



Une exploration des problèmes inverses par les représentations parcimonieuses et l'optimisation non lisse

Jalal M. Fadili

► To cite this version:

Jalal M. Fadili. Une exploration des problèmes inverses par les représentations parcimonieuses et l'optimisation non lisse. Traitement du signal et de l'image [eess.SP]. Université de Caen, 2010. tel-01071774

HAL Id: tel-01071774

<https://theses.hal.science/tel-01071774>

Submitted on 6 Oct 2014

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

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

Université de Caen Basse-Normandie
UFR Sciences
École doctorale SIMEM

MÉMOIRE

pour l'obtention de l'

HABILITATION À DIRIGER DES RECHERCHES

Spécialité: Traitement du Signal et des Images

Une exploration des problèmes inverses par les représentations parcimonieuses et l'optimisation non lisse

par

JALAL M. FADILI

Soutenue le 26 Mars 2010

Jury

Yves Meyer	Professeur, ENS Cachan	Président
Laure Blanc-Féraud	Directeur de Recherche, CNRS UNSA	Rapporteur
Mila Nikolova	Directeur de Recherche, CNRS ENS Cachan	Rapporteur
Pierre Vandergheynst	Professeur, EPFL	Rapporteur
Antonin Chambolle	Directeur de Recherche, CNRS Ecole Polytechnique	Examineur
Stéphane Mallat	Professeur, Ecole Polytechnique	Examineur
Jean-Luc Starck	Directeur de Recherche, CEA	Examineur
Marinette Revenu	Professeur, ENSICAEN	Garant

Table des matières

Introduction	1
1 Modélisation stochastique	1
1.1 Modèle de dégradation	1
1.2 Modèles d’images	1
2 Du choix de la transformée à l’a priori de parcimonie	2
3 Optimiser pour estimer	2
4 Organisation du document	3
Notations	5
Glossaire	7
Chapitre 1 Parcimonie et représentation des signaux et des images	9
1.1 Le monde de la parcimonie	9
1.1.1 Terminologie	9
1.1.2 Signaux strictement parcimonieux	10
1.1.3 Signaux compressibles et approximation non linéaire	11
1.1.4 Analyse harmonique computationnelle	11
1.2 Redondance et diversité morphologique	12
1.2.1 Apport de la redondance	12
1.2.2 La diversité morphologique	13
1.2.3 Choix du dictionnaire	13
1.3 Décomposition et approximation parcimonieuses dans des dictionnaires redondants	13
1.4 Récapitulatif et conclusion	15
Chapitre 2 Modélisation statistique des images	17
2.1 Lois leptokurtiques à queues lourdes	17
2.1.1 Statistiques marginales	17
2.1.2 Statistiques jointes	20
2.2 Invariance d’échelle et conséquences	21
2.2.1 Conséquence 1 : spectre en loi de puissance	22
2.2.2 Conséquence 2 : persistance à travers les échelles	22
2.2.3 Conséquence 3 : les images comme fonctions ou distributions ?	23

2.3	Epilogue	24
2.3.1	Modèle génératif sans occlusion	24
2.3.2	Atomes élémentaires	25
2.3.3	Et l'occlusion ?	25
2.4	Récapitulatif et conclusion	25
Chapitre 3 Optimisation convexe non lisse		27
3.1	Quelques définitions	27
3.2	Opérateur et calcul proximal	28
3.3	Décomposition d'opérateurs monotones maximaux	28
3.3.1	Origines de la décomposition	29
3.3.2	Schéma de décomposition implicite-implicite	29
3.3.3	Schéma de décomposition explicite-implicite	29
3.3.4	Schéma de décomposition de Douglas-Rachford	30
3.4	Premières applications	31
3.4.1	Opérateur proximal de $F \circ \mathbf{A}$	31
3.4.2	Opérateur proximal de $F_1 + F_2 \circ \mathbf{A}$	32
3.4.3	Opérateur proximal de pénalités favorisant la parcimonie	33
3.5	Récapitulatif et conclusion	34
Chapitre 4 Parcimonie et estimation		35
4.1	Régression non paramétrique, débruitage	35
4.1.1	Bruit additif Gaussien	35
4.1.2	Bruit de Poisson	37
4.1.3	Bruit mixte Poisson-Gaussien	41
4.1.4	Bruit multiplicatif	43
4.2	Estimation dans les séries temporelles	45
4.2.1	Régression semi-paramétrique	45
4.2.2	Régression paramétrique avec bruit à longue mémoire	46
4.3	Récapitulatif et conclusion	47
Chapitre 5 Parcimonie et problèmes inverses linéaires		49
5.1	Exemples de problèmes inverses	49
5.2	Problèmes inverses linéaires avec bruit additif Gaussien	50
5.2.1	Modèle de dégradation	50
5.2.2	Estimateur MAP avec a priori de parcimonie	50
5.2.3	Application à la déconvolution et l'inpainting	52
5.3	Problèmes inverses linéaires avec bruit de Poisson	52
5.3.1	Modèle de dégradation	52
5.3.2	Estimateur MAP avec parcimonie sous contraintes	52
5.4	Récapitulatif et conclusion	55

Chapitre 6 Décomposition parcimonieuse et séparation de sources	57
6.1 Décomposition parcimonieuse d'image	58
6.1.1 L'analyse en composantes morphologiques	58
6.1.2 Choix du dictionnaire : la clé pour une bonne séparation	59
6.1.3 Stratégies de variation du seuil	59
6.1.4 Identifiabilité du support et convergence	59
6.2 Séparation de sources parcimonieuses	61
6.2.1 Le modèle de mélange linéaire instantané	61
6.2.2 Décomposition parcimonieuse multispectrale	61
6.2.3 Séparation aveugle de sources parcimonieuses	62
6.3 Récapitulatif et conclusion	63
Chapitre 7 Echantillonnage compressé	67
7.1 Rappels essentiels	67
7.1.1 Incohérence et parcimonie	67
7.1.2 Paradigme d'échantillonnage	68
7.1.3 Propriété d'isométrie restreinte	68
7.1.4 Construire de bonnes matrices : Echantillonnage aléatoire	69
7.2 Algorithmes rapides pour les décodeurs convexes en échantillonnage compressé	69
7.3 Echantillonnage compressé et bruit de quantification	70
7.3.1 Modèle de mesure	70
7.3.2 Le décodeur BPDQ _p et faisabilité	70
7.3.3 Isométrie restreinte étendue	71
7.3.4 Consistance et réduction du bruit de quantification	71
7.3.5 Algorithme de minimisation	72
7.4 Exploration numérique des performances de l'échantillonnage compressé	72
7.4.1 Retour sur les critères topologiques d'identifiabilité ℓ_1	73
7.4.2 Faces intérieures et vecteurs non identifiables	73
7.4.3 Algorithme de poursuite glouton guidé par les faces	74
7.5 Récapitulatif et conclusion	75
Chapitre 8 Parcimonie et segmentation des textures	77
8.1 Contours actifs basés régions	77
8.1.1 Fonctionnelle à minimiser	77
8.1.2 Equation d'évolution et dérivées de forme	78
8.2 Descripteurs de textures	79
8.3 Segmentation des textures basée PDF	79
8.3.1 Segmentation supervisée	79
8.3.2 Segmentation non supervisée	81
8.4 Récapitulatif et conclusion	82

Chapitre 9 Applications et recherche reproductible	85
9.1 Applications	85
9.1.1 Imagerie médicale et biomédicale	85
9.1.2 Imagerie astronomique	89
9.1.3 Contrôle non destructif	94
9.1.4 Photographie computationnelle	97
9.2 La recherche reproductible	97
9.3 Récapitulatif et conclusion	98
Chapitre 10 Conclusion et perspectives	99
10.1 La modélisation des signaux et des images	99
10.2 La théorie de l'estimation	99
10.3 L'optimisation non lisse	100
10.4 Les problèmes inverses avec parcimonie	100
10.5 Analyse multiéchelle de fonctions non scalaires	101
10.6 Applications	101
Bibliographie	103
Annexes	113
Annexe A Curriculum Vitae	113
Liste des thèses encadrées	121
Liste de publications	123

Remerciements

Je tiens tout d'abord à adresser mes remerciements à tous les membres du jury qui m'ont fait l'honneur d'en faire partie. Je remercie Yves Meyer d'avoir accepté de présider ce jury, pour son enthousiasme et sa passion. Je remercie les rapporteurs pour leur lecture attentive et avisée de ce manuscrit.

Je remercie Marinette Revenu qui a accepté d'être garant de ce travail, pour son support le long de ces années. Je profite de ces quelques lignes pour remercier Daniel Bloyet, mon directeur de thèse maintenant à la retraite, de m'avoir initié au monde de la recherche et pour ce qu'il m'a transmis de rigueur et de passion.

La recherche est avant tout une aventure humaine faite de rencontres et souvent d'amitiés. J'ai la chance de la partager avec de nombreux chercheurs : Ed Bullmore, Jean-Luc Starck, Gabriel Peyré, Charles Dossal, Fionn Murtagh, Jean-François Aujol, Mila Nikolova, Sylvain Durand, Christophe Chesneau, David Tschumperlé, Stéphanie Jehan-Besson, Dimitri Vandeville, Pierre Vandergheynst, Laurent Jacques, David Hammond, Michael Elad, David Donoho, Bo Zhang, Seth Digel. Que toutes ces personnes reçoivent ici l'expression de ma profonde gratitude et ma haute considération. Je remercie aussi mes étudiants de thèse que j'ai eu l'opportunité d'encadrer. De nombreuses contributions dans ce manuscrit sont le fruit de ces collaborations et ces encadrements.

J'adresse par ailleurs mes remerciements aux membres du GREYC, permanents et doctorants, et l'équipe Image en particulier pour toutes années passées auprès d'eux, pour l'excellente ambiance et la bonne humeur. J'ai beaucoup appris et continue toujours à apprendre à leur contact.

Finalement, je ne pourrais terminer ces remerciements sans m'adresser à ma famille en les remerciant pour la confiance qu'ils m'accordent, leur patience et leur soutien indéfectible. Qu'ils voient en ces quelques lignes l'expression de ma profonde affection et ma reconnaissance.

Introduction

Ce mémoire résume mon parcours de recherche lors des dix dernières années. Ces travaux de recherche se trouvent à la croisée des chemins entre les mathématiques appliquées et le traitement du signal et des images. Ils s'articulent autour du triptyque : (i) modélisation stochastique-estimation statistique ; (ii) analyse harmonique computationnelle-représentations parcimonieuses ; (iii) optimisation. Ces trois piliers constituent le socle théorique de mes activités pour développer des approches originales capables de résoudre des problèmes classiques en traitement d'images comme les problèmes inverses en restauration et reconstruction, la séparation de sources, la segmentation, la détection, etc.

1 Modélisation stochastique

En traitement d'images, l'une des hypothèses les plus abondamment adoptées est celle où le signal d'intérêt est supposé a priori Gaussien, et où le bruit est additif blanc Gaussien indépendant du signal utile. Ce modèle est toutefois très simplificateur des situations réelles et ne reflète ni la complexité des propriétés statistiques des images naturelles, ni celle des dispositifs d'acquisition rencontrés en pratique. Il apparaît donc crucial d'établir à la fois des modèles adaptés de formation de l'image (c.à.d. types de dégradation et de bruit), mais aussi de concevoir des a priori plus subtiles pour capturer des propriétés statistiques complexes des images que l'on veut recouvrer.

1.1 Modèle de dégradation

Etablir une équation de dégradation nécessite de modéliser le problème direct pour en déduire une forme analytique la plus fidèle possible aux conditions physiques d'acquisition. C'est ici que la connaissance de la chaîne instrumentale d'acquisition revêt toute son importance. En effet, un signal ou une image observée en sortie d'un système d'acquisition n'est pas parfaite. Elle est sujette à des dégradations introduites par les lois de la physique d'acquisition et dont les origines sont multiples. On peut alors distinguer des causes déterministes engendrant une perte d'information en raison d'une transformation déterministe (pas nécessairement inversible). Les autres causes sont elles d'origine stochastique puisque les observations sont des mesures soumises à des fluctuations aléatoires non mesurables que l'on regroupe sous la dénomination de bruit. Ainsi, pour chacune des modalités d'imagerie à laquelle je me suis intéressé, la démarche systématique adoptée a été de se pencher sur la physique d'acquisition (dégradations, bruit, etc.) pour établir un modèle (direct) pertinent de formation de l'image observée.

1.2 Modèles d'images

L'inférence statistique pour la compréhension et la modélisation des images est largement utilisée depuis plus d'une trentaine d'années et a été initiée notamment par les travaux de Grenander [107] et Cooper [58].

Pour appliquer ces techniques, il est nécessaire d'avoir un modèle probabiliste d'une classe d'images ou d'une classe de structures dans les images. Plusieurs modèles stochastiques de ce type ont été introduits dans la littérature, comme les modèles des textures [104, 210], des contours [161, 108], des régions [24, 105, 163], des objets dans l'image, etc. La démarche commune de ces travaux est de décrire une classe d'images x par un ensemble de variables auxiliaires w décrivant la scène à l'origine de l'image, e.g. contours,

textures, objets, caractéristiques d'illumination de la scène, etc. L'inférence Bayésienne permet pour un x fixé, d'estimer ce que devrait être w en minimisant un risque moyenné par rapport à la loi $p(w|x)$. Cette dernière étant déduite par la règle de Bayes $p(w|x) \propto p(x|w)p(w)$. De fait, deux modèles stochastiques sont requis : d'une part un modèle avec une loi a priori $p(w)$ sur les variables "cachées" w spécifiant le type de scènes vraisemblables pour la classe d'images en question, et d'autre part un modèle d'imagerie avec une loi conditionnelle $p(x|w)$ traduisant ce que l'image observée devrait être sachant la scène.

Ce cadre est assez général, mais pour l'exploiter, il nécessite :

- (i) une théorie de modèles stochastiques suffisamment complète pour permettre d'exprimer toute la complexité et la variabilité des structures dans w et par conséquent dans x ;
- (ii) une méthode pour apprendre à partir de bases la multitude d'hyperparamètres mis en jeu dans une telle théorie ;
- (iii) une méthode pour inférer w sachant x .

L'approche la plus directe pour répondre au point (i) promue par D. Mumford et ses collaborateurs est de tenter de trouver de bons modèles de x en s'affranchissant des variables auxiliaires "cachées" w . Ceci signifie qu'il faut créer des bases d'images x qui sont des réalisations aléatoires représentatives de toutes les images appartenant à la classe en question. On peut alors étudier les statistiques marginales et jointes des réponses à des bancs de filtres linéaires (ou non linéaires¹). On gardera à l'esprit que caractériser la densité jointe de x ou les densités marginales de toutes ses combinaisons linéaires (c.à.d. versions filtrées linéairement) revient au même par le théorème de Cramér-Wold.

La propriété statistique frappante qui a été observée tout d'abord par D. Field dans [93] et ensuite confirmée par plusieurs auteurs [120, 110, 109, 139] est que la statistique marginale de la réponse à des bancs de filtres (typiquement passe-bande ou passe-haut) est leptokurtique avec des queues lourdes. C'est typiquement le cas pour les coefficients de détail de la transformée d'ondelettes des images régulières, et plus généralement pour toute transformée associée à un banc de filtres fournissant une description creuse ou parcimonieuse de l'image x . C'est ce constat qui m'a conduit à m'intéresser à la parcimonie marginale (terme-à-terme) et structurée (e.g. jointe par blocs) comme modèle d'a priori. Ceci sera revu avec plus de détails aux Chapitres 1 et 2.

2 Du choix de la transformée à l'a priori de parcimonie

Un signal est le support physique d'une information. La facilité de l'extraction de cette information se mesure à sa lisibilité : quitter l'espace original des données brutes pour passer à un espace transformé (au-delà du "Royaume de Fourier") contenant la même information pour mettre en évidence des caractéristiques spécifiques du signal, et de le scruter sous un autre angle. Le choix d'une transformée est crucial et sa pertinence se mesure à la capacité qu'elle a à fournir des descripteurs se prêtant bien à des tâches ultérieures. En particulier, il est désirable que l'essentiel de l'énergie du signal soit concentré sur très peu de coefficients dans le domaine de la transformée. En d'autres termes, le signal peut être expliqué par peu de degrés de liberté. Un signal possédant cette propriété est alors dit parcimonieux dans le domaine de la transformée en question. Comme il a été souligné dans la section ci-avant, la parcimonie est une source précieuse d'a priori sur les signaux et les images. On verra au Chapitre 2 qu'il existe des liens étroits entre cet a priori et une classe de modèles stochastiques.

3 Optimiser pour estimer

L'inférence Bayésienne consiste à estimer l'image inaccessible et non observée x à partir de sa version dégradée observée y en exploitant la loi a priori sur x et la loi conditionnelle de y sachant x déduite généralement de l'équation de dégradation. Ceci est accompli en minimisant un risque moyen au sens de la loi conditionnelle a posteriori de $x|y$. Typiquement, le risque quadratique correspond à l'estimateur de l'espérance conditionnelle a posteriori et nécessite une étape d'intégration. Quand au risque 0-1, il correspond à l'estimateur du maximum a posteriori (MAP), et nécessite de résoudre un problème

1. Les filtres non linéaires sont entendus au sens large, e.g. filtres d'ordre, analyse(s) topologique(s).

d'optimisation. De façon plus générale, bon nombre de problèmes en traitement d'images se ramènent à la formulation d'une énergie que l'on doit minimiser, cette énergie possédant un terme d'attache aux données et un ou plusieurs termes de régularisation². Dans l'essentiel des situations traitées dans ce document, ces termes de régularisation (voire parfois l'attache aux données) sont non différentiables. Ceci nous a conduit naturellement à nous intéresser à la théorie de l'optimisation non lisse. J'insisterai dans ce document sur le cas convexe même si nous nous consacrons de plus en plus avec mes collaborateurs à traiter le cas non convexe.

4 Organisation du document

Le plan de ce mémoire est le suivant :

- Les trois premiers chapitres tiennent lieu de mise en place des outils utilisés tout au long du document. En particulier, dans le Chapitre 1, je détaillerai les principaux concepts issus des représentations parcimonieuses. Le Chapitre 2 décrira dans un premier temps un modèle aléatoire d'images satisfaisant certains axiomes, et le lien entre l'a priori de parcimonie et ce modèle aléatoire sera établi. Les éléments essentiels issus de l'analyse convexe, le calcul proximal, et de la décomposition d'opérateurs monotones maximaux seront décrits dans le Chapitre 3. Je décrirai aussi certaines de nos contributions pour le calcul de certains opérateurs proximaux en s'appuyant sur la dualité de Fenchel-Rockafellar.
- Dans le Chapitre 4, on s'intéressera aux différentes approches que nous avons proposées pour exploiter la parcimonie à des fins d'estimation notamment pour le problème de la régression en présence de différents types de bruits aussi bien dans le cas non paramétrique (débruitage) que semi-paramétrique et paramétrique.
- L'extension aux problèmes inverses est traitée au Chapitre 5. Plusieurs problèmes inverses y sont formalisés dans un cadre commun et résolus en invoquant les outils développés au Chapitre 3 avec des garanties de convergence.
- Le problème de la décomposition parcimonieuse des signaux et des images ainsi que de la séparation aveugle de sources parcimonieuses fait l'objet du Chapitre 6. On mettra en exergue l'application à la décomposition d'image selon le modèle cartoon+texture ainsi qu'à l'analyse et la séparation de sources pour les images multispectrales.
- Dans le Chapitre 7, j'explicitai nos travaux menés pour mieux comprendre les performances du décodage par minimisation ℓ_1 en échantillonnage compressé sans bruit. Je décrirai aussi les contributions pour la restauration de signaux parcimonieux en échantillonnage compressé à partir de mesures quantifiées.
- La parcimonie a été aussi exploitée dans nos travaux pour la segmentation des textures en l'associant aux contours actifs basés régions. Ces contributions sont décrites au Chapitre 8.
- Je présenterai dans le Chapitre 9 l'ensemble de nos contributions dans plusieurs modalités d'imagerie auxquelles nous avons appliqué nos différents algorithmes d'estimation, de restauration, de séparation ou de segmentation. Au coeur de ma démarche scientifique a toujours été le souci de mettre à disposition de la communauté non seulement la production scientifique sous forme de publications, mais aussi des boîtes à outils et des composants logiciels permettant de reproduire les expériences qui y sont décrites. Cette démarche de recherche reproductible occupe de plus en plus le devant de la scène dans la communauté signal-image. La seconde partie du Chapitre 9 résume quelques réflexions et contributions que nous avons eues à ce sujet.

La présentation de ce travail renvoie à des articles publiés (cités avec le préfixe A) ou en cours (cités avec le préfixe S), des thèses (co-)encadrées soutenues (citées avec le préfixe TH), ainsi qu'à un livre à paraître [A1] qui reprend en détail une partie des travaux décrits dans ce document. La liste détaillée de ces publications est donnée en Annexe A.

2. Il existe bien évidemment une interprétation Bayésienne de ces formulations variationnelles.

Notations

Fonctions and signaux

$x(t)$	Signal à variable continue, $t \in \mathbb{R}$.
$x(\mathbf{t})$ ou $x(t_1, \dots, t_d)$	Signal d -dimensionnel sur un domaine continu, $\mathbf{t} \in \mathbb{R}^d$.
$x[k]$	Signal 1D discret, $k \in \mathbb{Z}$, où k -ème composante d'un vecteur en dimension finie.
$x[\mathbf{k}]$ ou $x[k_1, \dots, k_d]$	Signal d -dimensionnel discret, $\mathbf{k} \in \mathbb{Z}^d$.
$\hat{x}$	Transformée de Fourier de x .
$L_2(\Omega)$	Espace des fonctions carrées-intégrables sur un domaine continu Ω .
$\ell_2(\Omega)$	Espace des signaux carrés-sommables sur un domaine discret Ω .
$\Gamma_0(\mathcal{H})$	Classe des fonctions définies de $\mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ convexes semi-continues inférieurement et à domaine non vide.
$\iota_{\mathcal{C}}$	Indicatrice d'un convexe fermé $\mathcal{C}$.
F^*	Transformée (conjuguée) de Legendre-Fenchel de $F \in \Gamma_0(\mathcal{H})$.
${}^\rho F$	Enveloppe de Moreau d'ordre ρ de $F \in \Gamma_0(\mathcal{H})$.
$\text{dom}(F)$	Domaine de F .
$\text{int}(\mathcal{C})$	L'intérieur d'un convexe.
$\text{ri}(\mathcal{C})$	L'intérieur relatif d'un convexe.

Opérateurs sur signaux et fonctions

$*$	Produit de convolution en continu.
$\circledast$	Convolution discrète circulaire.
$\odot$	Composition de deux arguments.
$+$	
$\vee$	Inf-convolution de deux fonctions.

Matrices, opérateurs linéaires et normes

Symboles en gras et majuscules	Matrices ou opérateurs linéaires, e.g. $\mathbf{M}$.
$\cdot^T$	Transposé d'un vecteur ou d'une matrice.
$\mathbf{M}^*$	Adjoint de $\mathbf{M}$.
Matrice de Gram de $\mathbf{M}$	$\mathbf{M}^* \mathbf{M}$ (sur $\mathbb{C}$) où $\mathbf{M}^T \mathbf{M}$ (sur $\mathbb{R}$).
$\mathbf{M}[i, j]$	Composante (i, j) d'une matrice $\mathbf{M}$.
$\det(\mathbf{M})$	Déterminant d'une matrice $\mathbf{M}$.
$\text{rang}(\mathbf{M})$	Rang d'une matrice $\mathbf{M}$.
$\text{diag}(\mathbf{M})$	Matrice diagonale composée de la diagonale principale de $\mathbf{M}$.
$\text{trace}(\mathbf{M})$	Trace d'une matrice $\mathbf{M}$.
$\text{Im}(\mathbf{M})$	Image d'un opérateur linéaire $\mathbf{M}$.
$\text{Ker}(\mathbf{M})$	Noyau d'un opérateur linéaire $\mathbf{M}$.

$\text{vect}(\mathbf{M})$

$\mathbf{M}^+$

$\mathbf{I}$

$\langle \cdot, \cdot \rangle$

$\|\cdot\|$

$\|\cdot\|_p$

$\mathcal{B}_p^\rho$

$\|\cdot\|_0$

$\|\cdot\|_{w\ell_p}$

$\|\cdot\|_{\text{TV}}$

$\overline{\nabla}$

div

$\|\cdot\|$

$\|\cdot\|_{\text{F}}$

$\otimes$

Variables et vecteurs aléatoires

$\varepsilon \sim \mathcal{N}(\mu, \Sigma)$

$\varepsilon \sim \mathcal{N}(\mu, \sigma^2)$

$\varepsilon \sim \mathcal{P}(\lambda)$

$\varepsilon \sim \mathcal{U}([a, b])$

$\mathbb{E}[\cdot]$

$\text{Var}[\cdot]$

$\stackrel{d}{=}$

$\xrightarrow{d}$

Concatène les colonnes d'une matrice $\mathbf{M}$ dans un vecteur colonne 1D.

Pseudo-inverse de $\mathbf{M}$.

Opérateur ou matrice identité de dimension appropriée. $\mathbf{I}_n$ si la dimension n'est pas claire dans le contexte.

Produit scalaire.

Norme associée à $\langle \cdot, \cdot \rangle$.

$p \geq 1$, norme ℓ_p d'un signal avec l'adaptation usuelle pour $p = \infty$.

$p \geq 1$, boule ℓ_p fermée de rayon $\rho > 0$.

Quasi-norme ℓ_0 d'un signal; nombre de ses composantes non nulles.

Norme ℓ_p -faible.

Semi-norme de la variation totale discrète.

Gradient discret d'une image.

Opérateur divergence discret (adjoint de $\overline{\nabla}$).

Norme spectrale d'un opérateur linéaire.

Norme de Frobenius d'une matrice.

Produit de Kronecker.

Le vecteur ε suit une loi normale de moyenne μ et matrice de covariance Σ .

Les composantes de ε sont iid Gaussiennes de même variance σ^2 , c.à.d. ε est Gaussien blanc de moyenne μ et variance σ^2 .

Le vecteur ε suit une loi de Poisson d'intensité λ .

Le vecteur ε suit une loi uniforme dans $[a, b]$.

Opérateur d'espérance par rapport à une mesure de probabilité.

Opérateur de variance par rapport à une mesure de probabilité.

Egalité en distribution.

Limite en distribution.

Glossaire

AGWN	Bruit additif Gaussien blanc (Additive white Gaussian Noise)
AMGGD	Distribution GGD multivariée anisotrope (Anisotropic Multivariate GGD)
AIC	Akaike Information Criterion
ARFIMA	Processus auto-régressif à moyenne mobile et intégration fractionnaire (Auto-Regressive Fractionally Integrated Moving Average)
BCR	Relaxation par blocs (Block-Coordinate Relaxation)
BIC	Bayesian Information Criterion
BKF	Formes K de Bessel (Bessel K Form)
BP	Poursuite de base (Basis Pursuit)
BPDN	Basis Pursuit DeNoising
BSS	Séparation aveugle de sources (Blind Source Separation)
BV	Variation bornée (Bounded Variation)
CMB	Fond diffus cosmologique (Cosmic Microwave Background)
CS	Echantillonnage compressé (Compressed Sensing)
dB	decibel
DCT	Discrete Cosine Transform
DR	Schéma de décomposition de Douglas-Rachford
DWT	Discrete Wavelet Transform
EDP	Equation aux Dérivées Partielles
EEG	Electro-encephalographie
EFICA	Efficient Fast Independent Component Analysis
EM	Expectation Maximization
ERC	Exact Recovery Coefficient
FB	Schéma de décomposition de explicite-implicite (Forward-Backward)
FBM	Mouvement Brownien fractionnaire (Fractional Brownian Motion)
FDCT	Fast Discrete Curvelet Transform
FGP	Processus Gaussien fractionnaire (Fractional Gaussian Process)
GGD	Distribution Gaussienne Généralisée (Generalized Gaussian Distribution)
IRMf	Imagerie par Résonance Magnétique fonctionnelle
GCV	Validation croisée généralisée (Generalized Cross-Validation)
GMCA	Analyse en composantes morphologiques généralisée (Generalized Morphological Component Analysis)
ICA	Analyse en composantes indépendantes (Independent Component Analysis)
iid	indépendantes et identiquement distribuées
JADE	Joint Approximate Diagonalization of Eigen-matrices
KL	Kullback-Leibler
Lasso	Least absolute shrinkage and selection operator
lsc	Semi-continues inférieurement (lower semi-continuous)
MAD	Déviation Médiane Absolue (Median Absolute Deviation)
MAP	Maximum a posteriori
MBKF	Formes K de Bessel multivariées (Multivariate Bessel K Form)
MCA	Analyse en composantes morphologiques (Morphological Component Analysis)

MDL	Minimum Description Length
MMSE	Espérance conditionnelle a posteriori (Minimum Mean-Square Error)
MMV	Vecteurs à mesures multivariées (Multiple Measurements Vectors)
MOM	Mean of Max
MSE	Erreur quadratique moyenne (Mean-Square Error)
MSMG	Mélange d'échelle de Gaussiennes multivariées (Mutlivariate Scale-Mixture of Gaussians)
MS-VST	Stabilisation de la variance multiéchelle (MultiScale Variance Stabilization Transform)
NMSE	Erreur quadratique moyenne normalisée (Normalized Mean Square Error)
PDF	Densité de probabilité (Probability Density Function)
PSD	Densité spectrale de puissance (Power Spectral Density)
PSF	Fonction d'étalement du point (Point Spread Function)
PSNR	Rapport signal-sur-bruit en pic (Peak Signal to Noise Ratio)
RBAC	Contour actif basés région (Region-Bases Active Contour)
RIP	Propriété d'isométrie restreinte (Restricted Isometry Property)
RNA	Relative Newton Algorithm
SAR	Synthetic Aperture Radar
SMG	Mélange d'échelle de Gaussiennes (Scale-Mixture of Gaussians)
SNR	Rapport signal-sur-bruit (Signal-to-Noise Ratio)
SURE	Stein Unbiased Risk Estimator
t.q.	tel que
TV	Variation totale (Total Variation)
UDWT	Undecimated Discrete Wavelet Transform
VA	Variable aléatoire
VST	Transformée stabilisatrice de la variance (Variance Stabilizing Transform)
WMAP	Wilkinson Microwave Anisotropy Probe

Chapitre 1

Parcimonie et représentation des signaux et des images

La parcimonie est devenue indéniablement un concept majeur en traitement du signal et des images, et son utilisation comme source d'a priori a connu un essor considérable ces dernières années notamment pour des tâches comme la compression, l'estimation, la restauration, la séparation de sources ou encore la détection, etc. En témoigne la littérature exhaustive dont une revue a été dressée dans [A1, 148]. Au-delà de la théorie du traitement du signal et des images, la parcimonie a par ailleurs constitué une propriété attractive dans plusieurs domaines des mathématiques appliquées, notamment en analyse harmonique appliquée et computationnelle, en mathématiques statistiques ou en théorie de l'apprentissage.

L'intérêt de la parcimonie a sans doute culminé grâce à l'émergence plus récente de l'échantillonnage compressé (*compressed sensing*) qui renouvelle la vision de Shannon de l'échantillonnage pour l'acquisition des signaux analogiques qui ne sont pas nécessairement à bande limitée [43, 71, 37]. En établissant ainsi l'impact de la parcimonie sur les protocoles d'échantillonnage, l'échantillonnage compressé a eu des conséquences directes dans plusieurs domaines scientifiques comme la théorie de l'information, les codes correcteurs, l'imagerie médicale, la géophysique, l'imagerie astronomique. Toute proportion gardée, l'émulation créée par l'échantillonnage compressé est comparable à celle suscitée par les ondelettes il y a deux décennies, en favorisant une interaction forte de chercheurs provenant de différentes disciplines.

1.1 Le monde de la parcimonie

1.1.1 Terminologie

Trames

Définition 1 Une matrice Φ est un opérateur de synthèse d'une trame de $\mathbb{E}^N$, où $\mathbb{E} = \mathbb{C}$ ou $\mathbb{R}$, si son adjoint, c.à.d. l'opérateur d'analyse Φ^* , satisfait la relation de Parseval généralisée

$$c_1 \leq \|\Phi\Phi^*\| \leq c_2, \quad 0 < c_1 \leq c_2 < +\infty. \quad (1.1)$$

Autrement dit, le quotient de Rayleigh de $\Phi\Phi^*$ est dans $[c_1, c_2]$. L'opérateur $\Phi\Phi^*$ est appelé opérateur de trame, $\Phi^*\Phi$ est l'opérateur de Gram. La trame est dite ajustée si $c_1 = c_2 = c$, auquel cas l'opérateur de trame est diagonal $\Phi\Phi^* = c\mathbf{I}$.

Atome Un atome est un signal élémentaire entrant comme élément constituant d'un signal plus complexe par superposition linéaire. Des exemples incluent les sinusoides, les monômes, les ondelettes, etc.

Dictionnaire Un dictionnaire Φ est une collection d'atomes $\{\varphi_\gamma\}_{\gamma \in \Gamma}$ indexés par γ , où Γ est un ensemble dénombrable de cardinalité $|\Gamma| = T$. Généralement, tous les atomes sont de norme ℓ_2 unitaire.

L'interprétation de l'indice γ dépend du dictionnaire : fréquence pour Fourier, position pour Dirac (base canonique), position-échelle pour les ondelettes, position-durée-fréquence pour les paquets de cosinus, la position-échelle-orientation pour les curvelets en 2D, etc. En dimension finie, un dictionnaire est vu comme une matrice $N \times T$ dont les colonnes sont les atomes. Lorsque le dictionnaire est une matrice rectangulaire $T > N$, il est dit redondant (ou surcomplet). Dans la majeure partie de ce document, Φ correspondra à une base orthonormée ou une trame.

Analyse and synthèse Soit x un signal réel dans $\mathbb{R}^N$. Etant donné un dictionnaire, on distingue deux opérations : l'analyse et la synthèse. L'analyse est l'opération qui associe à chaque signal x un vecteur de coefficients α : $\alpha = \Phi^T x = \mathbf{T}x$ (bien évidemment, si le dictionnaire est complexe, Φ^* se substitue à Φ^T). La synthèse est l'opération consistant à reconstruire x en superposant les atomes : $x = \Phi\alpha$.

Dans la suite de ce document, étant donné un signal x , les algorithmes qui seront décrits mettent en jeu le calcul de la transformée de x en lui appliquant Φ^T (ou Φ^*). Reconstruire un signal à partir du vecteur de coefficients α consiste à lui appliquer la matrice de synthèse Φ . Ces matrices ne sont jamais construites explicitement en pratique. Leur application est plutôt implémentée sous forme implicite rapide. Dans un dictionnaire contenant une base orthonormée, il est clair que $\mathbf{T}^{-1} = \Phi$; et si Φ correspond à une trame ajustée, alors Φ n'est autre que l'inverse généralisée de Moore-Penrose $\mathbf{T}^+$ à une constante près : $\mathbf{T}^+ = c^{-1}\Phi$. Ce qui est remarquable c'est que beaucoup d'opérateurs de reconstruction implicite rapide implémentent justement $\mathbf{T}^+\alpha$.

Cohérence

Définition 2 La cohérence mutuelle μ_Φ de Φ (dont les colonnes sont de norme ℓ_2 unitaire) est la plus grande valeur absolue des éléments hors-diagonaux de sa matrice de Gram :

$$\mu_\Phi = \max_{i \neq j} |\langle \varphi_i, \varphi_j \rangle| . \quad (1.2)$$

Cette quantité est une mesure de ressemblance au pire cas entre paires d'atomes. La cohérence satisfait $1/\sqrt{N} \leq \mu_\Phi \leq 1$, la borne inférieure étant atteinte pour certaines paires de bases orthonormées. En général, $\mu_\Phi \geq \sqrt{\frac{T-N}{N(T-1)}}$, avec égalité si et seulement si Φ est une trame ajustée équi-angulaire [194]. Un dictionnaire est dit incohérent si sa cohérente μ_Φ est suffisamment petite.

1.1.2 Signaux strictement parcimonieux

Un signal x , considéré comme un vecteur dans un sous-espace de dimension finie de $\mathbb{R}^N$ ³, $x = [x[1], \dots, x[N]]$, est strictement ou exactement parcimonieux si la plupart de ses composantes sont nulles ; c.à.d. son support $\text{supp}(x) = \{1 \leq i \leq N \mid x[i] \neq 0\}$ est de cardinalité $|\text{supp}(x)| = K \ll N$. Un tel signal est dit K -parcimonieux.

Le signal peut ne pas être parcimonieux dans son domaine original, mais plutôt dans le domaine d'une transformée appropriée. C'est typiquement le cas pour un signal sinusoïdal qui est parcimonieux dans le domaine de Fourier. Un autre exemple intuitif est celui des images constantes par morceaux au-delà de contours de longueur finie qui ont un gradient parcimonieux.

Plus généralement, un signal x est supposé issu d'un modèle génératif de synthèse par combinaison linéaire de T atomes, tel que

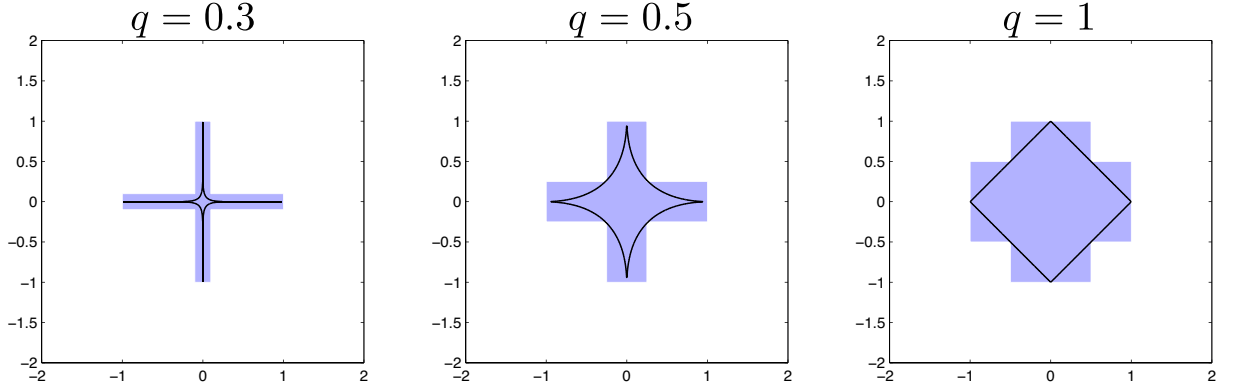
$$x = \Phi\alpha = \sum_{i=1}^T \alpha[i] \varphi_i , \quad (1.3)$$

où chaque $\alpha[i]$ est le coefficient (de synthèse) représentant la contribution de l'atome φ_i dans x .

Les signaux et images parcimonieux sont ceux que l'on peut écrire exactement comme

$$x = \Phi\alpha = \sum_{i \in \text{supp}(\alpha)} \alpha[i] \varphi_i , \quad \text{avec } \text{supp}(\alpha) = K \ll N . \quad (1.4)$$

3. La description peut être étendue au cas complexe, mais ça ne sera pas fait ici.

FIGURE 1.1 – Exemples dans $\mathbb{R}^2$ de boules unités $w\ell_q$ (plein) et ℓ_q (trait) pour $q \leq 1$.

1.1.3 Signaux compressibles et approximation non linéaire

La plupart des signaux et les images que l'on observe dans la pratique n'ont pas un support de petite taille. En revanche, ils peuvent vivre au voisinage d'un tel ensemble. Le modèle communément adopté est celui des signaux compressibles (faiblement parcimonieux). Une mesure de compressibilité consiste à considérer la vitesse de décroissance des coefficients de représentation α , typiquement en loi de puissance. Les signaux réguliers ou réguliers par morceaux, les images à variation bornée ou de norme de Besov bornée sont de ce type [82, 148].

Plus formellement, définissons la norme ℓ_q -faible $\|\alpha\|_{w\ell_q}^q := \sup_{\epsilon > 0} \epsilon^q |\{i \mid |\alpha[i]| > \epsilon\}|$. Supposons que α appartient à la boule ℓ_q -faible de rayon C , notée $w\ell_q(C)$ pour tout $0 < q < \infty$. En d'autres termes, si on trie les coefficients α par ordre décroissant en module, $|\alpha_{(1)}| \geq |\alpha_{(2)}| \geq \dots \geq |\alpha_{(T)}|$, alors

$$|\alpha_{(i)}| \leq C i^{-1/q}, \quad i = 1, \dots, T. \quad (1.5)$$

Autrement dit, la séquence des coefficients exhibe une décroissance en loi de puissance dont la vitesse est contrôlée par le paramètre q . Plus q est petit, plus la décroissance est rapide. Une boule $w\ell_q(C)$ est non convexe pour $q \leq 1$, et la boule ℓ_q y est strictement incluse (comme illustré par la fig. 1.1 dans $\mathbb{R}^2$). Par ailleurs, pour $T \rightarrow \infty$, $\ell_q \subset w\ell_q(C) \subset \ell_{q'}$ quelque soit $q' > q$ [64].

Un chapitre classique en théorie de l'approximation consiste à relier l'erreur d'approximation non linéaire par les K -premiers termes aux normes de séquences. En particulier, un objet $\alpha \in w\ell_q(C)$ satisfait

$$\|\alpha - \alpha_K\|_2 \leq C(2/q - 1)^{-1/2} K^{1/2 - 1/q}, \quad q < 2, \quad (1.6)$$

où α_K est la restriction de α à ses K plus grandes composantes. La même vitesse se répercute sur l'erreur d'approximation du signal x par $\sum_{i=1}^K \alpha_{(i)} \varphi_i$ puisque Φ est linéaire borné (en fait Φ correspond à une base orthonormée ou une trame dans la plupart des applications). En conclusion, x peut être bien approché par un signal K -parcimonieux dans Φ .

1.1.4 Analyse harmonique computationnelle

Dans ce document, on s'intéresse au point de vue offert par l'analyse harmonique computationnelle, une discipline dont les avancées récentes ont permis le développement des ondelettes, des paquets d'ondelettes ou de cosinus, des curvelets, des contourlets, des bandlets, des brushlets, des wave-atoms, et plusieurs autres représentations fournissant une description creuse d'images ayant différents contenus géométriques. La démarche classique en analyse harmonique computationnelle est (i) de trouver la représentation optimale décrivant de manière creuse des structures dans le signal ou l'image, et (ii) de concevoir un algorithme rapide pour calculer les transformées directe (analyse) et inverse (synthèse) correspondantes. La représentation optimale dépend de l'espace fonctionnel auxquels les signaux-images sont sensés appartenir. Par exemple,

- les séries de Fourier sont optimales pour les images vivant dans des espaces de Sobolev W_2^s ;
- les ondelettes sont optimales pour des images à variation bornée [52] ;
- les ridgelets sont optimales pour les images régulières au-delà de discontinuités le long de lignes [39] ;
- les curvelets [41, 40, A2], les contourlets [68], les bandlets [137, 172, 149] ou les shearlets [116] sont (presque) optimales pour les images C^2 - C^2 dites cartoon (c.à.d. images régulières par morceaux au-delà de contours eux mêmes réguliers) ;
- les brushlets [156] et les wave-atoms [62] sont optimales pour les textures localement oscillantes.

Une fois l'optimalité établie pour cette classe, les tâches ultérieures comme la compression ou l'estimation s'en trouvent facilitées.

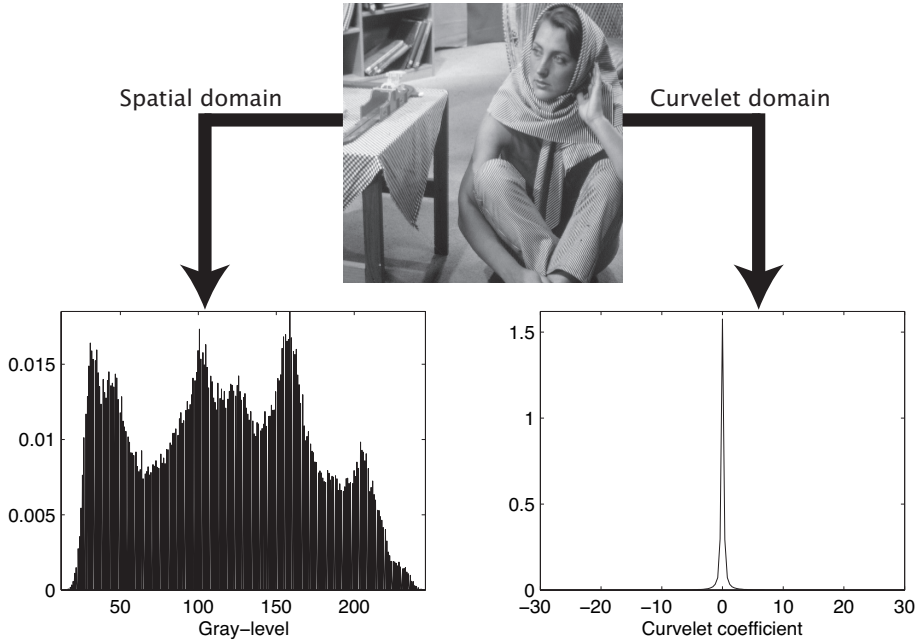


FIGURE 1.2 – Histogramme de l'image **Barbara** dans le domaine spatial, et celui de ses coefficients de curvelets.

La fig. 1.2 montre l'histogramme de l'image de **Barbara** dans le domaine spatial (gauche) et celui des curvelets (droite). Le second présente clairement un comportement typique d'une représentation creuse (unimodal, piqué en zéro avec des queues lourdes), où la majorité des coefficients est négligeable et seuls quelques uns sont significatifs et subsistent dans les queues.

1.2 Redondance et diversité morphologique

1.2.1 Apport de la redondance

Récemment, plusieurs auteurs d'horizons divers ont argumenté en faveur de l'utilisation des représentations redondantes. Ces représentations, en offrant un plus large panel d'atomes, permettent une description encore plus parcimonieuse de signaux complexes et une plus grande flexibilité pour s'adapter à leur contenu morphologique. Les neurosciences ont aussi souligné le rôle central de la redondance. Ceci a été mis en évidence dans plusieurs travaux sur le système visuel primaire chez le mammifère [94, 121, 169, 188].

Si la redondance n'est pas désirable pour des applications de type compression où elle est plutôt à éviter, elle s'avère très précieuse pour les problèmes inverses notamment pour le débruitage, la restauration ou la reconstruction.

1.2.2 La diversité morphologique

Le concept de diversité morphologique, avancé par [190, 191], pousse l'idée de la redondance encore plus loin en supposant que le signal d'intérêt x peut être écrit comme un mélange fini de K composantes x_k qui sont morphologiquement distinctes,

$$x = \sum_{k=1}^K x_k, \quad (1.7)$$

où chaque x_k est parcimonieuse dans un dictionnaire dédié Φ_k avec les opérateurs d'analyse et de synthèse rapides associés. Chaque x_k est appelée *composante morphologique*. Par exemple, si une image est composée de lignes globales et de textures localement oscillantes, il sera judicieux de l'analyser avec un dictionnaire contenant des ridgelets pour les lignes, et la DCT locale ou les wave-atoms optimaux pour les textures oscillantes (voir fig. 6.1, Chapitre 6).

Plusieurs des contributions décrites dans ce document se basent sur ce concept. Celles-ci incluent l'estimation, les problèmes inverses, la décomposition parcimonieuse et la séparation aveugle de sources.

1.2.3 Choix du dictionnaire

Le dictionnaire redondant global est choisi comme une union de sous-dictionnaires $\{\Phi_k\}_{k=1\dots K}$ (fig. 1.3). Le choix des sous-dictionnaires peut être guidé par les arguments empruntés à l'analyse harmonique computationnelle discutés en section 1.1.4. Par exemple, les ondelettes vont être bien adaptées pour des structures isotropes, les ridgelets pour des singularités le long de lignes globales, ou les curvelets pour des images avec des contours curvilignes réguliers, les wave-atoms ou la DCT locale pour les textures localement oscillantes, etc. En fait, en choisissant un sous-dictionnaire plutôt qu'un autre, un a priori est implicitement injecté sur ce que sont les constituants élémentaires du signal-image (voir section 2.3.2 pour un point de vue statistique).

Un autre point de vue sur ce choix est emprunté à un résultat très classique en traitement du signal : la filtrage adapté. Le théorème du filtre adapté affirme que pour détecter de façon optimale un signal en environnement bruité, l'observation doit être corrélée avec le signal et le rapport signal-à-bruit (SNR) sera maximal lorsque le signal en question est présent. L'enseignement tiré de ce résultat est que pour analyser efficacement un signal, il y a tout intérêt à choisir un dictionnaire dont les atomes ressemblent au mieux aux structures présentes dans ce signal.

1.3 Décomposition et approximation parcimonieuses dans des dictionnaires redondants

Le problème de la décomposition parcimonieuse d'un signal $x \in \mathbb{R}^N$ dans un dictionnaire redondant Φ consiste à retrouver les coefficients α dans (1.3), ayant observé x . Ceci revient donc à résoudre un système linéaire sous-déterminé. Bien que cette inversion soit notoirement un problème mal posé par les arguments élémentaires de l'algèbre linéaire, la situation change radicalement si le signal à recouvrer est parcimonieux suivant (1.4). En effet, parmi toutes les représentations de x dans Φ , nous voudrions la plus parcimonieuse, autrement dit la solution au problème d'optimisation

$$\min_{\alpha \in \mathbb{R}^T} \|\alpha\|_0 \quad \text{t.q.} \quad x = \Phi\alpha. \quad (1.8)$$

A première vue, cette tâche s'annonce ardue puisque (1.8) est un problème combinatoire NP-difficile. Pour contourner cette difficulté, tout un pan de littérature a vu le jour proposant des approximations (par des algorithmes de poursuite gloutons comme le matching pursuit et ses variantes) ou des relaxations convexes de (1.8). Dans [75], il est proposé de remplacer le problème ℓ_0 (1.8) par sa relaxation convexe ℓ_1 ,

$$\min_{\alpha \in \mathbb{R}^T} \|\alpha\|_1 \quad \text{t.q.} \quad x = \Phi\alpha. \quad (1.9)$$

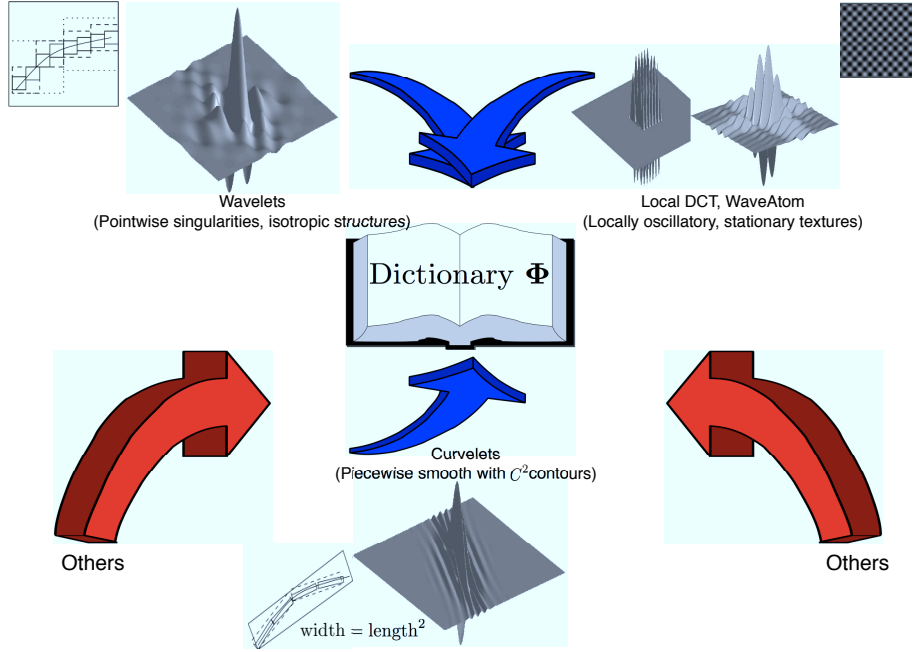


FIGURE 1.3 – Illustration du choix du dictionnaire reflétant la diversité morphologique d'une image.

Ce problème est connu sous l'appellation "poursuite de base" (Basis Pursuit, BP) [49]. Il sera montré aux chapitres 5-7 que (1.9) peut être résolu efficacement par des outils de décomposition d'opérateurs monotones maximaux. La régularisation ℓ_1 a été proposée dès la décennie 80-90 notamment par [182, 8] pour les problèmes inverses linéaires, et formalisée par D. Donoho et ses collaborateurs [79, 77, 69].

Toutefois, il est clair que les minimisations ℓ_0 et ℓ_1 ne sont en général pas équivalentes, puisque le minimiseur global de BP (si tant est qu'il soit unique) ne coïncide pas toujours avec celui de (1.8). Mais sous des hypothèses appropriées sur Φ et x , BP possède un unique minimiseur global qui de plus coïncide avec celui de (1.8); on parle alors d'identifiabilité ℓ_1 . Une littérature très importante s'est penchée sur les conditions suffisantes (parfois même nécessaires) pour assurer l'unicité de la solution de BP, son équivalence avec celle de (1.8) ainsi que sa robustesse aux signaux compressibles. En présence d'un bruit typiquement de variance bornée, (1.9) est modifiée en adaptant le terme de fidélité aux données. Des conditions ont été aussi établies assurant la stabilité par minimisation ℓ_1 à ce type de bruits de même que la consistance du support et du signe du vecteur retrouvé. Il n'est pas question ici de dresser une revue exhaustive ne serait-ce que parce que celle-ci évolue tellement vite qu'il serait illusoire d'être équitable et complet. Seuls les travaux les plus pertinents au regard de nos contributions du Chapitre 7 sont soulignés ci-après. Une revue assez complète est dressée dans [31] sur les critères de décomposition parcimonieuse aussi bien par minimisation ℓ_1 que par les algorithmes de poursuite gloutons.

Cohérence L'une des conditions suffisantes d'identifiabilité les plus connues est celle basée sur la cohérence mutuelle (1.2) [75, 73, 32, 111, 92, 74, 196]. Ces auteurs ont montré qu'un signal x K -parcimonieux dans Φ tel que

$$|\text{supp}(\alpha)| = K < C(\mu_{\Phi}^{-1} + 1), \quad (1.10)$$

pour une constante $C > 0$ (typiquement $C = 1/2$ [75]), alors la solution de BP est unique et coïncide avec le minimiseur global de (1.8) (en fait avec celui de la minimisation ℓ_q , $0 \leq q < 1$ [112]), assurant ainsi l'identifiabilité de la solution la plus parcimonieuse par minimisation ℓ_1 . En d'autres termes, un signal suffisamment parcimonieux dans un dictionnaire incohérent est identifiable par minimisation ℓ_1 . La cohérence dans (1.10) peut être légèrement raffinée par la cohérence cumulée ou la fonction de Babel [196]. Ce type de bornes assure par ailleurs la stabilité à un bruit borné et la robustesse à la compressibilité pour $C = 1/4$ dans (1.10) [74]. Cependant, le critère (1.10) correspond à une analyse au pire au cas et

conduit ainsi à des bornes pessimistes sur le niveau de parcimonie requis qui est au plus en $O(\sqrt{N})$, ou de façon équivalente, le nombre de mesures N nécessaire pour assurer l'identifiabilité ℓ_1 est $\gtrsim K^2$. Pour l'unicité du minimiseur du problème ℓ_0 (1.8), les critères type cohérence peuvent être raffinés en utilisant la notion de rang de Kruskal ou le spark de Φ [73, 74]. $\text{spark}(\Phi)$ est le nombre minimal de colonnes linéairement dépendantes, à ne pas confondre avec $\text{rang}(\Phi)$. Toutefois, non seulement le spark est aussi difficile à calculer que résoudre (1.8), mais en plus, le spark stricto sensu est très sensible aux perturbations sur Φ , et il n'est pas adapté pour étudier la stabilité de la minimisation ℓ_0 en présence de bruit additif de variance finie [74].

Support et signe Des conditions suffisantes plus fines ont été proposées en exploitant la structure du support soit en se basant sur le support uniquement (critère ERC) [196, 202], ou le support et le signe [101]. On reviendra sur le critère de [101] dans le Chapitre 7.

Critères topologiques En se basant sur les propriétés topologiques du polytope obtenu par projection aléatoire de la boule ℓ_1 unitaire par Φ , [72, 80] déterminent des bornes très fines sur le cardinal du support assurant l'identifiabilité ℓ_1 en résolvant BP,

$$K \leq \rho(N/T)N, \quad (1.11)$$

où $\rho(\eta) > 0$ est une fonction croissante de η délimitant une transition de phase entre vecteurs identifiables et ceux qui ne le sont pas. Ces auteurs distinguent plusieurs régimes de fonctions $\rho(\eta)$, dont l'un pour l'identifiabilité de tous les vecteurs (borne uniforme), l'autre pour la majorité des vecteurs. Pour une matrice Gaussienne et $T \gg N$, on a $\rho(N/T) = 1/(2 \log(T/N)) + o(1)$ [80]. Ces conditions sont à notre connaissance les plus fines de la littérature en absence de bruit et coïncident avec les limites de parcimonie assurant l'identifiabilité ℓ_1 observés en pratique, comme nous l'avons montré dans [A9, A49, A50], voir section 7.4. Mais ce critère n'est hélas pas robuste au bruit, ce qui au passage serait une bonne perspective future à examiner.

En s'inspirant de ces travaux et de ceux de [101], [83] établit qu'un vecteur α est ℓ_1 -identifiable si et seulement s'il appartient à l'adhérence de l'ensemble des vecteurs satisfaisant le critère de [101]. La condition de [83] est de surcroît robuste à un faible bruit.

Isométrie restreinte Depuis l'avènement de la théorie de l'échantillonnage compressé, des conditions d'identifiabilité ℓ_1 basées sur la propriété d'isométrie restreinte [42] (RIP pour restricted isometry property) ou des variantes proches ont été développées à la fois dans les communautés de la théorie de l'information et du signal, des statistiques ou de la théorie de l'apprentissage. L'hypothèse du RIP assure non seulement l'identifiabilité par BP uniformément sur tous les signaux, mais aussi la robustesse au bruit et aux signaux compressibles. Le RIP est de plus vérifié avec une grande probabilité pour plusieurs familles de matrices aléatoires, notamment l'ensemble sous-Gaussien isotrope dès lors que $\rho(N/T) = O(1/\log(T/N))$ [155], mais avec des constantes moins fines que les critères topologiques dans le cas sans bruit. On reviendra avec plus de détail sur ces conditions basées RIP dans la section 7.3.

1.4 Récapitulatif et conclusion

Ce chapitre décrit le premier concept central dans ces travaux de recherche : la parcimonie. Plusieurs des contributions décrites dans ce document se basent sur ce concept comme l'estimation, les problèmes inverses, la décomposition parcimonieuse et la séparation aveugle de sources, ou encore l'échantillonnage compressé. Quelques définitions ont été données, l'importance de la redondance et de la diversité morphologique a été mise en exergue, ainsi que le lien avec l'analyse harmonique computationnelle et la théorie de l'approximation non linéaire (pour les signaux compressibles). Ensuite, le problème de la décomposition parcimonieuse par minimisation ℓ_1 dans les dictionnaires redondants a été discuté avec une revue ciblée des conditions suffisantes (et parfois nécessaires) assurant l'identifiabilité par minimisation ℓ_1 . Certains de ces critères seront repris au Chapitre 7.

Chapitre 2

Modélisation statistique des images

Dans ce chapitre, nous allons établir les conséquences et le lien entre l'a priori de parcimonie et la modélisation aléatoire des images. Les travaux concernés par ce chapitre sont la première partie de la thèse de L. Boubchir [TH4], ainsi que [A30, A27, A73, A87, A83, A90, A112, A113, A114].

2.1 Lois leptokurtiques à queues lourdes

Comme il a été souligné en introduction, l'une des approches les plus directes pour établir les lois sur les images est d'étudier leurs statistiques marginales et jointes en réponses à des bancs de filtres linéaires. L'une des propriétés frappantes mise en évidence par de plusieurs études est que les statistiques en sortie de ces filtres ont un kurtosis élevé (voir (2.1) pour la définition du kurtosis), des queues lourdes et généralement unimodales en 0 et symétriques autour de ce mode. Elles sont ainsi loin d'avoir un comportement Gaussien. Ce constat est apparu dans les travaux de D. Field [93] et ensuite confirmé par plusieurs auteurs [120, 110, 109, 139].

2.1.1 Statistiques marginales

Si X est une variable aléatoire (VA) de moyenne nulle, son kurtosis centré (ou coefficient d'aplatissement) est défini comme

$$\text{Kurtosis}(X) = \frac{\mathbb{E}[X^4]}{\mathbb{E}[X^2]^2} - 3. \quad (2.1)$$

Une VA normale a un kurtosis nul ; une VA sans queues a un kurtosis négatif ; une VA ayant des queues lourdes et très piquée en sa moyenne a un kurtosis positif (on parle de VA leptokurtique). Le résultat empirique qui a été observé pour les images est que pour tout filtre h de moyenne nulle, l'image filtrée $x_f = x * h$ est justement leptokurtique à queues lourdes. C'est typiquement le cas pour un dérivateur discret, mais cet effet a été aussi observé pour des filtres aléatoires à support compact et de moyenne nulle [120], ou encore pour de nombreuses transformées comme la transformée en ondelettes (voir fig. 2.1).

Cette observation empirique doit aussi être interprétée à l'aune d'un résultat connu en théorie des probabilités stipulant que pour un processus aléatoire de Markov stationnaire 1D $X[t], t \in \Omega$ un ouvert inclus dans $\mathbb{Z}$, le fait que le kurtosis de l'incrément $X[t] - X[t']$ soit positif implique que $X[t]$ contient des sauts discrets. Ce résultat se généralise en dimension supérieure pour couvrir le cas des images par extension du processus incrément. Une image acquise est justement formée d'objets discrets séparés par leurs frontières. Ainsi, ce fait saillant sur la nature physique des images est clairement reflété par la simple statistique du kurtosis.

Par ailleurs, d'un point de vue de la parcimonie, la compressibilité d'une image dans une transformée donnée, c.à.d. la vitesse de décroissance de ses coefficients $\alpha[i]$, est étroitement liée à la vitesse de décroissance des queues de son histogramme. En effet, soit $p(\alpha)$ la densité de probabilité (PDF) marginale

d'un coefficient ⁴ α sous-jacente à l'histogramme empirique. A partir de la fonction de répartition de $|\alpha|$, on définit la probabilité de queue qui est une fonction décroissante sur $\mathbb{R}^+$:

$$P(\alpha) := 1 - \int_{-\alpha}^{\alpha} p(t)dt = 1 - 2 \int_0^{\alpha} p(t)dt \quad (2.2)$$

par symétrie de $p(\alpha)$. Par les résultats classiques sur les statistiques d'ordre, on peut observer que pour T grand, alors avec une grande probabilité le quantile d'ordre i/T est ⁵

$$|\alpha_{(i)}| = P^{-1}\left(\frac{i}{T}\right),$$

où on rappelle que T est le nombre total de coefficients et $\alpha_{(i)}$ sont les coefficients triés par ordre décroissant en valeur absolue. D'après la section 1.1.3, un signal compressible dans un domaine transformé est tel que ses coefficients $\alpha \in w\ell_q(C)$, et donc décroissent selon (1.5). En d'autres termes, $\alpha(P) = CP^{-1/q}$, et ainsi par dérivation

$$p(\alpha) = C'_q \alpha^{-(q+1)} \quad (2.3)$$

pour α grand. En clair, la parcimonie implique des PDF leptokurtiques à queues lourdes. Dans les travaux [A90, A27], issus de la thèse [TH4], nous avons montré que les distributions α -stables [141] exhibent exactement ce type de décroissance asymptotique et apparaissent comme de bons candidats pour modéliser ce comportement leptokurtique; voir aussi [6]. Cependant, la variable aléatoire sous-jacente α n'est de variance finie que pour $q \geq 1$ ce qui ne couvre qu'une partie du régime $q < 2$ pour lequel (1.6) est valide.

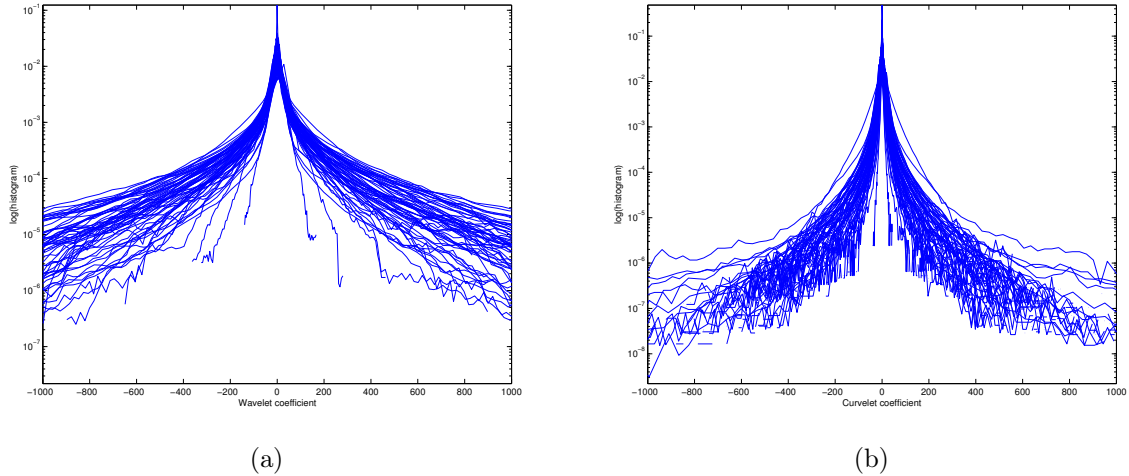


FIGURE 2.1 – Histogramme en échelle log des coefficients d'ondelettes (a) et de curvelets (b) de 100 images naturelles de la base [1]. Des exemples d'images de cette base sont montrés fig. 2.6. Les kurtosis estimés étaient entre 5 et 101 pour la transformée en ondelettes et entre 10 et 3000 pour les curvelets.

2.1.1.1 Mélange d'échelle de Gaussiennes univarié : cadre général

Pour ajuster des modèles théoriques sur les histogrammes marginaux observés, il faut répondre à une question importante : quelle est la taille des queues lourdes associées. Plusieurs modèles ont ainsi été proposés dans la littérature dont certains sur lesquels nous avons été actifs : les Gaussiennes généralisées

4. On fait l'hypothèse de la stationnarité au sens strict et de l'ergodicité.

5. En utilisant des inégalités de concentration sachant que $\mathbb{E}[\alpha_{(i)}] = P^{-1}\left(\frac{i}{T}\right)$ avec une variance qui décroît en $1/T$ [184].

(GGD)⁶ [147, 187], les formes K de Bessel (BKF) [110, A30, TH4], les distributions α -stables [6, A27]. Il s'avère que ces différentes lois sont des cas particuliers de la loi de mélange d'échelle de Gaussiennes (SMG) introduite en statistiques dans [9], et mise à profit pour la modélisation des images dans [201, TH4].

Définition 3 ([9]) *Soit X une VA à valeurs dans $\mathbb{R}$. Sous le modèle SMG univarié, il existe deux VAs indépendantes $U \geq 0$ et $Z \sim \mathcal{N}(0, 1)$ telles que :*

$$X \stackrel{d}{=} Z\sqrt{U}, \quad (2.4)$$

où $\stackrel{d}{=}$ est l'égalité en distribution.

La loi α -stable correspond au cas où U suit elle-même une loi α -stable positive. Si U suit une loi de Gamma, on trouve le modèle BKF. Notons aussi que la GGD est un cas particulier du SMG mais la PDF de U dans ce cas ne possède pas de forme analytique.

Dans [TH4], nous avons montré que la famille de lois SMG est très bien adaptée pour capturer le caractère leptokurtique des histogrammes des représentations creuses. Ce résultat se résume ainsi :

Théorème 1 *La VA X possède une décomposition (2.4) si et seulement si les dérivées $k^{\text{ème}}$ de sa PDF prises en $\sqrt{t}$ sont de signe alterné, c.à.d. :*

$$\left(-\frac{d}{dt}\right)^k p_X(\sqrt{t}) \geq 0 \quad \forall t > 0. \quad (2.5)$$

De plus, p_X est unimodale, symétrique autour de zéro et leptokurtique, c.à.d. $\text{Kurtosis}(X) > 0$.

Lien avec les espaces de Besov Dans [A83, TH4], nous avons montré que la loi SMG sur les statistiques marginales est adaptée pour caractériser les images appartenant aux espaces de Besov. Plus spécifiquement, en exploitant la propriété d'invariance d'échelle (section 2.2), nous avons établi la relation entre les paramètres de l'a priori SMG et ceux de l'espace de Besov auquel les réalisations d'un tel a priori sont sensés appartenir (presque sûrement).

2.1.1.2 Cas particuliers et comparaison

Dans nos contributions [A30, A27], rassemblées dans le travail de thèse [TH4], nous avons étudié en détail deux instances de la famille SMG : les formes BKF et les lois α -stables. Chacune d'entre elles implique deux hyperparamètres la décrivant complètement : un hyperparamètre de forme et un hyperparamètre d'échelle. Nous avons proposé des estimateurs efficaces pour ces hyperparamètres aussi bien en absence qu'en présence de bruit additif Gaussien (dont les estimateurs du maximum de vraisemblance par un algorithme EM, des cumulants et des quantiles). Pour la loi α -stable, le mélange d'échelle a été approché dans [A27] par un mélange fini de Gaussiennes pour stabiliser les calculs numériques. La fig. 2.2 montre un exemple où les trois modèles de PDFs (BKF, α -stable exact et par mélange fini, et GGD) ont été ajustés aux histogrammes observés des coefficients d'ondelettes de trois images tests de la base [1]. Des résultats similaires sont obtenus pour d'autres transformées comme les curvelets [TH4].

La comparaison quantitative entre les différents modèles de PDF sur l'intégralité de la base test est résumée dans la fig. 2.3⁷. En résumé, ces modèles diffèrent essentiellement dans l'allure de leurs queues. De ce fait, nous avons constaté que le modèle BKF présente des performances un peu en dessous du modèle α -stable par mélange fini pour les échelles les plus fines, mais devient meilleur aux autres échelles. Le modèle α -stable exact a parfois tendance à exagérer les queues, alors que le modèle GGD peut avoir tendance à les sous-estimer, et même s'il présente de bonnes performances, reste inférieur aux modèles BKF et α -stable par mélange fini, particulièrement aux échelles fines.

6. C'est ce modèle qui est sous-jacent à la pénalité ℓ_p , $0 < p < 2$ souvent utilisée comme mesure de parcimonie notamment pour la gamme convexe. Le cas $p = 1$ correspond à (1.9). Toutefois, cette observation doit être tempérée quelque peu car au vu de [20], on peut montrer qu'un vecteur issu d'une loi GGD est dans $w\ell_q(C)$ pour $q < 2$ (c.à.d. compressible et satisfaisant (1.6)) si son paramètre de forme $p < (2 - e^{-1})/\log T$, $T \geq 2e$.

7. Même si ces modèles de PDF ont été appliqués à plusieurs transformées multiéchelles, seul le résultat sur la transformée en ondelettes est montré ici.

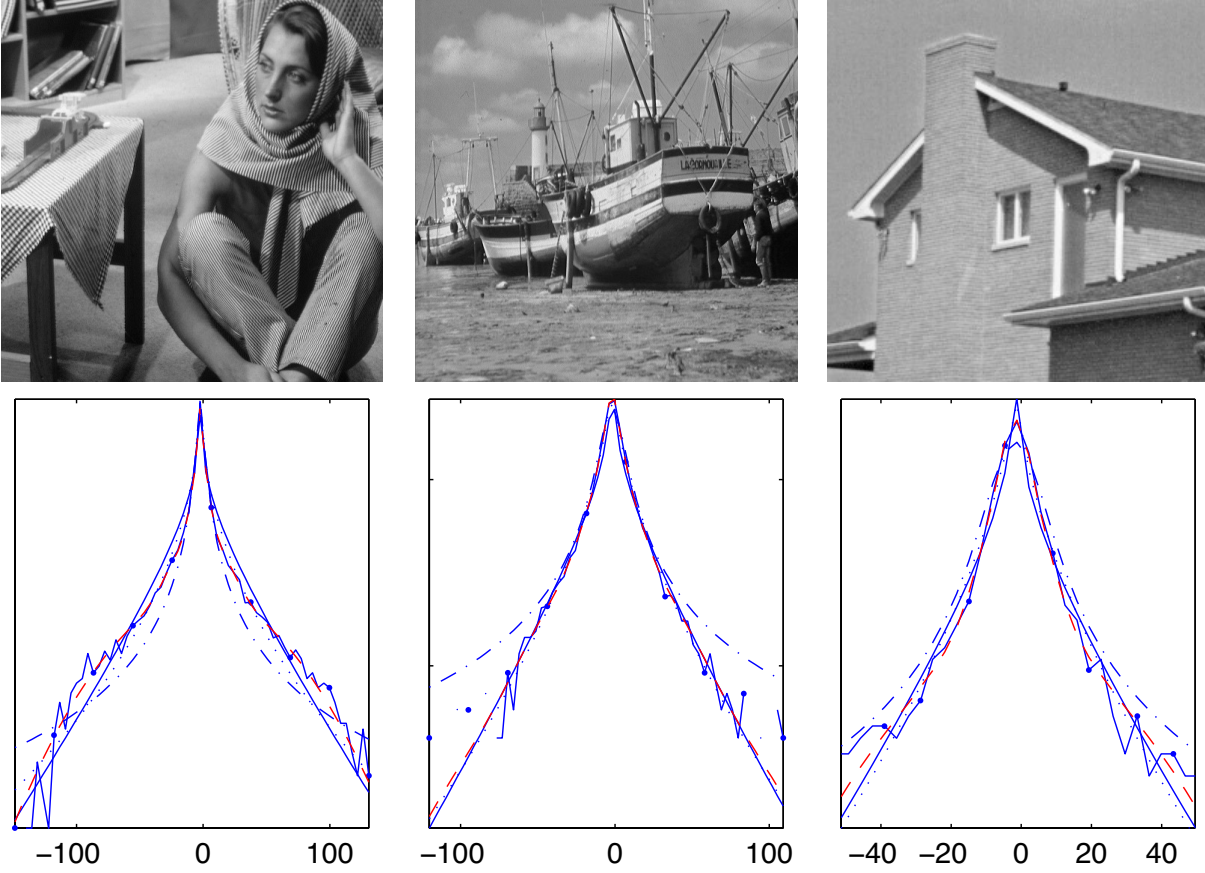


FIGURE 2.2 – Comparaison entre les PDFs ajustées à l’histogramme marginal observé (••, échelle log) des coefficients d’ondelettes d’une sous-bande pour trois images tests. Les trois modèles ajustés sont : BKF (trait plein), α -stable par mélange fini (trait pointillé), α -stable exact (alternance de points et de traits) et GGD (points).

2.1.2 Statistiques jointes

Dans une décomposition multiéchelle, dans les sous-bandes de détails, les coefficients de forte valeur tendent à se regrouper autour des singularités et ce quelle que soit l’orientation. Ces agglomérats de coefficients décrivent un chaînage géométrique local. Cette dépendance incite à introduire des modèles joints plus élaborés prenant en compte cette structuration de la parcimonie (e.g. par blocs, en arbre, etc.) lors de la modélisation, mais aussi pour les traitements ultérieurs notamment en inférence Bayésienne pour la résolution de nombreux problèmes inverses ; voir chapitres 4 et 5. Le travail de thèse [TH4] quantifie clairement cette dépendance intra-échelle en utilisant l’information mutuelle dans le cas de plusieurs transformées multiéchelles (e.g. ondelettes, curvelets) sur une grande base d’images.

2.1.2.1 Mélange d’échelle de Gaussiennes multivarié : cadre général

Dans [173, TH4], l’extension multivariée du modèle SMG univarié introduit dans Définition 3 est proposée. Il suffit de remplacer dans cette définition la VA normale Z par un vecteur de VA à valeurs dans $\mathbb{R}^{|\mathcal{V}|}$ de matrice de covariance Σ , où $|\mathcal{V}|$ est la taille du bloc isotrope $\mathcal{V}$ comprenant le coefficient et ses voisins. Les SMG multivariés (MSMG) font partie de la classe des distributions elliptiques [132].

De la même façon que pour le Théorème 1, dans la thèse de [TH4], il est montré que le modèle MSMG est un bon candidat pour modéliser le caractère leptokurtique avec dépendance. La conclusion de ce résultat est que la PDF jointe de la famille MSMG est symétrique et centrée en zéro, sa mesure du

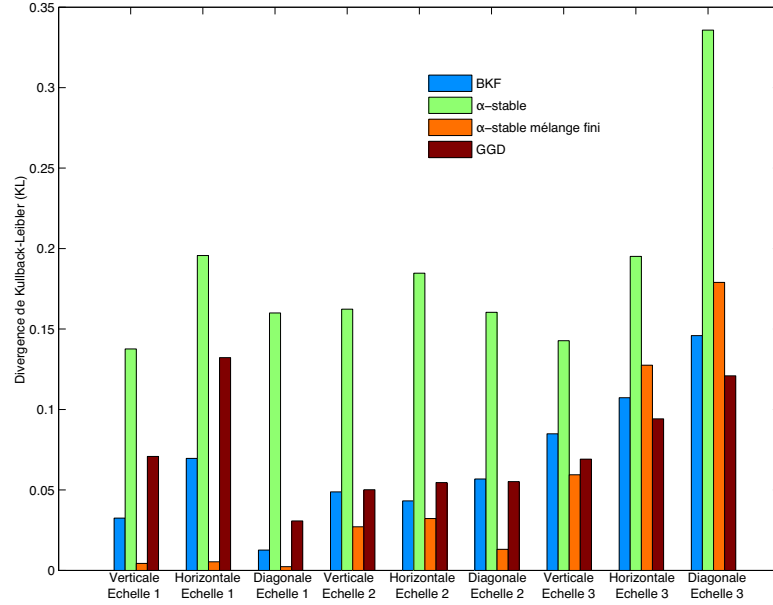


FIGURE 2.3 – Divergence de Kullback-Leibler (KL) entre l’histogramme marginal observé des coefficients d’ondelettes et les PDFs univariées ajustées correspondant aux lois BKF, α -stable exact et par mélange fini, et GGD. Les valeurs de KL ont été moyennées à chaque échelle et chaque orientation (verticale, horizontale et diagonale) sur la base [1] de 100 images.

kurtosis multivarié au sens de Mardia [152] est toujours strictement positive, et de plus, ses marginales sont leptokurtiques à queues lourdes.

2.1.2.2 Cas particuliers et comparaison

Plusieurs modèles proposés dans la littérature sont des cas particuliers des MSMG, chacun correspondant à un choix particulier de la loi de la variable de mélange U , e.g. l’a priori non informatif de Jeffrey [173], les distributions α -stables multivariées [5]. Dans [A83, A113, TH4], nous avons proposé les BKF multivariées (MBKF) et les GGD multivariées anisotropes (AMGGD). Ce dernier modèle n’est pas à symétrie elliptique mais est en mesure de capturer certaines anisotropies des statistiques jointes non accessibles aux distributions de la classe MSMG. De la même manière que dans le cas univarié, chacun de ces modèles met en jeu deux hyperparamètres, un hyperparamètre de forme contrôlant le comportement des queues et une matrice de covariance. Dans nos travaux, nous avons proposé plusieurs estimateurs de ces hyperparamètres.

La fig. 2.5 résume la comparaison entre les différents modèles de distributions jointes dans leurs performances à modéliser les histogrammes joints des coefficients d’ondelettes (sur 3 échelles et 3 orientations) et de curvelets (3 échelles et 4 quadrants) ; voir aussi fig. 2.4 pour une comparaison visuelle. La divergence KL est moyennée sur la base test de 100 images [1]. L’un des enseignements à retenir de cette étude est que le modèle MBKF dépasse le modèle de [173]. Par ailleurs, le modèle MBKF est généralement meilleur comparé à l’AMGGD dans le cas des ondelettes. Cette tendance s’inverse pour la cas des curvelets où le modèle AMGGD devient meilleur.

2.2 Invariance d'échelle et conséquences

L’autre propriété marquante des images est leur invariance par changement d’échelle⁸ (on parle aussi d’auto-similarité). Plusieurs auteurs ont mis en exergue cette propriété ; voir [180, 179, 135, 120] ainsi que

8. En fait, presque invariance car les effets de perspective sont supposés négligeables.

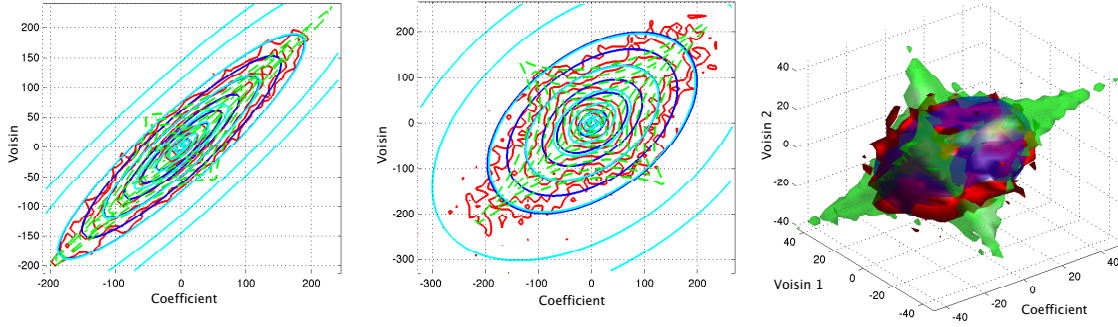


FIGURE 2.4 – Comparaison visuelle en échelle log entre les PDFs ajustées à l'histogramme joint observé (en rouge, échelle log) des coefficients d'ondelettes à une sous-bande pour une image test. À gauche et au milieu, statistique jointe (bivariée) d'un coefficient et un de ses voisins, et à droite statistique jointe (trivariée) d'un coefficient et deux de ses voisins. Les modèles multivariés observée ajustés sont : MBKF (en bleu), AMGGD (en vert) et celui de [173] (en bleu clair).

d'autres. De manière informelle, elle signifie que deux individus habitant le même environnement, dont le second porterait en permanence des lunettes grossissantes, auraient la même perception du monde, bien que le second reçoive systématiquement des images agrandies. En termes statistiques, la distribution marginale des images naturelles est sensée rester inchangée si les images sont agrandies ou réduites. Plus formellement, cette invariance peut être exprimée en termes de la mesure de probabilité μ munissant l'espace des images $\mathcal{X}$ sous l'action d'un changement d'échelle :

$$\mu(T_\phi(\mathcal{S})) = \mu(\mathcal{S}), \quad (2.6)$$

pour tout sous-ensemble mesurable $\mathcal{S}$ de $\mathcal{X}$, où $T_\phi(x) = \lambda^{-2}x \circ \phi$, $\lambda \in \mathbb{R}^+$, et $\phi(t_1, t_2) = (\lambda t_1, \lambda t_2)$.

2.2.1 Conséquence 1 : spectre en loi de puissance

La conséquence la plus simple de l'invariance d'échelle est la décroissance en loi de puissance de la densité spectrale de puissance (PSD) des images. De ce fait, la PSD en fonction de la fréquence radiale $\rho = \sqrt{\nu_1^2 + \nu_2^2}$ évolue selon :

$$\mathbb{E} \left[|\hat{x}(\nu_1, \nu_2)|^2 \right] \approx C / \rho^\gamma,$$

où $\gamma \geq 0$ est l'exposant caractéristique ; voir exemples fig. 2.6. Cette loi de puissance est une manifestation de la nature auto-similaire des images naturelles. Dans le domaine image, elle est équivalente à dire que la fonction d'auto-corrélation est $\approx C - \log \sqrt{t_1^2 + t_2^2}$. Cette observation a été faite dès les années 1950. Ensuite, plusieurs expériences ont validé cette loi sur des plages intermédiaires de la fréquence radiale⁹. Les valeurs trouvées pour l'exposant γ dépendent des scènes, et sont typiquement dans $[1, 3]$ avec une concentration autour de 2 pour certaines bases d'images comme des forêts ou des villes [120, 180]. Certains auteurs [162] ont observé des valeurs inférieures à 1 pour des scènes contenant de la végétation ou des nuages. Dans la base de test discutée ici, les valeurs trouvées vont de 0.85 à 2.

2.2.2 Conséquence 2 : persistance à travers les échelles

On vient d'examiner les conséquences de l'invariance d'échelle sur les statistiques du second ordre (la PSD). Or, cette invariance d'échelle se traduit aussi sur les statistiques d'ordres supérieurs (résumées par l'histogramme). En effet, il est naturel de faire appel à une décomposition multiéchelle pour analyser les images lorsqu'on évoque la notion d'invariance d'échelle. Et quand observe les modules des coefficients

9. La PSD n'étant pas intégrable sur $[0, \infty)$, l'énergie totale diverge.

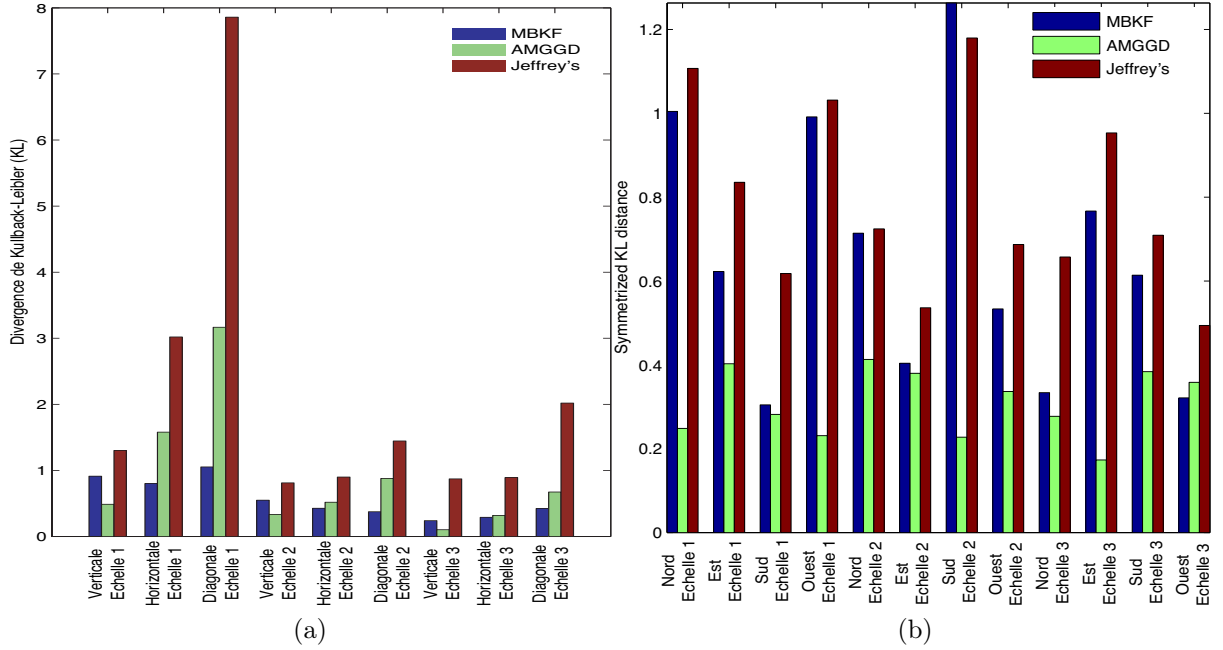


FIGURE 2.5 – Divergence de Kullback-Leibler (KL) entre l'histogramme joint observé des coefficients d'ondelettes (a) et de curvelets (b), et les PDFs multivariées ajustées pour les lois MBKF, AMGGD et celui de [173]. Les valeurs de KL ont été moyennées à chaque échelle et chaque orientation sur la base [1] de 100 images. Pour la transformée en ondelettes séparables, les trois sous-bandes classiques (horizontale, verticale et diagonale) sont considérées. Pour la transformée de curvelets, et pour une présentation synthétique des résultats, les valeurs KL sont moyennées sur toutes les orientations faisant partie d'un même quadrant : Est, Ouest, Nord, Sud.

de détail d'une décomposition à différentes échelles¹⁰, on note immédiatement une ressemblance. On constate en particulier une persistance des détails marquants entre les échelles, notamment des contours et des régions homogènes. Les contours, correspondant à des singularités de l'image, se manifestent par des coefficients élevés¹¹ qui se propagent à travers les échelles en s'atténuant. Les régions homogènes sont, elles, régulières et se traduisent par de faibles coefficients à toutes les échelles. De nombreuses applications en traitement d'images tirent profit de cette propriété de persistance entre échelles, notamment en compression ou en débruitage.

2.2.3 Conséquence 3 : les images comme fonctions ou distributions ?

Une autre conséquence de l'invariance d'échelle (approximative) est que les images sur un domaine continu, si elles ont une résolution infinie, ne sont pas des fonctions mais doivent être considérées comme des distributions au sens de Schwartz¹². L'idée étant que l'invariance d'échelle implique automatiquement des oscillations partout d'amplitudes comparables et les fonctions mesurables ne peuvent être aussi complexes. Cela signifie qu'à mesure que leur résolution augmente, les images n'ont pas une valeur numérique limite définie $x(t_1, t_2)$ pour presque tous les points (t_1, t_2) du domaine. Ce résultat est établi dans [162].

La conclusion de ce théorème peut prêter à controverse. En effet, d'un côté, ce résultat stipule qu'une image idéale d'une résolution infinie ne peut être une fonction. Elle est par exemple porteuse de tous les détails, a son énergie qui peut diverger à la limite, et par conséquent n'est pas dans L_2 . Toutefois, les

10. La transformée en ondelettes ici, mais la propriété que nous voulons mettre en évidence n'est pas propre à une transformée multiéchelle particulière.

11. Ceci est d'autant plus vrai que l'ondelette a un nombre suffisant de moments nuls.

12. En fait, il n'existe pas de mesure de probabilité invariante par changement d'échelle à support dans le sous-espace des fonctions mesurables (modulo des constantes) [162].

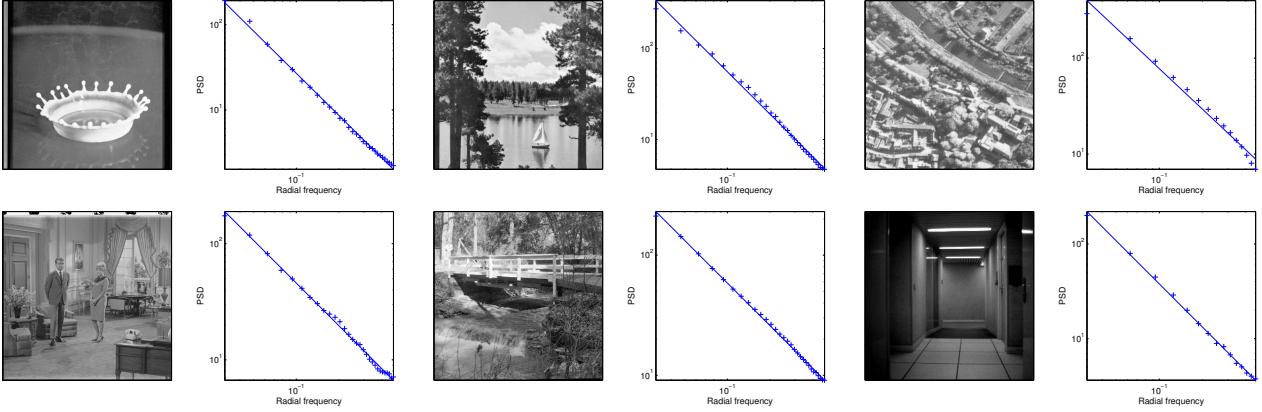


FIGURE 2.6 – Exemples de densité spectrale d’images tirées de la base [1]. Symbole ‘+’ : PSD observée en fonction de la fréquence radiale (échelle log-log). Trait plein : loi $1/\rho^\gamma$ ajustée.

systèmes d’acquisition ne peuvent résoudre que certains détails de ces images. En effet, une image acquise par un capteur perd une partie de ses détails à cause de la fonction d’étalement du point (PSF). D’un autre côté, plusieurs auteurs ont proposé d’autres espaces fonctionnels où les images sont sensées vivre. Par exemple, Mumford-Shah [163] ont proposé d’écrire les images observées comme :

$$x = u + \varepsilon, \quad (2.7)$$

où u est une composante régulière par morceaux, représentant la partie esquisse ("cartoon") de l’image, et ε un bruit de variance finie. Ceci a mené naturellement à considérer l’espace BV (de fonctions à variation bornée) comme l’espace des images (sans bruit). Toutefois, dans ce modèle, la texture est mal prise en compte puisqu’elle est indument absorbée par le bruit. Les approches plus récentes considèrent des modèles plus élaborés du type :

$$x = u + v + \varepsilon, \quad (2.8)$$

où u est la partie cartoon vivant dans BV, ε un bruit L_2 et v la texture. Remarquons au passage que ce modèle génératif linéaire fait écho à celui de (1.7) page 13. Depuis le livre d’Y. Meyer [157], plusieurs auteurs se sont intéressés à ce modèle et ont proposé des algorithmes pour calculer cette décomposition. Nous y reviendrons plus tard dans le Chapitre 6. Pour modéliser la texture, les choix typiques sont l’espace G de Meyer [157], l’espace de Sobolev négatif [17] ou certains espaces de Hilbert adaptés pour les textures localement oscillantes [190, 191, 16].

2.3 Epilogue

Existe-t-il un modèle stochastique simple pour les images qui incorpore les propriétés statistiques déclinées plus haut, à savoir la leptokurticité et l’invariance d’échelle ? La réponse à cette question que nous allons développer ci-après apporte un éclairage édifiant qui justifie élégamment le modèle génératif linéaire décrit en (1.3).

2.3.1 Modèle génératif sans occlusion

À notre connaissance, le modèle le plus simple avec ces deux propriétés statistiques semble être celui proposé par Gidas et Mumford [162], qu’ils ont nommé "le modèle des séries d’ondelettes aléatoires"¹³.

¹³. Notons que d’autres modèles capturant ces propriétés ont été proposés dans la littérature en particulier les cascades multiplicatives infiniment divisibles [46].

Dans ce modèle, une image aléatoire est une somme finie :

$$x(t_1, t_2) = \sum_{i=1}^T \varphi_i(e^{s_i} t_1 + a_i, e^{s_i} t_2 + b_i) , \quad (2.9)$$

où (a_i, b_i, s_i) est un processus de Poisson uniforme dans le groupe de transformations affines et les φ_i — la notation est volontairement similaire à celle de (1.3) afin de mettre en évidence le lien fort entre (1.3) et (2.9)— sont des réalisations indépendantes du processus auxiliaire de Lévy, dont la distribution est définie sur l'espace des composants ou atomes élémentaires d'une image normalisés en échelle et en position. En clair, il faut penser à ces composants élémentaires normalisés comme des atomes prototypes qui seront translatés et mis à l'échelle. Les auteurs dans [162] ont montré que cette série converge presque sûrement dans tous les espaces de Sobolev $W_2^{-\kappa}$, $\kappa > 0$.

2.3.2 Atomes élémentaires

L'ingrédient qui reste à développer sur ce modèle aléatoire est le vocabulaire relatif aux atomes élémentaires φ_i . Dès les années 1980, Julesz et Marr ont proposé certains constituants élémentaires. B. Julesz en se basant sur des études psychovisuelles, a proposé les textons comme composants élémentaires des textures [126]. Suite aux nombreux travaux de D. Hubel et T. Wiesel sur le cortex visuel primaire, la proposition de D. Marr [153] pour ces constituants fut d'extraire d'une image son esquisse primaire (primal sketch) qui consiste en des contours, des barres et des blobs.

Dans la section 1.1.4, nous avons adopté le point de vue de l'école de l'analyse harmonique computationnelle, où les atomes φ_i forment un dictionnaire fixe global. Ainsi, les atomes $\{\varphi_i\}_{1 \leq i \leq T}$ peuvent être des ondelettes, des ridgelets, des curvelets, des wave-atoms, des edgelets, des wedgelets, des bandlets, etc. Mais on pourrait bien évidemment aller plus loin en apprenant les atomes directement à partir d'un corpus d'images. C'est cette voie qui a été empruntée d'abord par Olshausen and Fields [169] qui ont montré qu'un apprentissage avec un codage parcimonieux à partir de patchs extraits d'images naturelles, révèle des atomes qui ressemblent à des ondelettes orientées. Leurs traces ont été ensuite suivies par plusieurs auteurs dont [94, 142, 134, 7]. Dans [A73, S3], nous avons proposé un algorithme réalisant conjointement l'apprentissage de dictionnaires et la décomposition parcimonieuse d'image. Nous y reviendrons au Chapitre 6.

2.3.3 Et l'occlusion ?

Il est à noter cependant que le modèle génératif stochastique (2.9) décrit ci-dessus ne tient pas compte d'un aspect pourtant primordial dans les images : les occlusions. En effet, ce modèle génératif linéaire (et donc (1.3)) est une simplification de la réalité. Suite au processus d'acquisition d'un monde 3D sur un capteur 2D, les objets sont ordonnés par leur distance au capteur et doivent donc être combinés par une opération non linéaire où les objets proches masquent les plus distants. Lee et al. [138] ont proposé une adaptation du modèle des "feuilles mortes" ("dead leaves model") [185], qui est similaire au modèle génératif (2.9) mais incluant des occlusions. L'analyse de ce modèle non linéaire s'avère beaucoup plus complexe et nous ne le développerons pas davantage dans ce manuscrit.

2.4 Récapitulatif et conclusion

Dans ce chapitre, nous avons mis en exergue l'importance de la modélisation stochastique pour établir des modèles d'a priori sur les images. En partant des propriétés de leptokurticité et d'invariance d'échelle, nous avons mis en évidence les liens étroits entre le modèle génératif linéaire parcimonieux (1.3) traité au chapitre précédent et le modèle des séries aléatoires (2.9) satisfaisant ces deux propriétés. En l'occurrence, il s'avère que (2.9) est le modèle le plus simple qui satisfait ces deux propriétés. Toutefois, en deux dimensions, l'occlusion reste une limite de ce type de processus génératif linéaire. Pour y pallier, il sera nécessaire à l'avenir de développer des modèles stochastiques directement en 3D s'appuyant par exemple sur les processus de branchement.

Chapitre 3

Optimisation convexe non lisse

Ce chapitre traitera des outils d'analyse convexe qui forment la matière première de notre cadre pour l'optimisation de fonctionnelles convexes mais non lisses. Les travaux les plus concernés par ce chapitre sont une partie de la thèse de F.-X. Dupé [TH1], ainsi que [A1, Chapitre 7][A12, A48, A63, A111, S1, S2].

3.1 Quelques définitions

Dans ce chapitre, nous nous plaçons dans un espace de Hilbert réel $\mathcal{H}$ typiquement de dimension finie, le plus souvent $\mathbb{R}^N$, même si un certain nombre de résultats peuvent être étendus au cas complexe et/ou en dimension infinie¹⁴. Nous notons $\langle \cdot, \cdot \rangle$ le produit scalaire dans $\mathcal{H}$ et $\|\cdot\|$ la norme associée. $\mathbf{I}$ représente l'opérateur identité de $\mathcal{H}$. La norme (induite) d'un opérateur $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{H}'$ est définie par $\|\mathbf{A}\| = \sup_{x \in \mathcal{H}} \frac{\|\mathbf{A}x\|}{\|x\|}$. $\mathcal{B}_\rho^\circ$ est la boule ℓ_p fermée de rayon $\rho > 0$.

Une fonction réelle $F : \mathcal{H} \rightarrow (-\infty, +\infty]$ est (faiblement-)coercive si $\lim_{\|x\| \rightarrow +\infty} F(x) = +\infty$. Le domaine de F est défini par $\text{dom } F = \{x \in \mathcal{H} : F(x) < +\infty\}$. Une fonction à valeurs réelles F est semi-continue inférieurement (lsc) si $\liminf_{x \rightarrow x_0} F(x) \geq F(x_0)$. $\Gamma_0(\mathcal{H})$ est la classe des fonctions convexes lsc et à domaine non vide.

La conjuguée (transformée de Legendre-Fenchel) d'une fonction $F \in \Gamma_0(\mathcal{H})$ est la fonction fermée F^* définie par

$$F^*(u) = \sup_{x \in \text{dom } F} \langle u, x \rangle - F(x) , \quad (3.1)$$

$\text{dom}(F^*)$ est l'ensemble des pentes des fonctions affines minorant F . F^* peut aussi être interprétée comme étant la transformée de Fourier au sens de l'algèbre $(\max, +)$, c.à.d. la conjuguée d'une inf-convolution est la somme des conjuguées et inversement [140], où l'inf-convolution de deux fonctions F_1 et $F_2 \in \Gamma_0(\mathcal{H})$ est

$$(F_1 \overset{+}{\vee} F_2)(x) = \inf_{y \in \mathcal{H}} F_1(y) + F_2(x - y) . \quad (3.2)$$

Pour un convexe non vide $\mathcal{C}$ de $\mathcal{H}$, son indicatrice $\iota_{\mathcal{C}}$ est

$$\iota_{\mathcal{C}}(x) = \begin{cases} 0, & \text{si } x \in \mathcal{C} , \\ +\infty, & \text{sinon.} \end{cases} \quad (3.3)$$

En particulier, toute fonction 1-homogène est la conjuguée de l'indicatrice d'un convexe fermé. Par exemple, la conjuguée d'une norme est l'indicatrice de la boule unité de sa norme duale.

Pour un convexe $\mathcal{C} \subseteq \mathcal{H}$, on note $\text{int}(\mathcal{C})$ son intérieur, L'intérieur relatif $\text{ri}(\mathcal{C})$ est défini comme l'intérieur de $\mathcal{C}$ dans la topologie relative à son enveloppe affine.

¹⁴. Lors de la généralisation en dimension infinie, la convergence des itérées dans les résultats décrits dans le reste du chapitre n'est pas toujours forte, loin s'en faut.

3.2 Opérateur et calcul proximal

L'opérateur proximal est une généralisation de la notion de projection sur un ensemble convexe fermé et forme le point de départ des schémas d'optimisation exposés ultérieurement.

Définition 4 ([158]) Soit $F \in \Gamma_0(\mathcal{H})$. Pour tout $x \in \mathcal{H}$, la fonction $\inf_{y \in \mathcal{H}} F(y) + \|x - y\|^2 / 2$ atteint son infimum en un point unique $\text{prox}_F(x)$. L'opérateur à valeur unique $\text{prox}_F : \mathcal{H} \rightarrow \mathcal{H}$ ainsi défini est l'opérateur proximal de F .

On peut remarquer que l'opérateur proximal de F est aussi l'inf-convolution de celle-ci par $x \mapsto \|x\|^2 / 2$. On retrouve le projecteur sur un convexe fermé $\mathcal{C}$ lorsque $F = \iota_{\mathcal{C}}$. Par la suite, nous utiliserons aussi l'opérateur de réflexion, défini par $\text{rprox}_F = 2 \text{prox}_F - \text{I}$.

Définition 5 L'enveloppe de Moreau d'ordre $\rho \in (0, +\infty)$ d'une fonction $F \in \Gamma_0(\mathcal{H})$ est la fonction convexe continue :

$${}^\rho F : \mathcal{H} \rightarrow \mathbb{R} : x \mapsto \inf_{y \in \mathcal{H}} F(y) + \frac{1}{2\rho} \|x - y\|^2 = \left(\frac{1}{2\rho} \|\cdot\|^2 \overset{+}{\vee} F \right) (x). \quad (3.4)$$

L'enveloppe de Moreau d'ordre ρ est une fonction Fréchet-différentiable à gradient $1/\rho$ -Lipschitz [160]. Du fait de la C^1 -continuité de ${}^\rho F$, l'enveloppe de Moreau s'appelle aussi la régularisation de Moreau-Yosida.

L'opérateur proximal dispose de plusieurs règles de calcul dont les plus simples sont déjà dans les premiers travaux de Moreau [159, 160]. D'autres résultats de calcul proximal qui ont été essentiels pour établir les résultats exposés dans la section 3.4 relient l'opérateur proximal de F^* et de ${}^\rho F$ à celui de F (voir [S1, A63] pour plus de détails et de références).

3.3 Décomposition d'opérateurs monotones maximaux

Un opérateur $\mathcal{A} : \mathcal{H} \rightarrow 2^{\mathcal{H}}$, est dit monotone si et seulement si

$$\forall (x, u), (y, v) \in \mathcal{H} \times \mathcal{H} \text{ tels que } u \in \mathcal{A}(x), v \in \mathcal{A}(y), \quad \langle x - y, u - v \rangle \geq 0.$$

Un opérateur monotone est maximal si en tant que graphe dans $\mathcal{H} \times \mathcal{H}$, il n'est pas inclus dans tout autre opérateur monotone.

La sous-différentielle $\partial F : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ d'une fonction $F \in \Gamma_0(\mathcal{H})$ est un opérateur monotone maximal. De plus, pour un opérateur maximal monotone $\mathcal{A}$, il existe un unique point $J_{\rho\mathcal{A}}(x)$ tel que

$$x - J_{\rho\mathcal{A}}(x) \in \mathcal{A}(J_{\rho\mathcal{A}}(x)),$$

où $J_{\rho\mathcal{A}} = (\text{I} + \rho\mathcal{A})^{-1}$ est la résolvante, un opérateur à valeur unique et de domaine plein [87]. Il clair que d'après la Définition 4, on a $x - \text{prox}_F(x) \in \partial F(\text{prox}_F(x))$, et ainsi la résolvante de la sous-différentielle n'est autre que l'opérateur proximal de F .

L'objectif dans bon nombre de problèmes d'optimisation que nous avons traités est de résoudre

$$(P) : \quad \min_{x \in \mathcal{H}} \sum_{l=1}^L F_l(x), \quad (3.5)$$

où $\forall l, F_l \in \Gamma_0(\mathcal{H})$ et $\cap_l \text{dom}(F_l) \neq \emptyset$. Soit $F(x) = \sum_{l=1}^L F_l(x)$ et $\mathcal{G}$ l'ensemble des solutions de (P), c.à.d. $\mathcal{G} = \{x \in \mathcal{H} | 0 \in \partial F(x)\}$. De la condition nécessaire et suffisante du minimum d'une fonction convexe, $x \in \mathcal{G}$ si et seulement si,

$$0 \in \partial F(x) \Leftrightarrow x = \text{prox}_{\beta F}(x). \quad (3.6)$$

où $\beta > 0$ est le pas proximal. L'itération proximale (et variantes) construite à partir de cette équation du point fixe est

$$x^{(t+1)} = \text{prox}_{\beta F}(x^{(t)}).$$

C'est un algorithme fondamental pour trouver un minimiseur global de F . Dans [154], il est montré que $x^{(t)}$ converge¹⁵ vers $x \in \mathcal{G}$. La principale difficulté qui subsiste pour implémenter une telle itération est que l'opérateur proximal de F est en général difficile à calculer sous forme analytique. D'un autre côté, les expressions analytiques des opérateurs proximaux des fonctions F_l peuvent être facilement accessibles (on parlera alors de fonctions simples selon la terminologie de [166]). Cette situation se produit typiquement dans les problèmes inverses en traitement du signal et des images.

3.3.1 Origines de la décomposition

Les méthodes de décomposition pour résoudre (P) sont des algorithmes itératifs qui tirent profit du fait que F_l sont des fonctions simples, en remplaçant l'évaluation explicite de $\text{prox}_{\beta F}$ par une séquence de calculs faisant intervenir individuellement les opérateurs proximaux des F_l . Les méthodes de décomposition des opérateurs monotones, qui trouvent leurs origines en analyse numérique, ont fait l'objet d'une intense activité de recherche notamment pour construire des algorithmes de décomposition en optimisation convexe et pour les inégalités variationnelles.

Une littérature importante est disponible sur le sujet surtout pour le cas $L = 2$ (voir [86] pour une revue). Le reste de cette section traite le cas $L = 2$. Ces algorithmes de décomposition peuvent être, classées en trois familles principales : explicite-implicite [102, 197], Douglas/Peaceman-Rachford [144] (bien que les racines remontent aux méthodes numériques pour la résolution des EDPs [84]), et le schéma implicite-implicite [143, 171]. Récemment dans [53, 88], un cadre théorique général et pertinent pour les trois classes d'algorithmes est décrit.

3.3.2 Schéma de décomposition implicite-implicite

L'itération implicite-implicite se décline comme suit :

$$x^{(t+1)} = \text{prox}_{\mu_t F_1} \left(\text{prox}_{\mu_t F_2} (x^{(t)}) \right), \quad (3.7)$$

pour une suite appropriée $(\mu_t)_{t \in \mathbb{N}}$. Lorsque F_1 et F_2 sont les indicatrices de deux convexes fermés, on retrouve la fameuse projection alternée sur les convexes très utilisée en traitement d'images. En dépit de sa simplicité, et à quelques exceptions près, cette itération présente le défaut que son ensemble de points fixes n'est pas toujours identique à $\mathcal{G}$. En effet, le schéma implicite-implicite peut être utilisé pour trouver des racines communes d'opérateurs monotones maximaux—c.à.d. éléments de $(\partial F_1)^{-1}(0) \cap (\partial F_2)^{-1}(0)$ si celui-ci est non vide—ou des problèmes mettant en jeu des enveloppes de Moreau. Un exemple de ce dernier type de problèmes apparaît dans la décomposition $u + v$ d'image tel que proposé par [15], ou encore celui traité dans la section 3.4.2.

Toutefois, en raison de ces subtilités de convergence, l'itération implicite-implicite est assez peu utilisée pour résoudre des problèmes d'optimisation généraux tels que (P).

3.3.3 Schéma de décomposition explicite-implicite

L'itération explicite-implicite (FB pour forward-backward) est essentiellement une généralisation du gradient projeté pour les problèmes avec contraintes, et de ce fait, hérite de ses restrictions. En particulier, il est nécessaire que l'une des fonctions (par exemple F_2) soit à gradient κ -Lipschitz.

3.3.3.1 Algorithme à un pas

Le schéma FB à un pas ne fait intervenir que l'itération précédente dans la descente. Sa forme simplifiée s'écrit

$$x^{(t+1)} = \text{prox}_{\mu_t F_1} \left(x^{(t)} - \mu_t \nabla F_2(x^{(t)}) \right), \quad (3.8)$$

où $0 < \inf_t \mu_t \leq \sup_t \mu_t < 2/\kappa$ pour que l'itération converge vers $x^* \in \mathcal{G} \neq \emptyset$. Lorsque $F_1(x) = \|x\|_p^p$, $p \geq 1$ et $F_2(x) = \frac{1}{2} \|y - \mathbf{A}x\|^2$, (3.8) a été redécouverte plusieurs fois [59, 96]. Les auteurs dans [54] ont été les

15. Pour être précis, la convergence est faible en dimension infinie.

premiers à proposer récemment le cadre élégant du FB en traitement d'images. Depuis, plusieurs auteurs l'ont mis à profit dans divers problèmes inverses, voir la revue détaillée de la littérature dans [A1, TH1] ainsi que le Chapitre 5 pour plus de détails sur nos contributions sur ce sujet.

Dans certaines situations, la constante de Lipschitz κ n'est pas connue ou mal estimée, rendant délicat le choix du pas μ_t . Pour contourner cette difficulté, l'auteur dans [198] propose une modification du schéma FB à un pas permettant d'estimer le pas de descente à chaque itération en utilisant la règle d'Armijo-Goldstein et en ajoutant une étape d'extra-gradient pour renforcer la robustesse de l'algorithme.

3.3.3.2 Algorithme à pas multiple

Le schémas FB à un pas mémorise uniquement l'itérée précédente pour ensuite mettre à jour la nouvelle itérée. Or, la succession des gradients et des itérées forment une information importante sur la variation de la fonction. Celle-ci peut être mise à profit pour obtenir un pas de descente mieux adapté. Ainsi, Nesterov propose dans [166] une version du schéma explicite-implicite à pas multiple utilisant la totalité des itérées précédentes. Sa description détaillée faisant intervenir l'opérateur proximal ainsi de nombreuses applications en traitement d'images sont données dans [204, TH1]. De la même façon que pour l'algorithme FB à un pas, le pas de descente peut aussi être adaptativement estimé dans l'algorithme à pas multiple de Nesterov. Très récemment, les auteurs dans [22] proposent un autre schéma à pas multiple utilisant uniquement les deux dernières itérées. Sa vitesse de convergence sur l'objectif F dans (P) (non pas sur les itérées) est la même que celle de l'algorithme de Nesterov, à savoir $O(1/t^2)$, alors que celle du FB à un pas n'est qu'en $O(1/t)$ [166, 30].

Bien que l'itération explicite-implicite (à un pas ou à pas multiple) ne nécessite que le calcul de l'opérateur proximal de F_1 , elle ne s'applique pas lorsque les deux fonctions sont non différentiables.

3.3.4 Schéma de décomposition de Douglas-Rachford

Le schéma de Douglas/Peaceman-Rachford forme la famille la plus générale des méthodes de décomposition d'opérateurs monotones maximaux avec des conditions d'utilisation plus faibles que précédemment. Pour l'appliquer, F_1, F_2 sont simples mais pas nécessairement différentiables. Pour $\gamma > 0$ une suite $\mu_t \in (0, 2)$, cette classe d'itération s'écrit :

$$x^{(t+1)} = \left(1 - \frac{\mu_t}{2}\right) x^{(t)} + \frac{\mu_t}{2} \text{rprox}_{\gamma F_1} \left(\text{rprox}_{\gamma F_2}(x^{(t)}) \right), \quad (3.9)$$

On peut remonter à [144] pour l'itération de Douglas-Rachford (DR) originale qui correspond à $\mu_t \equiv 1$. Sa forme la plus générale est due à [53] où sa convergence est établie. Plus précisément, supposons que $\mathcal{G} \neq \emptyset$ et $0 \in \text{ri}(\text{dom}(F_1)) - \text{ri}(\text{dom}(F_2))$ ¹⁶. Soit $(\mu_t)_{t \in \mathbb{N}}$ une suite dans $(0, 2)$ telle que $\sum_{t \in \mathbb{N}} \mu_t(2 - \mu_t) = +\infty$, et $x^{(0)} \in \mathcal{H}$. Alors $x^{(t)}$ converge vers x^* et $x^* = \text{prox}_{\gamma F_2}(\bar{x})$ est un minimiseur global de F .

Le schéma DR a été très récemment mis au exergue en traitement d'images par [57]. Nous l'avons aussi exploité pour résoudre plusieurs problèmes en débruitage, déconvolution et échantillonnage compressé [A12, A10, A62, A63, A58, A1, S1, TH1] (voir aussi les chapitres 4-5).

A notre connaissance, aucun résultat général sur la vitesse de convergence de l'itération de DR n'est connu aussi bien sur l'objectif F que sur les itérées. Dans le cas de la convexité forte de F_1 ou F_2 , les auteurs de [48] ont montré une vitesse linéaire sur les itérées.

3.3.4.1 Robustesse aux erreurs

L'une des propriétés saillantes des schémas FB à un pas (3.8) et DR (3.9) est leur robustesse aux erreurs sur le calcul de l'opérateur proximal de F_1 et F_2 ou le gradient de F_2 . Cette assertion de robustesse a été établie dans [53]. En substance, elle stipule qu'il suffit que les erreurs commises soient de normes sommables pour continuer à clamer la convergence. Il est important toutefois de noter qu'il n'existe pas de tel résultat pour l'itération à pas multiple de Nesterov.

En particulier, il arrive que l'un des opérateurs proximaux dans le schéma DR, ou l'opérateur proximal et/ou le gradient dans FB ne puissent pas être calculés exactement (e.g. fonctions non simples), mais

16. Cette condition de qualification sur les domaines est différente en dimension infinie [57].

plutôt par le biais d'une itération intérieure. Plusieurs exemples sont décrits dans ce chapitre ainsi que dans [A1, TH1] et les Chapitres 5 et 7. La propriété de robustesse permet d'affirmer la convergence même si l'itération intérieure est appliquée un nombre fini d'itérations. Dans [A48, TH1] nous avons donné une borne sur le nombre d'itérations intérieures suffisantes pour la robustesse lorsque la vitesse de convergence sur les itérées intérieures est connue.

3.4 Premières applications

3.4.1 Opérateur proximal de $F \circ \mathbf{A}$

Dans plusieurs situations, on est amené à devoir calculer l'opérateur proximal d'une fonction simple $F \in \Gamma_0(\mathcal{H})$ composée par un opérateur affine borné $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{H}', x \mapsto \mathbf{F}x - y, y \in \mathcal{H}'$. C'est typiquement le cas de l'a priori analyse que l'on verra au Chapitre 5 ou encore de la TV discrète. Le résultat que nous allons énoncé sous peu permet d'exprimer cet opérateur proximal à partir de celui de F . Ce résultat est précieux dans plusieurs algorithmes discutés dans le Chapitre 5. Bien évidemment, si $\mathbf{F}$ est unitaire, il est facile de voir que :

$$\text{prox}_{F \circ \mathbf{A}}(x) = y + \mathbf{F}^* \text{prox}_F(\mathbf{F}x - y) . \quad (3.10)$$

Dans tous les autres cas, nous avons montré dans [S1] (voir aussi [TH1]) le résultat suivant.

Théorème 2 *Soit $\mathbf{F}$ un opérateur linéaire borné tel que $\text{ri}(\text{dom}(F) \cap \text{Im}(\mathbf{A})) \neq \emptyset$. Alors $F \circ \mathbf{A} \in \Gamma_0(\mathcal{H})$ et de plus :*

(i) *Si $\mathbf{F}$ est une trame ajustée, alors*

$$\text{prox}_{F \circ \mathbf{A}}(x) = x + c^{-1} \mathbf{F}^* (\text{prox}_{cF} - \mathbf{I}) (\mathbf{F}x - y) . \quad (3.11)$$

(ii) *Si $\mathbf{F}$ est une trame générale : soit $\mu_t \in (0, 2/c_2)$, et définissons la récurrence suivante*

$$\begin{aligned} u^{(t+1)} &= \mu_t \left(\mathbf{I} - \text{prox}_{\mu_t^{-1}f} \right) \left(\mu_t^{-1} u^{(t)} + \mathbf{A}p^{(t)} \right) , \\ p^{(t+1)} &= x - \mathbf{F}^* u^{(t+1)} . \end{aligned} \quad (3.12)$$

Alors $u^{(t)}$ converge vers u^ et $p^{(t)}$ vers $\text{prox}_{F \circ \mathbf{A}}(x) = x - \mathbf{F}^* u^*$. Plus précisément, les suites $u^{(t)}$ et $p^{(t)}$ convergent linéairement et la meilleure vitesse de convergence est atteinte pour $\mu_t \equiv 2/(c_1 + c_2)$*

$$\left\| p^{(t)} - \text{prox}_{F \circ \mathbf{A}}(x) \right\| \leq \sqrt{\frac{c_2}{c_1}} \left(\frac{c_2 - c_1}{c_2 + c_1} \right)^t \left\| p^{(0)} - \text{prox}_{F \circ \mathbf{A}}(x) \right\| . \quad (3.13)$$

(iii) *Si $c_1 = 0$, appliquer (3.12) avec $\mu_t \in (0, 2/c_2)$. Dans ce cas, $p^{(t)}$ converge vers $\text{prox}_{F \circ \mathbf{A}}(x) = x - \mathbf{F}^* u^*$ à la vitesse $O(1/t)$.*

La preuve est basée sur la dualité de Fenchel-Rockafellar. Pour le cas où $\mathbf{F}$ est une trame ajustée, une démonstration différente a été proposée par [57]. Ce résultat peut être généralisé en dimension infinie avec quelques ajustements.

La vitesse du cas (ii) est une conséquence de la forte convexité de l'objectif dual et de la continuité de l'opérateur proximal. Pour le cas (iii), la vitesse obtenue est la conséquence du lemme général suivant dont une preuve est donnée dans [S2] :

Lemme 1 *Considérons le problème primal $\min_{x \in \mathcal{H}} F(x) = F_1(x) + F_2 \circ \mathbf{F}(x)$ où $F_1, F_2 \in \Gamma_0(\mathcal{H})$ avec F_1 fortement convexe, $\mathbf{F} : \mathcal{H} \rightarrow \mathcal{H}'$ est linéaire borné et $\text{ri}(\text{dom}(F_2) \cap \mathbf{F} \text{dom}(F_1)) \neq \emptyset$. Posons $u^{(t)}$ les itérées d'un algorithme convergeant qui résout le problème dual $\min_{u \in \mathcal{H}'} J(u) = F_1^* \circ \mathbf{F}^*(u) + F_2^*(u)$, et $x^{(t)}$ les itérées du problème primal. Si F_1 est à gradient Lipschitz ou si le domaine de F_2^* est borné, alors*

$$\left\| x^{(t)} - x^* \right\|^2 \leq C \left(J(u^{(t)}) - J(u^*) \right) . \quad (3.14)$$

En clair, ce lemme relie la vitesse de convergence des itérées primales à celle sur l'objectif dual. Ainsi, si comme dans (3.12), le problème dual est résolu par le schéma implicite-explicite dont la vitesse de convergence sur l'objectif est en $O(1/t)$, la même vitesse se répercute sur les itérées primales.

Au lieu du schéma implicite-explicite (3.12) à un pas, nous pouvons utiliser un algorithme de premier ordre à pas multiple. Cette modification de (3.12) est décrite dans [S1, TH1]. Il est par ailleurs montré dans [S1] que cette modification jouit d'une vitesse de convergence linéaire avec une meilleure constante que (3.13) lorsque $\mathbf{F}$ est une trame, et en $O(1/t^2)$ lorsque $c_1 = 0$, ce qui est bien mieux que la vitesse atteinte par (3.12). Ces vitesses de convergence apparaissent aussi dans [A63, A110, S2] pour la convergence d'un algorithme de projection sur la boule TV discrète.

3.4.2 Opérateur proximal de $F_1 + F_2 \circ \mathbf{A}$

Dans la thèse [TH1], le résultat ci-dessus a été généralisé dans un certain sens. En effet, dans plusieurs problèmes inverses, la régularisation peut regrouper deux termes. Les exemples sont nombreux dont typiquement des problèmes impliquant un terme de parcimonie sur les coefficients, et un second sur la positivité de l'image ; voir par exemple la section 5.3.

Soient F_1 et $F_2 \in \Gamma_0(\mathcal{H})$, $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{H}'$ est un opérateur linéaire, et posons $G = F_1 + F_2 \circ \mathbf{A}$. Supposons que les hypothèses suivantes sont vérifiées :

- (i) $\text{Im}(\mathbf{A}) \neq \emptyset$.
- (ii) $0 \in \text{ri}(\text{dom}(F_2) - \mathbf{A} \text{dom}(F_1))$.
- (iii) Les fonctions F_1 et F_2 sont simples, c.à.d. les opérateurs proximaux associés peuvent être calculés analytiquement.

Méthode 1 Une première approche pour le calcul de l'opérateur proximal de G est de se ramener au cas de la section 3.4.1. En effet, si on définit la fonction convexe à domaine non vide,

$$F : \mathcal{H} \times \mathcal{H}' \rightarrow \mathbb{R} \cup \{\infty\}, \quad z = \begin{pmatrix} v \\ w \end{pmatrix} \mapsto F_1(v) + F_2(w) .$$

Soit $\mathbf{B} = \begin{bmatrix} \mathbf{I} \\ \mathbf{A} \end{bmatrix} : \mathcal{H} \rightarrow \mathcal{H} \times \mathcal{H}'$. Alors on peut écrire le problème de minimisation proximale comme :

$$\inf_{y \in \mathcal{H}} \frac{1}{2} \|y - x\|^2 + F_1(y) + F_2 \circ \mathbf{A}(y) = \inf_{y \in \mathcal{H}} \frac{1}{2} \|y - x\|^2 + F \circ \mathbf{B}(y) . \quad (3.15)$$

Il est clair que cette dernière écriture a la forme se prêtant au cas (iii) du Lemme 2, en remarquant que la condition de qualification sur le domaine est vérifiée grâce à l'hypothèse (ii). Il suffit finalement de calculer la conjuguée de F qui est de surcroît séparable en v et w (l'opérateur proximal associé l'est aussi ; F est donc simple puisque F_1 et F_2 le sont par l'hypothèse (iii)). On peut alors appliquer le schéma implicite-explicite (3.12) (ou la version à pas multiple), et ainsi bénéficier de la vitesse de convergence qui y est énoncée (c.à.d. $O(1/t)$ ou $O(1/t^2)$ suivant le schéma utilisé). Cependant, le pas de descente sera moins favorable puisqu'il devra être borné supérieurement par $2 / \left(1 + \|\mathbf{A}\|^2\right)$. Cette première méthode s'étend facilement au cas où F_1 est elle aussi composée par un opérateur linéaire borné, voire même lorsque la somme porte sur plus de deux fonctions, pourvu que la condition de qualification sur le domaine soit vérifiée.

Méthode 2 En utilisant encore une fois la dualité de Fenchel-Rockafellar et le calcul proximal, nous avons :

Lemme 2 *Sous les hypothèses (i)-(ii), nous avons l'équivalence primale-duale :*

$$\inf_{y \in \mathcal{H}} \frac{1}{2} \|y - x\|^2 + F_1(y) + F_2 \circ \mathbf{A}(y) = - \min_{u \in \mathcal{H}'} {}^1 F_1^*(-\mathbf{A}^* u + x) + F_2^*(u) . \quad (3.16)$$

De plus, la solution primale $y^* = \text{prox}_F(x)$ est reliée à la solution duale u^* par

$$y^* = \text{prox}_{F_1}(-\mathbf{A}^*u^* + x) . \quad (3.17)$$

A partir de ce lemme et le schéma implicite-explicite, nous arrivons au résultat suivant :

Théorème 3 Soit $0 < \inf_t \mu_t \leq \mu_t \leq \sup_t \mu_t < 2/\|\mathbf{A}\|^2$. Soit la suite suivante :

$$u^{(t+1)} = \mu_t \left(\mathbf{I} - \text{prox}_{\mu_t^{-1}F_2} \right) \left(\mu_t^{-1}u^{(t)} + \mathbf{A} \circ \text{prox}_{F_1}(-\mathbf{A}^*u^{(t)} + x) \right) . \quad (3.18)$$

Alors, $u^{(t)}$ converge vers u^* et $y^{(t)} = \text{prox}_{F_1}(-\mathbf{A}^*u^{(t)} + x)$ vers $\text{prox}_F(x)$. De plus,

- (i) si F_2 est à gradient Lipschitz, alors le schéma converge en $O(1/t)$ sur les itérées duales et primales ;
- (ii) si F_1 est à gradient Lipschitz, alors la convergence est en $O(1/t)$ sur les itérées primales. Si de plus $\mathbf{A}$ est surjectif, alors la convergence est linéaire sur les itérées duales et primales ;
- (iii) si $\text{dom}(F_2^*)$ est borné, alors la convergence est en $O(1/t)$ sur les itérées primales.

Remarquons maintenant que le paramètre de descente dans cette méthode est plus favorable que dans la première méthode. De plus, la seconde méthode opère sur des itérées duales vivant dans $\mathcal{H}'$ alors que celles de la première sont dans $\mathcal{H} \times \mathcal{H}'$. Certaines vitesses ci-dessus sont une conséquence du Lemme 1 dont une preuve est dans [S2]. Le schéma implicite-explicite à un pas, peut être remplacé par un algorithme à pas multiple de type Nesterov auquel cas les vitesses exhibées ci-dessus deviennent en $O(1/t^2)$. A noter qu'un travail parallèle mais indépendant de [56] a montré un résultat similaire au nôtre en dimension infinie, mais avec plusieurs différences toutefois, notamment les vitesses de convergence résumées dans le Théorème 3.

3.4.3 Opérateur proximal de pénalités favorisant la parcimonie

Dans les Chapitres 1, 2 et 5, nous avons fait part d'une augmentation sensible du nombre d'algorithmes reposant sur la parcimonie comme a priori sur les images. Il serait donc utile de disposer d'une famille de pénalités promouvant la parcimonie qui soient simples dans le sens où les opérateurs proximaux associés puissent être exprimés analytiquement.

Par exemple, dans nos travaux, nous avons considéré la famille des pénalités additives :

$$\Psi(x) = \sum_{i=1}^T \psi_i(x[i]) ,$$

où on suppose que $\forall 1 \leq i \leq T$:

- (i) $\psi_i \in \Gamma_0(\mathbb{R})$.
- (ii) ψ_i est convexe, paire, positive, et croissante sur $[0, +\infty)$.
- (iii) ψ_i est continue sur $\mathbb{R}$, avec $\psi_i(0) = 0$.
- (iv) ψ_i est dérivable sur $(0, +\infty)$, mais pas nécessairement en 0 et admet une dérivée directionnelle à droite positive $\psi'_{i+}(0) = \lim_{t \downarrow 0} \frac{\psi_i(t)}{t} \geq 0$.

Proposition 1 Sous les hypothèses (i)–(iv), l'opérateur proximal de $\kappa\psi_i$, $\kappa > 0$, a une unique solution continue et impaire découpée en chaque coordonnée i :

$$\tilde{x}[i] = \text{prox}_{\kappa\psi_i}(x[i]) = \begin{cases} 0 & \text{if } |x[i]| \leq \kappa\psi'_{i+}(0) , \\ x[i] - \kappa\psi'_i(\tilde{x}[i]) & \text{if } |x[i]| > \kappa\psi'_{i+}(0) . \end{cases} \quad (3.19)$$

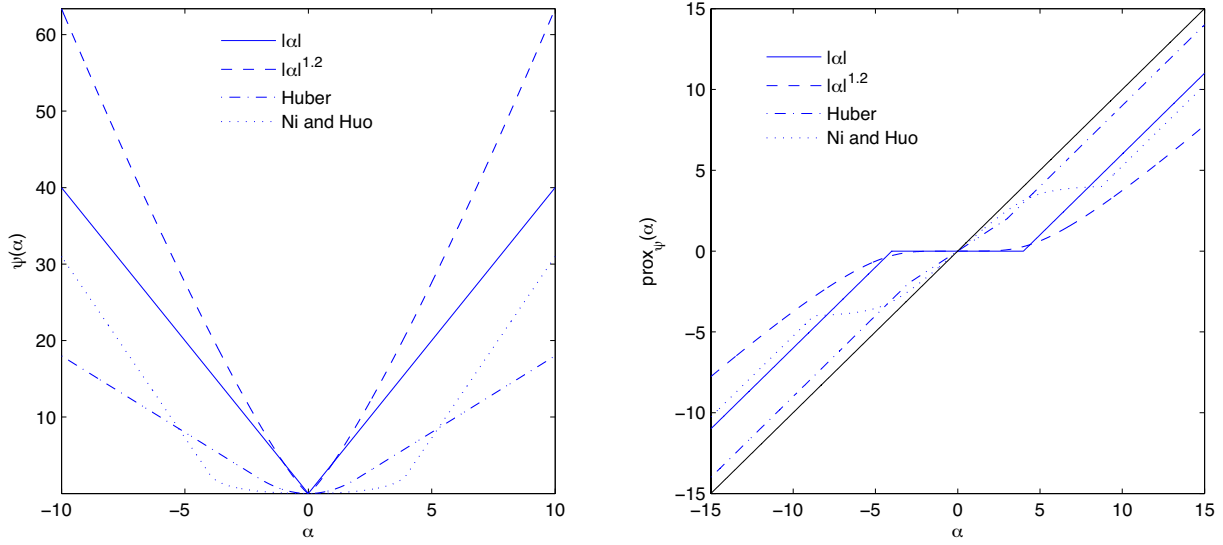


FIGURE 3.1 – Exemples de pénalité de parcimonie ψ_i (gauche) et l'opérateur proximal associé (droite).

Nous avons proposé deux preuves de ce résultat dont une dans [A31] et la seconde se trouve dans [S1]. Cette proposition peut se généraliser au cas des pénalités non convexes mais sous des conditions supplémentaires [A31, A26]. Des résultats similaires à la Proposition 3.19 apparaissent sous différentes formes dans d'autres travaux comme [11, 168, 55].

En clair, l'opérateur proximal dans le cas de ces pénalités revient à opérer un seuillage-contraction terme-à-terme. Pour $\psi_i(x[i]) = \lambda |x[i]|$, on obtient le désormais classique seuillage doux. D'autres pénalités ψ_i rentrent dans le cadre de la Proposition 3.19. A titre d'exemple, nous citons $\psi_i(x[i]) = \lambda |x[i]|^p$, $p \geq 1$, ce qui correspond à un a priori GGD comme expliqué au Chapitre 2, et la pénalité robuste de [167]. Ces exemples de fonction et l'opérateur proximal associé sont illustrés en fig. 3.1. D'autres pénalités intéressantes qui peuvent être considérées sont celles qui rentrent en jeu dans la parcimonie structurée avec des normes mixtes [195, 209]. Leur opérateur proximal peut aussi s'exprimer analytiquement dans de nombreux cas [A1, 133].

3.5 Récapitulatif et conclusion

Dans ce chapitre, les outils issus de l'analyse convexe et la décomposition d'opérateurs monotones maximaux, qui forment la pierre angulaire de nos algorithmes d'optimisation sont décrits. La notion d'opérateur proximal est introduite et des résultats de calcul proximal sont donnés. Ces résultats seront intensément exploités lors de la résolution de divers problèmes d'optimisation abordés dans les Chapitres 4, 5, 6 et 7.

Chapitre 4

Parcimonie et estimation

Dans ce chapitre, l'intérêt de la parcimonie pour l'estimation statistique sera mis en évidence. Et ce aussi bien pour s'attaquer au problème du débruitage en présence de différents types de bruit (additif Gaussien blanc, bruit de Poisson, bruit mixte Poisson-Gaussien et bruit multiplicatif), que pour l'estimation et l'inférence dans les séries temporelles avec le modèle linéaire paramétrique et le modèle partiellement linéaire sémi-paramétrique. Nos contributions concernées par les sujets traités dans ce chapitre sont la seconde partie de la thèse de L. Boubchir [TH4], ainsi que [A1, Chapitre 6] [A10, A15, A19, A20, A27, A31, A30, A36, A46, A62, A67, A73, A69, A76, A81, A87, A83, A89, A90, A100, A112, A113, A114, A117].

4.1 Régression non paramétrique, débruitage

En débruitage, le but est de restaurer x à partir de ses observations bruitées¹⁷ y ,

$$y[i] = x[i] \odot \varepsilon[k] \quad i = 1, \dots, N, \quad (4.1)$$

où ε est le bruit reflétant les fluctuations aléatoires non mesurées contaminant le signal utile, et $\odot$ est toute composition de deux arguments (e.g. $+$ pour un bruit additif, $\times$ pour un bruit multiplicatif, etc.). Ce problème est sans doute l'un des plus anciens à la fois en traitement du signal et de l'image, mais aussi en statistiques. Par ailleurs, les débruiteurs basés sur un seuillage dans le domaine des représentations parcimonieuses ont clairement fait leurs preuves aussi bien en termes de performances pratiques que théoriques [192, 118, 123, 148]. En particulier, il est établi que la qualité de l'estimation par seuillage des coefficients d'une transformée est étroitement liée à celle de la compressibilité du signal d'intérêt (voir section 1.1.3) dans cette transformée [76, 131]. En particulier, pour un bruit centré additif blanc Gaussien (AWGN), c.à.d. $\varepsilon[i] \sim_{\text{iid}} \mathcal{N}(0, \sigma^2)$, et pour un dictionnaire orthogonal Φ , si l'erreur d'approximation non linéaire ℓ_2 de x dans Φ par les K -termes les plus élevés décroît en $O(K^{-\kappa})$, alors l'erreur quadratique moyenne d'estimation par seuillage dans Φ est au mieux en $O\left(\sigma^{\frac{2\kappa}{\kappa+1}}\right)$.

4.1.1 Bruit additif Gaussien

4.1.1.1 Modèle d'observation

Dans ce cas, (4.1) devient

$$y[i] = x[i] + \varepsilon[i] \quad \varepsilon[i] \sim \mathcal{N}(0, \sigma^2). \quad (4.2)$$

La variance σ^2 est soit connue soit estimée à partir des observations en utilisant des estimateurs robustes tels que la déviation médiane absolue (MAD) calculée sur une version filtrée passe-haut des observations [76].

17. Quelle que soit la dimension de la grille d'échantillonnage des données, les images ou les données en dimension plus élevée sont traitées dans ce document en les réarrangeant sous forme de vecteurs 1D dans $\mathbb{R}^N$.

4.1.1.2 Estimation Bayésienne terme-à-terme

Dans [A30, A27, A90, A87, TH4], nous avons mis à profit le modèle du mélange d'échelle de Gaussiennes (SMG) univarié comme a priori pour du débruitage Bayésien terme-à-terme dans le domaine des transformées orthogonales dans un premier temps. Plus précisément les expressions analytiques des estimateurs de l'espérance conditionnelle a posteriori (MMSE) et du maximum a posteriori (MAP) ont été développées pour les formes K de Bessel (BKF) et l'a priori α -stable avec mélange fini (se reporter à la section 2.1.1.1 pour une description de ces a priori univariés). Plusieurs solutions ont été offertes pour l'estimation en contexte bruité des hyperparamètres mis en jeu dans les a priori. A noter au passage que l'estimateur MAP pour l'a priori BKF revient à calculer l'opérateur proximal associé à $-\log$ de la PDF BKF, et par conséquent rentre dans le cadre de l'extension non convexe de la Proposition 1.

Dans [A69], l'estimateur MMSE avec l'a priori BKF a été étendu au cas où Φ est la concaténation de plusieurs transformées orthogonales (typiquement ondelettes + DCT locale) pour prendre en compte la diversité morphologique des structures géométriques de l'image. La forme analytique obtenue du MMSE est une agrégation des MMSE de chacune des transformées avec des poids de pondération spatialement adaptatifs. Il a été clairement montré qu'un tel estimateur est meilleur que chacun des MMSE pris individuellement. Ceci a été récemment confirmé par d'autres auteurs [90, 146].

Les estimateurs MMSE et MAP ont aussi été étendus aux transformées redondantes, notamment la transformée en ondelettes non décimée invariante par translation (UDWT) et la transformée en curvelets rapide de seconde génération (FDCT) [40]. Le MAP pour le BKF est implémenté par un algorithme itératif qui repose sur le schéma explicite-implicite du Chapitre 3 où l'opérateur proximal est celui déduit du MAP dans le cas orthogonal.

Les performances observées de ces différents estimateurs terme-à-terme sur la base de test [1] aussi bien avec les transformées orthogonales que redondantes étaient meilleures que des procédures par seuillage simple ou avec des a priori moins élaborés. Toutefois, les résultats n'étaient pas encore au niveau de l'état de l'art des méthodes exploitant la dépendance entre coefficients. C'est ce qui nous a naturellement conduits à nous intéresser à des estimateurs Bayésiens utilisant les statistiques jointes se basant sur les a priori proposés en section 2.1.1.1.

4.1.1.3 Estimation Bayésienne par blocs

De manière analogue au cas univarié, l'a priori joint des formes K de Bessel multivarié (MBKF) a été utilisé pour le débruitage des images par blocs de coefficients dans le domaine des transformées redondantes (e.g. ondelettes invariantes par translations et curvelets) [TH4]. Ce travail s'est principalement focalisé sur le MMSE. Toutefois, aucune forme analytique n'étant disponible pour le MMSE joint, nous avons fait appel aux techniques d'intégration Monte-Carlo. Plusieurs approches pour estimer l'hyperparamètre de forme du bruit ainsi que les matrices de covariance du signal et du bruit ont été proposées. Des simulations exhaustives ont montré qu'associer ces estimateurs avec a priori joint et les transformées redondantes permet de se placer au moins au niveau de l'état de l'art en comparaison aux autres méthodes de l'époque reposant sur la parcimonie-seuillage, ou encore les approches non locales par patchs telles que les moyennes non locales (NL-means), voir fig. 4.2 pour un exemple de comparaison. Le point critique de ces approches Bayésiennes reste toutefois le temps de calcul et l'estimation des hyperparamètres à fort niveau de bruit.

4.1.1.4 Seuillage de Stein par blocs

C'est pour pallier ces limites que nous avons proposé dans [A15] un estimateur basé sur un seuillage par blocs de James-Stein [193] (voir fig. 4.1 pour une illustration graphique). Notre approche est une extension de [34] en d -dimensions et pour des transformées redondantes correspondant à des trames ajustées de $\mathbb{R}^N$.

Dans [A15], nous avons aussi étudié les vitesses minimax du seuillage par blocs de Stein en d dimensions. A cette fin, nous avons introduit une boule dans l'espace des coefficients de trame sur laquelle les vitesses sont obtenues. Le choix de cet espace est inspiré de la caractérisation donnée par [29] de famille d'espaces de régularité dans $\mathbb{R}^d$. Ces espaces couvrent le cas particuliers de espaces de Besov (isotropes) ainsi que ceux associés à des trames ajustées de curvelets dans leur construction de seconde génération

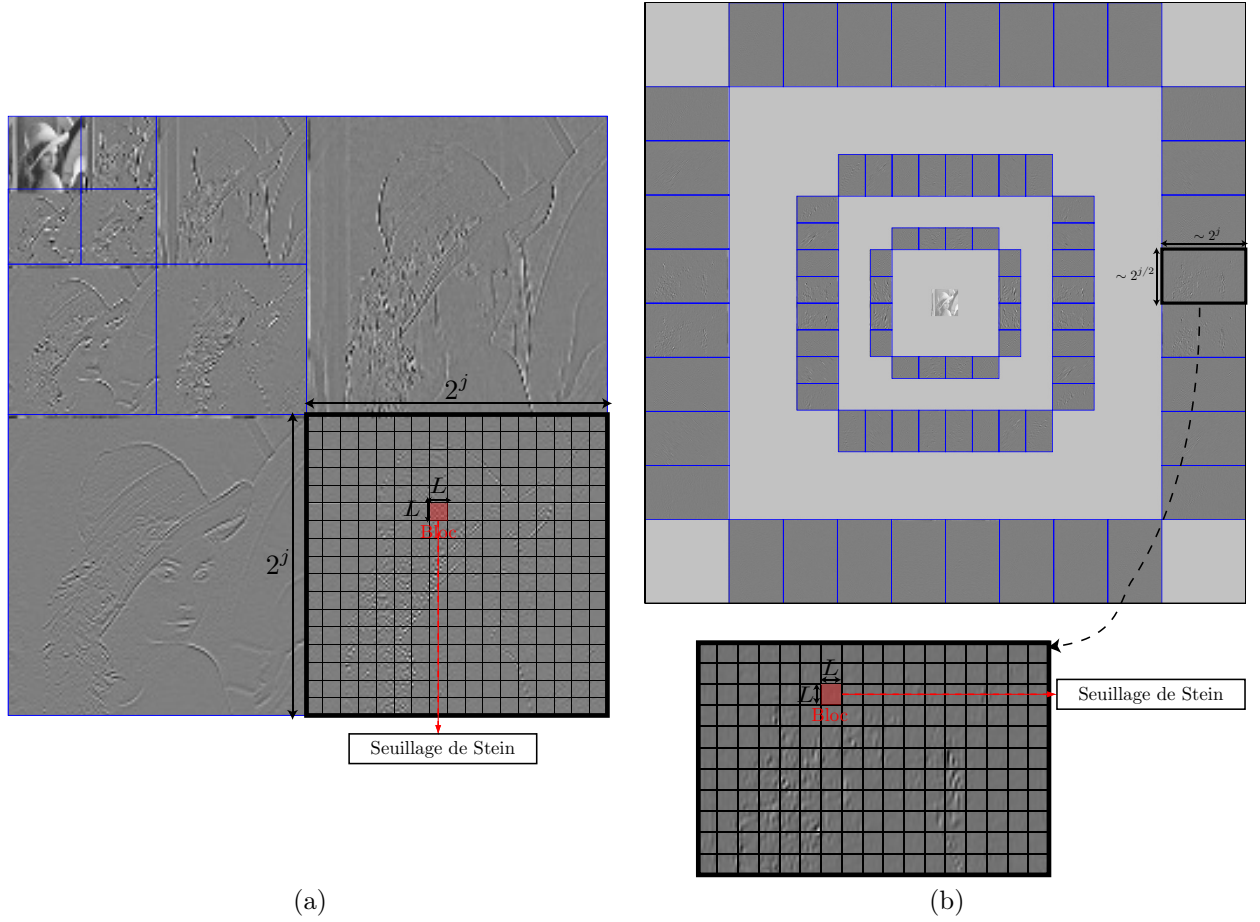


FIGURE 4.1 – Décomposition de sous-bandes en blocs de taille $L \times L$ sans recouvrement. (a) : cas de la transformée en ondelettes orthogonales (DWT). (b) : cas de la transformée de curvelets de seconde génération (FDCT).

[40]. Nous avons ainsi établi que l'estimateur du seuillage par bloc de Stein atteint les vitesses minimax sur ces boules, ou tout du moins quasi-minimax (à un facteur log près) dans le régime le plus défavorable. Ce résultat est valide pour un modèle de bruit qui n'est pas nécessairement AWGN (e.g. bruit coloré). Le choix du seuil est aussi discuté du point de vue théorique et sa valeur optimale est obtenue pour le cas du bruit additif Gaussien (pas nécessairement blanc). En plus de ses performances théoriques, la procédure pratique est très rapide et aisée à implémenter. Outre sa rapidité, elle donne des résultats se situant au niveau de l'état de l'art comparables (voire meilleurs à fort bruit) à ceux décrits ci-dessus dans un cadre Bayésien, voir fig. 4.2 pour une comparaison visuelle Tab. 4.1 pour les valeurs de PSNR.

4.1.2 Bruit de Poisson

4.1.2.1 Modèle d'observation

Dans bon nombre de systèmes d'acquisition des images, le bruit a pour origine des fluctuations aléatoires d'un processus de comptage (photodétecteurs dans les caméras, projections en tomographie, etc.). Dans ce cas, le bruit suit une loi de Poisson, où les observations sont issues de VA de Poisson indépendantes à valeur dans $\mathbb{N}$ telles que

$$y[i] \sim \mathcal{P}(x[i]), \quad i = 1, \dots, N,$$

où $x[i]$ est l'intensité sous-jacente avec la propriété classique $\mathbb{E}[y[i]] = \text{Var}[y[i]] = x[i]$. Le but du débruitage en présence de bruit de Poisson est d'estimer les intensités $x[i]$ à partir des observations $y[i]$. Les

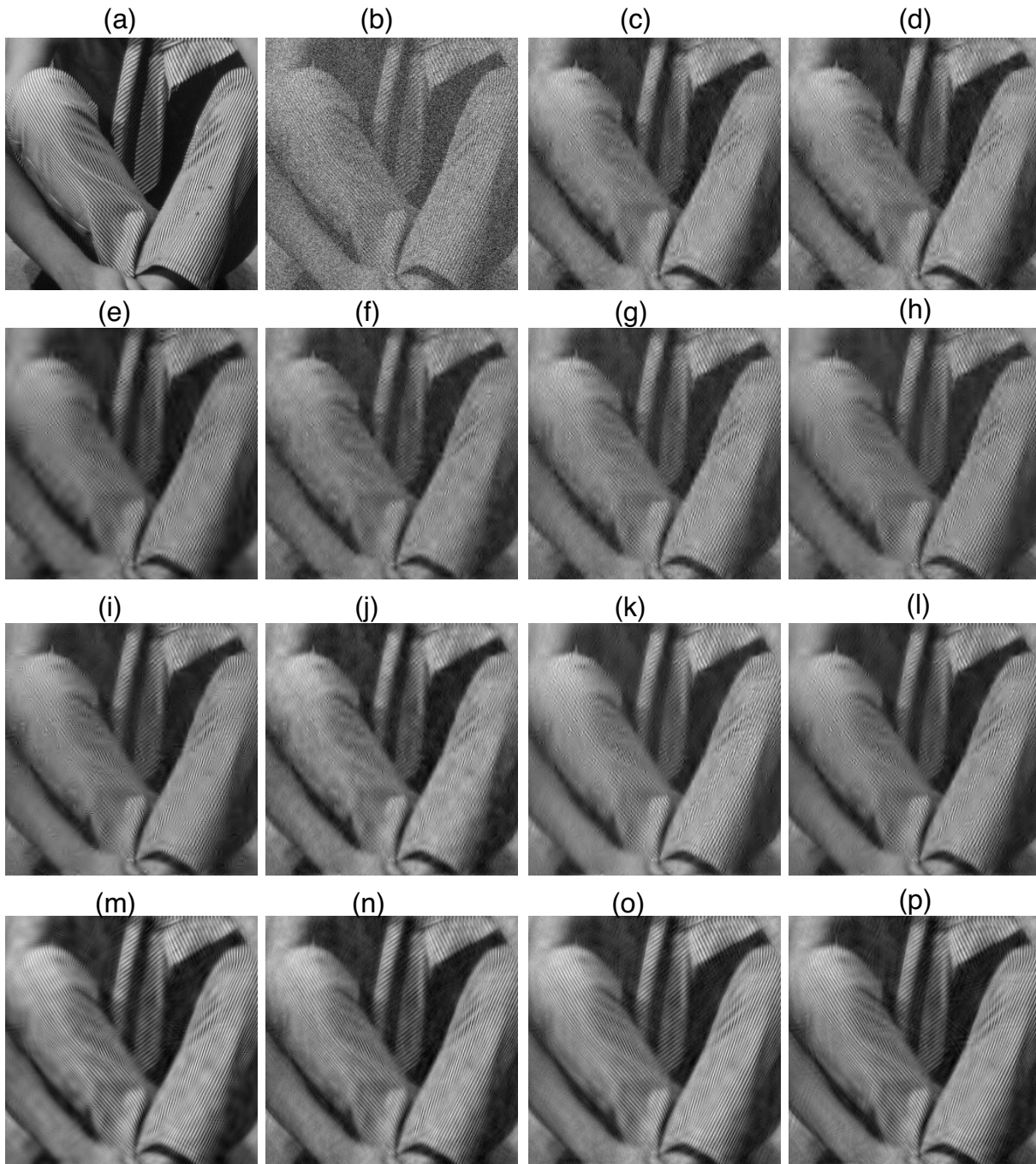


FIGURE 4.2 – Comparaison visuelle sur l'image **Barbara** des algorithmes de débruitage que nous avons proposés à quelques uns issus de la littérature. (a) Image originale. (b) Image contaminée par bruit AWGN de variance $\sigma = 40$ (PSNR=20.09dB). (c) NL-means [33]. (d) SURE-LET [145]. (e)-(i)-(m) Bloc-Stein avec respectivement la DWT, l'UDWT et la FDCT. (f)-(j)-(n) BKF avec les mêmes transformées. (g)-(k)-(o) MBKF avec les mêmes transformées. (h)-(l)-(p) MSMG-Jeffrey [173] avec les mêmes transformées.

Méthode	PSNR (dB)		
NL-means [33]	25.86		
SURE-LET [145]	24.29		
Bloc-Stein	DWT 24.50	UDWT 25.10	FDCT 26.33
BKF	DWT 23.94	UDWT 24.41	FDCT 25.79
MBKF	DWT 24.96	UDWT 25.37	FDCT 26.56
MSMG-Jeffrey [173]	DWT 24.86	UDWT 25.40	FDCT 26.53

TABLE 4.1 – Comparaison quantitative en termes de PSNR des algorithmes de la fig. 4.2 sur l'image de **Barbara**. L'image bruitée correspond à un PSNR=20.09dB.

champs d'applications sont nombreux et couvrent notamment l'imagerie astronomique et biomédicale. Cependant, la variance de ce bruit dépend du signal lui-même ce qui est source de nombreuses complications. Même par le théorème central limite auquel cas le bruit devient additif Gaussien $y[i] \xrightarrow[x[i] \rightarrow +\infty]{d} \mathcal{N}(x[i], x[i])$, la même difficulté subsiste.

Pour cette raison, ce bruit est beaucoup moins traité dans la littérature que le cas AWGN. Pour s'y atteler, deux grandes approches se dégagent de la littérature (voir [A1] pour une revue plus détaillée) :

- (i) Celle consistant à se ramener au bruit AWGN en utilisant une transformation stabilisatrice de la variance (VST), la plus connue étant celle d'Anscombe [10]

$$y_S[i] = \mathcal{A}(y[i]) = 2\sqrt{y[i] + \frac{3}{8}}. \quad (4.3)$$

On peut alors montrer que $y_S[i] \xrightarrow{d} \mathcal{N}(2\sqrt{x[i]}, 1)$ à mesure que $x[i] \rightarrow +\infty$. On peut alors déployer les méthodes de débruitage en présence de bruit AWGN, et appliquer la VST inverse pour obtenir la version débruitée (avec une correction du biais introduit par le carré d'une estimée). D'autres VST sont possibles comme celle de Haar-Fisz [97, 100, A46]. L'approche par VST se distingue par sa simplicité mais n'est performante que pour des intensités assez élevées (typiquement $x[i] \gtrsim 10$ [A19]).

- (ii) Celle consistant à tenir compte directement des propriétés statistiques de ce bruit lors de l'élaboration de l'estimateur. Par exemple dans un cadre Bayésien, l'attache aux données qui découle de la vraisemblance conditionnelle s'écrit

$$\begin{aligned} \ell_{\text{poisson}}(y|x) : x \in \mathbb{R}^N &\mapsto \sum_{i=1}^N f_{\text{poisson}}(x[i]), \\ \text{si } y[i] > 0, \quad f_{\text{poisson}}(x[i]) &= \begin{cases} -y[i] \log(x[i]) + x[i] & \text{si } x[i] > 0, \\ +\infty & \text{sinon,} \end{cases} \\ \text{si } y[i] = 0, \quad f_{\text{poisson}}(x[i]) &= \begin{cases} x[i] & \text{si } x[i] \in [0, +\infty], \\ +\infty & \text{sinon.} \end{cases} \end{aligned}$$

Toutefois, les obstacles à franchir pour mettre en place les estimateurs dans cette seconde catégorie d'approche sont en général plus complexes. Nous préciserons ces obstacles plus tard et les solutions que nous avons élaborées pour les contourner.

Dans nos travaux, nous avons proposé plusieurs méthodes de débruitage avec bruit de Poisson se fondant sur chacune de ces deux approches. C'est l'objet des deux sous-sections suivantes.

4.1.2.2 Test d'hypothèse sans stabilisation

Plusieurs auteurs ont mis en évidence l'intérêt du test d'hypothèse binaire dans le domaine de la transformée en ondelettes orthogonales de Haar pour restaurer des signaux/images corrompues par un

bruit de Poisson [128, 129, 26, A20, A76]. En effet, cet attrait part du constat que la PDF de la différence de deux VA indépendantes de Poisson $y[i] \sim \mathcal{P}(x[i])$ et $y[i'] \sim \mathcal{P}(x[i'])$ s'écrit analytiquement [189] :

$$\text{pdf}(y[i] - y[i'] = n) = e^{-(x[i] + x[i'])} (x[i]/x[i'])^{-n/2} I_n(2\sqrt{x[i]x[i']}) ,$$

où $I_n(t)$ est la fonction de Bessel modifiée de première espèce. Sous l'hypothèse H_0 de la constance de l'intensité x (c.à.d. $x[i] = x[i']$), on peut déduire de la PDF ci-dessus la distribution des différences sous H_0 , et ainsi obtenir l'expression analytique du seuil critique permettant de tester H_0 contre son alternative à un risque de 1ère espèce donné¹⁸. En utilisant la transformée en ondelettes orthogonales de Haar (non normalisée), l'auteur [128] se ramène exactement à ce cadre où l'hypothèse testée à chaque échelle est la constante de l'intensité dans un support de taille 2^j de l'ondelette.

Toutefois, bien que l'ondelette de Haar permette d'obtenir une forme analytique du seuil critique, elle produit les effets indésirables de stair-casing. On peut envisager d'introduire une forme d'invariance par translation par décalage périodique (cycle-spinning), mais cela devient très vite consommateur en temps de calcul dès que la dimension et la taille des données augmente, et pose des problèmes de dépendance des tests statistiques. D'un autre côté, d'autres ondelettes plus régulières ne permettent pas d'accéder à des formes analytiques des PDFs des coefficients de détail du bruit de Poisson.

Pour résoudre ce dilemme, nous avons proposé dans [A20] d'avoir recours à la transformée bi-orthogonale de Haar [148] avec un banc de filtres ayant une normalisation appropriée. La normalisation choisie est telle que les coefficients de détail à chaque échelle ont la même variance que les coefficients de Haar. On tire alors bénéfice du fait que la fonction d'échelle dans le cas bi-orthogonal est quasi-Lipschitz [176] ce qui assure une reconstruction plus régulière, tout en maintenant la simplicité des expressions des p -valeurs et des seuils critiques. En effet, nous avons montré dans [A20] que la p -valeur d'un coefficient de la transformée de Haar bi-orthogonale est très bien approchée (avec une vitesse contrôlée) par celle dans le cas orthogonal pour les intensités élevées. Pour les faibles intensités, l'approximation devient une borne supérieure. La conséquence immédiate de ces résultats est que l'on peut utiliser les formes analytiques de la PDF et des seuils critiques obtenus pour les coefficients de Haar pour tester ceux de la transformée bi-orthogonale tout en ayant le même contrôle sur l'erreur de 1ère espèce.

4.1.2.3 Stabilisation multiéchelle : MS-VST

Soit $(y[i])_i$ des VA de Poisson indépendantes chacune $\sim \mathcal{P}(x[i])$, $\forall i$, et $z[i] := \sum_k h[k]y[i - k]$ leur version filtrée par h . En fin de section 4.1.2.1, nous avons souligné que la VST d'Anscombe est critiquable pour les faibles intensités. Mais, si le filtre h agit comme un "moyenneur" (plus généralement un passe-bas), on est en mesure d'espérer que stabiliser $z[i]$ serait plus bénéfique puisque le SNR en sortie du filtre h serait meilleur. C'est sur cette intuition qu'est fondée notre approche de stabilisation multiéchelle (MS-VST) détaillée dans [A19, A81] et étendue dans [A11, A8] à des données multicanales et sur la sphère.

Pour ce faire, dans [A19], la VST d'Anscombe a tout d'abord été généralisée au processus de Poisson filtré $z[i]$ et les paramètres de la VST optimisés pour chaque filtre h pour atteindre la meilleure vitesse de convergence vers le comportement asymptotique à mesure que $x[i]$ grandit. Ensuite, pour plusieurs transformées multiéchelles (ondelettes, ridgelets, curvelets), la VST généralisée a été mise en oeuvre. Par exemple, pour la transformée en ondelettes, l'algorithme à trous est modifié en appliquant l'opérateur associé à la VST généralisée à chaque sous-bande d'approximation avant de calculer les coefficients de détail. La normalité asymptotique des coefficients de détail "stabilisés" sous l'hypothèse nulle H_0 de fond régulier a été établie avec une variance qui ne dépend que du banc de filtres utilisé. Grâce à ce résultat, la détection des coefficients significatifs peut être accomplie en contexte Gaussien. Compte tenu de la non-linéarité de la MS-VST, un algorithme d'inversion itératif promouvant la parcimonie a été proposé et sa convergence démontrée.

La fig. 4.3 compare les approches proposées à celles de l'état de l'art en 2008 sur une image biomédicale de microscopie de fluorescence. L'image originale fig. 4.3(a) a des intensités faibles à moyennes $x[i] \in [0.53, 16.93]$. Les méthodes comparées sont celle basée sur la VST d'Anscombe [10] (fig. 4.3(c)),

¹⁸. Ce risque peut être éventuellement corrigé contre le nombre de comparaisons dans le cas de tests multiples comme en image.

Méthode	NMSE
VST d'Anscombe	0.095
VST de Haar-Fisz+cycle-spinning [100, A46]	0.096
Test d'hypothèse dans le domaine de la transformée de Haar+cycle-spinning [129]	0.10
Platelets+cycle-spinning [205]	0.079
Test d'hypothèse dans le domaine de la transformée bi-orthogonale de Haar	0.094
MS-VST avec l'UDWT	0.090
MS-VST avec la transformée de curvelets	0.078

TABLE 4.2 – Comparaison quantitative en termes de NMSE des algorithmes de la fig. 4.3.

VST de Haar-Fisz [100, A46] (fig. 4.3(d)), test d'hypothèse dans le domaine de la transformée de Haar [129] (fig. 4.3(e)), les platelets [205] (fig. 4.3(f)), test d'hypothèse dans le domaine de la transformée bi-orthogonale de Haar (fig.4.3(g)), MS-VST associée à l'UDWT (fig.4.3(h)) et aux curvelets (fig.4.3(i)). En vue d'une comparaison équitable, pour les méthodes des fig. 4.3(d), (e) et (f), un décalage périodique (cycle-spinning) est utilisé pour leur conférer une quasi-invariance par translation et réduire leurs artefacts de blocs. La MS-VST associée à l'UDWT surpasse les autres méthodes basées sur les ondelettes. La MS-VST+curvelets conduit aux meilleurs résultats aussi bien visuellement que quantitativement en termes d'erreur quadratique moyenne normalisée (NMSE)¹⁹ comme le montre Tab. 4.2.

4.1.3 Bruit mixte Poisson-Gaussien

4.1.3.1 Modèle d'observation

Dans certains dispositifs d'acquisition, un bruit de lecture non nécessairement négligeable dû en général à l'agitation thermique vient se superposer au bruit de Poisson. C'est typiquement le cas dans les capteurs CCD. Ce bruit de lecture est supposé en général additif blanc Gaussien. Dans ce cas, l'observation $y[i]$ est la somme de deux VA indépendantes, l'une Gaussienne et la seconde une VA de Poisson amplifiée par le gain g_0 , ce qui donne

$$y[i] = g_0 z[i] + \varepsilon[i], \quad z[i] \sim \mathcal{P}(x[i]), \quad \varepsilon[i] \sim \mathcal{N}(\mu, \sigma^2).$$

Le gain g_0 représente le rapport d'amplification introduit par exemple par un photomultiplicateur ou par le détecteur CCD.

Débruiter en présence de ce bruit mixte Poisson-Gaussien (MPG) consiste à estimer les intensités $x[i]$ à partir des observations $y[i]$. Ceci nécessite d'estimer les paramètres (g_0, μ, σ^2) à partir des données en ayant remarqué que $\text{Var}[y[i]] = g_0 \mathbb{E}[y[i]] + (\sigma^2 - g_0 \mu)$. Ainsi, ces paramètres peuvent être estimés à partir de la droite de régression $(\text{Var}[y[i]], \mathbb{E}[y[i]])$ et/ou des cumulants jusqu'à l'ordre 4 [A67].

Ces paramètres connus, la VST d'Anscombe se généralise à [164]

$$y_S[i] = \mathcal{A}_{\text{MPG}}(y[i]) = \frac{2}{g_0} \sqrt{g_0 y[i] + \frac{3}{8} g_0^2 + \sigma^2 - g_0 \mu}. \quad (4.4)$$

Comme pour la VST d'Anscombe, il a été montré dans [164] que $y_S[i] \xrightarrow[x[i] \rightarrow +\infty]{d} \mathcal{N}(2\sqrt{x[i]/g_0}, 1)$. Cette approximation n'est toutefois valable en pratique que pour $x[i]$ suffisamment grand.

4.1.3.2 MS-VST

Dans [A67], nous avons étendu la MS-VST exposée en section 4.1.2 au bruit MPG en partant de la généralisation de la VST d'Anscombe dans (4.4). En l'occurrence, il a été montré que toutes les propriétés asymptotiques de la MS-VST pour le bruit MPG ne dépendent que du gain g_0 et du banc de filtres utilisé dans la transformée multiéchelle.

¹⁹. Le bruit étant Poissonnien, la normalisation de la MSE est nécessaire pour que le critère de comparaison prenne son sens.

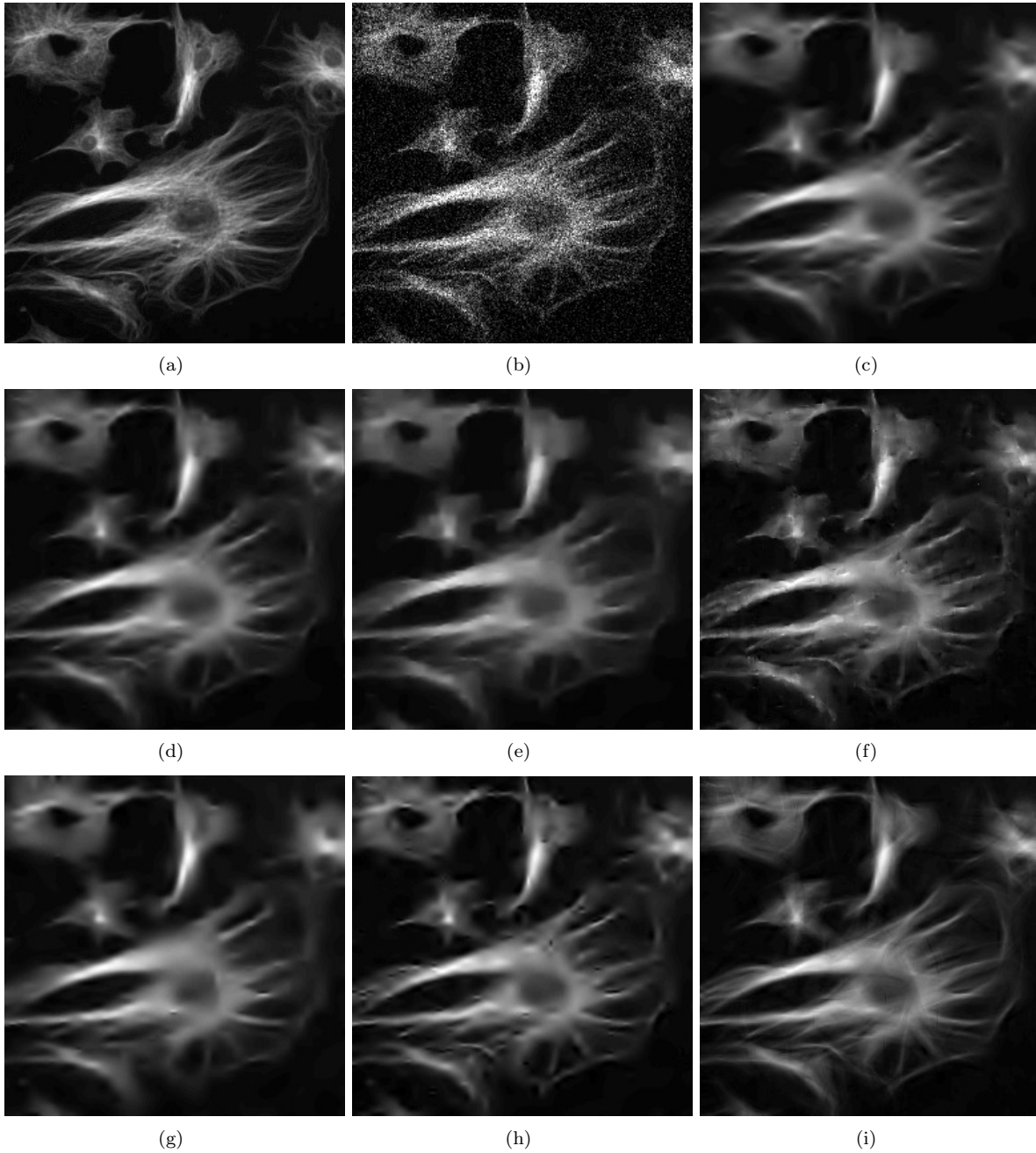


FIGURE 4.3 – Comparaison visuelle sur l'image de microscopie de fluorescence des algorithmes proposés de débruitage Poisson à ceux de l'état de l'art. (a) image des intensités x (régime d'intensités peu élevé avec $x[z] \in [0.53, 16.93]$); (b) image observée y avec un bruit de Poisson; (c) image débruitée après VST d'Anscombe; (d) image débruitée après VST de Haar-Fisz+cycle-spinning [100, A46]; (e) débruitée par test d'hypothèse dans le domaine de la transformée de Haar+cycle-spinning [129]; (f) débruitée par les platelets+cycle-spinning [205]; (g) test d'hypothèse dans le domaine de la transformée bi-orthogonale de Haar; (h) débruitée par la MS-VST associée à l'UDWT; (i) débruitée par la MS-VST associée à la transformée de curvelets.

4.1.4 Bruit multiplicatif

4.1.4.1 Modèle d'observation

Dans plusieurs systèmes d'imagerie active, comme l'imagerie SAR ou à ultra-sons dans certains régimes de diffusion, les observations sont corrompues par un bruit multiplicatif :

$$y[i] = x[i] \varepsilon[i], \quad (4.5)$$

où le bruit multiplicatif ε suit une loi de Gamma de paramètre $K \geq 1$ dont la PDF est

$$\text{pdf}(\varepsilon[i]) = \frac{K^K \varepsilon[i]^{K-1} \exp(-K\varepsilon[i])}{(K-1)!}. \quad (4.6)$$

Les approches classiques de débruitage pour un tel modèle d'observation²⁰ sont basées sur les méthodes variationnelles, l'inférence Bayésienne et la stabilisation de la variance, voir [A10] pour une revue détaillée de l'état de l'art. La stabilisation permet de se ramener à un bruit additif en prenant le log de (4.5) :

$$y_S[i] = \log y[i] = \log x[i] + \log \varepsilon[i] = \log x[i] + \epsilon[i]. \quad (4.7)$$

Le bruit $\epsilon[i]$ n'est ni à moyenne nulle ni Gaussien, mais l'on peut montrer que $\epsilon[i] \xrightarrow{d} \mathcal{N}(0, \psi_1(K))$ pour K grand, où $\psi_n(z) = \left(\frac{d}{dz}\right)^{n+1} \log \Gamma(z)$ est la fonction polygamma [4]. Cette idée est derrière bon nombre d'approches de débruitage en présence de bruit multiplicatif.

4.1.4.2 Estimation par une approche hybride

Dans [A10, A62], nous avons proposé une approche hybride combinant les représentations parcimonieuses et une méthode variationnelle pour la décomposition multiplicative. L'approche suggérée tente d'allier les avantages des méthodes de seuillage à ceux des méthodes variationnelles par régularisation TV tout en évitant leurs écueils respectifs, notamment les artefacts de blocs typiques de la régularisation TV. Celle-ci consiste en trois étapes :

- (i) Stabilisation de la variance par le log.
- (ii) Préestimation par seuillage dur terme-à-terme dans la trame ajustée des curvelets de seconde génération, où le seuil est volontairement sous-estimé de façon à préserver la majeure partie des structures géométriques de l'image. La contrepartie est la présence de coefficients aberrants (outliers).
- (iii) Du fait de ces outliers, il est montré qu'une attache aux données non lisse en zéro (e.g. ℓ_1) est la plus adaptée [85]. Ceci nous a conduit à proposer la minimisation du critère

$$\alpha^* \in \underset{\alpha}{\operatorname{argmin}} \lambda_1 \sum_{i \in \operatorname{supp}(\alpha_{\text{HT}})} |\alpha[i] - \alpha_{\text{HT}}[i]| + \lambda_0 \sum_{i \notin \operatorname{supp}(\alpha_{\text{HT}})} |\alpha[i]| + \|\Phi\alpha\|_{\text{TV}}, \quad (4.8)$$

où Φ est l'opérateur de synthèse de la trame ajustée des curvelets et α_{HT} est la préestimation obtenue après l'étape (ii). La log-image restaurée est $(\log(x))^* = \Phi\alpha^*$, et l'estimée x^* est obtenue en passant à l'exponentielle et en corrigeant le biais introduit par la stabilisation par le log.

Dans [A10], nous avons montré que le problème de minimisation (4.8) a une solution dans $\ell_2(\{1, \dots, T\})$, que les solutions images $\Phi\alpha^*$ ont les mêmes ensembles de niveaux (c.à.d. se déduisent les unes des autres par un changement local de contraste), et nous avons proposé des bornes théoriques pouvant guider le choix des paramètres λ_0 et λ_1 . Pour minimiser (4.8), nous avons proposé un algorithme itératif rapide basé sur le schéma de décomposition de Douglas-Rachford (voir section 3.3.4). Dans la fig. 4.4 et Tab. 4.3, sont affichés les résultats sur une image aérienne (un bruit simulé selon (4.5) avec $K = 4$) de plusieurs méthodes de débruitage de la littérature ainsi que celui de l'approche hybride. L'approche hybride est clairement moins sujette aux artefacts de blocs et restaure bien les détails géométriques complexes de l'image.

20. Le débruitage en présence de bruit multiplicatif est dénommé décomposition multiplicative par les auteurs [12].

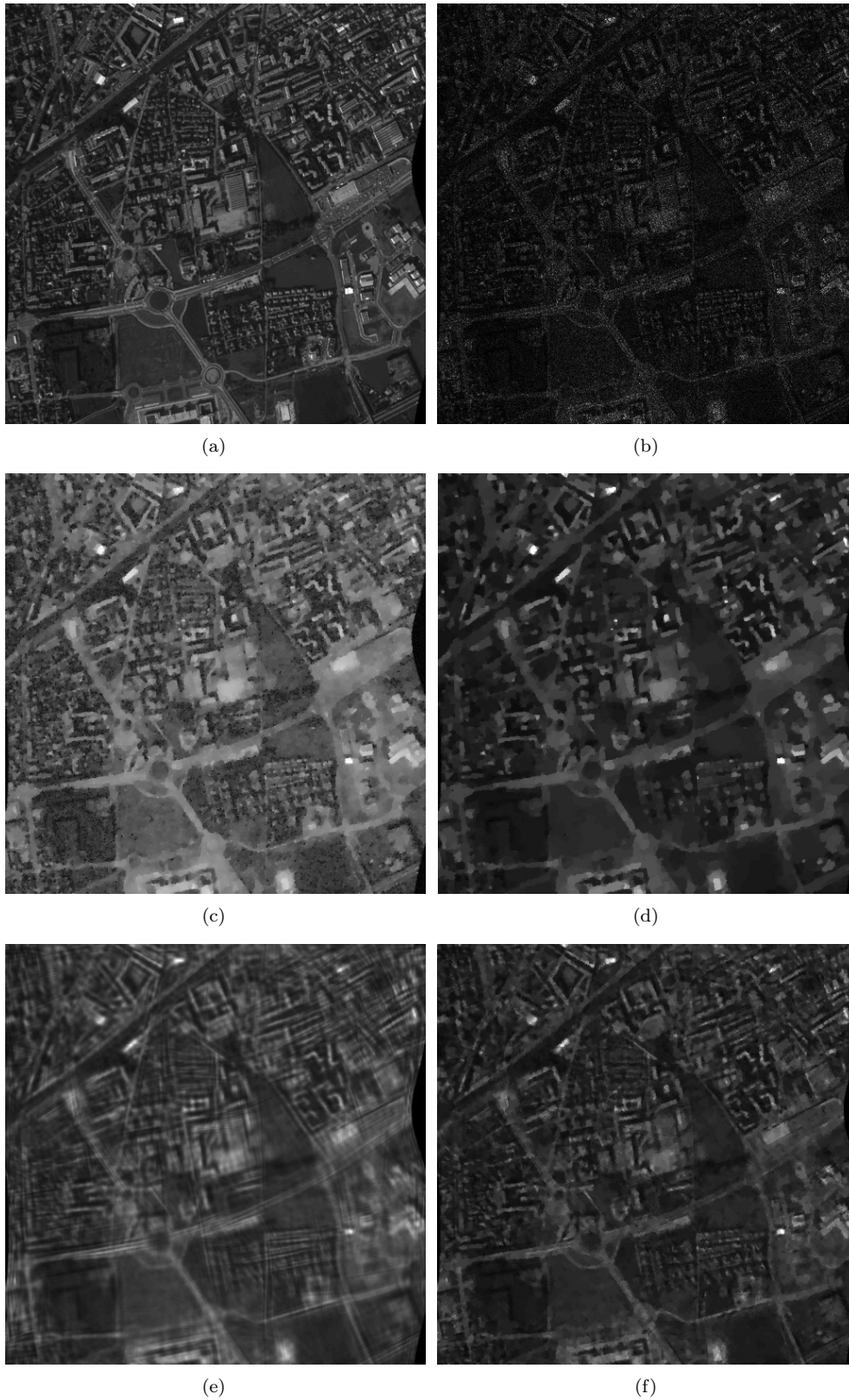


FIGURE 4.4 – Comparaison visuelle sur une image aérienne de la ville de Nîmes de l’algorithme hybride à ceux de la littérature. (a) Image originale. (b) Image bruitée selon (4.5) avec $K = 4$. Images débruitées par la méthode de : (c) [12]; (d) [186]; (e) seuillage de Stein par blocs (section 4.1.1.4) sur l’image log; (f) l’approche hybride.

Méthode	PSNR
[12]	24.55
[186]	24.40
Seuillage de Stein par blocs sur l'image log	24.92
Approche hybride	25.84

TABLE 4.3 – Comparaison quantitative en termes de PSNR des algorithmes de la fig. 4.4.

4.2 Estimation dans les séries temporelles

4.2.1 Régression semi-paramétrique

4.2.1.1 Modèle d'observation

Lors des deux dernières décennies, il y a eu un intérêt croissant de la communauté statistique dans les modèles partiellement linéaires (PLM). Dans de tels modèles, une série temporelle décrit l'évolution d'un système dynamique de la forme :

$$y[i] = \mathbf{X}[i, \cdot]\beta + g(t_i) + \varepsilon[i], \quad i = 1, \dots, N, \quad t_i \in [0, 1], \quad (4.9)$$

où $\beta \in \mathbb{R}^M$ est un vecteur de paramètres inconnus de la partie paramétrique linéaire, $\mathbf{X}[i, \cdot] = (\mathbf{X}[i, 1], \dots, \mathbf{X}[i, M])$ est une ligne de la matrice $\mathbf{X}$ de rang plein $M < N$ dont les colonnes sont les variables explicatives supposées connues, $g : [0, 1] \rightarrow \mathbb{R}$ est une fonction arbitraire inconnue, et les $\varepsilon[i]$ sont iid de moyenne nulle et de variance finie. Il existe de nombreuses situations pratiques où les séries temporelles sont régies par cette forme, comme par exemple en finance, économétrie, géophysique, biologie, imagerie médicale fonctionnelle, etc.

Les PLMs sont des modèles semi-paramétriques puisqu'ils contiennent à la fois une composante paramétrique (paramétrée par $\mathbf{X}$) et une autre non paramétrique au travers de g . Dans le contexte des PLMs, l'objectif est de construire des estimateurs de β et de g ayant observé y . Dans [119, A31], une revue détaillée de la littérature sur ces estimateurs est donnée.

4.2.1.2 Régularisation par la parcimonie

Dans [A31, A89], nous nous sommes focalisés sur la construction d'un estimateur de β et g en s'appuyant sur la parcimonie de g dans un dictionnaire Φ (typiquement un dictionnaire d'ondelettes). L'estimateur proposé est solution de

$$\min_{(\beta, \alpha)} \frac{1}{2} \|y - \mathbf{X}\beta - \Phi\alpha\|_2^2 + \Psi(\alpha), \quad (4.10)$$

où $\Psi(\alpha) = \sum_{i=1}^T \psi_i(\alpha[i])$ est une pénalité favorisant la parcimonie ressemblant à celles de la section 3.4.3, mais pas nécessairement convexe. Le problème d'optimisation (4.10) peut être réécrit en termes d'un dictionnaire augmenté $[\Phi \ \mathbf{X}] \in \mathbb{R}^{N \times (T+M)}$. Ainsi, dans le cas où Ψ satisfait les conditions de la Proposition 1, (4.10) peut-être efficacement résolu par les algorithmes de premier ordre décrits dans la section 3.3.3 avec une garantie de convergence vers un minimum global. L'écriture (4.10) peut aussi être réarrangée avec un dictionnaire modifié $\tilde{\Phi} = (\mathbf{I} - \mathbf{X}\mathbf{X}^+)\Phi \in \mathbb{R}^{N \times T}$ ²¹ et l'observation résiduelle $\tilde{y} = (\mathbf{I} - \mathbf{X}\mathbf{X}^+)y$. Dans le cas de la norme ℓ_1 comme pénalité Ψ , on peut décliner les conditions suffisantes en termes des propriétés de $\tilde{\Phi}$ (voir section 1.3) assurant l'unicité, la robustesse (aux signaux compressibles) et la consistance de l'estimateur de α , notamment les conditions basées sur la cohérence. A noter que ces conditions restent vraies même pour une régularisation ℓ_q , $0 \leq q < 1$.

Dans [A31], le cas où Φ est une trame ou une base orthonormée d'ondelettes a été détaillé avec diverses pénalités de parcimonie convexes et non convexes. Le choix du paramètre de régularisation (basé sur le GCV) ainsi que l'inférence statistique ont été étudiés. Ces travaux ont été repris récemment dans [103] pour démontrer la quasi-optimalité minimax de notre estimateur de g sur une boule de Besov si l'ondelette est suffisamment régulière.

21. C'est à dire la projection des atomes de Φ dans le sous-espace orthogonal à celui engendré par les colonnes de $\mathbf{X}$.

4.2.2 Régression paramétrique avec bruit à longue mémoire

4.2.2.1 Modèle d'observation

Le bruit à longue mémoire est présent dans plusieurs thématiques du traitement statistique du signal et de la théorie de l'information. Ce type de processus peut être rencontré dans divers domaines applicatifs comme la géophysique, l'électronique, l'hydrologie, l'économétrie, l'électrophysiologie, l'imagerie fonctionnelle cérébrale, etc. Sa présence et sa prise en compte rendent les traitements plus complexes. Dans [A36, A117], nous avons étudié son impact sur le modèle de régression linéaire paramétrique où l'on observe la série temporelle

$$y = \mathbf{X}\beta + \varepsilon, \quad (4.11)$$

avec $\mathbf{X} \in \mathbb{R}^{N \times M}$ la matrice connue de variables explicatives de rang plein en colonne, et ε est un processus Gaussien stationnaire à longue mémoire.

Un exemple bien connu de processus à longue mémoire est le mouvement Brownien fractionnaire (FBM) [151]. Ce dernier est un processus auto-similaire caractérisé par un paramètre scalaire appelé l'exposant de Hurst $0 < H < 1$. Le FBM n'est pas stationnaire au sens strict, tandis que l'incrément associé, qui est le processus Gaussien fractionnaire (FGP) [23], est stationnaire. Pour $\frac{1}{2} < H < 1$, le FGP a une dépendance à long terme ou une persistance; pour $H = \frac{1}{2}$ il est blanc; et pour $0 < H < \frac{1}{2}$, le processus a une dépendance à court terme ou une anti-persistance [23]. De plus, le FGP possède une densité spectrale de puissance (PSD) qui se comporte en C_H/ν^{2H-1} pour $\nu \rightarrow 0$ [23].

Une autre instance de bruit stationnaire à longue mémoire est le processus auto-régressif à moyenne mobile et intégration fractionnaire (ARFIMA) [106]. Le processus ARFIMA est une généralisation du processus ARIMA de Box et Jenkins au cas où le dérivateur devient fractionnaire d'ordre $d = H - 1/2 \in (-1/2, 1/2)$. Pour ces valeurs de d , le processus ARFIMA a une PSD de la forme $C |2 \sin \pi \nu|^{-2d} \propto_{\nu \rightarrow 0} 1/\nu^{2d}$ [23].

4.2.2.2 Régression linéaire dans le domaine des ondelettes

La présence d'un bruit à longue mémoire dans le modèle de régression linéaire (4.11) a un impact sur l'efficacité statistique de l'estimateur de β et le contrôle de l'erreur de 1ère espèce si une statistique de décision est construite pour tester une hypothèse sur β (e.g. $\beta[i] = 0$ contre $\beta[i] \neq 0$, $1 \leq i \leq M$). Le meilleur estimateur linéaire en statistique des séries temporelles correspond aux moindres carrés pondérés qui consistent à minimiser le critère fortement convexe

$$\min_{\beta \in \mathbb{R}^M} (y - \mathbf{X}\beta)^T \boldsymbol{\Sigma}^{-1} (y - \mathbf{X}\beta),$$

où $\boldsymbol{\Sigma} \succ 0$ est la matrice de covariance du processus ε . Le minimiseur unique de ce critère est donné par

$$\tilde{\beta} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} y, \quad (4.12)$$

de plus, $\tilde{\beta} \sim \mathcal{N}(\beta, (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1})$. En clair, cet estimateur est sans biais et à variance minimale. La difficulté ici est de construire et d'inverser la matrice de covariance, si tant est qu'elle soit connue, ou de l'estimer à partir des données si ce n'est pas le cas.

Dans [A36], nous avons introduit une approche fondée sur le fait que la transformée en ondelettes orthogonale est une (presque) base de Karhunen-Loève pour les processus en $1/f$ si l'ondelette possède une régularité suffisante [98, 207]. Ainsi, elle diagonalise la matrice de covariance $\boldsymbol{\Sigma}$ et lève par la même occasion la difficulté d'implémentation de l'expression (4.12). Nous avons alors construit l'estimateur du maximum de vraisemblance conjoint du vecteur β et des paramètres du bruit (l'exposant H et un paramètre d'échelle) opérant dans le domaine de la transformée en ondelettes orthogonales après avoir transformé l'observation y et les colonnes de $\mathbf{X}$. Les propriétés de cet estimateur, incluant le biais, les bornes de Cramér-Rao sur β et les paramètres du bruit, et leur distribution asymptotique ont été établies. Cet estimateur a été appliqué dans un large panorama d'études en imagerie fonctionnelle sur lesquels on reviendra au Chapitre 9.

4.3 Récapitulatif et conclusion

Dans ce chapitre, le rôle précieux de la parcimonie en théorie de l'estimation a été souligné. Nos contributions en débruitage (avec différents types de bruit au-delà du cas classique du bruit additif blanc Gaussien), en régression semi-paramétrique et paramétrique ont été décrites. En débruitage, les apports d'une modélisation adéquate du bruit et de la parcimonie structurée au travers des modèles multivariés par blocs (ou groupes) ont été mis en exergue. Les meilleures performances sont d'ailleurs obtenues lorsqu'on allie un estimateur basé sur la parcimonie structurée et une transformée adaptée aux structures de l'image, et ce pour tous les types de bruits traités. Les estimateurs construits dans ce chapitre ont été mis à profit dans plusieurs applications concrètes qui seront décrites au Chapitre 9.

Chapitre 5

Parcimonie et problèmes inverses linéaires

Comme extension naturelle du cadre du débruitage traité dans le chapitre précédent, les observations peuvent être des mesures indirectes bruitées d'un signal ou d'une image d'intérêt que l'on désire restaurer. C'est cet objectif qui sera poursuivi dans le présent chapitre. Plusieurs ingrédients seront empruntés aux chapitres précédents notamment ceux traitant de la modélisation et l'optimisation convexe non lisse. Nos travaux traitant des problèmes inverses sont une partie de la thèse de F.-X. Dupé [TH1], ainsi que [A1, Chapitre 7][A12, A26, A52, A55, A48, A66, A65, A74, A80, A84, A85, A111, S1, S2].

5.1 Exemples de problèmes inverses

Dans le cadre de la restauration d'images, une écriture générique possible de l'équation de dégradation, qui relie l'observation image y à l'image cible recherchée x , est ²²

$$y = (\mathbf{H}x) \odot \varepsilon, \quad (5.1)$$

où ²³ $\mathbf{H} : \mathbb{R}^N \rightarrow \mathbb{R}^M$ est un opérateur (possiblement non linéaire) borné de dégradation, $\odot$ est toute composition à deux arguments comme dans (4.1), ε représente le bruit ; des fluctuations aléatoires non mesurées autour des valeurs idéales ou une perturbation déterministe représentant l'imperfection du modèle.

Ce chapitre est dédié à nos travaux sur les problèmes inverses où $\mathbf{H}$ est linéaire. Les exemples typiques de problèmes linéaires inverses couverts par cette équation de dégradation sont :

- La déconvolution : $\mathbf{H}$ est un opérateur linéaire de convolution par une fonction d'étalement du point (PSF) généralement passe-bas impliquant que y ne contient pas les détails fins de x . Le bruit ε peut être AWGN ou un bruit de Poisson. Ces deux cas seront considérés dans ce chapitre.
- La super-résolution : $\mathbf{H}$ est une convolution par un noyau passe-bas de largeur s suivie d'un sous-échantillonnage d'un facteur s . Le nombre de mesures dans ce cas est $M = N/s$.
- L'interpolation de données manquantes (inpainting) : $\mathbf{H}$ est linéaire représentant l'action d'une matrice $\mathbb{R}^{M \times N}$ à entrées binaires ; 1 pour les échantillons observés et 0 sinon.
- Décodage en échantillonnage compressé : $\mathbf{H}$ est une matrice de mesure $\mathbb{R}^{M \times N}$ choisie dans un ensemble approprié (en général aléatoire), voir Chapitre 7 pour de plus amples détails.

En raison de la perte d'information induite par l'action de $\mathbf{H}$, le problème de restaurer x à partir de y est mal posé au sens de Hadamard. La démarche classique consiste donc à incorporer des connaissances supplémentaires permettant de réduire l'espace des solutions candidates. Ces informations a priori traduisent par exemple des propriétés de régularité de la solution (approche variationnelle), son appartenance à un modèle stochastique (approche Bayésienne), ou des propriétés plus complexes sur son contenu

²². Les images ou les données en dimension plus élevée sont traitées en les réarrangeant sous forme de vecteurs 1D.

²³. Comme dans le Chapitre 3, nous nous restreignons ici au cas réel en dimension finie, même si certains des résultats et algorithmes peuvent être adaptés aux cas complexe et/ou en dimension infinie.

géométrique. Dans nos travaux, nous avons cherché systématiquement à allier d'une part la modélisation statistique des images (Chapitre 2) et la diversité morphologique (Chapitre 1) pour construire l'attache au données et l'a priori au travers du dictionnaire et de la pénalité, et d'autre part l'inférence Bayésienne et les outils d'analyse convexe du Chapitre 3 pour résoudre les fonctionnelles qui en découlent.

5.2 Problèmes inverses linéaires avec bruit additif Gaussien

5.2.1 Modèle de dégradation

Dans ce cas, (5.1) se simplifie à

$$y = \mathbf{H}x + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2). \quad (5.2)$$

Les problèmes linéaires inverses en présence de bruit AWGN sont un domaine de recherche prolifique en traitement du signal et des images tant la littérature sur le sujet est abondante. L'utilisation de la parcimonie pour les résoudre a d'abord été amorcée par les méthodes de seuillage dans les bases d'ondelettes dans [70, décomposition ondelettes-vaguelettes (WVD)] et [3, décomposition vaguelettes-ondelettes (VWD)]. L'approche WVD a été ensuite raffinée dans [127, 165]. Nous avons récemment proposé dans [A52, S5] une extension multidimensionnelle de VWD avec seuillage de Stein par blocs avec les vitesses minimax associées.

La régularisation par des pénalités favorisant la parcimonie est encore plus récente [96, 95, 27, 59, 47, A26, A74, A84, A85] (voir [A1] pour une revue plus détaillée).

5.2.2 Estimateur MAP avec a priori de parcimonie

Partant du modèle de dégradation (5.2), on suppose ici que la solution $x = \Phi\alpha$ est parcimonieuse ou compressible dans le dictionnaire $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^T$. L'estimateur MAP dans ce cas prend la forme,

$$\min_{\alpha \in \mathbb{R}^T} \frac{1}{2\sigma^2} \|y - \mathbf{H}\Phi\alpha\|_2^2 + \lambda\Psi(\alpha). \quad (5.3)$$

où $\Psi \in \Gamma_0(\mathbb{R}^T)$ est une pénalité favorisant la parcimonie (voir section 3.4.3), et où l'on a supposé implicitement que l'a priori prenait la forme Gibbsienne $\propto \exp(-\lambda\Psi(\alpha))$. Cet a priori est dit a priori synthèse auquel cas la minimisation dans (5.3) porte sur les coefficients α . Ceci est à distinguer de l'a priori analyse où dans la formulation MAP, la variable du problème de minimisation devient x

$$\min_{x \in \mathbb{R}^N} \frac{1}{2\sigma^2} \|y - \mathbf{H}x\|_2^2 + \lambda\Psi(\Phi^T x). \quad (5.4)$$

Les deux formulations ne sont équivalentes que dans le cas où le dictionnaire correspond à une base orthonormée (voir les discussions dans [89, A1] de leurs avantages et inconvénients respectifs)²⁴.

Le problème de minimisation (5.3) est clairement un cas particulier de (P) en page 28, et son ensemble de solutions est non vide. L'attache aux données étant à gradient Lipschitz-continu de constante $\|\mathbf{H}\Phi\|^2/\sigma^2$, les schémas de décomposition explicite-implicite à un pas ou à pas multiple (voir section 3.3.3) apparaissent comme candidats naturels pour résoudre (5.3) et (5.4). Dans le cas de l'a priori analyse, l'opérateur proximal de $\Psi \circ \Phi^T$ est facilement déduit du Théorème 2 et de la Proposition 1. Les algorithmes ainsi obtenus aboutissent essentiellement à du seuillage itératif.

Le paramètre λ est généralement difficile à choisir (voir [A1, section 7.4.1.4] pour quelques solutions possibles). Une autre façon de contourner cette difficulté et de réécrire les problèmes (5.3)-(5.4) sous forme contrainte,

$$\min_{\alpha \in \mathbb{R}^T} \Psi(\alpha) \quad \text{t.q.} \quad \|y - \mathbf{H}\Phi\alpha\|_2 \leq \varsigma \quad (5.5)$$

$$\min_{x \in \mathbb{R}^N} \Psi(\Phi^T x) \quad \text{t.q.} \quad \|y - \mathbf{H}x\|_2 \leq \varsigma, \quad (5.6)$$

²⁴. D'autres cas sont couverts par cette formulation analyse comme la régularisation par la semi-norme TV discrète avec ∇ en lieu et place de Φ^T et div de Φ .

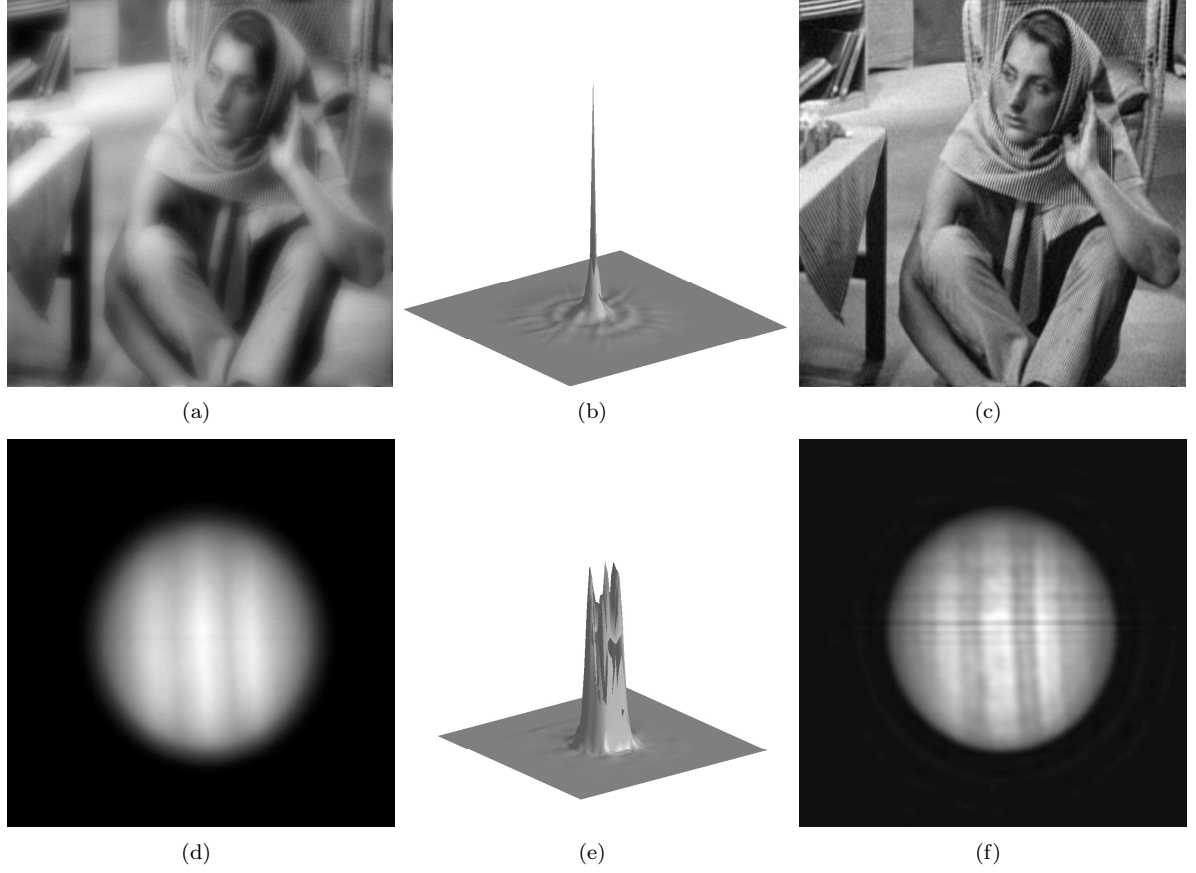


FIGURE 5.1 – Résultats de déconvolution avec a priori synthèse en résolvant (5.3) (ou sa forme contrainte (5.5)). (a) Image de **Barbara** dégradée par la PSF de (b) et un bruit AWGN BSNR=25dB ; (c) Image restaurée avec un dictionnaire contenant les curvelets. (d) Image de Jupiter observée lors de la calibration de Deep Impact. (e) PSF obtenue par la réponse de l'étoile 21 [192]. (f) Image restaurée avec un dictionnaire contenant la transformée en ondelettes (invariante par translation).

où Ψ est supposée coercive pour assurer l'existence d'une solution quelque soit $\mathbf{H}$. Compte tenu de la convexité, les formulations contraintes sont équivalentes à (5.3)-(5.4) dans le sens où pour tout λ , $\exists \varsigma(\lambda)$ tels que le minimiseur de (5.3) avec λ est aussi solution de la formulation analyse avec $\varsigma(\lambda)$ (idem pour l'a priori analyse). Le choix de la taille de la contrainte ς est guidé par des inégalités classiques de concentration sur les variables χ_M^2 pour assurer la faisabilité de la vraie solution. En l'occurrence, par l'inégalité de [136] sur les queues des distributions du χ^2 à M degrés de liberté, choisir $\varsigma^2 = M\sigma^2(1 + \kappa\sqrt{2/M})$ assure cette faisabilité avec une grande probabilité²⁵ supérieure à $1 - e^{-\kappa\sqrt{M/8}}$. Des arguments similaires peuvent être développés pour d'autres bruits comme le bruit de quantification uniforme, (voir section 7.3 et [A58, S4]).

Résoudre les formulations contraintes (5.5)-(5.6) est aussi possible grâce au schéma de décomposition de Douglas-Rachford (en ayant supposé que les contraintes satisfont la condition de qualification sur les domaines pour pouvoir appliquer ce schéma). L'opérateur proximal de Ψ est accessible par la Proposition 1, et celui de $\Psi \circ \Phi^T$ par le Théorème 2. Il reste ensuite à calculer le projecteur orthogonal sur le convexe fermé de la contrainte. Ce projecteur peut être facilement obtenu en utilisant encore une fois le Théorème 2.

25. Cette probabilité peut être améliorée à $1 - e^{-3\kappa^2/8}$ pour $0 \leq \kappa < \sqrt{M/8}$ par une borne due à [124].

5.2.3 Application à la déconvolution et l'inpainting

Dictionnaires globaux Dans [A26, A74, A84, A85], nous avons proposé des algorithmes de déconvolution et d'inpainting (interpolation avec données manquantes) avec bruit AWGN et a priori synthèse fondés sur ceux discutés plus haut. Il suffit de spécialiser l'opérateur $\mathbf{H}$ comme expliqué en section 5.1. Dans ces applications, le dictionnaire redondant global (fixe) était choisi en s'adaptant au contenu de l'image à restaurer et constitué en choisissant parmi les transformées présentées au Chapitre 1 (typiquement les curvelets ou les ondelettes pour le contenu cartoon, la DCT locale ou les wave-atoms pour les textures localement oscillantes).

Dans [A48, A111, S2], nous avons proposé de régulariser ces problèmes inverses par un a priori type analyse, en l'occurrence par une contrainte sur une boule TV (discrète). Le projecteur sur cette boule est calculé grâce au Théorème 2.

Dictionnaires globaux+locaux Pour mieux s'adapter à des contenus géométriques complexes de l'image (e.g. textures riches), nous avons traité dans [A73, S3] le cas où le dictionnaire contient un sous-dictionnaire global fixe, et un sous-dictionnaire local appris à partir de patchs extraits des données observées ou d'une référence. Le problème conjoint de la restauration et de l'apprentissage du dictionnaire induit une fonctionnelle à minimiser qui est une extension de (5.3). Ce problème de minimisation est résolu en adaptant les schémas de décomposition d'opérateurs monotones maximaux décrits plus haut. Toutefois, la fonctionnelle étant non convexe dans ce cas, nous n'avons démontré que la convergence vers un point stationnaire de l'algorithme proposé.

Les fig. 5.1 et fig. 5.2 illustrent visuellement les résultats obtenus pour la déconvolution et l'inpainting. Le problème résolu est celui avec l'a priori synthèse (5.3). Le dictionnaire fixe des curvelets a été utilisé à chaque fois sauf pour le résultat de la fig. 5.2(f). Dans cette dernière, le dictionnaire contenait un sous-dictionnaire global (ondelettes+curvelets) et un sous-dictionnaire local appris à partir des parties observées de l'image.

5.3 Problèmes inverses linéaires avec bruit de Poisson

5.3.1 Modèle de dégradation

En présence d'un bruit de Poisson, chaque observation est telle que

$$y[i] \sim \mathcal{P}((\mathbf{H}x)[i]) \quad i = 1, \dots, M, \quad (5.7)$$

où $(\mathbf{H}x)[i] \geq 0$ par définition. Ce problème est autrement plus difficile que celui en présence du bruit AWGN.

5.3.2 Estimateur MAP avec parcimonie sous contraintes

En adoptant un cadre Bayésien et l'estimateur usuel du MAP, le but est alors de minimiser la fonctionnelle suivante²⁶ :

$$J : \alpha \mapsto \underbrace{F_{\text{fid}} \circ \mathbf{H} \circ \Phi(\alpha)}_{F_1(\alpha)} + \underbrace{\lambda \Psi(\alpha)}_{F_2(\alpha)} + \underbrace{\iota_{\mathcal{C}} \circ \Phi(\alpha)}_{F_3(\alpha)}, \quad (5.8)$$

où $\Psi \in \Gamma_0(\mathbb{R}^T)$ est une pénalité additive de parcimonie, $F_{\text{fid}} \in \Gamma_0(\mathbb{R}^M)$, $\text{dom } F_{\text{fid}} \cap \text{Im}(\mathbf{H} \circ \Phi) \neq \emptyset$, $\lambda > 0$ est un paramètre de régularisation et $\iota_{\mathcal{C}}$ est l'indicatrice du convexe fermé $\mathcal{C}$. $\mathcal{C}$ est l'orthant positif, et le rôle du terme $\iota_{\mathcal{C}} \circ \Phi$ est d'imposer une contrainte de positivité sur l'image des intensités Poissonniennes restaurées qui le sont par nature. On suppose par la suite que les termes dans J sont tels que le domaine de J est non vide, c.à.d. $J \in \Gamma_0(\mathbb{R}^T)$.

²⁶ Ici, on se focalise sur l'a priori synthèse. Dans [TH1], le cas de l'a priori analyse est aussi traité.

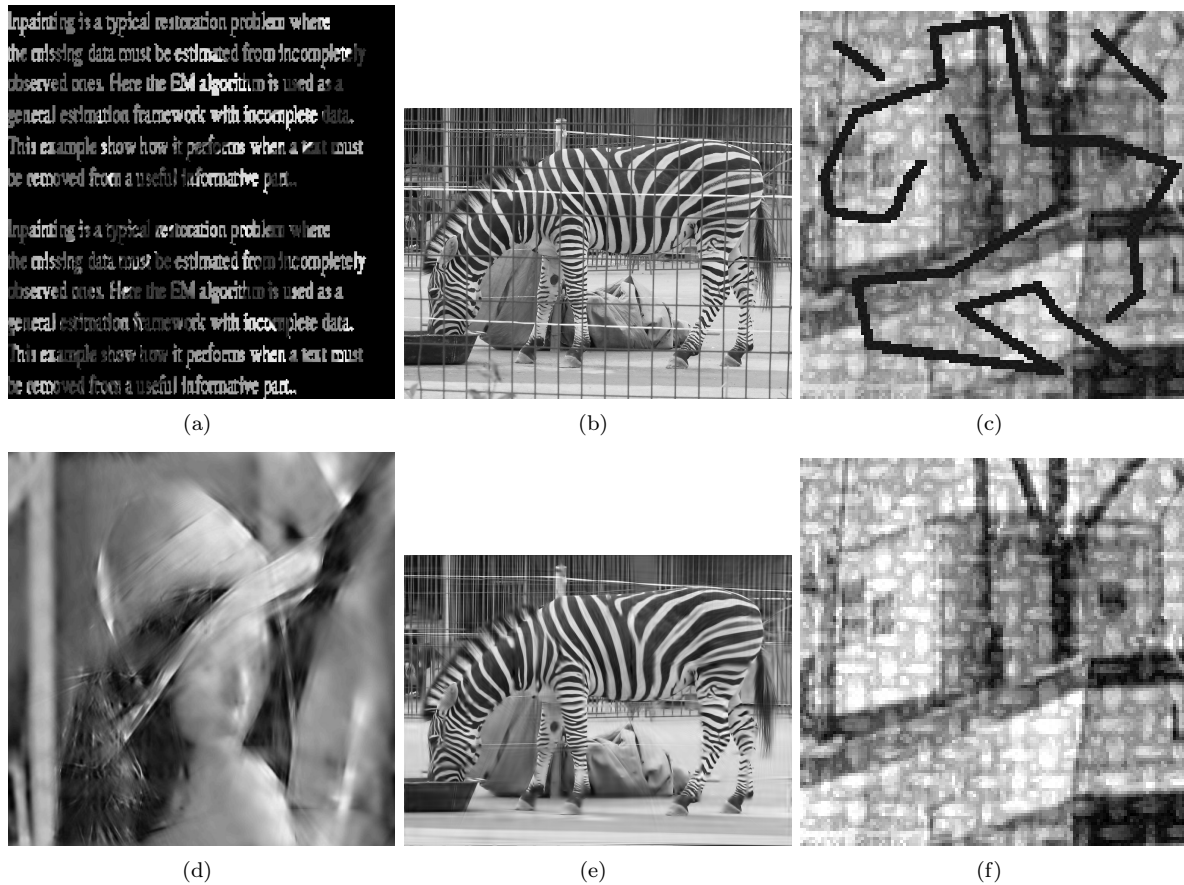


FIGURE 5.2 – Résultats d'inpainting avec a priori synthèse avec des parties manquantes de tailles différentes. (a)-(c) Images masquées. L'image du zèbre est réelle et on se propose d'en enlever la grille. Sur l'image du bateau, une texture artificielle complexe a été rajoutée. (d) Image de **Lena** restaurée avec un dictionnaire contenant les curvelets. (e) Image inpaintée du zèbre après élimination de la grille, le dictionnaire est formée encore une fois par les curvelets. (f) Image inpaintée avec un dictionnaire contenant un sous-dictionnaire global (ondelettes+DCT locale) et un sous-dictionnaire local conjointement appris lors de la restauration.

Le terme d'attache aux données est dépendant d'une fonction F_{fid} , reflétant la vraisemblance déduite du modèle de bruit issu du modèle de formation de l'image. Nous proposons deux méthodes de restauration : la première repose sur une stabilisation du bruit ; la deuxième consiste à formuler la vraisemblance relative au bruit de Poisson sans avoir à stabiliser les observations.

Attache aux données avec stabilisation Dans [A12, A65], et partant de (4.3), nous avons proposé de transformer (5.7) vers l'équation d'observation :

$$z[i] = 2\sqrt{(\mathbf{H}x)[i] + \frac{3}{8}} + \varepsilon[i], \quad \varepsilon[i] \sim \mathcal{N}(0, 1). \quad (5.9)$$

En toute rigueur, l'égalité est vraie pour des taux de comptage suffisamment élevés. Notons par ailleurs que la constante $3/8$ est celle optimisant la vitesse de convergence de la variance vers 1 à mesure que l'intensité croît. Cette constante peut être modifiée pour améliorer la vitesse du comportement asymptotique de l'espérance. Dans tous les cas, la VST d'Anscombe permet de se ramener à un bruit AWGN, mais au prix d'une dégradation non linéaire dans (5.9).

L'anti-log de vraisemblance correspondant à (5.9) engendre le terme d'attache aux données F_1 dans (5.8) où :

$$F_1 : \alpha \mapsto F_{\text{VST}} \circ \mathbf{H} \circ \Phi(\alpha), \text{ avec} \\ F_{\text{VST}} : \eta \in \mathbb{R}^M \mapsto \sum_{i=1}^M G_{\text{VST}}(\eta[i]), \quad G_{\text{VST}}(\eta[i]) = \frac{1}{2} \left(z[i] - 2\sqrt{\eta[i] + \frac{3}{8}} \right)^2. \quad (5.10)$$

Dans [A12, TH1], le problème d'optimisation (5.8) avec l'attache aux données (5.10) a été caractérisé. En observant que $F_1 \in \Gamma_0(\mathbb{R}^T)$ est à gradient Lipschitz-continu, des algorithmes pour le résoudre ont été proposés. Ces derniers sont fondés essentiellement sur les résultats du calcul proximal du Chapitre 3 et sur le schéma de décomposition explicite-implicite (à un pas ou pas multiple) qui y est décrit. Dans ce travail, le choix du paramètre de régularisation a été aussi discuté et quelques solutions déclinées, notamment celles fondées sur le rééchantillonnage aléatoire, les techniques de perturbation et le critère de validation croisée généralisée (GCV) comme estimateur non biaisé de l'erreur quadratique moyenne (voir aussi [A1]).

Attache aux données sans stabilisation De la PDF d'une VA de Poisson, on peut déduire de (5.7) le terme d'attache aux données suivant :

$$F_1 : \alpha \mapsto F_{\text{Poisson}} \circ \mathbf{H} \circ \Phi(\alpha), \text{ avec} \quad (5.11) \\ F_{\text{Poisson}} : \eta \in \mathbb{R}^M \mapsto \sum_{i=1}^M G_{\text{Poisson}}(\eta[i]), \\ \text{si } y[i] > 0, \quad G_{\text{Poisson}}(\eta[i]) = \begin{cases} -y[i] \log(\eta[i]) + \eta[i] & \text{si } \eta[i] > 0, \\ +\infty & \text{sinon,} \end{cases} \\ \text{si } y[i] = 0, \quad G_{\text{Poisson}}(\eta[i]) = \begin{cases} \eta[i] & \text{si } \eta[i] \in [0, +\infty], \\ +\infty & \text{sinon.} \end{cases}$$

Cette fois, $F_1 \in \Gamma_0(\mathbb{R}^T)$ mais n'est pas à gradient Lipschitz. De ce fait, pour résoudre (5.8) avec (5.11), des algorithmes basés sur le schéma de décomposition de type Douglas-Rachford ont été proposés et étudiés dans [TH1]. Pour calculer les opérateurs proximaux associés aux différents termes F_1 , F_2 et F_3 dans (5.8), nos résultats de calcul proximal du Chapitre 3 ont été mis à profit.

La fig 5.3 illustre les performances des algorithmes proposés comparativement à ceux de la littérature. Le dictionnaire contenait les curvelets pour nos méthodes. On peut remarquer en particulier que l'adaptation au bon modèle de bruit est cruciale comme l'illustre le résultat de la fig. 5.3(c). En effet, en supposant que le bruit est AWGN et en minimisant (5.3), le résultat n'est clairement pas satisfaisant, à l'inverse des résultats des fig. 5.3(g)-(h). Une application à des images réelles acquises par microscopie confocale est décrite au Chapitre 9.

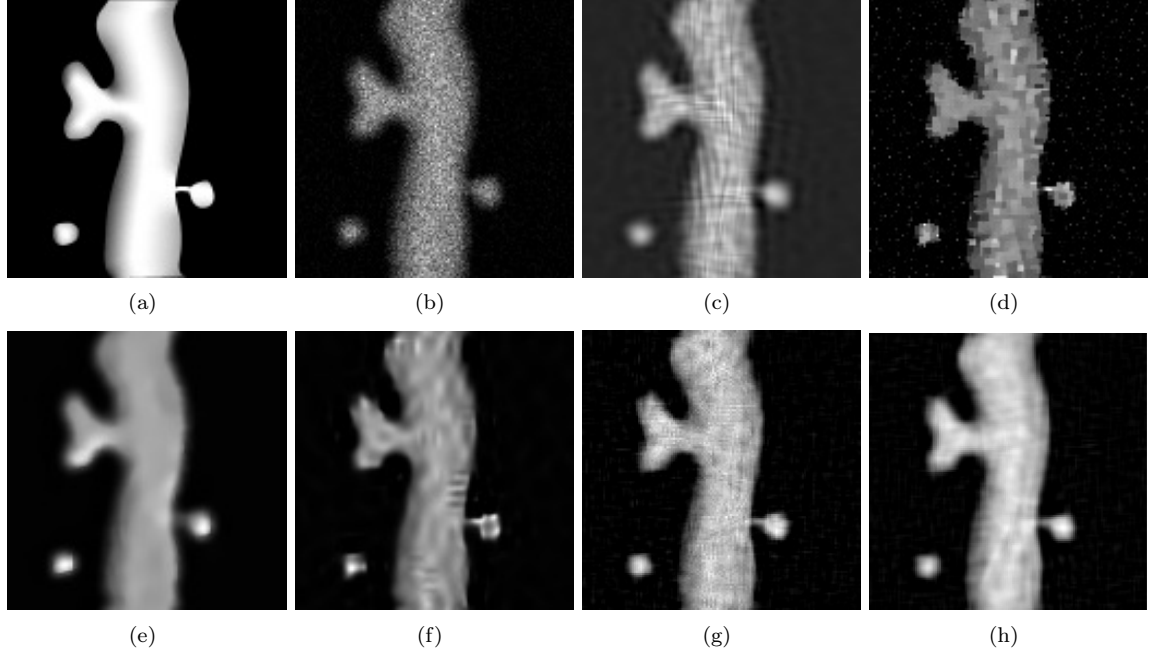


FIGURE 5.3 – Déconvolution d’une image simulée (intensités ≤ 30). (a) Image originale. (b) Image dégradée bruitée. (c) Résultat en résolvant (5.3) (c.à.d. en supposant que le bruit est additif blanc Gaussien). (d) Déconvolution par régularisation TV avec bruit de Poisson [66]. (e) Déconvolution par la méthode de [192]. (f) Déconvolution par la méthode de [206]. (g) Déconvolution en résolvant (5.8) avec stabilisation (5.10). (h) Déconvolution en résolvant (5.8) sans stabilisation (5.11).

5.4 Récapitulatif et conclusion

Dans ce chapitre, nos apports pour la résolution de problèmes inverses avec bruit additif Gaussien blanc et bruit de Poisson ont été détaillés. Ces contributions reposent sur la formulation d’un problème d’optimisation composé d’une attache aux données adéquate reflétant les propriétés statistiques du bruit (d’où l’intérêt crucial du modèle de formation de l’image), d’une pénalisation non différentiable promouvant la parcimonie dans un dictionnaire adapté à la diversité du contenu géométrique de l’image, et éventuellement d’une contrainte supplémentaire comme la positivité. Grâce au cadre de l’optimisation convexe et des résultats sur le calcul proximal développés au Chapitre 3, des algorithmes rapides et performants sont obtenus pour minimiser les diverses fonctionnelles.

Chapitre 6

Décomposition parcimonieuse et séparation de sources

Un des problèmes fondamentaux en traitement du signal est celui de décomposer un signal ou une image en une superposition de contributions issues de diverses sources (ce qu'on a appelé les composantes morphologiques au Chapitre 1). On peut penser à un morceau de musique impliquant la superposition de plusieurs signaux acoustiques générés par les différents instruments – et le but est de séparer ces contributions. Un obstacle se dresse toutefois immédiatement : pour un signal de N échantillons créés en superposant K composantes, nous avons $N \cdot K$ inconnues mais uniquement N équations. Les arguments élémentaires de l'algèbre linéaire nous inciterait à ne pas se lancer dans une telle entreprise. Mais comme nous l'avons expliqué en section 1.3, si des informations supplémentaires sont disponibles sur ces composantes – typiquement avec un a priori de parcimonie – des résultats rigoureux montrent que cette séparation (exacte ou approchée) est possible.

L'idée de la décomposition morphologique d'un signal en ses constituants, et les méthodes fructueuses pour y parvenir ont un impact dans un large spectre d'applications comme l'imagerie médicale et biomédicale, traitement du son, l'imagerie astronomique, imagerie aérienne, systèmes de communications, etc. Une instance intéressante de ce problème en vision/traitement d'image est celle visant la séparation d'une image en une esquisse (partie cartoon) et une partie texturée (voir (2.8), Chapitre 2). Depuis le livre d'Y. Meyer [157], ce modèle de décomposition a connu un fort regain d'intérêt. La partie cartoon est généralement supposée vivre dans BV. Pour modéliser la texture, les choix typiques sont l'espace G de Meyer [157], l'espace de Sobolev négatif [17] ou certains espaces de Hilbert adaptés pour les textures localement oscillantes [190, 191, 16]. Plusieurs critères variationnels de décomposition cartoon+texture ont été proposés dans la littérature notamment [200, 15, 16, 18] et d'autres.

Considérons maintenant la situation où l'on dispose d'une collection de signaux émis par différentes sources et mesurés par plusieurs capteurs. Chacun de ces derniers reçoit un mélange de contributions des sources, chacune pondérée différemment, et ce dépendamment de la configuration physique des capteurs/sources et du modèle de propagation direct. Les sources peuvent être des zones du cerveau émettant des signaux électriques mesurés par des électrodes disposées à la surface de la tête ; des personnes parlant dans un environnement avec des signaux sonores captés par plusieurs microphones (cocktail party problem) ; des sources de lumières émettant à différentes longueurs d'ondes ; etc. Le problème de la séparation de sources consiste à retrouver les signaux originaux des sources à partir des mélanges observés. Si les poids des mélanges sont inconnus, on parle alors de séparation aveugle de sources (BSS) où les poids doivent être estimés conjointement aux sources. Plusieurs questions posées par la BSS ont fait l'objet d'importants travaux, en témoigne la littérature extrêmement abondante sur le domaine (voir [A5, 50] pour une revue détaillée). L'enseignement principal à retenir dans un premier temps est que pour assurer la séparabilité des sources, il est fondamental de disposer de mesures de diversité ou de contraste entre elles (e.g. décorrélation, indépendance, etc.).

Dans ce chapitre, nous allons exposer nos contributions sur l'exploitation de la parcimonie et la

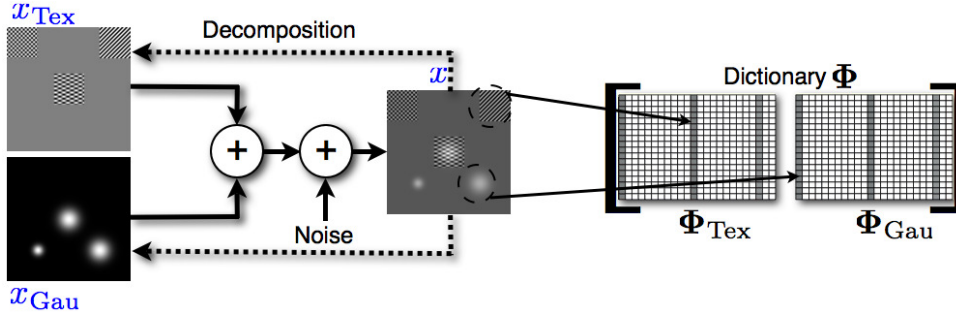


FIGURE 6.1 – Illustration du problème de la décomposition parcimonieuse d’images exploitant la diversité morphologique.

diversité morphologique²⁷ comme mesure de contraste pour la décomposition parcimonieuse des images et la BSS. Les travaux concernés par ce chapitre sont la thèse de J. Bobin [TH2], ainsi que [A1, Chapitres 8 et 9][A5, A13, A16, A17, A23, A24, A25, A51, A54, A61, A73, A68, A70, A112, S3].

6.1 Décomposition parcimonieuse d’image

En partant de (1.7) page 13, et en supposant un mélange linéaire bruité (voir l’illustration fig. 6.1) :

$$y = \sum_{k=1}^K x_k + \varepsilon = \sum_{k=1}^K \Phi_k \alpha_k + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \quad (6.1)$$

où chaque x_k est compressible dans un sous-dictionnaire dédié Φ_k avec les opérateurs d’analyse $\mathbf{T}_k$ et de synthèse rapides associés. L’algorithme de la décomposition en composantes morphologiques (MCA) proposé dans [190] vise à recouvrer toutes les composantes morphologiques $(x_k)_{k=1, \dots, K}$ ayant observé y , ce qui est éminemment un problème inverse mal posé. La MCA suppose comme postulat fondamental que l’on soit capable de construire un dictionnaire $\Phi = [\Phi_1, \dots, \Phi_K]$ tel que chaque composante x_k est parcimonieuse dans Φ_k , tout en l’étant beaucoup moins dans les autres $\Phi_l, l \neq k$. L’intuition dicte donc que les sous-dictionnaires doivent être mutuellement incohérents. En d’autres termes, chaque sous-dictionnaire joue le rôle de discriminant entre les composantes puisque chaque Φ_k affiche une préférence pour x_k . Ainsi, une fois que les sous-dictionnaires Φ_k ont été identifiés, l’utilisation d’algorithmes de poursuite cherchant la décomposition la plus parcimonieuse offre la séparation voulue (e.g. poursuite de base par minimisation ℓ_1 , algorithmes gloutons, ou algorithmes hybrides de seuillage itératif).

6.1.1 L’analyse en composantes morphologiques

Dans [190, 191], il a été proposé d’estimer les composantes $(x_k)_{1 \leq k \leq K}$ en résolvant le problème d’optimisation contraint :

$$\min_{\alpha_1, \dots, \alpha_K} \sum_{k=1}^K \|\alpha_k\|_p^p \quad \text{t.q.} \quad \left\| y - \sum_{k=1}^K \Phi_k \alpha_k \right\|_2 \leq \varsigma, \quad (6.2)$$

où $\|\alpha\|_p^p$ est la pénalité mesurant la parcimonie et correspondant à l’a priori GGD (avec $0 \leq p \leq 1$ comme régime intéressant), et ς est typiquement choisie comme $\sqrt{N} \sigma \sqrt{1 + 2\sqrt{2/N}}$ (par concentration des variables χ_N^2). Cette formulation est assez flexible pour autoriser une régularisation supplémentaire aidant le processus de séparation en favorisant une propriété particulière sur une composante. Par exemple, les

²⁷. Alors que la parcimonie est une propriété marginale, la diversité morphologique est une propriété relative n’ayant de sens que si deux composantes ou plus sont en jeu.

auteurs de [190] ont observé que pour la séparation cartoon+texture, la variation totale sur la partie cartoon permet de mieux renforcer sa régularité par morceaux.

Le problème (6.2) est loin d'être évident à résoudre pour $0 \leq p < 1$ (il est même NP-difficile pour $p = 0$). Pour le cas convexe $p \geq 1$, les schémas de décomposition du Chapitre 3 peuvent être utilisés sans difficulté. La MCA proposée dans [190], est un algorithme hybride et heuristique reposant sur une relaxation par blocs pour résoudre (6.2) pour le cas $0 \leq p \leq 1$ avec un accent mis sur $p = 0$. L'idée étant que si les Φ_k sont orthonormaux, et si tous les coefficients α_l sauf les k -èmes sont fixés, la solution est obtenue par seuillage (dur pour $p = 0$) appliqué au résidu marginal $r_k = y - \sum_{l \neq k} \Phi_l \alpha_l$ transformé dans Φ_k . De plus, la MCA utilise un seuil variable qui à la limite fait ressembler la MCA à un simple "matching pursuit" [150] (voir l'analyse en section 6.1.4). L'algorithme hybride de la MCA itère cycliquement sur les composantes de telle façon qu'à chaque itération, les atomes les plus saillants de chaque composante sont au fur à mesure intégrés et raffinés avec le seuil qui varie. Le seuil à la dernière itération est nul lorsque la décomposition est exacte (sans bruit) et $\tau\sigma$ lorsqu'un bruit AWGN contamine les observations, avec $\tau \in [3, 4]$ comme il est d'usage en débruitage.

6.1.2 Choix du dictionnaire : la clé pour une bonne séparation

Comme expliqué plus haut, le choix du dictionnaire joue un rôle central pour aboutir à une bonne séparation. Son choix comme union de sous-dictionnaires $\{\Phi_k\}_{k=1 \dots K}$ (voir fig. 1.3) peut être guidé par des arguments empruntés à l'analyse harmonique computationnelle discutés en section 1.1.4, et reflétant la diversité morphologique des structures géométriques dans l'image. Mais que faire si aucun dictionnaire fixe n'est disponible pour compresser le contenu d'une composante ; par exemple une texture naturelle complexe. Dans [A73, A112, S3], nous avons étendu le cadre de la MCA pour traiter le cas où le dictionnaire n'est pas entièrement connu a priori comme ci-dessus, mais constitué aussi de sous-dictionnaires locaux appris à partir d'exemples ou des données observées elles-mêmes.

6.1.3 Stratégies de variation du seuil

Dans [190, 191], un seuil variable selon une décroissance linéaire ou exponentielle a été préconisé. Toutefois, la vitesse de décroissance du seuil reste un choix qui n'est pas nécessairement trivial en pratique. Dans [A23], nous avons proposé deux stratégies adaptatives guidées par les observations :

- La première est basée sur un estimateur robuste de la variance du résidu marginal. Cette stratégie de seuil a été ensuite exploitée dans [81] dans le cas de l'échantillonnage compressé avec Φ aléatoire.
- Dans la seconde stratégie, et par simplicité, on considère $K = 2$; la généralisation pour $K \geq 2$ est directe. Soient $x_1^{(t)}$ et $x_2^{(t)}$ les composantes morphologiques estimées par l'algorithme de la MCA à l'itération t . Le résidu est alors $r^{(t)} = y - x_1^{(t)} - x_2^{(t)}$. Dans cette seconde stratégie appelée MOM (pour "Mean of Max"), le seuil à l'itération t est :

$$\lambda_t = \frac{1}{2} \left(\left\| \mathbf{T}_1 r^{(t)} \right\|_{\infty} + \left\| \mathbf{T}_2 r^{(t)} \right\|_{\infty} \right). \quad (6.3)$$

L'intuition sous-jacente est que les prochains atomes à sélectionner sont ceux du sous-dictionnaire Φ_k , et donc de la composante, où le résidu courant à l'itération t possède les plus grandes corrélations ; voir [A23] pour une explication plus détaillée.

6.1.4 Identifiabilité du support et convergence

L'algorithme de la MCA [190, 191, A23] s'est révélé performant pour la décomposition parcimonieuse des signaux et des images lorsque le dictionnaire est correctement choisi. Ces performances ont été confortées par un large panel d'expérimentations pratiques. Dans [A16], nous nous sommes attachés à en comprendre les raisons théoriques. Dans le cas sans bruit, une analyse des garanties (unicité, identifiabilité du support, convergence) de la MCA lorsque celle-ci est associée au seuillage MOM et que Φ est une union de bases orthonormées a été menée dans [A16]. Ces résultats peuvent être résumés comme suit.

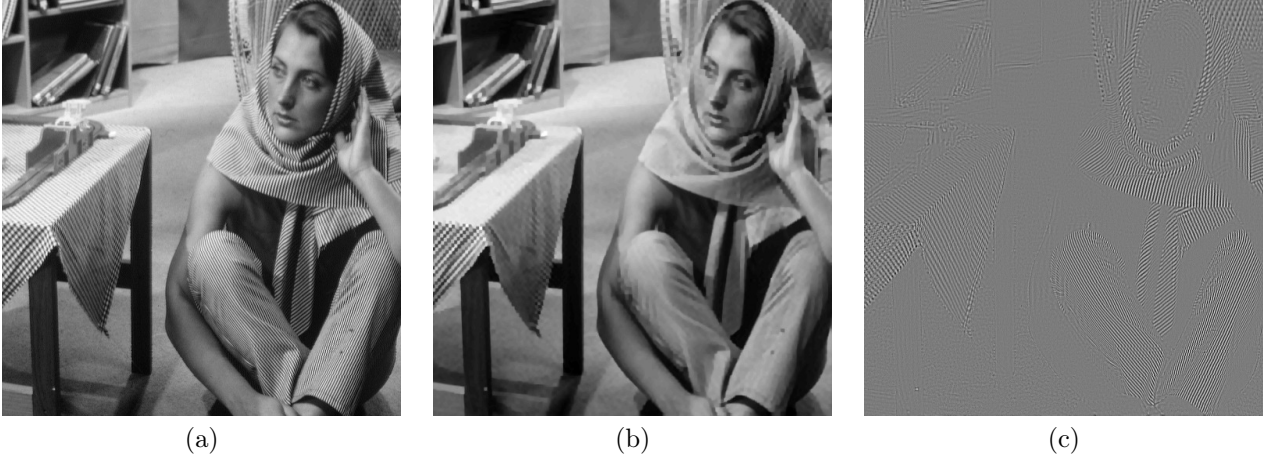


FIGURE 6.2 – Exemple d’application de la MCA à la décomposition cartoon+texture de l’image **Barbara** : (a) Image originale, (b) composante cartoon (curvelets), (c) composante texture (DCT locale).

Soit $y = \sum_{k=1}^K x_k = \sum_{k=1}^K \sum_{i \in \Lambda_k} \alpha_k[i] \Phi_k[\cdot, i]$, où Λ_k est le support de la k -ème composante morphologique et $\Lambda = \cup_{k=1}^K \Lambda_k$ (α_k) est le support total. A chaque itération, on peut écrire $r^{(t)} = \sum_{k=1}^K \tilde{\alpha}_k^{(t)}[i] \Phi_k[\cdot, i]$ et $r^{(0)} = y$. Notons

$$(k^*, i^*) = \underset{k, i \in \Lambda_k}{\operatorname{argmax}} \left| \tilde{\alpha}_k^{(t)}[i] \right|, \quad \tilde{\alpha}^* = \left| \tilde{\alpha}_{k^*}^{(t)}[i^*] \right|, \quad (k^\dagger, i^\dagger) = \underset{k \neq k^*, i \in \Lambda_k}{\operatorname{argmax}} \left| \tilde{\alpha}_k^{(t)}[i] \right|, \quad \tilde{\alpha}^\dagger = \left| \tilde{\alpha}_{k^\dagger}^{(t)}[i^\dagger] \right|.$$

Soit ρ tel que $\forall t \geq 0$, $\tilde{\alpha}^\dagger = \rho \tilde{\alpha}^*$ pour $0 < \rho < 1$. Le rapport ρ peut être interprété comme une mesure de contraste entre composantes morphologiques. On définit une généralisation du seuil MOM (6.3) comme la combinaison convexe :

$$\lambda_t = \gamma \left\| \mathbf{T}_1 r^{(t)} \right\|_\infty + (1 - \gamma) \left\| \mathbf{T}_2 r^{(t)} \right\|_\infty, \quad 0 < \gamma \leq 1. \quad (6.4)$$

Dans [A16], le résultat suivant est établi.

Théorème 4 Si $\operatorname{Card}(\Lambda) < \mu_\Phi^{-1}/4$ et $\rho \leq \min(1 - 2/(5\gamma), (1 - \gamma)/(3 - \gamma))$ pour $2/5 < \gamma \leq 1$, alors à chaque itération t et pour toute composante morphologique k^* , l’algorithme MCA avec le seuil (6.4) ne sélectionne aucun atome en dehors du vrai support Λ_{k^*} , et sélectionne au moins un atome du vrai support Λ_{k^*} pour entrer dans l’ensemble actif. De plus, l’algorithme converge linéairement à la bonne séparation des composantes et leur représentation la plus parcimonieuse dans $\Phi = [\Phi_1, \dots, \Phi_K]$.

Il est opportun de remarquer que la borne supérieure sur ρ atteint son maximum pour le seuil MOM ($\gamma = 1/2$). Ainsi, MCA-MOM est un corollaire du Théorème 4, auquel cas la condition suffisante sur ρ devient $\rho \leq 1/5$, tolérant de ce fait des composantes moins contrastées que pour les autres choix de γ . Par conséquent, MCA-MOM est capable de retrouver la bonne séparation si les composantes sont contrastées, et que leur représentation originale est suffisamment parcimonieuse dans un dictionnaire Φ incohérent.

Comme extension future de ce résultat, on peut penser à traiter le cas bruité où sous conditions suffisantes d’incohérence et de contraste, un résultat de consistance ℓ_2 sur les coefficients pourrait être établi. Nous pensons aussi qu’un résultat d’inclusion du support est à portée avec une condition supplémentaire de rapport signal sur bruit minimal (sur le(s) coefficient(s) le(s) plus faible(s)). Des raffinements de la condition d’incohérence vers d’autres critères (e.g. l’isométrie restreinte [42]) sont aussi envisageables.

La fig. 6.2 montre un exemple d’application de la décomposition parcimonieuse par l’algorithme MCA à l’image réelle **Barbara** (a)-(c). Le dictionnaire utilisé contient la DCT locale et les curvelets. Plusieurs autres exemples d’illustration et d’utilisation sont décrits dans [A13, A17].

6.2 Séparation de sources parcimonieuses

6.2.1 Le modèle de mélange linéaire instantané

Pour des données multicanales, le modèle de mélange linéaire instantané stipule que l'on dispose de N_c observations (canaux) $\{y_1, \dots, y_m\}$ où chaque y_j est un vecteur de taille N ; chaque canal est le mélange linéaire de N_s sources $\{s_1, \dots, s_{N_s}\}$, chacune de longueur N . Dans le cas bruité, ce modèle d'observation s'écrit

$$y_i[l] = \sum_{j=1}^{N_s} \mathbf{A}[i, j] s_j[l] + \varepsilon_i[l], \quad \forall i \in \{1, \dots, N_c\}, \forall l \in \{1, \dots, N\}, \quad (6.5)$$

où $\mathbf{A} \in \mathbb{R}^{N_c \times N_s}$ est la matrice de mélange de colonnes a_i , et ε_i est le bruit du canal i supposé borné. $\mathbf{A}$ définit le poids de contribution de chaque source à chaque mesure. Le modèle (6.5) prend aussi la forme matricielle

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{E} = \sum_{i=1}^{N_s} a_i s_i^T + \mathbf{E}, \quad (6.6)$$

où $\mathbf{Y} \in \mathbb{R}^{N_c \times N}$ est la matrice de mesures dont les lignes sont $y_i^T, i = 1, \dots, N_c$, et similairement pour la matrice des sources $\mathbf{S} \in \mathbb{R}^{N_s \times N}$ et de bruit $\mathbf{E}$.

Les techniques de séparation de sources visent à restaurer les sources originales $\{s_i\}_{i=1, \dots, N_s}$ ²⁸ en s'appuyant sur une mesure de diversité ou de contraste entre elles. En BSS, la matrice de mélange $\mathbf{A}$ et les sources $\mathbf{S}$ sont inconnues et doivent être estimées conjointement. En général, sans connaissances a priori supplémentaires, la séparation est un problème inverse mal posé. Nos travaux se sont attachés à étudier l'apport de la parcimonie et de la diversité morphologique pour le résoudre.

6.2.2 Décomposition parcimonieuse multispectrale

Supposons que $\mathbf{A}$ est connue. Dans ce cas, l'histoire de la décomposition parcimonieuse monocale discutée en section 1.3 peut être traduite dans le langage des données multicanales. Ceci est la conséquence d'une notion clé : le dictionnaire multicanal. En effet, supposons que chaque source s_i est parcimonieuse dans le dictionnaire Φ avec $s_i = \Phi \alpha_i$. Soit α la matrice $N_s \times T$ des coefficients dont les lignes sont α_i^T . De (6.5), les données $\mathbf{Y}$ peuvent être réécrites :

$$\mathbf{Y} = \mathbf{A}\alpha\Phi^T + \mathbf{E} = \sum_{i=1}^{N_s} \sum_{j=1}^T (a_i \varphi_j^T) \alpha_i[j] + \mathbf{E}. \quad (6.7)$$

Par conséquent

$$\text{vect}(\mathbf{Y}) = (\mathbf{A} \otimes \Phi) \text{vect}(\alpha) + \text{vect}(\mathbf{E}), \quad (6.8)$$

où $\otimes$ est le produit de Kronecker et l'opérateur vect concatène les colonnes de ses arguments dans un vecteur 1D. L'écriture (6.8) offre un éclairage intéressant : la parcimonie des sources dans Φ se traduit par une parcimonie des données multicanales dans le dictionnaire multicanal $\Psi = \mathbf{A} \otimes \Phi$. Ce concept de dictionnaire multicanal a aussi été remarqué dans [113].

En partant de (6.7) où le dictionnaire multicanal est supposé redondant, c.à.d. $N \cdot N_c < T \cdot N_s$, la décomposition parcimonieuse multicanale de $\mathbf{Y}$ consiste à résoudre

$$\min_{\alpha \in \mathbb{R}^{N_s \times T}} \sum_{i=1}^{N_s} \|\alpha_i\|_0 \quad \text{t.q.} \quad \|\mathbf{Y} - \mathbf{A}\alpha\Phi^T\|_F \leq \varsigma, \quad (6.9)$$

ou comme auparavant sa relaxation convexe

$$\min_{\alpha \in \mathbb{R}^{N_s \times T}} \sum_{i=1}^{N_s} \|\alpha_i\|_1 \quad \text{t.q.} \quad \|\mathbf{Y} - \mathbf{A}\alpha\Phi^T\|_F \leq \varsigma. \quad (6.10)$$

28. Ou de façon équivalente à décomposer une matrice $\mathbf{Y}$ en somme de N_s matrices de rang un comme le suggère (6.6).

Gâce à la forme vectorielle (6.8), certaines conditions d'identifiabilité dans le cas monocal de la section 1.3, peuvent être déclinées sans difficulté majeure dans le cas multicanal (6.10) [A16]. En particulier, pour les critères basés sur la cohérence, il suffit de remarquer que²⁹ :

$$0 \leq \mu_{\Psi} = \max(\mu_{\mathbf{A}}, \mu_{\Phi}) < 1. \quad (6.11)$$

D'autres critères d'identifiabilité peuvent aussi être établis, comme ceux basés sur le spark³⁰ en notant que $\text{spark}(\Psi) = \min(\text{spark}(\mathbf{A}), \text{spark}(\Phi))$, ou sur les constantes d'isométrie restreinte [125]. De la même façon, l'algorithme de la MCA peut prendre une forme multicanale [A16] et le Théorème 4 s'y applique directement en remplaçant verbatim Φ par Ψ .

Dans nos travaux jusqu'alors, il est opportun de remarquer que nous n'avons volontairement supposé aucune structuration de la parcimonie des différents sources. Cette structuration peut toutefois s'avérer fructueuse dans certaines applications où un atome actif l'est simultanément sur toutes les sources. Les conditions d'identifiabilité ℓ_1 basées sur une forme de cohérence prenant en compte cette forme de parcimonie jointe sont prouvées dans [113], voir aussi [114] pour des analyses au pire cas et en moyenne avec des algorithmes gloutons.

Remarque 1 *Il est tentant d'assimiler le problème de la décomposition parcimonieuse multicanale à celui des mesures multivariées (MMV) considéré par plusieurs auteurs, voir [A1, section 9.3.1] pour une revue détaillée. En dépit de certaines similarités, les deux modèles sont assez différents conceptuellement. Le modèle MMV n'a pas conscience d'un processus de mélange par $\mathbf{A}$. Son objectif est de restaurer α à partir de $\mathbf{Y} = \alpha\Phi^T + \mathbf{E}$ en favorisant des solutions où les lignes de α ont le même support commun en minimisant des pénalités de parcimonie sous la forme de normes mixtes $\ell_p - \ell_q$ pour $p \geq 1, 0 \leq q \leq +\infty$.*

6.2.3 Séparation aveugle de sources parcimonieuses

Dans ce cas, le modèle (6.5) est bilinéaire où $\mathbf{A}$ est inconnue et doit être estimée conjointement avec les sources dans $\mathbf{S}$. Le cadre de l'analyse en composantes morphologiques généralisée (GMCA) proposé dans une série de contributions [A5, A25, A68, A70], suppose a priori que chaque source est la superposition de K composantes morphologiques :

$$\forall i \in \{1, \dots, N_s\}; \quad s_i = \sum_{k=1}^K x_{i,k} = \sum_{k=1}^K \Phi_k \alpha_{i,k} = \Phi \alpha_i \quad \text{avec } \alpha_i = [\alpha_{i,1}^T, \dots, \alpha_{i,K}^T]^T, \quad (6.12)$$

où chaque $x_{i,k}$ est parcimonieuse dans Φ_k .

La GMCA pour la BSS vise à estimer $\mathbf{A}$ et les sources $\mathbf{S}$ les plus parcimonieuses dans le dictionnaire Φ . Ce qui correspond à la tâche ambitieuse de minimisation suivante

$$\min_{\mathbf{A}, \alpha_{1,1}, \dots, \alpha_{N_s,K}} \sum_{i=1}^{N_s} \sum_{k=1}^K \|\alpha_{i,k}\|_p^p \quad \text{t.q.} \quad \|\mathbf{Y} - \mathbf{A}\alpha\Phi^T\|_F \leq \varsigma \text{ et } \|a_i\|_2 \leq 1 \quad \forall i \in \{1, \dots, N_s\}, \quad (6.13)$$

où typiquement $p = 0$ ou $p = 1$ pour la forme relaxée convexe de la pénalité. En tout état de cause, que la pénalité soit convexe ou pas, le problème (6.13) reste non convexe. La contrainte ℓ_2 sur les colonnes de $\mathbf{A}$ lève l'ambiguïté classique de l'indétermination d'échelle dans le produit $\mathbf{AS}$ dans (6.6). Observons au passage que la MCA est un cas particulier du problème de la GMCA (6.13) lorsque $N_s = 1$ et $N_c = 1$.

Le programme (6.13) est non convexe et notoirement difficile à résoudre même pour des pénalités convexes, c.à.d. $p \geq 1$. L'algorithme de la GMCA tels que nous l'avons proposé dans [A25] s'inspire des mêmes préceptes que ceux de la MCA, à savoir la relaxation par blocs et le seuillage itératif. En effet, au vu de (6.6) et (6.12), en fixant $\mathbf{A}$, on se ramène à une décomposition parcimonieuse multicanale avec des composantes morphologiques multicanales $\left\{ (a_i \alpha_{i,k}^T) \Phi_k^T \right\}_{1 \leq i \leq N_s, 1 \leq k \leq K}$ pour laquelle l'analogie multicanal de la MCA s'applique. $\mathbf{A}$ est à son tour estimée par une descente de gradient projeté avec une solution unique dans le cas d'un mélange sur-déterminé correspondant à $N_c \geq N_s$.

²⁹. Les colonnes de $\mathbf{A}$ sont supposées normalisées.

³⁰. Voir la page 15 pour la définition du spark d'une matrice.

Pour $p = 1$, grâce à la séparabilité des pénalités sur α et des contraintes sur $\mathbf{A}$, on montre que l'algorithme de séparation converge vers un point stationnaire [A18]. Cependant, il n'est pas garanti que ce dernier soit un minimum local de (6.13), et ceci est encore moins le cas pour $p = 0$. Pour contourner ces limitations et afin d'améliorer la convergence en pratique et diminuer la sensibilité aux minima locaux, une méthode de continuation est appliquée avec un seuil décroissant. L'intuition est que les structures les plus significatives des sources sont d'abord restaurées et servent pour estimer la matrice de mélange. Ces estimations sont raffinées au fur et à mesure de la décroissance du seuil pour intégrer des détails de plus en plus fins sur les sources.

Toutefois, une compréhension plus profonde des propriétés théoriques de la GMCA pour la BSS (identifiabilité des sources et de la matrice de mélange, convergence et vitesse pour $0 \leq p \leq 1$ ou tout du moins pour $p = 1$, stabilité au bruit, etc.) reste encore un problème ouvert qu'il serait primordial d'examiner à l'avenir. La tâche s'annonce délicate compte tenu de la non-convexité et la non-différentiabilité du problème d'optimisation sous-jacent. Quelques éléments de réponse dans le cas sans bruit commencent à émerger sur le problème connexe de l'apprentissage parcimonieux de dictionnaire pour lequel une analyse de l'identifiabilité ℓ_1 locale est fournie dans [115]. Nous y reviendrons dans les perspectives.

Nombre de sources inconnu L'usage classique en BSS suppose que le nombre de sources N_s est connu et fixé a priori. Ceci peut s'avérer être une exception plutôt qu'une règle. Estimer N_s à partir des données est assez complexe et la littérature sur le sujet reste très limitée. On peut bien évidemment penser à des critères de sélection de modèle tels que le MDL utilisé dans [19], ou l'AIC, le BIC, etc. Dans [A5], nous avons décrit une méthode basée sur la parcimonie pour estimer N_s . L'approche consiste à résoudre une suite de problèmes (6.13) pour des rayons $\sigma(m)$ avec m croissant de 1 jusqu'au rang de $\mathbf{Y}$. Dans [A5], une discussion argumentée suggère de choisir $\sigma(m)$ comme la norme de Frobenius de l'erreur lorsque $\mathbf{Y}$ est approchée à partir de ses m plus grandes valeurs singulières.

La fig. 6.3 montre deux sources originales (gauche) et deux mélanges bruités SNR=10dB (milieu). La GMCA a été appliquée avec un dictionnaire contenant les curvelets et la DCT locale. Comme mesure quantitative de comparaison, nous montrons l'erreur sur l'estimation de la matrice de mélange. La fig. 6.3 droite compare l'erreur sur la matrice de mélange estimée par la GMCA à celles estimées par des algorithmes populaires de BSS en fonction du niveau de bruit en entrée : ICA-JADE [45], Relative Newton Algorithm [211] et EFICA [130]. Les algorithmes RNA et EFICA ont été appliqués sur les coefficients d'ondelettes des canaux. L'algorithme ICA-JADE montre rapidement ses limites en présence de bruit, alors que RNA et EFICA semblent assez similaires. GMCA les dépasse particulièrement en présence d'un fort bruit.

La GMCA peut aussi être appliquée au débruitage multicanal notamment d'images couleurs. Ceci est illustré en fig. 6.4 où l'image bruitée dans l'espace couleurs RGB est montrée en (a). L'image fig. 6.4(b) est celle débruitée en appliquant à chaque plan RGB indépendamment une méthode de seuillage dur dans le domaine de l'UDWT (voir Chapitre 4). Dans la fig. 6.4(c), la GMCA est d'abord appliquée avec un dictionnaire de curvelets pour estimer $\mathbf{A}$ et ainsi obtenir une estimation adaptative d'un espace couleur approprié à l'image. Ensuite, les sources obtenues par inversion de l'estimée de $\mathbf{A}$ (ici carrée de rang plein) sont débruitées indépendamment par la même méthode qu'avant, et finalement une estimée couleur est reconstruite en ré-appliquant $\mathbf{A}$. L'intérêt de débruiter dans l'espace couleur obtenu par la GMCA est clair et confirmé par plusieurs autres résultats décrits dans [A25, A5].

6.3 Récapitulatif et conclusion

Ce chapitre a été consacré à nos contributions exploitant la parcimonie et la diversité morphologique comme source de contraste pour la décomposition d'images et la BSS. En se basant sur ces deux ingrédients majeurs, deux algorithmes rapides à savoir la MCA et sa version généralisée multicanale GMCA ont été décrits et caractérisés. Ces algorithmes mènent à une meilleure séparation que d'autres alternatives de la littérature, et sont de surcroît robustes au bruit. Dans un cadre plus large, et même si ça n'a pas été décrit ici, la GMCA peut aussi servir de base pour résoudre efficacement certains problèmes inverses en

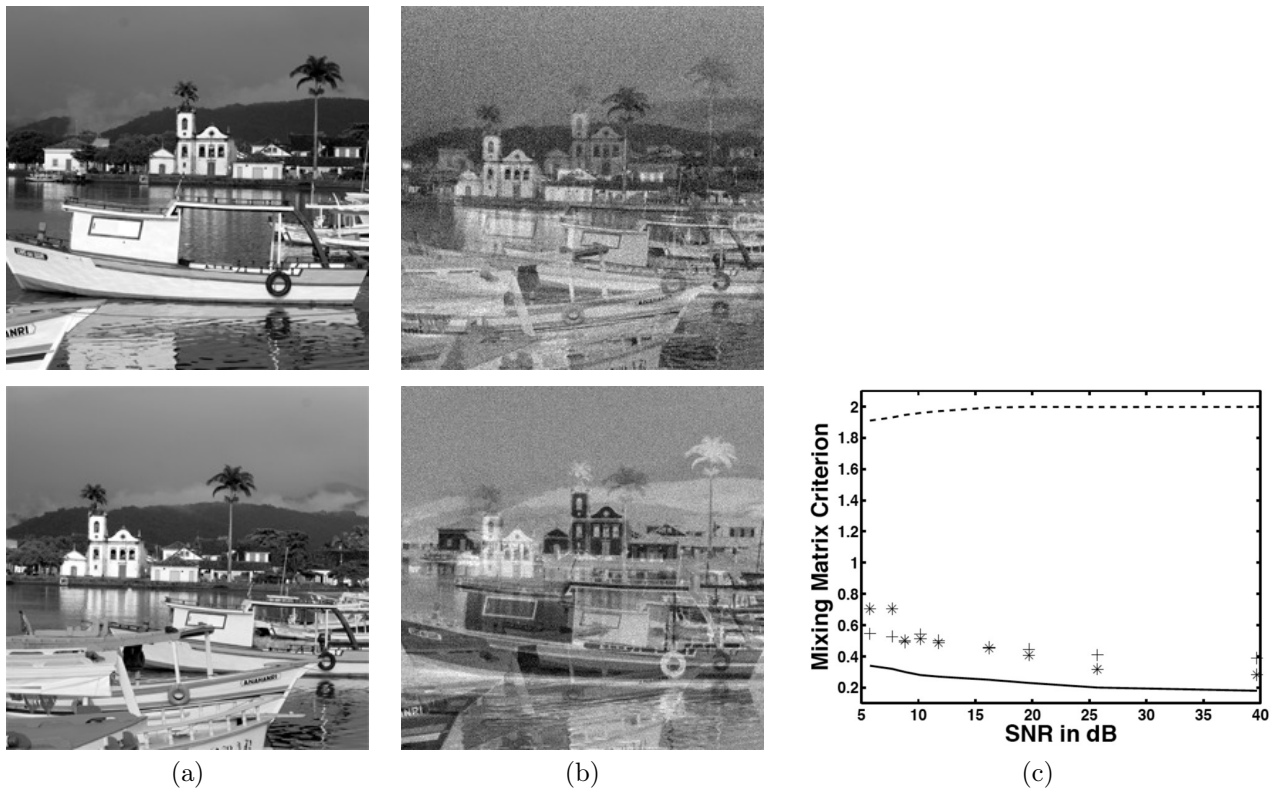


FIGURE 6.3 – Exemple de BSS avec $N_s = 2$ sources (a) et $N_c = 2$ mélanges bruités (b). (c) Evolution de l'erreur d'estimation de la matrice de mélange avec le SNR (GMCA (trait plein), ICA-JADE (trait pointillé), RNA (+), EFICA (*).

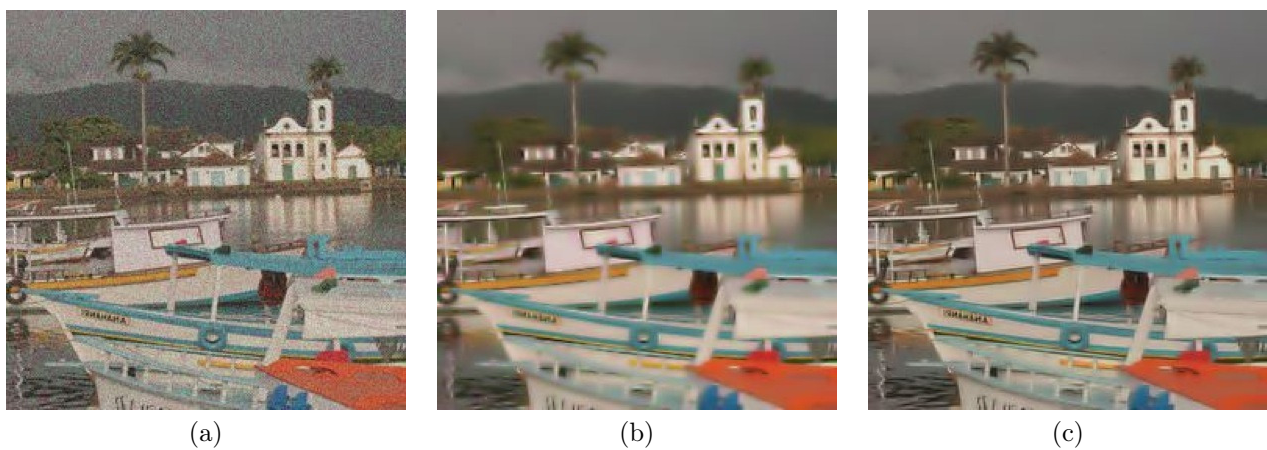


FIGURE 6.4 – (a) : Image RGB contaminée par un bruit AWGN SNR=15dB. (b) : Image débruitée en traitant indépendamment chaque composante couleur de l'espace RGB. (c) : Image débruitée dans l'espace couleur de la GMCA.

multicanal comme l'inpainting [A16, A70]. D'un point de vue applicatif, l'algorithme GMCA pour la BSS a été mis à profit en imagerie astronomique comme nous l'exposerons au Chapitre 9.

Il reste toutefois plusieurs problèmes ouverts dans le problème stimulant de la BSS avec parcimonie. Parmi eux, et comme nous l'avons souligné plus haut, une compréhension plus approfondie des garanties théoriques d'identifiabilité des sources et de la matrice de mélange à partir des mixtures observées dans la même veine que les résultats connus pour l'ICA, la robustesse en présence de bruit, ou encore l'extension à d'autres mélanges comme le modèle convolutif.

Chapitre 7

Echantillonnage compressé

Les travaux concernés par ce chapitre sont [A9, A49, A50, A58, A110, S4].

7.1 Rappels essentiels

La théorie de l'échantillonnage compressé (CS pour *compressed/ive sensing* ou *sampling*) a émergé en 2004 par les travaux parallèles accomplis indépendamment par [37, 43] et [71].

La théorie classique du traitement numérique du signal nous enseigne qu'un signal à bande limitée peut être reconstruit exactement à partir de ces échantillons si celui-ci est échantillonné à une cadence d'au moins deux fois sa fréquence maximale (théorème de (dés-)échantillonnage de Shannon(-Whittaker)). Ce principe est à la base de quasiment tous les protocoles d'acquisition utilisés dans les systèmes numériques modernes. Ce schéma d'acquisition exclut toutefois tout un pan de signaux qui ne sont pas à bande limitée mais qui restent explicables par relativement peu de degrés de liberté.

Le CS est un paradigme qui renouvelle la vision de Shannon de la théorie de l'échantillonnage pour les signaux parcimonieux ou compressibles. L'étape d'échantillonnage est rapide et linéaire ; elle consiste à corrélér le signal avec des vecteurs de mesures non adaptatives qui se partagent équitablement la structure du signal. La reconstruction ou le décodage du signal à partir de ces mesures est formulé comme un problème d'optimisation minimisant une pénalité de parcimonie sous contraintes. Nous nous focalisons ici sur les pénalités convexes (typiquement ℓ_1). D'autres algorithmes de décodage non linéaire avec des garanties théoriques existent, e.g. algorithmes de seuillage et gloutons, ou des décodeurs ℓ_p non convexes $0 \leq p < 1$, voir [A1, Chapitre 11][148] pour une revue plus détaillée. Ces derniers ne seront pas discutés dans la suite.

Le charme de la théorie du CS réside dans son approche interdisciplinaire voire transdisciplinaire. Elle emprunte des outils de plusieurs disciplines mathématiques incluant l'algèbre linéaire, la théorie des probabilités, la géométrie en haute dimension, l'analyse fonctionnelle, l'analyse harmonique appliquée, l'optimisation. Elle a aussi des ramifications et des conséquences en statistiques et en apprentissage, ou encore en théorie du signal et de l'information. A ce stade, nous voudrions aussi clairement couper court à une confusion récurrente dans ce domaine entre CS et reconstruction parcimonieuse (en anglais *sparse recovery*). En effet, bien que la parcimonie soit un ingrédient majeur du CS, ce dernier est d'abord un paradigme d'échantillonnage comme en témoigne son nom. La parcimonie comme régularisation est un domaine beaucoup plus large qui a existé avant le CS et a des implications bien au-delà comme nous avons tenté de le démontrer dans le long de ce document. De ce fait, il n'est pas rare de rencontrer des travaux de la littérature qui se réclament du CS sans en être, loin s'en faut.

7.1.1 Incohérence et parcimonie

La théorie du CS affirme qu'un signal peut être reconstruit à partir de nombre de mesures M bien en deçà de la dimension ambiante N , en s'appuyant sur deux principes :

- La parcimonie (stricte ou compressibilité) dans un dictionnaire Φ . Ce dernier est supposé orthonormé dans une grande majorité des travaux. La généralisation aux dictionnaires redondants est loin d'être triviale [175, 208].
- L'incohérence entre la matrice de mesure et Φ : ceci étend le principe d'incertitude d'Heinseberg, dans le sens où les signaux parcimonieux dans Φ doivent avoir été étalés dans le domaine des mesures. Les vecteurs de mesures ont ainsi une description dense dans Φ .

7.1.2 Paradigme d'échantillonnage

Le paradigme d'échantillonnage en CS consiste à acquérir la part importante du signal x sans le connaître à l'avance. Soit le modèle de mesure linéaire :

$$y = \mathbf{H}x = \mathbf{H}\Phi\alpha = \Psi\alpha, \quad (7.1)$$

où $y[i] = \langle h_i, x \rangle$, $\forall i = 1, \dots, M < N$ et les vecteurs de mesures $(h_i)_{1 \leq i \leq M}$ sont des lignes de la matrice $\mathbf{H}$ de rang M .

Puisque typiquement $M \ll N$ en CS, les questions fondamentales soulevées sont :

- Est-il possible d'avoir une reconstruction exacte (décompression ou décodage) à partir d'un petit nombre de mesures ? Si oui, combien de mesures sont nécessaires ?
- Si un tel décodeur existe, est-il numériquement implémentable ?
- Comment concevoir des modalités de mesure $\mathbf{H}$ capables de capturer non adaptativement l'information essentielle du signal ?
- Est-il possible d'avoir la stabilité et la robustesse lorsque l'on dévie (raisonnablement) du scénario idéal, e.g. signaux compressibles au lieu de signaux strictement parcimonieux, présence de bruit.

En fait, la réponse au point (i) consiste à résoudre un système linéaire sous-déterminé comme nous l'avons discuté en section 1.3 et section 6.1. Ceci peut se faire par exemple par la classe de décodeurs ℓ_1 :

$$\Delta_{\Psi}(y, \mathcal{C}) := \underset{\alpha \in \mathbb{R}^T}{\operatorname{argmin}} \|\alpha\|_1 \quad \text{t.q.} \quad \Psi\alpha - y \in \mathcal{C}, \quad (7.2)$$

où $\mathcal{C}$ est un convexe fermé traduisant les propriétés du bruit (c.à.d. $\mathcal{C} = \{0\}$ en absence de bruit, $\mathcal{C} = \mathcal{B}_2^S$ si le bruit est de variance finie, etc.). La section qui suit décrit une condition d'identifiabilité appelée la propriété d'isométrie restreinte (RIP) qui joue un rôle central en CS. Des variantes de cette condition apparaissent dans la littérature statistique sous d'autres dénominations, par exemple dans [25] et d'autres.

7.1.3 Propriété d'isométrie restreinte

Considérons Φ orthonormal et sans perte de généralité posons $\Phi = \mathbf{I} \Rightarrow \Psi = \mathbf{H}$ et $x = \alpha$ (parcimonie dans la base canonique). La propriété d'isométrie restreinte est définie comme suit.

Définition 6 ([42]) Ψ satisfait la propriété d'isométrie restreinte de constante $0 < \delta_K < 1$ si pour tout x tel que $|\operatorname{supp}(x)| \leq K$

$$(1 - \delta_K) \|\alpha\|_2^2 \leq \|\Psi\alpha\|_2^2 \leq (1 + \delta_K) \|\alpha\|_2^2, \quad (7.3)$$

où δ_K est la constante d'isométrie d'ordre K . (7.3) est équivalente à dire que pour tout α , la matrice de Gram de la restriction de Ψ au support de x possède un quotient de Rayleigh dans $[1 - \delta_K, 1 + \delta_K]$.

En d'autres termes, Ψ agit presque comme une isométrie sur les vecteurs K -parcimonieux, impliquant que de tels vecteurs ne peuvent vivre dans le noyau de Ψ . Plus récemment, des versions asymétriques de la RIP sont apparues, en particulier dans [99, 28]. Une condition de type RIP est assez forte puisqu'elle requiert de contrôler simultanément les supports de tous les vecteurs K -parcimonieux.

Dans [38], il a été montré que $\delta_{2K} < \sqrt{2} - 1$ ³¹ est une condition suffisante pour assurer non seulement l'identifiabilité par minimisation ℓ_1 , mais aussi la robustesse aux signaux compressibles et la stabilité au

³¹. La littérature du CS utilise δ_{2K} pour s'assurer que les éléments du noyau de Ψ seront au moins $2K + 1$ -parcimonieux, et ainsi toute autre solution candidate aura un support de cardinal au moins $K + 1$.

bruit, apportant ainsi une réponse positive à la question (iv) ci-dessus. Cette robustesse prend la forme d'une consistance dite ℓ_2 - ℓ_1 ³² pour reprendre la terminologie de [51] :

$$\|\alpha - \alpha^*\|_2 \leq \frac{C_1}{\sqrt{K}} \|\alpha - \alpha_K\|_1 + C_2 \sigma, \quad (7.4)$$

où $\{\alpha^*\} = \Delta_{\Psi}(y, \mathcal{B}_2^\sigma)$, α_K est la restriction de α à ses K plus grandes composantes, C_1 et C_2 sont des constantes finies. Ceci implique que les vecteurs $\alpha \in w\ell_q(C)$ pour $q < 1$, sont reconstruits avec une erreur $C_1 CK^{1/2-1/q} + C_2 \sigma$. La borne sur δ_{2K} a été améliorée par plusieurs auteurs en particulier $\delta_{2K} < 2(3 - \sqrt{2})/7 \approx 0.4531$ dans [99] ou $\delta_{2K} < 0.472$ dans [36] ³³. Dans [60], les auteurs montrent que pour $\delta_{2K} \geq 1/\sqrt{2}$, il existe au moins un vecteur K -parcimonieux non ℓ_1 -identifiable en absence de bruit. Ceci met en lumière que la marge de manoeuvre est assez réduite pour affiner les critères d'identifiabilité basés sur les constantes d'isométrie. C'est l'une des motivations derrière nos travaux [A9, A49] sur lesquels nous reviendrons en section 7.4.

7.1.4 Construire de bonnes matrices : Echantillonnage aléatoire

Pour répondre à la question (iii), il faut construire une matrice de mesure qui soit RIP d'ordre $2K$, c.à.d. qui vérifie (7.3) pour *toutes* les combinaisons $\binom{T}{2K}$ de vecteurs $2K$ -parcimonieux. C'est là où des arguments probabilistes rentrent en jeu, notamment sur les propriétés spectrales des matrices aléatoires. Les constructions déterministes quant à elles restent plus confidentielles hormis quelques tentatives [63, 117, 181].

En particulier, une matrice dont les composantes sont tirées indépendamment de l'ensemble (sous-) Gaussien isotrope satisfait avec grande probabilité les conditions d'identifiabilité ℓ_1 basées sur le RIP si $M \geq CK \log(T/K)$ [155]. D'autres constructions aléatoires sont possibles comme des sous-matrices constituées de lignes extraites aléatoirement d'une matrice orthonormale, e.g. matrice de Fourier [43, 178].

7.2 Algorithmes rapides pour les décodeurs convexes en échantillonnage compressé

Le domaine de l'échantillonnage compressé est très demandeur d'algorithmes efficaces pour résoudre des décodeurs du type (7.2). En absence de bruit, $\mathcal{C} = \{0\}$ et on retrouve la poursuite de base BP (1.9). En présence d'un bruit de variance finie σ^2 , $\mathcal{C} = \mathcal{B}_2^\sigma$, avec $\varsigma = \tau\sqrt{M}\sigma$, $1 \leq \tau < \|y\|_2 / (\sqrt{M}\sigma)$, ce qui est une instance particulière du problème (5.5) lorsque $\Psi(\alpha)$ est la norme ℓ_1 . Ce problème est aussi connu sous le nom BPDN [49] ³⁴, ou Lasso en statistiques. Peu de travaux se sont intéressés à résoudre exactement la forme contrainte du BPDN [38, 199]. Le sélecteur du Dantzig [44, 25] est aussi un autre problème difficile d'optimisation en CS qui est un cas particulier de (7.2) lorsque $\mathcal{C} = \{z \in \mathbb{R}^M \mid \|\Psi^T z\|_\infty \leq \lambda\}$ ³⁵, avec $0 < \lambda < \|\Psi^T y\|_\infty$. Le Dantzig peut être reformulé comme un programme linéaire.

Dans [A63, S1, A1], nous avons proposé de résoudre ces problèmes d'optimisation en utilisant le cadre proximal de décomposition du Chapitre 3. En effet, ceux-ci se résolvent facilement à l'aide du schéma de Douglas-Rachford. La norme ℓ_1 est coercive assurant que l'ensemble des minimiseurs est non vide, de domaine plein et son opérateur proximal est le seuillage doux. Si les contraintes sont telles que la condition de qualification est satisfaite, dans les cas discutés ici $0 \in \text{ri}(y + \mathcal{C}) - \text{Im}(\Psi)$, ceci assure la faisabilité. Il suffit finalement d'exprimer le projecteur sur $\mathcal{C}$ pour implémenter l'itération de Douglas-Rachford. Pour BP, le projecteur sur $\{\alpha \in \mathbb{R}^T \mid y = \Psi\alpha\}$ est bien évidemment :

$$\alpha + \Psi^T (\Psi\Psi^T)^{-1} (y - \Psi^T \alpha),$$

où Ψ est supposé surjectif pour que $\Psi\Psi^T$ soit inversible. Remarquons au passage que la surjectivité de Ψ rend inutile la condition de qualification pour BP, BPDN et Dantzig.

32. Nous n'évoquons pas ici la consistance $\ell_2 - \ell_2$ pour laquelle des conditions suffisantes existent mais uniquement dans un cadre probabiliste [65].

33. Très récemment, des conditions suffisantes sur δ_K (et non pas δ_{2K}) ont commencé à voir le jour [35].

34. Bien que ces auteurs l'aient considéré sous forme Lagrangienne augmentée.

35. C'est un convexe compact non vide correspondant au polytope dual de BPDN sous sa forme Lagrangienne.

Pour BPDN, le projecteur est obtenu par application directe du Théorème 2 avec $F = \iota_{\mathcal{B}_2^S}$. Quant au Dantzig, posons $\mathbf{G} = \Psi^T \Psi$ la matrice de Gram de Ψ et $u = \Psi^T y$. Par le Théorème 2 encore une fois, le projecteur sur $\{\alpha \in \mathbb{R}^T \mid \|u - \mathbf{G}\alpha\|_\infty \leq \lambda\}$ est obtenu avec $F = \iota_{\mathcal{B}_\infty^\lambda}$. Les projecteurs sur $\mathcal{B}_2^S$ et $\mathcal{B}_\infty^\lambda$ sont triviaux à calculer.

7.3 Echantillonnage compressé et bruit de quantification

7.3.1 Modèle de mesure

Dans un protocole d'acquisition réaliste, les mesures sont quantifiées par un convertisseur analogique-digital (ADC) donnant lieu au modèle de mesure :

$$y_q = Q_\theta(\mathbf{H}x) = \mathbf{H}x + \varepsilon, \quad (7.5)$$

où $y_q \in (\theta\mathbb{Z} + \frac{\theta}{2})^M$ est le vecteur des mesures quantifiées, $Q_\theta(\cdot)[i] = \theta \lfloor (\cdot)[i] / \theta \rfloor + \frac{\theta}{2}$ est l'opérateur de quantification dans $\mathbb{R}^M$ supposé uniforme de pas θ , et ε est le bruit de quantification uniforme $\varepsilon[i] \sim_{\text{iid}} \mathcal{U}([-\frac{\theta}{2}, \frac{\theta}{2}])$

L'approche classique en CS est de considérer ε comme un bruit Gaussien [43] ou un bruit uniforme [38] mais de se cantonner à résoudre BPDN avec une contrainte ℓ_2 [43, 38]. Cette approche n'est toutefois pas satisfaisante car elle n'est pas fidèle au modèle de mesure et ne garantit aucunement que le signal reconstruit requantifié reproduise exactement les mesures quantifiées observées. Quelques tentatives dans la littérature ont proposé de remplacer la contrainte ℓ_2 dans le décodeur BPDN par une forme de fidélité reflétant les propriétés du bruit de quantification (voir [S4] pour une revue détaillée). Mais aucune analyse des garanties théoriques de tels décodeurs n'a été donnée.

Dans [A58, S4], nous avons proposé une nouvelle classe de décodeurs convexes appelée BPDQ_p (Basis Pursuit DeQuantizer) d'ordre $2 \leq p \leq \infty$. Pour ce décodeur, l'analyse des garanties théoriques répondant aux quatre questions ci-dessus du CS ont été étudiées. La conclusion fondamentale de cette analyse tient dans le message suivant : pour un niveau de parcimonie donnée, si le nombre de mesures est plus grand qu'une valeur minimale qui croît avec p , résoudre BPDQ_p au lieu de BPDQ₂=BPDN engendre une réduction de l'erreur de quantification par un facteur $\sqrt{p+1}$. Le résultat surprenant de cette analyse est que BPDQ_p ne favorise pas $p = \infty$ contrairement à ce que dicte la conjecture habituelle puisque dans ce cas, nous avons établi que le nombre de mesures requis est exponentiel en la parcimonie.

7.3.2 Le décodeur BPDQ_p et faisabilité

Notre approche dans [A58, S4] consiste à retrouver un vecteur parcimonieux dans un dictionnaire orthonormal Φ ($\Phi = \mathbf{I}$ ici) à partir de mesures quantifiées au travers du décodeur BPDQ_p

$$\Delta_\Psi(y, p, \epsilon) = \underset{\alpha \in \mathbb{R}^T}{\operatorname{argmin}} \|\alpha\|_1 \quad \text{t.q.} \quad \|y - \Psi\alpha\|_p \leq \epsilon. \quad (\text{BPDQ}_p)$$

La contrainte de fidélité en norme ℓ_p autorise des bruits issus d'une distribution GGD, avec le cas uniforme pour $p = \infty$.

Le bruit étant supposé uniforme, l'inégalité de Hoeffding-Chernoff donne une estimation précise de ϵ assurant la faisabilité du vecteur α original avec une grande probabilité.

Lemme 3 *Pour $1 \leq p < \infty$, choisir*

$$\epsilon = \epsilon_p(\theta) := \frac{\theta}{2} (m/(p+1) + \kappa\sqrt{m})^{1/p}, \quad (7.6)$$

assure la faisabilité de α dans BPDQ_p avec une probabilité plus grande que $1 - e^{-2\kappa^2}$, avec $\lim_{p \rightarrow \infty} \epsilon_p(\theta) = \frac{\theta}{2}$.

7.3.3 Isométrie restreinte étendue

Avec la faisabilité établie, pour établir la consistance de de BPDQ_p, nous avons besoin de définir la propriété d'isométrie restreinte étendue d'ordre p (RIP_p) :

Définition 7 Une matrice $\Psi \in \mathbb{R}^{M \times T}$ satisfait la RIP_p (avec $p > 0$) d'ordre K et de constante $0 < \delta_K < 1$, s'il existe une constante $\mu_p > 0$ telle que

$$\mu_p \sqrt{1 - \delta_K} \|\alpha\|_2 \leq \|\Psi \alpha\|_p \leq \mu_p \sqrt{1 + \delta_K} \|\alpha\|_2, \quad (7.7)$$

pour tous les vecteurs K -parcimonieux $\alpha \in \mathbb{R}^T$.

Pour $p = 2$, on retrouve la définition originale de la RIP où μ_2 est absorbée dans la norme ℓ_2 des colonnes de Ψ .

Le résultat suivant prouvé dans [S4] donne une construction aléatoire de matrices satisfaisant la propriété RIP_p pour $2 \leq p < \infty$.

Proposition 2 Soit $\Psi \in \mathbb{R}^{m \times T}$ une matrice Gaussienne centrée réduite, c.à.d. ses entrées $\sim_{iid} \mathcal{N}(0, 1)$. Alors si $M \geq (p - 1)2^{p+1}$ pour $2 \leq p < \infty$ et $M \geq 0$ pour $p = \infty$, il existe une constante absolue $c > 0$ telle que

$$\Theta_p(M) \geq c \delta_K^{-2} (K \log(T/K) + K(1 + 12\delta_K^{-1}) + \log \frac{2}{\eta}), \quad (7.8)$$

avec $\Theta_p(M) = M^{2/p}$ pour $1 \leq p < \infty$, et $\Theta_p(M) = \log M$ pour $p = \infty$, Ψ est RIP_p d'ordre K et de constante δ_K avec une probabilité plus grande que $1 - \eta$. De plus, $\mu_p = \mathbb{E} [\|\xi\|_p]$ où $\xi \sim \mathcal{N}(0, \mathbf{I})$.

Nous avons montré par ailleurs que $\mathbb{E} [\|\xi\|_p]$, et donc μ_p , peut être évalué précisément et non asymptotiquement [S4, Lemme 1]. En conclusion, pour construire une matrice Gaussienne qui vérifie le RIP_p avec grande probabilité, il faut un nombre de mesures $\gtrsim (K \log(T/K))^{p/2}$ pour $2 \leq p < \infty$, tandis que cette borne devient exponentielle en $K \log(T/K)$ pour $p = \infty$. C'est cette évolution qui explique le fait que BPDQ_p ne favorise pas forcément $p = \infty$ même pour un bruit uniforme.

7.3.4 Consistance et réduction du bruit de quantification

Nous avons aussi établi dans [A58, S4] la consistance $\ell_2 - \ell_1$ du décodeur BPDQ_p sous la propriété RIP_p pour $2 \leq p < \infty$ [S4], répondant ainsi positivement à la question (iv) pour ce type de décodeur. Cette asertion est résumée dans le résultat suivant.

Théorème 5 Soit $\alpha \in \mathbb{R}^T$ un vecteur compressible et α_K sa restriction à ses K plus grandes composantes. Supposons que Ψ est RIP_p d'ordre S et de constante δ_S , pour $S \in \{K, 2K, 3K\}$ et $2 \leq p < \infty$. Alors l'unique minimiseur $\{\alpha_p^*\} = \Delta_{\Psi}(y, p, \epsilon)$ de BPDQ_p satisfait

$$\|\alpha_p^* - \alpha\|_2 \leq \frac{A_p}{\sqrt{K}} \|\alpha - \alpha_K\|_1 + B_p \frac{\epsilon}{\mu_p},$$

où A_p et B_p sont des constantes bornées dépendant uniquement de p et des δ_S , $S \in \{K, 2K, 3K\}$.

Ce théorème ne couvre toutefois pas le cas $p = \infty$ car la preuve mise en place utilise les propriétés géométriques et de régularité des espaces de Banach ℓ_p dans $\mathbb{R}^T$ pour $2 \leq p < \infty$. L'extension au cas $p = \infty$ nécessiterait d'autres outils.

Corollaire 1 Soit Ψ une matrice Gaussienne avec un nombre de mesures satisfaisant (7.8) pour $2 \leq p < \infty$. Soit $\{\alpha_p^*\} = \Delta_{\Psi}(y, p, \epsilon)$ l'unique minimiseur de BPDQ_p pour ϵ choisie selon (7.6). Alors

$$\|\alpha_p^* - \alpha\|_2 \leq \frac{A_p}{\sqrt{K}} \|\alpha - \alpha_K\|_1 + \frac{CB_p \theta}{\sqrt{p+1}}.$$

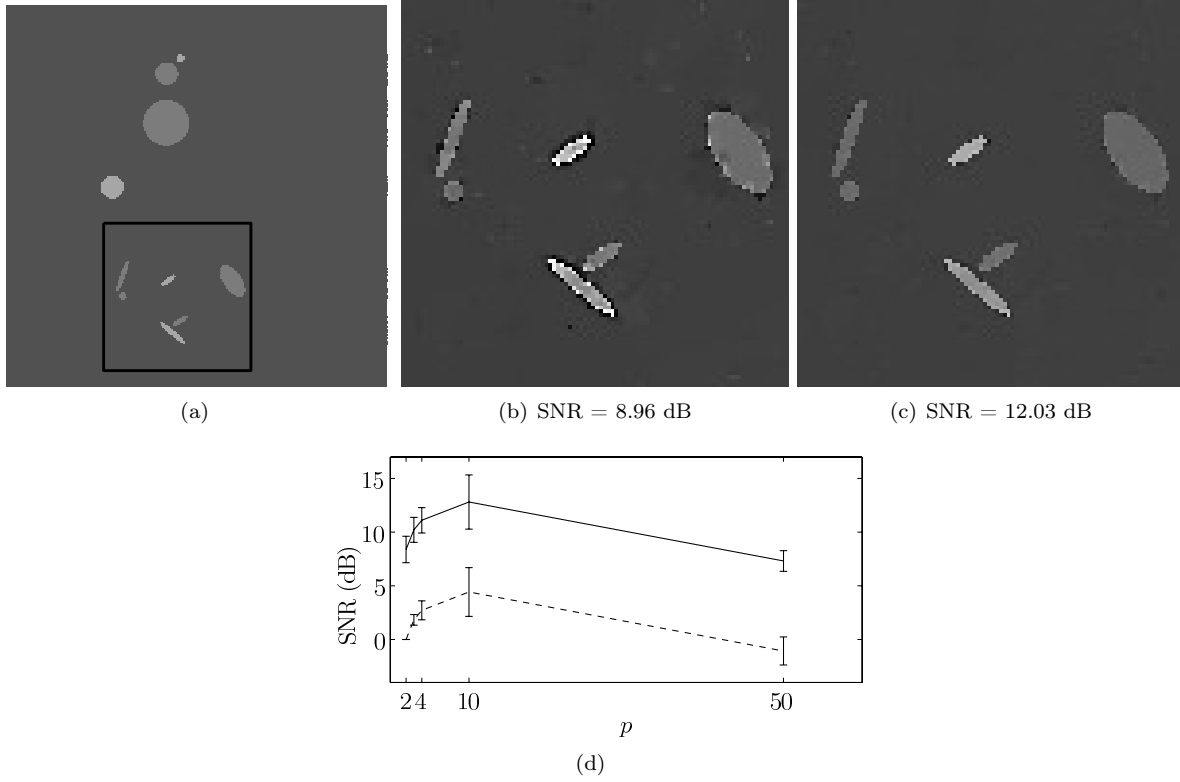


FIGURE 7.1 – Reconstruction d’une image simulée de $N = 512^2$ pixels à partir de $M = N/8$ coefficients aléatoires de Fourier, en minimisant la semi-norme TV. Les mesures ont été quantifiées avec un pas de $\theta = 50$. (a) Image originale, le rectangle montre la région du zoom de 100^2 pixels. (b) BPDN (zoom). (c) BPDQ₁₀ (zoom). (d) SNR moyen de BPDQ_p (trait plein) et amélioration par rapport à BPDN (pointillé) en fonction de p .

En clair, l’erreur de quantification décroît avec p . Il faut toutefois se garder de toute conclusion hâtive car par la Proposition 2, p ne peut pas être arbitrairement grand. En effet, pour un budget de mesures M/K donné, une matrice Gaussienne ne peut être RIP _{p} que pour $p \in [2, p_{\max}]$. Ainsi, pour M/K fixe, au-delà d’une valeur optimale $p_{\max}$, cette erreur se dégrade. C’est exactement cet effet que nous avons confirmé dans nos simulations (voir fig 7.1). Identifier automatiquement cette valeur de $p_{\max}$ pour un problème donné et un pas de quantification θ reste une question ouverte.

7.3.5 Algorithme de minimisation

Pour résoudre BPDQ_p, nous avons fait appel au schéma de Douglas-Rachford et au Théorème 2 comme nous l’avons fait aussi pour BPDN. Il suffit alors de pouvoir calculer le projecteur sur $\mathcal{B}_p^c$ pour $p \geq 2$. Hormis les cas $p = 2$ et $p = \infty$, ce projecteur ne s’exprime pas analytiquement. Pour $p \in (2, \infty)$, nous avons alors développé un algorithme de projection itératif très rapide résolvant les équation de Karush-Kuhn-Tucker par la méthode de Newton, ce qui évoque la stratégie en programmation quadratique séquentielle utilisée pour résoudre les problèmes contraints [2].

7.4 Exploration numérique des performances de l’échantillonnage compressé

Comme nous l’avons souligné en introduction, les conditions suffisantes d’identifiabilité basée sur une analyse au pire cas avec des critères du type RIP laissent peu de marge d’amélioration. De plus, les

bornes inférieures sur le nombre de mesures suffisantes pour l'identifiabilité basée sur le RIP cachent des constantes très pessimistes en pratique. D'ailleurs, ces bornes théoriques ne sont pas expliquées par les simulations numériques basées sur un échantillonnage Monte-Carlo comme il est d'usage dans la littérature.

Partant de ces constats, dans [A9, A49, A50], nous avons mis à l'épreuve les critères d'identifiabilité topologiques et RIP en explorant numériquement l'efficacité de la minimisation ℓ_1 pour retrouver des vecteurs parcimonieux depuis des mesures en CS, dans le cas sans bruit. Nous avons proposé un algorithme glouton qui calcule des vecteurs parcimonieux difficiles à retrouver par minimisation ℓ_1 . Cet algorithme est inspiré par des critères topologiques d'identifiabilité ℓ_1 . Contrairement à l'usage courant dans la littérature, nous avons évalué numériquement les performances théoriques (bornes de parcimonie, bornes inférieures sur les constantes d'isométrie) de ce type de critères d'identifiabilité ℓ_1 sans avoir à utiliser un échantillonnage Monte-Carlo, qui tend à éviter les cas pathologiques. Par exemple, pour $(T, M) = (4000, 1000)$ avec $\Phi = \mathbf{I}$ et $\mathbf{H} = \Psi$ Gaussienne, en utilisant la concentration des valeurs propres d'une matrice de Wishart, on peut montrer que la probabilité qu'une sous-matrice de 10 colonnes extraite aléatoirement de Ψ soit RIP avec $\delta_{10} \geq 0.58$ est inférieure à 4×10^{-6} . De telles matrices échappent à un échantillonnage Monte-Carlo classique alors que notre algorithme de recherche est capable de les trouver. Ceci révèle que l'analyse au pire cas basée RIP ne permet pas de prédire qu'un signal dont le support est de cardinal aussi petit que 5 est ℓ_1 -identifiable.

Dans la suite de cette section, nous supposons systématiquement que $\Phi = \mathbf{I}$. On notera ψ_i les colonnes de $\Psi = \mathbf{H}$.

7.4.1 Retour sur les critères topologiques d'identifiabilité ℓ_1

Le polytope centro-symétrique $\Psi(\mathcal{B}_1^1)$ est l'image de la boule ℓ_1 unitaire $\mathcal{B}_1^1$ par Ψ . C'est aussi l'enveloppe convexe de $\{\text{sign}(\alpha[i])\psi_i\}_i$. Soit $\Lambda = \text{supp}(\alpha)$. La $|\Lambda|$ -face $f_\alpha \subset \Psi(\mathcal{B}_1^1)$ sélectionnée par α est l'enveloppe convexe de $\{\pm\psi_i\}_{i \in \Lambda}$. Donoho [72] montre que

$$\alpha \text{ est } \ell_1\text{-identifiable} \iff f_\alpha \in \partial\Psi(\mathcal{B}_1^1) \quad (7.9)$$

où $\partial\Psi(\mathcal{B}_1^1)$ est la frontière du polytope $\Psi(\mathcal{B}_1^1)$ (voir fig. 7.2 pour une illustration pour $(T, M) = (3, 2)$). En utilisant (7.9), les auteurs dans [72, 80] déterminent, dans le cas sans bruit $y = \Psi\alpha$, une valeur asymptotique précise pour $\rho(\eta)$ dans (1.11) lorsque T and M sont grands. On a par exemple $\rho(1/2) \approx 0.089$ et $\rho(1/4) \approx 0.065$.

La fonction $\rho(\eta)$ induite par le critère topologique (7.9) est clairement plus fine que celle résultant du RIP. Ceci peut être interprété à la lumière de plusieurs arguments. Tout d'abord, celle du RIP résulte d'une analyse asymptotique avec une borne d'union, alors que celle du critère topologique utilise des résultats asymptotiques beaucoup plus fins. D'un autre côté, le critère topologique exploite pleinement la géométrie de la minimisation ℓ_1 alors que le RIP permet de conclure à l'identifiabilité avec d'autres décodeurs tels que les algorithmes gloutons. Le critère topologique n'est toutefois valable qu'en absence de bruit alors que le RIP assure la robustesse aux signaux compressibles et la stabilité au bruit si BP est remplacé par une variante appropriée.

7.4.2 Faces intérieures et vecteurs non identifiables

Une heuristique d'identifiabilité ℓ_1 La condition nécessaire et suffisante (7.9) dicte qu'un vecteur α non identifiable correspond à une facette f_α appartenant à l'intérieur du polytope $\Psi(\mathcal{B}_1^1)$. D'après [A9, Proposition 1], la distance de la $|\Lambda|$ -face f_α à l'origine est $1/\|d(\alpha)\|_2$, où

$$d(\alpha) = \Psi_\Lambda(\Psi_\Lambda^T \Psi_\Lambda)^{-1} \text{sign}(\alpha_\Lambda), \quad (7.10)$$

et Ψ_Λ est la sous-matrice de Ψ dont les colonnes sont indexées par Λ , et similairement pour α_Λ . Le vecteur $d(\alpha)$ joue un rôle central dans l'identifiabilité ℓ_1 exploitant le support et le signe [101, 83].

Ceci permet donc de conclure que le vecteur $d(\alpha)$ est bon indicateur d'identifiabilité ℓ_1 . Dans [A9, A50], nous avons montré par étude Monte-Carlo que la courbe de proportion de vecteurs identifiables en fonction de $d(\alpha)$ possède une transition de phase plus abrupte que la courbe paramétrée par la parcimonie. En

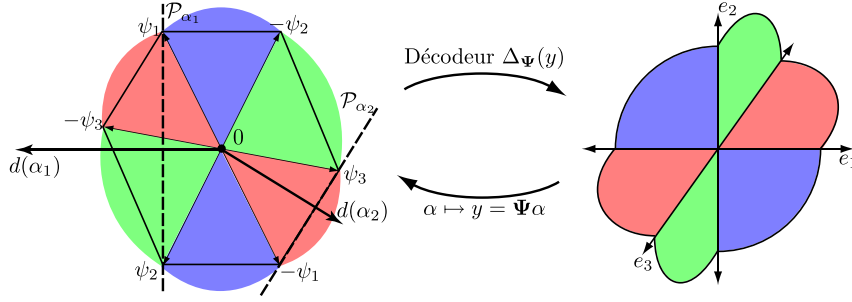


FIGURE 7.2 – Géométrie de l'identifiabilité ℓ_1 , pour $T = 3$ et $M = 2$. Le vecteur $\alpha_1 = (2, 3, 0)$ n'est pas identifiable car f_{α_1} est à l'intérieur du polytope $\Psi(\mathcal{B}_1^1)$, et a donc une grande valeur de $\|d(\alpha_1)\|_2$. Au contraire, $\alpha_2 = (-5, 0, 3)$ est identifiable car $f_{\alpha_2} \in \partial\Psi(\mathcal{B}_1^1)$, et a une petite valeur de $\|d(\alpha_2)\|_2$. Sur la droite, les vecteurs α 2-parcimonieux ℓ_1 -identifiables sont représentés par les orthants colorés. Sur la figure de gauche, leurs images par Ψ (des cônes) sont représentées par les mêmes couleurs.

conclusion, des vecteurs non identifiables peuvent être trouvés non seulement en augmentant la taille du support $\|\alpha\|_0 = |\Lambda|$, mais aussi en faisant décroître la valeur de $1/\|d(\alpha)\|_2$.

Une heuristique de conditionnement de sous-matrices Etant donné un vecteur $d \in \mathbb{R}^M$, on définit deux régions de $\mathbb{R}^M$:

$$\mathcal{C}_d = \{v \in \mathbb{R}^M \mid \langle d, v \rangle \geq 1\} \quad \mathcal{C}_d^c = \{v \in \mathbb{R}^M \mid \langle d, v \rangle \leq 1\}. \quad (7.11)$$

Dans le cas normalisé où les colonnes de Ψ sont de norme unité, $\mathcal{C}_d$ définit une double calotte sphérique, dont le rayon r vérifie $r^2 = 1 - 1/\|d\|_2^2$.

Dans [A9, A49], il est montré que le regroupement des vecteurs $\{\psi_i\}_{i \in \Lambda}$ dans ces régions permet de définir des bornes inférieures des constantes d'isométrie de Ψ_Λ . En particulier,

Proposition 3 Si $\{\psi_i\}_{i \in \Lambda} \subset \mathcal{C}_d$, alors

$$\delta_{|\Lambda|}^{\max} \geq |\Lambda| / \|d\|_2^2 - 1. \quad (7.12)$$

Si $\{\psi_i\}_{i \in \Lambda} \subset \mathcal{C}_d^c$ et si $d \in \text{Im}(\Psi_\Lambda)$ alors

$$\delta_{|\Lambda|}^{\min} \geq 1 - |\Lambda| / \|d\|_2^2. \quad (7.13)$$

où $\delta^{\min}$ et $\delta^{\max}$ sont les constantes d'isométrie gauche et droite dans (7.3).

Cependant, pour obtenir une estimée précise de $\delta_{|\Lambda|}^{\max}$ (respectivement $\delta_{|\Lambda|}^{\min}$), il faut identifier la région $\mathcal{C}_d$ contenant les colonnes de Ψ_Λ et qui correspond à la plus petite (respectivement la plus grande) valeur possible pour $\|d\|_2$. Le calcul de ce groupement optimal est difficile en grande dimension. Dans [A49], nous avons montré que le vecteur $d(\alpha)$ de (7.10) fournit une bonne approximation au vecteur d de ce groupement—voire optimale dans le cas normé. En clair, la région $\mathcal{C}_{d(\alpha)}$ est un choix optimal pour estimer les valeurs de $\delta_{|\Lambda|}^{\max}$ à l'aide de (7.12) si on se restreint à des régions passant par tous les vecteurs $\{\text{sign}(\alpha[i])\psi_i\}_{i \in \Lambda}$. De meilleures estimations sont possibles en utilisant une autre région $\mathcal{C}_d$ passant seulement par un sous-ensemble de ces vecteurs ou en utilisant un autre jeu de signes, mais leur calcul en haute dimension reste problématique.

7.4.3 Algorithme de poursuite glouton guidé par les faces

Extension de support minimale et maximale La discussion ci-dessus suggère que des bornes précises sur les constantes d'isométrie peuvent être obtenues en trouvant un vecteur α parcimonieux maximisant ou minimisant $\|d(\alpha)\|_2$. Parallèlement, maximiser $\|d(\alpha)\|_2$ est une bonne stratégie pour trouver des vecteurs non ℓ_1 -identifiables.

M	125	250	500	1000	M	250	500	1000	2000
$K_{\text{Topologique}}^*$	10	20	42	79	K_{RIP}^*	2	3	5	8
$\lceil \rho_{\text{Topologique}}(1/4)M \rceil$	9	17	33	65	$\lceil \rho_{\text{RIP}}(1/4)M \rceil$	1	2	3	6

TABLE 7.1 – Niveaux de parcimonie critiques révélés par nos algorithmes K^* , et ceux prédits par les bornes supérieures correspondant au critère topologique de [72] pour une redondance $T/M = 4$ avec $\rho_{\text{Topologique}}(1/4) \approx 0.065$, et au RIP avec $\rho_{\text{RIP}}(1/4) \approx 0.0027$ [28].

Dans [A49] puis [A9], nous avons proposé des algorithmes gloutons pour minimiser/maximiser la norme de $d(\alpha)$. Ces algorithmes consistent à partir d'un vecteur 1-parcimonieux et d'en étendre progressivement le support avec un choix judicieux du signe avec la mise à jour

$$\tilde{\alpha} = \alpha + \zeta \delta_i, \quad \zeta \in \{+1, -1\}, \quad (7.14)$$

pour $i \notin \Lambda$. La sélection de i et ζ optimaux se fait à la lumière du résultat suivant [A9] :

Proposition 4 Soit $\tilde{\Lambda} = \text{supp}(\tilde{\alpha})$, et $\tilde{\psi}_i \in \text{Im}(\Psi_{\tilde{\Lambda}})$ le vecteur dual tel que

$$\forall j \in \tilde{\Lambda}, \quad \langle \tilde{\psi}_i, \psi_j \rangle = \delta(i - j), \quad (7.15)$$

ou $\delta(i - j)$ est le delta de Kronecker. Alors

$$\|d(\tilde{\alpha})\|_2^2 = \|d(\alpha)\|_2^2 + \|\tilde{\psi}_i\|_2^2 |\langle d(\alpha), \psi_i \rangle - \zeta|^2. \quad (7.16)$$

Par conséquent, l'extension optimale qui minimise ou maximise $\|d(\tilde{\alpha})\|_2$ est celle qui minimise ou maximise $\|\tilde{\psi}_i\|_2 |\langle d(\alpha), \psi_i \rangle - \zeta|$ par rapport à i et ζ . Cependant, le calcul de $\|\tilde{\psi}_i\|_2$ pour tout $i \notin \Lambda$ est consommateur en temps de calcul. Nous avons alors suggéré de n'optimiser que le terme $|\langle d(\alpha), \psi_i \rangle - \zeta|$ ce qui a l'avantage d'être beaucoup plus rapide. Par exemple, pour maximiser $1/\|d(\tilde{\alpha})\|_2$ et ainsi trouver des vecteurs non identifiabiles, il suffit de choisir l'indice i minimisant $\langle d(\alpha), \psi_i \rangle$ et de le rajouter avec le même signe.

En se basant sur ces algorithmes gloutons d'extension du support, nous avons pu établir les performances numériques réelles du critère d'identifiabilité topologique et celui basé RIP. C'est ce qui est résumé dans les tableaux de Tab. 7.1 pour une redondance $T/M = 4$. Sur le tableau de gauche, sont montrées la borne supérieure de parcimonie prédite par [72] avec $\rho_{\text{Topologique}}(1/4) \approx 0.065$, et la limite numérique obtenue par notre algorithme minimisant $1/\|d(\alpha)\|_2$. A droite, sont données la borne de parcimonie du RIP avec $\rho_{\text{RIP}}(1/4) \approx 0.0027$ [28] et celle à partir des estimations de $\delta_{K^*}^{\min}$ et $\delta_{K^*}^{\max}$ fournies par nos algorithmes. Ces estimées sont montrées en fig. 7.3 et comparées à celles d'un algorithme glouton de poursuite des valeurs singulières des sous-matrices extraites. Ce dernier est lent car il procède par exploration gloutonne de l'ensemble des supports en exploitant sa structure de treillis ordonné par inclusion, en choisissant le chemin maximisant $\delta_{K^*}^{\min}$ et $\delta_{K^*}^{\max}$ des sous-matrices extraites à chaque étape. Les estimations de nos algorithmes sont remarquablement précises avec des temps 4 ordres de grandeur inférieurs à ceux de l'algorithme de poursuite des valeurs singulières.

7.5 Récapitulatif et conclusion

Ce chapitre a été consacré à l'échantillonnage compressé. Nos contributions dans ce domaine se situent à trois niveaux : les algorithmes rapides pour les décodeurs convexes, l'échantillonnage compressé avec mesures quantifiées et l'étude des performances numériques réelles des critères d'identifiabilité ℓ_1 . Plusieurs perspectives peuvent être esquissées pour chacun de ces sujets. Pour les décodeurs, il serait opportun de développer des algorithmes efficaces pour le CS mais avec des pénalités non convexes typiquement ℓ_q , $0 \leq q < 1$. Pour le bruit de quantification, nous avons déjà évoqué la nécessité d'adapter automatiquement la forme de la fidélité (c.à.d. valeur de $p_{\max}$) pour un problème donné et un budget de mesures ou de bits. Il serait tout aussi intéressant d'étendre l'analyse de (7.5) au cas plus complexe

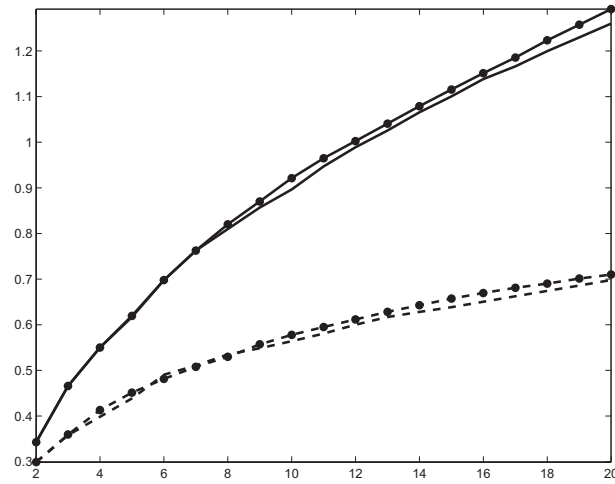


FIGURE 7.3 – Comparison des valeurs de $\delta_K^{\min}$ (pointillé) et $\delta_K^{\max}$ (plein) estimées par nos algorithmes et l’algorithme glouton force brute (●) en fonction de la parcimonie K pour $(T, M) = (2000, 500)$.

où un bruit thermique AWGN s’ajoute avant la quantification. L’exploitation des critères géométriques (notamment le vecteur $d(\alpha)$) et leur extension au cas compressible et/ou bruité est une autre perspective que nous avons commencé à explorer. Nous y reviendrons avec plus de détail au Chapitre 10.

Chapitre 8

Parcimonie et segmentation des textures

Dans ce chapitre, nous décrivons un cadre pour allier les représentations parcimonieuses et les contours actifs basés régions (RBAC) pour résoudre le problème de la segmentation des textures. Les contours actifs basés régions visent à obtenir une partition optimale du domaine d'une image en régions homogènes en minimisant une fonctionnelle composée d'un terme d'homogénéité (au sens large) donné de ces régions, et d'un second terme de régularité des frontières les séparant. Dans ce contexte, certaines transformées que nous avons utilisées dans les chapitres précédents ont été mises à profit comme descripteurs locaux pour les textures et servent de base pour construire le critère à minimiser dans les RBAC. De manière originale, le formalisme décrit est indépendant des transformées utilisées pour caractériser les textures, et s'applique ainsi de façon équivalente par exemple pour les ondelettes, la DCT locale ou les wave-atoms [62]. Ces travaux ont fait l'objet d'une partie de la thèse de F. Lecellier [TH3] ainsi que de [A64].

Même si ce n'est pas décrit ici en détail, nous mentionnons que le cadre général des contours actifs régions nous a permis de coupler élégamment des descripteurs statistiques régions et les RBAC. Ainsi, nous avons établi des résultats théoriques concernant la minimisation de fonctions de densités de probabilités paramétriques de la famille exponentielle. Nos travaux ont montré en particulier que la vitesse d'évolution dépend de la méthode utilisée pour estimer les hyperparamètres de la loi [A7]. Nous avons aussi proposé des critères de segmentation adaptés au modèle de bruit des images comme la vraisemblance [A7, A78], et d'autres critères statistiques comme l'entropie ou la divergence de Kullback-Leibler [A4, A59]. Ces critères photométriques (portant sur l'intensité du pixel) ont été avantageusement couplés à des termes d'a priori sur la forme utilisant les moments de Legendre [A79].

8.1 Contours actifs basés régions

8.1.1 Fonctionnelle à minimiser

La fonctionnelle générale dans le cadre de la segmentation variationnelle d'une image est une combinaison de termes basés région et de termes basés contour. Soit Ω_I un ouvert de $\mathbb{R}^2$ le domaine de l'image. Le problème de la segmentation en $K \geq 2$ régions disjointes se ramène à celui de la minimisation suivante

$$\min_{\Omega_1, \dots, \Omega_K \mid \bigcup_{k=1}^K \Omega_k \cup \partial\Omega = \Omega_I} F(\Omega_1, \dots, \Omega_K, \partial\Omega) = \sum_{k=1}^K \int_{\Omega_k} F_k(\mathbf{t}; \Omega_k) d\mathbf{t} + \int_{\partial\Omega} F_b(\mathbf{t}) ds(\mathbf{t}) , \quad (8.1)$$

où F_k est le critère associé à la région Ω_k , $\partial\Omega$ est la frontière des régions (contour en 2D, surface en 3D), et $s(\mathbf{t})$ est l'abscisse curviligne. Chaque critère basé région traduit une information globale sur celle-ci. Il peut par exemple exprimer l'homogénéité de la région comme l'écart quadratique à la moyenne, cette dernière étant fonction de la région. Le terme contour correspond à une information locale. Il peut être utilisé comme un terme de régularisation sur le contour $\partial\Omega$. Typiquement, ce terme correspond à la

longueur (la surface en 3D) du contour auquel cas $F_b(\mathbf{t}) = \lambda$, une constante positive qui joue le rôle d'un paramètre de régularisation.

8.1.2 Equation d'évolution et dérivées de forme

Outils de dérivation de domaine Une fois la fonctionnelle exprimée, il faut maintenant déduire une équation aux dérivées partielles (EDP) géométrique d'évolution qui régira le contour actif pour le conduire vers un point stationnaire de la fonctionnelle (qui peut s'avérer être un minimum local d'intérêt). Afin d'obtenir cette EDP, plusieurs méthodes de dérivation ont été développées dans la littérature. Nous nous pencherons particulièrement ici sur les outils de dérivation de domaine que nous avons utilisés. Nous en fournissons une description concise, le lecteur intéressé pourra se référer à [61] pour un ouvrage de référence et à [13] pour leur application aux RBAC.

Nous recherchons ici un domaine optimal, or l'ensemble des domaines de $\mathbb{R}^d$ n'a pas une structure d'espace vectoriel et les techniques classiques de dérivation ne s'appliquent donc pas. Les outils de dérivation de domaine consistent à reporter les variations de domaines sur des variations de fonctions appartenant à un espace vectoriel. On définit alors un champ de vitesse $\mathbf{V}$ correspondant aux variations de ces fonctions, et les variations d'une fonctionnelle de domaine $F_r(\Omega)$ sont calculées dans la direction $\mathbf{V}$.

Définition 8 La dérivée Eulérienne de la fonctionnelle de domaine $F_r(\Omega) = \int_{\Omega} F(\mathbf{t}; \Omega) d\mathbf{t}$ dans la direction du champ de vecteur $\mathbf{V}$ est définie comme :

$$\langle F'_r(\Omega), \mathbf{V} \rangle = \lim_{\tau \rightarrow 0} \frac{F_r(\Omega(\tau)) - F_r(\Omega)}{\tau} . \quad (8.2)$$

Définition 9 La dérivée de domaine de $F(\mathbf{t}; \Omega)$ dans la direction $\mathbf{V}$ notée $F'(\mathbf{t}; \Omega, \mathbf{V})$ est définie par :

$$F'(\mathbf{t}; \Omega, \mathbf{V}) = \lim_{\tau \rightarrow 0} \frac{F(\mathbf{t}; \Omega(\tau)) - F(\mathbf{t}; \Omega)}{\tau} . \quad (8.3)$$

Théorème 6 ([61, 122]) La dérivée Eulérienne de la fonctionnelle de domaine $F_r(\Omega)$ dans la direction $\mathbf{V}$ est reliée à la dérivée de domaine $F'(\mathbf{t}; \Omega, \mathbf{V})$ par :

$$\langle F'_r(\Omega), \mathbf{V} \rangle = \int_{\Omega} F'(\mathbf{t}; \Omega, \mathbf{V}) d\mathbf{t} - \int_{\partial\Omega} F(\mathbf{t}; \Omega) \langle \mathbf{V}, \mathbf{N} \rangle d\mathbf{a}(\mathbf{t}) , \quad (8.4)$$

où $\mathbf{N}$ est la normale unitaire intérieure à $\partial\Omega$ et $d\mathbf{a}(\mathbf{t})$ son élément d'aire.

Le premier terme provient de la variation du critère F par rapport à la région Ω , et le second correspond aux variations de la région elle-même.

EDP d'évolution Afin de déduire l'EDP d'évolution à partir des outils de dérivation de domaine, la dérivée Eulérienne de la fonctionnelle $F_r(\Omega)$ est supposée s'écrire :

$$\langle F'_r(\Omega), \mathbf{V} \rangle = - \int_{\partial\Omega} \text{speed}(\mathbf{t}; \Omega) \langle \mathbf{V}, \mathbf{N} \rangle d\mathbf{a}(\mathbf{t}) . \quad (8.5)$$

Il en est déduit l'EDP d'évolution du contour actif $\Gamma(\tau)$ assurant la descente de $F_r(\Omega)$ [122] :

$$\frac{\partial \Gamma}{\partial \tau} = \text{speed}(\mathbf{t}; \Omega) \cdot \mathbf{N} , \quad (8.6)$$

avec la condition initiale $\Gamma(\tau = 0) = \Gamma_0$. Dans le cadre des RBAC, dès lors que le premier terme dans (8.4) est transformé en une intégrale de contour, l'EDP d'évolution du contour actif se déduit directement.

8.2 Descripteurs de textures

La parcimonie et les dictionnaires globaux ou locaux d'atomes ont été utilisés récemment pour la synthèse des textures mais aussi comme descripteurs en vue de la segmentation ou la classification. Pour ces dernières tâches, une majorité des dictionnaires globaux mis en oeuvre correspondent à des bancs de filtres. ces bancs sont conçus afin d'être capables de discriminer les textures en fonction de leur réponse spectrale. Les ondelettes et les bancs filtres de Gabor sont certainement parmi les plus intensément utilisés pour la segmentation et la classification [67, 14], voir aussi l'état de l'art dressé dans [TH3].

Dans nos travaux, nous avons adopté un point de vue plus général en requérant que la transformée utilisée vérifie certaines propriétés, notamment, qu'elle soit multiéchelle, invariante par translation et éventuellement à sélectivité orientée. Les deux premières invariances ne sont pas sans rappeler les propriétés du modèle génératif aléatoire du Chapitre 2. Ainsi, nous considérons qu'une texture y est décrite de façon creuse par ses coefficients $\alpha_\gamma = \langle y, \varphi_\gamma \rangle$ dans un dictionnaire Φ , où $\gamma = (j, \theta, \mathbf{k})$, $j \geq 0$ est l'échelle, $\theta \in [0, 2\pi[$ est l'indice de l'orientation, et $\mathbf{k} = (k_1, k_2)$ est le vecteur de translation.

Par exemple, dans le cas de la transformée en ondelettes invariante par translation (UDWT), γ est tel que : $j \in \{0, \dots, J-1\}$, $\theta \in \{0, \pi/4, \pi/2\}$ et $\mathbf{k} \in \{0, \dots, N-1\}^2$. Pour les wave-atoms [62], on a : $j \in \{0, \dots, \lceil \frac{J-1}{2} \rceil + 1\}$, $\mathbf{k} \in \{0, \dots, N-1\}^2$ et θ est implicite à partir de la paramétrisation $\nu_{j,\mathbf{m}} = \pm \pi \mathbf{m} 2^j \sim 2^{2j}$, où $\nu_{j,\mathbf{m}}$ est la fréquence spatiale de la wave-atom, $\mathbf{m} = (m_1, m_2)$, (m_1, m_2) est une paire d'entiers tels que $\max_{i=1,2} m_i \sim 2^j$. Cette relation entre fréquence centrale et taille du support de la wave-atom traduit la loi d'échelle parabolique qui est au coeur du [62, Théorème 1.1] qui établit que les wave-atoms fournissent une représentation creuse optimale des textures localement oscillantes.

8.3 Segmentation des textures basée PDF

Les coefficients des transformées ont été utilisées dans notre cadre comme descripteurs capables d'isoler une texture parmi plusieurs en fonction de la réponse spectrale de celle-ci. Ainsi, pour comparer deux textures, il suffit de disposer d'une métrique discriminant leurs coefficients respectifs. Dans le cadre de nos travaux [A64, TH3], nous nous sommes focalisés sur celle mesurant l'écart entre les PDFs de ces coefficients. Sous hypothèse de stationnarité dans chaque sous-bande (j, θ) , ceci donne lieu à la fonctionnelle de domaine :

$$F_{j,\theta}(\Omega) = \int_{\mathcal{X}} \mathcal{D}(\text{pdf}_{j,\theta}^1(\alpha; \Omega), \text{pdf}_{j,\theta}^2(\alpha; \Omega)) d\alpha, \quad (8.7)$$

où $\mathcal{D} : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est la métrique comparant deux PDFs $\text{pdf}_{j,\theta}^1$ et $\text{pdf}_{j,\theta}^2$. Des choix typiques de $\mathcal{D}$ incluent la distance de Hellinger, le score χ^2 , la divergence de Kullback-Leibler (KL) ou sa version symétrique (auquel cas on parlera de distance de KL), etc.

En sommant sur toutes les sous-bandes le critère (8.7), la segmentation d'une image en zones texturées nécessitera donc d'optimiser par rapport au domaine des fonctionnelles de la forme :

$$F(\Omega) = \sum_{j,\theta} F_{j,\theta}(\Omega). \quad (8.8)$$

8.3.1 Segmentation supervisée

Nous entendons par segmentation supervisée le fait de disposer d'une texture de référence dont les PDFs des coefficients dans les différentes sous-bandes ont été apprises au préalable. Le but est donc de segmenter cette texture dans une nouvelle image. Dans ce cas, l'une des deux PDFs dans (8.7) (e.g. $\text{pdf}_{j,\theta}^2$) est celle de la référence $\text{pdf}_{j,\theta}^{\text{ref}}$ et de ce fait ne dépend pas de la région Ω .

8.3.1.1 Modèle paramétrique

Le chemin emprunté ici suppose que la transformée utilisée comme discriminant de la texture en fournit une description creuse sur *toutes* les sous-bandes. Par exemple, si on choisit comme PDF a priori

des coefficients la loi GGD discutée au Chapitre 2, et si la distance de KL (c.à.d. version symétrique de la divergence de KL) est utilisée, la fonctionnelle de domaine (8.8) devient :

$$F(\Omega) = \sum_{j,\theta} \left(\frac{\sigma_{j,\theta}^{\text{ref}}}{\sigma_{j,\theta}(\Omega)} \right)^{p_{j,\theta}(\Omega)} \frac{\Gamma\left(\frac{1+p_{j,\theta}^{\text{ref}}(\Omega)}{p_{j,\theta}^{\text{ref}}}\right)}{\Gamma\left(\frac{1}{p_{j,\theta}^{\text{ref}}}\right)} + \left(\frac{\sigma_{j,\theta}(\Omega)}{\sigma_{j,\theta}^{\text{ref}}} \right)^{p_{j,\theta}^{\text{ref}}} \frac{\Gamma\left(\frac{1+p_{j,\theta}^{\text{ref}}(\Omega)}{p_{j,\theta}^{\text{ref}}}\right)}{\Gamma\left(\frac{1}{p_{j,\theta}^{\text{ref}}}\right)} - \frac{1}{p_{j,\theta}(\Omega)} - \frac{1}{p_{j,\theta}^{\text{ref}}}, \quad (8.9)$$

où $p_{j,\theta}(\Omega)$ et $\sigma_{j,\theta}(\Omega)$ sont les paramètres de forme et d'échelle de la loi GGD imposée sur les coefficients à chaque sous-bande, $\Gamma(z)$ est la fonction Gamma [4]. Ils sont estimés dans le domaine d'où leur dépendance à Ω . Ces paramètres peuvent être estimés notamment par le biais de l'estimateur du maximum de vraisemblance ou la méthode des moments. Pour la méthode des moments, on a :

$$p_{j,\theta}^{\text{MO}}(\Omega) = h^{-1} \left(\frac{\left(M_{j,\theta}^{(1)}(\Omega) \right)^2}{M_{j,\theta}^{(2)}(\Omega)} \right), \quad \sigma_{j,\theta}^{\text{MO}}(\Omega) = \frac{M_{j,\theta}^{(1)}(\Omega) \Gamma\left(\frac{1}{p_{j,\theta}^{\text{MO}}(\Omega)}\right)}{\Gamma\left(\frac{2}{p_{j,\theta}^{\text{MO}}(\Omega)}\right)}, \quad (8.10)$$

où $h(p) = \frac{\Gamma^2(\frac{2}{p})}{\Gamma(\frac{1}{p})\Gamma(\frac{3}{p})}$, et $M_{j,\theta}^{(i)}(\Omega) := \frac{1}{|\Omega|} \int_{\Omega} |\alpha_{j,\theta}(\mathbf{t})|^i d\mathbf{t}$ est le moment empirique absolu d'ordre i . Des expressions similaires peuvent être développées pour l'estimateur du maximum de vraisemblance. Au-delà des questions d'existence et d'unicité de ces estimateurs³⁶, leur forme implicite rend l'obtention de la dérivée Eulérienne de $F(\Omega)$ dans le cas paramétrique très complexe. Avec les ondelettes, les auteurs de [14] obtiennent une EDP d'évolution par le calcul des variations et les fonctions annexes, mais en négligeant la dépendance des paramètres au domaine et sans s'attarder sur l'existence et l'unicité de l'estimateur. Par ailleurs, l'approche paramétrique montre ses limites dès que l'hypothèse de parcimonie est violée dans une sous-bande. C'est pour pallier ces limitations que nous avons opté pour une forme non paramétrique des PDFs.

8.3.1.2 Modèle non paramétrique

Estimateurs à noyaux L'estimation non paramétrique de densité est l'un des domaines les plus actifs de la théorie statistique. Le problème est le suivant : soit $\text{pdf}(\alpha)$ une densité (univariée ou multivariée) par rapport à la mesure de Lebesgue. On dispose d'un certain nombre d'échantillons qui sont des réalisations iid issues de $\text{pdf}(\alpha)$, et on aspire à estimer cette dernière. L'histogramme est typiquement l'estimateur du maximum de vraisemblance de pdf. De meilleurs estimateurs ont été proposés notamment les estimateurs à noyaux et ceux basés sur les séries orthogonales (e.g. ondelettes orthogonales). Les estimateurs à noyaux introduits dans les années 50 [177, 170] sont de loin les plus répandus. Soit K_{ς} un noyau positif d'intégrale unité et de largeur de bande ς (qui joue le rôle d'un paramètre de lissage). L'estimateur à noyaux de la $\text{pdf}_{j,\theta}^1$ dans (8.7) est :

$$\text{pdf}_{j,\theta}^1(\alpha; \Omega) = \frac{1}{|\Omega|} \int_{\Omega} K_{\varsigma}(\alpha_{j,\theta}(\mathbf{t}) - \alpha) d\mathbf{t}, \quad (8.11)$$

L'exemple typique de noyau est le noyau Gaussien auquel cas ς est l'écart-type. Dans le sens du risque quadratique, la forme du noyau d'Epanechnikov est optimale [91]. En réalité, le choix du noyau est bien moins important que celui de ς . En effet, le choix de ce paramètre résulte d'un compromis biais-variance dans l'étude du risque quadratique. En particulier, si la PDF est uniformément bornée et $(s+1)$ -continûment dérivable, alors le risque quadratique (global et ponctuel) de l'estimateur à noyaux a la vitesse classique $O\left(|\Omega|^{-2s/(2s+1)}\right)$ correspondant à un paramètre de lissage optimal $\varsigma_{\text{opt}} \asymp |\Omega|^{-1/(2s+1)}$.

Typiquement, $s = 2$ d'où une vitesse du risque quadratique en $O(|\Omega|^{-4/5})$ ce qui est plus lent que le taux paramétrique $O(|\Omega|^{-1})$, mais bien meilleur que le risque optimal en $O(|\Omega|^{-2/3})$ de l'histogramme.

Il convient de remarquer qu'un tel estimateur à noyaux peut paraître inapproprié pour des modèles de PDF parcimonieux tels que ceux discutés au Chapitre 1, puisque ces PDFs sont continues mais ne

36. Par exemple, l'estimé de p^{MO} existe si et seulement si $\frac{M^{(1)2}}{M^{(2)}} < 3/4$, auquel cas la solution est unique.

sont pas nécessairement différentiables en zéro, particulièrement à forte parcimonie ; voir section 2.1.1.1. Pour de telles densités irrégulières, des estimateurs par seuillage dans le domaine des ondelettes ont été récemment étudiées. Ils ne seront toutefois pas considérés ici à cause de leur forme non linéaire.

Expression des vitesses dans les RBACs Supposons que l'on désire segmenter une image contenant deux textures, où on dispose d'une référence pour chacune d'elles— les PDFs de ces références ayant été apprises au préalable. Ce problème de segmentation en deux régions Ω_{in} et Ω_{out} se ramène à la minimisation d'une fonctionnelle de la forme :

$$G(\Omega_{\text{in}}, \Omega_{\text{out}}) = \sum_{j,\theta} \int_{\mathcal{X}} \left(\underbrace{\mathcal{D}(\text{pdf}_{j,\theta}^{\text{in}}(\alpha; \Omega_{\text{in}}), \text{pdf}_{j,\theta}^{\text{ref},\text{in}}(\alpha))}_{F_{j,\theta}(\Omega_{\text{in}})} + \underbrace{\mathcal{D}(\text{pdf}_{j,\theta}^{\text{out}}(\alpha; \Omega_{\text{out}}), \text{pdf}_{j,\theta}^{\text{ref},\text{out}}(\alpha))}_{F_{j,\theta}(\Omega_{\text{out}})} \right) d\alpha \quad (8.12)$$

où $\text{pdf}_{j,\theta}^{\text{in}}$ et $\text{pdf}_{j,\theta}^{\text{out}}$ sont obtenues à partir de l'estimateur à noyaux (8.11). Dans [A64, TH3], le résultat suivant a été établi.

Théorème 7 *L'EDP d'évolution du contour actif $\Gamma(\tau)$ correspondant à (8.12) est :*

$$\begin{aligned} \frac{\partial \Gamma}{\partial \tau} = & \sum_{j,\theta} \left(F_{j,\theta}(\Omega_{\text{in}}) - F_{j,\theta}(\Omega_{\text{out}}) \right. \\ & + C_{j,\theta}(\Omega_{\text{out}}) - C_{j,\theta}(\Omega_{\text{in}}) \\ & \left. + (V_{j,\theta}(\Omega_{\text{in}}) - V_{j,\theta}(\Omega_{\text{out}})) * K_{\varsigma}(\alpha_{j,\theta}(\mathbf{t})) \right) \mathbf{N} , \end{aligned} \quad (8.13)$$

avec

$$\begin{aligned} C_{j,\theta}(\Omega_i) &= \int_{\mathcal{X}} \partial_1 \mathcal{D}(\text{pdf}_{j,\theta}^i(\alpha; \Omega_i), \text{pdf}_{j,\theta}^{\text{ref},i}(\alpha)) \text{pdf}_{j,\theta}^i(\alpha; \Omega_i) d\alpha, \\ V_{j,\theta}(\Omega_i) &= \partial_1 \mathcal{D}(\text{pdf}_{j,\theta}^i(\alpha; \Omega_i), \text{pdf}_{j,\theta}^{\text{ref},i}(\alpha)) , \end{aligned}$$

$\partial_1 \mathcal{D}(\cdot, \cdot)$ est la dérivée partielle de $\mathcal{D}(\cdot, \cdot)$ par rapport à sa première variable, et $\mathbf{N}$ est la normale unitaire au contour.

Les expressions littérales de $C_{j,\theta}$ et $V_{j,\theta}$ ont été établies dans [TH3] pour divers choix de $\mathcal{D}$, y compris la distance de Hellinger.

8.3.2 Segmentation non supervisée

Lorsqu'aucune référence n'est disponible sur les textures, il apparaît naturel de les mettre en compétition pour espérer aboutir à une segmentation satisfaisante. Avec le critère basé PDF, et dans le cas de la segmentation en deux régions, ceci conduit à la maximisation d'une fonctionnelle de la forme :

$$G(\Omega_{\text{in}}, \Omega_{\text{out}}) = \sum_{j,\theta} \int_{\mathcal{X}} \mathcal{D}(\text{pdf}_{j,\theta}^{\text{in}}(\alpha; \Omega_{\text{in}}), \text{pdf}_{j,\theta}^{\text{out}}(\alpha; \Omega_{\text{out}})) d\alpha. \quad (8.14)$$

La maximisation de cette fonctionnelle recherchera une partition dans laquelle les PDFs des coefficients dans les régions interne et externe sont les plus distinctes possibles, et ainsi permettra de segmenter les deux textures.

8.3.2.1 Modèle paramétrique

On peut adopter le point de vue paramétrique comme dans le cas non supervisé et déployer le même modèle paramétrique des PDFs. Toutefois, les difficultés auxquelles on s'était heurté dans le cas supervisé subsistent, notamment la dépendance implicite des hyperparamètres au domaine. A noter au passage que ceci reste vrai pour quasiment tous les modèles paramétriques de PDF connus tels que ceux décrits au Chapitre 2. C'est pour cette raison que notre effort s'est concentré sur le cas non paramétrique.

8.3.2.2 Modèle non paramétrique

Les PDFs dans les régions respectivement interne et externe sont remplacées par leurs estimées à noyaux (8.11). Avec les mêmes notations que celles du Théorème 7, la vitesse d'évolution du contour correspondant à la maximisation de (8.14) est résumée dans le résultat suivant prouvé dans [TH3].

Théorème 8 *L'EDP d'évolution du contour actif $\Gamma(\tau)$ pour maximiser (8.14) est :*

$$\begin{aligned} \frac{\partial \Gamma}{\partial \tau} = & \sum_{j,\theta} \left(\frac{1}{|\Omega_{\text{in}}|} \left(- (V_{j,\theta}(\Omega_{\text{in}}) * K_{\varsigma}) (\alpha_{j,\theta}(\mathbf{t})) + C_{j,\theta}(\Omega_{\text{in}}) \right) \right. \\ & \left. + \frac{1}{|\Omega_{\text{out}}|} \left((V_{j,\theta}(\Omega_{\text{out}}) * K_{\varsigma}) (\alpha_{j,\theta}(\mathbf{t})) - C_{j,\theta}(\Omega_{\text{out}}) \right) \right) \mathbf{N} , \end{aligned} \quad (8.15)$$

avec

$$\begin{aligned} C_{j,\theta}(\Omega_{\text{in}}) &= \int_{\mathcal{X}} \partial_1 \mathcal{D} \left(\text{pdf}_{j,\theta}^{\text{in}}(\alpha, \Omega_{\text{in}}), \text{pdf}_{j,\theta}^{\text{out}}(\alpha, \Omega_{\text{out}}) \right) \text{pdf}_{j,\theta}^{\text{in}}(\alpha, \Omega_{\text{in}}) d\alpha , \\ C_{j,\theta}(\Omega_{\text{out}}) &= \int_{\mathcal{X}} \partial_2 \mathcal{D} \left(\text{pdf}_{j,\theta}^{\text{in}}(\alpha, \Omega_{\text{in}}), \text{pdf}_{j,\theta}^{\text{out}}(\alpha, \Omega_{\text{out}}) \right) \text{pdf}_{j,\theta}^{\text{out}}(\alpha, \Omega_{\text{out}}) d\alpha , \end{aligned}$$

$$V_{j,\theta}(\Omega_{\text{in}}) = \partial_1 \mathcal{D} \left(\text{pdf}_{j,\theta}^{\text{in}}(\alpha, \Omega_{\text{in}}), \text{pdf}_{j,\theta}^{\text{out}}(\alpha, \Omega_{\text{out}}) \right) , \quad V_{j,\theta}(\Omega_{\text{out}}) = \partial_2 \mathcal{D} \left(\text{pdf}_{j,\theta}^{\text{in}}(\alpha, \Omega_{\text{in}}), \text{pdf}_{j,\theta}^{\text{out}}(\alpha, \Omega_{\text{out}}) \right) .$$

La distance de KL a été traitée en détail dans [A64].

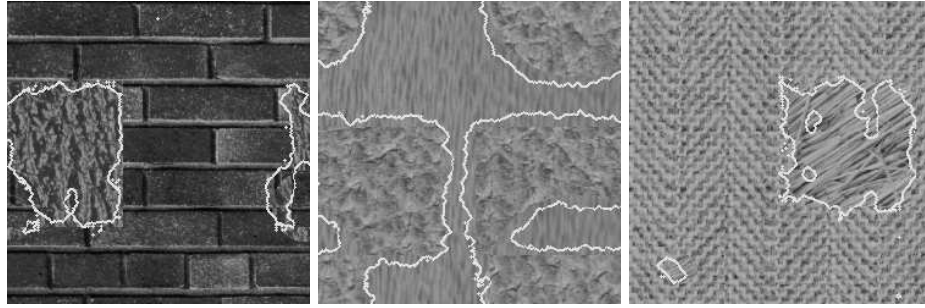
La fig.8.1 montre des exemples de segmentation non paramétrique d'images texturées avec deux transformées différentes (UDWT et wave-atoms) et les deux approches supervisée et non supervisée. Les images synthétiques sont générées en combinant aléatoirement des textures de la base de Brodatz. La segmentation est d'autant meilleure que le pouvoir discriminant de la transformée est plus grand. Nous avons remarqué que la meilleure sélectivité orientée des wave-atoms leur permet en général de segmenter des textures plus complexes que les ondelettes séparables. Ces dernières sont bien adaptées dès lors que les textures sont orientées verticalement, horizontalement ou diagonalement. Avec ce type de textures, la segmentation non supervisée (compétition de régions) associée à l'UDWT s'applique même à plus de deux textures dès lors qu'une texture orientée se distingue bien des autres. Ces constatations sont confirmées sur les images naturelles.

8.4 Récapitulatif et conclusion

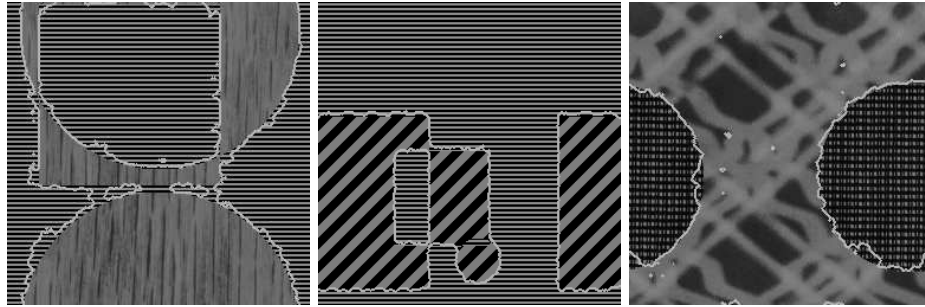
Dans ce chapitre, nous avons décrit une manière judicieuse combinant les représentations parcimonieuses pour obtenir des descripteurs, et les contours actifs basés régions pour la segmentation d'images texturées. Les vitesses dans les EDPs géométriques d'évolution sont obtenues à partir des fonctionnelles optimisées en utilisant les outils de dérivation de domaine. Notre cadre s'est focalisé essentiellement sur la segmentation basée sur un modèle non paramétrique des PDFs des coefficients qui servent de descripteurs des textures. Une perspective intéressante serait sans doute d'étendre l'approche à d'autres fonctionnelles ne mettant pas forcément en jeu des termes d'attache aux données relatifs aux PDFs des coefficients. Par exemple, pour une segmentation supervisée en K régions texturées, on peut minimiser une énergie du type

$$\min_{\{\alpha_k, \Omega_k\}_{k=1, \dots, K} \mid \bigcup_{k=1}^K \Omega_k \cup \partial\Omega = \Omega_I} \sum_{k=1}^K \left(\frac{1}{2} \int_{\Omega_k} \left| y(\mathbf{t}) - \sum_i \alpha_k[i] \varphi_{k,i}(\mathbf{t}) \right|^2 d\mathbf{t} + \lambda_k \|\alpha_k\|_1 \right) + \int_{\partial\Omega} F_b(\mathbf{t}) ds(\mathbf{t}) ,$$

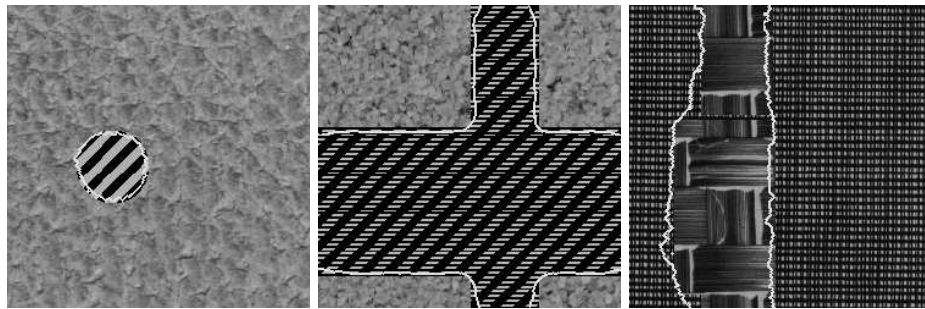
où les $(\varphi_{k,i})_i$ sont les atomes d'un dictionnaire Φ_k connu (d'où la supervision). En clair, cette énergie visera une partition où chaque texture dans la région Ω_k est parcimonieuse dans le dictionnaire Φ_k . Il serait par ailleurs important d'explorer le cas où la texture n'est plus efficacement décrite par un dictionnaire global mais plutôt local appris au préalable (voir section 6.1.2). Nous avons récemment amorcé un travail dans ces directions.



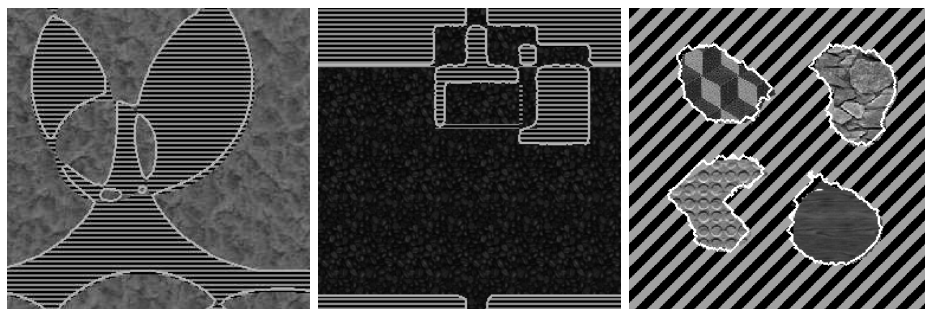
(a) Images synthétiques : Supervisée (wave-atoms).



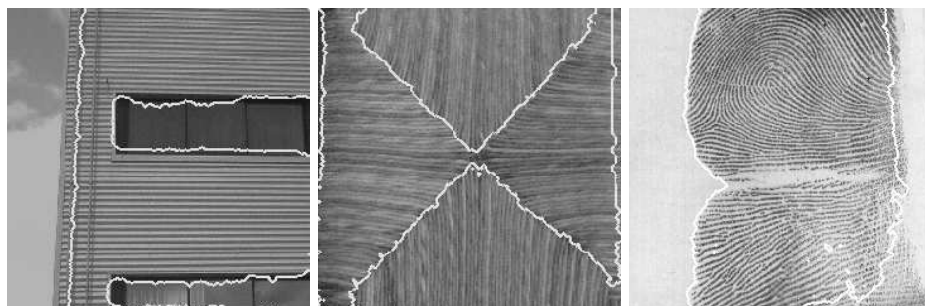
(b) Images synthétiques : Supervisée (UDWT).



(c) Images synthétiques : Non supervisée (wave-atoms).



(d) Images synthétiques : Non supervisée (UDWT).



(e) Images naturelles : Non supervisée (UDWT et wave-atoms).

FIGURE 8.1 – Exemples de segmentation non paramétrique de textures avec différentes transformées comme descripteurs. (a)-(d) : Images synthétisées à partir de la base de Brodatz. (e) : Images naturelles.

Chapitre 9

Applications et recherche reproductible

9.1 Applications

Dans une large mesure, plusieurs des travaux de recherche décrits dans ce document ont été initiés avec une application en ligne de mire. Ces applications ont été non seulement à l'origine de questions théoriques fondamentales, mais ont nécessité une approche qu'on peut qualifier d'interdisciplinaire voire de transdisciplinaire puisqu'il a souvent fallu s'imprégner aussi de la littérature du domaine d'application en question. Les applications auxquelles on s'est attaqué proviennent essentiellement de l'imagerie médicale, biomédicale et astronomique, et dans une moindre proportion du contrôle non destructif et de la photographie computationnelle. Ces applications ont été l'occasion de tisser plusieurs liens de collaborations avec des acteurs régionaux, nationaux et internationaux. Ces collaborations ont été formalisées par plusieurs projets de financement nationaux et internationaux (voir l'Annexe A pour un descriptif complet). Dans la suite nous ne décrivons volontairement que quelques exemples représentatifs. Les travaux concernés par ces applications sont les parties applicatives des thèses [TH1, TH2, TH3], ainsi que [A4, A11, A14, A17, A18, A21, A24, A28, A29, A31, A32, A34, A35, A36, A37, A38, A40, A41, A42, A45, A44, A39, A43, A47, A60, A56, A66, A67, A71, A75, A82, A79, A77, A86, A89, A92, A88, A93, A100, A98, A97, A99, A101, A103, A102, A104, A106, A105, A107, A108, A115, A118, A119, A120].

9.1.1 Imagerie médicale et biomédicale

Les applications médicales s'inscrivent dans plusieurs collaborations tissées avec des partenaires régionaux (Cyceron³⁷, CHU de Caen), nationaux (Institut Pasteur), et internationaux (Department of Psychiatry et Wolfson Brain Imaging Centre à l'Université de Cambridge, Biomedical Imaging Group à l'EPFL).

Analyse spatio-temporelle des séries d'IRM fonctionnelle L'imagerie par résonance magnétique fonctionnelle (IRMf) est une modalité non invasive désormais très répandue en neurosciences modernes pour l'étude indirecte de la dynamique spatio-temporelle de l'activité cérébrale. L'accomplissement d'une tâche cognitive ou sensorielle produit une augmentation localisée de l'activité neuronale qui se traduit par un changement local simultané du flux et du volume sanguins cérébraux, mais dans des proportions différentes. En effet, l'augmentation du flux, et donc de l'apport en oxygène, est bien supérieur à celui du volume sanguin cérébral. Ce déséquilibre engendre une diminution locale de la concentration de la déoxyhémoglobine qui sert alors d'agent de contraste endogène. En effet, cette dernière étant une substance paramagnétique, sa baisse de concentration se traduit par une augmentation relative du signal mesuré en IRM entre un état de tâche et un état de contrôle. C'est que l'on appelle le contraste BOLD (pour Blood Oxygenation Level Dependent).

Un protocole expérimental typique en IRMf consiste à acquérir par un imageur IRM une séquence rapide de volumes 3D à intervalles réguliers (dont la durée est appelée temps de répétition qui joue le rôle

37. Centre d'Imagerie Cérébrale et de Recherche en Neurosciences.

de la période d'échantillonnage, typiquement de l'ordre de la seconde) alors que le sujet est soumis à un paradigme expérimental qui alterne entre phases de stimulus et phases de repos ou de contrôle. La question neuroscientifique est alors de savoir si les phases de stimulus induisent des activations (significatives) dans le cortex cérébral et si oui dans quelles zones.

Beaucoup de signaux biomédicaux réels notamment ceux en IRMf sont non stationnaires et/ou de nature stochastique fractale. C'est justement dans cette non-stationnarité qu'est véhiculée l'information précieuse de ces signaux. Les représentations temps-fréquence et temps-échelle s'adaptent particulièrement à l'analyse de ce type de signaux. Depuis ma thèse et mon séjour post-doctoral, j'ai poursuivi mon effort en mariant la théorie statistique pour l'estimation et la détection et les approches temps-échelle, pour fournir des techniques puissantes et originales de traitement des séries spatio-temporelles. Ainsi, le modèle de régression paramétrique en présence de bruit à longue mémoire (section 4.2.2), le modèle semi-paramétrique partiellement linéaire (section 4.2.1), la caractérisation et la synthèse d'une classe de processus stochastiques localement stationnaires et à longue mémoire [A100] ont été intensément testés et validés sur une multitude de données IRMf, voir [A28, A29, A31, A32, A36, A40, A42, A89, A93, A117] et les deux articles de revue [A34, A35]. Ces travaux se sont inscrits dans le cadre du projet international Human Brain Project intitulé "Wavelets, fractals and fMRI of brain adaptivity", et financé par le NIH³⁸ de 2003-2008 (voir les détails en Annexe A et l'adresse <http://www-bmu.psychiatry.cam.ac.uk/projects/nih-wavelets/>). La fig. 9.1 résume une application poussée de ces méthodes d'estimation à l'étude de certaines pathologies chez l'homme.

Nous avons aussi récemment proposé une nouvelle approche exploratoire d'analyse des séries d'IRMf dites événementielles s'appuyant sur la parcimonie [A71, S7]. Pour ce faire nous avons introduit une nouvelle base d'ondelettes (en fait des splines exponentielles que nous avons dénommées activelets), construite à partir de la fonction de Green de l'opérateur traduisant la réponse hémodynamique impliquée dans le contraste BOLD. Ainsi cette réponse hémodynamique³⁹ observée en IRMf en réponse à un stimulus, a naturellement une représentation creuse dans cette base. Nous avons ensuite formulé le problème d'estimation de l'activité cérébrale événementielle comme un problème inverse mal posé avec une régularisation favorisant la parcimonie et résolu par les algorithmes du Chapitre 3. Les résultats obtenus se sont révélés très bons aussi bien sur des séries simulées que réelles, et ce sans connaissance a priori du stimulus. La fig. 9.2 résume l'application de cette approche à des données IRMf.

Correction d'artéfacts en IRMf-EEG L'algorithme de la MCA pour la décomposition parcimonieuse (Chapitre 6) a été appliqué à des signaux acquis lors d'expériences bi-modales de neuroimagerie. Lors d'une telle expérience, des données synchronisées d'électro-encéphalographie (EEG) et d'IRMf sont enregistrées pour l'étude de la dynamique spatio-temporelle de l'activité cérébrale. L'EEG apportant une meilleure résolution temporelle complémentaire de la résolution spatiale de l'IRMf. Toutefois, une telle acquisition se heurte à une difficulté majeure puisque des courants électriques perturbateurs sont induits sur les électrodes EEG à cause des gradients du champ magnétique lors de l'activation de l'imageur IRM. Ceci engendre des artéfacts importants qui dominent alors le signal EEG. Dans ce protocole expérimental, une étape de traitement des signaux EEG recueillis est nécessaire au préalable pour se débarrasser de ces artéfacts indésirables et rendre le signal EEG exploitable.

Un exemple d'un tel signal est montré en fig. 9.3(a). On peut clairement distinguer les périodes où l'imageur IRM est activé avec de fortes oscillations induites sur le signal EEG d'amplitude 10 à 20 fois celle du signal EEG attendu. La forme caractéristique localement périodique de ces courants induits par l'IRM sur l'EEG est issue des gradients de champs encodés dans la séquence IRM (les durées de ces oscillations sont connues). Par conséquent, en vue de "nettoyer" le signal EEG de ces oscillations IRM, l'algorithme de la décomposition parcimonieuse de la MCA a été appliqué avec un dictionnaire redondant contenant une TF locale et l'UDWT (pour le signal EEG très localisé). Les deux composantes sont montrées en fig. 9.3(c)-(d), et leur somme (débruitée donc) en fig. 9.3(b). Comme effet bénéfique secondaire, le signal a aussi été débruité.

38. NIH : National Institute of Health. L'agence américaine de recherche en sciences de la santé.

39. Cette réponse peut-être interprétée comme la PSF du système hémodynamique en contraste BOLD. Elle présente des variabilités spatiale, temporelle et inter-individuelle.

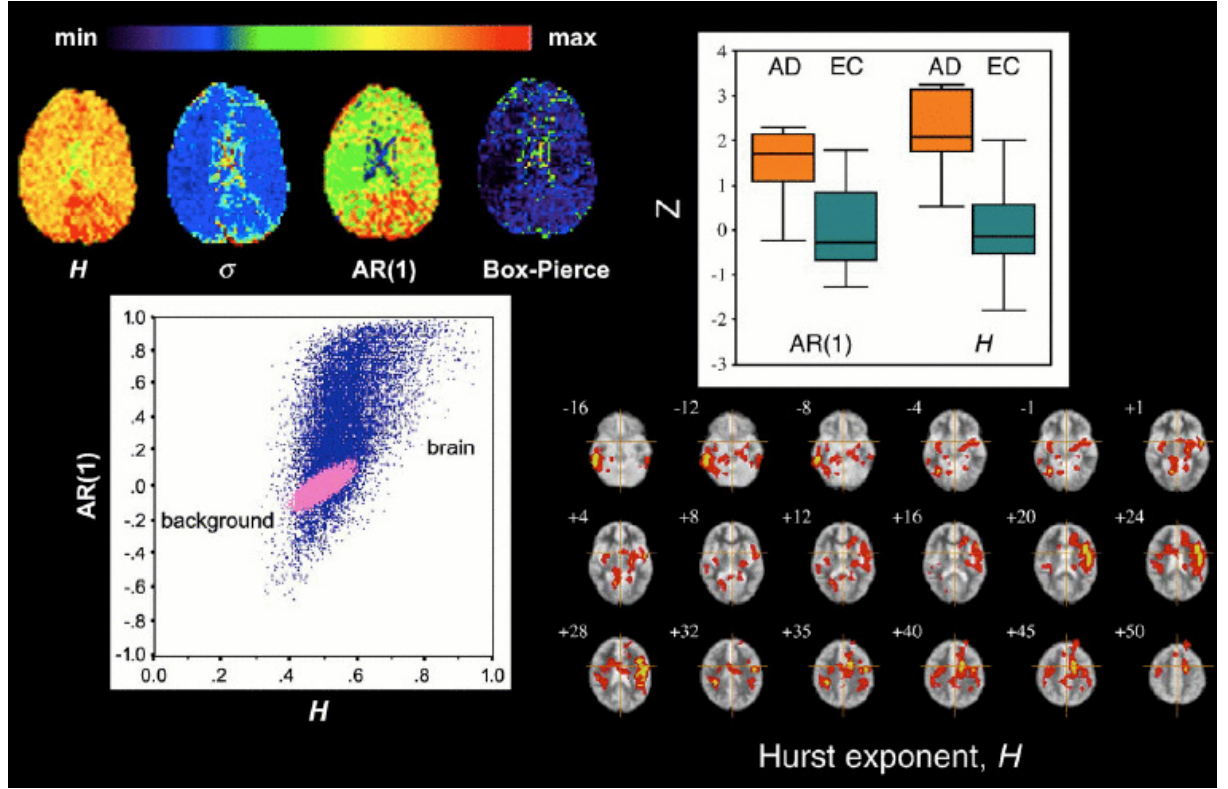
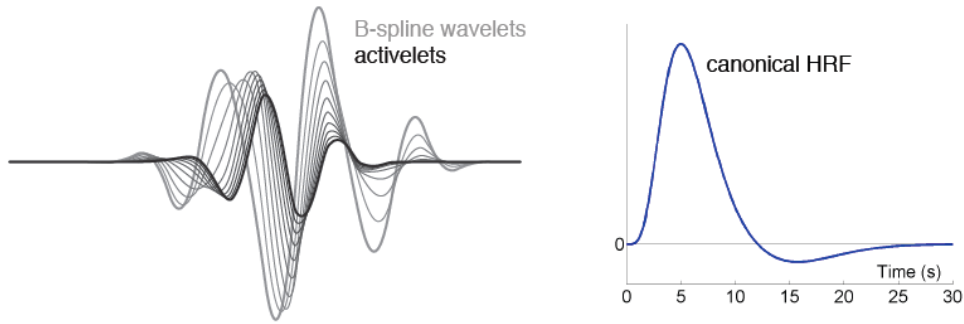
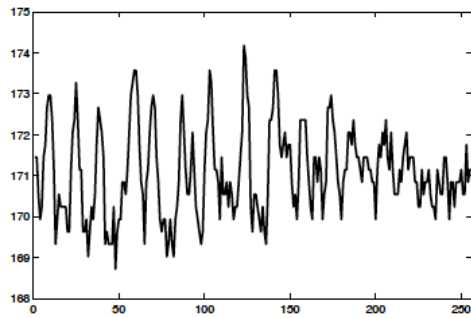


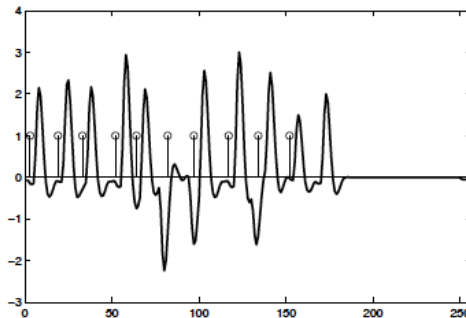
FIGURE 9.1 – Exemple d’application de l’estimation paramétrique multiéchelle avec bruit longue mémoire à l’étude de certaines pathologies neuro-dégénératives chez l’homme par IRMf. Les voxels avec un grand exposant de Hurst (voir section 4.2.2) (ou paramètre $AR(1)$) tendent à se concentrer symétriquement sur les régions corticales. Les nuages de points roses et bleus correspondent aux estimations de l’exposant de Hurst sur le fond et dans le cerveau, montrant ainsi que le bruit instrumental (fond) est blanc, et confirmant que le signal IRMf dans le cerveau est à longue mémoire. Les cartes avec zones colorées montrent les localisations des différences significatives de l’exposant de Hurst entre patients avec pathologies (AD) et un groupe de contrôle (EC). Ces zones se situent sur le lobe temporal latéral et médial et le cortex prémoteur, et sur le gyrus pré et postcentral. Ces anomalies révélées par notre modèle de régression et les estimateurs associés ont confirmé des études précédentes mais avec une sensibilité accrue. Cette application a par ailleurs montré l’intérêt de l’exposant de Hurst comme indicateur privilégié pour la compréhension de la pathophysiologie de certaines pathologies neuro-dégénératives [A29].



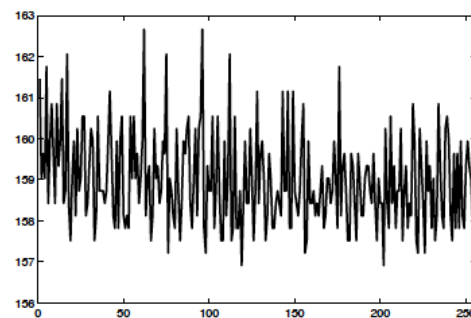
(a) Activelets et forme d'une réponse hémodynamique typique.



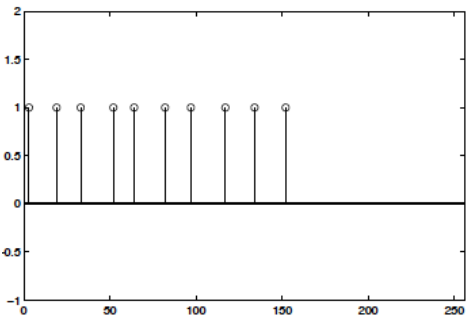
(b) Signal temporel d'un voxel dans une région cérébrale activée.



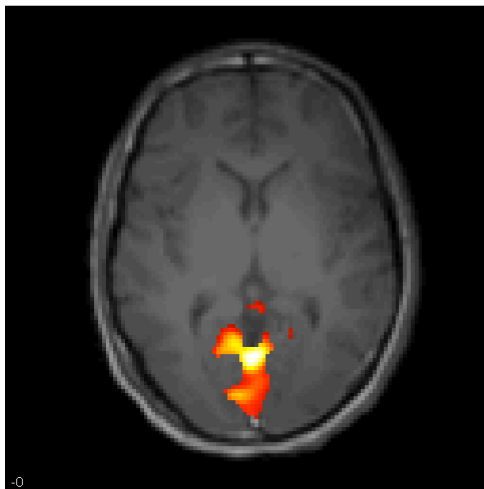
(c) Activité estimée par les activelets.



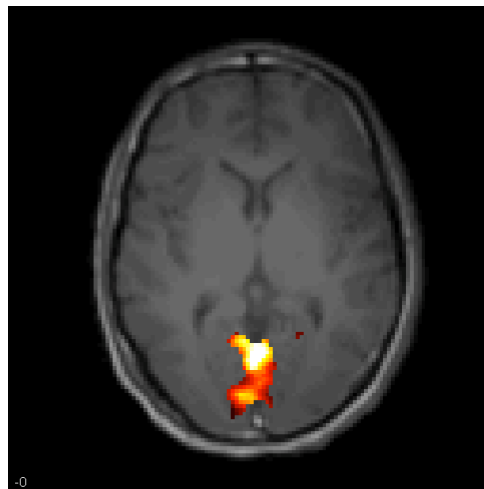
(d) Signal temporel d'un voxel dans une région cérébrale non activée.



(e) Activité estimée par les activelets.



(f) Carte d'activité cérébrale estimée par l'approche classique de la littérature (stimulus connu).



(g) Carte d'activité cérébrale estimée par les activelets (stimulus inconnu).

FIGURE 9.2 – Exemple d'application de la reconstruction parcimonieuse avec un dictionnaire d'activelets à l'analyse exploratoire et l'estimation de l'activité cérébrale dans les séries d'IRMf. Les impulsions de Dirac sur (d) et (e) correspondent au stimulus supposé inconnu dans notre approche.

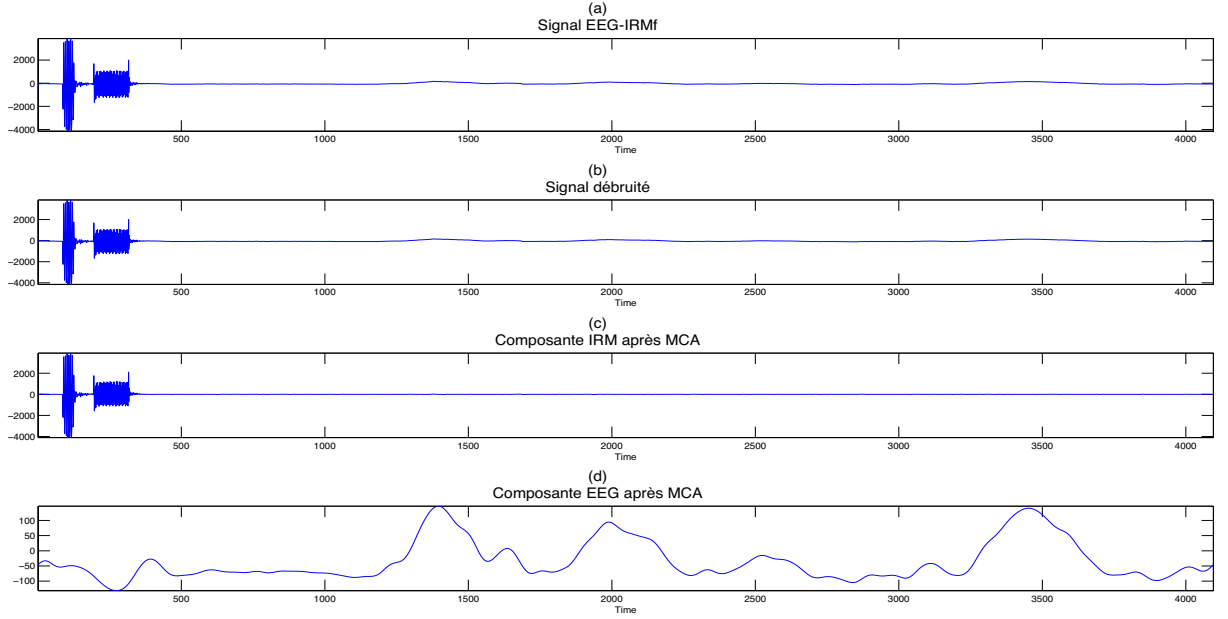


FIGURE 9.3 – Exemple d’application de la décomposition parcimonieuse à la corrections d’artéfacts sur un signal bi-modal IRMf-EEG. Le dictionnaire contenait la TF locale et l’UDWT.

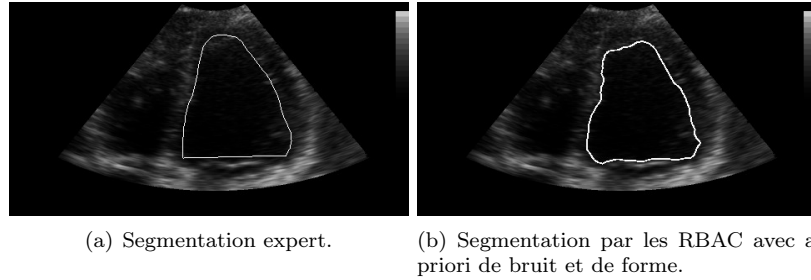


FIGURE 9.4 – Exemple d’application de la segmentation par RBAC sur des images d’échocardiographie.

Segmentation par contours actifs d’IRM et d’échocardiographies Les contours actifs basés régions (RBAC) associés à des descripteurs statistiques régions adaptés au modèle de bruit ont été appliqués à la segmentation de plusieurs modalités dont l’échocardiographie [A79, A109] et l’IRM [A4]. Ces exemples d’application sont illustrés en fig. 9.4 et 9.5.

Restauration en microscopie confocale La fig. 9.6 illustre une application des algorithmes de déconvolution proposés dans [A66, TH1] pour le bruit de Poisson (voir section 5.3) à une image réelle de neurones observés par microscopie confocale. Le dictionnaire utilisé dans cette application contenait les curvelets. Les épines dendritiques, montrées dans les régions zoomées, sont les structures d’intérêt. Leurs détails sont clairement restitués dans les images restaurées.

9.1.2 Imagerie astronomique

L’application sur les images astronomiques rentre dans le cadre d’une collaboration très étroite et soutenue avec le CEA, et ponctuellement avec le SLAC à l’Université de Stanford dans le cadre de la mission GLAST-Fermi. La collaboration avec le CEA a été concrétisée par le co-encadrement de plusieurs thèses et financée par un projet national (ANR NatImages) et européen (ERC SparseAstro), voir le détail en Annexe A. En particulier, nous traitons des problèmes d’estimation et de décision statistiques, de

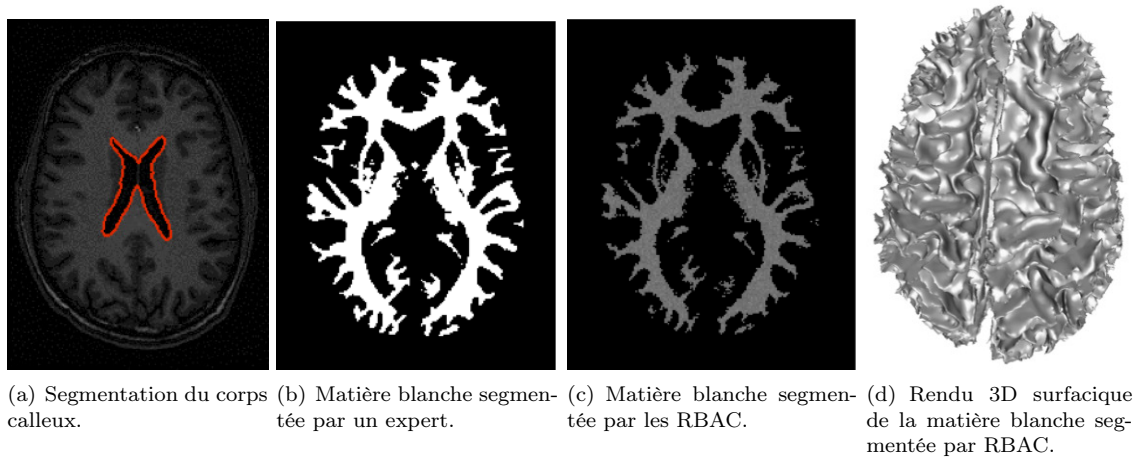


FIGURE 9.5 – Application des RBAC à la segmentation d'images IRM.

restauration, de reconstruction, de décomposition et de séparation des images mono et multispectrales en astronomie, aussi bien pour des données sur une grille cartésienne que sur la sphère.

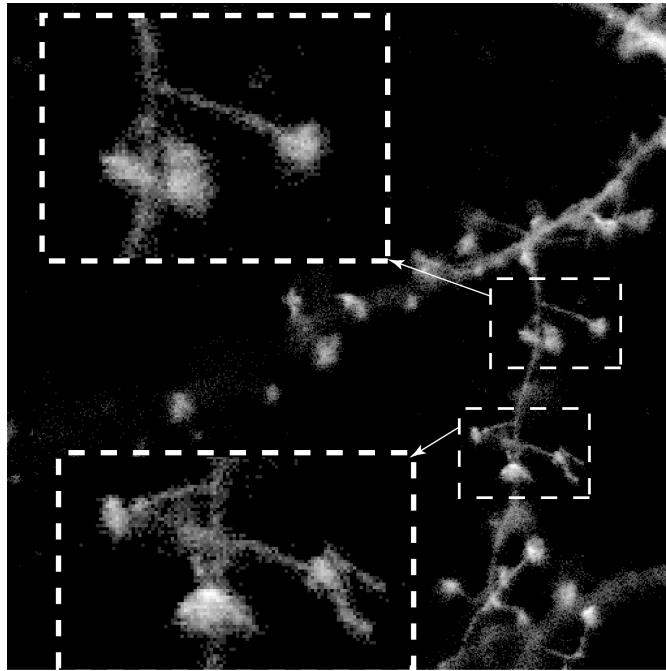
Débruitage des données GLAST-Fermi Le télescope LAT (Large Area Telescope), l'instrument principal de la mission Fermi-GLAST⁴⁰, permettra de détecter avec une sensibilité et une résolution spatio-spectrale inégalées des sources célestes de rayons gamma à hautes énergies. Il a été lancé par la NASA en Juin 2008. Ses deux principaux objectifs scientifiques, l'étude du fond diffus de la voie lactée et la détection des sources ponctuelles, sont rendus difficiles par le bruit de comptage photonique et la très faible intensité des taux comptage reçus. En effet, en dépit de sa large surface photosensible effective, le LAT reçoit peu de photons en dehors du plan galactique et quelques sources intenses. Pourtant, ce sont les sources peu photo-génératrices qui intéressent dans le cadre de cette mission. Les données issues du télescope peuvent être considérées comme un processus de Poisson hétérogène sur la sphère, et l'intensité sous-jacente est à estimer à partir de ses réalisations. Pour ce faire, nous avons proposé dans [A60, A8] une extension de la MS-VST de la section 4.1.2.3 pour pouvoir traiter des données sur la sphère, avec plusieurs transformées notamment la transformée d'ondelettes isotrope [A1]. Pour estimer les sources ponctuelles, l'algorithme de la MS-VST a été adapté pour permettre de s'affranchir de l'effet du fond galactique régulier. La fig. 9.7 illustre les résultats de la version de la MS-VST étendue sur un jeu de données GLAST-Fermi réaliste. Dans [A11, A53], une extension multispectrale de la MS-VST est décrite pour estimer les spectres des données Fermi acquises sur plusieurs longueurs d'ondes de la plage des rayons gamma.

Interpolation des données manquantes en lentillage faible La distorsion des images de galaxies lointaines par l'effet de lentillage gravitationnel (ou cisaillement gravitationnel) offre une voie directe pour accéder aux propriétés statistiques de la distribution de matière noire dans l'univers ; sans faire d'hypothèse sur la relation entre matières noire et visible. Cet effet appelé lentillage faible permet d'imposer des contraintes sur les paramètres cosmologiques des modèles de l'univers. L'analyse des données de lentillage faible nécessite inévitablement de masquer les régions contenant les étoiles brillantes. A ces données masquées, d'autres données peuvent manquer à cause par exemple des défauts du détecteur CCD. Ceci rend plus délicate l'étape d'estimation des statistiques en raison des données manquantes.

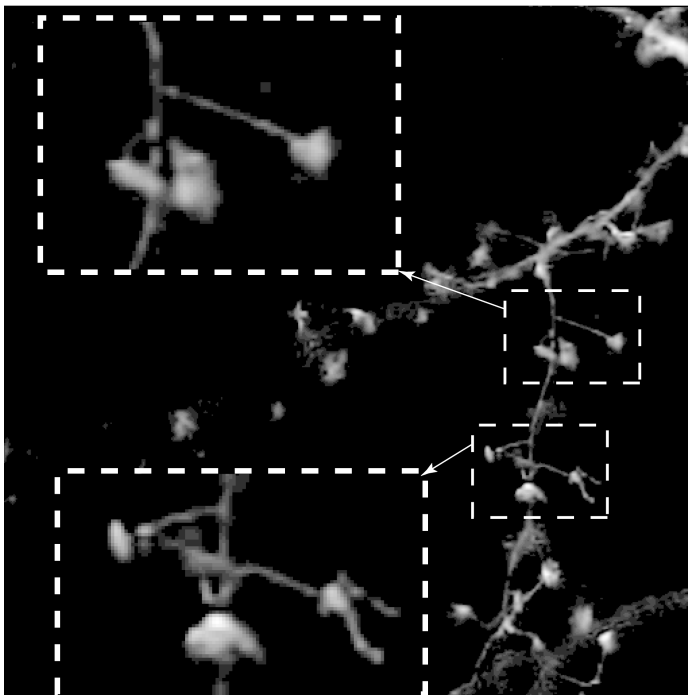
Dans [A14], nous avons montré que l'inpainting avec régularisation parcimonieuse apporte une solution efficace et rapide à ce problème. En particulier, les statistiques déduites à partir des images interpolées (notamment le spectre et le bi-spectre) se sont avérées très précises.

Sur les fig. 9.8(a), on peut voir deux cartes de masse obtenues par projection d'une simulation 3D très réaliste d'un volume de densité de matière du modèle cosmologique Λ CDM (Cold Dark Matter and

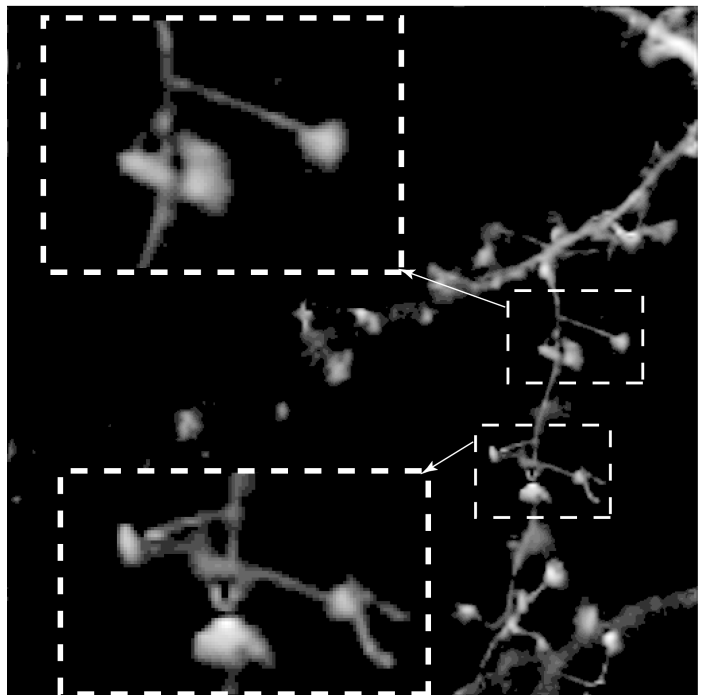
40. <http://fermi.gsfc.nasa.gov/>.



(a) Image observée dégradée.



(b) Déconvolution en résolvant (5.8) avec stabilisation (5.10).



(c) Déconvolution en résolvant (5.8) sans stabilisation (5.11).

FIGURE 9.6 – Application à la déconvolution d'une image réelle de neurones observés par microscopie confocale.

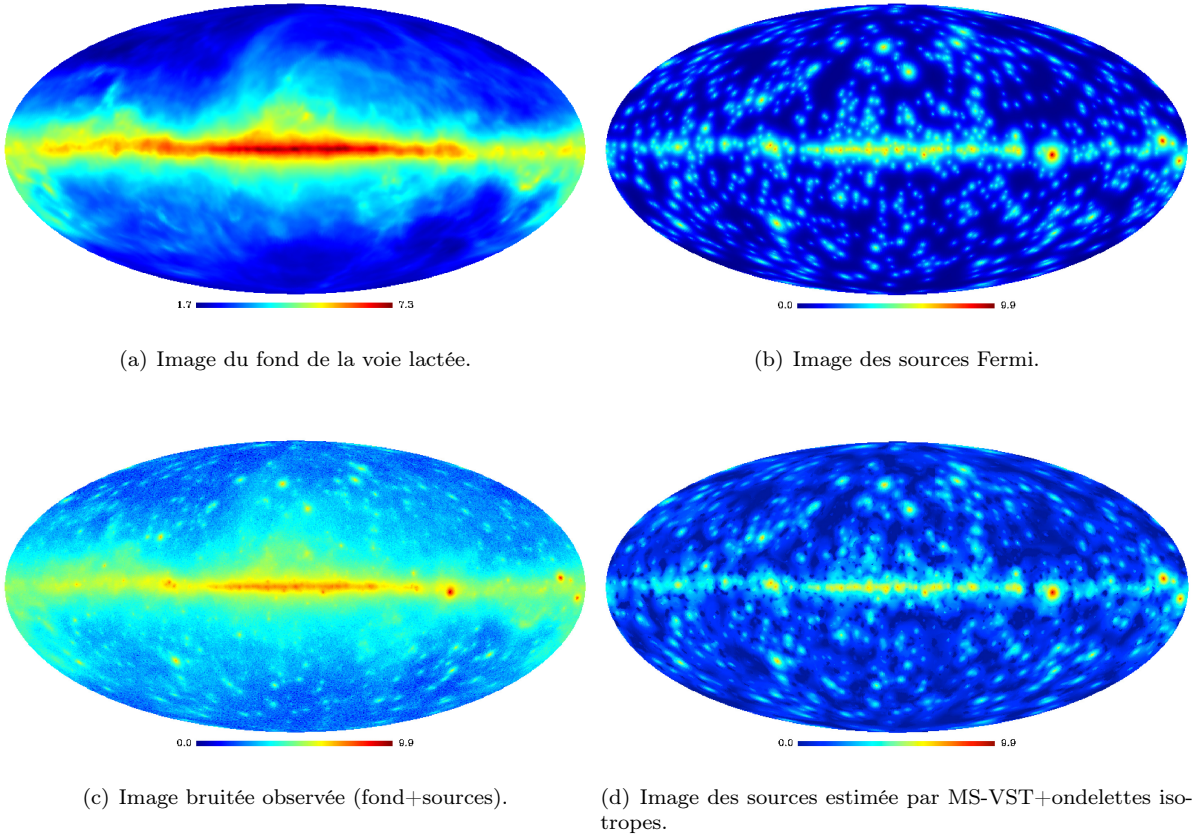


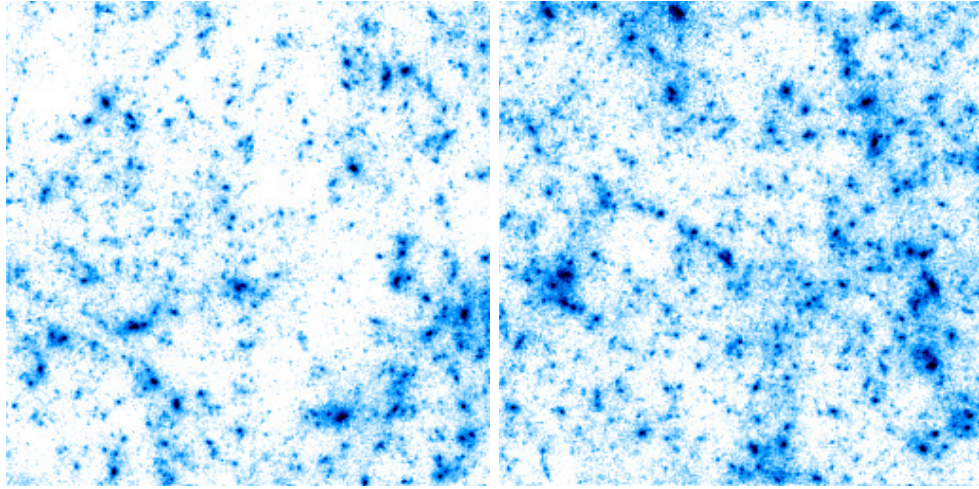
FIGURE 9.7 – Application de la MS-VST au débruitage des données GLAST-Fermi.

Energy). Les pics de surdensité de matière (en noir) correspondent notamment aux amas de galaxies. Les fig. 9.8(b) montrent deux masques typiques utilisés dans les observations de lentillage faible avec respectivement 20% et 10% de données manquantes. L'inpainting a été appliqué à des images masquées de cisaillement gravitationnel (deux cisaillements). Ces derniers sont reliés à la carte de masse par un opérateur linéaire connu qui doit aussi être inversé en même temps que l'inpainting. Ce problème inverse conjoint avec régularisation parcimonieuse a été résolu en utilisant les algorithmes décrits au Chapitre 5. Les cartes restaurées de matière noire correspondant à chacun des masques sont affichées en fig. 9.8(c).

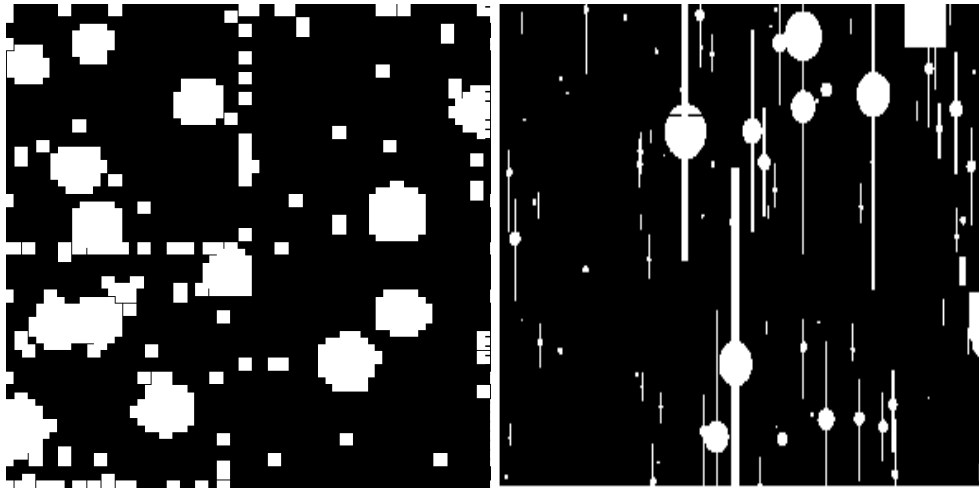
Analyse du fond diffus cosmologique La mission spatiale Planck⁴¹ lancée par l'ESA en Mai 2009 a pour objectif l'étude du fond cosmologique diffus (CMB). Le CMB est considéré par les cosmologistes comme l'une des clefs de voûte des théories actuelles de l'expansion de l'Univers (théorie du Big Bang). Le CMB est le reliquat du rayonnement le plus ancien observable actuellement (d'où le nom de rayonnement fossile). Selon la théorie du Big Bang, le CMB est la trace du passage rapide de l'état opaque (lié à une forte ionisation de l'Univers) à transparent de l'Univers, ce rayonnement a été refroidi et dilué par l'expansion de l'univers pour atteindre sa température actuelle de 2,735 K. La théorie du Big Bang prévoit que le CMB est un champ d'émission Gaussien stationnaire isotrope. Ainsi, cette lumière revêt une importance capitale pour les cosmologistes puisque selon la théorie du Big Bang, le spectre de puissance du CMB est une source d'information cruciale sur les paramètres intervenant dans les équations de Friedmann régissant l'évolution de notre Univers. De plus, tester la (non-)gaussianité du champ de fluctuation de température du CMB permet de tester certains modèles d'inflation.

Le rayonnement CMB observé par Planck sera mesuré sur une sphère dont l'observateur est le centre.

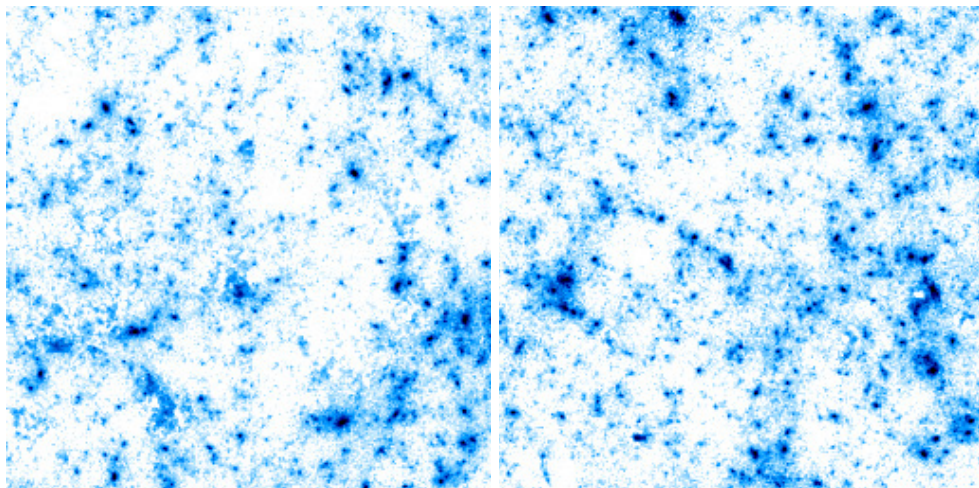
41. <http://planck.esa.int>.



(a) Cartes de masse originales simulées.



(b) Masques.



(c) Cartes restaurées de matière noire.

FIGURE 9.8 – Application de l’inpainting par régularisation parcimonieuse à la reconstruction des cartes de matière noire à partir de deux cartes masquées du cisaillement gravitationnel.

La mission Planck fait suite à une série de missions spatiales (COBE, WMAP) qui ont déjà apporté de nombreuses informations sur le CMB. La dernière de ces missions, WMAP⁴², a ainsi confirmé la théorie sur les deux premiers modes du spectre de puissance du CMB. Le défi porté par Planck est l'observation du ciel à une résolution très fine sur 9 canaux de 30 à 857 GHz. Outre le CMB, chaque canal Planck capturera différentes composantes astrophysiques : les sources galactiques (émissions Synchrotron, émissions Free-Free, émissions thermiques de poussières), Effet "Sunyaev Zel'Dovich". Hormis ces composantes astrophysiques, les données Planck sont entachées de sources ponctuelles (galaxies infra-rouge, radio, quasars, etc.) et contaminées par un bruit instrumental. Ainsi, chaque canal observé de Planck est supposé s'écrire comme un mélange linéaire instantané bruité des six sources astrophysiques évoquées plus haut.

Dans le cadre de la seconde série de challenges (CH2) du groupe de travail Planck (Planck G2) dédié à la séparation de sources, les données simulées fournies pour les tests 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 produites par le consortium Planck à l'issue d'une année d'observation. Chaque observation est une carte échantillonnée sur la sphère et le jeu de données total contient près de 450 millions de pixels. Les 9 canaux de ces données Planck sont visibles sur la fig. 9.9.

L'algorithme GMCA pour la séparation de sources décrit en section 6.2.3 a été modifié dans [A18, A82] pour l'estimation du CMB dans les données Planck, en particulier en contraignant certaines sources astrophysiques à vérifier leurs lois physiques d'émissions (c.à.d. contraintes sur la matrice de mélange). La fig. 9.10 représente la carte CMB originale et la source CMB estimée par GMCA à partir des données Planck CH2/G2. Notons que la carte du CMB estimée a été corrigée de la convolution par la PSF de l'instrument par un filtre de Wiener après séparation, d'où une perte légère de certains détails fins sur la carte estimée. Une voie d'amélioration envisageable à l'avenir est de procéder à une déconvolution jointe des sources conjointement avec la séparation. Ceci n'a pas été fait pour ces données pour des raisons de rapidité compte tenu de la taille très importante des données. Une validation exhaustive de la GMCA et une comparaison aux approches des autres laboratoires participant au groupe de travail Planck CH2/G2 a été menée dans [TH2].

Une autre démarche classique pour l'analyse des données CMB est de ne considérer que le canal correspondant au CMB. Toutefois, préalablement à toute exploitation, ce canal doit être nettoyé des contributions résiduelles, plus significativement la région galactique au centre et les sources ponctuelles intenses. Ces dernières sont en général simplement masquées. Dans [A21, A75], nous avons proposé une méthode d'interpolation de ces données manquantes en s'appuyant sur la parcimonie pour régulariser le problème inverse tout en imposant une contrainte statistique de stationnarité au second ordre sur la carte estimée (statistique suffisante puisque le CMB est un champ Gaussien isotrope). Un exemple de résultats de cette application sur les données WMAP est porté en fig. 9.11.

9.1.3 Contrôle non destructif

L'Institut Français du Pétrole (IFP) en collaboration avec le CEA-Saclay étudie et modélise le comportement mécanique de fibres de flexibles sous effort sur un banc d'essai mécanique muni d'une caméra X. Les séries temporelles d'images acquises étant de très mauvaise qualité et entachées de nombreuses dégradations, leur exploitation est impossible sans un prétraitement adéquat. Ainsi, notre apport dans la collaboration avec l'IFP (voir Annexe A pour les détails) a été d'appliquer nos outils de pointe de restauration, de rehaussement, de décomposition, et de tractographie de structures d'intérêt sur ces images [A72, A77].

La fig. 9.12 montre un exemple d'application de la séparation parcimonieuse par l'algorithme MCA à une image d'un tel flexible acquise par caméra X. Le flexible est formé d'une canalisation en matériau composite autour de laquelle sont enroulées deux tresses de fibres en acier d'angles opposés, et de marqueurs en plomb utilisés comme référence pour calibrer la caméra. Les structures d'intérêt sur cette image sont les fibres. Le dictionnaire utilisé est composé de la transformée en ondelettes invariante par translation (UDWT) + curvelets. A partir de cette figure, il est observé comment la MCA a pu se débarrasser des marqueurs tout en reconstituant les fibres curvilinéaires d'intérêt.

42. <http://map.gsfc.nasa.gov>.

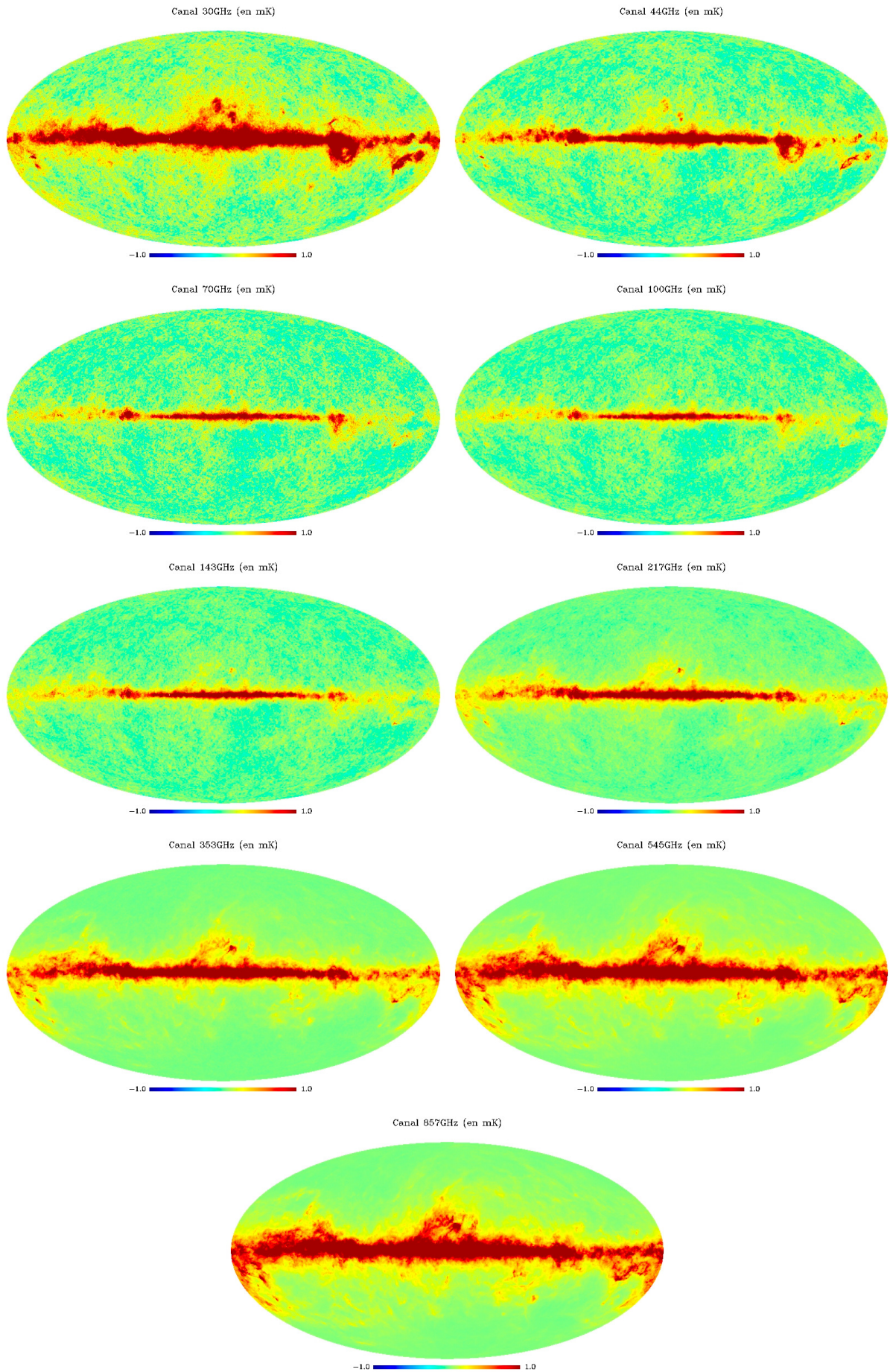


FIGURE 9.9 – Canaux observés des données Planck CH2/G2.

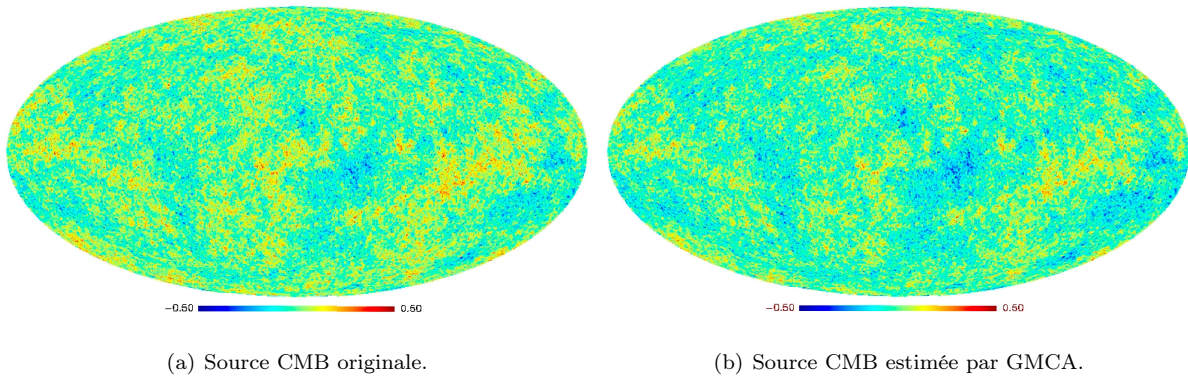


FIGURE 9.10 – Application de l'algorithme GMCA à la séparation de sources sur les données Planck. L'algorithme GMCA décrit au Chapitre 6 a été adapté pour contraindre certaines sources à vérifier les propriétés physiques prévues par les lois d'émission théoriques.

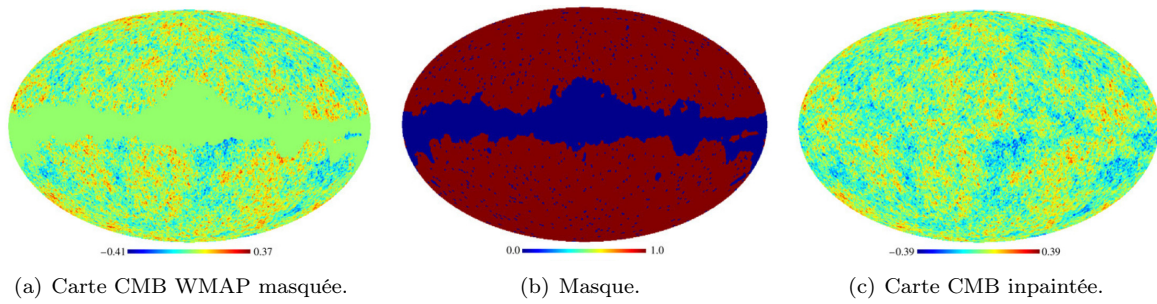


FIGURE 9.11 – (a) Carte CMB obtenue par WMAP masquée du plan galactique et des sources ponctuelles, et (c) sa version inpaintée.

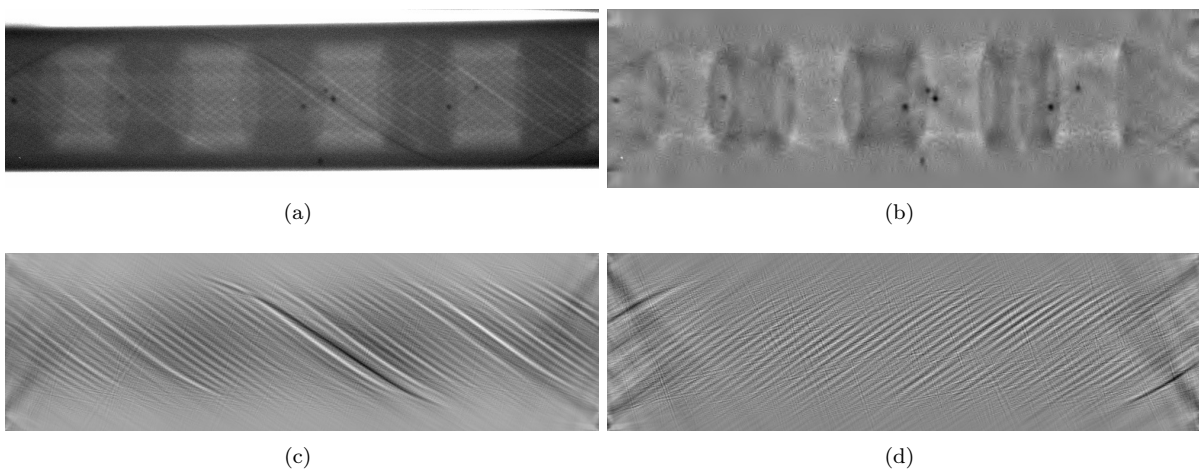


FIGURE 9.12 – Exemples d'application de la décomposition parcimonieuse à une image réelle d'un flexible acquise par rayons X : (a) Image observée, (b) structures isotropes et fond (ondelettes), (c)-(d) fibres curvilignes dans deux directions perpendiculaires (curvelets).

9.1.4 Photographie computationnelle

La mise au point automatique est un exemple de fonction critique du système d'imagerie d'une caméra numérique. La tendance actuelle est à la réalisation de cette fonction au travers de méthodes passives, e.g. ne faisant appel à aucun composant externe autre que le capteur lui-même pour des raisons de coût évidentes. Dans le cadre d'une collaboration avec Philips Semi-conducteurs (devenu depuis NxP Semiconductors), nous avons élaboré des algorithmes novateurs de mise au point passive (autofocus) utilisant des mesures de régularité basées sur la transformée d'ondelettes [A115]. Ces algorithmes se sont de plus montrés robustes vis-à-vis des bruits inhérents au capteur (notamment en technologie CMOS). L'algorithme a été implémenté en temps réel sur un prototype livré à la société NxP dans le cadre d'un projet régional (voir Annexe A pour les détails).

9.2 La recherche reproductible

La reproductibilité est au coeur de la méthodologie scientifique et les développements technologiques réussis. Dans les disciplines théoriques, la référence a été établie par les mathématiques, où la preuve formelle permet en principe de reproduire les étapes cognitives conduisant à la vérification d'un théorème. Alors qu'au début des mathématiques, des justifications plus ou moins rigoureuses étaient admises, dans le dernier siècle il a été accepté que même les grandes découvertes ne pouvaient être prises pour fait établi jusqu'à leur preuve rigoureuse formelle. Ceci a conduit aux prix de l'Institut Clay récompensant les preuves de conjectures célèbres.

Dans les disciplines expérimentales, comme en biologie, physique ou chimie, pour qu'un résultat soit admis par les pairs, une attention particulière est donnée à la possibilité de répliquer l'expérience. Bien sûr, une description de cette dernière avec tous les détails nécessaires est indispensable pour qu'un autre groupe puisse la reproduire.

Les sciences computationnelles (comme l'est le traitement du signal et des images) forment une discipline plus jeune que les mathématiques, mais déjà d'une importance primordiale ; mais la reproductibilité n'y a pas encore eu la place qu'elle mérite. Le slogan "recherche reproductible" est relativement récent, et essaye d'établir une référence pour la reproductibilité en sciences computationnelles. La recherche reproductible reconnaît que la production réelle d'un projet de recherche ne se restreint pas à l'article publié, mais s'étend à l'environnement entier qui y a conduit et qui permet de le reproduire. Ceci comprend les données, le logiciel, la documentation, etc.

Le résultat typique dans un article en sciences computationnelles est une figure. Mais il est illusoire pour un relecteur et/ou un lecteur de la vérifier en tentant de rebâtir le système sans doute complexe pour comprendre réellement ce que l'auteur a mis des mois voire des années à construire. Ceci pose des questions fondamentales y compris d'éthique scientifique. En effet, que signifie alors une évaluation par les pairs dans de telles conditions ? Un encadrant peut-il toujours affirmer que les résultats de son étudiant sont corrects ? Les sciences computationnelles peuvent-elles être une entreprise cumulative ? Comment instaurer un guide des bonnes pratiques de la recherche reproductible dans une telle discipline ?

La vision de J. Clearbout vers la fin des années 1980 [183] a énoncé le principe que les sciences computationnelles reposent fondamentalement sur une compréhension claire de l'environnement complet dans lequel un algorithme donné s'exécute y compris le code et les données, et l'idée qu'il faut tisser un lien étroit entre le code et l'article de façon à ce que les figures et le code soient inclus dans le même fichier. Cette discipline a été imposée dans son groupe pour garantir la reproductibilité de leurs publications.

Inspiré par Clearbout, D. Donoho au début des années 1990 a adopté la pratique de la recherche reproductible en analyse harmonique computationnelle dans l'environnement Matlab, en demandant à ses étudiants et ses co-auteurs de suivre la même démarche en mettant à disposition de la communauté scientifique intéressée sur une page web dédiée, des boîtes à outils contenant le code et les paramètres reproduisant les figures et les tableaux dans leurs articles, la documentation des outils et de l'environnement. Cette philosophie a été suivie avec diligence par Donoho et son groupe dans toutes leurs boîtes à outils dont [203]. Le lecteur peut aussi se reporter à l'article de revue [78] dans le numéro spécial sur la recherche reproductible publié par la revue IEEE Computing in Science and Engineering en 2009.

En traitement du signal et des images, l'intérêt en recherche reproductible a été revitalisé par [21].

A l'une des conférences majeures de ce domaine, ICASSP 2007, une session spéciale a été entièrement consacrée à la recherche reproductible en traitement du signal.

C'est en suivant cette philosophie que j'ai mis au coeur de ma démarche scientifique ainsi que celle de mes étudiants le souci de mettre à disposition de la communauté non seulement la production scientifique sous forme de publications, mais aussi des boîtes à outils et des composants logiciels permettant de reproduire les expériences qui y sont décrites. C'est le cas par exemple pour la séparation parcimonieuse de composantes et de sources [A49], ou encore le débruitage et les problèmes inverses (voir Annexe A pour un descriptif détaillé). La recherche reproductible a été aussi centrale dans l'élaboration du livre [A1].

Bien évidemment, il reste plusieurs questions majeures auxquelles il faut apporter des réponses pour que cette pratique puisse se généraliser. Parmi elles, un guide de bonnes pratiques pour encourager la mise en place et la promotion de la recherche reproductible en tant que contenu scientifique publiable dans les journaux académiques. Ce dernier aspect est en effet culturel puisque notre communauté scientifique favorise l'"innovation" avant tout, laissant peu de place à des contributions du type recherche reproductible. Toutefois, ce point de vue commence à changer par une prise de conscience de certains comités éditoriaux de journaux, de l'ouverture d'autres à ce type de travaux, mais aussi par la création de journaux dédiés, à l'image d'IPOL⁴³ à l'initiative du groupe de J.-M. Morel, et dans lequel j'ai pris part. Si cette pratique venait à se généraliser et se systématiser dans la communauté signal-image, les bénéfices à en tirer sont nombreux :

- L'éducation : les jeunes chercheurs, nouveaux dans le domaine, peuvent bénéficier plus facilement des expériences de leurs pairs.
- L'impact et la progrès de la recherche : dans le long terme, le travail publié a un impact plus large potentiellement au-delà de la communauté spécialisée. Ceci permet aussi d'éviter de réinventer les choses et d'accélérer le progrès.
- La réduction des risques : lorsqu'une expérience est reproduite avec succès et transparence, ses chances d'être erronées sont d'autant plus réduites.
- L'interaction : les collaborations entre groupes s'en trouvent facilitées.

9.3 Récapitulatif et conclusion

Dans ce chapitre sont présentées nos contributions dans plusieurs modalités d'imagerie auxquelles nous avons appliqué nos différents algorithmes d'estimation, de restauration, de séparation ou de segmentation. Ensuite, la philosophie de la recherche reproductible a été expliquée et nos réflexions et contributions sur le sujet ont été résumées.

43. <http://www.ipol.im>.

Chapitre 10

Conclusion et perspectives

Comme on peut s'en rendre compte, d'après ce qu'il vient d'être exposé, les approches stochastiques et l'analyse harmonique computationnelle, les représentations parcimonieuses et l'optimisation non lisse forment la pierre angulaire de mes recherches. En s'appuyant sur ce cadre, nous nous sommes attelés à résoudre plusieurs grandes problématiques du traitement d'images. Au regard de ces travaux, plusieurs perspectives se dégagent naturellement. Nous en esquissons ici celles qui paraissent les plus pertinentes dont certaines tenteront de répondre à des questions ouvertes qui restent à explorer.

10.1 La modélisation des signaux et des images

La modélisation et la synthèse de textures La modélisation des textures est sans doute l'un des problèmes les plus difficiles en traitement des images et la vision par ordinateur. Plusieurs modèles ont été proposés dans la littérature en particulier les modèles de textures oscillantes, les modèles stochastiques Markoviens, les textons ou plus récemment les approches non locales. Les outils mathématiques utilisés dans ces modèles sont issus essentiellement de l'analyse harmonique, des méthodes variationnelles et des statistiques. Toutefois, chacune de ces modélisations opère mieux pour une classe de texture particulière (par exemple, les textures localement périodiques pour le premier modèle). Par conséquent, il reste beaucoup à faire dans ce domaine et les outils mathématiques pour la synthèse et l'analyse des textures complexes sont encore à mettre au point. Une thèse en co-encadrement avec Jean-Luc Starck et en collaboration avec Gabriel Peyré débutera sur la modélisation et la synthèse de texture en alliant la théorie de l'apprentissage parcimonieux et la modélisation stochastique. Elle s'inscrit dans le cadre du projet SparseAstro.

La parcimonie structurée En plus de la parcimonie dans un dictionnaire donné, et comme nous l'avons montré au Chapitre 2, les signaux et images naturelles exhibent une structure plus riche comme le regroupement des coefficients, la structuration en arbre avec la persistance inter-échelle, etc. La prise en compte de cette information supplémentaire se révèle d'un apport précieux garantissant de meilleures performances aussi bien pour l'estimation, la restauration et les problèmes inverses, l'apprentissage ou l'échantillonnage compressé. La parcimonie structurée a d'ores et déjà commencé à être explorée par quelques auteurs dans la communauté des statistiques et celle du traitement théorique du signal, avec des modèles toutefois simples comme la structuration par blocs isotropes (taille connue), union de sous-espaces. Les garanties et les choix théoriques des paramètres du modèle restent encore relativement peu étudiés ou seulement pour les modèles simples évoqués ci-dessus. Ceci sera un thème principal du post-doctorant de l'ANR NatImages qui débutera sous mon encadrement avec Christophe Chesneau.

10.2 La théorie de l'estimation

L'agrégation d'estimateurs L'agrégation d'estimateurs est un cadre théorique pour établir les propriétés d'un estimateur formé en combinant plusieurs autres (appelés préestimateurs). Il est maintenant

bien établi que l'estimateur par agrégation avec poids exponentiels conduit à un risque moindre (au sens d'inégalité oracle) que le meilleur des estimateurs individuels. On peut par exemple penser à l'estimateur des moyennes non locales comme un agrégateur de patches. Notre intérêt dans ce sujet est dicté par le souci d'obtenir des estimateurs (en débruitage par exemple) encore plus efficaces et rapides tout en comprenant les performances théoriques que l'on peut atteindre par le biais de l'agrégation.

L'estimation par blocs Nous souhaitons en collaboration avec Christophe Chesneau d'accentuer nos efforts sur les performances minimax (ou d'autres sens comme le maxiset) de l'estimation par blocs et ses variantes (e.g. avec blocs adaptatifs anisotropes) en d -dimensions avec des transformées non nécessairement orthogonales et dans le cadre d'autres modèles, e.g. problèmes inverses, déconvolution aléatoire, déconvolution fonctionnelle, estimation de densité, au-delà du bruit additif Gaussien (bruit de Poisson, bruit additif à variance infinie, etc.).

10.3 L'optimisation non lisse

L'optimisation non lisse et/ou non convexe Alors que l'optimisation convexe a suscité un intérêt considérable, notamment grâce à l'analyse convexe, l'optimisation non lisse et non convexe demeure beaucoup moins développée. Beaucoup de problèmes en traitement du signal et des images impliquent cependant des fonctionnelles non lisses et non convexes à minimiser, e.g. en théorie de l'approximation, en échantillonnage compressé ou en segmentation.

Dans un travail de collaboration avec Mila Nikolova, nous nous intéressons à des problèmes inverses linéaires avec régularisation non lisse et non convexe comme la norme ℓ_p , $0 \leq p < 1$ ou des φ -fonctions non différentiables et non convexes. En plus de la caractérisation de ces problèmes, nous nous attelons à étudier les propriétés de leurs minimiseurs et à proposer des algorithmes déterministes pour les résoudre localement, et le cas échéant globalement (e.g. programmation DC). Ceci fait l'objet d'une proposition de projet ANR soumis porté par Mila Nikolova.

10.4 Les problèmes inverses avec parcimonie

La restauration La restauration reste l'une des préoccupations importantes de nos recherches. La déconvolution avec a priori de parcimonie dans les dictionnaires appris conjointement, la déconvolution aveugle ainsi que la déconvolution avec d'autres types de bruits constituent des exemples typiques de problèmes importants que nous comptons examiner dans le futur. Un post-doc avec Jean-Luc Starck débute sur ces sujets, ainsi qu'une collaboration avec Mila Nikolova et Sylvain Durand dans le cadre d'une thèse qu'ils co-encadrent.

L'échantillonnage compressé et applications En collaboration avec Gabriel Peyré et Charles Dossal, nous allons poursuivre et intensifier nos efforts pour une compréhension plus fine des conditions suffisantes (et nécessaires) garantissant la restauration (au sens support, signe, consistance, etc.) par minimisation ℓ_1 aussi bien dans un cadre bruité que non. Nous pensons en particulier aux extensions-adaptations des critères topologiques discutés au Chapitre 7.

La super-résolution La super-résolution de sources ponctuelles avec régularisation parcimonieuse est un autre axe que nous désirons poursuivre en collaboration avec Charles Dossal et Gabriel Peyré. Il découle assez naturellement de celui de la restauration parcimonieuse en spécialisant la matrice de mesure. Ce sujet soulève plusieurs difficultés liées aux "mauvaises" propriétés spectrales de la matrice en particulier, mais est porteur de plusieurs applications potentielles notamment la spectroscopie et l'imagerie sismique.

Estimation des hyperparamètres Comme nous l'avons souligné dans le manuscrit, les fonctionnelles que nous devons minimiser pour différents problèmes inverses mettent en jeu des hyperparamètres parfois difficiles à choisir. Ils peuvent néanmoins être estimés à partir des données elles-mêmes en utilisant des critères de sélection de modèles (AIC, BIC), des estimateurs non biaisés du risque quadratique

(GCV, SURE), les techniques de perturbation, ou les méthodes de rééchantillonnage aléatoire (Bootstrap, MCMC). Certaines de ces méthodes ont été employées et étudiées pour la déconvolution dont certaines avec des heuristiques. Nous devons continuer cet effort pour une meilleure compréhension et proposer des choix avec des garanties théoriques si possible.

Analyse versus synthèse Au Chapitre 5, la distinction entre a priori de parcimonie analyse et synthèse a été présentée, et seuls les aspects reliés à l’optimisation ont été discutés et résolus. Concernant les garanties de restauration parcimonieuse (unicité, support, signe, consistance, etc.), les travaux existants s’intéressent presque exclusivement à l’a priori synthèse pour en expliquer les performances. D’un point de vue pratique, plusieurs travaux récents s’accordent à conclure que la régularisation parcimonieuse avec a priori analyse peut s’avérer bien supérieur à l’a priori synthèse pour résoudre bon nombre de problèmes inverses comme la déconvolution ou l’inpainting par exemple. En revanche, ces observations pratiques ne sont pas appuyées par des garanties théoriques qui restent moins développées (et de loin) qu’avec l’a priori synthèse. Il reste donc beaucoup de chemin à accomplir pour comprendre le problème difficile des performances de l’a priori analyse.

10.5 Analyse multiéchelle de fonctions non scalaires

Les imageurs modernes fournissent des données de plus en plus complexes où les échantillons sont acquis sur des grilles non nécessairement Cartésiennes équi-espacées, et où le type des données n’est pas forcément un scalaire. En plus des données couleurs, nous assistons à l’apparition de données très riches où les échantillons ont des valeurs structurées obéissant à des contraintes non linéaires. On peut penser à ces données comme une application $f : \Omega \rightarrow \mathcal{V}$, où $\mathcal{V}$ est une variété. Des exemples typiques de $\mathcal{V}$ sont : S^1 ou S^2 (données de directions, polarisation) ; $SO(3)$ (données d’orientation 3D) ; matrices définies-positives (typiquement en IRM de diffusion) ; variété Grassmanienne (données projetées dans un sous-espace de $\mathbb{R}^d$) ; patches comme dans les méthodes non locales ; etc. Pour Ω , outre la grille Cartésienne, les exemples couvrent le cas de la sphère, les sections coniques, les surfaces ou encore les graphes.

Des constructions de la transformée en ondelettes ont été proposées pour certaines variétés non Euclidiennes Ω dont la sphère, les variétés compactes différentiables, les variétés lisses (diffusion wavelets), ou encore les graphes.

L’analyse par ondelettes dans le cas où Ω est une grille Cartésienne, mais $\mathcal{V}$ est une variété quelconque a été considérée dans la littérature par très peu d’auteurs dont [174]. Leur représentation multiéchelle s’appuie sur le schéma de raffinement (par interpolation) de Deslauriers-Dubuc, agissant dans le plan tangent de la variété. L’approche s’applique à des variétés plutôt générales mais s’avère plus facilement implémentable pour des espaces Riemanniens symétriques comme S^{d-1} ou $SO(d)$.

Dans cette lignée de travaux, les perspectives qui nous apparaissent pertinentes à explorer sont multiples. En l’occurrence, la définition d’une analyse multiéchelle pour d’autres variétés $\mathcal{V}$ comme les patches, l’étude d’autres schémas de raffinement sur les variétés, ou encore les transformées multiéchelles pour des fonctions définies sur des variétés non Euclidiennes (e.g. S^2) et à valeurs dans des variétés comme peuvent l’être des données de polarisation sur la sphère en imagerie astronomique.

10.6 Applications

Les collaborations tissées avec les partenaires dans le cadre d’applications continueront à être fructifiées et renforcées, notamment dans les domaines de l’imagerie médicale et astronomique. Des discussions ont été aussi plus récemment amorcées avec Christophe Zimmer, chef du Groupe Imagerie et Modélisation à l’Institut Pasteur pour la mise en place de nouveaux protocoles d’acquisition inspirés par l’échantillonnage compressé et la reconstruction parcimonieuse pour l’imagerie rapide en microscopie super-résolutive. Ces voies de développements s’avèrent très prometteuses avec des challenges de réalisation physique du processus de mesure, ainsi que des questions théoriques ouvertes notamment compte tenu du bruit contaminant les données.

Bibliographie

- [1] <http://sipi.usc.edu/services/database/database.html>.
- [2] A. Nemirovski A. Ben-Tal. *Optimization III : Convex Analysis, Nonlinear Programming Theory, Standard Nonlinear Programming Algorithms*. Lecture notes, 2004.
- [3] F. Abramovich and B. W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85 :115–129, 1998.
- [4] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, New York, 1972.
- [5] A. Achim and E. E. Kuruoglu. Image denoising bivariate α -stable distributions in the complex wavelet domain. *IEEE Signal Processing letters.*, 12(1) :17–20, 2004.
- [6] A. Achim, P. Tsakalides, and A. Bezerianos. SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 41(8), 2003. 1773–1784.
- [7] 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), 2006. 4311–4322.
- [8] S. Alliney and S. Ruzinsky. An algorithm for the minimization of mixed l1 and l2 norms with application to bayesian estimation. *IEEE Transactions on Signal Processing*, 42(3), 1994. 618 - 627.
- [9] D. F. Andrews and C. L. Mallows. Scale mixtures of normality. *Journal of the Royal Statistical Society, Series B* 36, pages 99–102, 1974.
- [10] F.J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 15, 1948. 246–254.
- [11] A. Antoniadis and J. Fan. Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455), 2001. 939–963.
- [12] G. Aubert and J.-F. Aujol. A variational approach to remove multiplicative noise. *SIAM Journal on Applied Mathematics*, 68(4) :925–946, January 2008.
- [13] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson. Image segmentation using active contours : Calculus of variations or shape gradients? *SIAM Applied Mathematics*, 63(6) :2128–2154, 2003.
- [14] J.-F. Aujol, G. Aubert, and L. Blanc-Féraud. Wavelet-based level set evolution for classification of textured images. *IEEE Transaction on Image Processing*, 12(12) :1634–1641, 2003.
- [15] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22, 2005. 71–88.
- [16] J.-F. Aujol and G. Gilboa. Constrained and snr-based solutions for tv-hilbert space image denoising. *Journal of Mathematical Imaging and Vision*, 26(1-2), 2006. 217-237.
- [17] J.F. Aujol and A. Chambolle. Dual norms and image decomposition models. *International Journal of Computer Vision*, 63(1), 2005. 85–104.

- [18] J.F. Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. Structure and texture compression – modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1), 2006. 111–136.
- [19] R. 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.
- [20] R. G. Baraniuk, V. Cevher, and M. B. Wakin. Low-dimensional models for dimensionality reduction and signal recovery : A geometric perspective. *IEEE Signal Processing Magazine*, 2010.
- [21] M. Barni and F. Perez-Gonzales. Pushing science into signal processing. *IEEE Signal Proc. Mag.*, 22, 2005. 119–120.
- [22] A. Beck and M. Teboulle. Fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2, 2009. 183–202.
- [23] J. Beran. *Statistics for long-memory processes*. Monographs on statistics and applied probability 61. London, Chapman & Hall, 1994.
- [24] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48 :259–302, 1986.
- [25] P. J. Bickel, Y. Ritov, and A. Tsybakov. Simultaneous analysis of lasso and Dantzig selector. *Annals of Statistics*, 37, 2009. 1705–1732.
- [26] A. Bijaoui and G. Jammal. On the distribution of the wavelet coefficient for a Poisson noise. *Signal Processing*, 81, 2001. 1789–1800.
- [27] J. Bioucas-Dias. Bayesian wavelet-based image deconvolution : a GEM algorithm exploiting a class of heavy-tailed priors. *IEEE Transactions on Image Processing*, 15(4), 2006. 937–951.
- [28] J.D. Blanchard, C. Cartis, and J. Tanner. Compressed sensing : How sharp is the RIP ? *Preprint*, pages , 2009.
- [29] L. Borup and M. Nielsen. Frame decomposition of decomposition spaces. *Journal of Fourier Analysis and Applications*, 13(1) :39–70, 2007.
- [30] K. Bredies and D.A. Lorenz. Linear convergence of iterative soft-thresholding. *Journal of Fourier Analysis and Applications*, 14, Dec 2008. 813–837.
- [31] 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*, 51(1), 2009. 34–81.
- [32] A.M. Bruckstein and M. Elad. A generalized uncertainty principle and sparse representation in pairs of $\mathbf{r}^n$ bases. *IEEE Transactions on Information Theory*, 48, 2002. 2558–2567.
- [33] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2) :490–530, 2005.
- [34] T. Cai. Adaptive wavelet estimation : a block thresholding and oracle inequality approach. *Annals of Statistics*, 27, 1999. 898–924.
- [35] T. Cai, L. Wang, and G. Xu. New bounds for restricted isometry constants. Technical report, University of Pennsylvania, Department of Statistics, 2009.
- [36] T. Cai, L. Wang, and G. Xu. Shifting inequality and recovery of sparse signals. *IEEE Transactions on Signal Processing*, 2009. to appear.
- [37] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2), 2006. 489–509.
- [38] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8), 2006. 1207–1223.
- [39] E. J. Candès. Ridgelets and the representation of mutilated Sobolev functions. *SIAM Journal on Mathematical Analysis*, 33, 2001. 347–368.

- [40] E.J. Candès, L. Demanet, D.L. Donoho, and L. Ying. Fast discrete curvelet transforms. *Multiscale Modeling and Simulation*, 5(3), 2006. 861–899.
- [41] E.J. Candès and D.L. Donoho. Curvelets and curvilinear integrals. *Journal of Approximation Theory*, 113, 2001. 59–90.
- [42] E.J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12), 2005. 4203–4215.
- [43] E.J. Candès and T. Tao. Near optimal signal recovery from random projections : Universal encoding strategies? *IEEE Transactions on Information Theory*, 52, 2006. 5406–5425.
- [44] E.J. Candès and T. Tao. The Dantzig selector : Statistical estimation when p is much larger than n . *Annals of Statistics*, 35, 2007. 2392–2404.
- [45] J.-F. Cardoso. Blind signal separation : statistical principles. *Proceedings of the IEEE. Special issue on blind identification and estimation*, 9(10) :2009–2025, October 1998.
- [46] P. Chainais. Infinitely divisible cascades to model the statistics of natural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2) :2105–2119, 2007.
- [47] C. Chaux, P.L. Combettes, J.-C. Pesquet, and V.R. Wajs. A variational formulation for frame based inverse problems. *Inverse Problems*, 23, June 2007. 1495–1518.
- [48] C. Chaux, J.-C. Pesquet, and . Pustelnik. Nested iterative algorithms for convex constrained image recovery problems. *SIAM Journal on Imaging Sciences*, 2(2), 2009. 730–762.
- [49] S. S. Chen, D.L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1), 1999. 33–61.
- [50] A. Cichocki and S. Amari. *Adaptive Blind Signal and Image Processing : Learning Algorithms and Applications*. Wiley, New York, 2002.
- [51] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k -term approximation. *Journal of the American Mathematical Society*, 22, 2009. 211–231.
- [52] A. Cohen, R.A. DeVore, P. Petrushev, and H. Xu. Nonlinear approximation and the space $BV(R^2)$. *American Journal of Mathematics*, 121, 1999. 587–628.
- [53] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6), December 2004. 475–504.
- [54] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4), 2005. 1168–1200.
- [55] Patrick L. Combettes and Jean-Christophe Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM Journal on Optimization*, 18(4), 2007. 1351–1376.
- [56] P.L. Combettes, D. Dung, and B. C. Vu. Fenchel-rockafellar dualization of signal recovery problems. July 2009. submitted.
- [57] P.L. Combettes and J.-C. Pesquet. A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(2), 2007. 564–574.
- [58] D. Cooper. Maximum likelihood estimation of markov process blob boundaries in noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1 :372–384, 1979.
- [59] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57, 2004. 1413–1541.
- [60] M. E. Davies and R. Gribonval. Restricted isometry constants where l_p sparse recovery can fail for $0 < p \leq 1$. *IEEE Trans. on Information Theory*, 5(55) :2203–2214, 2009.
- [61] M.C. Delfour and J.P. Zolésio. *Shape and geometries*. Advances in Design and Control, SIAM, 2001.
- [62] L. Demanet and L. Ying. Wave atoms and sparsity of oscillatory patterns. *Applied and Computational Harmonic Analysis*, 23(3), 2007. 368–387.

- [63] R.A. DeVore. Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 23, 2007. 918–925.
- [64] R.A. DeVore. Foundations of compressed sensing. Lecture notes, 2009.
- [65] R.A. DeVore, G. Petrova, and P. Wojtaszczyk. Instance-optimality in probability with an ℓ_1 -minimization decoder. *Applied and Computational Harmonic Analysis*, 27(3) :275 – 288, 2009.
- [66] N. Dey, L. Blanc-Féraud, C. Zimmer, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia. A deconvolution method for confocal microscopy with total variation regularization. In *IEEE ISBI*, 2004.
- [67] M.N. Do and M. Vetterli. Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance. *IEEE Transactions on Image Processing*, 11(2) :146–158, 2002.
- [68] M.N. Do and M. Vetterli. The contourlet transform : an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12), 2005. 2091–2106.
- [69] David Donoho, Iain Johnstone, Jeffrey Hoch, and Alan Stern. Maximum entropy and the nearly black object. *J. R. Stat. Soc. B*, 54(1) :41–81, 1992.
- [70] D.L. Donoho. Nonlinear solution of inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2, 1995. 101–126.
- [71] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 2006. 1289–1306.
- [72] D.L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete and Computational Geometry*, 35(4), 2006. 617–652.
- [73] D.L. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization. *Proceedings of the National Academy of Sciences*, 100, 2003. 2197–2202.
- [74] D.L. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1), 2006. 6–18.
- [75] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7), 2001. 2845–2862.
- [76] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81, 1994. 425–455.
- [77] D.L. Donoho and B. F. Logan. Signal recovery and the large sieve. *SIAM J. Appl. Math*, 52 :577–591, 1992.
- [78] D.L. Donoho, A. Maleki, I. Ur Rahman, M. Shahram, and Stodden. Reproducible research in computational harmonic analysis. *Computing in Science and Engineering*, 11(1) :8–18, Jan./Feb. 2009.
- [79] D.L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math*, 49 :906–931, 1989.
- [80] D.L. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. *Proceedings of the National Academy of Sciences*, 27(102), 2005. 9452–9457.
- [81] D.L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2006. submitted.
- [82] D.L. Donoho, M Vetterli, R.A. DeVore, and I. Daubechies. Data compression and harmonic analysis. *IEEE Trans. Inform. Theory*, 44, 1998. 2435–2476.
- [83] C. Dossal. A necessary and sufficient condition for exact recovery by ℓ_1 minimization. *Foundations of Computational Mathematics*, 2009. revised.
- [84] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82, 1956. 421–439.
- [85] S. Durand and M. Nikolova. Denoising of frame coefficients using l_1 data-fidelity term and edge-preserving regularization. *SIAM Journal on Multiscale Modeling and Simulation*, 6(2) :547–576, 2007.

- [86] J. Eckstein. *Splitting methods for monotone operators with application to parallel optimization*. PhD thesis, MIT, 1989.
- [87] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55, 1992. 293–318.
- [88] J. Eckstein and B. F. Svaiter. A family of projective splitting methods for the sum of two maximal monotone operators. *Mathematical Programming, Series B*, 111, 2008. 173–199.
- [89] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3), 2007. 947–968.
- [90] M. Elad and I. Yavneh. A plurality of sparse representations is better than the sparsest one alone. *IEEE Transactions on Information Theory*, 2009. to appear.
- [91] V.A. Epanechnikov. Nonparametric estimates of a multivariate probability density. *Theory Prob. Appl.*, 14 :153–158, 1969.
- [92] A. Feuer and A. Nemirovsky. On sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 49(6), 2003. 1579–1581.
- [93] D. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal Optical Society America, A*, 4 :2379–2394, 1987.
- [94] D.J. Field. Wavelets, vision and the statistics of natural scenes. *Philosophical Transactions of the Royal Society of London A*, 357, 1999. 2527–2542.
- [95] M. Figueiredo, J. L. Bioucas-Dias, and R. Nowak. Majorization–minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 16(12), 2007. 2980–2881.
- [96] M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8), 2003. 906–916.
- [97] M. Fisz. The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, 3, 1955. 138–146.
- [98] P. Flandrin. Wavelet analysis and synthesis of fractional brownian motion. *IEEE Trans. Inf. Theory*, 38 :910–917, 1992.
- [99] S. Foucart and M.-J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$. *Applied and Computational Harmonic Analysis*, 26(3), 2009. 395–407.
- [100] P. Fryżlewicz and G. P. Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 13, 2004. 621–638.
- [101] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6), 2004. 1341–1344.
- [102] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods : Applications to the Solution of Boundary-Value Problems*. North-Holland, Amsterdam, 1983.
- [103] I. Gannaz. Robust estimation and wavelet thresholding in partially linear models. *Statistics and Computing*, 10(4) :293–310, 2007.
- [104] D. Geman, S. Geman, C. Gragne, and P. Dong. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 1990.
- [105] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 1984. 721–741.
- [106] C.W. Granger and R. Joyeux. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1 :15–29, 1980.
- [107] U. Grenander. *Lectures in Pattern Theory I, II and III*. Springer, 1976–1981.
- [108] U. Grenander, Y. Chow, and D. Keenan. *HANDS A Pattern-theoretic Study*. Springer, 1991.
- [109] U. Grenander, Xiuwen Liu, and A. Srivastava. Universal analytical forms for modeling image probabilities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9) :1200–1214, 2002.

- [110] U. Grenander and A. Srivastava. Probability models for clutter in natural images. *IEEE Trans. PAMI*, 23(4) :424–429, 2001.
- [111] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12), 2003. 3320–3325.
- [112] R. Gribonval and M. Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure. *Applied and Computational Harmonic Analysis*, 22(3), 2007. 335–355.
- [113] R. Gribonval and M. Nielsen. Beyond sparsity : recovering structured representations by 1-minimization and greedy algorithms. Application to the analysis of sparse underdetermined ICA. *Journal of Advances in Computational Mathematics*, 28, 2008. 23–41.
- [114] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst. Atoms of all channels, unite ! Average case analysis of multi-channel sparse recovery using greedy algorithms. *Journal of Fourier Analysis and Applications*, 14(5), 2008. 655–687.
- [115] R. Gribonval and K. Schnass. Dictionary identification - sparse matrix-factorisation via ℓ_1 -minimisation. arXiv preprint 0904.4774. submitted.
- [116] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM Journal on Numerical Analysis*, 39 :298–318, 2007.
- [117] S. Gurevich, R. Hadani, and N. Sochen. On some deterministic dictionaries supporting sparsity. *Journal of Fourier Analysis and Applications*, 14, 2008. 859–876.
- [118] W. Härdle, G. Kerkycharian, D. Picard, and A. B. Tsybakov. *Wavelets, Approximation and Statistical Applications*. Lecture Notes in Statistics. Springer, New York, 1998.
- [119] W. Härdle, H. Liang, and J. Gao. *Partially Linear Models*. Springer Verlag, 2000.
- [120] J. Huang. *Statistics of Natural Images and Models*. PhD thesis, Division Applied Mathematics, Brown University, 2000.
- [121] A. Hyvärinen and P. Hoyer. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 41, 2001. 2413–2433.
- [122] S. Jehan-Besson. *Modèles de contours actifs basés régions pour la segmentation d'images et de vidéos*. PhD thesis, Université de Nice - Sophia Antipolis, january 2003.
- [123] I. Johnstone. Wavelets and the theory of non-parametric function estimation. *Philosophical Transactions of the Royal Society of London A*, 357, 1999. 2475–2494.
- [124] I. Johnstone. Chi-square oracle inequalities. In C. Klaassen M. de Gunst and A. van der Vaart, editors, *State of the Art in Probability and Statistics*, volume 37 of *IMS Lecture Notes*, pages 399–418. Institute of Mathematical Statistics, 2001.
- [125] Sadegh Jokar and Volker Mehrmann. Sparse solutions to underdetermined kronecker product systems. *Linear Algebra and its Applications*, 431 :2437—2447, 2009.
- [126] B. Julesz. Textons, the elements of texture perception. *Nature*, 290(91-97), 1981.
- [127] J. Khalifa, S. Mallat, and B. Rougé. Deconvolution by thresholding in mirror wavelet bases. *IEEE Transactions on Image Processing*, 12(4), 2003. 446–457.
- [128] E. Kolaczyk. Nonparametric estimation of gamma-ray burst intensities using Haar wavelets. *Astrophysical Journal*, 483, 1997. 340–349.
- [129] E. Kolaczyk and D. Dixon. Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics. *Astrophysical Journal*, 534, 2000. 490–505.
- [130] Z. Koldovsky, P. Tichavsky, and E. Oja. Efficient variant of algorithm FastICA for Independent Component Analysis attaining the Cramér-Rao lower bound. *IEEE Transactions on Neural Networks*, 17, 2006. 1265–1277.
- [131] A. P. Korostelev and A. B. Tsybakov. *Minimax Theory of Image Reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993.
- [132] S. Kotz. *Continuous multivariate distributions, models and applications*. John Wiley & Sons, 2004.

- [133] M. Kowalski. Sparse regression using mixed norms. *Applied and Computational Harmonic Analysis*, 27(3) :303 – 324, 2009.
- [134] K. Kreutz-Delgado, J.F. Murray, B.D. Rao, K. Engan, T. Lee, and T.J. Sejnowski. Dictionary learning algorithms for sparse representation. *Neural Computation*, 15(2), 2003. 349–396.
- [135] Y. Gousseau et J.-M. Morel L. Alvarez. The size of objects in natural and artificial image. In *Advances in Imaging and Electron Physics*, volume 111. Academic Press, Elsevier, 1999.
- [136] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, 28(5) :1303–1338, 1998.
- [137] E. Le Pennec and S. Mallat. Sparse geometric image representations with bandelets. *IEEE Transactions on Image Processing*, 14(4), 2005. 423–438.
- [138] A. Lee, D. Mumford, and J. Huang. Occlusion models for natural images. *International Journal of Computer Vision*, 41(35-59), 2001.
- [139] A. B. Lee, A. Srivastava, E.P. Simoncelli, and S.C. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18 :17–33, 2003.
- [140] C. Lemaréchal and J.-B. Hiriart-Urruty. *Convex Analysis and Minimization Algorithms I and II*. Springer, 2nd edition, 1996.
- [141] P. Lévy. *Théorie de l'addition des variables aléatoires*. Gautier-Villars, Paris, 1954.
- [142] M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2), 2000. 337–365.
- [143] P.-L. Lions. Une méthode itérative de resolution d'une inéquation variationnelle. *Israel Journal of Mathematics*, 31, 1978. 204–208.
- [144] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16, 1979. 964–979.
- [145] F. Luisier, T. Blu, and M. Unser. A new SURE approach to image denoising : Interscale orthonormal wavelet thresholding. *IEEE Transactions on Image Processing*, 16(3), March 2007. 593–606.
- [146] F. Luisier, T. Blu, and M. Unser. *The SURE-LET Approach to Image Denoising*. PhD thesis, EPFL, Lausanne, Switzerland, November 2009.
- [147] S. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 1989. 674–693.
- [148] S. Mallat. *A Wavelet Tour of Signal Processing, The Sparse Way*. Academic Press, 3rd edition, December 2008.
- [149] S. Mallat and G. Peyré. Orthogonal bandlet bases for geometric images approximation. *Communications on Pure and Applied Mathematics*, 61(9), 2008. 1173–1212.
- [150] S. Mallat and Z. Zhang. Atomic decomposition by basis pursuit. *IEEE Transactions on Signal Processing*, 41, 1993. 3397–3415.
- [151] B. B. Mandelbrot and J. W. van Ness. Fractional brownian motion, fractional noises and applications. *SIAM Rev.*, 10(4) :422–437, 1968.
- [152] K. V. Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3) :519–530, 1970.
- [153] D. Marr. *Vision*. W.H. Freeman, 1982.
- [154] B. Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. *Comptes Rendus de l'Académie des Sciences de Paris*, 1274, 1972. 163–165.
- [155] Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Uniform uncertainty principle for bernoulli and subgaussian ensembles. *Constructive Approximation*, 28 :277–289, 2008.
- [156] F. G. Meyer and R. R. Coifman. Brushlets : A tool for directional image analysis and image compression. *Applied and Computational Harmonic Analysis*, 4(2) :147 – 187, 1997.

- [157] Y. Meyer. Oscillating patterns in image processing and in some nonlinear evolution equations. *The Fifteenth Dean Jacqueline B. Lewis Memorial Lecture*, 2001.
- [158] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes Rendus de l'Académie des Sciences Série A Mathématiques*, 255, 1962. 2897–2899.
- [159] J.-J. Moreau. Propriétés des applications “prox”. *Comptes Rendus de l'Académie des Sciences Série A Mathématiques*, 256, 1963. 1069–1071.
- [160] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93, 1965. 273–299.
- [161] D. Mumford. *Algebraic Geometry and its Applications*, chapter Elastica and Computer Vision, pages 507–518. Springer-Verlag, 1993.
- [162] D. Mumford and B. Gidas. Stochastic models for generic images. *Quarterly of Applied Mathematics*, 4(1) :85–111, 2001.
- [163] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions. *Communications in Pure and Applied Mathematics*, 42 :577–685, 1989.
- [164] F. Murtagh, J.-L. Starck, and A. Bijaoui. Image restoration with noise suppression using a multiresolution support. *Astronomy and Astrophysics, Supplement Series*, 112, 1995. 179–189.
- [165] R. Neelamani, H. Choi, and R. G. Baraniuk. ForWard : Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on Signal Processing*, 52(2), 2004. 418–433.
- [166] Yu. Nesterov. Gradient methods for minimizing composite objective function. CORE Discussion Papers 2007076, Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE), September 2007.
- [167] X. S. Nia and X. Huo. Another look at Huber’s estimator : A new minimax estimator in regression with stochastically bounded noise. *Journal of Statistical Planning and Inference*, 139(2), 2009. 503–515.
- [168] M. Nikolova. Local strong homogeneity of a regularized estimator. *SIAM Journal on Applied Mathematics*, 61(2), 2000. 633–658.
- [169] B.A. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set : A strategy employed by V1? *Vision Research.*, 37, 1996. 3311–3325.
- [170] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33 :1065–1076, 1962.
- [171] G.B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *Journal of Mathematical Analysis and Applications*, 72, 1979. 383–390.
- [172] Gabriel Peyré and Stéphane Mallat. A review of bandlet methods for geometrical image representation. *Numerical Algorithms*, 44(3), 2007. 205–234.
- [173] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11), November 2003. 1338–1351.
- [174] I. Ur Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and Peter Schröder. Multiscale representations for manifold-valued data. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4), 2005. 1201–1232.
- [175] H. Rauhut, K. Schass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory*, 54(5), 2008. 2210–2219.
- [176] O. Rioul and P. Duhamel. Fast algorithms for discrete and continuous wavelet transforms. *IEEE Transactions on Information Theory*, 2(38), 1992. 569–586.
- [177] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, 27 :832–835, 1956.
- [178] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8), 2007. 1025–1045.

- [179] D. Ruderman. Origins of scaling in natural images. *Vision Research*, 37(23) :3385–3398, 1997.
- [180] D. Ruderman and W. Bialek. Statistics of natural images : Scaling in the woods. *Physical Review Letters*, 73, 1994.
- [181] V. Saligrama. Deterministic designs with deterministic guarantees : Toeplitz compressed sensing matrices, sequence designs and system identification. ArXiv preprint, 2008.
- [182] F. Santosa and W.W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Statist. Comput.*, 7 :1307–1330, 1986.
- [183] M. Schwab, N. Karrenbach, and J. Claerbout. Making scientific computations reproducible. *Computing in Science and Engineering*, 2, 2000. 61–67.
- [184] R. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley, New York, 1980.
- [185] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [186] J. Shi and S. Osher. A nonlinear inverse scale space method for a convex multiplicative noise model. *SIAM Journal on Imaging Sciences*, 1(3) :294–321, 2008.
- [187] E.P. Simoncelli. Bayesian denoising of visual images in the wavelet domain. In P Müller and B Vidakovic, editors, *Bayesian Inference in Wavelet Based Models*, chapter 18. Springer-Verlag, New York, 1999. 291–308.
- [188] E.P. Simoncelli and B.A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 2001. 1193–1216.
- [189] J. G. Skellman. The frequency distribution of the difference between two Poisson variates belonging to different populations. *J. Roy. Statist. Soc. ser. A*, 109, 1946. 296.
- [190] 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, 2004.
- [191] 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), 2005. 1570–1582.
- [192] J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer, 2006. 2nd edn.
- [193] C. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9, 1981. 1135–1151.
- [194] T. Strohmer and R. W. Heath. Grassmannian frames with applications to coding and communication. *Applied and Computational Harmonic Analysis*, 14 :257–275, 2003.
- [195] J.A. Tropp. Algorithms for simultaneous sparse approximation. Part II : Convex relaxation. *Signal Processing*, 86(589-602), 2006.
- [196] J.A. Tropp. Just relax : Convex programming methods for subset selection and sparse approximation. *IEEE Transactions on Information Theory*, 52(3), 2006. 1030-1051.
- [197] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control and Optimization*, 29(1), 1991. 119–138.
- [198] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38, 2000. 431–446.
- [199] E. Van Den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2), 2008. 890–912.
- [200] L.A. Vese and S. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19, 2003. 553–577.
- [201] M. Wainwright and E.P. Simoncelli. Scale mixtures of gaussians and the statistics of natural images. In *Advances in Neural Information Processing Systems*, volume 12, pages 855–861. MIT Press, 2000.
- [202] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso). *IEEE Transactions on Information Theory*, 55(5), 2009. 2183–2202.

- [203] Wavelab 802. Wavelab toolbox. <http://www-stat.stanford.edu/~wavelab>, 2001.
- [204] P. Weiss. *Algorithmes rapides d'optimisation convexe : application à la restauration d'images et à la détection de changements*. PhD thesis, Université Sophia-Antipolis, Nice, France, 2008.
- [205] R. Willett and R. Nowak. Platelets : a multiscale approach for recovering edges and surfaces in photon-limited medical imaging. *IEEE Transactions on Medical Imaging*, 22(3), 2003. 332–350.
- [206] R. Willett and R. Nowak. Fast multiresolution photon-limited image reconstruction. In *IEEE ISBI*, 2004.
- [207] G. W. Wornell. A karhunen-loeve-like expansion for 1/f process via wavelets. *IEEE Trans. Inf. Theory*, 36(4) :859–861, 1990.
- [208] L. Ying and Y. M. Zou. Linear transformations and restricted isometry property. *CoRR*, abs/0901.0541, 2009.
- [209] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society B*, 68(1), 2006. 49–67.
- [210] S.C. Zhu, Y.Wu, and D.Mumford. Minimax entropy principle and its application to texture modeling. *Neural Computation*, 9 :1627–1660., 1997.
- [211] M. Zibulevski. Blind source separation with relative Newton method. *Proceedings of ICA, Independent Component Analysis, 2003*, 2003. 897-902.

Annexe A

Curriculum Vitae

ETAT CIVIL

Nom	FADILI
Prénom(s)	Mohamed JALAL
Date et lieu de naissance	07/02/73 Casablanca (Maroc).
Nationalité	Franco-Marocaine.
Adresse	GREYC UMR CNRS 6072 ENSICAEN-6 Bd Maréchal Juin 14050 Caen France.
Téléphone	02-31-45-29-20
Fax	02-31-45-26-98
Mél	Jalal.Fadili@greyc.ensicaen.fr
Page web	http ://www.greyc.ensicaen.fr/~jfadili/

FORMATION

Déc. 1999	Thèse de doctorat spécialité Traitement du Signal et des Images de l'Université de Caen. Mention "Très honorable et félicitations du jury".
1996	DEA Mesures, Capteurs et Images (Université de Caen) - Mention Bien.
1996	Diplôme d'ingénieur ENSICAEN - Spécialisation : Instrumentation (Major de filière).

PARCOURS SCIENTIFIQUE

Depuis Sept 2001	Maître de Conférences (61 ^{ème} section) à l'ENSICAEN affecté au GREYC UMR CNRS 6072. Responsable du groupe "Approches géométrique, statistique et variationnelle en traitement d'images".
2000-2001	ATER (61 ^{ème} section) IUT de Caen affecté au GREYC UMR CNRS 6072.
1999-2000	Post-doctorat à l'Univ. de Cambridge, (Royaume Uni) (Bourse McDonnell Foundation).

PRIMES ET DISTINCTIONS

Depuis 2003	PEDR (Prime d'encadrement doctorale et de recherche).
1999-2000	McDonnell Pew Fellow.

ENSEIGNEMENT

Type et discipline Je suis en charge d'enseigner le traitement, l'acquisition et la reconstruction des images (sous forme CM-TP), les mathématiques pour le signal et les images (sous forme TD-TP), la modélisation et traitement statistiques des images (CM-TP), la modélisation des signaux et des systèmes (TD-TP), ainsi que différentes disciplines en informatique dont la programmation et l'algorithmique (sous forme TD-TP).

Public concerné J'enseigne principalement en école d'ingénieurs (tout niveau). Je suis co-responsable de l'organisation des enseignements de l'informatique en seconde année dans la spécialité Electronique. En outre, j'ai assuré la mise en place de deux nouveaux cours magistraux pour les 3^{ème} années Electronique (options Optique et Nucléaire) et Informatique.

J'interviens aussi dans différents Masters professionnel et de recherche habilités à l'Université de Caen. J'assure ainsi le CM "Acquisition, reconstruction et prétraitements des images" dans le cadre du Master Recherche ESCI, le CM de "Traitement et modélisation statistiques des images" en Master Recherche LID (cours commun aux 3^{ème} années Informatique option Image) ainsi que le CM-TP "Classification et segmentation des images" en Master Pro RADI.

En 2008, j'ai assuré deux cours lors de la série organisée à l'IHP sur les "Méthodes Variationnelles et Parcimonieuses en Traitement des Signaux et des Images" avec Gabriel Peyré (Céramade, Paris Dauphine) et Jean-François Aujol (LATP, Marseille).

Documents pédagogiques produits

- Polycopié et diapositives "Acquisition et prétraitement des images" (M2 Recherche ESCI).
- Polycopié et diapositives "Eléments fondamentaux du prétraitement des images" (3A Electronique).
- Polycopié "Traitement et modélisation statistiques des images" (3A Informatique et M2 Recherche LID).
- Diapositives "Classification et segmentation des images" (M2 Pro. RADI).

THÈMES DE RECHERCHE

Mots clés : Traitement du signal et des images. Approches statistiques pour la modélisation, l'estimation et la détection. Analyse harmonique computationnelle. Représentations parcimonieuses. Approches variationnelles. Analyse convexe et optimisation non lisse. Problèmes inverses et restauration. Imagerie médicale et astronomique.

Description : Mes activités de recherche se trouvent à la croisée des chemins entre le traitement du signal et des images et les mathématiques appliquées. Plus précisément, mes champs de recherche concernent les théories de l'estimation et de la détection statistiques, la modélisation stochastique, l'analyse harmonique computationnelle, les représentations parcimonieuses, l'optimisation non lisse et la résolution des problèmes d'optimisation dans les formulations variationnelles.

Résultats marquants : Ces travaux théoriques trouvent naturellement des applications pour nombre de problèmes inverses en traitement d'images tels que la restauration, la super-résolution, l'inpainting, la séparation de composantes ou encore la théorie de l'échantillonnage compressé (compressed sensing). Avec mes collaborateurs, nous avons ainsi proposé de nouveaux algorithmes de restauration avec des bruits au-delà du cas Gaussien. Nous avons aussi proposé des algorithmes de résolution de problèmes inverses sous contraintes dans un cadre unifié par décomposition d'opérateurs monotones maximaux. Je travaille par ailleurs en partenariat avec plusieurs collaborateurs académiques et industriels où nous mettons à profit nos compétences dans plusieurs domaines applicatifs comme l'imagerie médicale et biomédicale (IRM fonctionnelle, échographie, microscopie confocale), l'imagerie astronomique avec le CEA et le contrôle non destructif.

COLLABORATIONS ACADÉMIQUES

Collaborations nationales

- Je collabore très étroitement depuis 2003 avec Jean-Luc Starck (CEA Saclay) sur les problèmes inverses, la parcimonie, la séparation de sources, l'estimation statistique et leur application en imagerie astronomique. Nous avons co-encadré une thèse (Jérôme Bobin), et deux autres sont en cours. Nous sommes aussi associés à plusieurs actions communes dont l'ANR NatImages et l'ERC SparseAstro. Dans ces travaux, nous avons aussi collaboré régulièrement avec Yassir Moudden (CEA Saclay). Cette collaboration avec J.-L. Starck a récemment abouti à un livre co-rédigé en association avec F. Murtagh, et qui résume notre activité des six dernières années dans le domaine du traitement parcimonieux des signaux et des images.
- Je collabore intensivement avec Gabriel Peyré (CEREMADE-Univ. Paris Dauphine), Charles Dossal (LaBAG, Univ. Bordeaux 1), et Jean-François Aujol (CMLA-ENS Cachan) sur les sujets liés à la modélisation des images et des textures, l'optimisation non lisse et l'échantillonnage compressé. Nous sommes associés dans le cadre de l'ANR NatImages et d'un projet jeunes chercheurs du GDR ISIS.
- Avec Christophe Chesneau du laboratoire de mathématiques (LMNO-Univ. de Caen), nous travaillons sur les problèmes statistiques liés à la théorie de l'estimation. C. Chesneau fait aussi partie de l'ANR NatImages.
- Avec Mila Nikolova (CMLA-ENS Cachan) et Sylvain Durand (MAP5-Univ. Paris 5), nous collaborons sur des thématiques liées à l'optimisation non lisse et ses applications en restauration des images.
- Collaboration avec David Tschumperlé (GREYC) sur des problématiques de restauration et de décomposition d'images avec des applications en contrôle non destructif.
- Avec Stéphanie Jehan-Besson (LIMOS-Univ. Clermont) et Marinette Revenu (GREYC), nous avons co-encadré une thèse (François Lecellier) alliant les contours actifs basés régions avec un cadre statistique d'une part et les représentations parcimonieuses d'autre part pour la segmentation des régions bruitées et texturées.

Collaborations internationales

- J'ai initié ma collaboration avec le Pr. Ed Bullmore (Univ. de Cambridge) lors de mon séjour post-doctoral, sur les méthodes multiéchelles (e.g. ondelettes) pour l'analyse des séries spatio-temporelles en IRM fonctionnelle. Cette collaboration a donné lieu à un projet international commun financé par le NIH (USA), dans le cadre duquel nous avons co-encadré des post-docs.
- Je collabore depuis 2005 avec Michael Elad (Technion) et le Pr. David Donoho (Univ. Stanford) sur les méthodes de décompositions parcimonieuses structurées et la séparation de sources.
- Collaboration avec le Pr. Fionn Murtagh (Royal Univ. London) sur les approches statistiques et multiéchelles pour l'analyse des images avec une application en imagerie astronomique.
- Je collabore régulièrement avec Dimitri VandeVelle (Univ. de Genève et EPFL) sur les approches multiéchelles (e.g. ondelettes) pour l'analyse des séries spatio-temporelle en IRM fonctionnelle.
- Je collabore avec le Pr. Pierre Vandergheynst (EPFL) sur l'application des méthodes d'optimisation non lisse pour la résolution de problèmes inverses avec régularisation parcimonieuse.
- Collaboration avec Laurent Jacques (EPFL et UCL) et David Hammond (EPFL) sur l'échantillonnage compressé.
- Collaboration avec Seth Digel (SLAC-Univ. Stanford) sur le débruitage d'images de rayonnement gamma en astronomie dans le cadre du projet Herschel/Fermi.

ENCADREMENTS POUR LA RECHERCHE

- 2 post-docs.
- 6 thèses (dont 4 soutenues).
- Plusieurs stages de Master.

Thèses

- Larbi Boubchir (2002-Soutenue 2007) : Approches bayésiennes pour le débruitage des images dans le domaine des transformées multiéchelles orientées et non orientées. Co-encadrement à 80% avec M. Revenu au GREYC à Caen. Bourse MNERT. Actuellement ATER.
- François Lecellier (2005-Soutenue 2009) : Segmentation par contours actifs basés régions avec a priori de bruit, de forme et de texture : application à l'échocardiographie. Co-encadrement à 30% avec S. Jéhan-Besson et M. Revenu au GREYC à Caen. Bourse BDI CNRS-Région. Actuellement post-doctorant à l'IRISA.
- Jérôme Bobin (2005-Soutenue 2008) : Diversité morphologique et analyse de données multivaluées. Co-encadrement à 50% avec J.-L. Starck au CEA-Univ. Paris XI. Bourse CEA. Actuellement post-doctorant ACM-Caltech et Stanford.
- François-Xavier Dupé (2006-Soutenue Jan. 2010) : Restauration et segmentation d'images avec bruit de Poisson : application à la microscopie confocale. Co-encadrement à 30% avec L. Brun et A. Buisson au GREYC à Caen. Bourse BDI CNRS-Région.
- Arnaud Woiselle (depuis Oct. 2007) : Représentations multiéchelles géométriques 3D : application à l'analyse et la restauration des vidéos. Co-encadrement à 50% avec J.-L. Starck au CEA-Univ. Paris 7. Bourse Cifre de SAGEM DS.
- Lila Hadji (depuis Oct. 2009) : Construction de dictionnaires adaptatifs et applications en astronomie. Co-encadrement à 50% avec J.-L. Starck au CEA-Univ. Paris 7. Bourse CEA.

Post-docs

- Hua Li (Fév. 2003-Janv. 2004) : Fast 3D level set methods : Application to object segmentation and tracking. Co-encadrement à 50% avec A. Elmoataz au GREYC à Caen. Financement région.
- Voichita Maxim (2003-2004) : Analyse par ondelettes des séries d'IRM fonctionnelle. Co-encadrement à 50% avec E. Bullmore à l'univ. de Cambridge. Financement région.

Stages

- François Lecellier (2005, M2 Recherche LID à Caen) : Segmentation par contours actifs basés régions avec modèle de bruit et a priori de forme. Co-encadré à 30% avec S. Jéhan-Besson et M. Revenu au GREYC. Collaboration CHU de Caen.
- Bo-Zhang (2004, M2 Recherche MVA ENS Cachan et Ingénieur ENST) : Débruitage basé ondelettes d'images avec bruit de Poisson. Co-encadré à 50% avec J.-L. au CEA.
- Mohamed Boukelia (2004, M2 Recherche ESCI à Caen) : Segmentation par contours actifs basés régions avec modèle de bruit. Co-encadré à 50% avec S. Jéhan-Besson au GREYC.
- Frédéric Bataille (2002, M2 Recherche ESCI à Caen et Ingénieur ENSICAEN) : Mesure de netteté d'un autofocus par transformée d'ondelettes et bruit des caméras. Co-encadrement à 50% M. Desvignes au GREYC. Collaboration Philips Semi-conducteurs.
- Jérôme Mathieu (2002, M2 Recherche ESCI à Caen et Ingénieur ENSICAEN) : Débruitage bayésien dans le domaine des ondelettes avec a priori à queues lourdes. Co-encadrement à 50% M. Desvignes au GREYC.
- Samuel Planchot (2003, Ingénieur ENSI-CAEN) : Décomposition parcimonieuse de signaux : application à la réjection d'artéfacts EEG en IRM fonctionnelle. Co-encadrement à 50% M. Joliot entre le GREYC et Cyceron. Projet CNRS SDV-STIC.

H.L. Leyva (1998-1999, Ingénieur Univ. polytechnique de Valence Espagne) :

Implémentation temps-réel de débruitage par ondelettes pour un capteur magnétique portable ultra-sensible. Co-encadrement à 50% Ch. Dolabdjian au GREYC.

JURYS DE THÈSES EXTÉRIEURES

- Novembre 2009 : Rapporteur de la thèse de Florian Luisier (directeurs : T. Blu et M. Unser, EPFL, Lausanne).
- Avril 2007 : Membre du jury de thèse de Sylvain Lesage (directeur : R. Gribonval, IRISA, Rennes).
- Décembre 2007 : Membre du jury de thèse de Irène Gannaz (directeur : A. Antoniadis, Laboratoire Jean Kuntzmann, INPG, Grenoble).

ORGANISATION DE COLLOQUES

Organisation de conférences et de workshops

- Novembre 2009 : Co-organisateur avec Jean-Luc Starck d'une session spéciale "Sparsity Tour of Inverse Problems" à l'occasion de IEEE ICIP .
- Août 2009 : Co-organisateur avec Jean-Luc Starck d'une session spéciale dans SPIE Wavelets XIII "Sparsity in Multidimensional Data Analysis".
- Mars-Avril 2008 : Co-organisateur avec Gabriel Peyré (Paris Dauphine) et Jean-François Aujol (ENS Cachan) de la série de cours "Méthodes Variationnelles et Parcimonieuses en Traitement des Signaux et des Images" à l'Institut Henri Poincaré (IHP), Paris.
- Janvier 2007 : Co-organisateur avec Gabriel Peyré (Paris Dauphine) et Jean-François Aujol (ENS Cachan) du workshop "An interdisciplinary approach to Textures and Natural Images Processing" à l'Institut Henri Poincaré (IHP), Paris.
- Membre du comité d'organisation du colloque "Imagerie Médicale et Santé" de l'IFR 47 en 2002 et 2003.

Comités de programme et de pilotage

- Comité de programme RFIA'09, Caen, France.
- Comité de programme Wavelets XII et XIII (SPIE-Mathematical Imaging) 2007, 2009 à San Diego, USA.
- Comité de programme SPARS'09, St Malo, France.
- Comité de pilotage de SPARS.
- Comité de programme International Conference on Image and Signal Processing ICISP 2008 à Cherbourg, France).

EXPERTISE, COMITÉS DE LECTURE ET ÉDITION DE NUMÉROS SPÉCIAUX

Revue internationale IEEE Trans. on Image Processing, IEEE Trans. on Signal Processing, IEEE Trans. on Signal Processing Letters, IEEE Trans. PAMI, IEEE Trans. on Medical Imaging, International Journal of Computer Vision, Neurocomputing, Journal of the American Optical Society, Computer Vision and Image Understanding, Signal Processing, EURASIP Journal of Applied Signal Processing, NeuroImage, Human Brain Mapping, Medical Image Analysis, Applied and Computational Harmonic Analysis, Journal of Fourier Analysis and its Application, SIAM Journal on Imaging Science, SIAM journal on Applied Mathematics, SIAM journal on Scientific Computing.

Conférences EUSIPCO, ICIP, SPARS, ICISP, GRETSI.

Edition de numéros spéciaux de revues internationales

- 2009-2010 : éditeur avec J.-L. Starck d'un numéro spécial "Compressed Sensing" pour la revue "Sampling Theory in Image and Signal Processing".
- 2010-2011 : éditeur avec J.-L. Starck, M. Elad d'un numéro spécial "Adaptive sparse representation of data" pour la revue "IEEE Journal on Selected Topics in Signal Processing".

CHARGES ADMINISTRATIVES ET COLLECTIVES

-
- Depuis Janv. 2010 : Directeur Scientifique Adjoint du GDR ISIS.
 - Depuis Sept. 2002 : Responsable du groupe "Approches géométrique, statistique et variationnelle en traitement d'images" de l'équipe Image.
 - Depuis 2003 : Membre élu collègue B de l'équipe Image au sein du conseil de laboratoire.
 - Depuis Sept. 2006 : Membre élu du conseil scientifique de l'ENSICAEN.
 - Depuis 2004 : Membre CS puis CoS 61^{ème} section à Caen et à Nice-Sophia Antipolis.

SÉJOURS INVITÉS

Juillet-Août 2005	Queensland University of Technology, School of Economics and Finance (Jérôme Collet et Rodney Wolff), Brisbane, Queensland, Australie.
Avril-Mai 2005	Université de Stanford, Département de Statistiques (Jean-Luc Starck et David Donoho), CA, USA.
Avril 2006	EPFL, Biomedical Imaging Group, (Dimitri VanDeVille), Lausanne, Suisse.
Juillet 2008	Université de Stanford, SLAC (Seth Digel), CA, USA.
Septembre 2008	EPFL, LTSI, (Pierre Vandergheynst), Lausanne, Suisse.

CONFÉRENCES INVITÉES

-
- Workshop "Geometrical Methods in Mathematical Imaging", Munich, Octobre 2009.
 - Workshop "Approximation and optimization in image restoration and reconstruction", Porquerolles, 2009.
 - SIAM International conference on Imaging Science en 2004, 2006, 2008, 2010.
 - International Conference on Astronomical Data Analysis, 2006 et 2008.
 - Workshop "Mathematical Image Processing", Luminy, Marseille, Septembre 2007.
 - Journées scientifiques "Images et Modélisations Mathématiques", IRMAR et IRISA Rennes, Décembre 2006.
 - Journées Mathématiques pour l'Image, INRIA et Univ. Nice France, Janvier 2006.
 - Wavelets and Applications Semester, Ecole Polytechnique Fédérale de Lausanne, Avril 2006.
 - IEEE International Symposium on Biomedical Imaging, Arlington USA, Avril 2004.
 - Wavelets X, SPIE-Mathematical Imaging : Wavelet Applications in Signal and Image Processing X, San Diego USA, Août 2003.

SÉMINAIRES ET GROUPES DE TRAVAIL

-
- De nombreux séminaires dans différents organismes et universités françaises et étrangères.
 - Participation à plusieurs groupes de travail dans le cadre du GDR ISIS, GDR MSPC, GT statistiques à l'université de Caen.

ACTIONS DE FINANCEMENT ET RELATIONS INDUSTRIELLES

Actions régionales

Projet régional CRITIC : GREYC - Philips Semi-conducteurs.

- Titre : Identification des bruits et optimisation de l'autofocus d'une caméra numérique.
- Responsables : M. Desvignes, M.J. Fadili, D. Bloyet.
- Dates : Oct. 2001-Déc. 2003.
- Budget : 80 000 Euros TTC (Feder), 20 000 Euros TTC (Etat).

Actions nationales

PICTI-CNRS Programme interdisciplinaire Cognition et Traitement de l'Information PICTI-CNRS STIC ⁴⁴/SDV ⁴⁵

- Sujet : Modulation attentionnelle des aires visuelles et auditives primaires : Etude chez le sujet sain en IRMf et EEG.
- Responsables : L. Petit (CR1 CNRS GIN ⁴⁶-Cyceron) et M.J. Fadili.
- Dates : 2002-2005.
- Budget : 88 600 Euros.

Jeunes Chercheurs GDR ISIS Programme Jeunes Chercheurs du GDR Information, Signal, Images et ViSion.

- Sujet : Représentations parcimonieuses et adaptatives pour les textures naturelles.
- Responsables : Gabriel Peyré (CR2 CNRS CEREMADE, Paris Dauphine), Jalal Fadili et Jean-François Aujol (CR2 CNRS CMLA, ENS Cachan).
- Dates : 18 mois depuis Fév. 2007.
- Budget : 6000 Euros.

Projet ANR NatImages

- Sujet : Adaptivity for Natural Images and Textures Representation.
- Porteur : Gabriel Peyré (CR2 CNRS CEREMADE, Univ. Paris Dauphine).
- Laboratoires participants : CEREMADE (CNRS Univ. Dauphine), GREYC (CNRS-ENSICAEN-Univ. de Caen), CMLA (CNRS-ENS Cachan), LAIM (CEA/DSM-CNRS-Univ. Paris 7), LMNO (CNRS-Univ. de Caen), IMB (CNRS-Univ. Bordeaux I).
- Statut GREYC : partenaire.
- Dates : 4 ans à compter du 1er Janvier 2009.
- Budget : 309 878 Euros dont 74 572 Euros pour le GREYC.

Actions internationales

Human Brain Project Human Brain Project (NIH USA) Grant from the National Institute of Biomedical Imaging and Bioengineering and the National Institute of Mental Health.

- Sujet : L'IRM pharmacologique et les fonctions cérébrales supérieures chez l'homme.
- Mots-clés : IRMf, neurosciences, ondelettes, rééchantillonnage aléatoire.
- Porteur : Pr. Ed Bullmore, Cambridge University.
- Nombre de participants : 7.
- Statut GREYC : partenaire.
- Rôle et expertise : traitements statistiques et multiéchelles des images.
- Dates : Avril 2003-Avril 2008.
- Budget : USD 1 350 000.
- Bénéfice : post-docs, missions.
- Page web : <http://www-bmu.psychiatry.cam.ac.uk/projects/nih-wavelets/>.

44. Sciences et Technologie de l'Information et de la Communication.

45. Sciences du Vivant.

46. Groupe d'Imagerie Neuro-fonctionnelle UMR CNRS 6194.

ERC grant SparseAstro Partenaire dans le cadre du programme European Research Council Advanced Researcher Grant.

- Sujet : Sparse Representation of Multi-Valued Images : Application in Astrophysics.
- Porteur : Jean-Luc Starck, LAIM CEA/DSM-CNRS.
- Statut GREYC : partenaire.
- Rôle et expertise : co-encadrement de thèses et de post-docs avec J.-L. Starck en traitements statistiques et parcimonieux des images.
- Dates : A partir de 2009 sur 5 ans.
- Bénéfice : co-encadrement de thèses et post-docs, missions.

Contrats et transferts de technologie

- Convention de collaboration de recherche avec l'Institut Français du Pétrole portant sur une "Etude de faisabilité pour la détermination numérique de fils d'armures à partir de radioscopie X de flexibles". Responsables : David Tschumperlé (CR GREYC), Jalal Fadili, Marinette Revenu (PR GREYC). Dates : avril 2005 à décembre 2005. Budget : 10.3 kEuros HT.
- Convention de collaboration de recherche avec l'Institut Français du Pétrole portant sur le "Suivi du mouvement des fils d'armures dans les séquences d'images du flexible sous radioscopie X". Responsables : David Tschumperlé (CR GREYC), Jalal Fadili, Marinette Revenu (PR GREYC). Dates : avril 2006 à décembre 2006. Budget : 17 kEuros HT.
- Convention de collaboration de recherche avec l'Institut Français du Pétrole portant sur le "Suivi du mouvement des fils d'armures dans les séquences d'images du flexible sous radioscopie X". Responsables : David Tschumperlé (CR GREYC), Jalal Fadili, Marinette Revenu (PR GREYC). Dates : avril 2007 à décembre 2007. Budget : 5 kEuros HT.
- Contrat Cifre avec SAGEM DS pour l'encadrement de la thèse d'Arnaud Woiselle. Budget : 13 kEuros. Dates : depuis octobre 2007.

Production de logiciels

- Fispex : Logiciel de restauration, de décomposition et de suivi des fibres de flexibles. Ce produit s'inscrit dans le cadre du contrat avec l'IFP pour une licence d'utilisation de 6 mois. Concepteurs : Jalal Fadili et David Tschumperlé (CR GREYC).
- MCALab : boîte à outils Matlab pour la décomposition en composantes morphologiques et l'inpainting des signaux 1D et des images 2D. Elle est disponible au téléchargement académique libre depuis 2005.
- SparseLab : logiciel C++ pour le traitement d'images avec des représentations parcimonieuses.
- DBlockToolbox : boîte à outils Matlab pour le débruitage et les problèmes inverses inverses par seuillage par blocs.
- MSVSTLab : boîte à outils Matlab pour le débruitage avec bruit de Poisson.
- SplittingSolvers : boîte à outils Matlab pour la résolution des problèmes linéaires inverses avec régularisations parcimonieuses.

Liste des thèses encadrées

Thèses encadrées

- [TH1] F.-X. Dupé. *Restauration et reconnaissance des formes dans des images avec bruits de Poisson : application à l'analyse des neurones en microscopie de fluorescence*. PhD thesis, Université de Caen, GREYC CNRS-ENSICAEN-Univ. de Caen, 2010.
- [TH2] J. Bobin. *Diversité morphologique et analyse de données multivaluées*. PhD thesis, Université Paris XI, LAIM CEA-CNRS, 2008.
- [TH3] F. Lecellier. *Les contours actifs basés région avec a priori de bruit, de texture et de forme : Application à l'échocardiographie*. PhD thesis, Université de Caen, GREYC CNRS-ENSICAEN-Univ. de Caen, 2009.
- [TH4] L. Boubchir. *Approches bayésiennes pour le débruitage des images dans le domaine des transformées multi-échelles parcimonieuses orientées et non orientées*. PhD thesis, Université de Caen, GREYC CNRS-ENSICAEN-Univ. de Caen, 2007.

Liste de publications

Tableau récapitulatif

Type de publication	Nombre
Livre	1
Chapitres de livres	4
Revue internationale avec comité de lecture	40
Revue nationale avec comité de lecture	2
Conférences internationales avec actes et comité de lecture	61
Conférences nationales avec actes et comité de lecture	12
Revue internationale soumise ou révisée	7
Rapports techniques	3

Livre

- [A1] J.-L. Starck, F. Murtagh, and M.J. Fadili. *Sparse Signal and Image Processing : Wavelets, Curvelets and Morphological Diversity*. Cambridge University Press, Cambridge, UK, 2010. in press.

Chapitres de livres

- [A2] M.J. Fadili and J.-L. Starck. Curvelets and ridgelets. In R.A. Meyers, editor, *Encyclopedia of Complexity and System Science*. Springer, 2009.
- [A3] J.-L. Starck and M.J. Fadili. Numerical issues when using wavelets. In R.A. Meyers, editor, *Encyclopedia of Complexity and System Science*. Springer, 2009.
- [A4] S. Jehan-Besson, F. Lecellier, M.J. Fadili, G. Née, and G. Aubert. *Biomedical diagnostics and clinical technologies : Applying high performance clu*, chapter Medical image segmentation and tracking through the maximisation or the minimisation of divergence between pdfs. 2009.
- [A5] J. Bobin, J.-L. Starck, Y. Moudden, and M.J. Fadili. Blind source separation : The sparsity revolution. In Peter Hawkes, editor, *Advances in Imaging and Electron Physics*, volume 152, pages 221–298. Academic Press, Elsevier, 2008.

Revue internationale

- [A6] A. Woiselle, J.-L. Starck, and M.J. Fadili. 3D curvelet transforms and astronomical data restoration. *Applied and Computational Harmonic Analysis*, 28(2) :171–188, 2010.
- [A7] F. Lecellier, M.J. Fadili, S. Jehan-Besson, G. Aubert, M. Revenu, and E. Saloux. Region-based active contours with exponential family observations. *Journal of Mathematical Imaging and Vision*, 36(1) :28–45, 2010.

- [A8] J. Schmitt, J.-L. Starck, J.-M. Casandjian, M.J. Fadili, and I. Grenier. Poisson denoising on the sphere. *Astronomy and Astrophysics*, 2010.
- [A9] C. Dossal, G. Peyré, and M.J. Fadili. A numerical exploration of compressed sampling recovery. *Linear Algebra and its Applications*, 432(7) :1663–1679, 2010.
- [A10] S. Durand, M.J. Fadili, and M. Nikolova. Multiplicative noise removal using l_1 fidelity on frame coefficients. *Journal of Mathematical Imaging and Vision*, 2009. in press.
- [A11] J.-L. Starck, M.J. Fadili, S. Digel, B. Zhang, and J. Chiang. Source detection using a 3d sparse representation : Application to the fermi gamma-ray space telescope. *Astronomy and Astrophysics*, 504(2) :641–652, 2009.
- [A12] F.-X. Dupé, M.J. Fadili, and J.-L. Starck. A proximal iteration for deconvolving poisson noisy images using sparse representations. *IEEE Transactions on Image Processing*, 18(2) :310–321, 2009.
- [A13] M.J. Fadili, J.-L. Starck, J. Bobin, and Y. Moudden. Image decomposition and separation using sparse representations : an overview. *Proceedings of the IEEE, Special Issue : Applications of Sparse Representation*, 2009. in press.
- [A14] S. Pires, J.-L. Starck, A. Amara, R. Teyssier, A. Réfrégier, and M.J. Fadili. Fastlens (fast statistics for weak lensing) : Fast method for weak lensing statistics and map making. *Monthly Notices of the Royal Astronomical Society*, 2009. in press.
- [A15] C. Chesneau, M.J. Fadili, and J.-L. Starck. Stein block thresholding for image denoising. *Applied and Computational Harmonic Analysis*, 28 :67–88, 2009.
- [A16] 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*, 33(2) :149–168, 2009.
- [A17] M.J. Fadili, J.-L. Starck, M. Elad, and D.L. Donoho. Mcalab : Reproducible research in signal and image decomposition and inpainting. *IEEE Computing in Science and Engineering*, 12(1) :44–62, 2010.
- [A18] J. Bobin, Y. Moudden, J.-L. Starck, M.J. Fadili, and J. Robin. Sz and cmb reconstruction using generalized morphological component analysis. *Statistical Methodology*, 5(4) :307–317, 2008.
- [A19] B. Zhang, M.J. Fadili, and J.-L. Starck. Wavelets, ridgelets and curvelets for poisson noise removal. *IEEE Transactions on Image Processing*, 17(7) :1093–1108, 2008.
- [A20] B. Zhang, M.J. Fadili, and J.-L. Starck. Fast poisson noise removal by biorthogonal haar domain hypothesis testing. *Statistical Methodology*, 5(4) :387–396, 2008.
- [A21] P. Abrial, Y. Moudden, J.-L. Starck, M.J. Fadili, J. Delabrouille, and M.K. Nguyen. Cmb data analysis and sparsity. *Statistical Methodology*, 5(4) :289–298, 2008.
- [A22] J.-L. Starck, M.J. Fadili, and F. Murtagh. The undecimated wavelet decomposition and its reconstruction. *IEEE Transactions on Image Processing*, 16(2) :297–309, 2007.
- [A23] J. Bobin, J.-L. Starck, M.J. Fadili, Y. Moudden, and D.L. Donoho. Morphological component analysis : An adaptive thresholding strategy. *IEEE Transactions on Image Processing*, 16(11) :2675–2681, 2007.
- [A24] P. Abrial, Y. Moudden, J.-L. Starck, B. Afeyan, J. Bobin, M.J. Fadili, and M.K. Nguyen. Morphological component analysis and inpainting on the sphere : Application in physics and astrophysics. *Journal of Fourier Analysis and Applications*, 13(6) :729–748, 2007.
- [A25] 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, 2007.
- [A26] M.J. Fadili, J.-L. Starck, and F. Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal*, 52(1) :64–79, 2007.
- [A27] L. Boubchir and M.J. Fadili. A closed-form nonparametric bayesian estimator in the wavelet-domain of images using an approximate alpha-stable prior. *Pattern Recognition Letters*, 27(12) :1370–1382, 2006.

- [A28] J. Suckling, M. Davis, C. Ooi, A. Meije Wink, M.J. Fadili, R. Salvador, D. Welchew, L. Sendur, V. Maxim, and E. Bullmore. Permutation testing of orthogonal, factorial effects in a language processing experiment using fmri. *Human Brain Mapping*, 27(5) :425–433, 2006.
- [A29] V. Maxim, L. Sendur, M.J. Fadili, J. Suckling, R. Gould, and R. Howard. Fractional gaussian noise, functional mri and alzheimer’s disease. *NeuroImage*, 25 :141–158, 2005.
- [A30] M.J. Fadili and L. Boubchir. Analytical form for a bayesian wavelet estimator of images using the bessel k form densities. *IEEE Transactions on Image Processing*, 14(2) :231–240, 2005.
- [A31] M.J. Fadili and E.T. Bullmore. Penalized partially linear models using sparse representations with an application to fmri time series. *IEEE Transactions on Signal Processing*, 53(9) :3436–3448, 2005.
- [A32] M.J. Fadili and E. Bullmore. A comparative evaluation of wavelet-based methods for hypothesis testing of brain activation maps. *NeuroImage*, 23(3) :1112–1128, 2004.
- [A33] M.J. Fadili, M. Melkemi, and A. Elmoataz. Non-convex onion peeling using a shape hull algorithm. *Pattern Recognition Letters*, 25(14) :1577–1585, 2004.
- [A34] E. Bullmore, M.J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear. Wavelets and functional magnetic resonance imaging of the human brain. *NeuroImage (invited), Mathematics in Brain Imaging*, 23 :234–249, 2004.
- [A35] E.T. Bullmore, M.J. Fadili, M. Breakspear, R. Salvador, J. Suckling, and M. Brammer. Wavelets and statistical analysis of functional magnetic resonance images of the human brain. *Statistical Methods in Medical Research (invited)*, 12(5) :375–399, 2003.
- [A36] M.J. Fadili and E.T. Bullmore. Wavelet-generalised least squares : a new blu estimator of linear regression models with 1/f errors. *NeuroImage*, 15 :217–232, 2002.
- [A37] C. Dolabdjian, M.J. Fadili, and E. Huertas Leyva. Classical low-pass filter and real-time wavelet-based denoising technique implemented on a dsp : a comparison study. *The European Physical Journal Applied Physics*, 20 :135–140, 2002.
- [A38] L.K. Pilgrim, M.J. Fadili, P. Fletcher, and L.K. Tyler. Overcoming confounds of stimulus blocking : An event-related fmri design of semantic processing. *NeuroImage*, 16 :713–723, 2002.
- [A39] S. Ruan, B. Moretti, M.J. Fadili, and D. Bloyet. Fuzzy markovian segmentation in application of magnetic resonance images. *Computer Vision and Image Understanding*, 85 :54–69, 2002.
- [A40] S. Gustard, M.J. Fadili, E.J. Williams, L.D. Hall, T.A. Carpenter, M. Brett, and E.T. Bullmore. Effect of slice orientation on reproducibility of fmri motor activation at 3 tesla. *Magnetic Resonance Imaging*, 19(10) :1323–1331, 2001.
- [A41] M.J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer. On the number of clusters and the fuzziness index for unsupervised fca application to bold fmri time series. *Medical Image Analysis*, 5(1) :55–67, 2001.
- [A42] E.T. Bullmore, C. Long, J. Suckling, M.J. Fadili, G. Calvert, F. Zelaya, T.A. Carpenter, and M. Brammer. Colored noise and computational inference in neurophysiological fmri time series analysis : Resampling methods in time and wavelet domains. *Human Brain Mapping*, 12(2) :61–78, 2001.
- [A43] M.J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer. A multistep unsupervised fuzzy clustering analysis of fmri time series. *Human Brain Mapping*, 10(4) :160–178, 2000.
- [A44] S. Ruan, C. Jaggi, J. Xue, M.J. Fadili, and D. Bloyet. Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Transactions on Medical Imaging*, 19(12) :1179–1187, 2000.
- [A45] B. Moretti, M.J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer. Phantom-based performance evaluation - application to brain segmentation from magnetic resonance images. *Medical Image Analysis*, 4(4) :303–316, 2000.

Revue nationale

- [A46] M.J. Fadili, J. Mathieu, and M. Desvignes. La transformation de fisz pour l'estimation de l'image des intensités d'un bruit poissonien dans le domaine des ondelettes. *Traitement du Signal*, 21(4) :313–328, 2004.
- [A47] M.J. Fadili, B. Moretti, S. Ruan, and D. Bloyet. Méthode coopérative et quantification de la segmentation de l'encéphale en imagerie par résonance magnétique. *Traitement du signal*, 15(5) :353–363, 1998.

Conférences internationales

- [A48] M.J. Fadili and G. Peyré. Total variation projection with first order schemes. In *IEEE ICIP*, Cairo, 2009.
- [A49] C. Dossal, G. Peyré, and M.J. Fadili. Challenging restricted isometry constants with greedy pursuit. In *IEEE Information Theory Workshop*, Taormina, 2009.
- [A50] C. Dossal, G. Peyré, and M.J. Fadili. A numerical exploration of compressed sampling recovery. In *SPARS'09*, St Malo, 2009.
- [A51] Y. Moudden, J. Bobin, J.-L. Starck, and M.J. Fadili. Dictionary learning with spatio-spectral sparsity constraints. In *SPARS'09*, St Malo, 2009.
- [A52] C. Chesneau, M.J. Fadili, and J.-L. Starck. Image deconvolution by block stein thresholding. In *IEEE ICIP*, Cairo, 2009.
- [A53] M.J. Fadili, J.-L. Starck, B. Zhang, and S. Digel. Poisson noise removal in multivariate count data. In *SPIE Wavelets XIII*, San Diego, 2009.
- [A54] J. Bobin, Y. Moudden, J.-L. Starck, and M.J. Fadili. Sparsity and morphological diversity for hyperspectral data analysis. In *IEEE ICIP*, Cairo, 2009.
- [A55] A. Waiselle, J.-L. Starck, and M.J. Fadili. Three-dimensional inpainting using sparse representations. In *SPIE Wavelets XIII*, San Diego, 2009.
- [A56] Y. Boursier, L. Jacques, D. Raboud, P. Frossard, M.J. Fadili, and P. Vandergheynst. Tv-regularized generation of planar images from omniscams. In *IEEE ICIP*, Cairo, 2009.
- [A57] J.-L. Starck and M.J. Fadili. An overview of inverse problem regularization using sparsity. In *IEEE ICIP*, Cairo, 2009.
- [A58] L. Jacques, D. K. Hammond, and M. J. Fadili. Dequantizing compressed sensing with non-gaussian constraints. In *IEEE ICIP*, Cairo, 2009.
- [A59] F. Lecellier, S. Jehan-Besson, M.J. Fadili, G. Aubert, and M. Revenu. Optimization of divergences within the exponential family for image segmentation. In *2nd International Conference on Scale Space and Variational Methods in Computer Vision, SSVM*, Voss, 2009.
- [A60] J. Schmitt, J.L. Starck, M.J. Fadili, I. Grenier, and J.M. Casandjian. Poisson denoising on the sphere. In *SPIE Wavelets XIII*, San Diego, 2009.
- [A61] J. Bobin, Y. Moudden, J.-L. Starck, and M.J. Fadili. Sparsity and morphological diversity for multivalued data analysis. In *SPIE Wavelets XIII*, San Diego, 2009.
- [A62] S. Durand, M.J. Fadili, and M. Nikolova. Multiplicative noise cleaning via a variational method involving curvelet coefficients. In *2nd International Conference on Scale Space and Variational Methods in Computer Vision, SSVM*, Voss, 2009.
- [A63] M.J. Fadili and J.-L. Starck. Monotone operator splitting for optimization problems in sparse recovery. In *IEEE ICIP*, Cairo, 2009.
- [A64] F. Lecellier, M.J. Fadili, S. Jehan-Besson, G. Aubert, and M. Revenu. Region-based active contours and sparse representations for texture segmentation. In *ICPR 2008*, Tampa, 2008.
- [A65] F.-X. Dupé, M.J. Fadili, and J.-L. Starck. Image deconvolution under poisson noise using sparse representations and proximal thresholding iteration. In *ICASSP 2008*, Las Vegas, 2008.

- [A66] F.-X. Dupé, M.J. Fadili, and J-L. Starck. Deconvolution of confocal microscopy images using proximal iteration and sparse representations. In *ISBI 2008*, Paris, 2008.
- [A67] B. Zhang, M.J. Fadili, J-L. Starck, and J-C. Olivo-Marin. Multiscale variance-stabilizing transform for mixed-poisson-gaussian processes and its applications in bioimaging. In *IEEE ICIP*, San Antonio, 2007.
- [A68] J. Bobin, Y. Moudden, M.J. Fadili, and J-L. Starck. Morphological diversity and sparsity in blind source separation. In *Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation ICA'07*, London, 2007.
- [A69] M.J. Fadili, J-L. Starck, and L. Boubchir. Morphological diversity and sparse image denoising. In *IEEE ICASSP*, volume I, pages 589–592, Honolulu, 2007.
- [A70] J. Bobin, M.J. Fadili, Y. Moudden, and J-L. Starck. Morphological diversity and sparsity : new insights into multivariate data analysis. In *SPIE Wavelets XII*, San Diego, 2007.
- [A71] I. Khalidov, D. Van De Ville, M.J. Fadili, and M. Unser. Activelets and sparsity : a new way to detect brain activation from fmri data. In *SPIE Wavelets XII*, San Diego, 2007.
- [A72] D. Tschumperlé, Y. Bontolila, J. Martinot, and M.J. Fadili. Fast time-space tracking of smoothly moving fine structures in image sequences. In *IEEE ICIP*, San Antonio, 2007.
- [A73] G. Peyré, M.J. Fadili, and J-L. Starck. Learning adapted dictionaries for geometry and texture separation. In *SPIE Wavelets XII*, San Diego, 2007.
- [A74] M.J. Fadili and J-L. Starck. Sparse representation-based image deconvolution by iterative thresholding. In *Astronomical Data Analysis ADA'06*, Marseille, 2006.
- [A75] P. Abrial, J-L. Starck, Y. Moudden, M.J. Fadili, and J. Delabrouille. Cmb data analysis and sparsity. In *Astronomical Data Analysis ADA'06*, Marseille, 2006.
- [A76] B. Zhang, M.J. Fadili, J-L. Starck, and S.W. Digel. Fast poisson noise removal by biorthogonal haar domain hypothesis testing. In *Astronomical Data Analysis ADA'06*, Marseille, 2006.
- [A77] D. Tschumperlé, M.J. Fadili, and Y. Bontolila. Wire structure pattern extraction and tracking from x-ray images of composite mechanisms. In *CVPR'06 (IEEE Computer Society Conference on)*, New York, 2006.
- [A78] F. Lecellier, S. Jehan-Besson, M.J. Fadili, G. Aubert, and M. Revenu. Statistical region-based active contours with exponential family observations. In *IEEE ICASSP*, volume 2, pages 113–116, Toulouse, 2006.
- [A79] F. Lecellier, S. Jehan-Besson, M.J. Fadili, G. Aubert, M. Revenu, and E. Saloux. Region-based active contours with noise and shape priors. In *IEEE ICIP*, pages 1649–1652, Atlanta, 2006.
- [A80] M.J. Fadili and J-L. Starck. Em algorithm and sparse representations for joint image decomposition and inpainting. In *SIAM International conference on Imaging Science (invited presentation)*, Minneapolis, 2006.
- [A81] B. Zhang, M.J. Fadili, and J-L. Starck. Multi-scale variance stabilizing transform for multi-dimensional poisson count image denoising. In *IEEE ICASSP*, Toulouse, 2006.
- [A82] J. Bobin, Y. Moudden, J-L. Starck, M.J. Fadili, and J. Robin. Sz and cmb reconstruction using generalized morphological component analysis. In *Astronomical Data Analysis ADA'06*, Marseille, 2006.
- [A83] L. Boubchir and M.J. Fadili. Multivariate statistical modeling of images with the curvelet transform. In *Eighth International Conference on Signal Processing and Its Applications - IEEE ISSPA 2005*, pages 747–750, Sydney, 2005.
- [A84] M.J. Fadili and J-L. Starck. Em algorithm for sparse representation-based image inpainting. In *IEEE International Conference on Image Processing ICIP'05*, volume II, pages 61–63, Genoa, 2005.
- [A85] M.J. Fadili and J-L. Starck. Sparse representations and bayesian image inpainting. In *SPARS'05*, volume I, Rennes, 2005.

- [A86] S.W. Digel, B. Zhang, J. Chiang, M.J. Fadili, and J-L. Starck. Application of a multidimensional wavelet denoising algorithm for the detection and characterization of astrophysical sources of gamma rays. In *PHYSTAT 05, Statistical Problems In Particle Physics, Astrophysics And Cosmology*, Oxford, 2005.
- [A87] L. Boubchir and M.J. Fadili. Bayesian denoising based on the map estimation in wavelet-domain using bessel k form prior. In *IEEE International Conference on Image Processing ICIP'05*, volume I, pages 113–116, Genoa, 2005.
- [A88] H. Li, A. Elmoataz, M.J. Fadili, and S. Ruan. Dual front evolution model and its application in medical imaging. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI'2004*, volume 3216, pages 103–110, Saint Malo, 2004.
- [A89] M.J. Fadili and E.T. Bullmore. Penalized partially linear models using orthonormal wavelet bases with an application to fmri time series. In *IEEE International Symposium on Biomedical Imaging (invited paper)*, pages 1171–1174, Arlington VA, 2004.
- [A90] L. Boubchir, M.J. Fadili, and D. Bloyet. Bayesian denoising in the wavelet-domain using an analytical approximate alpha-stable prior. In *ICPR 2004 - 17 th International Conference On Pattern Recognition*, volume 4, pages 889–892, Cambridge, 2004.
- [A91] H. Li, A. Elmoataz, M.J. Fadili, and S. Ruan. A multi-label front propagation approach for object segmentation. In *Proceedings of the 17th International Conference of Pattern Recognition (ICPR 2004)*, Cambridge, 2004.
- [A92] M.J. Fadili. Estimating fmri time series models in the wavelet domain. In *SIAM International conference on Imaging Science (invited presentation)*, Salt Lake city, 2004.
- [A93] M.J. Fadili and E.T. Bullmore. Wavelet-based approaches for multiple hypothesis testing in activation mapping of functional magnetic resonance images of the human brain. In *SPIE-Mathematical Imaging : Wavelet Applications in Signal and Image Processing X (invited paper)*, volume 5207, pages 405–416, San Diego, 2003.
- [A94] M.J. Fadili, J. Mathieu, and M. Desvignes. Bayesian wavelet-based poisson intensity estimation of images using the fisz transformation. In *International Conference on Image and Signal Processing*, volume 1, pages 242–253, Agadir, 2003.
- [A95] H. Li, A. Elmoataz, M.J. Fadili, and S. Ruan. An improved image segmentation method based on level set and mathematical morphology. In *Proceedings of SPIE 3rd International Symposium on Multispectral Image Processing and Pattern Recogn*, volume 5286, pages 851–854, Pekin, 2003.
- [A96] S. Schupp, A. Elmoataz, M.J. Fadili, and D. Bloyet. Fast statistical level sets image segmentation. In *Scale-Space01 : Third International Conference on Scale-Space and Morphology*, pages 380–288, Vancouver, 2001.
- [A97] S. Schupp, A. Elmoataz, M.J. Fadili, and D. Bloyet. Discrete approach for active contours for biomedical applications. In *Scandinavian Conference on Image Analysis (SCIA 01)*, Bergen, 2001.
- [A98] S. Ruan, B. Moretti, M.J. Fadili, and D. Bloyet. Segmentation of magnetic resonance images using fuzzy markov random fields. In *IEEE International Conference on Image Processing*, pages 1051–1054, Thessaloniki, 2001.
- [A99] S. Schupp, A. Elmoataz, M.J. Fadili, and D. Bloyet. Pde based image segmentation for biomedical applications. In *International Conference on Image and Signal Processing*, volume 3, page 204, Agadir, 2001.
- [A100] M.J. Fadili, E.T. Bullmore, and M. Brett. Wavelet methods for characterising mono- and poly-fractal noise structures in shortish time series : An application to functional mri. In *IEEE International Conference on Image Processing*, pages 225–228, Thessaloniki, 2001.
- [A101] S. Ruan, M.J. Fadili, J. Xue, and D. Bloyet. Unsupervised segmentation of 3d mr brain images. In *Int. Conf. Patten Recognition*, volume 3, pages 409–412, Barcelona, 2000.
- [A102] S. Schupp, A. Elmoataz, M.J. Fadili, P. Herlin, and D. Bloyet. Image segmentation via multiple active contour models and fuzzy clustering with biomedical applications. In *ICPR*, volume 1, pages 622–625, Barcelona, 2000.

- [A103] S. Ruan, M.J. Fadili, J. Xue, and D. Bloyet. Brain tissue classification based on a mixel model and markov random fields. In *First Int. Conf. On Image and Graphics*, volume 1, pages 369–372, Tianjing, 2000.
- [A104] M.J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer. Prototype characterization and unsupervised fuzzy clustering of fmri time series. In NeuroImage, editor, *Fifth International Conference on Functional Mapping of the Human Brain*, volume 9, page S30, Dusseldorf, 1999.
- [A105] B. Moretti, S. Ruan, M.J. Fadili, and D. Bloyet. Phantom based segmentation assessment for mri images. In *IEEE Engineering in Medicine and Biology Society*, page 613, Hong Kong, 1998.
- [A106] M.J. Fadili, S. Ruan, and B. Mazoyer. Fiatool : a software for fmri data analysis and visualization. In *Ed. Neuroimage*, volume 7, page S800, Montreal, 1998.
- [A107] M.J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer. Unsupervised fuzzy clustering analysis of fmri series. In *IEEE Engineering in Medicine and Biology Society*, page 696, Hong Kong, 1998.
- [A108] L. V  rard, M.J. Fadili, S. Ruan, and D. Bloyet. 3d mri segmentation of brain structures. In *IEEE Engineering in Medicine and Biology Society*, Amsterdam, 1996.

Conf  rences nationales

- [A109] F. Lecellier, S. Jehan-Besson, M.J. Fadili, G. Aubert, M. Revenu, and E. Saloux. Les contours actifs bas  s r  gion avec a priori de bruit et de forme. In *RITS 2009*, 2009.
- [A110] C. Dossal, G. Peyr  , and M.J. Fadili. Une exploration num  rique des performances de l’  chantillonnage compress  . In *22nd GRETSI Symposium on Signal and Image Processing*, Dijon, 2009.
- [A111] G. Peyr   and M.J. Fadili. Algorithmes de premier ordre pour la projection sur une contrainte de variation totale. In *22nd GRETSI Symposium on Signal and Image Processing*, Dijon, 2009.
- [A112] G. Peyr  , M.J. Fadili, and J-L. Starck. Apprentissage de dictionnaires parcimonieux adapt  s pour la s  paration d’images. In *21st GRETSI Symposium on Signal and Image Processing*, Troyes, 2007.
- [A113] L. Boubchir and M.J. Fadili. Mod  lisation statistique multivari  e des images dans le domaine de la transform  e de curvelet. In *20th GRETSI Symposium on Signal and Image Processing*, pages 233–236, Louvain-la-Neuve, Belgique, 2005.
- [A114] L. Boubchir, M.J. Fadili, and D. Bloyet. Le processus alpha-stable pour l’estimation bay  sienne non param  trique des images dans le domaine des ondelettes. In *Journ  es d’  tude sur les m  thodes pour les signaux complexes en traitement d’image 2003*, pages 105–118, INRIA-Rocquencourt, 2003.
- [A115] M. Desvignes, M.J. Fadili, and F. Bataille. Mesure de nettet   par transform  e en ondelettes. d  finition et comparaison pour l’autofocus de cam  ra. In *19th GRETSI Symposium on Signal and Image Processing*, volume II, pages 271–274, Paris, 2003.
- [A116] M.J. Fadili, J. Mathieu, and M. Desvignes. La transformation de fiz pour l’estimation d’images d’intensit   poissonnienne dans le domaine des ondelettes. In *19th GRETSI Symposium on Signal and Image Processing*, volume I, pages 317–320, Paris, 2003.
- [A117] M.J. Fadili and E.T. Bullmore. Un mod  le de r  gression lin  aire g  n  ralis   dans le domaine des ondelettes pour des erreurs de mesure en 1/f. In *18e colloque GRETSI*, pages 405–408, Toulouse, 2001.
- [A118] C. Dolabdjian, M.J. Fadili, and E.H. Leyva. Comparaison d’un filtrage num  rique    l’aide de la transform  e d’ondelettes discr  te, mis en oeuvre en temps r  el sur un dsp,    un filtrage classique de type p-b. In *Congr   Interdisciplinaire en Instrumentation*, volume 2, pages 397–404, Paris, 2001.
- [A119] C. Jaggi, S. Ruan, M.J. Fadili, and D. Bloyet. Approche markovienne pour la segmentation 3d des tissus c  r  braux en irm. In *16e Colloque GRETSI*, pages 327–330, Grenoble, 1997.
- [A120] M.J. Fadili, S. Ruan, L. V  rard, C. Jaggi, and D. Bloyet. Extraction et partitionnement de l’enc  phale par un mod  le de contours actifs. In *16e Colloque GRETSI*, pages 439–442, Grenoble, 1997.

Revues internationales soumises ou révisées

- [S1] M.J. Fadili and J.-L. Starck. Monotone operator splitting for fast sparse solutions of inverse problems. *SIAM Journal on Imaging Sciences*, 2009. submitted.
- [S2] M.J. Fadili and G. Peyré. Total variation projection with first order schemes. *IEEE Transactions on Image Processing*, 2009. revised.
- [S3] G. Peyré, M.J. Fadili, and J.-L. Starck. Learning the morphological diversity. *SIAM Journal on Imaging Sciences*, 2009. revised.
- [S4] L. Jacques, D. K. Hammond, and M.J. Fadili. Dequantizing compressed sensing : When oversampling and non-gaussian constraints combine. *IEEE Trans. on Information Theory*, 2009. submitted.
- [S5] C. Chesneau and M.J. Fadili. Stein block thresholding for wavelet-based image deconvolution. *Electronic Journal of Statistics*, 2009. revised.
- [S6] C. Chesneau and M.J. Fadili. A note on adaptive wavelet estimation in a shifted curves model via block thresholding. *Mathematical Methods of Statistics*, 2009. submitted.
- [S7] I. Khalidov, D. Van De Ville, M.J. Fadili, F. Lazeyras, and M. Unser. Activelets : Wavelets for sparse representation of hemodynamic responses. *Signal Processing*, 2010. submitted.

Rapports techniques

- [RR1] M.J. Fadili, Bo Zhang, and J.-L. Starck. Intensity estimation based on the wavelet domain hypothesis testing. Technical report, GREYC CNRS-ENSICAEN-Univ. de Caen, France, 25-26 Nov. 2004.
- [RR2] J.J. Collet and M.J. Fadili. Simulation of gegenbauer processes using wavelet packets. Technical report, GREYC CNRS-ENSICAEN-Univ. de Caen (France) and Queensland University of Technology (Australia), 2005.
- [RR3] F. Lecellier, M.J. Fadili, S. Jehan-Besson, G. Aubert, M. Revenu, and E. Saloux. Region-based active contours with exponential family observations. Technical Report 2, GREYC, 2008.