



HAL
open science

Apports méthodologiques basés sur la parcimonie pour les problèmes inverses

Aurélia Fraysse

► **To cite this version:**

Aurélia Fraysse. Apports méthodologiques basés sur la parcimonie pour les problèmes inverses. Mathématiques [math]. Université Paris Sud, 2017. tel-01658054

HAL Id: tel-01658054

<https://hal.science/tel-01658054>

Submitted on 7 Dec 2017

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 PARIS-SUD

Laboratoire des Signaux et Systèmes

Spécialité : Traitement du signal et des images

HABILITATION A DIRIGER DES RECHERCHES

soutenue le 07/11/2017

par

Aurélia FRAYSSE

<p>APPORTS MÉTHODOLOGIQUES BASÉS SUR LA PARCIMONIE POUR LES PROBLÈMES INVERSES</p>

Composition du jury :

Rapporteurs :

Laure BLANC-FERRAUD	Directeur de recherche CNRS
Jérôme IDIER	Directeur de Recherche CNRS
Bruno TORRESANI	Professeur de l'Université Aix-Marseille

Examineurs :

Alain ABERGEL	Professeur de l'Université Paris Sud
Jean-Francois AUJOL	Professeur de l'Université de Bordeaux
Rémi GRIBONVAL	Directeur de recherche INRIA

Table des matières

I Parcours professionnel

1	Curriculum Vitae	9
----------	-------------------------------	----------

II Publications

2	Publications	17
2.1	Articles de revue	17
2.2	Articles de conférences internationales avec comité de lecture	18
2.3	Articles de conférences nationales avec comité de lecture	19
2.4	Chapitres de livre	19

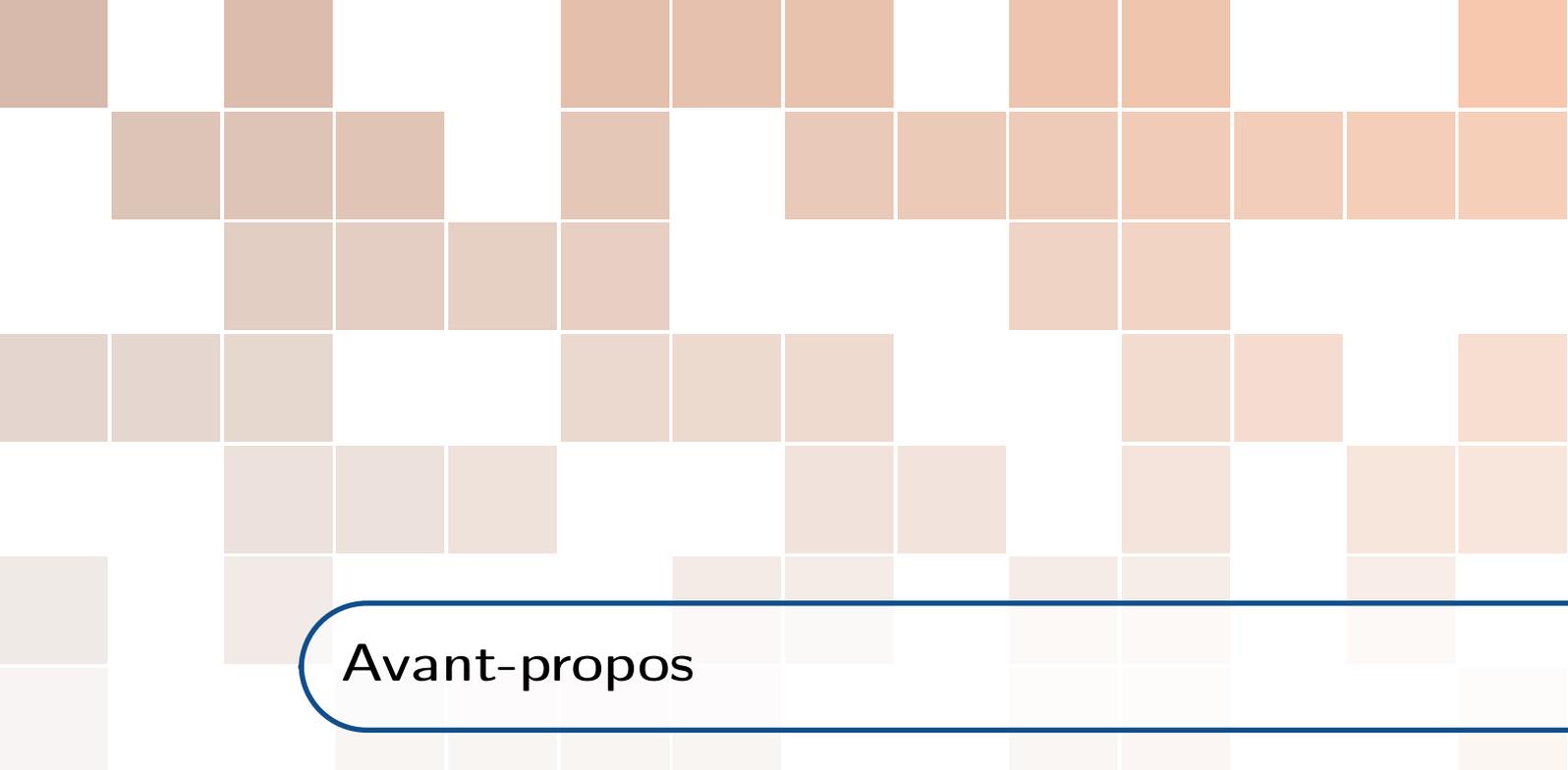
III Synthèse des travaux de recherche

3	Codage et compression	25
3.1	Performances débit-distorsion d'un quantificateur scalaire uniforme	25
3.2	Allocation de débits	31
3.2.1	Principe de l'algorithme proposé	32
3.3	Codage avec information adjacente	34
3.3.1	Modèle de mélange scalaire de gaussiennes	35
3.3.2	Fonctions débit-distorsion	36
3.3.3	Principaux résultats obtenus	36
3.4	Conclusion	37
4	Estimation et problèmes inverses	39
4.1	Problèmes inverses	39
4.1.1	Modèle direct	40
4.1.2	Problème inverse	40

4.2	Méthodes pénalisées en problèmes inverses	41
4.3	Ondelettes	42
4.4	Estimation non paramétrique	43
4.4.1	Principaux résultats obtenus	46
4.5	Détection d'ondes gravitationnelles	47
4.6	Problèmes inverses des ondes	49
4.6.1	Inversion en deux étapes	51
4.7	Application au JWST	52
4.8	Conclusion	54
5	Méthodologie du bayésien variationnel	55
5.1	Inférence bayésienne	56
5.1.1	Loi <i>a priori</i>	57
5.2	Le Bayésien variationnel	58
5.2.1	Méthode classique	58
5.2.2	Principe	58
5.2.3	L'algorithme	59
5.2.4	Choix de la séparabilité	60
5.3	Bayésien variationnel de type gradient	61
5.4	Bayésien variationnel de type gradient à mémoire	65
5.4.1	Sous-espaces considérés	67
5.5	Application aux problèmes inverses	68
5.5.1	Loi <i>a posteriori</i>	68
5.5.2	Exemple	74
5.6	Application à la reconstruction d'images par ondelettes	75
5.6.1	Modèle utilisé	75
5.6.2	Résultats	78
5.7	Conclusion	79

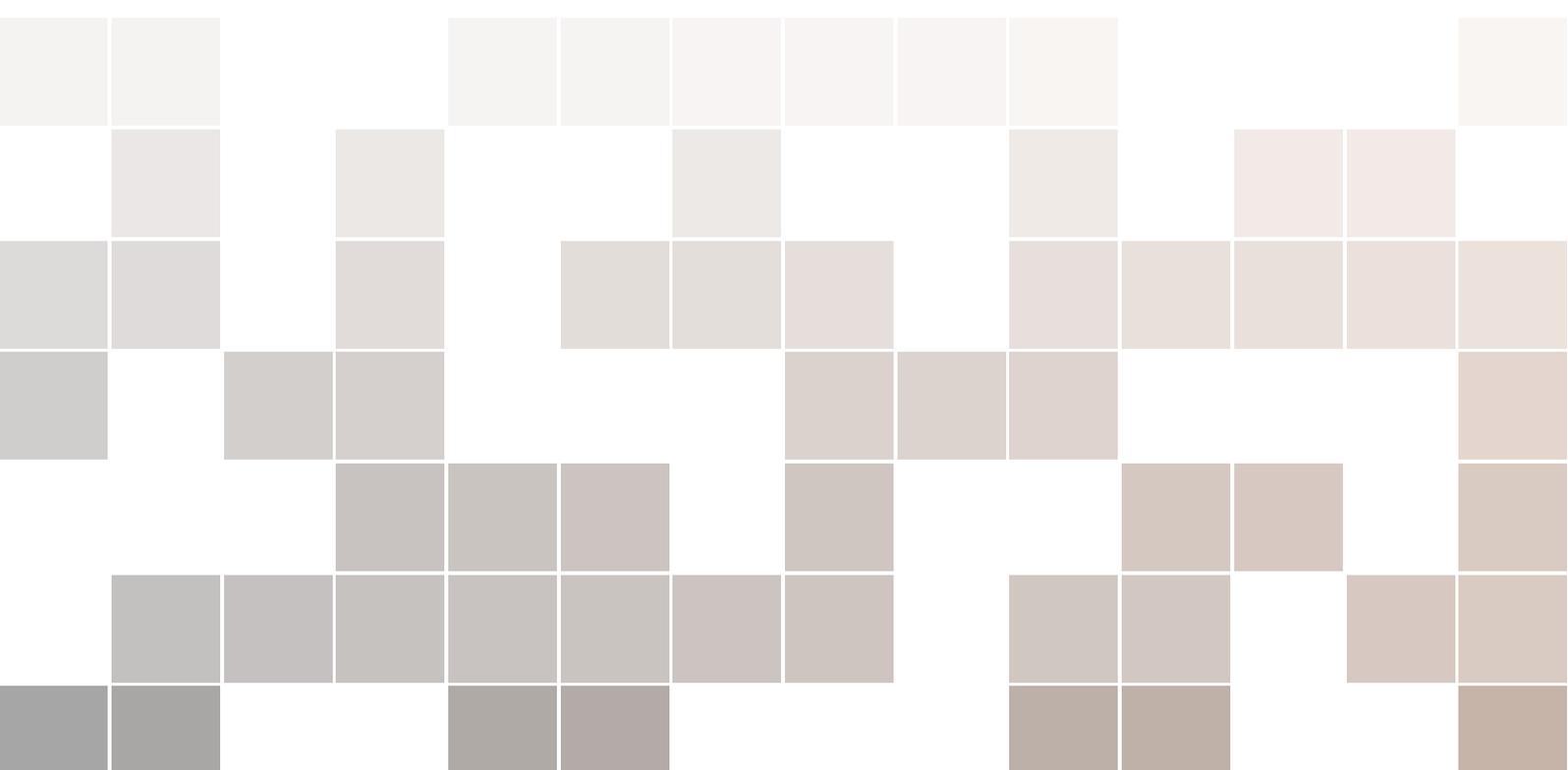
IV Projet de recherche

6	Projet de recherche	83
6.1	Aspects méthodologiques	83
6.1.1	Comportement générique des estimateurs non paramétriques	83
6.1.2	Généralisation du bayésien variationnel exponentialisé	84
6.1.3	Bayésien variationnel non-paramétrique	85
6.2	Aspects applicatifs	85
6.2.1	Domaine de l'astrophysique	86
6.2.2	Domaine de l'imagerie bio-médicale	86



Avant-propos

Ce manuscrit décrit mon activité de recherche retraçant mon post-doctorat et mes années à l'Université Paris-Sud. Dans un premier temps je présente un *curriculum vitae* détaillant des éléments factuels de mon activité. Dans une deuxième partie je précise les différents résultats obtenus au cours de ma carrière et les pistes que j'envisage d'explorer par la suite. Une partie de mon activité, plus proche de mes thématiques de thèse, n'est pas présentée explicitement dans cette partie mais elle l'est dans l'annexe. En effet celle-ci présente des articles significatifs de mon activité de recherche.



Première partie

Parcours professionnel

Curriculum Vitae

Postes Occupés

2008-	Maître de conférences à l'IUT de Cachan, Université Paris-Sud.
2006-2008	Post-doctorant à l'Ecole Nationale Supérieure des Télécommunications (Telecom ParisTech)
2004-2006	ATER à temps partiel à l'Université Paris XII, Créteil.
2001-2004	Allocataire Moniteur à l'Université Paris XII, Créteil.

Cursus Universitaire

2001-2005	Thèse de Mathématiques Appliquées sous la direction de Stéphane Jaffard au sein du Laboratoire d'Analyse et de Mathématiques Appliquées de l'Université Paris XII, Créteil. Titre : <i>Résultats de généricité en analyse multifractale.</i> Soutenue à Créteil le 14 Décembre 2005 devant le jury composé de : M. Julien Barral, CR INRIA, Rocquencourt, Rapporteur M. Yanick Heurteaux, Professeur des Universités, Clermont Ferrand, Examineur M. Stéphane Jaffard, Professeur des Universités, Paris 12, Directeur de thèse M. Jean-Pierre Kahane, Professeur des Universités, Paris 11, Président du jury M. Yves Meyer, Professeur des Universités, ENS Cachan, Examineur M. Claude Tricot, Professeur des Universités, Clermont Ferrand, Rapporteur
2000-2001	DEA d'Analyse et Systèmes aléatoires Université de Marne La Vallée, mention bien.
1999-2000	Maîtrise de mathématiques de l'Université Paris XII, Créteil, mention bien.
1998-1999	Licence de mathématiques de l'Université Paris XII, Créteil, mention bien.
1996-1998	DEUG MIAS de l'Université Paris XII, Créteil, mention bien.

Production scientifique

- 11 articles dans des revues internationales.
- 12 articles dans des conférences internationales.
- 6 articles dans des conférences nationales (GRETSI).
- 1 chapitre de livre.

Encadrements

— Thèses :

- **Sept.2011- Dec. 2014** : Encadrement à 50% de la thèse de Yuling Zheng (co-encadrement avec T. Rodet, SATIE) :
 - Thèse financée par une bourse ministère.
 - Soutenue le 4 Décembre 2014.
 - Sujet de thèse : Algorithmes bayésiens variationnels accélérés et application aux problèmes inverses de grande taille.
 - 5 articles de conférences, 1 article de revue.
 - Etudiante actuellement ingénieure à IBM Chine.
- **Sept.2014- ?** : Encadrement à 30% de la thèse d’Hidayet Zaimaga (co-encadrement avec M. Lambert, GEEPS et A. Massa, L2S) :
 - Thèse financée par une bourse DIGITEO liée à la chaire SIRENA.
 - Sujet de thèse : Bayesian inversion and sparsity in the framework of wavefield inverse problems.
 - 2 articles de conférences.
- **Sept.2015- ?** : Encadrement à 50% de la thèse de Amine Hadj-Youcef (co-encadrement avec F. Orioux, L2S et A. Abergel, IAS) :
 - Thèse financée par une bourse ministère.
 - Sujet de thèse : Approches convexes non supervisées pour l’imagerie haute résolution en astronomie.
 - 1 article de conférence.

— Stages :

- **Fev.2010-Sept.2010** : Encadrement du stage de M2R de Ali Komaty (co-encadrement avec C. Delpha, L2S) :
 - Stage de M2R SC (Signaux et Circuits) de l’Université de Bretagne Occidentale.
 - Soutenue le 4 Décembre 2014.
 - Sujet du stage : Schéma de Costa à structure fractale pour le watermarking.
 - 2 articles de conférences.
- **2. Mai 2011-Sept.2011** : Encadrement du stage de M2R de Leïla Gharsalli :
 - Stage de M2R MVA (Mathématiques Vision Apprentissage) de l’ENS Cachan.
 - Sujet du stage : Méthodes d’optimisation dans le cadre Bayésien
 - 1 article de conférence.
- **Mai 2013 - Juil. 2013** : Encadrement du stage de M1 de Kévin Malleron :
 - Stage de l’ENS Cachan.
 - Sujet du stage : Réglage des paramètres dans un seuillage itératif.
- **Janv. 2015 - Juil. 2015** : Encadrement du stage de Bachelor de Matthew Tivnan (co-encadrement avec D. Lesselier, L2S et M. Lambert, GeePS) :
 - Stage de Northeastern University à Boston.
 - Sujet du stage : Fusion de données pour l’imagerie micro ondes.
- **Juin 2015 - Aout 2015** : Encadrement du stage de l’ENSEIHT de Dylan Fagot (co-encadrement avec F. Orioux, L2S) :
 - Stage de 2eme année ingénieur de l’ENSEIHT.
 - Sujet du stage : Méthodes de parcimonie pour l’imagerie stellaire.
- **Mai 2016 - Septembre 2016** : Encadrement du stage de M2 de Amal Fethi (co-encadrement avec E. Chassande-Mottin, APC et S. Jaffard, LAMA) :
 - Stage du M2 FESup de l’ENS Cachan.

- Sujet du stage : Choix de l'ondelette dans l'algorithme Coherent WaveBurst pour la détection d'ondes gravitationnelles.
- **Avril 2017 - Septembre 2017** : Encadrement du stage de M2R de Quentin Bammey (co-encadrement avec E. Chassande-Mottin, APC et S. Jaffard, LAMA) :
 - Stage du M2R MVA de l'ENS Cachan.
 - Sujet du stage : Améliorations Matching poursuit pour les ondes gravitationnelles.

Activités d'organisation

- Membre du comité d'organisation de MaxEnt en 2014.
- Membre du comité d'organisation du colloque pédagogique GEII à l'IUT de Cachan en 2009.
- Responsable du séminaire Multimédia à Telecom Paris Tech, en 2006-2008.
- Responsable du séminaire des étudiants du Laboratoire d'Analyse et de Mathématiques Appliquées de l'Université Paris XII, pour les années 2003-2004 et 2004-2005.

Participations à des comités de recrutements

- En tant que membre extérieur :
 - deux comités de recrutement en section 61, à l'IUT de Ville d'Avray et à l'Université de Nantes en 2016.
 - un comité de recrutement MCF en section 61 à l'Université de Marne La Vallée en 2011.
- En tant que membre interne :
 - membre un comité de recrutement MCF en section 61 à l'Université Paris-Sud en 2015.
 - présidente d'une commission de recrutement PRAG en mathématiques à l'IUT de Cachan en 2014.
 - membre de 3 comités de recrutement MCF en section 61 à l'Université Paris-Sud en 2014. Membre d'un comité de recrutement MCF en section 61 à l'Université Paris-Sud en 2013.
 - vice présidente d'une commission de recrutement de 2 PRAG en génie électrique à l'IUT de Cachan en 2012.
 - membre de 3 comités de recrutement PRAG en mathématiques à l'IUT de Cachan de 2010 à 2012.

Responsabilités administratives

- Responsable du Groupe Problèmes Inverses du L2S depuis Septembre 2013, groupe composé actuellement de 14 permanents, 14 doctorants et 4 post-doctorants.
- Responsable de la Poursuite d'études du département GEII2 de l'IUT de Cachan de Septembre 2014 à Septembre 2016.
- Assesseur B de la sous-CCSU 61 depuis Décembre 2014.
- Responsable des mathématiques pour le département GEII2 de 2008 à 2012.
- Membre du jury VAE de l'IUT de Cachan de 2010 à 2014.

Mandats électifs

- Membre élu du conseil de laboratoire de 2013 à 2015.
- Membre élu de la CCSU 60-61-62 de l'Université Paris-Sud depuis 2013.
- Membre élu du Conseil d'IUT depuis 2015.
- Membre élu du CNU 61 depuis 2015.
- Membre du CHSCT spécial de l'IUT de Cachan de 2010 à 2017.

Travail d'expertise

- Membre du jury du prix de thèse EEA/ISIS/GRETSI depuis 2017.
- Rapporteur d'articles pour IEEE Signal Processing (4 rapports).
- Rapporteur d'articles pour IEEE Image Processing (3 rapport).
- Rapporteur d'articles pour Elsevier Signal Processing (5 rapports).
- Rapporteur d'articles pour des conférences internationales : ICASSP 2017, SSP 2016, Eusipco 2016, ICASSP 2016, ICASSP 2015, ICASSP 2014, ICASSP 2013, SSP 2011, Eusipco 2011, ICIP 2012.
- Rapporteur de livre pour Wiley (1 rapport).

Contrats de recherche

- Participation au projet européen MUSCLE de 2006 à 2008.
- Membre du projet MAORI (Mathématiques de l'Optimisation pour l'Imagerie) financé par le Programme Gaspard Monge pour l'Optimisation de 2013 à 2015.
- Porteur local du projet ANR Wavegraph démarré en Septembre 2015.

Rayonnement international

- Responsable scientifique français d'un projet franco chilien (CSCS-L2S) sur le radiotélescope 21cm.
- Membre du projet de coopération "échange enseignement-recherche" avec la Northeastern University à Boston.
- Collaboration avec l'université de Trente, Italie (Invitation de Paolo Rocca pour un mois en Janvier 2015).

Activités d'enseignement

Mon expérience d'enseignante a commencé en 1998 avec le tutorat pédagogique que j'ai effectué à l'université Paris 12. Cette expérience s'est ensuite enrichie par trois ans de monitorat et deux années d'ATER.

Depuis mon recrutement comme Maître de Conférences en 2008, mon enseignement s’effectue essentiellement à l’IUT de Cachan, au sein du département GEII 2. Dans ce cadre j’effectue entre 250 HETD et 300 HETD par an. Je suis aussi intervenue à l’école d’ingénieurs Polytech Paris-Sud (PPS) ainsi qu’en Master 2 ATSI (Automatique et Traitement du Signal et des Images).

En ce qui concerne mon enseignement à l’IUT de Cachan, il concerne essentiellement les mathématiques en première et deuxième année de DUT. Le contenu de ces enseignements part des bases telles que la trigonométrie et les bases de l’analyse en première année de DUT. Les cours dispensés en deuxième année sont plus axés sur les notions liées au traitement du signal et au génie électrique, telles que la transformée de Fourier ou la transformée en Z , ainsi que l’optimisation. J’étais aussi, depuis mon recrutement jusqu’en 2012, responsable du module de mathématiques et à ce titre je gérais les vacataires, le contrôle des connaissances ainsi que le contenu des enseignements et les différents photocopiés de ce cours. Chaque promotion est composée d’environ 230 étudiants répartis en 5 groupes. Depuis Septembre 2016, je fais aussi les Cours/TD de Traitement Numérique du Signal pour la Licence Professionnelle MECSE (Métiers de l’Electronique : Communication et Systèmes Embarqués) qui concerne environ 20 étudiants.

Ces enseignements sont dispensés à un public venant en majorité de bac STI ou STI2D depuis 2013. Ce public disparate et peu attiré par l’abstraction en général et les mathématiques en particulier m’a amenée à tenter différentes techniques pour arriver à ce qu’ils aient au moins des bases de calcul. Dans ce cadre, j’ai essentiellement travaillé avec l’outil WIMS (WWW Interactive Multipurpose Server) qui permet en particulier de créer des feuilles d’exercices et d’évaluations dont les données numériques sont prises aléatoirement. J’ai ainsi utilisé cet outil pour reprendre les notions de base non vues en cours mais aussi comme complément des photocopiés, pour manipuler de façon plus pragmatique certains concepts enseignés.

Les enseignements que j’ai dispensés à Polytech Paris-Sud de 2010 à 2013 sont des Cours-TP sur les notions de base de Traitement d’images sous Matlab destinés à des élèves ingénieurs en apprentissage de cinquième année de formation. J’ai aussi monté et enseigné des cours d’introduction à Matlab en lien avec différents aspects des mathématiques, tels que l’analyse, l’algèbre linéaire ou les probabilités, pour les étudiants en 3eme année. J’étais aussi, entre 2011 et 2015, responsable du module “Ondelettes et représentations parcimonieuses” du M2R ATSI de la Faculté des Sciences d’Orsay.

Le détail des principales activités est présenté dans le tableau ci-après, auxquelles s’ajoutent les charges pédagogiques de jury, d’encadrement de stages ou d’apprentis.

	Enseignement	Niveau	Période	Cours/an	TD/an	TP/an	H éq. TD/
Université Paris-Sud	Traitement Numérique du Signal	LPro	2016/2017 → ...	9	16		16
	Ondelettes et parcimonie	M2	2009/2010 → ...		13.5		
	Traitement d’images	M2 PPS 5	2010/2011 → 2012/2013		25		
	Introduction à Matlab	L3 PPS 3	2013/2014		20		
	Bases de Mathématiques	L1 DUT 1	2008/2009 → ...		190	6	196
	Mathématiques pour le signal	L2 DUT 2	2008/2009 → ...		64		64
	Préparation aux concours	L2 DUT 2	2014/2015 → ...		20		20
Université Paris 12	Analyse	L1-L2	2001/2002 → 2004/2005	12	32		32
	Algèbre linéaire	L1-L2	2001/2002 → 2005/2006		32		32
	Optimisation	L1	2005/2006		18		
	Optimisation dynamique	L3	2004 /2005		20		20
	Probabilités	L2	2005 /2006		24		24
ingénieur IMAC	Mathématiques pour l’ingénieur	L3	2007/2008		36		36

Ces enseignements m’ont amenée à produire les supports suivants :

- 4 photocopiés de TD et 1 photocopiés de cours de mathématiques en DUT par an.
- 1 photocopié de TD pour le Cours/TD de TNS en Licence Professionnelle.

- 1 polycopié de TP pour le traitement d'images.
- 3 polycopiés de TP pour l'introduction à Matlab.
- 1 polycopié de cours "Ondelettes et représentations parcimonieuses".

En complément de cette activité d'enseignement devant étudiants je me suis aussi investie dans la représentation de l'IUT, participant en particulier aux journées portes ouvertes et aux différents salons, salon de l'étudiant, de l'éducation et APB. J'ai aussi monté un forum Poursuite d'Etudes au sein de l'IUT.

Deuxième partie

Publications

- 2.1 Articles de revue
- 2.2 Articles de conférences internationales avec comité de lecture
- 2.3 Articles de conférences nationales avec comité de lecture
- 2.4 Chapitres de livre

2 — Publications

2.1 Articles de revue

- [J1] Y. Zheng, **A. Fraysse**, T. Rodet, *Efficient Variational Bayesian Approximation Method Based on Subspace Optimization*, "IEEE Trans. on Image Processing", Vol. 24 (2015), pages 681-693.
- [J2] F. Bassi, **A. Fraysse**, E. Dupraz, M. Kieffer, *Rate-distortion bounds for Wyner-Ziv coding with Gaussian Scale Mixture correlation noise*, "IEEE Trans. on Information Theory ", Vol. 20 (2014), pages 7540-7546.
- [J3] **A. Fraysse**, T. Rodet, *A measure-theoretic variational Bayesian algorithm for large dimensional problems*, "SIAM Journal in Imaging Science", vol. 7 (2014), pages 2591-2622.
- [J4] M. Kaaniche, **A. Fraysse**, B. Pesquet-Popescu, J.-C. Pesquet, *A Bit Allocation Method for Sparse Source Coding*, "IEEE Trans. on Image Processing", Vol. 23 (2014), pages 137-152.
- [J5] **A. Fraysse**, *Why minimax is not that pessimistic*, "ESAIM PS", Vol.17 (2013), pages 472-484.
- [J6] **A. Fraysse**, *Regularity criteria of almost every function in a Sobolev space*, "Journal of functional analysis", Vol. 258 (2010), pages 1806-1821.
- [J7] **A. Fraysse**, B. Pesquet-Popescu, J.-C. Pesquet, *On the uniform quantization of a class of sparse sources*, "IEEE Trans. on Information Theory ", Vol. 55 (2009), pages 3243-3263.
- [J8] **A. Fraysse**, *Generic validity of the multifractal formalism*, "SIAM J. Math. Anal.", Vol. 37 (2007), pages 593-607.
- [J9] **A. Fraysse**, S. Jaffard, *How smooth is almost every function in a Sobolev space ?*, "Rev. Math. Iber. Amer.", Vol 22 (2006), pages 663-682.

- [J10] **A. Fraysse**, S. Jaffard, *The Sobolev embeddings are usually sharp*, "Abstract and Applied Analysis", Vol. 4 (2005), pages 437-448.
- [J11] **A. Fraysse**, S. Jaffard, J.-P. Kahane, *Quelques propriétés génériques en analyse*, "C. R. Acad. Sci. Paris, Ser. I" Vol. 340 (2005), pages 645-651.

2.2 Articles de conférences internationales avec comité de lecture

- [C1] H. Zaimaga, **A. Fraysse**, M. Lambert, *Sparse Reconstruction Algorithms for Non-linear Microwave Imaging*, Proceedings EUSIPCO 2017, to appear.
- [C2] A. Hadj-Youcef, F. Orieux, **A. Fraysse**, A. Abergel, *Restoration from Multispectral Blurred Data with Non-Stationary Instrument Response*, Proceedings EUSIPCO 2017, to appear.
- [C3] H. Zaimaga, **A. Fraysse**, M. Lambert, *Nonlinear imaging of 3D defect in anisotropic laminate using joint sparsity constraints*, Proceedings ISEM 2017, to appear.
- [C4] Y. Zheng, **A. Fraysse**, T. Rodet, *Wavelet based unsupervised variational Bayesian image reconstruction approach*, Proceedings EUSIPCO 2015, pages 2212-2216 (2015).
- [C5] M. Kaaniche, **A. Fraysse**, B. Pesquet-Popescu, J.-C. Pesquet, *Accurate rate-distortion approximation for sparse Bernoulli-Generalized Gaussian models*, Proceedings ICASSP 2014, pages 2020-2024 (2014).
- [C6] Y. Zheng, **A. Fraysse**, T. Rodet, *A subspace-based variational Bayesian method*, Proceedings ICASSP 2013, vol. p.id 4489 (2013)
- [C7] Y. Zheng, **A. Fraysse**, T. Rodet, *Fast variational Bayesian approaches applied to large dimensional problems*, Proceedings ICIP 2013, vol. p.id 1240 (2013).
- [C8] M. Kaaniche, **A. Fraysse**, B. Pesquet-Popescu, J.-C. Pesquet, *A Convex Programming Bit Allocation Method for Sparse Sources*, PCS2012, (2012).
- [C9] L. Gharsalli, A. Djafari, **A. Fraysse**, T. Rodet, *Variation Bayesian Approximation with scale mixture priors : A comparison between three algorithms*; Proceedings MaxEnt 12 (2012).
- [C10] A. Komaty, C. Delpha, **A. Fraysse**, *Floating Costa Scheme with Fractal Structure for Information Embedding*, IEEE International Conference on Telecommunications (ICT 2012), (2012).
- [C11] **A. Fraysse**, T. Rodet, *A gradient-like variational Bayesian algorithm*, Proceedings SSP 11 (2011).
- [C12] **A. Fraysse**, B. Pesquet-Popescu, J.-C. Pesquet, *Rate-distortion results for Generalized Gaussian distributions*, Proceedings ICASSP 08 (2008).

2.3 Articles de conférences nationales avec comité de lecture

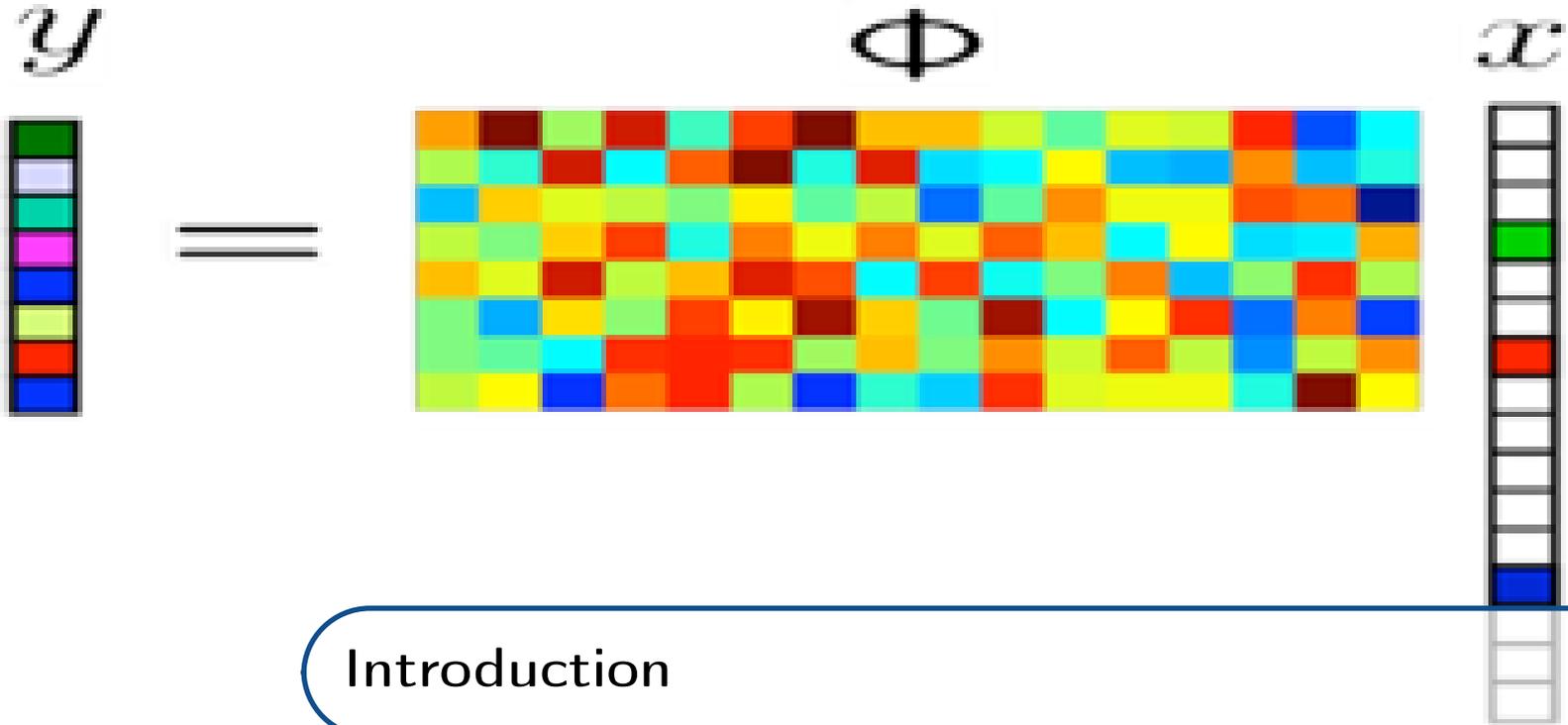
- [CN1] Y. Zheng, T. Rodet, **A. Fraysse**, *Approche bayésienne variationnelle non-supervisée basée sur les ondelettes*, Proceedings GRETSI 15 (2015).
- [CN2] Y. Zheng, T. Rodet, **A. Fraysse**, *Approches rapides du bayésien variationnel pour des problèmes de grande dimension*, Proceedings GRETSI 13 (2013).
- [CN3] A. Komaty, C. Delpha, **A. Fraysse**, *Tatouage Numérique par Schéma de Costa à Structure Fractale Flottante*, Proceedings GRETSI 11 (2011).
- [CN4] **A. Fraysse**, T. Rodet, *Sur un nouvel algorithme Bayésien variationnel*, Proceedings GRETSI 11 (2011).
- [CN5] **A. Fraysse**, B. Pesquet-Popescu, J.-C. Pesquet, *Une méthode d'allocation de débit basse-résolution pour des données parcimonieuses*, Proceedings GRETSI 09 (2009).
- [CN6] T. Petrisor, **A. Fraysse**, B. Pesquet-Popescu, J.-C. Pesquet, *Une nouvelle approche du codage par descriptions multiples, utilisant des représentations linéaires creuses*, Proceedings GRETSI 07 (2007).

2.4 Chapitres de livre

- [O1] T. Rodet, **A. Fraysse**, H. Ayasso, *Méthodes d'inversion appliquées au traitement du signal et de l'image, chapitre Approche bayésienne variationnelle et bi-modèle pour la reconstruction- séparation de composantes astrophysiques*, pp. 253-278, Hermès, 2012.

Troisième partie

Synthèse des travaux de recherche



Introduction

L'utilisation des ondelettes en traitement du signal a connu un essor important dans les années 90. Un des intérêts majeurs de cet outil est de permettre une représentation de signaux, par exemple les images naturelles, en peu de coefficients non nuls, autrement dit parcimonieuse. Cet intérêt s'est encore accru avec l'apparition de l'échantillonnage compressé [Don06, CRT06]. Cette théorie permet en effet de reconstruire un bon nombre de signaux à partir d'un jeu réduit d'échantillons, du moment qu'ils ont une représentation qui requiert peu de coefficients. Un autre intérêt des ondelettes est porté par l'estimation non paramétrique et le lien entre la représentation des signaux naturels et des espaces fonctionnels donnés. Même si dans les faits les signaux ont tous une représentation finie, pouvoir les considérer comme des objets de dimension infinie offre plus de flexibilité et permet de s'affranchir de contraintes telles que le choix de l'échantillonnage ou du nombre de paramètres à considérer.

Depuis le début de ma carrière post-thèse j'ai étudié trois grands axes thématiques du traitement du signal présentés dans les trois chapitres de ce mémoire. Le premier présenté ici porte sur la théorie de l'information pour le codage de données parcimonieuses. En effet, la parcimonie joue un rôle important à la fois pour réduire la quantité d'information utile dans un schéma de compression mais aussi pour modéliser éventuellement les effets des pertes de paquets dans les canaux de transmission. Nous avons ainsi étudié les performances débit-distorsion dans différents schémas de compression en nous appuyant sur des modèles de parcimonie. Cela a en particulier permis de construire un schéma de compression simple pour les coefficients d'ondelette des images naturelles. Les résultats montrés s'appuient aussi sur l'utilisation de méthodes d'optimisation en dimension infinie. Le second axe présenté est centré sur les problèmes inverses et découpé en deux parties. Après une présentation générale des problèmes inverses, je vais me concentrer

sur les méthodes de reconstruction. Dans un premier temps je vais présenter les travaux basés sur les méthodes déterministes de reconstruction, en particulier des méthodes linéaires et par seuillage. Dans ce cadre je présenterai à la fois des résultats théoriques d'étude de performances d'estimateurs dans un cadre non paramétriques ainsi que des applications pratiques de ces deux types d'estimateurs. La dernière partie de ce mémoire traite du développement de méthodes bayésiennes variationnelles. C'est celle que je développerai plus ici puisque c'est en grande partie sur elle que se basent mes perspectives de recherche. Le développement de ces méthodes s'appuie sur l'optimisation dans des ensembles de mesures, donc de dimension infinie. Je présente aussi leur mise en œuvre pratique pour des données parcimonieuses et en particulier pour la restauration des coefficients d'ondelettes des images naturelles.

Comme on le verra tout au long de ce mémoire, la prise en compte, sous différents aspects, de l'information de parcimonie m'a permis de contribuer à des domaines très différents du traitement du signal. Un autre apport de mon travail est l'utilisation de méthodes d'optimisation ou d'estimation en dimension infinie. Cette approche permet en effet de relaxer les contraintes liées au modèle et fournit ainsi des résultats plus généraux et plus facilement adaptables.

- 3.1 Performances débit-distorsion d'un quantificateur scalaire uniforme
- 3.2 Allocation de débits
- 3.3 Codage avec information adjacente
- 3.4 Conclusion

3 — Codage et compression

Une partie de mon activité de recherche, développée plus particulièrement entre 2006 et 2010, est la compression pour des données parcimonieuses. Ce travail s'articule tout d'abord sur les effets de la quantification scalaire uniforme dans [FPPP09] et son utilisation pour l'allocation de débits valable à la fois à haute et basse résolution dans [KFPPP14]. Je me suis aussi intéressée au codage avec information adjacente, [BFDK14], toujours sous hypothèse de parcimonie.

3.1 Performances débit-distorsion d'un quantificateur scalaire uniforme

En collaboration avec Béatrice Pesquet-Popescu de Telecom Paristech et Jean-Christophe Pesquet de l'Université Paris Est, nous nous sommes intéressés dans [FPPP09] au codage de sources parcimonieuses. En particulier, nous avons regardé les performances débit-distorsion d'un quantificateur scalaire uniforme pour ce type de signaux, ici modélisés par des lois gaussiennes généralisées.

Supposons que le signal source corresponde à la trajectoire d'une variable aléatoire X suivant une loi de probabilité donnée par :

$$\forall \xi \in \mathbb{R}, \quad g(\xi) = (1 - \epsilon)\delta(\xi) + \epsilon f(\xi) \quad (3.1)$$

où ϵ est un paramètre de mélange, δ est la masse de Dirac et f est la densité d'une loi gaussienne généralisée :

$$\forall \xi \in \mathbb{R}, \quad f(\xi) = \frac{\beta\omega^{1/\beta}}{2\Gamma(1/\beta)} e^{-\omega|\xi|^\beta} \quad (3.2)$$

où $\beta \in]0, 2]$ et $\omega > 0$ sont les paramètres de la loi. L'intérêt de ce modèle, comme on peut le voir dans [Mal89, ML99, ALP02], vient de sa bonne représentation des coefficients d'ondelette des images naturelles. On peut voir de plus que les lois gaussiennes généralisées sont des lois à queue lourde quand β est petit.

On suppose que cette variable aléatoire X est quantifiée uniformément. On étudie alors la variable aléatoire quantifiée \bar{X} donnée pour un pas de quantification $q > 0$ constant par :

$$\bar{X} = r_0 = 0, \quad \text{si } |X| < \frac{q}{2}, \quad (3.3)$$

et pour tout $i \in \mathbb{Z}$ non nul,

$$\bar{X} = r_i, \quad \text{si } (|i| - \frac{1}{2})q \leq |X| < (|i| + \frac{1}{2})q \quad (3.4)$$

où

$$\forall i \geq 1, \quad r_i = -r_{-i} = (i + \zeta)q \quad (3.5)$$

$-1/2 \leq \zeta \leq 1/2$ étant le niveau de reconstruction.

Dans le cas de sources sans mémoire, le débit est approché par l'entropie H_g de la variable aléatoire quantifiée. Si on renormalise le pas de quantification en posant $\bar{q} = q\omega^{1/\beta}$ et qu'on note

$$p_0 = 2 \int_0^{\frac{q}{2}} f(\xi) d\xi, \quad (3.6)$$

cette quantité H_g s'écrit

$$H_g(\epsilon, \bar{q}) = \epsilon H_f(\bar{q}) + \Phi(p_0, \epsilon), \quad \epsilon \in [0, 1] \quad (3.7)$$

où

$$\Phi(p_0, \epsilon) = -(1 - \epsilon(1 - p_0)) \ln(1 - \epsilon(1 - p_0)) - \epsilon(1 - p_0) \ln \epsilon + \epsilon p_0 \ln p_0, \quad (3.8)$$

et

$$H_f(q) = - \sum_{i=-\infty}^{\infty} P(\bar{X} = r_i) \log_2 P(\bar{X} = r_i). \quad (3.9)$$

Nous avons étudié ces entropies en en donnant une approximation à haute résolution ($q \rightarrow 0$) et à basse résolution. Ainsi nous avons obtenu le résultat suivant.

Résultat 3.1.1 Soit $\epsilon \in (0, 1]$ fixé. Quand $q \rightarrow 0$, on a

$$H_g(\epsilon, \bar{q}) = H_\epsilon + \epsilon \left(h_\beta(1) - \ln \bar{q} + \mathbf{1}_{(0,1)}(\epsilon) \frac{\beta \bar{q}}{2\Gamma(1/\beta)} \ln \bar{q} \right) + O(\bar{q}) \quad (3.10)$$

où

$$H_\epsilon = -\epsilon \ln \epsilon - (1 - \epsilon) \ln(1 - \epsilon) \quad (3.11)$$

est l'entropie d'une variable suivant une loi de Bernoulli de paramètre $(1 - \epsilon, \epsilon)$, tandis que $h_\beta(1) = \ln \left(\frac{2\Gamma(1/\beta)}{\beta} \right) + 1/\beta$ est l'entropie différentielle d'une variable aléatoire suivant une loi gaussienne généralisée.

Pour $\bar{q} \rightarrow \infty$,

$$H_g(\epsilon, \bar{q}) = -(1 - \epsilon(1 - \rho_0)) \ln(1 - \epsilon(1 - \rho_0)) - \epsilon(1 - \rho_0) \ln \left(\frac{\epsilon(1 - \rho_0)}{2} \right) \quad (3.12)$$

$$+ O(\bar{q} e^{-(3\bar{q}/2)^\beta}) \quad (3.13)$$

$$= \frac{\epsilon \bar{q} e^{-\bar{q}^\beta/2^\beta}}{2\Gamma(1/\beta) \bar{q}^\beta} \left(1 + \frac{1 - 1/\beta}{\bar{q}^{2\beta}} \right) \left(\left(\frac{\bar{q}}{2} \right)^\beta + \ln \left(\frac{\bar{q}^\beta}{\bar{q}} \right) + \ln \left(\frac{4e\Gamma(1/\beta)}{\epsilon} \right) \right) \quad (3.14)$$

$$+ O \left(\frac{1}{\bar{q}^{2\beta}} \right), \quad (3.15)$$

où $\tilde{q} = (2^{-\beta} \bar{q}^\beta + 1 - 1/\beta)^{1/\beta}$.

Ce résultat est représenté par la figure 3.1.

Nous avons de même regardé les approximations de la distorsion à haute et basse résolution. Pour calculer cette quantité nous avons considéré les normes L^p , pour $1 \leq p \leq 2$, de l'erreur due à la quantification. Ainsi

$$d_{p,\zeta}(\omega, \epsilon, q) = 2\epsilon \left(\int_0^{\frac{q}{2}} \xi^p f(\xi) d\xi + \sum_{i=1}^{\infty} \int_{(i-\frac{1}{2})q}^{(i+\frac{1}{2})q} |\xi - r_i|^p f(\xi) d\xi \right). \quad (3.16)$$

Et sa version normalisée :

$$\bar{d}_{p,\zeta}(\bar{q}) = \omega^{p/\beta} d_{p,\zeta}(\omega, \epsilon, q). \quad (3.17)$$

Nous avons alors obtenu le résultat suivant.

Résultat 3.1.2 Soit $\nu = \left(\frac{1}{2} + \zeta \right)^{p+1} + \left(\frac{1}{2} - \zeta \right)^{p+1}$. Quand $\bar{q} \rightarrow 0$,

$$\bar{d}_{p,\zeta}(\bar{q}) = \frac{\nu \bar{q}^p}{p+1} (1 + O(\bar{q})). \quad (3.18)$$

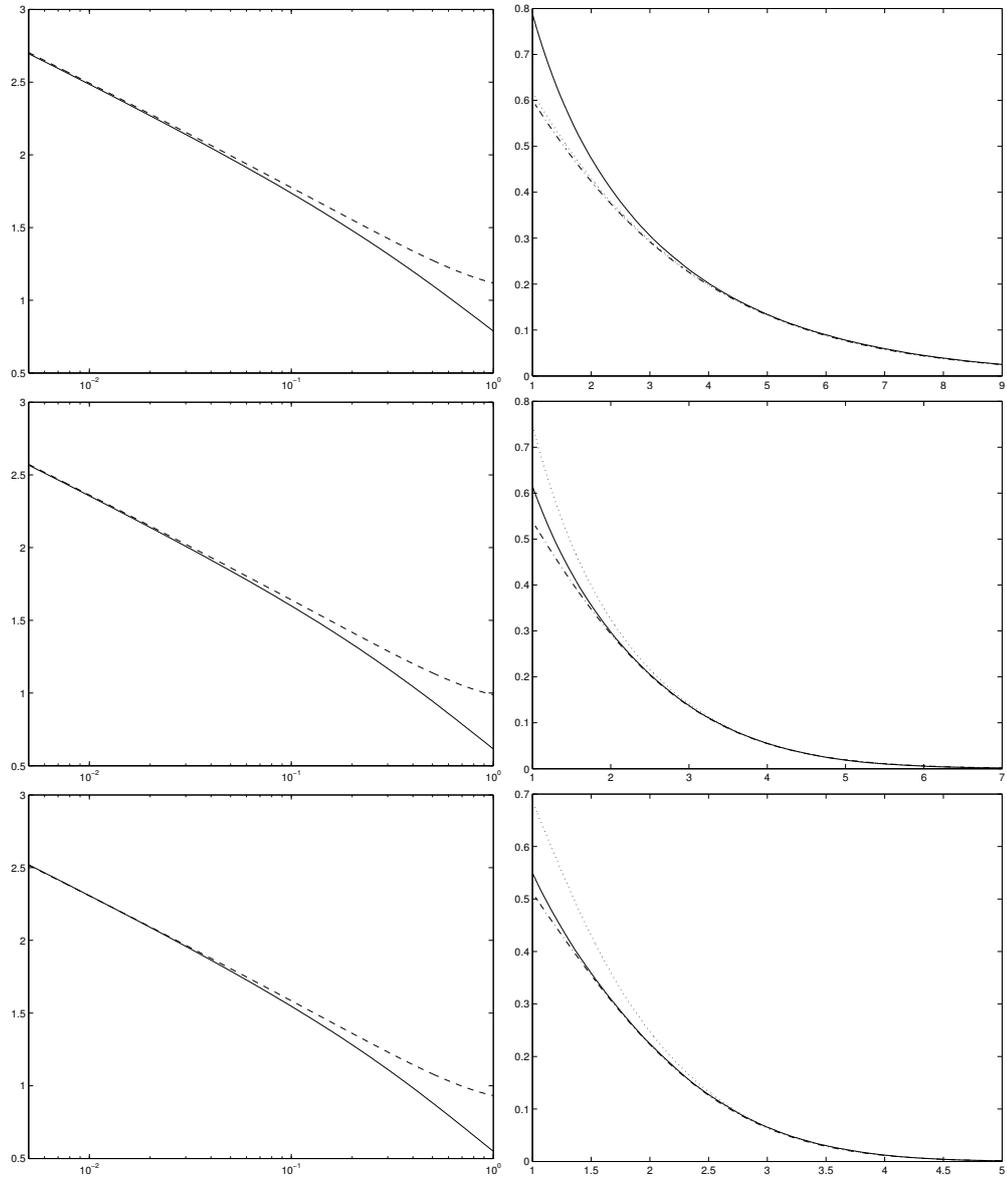


FIGURE 3.1 – Approximations de l'entropie d'une source BGG quantifiée uniformément à haute (gauche) et basse (droite) résolution pour $\epsilon = 0.3$, $\beta = 1$ (haut), $\beta = 3/2$ (milieu) et $\beta = 2$ (bas). H_f est en traits pleins, son approximation haute résolution en tirets et ses approximations basse résolutions $\bar{q} \mapsto -(1 - \epsilon(1 - p_0)) \ln(1 - \epsilon(1 - p_0)) - \epsilon(1 - p_0) \ln(\epsilon(1 - p_0)/2)$ et $\bar{q} \mapsto \frac{\epsilon \bar{q} \exp(-2^{-\beta} \bar{q}^\beta)}{2\Gamma(1/\beta) \bar{q}^\beta} (1 + (1 - 1/\beta) \bar{q}^{-2\beta}) (2^{-\beta} \bar{q}^\beta + \ln(\bar{q}^\beta/\bar{q}) + \ln(4e\Gamma(1/\beta)/\epsilon))$ sont en tirets-points.

uniformément en ζ . Tandis que quand $\bar{q} \rightarrow \infty$, nous avons plusieurs cas :

$$\bar{d}_{p,\zeta}(\bar{q}) = \begin{cases} \mu_p - \frac{\bar{q}^{p+1} e^{-\bar{q}^\beta/2^\beta}}{2^{p+1}\Gamma(1/\beta)\bar{q}^\beta} \left(1 - (1+2\zeta)^p + \frac{p}{\beta\bar{q}^\beta} (1 + (1+2\zeta)^{p-1}) + O\left(\frac{1}{\bar{q}^{2\beta}}\right) \right) & \text{si } \zeta \neq -1/2 \text{ ou } p \geq 2 \\ \mu_1 - \frac{\bar{q}^2 e^{-\bar{q}^\beta/2^\beta}}{4\Gamma(1/\beta)\bar{q}^\beta} \left(1 + O\left(\frac{1}{\bar{q}^{2\beta}}\right) \right) & \text{si } \zeta = -1/2 \text{ et } p = 1 \\ \mu_p - \frac{\bar{q}^{p+1} e^{-\bar{q}^\beta/2^\beta}}{2^{p+1}\Gamma(1/\beta)\bar{q}^\beta} \left(1 + \frac{p}{\beta\bar{q}^\beta} + O\left(\frac{1}{\bar{q}^{p\beta}}\right) \right) & \text{si } \zeta = -1/2 \text{ et } 1 < p < 2 \end{cases} \quad (3.19)$$

Ici,

$$\mu_p = \omega^{p/\beta} \frac{\mathbb{E}[|X|^p]}{\epsilon} = \frac{\Gamma((p+1)/\beta)}{\Gamma(1/\beta)}.$$

Une fois encore, on peut représenter ce résultat par la figure 3.2.

Finalement, nous avons étudié la fonction débit distorsion fonctionnelle, donnée pour tout $D > 0$ par :

$$\forall \zeta \in [-1/2, 1/2], \quad R_{p,\zeta}(\epsilon, D) = \inf_{\{q>0 \mid \epsilon \bar{d}_{p,\zeta}(q) \leq D\}} H_g(\epsilon, q) \quad (3.20)$$

et

$$R_p(\epsilon, D) = \inf_{\zeta \in [-1/2, 1/2]} R_{p,\zeta}(\epsilon, D). \quad (3.21)$$

Grâce aux résultats précédents, nous obtenons finalement :

Résultat 3.1.3 Pour $D \rightarrow 0$,

$$R_p(\epsilon, D) = H_\epsilon + \epsilon \left(h_\beta(1) - \ln 2 - \frac{1}{p} \ln \left(\frac{(p+1)D}{\epsilon} \right) + \mathbf{1}_{(0,1)}(\epsilon) \left(\frac{(p+1)D}{\epsilon} \right)^{1/p} \frac{\beta \ln D}{p\Gamma(1/\beta)} \right) + O(D^{1/p}).$$

Tandis qu'à basse résolution dans le cas extrême où $D \rightarrow \epsilon\mu_p$, nous avons

$$\lim_{D \rightarrow (\epsilon\mu_p)^-} \frac{R_p(\epsilon, D)}{\epsilon\mu_p - D} = \begin{cases} \infty & \text{si } p < \beta \\ 1 & \text{si } p = \beta \\ 0 & \text{si } p > \beta. \end{cases} \quad (3.22)$$

En comparant les résultats obtenus avec les bornes optimales de Shannon, nous avons ainsi montré qu'à basse résolution la quantification scalaire uniforme est optimale à la

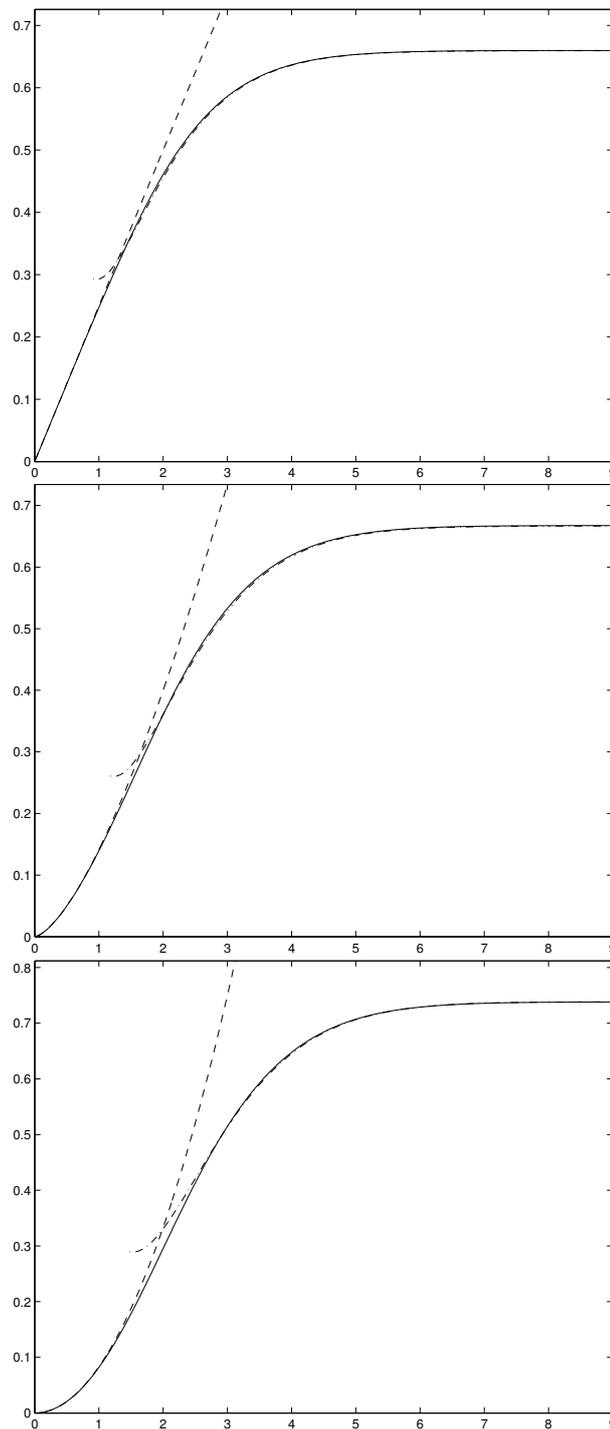


FIGURE 3.2 – Distorsion pour $\beta = 3/2$ et : $p = 1$ (haut), $p = 3/2$ (milieux) ou $p = 2$ (bas). $\bar{d}_{p,0}$ est en trait plein tandis que ses approximations à haute et basse résolution sont en tirets et points.

condition d'avoir $p \geq \beta$ et en prenant le niveau de reconstruction ζ égal à $-\frac{1}{2}$ et qu'elle est sous optimale dans les autres cas. Ce résultat montre notamment que cette méthode simple de quantification n'est pas aussi mauvaise que l'on pourrait le croire, mais que pour qu'elle soit optimale on ne doit pas centrer la reconstruction à $\zeta = 0$ comme cela est fait généralement.

Les démonstrations de ces différents résultats sont dans [FPPP09] présenté en annexe.

3.2 Allocation de débits

Récemment dans la continuité de ces travaux, dans le cadre du post doctorat de Mounir Kaaniche actuellement Maître de conférences à l'Université Paris 13, nous avons dans [KFPPP14] présenté un schéma de compression avec pertes de données basé sur les résultats précédents. Dans ce contexte les méthodes analytiques classiques, où la solution prend une forme explicite, s'appuient sur une optimisation à haute résolution de critères débit-distorsion, utilisant généralement une quantification scalaire uniforme. L'inconvénient majeur de cette approche est qu'elle repose sur des approximations haut débit et est par conséquent limitée dans son champ d'applications.

Pour pouvoir proposer une méthode d'allocation de débit par sous bandes plus générale, les résultats de [FPPP09] sur l'approximation de la distorsion ont été améliorés, donnant une estimation plus précise de la courbe débit-distorsion quel que soit le débit considéré. A l'aide de cette approximation, valable à haute et basse résolution, nous avons formulé le problème d'allocation de débits comme un problème d'optimisation convexe, ce qui nous a permis de proposer un algorithme applicable à des signaux présentant une représentation creuse lors d'une analyse en sous-bandes. Ce problème d'optimisation donne une solution quasi-explicite.

Nous avons ainsi gardé le même modèle de signal que dans la partie 3.1 mais en supposant que nous avons plusieurs jeux de paramètres suivant les sous-bandes. Ainsi, si le signal est décomposé sur J sous-bandes, les coefficients dans chacune d'elles peuvent être caractérisés par un modèle de mélange Bernoulli-Gaussien Généralisé (BGG). Pour chaque sous-bande $j \in \{1, \dots, J\}$, il y a alors n_j (tels que $\sum_{j=1}^J n_j = n$) coefficients dont la loi est donnée par :

$$\forall \xi \in \mathbb{R}, \quad g_j(\xi) = (1 - \epsilon_j)\delta(\xi) + \epsilon_j f_j(\xi) \quad (3.23)$$

où $\epsilon_j \in [0, 1]$ est un paramètre de mélange et f_j est une densité de probabilité gaussienne généralisée d'exposant β_j et de facteur d'échelle ω_j . Les données ainsi modélisées sont quantifiées à l'aide d'un quantificateur scalaire uniforme de pas de quantification $q_j > 0$.

Le problème que nous avons alors étudié consiste à déterminer les pas de quantification

$\mathbf{q} = (q_1, q_2, \dots, q_J) \in [0, +\infty[^J$ minimisant l'erreur

$$D(\mathbf{q}) = \sum_{j=1}^J \rho_j d_j(q_j), \quad (3.24)$$

où $(\rho_j)_{j=1, \dots, J}$ est une suite de poids positifs tels que $\sum_{j=1}^J \rho_j = 1$, sous la contrainte que l'entropie globale soit inférieure à une valeur cible R_{\max} :

$$H(\mathbf{q}) = \sum_{j=1}^J \frac{n_j}{n} H_{f_j}(q_j) \leq R_{\max}. \quad (3.25)$$

3.2.1 Principe de l'algorithme proposé

La méthode analytique classique d'allocation de débit utilise la formule de Bennett pour approcher l'entropie. Cette formule exprime l'entropie à haute-résolution comme une application affine de \mathbf{l} , où $\mathbf{l} = (l_1, \dots, l_J)$, avec $\forall j \in \{1, \dots, J\} l_j = \log_2(q_j)$. Malheureusement cette formule correspond à un développement asymptotique, pour des pas de quantification tendant vers zéro. Son utilisation n'est donc parfaitement valide que dans ces conditions idéales. Dans notre étude, nous avons remplacé cette approximation affine par une fonction convexe de cette même variable \mathbf{l} , donnée par une fonction affine par morceaux. Ainsi :

$$\forall j \in \{1, \dots, J\}, \quad g_j(l_j) = a_j^k l_j + c_j^k \quad \text{si} \quad l_j^{(h, k-1)} \leq l_j \leq l_j^{(h, k)} \quad (3.26)$$

où $k \in \{1, 2, \dots, m^{(h)}\}$ et $m^{(h)}$ correspond au nombre d'intervalles choisis. Ainsi

$$l_j^{(h, 0)} \leq l_j^{(h, 1)} \leq l_j^{(h, 2)} \leq \dots < l_j^{(h, m^{(h)})}$$

et $(a_j^k)_{1 \leq k \leq m^{(h)}}$ forme une suite de nombres négatifs, ce qui permet d'avoir une fonction décroissante tandis que la suite $(c_j^k)_{k \in \{1, 2, \dots, m^{(h)}\}}$ est choisie de façon à ce que g_j soit continue. De plus afin de conserver la positivité de l'entropie, $g_j(l_j) = 0, \forall l_j \geq l_j^{(h, m^{(h)})}$. Ainsi $l_j^{(h, m^{(h)})}$ est le premier point pour lequel g_j s'annule. En outre, en cohérence avec (3.25) :

$$\forall j \in \{1, \dots, J\}, \quad l_j^{(h, 0)} = \min\left(\frac{nn_j^{-1}R_{\max} - c_j^1}{a_j^1}, l_j^{(h, 1)}\right) \quad (3.27)$$

Les autres points $(l_j^{(h, k)})_{1 \leq k \leq m^{(h)}-1}$ sont choisis afin d'avoir la meilleure approximation affine par morceaux de H_{f_j} .

Pour ce qui est de la distorsion utilisée, nous nous basons sur l'approximation (3.18), qui pour $p_j \geq 1$ nous donne :

$$d_j(l_j) = \epsilon_j \frac{\nu_j}{p_j + 1} 2^{l_j p_j}, \quad (3.28)$$

que nous avons de même étendue à des pas de quantification potentiellement plus grands en considérant $\sum_{j=1}^J \frac{n_j}{n} e_j(l_j)$, où $e_j(l_j)$ est donné par :

$$\forall j \in \{1, \dots, J\}, \quad e_j(l_j) = \frac{\epsilon_j \eta_j (p_j + 1)}{\nu_j} (\alpha_j^k 2^{l_j \gamma_j^k} + \delta_j^k) \quad \text{pour } l_j^{(d,k-1)} < l_j \leq l_j^{(d,k)} \quad (3.29)$$

où $k \in \{1, 2, \dots, m^{(d)}\}$ et $m^{(d)}$ correspond au nombre d'intervalles considérés pour approcher au mieux la distorsion et $\forall k \in \{1, 2, \dots, m^{(d)}\}$, $\alpha_j^k > 0$, $\gamma_j^k > 0$, δ_j^k et $\eta_j > 0$ sont des poids. Il est à noter que pour approcher au mieux l'entropie et la distorsion, les différents intervalles $\left(l_j^{(h,k)}\right)_{1 \leq k \leq m^{(h)}}$ et $\left(l_j^{(d,k)}\right)_{1 \leq k \leq m^{(d)}}$ sont choisis indépendamment. Notons que dans **[KFPPP14]** nous avons testé l'influence du choix de $m^{(h)}$ et $m^{(d)}$ sur la qualité de l'approximation.

Une fois ces approximations faites, nous avons réécrit le problème d'allocation de débit comme un problème d'optimisation convexe : trouver $\tilde{\mathbf{l}}$ minimisant la distorsion

$$\forall \mathbf{l} = (l_1, \dots, l_J) \in \mathbb{R}^J \mapsto D(\mathbf{l}) = \sum_{j=1}^J \frac{n_j}{n} e_j(l_j)$$

sur l'ensemble C défini par

$$C := \{\mathbf{l} = (l_1, \dots, l_J) \in \mathbb{R}^J \mid \sum_{j=1}^J \frac{n_j}{n} g_j(l_j) \leq R_{\max}\}. \quad (3.30)$$

Notons que par construction, dans ce problème les fonctions g_j ne sont pas différentiables. Nous avons alors réarrangé les $\left(l_j^{(h,k)}\right)_{0 \leq k \leq m^{(h)}}$ et $\left(l_j^{(d,k)}\right)_{1 \leq k \leq m^{(d)}}$ par ordre croissant pour chaque $j \in \{1, \dots, J\}$, ce qui nous donne une suite croissante de points (l_j^0, \dots, l_j^m) . Puis nous séparons le problème d'optimisation en une optimisation sur chacun des domaines $[l_1^{b_1}, l_1^{b_1+1}] \times \dots \times [l_J^{b_J}, l_J^{b_J+1}]$ où $\mathbf{b} = (b_1, \dots, b_J) \in \{0, \dots, m-1\}^J$. Nous obtenons ainsi m^J sous-domaines sur lesquels nous pouvons utiliser les méthodes d'optimisation convexe différentiable pour trouver les pas de quantification optimaux. Notons donc $\mathbf{P}_{\mathbf{b}} = [l_1^{b_1}, l_1^{b_1+1}] \times \dots \times [l_J^{b_J}, l_J^{b_J+1}]$ l'un de ces domaines et $\mathcal{P}_{\mathbf{b}}$ le problème d'optimisation associé sur $\mathbf{P}_{\mathbf{b}}$. Afin de simplifier les écritures, notons pour chaque $j \in \{1, \dots, J\}$:

$$a_j = a_j^{b_j}, \quad \gamma_j = \gamma_j^{b_j}, \quad \alpha_j = \alpha_j^{b_j}, \quad \delta_j = \delta_j^{b_j}, \quad c_j = c_j^{b_j} \quad (3.31)$$

$$N_j = -\frac{n_j a_j}{\gamma_j}, \quad (3.32)$$

$$\underline{\lambda}_j = \kappa_j 2^{\gamma_j l_j^{b_j}}, \quad (3.33)$$

$$\bar{\lambda}_j = \kappa_j 2^{\gamma_j l_j^{b_j+1}} \quad (3.34)$$

où $\kappa_j = -\epsilon_j \eta_j \alpha_j \gamma_j \ln 2 (p_j + 1) / (a_j \nu_j)$.

Alors la solution de $(\mathcal{P}_{\mathbf{b}})$ est donnée par le résultat suivant.

Résultat 3.2.1 [O1] Si $\sum_{j=1}^J \frac{n_j}{n} g_j(l_j^{b_j+1}) > R_{\max}$, le problème $(\mathcal{P}_{\mathbf{b}})$ n'a pas de solution.

[O2] Si $\sum_{j=1}^J \frac{n_j}{n} g_j(l_j^{b_j}) \leq R_{\max}$, la solution est $\tilde{\mathbf{I}} = (l_1^{b_1}, \dots, l_J^{b_J})$.

[O3] Sinon la solution est donnée par $\tilde{\mathbf{I}}_{\mathbf{b}}$ où

$$\forall j \in \{1, \dots, J\}, \tilde{l}_{j,\mathbf{b}} = \begin{cases} l_j^{b_j} & \text{si } j \in \mathbb{I} \\ \frac{1}{\gamma_j} \log_2 \left(\frac{\tilde{\lambda}}{\kappa_j} \right) & \text{si } j \in \mathbb{J} \\ l_j^{b_j+1} & \text{si } j \in \mathbb{K} \end{cases} \quad (3.35)$$

et

$$\tilde{\lambda}^{N_{\mathbb{J}}} = 2^{-nR_{\max} + \sum_{j=1}^J n_j c_j - \sum_{j \in \mathbb{I}} N_j \gamma_j l_j^{b_j} - \sum_{j \in \mathbb{K}} N_j \gamma_j l_j^{b_j+1}} \prod_{j \in \mathbb{J}} \kappa_j^{N_j} \quad (3.36)$$

$$N_{\mathbb{J}} = \sum_{j \in \mathbb{J}} N_j \quad (3.37)$$

$$\mathbb{I} = \{j \in \{1, \dots, J\} \mid \Phi'(\underline{\lambda}_j) \leq 0\} \quad (3.38)$$

$$\mathbb{K} = \{j \in \{1, \dots, J\} \mid \Phi'(\bar{\lambda}_j) > 0\} \quad (3.39)$$

$$\mathbb{J} = \{1, \dots, J\} \setminus (\mathbb{I} \cup \mathbb{K}) \quad (3.40)$$

$$\forall j \in \{1, \dots, J\}, \forall \lambda \in \mathbb{R}_+, \Phi(\lambda) = \lambda \left(\sum_{j=1}^J \frac{n_j}{n} c_j - R_{\max} \right) - \sum_{j=1}^J \varphi_j(\lambda), \quad (3.41)$$

$$\varphi_j(\lambda) = \begin{cases} \frac{N_j}{n} (\gamma_j l_j^{b_j} \lambda - \frac{\lambda_j}{\ln 2}) - \frac{n_j \eta_j (p_j+1)}{\nu_j n} \delta_j & \text{pour } \lambda \leq \underline{\lambda}_j \\ \frac{N_j \lambda}{n \ln 2} (\ln(\frac{\lambda}{\kappa_j}) - 1) - \frac{n_j \eta_j (p_j+1)}{\nu_j n} \delta_j & \text{pour } \underline{\lambda}_j < \lambda < \bar{\lambda}_j \\ \frac{N_j}{n} (\gamma_j l_j^{b_j+1} \lambda - \frac{\bar{\lambda}_j}{\ln 2}) - \frac{n_j \eta_j (p_j+1)}{\nu_j n} \delta_j & \text{pour } \lambda \geq \bar{\lambda}_j. \end{cases}$$

Une fois obtenus les pas de quantification optimaux sur chaque sous-domaine, nous avons un nombre fini de valeurs possibles pour la distorsion. Nous choisissons alors la distorsion minimale parmi ces candidats.

La démonstration de ce résultat ainsi que la validation numérique de cette démarche et son application à la compression d'images sont dans [KFPPP14].

3.3 Codage avec information adjacente

Un autre aspect des méthodes de codage auquel je me suis intéressée, en collaboration avec Michel Kieffer, Elsa Dupraz et Francesca Bassi du L2S, est celui d'un codage avec pertes d'une source continue avec information adjacente. Le principe de ces méthodes de type Wyner-Ziv est de profiter de l'information donnée par un deuxième vecteur corrélé

à la source pour améliorer la reconstruction dans le cas sans mémoire. L'information adjacente est alors connue au niveau du décodeur mais pas de l'encodeur. Elle permet d'améliorer la reconstruction des données sans être prise en compte au niveau de l'encodage. Dans les cas gaussiens, la fonction débit-distorsion optimale correspondant à ce schéma a une forme explicite bien connue et est largement utilisée comme référence pour mesurer les performances d'un schéma de codage. Cependant dans beaucoup d'applications l'hypothèse de gaussianité n'est pas réaliste.

C'est pourquoi nous nous sommes intéressés à la caractérisation d'une fonction débit-distorsion de type Wyner-Ziv dans le cadre d'un modèle plus général d'information adjacente. Ce cadre permet en particulier de considérer les pertes d'information qui peuvent intervenir au cours de la transmission. Cette perte potentielle est considérée sur le canal de corrélation entre la source et l'information adjacente. Dans ce cas, la corrélation est modélisée par une variable d'état cachée correspondant à cet aspect erratique. Naturellement on pense dans ce cas au mélange de gaussiennes mais plus généralement on peut considérer des modèles à variables cachées de types mélanges scalaires de gaussiennes qui permettent de généraliser ce modèle à un canal portant une information de parcimonie.

3.3.1 Modèle de mélange scalaire de gaussiennes

Dans l'étude effectuée dans [BFDK14], nous avons modélisé la dépendance entre la source S et l'information adjacente Y en prenant le modèle additif suivant :

$$Y = S + X,$$

où S et X sont indépendants et $S \sim \mathcal{N}(0, \sigma_S^2)$. La corrélation entre S et Y est modélisée par X qui suit une loi de mélange de gaussiennes. Ces lois ont été présentées dans [WS00] comme un modèle général de distributions à queues lourdes.

Ici $X = \sqrt{Z}V$, où Z et V sont indépendants et $V \sim \mathcal{N}(0, 1)$. La variable d'état cachée Z a une distribution \mathbb{P}_Z dont le support est dans \mathbb{R}^+ . La distribution de X peut donc s'écrire

$$p_X(x) = \int \frac{1}{(2z\pi)^{1/2}} \exp\left(-\frac{x^2}{2z}\right) d\mathbb{P}_Z(z). \quad (3.42)$$

La variance de X est $\sigma_X^2 = \mathbb{E}[Z]$.

L'intérêt de cette représentation est qu'elle permet de garder un modèle de parcimonie tout en maintenant une forme pour X gaussienne, sachant une variable cachée. En effet, la densité (3.42) peut prendre différentes formes en fonction de la distribution de la variable cachée, \mathbb{P}_Z . On peut citer le mélange de gaussiennes quand Z a un support fini mais aussi la loi de Student-t, pour $Z \sim \Gamma(\alpha, \alpha)$ ou la distribution de Laplace pour laquelle Z suit une loi exponentielle [EKL06]. Plus généralement, ce type de modèle est très utilisé pour différents types de lois à queues lourdes telles que les gaussiennes généralisées ou les

lois α -stables [CGLS08, PSWS03, ATB03, BD06]. Ainsi ce modèle peut être considéré comme une généralisation du modèle précédent. L'inconvénient dans certains cas est la variable cachée dont la loi n'est pas forcément connue explicitement.

3.3.2 Fonctions débit-distorsion

On note dans la suite W_e et W_d les informations adjacentes disponibles respectivement à l'encodeur et au décodeur. La mesure de distorsion considérée est l'erreur quadratique moyenne $d(S, \hat{S}) = \|S - \hat{S}\|^2$.

La fonction débit-distorsion conditionnelle, [Gra72], décrit les performances optimales que l'on peut espérer obtenir quand l'information adjacente est connue à la fois à l'encodeur et au décodeur, c'est-à-dire pour $W_e = W_d = W$:

$$R_{S|W}(D) = \inf_{\hat{S} : \mathbb{E}[d(S, \hat{S})] \leq D} I(S; \hat{S}|W). \quad (3.43)$$

A contrario, la fonction débit-distorsion de Wyner-Ziv est obtenue en considérant que cette information adjacente n'est disponible qu'au décodeur. On a alors

$$R_{S|W_d}^{\text{WZ}}(D) = \inf_{U \in \mathcal{M}(D)} (I(S; U) - I(U; W_d)). \quad (3.44)$$

où $\mathcal{M}(D)$ est l'ensemble des variables aléatoires U telles qu'il existe une fonction \hat{S} de U vérifiant $\mathbb{E}[d(S, \hat{S})] \leq D$.

3.3.3 Principaux résultats obtenus

Le but de [BFDK14] est de déterminer la fonction débit-distorsion $R_{S|Y}^{\text{WZ}}(D)$, définie dans (3.44) pour une information adjacente $W_d = Y$ et un signal dont le modèle est donné dans la partie 3.3.1. Le terme $I(S; U) - I(U; Y)$ apparaissant dans (3.44) ne peut en général pas être exprimé de manière explicite rendant le problème de minimisation (3.44) difficile à résoudre dès qu'on sort du cas gaussien donné dans [Wyn78]. C'est pourquoi plutôt que de chercher à déterminer exactement cette fonction débit-distorsion nous l'avons encadrée par deux quantités facilement calculables.

En ce qui concerne la borne inférieure, un candidat naturel est la fonction débit-distorsion conditionnelle $R_{S|Y}(D)$ [Zam96], définie par (3.43), où $W = Y$. Cependant même cette quantité est difficilement calculable dans le cas présent puisque l'estimateur des moindres carrés de S ne l'est pas. C'est pourquoi nous considérons plutôt une autre borne inférieure pour laquelle nous supposons connue la variable cachée Z à l'encodeur et au décodeur, $R_{S, Z|Y, Z}^{\text{WZ}}(D)$. En ce qui concerne la borne supérieure elle est obtenue en prenant une stratégie de canal test.

Résultat 3.3.1 Considérons le modèle donné dans la partie 3.3.1. La fonction débit-

distorsion de Wyner-Ziv $R_{S|Y}^{\text{wz}}(D)$ est bornée inférieurement dans ce cas par

$$R_{S,Z|Y,Z}^{\text{wz}}(D) \leq R_{S|Y}^{\text{wz}}(D), \quad (3.45)$$

La borne inférieure $R_{S,Z|Y,Z}^{\text{wz}}(D)$ vérifie alors

$$R_{S,Z|Y,Z}^{\text{wz}}(D) = \frac{1}{2} \mathbb{E} \left[\log_2 \left(\frac{\varphi(Z)}{\Delta^*(Z)} \right) \right], \quad (3.46)$$

où $\varphi(Z) = \sigma_X^2 Z / (\sigma_X^2 + Z)$ et

$$\Delta^*(z) = \begin{cases} \lambda & \text{si } \varphi(z) > \lambda \\ \varphi(z) & \text{sinon,} \end{cases} \quad (3.47)$$

où $\lambda = \frac{1}{2 \ln 2 \mu}$, $\mathbb{E}[\Delta^*(Z)] = D$.

Résultat 3.3.2 Considérons le modèle donné dans la partie 3.3.1. La fonction débit-distorsion de Wyner-Ziv $R_{S|Y}^{\text{wz}}(D)$ est bornée supérieurement dans ce cas par

$$R_{S|Y}^{\text{wz}}(D) \leq \min \left\{ R_{S,Z|Y,Z}^{\text{wz}}(D) + L(D) + I(Z; X), R_{S|\tilde{Y}}^{\text{wz}}(D) \right\}. \quad (3.48)$$

Où $L(D)$ in (3.48) est donné par

$$L(D) = \frac{1}{2} \mathbb{E} \left[\log_2 \left(D \left(\frac{1}{\varphi(Z)} - \frac{1}{\sigma_S^2} \right) + 1 \right) \right], \quad (3.49)$$

et $R_{S|\tilde{Y}}^{\text{wz}}(D)$ est la fonction débit-distorsion de Wyner-Ziv pour laquelle $\tilde{Y} = S + \tilde{X}$ et $\tilde{X} \sim \mathcal{N}(0, \mathbb{E}(Z))$.

On peut remarquer que le terme $L(D)$ dans l'équation (3.48) décroît vers 0 quand $D \rightarrow 0$. Ainsi à haut débit la différence entre la borne inférieure et la borne supérieure est inférieure à une constante $I(Z; X)$.

La démonstration détaillée de ces résultats n'est pas présentée ici mais se trouve dans [BFDK14], présenté en annexe. On peut noter que la démonstration du Résultat 3.3.1 s'appuie sur le théorème de Karush-Kuhn-Tucker généralisé à la dimension infinie, qui se trouve par exemple dans [Lue69].

3.4 Conclusion

Ce travail initié durant mon post-doctorat a été développé surtout dans mes premières années au L2S. Il s'est fortement focalisé sur les fonctions débit-distorsion pour des signaux parcimonieux et en particulier sur leur codage à haut et à bas débit. Les résultats

obtenus dans la première partie ont permis de voir qu'une quantification simple pouvait donner de bons résultats de reconstruction même à bas débit sous certaines conditions. Grâce à cela nous avons construit un schéma de codage simple mais efficace, en particulier pour la compression d'images naturelles. Ces travaux ont ensuite été généralisés par exemple au cas stéréo dans [HKBC15]. En parallèle, les travaux plus récents, menés au sein de mon laboratoire ont permis de prolonger ces résultats à des schémas de transmission plus sophistiqués, en utilisant une information adjacente au décodage. Là encore nous avons obtenu des résultats intéressants pour la communauté puisqu'ils donnent un ordre de grandeur des performances qu'un schéma de codage peut espérer atteindre.

BILAN (THÈSE/PUBLICATION)

- 3 publications dans des revues internationales à comité de lecture ([J4],[J2],[J7]).
- 4 articles dans des conférences internationales à comité de lecture ([C5],[C8],[C10],[C12]) et 3 dans des conférences nationales ([CN3],[CN5],[CN6]).
- 1 stage de M2.

4.1	Problèmes inverses
4.2	Méthodes pénalisées en problèmes inverses
4.3	Ondelettes
4.4	Estimation non paramétrique
4.5	Détection d'ondes gravitationnelles
4.6	Problèmes inverses des ondes
4.7	Application au JWST
4.8	Conclusion

4 — Estimation et problèmes inverses

Dans ce chapitre nous nous intéressons à deux méthodes classiques d'estimation dans le cadre de problèmes inverses, à savoir les estimateurs linéaires et par seuillage. Nous étudions les performances de ces estimateurs d'un point de vue théorique d'une part et applicatif d'autre part.

Ma thèse et les premiers résultats postérieurs, que je ne rappellerai pas ici, portaient sur des résultats de prévalence pour l'analyse multifractale, autrement dit sur la généralité de certaines propriétés de régularité. J'ai ensuite montré comment ces résultats pouvaient être utilisés pour déterminer les performances d'estimateurs classiques, à savoir les estimateurs linéaires et par seuillage dans le cadre du débruitage. Ces travaux théoriques se sont suivis plus récemment, avec les thèses de Hidayet Zaimaga et d'Amine Hadj-Youcef, par l'application pratique de ces méthodes pour les problèmes inverses. Avant de présenter ces travaux, introduisons un bref rappel sur les problèmes inverses et leur résolution par approches variationnelles, auxquelles sont liés les estimateurs considérés. Les méthodes bayésiennes pour les problèmes inverses seront elles développées dans le chapitre 5.

4.1 Problèmes inverses

Les performances des appareils actuels d'acquisition ont entraîné une augmentation importante de la taille des données à traiter. Qui plus est dans bon nombre de problèmes de reconstruction actuels, en imagerie médicale ou en astrophysique par exemple, on s'intéresse à des problèmes en grande dimension pour lesquels on a besoin d'une précision importante. Les algorithmes de reconstruction doivent donc être de plus en plus capables de reconstruire rapidement des données de grande dimension. Les problèmes

de reconstruction qui nous intéressent ici sont des problèmes inverses mal posés au sens d’Hadamard. Dans ce genre de problèmes il n’y a pas d’unicité de la solution, d’où l’introduction d’information *a priori* sur le signal inconnu permettant d’obtenir une solution acceptable. C’est le cas par exemple des problèmes de tomographie, de super-résolution, entre autres, dans lesquels le processus d’acquisition est relativement complexe et ne permet pas de reconstruction simple du signal source. Afin d’introduire cette information, plusieurs types de méthodes sont envisageables. Les deux grandes familles de méthodes de reconstruction dans le cadre de problèmes inverses sont les méthodes variationnelles et les méthodes probabilistes.

En pratique, la résolution des problèmes inverses consiste à regarder successivement un problème direct et un problème inverse. La résolution du problème direct consiste à construire un modèle direct qui met en relation la grandeur d’intérêt et les données observables. La résolution du problème inverse associé vise à déterminer un signal source à partir de ce modèle.

4.1.1 Modèle direct

Déterminer un modèle direct au plus proche de la physique et permettant une reconstruction adaptée est le premier défi de la résolution d’un problème inverse. Cependant les méthodes d’acquisition de plus en plus sophistiquées nécessitent une modélisation adaptée qui n’est ni forcément triviale ni toujours manipulable.

En règle générale, si on utilise un vecteur $\mathbf{y} \in \mathbb{R}^M$ pour représenter les données et un vecteur $\mathbf{x} \in \mathbb{R}^N$ pour les paramètres de l’objet d’intérêt, et si on se place dans un cadre linéaire on obtient le modèle :

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (4.1)$$

où $\mathbf{A} \in \mathbb{R}^{M \times N}$ modélise l’interaction entre \mathbf{x} et \mathbf{y} et $\mathbf{n} \in \mathbb{R}^M$ représente le bruit de mesure et/ou les incertitudes sur le modèle.

4.1.2 Problème inverse

Le principe du problème inverse associé à (4.1) est de construire un estimateur $\hat{\mathbf{x}}$ des paramètres d’intérêt \mathbf{x} à partir de mesures \mathbf{y} [Idi01]. Même si les problèmes directs sont généralement *bien posés*, les problèmes inverses associés sont souvent *mal-posés*. Un problème est dit bien posé au sens d’Hadamard s’il satisfait les conditions suivantes :

- la solution existe,
- la solution est unique,
- la solution dépend de façon continue des données (i.e. une petite variation dans les données n’introduit pas de grande variation dans la solution).

Un problème est donc dit mal posé au sens d’Hadamard lorsque l’une de ces conditions n’est pas respectée. Malheureusement, la plupart des problèmes inverses rencontrés en

traitement du signal et des images sont mal-posés. Dans les problèmes qui nous ont intéressés, décrits par (4.1), \mathbf{A} est généralement non-inversible ou mal conditionnée, d'où le caractère mal-posé.

Pour résoudre de tels problèmes les méthodes d'inversion nécessitent des informations supplémentaires concernant le paramètre d'intérêt \mathbf{x} [Idi01].

4.2 Méthodes pénalisées en problèmes inverses

Les approches pénalisées s'appuient sur l'introduction d'informations supplémentaires sur le signal d'intérêt grâce à des termes de régularisation qui sont représentés par $R(\mathbf{x})$. Dans ce cas, l'estimateur est obtenu en minimisant un critère pénalisé :

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J(\mathbf{x}) = \arg \min_{\mathbf{x}} J_0(\mathbf{x}) + \lambda R(\mathbf{x}) \quad (4.2)$$

où J_0 est un critère d'adéquation aux données qui mesure la distance entre les données mesurées \mathbf{y} et les données prédites par le modèle $\mathbf{A}\mathbf{x}$ tandis que λ représente le paramètre de régularisation qui contrôle le compromis entre cette fidélité aux données et l'importance des connaissances *a priori* sur le vecteur d'inconnus apporté par la fonction $R(\mathbf{x})$. En ce qui concerne J_0 , le critère des moindres carrés est un choix naturel.

La qualité de la reconstruction $\hat{\mathbf{x}}$ dépend fortement du terme de régularisation choisi ainsi que du paramètre de régularisation λ . En ce qui concerne le terme de régularisation, la régularisation de Tikhonov [Tik63] a été la plus utilisée. Elle possède la forme suivante :

$$R(\mathbf{x}) = \|\mathbf{\Gamma}\mathbf{x}\|^2, \quad (4.3)$$

où $\mathbf{\Gamma}$ représente la *matrice de Tikhonov* qui influence la solution pour le problème considéré. Par exemple, dans le cas où $\mathbf{\Gamma}$ est la matrice identité, des solutions dont la norme euclidienne est petite sont favorisées. On peut aussi, comme nous l'avons fait dans la thèse d'Amine Hadj-Youcef, utiliser un opérateur de différence finie qui permet de favoriser les solutions lisses. L'avantage d'une régularisation quadratique, est qu'elle permet aisément de trouver un estimateur $\hat{\mathbf{x}}$ sous une forme analytique linéaire vis-à-vis des mesures \mathbf{y} . L'inconvénient de cet estimateur est qu'il permet difficilement de prendre en compte d'autres types d'information comme l'information de parcimonie.

Pour surmonter cette limitation, les estimations non-linéaires basées sur des termes de régularisation non-quadratiques préservant les discontinuités ont été proposées. De nombreuses fonctions ont été considérées dans la littérature incluant des fonctions convexes et non-convexes. En ce qui nous concerne, nous nous sommes essentiellement concentrés sur l'information de parcimonie. Dans ce cas, le terme de régularisation R doit correspondre au nombre de pixels non nuls. Cependant dans ce cas le problème (4.2) est NP-complet et donc impossible à résoudre. C'est pourquoi en pratique ce critère est remplacé par une

norme ℓ_1

$$R(\mathbf{x}) = \sum_i |x_i|, \quad (4.4)$$

qui a l'avantage d'être convexe. Dans le cas d'un problème de débruitage, où $\mathbf{A} = Id$, l'estimateur associé à (4.2) correspond à un estimateur de seuillage doux.

Le principal inconvénient de ce type de méthodes, surtout en comparaison avec les méthodes bayésiennes, est qu'il n'y a pas en général de méthode simple pour régler automatiquement le compromis entre les deux termes. De nombreuses approches déterminent la valeur de λ de manière empirique, en choisissant les valeurs selon les expériences ou en considérant la meilleure valeur en faisant plusieurs tentatives. Cette manière de choisir λ s'avère peu rapide et peu pratique. Pour s'affranchir de cette difficulté, de nombreuses méthodes ont été proposées pour les sélections automatiques du paramètre de régularisation, par exemple la validation croisée [GHW79] ou les méthodes L-curve [Han92]. La validation croisée est une bonne approche pour de petits ensembles de données et les cas monodimensionnels mais elle est souvent très coûteuse pour les problèmes de grande taille à cause des inversions de matrices qu'elle nécessite. La méthode L-curve est plus rapide que la validation croisée mais possède des maxima locaux. Dans le cas du débruitage, une autre technique pour régler automatiquement ce paramètre λ est la méthode de Stein, [Ste81], qui relie la valeur de ce paramètre à la variance du bruit. Depuis cet estimateur de Stein a été généralisé dans le cadre de problèmes inverses dans [Eld09, VDP⁺13] mais reste difficile à mettre en œuvre en général.

Dans beaucoup d'applications, ces méthodes sont basées sur une représentation différente des signaux, permettant d'avoir de la parcimonie. Une représentation sur laquelle je me suis concentrée tout au long de ma carrière est basée sur des ondelettes. En effet, cet outil permet d'avoir des représentations de signaux parcimonieuses dans de nombreuses applications tout en gardant un lien avec les espaces fonctionnels classiques.

4.3 Ondelettes

Une base d'ondelettes est une base orthonormée de $L^2(\mathbb{R}^N)$ ayant une bonne localisation dans le plan temps-fréquence. Grâce aux ondelettes, nous obtenons des conditions simples qui permettent d'estimer la régularité ponctuelle des signaux ainsi que l'appartenance à certains espaces fonctionnels liés en particulier à la régularité globale ou à la parcimonie. Historiquement, les bases d'ondelettes remontent aux travaux de Haar [Haa10] et de Stromberg [Str83]. Largement utilisées ensuite en traitement du signal elles offrent l'avantage, par rapport à l'analyse de Fourier, d'être bien localisées à la fois en temps et en fréquence. Elles permettent également de "voir" directement sur les coefficients les singularités des fonctions, ce que ne permet pas la transformée de Fourier.

Une fonction f appartenant à $L^2(\mathbb{R}^N)$ peut se décomposer sous la forme

$$f = \sum_{i,j,k} c_{j,k}^{(i)} \psi_{j,k}^{(i)} \quad (4.5)$$

où $\psi_{j,k}^{(i)} = 2^{Nj/2} \psi^{(i)}(2^j x - k)$ et

$$c_{j,k}^{(i)} = 2^{Nj/2} \int_{\mathbb{R}^N} f(x) \psi_{j,k}^{(i)}(x) dx. \quad (4.6)$$

On peut construire différents types de bases d'ondelettes, ayant des propriétés de régularité différentes. Parmi les plus connues on peut citer l'ondelette de Haar qui n'est pas continue mais dont le support est petit et l'ondelette de Meyer C^∞ à support infini.

Comme on peut le voir sur ces deux exemples, les propriétés de régularité et de taille de support des ondelettes peuvent être très variées. Il n'est par conséquent pas évident de déterminer l'ondelette la mieux adaptée à un problème donné. Un résultat important pour le choix d'une ondelette est alors le théorème suivant, [Dau88].

Théorème 1 — Daubechies. Si ψ est une ondelette qui engendre $L^2(\mathbb{R})$ ayant p moments nuls alors son support est de taille supérieure ou égale à $2p - 1$.

Une conséquence importante de ce résultat est qu'on ne peut pas trouver d'ondelettes C^∞ à support compact.

Un intérêt majeur des ondelettes est qu'elles permettent une représentation parcimonieuse d'un grand nombre de signaux, des images naturelles par exemple. En effet, un grand nombre de signaux, même de grande dimension, peuvent être représentés par un faible nombre de coefficients d'ondelettes. Cette propriété de parcimonie est une caractéristique importante des signaux que ce soit pour la compression, l'estimation ou la reconstruction. Cette propriété est aussi possible et avantageuse par d'autres représentations, en particulier sur des trames redondantes. L'intérêt des ondelettes par rapport à ces représentations redondantes est qu'elles permettent aisément de caractériser certains espaces fonctionnels, de régularité par exemple. Cette propriété est utile notamment pour l'estimation non paramétrique.

4.4 Estimation non paramétrique

Depuis l'essor des ondelettes dans les années 80, on s'est intéressé à la meilleure approximation de signaux dans une base d'ondelettes et au lien entre cette approximation et la régularité de ces signaux. Cette étude a conduit à l'utilisation des espaces de Besov d'une part et à celle d'estimateurs non paramétriques d'autre part. Un des inconvénients de l'utilisation d'estimateurs non paramétriques est qu'ils concernent par définition l'estimation d'objets de dimension infinie et sont donc difficiles à comparer entre eux. La méthode la plus connue pour déterminer les performances de ces estimateurs est alors la recherche de méthodes atteignant un risque minimax, autrement dit minimisant le risque

maximum atteint dans un espace donné. Cette méthode permet de relier les performances d'un estimateur à la régularité du signal recherché.

Le premier résultat faisant intervenir à la fois des problèmes de régularité et la notion de généricité est apparu dans les années 30 avec les travaux de Banach [Ban31] et Mazurkiewicz [Maz31]. Ceux-ci prouvaient que la plupart des fonctions continues étaient nulle part dérivables. Cette notion de « plupart » fondée sur les catégories de Baire donne un aperçu, topologique, de l'importance des objets considérés.

Il faut attendre 1972 avec les travaux de Christensen [Chr72] pour avoir un équivalent de la notion d'ensembles de mesure de Lebesgue nulle dans un espace de dimension infinie, et donc une définition probabiliste de la généricité. Le principe de la théorie de la prévalence introduite par Christensen est de considérer comme petit un ensemble tel que l'on puisse trouver une mesure vérifiant que toutes les translatées de cet ensemble sont de mesure nulle. Si E est un sous espace fonctionnel et si on se donne une propriété \mathcal{P} et un processus X dont les trajectoires sont presque sûrement dans E , pour montrer que \mathcal{P} est une propriété valable sur un ensemble prévalent, il suffit de prouver que :

$$\forall f \in E, \quad \text{presque sûrement } f + X \text{ vérifie } \mathcal{P}.$$

Dans [Fra13] je me suis placée dans le cadre statistique donné par le modèle de bruit blanc gaussien dans $L^2((0,1)^N)$ où le signal observé correspond à la trajectoire d'un processus $(Y_t)_{t \in [0,1]^N}$ donné par :

$$dY_t = f(t)dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in (0,1)^N, \quad (4.7)$$

où dW_t est un processus de Wiener, n est donné et f est la fonction que l'on cherche à estimer.

Dans ce contexte, si Θ est un espace fonctionnel auquel f est supposé appartenir et que Ri est une fonction de risque donnée, souvent une norme L^p , le risque minimax associé à une famille d'estimateurs \mathcal{T}_n définis à partir de (4.7), est :

$$Ri(\Theta) = \inf_{\hat{f} \in \mathcal{T}_n} \sup_{f \in \Theta} \mathbb{E}(Ri(\hat{f}_n, f)). \quad (4.8)$$

Ce risque minimax donne alors une vitesse de convergence optimale sur Θ . Le principal inconvénient de cette technique est qu'elle permet seulement d'optimiser le plus grand risque atteint sur un ensemble. Ce point de vue peut alors être considéré comme pessimiste. Le problème qui se pose est de savoir si ce pire risque correspond à un comportement générique ou si l'on ne risque pas de rejeter ainsi des procédures dont le risque minimax serait mauvais mais qui seraient par ailleurs très bonnes pour un grand nombre de fonctions.

Dans [Fra13] je me suis concentrée sur les espaces de Besov qui peuvent se caractériser par les ondelettes. Ainsi, les espaces de Besov homogènes sont définis, pour $p, q > 0$ et $s \in \mathbb{R}$, par :

$$f \in B_p^{s,q}([0, 1]^N) \iff \exists c > 0, \quad \|f\|_{B_p^{s,q}} := \sum_{j \geq 0} \left(\sum_{k \in \{0, \dots, 2^j - 1\}^N} |c_{j,k}|^p 2^{(sp-d+\frac{pN}{2})j} \right)^{q/p} \leq c. \quad (4.9)$$

Cette caractérisation est indépendante de l'ondelette choisie ψ tant qu'elle a r moments nuls, r vérifiant $r \geq s$. On note aussi $B_{p,c}^{s,q}(\mathbb{R}^N)$, la boule dans $B_p^{s,q}(\mathbb{R}^N)$ de rayon $c > 0$.

En ce qui concerne les estimateurs étudiés ils sont de deux types, les estimateurs linéaires et les estimateurs par seuillage.

Définition 1 Les estimateurs linéaires \hat{f}_n^L pour le modèle (4.8) sont définis par

$$\hat{f}_n^L(x) = \sum_{j \geq 0} \sum_{k \in \{0, \dots, 2^j - 1\}^N} \hat{\theta}_{j,k}^{(n)} \psi_{j,k}(x), \quad (4.10)$$

où

$$\hat{\theta}_{j,k}^{(n)} = \lambda_{j,k}^{(n)} y_{j,k}.$$

Les paramètres $(\lambda_{j,k}^{(n)})_{j,k}$ sont des poids appartenant à $[0, 1]$.

Une autre procédure d'estimation naturelle dans le modèle du bruit blanc, définie par exemple dans [DJ94] et largement utilisée dans la communauté du signal, est de supprimer de petits coefficients afin d'avantager une représentation parcimonieuse. C'est le principe du seuillage.

Définition 2 Le seuillage par ondelettes, dans le cadre du modèle de bruit blanc défini par (4.8) est défini par

$$\hat{f}_n^T(x) = \sum_{j=0}^{j(n)} \sum_{k \in \{0, \dots, 2^j - 1\}^N} \theta_{j,k}^T \psi_{j,k}(x). \quad (4.11)$$

Où les coefficients sont donnés par :

$$\theta_{j,k}^T = y_{j,k} \mathbb{1}_{\{|y_{j,k}| \geq \kappa t_n\}}, \quad (4.12)$$

dans le cas d'un seuillage dur, ou

$$\theta_{j,k}^T = \text{sign}(y_{j,k}) (|y_{j,k}| - \kappa t_n)_+, \quad (4.13)$$

pour le seuillage doux. De plus,

$$t_n = \sqrt{\frac{\log n}{n}},$$

est ce qu'on appelle le seuil universel tandis que $j(n)$ est tel que

$$2^{-j(n)} \leq \frac{\log n}{n} < 2^{-j(n)+1},$$

où κ est une constante assez grande.

En ce qui concerne les résultats d'optimalité minimax de ces différents estimateurs ils remontent à Pinsker [Pin80] qui montra que les estimateurs linéaires sont optimaux si on considère un risque L^2 sur un espace de Sobolev L^2 . Si on généralise la fonction de risque à une norme L^p , [IH81, Bir83] ont montré qu'on pouvait encore trouver des estimateurs linéaires optimaux dans ces espaces de Sobolev.

4.4.1 Principaux résultats obtenus

En ce qui concerne les résultats minimax les plus généraux dans les espaces de Besov, obtenus avec une norme L^p , où $1 \leq p < \infty$, comme fonction de risque, ils remontent à [DJKP97] avec la proposition suivante.

Proposition 1 Soit $1 \leq r \leq \infty$, $1 \leq p < \infty$ et $s > \frac{N}{r}$. Il existe alors $C > 0$ tel que

$$Ri^n(B_{r,c}^{s,\infty}) = \inf_{\widehat{f}_n} \sup_{f \in B_{r,c}^{s,\infty}} \mathbb{E} \|\widehat{f}_n - f\|_{L^p}^p \geq Cr_n(s, r, p)$$

où

$$r_n(s, r, p) = \begin{cases} n^{-\frac{ps}{2s+N}} & \text{si } r > \frac{Np}{2s+N}, \\ \left(\frac{n}{\log n}\right)^{-\frac{p(s-\frac{N}{r}+\frac{N}{p})}{2(s-\frac{N}{r})+N}} & \text{sinon.} \end{cases}$$

Il est de plus démontré dans [DJKP97] que les algorithmes de seuillage atteignent ce risque minimax à une correction logarithmique près contrairement aux estimateurs linéaires. En particulier, comme on peut le voir dans [DJ98], les estimateurs linéaires ont une très mauvaise qualité d'approximation pour des signaux parcimonieux. Ce résultat est généralisé dans [DJ98] qui donne le résultat optimal pour ce type d'estimateurs.

Proposition 2 Soit $1 \leq r \leq \infty$, $1 \leq p < \infty$ et $s > \frac{N}{r}$. Il existe $C > 0$ tel que

$$R_{lin}^n(B_{r,c}^{s,\infty}) = \inf_{\widehat{f}_n \text{ linéaire}} \sup_{f \in B_{r,c}^{s,\infty}} \mathbb{E} \|\widehat{f}_n - f\|_{L^p}^p \geq C\tilde{r}_n(s, r, p)$$

où

$$\tilde{r}_n(s, r, p) = \begin{cases} n^{-\frac{ps}{2s+N}} & \text{si } r > p \\ \left(\frac{n}{\log n}\right)^{-\frac{ps'}{2s'+N}} & \text{sinon,} \end{cases}$$

et $s' = s - \frac{N}{r} + \frac{N}{p}$.

Les résultats obtenus dans [Fra13] prouvent que ces résultats restent valables si on remplace le risque minimax par un risque générique, au sens de la prévalence.

Résultat 4.4.1 Soit $1 \leq r \leq \infty$, $1 \leq p < \infty$ et $s > \frac{N}{r}$. Alors dans le contexte de (4.7) :

— Pour presque toute fonction f de $B_r^{s,\infty}([0,1]^N)$,

$$\inf_{\hat{f}_n^L \text{ linéaire}} \mathbb{E} \|\hat{f}_n^L - f\|_{L^p}^p \approx n^{-\alpha p}, \quad (4.14)$$

où

$$\alpha = \begin{cases} \frac{s}{2s+N} & \text{si } r \geq p, \\ \frac{s - \frac{N}{r} + \frac{N}{p}}{2(s - \frac{N}{r} + \frac{N}{p}) + N} & \text{sinon.} \end{cases} \quad (4.15)$$

— Pour presque toute fonction de $B_r^{s,\infty}([0,1]^N)$, et pour l'estimation par seuillage \hat{f}_n^T

$$\mathbb{E} \|\hat{f}_n^T - f\|_{L^p}^p \approx n^{-\alpha p} \quad (4.16)$$

où

$$\alpha = \begin{cases} \frac{s}{2s+N} & \text{si } r > \frac{pN}{2s+N} \\ \frac{(s - \frac{N}{r} + \frac{N}{p})}{2(s - \frac{N}{r}) + N} & \text{sinon.} \end{cases} \quad (4.17)$$

où $a_n \approx b_n$ signifie que $\frac{\log a_n}{\log b_n} \rightarrow 1$.

Ce résultat, tout comme le théorème de Doob dans le cas paramétrique, ne permet pas de prévoir le comportement de ces estimateurs pour un signal donné. Cependant, il donne une vision générique de ce à quoi on pourrait s'attendre dans les espaces de Besov, largement utilisés en traitement du signal. Il prouve aussi le bien fondé de l'approche minimax dans certains cas puisque celle-ci correspond ici avec des résultats génériques. On constate aussi, comme on pouvait naturellement s'en douter, que dans les espaces de Besov correspondants à des signaux très réguliers (r grand), les estimateurs linéaires sont optimaux alors que pour des signaux parcimonieux, r proche de 1, ils ne le sont plus alors que les estimateurs par seuillage le sont encore.

La démonstration de ce théorème se trouve dans [Fra13] en annexe, et n'est pas rappelée ici mais se base d'une part sur les approches maxisets de [KP02] et d'autre part sur l'étude de la régularité dans les espaces de Besov menée dans [Fra07].

4.5 Détection d'ondes gravitationnelles

Un autre travail plus récent auquel je me suis intéressée est l'analyse des données des interféromètres terrestres LIGO-VIRGO. Ce travail fait l'objet d'un projet ANR en cours avec Eric Chassande-Mottin de l'Université Paris Diderot et est aussi en collaboration avec Stéphane Jaffard de l'Université Paris-Est. Dans le cadre de ce projet, nous nous

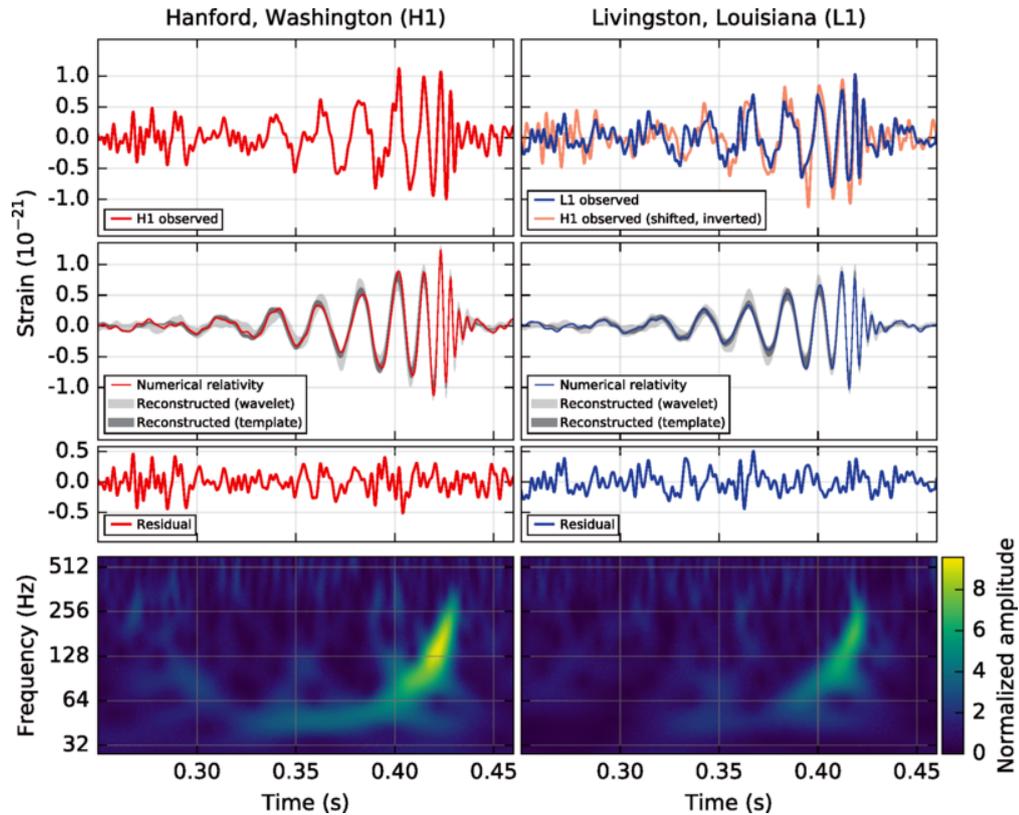


FIGURE 4.1 – Signaux reçus par LIGO en Septembre 2015, image issue de [AAA⁺16]

intéressons aux méthodes d’analyse par ondelettes des signaux reçus par ces interféromètres afin d’améliorer la détection d’ondes gravitationnelles, telles que celles annoncées dans [AAA⁺16].

L’existence d’ondes gravitationnelles a été prédite par la théorie de la relativité générale d’Einstein il y a un siècle et confirmée par des observations en Septembre 2015. Ces ondes sont des déformations de l’espace temps liées à des phénomènes astrophysiques tels que la coalescence de binaires d’étoiles à neutrons ou la formation de trous noirs. Les signaux reçus reportés sur la figure 4.1 sont alors des signaux de type “chirps”, oscillants et dont la fréquence varie au cours du temps. Suivant la masse et la distance des objets entre eux et avec la Terre les caractéristiques de ces signaux vont changer.

Les signaux détectés en 2015 sont issus d’une coalescence de trous noirs ayant une masse importante. L’outil naturel pour l’analyse de tels signaux est la transformation sur des bases temps-fréquence. Une des méthodes d’analyse qui a permis de mettre en avant l’existence de ces ondes gravitationnelles est basée sur l’algorithme Coherent WaveBurst, [KYMM08]. Le principe de cet algorithme est de décomposer un signal sur des bases de Wilson, [DJJ91], de faire un seuillage dans chacune de ces bases puis de reconstruire. Les premiers signaux caractérisés sont issus de la coalescence de trous noirs, donc des événe-

ments ayant une très forte énergie et aisément repérables dans le plan temps-fréquence. D'autres signaux d'amplitude plus faible risquent d'être moins bien repérés par cet algorithme. C'est pourquoi les premières travaux que nous avons menés ont été concentrés sur son amélioration suivant deux directions.

La première piste que nous avons suivie est l'étude de la base considérée. La transformée de Wilson est une transformée temps-fréquence proche de la transformée de Gabor dont on a remplacé l'exponentielle complexe par des fonctions sinus et cosinus. Ainsi si g est une fonction vérifiant certaines conditions d'admissibilité les ondelettes de Wilson correspondantes sont définies pour l et n entiers naturels par :

$$\psi_{l,n} = \begin{cases} g(x - n) & \text{si } l = 0 \\ \sqrt{2} \cos(2\pi lx) g(x - \frac{n}{2}) & \text{si } l > 0 \text{ et } l + n \text{ pair} \\ i\sqrt{2} \sin(2\pi lx) g(x - \frac{n}{2}) & \text{si } l > 0 \text{ et } l + n \text{ impair} \end{cases} \quad (4.18)$$

L'intérêt de cette transformée est qu'elle n'entre pas dans le cadre des hypothèses de Balian-Low. Ainsi, si g vérifie les bonnes propriétés, les fonctions $(\psi_{l,n})$ forment une base de $L^2(\mathbb{R})$. Dans le cadre de Coherent WaveBurst, la fonction g utilisée est l'ondelette de Meyer qui est bien localisée en fréquence mais a l'inconvénient d'être oscillante dans le domaine temporel, ce qui peut conduire à des artefacts dans la représentation temps-fréquence. Aussi avons nous regardé quelle fonction serait mieux adaptée dans le plan temps-fréquence. Nous avons ainsi obtenu qu'en remplaçant cette fenêtre par un sinus itéré défini dans [WW93] nous supprimons les artefacts et obtenons une signature temps-fréquence mieux localisée. L'intérêt de cette fenêtre est qu'elle est encore admissible mais beaucoup moins oscillante en temps.

Une autre partie du travail effectué dans le cadre de ce projet est l'amélioration de l'algorithme au niveau de la synthèse du signal. Pour cela nous avons remplacé les seuillages dans chaque base par un algorithme glouton de type Matching Pursuit, [BM95], sur l'union des bases. Grâce à ces deux évolutions, nous arrivons à améliorer la signature temps-fréquence, tant en termes d'intensité que de pixels détectés, des ondes gravitationnelles.

4.6 Problèmes inverses des ondes

Un autre aspect de l'application des méthodes classiques d'estimation pour des données parcimonieuses auquel je me suis intéressée dans le cadre de la thèse d'Hidayet Zaimaga en collaboration avec Andrea Massa du L2S et Marc Lambert du GeePs est la reconstruction en imagerie des ondes.

Le principe de l'imagerie micro-ondes est d'estimer une cartographie des paramètres électromagnétiques dans une région donnée à partir de la mesure du champ diffracté dû à l'interaction entre une onde incidente connue et un milieu affecté de défauts à identifier. Ce problème revient à étudier la diffusion d'ondes électromagnétiques afin de

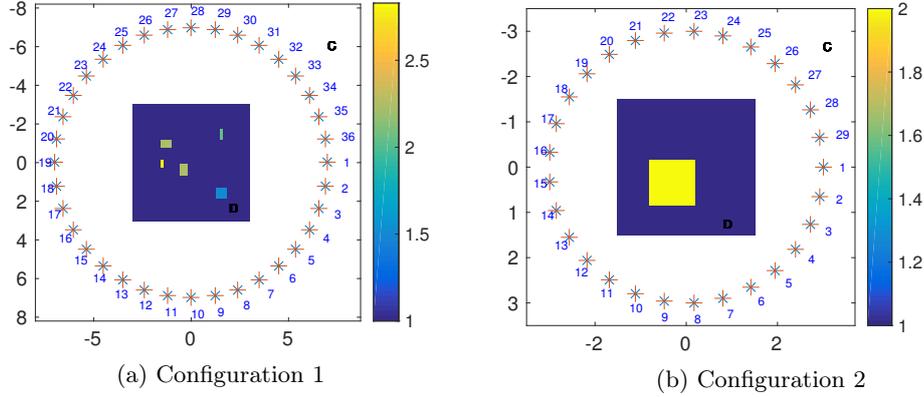


FIGURE 4.2 – Configuration des mesures présentant les profils étudiés ainsi que les positions des émetteurs et récepteurs suivant les axes (x, y) .

reconstruire le champ de contraste correspondant à l'objet d'intérêt. Le développement de méthodes stables et robustes adaptées à ce type d'imagerie est rendu difficile par le caractère mal posé et non linéaire du problème [POM12, DBENP09, RT06]. Des méthodes d'approximation telles que la tomographie par diffraction ou les approximations de Born et de Rytov ont été appliquées afin de linéariser le problème [POM12, DBENP09]. Cependant ces dernières sont limitées à des contrastes faibles et font apparaître des erreurs de reconstruction qui peuvent être importantes sinon. Récemment, des méthodes de représentations parcimonieuses pour l'imagerie micro-ondes ont été développées dans [JXGJ12, WZSL13].

Dans le cadre de cette thèse, nous nous sommes intéressés à la résolution de ce problème inverse utilisant une information de parcimonie, et plus particulièrement de parcimonie structurée. Cette idée de parcimonie structurée a été développée récemment, avec les algorithmes de type Group-Lasso [YL06]. Le principe que nous avons utilisé est de considérer que si les différentes composantes sont parcimonieuses, elles doivent néanmoins avoir un support commun [DYZ13].

Considérons le problème 2D présenté dans la figure 4.2. Chaque source en rouge génère un champ électrique E^{inc} polarisé le long de l'axe z . L'objet d'intérêt est dans un domaine D dont les propriétés diélectriques sont décrites par une fonction de contraste non homogène, $\chi(\mathbf{x}) = (k(\mathbf{x})^2 - k_B^2)$, où k_B est la constante de propagation du milieu considéré. Les positions des N_s sources et N_r récepteurs sont notées respectivement \mathbf{x}_i et \mathbf{x}_r .

Le champ électrique diffracté $E^{\text{diff}}(\mathbf{x}_r, \mathbf{x}_i)$ mesuré en \mathbf{x}_r et dû à la source incidente située en \mathbf{x}_i vérifie [Pas10] :

$$E^{\text{diff}}(\mathbf{x}_r, \mathbf{x}_i) = \int_D G(\mathbf{x}_r, \mathbf{y}) J(\mathbf{y}, \mathbf{x}_i) d\mathbf{y} \quad (4.19)$$

où

$$J(\mathbf{x}, \mathbf{x}_i) = \chi(\mathbf{x})E(\mathbf{x}, \mathbf{x}_i). \quad (4.20)$$

Ici $J(\mathbf{x}, \mathbf{x}_i)$ et $E(\mathbf{x}, \mathbf{x}_i)$ sont la densité de courant équivalente et le champ total, tous deux induits par l'objet d'intérêt. De plus $G(\mathbf{x}, \mathbf{y}) = \frac{-1}{4\omega\epsilon_0} H_0^{(1)}(k_B \|\mathbf{x} - \mathbf{y}\|)$ où $H_0^{(1)}$ est la fonction de Hankel de première espèce. En outre $E(\mathbf{x}, \mathbf{x}_i)$ est obtenu par

$$E(\mathbf{x}, \mathbf{x}_i) = E^{\text{inc}}(\mathbf{x}, \mathbf{x}_i) + \int_D G(\mathbf{x}, \mathbf{y}) J(\mathbf{y}, \mathbf{x}_i) d\mathbf{y} \quad \forall \mathbf{x} \in D. \quad (4.21)$$

En discrétisant le problème par la méthode des moments, [Ric65], et en posant $\mathbf{J}_i^{\text{inc}} = \text{diag}(\boldsymbol{\chi}) \mathbf{E}_i^{\text{inc}}$ pour $i = 1, \dots, N_s$ le problème inverse correspondant devient

$$\boldsymbol{\chi}^* = \arg \min_{\boldsymbol{\chi}} \|\boldsymbol{\zeta}_i - \mathbf{G}_{\text{or}} \text{diag}(\boldsymbol{\chi}) [\mathbf{I} - \text{diag}(\boldsymbol{\chi}) \mathbf{G}_{\text{oo}}]^{-1} \mathbf{J}_i^{\text{inc}}\|_2 \quad (4.22)$$

où $\boldsymbol{\zeta}_i$ est un vecteur de taille N_r qui contient le signal émis par la source $\#i$ et mesuré par les N_r récepteurs. Comme on peut le voir dans (4.22) ce problème est non linéaire et mal posé.

4.6.1 Inversion en deux étapes

Pour résoudre (4.22) nous avons utilisé deux méthodes. La première consiste à prendre une régularisation ℓ^1 pour le problème non-linéaire dans [ZL16]. L'autre est d'utiliser une minimisation alternée et une parcimonie structurée sur les courants équivalents. Dans ce cas, nous cherchons \mathbf{J}_i où $i = 1, \dots, N_s$ puis nous déterminons $\boldsymbol{\chi}$.

Pour déterminer dans un premier temps les courants équivalents nous résolvons

$$\mathbf{J}_i^* = \arg \min_{\mathbf{J}_i} \left[\frac{1}{2} \|\boldsymbol{\zeta}_i - \mathbf{G}_{\text{or}} \mathbf{J}_i\|_2 \right] \quad i = 1, \dots, N_s. \quad (4.23)$$

Dans cette étape nous considérons que \mathbf{J}_i et $\boldsymbol{\chi}$ sont parcimonieux et partagent un même support. Ainsi $\chi^j = 0$ implique que $J_i^j = 0, \forall i = 1, \dots, N_s$. Ici χ^j et J_i^j correspondent à la j -ème composante de chaque vecteur $\boldsymbol{\chi}$, respectivement \mathbf{J}_i . A l'aide de cette hypothèse de parcimonie jointe, on peut réécrire (4.23) comme un problème d'optimisation sous contraintes. En effet, considérer que les \mathbf{J}_i ont le même support quel que soit i peut s'écrire à l'aide d'une norme mixte $\ell_{2,1}$ comme

$$\min_{\mathbf{J}} \|\mathbf{J}\|_{2,1} := \sum_{j=1}^N \|\mathbf{J}^j\|_2 \quad \text{et} \quad \mathbf{G}_{\text{or}} \mathbf{J}_i = \boldsymbol{\zeta}_i. \quad (4.24)$$

Dans notre étude, nous avons utilisé l'algorithme YALL1 [DYZ13] pour résoudre (4.24).

Une fois \mathbf{J}_i connu, nous déterminons la fonction de contraste [AvdB04].

$$\boldsymbol{\chi}^* = \arg \min_{\boldsymbol{\chi}} \sum_{i=1}^{N_s} \|\mathbf{J}_i - \text{diag}(\boldsymbol{\chi}) \mathbf{E}_i\|_2 = \frac{\sum_{i=1}^{N_s} \mathbf{J}_i \cdot \bar{\mathbf{E}}_i}{\sum_{i=1}^{N_s} \mathbf{E}_i \cdot \bar{\mathbf{E}}_i}, \quad (4.25)$$

TABLE 4.1 – Erreur et temps de calcul (en s) de la méthode

	Méthode	Temps	err $_{\chi}$	err $_{\mathbf{E}^{\text{diff}}}$
Fig. 4.2a	Sans parcimonie	367.1	0.17	0.13
	Seuillage doux	131.8	0.14	0.12
	Parcimonie structurée	25.1	0.14	0.20
Fig. 4.2b	Sans parcimonie	64.5	0.25	0.10
	Seuillage doux	16.2	0.06	0.08
	Parcimonie structurée	8.3	0.17	0.57

où \mathbf{E}_i est obtenu en discrétisant (4.21) tandis que $\bar{\mathbf{E}}_i$ est son conjugué.

Nous avons ensuite testé notre méthode sur les deux configurations de la figure 4.2. Nous avons montré que dans le cas réellement parcimonieux de la figure 4.2a cette méthode était plus efficace, tant en temps de calcul et en amélioration de la restauration que la méthode plus directe de [ZL16]. Cependant dans le deuxième cas, figure 4.2b, la qualité de la restauration, en terme d’erreur quadratique moyenne, était moindre comme on peut le voir dans le tableau 4.1.

4.7 Application au JWST

Dans le cadre de la thèse d’Amine Hadj-Youcef co-encadrée avec François Orioux du L2S et Alain Abergel de l’IAS, nous nous intéressons à la reconstruction d’images qui seront acquises par le télescope JWST (James Webb Space Telescope), qui devrait être lancé en 2018. Ce projet se présente dans un cadre collaboratif avec l’IAS qui est lui même fortement impliqué, avec l’Université Paris-Saclay, dans les recherches sur l’instrument infrarouge MIRI (Mid-InfraRed Instrument), [BGML⁺15], de ce télescope. Le but de cette thèse est de construire des algorithmes de reconstruction multi-longueur d’ondes adaptés à MIRI.

Dans ce cadre, nous nous sommes concentrés essentiellement sur la modélisation de l’instrument et la définition d’un modèle direct adapté. Ainsi dans le cas de MIRI, le faisceau de lumière 3D (2 dimensions spatiales et 1 spectrale) à l’entrée du télescope est diffracté et focalisé sur le détecteur. Le résultat est modélisé par un produit de convolution spatial 2D de l’objet avec la réponse impulsionnelle, ou PSF (Point Spread Function) [Goo05]. Cette PSF floute l’objet différemment en fonction de la longueur d’onde. De plus, la distribution spectrale de l’objet disparaît lors de l’intégration sur de larges fenêtres spectrales.

Ainsi si on considère un flux lumineux $\phi(\alpha, \beta, \lambda) : \mathbb{R}^3 \rightarrow \mathbb{R}$ ayant deux dimensions spatiales $(\alpha, \beta) \in \mathbb{R}^2$ et une dimension spectrale $\lambda \in \mathbb{R}^+$ discrétisé en espace sur une base

de pixels, la sortie de l'instrument peut s'écrire

$$y^{(f)}(i, j) = \int_{\mathbb{R}_+} PCE(\lambda) \phi_{i,j}(\lambda) *_{(i,j)} h(i, j, \lambda) d\lambda, \quad (4.26)$$

où PCE correspond à une fonction de la longueur d'onde qui dépend de l'instrument et h est la PSF de l'instrument. La principale difficulté ici est que cette PSF est non stationnaire en longueur d'onde.

Afin de résoudre (4.26) en prenant au maximum en compte l'invariance en longueur d'ondes, nous avons choisi pour chaque pixel un spectre continu, linéaire par morceaux tel qu'illustré sur la figure 4.3. Ce modèle présente l'avantage de permettre une prise en compte de la variabilité en longueur d'ondes ce que ne font pas les méthodes de l'état de l'art telles que [ADGS11, GL10].

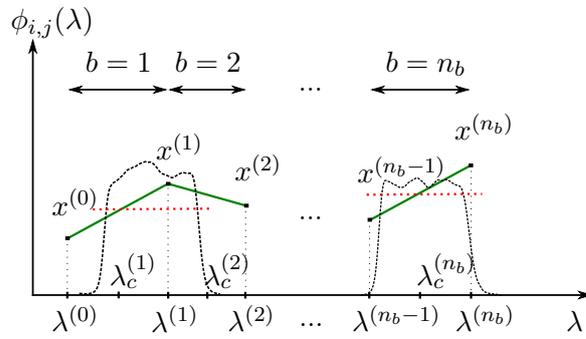


FIGURE 4.3 – Illustration du modèle linéaire par morceaux du spectre de l'objet (en trait continu). Spectre d'objet constant sur la bande spectrale du filtre (en pointillés).

En réarrangeant les termes de l'équation résultante, ceux dépendant de la longueur d'onde restent dans l'intégration spectrale et le problème peut s'écrire comme un problème inverse linéaire

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (4.27)$$

que nous avons résolu par une méthode de moindres carrés régularisés.

Nous avons ensuite testé notre méthode de reconstruction sur un exemple simple. Nous avons donc simulé un cube pour l'objet original de taille $64 \times 64 \times 9$, ayant une source spatialement gaussienne et un spectre linéaire. Ensuite, neuf observations sont simulées pour les neuf bandes de MIRI. Toutes les données sont corrompues avec un bruit blanc gaussien centré avec un rapport signal sur bruit $SNR \approx 30$ dB. Nous voyons dans la figure 4.4 que dans ce cas nous reconstruisons parfaitement la composante spectrale de l'objet original. De plus, pour la reconstruction spatiale nous obtenons une erreur de 6% au lieu des 21% obtenus avec la méthode standard large bande.

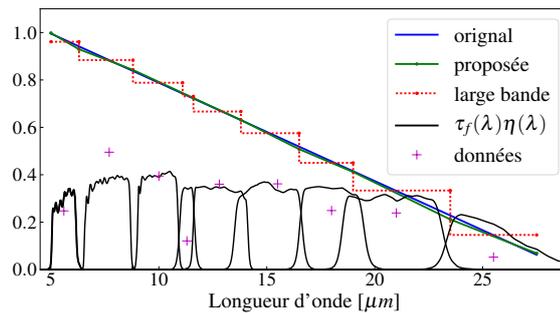


FIGURE 4.4 – Illustration de la reconstruction spectrale pour le pixel central de la source.

4.8 Conclusion

Ce chapitre rassemble un travail théorique sur les performances d'estimateurs non-paramétriques et des apports plus appliqués sur l'utilisation des algorithmes d'optimisation dans le cadre de problèmes inverses. Ces différents travaux peuvent être vus comme différentes études des estimateurs classiques que sont les estimateurs linéaires et par seuillage.

Le premier travail cité ici permet de redorer le blason du minimax dont le caractère pessimiste est souvent décrié. C'est un premier pas vers une nouvelle méthode de comparaison que nous développerons dans le projet. En ce qui concerne les autres projets plus concrets, et plus récents, ce sont essentiellement des travaux en cours qui correspondent à une dynamique axée vers les applications.

BILAN (THÈSE/PUBLICATION)

- 1 projet ANR.
- 1 publication dans une revue internationale à comité de lecture ([J5]).
- 3 publications dans des conférences internationales à comité de lecture ([C1], [C2],[C3]).
- 2 encadrements de stages M2, 1 encadrement de stage M1.
- 2 encadrements de thèse.

5.1	Inférence bayésienne
5.2	Le Bayésien variationnel
5.3	Bayésien variationnel de type gradient
5.4	Bayésien variationnel de type gradient à mémoire
5.5	Application aux problèmes inverses
5.6	Application à la reconstruction d'images par ondelettes
5.7	Conclusion

5 — Méthodologie du bayésien variationnel

Les approches bayésiennes rencontrent depuis [Dem89] un grand succès dans la reconstruction pour les problèmes inverses. Elles ont l'avantage de permettre aisément un réglage automatique des différents paramètres de la méthode tels que le compromis entre les informations qui viennent des données et les informations *a priori*. Dans le cadre bayésien, l'information *a priori* ainsi que la modélisation des paramètres de la méthode sont données par des lois de probabilité. Les estimateurs des sources inconnues sont alors souvent basés sur la loi de probabilité *a posteriori* qui permet de prendre en compte les différents types d'informations. Malheureusement dans la majorité des cas, cette loi n'est pas connue explicitement et ces estimateurs sont par conséquent difficiles à mettre en œuvre directement.

Les méthodes les plus courantes pour obtenir un estimateur bayésien sont les méthodes numériques de type Monte-Carlo par Chaînes de Markov (MCMC). Le principe de ces méthodes est de générer une chaîne d'échantillons qui converge asymptotiquement vers des échantillons sous la vraie loi *a posteriori*. La principale limitation de ces méthodes est leur temps de calcul très important, en particulier à cause du temps de chauffe de la méthode, qui rend leur utilisation en très grande dimension quasiment impossible en général. Un autre type de méthodes, basé sur une approximation analytique de la loi *a posteriori*, est apparu dans les années 90 avec le bayésien variationnel. Le principe est de considérer des approximations de la loi *a posteriori* par des lois plus simples à manipuler, telles que des lois séparables. On cherche donc la loi "simple" la plus proche de la loi cherchée au sens de la divergence de Kullback-Leibler. On se ramène ainsi à un problème d'optimisation convexe. Cette méthode a l'avantage, par rapport aux méthodes MCMC, d'avoir une solution analytique plus simple à déterminer. Ces approches bayésiennes variationnelles permettent de résoudre des problèmes inverses de grande di-

mension de manière non supervisée en cassant les dépendances apparaissant dans l'*a posteriori* entre les hyperparamètres et les paramètres d'intérêts. L'inconvénient de ces méthodes est que, bien que la loi approchante ait une expression analytique, celle-ci n'est pas explicite et nécessite la mise en œuvre d'algorithmes de minimisation alternée. Une fois encore cette solution peut se révéler coûteuse en très grande dimension.

Afin de contourner ce problème j'ai développé dans [FR14], en collaboration avec Thomas Rodet de l'ENS Paris Saclay, un algorithme bayésien variationnel, basé sur de l'optimisation dans un espace non vectoriel de dimension infinie. Dans un second temps dans la thèse de Yuling Zheng, et plus particulièrement dans [ZFR15a] nous avons utilisé d'autres algorithmes d'optimisation, les méthodes par sous-espaces, qui généralisent la descente de gradient. Ces méthodes ont été transposées à la structure de l'ensemble des densités de probabilités généralisant ainsi le gradient exponentialisé que nous avons utilisé. Grâce à cette généralisation, le temps de calcul de notre méthode a été largement amélioré. Nous avons appliqué ces différentes méthodes à la reconstruction d'images en utilisant différents *a priori*, dans le domaine direct et dans le domaine des ondelettes. Je ne présenterai ici que cette dernière application.

5.1 Inférence bayésienne

L'inférence bayésienne est basée sur la modélisation probabiliste du vecteur des paramètres inconnus \mathbf{x} et du vecteur du bruit \mathbf{n} . Les informations apportées par les mesures ainsi que les connaissances *a priori* sur l'objet cherché sont toutes modélisées sous forme de densités de probabilité. L'inférence bayésienne s'appuie principalement sur la règle de Bayes qui nous permet d'obtenir une loi *a posteriori* des paramètres inconnus \mathbf{x} sachant les mesures \mathbf{y} . Cette loi résume toute l'information disponible sur les paramètres inconnus et permet la mise en œuvre d'estimateurs tels que l'espérance *a posteriori* ou le maximum *a posteriori*. Dans les travaux qui sont développés ici nous avons utilisé comme estimateur l'espérance *a posteriori* qui minimise l'erreur quadratique moyenne, [Idi01].

En utilisant la règle de Bayes, la loi *a posteriori* est donnée par

$$p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1)p(\mathbf{x}|\boldsymbol{\theta}_2)}{p(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} \quad (5.1)$$

où $\boldsymbol{\theta}_1$ et $\boldsymbol{\theta}_2$ représentent des paramètres liés à la modélisation du problème et au compromis entre les deux sources d'information, celle venant des données et celle venant de l'*a priori*. Ils sont souvent appelés hyperparamètres dans la littérature. On note $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ l'ensemble de ces hyperparamètres. Par ailleurs, $p(\mathbf{y}|\mathbf{x})$ est la vraisemblance qui rend compte des incertitudes sur les données observées. Cette probabilité est obtenue à partir du modèle direct et de la modélisation probabiliste du bruit tandis que $p(\mathbf{x})$ est la loi *a priori* qui représente les connaissances *a priori* sur l'objet. Dans ce cas, $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}$ est la loi marginale qui assure la normalisation de la loi *a posteriori*. Elle est connue aussi sous le nom de *fonction de partition* dans la littérature

de physique statistique et sous le nom d'*évidence du modèle* lorsqu'elle est utilisée pour la sélection de modèle. Cette loi marginale n'est généralement pas explicitement connue.

La vraisemblance se déduit de la modélisation du problème direct et d'une modélisation statistique de la composante de bruit \mathbf{n} . Dans ce travail, on n'étudiera que le bruit gaussien. Supposons donc que le bruit est un bruit gaussien de moyenne nulle et de matrice de covariance définie positive Σ . Selon l'hypothèse du bruit et le modèle direct (4.1), la loi conditionnelle de \mathbf{y} sachant \mathbf{x} , qui est appelée la *vraisemblance*, est de la forme suivante :

$$p(\mathbf{y}|\mathbf{x}, \Sigma) = (2\pi)^{-M/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{Ax})^T \Sigma^{-1}(\mathbf{y} - \mathbf{Ax})\right). \quad (5.2)$$

Si on suppose en plus que les composantes du vecteur de bruit sont indépendantes et identiquement distribuées (*i.i.d.*), la matrice de covariance Σ devient diagonale et ses composantes diagonales sont donc égales à une même valeur, notée ici σ_n^2 . En conséquence, la vraisemblance devient

$$p(\mathbf{y}|\mathbf{x}, \sigma_n^2) \propto \exp\left(-\frac{\|\mathbf{y} - \mathbf{Ax}\|^2}{2\sigma_n^2}\right), \quad (5.3)$$

où \propto signifie " est proportionnel à ".

5.1.1 Loi *a priori*

Dans le cadre bayésien, l'information *a priori* sur l'objet cherché prend aussi la forme d'une probabilité :

$$p(\mathbf{x}) \propto \exp(-\Phi(\mathbf{x})). \quad (5.4)$$

où $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}$ est une fonction dont les propriétés restent à définir. Ainsi cet *a priori* p est une version probabiliste du terme de régularisation que nous avons vu dans le chapitre précédent. Le choix de la loi *a priori* est un problème non-trivial et lié au problème à résoudre. Dans la littérature, de nombreuses possibilités ont été proposées. La loi *a priori* gaussienne a été premièrement introduite et largement utilisée [Tik63, Dem89]. Néanmoins, sa capacité à reconstruire des composantes non-linéaires, tels que l'emplacement et la grandeur des discontinuités, est limitée. Pour pouvoir décrire les propriétés non-linéaires de l'objet à chercher, les champs aléatoires de Markov composés [JW91], les champs aléatoires de Markov avec des fonctions de potentiel non-quadratiques telles que la valeur absolue (une loi de Laplace) [CDS01], et les distributions à queue lourde, par exemple, la loi de Student, ont été aussi utilisés. Par ailleurs, dans diverses applications du traitement d'image ou traitement du signal, une caractéristique présente souvent dans notre objet d'intérêt est la régularité par morceau. Pour cela, un *a priori* de variation totale [ROF92] est souvent utilisé. Dans la suite de la présentation nous nous concentrons sur des *a priori* de type MSG, définis dans 3.3.1. Néanmoins dans la thèse de Yuling Zheng, nous avons étudié d'autres lois comme celle basée sur la variation totale.

L'un des intérêts majeurs de l'inférence bayésienne est qu'elle permet d'estimer les hyperparamètres en même temps que la source inconnue et de définir ainsi des méthodes dites non-supervisées.

5.2 Le Bayésien variationnel

Comme évoqué précédemment, le principal inconvénient des estimateurs bayésiens, et plus particulièrement de la moyenne *a posteriori* est que dans la majorité des cas la fonction de partition n'est pas calculable explicitement. La loi *a posteriori* n'est pas totalement connue. Afin de remédier à ce travers des approximations sont nécessaires. Ces approximations peuvent être numériques, c'est le cas des méthodes MCMC, ou analytiques comme dans le cas du Bayésien variationnel (BV) défini dans [Mac95]. A l'heure actuelle, les approximations analytiques sont les plus efficaces en grande dimension et ce sont sur celles-ci que nous nous sommes concentrés.

Dans la suite, nous utilisons \mathbf{w} pour représenter un vecteur de taille J regroupant toutes les variables à estimer, autrement dit $\mathbf{w} = (\mathbf{x}, \boldsymbol{\theta})$. La loi *a posteriori* à approcher est représentée par $p(\cdot|\mathbf{y})$ et son approximation analytique est donnée par q .

5.2.1 Méthode classique

La méthode d'approximation bayésienne variationnelle provient de la physique statistique et plus particulièrement de la théorie du champ moyen [Par88]. Le principe est d'approcher la loi d'intérêt par une fonction séparable qui est plus facile à étudier. Une telle fonction séparable optimale peut être choisie en minimisant une mesure de dissemblance avec la probabilité originale compliquée. L'approximation en champ moyen exploite différentes fonctions de dissemblance tandis que l'approche bayésienne variationnelle se base pour de telles approximations uniquement sur la divergence de *Kullback-Leibler*.

La méthode d'approximation BV a été introduite dans le cadre bayésien par MacKay [Mac95]. Elle a depuis été largement utilisée dans diverses applications telles que l'apprentissage des modèles graphiques [JGJS99], la restauration d'images [LG04, CGLS08], la séparation de sources [Cho02] et la super-résolution [BMK11].

5.2.2 Principe

L'idée du bayésien variationnel est d'approcher la loi *a posteriori* cible $p(\cdot|\mathbf{y})$ par une loi séparable q^{opt} . Cette loi approchante doit être la plus proche possible de la loi cible au sens où elle doit minimiser la divergence de Kullback-Leibler entre une loi approchante q et la loi $p(\cdot|\mathbf{y})$.

Définissons Ω l'ensemble des densités de probabilité séparables,

$$\Omega = \{q : \text{densité de probabilité et } q(\mathbf{w}) = \prod_{i=1}^P q_i(\mathbf{w}_i) \text{ avec } \mathbf{w} = (\mathbf{w}_1, \dots, \mathbf{w}_P)\}, \quad (5.5)$$

où $(\mathbf{w}_i)_{i=1,\dots,P}$ représentent les sous-ensembles de composantes de \mathbf{w} , $1 \leq P \leq J$ étant le degré de séparabilité considéré. Le problème de recherche de la loi approchante optimale est un problème d'optimisation fonctionnelle dans l'ensemble Ω :

$$q^{opt} = \arg \min_{q \in \Omega} \mathcal{KL}(q \| p(\cdot | \mathbf{y})). \quad (5.6)$$

Néanmoins, la divergence \mathcal{KL} n'est pas numériquement calculable car elle dépend de la loi *a posteriori* qui n'est pas connue. Mais on peut développer l'expression de la divergence :

$$\begin{aligned} \mathcal{KL}[q \| p(\cdot | \mathbf{y})] &= \int_{\mathbb{R}^J} q(\mathbf{w}) \ln \frac{q(\mathbf{w})}{p(\mathbf{w} | \mathbf{y})} d\mathbf{w} \\ &= \int_{\mathbb{R}^J} q(\mathbf{w}) \ln \frac{q(\mathbf{w})p(\mathbf{y})}{p(\mathbf{w}, \mathbf{y})} d\mathbf{w} \\ &= \ln p(\mathbf{y}) - \int_{\mathbb{R}^J} q(\mathbf{w}) \ln \frac{p(\mathbf{w}, \mathbf{y})}{q(\mathbf{w})} d\mathbf{w} \\ &= \ln p(\mathbf{y}) - \mathcal{F}(q) \end{aligned} \quad (5.7)$$

où $p(\mathbf{y})$ est l'évidence du modèle et

$$\mathcal{F}(q) = \int_{\mathbb{R}^J} q(\mathbf{w}) \ln \frac{p(\mathbf{w}, \mathbf{y})}{q(\mathbf{w})} d\mathbf{w} \quad (5.8)$$

est appelée l'énergie libre négative. Ici, la log-évidence $\ln p(\mathbf{y})$ est constante par rapport à q . Par conséquent la minimisation de la divergence \mathcal{KL} est équivalente à la maximisation de l'énergie libre négative $\mathcal{F}(q)$. Nous voyons que cette énergie dépend de la loi jointe $p(\mathbf{y}, \mathbf{w})$ qui peut être aisément obtenue par le produit de la vraisemblance et de la loi *a priori*. En pratique nous utilisons donc plutôt l'énergie libre négative $\mathcal{F}(q)$ au lieu de la divergence de Kullback-Leibler. Ainsi le problème d'optimisation résolu dans le bayésien variationnel est :

$$q^{opt} = \arg \max_{q \in \Omega} \mathcal{F}(q), \quad (5.9)$$

avec \mathcal{F} l'énergie libre négative définie dans (5.8). La divergence de Kullback-Leibler étant convexe, \mathcal{F} est naturellement une fonctionnelle concave.

5.2.3 L'algorithme

Le problème d'optimisation (5.9) admet une solution analytique qui est donnée par $q^{opt}(\mathbf{w}) = \prod_{i=1}^P q_i(\mathbf{w}_i)$ avec

$$q_i(\mathbf{w}_i) = K_i \exp \left(\langle \ln p(\mathbf{y}, \mathbf{w}) \rangle_{\prod_{j \neq i} q_j(\mathbf{w}_j)} \right), \quad \forall i = 1, \dots, P, \quad (5.10)$$

où $\langle \cdot \rangle_q = \mathbb{E}_q[\cdot]$ et K_i représente la constante de normalisation. Cette solution analytique a été trouvée de plusieurs manières différentes dans [ŠQ06] ou [Mis00]. Les approches d'approximation bayésienne variationnelle classiques sont basées sur cette solution analytique.

Néanmoins, nous pouvons voir dans (5.10) que chaque composante séparable q_i dépend de toutes les autres composantes q_j pour j différent de i . Par conséquent, on ne peut pas obtenir une expression explicite pour q^{opt} . En pratique, ce problème est généralement abordé en utilisant un algorithme alterné, qui met à jour une composante séparable en fixant les autres à chaque étape. On peut résumer l'algorithme d'approximation bayésien variationnel classique dans l'Algorithme 1.

Algorithm 1 Algorithme Bayésien Variationnel classique (BVClass)

- A) Initialiser ($q^0 \in \Omega$)
- B) Mettre à jour $q_1^{k+1}(\mathbf{w}_1)$
- C) Mettre à jour $q_2^{k+1}(\mathbf{w}_2)$
- D) ...
- E) Mettre à jour $q_P^{k+1}(\mathbf{w}_P)$ avec

$$q_i^{k+1}(\mathbf{w}_i) = K_i \exp \left(\langle \ln p(\mathbf{y}, \mathbf{w}) \rangle_{(\prod_{j=1}^{i-1} q_j^{k+1}(\mathbf{w}_j)) (\prod_{j=i+1}^P q_j^k(\mathbf{w}_j))} \right), \forall i = 1, \dots, P. \quad (5.11)$$

- F) Retourner à l'étape 2 jusqu'à convergence
-

Théoriquement, l'algorithme 1 peut être utilisé pour approcher n'importe quelle loi *a posteriori*. Néanmoins, en pratique, la faisabilité d'un point de vue numérique impose d'utiliser des lois conjuguées. En effet, les lois *a priori* conjuguées assurent que la loi *a posteriori* appartient à la même famille paramétrique que la loi *a priori*. Les lois approchantes sont alors choisies dans cette famille et la mise à jour de (5.11) revient à une mise à jour des paramètres.

5.2.4 Choix de la séparabilité

Le principe du bayésien variationnel est basé sur une approximation séparable des densités. L'inconvénient de cette hypothèse de séparabilité est qu'elle enlève la corrélation sur chaque ensemble de paramètres $(\mathbf{w}_i)_{i=1, \dots, P}$. En général, plus le degré de séparabilité utilisé est grand, plus la loi approchante est potentiellement éloignée de la loi cible. En pratique, il n'y a pas de règles pour le choix de la séparabilité et on peut choisir la séparabilité selon le problème. Notons qu'une séparation partielle, où $P < J$ permettra de garder des corrélations entre les composantes mais fera intervenir des matrices de covariance potentiellement importantes. Au contraire, la séparation totale où $P = J$ fait perdre ces corrélations mais facilite le calcul des lois approchantes. En effet dans ce cas le calcul matriciel est remplacé par un grand nombre de mises à jour uni-dimensionnelles. Cependant dans ce cas le caractère alterné de l'algorithme peut entraîner un temps de calcul important, surtout en grande dimension.

En résumé, même en ne considérant que des lois *a priori* conjuguées, l'algorithme d'approximation BV classique est limité en termes de convergence et ce quelle que soit

la séparabilité considérée. Par conséquent, le développement d'algorithmes accélérés est nécessaire pour une utilisation en très grande dimension.

5.3 Bayésien variationnel de type gradient

Pour remédier à ce défaut du bayésien variationnel classique, nous avons proposé dans [FR14] de construire un algorithme de descente de gradient adapté. L'idée est donc de transposer l'algorithme classique, défini dans un espace de Hilbert, à l'ensemble des densités de probabilités. Comme détaillé dans la partie précédente, les approches BV classiques sont basées sur une solution analytique mais non-explicite du problème d'optimisation fonctionnelle (5.9), donnée par (5.10). L'idée est alors d'approcher cette solution non-explicite à l'aide de méthodes itératives de type Gauss-Seidel. La méthode proposée s'appuie, elle, sur des équations de mise à jour explicites pour chaque composante séparable, permettant ainsi une mise à jour en parallèle et donc une accélération significative des approches BV classiques.

L'approximation BV a pour but de trouver une distribution séparable optimale q^{opt} pour approcher la loi *a posteriori* $p(\cdot|\mathbf{y})$ en maximisant l'énergie libre négative $\mathcal{F}(q)$ donnée par (5.8). Ce problème est un problème d'optimisation fonctionnelle dans un ensemble qui n'a pas de structure vectorielle. Le point de vue adopté dans [FR14] est de considérer le problème équivalent dans un espace de mesures de probabilité séparables $\mathcal{A} = \bigotimes_{i=1}^P \mathcal{A}_i$, le produit cartésien des \mathcal{A}_i où

$$\mathcal{A}_i = \{\mu_i : \text{mesure de probabilité} \\ \text{et } \mu_i(d\mathbf{w}_i) = q_i(\mathbf{w}_i)d\mathbf{w}_i \text{ avec } q_i \text{ une densité de probabilité}\}.$$

L'espace \mathcal{A} peut être considéré comme un sous-ensemble de l'espace des mesures de Radon signées \mathcal{M} doté de la norme de la variation totale, qui est un espace de Banach. Par conséquent, l'optimisation de (5.9) est équivalente à la résolution du problème d'optimisation suivant :

$$\mu^{opt} = \arg \max_{\mu \in \mathcal{A}} F(\mu), \quad (5.12)$$

où la fonctionnelle $F : \mathcal{M} \rightarrow \mathbb{R}$ satisfait que $\forall \mu \in \mathcal{A}$ ayant une densité q , $F(\mu) = \mathcal{F}(q)$. Ce problème d'optimisation est un problème d'optimisation sous-contraintes dans un espace de Banach de dimension infinie \mathcal{M} .

Comme \mathcal{M} est un espace de Banach, on peut calculer la différentielle de Gâteaux de F qui s'écrit comme

$$\forall \nu \in \mathcal{M}, \quad \partial F_\mu(\nu) = \lim_{t \rightarrow 0} \frac{F(\mu + t\nu) - F(\mu)}{t}. \quad (5.13)$$

Dans notre cas on peut trouver une fonction continue et bornée supérieurement $\partial f : \mathcal{M} \times \mathbb{R}^P \rightarrow \mathbb{R}$ de sorte que la différentielle de Gâteaux soit donnée par

$$\forall \nu \in \mathcal{M}, \quad \partial F_\mu(\nu) = \int_{\mathbb{R}^P} \partial f(\mu, \mathbf{w}) d\nu(\mathbf{w}). \quad (5.14)$$

Dans le cas où $\mu \in \mathcal{A}$, nous avons alors $\partial f(\mu, \mathbf{w}) = df(q, \mathbf{w})$ où $df : \Omega \times \mathbb{R}^P \rightarrow \mathbb{R}$. Par ailleurs, comme nous considérons q séparable, $df(q, \mathbf{w}) = \sum_i df(q_i, \mathbf{w}_i)$, où $\forall i = 1, \dots, P$ $df(q_i, \mathbf{w}_i)$ est exprimée par

$$df(q_i, \mathbf{w}_i) = \langle \log p(\mathbf{y}, \mathbf{w}) \rangle_{\prod_{j \neq i} q_j(\mathbf{w}_j)} - \log q_i(\mathbf{w}_i) - 1. \quad (5.15)$$

L'algorithme bayésien variationnel proposé par [FR14] est un algorithme itératif qui fournit une densité à chaque itération et approche la solution de (5.12) progressivement. Il s'appuie sur le théorème de Radon-Nikodym [Rud87a].

Soit $k \in \mathbb{N}$, initialement fixé à zéro, l'indice d'itération et supposons que μ^k est une mesure de probabilité ayant une densité q^k , c'est-à-dire telle que

$$d\mu^k(\mathbf{w}) = q^k(\mathbf{w})d\mathbf{w}. \quad (5.16)$$

Comme nous sommes dans l'espace des mesures de probabilité à densité, la prochaine itération devrait donner également une mesure de probabilité absolument continue par rapport à μ^k . Le théorème de Radon-Nikodym assure que cette mesure devrait être écrite comme

$$d\mu^{k+1}(\mathbf{w}) = h^k(\mathbf{w})d\mu^k(\mathbf{w}), \quad (5.17)$$

où $h^k \in L^1(\mu^k)$ est une fonction positive. De manière équivalente, nous pouvons écrire que

$$q^{k+1}(\mathbf{w}) = h^k(\mathbf{w})q^k(\mathbf{w}), \quad (5.18)$$

est l'équation de mise à jour pour la densité approchante. En ce qui concerne la fonction h^k , suivant le schéma itératif donné par la méthode du gradient :

$$\begin{aligned} h^k(\mathbf{w}) &= K^k(\alpha^k) \exp \left[\alpha^k \partial f(\mu^k, \mathbf{w}) \right] \\ &= K^k(\alpha^k) \exp \left[\alpha^k df(q^k, \mathbf{w}) \right] \end{aligned} \quad (5.19)$$

où $\alpha^k > 0$ est le pas de l'algorithme et $df(q^k, \mathbf{w})$ est définie par (5.15). Par ailleurs, $K^k(\alpha^k)$ représente la constante de normalisation exprimée par

$$K^k(\alpha^k) = \left[\int_{\mathbb{R}^J} \exp \left[\alpha^k df(q^k, \mathbf{w}) \right] q^k(\mathbf{w})d\mathbf{w} \right]^{-1}. \quad (5.20)$$

Ce choix pour la fonction h^k (5.19) est motivé par l'hypothèse de positivité et d'intégrabilité ainsi que par la cohérence avec la structure de la méthode du gradient. Ce gradient exponentialisé est d'ailleurs déjà connu dans le cas discret [KW97a]. Une autre façon de présenter cette mise à jour est celle de [TB93] qui montre comment elle peut être vue comme le dual d'un algorithme proximal basé sur une pénalité d'entropie.

L'algorithme d'approximation proposé est donc basé sur l'équation itérative donnée par (5.18) et (5.19). Concernant le pas de l'algorithme, on définit tout d'abord une fonction de α :

$$g^k : \mathbb{R} \rightarrow \mathbb{R}$$

$$g^k(\alpha) = \mathcal{F} \left(K^k(\alpha) q^k(\mathbf{w}) \exp \left[\alpha df(q^k, \mathbf{w}) \right] \right), \quad (5.21)$$

ce qui nous permet de définir le pas optimal par :

$$\alpha^{opt} = \arg \max_{\alpha \in \mathbb{R}} g^k(\alpha). \quad (5.22)$$

En pratique nous avons remplacé ce pas optimal par un pas sous-optimal calculé à partir du développement de Taylor à l'ordre 2 de g^k ou par une méthode de "backtracking", [NW00], quand ce dernier est trop grand.

L'algorithme défini dans [FR14] peut finalement se résumer comme suit.

Algorithme 5.3.1 — Bayésien Variationnel Gradient exponentialisé (BV-Grad).

[O1] Initialiser ($q^0 \in \Omega$)

[O2] Répéter

a. déterminer $df(q^k, \mathbf{w})$

b. déterminer le pas de l'algorithme α^{subopt}

c. calculer $q^{k+1}(\mathbf{w}) = K^k(\alpha^{subopt}) \exp[\alpha^{subopt} df(q^k, \mathbf{w})]$

jusqu'à convergence

De plus nous avons aussi démontré le théorème de convergence suivant.

Résultat 5.3.1 Soit F une fonctionnelle concave sur \mathcal{M} vérifiant :

- A) F peut s'écrire sous la forme $F = G + H$ où G est une fonctionnelle concave et L -Lipschitz Fréchet-différentiable tandis que H correspond à l'entropie de la mesure considérée.
- B) Pour $\mu \in \mathcal{M}$, G est deux fois différentiable au sens de Fréchet en μ et sa différentielle première vérifie

$$\forall \nu \in \mathcal{M}, \quad \partial F_\mu(\nu) = \int_{\mathbb{R}^N} df(\mu, \mathbf{x}) d\nu(\mathbf{x}). \quad (5.23)$$

- C) $\lim_{\|\mu\| \rightarrow \infty} F(\mu) = -\infty$.

Soit, pour tout $k \geq 0$, α_k le pas obtenu par (5.22). La suite $(\mu^k)_{k \geq 0}$ définie par $\mu^{k+1} = K_k(\alpha_k) \exp(\alpha_k df(\mu^k, \cdot)) \mu^k$ converge vers un maximum de F sur \mathcal{A} .

La démonstration de ce résultat se trouve dans [FR14], présenté en annexe. Sans entrer dans les détails techniques, je vais en présenter ici les grandes lignes.

Démonstration :

Tout d'abord, il est à noter que l'ensemble \mathcal{A} défini plus haut n'est pas un ensemble convexe. Ainsi le problème (5.12) peut ne pas avoir de solution unique. Afin de contourner ce problème, nous considérons plutôt dans la suite l'ensemble Θ qui est l'enveloppe convexe fermée de \mathcal{A} . Dans ce cas, le problème plus général

$$\mu^{opt} = \arg \max_{\mu \in \Theta} F(\mu) \quad (5.24)$$

est un problème d'optimisation convexe sous contraintes dans un espace de Banach de dimension infinie. Afin de résoudre ce problème nous avons procédé en deux étapes. La première consiste à montrer que la suite des itérées $(F(\mu^k))_{k \in \mathbb{N}}$ est une suite croissante et bornée dans \mathbb{R} . Ce qui signifie que la suite $F(\mu^{k+1}) - F(\mu^k)$ converge vers zéro. Cette partie de la démonstration est technique et n'est pas présentée ici. Dans un second temps nous avons montré que la suite $(\mu^k)_{k \in \mathbb{N}}$ converge vers un maximum de F dans le cas d'un pas α optimal.

Pour cela, nous nous sommes d'abord appuyés sur le théorème de Banach-Alaoglu, voir [Rud87b], qui dit qu'un ensemble fermé et borné est compact pour la topologie faible*. Ainsi dans notre cas il existe $\mu^{lim} \in \mathcal{M}$ et une sous-suite $(\mu^{k_n})_{n \in \mathbb{N}}$ telle que pour n'importe quelle fonction f qui tend vers zéro à l'infini :

$$\int f(\mathbf{x}) \mu^{k_n}(d\mathbf{x}) \rightarrow \int f(\mathbf{x}) \mu^{lim}(d\mathbf{x}).$$

De plus, comme nous avons choisi le pas optimal pour α nous avons donc que

$$F(\mu^{k+1}) = g_k(\alpha_{opt}) \geq g_k(\alpha), \quad \forall \alpha > 0,$$

où pour chaque $k \in \mathbb{N}$ nous avons défini $g_k(\alpha)$ par (5.21). Au lieu de travailler directement avec ce pas optimal α_{opt} qui n'a pas de forme analytique, nous avons considéré le pas sous-optimal obtenu en considérant le développement de Taylor à l'ordre 2 de g_k . Ainsi $\alpha_{subopt} = -\frac{g'_k(0)}{g''_k(0)}$, dès que $g''_k(0) \neq 0$ existe et

$$g_k(\alpha_{subopt}) = g_k(0) - \frac{g'_k(0)^2}{2g''_k(0)} + \alpha_{subopt}^2 \varepsilon(\alpha_{subopt}), \quad (5.25)$$

ce qui donne pour μ^{k+1} ,

$$\begin{aligned} F(\mu^{k+1}) &\geq g_k(\alpha_{subopt}) = g_k(0) - \frac{g'_k(0)^2}{2g''_k(0)} + \alpha_{subopt}^2 \varepsilon(\alpha_{subopt}) \\ &= F(\mu^k) - \frac{g'_k(0)^2}{2g''_k(0)} + \alpha_{subopt}^2 \varepsilon(\alpha_{subopt}). \end{aligned}$$

Comme $F(\mu^{k+1}) - F(\mu^k) \rightarrow 0$, cela implique que $\lim_{k \rightarrow \infty} \frac{g'_k(0)^2}{2g''_k(0)} = 0$. Si on considère une sous-suite convergente $(k_n)_{n \in \mathbb{N}}$ et que l'on définit la suite $(\gamma_{k_n})_{n \in \mathbb{N}}$ par $\forall n \in \mathbb{N}$, $\gamma_{k_n} = -\frac{g'_{k_n}(0)}{g''_{k_n}(0)}$ cette convergence implique :

$$\begin{aligned} g_{k_n}(\gamma_{k_n}) - g_{k_n}(0) &= \gamma_{k_n} g'_{k_n}(0) + \frac{\gamma_{k_n}^2}{2} g''_{k_n}(0) + \gamma_{k_n}^2 \varepsilon(\gamma_{k_n}) \\ &= \frac{1}{2} \gamma_{k_n} g'_{k_n}(0) + \gamma_{k_n}^2 \varepsilon(\gamma_{k_n}). \end{aligned}$$

Ce qui n'est possible que si $g'_{k_n}(0) \rightarrow 0$ quand $n \rightarrow \infty$.

En calculant cette dérivée g'_k , nous obtenons finalement que

$$\begin{aligned} g'_k(0) &= \partial F_{\mu^k}(\tilde{g}'(0)\mu^k) = \int_{\mathbb{R}^N} df(\mu^k, \mathbf{x}) \left(df(\mu^k, \mathbf{x}) - \int_{\mathbb{R}^N} df(\mu^k, \mathbf{y}) d\mu^k(\mathbf{y}) \right) d\mu^k(\mathbf{x}) \\ &= \int_{\mathbb{R}^N} df(\mu^k, \mathbf{x})^2 \mu^k(d\mathbf{x}) - \left(\int_{\mathbb{R}^N} df(\mu^k, \mathbf{y}) d\mu^k(\mathbf{y}) \right)^2 \end{aligned} \quad (5.26)$$

Autrement dit pour n assez grand

$$\|df(\mu^{k_n}, \cdot)\|_{L^2(\mu^{k_n})}^2 - \|df(\mu^{k_n}, \cdot)\|_{L^1(\mu^{k_n})}^2 \rightarrow 0,$$

et $df(\mu^{k_n}, \cdot)$ converge vers une constante λ , indépendante de $\mathbf{x} \in \mathbb{R}^N$.

Soit ν un élément quelconque de Θ , alors $\partial F_{\mu^{lim}}(\nu - \mu^{lim}) = 0$ et la concavité de F donne

$$F(\nu) \leq F(\mu^{lim}) \quad \forall \nu \in \Theta.$$

□

Comme nous le verrons dans la partie 5.5.2, l'intérêt majeur de cet algorithme est la forte accélération du bayésien variationnel avec séparabilité totale et son adaptation à des dimensions importantes. Cependant, cet algorithme se base sur une méthode simple d'optimisation, la descente de gradient.

5.4 Bayésien variationnel de type gradient à mémoire

Dans la thèse de Yuling Zheng et dans [ZFR15a], le premier travail qui a été fait était de déterminer une méthode d'optimisation plus rapide que le gradient et qui puisse s'adapter au gradient exponentialisé. Pour cela, nous avons regardé les méthodes par

sous-espaces. Celles-ci généralisent en effet beaucoup de méthodes de descente à pas optimaux existantes.

Le principe des méthodes par sous-espaces est de chercher la direction de descente dans un sous-espace de dimension I au lieu de privilégier une unique direction. Ainsi, dans le cas de l'optimisation dans un espace de Hilbert on considère

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}^k = \mathbf{x}^k + \mathbf{D}^k \mathbf{s}^k, \quad (5.27)$$

où \mathbf{x}^k et \mathbf{x}^{k+1} représentent respectivement l'estimation à la k ème et à la $(k+1)$ ème iteration, $\mathbf{d}^k = \mathbf{x}^{k+1} - \mathbf{x}^k$ est la variation du vecteur \mathbf{x} correspondant à la donnée des I directions $\mathbf{D}^k = [\delta_1^k, \dots, \delta_I^k]$ et du vecteur $\mathbf{s}^k = [s_1^k, \dots, s_I^k]^T$ correspondant aux pas suivant chaque direction $\{\delta_i^k\}_{i=1, \dots, I}$. La méthode itérative de sous-espace offre ainsi une plus grande flexibilité que les méthodes de descente ayant une direction imposée. Plusieurs ensembles de directions ont été étudiés. Ainsi dans [MC69] ils considèrent un sous-espace engendré par l'opposé du gradient et la direction obtenue à l'itération précédente, sous le nom de *mémoire de gradient* (MG), généralisé dans [CL69], tandis que dans [SS05] le sous-espace considéré est constitué de gradients aux itérations précédentes.

Dans la thèse de de Yuling Zheng ainsi que dans [ZFR15a], présenté en annexe, nous avons donc adapté ces méthodes d'optimisation au cadre donné par la partie précédente. Pour cela, nous reprenons le même schéma

$$q^{k+1}(\mathbf{w}) = h^k(\mathbf{w})q^k(\mathbf{w}), \quad (5.28)$$

mais cette fois-ci h^k est donnée par

$$h^k(\mathbf{w}) = K^k(\mathbf{s}^k) \exp \left[\mathbf{D}^k(\mathbf{w}) \mathbf{s}^k \right], \quad (5.29)$$

où $K^k(\mathbf{s}^k)$ représente la constante de normalisation qui s'écrit

$$K^k(\mathbf{s}^k) = \left[\int_{\mathbb{R}^J} \exp \left[\mathbf{D}^k(\mathbf{w}) \mathbf{s}^k \right] q^k(\mathbf{w}) d\mathbf{w} \right]^{-1}, \quad (5.30)$$

et $\mathbf{D}^k(\mathbf{w}) = [\delta_1^k(\mathbf{w}), \dots, \delta_I^k(\mathbf{w})]$ est l'ensemble des I directions qui engendrent le sous-espace. Grâce à la forme exponentielle, (5.29) peut être aussi écrite :

$$h^k(\mathbf{w}) = K^k(\mathbf{s}^k) \left[\phi_1^k(\mathbf{w}) \right]^{s_1^k} \dots \left[\phi_I^k(\mathbf{w}) \right]^{s_I^k} \quad (5.31)$$

où $\phi_l^k(\mathbf{w}) = \exp[\delta_l^k(\mathbf{w})]$, pour $l = 1, \dots, I$, ce qui donne comme équation de mise à jour :

$$q^{k+1}(\mathbf{w}) = K^k(\mathbf{s}^k) q^k(\mathbf{w}) \left[\phi_1^k(\mathbf{w}) \right]^{s_1^k} \dots \left[\phi_I^k(\mathbf{w}) \right]^{s_I^k}. \quad (5.32)$$

5.4.1 Sous-espaces considérés

Les sous-espaces employés dans notre problème d'optimisation fonctionnelle peuvent être construits de la même façon que dans les espaces de Hilbert. Dans [ZFR15a] nous avons considéré le sous-espace de gradient à mémoire (MG) dont la structure est

$$\mathbf{D}_{SMG}^k(\mathbf{w}) = [df(q^k, \mathbf{w}), d^{k-1}(\mathbf{w}), \dots, d^{k-I+1}(\mathbf{w})], \quad (5.33)$$

où $(d^{k-l}(\mathbf{w}))_{l=0, \dots, I-1}$ représentent les directions aux itérations précédentes. Elles sont données par

$$d^{k-l}(\mathbf{w}) = \ln \left(\frac{\prod_i q_i^{k-l+1}(w_i)}{K^{k-l}(\mathbf{s}^{k-l}) \prod_i q_i^{k-l}(w_i)} \right) \cdot d^{k-l}(\mathbf{w}) = \ln \left(\frac{q^{k-l+1}(w_i)}{K^{k-l}(\mathbf{s}^{k-l}) q^{k-l}(w_i)} \right). \quad (5.34)$$

Dans la thèse de Yuling Zheng nous avons aussi considéré le cas de sous-espace généré uniquement par des gradients. Cependant le sous-espace gradient à mémoire est en pratique plus rapide que le sous-espace par descente de gradients.

Ainsi l'algorithme défini dans [ZFR15a] est le suivant.

Algorithme 5.4.1 — Bayésien variationnel basé sur le sous-espace.

[O1] Initialiser $(q^0 \in \Omega)$

[O2] répéter

- a. déterminer le sous-espace $\mathbf{D}^k(\mathbf{w})$ en utilisant (5.33)
- b. déterminer le pas multi-dimensionnel de l'algorithme \mathbf{s}^k
- c. calculer $q^{k+1}(\mathbf{w}) = K^k(\mathbf{s}^k) q^k(\mathbf{w}) \exp[\mathbf{D}^k(\mathbf{w})\mathbf{s}^k]$

jusqu'à convergence

Dans les faits, en suivant l'analyse de [Cho10] la dimension du sous-espace que nous avons considéré est $I = 2$. En effet, la complexité de l'algorithmique augmente rapidement avec cette dimension.

En ce qui concerne la performance des algorithmes proposés, elle dépend aussi du pas multi-dimensionnel \mathbf{s}^k . Concernant ce pas, nous définissons d'abord la fonction

$$g^k : \mathbb{R}^2 \rightarrow \mathbb{R} \\ g^k(\mathbf{s}) = \mathcal{F} \left(K^k(\mathbf{s}) q^k(\mathbf{w}) \exp \left[\mathbf{D}^k(\mathbf{w})\mathbf{s} \right] \right), \quad (5.35)$$

puis le pas sous-optimal défini par

$$(\hat{\mathbf{s}}^{opt})^k = \arg \max_{\mathbf{s} \in \mathbb{R}^2} g^k(\mathbf{s}). \quad (5.36)$$

Encore une fois, ce choix est dicté par la recherche du meilleur compromis entre rapidité de l'algorithme et complexité algorithmique. Dans les faits, ce pas peut être trop grand pour permettre à l'algorithme de converger. Dans ce cas, nous utilisons encore une fois une stratégie de « backtracking ».

5.5 Application aux problèmes inverses

En ce qui concerne l'application de nos algorithmes bayésiens variationnels, nous avons considéré le modèle linéaire (4.1). Quant à la loi *a priori*, dans un souci de clarté je ne présenterai ici que l'application à un *a priori* de type mélange scalaire de gaussiennes (MSG) tel que défini dans la partie 3.3.

Nous considérons donc le problème inverse linéaire associé au modèle direct

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (5.37)$$

où $\mathbf{y} \in \mathbb{R}^M$ représente le vecteur des mesures obtenues, $\mathbf{x} \in \mathbb{R}^N$ est le vecteur des paramètres à estimer et $\mathbf{A} \in \mathbb{R}^{M \times N}$ est une matrice connue. Par ailleurs, $\mathbf{n} \in \mathbb{R}^M$ est le vecteur du bruit. Pour résoudre ce type de problèmes, nous utilisons dans la suite nos approches bayésiennes variationnelles en considérant des lois *a priori* de type MSG. Plus exactement, nous suivons la définition de [JM06] et supposons que les éléments de \mathbf{x} sont indépendants et suivent une loi MSG. Par conséquent, pour $i = 1, \dots, N$ nous considérons que $X_i = V_i/\sqrt{Z_i}$, où Z_i et V_i sont indépendants et $V_i \sim \mathcal{N}(0, 1/\gamma_p)$. Par rapport à la définition classique, nous rajoutons aussi un hyperparamètre γ_p qui nous permettra dans les algorithmes de régler le compromis entre l'information venant de l'*a priori* et celle venant de la vraisemblance. Finalement, la distribution de X_i , pour $i = 1, \dots, N$, peut s'écrire

$$p_{X_i}(x_i) = \int \frac{z_i}{(2\pi)^{1/2}} \exp\left(-\frac{\gamma_p z_i x_i^2}{2}\right) d\mathbb{P}_{Z_i}(z_i), \quad (5.38)$$

où $\mathbb{P}_{Z_i}(z_i)$ correspond à la loi de Z_i qui peut être continue ou discrète. Quand Z_i est une variable continue, on note $d\mathbb{P}_{Z_i}(z_i) = p_{Z_i} dz_i$. Notons que dans un souci de simplicité des notations nous avons mis z comme un paramètre de précision et non de variance.

5.5.1 Loi *a posteriori*

L'inférence bayésienne est basée sur la loi *a posteriori* obtenue dans (5.1). En considérant que le bruit \mathbf{n} est un bruit blanc gaussien de moyenne nulle et de variance $\sigma_n^2 = 1/\gamma_n$, la vraisemblance devient

$$p(\mathbf{y}|\mathbf{x}, \gamma_n) = C \gamma_n^{M/2} \exp\left(-\frac{\gamma_n \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2}{2}\right) \quad (5.39)$$

et grâce à la règle de Bayes la loi *a posteriori* vérifie :

$$p(\mathbf{x}, \mathbf{z}|\mathbf{y}) \propto \gamma_n^{M/2} \exp\left[-\frac{\gamma_n \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2}{2}\right] \prod_{i=1}^N \sqrt{\gamma_p z_i} \exp\left[-\frac{1}{2} \gamma_p z_i x_i^2\right] \mathbb{P}_{Z_i}(z_i). \quad (5.40)$$

Notons que dans cette équation nous nous situons dans le cadre supervisé puisque les paramètres de la méthode, ici (γ_n, γ_p) , ne sont pas estimés. Si nous généralisons au cadre

non-supervisé, il faut introduire des lois *a priori* aussi sur les hyperparamètres. Notons p_{γ_n} et p_{γ_p} ces lois *a priori*. Comme pour le signal d'intérêt, \mathbf{x} , nous devons prendre des lois conjuguées. Nous avons donc considéré des lois gamma

$$p(\gamma_n) = \mathcal{G}(\gamma_n | \tilde{a}_n, \tilde{b}_n) \quad (5.41)$$

$$p(\gamma_p) = \mathcal{G}(\gamma_p | \tilde{a}_p, \tilde{b}_p) \quad (5.42)$$

où pour $a > 0$ et $b > 0$

$$\mathcal{G}(z | a, b) = \frac{b^a}{\Gamma(a)} z^{a-1} \exp(-bz). \quad (5.43)$$

Comme nous n'avons aucune information sur ces paramètres, nous considérons pour ces lois des cas dégénérés, $a = b = 0$, ce qui nous permet d'obtenir des lois non-informatives de Jeffreys.

Finalement, l'*a posteriori* que nous avons est

$$p(\mathbf{x}, \mathbf{z}, \gamma_n, \gamma_p | \mathbf{y}) \propto \gamma_n^{M/2} \exp\left[-\frac{\gamma_n \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2}{2}\right] \prod_{i=1}^N \sqrt{\gamma_p z_i} \exp\left[-\frac{1}{2} \gamma_p z_i x_i^2\right] \mathbb{P}_{Z_i}(z_i) p_{\gamma_n}(\gamma_n) p_{\gamma_p}(\gamma_p). \quad (5.44)$$

C'est cette loi que nous cherchons à approcher par nos algorithmes bayésiens variationnels. Une première étape est de choisir la séparabilité considérée. En ce qui concerne les travaux que nous avons menés, nous avons pris à chaque fois une séparabilité totale

$$\begin{aligned} q_{\mathbf{w}}(\mathbf{w}) &= q_{\mathbf{x}}(\mathbf{x}) q_{\mathbf{z}}(\mathbf{z}) q_{\gamma_n}(\gamma_n) q_{\gamma_p}(\gamma_p) \\ &= \prod_{i=1}^N q_i(x_i) \prod_{j=1}^N \tilde{q}_j(z_j) q_{\gamma_n}(\gamma_n) q_{\gamma_p}(\gamma_p). \end{aligned} \quad (5.45)$$

Ce type de séparabilité est d'après la partie 5.2.4 la mieux adaptée à nos algorithmes.

Sous cette hypothèse, nous menons une optimisation alternée par rapport à $q_{\mathbf{z}}$, $q_{\mathbf{x}}$, q_{γ_n} et q_{γ_p} , qui nous conduit à

$$q_{\mathbf{z}}^{k+1} = \arg \max_{q_{\mathbf{z}}} \mathcal{F} \left(q_{\mathbf{x}}^k q_{\mathbf{z}}^k q_{\gamma_n}^k q_{\gamma_p}^k \right), \quad (5.46)$$

$$q_{\mathbf{x}}^{k+1} = \arg \max_{q_{\mathbf{x}}} \mathcal{F} \left(q_{\mathbf{x}}^{k+1} q_{\mathbf{z}}^k q_{\gamma_n}^k q_{\gamma_p}^k \right), \quad (5.47)$$

$$q_{\gamma_n}^{k+1} = \arg \max_{q_{\gamma_n}} \mathcal{F} \left(q_{\mathbf{x}}^{k+1} q_{\mathbf{z}}^{k+1} q_{\gamma_n}^k q_{\gamma_p}^k \right), \quad (5.48)$$

$$q_{\gamma_p}^{k+1} = \arg \max_{q_{\gamma_p}} \mathcal{F} \left(q_{\mathbf{x}}^{k+1} q_{\mathbf{z}}^{k+1} q_{\gamma_n}^{k+1} q_{\gamma_p}^k \right) \quad (5.49)$$

A chaque étape nous déduisons de la loi approchante une estimation du paramètre d'intérêt correspondant par moyenne *a posteriori*.

Le choix de lois conjuguées pour l'*a priori* permet de ramener cette mise à jour des lois approchantes à celle des paramètres.

Estimation de \mathbf{x}

En ce qui concerne \mathbf{x} sa loi est gaussienne sachant la variable cachée \mathbf{z} . Par conséquent pour sa loi approchante, nous considérons une variable gaussienne. Ainsi pour $k \in \mathbb{N}^*$, la variable approchante à l'itération k est

$$q^k \sim \prod_i \mathcal{N}(\mathbf{m}_k(i), \boldsymbol{\sigma}_k^2(i))$$

et concernant l'initialisation

$$q^0 = \mathcal{N}(\mathbf{m}_0, \text{Diag}(\boldsymbol{\sigma}_0^2))$$

où $\text{Diag}(\mathbf{v})$ est une matrice diagonale dont les composantes sont données par le vecteur \mathbf{v} tandis que $\boldsymbol{\sigma}_0^2 \in \mathbb{R}^N$ est le vecteur d'initialisation des variances.

Pour chacun des algorithmes que nous avons définis, que ce soit l'algorithme 5.3.1 ou 5.4.1, la première étape est de calculer une fonction intermédiaire q^r définie, pour $i = 1, \dots, N$, par

$$q_i^r(x_i) \propto \exp \left(\langle \ln p(\mathbf{w}, \mathbf{y}) \rangle_{\prod_{j \neq i} q_j^k(x_j) q_{\mathbf{z}}^k(\mathbf{z}) q_{\gamma_n}(\gamma_n) q_{\gamma_p}(\gamma_p)} \right). \quad (5.50)$$

On obtient après développement

$$q_i^r(x_i) \propto \exp \left[-\frac{\langle \gamma_n \rangle_k}{2} \left(x_i^2 \text{diag}(\mathbf{A}^T \mathbf{A})_i - 2x_i (\mathbf{A}^T \mathbf{y})_i + 2x_i (\mathbf{A}^T \mathbf{A} \mathbf{m}_k)_i - 2x_i \text{diag}(\mathbf{A}^T \mathbf{A})_i \mathbf{m}_k(i) \right) + \frac{\langle \gamma_p \rangle_k x_i^2}{2} \langle z_i \rangle_k \right], \quad (5.51)$$

où $\text{diag}(A)$ est le vecteur composé des éléments diagonaux de A tandis que $\langle z_i \rangle_k$ est l'espérance de z_i sous la loi approchante $q_{z_i}^k$, $\langle \gamma_n \rangle_k$ l'espérance de γ_n sous la loi $q_{\gamma_n}^k$ et $\langle \gamma_p \rangle_k$ celle de γ_p sous la loi $q_{\gamma_p}^k$. Cette fonction intermédiaire q^r correspond, à une normalisation près, à une gaussienne de paramètres :

$$\boldsymbol{\sigma}_r^2(i) = (\text{diag}(\mathbf{A}^T \mathbf{A})_i \langle \gamma_n \rangle_k + \langle z_i \rangle_k \langle \gamma_p \rangle_k)^{-1} \quad (5.52)$$

et

$$\mathbf{m}_r(i) = \boldsymbol{\sigma}_r^2(i) \times (\langle \gamma_n \rangle_k (\mathbf{A}^T \mathbf{y})_i - (\mathbf{A}^T \mathbf{A} - \text{diag}(\mathbf{A}^T \mathbf{A})) \mathbf{m}_k)_i. \quad (5.53)$$

A partir de cette fonction intermédiaire, nous pouvons construire les mises à jour des différents algorithmes en remarquant que dans (5.19), la fonction de mise à jour peut aussi s'écrire

$$h^k(\mathbf{w}) = \tilde{K}^k(\alpha^k) \left(\frac{q^r(\mathbf{w})}{q^k(\mathbf{w})} \right)^\alpha. \quad (5.54)$$

Ainsi dans 5.3.1 la mise à jour se fait en prenant $q^{k+1} = q^\alpha$ avec q^α une gaussienne de paramètres

$$\sigma_\alpha^2(i) = \frac{\sigma_r^2(i)\sigma_k^2(i)}{\sigma_r^2(i) + \alpha(\sigma_k^2(i) - \sigma_r^2(i))} \quad (5.55)$$

$$\mathbf{m}_\alpha(i) = \frac{\mathbf{m}_k(i)\sigma_r^2(i) + \alpha(\mathbf{m}_r(i)\sigma_k^2(i) - \mathbf{m}_k(i)\sigma_r^2(i))}{\sigma_r^2(i) + \alpha(\sigma_k^2(i) - \sigma_r^2(i))}, \quad (5.56)$$

et α obtenu par le pas sous-optimal.

De manière équivalente, dans (5.31), on peut remarquer que

$$\phi_1^k(\mathbf{w}) = \prod_i \frac{q_i^r(\mathbf{x}_i)}{q_i^k(\mathbf{x}_i)}, \quad (5.57)$$

tandis que l'autre fonction intervenant dans le gradient à mémoire peut s'écrire

$$\phi_2^k(\mathbf{w}) = \prod_i \frac{q_i^k(\mathbf{x}_i)}{q_i^{k-1}(\mathbf{x}_i)}. \quad (5.58)$$

Ainsi une fois encore, après le calcul de q^r , la mise à jour se fait en prenant la gaussienne q^s de paramètres :

$$\sigma_s^2 = \left[\frac{1}{\sigma_k^2} + s_1 \left(\frac{1}{\sigma_{r_1}^2} - \frac{1}{\sigma_k^2} \right) + s_2 \left(\frac{1}{\sigma_k^2} - \frac{1}{\sigma_{k-1}^2} \right) \right]^{-1}, \quad (5.59)$$

$$\mathbf{m}_s = \sigma_s^2 \left[\frac{\mathbf{m}_k}{\sigma_k^2} + s_1 \left(\frac{\mathbf{m}_{r_1}}{\sigma_{r_1}^2} - \frac{\mathbf{m}_k}{\sigma_k^2} \right) + s_2 \left(\frac{\mathbf{m}_k}{\sigma_k^2} - \frac{\mathbf{m}_{k-1}}{\sigma_{k-1}^2} \right) \right]. \quad (5.60)$$

Une fois encore la dernière étape consiste à trouver le pas.

Finalement dans ce contexte de MSG, les mises à jour de nos différents algorithmes sont assez similaires et peuvent se résumer de la façon suivante

Algorithme 5.5.1 — Mises à jour de \mathbf{x} dans le cas de mélange scalaire de gaussiennes.

[O1] Initialiser $\mathbf{m}_0, \text{Diag}(\sigma_0^2)$

[O2] Répéter

(a) Calculer $\mathbf{m}_r, \text{Diag}(\sigma_r^2)$ grâce à (5.52) et (5.53).

(b) Utiliser (5.55) et (5.56) ou (5.59) et (5.59) pour déterminer la loi de mise à jour en fonction des pas.

(c) Calculer les pas sous-optimaux.

Estimation de \mathbf{z}

En ce qui concerne l'estimation de \mathbf{z} , nous avons étudié plusieurs cas. Dans un premier temps, nous avons considéré le cas particulier où \mathbf{x} suivait une loi Student-t. Dans ce cas, on peut voir dans [WS00] que la variable cachée \mathbf{z} suit une loi gamma. On peut alors calculer explicitement la loi approchante $q_{\mathbf{z}}^k$ à l'aide de la méthodologie du bayésien variationnel. Cependant pour la plupart des distributions dans la famille MSG, on ne connaît pas l'expression de $p_{z_i}(z_i)$ et on ne peut pas non plus déterminer cette loi approchante. En reprenant (5.11), on peut remarquer que

$$\begin{aligned}
\forall i, \quad q_{z_i}^{k+1}(z_i) &\propto \exp \left[\langle \ln p(\mathbf{w}, \mathbf{y}) \rangle_{q_{\mathbf{x}}^{k+1} \prod_{j \neq i} q_{z_j}^k q_{\gamma_n}^k q_{\gamma_p}^k} \right] \\
&\propto \exp \left(\frac{1}{2} \ln(z_i) - \frac{\gamma_p^k}{2} z_i \langle c_i^2 \rangle^k + \ln p_{z_i}(z_i) \right) \\
&\propto \sqrt{z_i} p_{z_i}(z_i) \exp \left(-\frac{\gamma_p^k}{2} \langle x_i^2 \rangle^k z_i \right) \\
&= p(z_i | \sqrt{\langle x_i^2 \rangle^k}, \gamma_p^k).
\end{aligned} \tag{5.61}$$

Cette loi n'a pas d'expression explicite. Néanmoins, notre objectif n'est pas d'obtenir l'expression explicite de $q_{\mathbf{z}}$, mais d'obtenir $q_{\mathbf{x}}$. Comme on peut le voir dans (5.52), il suffit alors de connaître l'espérance de \mathbf{z} sous la loi $q_{\mathbf{z}}^{k+1}$:

$$\begin{aligned}
\forall i, \quad \langle z_i \rangle_{k+1} &= \int z_i q_{z_i}^{k+1}(z_i) dz_i \\
&= \int z_i p \left(z_i | \sqrt{\langle x_i^2 \rangle^k}, \gamma_p^k \right) dz_i \\
&= \mathbb{E}_{p(z_i | \sqrt{\langle x_i^2 \rangle^k}, \gamma_p^k)}(z_i).
\end{aligned} \tag{5.62}$$

Comme on peut le voir dans [PKDRW06], dans le cas de MSG on a alors

$$\begin{aligned}
\forall i, \quad p'_{x_i}(x_i) &= \frac{\partial}{\partial x_i} \int_0^\infty p(x_i | z_i, \gamma_p) p_{z_i}(z_i) dz_i \\
&= - \int_0^\infty \gamma_p x_i z_i p(x_i | z_i, \gamma_p) p_{z_i}(z_i) dz_i \\
&= - \int_0^\infty \gamma_p x_i z_i p(z_i | x_i, \gamma_p) p_{x_i}(x_i) dz_i = -\gamma_p x_i p_{x_i}(x_i) \mathbb{E}_{p(z_i | x_i, \gamma_p)}(z_i),
\end{aligned} \tag{5.63}$$

qui nous amène à

$$\mathbb{E}_{p(z_i | x_i, \gamma_p)}(z_i) = -\frac{p'_{x_i}(x_i)}{\gamma_p x_i p_{x_i}(x_i)}. \tag{5.64}$$

On peut ainsi à chaque itération calculer l'estimation des variables cachées sans déterminer leur loi approchante par :

$$\langle z_i \rangle^{k+1} = - \frac{p'_{X_i}(x_i)}{\gamma_p x_i p_{X_i}(x_i)} \Big|_{x_i = \sqrt{\langle x_i^2 \rangle^k}, \gamma_p = \langle \gamma_p \rangle^k}. \quad (5.65)$$

Estimation des hyperparamètres

Grâce à l'utilisation des lois *a priori* conjuguées, l'optimum pour γ_n et γ_p sera obtenu avec des lois gamma. Nous prenons donc des lois approchantes dans la même famille

$$q_{\gamma_n}^k(\gamma_n) = \mathcal{G}(\gamma_n | a_{\gamma_n}^k, b_{\gamma_n}^k), \quad (5.66)$$

$$q_{\gamma_p}^k(\gamma_p) = \mathcal{G}(\gamma_p | a_{\gamma_p}^k, b_{\gamma_p}^k). \quad (5.67)$$

Ainsi, l'optimisation des lois approchantes est équivalente à l'optimisation de ses paramètres.

En utilisant la formule (5.11), nous obtenons

$$\begin{aligned} q_{\gamma_n}^{k+1}(\gamma_n) &\propto \exp \left[\langle \ln p(\mathbf{w}, \mathbf{y}) \rangle_{q_{\mathbf{x}}^{k+1} q_{\mathbf{z}}^{k+1} q_{\gamma_p}^k} \right] \\ &\propto \exp \left[\left(\frac{M}{2} + \tilde{a}_n - 1 \right) \ln(\gamma_n) - \gamma_n \left(\int \frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2}{2} q_{\mathbf{x}}^{k+1}(\mathbf{x}) d\mathbf{x} + \tilde{b}_n \right) \right] \\ &\propto \exp \left[\left(\frac{M}{2} + \tilde{a}_n - 1 \right) \ln(\gamma_n) \right. \\ &\quad \left. - \gamma_n \left(\frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{m}_{k+1}\|^2 + \frac{1}{2} \sum_{i=1}^N \text{diag}(\mathbf{A}^T \mathbf{A})_i (\sigma_{k+1}^2)_i + \tilde{b}_n \right) \right]. \end{aligned} \quad (5.68)$$

Donc $q_{\gamma_n}^{k+1}$ est une loi gamma de paramètres

$$a_{\gamma_n}^{k+1} = \frac{M}{2} + \tilde{a}_n = a_{\gamma_n} \quad (5.69)$$

$$b_{\gamma_n}^{k+1} = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{m}_{k+1}\|^2 + \frac{1}{2} \sum_{i=1}^N \text{diag}(\mathbf{A}^T \mathbf{A})_i (\sigma_{k+1}^2)_i + \tilde{b}_n \quad (5.70)$$

En prenant la moyenne *a posteriori* selon la loi approchante comme estimateur nous avons donc

$$\langle \gamma_n \rangle^{k+1} = a_{\gamma_n} / b_{\gamma_n}^{k+1}. \quad (5.71)$$

La distribution q_{γ_p} est déterminée de façon analogue. En utilisant encore une fois

(5.11), nous avons

$$\begin{aligned}
q_{\gamma_p}^{k+1}(\gamma_p) &\propto \exp \left[\langle \ln p(\mathbf{x}, \mathbf{z}, \gamma_n, \gamma_p, \mathbf{y} | \nu) \rangle_{q_{\mathbf{x}}^{k+1} q_{\mathbf{z}}^{k+1} q_{\gamma_n}^{k+1}} \right] \\
&\propto \exp \left[\left(\frac{N}{2} + \tilde{a}_p - 1 \right) \ln(\gamma_p) - \gamma_p \left(\int \sum_{i=1}^N \frac{z_i x_i^2}{2} \prod_i q_i^{k+1}(x_i) \tilde{q}_i^{k+1}(z_i) dx dz + \tilde{b}_p \right) \right] \\
&\propto \exp \left[\left(\frac{N}{2} + \tilde{a}_p - 1 \right) \ln(\gamma_p) - \gamma_p \left(\frac{1}{2} \sum_{i=1}^N \frac{(\mathbf{a})_i}{(\mathbf{b}_{k+1})_i} ((\mathbf{m}_{k+1})_i^2 + (\boldsymbol{\sigma}_{k+1}^2)_i) + \tilde{b}_p \right) \right]
\end{aligned} \tag{5.72}$$

qui est identifiée comme une loi gamma de paramètres

$$a_{\gamma_p}^{k+1} = N/2 + \tilde{a}_p \tag{5.73}$$

$$b_{\gamma_p}^{k+1} = \frac{1}{2} \sum_{i=1}^N \frac{(\mathbf{a})_i}{(\mathbf{b}_{k+1})_i} ((\mathbf{m}_{k+1})_i^2 + (\boldsymbol{\sigma}_{k+1}^2)_i) + \tilde{b}_p \tag{5.74}$$

et l'estimation de γ_p est

$$\langle \gamma_p \rangle^{k+1} = a_{\gamma_p} / b_{\gamma_p} \tag{5.75}$$

5.5.2 Exemple

Afin de présenter l'intérêt des méthodes que nous avons développées dans cette partie, nous avons étudié un exemple simple de tomographie pour lequel nous avons comparé nos méthodes avec l'état de l'art. Pour cela, nous avons considéré une image 64×64 , pour pouvoir comparer avec les autres méthodes, composée de 7 pics d'amplitude variant entre 0.5 et 1.

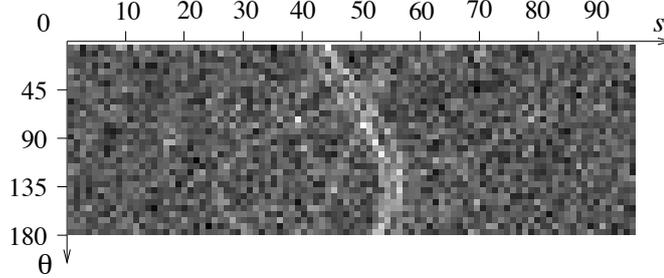


FIGURE 5.1 – Données : sinogramme obtenue avec 32 angles de projections et 95 détecteurs.

Les données simulées sont des projections obtenues sous une géométrie parallèle. Ces projections sont collectées à 32 angles différents θ , qui sont uniformément distribués sur $[0, \pi[$ et obtenus avec 95 détecteurs. Nous avons ajouté aussi un bruit blanc gaussien de moyenne nulle et d'écart-type 0.3. Par conséquent, les données (voir Figure 5.1) présentent un faible rapport signal sur bruit. Nous avons alors comparé nos différentes méthodes avec une approche directe de retroprojection filtrée (FBP), un algorithme MCMC

de type échantillonneur de Gibbs et les algorithmes bayésien variationnels, avec séparabilité partielle (BVClass-SP) pour laquelle nous avons gardé les corrélations entre les pixels et séparabilité totale (BVClass-ST). Tous les algorithmes itératifs ont été initialisés avec une moyenne nulle pour toutes les composantes et une variance à 1.

Les résultats de simulation présentés en figure 5.2 donnent des qualités de reconstruction similaires pour les méthodes basées sur du bayésien variationnel et meilleures que la méthode directe ou la méthode par MCMC. Théoriquement, la reconstruction MCMC est la meilleure car elle converge asymptotiquement vers la loi vraie *a posteriori* plutôt de chercher à l'estimer. Cependant, dans un temps limité, les échantillons obtenus par cette méthode n'ont pas suffisamment convergé, ce qui explique la reconstruction de qualité relativement mauvaise montrée figure 5.2 (c). En ce qui concerne le gain de nos méthodes par rapport au bayésien variationnel classique il faut regarder le temps de calcul pour le voir apparaître.

TABLE 5.1 – COMPARAISON DE PERFORMANCE EN TERMES DE PSNR(dB)/TEMPS CPU (s).

Méthode	FBP	MCMC	BVClass-SP	BVClass-ST	BV-Grad	BV-MG
PSNR(dB)	27.32	28.83	35.12	35.24	35.90	35.92
Temps(s)	0.05	69313.8	723.5	327.6	23.4	3.2

Dans le tableau 5.1, nous comparons la qualité de reconstruction des différentes méthodes ainsi que le temps de calcul. Les méthodes basées sur le bayésien variationnel montrent des résultats équivalents. Cependant, les méthodes proposées sont bien plus rapides. Ainsi on constate que sur cet exemple simple, le bayésien variationnel basé sur le gradient est déjà 14 fois plus rapide que les méthodes classiques, 100 fois pour le gradient à mémoire.

5.6 Application à la reconstruction d'images par ondelettes

Afin de tester notre approche sur des problèmes inverses de grande dimension, nous avons, entre autres, considéré un problème de déconvolution d'images dans le domaine des ondelettes.

5.6.1 Modèle utilisé

Supposons que l'image inconnue $\mathbf{x} \in \mathbb{R}^P$ puisse être représentée par une décomposition sur un dictionnaire $\mathbf{P} \in \mathbb{R}^{P \times N}$ sous la forme $\mathbf{x} = \mathbf{P}\mathbf{u}$ où $\mathbf{u} \in \mathbb{R}^N$ représente les coefficients associés. Nous considérons ici un modèle linéaire

$$\mathbf{y} = \mathbf{A}\mathbf{P}\mathbf{u} + \mathbf{n}, \quad (5.76)$$

où $\mathbf{y} \in \mathbb{R}^M$ représente les données, l'opérateur $\mathbf{A} \in \mathbb{R}^{M \times P}$ est supposé connu et \mathbf{n} est un bruit blanc gaussien, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \gamma_n^{-1}\mathbf{I})$, avec γ_n l'inverse de la variance du bruit. Par conséquent, $p(\mathbf{y}|\mathbf{u}, \gamma_n) = \mathcal{N}(\mathbf{A}\mathbf{P}\mathbf{u}, \gamma_n^{-1}\mathbf{I})$.

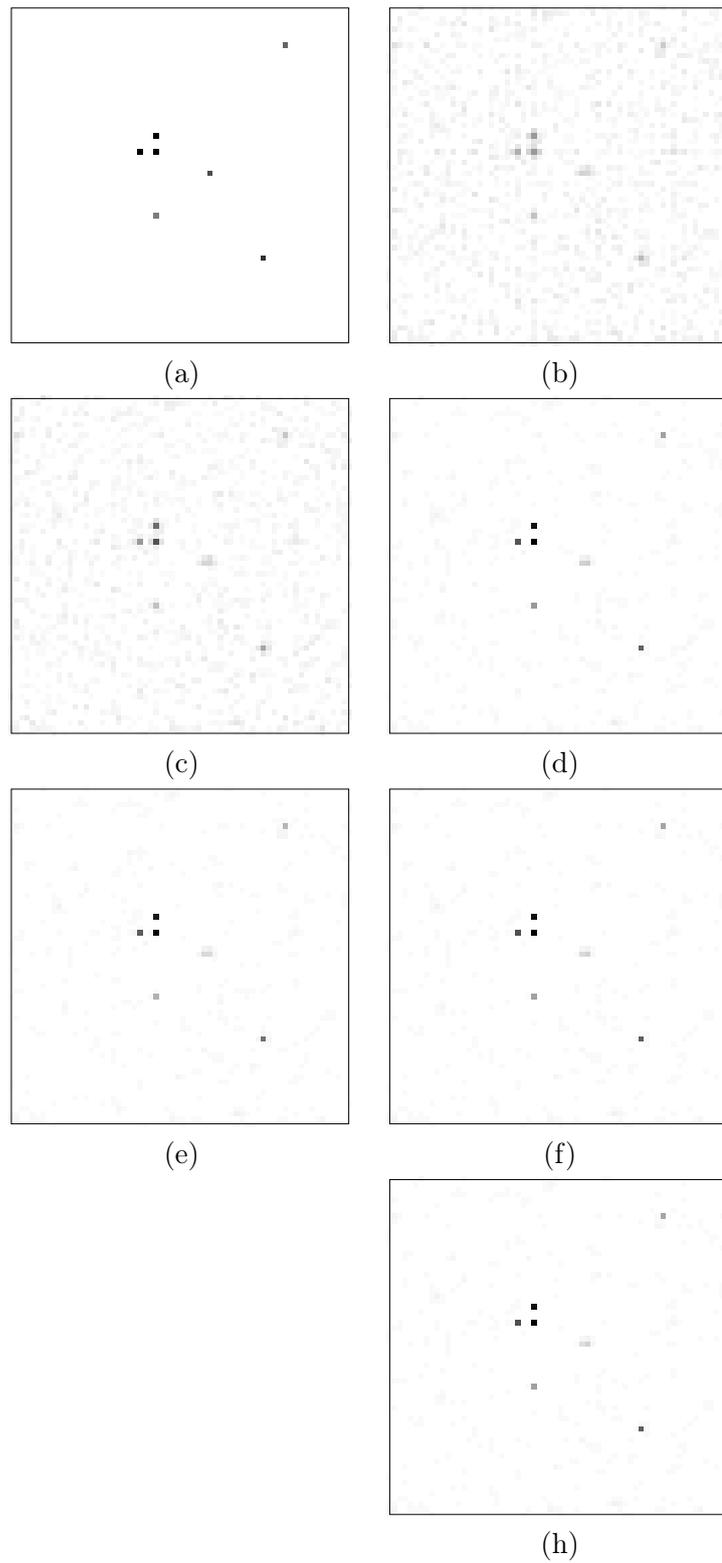


FIGURE 5.2 – Les images sont présentées avec le même niveau de gris inversé. (a) L'image Vrai, (b) FBP avec un filtre rampe, (c) l'échantillonnage de Gibbs, méthode d'approximation (d) BVClass-SP, (e) BVClass-ST, (f) BV-Grad, (h) BV-MG.

Concernant les coefficients \mathbf{u} , nous considérons une loi *a priori* séparable gaussienne généralisée définie par (3.2). Comme nous l'avons déjà vu dans la partie 3.3.1, cette densité peut s'écrire sous la forme d'un MSG pour lequel on ne connaît pas la loi de la variable cachée. De plus ici, les coefficients \mathbf{u} sont divisés en L sous-bandes et les paramètres des lois *a priori* des coefficients dans les sous-bandes différentes sont supposés différents. Notons l'ensemble des indices des coefficients dans la l^e sous-bande par $(I_l)_{l=1,\dots,L}$. On écrit la loi *a priori* de \mathbf{u} comme

$$\begin{aligned} p(\mathbf{u}) &= \prod_{l=1}^L \prod_{i \in I_l} p(u_i | \gamma_p^l, \beta^l) \\ &= \prod_{l=1}^L \prod_{i \in I_l} \int_{\mathbb{R}} \mathcal{N}(u_i | 0, (z_i \gamma_p^l)^{-1}) p(z_i | \beta^l) dz_i \end{aligned}$$

où $\mathbf{z} = (z_1, \dots, z_N)$ est le vecteur des variables cachées qui suivent des distributions p_{z_i} de paramètres β^l . Ici, $\gamma_p = (\gamma_p^1, \dots, \gamma_p^L)$ est le vecteur des paramètres d'échelle de notre *a priori* GSM, qui correspond à $\omega^{2/\beta}$ dans (3.2).

Au total, on introduit un hyperparamètre γ_n pour la loi du bruit, un vecteur d'hyperparamètres γ_p associé à l'échelle des lois *a priori* des coefficients et un vecteur d'hyperparamètres β associé aux paramètres de forme de ces mêmes lois. Les paramètres de forme permettent de régler le degré de parcimonie introduit *a priori*. Afin d'introduire un degré de parcimonie qui varie avec l'échelle des coefficients en ondelettes nous choisissons de fixer β . Néanmoins, nous estimons γ_n et γ_p car ils règlent le compromis entre la fidélité aux données et la confiance à l'information *a priori*. Pour ce faire, des lois de Jeffreys non-informatives sont affectées à γ_n et γ_p .

La loi *a posteriori* correspondante est :

$$p(\mathbf{w} | \mathbf{y}) \propto \gamma_n^{M/2} \exp \left[-\frac{\gamma_n \|\mathbf{y} - \mathbf{A}\mathbf{P}\mathbf{u}\|^2}{2} \right] \prod_{l=1}^L \prod_{i \in I_l} \sqrt{z_i \gamma_p^l} \exp \left[-\frac{\gamma_p^l}{2} z_i u_i^2 \right] p_{z_i}(z_i) \gamma_n^{-1} \prod_{l=1}^L (\gamma_p^l)^{-1}.$$

Cette distribution est de forme compliquée. Pour s'affranchir de la difficulté de calcul, nous utilisons les mises à jour des approches d'approximation bayésienne variationnelle définies dans la partie 5.5.

Ainsi pour chaque $i = 1, \dots, N$, les q_{u_i} sont mis à jour en utilisant l'algorithme 5.5.1. En ce qui concerne la mise à jour des Z_i nous remarquons que dans le cas d'une loi gaussienne généralisée, nous pouvons calculer la dérivée de la densité :

$$\begin{aligned} \forall u_i \neq 0 \quad p'(u_i) &= -\frac{\sqrt{\gamma_p} \beta}{2\Gamma(1/\beta)} e^{-|\sqrt{\gamma_p} u_i|^\beta} \gamma_p^{\beta/2} \frac{\beta}{2} (u_i^2)^{\beta/2-1} 2u_i \\ &= -p(u_i) \beta u_i \gamma_p^{\beta/2} (u_i^2)^{\beta/2-1}. \end{aligned} \tag{5.77}$$

En utilisant (5.65) et (5.77), nous pouvons en déduire que

$$\begin{aligned}
\langle z_i \rangle^{k+1} &= \beta \gamma_p^{\beta/2-1} (u_i^2)^{\beta/2-1} \Big|_{u_i = \sqrt{\langle u_i^2 \rangle^k}, \gamma_p = \langle \gamma_p \rangle^k} \\
&= \beta (\langle \gamma_p \rangle^k)^{\beta/2-1} [(\mathbf{m}_k)_i^2 + (\boldsymbol{\sigma}_k)_i^2]^{\frac{\beta}{2}-1} \\
&= \beta \left[\langle \gamma_p \rangle^k ((\mathbf{m}_k)_i^2 + (\boldsymbol{\sigma}_k)_i^2) \right]^{\frac{\beta}{2}-1}
\end{aligned} \tag{5.78}$$

où $\beta = \beta^l$, $\gamma_p = \gamma_p^l$ pour $i \in \mathbf{I}_l$.

En ce qui concerne γ_n et γ_p^l nous utilisons (5.68) et (5.75). Comme nous avons reconstruit les coefficients \mathbf{u} , une étape de synthèse est appliquée à la fin pour obtenir une estimation de \mathbf{x} : $\hat{\mathbf{x}} = \mathbf{P}\hat{\mathbf{m}}$.

5.6.2 Résultats

Pour évaluer notre approche, nous l'avons appliqué à un problème de déconvolution et comparé les résultats avec deux approches existantes : l'approche SURE-LET [PBBC09] dans le domaine des ondelettes et une approche non-supervisée basée sur la variation totale (TV) dans le domaine de l'image [ZFR15b].

Dans toutes les simulations, nous avons utilisé une trame redondante constituée de neuf bases d'ondelettes orthogonales. Nous avons utilisé pour cela les ondelettes Daubechies 4 et nous avons fait une décomposition sur 3 niveaux, ce qui conduit à $L = 10$ sous-bandes différentes.

Les trois approches ont été testées avec les données générées à partir de deux images standard : *Lena* et *Cameraman*. Pour générer des données, un noyau de convolution uniforme de taille 9×9 a été utilisé. Ensuite, nous avons ajouté un bruit gaussien aux images convoluées à un SNR = 40 dB.

Notre approche a été mise en œuvre avec les initialisations suivantes : la transformée des données comme moyenne, 1 comme variance des coefficients à estimer et l'inverse de la vraie variance du bruit comme initialisation pour γ_n et γ_p^l . Concernant les paramètres de forme de l'*a priori* GG, nous avons utilisé 2, qui conduit à une loi gaussienne, pour les coefficients d'approximation, et [1, 0.9, 0.8], qui conduisent à des distributions à queue lourde, pour les coefficients de détails de l'échelle la plus grossière à l'échelle la plus fine.

Nous montrons dans la figure 5.3 (a) une image de *Lena*. Les reconstructions obtenues par les trois approches sont montrées dans la figure 5.3 (b)-(d). Nous pouvons voir que le résultat de notre approche (figure 5.3 (d)) est meilleur que celui obtenu par SURE-LET (figure 5.3 (b)). Par ailleurs, en comparant la figure 5.3 (c) et (d), nous pouvons voir que les détails de *Lena* sont mieux reconstruits par notre approche que par l'approche basée sur TV. Par exemple, le bord du chapeau et la plume sur le chapeau sont plus nets que ceux dans la figure 5.3 (c).

Nous montrons dans le tableau 5.2 les PSNR des reconstructions obtenues par les trois approches. Nous pouvons voir que dans les deux cas, l'approche proposée donne un meilleur PSNR que l'approche SURE-LET : pour *Lena*, notre approche donne 29.69 dB qui est 0.47 dB de plus que celui obtenu par SURE-LET, et pour *Cameraman*, notre

TABLE 5.2 – PSNR (dB) OBTENU PAR SURE-LET, L'APPROCHE BASÉE SUR TV ET L'APPROCHE PROPOSÉE

	SURE-LET	TV	proposée
Lena	29.04	29.22	29.69
Cameraman	28.17	29.61	29.47

approche donne 29.47 dB qui est 1.3 dB de plus que celui obtenu par SURE-LET. La bonne performance de notre approche par rapport à SURE-LET peut être expliquée par le fait qu'elle estime dix paramètres différents pour les coefficients des dix sous-bandes.

En comparant les PSNR de notre approche et de l'approche TV, nous pouvons voir que pour *Lena*, notre approche donne un PSNR 0.47 dB de plus que celui de l'approche TV, mais pour *Cameraman*, c'est l'approche TV qui donne un PSNR 0.14 dB de plus que celui de notre approche. En fait, l'image *Cameraman* est plutôt régulière par morceaux et l'*a priori* TV marche très bien pour ce type d'images. Par contre, dans *Lena*, il y a plus de détails. Dans ce cas, c'est l'approche basée sur un *a priori* dans le domaine des ondelettes qui est le plus adapté.

5.7 Conclusion

La première contribution de ce travail est l'algorithme 5.3.1 qui nous a permis d'introduire une autre vision du bayésien variationnel. Grâce à ce résultat nous avons pu définir de nouveaux algorithmes beaucoup plus rapides et plus aisément utilisables en grande dimension que l'état de l'art, comme on peut le voir dans la thèse de Yuling Zheng. Cette méthode de reconstruction est de plus aisément utilisable pour des mélanges scalaires de gaussiennes. Elle ouvre un grand champ d'investigations possibles dont une partie sera détaillée dans le projet.

BILAN (THÈSE/PUBLICATION)

- 2 publications dans des revues internationales à comité de lecture ([J1],[J3]).
- 5 articles dans des conférences internationales à comité de lecture([C4],[C6],[C7],[C9],[C11]), 3 dans des conférences nationales ([CN1],[CN2],[CN4]) et 1 chapitre de livre ([O1]).
- Encadrement d'une thèse.
- Encadrement de 2 stages de M1 et un stage de M2.

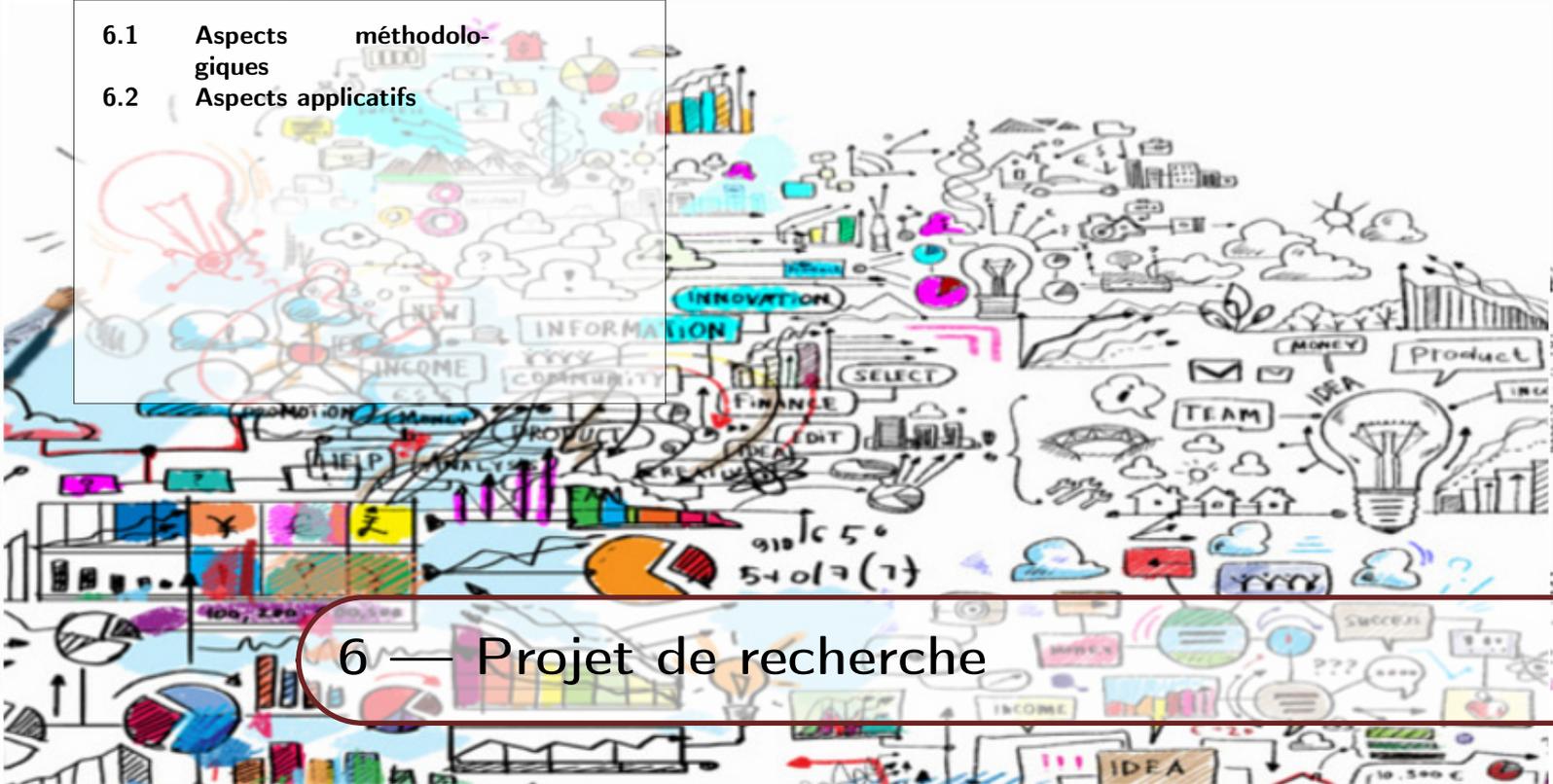


FIGURE 5.3 – (a) Image floue, images reconstruites par (b) SURE-LET [PBBC09] (c) TV [ZFR15b] (d) l'approche proposée.

Quatrième partie

Projet de recherche

- 6.1 Aspects méthodologiques
- 6.2 Aspects applicatifs



6 — Projet de recherche

Mon projet de recherche se situe dans la continuité des deux derniers axes présentés ici, à savoir le développement et l'application de méthodes de reconstruction dans le cas de problèmes inverses mal-posés. Afin de clarifier la présentation je séparerai ce projet en deux parties, l'une centrée sur les développements théoriques et méthodologiques, en particulier sur les aspects non paramétriques, et l'autre sur la partie plus applicative.

6.1 Aspects méthodologiques

En ce qui concerne les pistes théoriques sur lesquelles s'appuie ce projet, elles sont de deux types, correspondant aux deux approches théoriques que j'ai étudiées jusqu'à aujourd'hui. La première est la généralisation de l'étude générique du comportement des estimateurs non paramétriques dans des espaces fonctionnels donnés. L'autre partie importante est basée essentiellement sur le développement des méthodes bayésiennes de la partie 5.

6.1.1 Comportement générique des estimateurs non paramétriques

Les travaux présentés dans la partie 4.4 sont de premiers pas vers le développement d'une théorie générale permettant de comparer les estimateurs non-paramétriques entre eux. En effet, comme indiqué plus haut, étudier les performances minimax peut s'avérer mal adapté dans certains cas. Une approche plus récente est celle des maxisets, [KP02], pour laquelle on détermine les espaces maximaux sur lesquels les estimateurs atteignent une vitesse de convergence donnée. Là encore, cette théorie n'est pas complètement satisfaisante puisqu'elle nécessite le choix d'une vitesse de convergence.

Dans la définition du minimax, le critère d'optimalité porte sur tous les estimateurs définis dans un espace fonctionnel. Le problème de ce critère d'optimalité est qu'il donne

le risque maximum atteint sur cet espace. Mais nous avons vu que dans certains cas ce risque "pire" et le risque générique coïncident. Il serait pour autant naturel de se demander si ces résultats sont fortuits ou s'ils correspondent, comme on peut s'y attendre, pour tous les estimateurs. Pour répondre à cette question, et pour définir aussi un critère théorique d'optimalité plus réaliste, je me propose de définir un critère correspondant à la meilleure vitesse atteinte pour presque toutes les fonctions de l'espace.

Pour cela, je pense pouvoir m'appuyer sur les lemmes généraux de la théorie du minimax. En effet, pour déterminer une vitesse minimax dans un espace donné on peut s'appuyer soit sur le lemme d'Assouad, [Ass83] qui se base sur les tests d'hypothèses soit sur la minoration du risque minimax par le risque bayésien. Il semble alors logique de démontrer des équivalents dans un cadre générique, et d'en déduire des minoration du risque valables pour presque toutes les fonctions d'un espace donné.

Une telle théorie, pourra ensuite se généraliser à des cadres plus proches des phénomènes observés dans les applications, tels que la théorie du minimax adaptatif et surtout la généralisation aux problèmes inverses.

6.1.2 Généralisation du bayésien variationnel exponentialisé

Une première piste de recherche liée à la méthodologie du bayésien variationnel que je vais exploiter est la généralisation de cette théorie à d'autres types de lois approchantes. En effet dans les travaux effectués jusqu'à présent, les lois de probabilité considérées étaient soit discrètes dans [KW97b], soit à densité dans nos travaux. Mais l'ensemble des mesures de probabilité est beaucoup plus grand que cela. Par exemple pour de nombreuses applications on a besoin d'un modèle prenant en compte exactement la parcimonie, modélisée par un mélange Bernoulli-Gaussien. Ce type de lois mixtes nécessitent donc une généralisation des travaux antérieurs.

Une autre généralisation envisagée est l'étude et l'amélioration de l'algorithme bayésien variationnel présenté dans la prise en compte des corrélations. En effet, dans les travaux que nous avons menés, nous avons obtenu de fortes accélérations du bayésien variationnel dans le cas d'une séparabilité totale. Dans le cas de la séparation partielle présentée dans l'exemple 5.5.2 pour laquelle les pixels de l'image ne sont pas du tout séparés, l'utilisation de cet algorithme n'apporte rien par rapport à la méthode standard. Néanmoins, une étude intéressante serait de regarder différents niveaux de corrélation afin de déterminer à partir duquel l'apport de la méthode est significatif. Cette étude pourrait être menée en appliquant notre méthode à des champs de Markov pour lesquels on peut jouer sur la corrélation.

Finalement, la principale limitation de la méthode du bayésien variationnel est le cadre de lois *a priori* conjuguées avec la vraisemblance. Même dans les méthodes développées, cette hypothèse est nécessaire en pratique puisqu'elle permet de passer de l'optimisation d'une loi à une optimisation de paramètres, plus simple à mettre en œuvre.

Plusieurs travaux ont été développés pour essayer de sortir de cette contrainte, mais ils faisaient appel à des approximations ou des contraintes supplémentaires sur les lois approchantes. Une piste de recherche à moyen terme pour sortir de ce cadre contraint serait d'utiliser les algorithmes récents d'optimisation basés sur l'opérateur proximal qui permettent de considérer des fonctionnelles plus complexes. Une autre piste envisagée est d'échantillonner les lois approchantes qui permettrait d'utiliser des méthodes numériques, moins limitées par ces contraintes.

6.1.3 Bayésien variationnel non-paramétrique

Enfin une piste qui m'intéresse, vraisemblablement à plus long terme, est le développement de méthodes bayésiennes variationnelles dans un cadre non-paramétrique. Les méthodes développées jusqu'à présent s'appliquent seulement à des densités de probabilités dans un cadre paramétrique, pour l'estimation d'un nombre fini de paramètres. J'aimerais adapter ces travaux à un cadre non paramétrique, celui-ci offrant une plus grande flexibilité mais faisant aussi intervenir une infinité de paramètres. De plus l'approche non-paramétrique permet de s'affranchir du choix du nombre de pixels, de coefficients de base etc... permettant de représenter l'objet. La principale difficulté dans ce cadre est la nature même des objets étudiés. Cette fois les mesures concernées sont données par des processus continus, plus délicats à manipuler. Pour l'instant la grande majorité des travaux sur le bayésien non-paramétrique font intervenir des méthodes de reconstruction de type MCMC plus coûteuses en temps de calcul. Dans le domaine de l'apprentissage, des travaux de [BJ⁺06, DMGT09] présentent l'application du bayésien variationnel pour certaines lois non-paramétriques qui peuvent être vues comme des modèles de mélange ayant une infinité de composantes. Cependant dans ces deux cas, la loi approchante considérée est finie-dimensionnelle, le nombre de composantes considérées dans le mélange étant choisi à l'avance. *A contrario* dans [MTC11], le modèle est plus simple mais le nombre de composantes est estimé aussi. Un premier travail envisagé est de partir de [BJ⁺06] et d'estimer le nombre de composantes dans le modèle du processus de Dirichlet puis d'appliquer ces méthodes à la résolution de problèmes inverses en traitement du signal. Dans un deuxième temps, l'idée serait de déterminer une approche générale valable pour une grande classe de processus stochastiques. J'envisage donc de développer un algorithme de bayésien variationnel du type présentés dans la partie 5 mais adapté à des *a priori* non paramétriques.

6.2 Aspects applicatifs

Ces dernières années je me suis intéressée plus spécifiquement aux apports du traitement du signal dans différents domaines applicatifs, grâce à des collaborations avec des experts de ces différents domaines. Je souhaite non seulement continuer à développer les pistes actuelles, présentées dans la partie 4, qui sont essentiellement portées par l'astrophysique ou les problèmes inverses des ondes mais aussi développer de nouvelles

collaborations dans des applications plus proches du bio-médical en imagerie TEP (Tomographie par Emission de Positons) ou IRM (Imagerie par Résonance Magnétique).

6.2.1 Domaine de l'astrophysique

Dans ce domaine j'ai développé deux grands axes de collaboration qui sont liés à deux projets phares, le JWST avec le L2S et l'IAS et la détection d'ondes gravitationnelles avec le LAMA et l'APC. Les travaux déjà menés ouvrent d'autres pistes de généralisation et d'amélioration des méthodes existantes.

Imagerie multi-spectrale

Actuellement le travail effectué dans la thèse d'Amine Hadj-Youcef est surtout porté par la construction d'un modèle direct pour les données attendues dans l'imageur MIRI prenant en compte le côté multi-spectral. Ce travail ouvre la voie au développement de méthodes de reconstruction dédiées, en particulier des méthodes parcimonieuses. On pourra ensuite généraliser ces méthodes à des bi-modèles permettant de prendre en compte différents types d'objets spatiaux tels que les étoiles et les nuages gazeux. Le modèle présenté actuellement est basé sur une approximation linéaire par morceaux de la composante spectrale. Le modèle étudié pourrait aussi être généralisé par exemple à des approches de type décomposition sur des dictionnaires.

Détection d'ondes gravitationnelles

Le travail effectué dans le cadre du projet ANR Wavegraph a pour objet la détection d'ondes gravitationnelles dans le plan temps-fréquence grâce en particulier à des formes d'ondes particulières. En plus de la continuation de ce travail pour la détection d'événements de plus faibles signatures temps-fréquence, une autre approche est basée sur l'utilisation des méthodes bayésiennes. En effet, un défi qui intéresse actuellement les spécialistes est le développement de méthodes rapides permettant une détection et une analyse temps réel. L'intérêt est de pouvoir pointer des télescopes dans la direction d'arrivée de l'onde gravitationnelle et ainsi de voir directement l'événement correspondant. Dans ce contexte, les méthodes privilégiées actuellement sont des méthodes bayésiennes, en particulier l'algorithme BayesWave [CL15], basées sur du MCMC. Plusieurs pistes d'accélération utilisant le bayésien variationnel sont ouvertes. Une première idée serait d'adapter l'algorithme BayesWave, en utilisant par exemple les mêmes *a priori* dans un cadre bayésien variationnel. Une autre approche serait de regarder une décomposition sur des dictionnaires temps-fréquence afin de déterminer les caractéristiques du signal reçu.

6.2.2 Domaine de l'imagerie bio-médicale

En ce qui concerne les applications en imagerie liées au bio-médical, là encore plusieurs pistes sont envisagées. L'une d'elle en collaboration avec le GEEPS traite de l'imagerie micro-ondes tandis que l'autre, sans doute à plus long terme, concerne essentiellement l'imagerie IRM.

Problèmes inverses des ondes

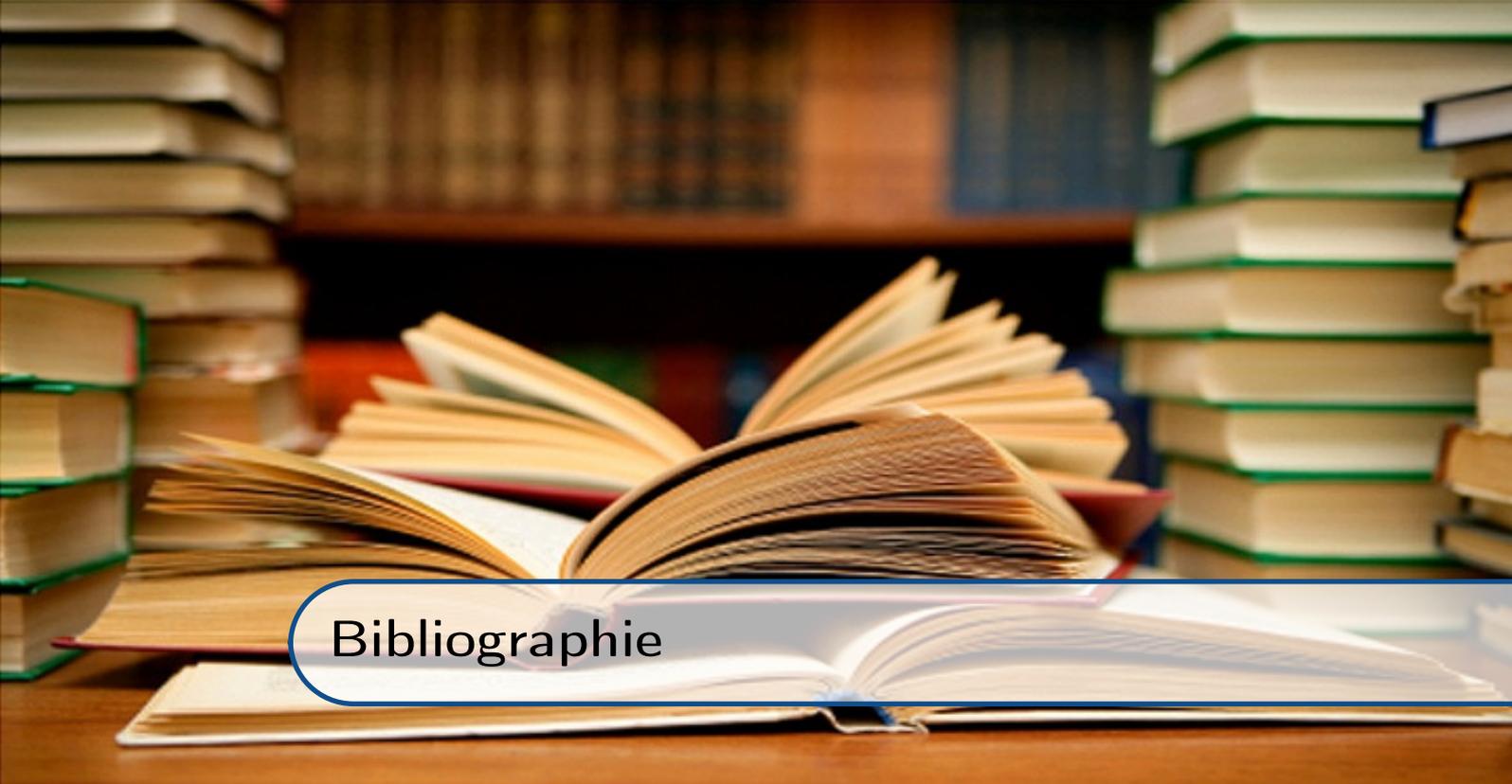
Les travaux entrepris dans le cadre de la thèse d'Hidayet Zaimaga permettent de tester les méthodes basées sur la parcimonie à la reconstruction en imagerie micro-ondes. Cependant ils concernent la parcimonie dans le domaine direct des pixels. Une piste d'amélioration de ces résultats est l'utilisation de dictionnaires, basés par exemple sur des ondelettes, afin de trouver la représentation la plus parcimonieuse possible. On s'intéresse également à la généralisation 3D prenant en particulier en compte les problèmes d'anisotropie.

Une autre piste d'amélioration de la reconstruction dans ce contexte est l'utilisation des méthodes bayésiennes et en particulier les méthodes développées dans la thèse de Yuling Zheng. Une fois encore, ces méthodes devront être adaptées, d'une part, au problème non-linéaire complexe inhérent et, d'autre part, aux différentes applications, telles que la détection du cancer du sein ou le contrôle non-destructif.

Imagerie TEP/IRM

Finalement, un problème que j'aborderai est l'amélioration des images IRM et TEP afin de diminuer à la fois le temps d'exposition des patients et l'injection d'agents de contraste. Dans ce cas, et plus spécialement dans les applications à l'imagerie fonctionnelle cérébrale, l'évolution temporelle est porteuse d'informations très importantes. Or dans les approches classiques utilisées pour reconstruire les données TEP, la résolution temporelle est très mauvaise à cause du faible nombre de photons détectés. Dans ce contexte plusieurs pistes peuvent être explorées. Une première est l'utilisation des méthodes bayésiennes ainsi que leurs développements non paramétriques afin de tirer au mieux profit du faible nombre de données pour reconstruire un grand nombre de pixels, comme dans [FBC⁺13], en contournant la question de l'échantillonnage.

Une autre piste envisagée est une prise en compte de la parcimonie pour ce type d'imagerie. Ce problème est largement étudié, en particulier par [BCC⁺16] où ils proposent des schémas d'acquisition basés sur de l'échantillonnage compressé. Une autre façon d'intégrer la parcimonie dans ce type d'imagerie que l'on pourrait envisager est de l'utiliser directement dans les mécanismes de restauration du contraste. Ces approches pourraient être développées aussi pour des problèmes d'imagerie multi-modale.



Bibliographie

- [AAA⁺16] B. P Abbott, R. Abbott, TD Abbott, MR Abernathy, F. Acernese, et al. Observation of gravitational waves from a binary black hole merger. *Physical review letters*, 116(6) :061102, 2016.
- [ADGS11] G Aniano, BT Draine, KD Gordon, and K Sandstrom. Common-resolution convolution kernels for space-and ground-based telescopes. *Publications of the Astronomical Society of the Pacific*, 123(908) :1218, 2011.
- [ALP02] A. Antoniadis, D. Leporini, and J.-C. Pesquet. Wavelet thresholding for some classes of non-Gaussian noise. *Statistica Neerlandica*, 56(4) :434–453, 2002.
- [Ass83] P. Assouad. Deux remarques sur l'estimation. *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique*, 296(23) :1021–1024, 1983.
- [ATB03] A. Achim, P. Tsakalides, and A. Bezerianos. SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. Remote Sens.*, 41(8) :1773–1784, 2003.
- [AvdB04] A. Abubakar and P. van den Berg. Iterative forward and inverse algorithms based on domain integral equations for three-dimensional electric and magnetic objects. *J. Comput. Phys.*, 195(1) :236–262, 2004.
- [Ban31] S. Banach. Über die Baire'sche Kategorie gewisser Funktionenmengen. *Studia Math.*, 3 :174–179, 1931.
- [BCC⁺16] C. Boyer, N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss. On the generation of sampling schemes for magnetic resonance imaging. *SIAM Journal on Imaging Sciences*, 9(4) :2039–2072, 2016.

- [BD06] J. M. Bioucas-Dias. Bayesian wavelet-based image deconvolution : a GEM algorithm exploiting a class of heavy-tailed priors. *IEEE Trans. Image Process.*, 15(4) :937–951, 2006.
- [BFDK14] F. Bassi, A. Fraysse, E. Dupraz, and M. Kieffer. Rate-distortion bounds for wyner–ziv coding with gaussian scale mixture correlation noise. *IEEE Transactions on Information Theory*, 60(12) :7540–7546, 2014.
- [BGML⁺15] P. Bouchet, M. García-Marín, P-O Lagage, J. Amiaux, J-L Auguères, E. Bauwens, J. Blommaert, CH Chen, ÖH Detre, D. Dicken, et al. The Mid-Infrared Instrument for the James Webb Space Telescope, III : MIRIM, The MIRI Imager. *Publications of the Astronomical Society of the Pacific*, 127(953) :612, 2015.
- [Bir83] L. Birgé. Approximation dans les espaces métriques et théorie de l’estimation. *Z. Wahrsch. Verw. Gebiete*, 65(2) :181–237, 1983.
- [BJ⁺06] D. Blei, M. Jordan, et al. Variational inference for dirichlet process mixtures. *Bayesian analysis*, 1(1) :121–143, 2006.
- [BM95] F. Bergeaud and S. Mallat. Matching pursuit of images. In *Image Processing, 1995. Proceedings., International Conference on*, volume 1, pages 53–56. IEEE, 1995.
- [BMK11] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational Bayesian super resolution. *IEEE Trans. Image Process.*, 20(4) :984–999, 2011.
- [CDS01] S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1) :129–159, 2001.
- [CGLS08] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders. Variational Bayesian image restoration based on a product of t -distributions image prior. *IEEE Trans. Image Process.*, 17(10) :1795–1805, October 2008.
- [Cho02] R. A. Choudrey. *Variational Methods for Bayesian Independent Component Analysis*. PhD thesis, University of Oxford, 2002.
- [Cho10] E. Chouzenoux. *Recherche de pas par Majoration-Minoration. Application à la résolution de problème inverse*. PhD thesis, École Centrale de Nantes, 2010.
- [Chr72] J.P.R. Christensen. On sets of Haar measure zero in Abelian Polish groups. *Israel J. Math.*, 13 :255–260, 1972.
- [CL69] E. E. Cragg and A. V. Levy. Study on a supermemory gradient method for the minimization of functions. *Journal of Optimization Theory and Applications*, 4 :191–205, 1969. 10.1007/BF00930579.
- [CL15] N/ Cornish and T. Littenberg. Bayeswave : Bayesian inference for gravitational wave bursts and instrument glitches. *Classical and Quantum Gravity*, 32(13) :135012, 2015.
- [CRT06] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2) :489–509, 2006.

- [Dau88] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7) :909–996, 1988.
- [DBENP09] F. Di Benedetto, C. Estatico, J. G Nagy, and M. Pastorino. Numerical linear algebra for nonlinear microwave imaging. *Electron. Trans. Numer. Anal.*, 33 :105–125, 2009.
- [Dem89] G. Demoment. Image reconstruction and restoration : Overview of common estimation structure and problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12) :2024–2036, 1989.
- [DJ94] D. Donoho and I. Johnstone. Minimax risk over l_p -balls for l_q -error. *Probab. Theory Related Fields*, 99(2) :277–303, 1994.
- [DJ98] D. Donoho and I. Johnstone. Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26(3) :879–921, 1998.
- [DJJ91] I. Daubechies, S. Jaffard, and J-L. Journé. A simple wilson orthonormal basis with exponential decay. *SIAM Journal on Mathematical Analysis*, 22(2) :554–573, 1991.
- [DJKP97] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam*, pages 183–218. Springer, New York, 1997.
- [DMGT09] F. Doshi, K.t Miller, J. Gael, and Y. Teh. Variational inference for the indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, pages 137–144, 2009.
- [Don06] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4) :1289–1306, 2006.
- [DYZ13] W. Deng, W. Yin, and Y. Zhang. Group sparse optimization by alternating direction method. In *SPIE Optical Engineering+ Applications*, pages 88580R–88580R-15. International Society for Optics and Photonics, 2013.
- [EKL06] T. Eltoft, T. Kim, and T. Lee. On the multivariate Laplace distribution. *IEEE Signal Processing Letters*, 13(5) :300–303, 2006.
- [Eld09] Y. C. Eldar. Generalized SURE for exponential families : Applications to regularization. *IEEE Trans. Signal Process.*, 57(2) :471–481, 2009.
- [FBC⁺13] M. Fall, E. Barat, C. Comtat, T. Dautremer, T. Montagu, and S. Stute. Dynamic and clinical PET data reconstruction : A nonparametric Bayesian approach. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 345–349. IEEE, 2013.
- [FPPP09] A. Fraysse, B. Pesquet-Popescu, and J-C. Pesquet. On the uniform quantization of a class of sparse sources. *IEEE Transactions on Information Theory*, 55(7) :3243–3263, 2009.
- [FR14] A. Fraysse and T. Rodet. A measure-theoretic variational Bayesian algorithm for large dimensional problems. *SIAM Journal on Imaging Sciences.*, 7(4) :2591–2622, 2014.

- [Fra07] A. Fraysse. Generic validity of the multifractal formalism. *SIAM J. Math. Anal.*, 37(2) :593–607, 2007.
- [Fra13] A. Fraysse. Why minimax is not that pessimistic. *ESAIM : Probability and Statistics*, 17 :472–484, 2013.
- [GHW79] Gene H. Golub, M. Heath, and Grace Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. 21(2) :215–223, May 1979.
- [GL10] N Geis and D Lutz. Herschel/PACS modelled point-spread functions. http://herschel.esac.esa.int/twiki/pub/Public/PacsCalibrationWeb/PACSPSF_PICC-ME-TN-029_v2.0.pdf, 2010.
- [Goo05] J. Goodman. *Introduction to Fourier optics*. Roberts and Company Pub., 2005.
- [Gra72] R. Gray. Conditional rate-distortion theory. Technical report, Stanford University, 1972.
- [Haa10] A. Haar. Zur Theorie der orthogonalen Funktionensysteme. (Erste Mitteilung.). *Math. Ann.*, 69 :331–371, 1910.
- [Han92] P.C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. 34 :561–580, 1992.
- [HKBC15] W. Hachicha, M. Kaaniche, A. Beghdadi, and F. Cheikh. Efficient inter-view bit allocation methods for stereo image coding. *IEEE Transactions on Multimedia*, 17(6) :765–777, 2015.
- [Idi01] J. Idier, editor. *Approche bayésienne pour les problèmes inverses*. Traité IC2, Série traitement du signal et de l’image, Hermès, Paris, 2001.
- [IH81] I. Ibragimov and R. Has’minskii. *Statistical estimation. Asymptotic theory*. Springer Verlag, 1981.
- [JGJS99] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An Introduction to variational Methods for Graphical Models. *Machine Learning*, 37(2) :183–233, 1999.
- [JM06] S. Jana and P. Moulin. Optimality of klt for high-rate transform coding of gaussian vector-scale mixtures : Application to reconstruction, estimation, and classification. *IEEE Trans. on Info. Theory*, 52(9) :4049–4067, 2006.
- [JW91] F. C. Jeng and J. W. Woods. Compound Gauss-Markov random fields for image estimation. *IEEE Trans. Signal Process.*, 39(3) :683–697, 1991.
- [JXGJ12] Shi Jun, Zhang Xiaoling, Xiang Gao, and Jianyu Jianyu. Signal processing for microwave array imaging : Tdc and sparse recovery. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11) :4584–4598, 2012.
- [KFPPP14] M. Kaaniche, A. Fraysse, B. Pesquet-Popescu, and JC. Pesquet. A bit allocation method for sparse source coding. *IEEE Transactions on Image processing*, 23(1) :137–152, 2014.

- [KP02] G. Kerkyacharian and D. Picard. Minimax or maxisets? *Bernoulli*, 8(2) :219–253, 2002.
- [KW97a] J. Kivinen and M. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Inform. and Comput.*, 132(1) :1–63, 1997.
- [KW97b] J. Kivinen and M. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1) :1–63, 1997.
- [KYMM08] S. Klimentko, I Yakushin, A Mercer, and G. Mitselmakher. A coherent method for detection of gravitational wave bursts. *Classical and Quantum Gravity*, 25(11) :114029, 2008.
- [LG04] C. L. Likas and N. P. Galatsanos. A variational approach for Bayesian blind image deconvolution. *IEEE Trans. Signal Process.*, 52(8) :2222–2233, 2004.
- [Lue69] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley, 1969.
- [Mac95] D. J. C. MacKay. Ensemble learning and evidence maximization. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.54.4083>, 1995.
- [Mal89] S. Mallat. A theory for multiresolution signal decomposition : The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7) :674–693, 1989.
- [Maz31] S. Mazurkiewicz. Sur les fonctions non d?rivables. *Stud. Math.*, 3 :92–94, 1931.
- [MC69] A. Miele and J. W. Cantrell. Study on a memory gradient method for the minimization of functions. *J. Optimiz. Theory App.*, 3(6) :459–470, 1969.
- [Mis00] J. W. Miskin. Ensemble learning for independent component analysis. Technical report, in *Advances in Independent Component Analysis*, 2000.
- [ML99] P. Moulin and J. Liu. Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE Trans. on Information Theory*, 45(3) :909–919, 1999.
- [MTC11] K. Morton, P. Torrione, and L. Collins. Variational bayesian learning for mixture autoregressive models with uncertain-order. *IEEE Transactions on Signal Processing*, 59(6) :2614–2627, 2011.
- [NW00] J. Nocedal and S. Wright. *Numerical Optimization*. Series in Operations Research. Springer Verlag, New York, 2000.
- [Par88] G. Parisi. *Statistical field theory*. Addison Wesley, Redwood City, CA, 1988.
- [Pas10] M. Pastorino. *Microwave Imaging*, volume 208. John Wiley & Sons, 2010.
- [PBBC09] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chau. A SURE approach for digital signal/image deconvolution problems. *IEEE Trans. Signal Process.*, 57(12) :4616–4632, 2009.

- [Pin80] M. S. Pinsker. Optimal filtration of square-integrable signals in Gaussian noise. *Problems Inform. Transmission*, 16(2) :52–68, 1980.
- [PKDRW06] J. Palmer, K. Kreutz-Delgado, B. D. Rao, and D. P. Wipf. Variational EM algorithms for non-Gaussian latent variable models. In *Adv. Neural. Inf. Process. Syst. 18*, pages 1059–1066. MIT Press, 2006.
- [POM12] L. Poli, G. Oliveri, and A. Massa. Microwave imaging within the first-order born approximation by means of the contrast-field bayesian compressive sensing. *IEEE Trans. Antennas Propagat.*, 60(6) :2865–2879, 2012.
- [PSWS03] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.*, 12(11) :1338–1351, 2003.
- [Ric65] J. Richmond. Scattering by a dielectric cylinder of arbitrary cross section shape. *IEEE Trans. Antennas Propagat.*, 13(3) :334–341, May 1965.
- [ROF92] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D : Nonlinear Phenomena*, 60(1) :259–268, 1992.
- [RT06] R. Ramlau and G. Teschke. A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numer. Math.*, 104(2) :177–203, 2006.
- [Rud87a] W. Rudin. *Real and Complex Analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
- [Rud87b] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, 1987.
- [ŠQ06] V. Šmídl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer, 2006.
- [SS05] Z. J. Shi and J. Shen. A new super-memory gradient method with curve search rule. *Appl. Math. Comput.*, 170(1) :1–16, 2005.
- [Ste81] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, pages 1135–1151, 1981.
- [Str83] J.O. Stromberg. A modified Franklin system and higher-order spline systems on \mathbb{R}^n as unconditional bases for Hardy spaces. In Beckner, editor, *Conference in Harmonic Analysis in honor of Antony Zygmund*, volume 2, pages 475–493, 1983.
- [TB93] P. Tseng and D. P. Bertsekas. On the convergence of the exponential multiplier method for convex programming. *Mathematical Programming*, 60(1-3) :1–19, 1993.
- [Tik63] A. Tikhonov. Regularization of incorrectly posed problems. *Soviet. Math. Dokl.*, 4 :1624–1627, 1963.
- [VDP⁺13] S. Vaiter, C. A. Deledalle, G. Peyré, C. Dossal, and J. Fadili. Local behavior of sparse analysis regularization : Applications to risk estimation. *Appl. Comput. Harmon. Anal.*, 35(3) :433–451, 2013.

- [WS00] M. J. Wainwright and E. P. Simoncelli. Scale mixtures of gaussians and the statistics of natural images. *Advances in neural information processing systems*, 12(1) :855–861, 2000.
- [WW93] E. Wesfreid and M. Wickerhauser. Adapted local trigonometric transforms and speech processing. *IEEE Transactions on Signal Processing*, 41(12) :3596–3600, 1993.
- [Wyn78] A. Wyner. The rate-distortion function for source coding with side information at the decoder-ii : General sources. *Information and Control*, 38(1) :60 – 80, 1978.
- [WZSL13] S. Wei, X. Zhang, J. Shi, and K. Liao. Sparse array microwave 3-d imaging : Compressed sensing recovery and experimental study. *Progress In Electromagnetics Research*, 135 :161–181, 2013.
- [YL06] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J R Stat Soc Series B Stat Methodol.*, 68(1) :49–67, 2006.
- [Zam96] R. Zamir. The rate loss in the Wyner-Ziv Problem. *IEEE Trans. on Information Theory*, 42(6) :2073–2084, 1996.
- [ZFR15a] Y. Zheng, A. Fraysse, and T. Rodet. Efficient variational bayesian approximation method based on subspace optimization. *IEEE Transactions on Image Processing*, 24(2) :681–693, 2015.
- [ZFR15b] Y. Zheng, A. Fraysse, and T. Rodet. Efficient variational Bayesian approximation method based on subspace optimization. *IEEE Trans. Image Process.*, 24(2) :681–693, 2015.
- [ZL16] H. Zaimaga and M. Lambert. Soft shrinkage thresholding algorithm for nonlinear microwave imaging. *J. Phys. Conf. Ser.*, 756(1) :012011, 2016.

Résumé :

Les activités de recherche présentées dans cette habilitation sont portées par les méthodes temps-fréquence et leur apport méthodologique au traitement du signal et de l'image, plus particulièrement dans le contexte de problèmes inverses mal posés. Dans ce cadre, je me suis intéressée à deux apports importants de cet outil. Le premier est évidemment la parcimonie, qui est l'un des intérêts majeurs de la décomposition temps-fréquence. Un autre intérêt est porté par l'estimation non paramétrique et le lien entre les signaux naturels et des espaces fonctionnels donnés. Même si dans les faits les signaux ont tous une représentation finie, pouvoir les considérer comme des objets de dimension infinie offre plus de flexibilité et permet de s'affranchir de contraintes telles que le choix de l'échantillonnage ou du nombre de paramètres à considérer.

C'est donc par le biais des apports de ces deux aspects méthodologiques que j'ai étudié trois grands axes thématiques du traitement du signal. Le premier axe porte sur la théorie de l'information pour le codage de données parcimonieuses. Le second axe s'appuie sur l'étude et l'application de méthodes déterministes d'estimation et de reconstruction, en particulier des méthodes par seuillage d'ondelettes, dans le cadre de problèmes inverses. Finalement, la dernière partie de ce mémoire traite du développement de méthodes bayésiennes variationnelles pour les problèmes inverses.