point source as illustrated in Figure B.6. For a fixed intensity  $f$ , the resolution limit is evaluated by seeking the minimal distance between the point sources for which the Rayleigh criterion is verified. For intensities varying from  $f = 100$  to  $f = 2000$ , the resolution limit is reported in Table II.

The CS-based compression scheme provides a solution with better spatial resolution. At high intensity, the resolution gain (in comparison with MO6) is equal to a third of the instrumental FWHM (1 pixel). At low intensity, the resolution gain provided by the CS-based method slightly decreases.

This experiment shows that CS mitigates the resolution loss resulting from the joint compression of 6 consecutive images.

SNR (in dB)	-17.3	-9.35	-3.3	0.21	2.7
Intensity (a.u. <sup>7</sup> )	100	250	500	750	1000
MO6	2.7	2.7	2.7	2.7	2.7
CS	2	2	1.7	1.7	1.7
SNR	4.7	6.2	7.6	8.7	
Intensity	1250	1500	1750	2000	
MO6	2.7	2.7	2.7	2.7	
CS	1.7	1.7	1.7	1.7	

**TAB. B.2: Spatial resolution in pixels :** for varying datum flux, the resolution limit of each compression technique is reported. The CS-based compression entails a resolution gain equal to a 30% of the spatial resolution provided by MO6.

### B.3.4 Realistic data

#### The data

Real Herschel/PACS data are more complex than those we simulated in the previous experiments. The original datum  $x$  is contaminated with a slowly varying “flat field” component  $c_f$ . In a short sequence of 6 consecutive images, the flat field component is almost

fixed. In this context, the data  $\{x_i\}_{i=0,\dots,1}$  can then be modeled as follows :

$$x_i = \mathcal{T}_{\lambda_i}(x) + n_i + c_f \quad (\text{B.20})$$

If  $c_f$  is known (which will be the case in the forthcoming experiments),  $\mathcal{T}_{\lambda_i}(x^{(h-1)})$  is replaced by  $\mathcal{T}_{\lambda_i}(x^{(h-1)}) + c_f$  in Equation B.18. If  $c_f$  is unknown, it can be estimated within the iterative algorithm. The next Section focuses on the resolution gain provided by the CS-based method in the scope of real Herschel/PACS data. The data have been designed by adding realistic pointwise sources to real calibration measurements performed in mid-2007.

## Resolution

Similarly to the experiments performed in Section B.3.3, we added a couple of point sources to Herschel/PACS data. The top-left picture of Figure B.7 features the original signal  $x$ . In the top-right panel of Figure B.7, the intensity of the point sources is set to  $f = 4500$ . The “flat field” component overwhelms the useful part of the data so that  $x$  has at best a level that is 30 times lower than the “flat field” component. The MO6 solution (*resp.* the CS-based solution) is shown on the left (*resp.* right) and at the bottom of Figure B.7 and all the results are presented in Table III. Similarly to the previous fully simulated experiment, the CS-based algorithm provides better resolution performances. The resolution gain can reach 30% of the FWHM of the instrument’s PSF for a wide range of signal intensities. This experiment illustrates the reliability of the CS-based compression to deal with real-world data compression.

SNR (in dB)	−17.3	−9.35	−3.3	0.21	2.7
Intensity (a.u.)	900	2250	4500	6750	9000
MO6	3	3	3	3	3
CS	2.33	2.33	2	2	2
SNR	4.7	6.2	7.6	8.7	
Intensity	11250	13500	15750	18000	
MO6	3	3	3	3	
CS	2	2	2	2	

**TAB. B.3: Spatial resolution in pixels :** for varying datum flux, the resolution limit of each compression technique is reported. The CS-based compression entails a resolution gain equal to a 30% of the spatial resolution provided by MO6.

## B.4 Conclusion

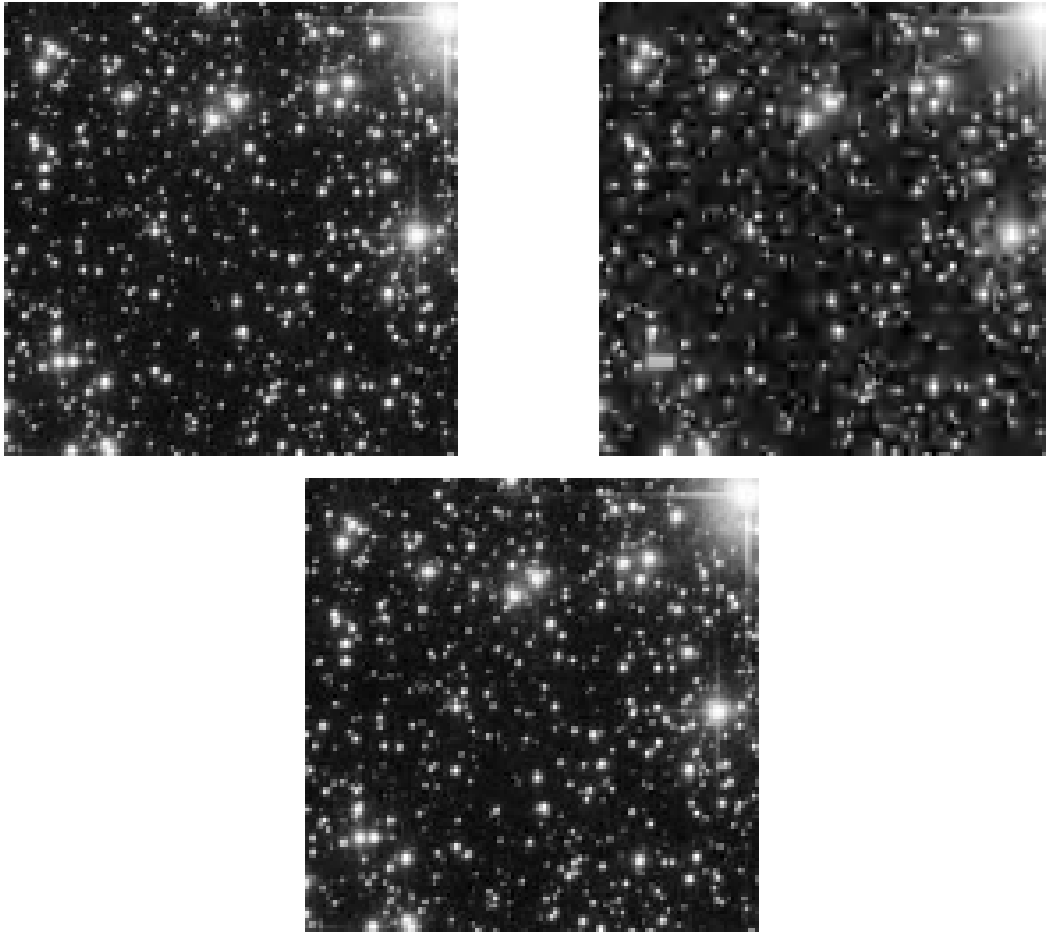
In this paper, we overview the potential applications of compressed sensing (CS) in astronomical imaging. The CS appeal in astronomy is twofold : i) it provides a very easy and

computationally cheap coding scheme for onboard astronomical remote sensing, ii) from a data fusion perspective, the decoding stage is flexible enough to account for the redundancy of the data thus leading to significant recovery enhancements. We particularly point out the huge advantage of compressed sensing over standard compression techniques in the scope of multiple scanning observations (observing the same sky area several times). We have shown that compressed sensing data fusion can lead to improvements compared to standard techniques. Preliminary numerical experiments illustrate the reliability of a CS-based compression scheme in the scope of astronomical remote sensing such as the Herschel space mission. We show that compressed sensing provides an elegant and effective compression technique that overcome the compression issue ESA is faced with. In the next step we will focus on performing more realistic experiments in the scope of the Herschel space mission by adding some physical information : calibration, statistical noise models, flat field estimation . . . to name a few. This paper show that applying CS to the Herschel space mission is of major interest. Indeed, CS is the only existing alternative solution to the averaging solution, and CS enables to recover data with a spatial resolution enhanced up to 30% with similar sensitivity compared to the averaging technique. CS will probably be implemented onboard thus being the first application of compressed sensing for a real-world space mission.

## **B.5 Acknowledgment**

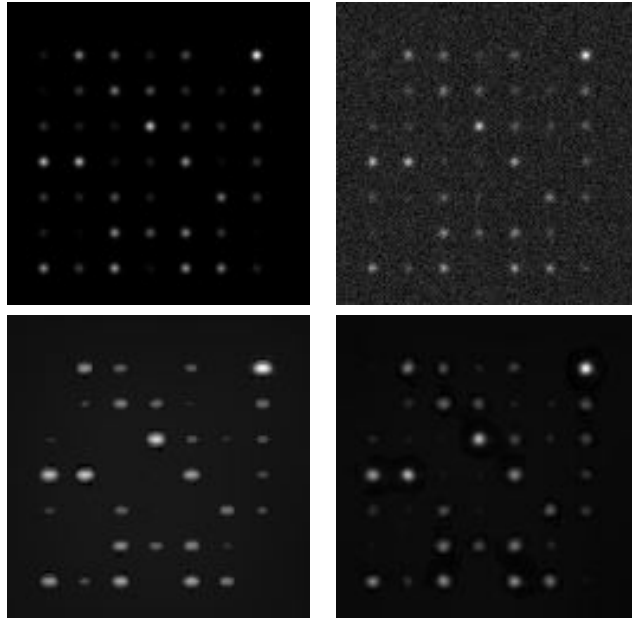
The authors are very grateful to E. Candès for useful discussions and for having provided the noiselet code.

---

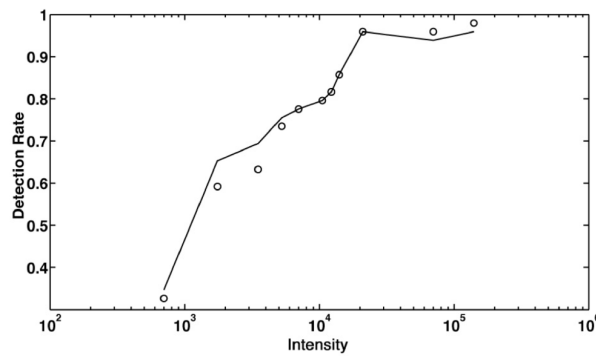


**FIG. B.2: Zoomed versions** - **Left** : Input image  $x$  of size  $512 \times 512$ . **Middle** : Estimate from the average of 100 images compressed by JPEG2000 with a compression rate  $\rho = 0.1$ . **Right** : Estimate from 100 pictures compressed by CS with a compression rate  $\rho = 0.1$ .

---

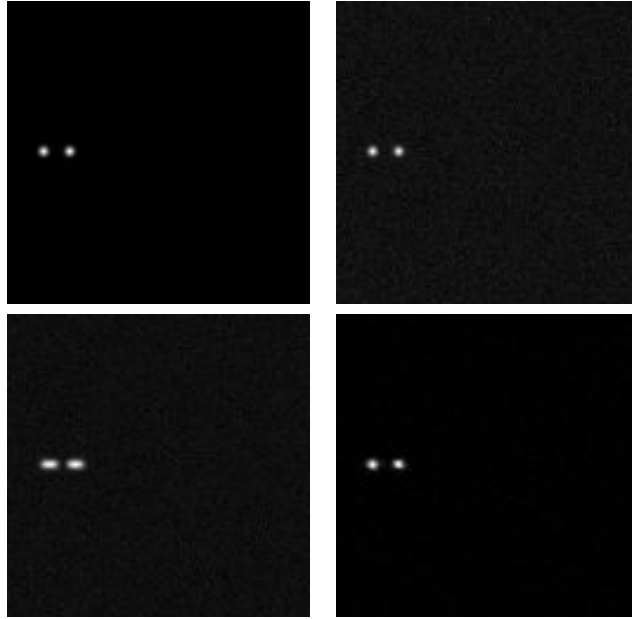


**FIG. B.3:** **Top left :** Original image of size  $128 \times 128$  the total intensity of which is  $f = 3500$ . **Top right :** First input noisy map (out of 6). White Gaussian with variance  $\sigma_n^2 = 1$  was added. **Bottom left :** Mean of the 6 input images. **Bottom right :** Reconstruction from noiselet-based CS projections. The iterative algorithm described in Section B.2 has been used with 100 iterations.

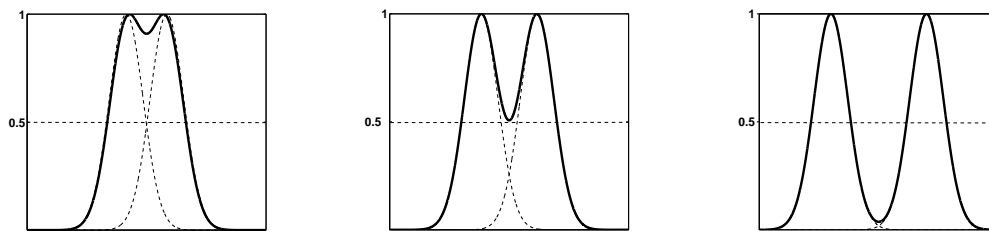


**FIG. B.4:** Detection rate when the intensity of the input data varies : *Solid line* Resolution defined by the Rayleigh criterion of the CS-based reconstruction.  $\circ$  : Resolution of the solution provided by the mean of 6 images.

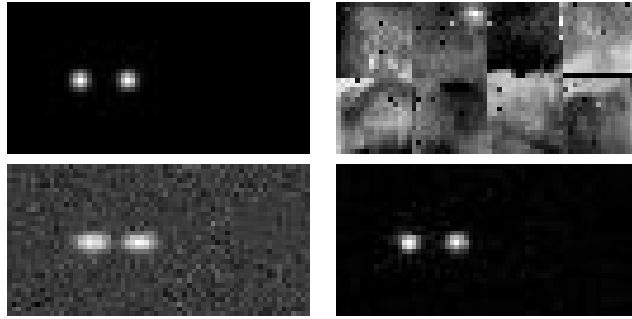




**FIG. B.5:** **Top left :** Original image of size  $128 \times 128$  the total intensity of which is  $f = 1000$ . **Top right :** First input noisy map (out of 6). White Gaussian with variance  $\sigma_n^2 = 1$  was added. **Bottom left :** Mean of the 6 input images. **Bottom right :** Reconstruction from noiselet-based CS projections. The iterative algorithm has been used with 100 iterations.



**FIG. B.6: The Rayleigh criterion - Left :** The point sources are not resolved. **Middle :** Resolution limit. **Right :** Fully resolved point sources.



**FIG. B.7:** **Top left :** Original image of size  $32 \times 64$  with a total intensity of  $f = 4500$ . **Top right :** First input noisy map (out of 6). The PACS data already contains approximately Gaussian noise. **Bottom left :** Mean of the 6 input images. **Bottom right :** Reconstruction from noiselet-based CS projections. The iterative algorithm has been used with 100 iterations.

---

# Bibliographie

- [AC96] S.I. Amari and J.-F. Cardoso. Blind source separation : semiparametric statistical approach. *IEEE Tr. on Signal Processing*, 45(11), November 1996.
  - [AEB06] M. Aharon, M. Elad, and A. Bruckstein. k-SVD : An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11) :4311–4322, 2006.
  - [AFB01] N. Aghanim, O. Forni, and F. R. Bouchet. Using the COBE/DMR data as a test-bed for normality assessments. *Astronomy and Astrophysics*, 365 :341–346, January 2001.
  - [Aka70] H. Akaike. Statistical predictor estimation. *Annals Inst. Stat. Math.*, 22 :203, 1970.
  - [AMS<sup>+</sup>07] P. Abrial, Y. Moudden, J.L. Starck, J. Bobin, M.J. Fadili, B. Afeyan, and M. Nguyen. Morphological component analysis and inpainting on the sphere : Application in physics and astrophysics. *Journal of Fourier Analysis and Applications (JFAA)*, 13(6) :729–748, 2007.
  - [Att54] F Attneave. Some informational aspects of visual perception. *Psychological Review*, 3(61) :183–193, 1954.
  - [Bal07] Radu Balan. Estimator for number of sources using minimum description length criterion for blind sparse source mixtures. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D Plumbley, editors, *Independent Component Analysis and Signal Separation*, volume 4666 of *LNCS*, pages 333–340. Springer, 2007.
  - [Bar61] H.B Barlow. *Possible principles underlying the transformation of sensory messages*, pages 217–234. Sensory Communications - W.Rosenblith, 1961.
  - [Bar07] R. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, July 2007.
  - [BBO<sup>+</sup>05] A. N. Belbachir, H. Bischof, R. Ottensamer, F. Kerschbaum, and C. Reimers. On-board data processing to lower bandwidth requirements on an infrared
-

- astronomy satellite : Case of herschel-pacs camera. *EURASIP Journal for Applied Signal Processing*, 15 :2585–2594, 2005.
- [BBS<sup>+</sup>06] A. Bonaldi, L. Bedini, E. Salerno, C. Baccigalupi, and G. De Zotti. Estimating the spectral indices of correlated astrophysical foregrounds by a second-order statistical approach. *Mon.Not.Roy.Astron.Soc.*, 373 :263–270, 2006.
- [BBZZ05] A.M Bronstein, M.M Bronstein, M. Zibulevsky, and Y.Y. Zeevi. Sparse ica for blind separation of transmitted and reflected images. *Intl. Journal of Imaging Science and Technology (IJIST)*, 15/1 :84–91, 2005.
- [BD07] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications* - submitted, 2007.
- [BDE07] A.M. Bruckstein, D.L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 2007. to appear.
- [BG99] R. Bouchet and R. Gispert. Foregrounds and cmb experiments : I. semi-analytical estimates of contamination. *New Astronomy*, 4(443), 1999.
- [BL97] Pearlmutter B. and Parra L.C. Maximum likelihood blind source separation : A context-sensitive generalization of ica. *Advances in Neural Information Processing Systems*, 9, 1997.
- [BL07] K. Bredies and D. A. Lorenz. Iterated hard shrinkage for minimization problems with sparsity constraints. In *Preprint*, 2007.
- [BMCM97] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2) :434–444, 1997.
- [BMFS07a] J. Bobin, Y. Moudden, J. Fadili, and J.-L. Starck. Morphological diversity and sparsity in blind source separation. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D Plumbley, editors, *Independent Component Analysis and Signal Separation*, volume 4666 of *LNCS*, pages 349–356. Springer, 2007.
- [BMFS07b] J. Bobin, Y. Moudden, M. J. Fadili, and J.-L. Starck. Morphological diversity and sparsity for multichannel data restoration. *Journal of Mathematical Imaging and Vision* - in press., 2007.
- [BMS<sup>+</sup>08] J. Bobin, Y. Moudden, J.-L. Starck, J. Fadili, and N.Aghanim. Sz and cmb reconstruction using gmca. *Statistical Methodology*, (5) :307–317, 2008.
- [BMSE06] J. Bobin, Y. Moudden, J.-L. Starck, and M. Elad. Morphological diversity and source separation. *IEEE Trans. on Signal Processing*, 13(7) :409–412, 2006.
-

- [Bob07] J. Bobin. Gmcalab. <http://pagesperso-orange.fr/jbobin/gmcalab.html>, 2007.
- [BS95] A.J. Bell and T.J. Sejnowski. An information maximisation approach to blind separation and blind deconvolution,. *Neural Computation*, 7(6) :1129–1159, 1995.
- [BS97] A.J. Bell and T.J. Sejnowski. The ‘independent components’ of natural scenes are edge filters. *Vision Research.*, 23(37) :3327–3338, 1997.
- [BSF<sup>+</sup>07] J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, and D.L. Donoho. Morphological component analysis : An adaptive thresholding strategy. *IEEE Trans. On Image Processing*, 16(11) :2675 – 2681, November 2007.
- [BSFM07] J. Bobin, J.-L. Starck, M. J. Fadili, and Y. Moudden. Sparsity and morphological diversity in blind source separation. *IEEE Transactions on Image Processing*, 16(11) :2662 – 2674, November 2007.
- [BSO08] J. Bobin, J.-L. Starck, and R. Ottensamer. Compressed sensing in astronomy. *IEEE Journal of Selected Topics in Signal Processing - revised*, 2008.
- [BST98] A.G. Bruce, S. Sardy, and P. Tseng. Block coordinate relaxation methods for nonparametric signal de-noising. *Proceedings of the SPIE - The International Society for Optical Engineering*, 3391 :75–86, 1998.
- [Can05] E.J. Candès. Modern statistical estimation via oracle inequalities. *Acta Numerica*, 15 :257–325, 2005.
- [Can06] Emmanuel Candès. Compressive sampling. *International Congress of Mathematics, Madrid*, 2006.
- [Can07] E. J. Candès.  $\ell_1$ -magic. Technical report, Caltech, 2007.
- [Can08] E. Candès. The restricted isometry property and its implications for compressed sensing. Technical report, Caltech, 2008.
- [Car] J. F. Cardoso. Cocktail party et autres problèmes compliqués. Technical report, ENST, <http://tsi.enst.fr/EIJcardoso/stuff.html>.
- [Car97] J. F. Cardoso. Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing*, 4(4) :112–114, 1997.
- [Car99] Jean-François Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1) :157–192, January 1999.
- [Car03] Jean-François Cardoso. Dependence, correlation and non Gaussianity in independent component analysis. *Journal of Machine Learning Research*, 4 :1177–1203, December 2003.
-

- [CCPW07] C. Chaux, P.L. Combettes, J.-C. Pesquet, and V.R. Wajs. A variational formulation for frame based inverse problems. *Inverse Problems*, 23 :1495–1518, 2007.
  - [CD99] E.J. Candès and D.L. Donoho. Curvelets. Technical report, Statistics, Stanford University, 1999.
  - [CDDY06] E. Candès, L. Demanet, D.L. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Model. Simul.*, 5/3 :861–899, 2006.
  - [CDS98] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20 :33–61, 1998.
  - [CGM01] R. Coifman, F. Geshwind, and Y Meyer. Noiselets. *Appl. Comput. Harmon. Anal.*, 10(1) :27–44, 2001.
  - [CH05] J. Chen and X. Huo. Sparse representations for multiple measurement vectors (MMV) in an over-complete dictionary. In *ICASSP '05*, 2005.
  - [Cic05] A. Cichocki. New tools for extraction of source signals and denoising. In *Proc. SPIE 2005, Bellingham*, volume 5818, pages 11–24, 2005.
  - [Com94] P. Comon. Independent component analysis, a new concept ? *Signal Processing*, 36(3) :287–314, April 1994.
  - [CR03] R.A. Choudrey and S.J. Roberts. Variational mixture of bayesian independent component analysers. *Neural Computation*, 15(1), 2003.
  - [CR06] Emmanuel Candès and Justin Romberg. Sparsity and incoherence in compressive sampling. *Preprint - available at <http://www.dsp.ece.rice.edu/cs/>*, 2006.
  - [CREKD05] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing*, 53 :2477–2488, 2005.
  - [CRT06] Emmanuel Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory*, 52(2) :489–509, 2006.
  - [CS96] Jean-François Cardoso and Antoine Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM Journal of Matrix Analysis and Applications*, 17(1) :161–164, 1996.
  - [CT06] E. Candès and T.Tao. The dantzig selector : Statistical estimation when p is much larger than n. *Annals of Statistics*, 2006. to be published.
  - [Cur06] Curvelab 2.01 for Matlab7.x. Curvelab. (<http://www.curvelet.org/>), 2006.
-

- [CW92] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. on Information Theory*, 38(2) :713–718, 1992.
  - [CW05] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4) :1168–1200, 2005.
  - [CWB08] E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted  $\ell_1$  minimization. Technical report, Caltech, 2008.
  - [Dar53] G. Darmais. Analyse générale des liaisons stochastiques. *Rev. Inst. Internat. Stat.*, pages 2–8, 1953.
  - [Dav04] M. Davies. Identifiability issues in noisy ica. *IEEE Signal processing letters*, 11 :470–473, 2004.
  - [DC05] J. Delabrouille and J-F. Cardoso. Lecture paper for the international summer school 'data analysis in cosmology'. 2005.
  - [DCP03] J. Delabrouille, J.-F. Cardoso, and G. Patanchon. Multi-detector multi-component spectral matching and applications for CMB data analysis. *Monthly Notices of the Royal Astronomical Society*, 346(4) :1089–1102, December 2003. to appear, also available as <http://arXiv.org/abs/astro-ph/0211504>.
  - [DDM04] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. 57(11) :1413–1457, Aug 2004.
  - [DE03] D.L. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via  $\ell^1$  minimization. *Proc. Nat. Aca. Sci.*, 100 :2197–2202, 2003.
  - [DET06] D.L. Donoho, M. Elad, and V Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. On Information Theory*, 52 :6–18, 2006.
  - [DF01] D.L. Donoho and G. Flesia. Can recent innovations in harmonic analysis explain key findings in natural image statistics ? *Network : Computation in Neural Systems*, 12 :371–93, 2001.
  - [DH01] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7) :2845–2862, 2001.
  - [Don93] D. L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1 :100–115, 1993.
-

- [Don00] D.L. Donoho. High-dimensional data analysis. the curse and blessing of dimensionality. *Lecture delivered at the Conference Math Challenges of the 21st Century*, 2000.
- [Don06] D.L. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52(4) :1289–1306, 2006.
- [DT06] D. L. Donoho and Y. Tsaig. Fast solution of  $\ell_1$  minimization problems when the solution may be sparse. 2006. submitted.
- [DTDS06] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse solution of under-determined linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions On Information Theory*, 2006. submitted.
- [DVDD98] D. L. Donoho, M Vetterli, R. A. DeVore, and I. Daubechies. Data compression and harmonic analysis. *IEEE Trans. Inform. Theory*, 44 :2435–2476, 1998.
- [DY06] D.L. Donoho and Y.Tsaig. Extensions of compressed sensing. *Signal Processing*, 86(3) :5433–548, 2006.
- [EB02] M. Elad and A.M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 48 :2558–2567, 2002.
- [EHJT04] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2) :407–499, 2004.
- [Ela06] M. Elad. Why simple shrinkage is still relevant for redundant representations ? *IEEE Transactions on Information Theory*, 52(12) :5559–5569, 2006.
- [ESDQ05] M. Elad, J.-L Starck, D.L. Donoho, and P. Querre. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *ACHA*, 19(3) :340–358, 2005.
- [EY08] M. Elad and I. Yavneh. A weighted average of sparse representations is better than the sparsest one alone. *IEEE Trans. on Information Theory*, preprint, 2008.
- [EZ04] M. Elad and M. Zibulevsky. A probabilistic study of the average performance of the basis pursuit. *submitted to IEEE Transactions on Information Theory*, 2004.
- [FG06] C. Févotte and S. J. Godsill. A bayesian approach for blind separation of sparse sources. *IEEE Trans. Speech and Audio Processing*, 2006. in press.
- [Fie99] D.J. Field. Wavelets, vision and the statistics of natural scenes. *Phil. Trans. R. Soc. Lond. A*, 357 :2527–2542, 1999.
-



- [FN03a] A. Feuer and A. Nemirovsky. On sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 49(6) :1579–1581, 2003.
- [FN03b] M.A. Figueiredo and R.D. Nowak. An em algorithm fo wavelet-based image restoration. *IEEE Trans. On Image Processing*, 12(8) :906–916, August 2003.
- [FNW07] M.A. Figueiredo, R. Nowak, and S.J. Wright. Gradient projection for sparse reconstruction : Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing - To appear*, 2007.
- [FR06] Massimo Fornasier and Holger Rauhut. Recovery algorithms for vector valued data with joint sparsity constraints. *Preprint - available at <http://www.dsp.ece.rice.edu/cs/>*, 2006.
- [FR07] M. Fornasier and H. Rauhut. Iterative thresholding algorithms. In *Preprint*, 2007.
- [FS05] J. Fadili and J-L. Starck. EM algorithm for sparse representation - based image inpainting. *IEEE International Conference on Image Processing ICIP'05*, 2 :61–63, 2005. Genoa,Italy.
- [FSM06] J. Fadili, J.-L. Starck, and F. Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal - in press*, 2006. in press.
- [Fuc06] J.-J. Fuchs. Recovery conditions of sparse representations in the presence of noise. *ICASSP '06*, 3(3) :337–340, 2006.
- [GN03] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12) :3320–3325, 2003.
- [G.P06] G.Peyré. Best basis compressed sensing. In *SSVM*, 2006. Preprint - available at <http://www.dsp.ece.rice.edu/cs/>.
- [GTC05] P. G. Georgiev, F. Theis, and A. Cichocki. Sparse component analysis and blind source separation of underdetermined mixtures. *IEEE Transactions on Neural Networks*, 16(4) :992–996, 2005.
- [Gul06a] O.G. Guleryuz. Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part I : theory. *IEEE Trans. On Image Processing*, 15(3) :539 – 554, 2006.
- [Gul06b] O.G. Guleryuz. Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part II : adaptive algorithms. *IEEE Trans. On Image Processing*, 15(3) :555 – 571, 2006.
-

- [GV06] R. Gribonval and P. Vandergheynst. On the exponential convergence of matching pursuits in quasi-incoherent dictionaries. *IEEE Trans. Information Theory*, 52(1) :255–261, 2006.
- [H74] J.A. Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 15 :417–426, 1974.
- [Hin06] G Hinshaw. Three-year wilkinson microwave anisotropy probe (wmap) observations : Temperature analysis. *ArXiv Astrophysics e-prints*, *arXiv :astro-ph/0603451*, 2006.
- [HK02] A. Hyvärinen and R. Karthikesh. Imposing sparsity on the mixing matrix in independent component analysis. *Neurocomputing*, 49 :151–162, 2002.
- [HKO01] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.
- [Hoy04] P.O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5 :1457–1469, 2004.
- [HSSW82] C.G.T. Haslam, C.J. Salter, H. Stoffel, and Wilson W.E. A 408 mhz all-sky continuum survey ii - the atlas of contour maps. *Astronomy and Astrophysics Supplement Series*, 47 :1–143, 1982.
- [HTF01] T. Hastie, R. Tibshirani, and J Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2001.
- [Hyv99] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3) :626–634, 1999.
- [IMD06] M.M. Ichir and A. Mohammad-Djafari. Hidden markov models for wavelet-based blind source separation. *IEEE Transactions on Image Processing*, 15(7) :1887–1899, 2006.
- [Jay03] E.T. Jaynes. *Probability Theory, The Logic of Science*. Cambridge University Press, 2003.
- [JH91] C. Jutten and J. Herault. Blind separation of sources, part i : An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24 :1–10, 1991.
- [JMS07] C. Jutten, S. Moussaoui, and F. Schmidt. How to apply ica on actual data ? example of mars hyperspectral image analysis. In *Digital Signal Processing, 2007 15th International Conference on*, pages 3–12, 2007.
-

- [JSD<sup>+</sup>05] J. Jin, J.-L. Starck, D.L. Donoho, N. Aghanim, and O. Forni. Cosmological non-gaussian signatures detection : Comparison of statistical tests. *Eurasip Journal*, 15 :2470–2485, 2005.
- [Jun96] G. Jungman. Cosmological parameter determination with microwave background maps. *Phys. Rev. D*, 54 :1332–1344, 1996.
- [KKB07] Kwangmoo Koh, Seung-Jean Kim, and Stephen Boyd. Solver for l1-regularized least squares problems. Technical report, Stanford University, 2007.
- [Knu99] K.H. Knuth. A bayesian approach to source separation. In *ICA99 - Aussois, France*, 1999.
- [LAM97] A. Lannes, E. Anterrieu, and P. Marechal. Clean and Wipe. *AAS*, 123 :183–198, May 1997.
- [LGBS98] Te-Won Lee, Mark Girolami, Anthony J. Bell, and Terrence J. Sejnowski. A unifying information-theoretic framework for independent component analysis, 1998.
- [LM05] E. LePennec and S. Mallat. Sparse geometric image representations with bandelets. *IEEE Transactions on Image Processing*, 14(4) :423–438, 2005.
- [LR05] M. Lachièze-Rey. *Initiation à la cosmologie*. Dunod, 4ème edition, 2005.
- [Mal98] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [McG97] O. Moulines, J.F. cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *ICASSP 97*, 1997.
- [MCSD05] Y. Moudden, J.-F. Cardoso, J.-L. Starck, and J. Delabrouille. Blind component separation in wavelet space : Application to CMB analysis. *Eurasip Journal*, 15 :2437–2454, 2005.
- [MCW05] D. M. Malioutov, M. Cetin, and A. S. Willsky. Homotopy continuation for sparse signal representation. In *ICASSP '05*, volume 5, pages 733–736, 2005.
- [MES07] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *ITIP*, 2007. submitted.
- [Mey03] Y. Meyer. Images et vision. *Personal Communication*, 2003.
- [MHS<sup>+</sup>08] S. Moussaoui, H. Hauksdottir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J.A. Benediktsson. On the decomposition of mars hyperspectral data by ica and bayesian positive source separation. *Neurocomputing*, in press, 2008.
-

- [MW08] R. Marcia and R. Willett. Compressive coded aperture video reconstruction. In *European Signal Processing Conference — EUSIPCO 2008*, 2008.
- [MZ93] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12) :3397–3415, 1993.
- [MZ07] F. Malgouyres and T. Zeng. A proximal point algorithm for a non negative basis pursuit denoising model. In *Preprint*, 2007.
- [Nat95] B.K Natarayan. Sparse approximate solutions to linear systems. *SIAM Journal on Scientific Computing*, 24 :227–234, 1995.
- [NCB04] R. Neelamani, H. Choi, and R. G. Baraniuk. ForWard : fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Trans. on Signal Processing*, 52(2) :418–433, 2004.
- [OF04] B.A. Olshausen and D.J. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14 :481–487, 2004.
- [OF06] B.A. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set : A strategy employed by v1 ? *Vision Research.*, 37 :3311–3325, 2006.
- [OPT00] M. R. Osborne, B. Presnell, and B. A. Turlach. A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, 20(3) :389–403, 2000.
- [Pey07] Gabriel Peyré. Texture synthesis and modification with a patch-valued wavelet transform. In *SSVM 07*, 2007.
- [PGJ92] D.-T. Pham, P. Garrat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In *Proc. EUSIPCO*, pages 771–774, 1992.
- [Plu07] M.D. Plumbley. On polar polytopes and recovery of sparse representations. *IEEE Trans. on Information Theory*, 53(9) :3188–3195, September 2007.
- [PRK93] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad. Orthogonal matching pursuit : recursive function approximation with applications to wavelet decomposition. In *27th Asilomar Conf. on Signals, Systems and Comput.*, 1993.
- [PWB<sup>+</sup>06] A. Poglitscha, C. Waelkensb, O.H. Bauera, J. Cepac, H. Feuchtgrubera, T. Henning, C. van Hoofe, F. Kerschbaumf, D. Lemked, E. Renotteg, L. Rodriguez, P. Saraceno, and B. Vandenbussche. The photodetector array camera and spectrometer (pacs) for the herschel space observatory. In *SPIE*, 2006.
-

- [RM05] R.Gribonval and M.Nielsen. Beyond sparsity : recovering structured representations by  $\ell_1$ -minimization and greedy algorithms. – application to the analysis of sparse underdetermined ica –. *to appear in J. Advances in Computational Mathematics*, 2005.
- [Rob01] C. Robert. *The Bayesian Choice*. Springer Texts In Statistics. Springer, 2001.
- [Roc70] Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1970.
- [SAF04] J.-L. Starck, N. Aghanim, and O. Forni. Detecting cosmological non-gaussian signatures by multi-scale methods. *Astronomy and Astrophysics*, 416 :9–17, 2004.
- [Sai04] N. Saito. The generalized spike process, sparsity, and statistical independence. in *Modern Signal Processing (D. Rockmore and D. Healy, Jr. eds.)*, MSRI Publications, 46 :317–340, 2004.
- [SAP00] S.Sardy, A.Bruce, and P.Tseng. Block coordinate relaxation methods for non-parametric wavelet denoising. *Journal of Computational and Graphical Statistics*, 9(2) :361–379, 2000.
- [SB03] N. Saito and B. Benichou. Sparsity vs. statistical independence in adaptive signal representations : A case study of the spike process. in *Beyond Wavelets (G. V. Welland ed.)*, Studies in Computational Mathematics, 10/9 :225–257, 2003.
- [SCD02] J.-L. Starck, E.J. Candès, and D.L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6) :670–684, 2002.
- [Sch78] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6 :461–464, 1978.
- [SED04] J.-L. Starck, M. Elad, and D.L. Donoho. Redundant multiscale transforms and their application for morphological component analysis. *Advances in Imaging and Electron Physics*, 132 :287–348, 2004.
- [SED05] J.-L. Starck, M. Elad, and D.L. Donoho. Image decomposition via the combination of sparse representation and a variational approach. *IEEE Transactions on Image Processing*, 14(10) :1570–1582, 2005.
- [SFM07] J.-L. Starck, J. Fadili, and F. Murtagh. The undecimated wavelet decomposition and its reconstruction. *IEEE Transactions on Image Processing*, 16 :297–309, 2007.
- [SH04] T. Strohmer and R.W Heath. Grassmanian frames with applications to coding and communication. *Applied and Computational Harmonic Analysis*, 14 :131–149, 2004.
-

- [SJFC<sup>+</sup>08] S.M.Leach, J.-F.Cardoso, C.Baccigalupi, R.B.Barreiro, M.Betoule, J.Bobin, A.Bonaldi, G.de Zotti, J.Delabrouille, C.Dickinson, H.K.Eriksen, J.González-Nuevo, F.K.Hansen, D.Herranz, M.LeJeune, M.López-Caniego, E.Martinez-González, M.Massardi, J.-B.Melin, M.-A.Miville-Deschênes, G.Patanchon, S.Prunet, S.Ricciardi, E.Salerno, J.L.Sanz, J.-L.Starck, F.Stivoli, V.Stolyarov, R.Stompor, and P.Vielva. Title : Component separation methods for the planck mission. *Astronomy and Astrophysics*, 2008.
  - [Ski02] G.K. ; Skinner. Coded-mask imaging in gamma-ray astronomy - separating the real and imaginary parts of a complex subject. In *Proceedings of the 22nd Moriond Astrophysics Meeting "The Gamma-Ray Universe"*, 2002.
  - [SM06] J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Astronomical image and data analysis, by J.-L. Starck and F. Murtagh. Astronomy and astrophysics library. Berlin : Springer, 2006, 2006.
  - [SMB06] J.-L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis : The Multiscale Approach*. Cambridge University Press, 2006.
  - [SMPA96] J.-L. Starck, F. Murtagh, B. Pirenne, and M. Albrecht. Astronomical image compression based on noise suppression. *Publications of the Astronomical Society of the Pacific*, 108 :446–455, 1996.
  - [SNM03] J.-L. Starck, M.K. Nguyen, and F. Murtagh. Wavelets and curvelets for image deconvolution : a combined approach. *Signal Processing*, 83(10) :2279–2283, 2003.
  - [SO01] E.P. Simoncelli and B.A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24 :1193–1216, 2001.
  - [SO03] P. Sallee and B. Olshausen. Learning sparse multiscale image representations. In *Advances Neural Information Processing Systems*, 2003.
  - [SQ01] J.-L. Starck and P. Querre. Multispectral data restoration by the wavelet-karhunen-loève transform. *Signal Processing*, 81 :2449–2459, 2001.
  - [Sto77] C. J. Stone. Consistent nonparametric regression, with discussion. *Ann. Statist.*, 5 :595–645, 1977.
  - [TGS06] J.A. Tropp, A.C. Gilbert, and M.J Strauss. Algorithms for simultaneous sparse approximation. part i : greedy pursuit. *Signal Processing - special issue "Sparse approximations in signal and image processing"*, 86 :572–588, 2006.
  - [Tib96] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B.*, 58(1) :267–288, 1996.
-

- [TLW<sup>+</sup>06] Dharmpal Takhar, Jason Laska, Michael Wakin, Marco Duarte, Dror Baron, Shriram Sarvotham, Kevin Kelly, and Richard Baraniuk. A new compressive imaging camera architecture using optical-domain compression. *Proc. of Computational Imaging IV at SPIE Electronic Imaging, San Jose, California*, 2006.
  - [Tro04] T.A. Tropp. Greed is good : algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10) :2231–2242, 2004.
  - [Tro06a] J.A. Tropp. Algorithms for simultaneous sparse approximation. part ii : convex relaxation. *Signal Processing - special issue "Sparse approximations in signal and image processing"*, 86 :589–602, 2006.
  - [Tro06b] J.A. Tropp. Just relax : Convex programming methods for subset selection and sparse approximation. *IEEE Transactions on Information Theory*, 52(3) :1030–1051, 2006.
  - [Tse01] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimizations. *J. of Optim. Theory and Appl.*, 109(3) :457–494, 2001.
  - [Vin07] Emmanuel Vincent. Complex nonconvex  $\ell_p$  norm minimization for underdetermined source separation. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D Plumbley, editors, *Independent Component Analysis and Signal Separation*, volume 4666 of *LNCS*, pages 430–437. Springer, 2007.
  - [Wav05] Wavelab 850 for Matlab7.x. Wavelab. (<http://www-stat.stanford.edu/~wavelab/>), 2005.
  - [WPL92] R. White, M. Postman, and M. Lattanzi. Compression of the Guide Star digitised Schmidt plates. In H.T. MacGillivray and E.B. Thompson, editors, *Digitized Optical Sky Surveys*, pages 167–175. Kluwer, 1992.
  - [WS88] B.P. Wakker and U.J. Schwarz. The multi-resolution Clean and its application to the short-spacing problem in interferometry. *Annual Reviews of Astronomy and Astrophysics*, 200 :312, 1988.
  - [YW08] A.P.S Yadav and B.J Wandelt. Detection of primordial non-gaussianity (fnl) in the wmap 3-year data at above 99.5 *Physical Review Letters - revised*, 2008.
  - [ZCA99] L. Zhang, A. Cichocki, and S. Amari. Natural gradient algorithm for blind separation of overdetermined mixture with additive noise. *IEEE Signal Processing Letters*, 6(11) :293–295, 1999.
-

- [Zib03] M. Zibulevski. Blind source separation with relative newton method. *Proceedings ICA2003*, pages 897–902, 2003.
- [ZK06] P. Tichavsky Z. Koldovsky. Methods of fair comparison of performance of linear ica techniques in presence of additive noise. Number 5, 2006.
- [ZKO06] P. Tichavsky Z. Koldovsky and E. Oja. Efficient variant of algorithm fastica for independent component analysis attaining the cramer-rao lower bound. *IEEE Transactions on neural networks*, 17 :1265–1277, 2006.
- [ZP01] M. Zibulevsky and B.B. Pearlmutter. Blind source separation by sparse decomposition. *Neural Computations*, 13/4, 2001.
-