



MÉMOIRE

pour obtenir

l'Habilitation à Diriger des Recherches

soutenue le 09/01/2019

par

Caroline CHAUX

**De la modélisation de problèmes directs en
traitement d'image à la résolution des problèmes
inverses**

Composition de la commission d'examen :

Rapporteurs : Jean-François AUJOL
Julie DELON
Nick KINGSBURY

Examineurs : Rémi GRIBONVAL
Gabriele STEIDL
Bruno TORRÉSANI

*À mes grands-parents
À Baptiste, Pauline et Xavier*

Table des matières

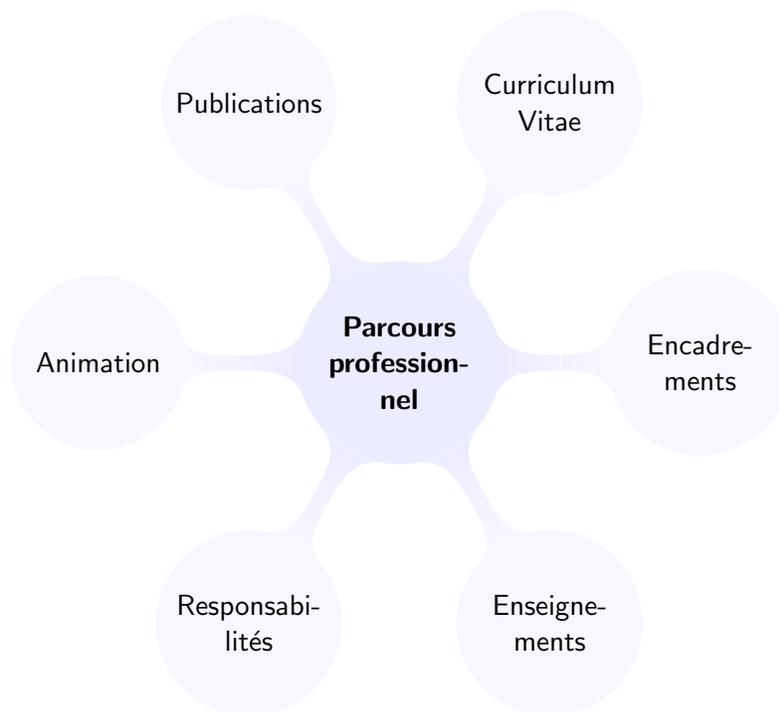
I	Parcours professionnel	9
1	Curriculum vitae	11
1.1	Renseignements généraux	11
1.2	Formation initiale	12
1.3	Expérience professionnelle	12
1.4	Activités d'enseignements	12
1.5	Activités d'encadrement	14
1.5.1	Encadrement de stagiaires	14
1.5.2	Encadrement de doctorants	14
1.6	Management de la recherche	15
1.7	Activités d'animation scientifiques	16
1.7.1	Organisation de workshop	16
1.7.2	Organisation d'évènement de vulgarisation scientifique	16
1.7.3	Participation (cours) à des écoles	16
1.7.4	Organisation de journées GdR	16
1.7.5	Organisation de formations	17
1.7.6	Autres	17
1.8	Activités de relecture	17
1.9	Activités administratives	17
1.10	Distinction	18
1.11	Activités de recherche	18
1.11.1	Parcours dans la recherche	18
1.11.2	Collaborations	19
1.11.3	Participation à des contrats de recherches	20
1.11.4	Dépôt de brevet	21
1.11.5	Logiciels	21
2	Liste de publications	23
2.1	Articles de revues à comité de lecture publiés ou acceptés	23
2.2	Articles de conférences invités avec actes	24
2.3	Articles de conférences avec actes	25
2.4	Chapitre de livre	28
II	Travaux de recherche	29
3	Introduction	31

4	Modèles et formulation du problème direct	33
4.1	Formulation directe et objectif	33
4.1.1	Un modèle général	33
4.1.2	... et de multiples problèmes sous-jacents	34
4.2	Cadre bayésien / modélisation variationnelle du problème inverse	37
4.3	Approches proximales	39
4.3.1	Introduction aux opérateurs proximaux	39
4.3.2	Un aperçu des algorithmes proximaux	40
4.4	Conclusion	42
4.5	Quelques publications personnelles en lien avec ce chapitre	43
5	Estimation de paramètres et hyperparamètres	45
5.1	Estimation des paramètres du bruit	45
5.1.1	Contexte	45
5.1.2	Estimation à partir de trames temporelles	46
5.1.3	Estimation à partir d'une seule image	48
5.1.4	Simulations numériques	50
5.1.5	Conclusion	52
5.2	Estimation des paramètres de régularisation	53
5.2.1	Contexte	53
5.2.2	Approche proposée : estimateur au sens du maximum de vraisemblance	53
5.2.3	Difficultés	53
5.2.4	Introduction d'une variable intermédiaire	55
5.2.5	Algorithme de gradient	56
5.2.6	Accélération	56
5.2.7	Simulations numériques	57
5.2.8	Conclusion	60
5.3	Estimation des paramètres du modèle d'acquisition	60
5.3.1	Présentation du problème	60
5.3.2	Formulation variationnelle du problème	62
5.3.3	Algorithme proximal primal-dual	63
5.3.4	Extension aux images sismiques	66
5.3.5	Simulations numériques	69
5.4	Conclusion	70
5.5	Quelques publications personnelles en lien avec ce chapitre	71
6	Optimisation en données complètes	81
6.1	Application : reconstruction d'images	81
6.1.1	Introduction	81
6.1.2	Méthodes proposées	82
6.1.3	Résultats de simulation	84
6.1.4	Conclusion	87
6.2	Application : factorisation tensorielle non-négative	87
6.2.1	Choix de la fonction coût et calcul des dérivées partielles	89
6.2.2	Calcul du préconditionnement et des opérateurs proximaux associés	89
6.2.3	Simulations numériques	90
6.3	Quelques publications personnelles en lien avec ce chapitre	99

III Conclusions et perspectives	101
7 Conclusion	103
7.1 Conclusions	103
8 Perspectives	105
8.1 La suite	105
8.2 Quelques directions en cours d'étude ou de développement	106
8.2.1 Données manquantes en audio	106
8.2.2 Modèles graphiques	107
8.2.3 Spectroscopie résonance magnétique nucléaire (RMN)	108
8.2.4 Stein pour l'estimation des paramètres	109
Liste des acronymes	115
Bibliographie	117

Partie I

Parcours professionnel



Chapitre 1

Curriculum vitae

Sommaire

1.1 Renseignements généraux	11
1.2 Formation initiale	12
1.3 Expérience professionnelle	12
1.4 Activités d'enseignements	12
1.5 Activités d'encadrement	14
1.6 Management de la recherche	15
1.7 Activités d'animation scientifiques	16
1.8 Activités de relecture	17
1.9 Activités administratives	17
1.10 Distinction	18
1.11 Activités de recherche	18

1.1 Renseignements généraux

Nom patronymique : CHAUX

Prénom : Caroline

Date et lieu de naissance : 31 Décembre 1980 à Lyon 9ème

Nationalité : Française

Situation de famille : Mariée, deux enfants.

Fonctions actuelles : Chargée de recherche CNRS, I2M UMR CNRS 7373, Aix-Marseille Université

Adresse postale : CMI, 39 Rue Frédéric Joliot Curie, 13453 Marseille Cedex 13

Adresse mél : caroline.chaux@univ-amu.fr

Page web : <https://www.i2m.univ-amu.fr/~caroline.chaux/>

1.2 Formation initiale

2003 à 2006 : Doctorat en traitement d'images

Université de Marne-la-Vallée, Laboratoire d'Informatique de l'Institut Gaspard Monge (CNRS-UMR 8049) sous la direction du Prof. Jean-Christophe Pesquet.

Sujet : "Analyse en ondelettes M-bandes en arbre dual ; application à la restauration d'images".

Soutenance le 13 décembre 2006 à l'Université de Marne-la-Vallée.

Jury : M. Barlaud (Président), J. Zerubia et P. Abry (rapporteurs), A. Benazza, L. Duval et A. Mohammad-Djafari (examineurs), J.-C. Pesquet (directeur de thèse).

2002 à 2003 : DEA Signal et communications numériques

Université de Nice Sophia-Antipolis, mention Bien.

2000 à 2003 : Diplôme d'ingénieur en télécommunications

I.S.I.T.V. (Institut des Sciences de l'Ingénieur de Toulon et du Var).

1998 à 2000 : Classes préparatoires MP SI

Lycée Lalande, Bourg en Bresse.

1.3 Expérience professionnelle

depuis Sept. 2012 : Chargée de recherche (CR1) CNRS (section 07)

Aix-Marseille Université, Institut de Mathématiques de Marseille (I2M ex LATP/IML) (UMR CNRS 7373).

Fév. 2012 à Août 2012 : Chargée de recherche (CR1) CNRS (section 07)

Université Paris-Est, Laboratoire d'Informatique Gaspard Monge (CNRS-UMR 8049).

2007 à 2011 : Chargée de recherche (CR2) CNRS (section 07)

Université Paris-Est, Laboratoire d'Informatique Gaspard Monge (CNRS-UMR 8049).

2006 à 2007 : Post-Doctorat en traitement d'images

INRIA de Sophia-Antipolis, projet ARIANA, sujet : restauration d'images 3D de microscopie confocale.

Mars à Août 2003 : Stage de DEA, Institut Français du Pétrole, Rueil-Malmaison

Sujet : Analyse et débruitage d'images sismiques.

Juin à Août 2002 : Stage ingénieur, Laboratoire de traitement du signal de l'ETSI

Industriales y Telecomunicación, Université de Cantabria, Santander (Espagne)

Sujet : Analyse numérique d'enregistrements polysomnographiques.

1.4 Activités d'enseignements

Cours magistraux et travaux pratiques en traitement du signal (2016-auj.) : À l'université de Toulon

Niveau : CNAM - formation d'ingénieur informatique Sciences et Technologies des Médias Numériques - 1ère année.

Volume horaire : 14h ETD /an

Cours magistraux et travaux pratiques sur le thème "Problèmes inverses" (2015-auj.) :

À Aix-Marseille Université

Niveau : Master mathématiques et applications 2ème année.

Volume horaire : 15h ETD /an

Travaux dirigés et projet d'algorithmique (langage C) (2015-2016) : **À Aix-Marseille Université**

Niveau : Licence mathématiques 2ème année

Volume horaire : 30h ETD / an

Cours magistraux et travaux dirigés sur le thème "Traitement numérique des signaux"

(2014-auj.) : **À Aix-Marseille Université**

Niveau : Master mathématiques et applications 1ère année

Volume horaire : 36h ETD/an

Cours magistraux et travaux dirigés sur le thème "Image, ondelettes et compression"

(2012-2016) : **À l'Institut des Sciences de l'Ingénieur de Toulon et du Var**

Niveau : école d'ingénieur (ISITV) 3ème année

Volume horaire : 25.5h ETD/an

Cours magistraux sur le thème "Problèmes inverses" (2013) : **À Aix-Marseille Université**

Niveau : Master (MI3S) 2ème année

Volume horaire : 6h ETD/an

Cours magistraux de C++ (2007-2011) : **À l'Université Paris-Est Marne-la-Vallée**

Niveau : école d'ingénieur (IMAC) 2ème année

Volume horaire : 18h ETD/an

Cours magistraux de représentation et filtrage 1D/2D (2010-2011) : **À l'Université Paris-Est Marne-la-Vallée**

Niveau : M2 Recherche

Volume horaire : 30h ETD/an

Travaux dirigés et travaux pratiques de traitement du signal aléatoire (2007-2011) :

À l'Université Paris-Est Marne-la-Vallée

Niveau : master 1ère année

Volume horaire : 16h ETD/an + 6h ETD/an de TP

Travaux dirigés de traitement statistique du signal (2008-2011) : **À l'Université Paris-Est Marne-la-Vallée**

Niveau : master 1ère année

Volume horaire : 15h ETD/an

Cours magistraux et Travaux dirigés de traitement d'images pour l'imagerie médicale

(2011) **À l'Université Paris-Est Créteil**

Niveau : M2 Recherche

Volume horaire : 6 ETD/an

Cours magistraux et Travaux dirigés sur les représentations multirésolution (2009-

2011) **À Telecom SudParis**

Niveau école d'ingénieur 2ème année

Volume horaire : env. 18 ETD/an

Formation continue (2008,2010,2014) : Intervention d'une journée à l'ENSTA pour le stage "Ondelettes et applications" (formation SAE).

1.5 Activités d'encadrement

1.5.1 Encadrement de stagiaires

2018 **Participation à l'encadrement d'un stagiaire de 1ère année de l'ENSTA Bretagne** (1 mois).

Sujet : Traitement de données de spectroscopie à RMN : mise en forme et compression.

2018 **Participation à l'encadrement d'un stagiaire de 2ème année de master Université de Bordeaux** (4 mois).

Sujet : Source separation algorithms for NMR spectroscopy.

2017 **Participation à l'encadrement d'une stagiaire de 2ème année de master Aix-Marseille Université** (6 mois).

Sujet : Reconstruction de phase dans le plan temps-fréquence. Application à l'inpainting audio.

2015 **Participation à l'encadrement d'une stagiaire de 2ème année de master Aix-Marseille Université** (3 mois).

Sujet : Covariance estimation by penalty criterion.

2013 **Encadrement (avec S. Anthoine) d'une stagiaire de 2ème année de master à Aix-Marseille Université.**

Sujet : Déconvolution d'images issues de microscopie biphoton.

2011 **Participation à l'encadrement d'un stagiaire de 3ème année de l'Université de Naples** (6 mois).

Sujet : Elimination des multiples dans les signaux sismiques.

2009 : **Participation à l'encadrement d'un stagiaire de 3ème année Université Polytechnique de Bucarest** (6 mois).

Sujet : Reconstruction d'images en Imagerie par Résonance Magnétique (IRM) parallèle à l'aide d'estimateurs statistiques et d'ondelettes.

2008-2009 : **Encadrement d'un stagiaire de master 2ème année SUP'COM Tunis** (6 mois).

Sujet : Déconvolution d'images satellitaires dans le cas d'une Point Spread Function (PSF) (flou) non stationnaire spatialement.

En collaboration avec Thalès Alenia Space Cannes.

1.5.2 Encadrement de doctorants

2017- : **Participation à l'encadrement** (30%) d'une doctorante (A. M. Kreme), bourse région PACA

Sujet : Inpainting audio dans le plan temps-fréquence.

2015- : **Participation à l'encadrement** (50%) d'une doctorante (T. K. Le), bourse ministérielle ED 184

Sujet : Statistiques et optimisation pour l'estimation de covariances et application à l'imagerie médicale.

- 2013-2017** : **Participation à l'encadrement** (30%) d'une doctorante (T. Thanh Xuan Vu), thèse LabEx Archimède LSIS/I2M
Sujet : Optimisation déterministe et stochastique pour des problèmes de traitement d'images en grande dimension.
Soutenance de thèse : 13 novembre 2017.
- 2011-2015** : **Participation à l'encadrement** (10%) d'une doctorante (M. Quyen Pham), thèse CIFRE IFPEN/LIGM
Sujet : Méthodes d'optimisation convexe pour l'élimination de multiples dans les signaux sismiques.
Soutenance de thèse : 15 juillet 2015.
Situation actuelle : post-doctorat à l'université de Grenoble.
- 2010-2014** : **Participation à l'encadrement** (10%) d'une doctorante (M. El Gheche) du LIGM
Sujet : Méthodes d'optimisation parallèle pour l'analyse et la compression de flux multivues.
Soutenance de thèse : 27 mai 2014.
Situation actuelle : post-doctorat à l'université de Bordeaux.
- 2009-2013** : **Participation à l'encadrement** (30%) d'une doctorante (A. Jezierska) du LIGM (dans le cadre de l'ANR DIAMOND)
Sujet : Restauration d'images de microscopie confocale.
Soutenance de thèse : 13 mai 2013.
Situation actuelle : chercheur au Systems Research Institute, Polish Academy of Sciences.
- 2007-2010** : **Participation à l'encadrement** (40%) d'une doctorante (N. Pustelnik) du LIGM (dans le cadre de l'ANR OPTIMED)
Sujet : Reconstruction d'images de tomographie par émission de positrons (TEP)
Soutenance de thèse : 13 décembre 2010.
Situation actuelle : chargée de recherche CNRS à l'ENS Lyon.

1.6 Management de la recherche

- Editeur associé de la revue IEEE Transactions On Signal Processing depuis Mar. 2018.
- Membre du comité d'évaluation scientifique (CES) 40 (Mathématiques, informatique) de l'ANR en 2016-2017 et 2017-2018.
- Membre (nommé) du Conseil National des Universités (CNU) de Nov. 2014 à Nov. 2015 - Section 61.
- Membre du comité technique IEEE SPTM (Signal Processing Theory and Method) depuis Jan. 2015.
- Membre du conseil d'administration de l'association GRETSI depuis 2012, responsable de la branche recherche depuis 2016.

1.7 Activités d'animation scientifiques

1.7.1 Organisation de workshop

- Co-organisation d'un workshop à Porquerolles "Approximation and optimization in image restoration and reconstruction" <http://www-syscom.univ-mlv.fr/porquerolles2009/> qui a eu lieu du 8 au 12 juin 2009.
- Co-organisation avec J. Idier d'une session spéciale à EUSIPCO 2009 portant sur le thème "Optimization and inverse problems".
- Présidente de session au GRETSI 2011.

1.7.2 Organisation d'évènement de vulgarisation scientifique

- Co-organisation des "Treize Minutes" Marseille qui ont eu lieu les 3 décembre 2013, 4 décembre 2014, 1er décembre 2015, 1er décembre 2016 et 30 novembre 2017 à l'Alcazar, Marseille et qui auront lieu le 4 décembre 2018 au théâtre Joliette. <http://treize.lif.univ-mrs.fr/>

1.7.3 Participation (cours) à des écoles

- Participation à l'école **Analyse Multirésolution du club EEA** au Creusot, Mar. 2009.
Titre du cours (1h30) : Ondelettes et problèmes inverses
- Participation à l'école **Analyse Multirésolution du club EEA** à Auxerre, Jun. 2012.
Titre du cours (1h30) : Débruitage, déconvolution et problèmes inverses
- Participation à l'école **Computational Harmonic Analysis - with Applications to Signal and Image Processing** à Marseille, Oct. 2014.
Titre du cours (1h30) : Inverse problems and optimization.
- Participation à l'après-midi **Hommage à Claude Shannon, le père de la théorie de l'information**, au CIRM, Marseille, Nov. 2016.
<http://www.fr-cirm-math.fr/hommage-claude-shannon.html>
Titre du cours (30min) : L'échantillonnage
- Participation à l'école **thématique GEOMDATA** à Fréjus, 10-14 Sep. 2018.
Titre du cours (3h cours - 3h TP) : Méthodes variationnelles pour l'imagerie

1.7.4 Organisation de journées GdR

- Co-organisation d'une journée conjointe GDR ISIS et GDR MSPC *Optimisation de critères convexes non différentiables pour la résolution de problèmes inverses* qui a eu lieu le 7 décembre 2010 à Paris, <http://www.ceremade.dauphine.fr/~peyre/mspc/mspc-isis-10/>.
- Co-organisation avec R. Gribonval d'une journée GDR ISIS *Transformées multirésolution géométriques* qui a eu lieu le 1 avril 2011 à Paris.
- Co-organisation avec N. Bertaux, J. Fadili et J. Idier d'une journée GDR ISIS *Résolution de problèmes inverses : optimisation et parallélisation* qui a eu lieu le 29 juin 2012 à Paris.

- Co-organisation avec S. Anthoine, V. Emiya, C. Mélot et F. Richard d'une journée GDR ISIS *Traitement du signal de données à valeurs complexes* qui a eu lieu le 12 juin 2013 à Marseille.
- Co-organisation avec V. Emiya et C. Févotte d'une journée GDR ISIS *Estimation de données manquantes* qui a eu lieu le 8 octobre 2015 à Marseille.
- Co-organisation d'une journée inter-GDR avec le GDR ISIS et le GDR OG, qui aura lieu le 8 octobre 2018 à Paris.

1.7.5 Organisation de formations

- Organisation d'une formation "Signal sur graphes ; application aux neurosciences", FRUMAM, Marseille, 6-7 Fev. 2017 et 3-4 Avr. 2017.
http://www.i2m.univ-amu.fr/seminaires_signal_apprentissage/Conf/Feb2017/
- Proposition de formation dans le cadre du CNRS formation entreprises sur "Problèmes inverses : méthodes et applications", Sep. 2018.
<http://cnrsformation.cnrs.fr/stage-18290-Problemes-inverses-methodes-et-applications.html?axe=85>

1.7.6 Autres

- Auteur de fiches PLUME (Promouvoir les Logiciels Utiles Maîtrisés et Economiques dans l'Enseignement Supérieur et la Recherche), <http://www.projet-plume.org>.

1.8 Activités de relecture

- **Experte** pour les principales revues de mon domaine (IEEE Trans. on Image Proc., IEEE Trans. on Signal Proc., IEEE Signal Proc. Letters, IEEE Trans. on Circuits and Systems et Signal Processing) et les principaux congrès (ICASSP, ICIP, EUSIPCO, ISBI, GRETSI). Membre du comité de programme pour RFIA 2012 et EUSIPCO 2012.
- **Experte** pour l'ANR (évaluation de proposition de projet 2013).
- **Expert-évaluateur** - Appel "Crédits et Projets 2013" du Fonds de la Recherche Scientifique - FNRS (agence de financement de la recherche pour la Belgique francophone).
- **Experte** pour l'ANR pour la première étape du processus de sélection de l'appel à projets générique 2016.

1.9 Activités administratives

Comité de sélection :

- 2011 : Membre du comité de sélection (Section CNU 61), Université Paris-Est.
- 2011 : Membre du comité de sélection (Section CNU 61), IUT de Cachan.
- 2011 : Membre du comité de sélection (Section CNU 26), Université Paris-Descartes.
- 2013 : Membre du comité de sélection (Section CNU 27), Université Paris-Descartes.
- 2013 : Membre du comité de sélection (Section CNU 61), Université du Sud Toulon-Var.

2015 : Membre du comité de sélection (Section CNU 61), Université Paris-Sud.

2016 : Membre du comité de sélection (Section CNU 26), Université Paris-Descartes.

2018 : Membre du comité de sélection (Section CNU 61), Université de Bordeaux.

2018 : Membre du comité de sélection (Section CNU 61), Aix-Marseille Université.

Jury de thèse :

- Membre du Jury de thèse (examinatrice) de Guillaume FUMAT, soutenue le 2 Déc. 2011, INSA Toulouse.
- Membre du Jury de thèse (examinatrice) de Anthony GIROUD, soutenue le 18 Déc. 2012, Univ. Paris-Est.
- Membre du Jury de thèse (examinatrice) de Mickael CARLAVAN, soutenue le 10 Juin 2013, Univ. Nice.
- Membre du Jury de thèse (examinatrice) de Ichrak TOUMI, soutenue le 28 Oct. 2013, Aix-Marseille Univ.
- Membre du Jury de thèse (examinatrice) de Yoann LE MONTAGNER, soutenue le 12 Nov. 2013, Telecom Paris-Tech.
- Membre du Jury de thèse (rapporteur) de Theofanis APOSTOLOPOULOS, soutenue le 25 Sep. 2015, King's College London, Londres.
- Membre du Jury de thèse (examinatrice) de Elodie PUYBAREAU, soutenue le 28 Nov. 2016, ESIEE Paris.
- Membre du Jury de thèse (examinatrice) de Lola BAUTISTA, soutenue le 30 Juin 2017, Université Côte d'Azur, Nice.
- Membre du Jury de thèse (examinatrice) de Lucie THIEBAUT LONJARET, soutenue le 15 Déc. 2017, Aix-Marseille Université, Marseille.
- Membre du Jury de thèse (examinatrice) de Afef CHERNI, soutenance prévue le 20 Sep. 2018, Université de Strasbourg, Strasbourg.

1.10 Distinction

2008 : **Prix de thèse** (en traitement du signal et des images du club EEA).

2011 : **Prime d'Excellence Scientifique** (PES).

1.11 Activités de recherche

1.11.1 Parcours dans la recherche

Mes recherches en tant que chargée de recherche ont démarré à l'université de Marne-la-Vallée, dans le laboratoire d'informatique Gaspard Monge (LIGM). Même si les outils d'analyse multirésolution m'ont toujours accroché, et que j'en ai développé durant mon doctorat, j'ai orienté mes recherches vers l'optimisation convexe pour la résolution de problèmes inverses. C'est ainsi que j'ai étudié et développé de nouveaux outils pour résoudre des problèmes d'optimisation impliquant des opérateurs linéaires tels que les opérateurs de trames mais aussi des régularisations multiples (e.g. parcimonie dans une

trame d'ondelettes et positivité). J'ai également étudié l'association d'approches d'optimisation continues et discrètes. Parmi les problèmes inverses considérés, on peut citer des problèmes de reconstruction en TEP, de déconvolution en microscopie ou macroscopie, d'estimation de cartes de profondeurs à partir d'images stéréoscopiques, de suppression d'échos dans des données de sismique.

J'ai ensuite effectué une mobilité à l'université d'Aix-Marseille, dans le laboratoire de mathématiques (I2M). Même si cette transition s'est faite sans rupture (continuité de l'encadrement de doctorant, des projets en cours), ce changement m'a à la fois permis d'élargir l'ensemble de mes collaborateurs mais aussi mes thématiques de recherche. En effet, j'ai assez rapidement démarré des travaux sur la factorisation non négative de tenseurs pour lesquels des approches stochastiques mais aussi proximales ont été proposées et appliquées avec succès à la spectroscopie de fluorescence. J'ai également initié des travaux en traitement du signal audio en collaboration avec l'équipe QARMA du LIS. Ces travaux ont, entre autre, conduit à l'étude des coefficients (propriété de rang faible) issus d'une transformée temps-fréquence mais aussi à l'étude d'approches de reconstruction de phase (approches d'optimisation semi-définie positive (SDP)) dans le plan temps-fréquence. Cela m'a également menée à travailler avec des collègues sur les modèles graphiques, assez largement étudiés par les communautés statistiques et traitement du signal, pour lesquels nous étudions de nouveaux modèles pour des tâches de classification (avec en vue une application à la classification de patients atteints de maladies neurodégénératives (collab. Hôp. Timone)). Enfin, plus récemment, j'ai entamé des collaborations avec des chercheurs chimistes avec lesquels nous allions nos connaissances afin d'analyser des données issues de chimie séparative. L'objectif est, par des approches de séparation de sources (ou factorisation en matrices nonnégatives), analyser des mélanges complexes et déterminer leur composition (composés et proportions).

1.11.2 Collaborations

- Sandrine Anthoine (I2M, Marseille) : problèmes inverses, optimisation
- Amel Benazza (SupCom Tunis, Tunisie) : ondelettes et estimateurs statistiques
- Laure Blanc-Féraud (I3S, Nice) : estimation d'hyperparamètres
- E. M. Carstea (National Institute of R&D for Optoelectronics, Roumanie) : spectroscopie de fluorescence pour la surveillance de la qualité de l'eau
- Lotfi Chaari (IRIT - INT-ENSEEIH, Toulouse) : approches bayésiennes, MCMC
- Emilie Chouzenoux (LIGM, Paris-Est) : problèmes inverses, optimisation
- Franck Debarbieux (INT/CERIMED, Marseille) : imagerie biphotonique
- Laurent Duval (IFPEN, Rueil-Malmaison) : géophysique
- Valentin Emiya (LIF, Marseille) : données manquantes en audio
- Eric Guedj (Professeur des universités - Praticien hospitalier, Hôpital de la Timone, Marseille) : neurologie nucléaire
- Sylvain Maire (LSIS, Toulon) : optimisation stochastique
- Jean-Christophe Pesquet (LIGM, Paris-Est) : problèmes inverses, optimisation
- Frédéric Richard (I2M, Marseille) : statistiques
- Nadège Thirion-Moreau (LSIS, Toulon) : décomposition tensorielle

- Jean-Yves Tourneret (IRIT - INT-ENSEEIH, Toulouse) : approches bayésiennes, MCMC
- Konstantin Usevich (CRAN, Nancy) : temps-fréquence, décomposition tensorielle
- Luca Zanni (University of Modena and Reggio Emilia, Italy) : estimation d'hyperparamètres

1.11.3 Participation à des contrats de recherches

1.11.3-a) Valorisation industrielle

- **1 Contrat avec l'Institut Français du Pétrole** (2007) d'un montant de 3000 euros HT pour la conception d'une Boîte à outils Matlab sur le débruitage d'images multi-canaux.
- **1 Contrat avec l'Institut Français du Pétrole** (Jan.-Sep. 2011) d'un montant de 11500 euros HT afin proposer une méthode fondée sur les ondelettes et l'analyse convexe permettant de supprimer des échos du signal d'intérêt dans des images de sismique.
- **1 Contrat avec Thalès Alenia Space** (Sep. 2011 - Juin 2012) d'un montant de 10000 euros HT afin proposer une méthode permettant de corriger le flou de bougé en observation haute résolution.

1.11.3-b) Participation à des projets publics

- Participation à l'**ANR OPTIMED** (porteur : J.-C. Pesquet) *Algorithmes d'optimisation décomposés pour les problèmes d'imagerie médicale de grande taille* (2006-2009). Ce projet vise à proposer des solutions algorithmiques originales aux problèmes de reconstruction en IRM et TEP.
- Participation à l'**ANR DIAMOND** (porteur : L. Blanc-Féraud) *Déconvolution d'Images Augmentée en Microscopie Optique N Dimensions* (2009-2013) pour laquelle nous avons obtenu 146663 euros. L'objectif de cette ANR est l'étude exhaustive (du signal à l'information) de deux nouvelles techniques d'imagerie optique : la microscopie tomographique et la macroscopie confocale.
- **1 Contrat de sous-traitance avec Telecom Paris** (resp. : C. Chaux) (Nov. 2010-Oct. 2012) d'un montant de 134884 euros HT dans le cadre du projet européen ACDC (porteur : P. Schwartz). L'objectif est d'une part de proposer des méthodes novatrices d'allocation du débit dans le cadre du codage vidéo, entre les images composant la vidéo ou entre les macroblocs, permettant l'encodage de ces différents éléments sur l'architecture ciblée et d'autre part, de fournir des algorithmes parallélisables au niveau multi-coeur et GPU (Graphics Processing Unit) pour l'estimation des vecteurs de mouvement et de disparité (dans le cas des séquences stéréoscopiques ou multi-vues).
- Participation au **SUPER BQR RICO 3D** (porteur : V. Biri) portant sur l'*Optimisation Continue pour l'Illumination et la Reconstruction 3D* (2011-2013). L'objectif est de reformuler certains problèmes d'illumination, du domaine de la vision et de la reconstruction 3D à l'aide de critères dont la minimisation sera réalisée efficacement par des techniques récentes. Les opérations de reconstruction devant être effectuées en temps réel, des techniques parallèles et GPU seront donc considérées.

- Participation au **Projet jeune chercheur GdR ISIS - GRETSI** (porteur : E. Chouzenoux) portant sur les *Méthodes d'optimisation pour l'imagerie biphotonique* (2013-2015). Ce projet porte sur la restauration de données de microscopie bi-photon pour lesquelles des approches de type Majoration-Minimisation et des approches de type proximales seront considérées.
- Participation à l'**ANR MAD** (porteur : V. Emiya) *Inpainting de données audio manquantes* (2014-2017). L'objectif de cette ANR est de travailler dans le cadre général de l'inpainting audio pour des tâches impliquant des données audio manquantes (declipping, perte de paquets ...).
- Participation au **Projet Défi Imag'in OPTIMISME** (porteur : J.-C. Pesquet) portant sur l'*OPTIMisation Stochastique en imagerie Multispectrale* (2015-2017). Ce projet porte sur la résolution de problèmes de déconvolution multispectrale dans de nouvelles modalités de microscopie biphotonique pour lesquels une nouvelle génération de méthodes parallèles exploitant des avancées très récentes en optimisation stochastique sera développée.
- Participation au **Projet Amidex Interdisciplinarité BIFROST** (porteur : C. Chaux) portant sur le *développement d'une chaîne de traitements pour la détermination directe de la composition moléculaire d'un mélange à partir de ses spectres RMN et SM (Spectroscopie de Masse) (Blind Identification, Filtering & Restoration On Spectral Techniques)* (2018-2021). De nouvelles méthodes bioanalytiques exploratoires ainsi que des algorithmes novateurs pour la décomposition de signaux de mélanges seront développés dans ce projet.
- Participation au **Projet PEPS I3A** (porteur : K. Usevich) portant sur la *simplification des réseaux de neurones par approche tensorielle* (2018).

1.11.4 Dépôt de brevet

- Suite au contrat avec l'Institut Français du Pétrole (Jan.-Sep. 2011), un brevet a été déposé le 24 août 2012.
Numéro de dépôt : 12/02.299
Titre : Procédé de traitement de réflexions multiples.

1.11.5 Logiciels

- Conception d'une **Boîte à outils Matlab en libre téléchargement** sur ma page web permettant de faire une analyse en ondelettes M -bandes monodimensionnelle en arbre dual.
- Conception d'une **Boîte à outils Matlab en libre téléchargement** sur ma page web permettant de restaurer des images en utilisant le principe de Stein (SURE) et plus précisément des SURE-LET (Linear Expansion of Threshold).
- Participation à la conception d'un **PlugIn ImageJ/Fiji en libre téléchargement** sur GitHub <https://github.com/dbenielli/optimisme/wiki>.

Chapitre 2

Liste de publications

2.1 Articles de revues à comité de lecture publiés ou acceptés

- [R1] L. Duval and C. Chaux. Lapped transforms and hidden Markov models for seismic data filtering. *Int. J. Wavelets Multidim. Inform. Proc.*, 2(4) :455–476, Dec. 2004.
- [R2] C. Chaux, L. Duval, and J.-C. Pesquet. Image analysis using a dual-tree M -band wavelet transform. *IEEE Trans. Image Process.*, 15(8) :2397–2412, Aug. 2006.
- [R3] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame based inverse problems. *Inverse Probl.*, 23(4) :1495–1518, Aug. 2007.
- [R4] C. Chaux, J.-C. Pesquet, and L. Duval. Noise covariance properties in dual-tree wavelet decompositions. *IEEE Trans. Inform. Theory*, 53(12) :4680–4700, Dec. 2007.
- [R5] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet. A nonlinear Stein based estimator for multichannel image denoising. *IEEE Trans. Signal Process.*, 56(8) :3855–3870, Aug. 2008.
- [R6] C. Chaux, J.-C. Pesquet, and N. Pustelnik. Nested iterative algorithms for convex constrained image recovery problems. *SIAM Journal on Imaging Sciences*, 2(2) :730–762, Jun. 2009.
- [R7] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A SURE approach for digital signal/image deconvolution problems. *IEEE Trans. Signal Process.*, 57(12) :4616–4632, Dec. 2009.
- [R8] N. Hajlaoui, C. Chaux, G. Perrin, F. Falzon, and A. Benazza-Benyahia. Satellite image restoration in the context of a spatially varying point spread function. *J. Opt. Soc. Am. A*, 27(6) :1473–1481, 2010.
- [R9] L. Chaari, E. Chouzenoux, N. Pustelnik, C. Chaux, and S. Moussaoui. OPTIMED : Optimisation itérative pour la résolution de problèmes inverses de grande taille. *Traitement du signal*, 28(3-4) :329–374, 2011.

- [R10] C. Chaux, A. Jezierska, J.-C. Pesquet, and H. Talbot. A spatial regularization approach for vector quantization. *Journal of Mathematical Imaging and Vision*, 41 :23–38, 2011.
- [R11] L. Jacques, L. Duval, C. Chaux, and G. Peyré. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Process.*, 91(12) :2699–2730, Dec. 2011.
- [R12] N. Pustelnik, C. Chaux, and J.-C. Pesquet. Parallel proximal algorithm for image restoration using hybrid regularization. *IEEE Trans. Image Process.*, 20(9) :2450–2462, Sep. 2011.
- [R13] N. Pustelnik, J.-C. Pesquet, and C. Chaux. Relaxing tight frame condition in parallel proximal methods for signal restoration. *IEEE Trans. Signal Process.*, 60(2) :968–973, feb. 2012.
- [R14] C. Chaux, M. El Gheche, J. Farah, J.-C. Pesquet, and B. Pesquet-Popescu. A parallel proximal splitting method for disparity estimation from multicomponent images under illumination variation. *Journal of Mathematical Imaging and Vision*, 47(3) :167–178, Nov. 2013.
- [R15] A. Jezierska, C. Chaux, J.-C. Pesquet, H. Talbot, and G. Engler. An EM approach for time-variant Poisson-Gaussian model parameter estimation. *IEEE Trans. Signal Process.*, 62(1) :17–30, Jan. 2014.
- [R16] M. Q. Pham, L. Duval, C. Chaux, and J.-C. Pesquet. A primal-dual proximal algorithm for sparse template-based adaptive filtering : Application to seismic multiple removal. *IEEE Trans. Signal Process.*, 62(16) :4256–4269, Aug. 2014.
- [R17] X. T. Vu, S. Maire, C. Chaux, and N. D. Thirion-Moreau. A new stochastic optimization algorithm to decompose large nonnegative tensors. *Signal Process. Lett.*, 22(10) :1713–1717, Oct. 2015.
- [R18] L. Chaari, J.-Y. Tourneret, C. Chaux, and H. Batatia. A hamiltonian monte carlo method for non-smooth energy sampling. *IEEE Trans. Signal Process.*, 64(21) :5585–5594, Nov. 2016.
- [R19] X. T. Vu, C. Chaux, N. D. Thirion-Moreau, S. Maire, and E. M. Carstea. A new penalized nonnegative third order tensor decomposition using a block coordinate proximal gradient approach : application to 3d fluorescence spectroscopy. *J. Chemometrics*, 31(4), Apr. 2017.

2.2 Articles de conférences invités avec actes

- [CI1] C. Chaux, A. Benazza-Benyahia, and J.-C. Pesquet. A block-thresholding method for multispectral image denoising. In *Proc. SPIE, Wavelets : Appl. Signal Image Process.*, volume 5914, pages 1–H1,1–H13, San Diego, CA, USA, Aug. 2005.
- [CI2] C. Chaux, A. Benazza-Benyahia, and J.-C. Pesquet. Use of Stein's principle for multichannel image denoising. In *International Symposium on Communications, Control and Signal Processing*, March 2006. 6 pp.

- [CI3] L. Duval, C. Chaux, and J.-C. Pesquet. m -band filter banks and dual-tree wavelets for engine combustion and geophysical image analysis. In *Wavelet Applications in Industrial Processing IV, on SPIE Symposium Optics East*, Boston, MA, USA, 2006. 12 pp.
- [CI4] C. Chaux, L. Blanc-Féraud, and J. Zerubia. Wavelet-based restoration methods : application to 3D confocal microscopy images. In *Proc. SPIE, Wavelets*, San Diego, CA, August 2007. 11 pp.
- [CI5] A. Benazza-Benyahia, J.-C. Pesquet, and C. Chaux. Image denoising in the wavelet transform domain based on Stein's principle (tutorial). In *1st IEEE international Workshops on Image Processing Theory, Tools and Applications (IPTA'08)*, Sousse, Tunisia, Nov. 23-26 2008. 9 pp.
- [CI6] L. Chaari, N. Pustelnik, C. Chaux, and J.-C. Pesquet. Solving inverse problems with overcomplete transforms and convex optimization techniques. In *Proc. SPIE, Wavelets*, San Diego, California, USA, Aug. 2-6 2009.
- [CI7] A. Marin, C. Chaux, Pesquet J.-C., and P. Ciuciu. Image reconstruction from multiple sensors using Stein's principle. application to parallel MRI. In *ISBI 2011*, Chicago, USA, 30 Mar-2 Avr. 2011. 4 pp.
- [CI8] R. Gaetano, B. Pesquet-Popescu, and C. Chaux. A convex optimisation approach for image resolution enhancement from compressed representations. In *International Conference on Digital Signal Processing (DSP)*, page 8 pp., Santorini, Greece, Jul. 1-3 2013.
- [CI9] M. Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet. A constrained-based optimization approach for seismic data recovery problems. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, pages 2377–2381, Florence, Italie, 2014.
- [CI10] A. Jezierska, J.-C. Pesquet, H. Talbot, and C. Chaux. Iterative poisson-gaussian noise parametric estimation for blind image denoising. In *Proc. Int. Conf. Image Process.*, Paris, France, Oct. 2014.
- [CI11] A. M. Krémé, V. Emiya, and C. Chaux. Phase reconstruction for time-frequency inpainting. In *14th International Conference on Latent Variable Analysis and Signal Separation*, Guildford, UK, Jul. 2-6 2018.

2.3 Articles de conférences avec actes

- [C1] C. Chaux, L. Duval, and J.-C. Pesquet. Hilbert pairs of M -band orthonormal wavelet bases. In *Proc. Eur. Sig. Image Proc. Conf.*, pages 1187–1190, Vienna, Austria, Sep. 6-10, 2004.
- [C2] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A forward-backward algorithm for image restoration with sparse representations. In *Signal Processing with Adaptive Sparse Structured Representations (SPARS'05)*, pages 49–52, Rennes, France, November 2005.

- [C3] C. Chaux, L. Duval, and J.-C. Pesquet. 2D dual-tree M -band wavelet decomposition. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, volume 4, pages 537–540, Philadelphia, USA, Mar. 18-23, 2005.
- [C4] C. Chaux, L. Duval, and J.-C. Pesquet. Étude du bruit dans une analyse M -bandes en arbre dual. In *Proc. GRETSI*, pages 229–232, Louvain, Belgique, Sep. 2005.
- [C5] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. Iterative image deconvolution using overcomplete representations. In *European Signal Processing Conference (EUSIPCO'06)*, September 2006. electronic version (5 pp.).
- [C6] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet. A new estimator for image denoising using a 2D dual-tree M -band wavelet decomposition. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, volume 3, pages 249–252, Toulouse, France, May 21-24, 2006.
- [C7] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. Opérateurs proximaux pour la restauration bayésienne de signaux. In *(GRETSI'07)*, pages 1277–1280, Troyes, France, Sep. 2007.
- [C8] C. Chaux, J.-C. Pesquet, and L. Duval. 2D dual-tree complex biorthogonal M -band wavelet transform. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, volume 3, pages 845–848, Honolulu, HI, USA, Apr. 15-20, 2007.
- [C9] L. Duval, C. Chaux, and S. Ker. Seismic data analysis with dual-tree M -band wavelet transforms. In *Proc. EAGE Conf. Tech. Exhib.*, 2007.
- [C10] L. Duval, C. Chaux, and S. Ker. Coherent noise removal in seismic data with dual-tree M -band wavelets. In *Proc. SPIE, Wavelets : Appl. Signal Image Process.*, San Diego, CA, USA, Aug. 2007.
- [C11] N. Pustelnik, C. Chaux, and J.-C. Pesquet. A constrained forward-backward algorithm for image recovery problems. In *16th European Signal Processing Conference (EUSIPCO'08)*, 2008. electronic version (5 pp.) To appear.
- [C12] C. Chaux and L. Blanc-Féraud. Estimation d'hyperparamètres pour la résolution de problèmes inverses à l'aide d'ondelettes. In *GRETSI 2009*, Dijon, Sep. 8-11 2009.
- [C13] N. Pustelnik, C. Chaux, and J.-C. Pesquet. Hybrid regularization for data restoration in the presence of Poisson noise. In *EUSIPCO 2009*, Glasgow, UK, Aug. 24-28 2009. 5 pp.
- [C14] N. Pustelnik, C. Chaux, and J.-C. Pesquet. Extension des algorithmes imbriqués pour la résolution de problèmes d'optimisation convexe en imagerie. In *GRETSI 2009*, Dijon, Sep. 8-11 2009.
- [C15] N. Pustelnik, C. Chaux, and J.-C. Pesquet. A wavelet-based quadratic extension method for image deconvolution in the presence of Poisson noise. In *ICASSP 2009*, Taipei, Taiwan, Apr. 19-24 2009. 4 pp.

- [C16] N. Pustelnik, C. Chaux, J.-C. Pesquet, F. C. Sureau, E. Dusch, and C. Comtat. Adapted convex optimization algorithm for wavelet-based dynamic PET reconstruction. In *Fully3D*, Beijing, Chine, Sep. 5-10 2009.
- [C17] A. Jezierska, C. Chaux, H. Talbot, and J.-C. Pesquet. Image quantization under spatial smoothness constraints. In *ICIP 2010*, Hong Kong, Sep. 26-29 2010.
- [C18] N. Pustelnik, C. Chaux, J.-C. Pesquet, and C. Comtat. Parallel algorithm and hybrid regularization for dynamic PET reconstruction. In *IEEE Medical Imaging Conference*, Knoxville, Tennessee, Oct. 30 - Nov. 6 2010.
- [C19] N. Pustelnik, J.-C. Pesquet, and C. Chaux. Proximal methods for image restoration using a class of non-tight frame representations. In *EUSIPCO 2010*, pages 611–615, Aalborg, Danemark, Aug. 23-27 2010.
- [C20] M. El Gheche, J.-C. Pesquet, J. Farah, C. Chaux, and B. Pesquet-Popescu. Disparity map estimation under convex constraints using proximal algorithms. In *IEEE Workshop on Signal Processing Systems (SiPS)*, Beirut, Liban, Oct. 4-7 2011.
- [C21] A. Jezierska, C. Chaux, J.-C. Pesquet, and H. Talbot. An EM approach for Poisson-Gaussian noise modeling. In *European Signal Processing Conference (EUSIPCO)*, Barcelone, Espagne, 29 Aug. - 2 Sep. 2011.
- [C22] N. Pustelnik, J.-C. Pesquet, and C. Chaux. Bancs de filtres et méthodes proximales pour la restauration d'images. In *GRETSI*, Bordeaux, France, Sep. 5-8 2011.
- [C23] D. Gagnaniello, C. Chaux, J.-C. Pesquet, and L. Duval. A convex variational approach for multiple removal in seismic data. In *European Signal Processing Conference (EUSIPCO)*, Bucarest, Roumanie, Aug. 27 - 31 2012.
- [C24] A. Jezierska, H. Talbot, C. Chaux, J.-C. Pesquet, and G. Engler. Poisson-Gaussian noise parameter estimation in fluorescence microscopy imaging. In *International Symposium on Biomedical Imaging (ISBI)*, Barcelona, Spain, May 2012.
- [C25] R. Cavicchioli, C. Chaux, L. Blanc-Féraud, and L. Zanni. ML estimation of wavelet regularization hyperparameters in inverse problems. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, page 5 pp., Vancouver, Canada, May, 26-31 2013.
- [C26] M. Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet. Seismic multiple removal with a primal-dual proximal algorithm. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, page 5 pp., Vancouver, Canada, May, 26-31 2013.
- [C27] M. Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet. Filtrage de multiples sismiques par ondelettes et optimisation convexe. In *GRETSI*, Brest, France, 3-6 Sep. 2013.
- [C28] X. T. Vu, C. Chaux, S. Maire, and N. D. Thirion-Moreau. Study of different strategies for the canonical polyadic decomposition of nonnegative third order tensors with application to the separation of spectra in 3d fluorescence spectroscopy. In *IEEE Int. Workshop Mach. Learn. Signal Process.*, Reims, France, 21-24 Sep. 2014.
- [C29] L. Chaari, J.-Y. Tourneret, and C. Chaux. Sparse signal recovery using a bernoulli generalized gaussian prior. In *Proc. Eur. Sig. Image Proc. Conf.*, Nice, France, 31 Aug. - 4 Sep. 2015.

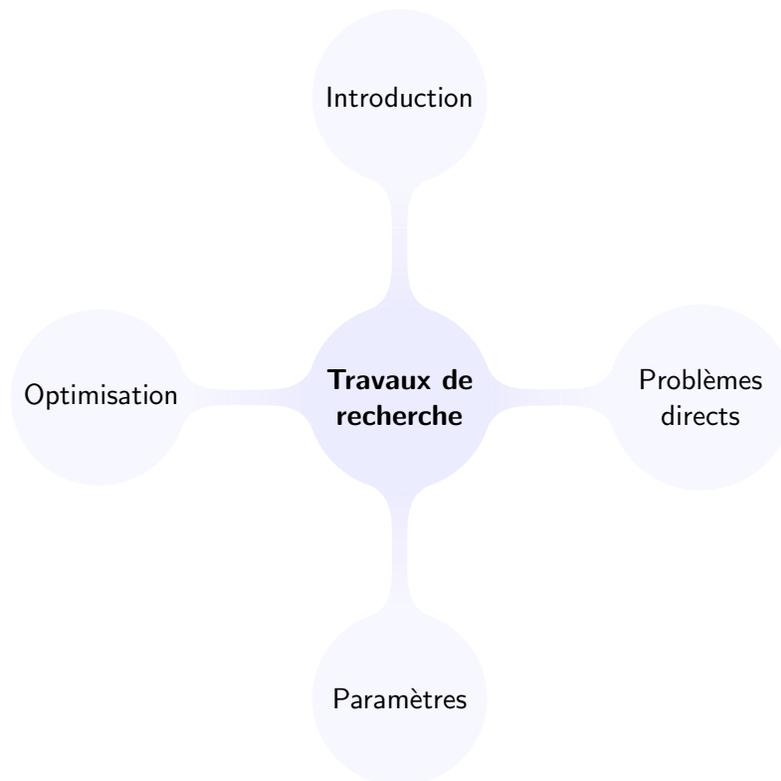
- [C30] E. Chouzenoux, L. Lamassé, S. Anthoine, C. Chaux, A. Jaouen, I. Vanzetta, and F. Debarbieux. Approche variationnelle pour la déconvolution rapide de données 3D en microscopie biphotonique. In *Proc. GRETSI*, Lyon, France, 8-11 Sep. 2015.
- [C31] M. Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet. Sparse adaptive template matching and filtering for 2d seismic images with dual-tree wavelets and proximal methods. In *Proc. Int. Conf. Image Process.*, Québec city, Canada, 27-30 Sep. 2015.
- [C32] S. Cadoni, E. Chouzenoux, Pesquet J.-C., and C. Chaux. A block parallel majorize-minimize memory gradient algorithm. In *Proc. Int. Conf. Image Process.*, Phoenix, USA, 25-28 Sep. 2016.
- [C33] X. T. Vu, C. Chaux, N. D. Thirion-Moreau, and S. Maire. A proximal approach for nonnegative tensor decomposition. In *13th International Conference on Latent Variable Analysis and Signal Separation*, Grenoble, France, Feb. 21-23 2017.
- [C34] R. Hamon, C. Chaux, and V. Emiya. De la faiblesse de rang en temps-fréquence. In *Proc. GRETSI*, Sep. 5-8 2017.
- [C35] V. Emiya, R. Hamon, and C. Chaux. Being low-rank in the time-frequency plane. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, Calgary, Canada, Apr. 15-20 2018.
- [C36] K. Usevich, V. Emiya, D. Brie, and C. Chaux. Characterization of finite signals with low-rank STFT. In *Proc. IEEE Workshop Stat. Signal Process.*, Freiburg, Germany, June 2018.

2.4 Chapitre de livre

- [Ch1] C. Chaux, A. Benazza-Benyahia, J.-C. Pesquet, and L. Duval. Wavelet transform for the denoising of multivariate images. In *Multivariate Image Processing*, pages 203–238. ISTE Ltd and John Wiley & Sons Inc., 2009.

Partie II

Travaux de recherche



Chapitre 3

Introduction

Ce manuscrit d'habilitation traite de problèmes inverses. Durant ces 10 années de recherche, je me suis intéressée à toutes les facettes que comporte la résolution de problèmes inverses à savoir

- leur modélisation
 - comment sont obtenues les données que j'observe ?
 - par quel(s) processus (nature) sont elles affectées ?
 - ...
- leur résolution
 - qu'est ce que je connais sur les données observées : sont elles parcimonieuses ou plutôt lisses ? Ai-je plus d'informations sur une représentation de mes données dans un autre domaine (e.g. ondelettes) plutôt que sur les données elles même ?
 - le problème d'optimisation formulé est-il convexe ? différentiable ?
- l'estimation des paramètres modélisant l'observation ou gouvernant la qualité de reconstruction
 - quelle est l'intensité du bruit ?
 - est-ce un bruit structuré ou aléatoire ?
 - comment fixer les paramètres de régularisation du problème d'optimisation ainsi formulé ?

Même s'il est vrai que durant ma thèse je me suis beaucoup intéressée aux transformées en ondelettes géométriques [1], ces dernières années je m'en suis plutôt servis en tant qu'outils d'analyse. Les différentes questions ainsi formulées m'ont amenée, accompagnée de doctorants, à mettre en place des algorithmes de gradient et de les accélérer, à mettre en œuvre des algorithmes de type espérance-maximisation (*expectation-maximization*) (EM) dont les étapes E et M étaient elle mêmes constituées de problèmes d'optimisation continue ou discrète et à faire appel aux algorithmes d'optimisation convexe proximaux. L'optimisation est vraiment au cœur de mes activités de recherche.

La majorité des travaux que j'ai entrepris sont en lien avec des applications réelles telles que la géophysique, la microscopie (MACROscopie, microscopie biphotonique), la TEP, la spectroscopie de fluorescence ou encore la spectroscopie RMN.

J'ai choisi de construire le premier chapitre 4 en lien avec ces applications afin de bien illustrer les spécificités de chacune. J'aborde dans un second temps le problème d'estimation des paramètres (des perturbations) ou hyperparamètres (des méthodes d'optimisation) dans le chapitre 5. Cela permet déjà d'appréhender la résolution du problème inverse en soit par la mise en place de stratégie d'optimisation. Ensuite, dans le chapitre 6 je m'intéresse à la résolution de problèmes inverses en données complètes c'est à dire lorsque tout est supposé connu. Enfin, je conclus ce manuscrit d'habilitation par le chapitre 7 afin de dresser le bilan du travail effectué et j'en profite pour donner dans le chapitre 8 quelques pistes de recherche sur lesquelles je commence à réfléchir.

Chapitre 4

Modèles et formulation du problème direct

Avant-propos

Dans ce chapitre, nous introduisons quelques modèles directs qui sont d'intérêt dans ce manuscrit et nous illustrons ces modèles par des applications réelles. Les problèmes inverses associés étant formulés sous formes variationnelles, nous introduisons ensuite les formes générales des problèmes d'optimisation auxquels nous serons confrontés dans la suite de ce manuscrit. Une brève introduction aux opérateurs et algorithmes proximaux est également effectuée à la fin de ce chapitre.

Sommaire

4.1	Formulation directe et objectif	33
4.2	Cadre bayésien / modélisation variationnelle du problème inverse .	37
4.3	Approches proximales	39
4.4	Conclusion	42
4.5	Quelques publications personnelles en lien avec ce chapitre	43

4.1 Formulation directe et objectif

4.1.1 Un modèle général ...

Nous nous intéressons dans un premier temps au modèle direct, c'est à dire au modèle (physique) de formation des données. Le modèle direct peut s'exprimer de manière générale sous la forme

$$z = \mathcal{D}_\alpha(Hy) \tag{4.1}$$

où $z \in \mathcal{G}$ constituent les observations, $y \in \mathcal{H}$ les données originales (inconnues, à estimer), $H : \mathcal{H} \rightarrow \mathcal{G}$ un opérateur linéaire (e.g. opérateur de flou, opérateur de projection,

masque, ...) et \mathcal{D}_α une dégradation (bruit) de paramètre α (bruit additif Gaussien de variance α , bruit de poisson de facteur d'échelle α , ...). \mathcal{H} et \mathcal{G} constituent des espaces de Hilbert de dimensions finies.

Résoudre le problème inverse associé requiert à prime abord la détermination de H et \mathcal{D}_α . Bien que H puisse être supposé inconnu (problème aveugle) il sera supposé connu ici. En ce qui concerne la dégradation \mathcal{D}_α , le type de dégradation (nature du bruit) sera supposé connu mais le (ou les) paramètre(s) qui le caractérisent pourront être supposés inconnus et donc estimés.

4.1.2 ... et de multiples problèmes sous-jacents

Tous les problèmes auxquels je me suis intéressée durant ces 10 années de recherche sont issus de domaines divers et variés mais peuvent tous s'écrire sous la forme de l'Eq. (4.1). Je les détaille individuellement dans les paragraphes ci-après.

4.1.2-a) Débruitage

Un premier problème qui m'a longuement intéressé est le problème du débruitage. Dans ce cas, l'opérateur linéaire H se réduit à l'identité. Le problème direct s'exprime donc comme

$$z = \mathcal{D}_\alpha(y) \quad (4.2)$$

et l'enjeu consiste à débruiter les observations z .

Ce problème m'a particulièrement intéressé lorsque j'ai travaillé sur des données de macroscopie. Dans ce cas, la nature du bruit \mathcal{D}_α s'est révélée être un mélange de bruit Poisson et de bruit Gaussien [2]. Les observations peuvent s'écrire dans ce cas

$$z = \alpha \mathcal{P}(y) + \mathcal{N}(c, \sigma^2) \quad (4.3)$$

où $\mathcal{P}(y)$ représente un bruit de Poisson et $\mathcal{N}(c, \sigma^2)$ un bruit Gaussien de moyenne c et de variance σ^2 . Lorsque l'acquisition se fait au cours du temps, on peut souligner l'apparition d'un phénomène de photoblanchiment (voir Fig. 4.1), ou évanescence de fluorescence au cours du temps, se traduisant par l'introduction d'une exponentielle décroissante dans le modèle ($y \rightarrow y \exp(-at)$, $a \in \mathbb{R}_+^*$) amenuisant l'intensité des données observées au cours du temps [3].

4.1.2-b) Déconvolution

Un deuxième problème qui m'a fortement intéressé est le problème de la déconvolution. Dans ce cas, l'opérateur linéaire H représente un opérateur de convolution (e.g. un flou). Le problème direct s'exprime donc comme

$$z = \mathcal{D}_\alpha(\{h * y\}) \quad (4.4)$$

où $\{h * y\}$ dénote le produit de convolution entre les signaux h et y (on pourra avoir à faire parfois à des noyaux de convolution non stationnaires comme illustré Fig. 4.2). La nature du bruit \mathcal{D}_α s'adapte au problème considéré.

Mes travaux ont majoritairement porté sur des bruits Gaussiens (e.g. images satellitaires cf. Fig. 4.2) et des bruits de Poisson (e.g. images de microscopie).

L'enjeu ici est double puisqu'il consiste à débruiter les observations z mais aussi à supprimer l'effet de la convolution avec le noyau h .

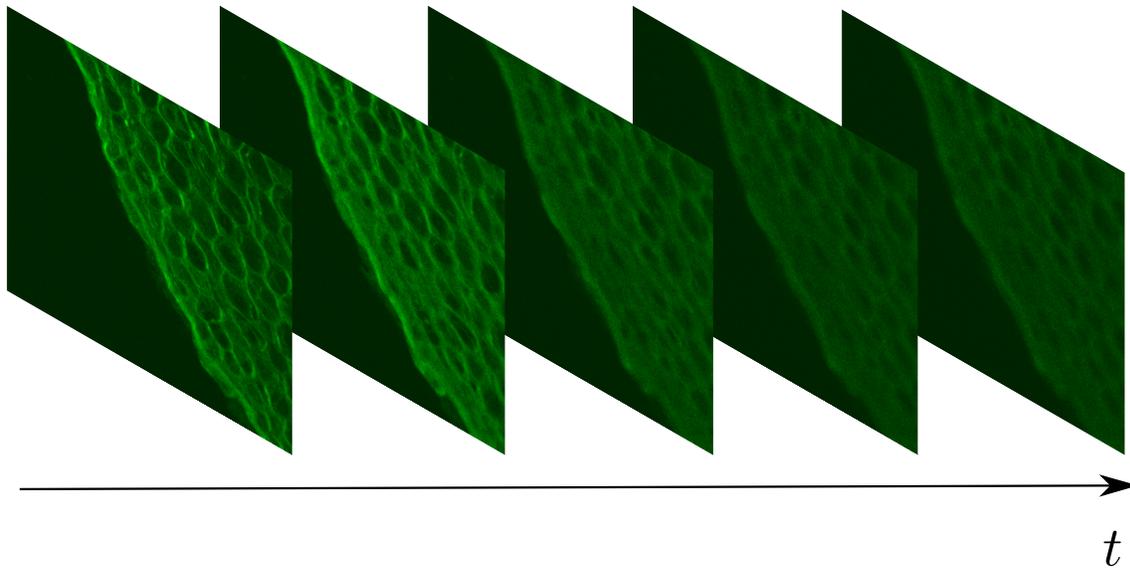
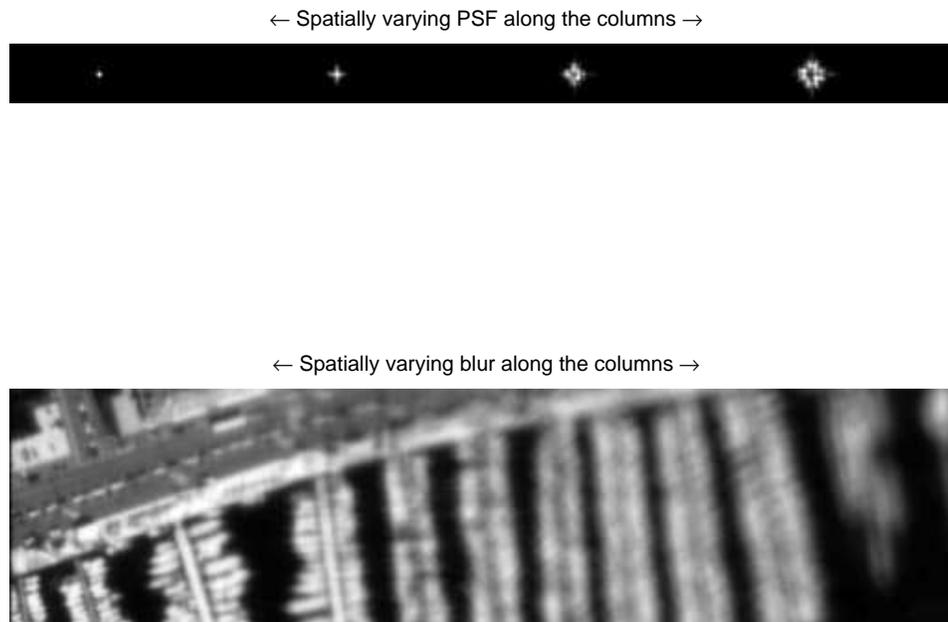


FIGURE 4.1 – Images de macroscopie - bruit et phénomène de photoblanchiment.

FIGURE 4.2 – Variation de la PSF (h) (haut) et son impact sur la scène observée (©CNES) (bas)

D'une manière générale, les problèmes inverses impliquant des opérateurs linéaires sont souvent mal posés au sens de Hadamard [4] (existence, unicité, stabilité de la solution) et leur résolution requiert des attentions particulières comme nous le verrons par la suite.

4.1.2-c) Reconstruction

Un autre problème d'intérêt est la reconstruction d'images. L'opérateur linéaire H est dans ce cas un opérateur de projection, projetant l'image dans un autre domaine (sinogramme dans le cas de la TEP cf. Fig. 4.3 ou Fourier dans le cas de l'IRM). Encore une fois, le bruit est de nature poissonnienne pour la TEP et gaussien pour l'IRM.

On a donc

$$z = \alpha \mathcal{P}(Hy) \quad (4.5)$$

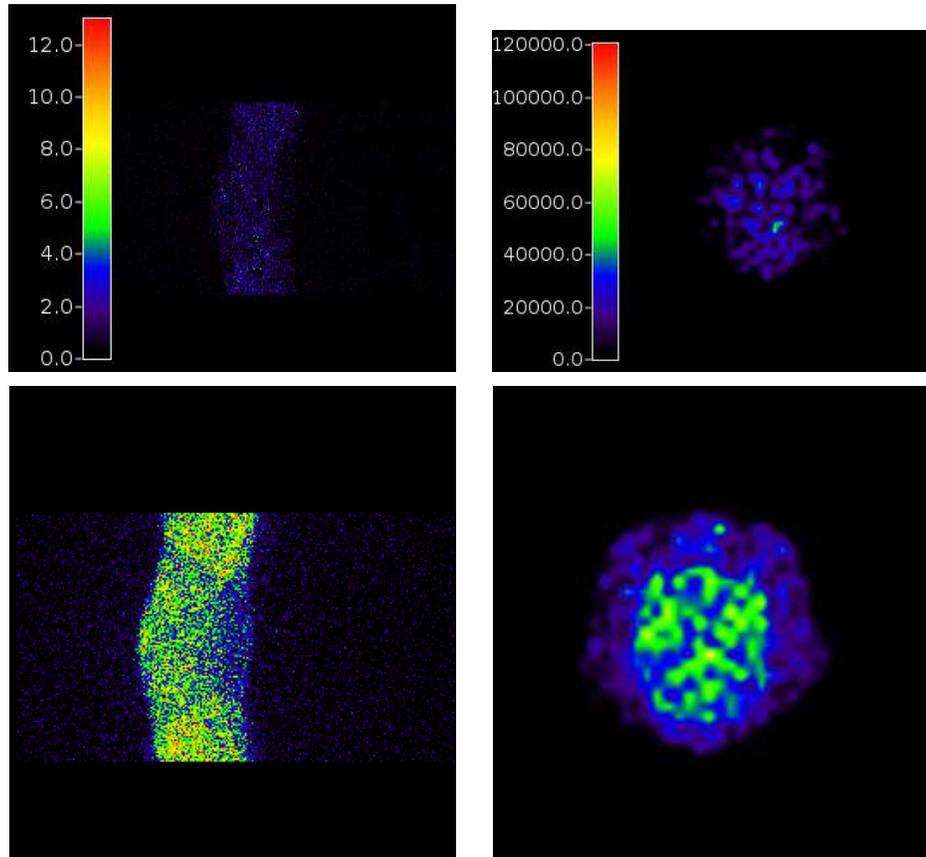


FIGURE 4.3 – Images TEP. Sinogrammes observés (gauche) et images associées (droite).

4.1.2-d) Suppression d'échos

Ma collaboration avec l'IFP Energies Nouvelles m'a amené à travailler sur des signaux/images dans lesquels on souhaite supprimer les échos (appelés multiples). Dans ce cas, le modèle direct s'exprime comme

$$z = y + s + b \quad (4.6)$$

où y est le signal primaire (inconnu), s sont les multiples et b une perturbation aléatoire. Le schéma donné dans la figure 4.4 donne un aperçu d'une acquisition mêlant signaux primaires et réflexions multiples.

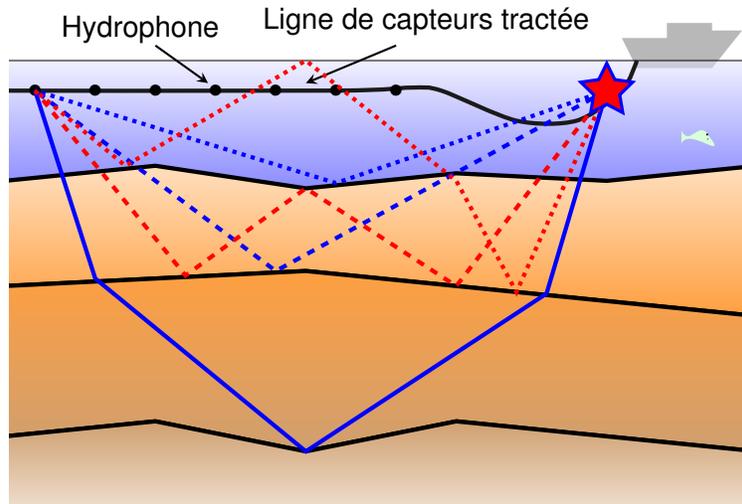


FIGURE 4.4 – Illustration d’une acquisition sismique : les signaux primaires y sont représentés en bleu et les réflexions multiples s en rouge.

L’enjeu ici est également double puisqu’il faut à la fois supprimer les échos s et le bruit b .

4.1.2-e) Décomposition (CP) de tenseur

Plus tardivement, je me suis intéressée à la décomposition tensorielle. A prime abord le modèle peut paraître différent mais cette problématique constitue elle aussi un problème inverse dont le modèle direct s’écrit comme suit :

$$\mathcal{T} = \bar{\mathcal{T}} + \mathcal{B} \quad (4.7)$$

où $\bar{\mathcal{T}} = (\bar{t}_{i_1 \dots i_N})_{i_1, \dots, i_N} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ et $\{I_1, \dots, I_N\} \subset \mathbb{N}, N \in \mathbb{N}$. La décomposition Canonique Polyadique (Canonical Polyadic Decomposition (CPD)) consiste à décomposer $\bar{\mathcal{T}}$ comme $\bar{\mathcal{T}} = \sum_{r=1}^R \bar{\mathbf{a}}_r^{(1)} \circ \bar{\mathbf{a}}_r^{(2)} \circ \dots \circ \bar{\mathbf{a}}_r^{(N)}$ où \circ désigne le produit tensoriel.

Ce qui change ici est la relation multilinéaire entre les variables. L’objectif est à la fois de débruiter le tenseur mais également d’estimer les facteurs latents $\bar{\mathbf{a}}_r^{(1)}, \bar{\mathbf{a}}_r^{(2)}, \dots, \bar{\mathbf{a}}_r^{(N)}$.

Ce modèle est particulièrement bien adapté pour modéliser des données de spectroscopie de fluorescence. Un exemple de telles données (tenseur d’ordre 3) est donné Fig. 4.5. On peut le décomposer selon 3 facteurs : le spectre d’émission, le spectre d’excitation et la concentration.

4.2 Cadre bayésien / modélisation variationnelle du problème inverse

Une fois le modèle identifié et défini, on cherche à résoudre le problème inverse, c’est à dire à estimer les données originales à partir des observations. Bien que cela puisse être réalisé de différentes manières, je me suis intéressée à la formulation du problème

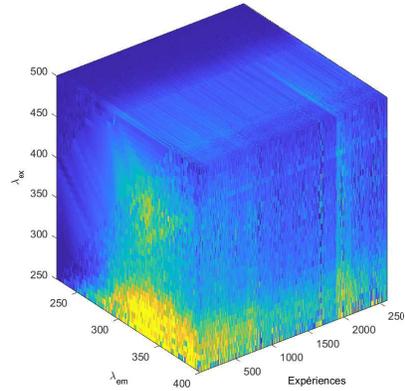


FIGURE 4.5 – Données de spectroscopie de fluorescence 3D.

sous forme variationnelle. On peut remarquer que cette formulation peut également se faire dans un contexte bayésien [5].

D'une manière générale, cela revient donc à considérer des problèmes d'optimisation de la forme

$$\underset{y \in \mathcal{H}}{\text{minimiser}} \quad \underbrace{f_1(y)}_{\text{Terme de fidélité}} + \underbrace{f_2(y)}_{\text{Terme de régularisation}} . \quad (4.8)$$

où $f_1 : \mathcal{H} \rightarrow \mathbb{R}$ et $f_2 : \mathcal{H} \rightarrow \mathbb{R}$ sont des fonctions semi-continues inférieurement, propres et convexes [6].

Le terme de fidélité aux données découle directement du problème inverse considéré et est étroitement lié à la statistique du bruit. Le terme de régularisation correspond lui à un a priori que l'on peut avoir sur la solution cible. D'une manière générale, ce terme intègre un paramètre de régularisation permettant de lui attribuer plus ou moins d'importance.

On peut à ce stade avoir plusieurs remarques :

- Dans de nombreuses situations, on peut avoir non pas un mais plusieurs a priori sur la solution cible. Le problème de minimisation (4.8) devient alors

$$\underset{y \in \mathcal{H}}{\text{minimiser}} \quad \underbrace{f_1(y)}_{\text{Terme de fidélité}} + \underbrace{\sum_{p=2}^P f_p(y)}_{P-1 \text{ termes de régularisation}} . \quad (4.9)$$

Ce sont alors $P - 1$ paramètres de régularisation qui sont intégrés à chaque terme $(f_p)_p \in \{2, \dots, P\}$ permettant de leur attribuer individuellement plus ou moins d'importance.

- D'autre part, il est fréquent d'avoir non pas un a priori sur la solution cible directement mais un a priori sur une transformation de celle-ci par un opérateur linéaire $F : \mathcal{H} \rightarrow \mathcal{G}$. Cet opérateur peut correspondre, par exemple, à un opérateur de gradient ou encore un opérateur de trame [7]. D'une manière générale, le problème de

minimisation (4.9) peut s'écrire comme suit :

$$\underset{y \in \mathcal{H}}{\text{minimiser}} \sum_{p=1}^P f_p(F_p y). \quad (4.10)$$

- Dans le cas où F représente un opérateur de trame [7, 8] ($F^* : \mathcal{G} \rightarrow \mathcal{H}$ désigne alors l'opérateur de synthèse de trame), on peut formuler le problème de minimisation de deux manières. En effet, soit $y = F^* x$, on a alors la possibilité de résoudre (4.10) qui correspond à la formulation à l'analyse ou

$$\underset{x \in \mathcal{G}}{\text{minimiser}} \sum_{p=1}^P f_p(F_p^* x). \quad (4.11)$$

qui correspond à la formulation à la synthèse et où y est obtenu en appliquant l'opérateur de synthèse de trame à x .

On a démontré dans la thèse de Nelly Pustelnik [9] que

Proposition 4.1 *La formulation à l'analyse (4.10) est un cas particulier de la formulation à la synthèse (4.11).*

Cela signifie que tout problème formulé à l'analyse peut être réécrit comme un problème formulé à la synthèse conduisant au même résultat.

- Enfin, on peut vouloir exprimer une régularisation sur un sous-ensemble de données plutôt que sur toutes les données simultanément (régularisation par blocs de coordonnées). Cela revient donc à considérer le problème d'optimisation suivant :

$$\underset{y \in \mathcal{H}}{\text{minimiser}} \sum_{p=1}^P f_p(F_p y^{(p)}). \quad (4.12)$$

où $f_p : \mathcal{H}_p \rightarrow \mathbb{R}$ et $y = (y^{(p)})_{p=1}^P$ (les blocs $(y^{(p)})_{p=1}^P$ représentent une partition de y).

4.3 Approches proximales

L'objectif de cette section n'est pas de décrire exhaustivement les approches proximales mais plutôt de donner au lecteur les définitions et outils nécessaires pour la bonne compréhension de ce manuscrit et de rendre ainsi ce manuscrit "self-contained". Pour plus de détails, les lecteurs sont invités à se reporter aux références suivantes : pour les opérateurs et algorithmes proximaux [10] et pour les algorithmes primaux-duaux [11].

4.3.1 Introduction aux opérateurs proximaux

Dans ce manuscrit nous travaillons principalement avec des fonctions φ appartenant à $\Gamma_0(\mathbb{R})$ l'ensemble des fonctions propres convexes, semi-continues inférieurement de $\mathbb{R} \rightarrow]-\infty, +\infty]$. L'opérateur proximal [12] de $\varphi \in \Gamma_0(\mathbb{R})$ est défini par

$$\text{prox}_{\varphi} : \mathbb{R} \rightarrow \mathbb{R} : v \mapsto \arg \min_{u \in \mathbb{R}} \frac{1}{2} \|u - v\|^2 + \varphi(u), \quad (4.13)$$

Cet opérateur peut être vu comme une généralisation de la notion de projection. En effet, lorsque φ est la fonction indicatrice ι_C d'un ensemble convexe fermé non vide C de \mathbb{R} , *i.e.* une fonction qui vaut 0 dans C et $+\infty$ dans $\mathbb{R} \setminus C$, prox_{ι_C} est le projecteur Π_C sur C . Des formes explicites d'opérateurs proximaux¹ existent pour de nombreuses fonctions $\varphi \in \Gamma_0(\mathbb{R})$ [13][5][14].

On peut citer par exemple l'opérateur proximal de la norme ℓ_1 . Plus précisément, soit $\alpha > 0$, et soit $\varphi: \mathbb{R} \rightarrow \mathbb{R}: \xi \mapsto \alpha|\xi|$. Alors pour tout $\xi \in \mathbb{R}$,

$$\text{prox}_{\varphi} \xi = \text{sign}(\xi) \max\{|\xi| - \alpha, 0\} \quad \text{with} \quad \text{sign}(\xi) = \begin{cases} +1 & \text{si } \xi > 0 \\ -1 & \text{si } \xi < 0 \\ 0 & \text{si } \xi = 0 \end{cases} \quad (4.14)$$

où $\max\{\cdot, \cdot\}$ retourne la valeur maximum entre ses deux arguments. L'opérateur proximal de la norme ℓ_1 se résume donc à une opération de seuillage doux.

Si l'on considère à présent la norme ℓ_2 ou plus précisément, la fonction $\varphi: \mathbb{R} \rightarrow \mathbb{R}: \xi \mapsto \alpha|\xi|^2$ où $\alpha > 0$. Alors, pour tout $\xi \in \mathbb{R}$,

$$\text{prox}_{\varphi} \xi = \frac{\xi}{2\alpha + 1}. \quad (4.15)$$

Plus récemment, des auteurs [15] ont introduit la notion d'opérateur proximal associé à une matrice symétrique définie positive (PSD) \mathbf{P} c'est à dire intégrant la métrique induite par \mathbf{P} dans sa définition. Tout d'abord, on rappelle que $\forall \mathbf{x} \in \mathbb{R}^L$, $\|\mathbf{x}\|_{\mathbf{P}}^2 = \langle \mathbf{x}, \mathbf{P}\mathbf{x} \rangle$, où $\langle \cdot, \cdot \rangle$ est le produit scalaire. L'opérateur proximal de la fonction $\varphi \in \Gamma_0(\mathbb{R}^L)$ associé à une matrice SPD \mathbf{P} est défini par [15]

$$\text{prox}_{\mathbf{P}, \varphi}: \mathbb{R}^L \rightarrow \mathbb{R}^L: \mathbf{v} \mapsto \arg \min_{\mathbf{u} \in \mathbb{R}^L} \frac{1}{2} \|\mathbf{u} - \mathbf{v}\|_{\mathbf{P}}^2 + \varphi(\mathbf{u}). \quad (4.16)$$

On remarque que si \mathbf{P} est l'identité, alors la définition (4.16) se réduit à la définition de l'opérateur proximal classique (4.13).

Ces opérateurs sont très utiles pour la minimisation des fonctionnelles formulées pour la résolution de problème inverse.

4.3.2 Un aperçu des algorithmes proximaux

Plaçons nous dans le cadre des problèmes d'optimisation exposés dans la section 4.2.

Intéressons nous tout d'abord au problème 4.8. Ce problème consiste en la minimisation d'une fonctionnelle composée de la somme de deux fonctions de $\Gamma_0(\mathcal{H})$. Si l'on suppose que l'un des deux termes est différentiable alors le problème 4.8 peut être résolu à l'aide d'un algorithme de type implicite-explicite [16, 17, 18, 13, 5, 19] dont nous donnons le schéma dans l'algorithme 1. Si aucun des deux termes n'est différentiable mais que les opérateurs proximaux de f_1 et f_2 sont calculables, alors on pourra utiliser l'algorithme de Douglas-Rachford [20, 21].

Lorsque plusieurs termes de régularisation sont considérés, on peut utiliser l'algorithme parallèle proximal (PPXA) [14] pour résoudre 4.9.

Plaçons nous à présent dans un cadre plus général en considérant une fonctionnelle composée de P termes et impliquant des opérateurs linéaires (soient les problèmes

1. <http://proximity-operator.net/>

Algorithm 1 Algorithme proximal Implicite-Explicite pour la résolution du problème 4.8 dans le cas où f_1 est supposée différentiable de gradient continu β -Lipschitz.

- 1: Soit $(\gamma_k)_{k \in \mathbb{N}}$ une suite de $]0, +\infty[$ telle que $0 < \inf_{k \in \mathbb{N}} \gamma_k \leq \sup_{k \in \mathbb{N}} \gamma_k < 2/\beta$ // *Choix pas de l'algorithme*
 - 2: Soit $(\lambda_k)_{k \in \mathbb{N}}$ une suite dans $]0, 1]$ telle que $\inf_{k \in \mathbb{N}} \lambda_k > 0$ // *Choix du paramètre de relaxation*
 - 3: Soient $(a_k)_{k \in \mathbb{N}}$ et $(b_k)_{k \in \mathbb{N}}$ des suites de \mathcal{H} telles que $\sum_{k \in \mathbb{N}} \|a_k\| < +\infty$ et $\sum_{k \in \mathbb{N}} \|b_k\| < +\infty$ // *Robustesse de l'algorithme aux erreurs numériques*
 - 4: On fixe $y_0 \in \mathcal{H}$
 - 5: **for** $k = 0, 1, \dots, k_{\max}$ **do** // *k-ème itération*
 - 6: $y_{k+1} = y_k + \lambda_k \left(\text{prox}_{\gamma_k f_2} (y_k - \gamma_h (\nabla f_1(x_k) + b_k)) + a_k - x_k \right)$ // *Une étape de gradient et une étape de prox à chaque itération*
 - 7: **end for**
-

4.10.4.11). Une première solution consiste à calculer l'opérateur proximal d'une fonction $f \in \Gamma_0(\mathcal{G})$ composé avec un opérateur linéaire $L : \mathcal{H} \rightarrow \mathcal{G}$.

Proposition 4.2 [21, Proposition 11] Soit \mathcal{G} un espace de Hilbert réel, $f \in \Gamma_0(\mathcal{G})$ et $L : \mathcal{H} \rightarrow \mathcal{G}$ un opérateur linéaire borné tel que $L \circ L^* = \mu \text{Id}$, pour $\mu \in]0, +\infty[$, alors $f \circ L \in \Gamma_0(\mathcal{H})$ et

$$\text{prox}_{f \circ L} = \text{Id} + \mu^{-1} L^* \circ (\text{prox}_{\mu f} - \text{Id}) \circ L \quad (4.17)$$

Nous avons étendu cette proposition à un cadre plus général dans [22, Proposition 3.4]. Il est cependant intéressant de noter que la proposition 4.2 s'applique en particulier aux trames ajustées et μ représente alors la constante de trame.

La deuxième solution, consiste à employer d'autres algorithmes comme les algorithmes primaux PPXA+ [23] et le forward-backward généralisé [24] ou les algorithmes primaux-duaux comme l'algorithme Monotone + Skew Forward-Backward-Forward (M+S FBF) [25] ou Monotone + Lipschitz Forward-Backward-Forward (M+L FBF) [26], ou encore ceux basés sur l'algorithme Explicite-Implicite [27, 28, 29]. La majorité de ces algorithmes requiert le calcul de l'opérateur proximal de chacune des fonctions présentes dans la fonctionnelle à minimiser, certains prennent en compte le caractère Lipschitz différentiable de certaines fonctions, d'autres requièrent l'inversion d'opérateurs linéaires de grandes tailles ... le choix de l'algorithme se fait donc suivant le contexte. J'ai choisi de décrire à présent le fonctionnement de l'algorithme M+L FBF puisque je vais l'utiliser à plusieurs reprises par la suite. Nous sommes donc dans le contexte du problème 4.10 où sans perte de généralité on suppose que : $F_1 = \text{Id}$, f_1 est différentiable de gradient μ -Lipschitz et $F_2 = \text{Id}$. Le problème considéré ici est donc de la forme

$$\underset{y \in \mathcal{H}}{\text{minimiser}} \quad f_1(y) + f_2(y) + \sum_{p=3}^P f_p(F_p y) \quad (4.18)$$

et peut être résolu grâce à l'algorithme 2.

Enfin, lorsque plusieurs termes de régularisation sont considérés et qu'en plus ces régularisations portent sur des sous-ensembles de coordonnées, c'est à dire dans le cas du problème 4.12, on a alors recourt à des algorithmes fonctionnant par blocs de coordonnées. Je me suis en particulier intéressée à l'algorithme Explicite-Implicite à métrique

Algorithm 2 Algorithme proximal M+L FBF pour la résolution du problème 4.18.

```

1: Soit  $\beta = \mu + \sqrt{\sum_{p=3}^P \|L_p\|^2}$ 
2: On choisit  $\epsilon \in ]0, \frac{1}{\beta+1}[$ 
3: Soit  $(\gamma_k)_{k \in \mathbb{N}}$  une suite de  $[\epsilon, \frac{1-\epsilon}{\beta}]$  // Choix du pas de l'algorithme
4: On choisit  $y_0 \in \mathcal{H}$  et  $(v_{3,0}, \dots, v_{P,0}) \in \mathcal{G}_3 \oplus \dots \oplus \mathcal{G}_P$  // Initialisation des variables
   primales et duales
5: for  $k = 0, 1, \dots, k_{\max}$  do // k-ème itération
6:    $x_{1,k} = y_k - \gamma_k \nabla f_1(y_k) + \sum_{p=3}^P L_p^* v_{p,k}$ 
7:    $r_{1,k} = \text{prox}_{\gamma_n f_2}(x_{1,k})$ 
8:   for  $p = 3, 4, \dots, P$  do // Mise à jour des variables duales
9:      $x_{2,p,k} = v_{1,k} + \gamma_k L_p y_k$ 
10:     $r_{2,p,k} = x_{2,p,k} - \gamma_n \text{prox}_{\gamma_n^{-1} f_p}(\gamma_n^{-1} x_{2,p,k})$ 
11:     $q_{2,p,k} = r_{2,p,k} + \gamma_k L_p r_{1,k}$ 
12:     $v_{p,k+1} = v_{p,k} - x_{2,p,k} + q_{2,p,k}$ 
13:   end for
14:    $q_{1,k} = r_{1,k} - \gamma_k \nabla f - 1(r_{1,k}) + \sum_{p=3}^P L_p^* r_{2,p,k}$ 
15:    $y_{k+1} = y_k - x_{1,k} + q_{1,k}$  // Mise à jour de la variable primale
16: end for

```

variable par bloc de coordonnées (*Block Coordinate Variable Metric Forward-Backward* (BC-VMFB) [30, 31] qui constitue une généralisation de l'algorithme PALM (Proximal Alternating Linearized Minimization) [32]. Cet algorithme est fondé sur les opérateurs proximaux associés à une matrice PSD \mathbf{P} (cette dernière jouant alors le rôle de matrice de préconditionnement). Reprenons le problème 4.12 et explicitons ce dernier dans un formalisme plus précis et considérons le problème de minimisation suivant

$$\underset{y \in \mathcal{H}}{\text{minimiser}} \left\{ f_1(y) + \sum_{p=2}^P f_p(y^{(p)}) \right\}, \quad (4.19)$$

où, $f_1 \in \Gamma_0(\mathcal{H})$ est différentiable de gradient β -Lipschitz et chaque fonction $f_p \in \Gamma_0(\mathcal{H}_p)$ est supposée être minorée par une fonction affine, et sa restriction à son domaine est continue pour tout $p = 2, \dots, P$.

Ce problème d'optimisation peut être résolu via l'algorithme BC-VMFB. Cet algorithme consiste en deux étapes comme détaillé dans l'algorithme 3.

Ce type d'algorithme a été utilisé pour de nombreuses applications comme la factorisation en matrices non-négatives [33].

4.4 Conclusion

Une fois le problème d'optimisation posé, il convient de le résoudre à l'aide d'algorithmes performants et adapté au contexte. Nous le verrons dans le chapitre 6 tout en ayant déjà un aperçu dans le chapitre 5.

Mais comme vous vous en serez sûrement aperçu et comme évoqué plus haut, avant même de procéder à l'optimisation de ces différents critères, encore faut-il avoir déterminé les hyperparamètres, que ce soit les paramètres du modèle (dégradation) ou les paramètres de régularisation qui permettent de donner un poids plus ou moins important

Algorithm 3 Algorithme BC-VMFB pour la résolution du problème 4.19

```

1: Soient  $y_0 \in \mathcal{H}$ ,  $k \in \mathbb{N}$  et  $\mu_k \in ]0, +\infty[$  // Initialisation
2: for  $k = 0, 1, \dots$  do //  $k$ -ème itération
3:   Soit  $p_k \in \{2, \dots, P\}$  // Le bloc  $p_k$  est considéré (suivant une certaine règle)
4:   Soit  $\mathbf{P}_{p_k}(y_k)$  une matrice PSD // Construction des preconditionneurs  $\mathbf{P}_{p_k}(y_k)$ 
5:   Soit  $\nabla_{p_k} f_1(y_k)$  le gradient de  $f_1$  // Calcul des gradients partiels
6:    $\tilde{y}_k^{(p_k)} = \mathbf{x}_k^{(p_k)} - \mu_k \mathbf{P}_{p_k}(y_k)^{-1} \nabla_{p_k} f_1(y_k)$  // Etape de gradient pour la mise à jour du
   bloc  $p_k$ 
7:    $y_{k+1}^{(p_k)} \in \text{prox}_{\mu_k^{-1} \mathbf{P}_{p_k}(y_k), f_{p_k}}(\tilde{y}_k^{(p_k)})$  // Etape proximale appliquée au bloc  $p_k$ 
8:    $y_{k+1}^{\bar{p}} = y_k^{\bar{p}}$  où  $\bar{p} = \{1, \dots, P\} \setminus \{p\}$  // Les autres blocs restent inchangés
9: end for

```

au terme de fidélité ou au(x) terme(s) de régularisation. C'est ce à quoi nous allons nous intéresser dans le chapitre suivant.

4.5 Quelques publications personnelles en lien avec ce chapitre

- [CCm1] N. Hajlaoui, C. Chaux, G. Perrin, F. Falzon, and A. Benazza-Benyahia. Satellite image restoration in the context of a spatially varying point spread function. *J. Opt. Soc. Am. A*, 27(6) :1473–1481, 2010.
- [CCm2] L. Chaari, E. Chouzenoux, N. Pustelnik, C. Chaux, and S. Moussaoui. OPTI-MED : Optimisation itérative pour la résolution de problèmes inverses de grande taille. *Traitement du signal*, 28(3-4) :329–374, 2011.
- [CCm3] A. Jezierska, C. Chaux, J.-C. Pesquet, H. Talbot, and G. Engler. An EM approach for time-variant Poisson-Gaussian model parameter estimation. *IEEE Trans. Signal Process.*, 62(1) :17–30, Jan. 2014.
- [CCm4] M. Q. Pham, L. Duval, C. Chaux, and J.-C. Pesquet. A primal-dual proximal algorithm for sparse template-based adaptive filtering : Application to seismic multiple removal. *IEEE Trans. Signal Process.*, 62(16) :4256–4269, Aug. 2014.
- [CCm5] X. T. Vu, C. Chaux, N. D. Thirion-Moreau, S. Maire, and E. M. Carstea. A new penalized nonnegative third order tensor decomposition using a block coordinate proximal gradient approach : application to 3d fluorescence spectroscopy. *J. Chemometrics*, 31(4), Apr. 2017.

Chapitre 5

Estimation de paramètres et hyperparamètres

Avant-propos

Nous nous intéressons dans ce chapitre à l'estimation des hyperparamètres qui peuvent être de différentes natures : des paramètres liés au modèle direct comme les paramètres du bruit ou encore des paramètres liés au choix de la fonction de régularisation. Après un bref rappel des problématiques générales, je me focaliserai sur trois problématiques : l'estimation des paramètres d'un bruit de type Poisson-Gaussien, l'estimation des paramètres de régularisation lorsqu'elle est effectuée dans le domaine des ondelettes et enfin, l'estimation d'un bruit structuré (échos) et sa suppression dans des signaux sismiques.

Sommaire

5.1 Estimation des paramètres du bruit	45
5.2 Estimation des paramètres de régularisation	53
5.3 Estimation des paramètres du modèle d'acquisition	60
5.4 Conclusion	70
5.5 Quelques publications personnelles en lien avec ce chapitre	71

5.1 Estimation des paramètres du bruit

5.1.1 Contexte

Ces travaux ont été effectués dans le cadre de la thèse d'Ania Jezierska. Nous avons travaillé sur des images de MACROscopie, qui est un type de microscopie permettant d'imager des spécimens allant de la taille de la cellule à celle d'un petit animal.

Ces images $(u_s)_{1 \leq s \leq S}$ où s représente un indice de localisation (e.g. un pixel (s_1, s_2) en 2D ou (s_1, s_2, s_3) en 3D), sont dégradées par un bruit résultant du mélange d'un bruit de Poisson de paramètre d'échelle $\alpha \in (0, +\infty)$ et d'un bruit Gaussien de moyenne $c \in \mathbb{R}$ et de variance $\sigma^2 > 0$.

5.1.2 Estimation à partir de trames temporelles

Nous avons supposé dans un premier temps que nous disposions de T réalisations du bruit qui correspondent à effectuer une acquisition temporelle pour $1 \leq t \leq T$.

Le problème direct s'écrit donc $(\forall s \in \{1, \dots, S\})(\forall t \in \{1, \dots, T\})$

$$R_{s,t} = \alpha Q_{s,t} + W_{s,t} \quad (5.1)$$

où $Q_{s,t} \sim \mathcal{P}(u_s)$, $W_{s,t} \sim \mathcal{N}(c, \sigma^2)$.

L'objectif est donc d'estimer $u = (u_s)_{1 \leq s \leq S}$, α , c et σ à partir des observations $r = (r_{s,t})_{1 \leq s \leq S, 1 \leq t \leq T}$, qui constituent une réalisation d'un champ aléatoire $R = (R_{s,t})_{1 \leq s \leq S, 1 \leq t \leq T}$. Nous avons donc $S + 3$ paramètres à estimer et nous noterons le vecteur de paramètres $\theta = (u, \alpha, c, \sigma)$.

On suppose ici que u est déterministe, que $Q = (Q_{s,t})_{1 \leq s \leq S, 1 \leq t \leq T}$ et $W = (W_{s,t})_{1 \leq s \leq S, 1 \leq t \leq T}$ sont mutuellement indépendants et que les composantes de W (resp. Q) sont indépendantes.

L'estimation au sens du maximum de vraisemblance (MV) des paramètres est donnée par

$$(\hat{u}, \hat{\alpha}, \hat{c}, \hat{\sigma}) = \underset{(u, \alpha, c, \sigma)}{\operatorname{argmax}} f_R(r \mid u, \alpha, c, \sigma). \quad (5.2)$$

où $f_R(\cdot \mid u, \alpha, c, \sigma)$ est la fonction densité de probabilité de R .

Soit $p_{R,Q}(\cdot, \cdot \mid u, \alpha, c, \sigma)$ la distribution de probabilité mixte continue/discrète de (R, Q) . En utilisant la règle de Bayes, on obtient

$$\begin{aligned} p_{R,Q}(r, q \mid u, \alpha, c, \sigma) &= f_{R|Q=q}(r \mid u, \alpha, c, \sigma) P(Q = q \mid u) \\ &= f_W(r - \alpha q \mid c, \sigma) P(Q = q \mid u) \end{aligned} \quad (5.3)$$

où $f_{R|Q=q}(\cdot \mid u, \alpha, c, \sigma)$ est la fonction densité de probabilité de R sachant que $Q = q$ et $f_W(\cdot \mid c, \sigma)$ est la fonction densité de probabilité de W .

L'estimation par maximum de vraisemblance s'annonce compliquée dans ce cas. C'est pourquoi nous avons opté pour une approche de type EM qui s'exprime comme le problème d'optimisation suivant :

$$(\forall n \in \mathbb{N}) \quad \theta^{(n+1)} = \underset{\theta}{\operatorname{argmin}} J(\theta \mid \theta^{(n)}), \quad (5.4)$$

où

$$\begin{aligned} J(\theta \mid \theta^{(n)}) &= \frac{1}{2\sigma^2} \sum_{s=1}^S \sum_{t=1}^T \mathbb{E}_{Q|R=r, \theta^{(n)}} [(r_{s,t} - \alpha Q_{s,t} - c)^2] \\ &\quad + \frac{ST}{2} \ln(\sigma^2) + T \sum_{s=1}^S u_s - \sum_{s=1}^S \ln u_s \sum_{t=1}^T \mathbb{E}_{Q|R=r, \theta^{(n)}} [Q_{s,t}]. \end{aligned} \quad (5.5)$$

La solution de ce problème s'écrit itérativement comme

$$(\forall s \in \{1, \dots, S\}) \quad u_s^{(n+1)} = \frac{1}{T} \sum_{t=1}^T \mathbb{E}_{Q|R=r, \theta^{(n)}} [Q_{s,t}] \quad (5.6)$$

$$\begin{bmatrix} ST & \sum_{s,t} \mathbb{E}_{Q|R=r, \theta^{(n)}} [Q_{s,t}] \\ \sum_{s,t} \mathbb{E}_{Q|R=r, \theta^{(n)}} [Q_{s,t}] & \sum_{s,t} \mathbb{E}_{Q|R=r, \theta^{(n)}} [Q_{s,t}^2] \end{bmatrix} \begin{bmatrix} c^{(n+1)} \\ \alpha^{(n+1)} \end{bmatrix} = \begin{bmatrix} \sum_{s,t} r_{s,t} \\ \sum_{s,t} r_{s,t} \mathbb{E}_{Q|R=r, \theta^{(n)}} [Q_{s,t}] \end{bmatrix} \quad (5.7)$$

$$(\sigma^2)^{(n+1)} = \frac{1}{ST} \sum_{s=1}^S \sum_{t=1}^T r_{s,t} \left(r_{s,t} - \alpha^{(n+1)} \mathbb{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] - c^{(n+1)} \right). \quad (5.8)$$

Toute la difficulté repose donc sur le calcul de $\mathbb{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}]$ et $\mathbb{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}^2]$.
Pour tout $(t, s) \in \{1, \dots, T\} \times \{1, \dots, S\}$, nous avons

$$\mathbb{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] = \sum_{q_{s,t}=1}^{+\infty} q_{s,t} \mathbb{P}(Q_{s,t} = q_{s,t} \mid R = r, \theta^{(n)}). \quad (5.9)$$

De plus,

$$\mathbb{P}(Q_{s,t} = q_{s,t} \mid R = r, \theta^{(n)}) = \frac{p_{R_{s,t}, Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta^{(n)})}{f_{R_{s,t}}(r_{s,t} \mid \theta^{(n)})}. \quad (5.10)$$

D'après (5.3), on en déduit que

$$\mathbb{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] = \frac{\zeta_{s,t}^{(n)}}{\eta_{s,t}^{(n)}}, \quad (5.11)$$

où

$$\zeta_{s,t}^{(n)} = \sum_{q_{s,t}=1}^{+\infty} \exp\left(-\frac{(r_{s,t} - \alpha^{(n)} q_{s,t} - c^{(n)})^2}{2(\sigma^2)^{(n)}}\right) \frac{(u_s^{(n)})^{q_{s,t}}}{(q_{s,t} - 1)!} \quad (5.12)$$

$$\eta_{s,t}^{(n)} = \sum_{q_{s,t}=0}^{+\infty} \exp\left(-\frac{(r_{s,t} - \alpha^{(n)} q_{s,t} - c^{(n)})^2}{2(\sigma^2)^{(n)}}\right) \frac{(u_s^{(n)})^{q_{s,t}}}{q_{s,t}!}. \quad (5.13)$$

De même,

$$\mathbb{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}^2] = \frac{\xi_{s,t}^{(n)}}{\eta_{s,t}^{(n)}}, \quad (5.14)$$

où

$$\xi_{s,t}^{(n)} = \sum_{q_{s,t}=1}^{+\infty} q_{s,t} \exp\left(-\frac{(r_{s,t} - \alpha^{(n)} q_{s,t} - c^{(n)})^2}{2(\sigma^2)^{(n)}}\right) \frac{(u_s^{(n)})^{q_{s,t}}}{(q_{s,t} - 1)!}. \quad (5.15)$$

La difficulté ici vient de la sommation infinie sur $q_{s,t}$. Il va donc falloir tronquer cette sommation. Pour cela, nous proposons une stratégie où nous dérivons une borne inférieure $q_{s,t}^-$ et une borne supérieure $q_{s,t}^+$ de sommation données par

$$q_{s,t}^- = \lfloor q_{s,t}^* - \Delta \frac{\sigma}{\alpha} \rfloor, \quad q_{s,t}^+ = \lceil q_{s,t}^* + \Delta \frac{\sigma}{\alpha} \rceil \quad (5.16)$$

où $\Delta > 0$ et

$$q_{s,t}^* = \frac{\sigma^2}{\alpha^2} W\left(\frac{\alpha^2}{\sigma^2} u_s e^{\frac{\alpha}{\sigma^2}(r_{s,t} - c - d\alpha)}\right) \quad (5.17)$$

et W est la fonction W de Lambert. Cette stratégie nous permet de borner l'erreur d'approximation [3, Proposition A.2].

Je n'en parlerai pas ici mais je tenais à mentionner que nous avons également travaillé avec un modèle plus complexe où

$$Q_{s,t} \sim \mathcal{P}(u_s e^{-k_s t}). \quad (5.18)$$

L'inclusion de l'exponentielle décroissante $e^{-k_s t}$ modélise les phénomènes de photoblanchiment c'est à dire les phénomènes de décroissance de la fluorescence au cours du temps. Les paramètres k_s peuvent alors être ajoutés au vecteur de paramètres θ et estimés [3].

5.1.3 Estimation à partir d'une seule image

Dans certains cas, l'acquisition ne se fait pas au cours du temps et nous ne disposons alors que d'une réalisation du bruit. L'idée ici a été de se dire qu'il était possible de découper l'image en zones possédant des propriétés communes et d'utiliser ces différentes zones partageant la même propriété comme étant différentes réalisations du bruit. L'approche étudiée ici utilise donc deux étapes fondamentales : une étape de segmentation et une étape d'estimation des paramètres basées sur l'algorithme EM présenté dans la partie précédente.

La modélisation du problème est donc légèrement modifiée, les T trames temporelles étant remplacées par une partition \mathbb{P} des observations notées ici y .

Soit x l'image originale de taille $N \times M$. On suppose que x peut prendre $K \in \mathbb{N}^*$ valeurs d'intensité différentes $\{u_1, \dots, u_K\}$. On notera par la suite u le vecteur $[u_1, \dots, u_K]^\top \in \mathbb{R}^K$. La partition \mathbb{P} définie plus haut peut s'écrire comme $\mathbb{P} = (\mathbb{D}_k)_{1 \leq k \leq K}$ et elle peut être reliée à x par la relation suivante

$$x = (u_{i_{\mathbb{P}}(s)})_{s \in \mathbb{S}} \in \{u_1, \dots, u_K\}^{N \times M}, \quad (5.19)$$

où $(i_{\mathbb{P}}(s))_{s \in \mathbb{S}} \in \mathbb{I} = \{1, \dots, K\}^{N \times M}$ représente l'image des labels

$$(\forall s \in \mathbb{S})(\forall k \in \{1, \dots, K\}) \quad i_{\mathbb{P}}(s) = k \quad \Leftrightarrow \quad s \in \mathbb{D}_k. \quad (5.20)$$

x étant liée aux labels $i_{\mathbb{P}}$ ainsi qu'aux intensités u , on la notera $x_{i_{\mathbb{P}}, u}$ par la suite.

L'observation bruitée $y \in \mathbb{R}^{N \times M}$ est la réalisation d'un vecteur aléatoire Y et est définie comme suit

$$y = \alpha q + w \quad (5.21)$$

où α est le facteur d'échelle, et $q = (q(s))_{s \in \mathbb{S}}$ (resp. w) sont des réalisations mutuellement indépendantes des vecteurs aléatoires Q (resp. $W = (W(s))_{s \in \mathbb{S}}$) de composantes indépendantes.

De manière très similaire à la partie précédente, nous avons

$$Q \sim \mathcal{P}(x_{i_{\mathbb{P}}, u}) \quad (5.22)$$

$$(\forall s \in \mathbb{S}) \quad W(s) \sim \mathcal{N}(c, \sigma^2), \quad (5.23)$$

Comme précédemment nous cherchons à estimer les paramètres c, σ, α et u . Soit $\theta \in \mathcal{T}$ le vecteur de paramètres défini par $\theta = [c, \sigma^2, \alpha, u^\top]^\top$ et $\mathcal{T} = \mathbb{R} \times]0, +\infty[^2 \times C$ où C est un sous-ensemble convexe fermé de \mathbb{R}^K (contrainte d'ordre sur les valeurs de u).

Le vecteur de paramètres θ est lié au vecteur d'observations y à travers la fonction densité de probabilité Y , définie par : $\forall y = (y(s))_{s \in \mathbb{S}} \in \mathbb{R}^{N \times M}$,

$$p_Y(y; i_{\mathbb{P}}, \theta) = \prod_{k=1}^K \prod_{s \in \mathbb{D}_k} \left(\sum_{q_s=0}^{+\infty} \frac{e^{-u_k} (u_k)^{q_s}}{q_s!} \frac{e^{-\frac{(y(s) - \alpha q_s - c)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \right). \quad (5.24)$$

Nous adoptons ici une approche variationnelle où le vecteur de paramètres θ est obtenu par la minimisation d'un critère pénalisé. Cela revient à résoudre

$$\underset{(i_{\mathbb{P}}, \theta) \in \mathbb{I} \times \mathcal{T}}{\text{minimiser}} \quad \Phi(\theta, i_{\mathbb{P}}, y) + \rho(i_{\mathbb{P}}). \quad (5.25)$$

Le terme Φ ci-dessus est l'antilog vraisemblance de y :

$$\Phi(\theta, i_{\mathbb{P}}, y) = -\log(p_Y(y; i_{\mathbb{P}}, \theta)) \quad (5.26)$$

et ρ est une fonction de régularisation permettant d'introduire de l'information a priori sur les ensembles de niveaux de l'image. Le problème (5.25) est non convexe car $i_{\mathbb{P}}$ appartient à un ensemble discret de valeurs (non convexe).

Nous proposons assez naturellement un algorithme d'optimisation alternée qui s'écrit comme suit

Algorithm 4 Procédure alternée

Initialisation :

Fixons $K \in \mathbb{N}^*$ et $\theta^{(0)} \in \mathcal{T}$.

Boucle principale :

For $\ell = 0, 1, \dots$

$$\left[\begin{array}{l} i_{\mathbb{P}}^{(\ell)} \in \underset{i_{\mathbb{P}} \in \mathbb{I}}{\text{argmin}} \quad \Phi(\theta^{(\ell)}, i_{\mathbb{P}}, y) + \rho(i_{\mathbb{P}}) \\ \theta^{(\ell+1)} \in \underset{\theta \in \mathcal{T}}{\text{argmin}} \quad \Phi(\theta, i_{\mathbb{P}}^{(\ell)}, y) \end{array} \right.$$

La première étape de l'algorithme 4, qui consiste à estimer $i_{\mathbb{P}}$, peut être résolue grâce à l'algorithme de quantification proposé dans [34] (et non détaillé dans ce manuscrit, alliant optimisation continue et discrète). La deuxième étape de l'algorithme 4, quand à elle s'avère plus complexe car non convexe. Ceci dit, l'idée ici est de s'appuyer sur ce qui a été fait précédemment pour l'estimation des paramètres du bruit dans le cas d'une série d'images. Cette étape est résolue grâce à un algorithme EM dans le même esprit que celui présenté dans la partie précédente. Nous résumons ci-après les étapes clés. L'objectif est donc de générer des estimés de notre vecteur de paramètres, $(\theta^{(\ell)})_{\ell \in \mathbb{N}}$, en résolvant

$$(\forall \ell \in \mathbb{N}) \quad \theta^{(\ell+1)} = \underset{\theta \in \mathcal{T}}{\text{argmin}} \quad \Phi(\theta | \theta^{(\ell)}) \quad (5.27)$$

où

$$\begin{aligned} \Phi(\theta | \theta^{(\ell)}) = & \frac{1}{2\sigma^2} \sum_{s \in \mathbb{S}} E_{Q|Y=y, \theta^{(\ell)}} [(y(s) - \alpha Q(s) - c)^2] + \frac{NM}{2} \ln(\sigma^2) \\ & + \sum_{k=1}^K \text{card}(\mathbb{D}_k^{(\ell)}) u_k - \sum_{k=1}^K \ln u_k \sum_{s \in \mathbb{D}_k^{(\ell)}} E_{Q|R=r, \theta^{(\ell)}} [Q(s)]. \end{aligned} \quad (5.28)$$

et $(Q(s))_{s \in \mathbb{S}}$ sont les composantes de Q définies dans (5.22) avec $i_{\mathbb{P}} = i_{\mathbb{P}}^{(\ell)}$. On voit ainsi que $u^{(\ell+1)}$ est le résultat d'un problème d'optimisation sous contrainte (ensemble convexe C) résolu par une approche proximale primale-duale [26], $c^{(\ell+1)}$ et $\alpha^{(\ell+1)}$ sont obtenus par résolution d'un système d'équations et $(\sigma^2)^{(\ell+1)}$ a une forme explicite. Cependant comme précédemment, les espérances conditionnelles impliquent des sommes infinies qui doivent être tronquées comme nous l'avons vu pour l'estimation des paramètres à partir d'une série d'images.

5.1.4 Simulations numériques

5.1.4-a) Estimation à partir d'une série d'images

Le tableau 5.1 fournit quelques résultats numériques effectués sur données simulées. On y présente le biais et l'écart-type des paramètres estimés calculés pour 20 réalisations de bruit. On peut remarquer que l'algorithme EM ainsi proposé fournit une très bonne précision d'estimation. Les conditions des tests qui ont été réalisés ici sont volontairement défavorables pour la méthode fondée sur les cumulants. On peut fournir les conclusions suivantes : 1) la fiabilité de l'algorithme EM croît avec T , 2) l'algorithme EM fournit une bonne estimation même si T est petit, 3) il n'est pas sensible aux petites valeurs de α et c et 4) les résultats sont moins fiables en présence d'un fort bruit.

No.	Param.					Algorithme EM						Erreur	
	S	T	σ^2	c	α	$\hat{\sigma}^2$		\hat{c}		$\hat{\alpha}$		$\overline{\text{err}}_{\text{cum}}$	$\overline{\text{err}}_{\text{EM}}$
						biais	std	biais	std	biais	std		
1	1024	50	100	10	10	-1.86	14.02	-0.35	1.79	-0.21	0.05	0.22	0.11
2	1024	100	100	10	10	0.54	10.92	0.27	1.38	-0.08	0.02	0.16	0.08
3	1024	200	100	10	10	-0.91	7.89	0.09	0.88	-0.04	0.03	0.14	0.05
4	1024	50	100	5	10	1.30	13.16	0.08	1.61	-0.20	0.05	0.33	0.15
5	1024	50	25	10	1	-0.28	0.50	0.00	0.48	-0.02	0.01	0.64	0.03
6	1024	50	400	10	10	7.51	45.24	0.71	5.13	-0.22	0.07	0.34	0.21

TABLE 5.1 – Performance de l'algorithme EM pour des cas défavorables pour la méthode basée sur les cumulants.

Nous avons également appliqué notre méthode à des séries temporelles réelles, acquises avec un microscope confocal (Leica TCS-LSI), correspondant à une coupe de rhizome de *Convallaria majalis* (muguet).

La série d'images traitée se compose de 300 images ayant une résolution de 12-bit de taille 190×190 , ce qui correspond à $T = 300$ et $S = 36100$. Les Fig. 5.1 (a,b) illustrent la première et la dernière image de la série considérée. L'image estimée est représentée en Fig. 5.1 (c) où les paramètres du modèle ont été estimés comme $\hat{\alpha} = 168$, $\hat{c} = 114$, $\hat{\sigma}^2 = 64.1^2$.

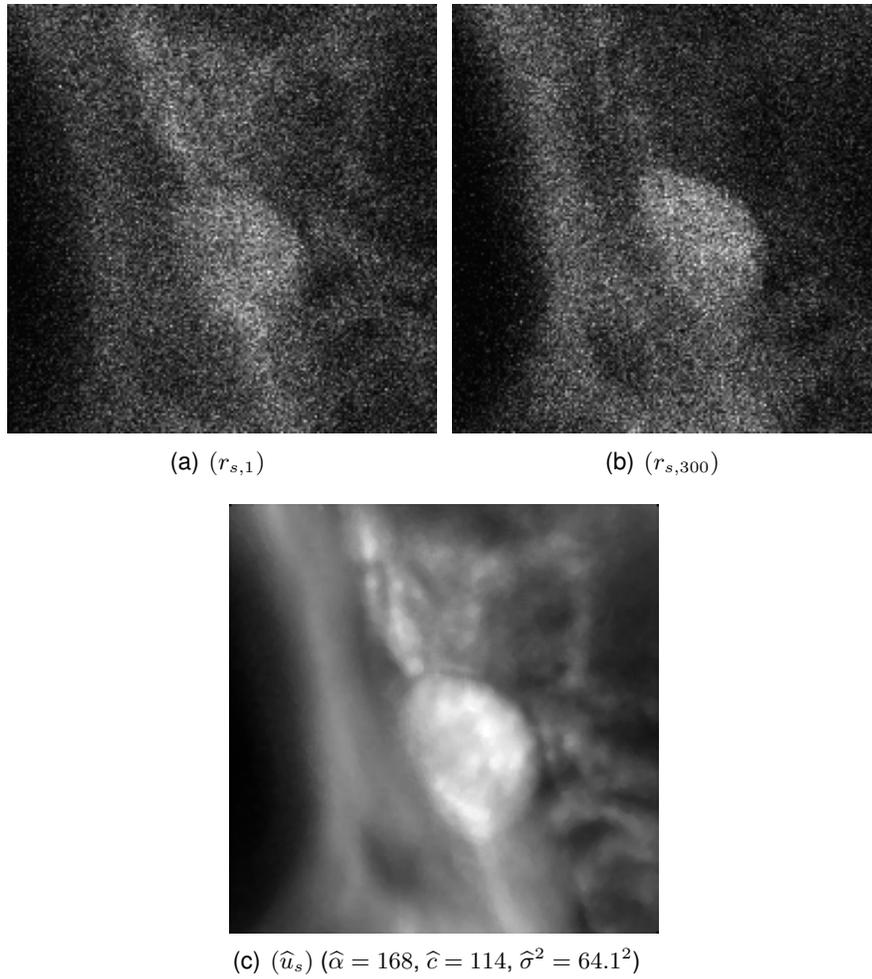


FIGURE 5.1 – Résultats sur données réelles.

5.1.4-b) Estimation à partir d'une seule image

Nous nous intéressons maintenant à l'estimation des mêmes paramètres mais à partir d'une seule image. Nous avons vu qu'il nous fallait alors définir des patches choisis ici de taille 12×12 . Par ailleurs, pour tous les tests nous avons fixé $Q = 20$, $\epsilon = 0.2$ et ρ comme étant la variation totale anisotrope soit

$$\rho(i_{\mathcal{P}}) = \mu \left(\sum_{n=1}^{N-1} \sum_{m=1}^M \psi(|i_{\mathcal{P}}(n+1, m) - i_{\mathcal{P}}(n, m)|) + \sum_{n=1}^N \sum_{m=1}^{M-1} \psi(|i_{\mathcal{P}}(n, m+1) - i_{\mathcal{P}}(n, m)|) \right), \quad \mu \geq 0. \quad (5.29)$$

De plus, le paramètre de régularisation a été fixé à $\mu = 1.3$ et ψ est la fonction ℓ_0 définie par

$$(\forall x \in [0, +\infty[) \quad \psi(x) = \begin{cases} 0 & \text{si } x = 0 \\ 1 & \text{sinon.} \end{cases} \quad (5.30)$$

Les résultats sont présentés dans la Fig. 5.2. Pour la première image test, nous avons utilisé un fantôme biologique [35] comme image originale (Fig. 5.2 (a)) pour laquelle les intensités varient de 0.25 à 17.25. On a donc fixé $u_{\max} = 20$. L'image est ensuite dégradée par un bruit de paramètres $\alpha = 20$, $c = 100$, $\sigma^2 = 1000$ (Fig. 5.2 (b)). La procédure d'identification du bruit est appliquée en utilisant la fonction cout binaire et l'on estime alors $\hat{\alpha} = 19.8$, $\hat{c} = 103$, $\hat{\sigma}^2 = 1039$.

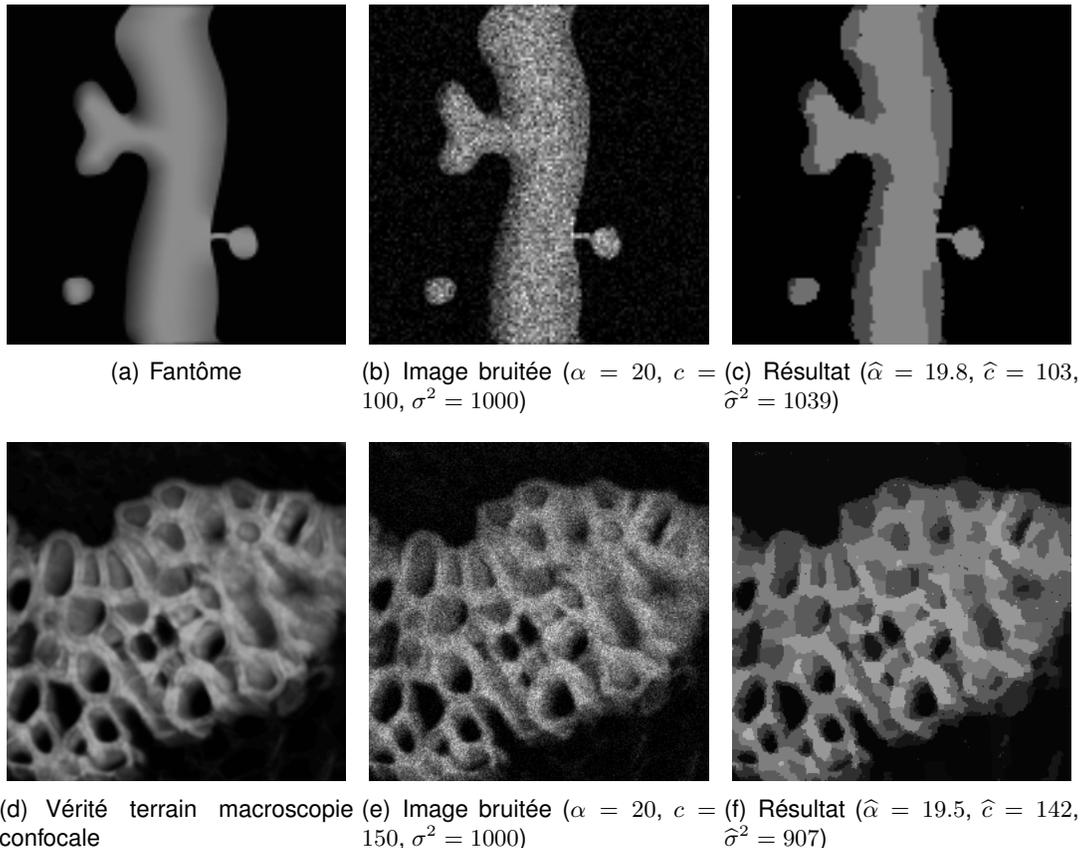


FIGURE 5.2 – Résultats d'identification du bruit à partir d'une seule image

Pour la deuxième expérience, la vérité terrain est obtenue en appliquant la méthode développée pour une série d'images. L'image bruitée est représentée en Fig. 5.2 (e) et les valeurs estimées de $\alpha = 20$, $c = 150$, $\sigma^2 = 1000$ sont $\hat{\alpha} = 19.5$, $\hat{c} = 142$, $\hat{\sigma}^2 = 907$. L'image estimée est représentée dans la figure 5.2 (f).

5.1.5 Conclusion

Nous avons vu dans cette première partie une manière de déterminer les paramètres du modèle de dégradation, dans le cas d'un bruit mixte Poisson-Gaussien. L'approche proposée, basée sur un algorithme EM, fournit également une estimation des données originales, qui n'était pas le but premier de la méthode. L'estimation de ces données est donc sous-optimale mais elle donne déjà une première approximation (qui peut par exemple servir d'initialisation à des méthodes dédiées au débruitage d'image dégradées par un tel bruit). Par ailleurs, les paramètres estimés permettent d'être insérés dans

une seconde procédure dont le but est de restaurer les données dégradées [36, 37]. Nous allons à présent nous intéresser à une toute autre catégorie de paramètres : les paramètres de régularisation.

5.2 Estimation des paramètres de régularisation

Comme je l'évoquais dans le chapitre précédent, la résolution d'un problème inverse passe par la formulation d'un problème de minimisation impliquant un terme de régularisation. Qui dit régularisation dit paramètre(s) de régularisation. Sa détermination est importante puisqu'elle impacte directement la solution. Il est souvent supposé connu et déterminé de manière exhaustive/empirique. Lorsqu'un problème de déconvolution est considéré, je me suis intéressée à l'estimation des paramètres de régularisation lorsque cette dernière est formulée dans le domaine des ondelettes.

5.2.1 Contexte

Plus précisément, le modèle direct considéré ici est le suivant :

$$g = Au + n.$$

où g (vecteur de dimension k) est l'image observée, dégradée par un opérateur de flou A connu et un bruit additif supposé blanc et gaussien $n \sim \mathcal{N}(0, \sigma^2 I)$ (σ connu).

Afin d'estimer l'image u à partir de l'observation dégradée g , on peut adopter une approche variationnelle avec une régularisation définie par une norme ℓ^1 sur des coefficients en ondelettes [18, 21, 16, 17, 38, 5]. Le critère $J(u)$ qui est minimisé est :

$$J(u) = \frac{\|g - Au\|^2}{2\sigma^2} + \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1 \quad (5.31)$$

où $F_{\mathbf{m}}u$ désigne une sous-bande et F représente l'opérateur de transformée en ondelettes. Nous considérons dans ce travail une transformée en ondelettes sur une base orthogonale [39].

La minimisation du critère requiert la connaissance de $\lambda = (\lambda_{\mathbf{m}})_{\mathbf{m}}$ et la question est de savoir comment fixer ces hyperparamètres. Notre objectif est donc l'estimation automatique des hyperparamètres. Estimer λ ne connaissant que l'observation dégradée g est le problème difficile de l'estimation de paramètres en données incomplètes [40].

5.2.2 Approche proposée : estimateur au sens du maximum de vraisemblance

5.2.3 Difficultés

Dans le but d'estimer les paramètres, le critère (5.31) est interprété au sens stochastique. Il correspond à l'énergie de la probabilité *a posteriori* dans un cadre bayésien où le premier terme est l'énergie de la probabilité $p(g|u)$ donnée par la probabilité du bruit n supposé de loi gaussienne

$$p(g|u) = \frac{1}{K_{\sigma}} \exp\left(\frac{-\|g - Au\|^2}{2\sigma^2}\right) \quad (5.32)$$

où la constante vaut $K_\sigma = (2\pi)^{k/2}\sigma^k$ et le deuxième terme correspond à l'énergie du modèle *a priori* donnée par

$$p_\lambda(u) = \frac{1}{Z_\lambda} \prod_{\mathbf{m}} \exp(-\lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1) \quad (5.33)$$

où la constante Z_λ est la constante de normalisation définie par

$$Z_\lambda = \int_u \exp\left(-\sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1\right) du \quad (5.34)$$

La variable u est un champ discret de k variables aléatoires réelles. Ainsi la variable u prend ses valeurs dans \mathbb{R}^k , qui est le domaine d'intégration de (5.34). F étant un changement de base, $\text{Ker}(F) = \{0\}$ et l'intégrale (5.34) converge en supposant que $\exists \mathbf{m}, \lambda_{\mathbf{m}} \geq \epsilon > 0$.

L'estimation de λ au sens du maximum de vraisemblance consiste à maximiser $p_\lambda(g)$ en λ . Par inférence puis dérivation, l'estimateur peut être calculé par une méthode de gradient [40]. On a :

$$p_\lambda(g) = \int_u p_\lambda(g, u) du = \int_u p(g|u) p_\lambda(u) du = \frac{Z_{\sigma, \lambda}(g)}{K_\sigma Z_\lambda} \quad (5.35)$$

où $Z_{\sigma, \lambda}(g) = \int_u \exp\left(-\frac{\|g - Au\|^2}{2\sigma^2} - \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1\right) du$.

Pour maximiser $p_\lambda(g)$ en λ on cherche la dérivée de (5.35) par rapport à chaque $\lambda_{\mathbf{m}}$. On note que, étant donné qu'il n'est pas possible de montrer la convexité de $p_\lambda(g)$ en λ , un algorithme de gradient donnera un minimum local, dépendant des valeurs initiales de λ .

La dérivée de (5.35) par rapport à chaque $\lambda_{\mathbf{m}}$ est (en appliquant le théorème de convergence dominée de Lebesgue afin d'inverser dérivée et signe intégral) :

$$\frac{\partial \log p_\lambda(g)}{\partial \lambda_{\mathbf{m}}} = \frac{\partial \log Z_{\sigma, \lambda}(g)}{\partial \lambda_{\mathbf{m}}} - \frac{\partial \log(Z_\lambda)}{\partial \lambda_{\mathbf{m}}}$$

et

$$\begin{aligned} \frac{\partial \log Z_{\sigma, \lambda}(g)}{\partial \lambda_{\mathbf{m}}} &= \frac{-1}{Z_{\sigma, \lambda}(g)} \int_u \|F_{\mathbf{m}}u\|_1 \exp\left(\frac{-\|g - Au\|^2}{2\sigma^2} - \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1\right) du \\ &= -E_{\sigma, \lambda}[\|F_{\mathbf{m}}u\|_1] \end{aligned} \quad (5.36)$$

$$\begin{aligned} \frac{\partial \log(Z_\lambda)}{\partial \lambda_{\mathbf{m}}} &= \frac{-1}{Z_\lambda} \int_u \|F_{\mathbf{m}}u\|_1 \exp\left(-\sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1\right) du \\ &= -E_\lambda[\|F_{\mathbf{m}}u\|_1] \end{aligned} \quad (5.37)$$

La première espérance, en (5.36), dépend de la loi *a posteriori* $p_\lambda(u|g)$ et la seconde espérance, en (5.37) dépend de la loi *a priori* $p_\lambda(u)$.

Pour mettre en œuvre une méthode de gradient, il faut savoir approcher numériquement les espérances (5.36) et (5.37). La difficulté vient du terme d'espérance selon la loi de probabilité *a posteriori* $p_\lambda(u|g)$ dont l'énergie est (5.31). Cette espérance n'est pas calculable analytiquement et il faut l'approcher par une moyenne empirique, calculée sur des échantillons tirés selon cette loi. Mais l'échantillonnage par un algorithme de Gibbs

ou de Metropolis-Hastings est difficile d'un point de vue pratique (coût de calcul élevé) car l'opérateur de flou A induit une interaction des pixels entre eux sur un voisinage très large [41]. Une manière d'éliminer cette difficulté est de passer dans le domaine de Fourier (ou transformée en Cosinus) afin de diagonaliser l'opérateur A . Le problème est que F ne se diagonalise pas par transformée de Fourier, mais devient au contraire un opérateur agissant sur tous les pixels. C'est alors le terme de régularisation (norme ℓ^1) qui pose problème pour l'échantillonnage. La difficulté vient de la présence simultanée des opérateurs A et F qui ne sont pas diagonalisables dans le même espace. Il faut donc dissocier les deux opérateurs dans l'énergie de la probabilité.

5.2.4 Introduction d'une variable intermédiaire

La dissociation des deux opérateurs peut être obtenue en introduisant une variable intermédiaire w . On considère alors un nouveau critère $J(u, w)$ donné par [17]

$$J(u, w) = \frac{1}{2\sigma^2\mu} \left(\|u - w\|^2 + \langle Cw, w \rangle \right) + \frac{1}{2\sigma^2} \left(\|g\|^2 - 2\langle Au, g \rangle \right) + \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1$$

où $C = B(I - B)^{-1}$ et $B = \mu A^*A$ (μ tel que $\mu \|A^*A\| < 1$). Il est équivalent de minimiser $J(u)$ en u et $J(u, w)$ en (u, w) [17]. L'opérateur linéaire peut ainsi se trouver appliqué à g et non à u puisque $\langle Au, g \rangle = \langle u, A^*g \rangle$.

Afin d'interpréter ce critère en terme de probabilités, nous le réécrivons de manière équivalente :

$$J(u, w) = \frac{1}{2\sigma^2\mu} \left((w - (I + C)^{-1}u)^T (I + C) (w - (I + C)^{-1}u) \right) + \frac{1}{2\sigma^2} \left(\|g - Au\|^2 \right) + \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1$$

puisque $(I + C)^{-1} = I - \mu A^*A$.

On remarque que les variables g et w sont indépendantes conditionnellement à u et

$$p(w|u) = \frac{1}{K_{\mu}} \exp \left(- \frac{(w - (I + C)^{-1}u)^T (I + C) (w - (I + C)^{-1}u)}{2\sigma^2\mu} \right) \quad (5.38)$$

avec $K_{\mu} = (2\pi\sigma^2\mu)^{k/2} (\det(I + C))^{-1/2}$, est une loi gaussienne $\mathcal{N}((I + C)^{-1}u, \sigma^2\mu(I + C)^{-1})$.

A présent, nous allons procéder de manière identique à la section précédente. Afin d'estimer les paramètres au sens du maximum de vraisemblance, il faut maximiser en λ la loi de probabilité

$$\begin{aligned} p_{\lambda}(g) &= \int_{u,w} p_{\lambda}(g, u, w) du dw \\ &= \int_{u,w} p(g|u) p(w|u) p_{\lambda}(u) du dw \end{aligned} \quad (5.39)$$

$$\begin{aligned} &= \frac{1}{K_{\sigma} K_{\mu} Z_{\lambda}} \int_{u,w} \exp \left(- \frac{1}{2\sigma^2\mu} \left(\|u - w\|^2 + \langle Cw, w \rangle \right) - \frac{1}{2\sigma^2} \left(\|g\|^2 - 2\langle u, A^*g \rangle \right) \right. \\ &\quad \left. - \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1 \right) du dw \end{aligned} \quad (5.40)$$

Calculons les dérivées partielles de la log-vraisemblance par rapport à chaque paramètre $\lambda_{\mathbf{m}}$ en invoquant de nouveau de théorème de convergence dominée de Lebesgue :

$$\frac{\partial \log p_{\lambda}(g)}{\partial \lambda_{\mathbf{m}}} = E_{\lambda}[\|F_{\mathbf{m}}u\|_1] - E_{\sigma, \lambda, \mu}[\|F_{\mathbf{m}}u\|_1].$$

La première espérance dépend de la loi *a priori* $p_{\lambda}(u)$ qui est une loi de Laplace (5.33) dans les sous-bandes et donc l'espérance se calcule facilement et vaut $1/\lambda_{\mathbf{m}}$. La seconde espérance dépend de la loi *a posteriori* $p_{\lambda}(u, w|g)$. Contrairement au cas précédent où l'opérateur A était appliqué à u , on se rend compte qu'il est cette fois-ci appliqué à g ce qui va nous permettre de pouvoir échantillonner plus facilement selon cette loi et comme décrit dans le paragraphe suivant.

5.2.5 Algorithme de gradient

Afin de calculer les paramètres $\lambda_{\mathbf{m}}$ au sens du maximum de vraisemblance, on utilise une méthode de gradient [40] :

$$\lambda_{\mathbf{m}}^{(n+1)} = \lambda_{\mathbf{m}}^{(n)} + \alpha_n \left[1/\lambda_{\mathbf{m}}^{(n)} - \frac{2}{I} \sum_{i=I/2+1}^I \|F_{\mathbf{m}}(u^{(n)})_{\sigma, \lambda^{(n)}, \mu}_i\|_1 \right] \quad (5.41)$$

où $(u^{(n)})_{\sigma, \lambda^{(n)}, \mu}_i$ est le i ème échantillon tiré selon la loi de probabilité *a posteriori* $p_{\lambda^{(n)}}(u, w|g)$. α_n est ici le pas de l'algorithme et il peut varier au cours des itérations.

L'échantillonnage devra donc se faire suivant la loi $p_{\lambda}(u, w|g)$ pour $\lambda = \lambda^{(n)}$. Pour cela, on tirera (échantillonneur de Gibbs) successivement $p(w|u)$ donnée par (5.38) (loi gaussienne) et $p_{\lambda}(u|w, g)$ où :

$$p_{\lambda}(u|w, g) \propto \exp \left(-\frac{1}{2\sigma^2\mu} \|u - w\|^2 + \frac{1}{\sigma^2} \langle u, A^*g \rangle - \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1 \right) \quad (5.42)$$

Concernant le tirage selon $p(w|u)$, on effectue le tirage dans le domaine de Fourier car la matrice de covariance de la loi se diagonalise facilement. Pour le tirage selon $p_{\lambda}(u|w, g)$, on utilise un algorithme de type Metropolis-Hastings directement dans le domaine ondelettes ce qui revient exprimer (5.42) comme

$$p_{\lambda}(u|w, g) \propto \exp \left(-\frac{1}{2\sigma^2\mu} \|Fu - Fw\|^2 + \frac{1}{\sigma^2} \langle Fu, FA^*g \rangle - \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \|F_{\mathbf{m}}u\|_1 \right)$$

Cette opération est effectuée I fois et une moyenne sur les $I/2$ derniers échantillons générés est prise en compte.

5.2.6 Accélération

Afin d'accélérer l'algorithme de gradient, il est possible de recourir à des choix de pas adaptatifs. Nous avons mis en œuvre une stratégie de choix de pas basée sur une technique de recherche linéaire (line search). La difficulté à laquelle nous sommes confrontés ici est le fait que la fonction objectif (5.39) ne peut pas être évaluée et que par conséquent, les recherches de pas optimales classiques ne peuvent pas être utilisées. Cette difficulté a pu être contournée en proposant une méthode de gradient en 2 phases (2Ph) pour laquelle 1) des étapes de gradients classiques sont effectuées afin de minimiser la

fonction coût et 2) une technique de recherche linéaire ne requérant pas le calcul de la fonction coût est introduite dans le processus itératif afin d'assurer une diminution de la norme du gradient. La première étape consiste à effectuer des étapes de la forme (5.41) dans lesquelles la valeur du pas est obtenue en alternant les valeurs obtenues par Barzilai-Borwein [42] :

$$\alpha_n^{BB1} = -\frac{s_n^T s_n}{s_n^T y_n}, \quad \alpha_n^{BB2} = -\frac{s_n^T y_n}{y_n^T y_n},$$

où $s_n = \lambda^{(n)} - \lambda^{(n-1)}$ et $y_n = \nabla p_{\lambda^{(n)}}(g) - \nabla p_{\lambda^{(n-1)}}(g)$. Quand à l'alternance adaptative, elle est dérivée de [43, 44] et décrite dans [45]. Dans la seconde phase de la stratégie de recherche linéaire, une approche similaire à celle proposée dans [46] est utilisée. Soient $G(\lambda) = \nabla p_{\lambda}(g)$ et $f(\lambda) = \frac{1}{2}\|G(\lambda)\|^2$, cette stratégie de recherche linéaire suffit à suffisamment réduire $f(\lambda)$ en effectuant une étape de gradient ou en bougeant dans la direction du gradient pour $f(\lambda)$ défini par $\bar{q}_n(\alpha_n) = -(G(\lambda^{(n)} + \alpha_n G(\lambda^{(n)})) - G(\lambda^{(n)}))$ (pour des choix de α_n suffisamment petits). Grâce à cette propriété de réduction, il est prouvé dans [46] qu'il existe un point limite λ^* de $\{\lambda^{(n)}\}$ tel que $G(\lambda^*) = 0$, et que tous les points limite de $\{\lambda^{(n)}\}$ résolvent $G(\lambda) = 0$. Par conséquent, la seconde phase peut être utilisée pour stabiliser notre processus itératif en forçant l'approximation d'un point stationnaire de $p_{\lambda}(g)$.

La commutation entre les deux phases s'effectue dans le but d'activer la recherche de pas linéaire lorsque les dernières itérations se trouvent dans une région dans laquelle $p_{\lambda}(g)$ est concave. Rappelons que lorsque $s_n^T y_n > 0$ on peut conclure que la fonction $p_{\lambda}(g)$ n'est pas concave dans un ensemble contenant $\lambda^{(n)}, \lambda^{(n-1)}$, on active avec la deuxième phase lorsqu'une suite de N_2 itérations consécutives fournissent toujours des pas BB positifs. De plus, le passage à la deuxième phase est assuré après un nombre préfixé $N_1 > N_2$ d'itérations. L'algorithme ainsi proposé est décrit dans l'Algorithme 5.

5.2.7 Simulations numériques

Afin de tester notre approche, nous avons tout d'abord travaillé sur des données purement synthétiques. En effet, nous avons généré des coefficients d'ondelettes aléatoirement suivant des lois a priori (5.33) en fixant des paramètres par sous-bande réalistes λ_m . Nous considérons des ondelettes de [47] de longueur 8 sur $J = 2$ niveaux de résolution et chaque sous-bande m est représentée par un triplet (j, l, c) où j est l'indice de niveau de résolution, et $(l, c)_{l \in \{0,1\}, c \in \{0,1\}}$ représente les sous-bandes filtrées passe-bas/haut (le couple $(1, 1)$ représente les coefficients diagonaux). La taille de l'image générée est 128×128 , et elle est dégradée par un opérateur de flou A d'écart-type 0.5, et par un bruit additif Gaussien de variance $\sigma^2 = 25$.

La descente de gradient classique est faite sur 400 itérations en utilisant un pas fixe $\alpha_n = 10^{-4}$ et les paramètres estimés sont calculés comme étant la valeur moyenne des 50 dernières itérations (à cause des oscillations (voir Fig. 5.4)). La méthode à 2 phases utilise les paramètres suivant : $P = 3, \theta = 0.5, \gamma = 10^{-4}, \alpha_0 = 10^{-4}$; $N_1 = 50, N_2 = 10$ pour la règle de commutation et $\tau_g = 10^{-2}, \tau_{\lambda} = 10^{-7}$ pour la règle d'arrêt. Les deux algorithmes sont initialisés par l'observation g filtrée Wiener. Les Fig. 5.3 et Fig. 5.4 illustrent le comportement de deux hyperparamètres estimés pour les 2 approches. Par ailleurs, dans le tableau 5.2 nous reportons le nombre d'itérations effectué (it.), le nombre d'évaluations de gradients (Grad.), les erreurs relatives (norme euclidienne) entre les estimations et les valeurs théoriques ($E_{err_{th.}}$) et entre les estimations et les valeurs obtenues

Algorithm 5 Méthode de gradient à 2 phases (2Ph)

Initialisation : choisir $\lambda^{(0)}$, α_0 et $\theta, \gamma \in (0, 1)$; soient $\text{flag}(N_1, N_2) = 0$, $n = 1$, $\lambda^{(1)} = \lambda^{(0)} + \alpha_0 G(\lambda^{(0)})$, $gr_0 = f(\lambda^{(0)})$, $gr = f(\lambda^{(1)})$ et un entier $P \geq 1$.

Phase 1 : (BB-like étape de gradient)

TANT QUE $\left(\frac{gr}{gr_0} > \tau_g \text{ ou } \frac{\|\lambda^{(n)} - \lambda^{(n-1)}\|}{\|\lambda^{(n)}\|} > \tau_\lambda \right)$ et $\text{flag}(N_1, N_2) = 0$

1.1 Choisir α_n et mettre à jour $\text{flag}(N_1, N_2)$;

1.2 Etape de gradient : $\lambda^{(n+1)} = \lambda^{(n)} + \alpha_n G(\lambda^{(n)})$, $n = n + 1$;

1.3 Calculer $gr = \max_{0 \leq j \leq \min(n, P-1)} f(\lambda^{(n-j)})$;

FIN TANT QUE

Phase 2 : (Stabilisation avec recherche linéaire)

TANT QUE $\left(\frac{gr}{gr_0} > \tau_g \text{ ou } \frac{\|\lambda^{(n)} - \lambda^{(n-1)}\|}{\|\lambda^{(n)}\|} > \tau_\lambda \right)$

2.1 Choisir α_n ;

2.2 Recherche linéaire :

SI $f(\lambda^{(n)} + \alpha_n G(\lambda^{(n)})) \leq gr - \gamma \|\alpha_n^2 G(\lambda^{(n)})\|^2$ ALORS
prendre $\lambda^{(n+1)} = \lambda^{(n)} + \alpha_n G(\lambda^{(n)})$ et $n = n + 1$;

SINON SI $f(\lambda^{(n)} - \alpha_n \bar{q}_n(\alpha_n)) \leq gr - \gamma \|\alpha_n^2 G(\lambda^{(n)})\|^2$ ALORS
prendre $\lambda^{(n+1)} = \lambda^{(n)} - \alpha_n \bar{q}_n(\alpha_n)$ et $n = n + 1$;

SINON prendre $\alpha_n = \theta \alpha_n$ et aller à l'étape 2.2;

FIN SI

2.3 Calculer $gr = \max_{0 \leq j \leq \min(n, P-1)} f(\lambda^{(n-j)})$;

FIN TANT QUE

par maximum de vraisemblance sur l'image originale (Err_{ML}) et enfin, le temps de calcul (temps) en seconde. Nous avons également considéré deux images classiques (Mandrill et Barbara) que nous avons dégradées avec un flou Gaussien A d'écart-type 2, et un bruit additif Gaussien de variance $\sigma^2 = 25$. Nous utilisons les méthodes proposées pour déterminer les hyperparamètres et restaurons les images en appliquant un algorithme explicite-implicite [5]. Le tableau 5.3 reporte les valeurs de rapport signal à bruit (SNR) calculés par rapport à l'image dégradée (SNR_{init}) et par rapport à l'image estimée (SNR_{fin}).

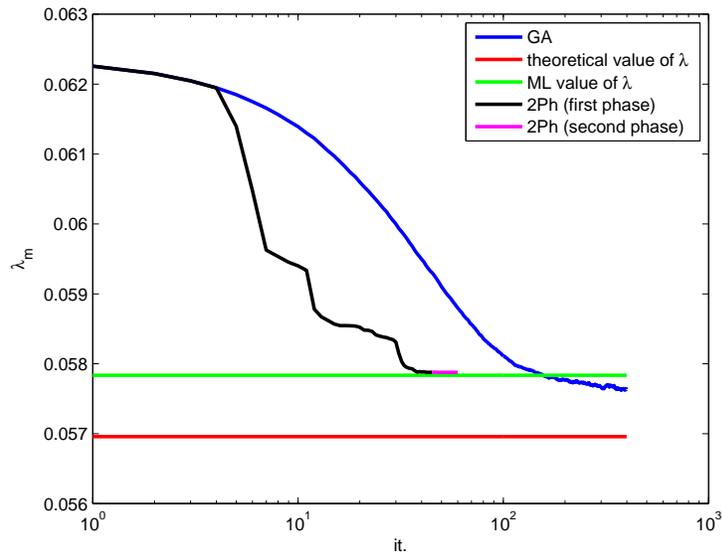
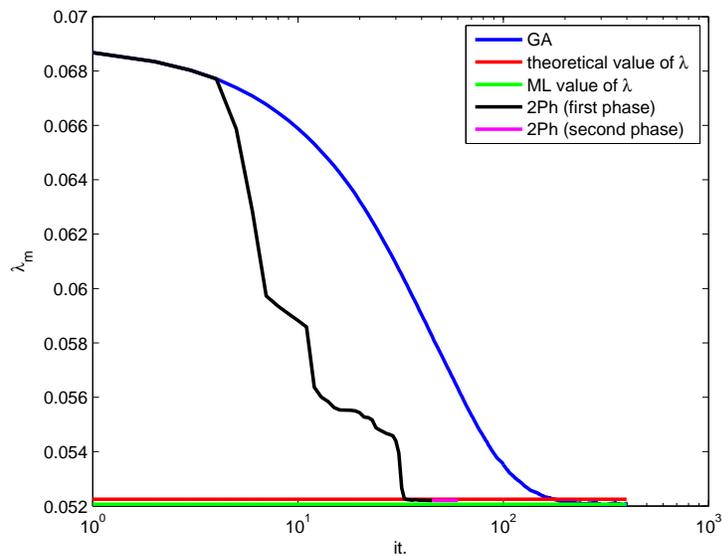
FIGURE 5.3 – Valeur de λ_m ($\mathbf{m} = (1, 1, 0)$) au cours des itérationsFIGURE 5.4 – Valeur de λ_m ($\mathbf{m} = (1, 1, 1)$) au cours des itérations

TABLE 5.2 – Données simulées

Alg.	It.	Grad.	Err_{ML}	$\text{Err}_{th.}$	temps
2Ph	61	90	0.007	0.021	431.8
GA	400	400	0.007	0.020	1920.5

TABLE 5.3 – Données réelles

Problème	Alg.	It.	SNR _{init}	SNR _{fin}
Mandrill 256 ²	2Ph	883	12.1616	14.2324
	GA	1000	12.1616	14.0051
Barbara 512 ²	2Ph	785	18.5272	19.5316
	GA	1000	18.5272	19.3655

Les résultats obtenus avec l'algorithme proposé sont proches de ceux obtenus par maximum de vraisemblance sur l'image originale ce qui est vraiment encourageant. Par ailleurs, l'introduction d'un algorithme à 2 phases permet d'accélérer la méthode tout en conservant sa précision.

5.2.8 Conclusion

Nous avons vu ici, à modèle de dégradation connu, une manière d'estimer les paramètres de régularisation lorsque la régularisation concernée est exprimée dans le domaine ondelettes. La méthode proposée basée sur une descente de gradient a été accélérée grâce à l'introduction d'une recherche de pas optimal.

Le domaine ondelettes considéré ici est celui engendré par une base orthogonale. Or, il serait intéressant d'étendre cette approche au cadre plus général des trames. Le point clé pour cela est de voir s'il est possible de mettre en place l'algorithme de Metropolis-Hastings associé, qui travaille dans le domaine transformé et donc implique de générer directement des coefficients de trame.

5.3 Estimation des paramètres du modèle d'acquisition

Cette dernière section me permet également de faire le lien avec le chapitre suivant car elle mêle à la fois un problème d'estimation de modèle et un problème d'optimisation complexe à résoudre.

Nous nous intéressons dans cette dernière partie à la suppression d'échos dans des signaux sismiques, dans un premier temps à une dimension, puis, dans un second temps, à deux dimensions. Ces travaux ont été effectués dans le cadre de la thèse de Mai Quyen Pham.

5.3.1 Présentation du problème

Nous considérons le problème de filtrage adaptatif de données sismiques composées : de signaux d'intérêts ("primaires"), de perturbations structurées correspondant à des propagations d'ondes présentant des réflexions multiples ("multiples") et de bruit aléatoire. Plus précisément, soit $n \in \{0, \dots, N-1\}$ l'indice temporel d'une trace sismique observée z , enregistrée par un capteur donné. On fait l'hypothèse d'additivité des contributions soit :

$$z^{(n)} = \bar{y}^{(n)} + \bar{s}^{(n)} + b^{(n)}. \quad (5.43)$$

Le signal d'intérêt (inconnu) (primaire, en bleu) et la somme des signaux réfléchis indésirables (multiples, en rouge) sont désignés, respectivement, par $\bar{y} = (\bar{y}^{(n)})_{0 \leq n < N}$ et $\bar{s} = (\bar{s}^{(n)})_{0 \leq n < N}$. D'autres contributions non structurées sont regroupées sous un terme de bruit $b = (b^{(n)})_{0 \leq n < N}$. On suppose ici que plusieurs approximations de templates (=

modèles) pour modéliser les multiples sont disponibles. Concrètement, un ou plusieurs modèles réalistes d'un multiple sont déterminés hors ligne ; on supposera ici qu'ils ont été déterminés par un pré-traitement sismique ou par une modélisation.

Le signal multiple possède un comportement local lié au contexte géologique. On suppose alors disposer de J templates $(r_j^{(n)})_{0 \leq n < N, 0 \leq j < J}$, reliés à $(\bar{s}^{(n)})_{0 \leq n < N}$ via un modèle linéaire éventuellement non causal par la relation suivante :

$$\bar{s}^{(n)} = \sum_{j=0}^{J-1} \sum_{p=p'}^{p'+P_j-1} \bar{h}_j^{(n)}(p) r_j^{(n-p)} \quad (5.44)$$

où $\bar{h}_j^{(n)}$ représente une réponse impulsionnelle finie (de longueur P_j) associée au template j et au temps n , et où $p' \in \{-P_j + 1, \dots, 0\}$ est son indice de départ ($p' = 0$ correspond donc au cas causal). On peut souligner ici la dépendance des réponses impulsionnelles au temps n impliquant que les réponses impulsionnelles varient au cours du temps et que donc le processus est variant dans le temps, même si cette variation est lente en pratique. En effet, les ondes sismiques évoluent progressivement avec la profondeur de propagation, contrairement aux variations plus prononcées autour des contours des images naturelles. Les modèles sont générés avec une modélisation géophysique standard basée sur les primaires décrits ci dessus.

Soient \bar{s} et $(\bar{h}_j)_{0 \leq j < J}$ deux vecteurs définis par :

$$\bar{s} = [\bar{s}^{(0)} \quad \dots \quad \bar{s}^{(N-1)}]^\top,$$

$$\bar{h}_j = [\bar{h}_j^{(0)}(p') \quad \dots \quad \bar{h}_j^{(0)}(p' + P_j - 1) \quad \dots \quad \bar{h}_j^{(N-1)}(p') \quad \dots \quad \bar{h}_j^{(N-1)}(p' + P_j - 1)]^\top,$$

et soient $(R_j)_{0 \leq j < J}$ des matrices bloc diagonales de taille $N \times NP_j$ définies par :

$$R_j = \begin{bmatrix} R_j^{(0)} & 0 & \dots & 0 \\ 0 & R_j^{(1)} & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & R_j^{(N-1)} \end{bmatrix},$$

où $(R_j^{(n)})_{0 \leq n \leq N-1}$ sont des vecteurs de dimensions P_j tels que

$$\left[(R_j^{(0)})^\top (R_j^{(1)})^\top \dots (R_j^{(N-1)})^\top \right]^\top = \begin{bmatrix} r_j^{(-p')} & \dots & r_j^{(0)} & 0 & \dots & 0 \\ r_j^{(-p'+1)} & \dots & & r_j^{(0)} & 0 & \dots & 0 \\ \vdots & & & & & & \\ r_j^{(N-1)} & r_j^{(N-2)} & \dots & & & & r_j^{(N-P_j)} \\ 0 & r_j^{(N-1)} & \dots & & & & r_j^{(N-P_j+1)} \\ \vdots & & & & & & \\ 0 & \dots & 0 & r_j^{(N-1)} & \dots & & r_j^{(N-P_j-p')} \end{bmatrix}.$$

L' eq. (5.44) peut être écrite de manière plus concise comme

$$\bar{s} = \sum_{j=0}^{J-1} R_j \bar{h}_j.$$

Pour plus de concision encore, on peut écrire $\bar{s} = \mathbf{R}\bar{\mathbf{h}}$ en définissant $\mathbf{R} = [R_0 \dots R_{J-1}] \in \mathbb{R}^{N \times Q}$ où $Q = NP$ avec $P = \sum_{j=0}^{J-1} P_j$ et $\bar{\mathbf{h}} = [\bar{h}_0^\top \dots \bar{h}_{J-1}^\top]^\top \in \mathbb{R}^Q$.

5.3.2 Formulation variationnelle du problème

5.3.2-a) Contexte bayésien

Nous avons choisi ici de ne pas modéliser directement les primaires mais les coefficients de représentation des primaires dans une trame (redondante ou non), par exemple une trame d'ondelettes [48].

Soit \bar{x} le vecteur de coefficients de trame, et $F \in \mathbb{R}^{K \times N}$ l'opérateur d'analyse de trame associé. On a alors [5] : $\bar{x} = F\bar{y}$. De plus, on suppose que \bar{y} est une réalisation d'un vecteur aléatoire Y , la fonction densité de probabilité duquel est donnée par

$$(\forall y \in \mathbb{R}^N) \quad f_Y(y) \propto \exp(-\varphi(Fy)) \quad (5.45)$$

où $\varphi: \mathbb{R}^N \rightarrow]-\infty, +\infty]$ est le potentiel associé, supposé décroître suffisamment rapidement.

D'un autre côté, afin de prendre en compte les informations disponibles sur les filtres inconnus, on suppose que pour tout $j \in \{0, \dots, J-1\}$, \bar{h}_j est une réalisation d'un vecteur aléatoire H_j . Soit $\mathbb{H} = \mathbb{R}^{NP_1} \times \dots \times \mathbb{R}^{NP_J}$. La fonction densité de probabilité jointe des coefficients de filtre peut s'écrire :

$$(\forall h \in \mathbb{H}) \quad f_{H_0, \dots, H_{J-1}}(h) \propto \exp(-\rho(h)),$$

où (H_0, \dots, H_{J-1}) est indépendant de Y . On suppose de plus que le vecteur de bruit b est la réalisation d'un vecteur aléatoire B de fonction densité de probabilité

$$(\forall b \in \mathbb{R}^N) \quad f_B(b) \propto \exp(-\psi(b)),$$

où $\psi: \mathbb{R}^N \rightarrow]-\infty, +\infty]$, et que B est indépendant de Y et H_0, \dots, H_{J-1} . La distribution a posteriori de (Y, H_0, \dots, H_{J-1}) conditionnellement à $Z = Y + \sum_{j=0}^{J-1} R_j H_j + B$ est donnée par $(\forall y \in \mathbb{R}^N)(\forall h \in \mathbb{H})$

$$f_{Y, H_0, \dots, H_{J-1} | Z=z}(y, h) \propto \exp\left(-\psi\left(z - y - \sum_{j=0}^{J-1} R_j h_j\right)\right) f_Y(y) f_{H_0, \dots, H_{J-1}}(h).$$

Une estimation au sens du maximum a posteriori (MAP) de $(\bar{y}, \bar{h}_0, \dots, \bar{h}_{J-1})$ mène au problème de minimisation suivant :

$$\underset{y \in \mathbb{R}^N, h \in \mathbb{H}}{\text{minimiser}} \psi\left(z - y - \sum_{j=0}^{J-1} R_j h_j\right) + \varphi(Fy) + \rho(h).$$

5.3.2-b) Formulation du problème

Pour plus de simplicité, nous proposons de choisir des a priori uniformes pour Y et (H_0, \dots, H_{J-1}) en choisissant pour φ et ρ des fonctions indicatrices d'ensembles convexes fermés. L'estimation MAP associée se transforme alors en un problème d'optimisation sous contraintes :

$$\underset{Fy \in D, \mathbf{h} \in C}{\text{minimiser}} \Psi(y, \mathbf{h}) \quad (5.46)$$

où le terme de fidélité est défini par $\Psi: (y, \mathbf{h}) \mapsto \psi(z - y - \mathbf{R}\mathbf{h})$, et les a priori formulés sur les filtres et les primaires sont exprimés en termes de contraintes dures modélisés par des ensembles convexes fermés C et D . Un des avantages de formuler le problème sous contrainte et non sous forme régularisé [49, 50] est d'éviter l'introduction de paramètres de régularisations parfois difficiles à déterminer de manière optimale.

Nous allons à présent discuter du choix de la fonction Ψ , et des ensembles C et D .

5.3.2-c) Terme de fidélité et contraintes

5.3.2-c)-1 Terme de fidélité La fonction Ψ modélise la statistique du bruit. Dans ce travail, le bruit est supposé additif, centré, blanc et Gaussien. Par conséquent, une forme quadratique est adoptée $\psi = \|\cdot\|^2$.

5.3.2-c)-2 Information a priori sur \mathbf{h} Les filtres sont supposés variants dans le temps. Cependant, afin d'assurer des variations lentes au cours du temps, nous proposons d'introduire les ensembles de contraintes suivants

$$C_1 = \left\{ \mathbf{h} \in \mathbb{R}^Q \mid \forall (j, p), \forall n \in \left\{ 0, \dots, \left\lfloor \frac{N}{2} \right\rfloor - 1 \right\}, |h_j^{(2n+1)}(p) - h_j^{(2n)}(p)| \leq \varepsilon_{j,p} \right\} \quad (5.47)$$

$$C_2 = \left\{ \mathbf{h} \in \mathbb{R}^Q \mid \forall (j, p), \forall n \in \left\{ 1, \dots, \left\lfloor \frac{N-1}{2} \right\rfloor \right\}, |h_j^{(2n)}(p) - h_j^{(2n-1)}(p)| \leq \varepsilon_{j,p} \right\}. \quad (5.48)$$

Ces contraintes permettent d'éviter des variations trop brutales des coefficients de la réponse impulsionnelle pris à deux temps consécutifs.

Par ailleurs, des informations a priori sur les coefficients du filtre \mathbf{h} peuvent être ajoutées par le biais d'un nouvel ensemble convexe C_3 défini par $C_3 = \{ \mathbf{h} \in \mathbb{R}^Q \mid \tilde{\rho}(\mathbf{h}) \leq \lambda \}$ où $\lambda \in]0, +\infty[$. $\tilde{\rho}: \mathbb{R}^Q \rightarrow [0, +\infty[$ peut-être choisie comme une simple norme telle que les normes ℓ_1 ou ℓ_2 ou comme une norme plus sophistiquée comme les normes mixtes $\ell_{1,2}$ [51]. Par conséquent, l'ensemble convexe C est défini par $C = C_1 \cap C_2 \cap C_3$.

5.3.2-c)-3 Information a priori sur y Comme mentionné précédemment dans le paragraphe 5.3.2-a), on suppose que le signal primaire est parcimonieux au travers d'un opérateur d'analyse de trame $F \in \mathbb{R}^{K \times N}$ [52], ce qui peut faciliter son traitement, en augmentant l'écart de domaine de données entre les primaires et les multiples. La contrainte associée peut-être éclatée en définissant une partition de $\{1, \dots, K\}$ que l'on notera $\{\mathbb{K}_\ell \mid \ell \in \{1, \dots, \mathcal{L}\}\}$. Par exemple, pour des trames d'ondelettes, \mathcal{L} peut correspondre au nombre de sous-bandes et ainsi \mathbb{K}_ℓ correspond à la ℓ -ème sous-bande. Par conséquent, on peut prendre $D = D_1 \times \dots \times D_{\mathcal{L}}$ avec $D_\ell = \{(x_k)_{k \in \mathbb{K}_\ell} \mid \sum_{k \in \mathbb{K}_\ell} \tilde{\varphi}_\ell(x_k) \leq \beta_\ell\}$, où, pour tout $\ell \in \{1, \dots, \mathcal{L}\}$, $\beta_\ell \in]0, +\infty[$, et $\tilde{\varphi}_\ell: \mathbb{R}^{\text{card}(\mathbb{K}_\ell)} \rightarrow [0, +\infty[$ est une fonction convexe semi-continue inférieurement.

5.3.3 Algorithme proximal primal-dual

L'objectif est maintenant le problème (5.46) ce qui revient à minimiser la fonction Ψ par rapport à y et \mathbf{h} sous les contraintes définies par les ensembles D et C . Les algorithmes primaux-duaux [53, 27, 28], comme l'algorithme M+L FBF [26], sont bien adaptés à notre problème car ils ne requièrent pas d'inversion de matrice de grande taille. Par

ailleurs Ψ est une fonction quadratique (voir le paragraphe 5.3.2-c)-1) et de gradient Lipschitz, propriétés qui seront directement exploitées par l'algorithme M+L FBF. Quand aux contraintes, des projections sur les ensembles fermés $(C_m)_{1 \leq m \leq 3}$ et D sont effectuées. Quelques détails sont donnés dans le paragraphe suivant.

5.3.3-a) Calcul du gradient et des projections

Etant donné que le bruit est supposé centré, on en déduit que Ψ est différentiable de gradient μ -Lipschitz, i.e. $(\forall \begin{bmatrix} y \\ \mathbf{h} \end{bmatrix} \in \mathbb{R}^{N+Q})(\forall \begin{bmatrix} y' \\ \mathbf{h}' \end{bmatrix} \in \mathbb{R}^{N+Q})$:

$$\|\nabla\Psi\left(\begin{bmatrix} y \\ \mathbf{h} \end{bmatrix}\right) - \nabla\Psi\left(\begin{bmatrix} y' \\ \mathbf{h}' \end{bmatrix}\right)\| \leq \mu\left\|\begin{bmatrix} y \\ \mathbf{h} \end{bmatrix} - \begin{bmatrix} y' \\ \mathbf{h}' \end{bmatrix}\right\|$$

et

$$\nabla\Psi = 2[\mathcal{I} \mathbf{R}]^\top([\mathcal{I} \mathbf{R}] \cdot -z). \quad (5.49)$$

Le gradient de Ψ est donc μ -Lipschitz avec

$$\mu = 2\|\|\mathcal{I} \mathbf{R}\|\|^2 \quad (5.50)$$

et où $\|\cdot\|$ représente la norme spectrale.

Regardons à présent ce qu'il se passe pour les espaces de contraintes C et D . C modélise les contraintes imposés aux filtres, qui sont éclatées en 3 termes (voir paragraphe 5.3.2-c)). Il faut donc projeter sur chaque ensemble C_m avec $m \in \{1, 2, 3\}$. Les projections sur C_1 et C_2 — qui imposent, je le rappelle, une variation lente au court du temps de la réponse impulsionnelle — sont équivalentes à des projections sur des hyperplans de \mathbb{R}^2 comme l'illustre la figure 5.5.

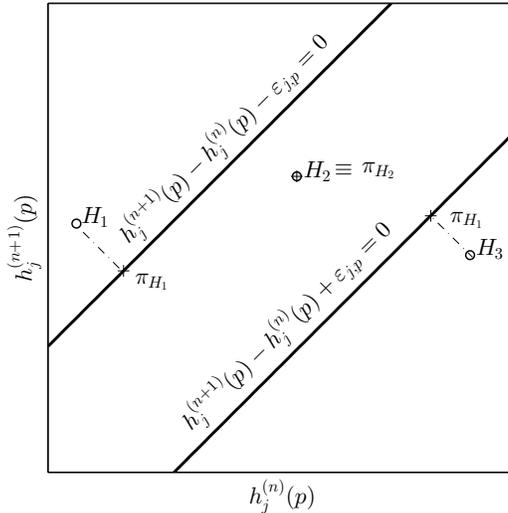


FIGURE 5.5 – Projection sur C_1/C_2 des points H_1, H_2 et H_3 de \mathbb{R}^2 .

Plus précisément, la projection sur C_1 (la projection sur C_2 menant à des expressions similaires) est calculée de la manière suivante : soit $\mathbf{h} \in \mathbb{R}^Q$ et soit $\mathbf{g}_1 = \Pi_{C_1}(\mathbf{h})$; alors pour tout $j \in \{0, \dots, J-1\}$, $p \in \{p', \dots, p'+P-1\}$ et $n \in \{0, \dots, \lfloor \frac{N}{2} \rfloor - 1\}$,

1. si $|h_j^{(2n+1)}(p) - h_j^{(2n)}(p)| < \varepsilon_{j,p}$, alors

$$g_{j,1}^{(2n)}(p) = h_j^{(2n)}(p), \quad g_{j,1}^{(2n+1)}(p) = h_j^{(2n+1)}(p);$$

2. si $h_j^{(2n+1)}(p) - h_j^{(2n)}(p) > \varepsilon_{j,p}$, alors

$$g_{j,1}^{(2n)}(p) = \frac{h_j^{(2n+1)}(p) + h_j^{(2n)}(p)}{2} - \frac{\varepsilon_{j,p}}{2}$$

$$g_{j,1}^{(2n+1)}(p) = \frac{h_j^{(2n+1)}(p) + h_j^{(2n)}(p)}{2} + \frac{\varepsilon_{j,p}}{2};$$

3. si $h_j^{(2n+1)}(p) - h_j^{(2n)}(p) < -\varepsilon_{j,p}$, alors

$$g_{j,1}^{(2n)}(p) = \frac{h_j^{(2n+1)}(p) + h_j^{(2n)}(p)}{2} + \frac{\varepsilon_{j,p}}{2}$$

$$g_{j,1}^{(2n+1)}(p) = \frac{h_j^{(2n+1)}(p) + h_j^{(2n)}(p)}{2} - \frac{\varepsilon_{j,p}}{2}.$$

C_3 introduit de l'information a priori sur le vecteur de filtres \mathbf{h} via la fonction convexe semi-continue inférieurement $\tilde{\rho}$. Cette fonction peut être choisie séparable par rapport à $j \in \{0, \dots, J-1\}$ dans le sens où

$$\tilde{\rho}(\mathbf{h}) = \sum_{j=0}^{J-1} \tilde{\rho}_j(h_j).$$

Nous avons considéré 3 choix possibles pour $\tilde{\rho}_j$:

1. norme ℓ_1 :

$$\tilde{\rho}_j(h_j) = \|h_j\|_{\ell_1} = \sum_{n=0}^{N-1} \sum_{p=p'}^{P_j-1} |h_j^{(n)}(p)|.$$

Ce choix requiert d'effectuer des projections sur la boule ℓ_1 , ce qui peut être fait itérativement en utilisant la procédure décrite dans [54] (la projection est obtenue en un nombre fini d'itérations).

2. norme ℓ_2 au carré :

$$\tilde{\rho}_j(h_j) = \|h_j\|_{\ell_2}^2 = \sum_{n=0}^{N-1} \sum_{p=p'}^{P_j-1} |h_j^{(n)}(p)|^2.$$

Dans ce cas, la projection est immédiate.

3. norme mixte $\ell_{1,2}$:

$$\tilde{\rho}_j(h_j) = \|h_j\|_{\ell_{1,2}} = \sum_{n=0}^{N-1} \left(\sum_{p=p'}^{P_j-1} |h_j^{(n)}(p)|^2 \right)^{1/2}.$$

Un algorithme similaire à celui utilisé pour effectuer la projection sur la boule ℓ_1 peut être facilement mis en oeuvre.

Enfin, comme mentionné précédemment, l'information a priori sur le primaire y s'exprime en fonction d'un opérateur d'analyse de trame F et la contrainte est éclatée en un ensemble de contraintes portant sur chacune des sous-bandes. Dans le but de promouvoir la parcimonie de la représentation, la fonction employée pour la ℓ ème sous-bande, $\ell \in \{1, \dots, \mathcal{L}\}$, peut être choisie comme $\tilde{\varphi}_\ell = |\cdot|$. La projection résultante peut également être effectuée avec la procédure itérative décrite dans [54].

5.3.3-b) Algorithme proximal M+LFBF

L'approche primal-duale choisie pour résoudre le problème de minimisation (5.46) est détaillée dans l'algorithme 6. Elle alterne entre le calcul du gradient de Ψ , et les projections sur les $(C_m)_{1 \leq m \leq 3}$ et $(D_\ell)_{1 \leq \ell \leq \mathcal{L}}$.

Le choix du pas est crucial pour assurer la convergence de l'algorithme et pour garantir une convergence rapide. Pour le choisir judicieusement, nous avons besoin

- de connaître la norme de chaque opérateur linéaire impliqué (ou au moins une borne supérieure). Dans notre cas, nous avons

$$\|[\mathcal{I} \mathbf{R}]\| \leq \sqrt{1 + \|R_0\|^2 + \dots + \|R_{J-1}\|^2}$$

où $\|R_j\| = \max_{n \in \{0, \dots, N-1\}} \|R_j^{(n)}\|$ pour tout $j \in \{0, \dots, J-1\}$.

- que $\gamma^{[i]}$ à chaque itération i soit choisi de telle sorte que : soit μ la constante de Lipschitz définie dans (5.50), soit $\beta = \mu + \sqrt{\|F\|^2 + 3}$ et soit $\epsilon \in]0, \frac{1}{\beta+1}[$, alors $\gamma^{[i]} \in [\epsilon, \frac{1-\epsilon}{\beta}]$. $\|F\|^2$ peut être calculée facilement pour les trames ajustées. Sinon, on peut recourir à une approche itérative pour l'estimer [55, Algorithme 4].

La convergence de cet algorithme vers une solution optimale du problème (5.46) est garantie grâce à [26, Théorème 4.2].

5.3.4 Extension aux images sismiques

L'approche que nous avons décrite précédemment est adaptée aux signaux monodimensionnels. Elle peut cependant être appliquée aux signaux sismiques bidimensionnels dans la mesure où ces derniers consistent en une concaténation de plusieurs traces sismiques. Cependant, en ne les considérant pas comme tels mais plutôt comme une image classique à deux dimensions, il est possible de proposer une variante de l'approche précédente, adaptée aux données sismiques anisotropes 2D. Je vais décrire ci-après les extensions que nous avons proposées.

5.3.4-a) Formulation du problème à deux dimensions

Tout d'abord, le modèle change et devient

$$z^{(\mathbf{n})} = \bar{s}^{(\mathbf{n})} + \bar{y}^{(\mathbf{n})} + b^{(\mathbf{n})} \quad (5.51)$$

où les données sont maintenant double indicées $\mathbf{n} = (n_t, n_x)$, $n_t \in \mathbb{N}_t \triangleq \{0, \dots, N_t - 1\}$ est l'indice temps, $n_x \in \mathbb{N}_x \triangleq \{0, \dots, N_x - 1\}$ est l'indice capteur et $\mathcal{N} \triangleq \{(n_t, n_x) \mid n_t \in \mathbb{N}_t, n_x \in \mathbb{N}_x\}$. Comme dans le cas monodimensionnel, les données observées $z = (z^{(\mathbf{n})})_{\mathbf{n} \in \mathcal{N}}$ sont composées des primaires $\bar{y} = (\bar{y}^{(\mathbf{n})})_{\mathbf{n} \in \mathcal{N}}$, des multiples $(\bar{s}^{(\mathbf{n})})_{\mathbf{n} \in \mathcal{N}}$

Algorithm 6 Algorithme proximal primal-dual M+LFBF pour résoudre (5.46)

Let $\gamma^{[i]} \in [\epsilon, \frac{1-\epsilon}{\beta}]$

Let $\begin{bmatrix} y^{[0]} \\ \mathbf{h}^{[0]} \end{bmatrix} \in \mathbb{R}^{N+Q}$, $v^{[0]} \in \mathbb{R}^K$, $(u_m^{[0]})_{m \in \{1,2,3\}} \in (\mathbb{R}^Q)^3$, $s_2^{[0]} \in \mathbb{R}^K$, $w_1^{[0]} \in \mathbb{R}^K$

for $i = 0, 1, \dots$ **do**

Calcul du gradient

$$\begin{bmatrix} s_1^{[i]} \\ t_1^{[i]} \end{bmatrix} = \begin{bmatrix} y^{[i]} \\ \mathbf{h}^{[i]} \end{bmatrix} - \gamma^{[i]} \left(\nabla \Psi \left(\begin{bmatrix} y^{[i]} \\ \mathbf{h}^{[i]} \end{bmatrix} \right) + \begin{bmatrix} F^* v^{[i]} \\ \sum_{m=1}^3 u_m^{[i]} \end{bmatrix} \right)$$

Calcul des projections

$$x_1^{[i]} = F y^{[i]}$$

for $\ell = 1 : \mathcal{L}$ **do**

$$\left(s_2^{[i]}(k) \right)_{k \in \mathbb{K}_\ell} = \left(v^{[i]}(k) + \gamma^{[i]} x_1^{[i]}(k) \right)_{k \in \mathbb{K}_\ell}$$

$$\left(w_1^{[i]}(k) \right)_{k \in \mathbb{K}_\ell} = \left(s_2^{[i]}(k) \right)_{k \in \mathbb{K}_\ell} - \gamma^{[i]} \Pi_{D_\ell} \left(\frac{\left(s_2^{[i]}(k) \right)_{k \in \mathbb{K}_\ell}}{\gamma^{[i]}} \right)$$

end for

for $m = 1 : 3$ **do**

$$t_{2,m}^{[i]} = u_m^{[i]} + \gamma^{[i]} \mathbf{h}^{[i]}$$

$$w_{2,m}^{[i]} = t_{2,m}^{[i]} - \gamma^{[i]} \Pi_{C_m} \left(\frac{t_{2,m}^{[i]}}{\gamma^{[i]}} \right)$$

end for

Moyennage

$$x_2^{[i]} = F s_1^{[i]}$$

for $\ell = 1 : \mathcal{L}$ **do**

$$\left(q_1^{[i]}(k) \right)_{k \in \mathbb{K}_\ell} = \left(w_1^{[i]}(k) + \gamma^{[i]} x_2^{[i]}(k) \right)_{k \in \mathbb{K}_\ell}$$

$$\left(v^{[i+1]}(k) \right)_{k \in \mathbb{K}_\ell} = \left(v^{[i]}(k) - s_2^{[i]}(k) + q_1^{[i]}(k) \right)_{k \in \mathbb{K}_\ell}$$

end for

for $m = 1 : 3$ **do**

$$q_{2,m}^{[i]} = w_{2,m}^{[i]} + \gamma^{[i]} t_1^{[i]}$$

$$u_m^{[i+1]} = u_m^{[i]} - t_{2,m}^{[i]} + q_{2,m}^{[i]}$$

end for

Mise-à-jour

$$\begin{bmatrix} y^{[i+1]} \\ \mathbf{h}^{[i+1]} \end{bmatrix} = \begin{bmatrix} y^{[i]} \\ \mathbf{h}^{[i]} \end{bmatrix} - \gamma^{[i]} \left(\nabla \Psi \left(\begin{bmatrix} s_1^{[i]} \\ t_1^{[i]} \end{bmatrix} \right) + \begin{bmatrix} F^* w_1^{[i]} \\ \sum_{m=1}^3 w_{2,m}^{[i]} \end{bmatrix} \right)$$

end for

et d'un bruit additif $(b^{(\mathbf{n})})_{\mathbf{n} \in \mathcal{N}}$. Les multiples $(\bar{s}^{(\mathbf{n})})_{\mathbf{n} \in \mathcal{N}}$ peuvent être estimés localement en les supposant égaux à une somme pondérée de modèles :

$$\bar{s}^{(\mathbf{n})} = \sum_{j=0}^{J-1} \sum_{p=p'}^{p'+P_j-1} \bar{h}_j^{(\mathbf{n})}(p) r_j^{(n_t-p, n_x)} \quad (5.52)$$

Nous supposons avoir J templates à notre disposition $(r_j^{(\mathbf{n})})_{\mathbf{n} \in \mathcal{N}, 0 \leq j < J}$ et $\bar{h}_j^{(\mathbf{n})}$ représentent donc les réponses impulsionnelles des filtres (avec P_j coefficients) pour le template r_j au temps n_t et capteur n_x (le filtrage est variant en temps et en espace) et où $p' \in \{-P_j + 1, \dots, 0\}$.

On peut également réécrire l'Eq. (5.52) comme

$$\bar{s} = \sum_{j=0}^{J-1} R_j \bar{h}_j \quad (5.53)$$

où

$$\bar{s} = \left[\bar{s}^{(0,0)}, \dots, \bar{s}^{(N_t-1,0)}, \dots, \bar{s}^{(0,N_x-1)}, \dots, \bar{s}^{(N_t-1,N_x-1)} \right]^\top, \quad (5.54)$$

$$\bar{h}_j = \left[\bar{h}_j^{(0)\top}, \dots, \bar{h}_j^{(N_x-1)\top} \right]^\top$$

et

$$\bar{h}_j^{(n_x)} = \left[\bar{h}_j^{(0,n_x)}(p'), \dots, \bar{h}_j^{(0,n_x)}(p' + P_j - 1), \dots, \bar{h}_j^{(N_t-1,n_x)}(p'), \dots, \bar{h}_j^{(N_t-1,n_x)}(p' + P_j - 1) \right]^\top. \quad (5.55)$$

La matrice R_j est bloc diagonale et ses éléments diagonaux notés $(R_j^{(n_t, n_x)})_{n_t \in \mathbb{N}_t}$ sont donnés par

$$\begin{bmatrix} R_j^{(0, n_x)} \\ \vdots \\ R_j^{(N_t-1, n_x)} \end{bmatrix} = \begin{bmatrix} r_j^{(-p', n_x)} & \dots & r_j^{(0, n_x)} & 0 & \dots & 0 \\ r_j^{(-p'+1, n_x)} & \dots & & r_j^{(0, n_x)} & 0 & \dots & 0 \\ \vdots & & & & & & \\ r_j^{(N_t-1, n_x)} & r_j^{(N_t-2, n_x)} & & \dots & & r_j^{(N_t-P_j, n_x)} \\ 0 & r_j^{(N_t-1, n_x)} & & \dots & & r_j^{(N_t-P_j+1, n_x)} \\ \vdots & & & & & \\ 0 & \dots & 0 & r_j^{(N_t-1, n_x)} & \dots & r_j^{(N_t-P_j-p', n_x)} \end{bmatrix}. \quad (5.56)$$

Soient $\mathbf{R} = [R_0 \dots R_{J-1}] \in \mathbb{R}^{N_t N_x \times N_t N_x P}$ et $\bar{\mathbf{h}} = [\bar{h}_0^\top \dots \bar{h}_{J-1}^\top]^\top \in \mathbb{R}^{N_t N_x P}$, avec $P = \sum_{j=0}^{J-1} P_j$, alors le modèle 2D (5.51) peut être réécrit comme

$$z = \mathbf{R} \bar{\mathbf{h}} + \bar{y} + b. \quad (5.57)$$

Dans un contexte variationnel, le problème d'optimisation à résoudre s'écrit

$$\underset{y \in \mathbb{R}^{N_t N_x}, \mathbf{h} \in \mathbb{R}^{N_t N_x P}}{\text{minimize}} \quad \|z - y - \mathbf{R} \mathbf{h}\|_2^2 + \iota_D(Fy) + \iota_C(\mathbf{h}), \quad (5.58)$$

où $F \in \mathbb{R}^{KN_x \times N_t N_x}$ représente un opérateur de trame (e.g. ondelettes directionnelles [52]). Les fonctions ι_C et ι_D représentent les fonctions indicatrices des ensemble convexes C et D , respectivement i.e. $\iota_C(x) = 0$ si $x \in C$ et $\iota_C(x) = +\infty$ si $x \notin C$.

5.3.4-b) Définitions des nouveaux ensembles de contraintes C et D

Comme précédemment, la contrainte sur l'ensemble D peut être éclatée en plusieurs contrainte en définissant une partition de $\{1, \dots, KN_x\}$ notée $\{\mathbb{K}_l \cup \mathbb{K}_{l+\mathcal{L}} \mid l \in \{1, \dots, \mathcal{L}\}\}$, où \mathcal{L} correspond au nombre de sous-bandes, \mathbb{K}_l et $\mathbb{K}_{l+\mathcal{L}}$ sont les l ème sous-bandes correspondant aux coefficients primaux et duaux respectivement. On peut donc choisir

$D = D_1 \times \dots \times D_{\mathcal{L}}$ où, pour tout $l \in \{1, \dots, \mathcal{L}\}$, $D_l = \{(x_k)_{k \in \mathbb{K}_l \cup \mathbb{K}_{l+\mathcal{L}}} \mid \sum_{k \in \mathbb{K}_l} \varphi_l(x_k) \leq \beta_l \text{ et } \sum_{k \in \mathbb{K}_{l+\mathcal{L}}} \varphi_l(x_k) \leq \beta_l\}$, $\varphi_l : \mathbb{R} \rightarrow [0, +\infty[$ est une fonction convexe semi-continue inférieurement (pour laquelle on peut naturellement choisir une norme ℓ_1).

A nouveau, C introduit des a priori sur les filtres et peut s'exprimer comme $C = C_1 \cap C_2 \cap C_3$ où C_1 et C_2 (variations lentes) deviennent

$$\begin{aligned} C_1 &= \left\{ h \in \mathbb{R}^{N_t N_x P} \mid \forall (j, n_t, n_x, p), |h_j^{(n_t+1, n_x)}(p) - h_j^{(n_t, n_x)}(p)| \leq \varepsilon_{j,p}^{n_x} \right\} \\ C_2 &= \left\{ h \in \mathbb{R}^{N_t N_x P} \mid \forall (j, n_t, n_x, p), |h_j^{(n_t, n_x+1)}(p) - h_j^{(n_t, n_x)}(p)| \leq \varepsilon_{j,p}^{n_t} \right\}, \end{aligned} \quad (5.59)$$

avec $(\varepsilon_{j,p}^{n_x}, \varepsilon_{j,p}^{n_t}) \in [0, +\infty[^2$. Quand à C_3 , il devient $C_3 = \{\mathbf{h} \in \mathbb{R}^{N_t N_x P} \mid \rho(\mathbf{R}\mathbf{h}) \leq \lambda\}$, où $\lambda \in]0, +\infty[$ et $\rho : \mathbb{R}^{N_t N_x} \rightarrow [0, +\infty[$ est une fonction convexe semi-continue inférieurement.

5.3.4-c) Algorithme

Je vais passer assez vite puisque l'algorithme permettant de résoudre le problème de minimisation ainsi défini est très proche de celui développé dans le paragraphe 5.3.3-b). Il est possible donc d'utiliser un algorithme primal-dual M+L FBF alternant une étape de gradient et des projections sur les différents ensembles de contraintes (voir paragraphe 5.3.3-a)).

5.3.5 Simulations numériques

Nous considérons ici deux familles de filtres de longueurs respectives $P_0 = 10$ et $P_1 = 14$. Les filtres évoluent de manière complémentaire au cours du temps, simulant un mélange multiple bimodal à deux profondeurs différentes. Ils sont combinés à deux modèles pour donner les multiples représentés à la cinquième ligne en partant du haut de la figure 5.6. Les données et les modèles ont été construits de manière à reproduire les contenus temporels et fréquentiels d'un signal de sismique. Cette figure représente tous les signaux d'intérêt (dont certains sont connus d'autres inconnus) pour un écart type de bruit $\sigma = 0.08$. L'objectif ici est donc de retrouver le signal primaire, souvent caché par les multiples et les perturbations aléatoires, à partir des observations z . L'estimation obtenue est représentée sur la deuxième ligne en rose. Concentrons nous à présent sur ces signaux, pour des indices temporels compris entre 350 et 700. Ces signaux sont représentés dans la figure 5.7 où un deuxième zoom est effectué pour les indices 540-600. Nous pouvons ici faire deux observations : le premier événement (indices 400-500) est principalement affecté par le bruit aléatoire qui est assez bien éliminé de l'estimation ; le deuxième événement (indices 530-600) est à la fois affecté par le bruit aléatoire mais aussi par un multiple d'amplitude significative qui sont tout deux éliminés de l'estimation. L'algorithme proposé prouve ici sa grande efficacité.

Nous avons également testé notre méthode sur des données sismiques synthétiques 2D constituées de $N_x = 512$ traces sismiques de longueur $N_t = 512$ avec deux filtres (liés aux multiples donc) de longueurs $P_0 = 4$ et $P_1 = 5$.

Les ensembles de contraintes C_1 et C_2 sont définis par (5.59) où, pour tout (j, p, n_x, n_t) , $\varepsilon_{j,p}^{(n_x)} = 0.05$ et $\varepsilon_{j,p}^{(n_t)} = 0.0001$. L'image observée z (avec $\sigma = 0.04$) est représentée dans en haut de la figure 5.8, où les hyperboles en haut et en bas correspondent aux primaires alors que celle du milieu représente les multiples. Les ensembles de contraintes D et C_3 , sont choisis en prenant $\varphi_l = |\cdot|$, $L = \mathbf{R}$ et nous étudions plusieurs possibilités pour ρ :

1. norme ℓ_1

$$(\forall d \in \mathbb{R}^{N_t \times N_x}) \quad \rho(d) = \ell_1(d) = \sum_{n_x=0}^{N_x-1} \sum_{n_t=0}^{N_t-1} |d^{(n_t, n_x)}|,$$

2. norme ℓ_2

$$(\forall d \in \mathbb{R}^{N_t \times N_x}) \quad \rho(d) = \ell_2(d) = \left(\sum_{n_x=0}^{N_x-1} \sum_{n_t=0}^{N_t-1} (d^{(n_t, n_x)})^2 \right)^{1/2},$$

3. norme mixte $\ell_{1,2}$

$$(\forall d \in \mathbb{R}^{N_t \times N_x}) \quad \rho(d) = \ell_{1,2}(d) = \sum_{n_x=0}^{N_x-1} \left(\sum_{n_t=0}^{N_t-1} (d^{(n_t, n_x)})^2 \right)^{1/2}.$$

Les projections sur les ensembles C et D sont effectuées de la même manière que dans le cas 1D (c'est à dire basée sur la procédure décrite dans [54]). Les transformées sont choisies de la manière suivante : pour la dimension temps (verticalement) nous avons choisis d'utiliser des Symlet de longueur 8 alors que pour la dimension spatiale, nous avons choisis des ondelettes de Daubechies de longueur 4. La transformée F ainsi définie peut-être soit : 1D invariant par translation (SI, Symlet longueur 8 suivant la dimension temps uniquement), soit 2D en arbre dual (redondance d'un facteur 2). La figure 5.9, représente les primaires estimés : (haut) avec une méthode de l'état de l'art [56], (milieu) estimés en utilisant la méthode 1D avec des trames SI, et enfin, (bas) en utilisant la méthode 2D avec une décomposition en arbre dual. La meilleure reconstruction est obtenue avec la méthode 2D.

Nous avons également testé nos algorithmes sur données réelles affichées en figure 5.10. Les templates utilisés sont donnés en figure 5.11 et les résultats obtenus sont présentés dans la figures 5.12 pour la méthode de l'état de l'art et la méthode 1D et dans la figure 5.13 pour la méthode 2D.

5.4 Conclusion

Comme nous l'avons vu dans ce chapitre, l'estimation des paramètres du modèle s'associe la plupart du temps avec l'estimation de la quantité recherchée elle-même mais souvent de manière sous-optimale ... Nous allons voir dans le chapitre suivant des stratégies d'optimisation évoluées permettant de résoudre des problèmes d'optimisation complexes et parfois de grande taille.

5.5 Quelques publications personnelles en lien avec ce chapitre

- [CCp1] A. Jeziarska, C. Chaux, J.-C. Pesquet, H. Talbot, and G. Engler. An EM approach for time-variant Poisson-Gaussian model parameter estimation. *IEEE Trans. Signal Process.*, 62(1) :17–30, Jan. 2014.
- [CCp2] A. Jeziarska, J.-C. Pesquet, H. Talbot, and C. Chaux. Iterative poisson-gaussian noise parametric estimation for blind image denoising. In *Proc. Int. Conf. Image Process.*, Paris, France, Oct. 2014.
- [CCp3] C. Chaux and L. Blanc-Féraud. Estimation d’hyperparamètres pour la résolution de problèmes inverses à l’aide d’ondelettes. In *GRETSI 2009*, Dijon, Sep. 8-11 2009.
- [CCp4] R. Cavicchioli, C. Chaux, L. Blanc-Féraud, and L. Zanni. ML estimation of wavelet regularization hyperparameters in inverse problems. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, page 5 pp., Vancouver, Canada, May, 26-31 2013.
- [CCp5] M. Q. Pham, L. Duval, C. Chaux, and J.-C. Pesquet. A primal-dual proximal algorithm for sparse template-based adaptive filtering : Application to seismic multiple removal. *IEEE Trans. Signal Process.*, 62(16) :4256–4269, Aug. 2014.
- [CCp6] M. Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet. Sparse adaptive template matching and filtering for 2d seismic images with dual-tree wavelets and proximal methods. In *Proc. Int. Conf. Image Process.*, Québec city, Canada, 27-30 Sep. 2015.

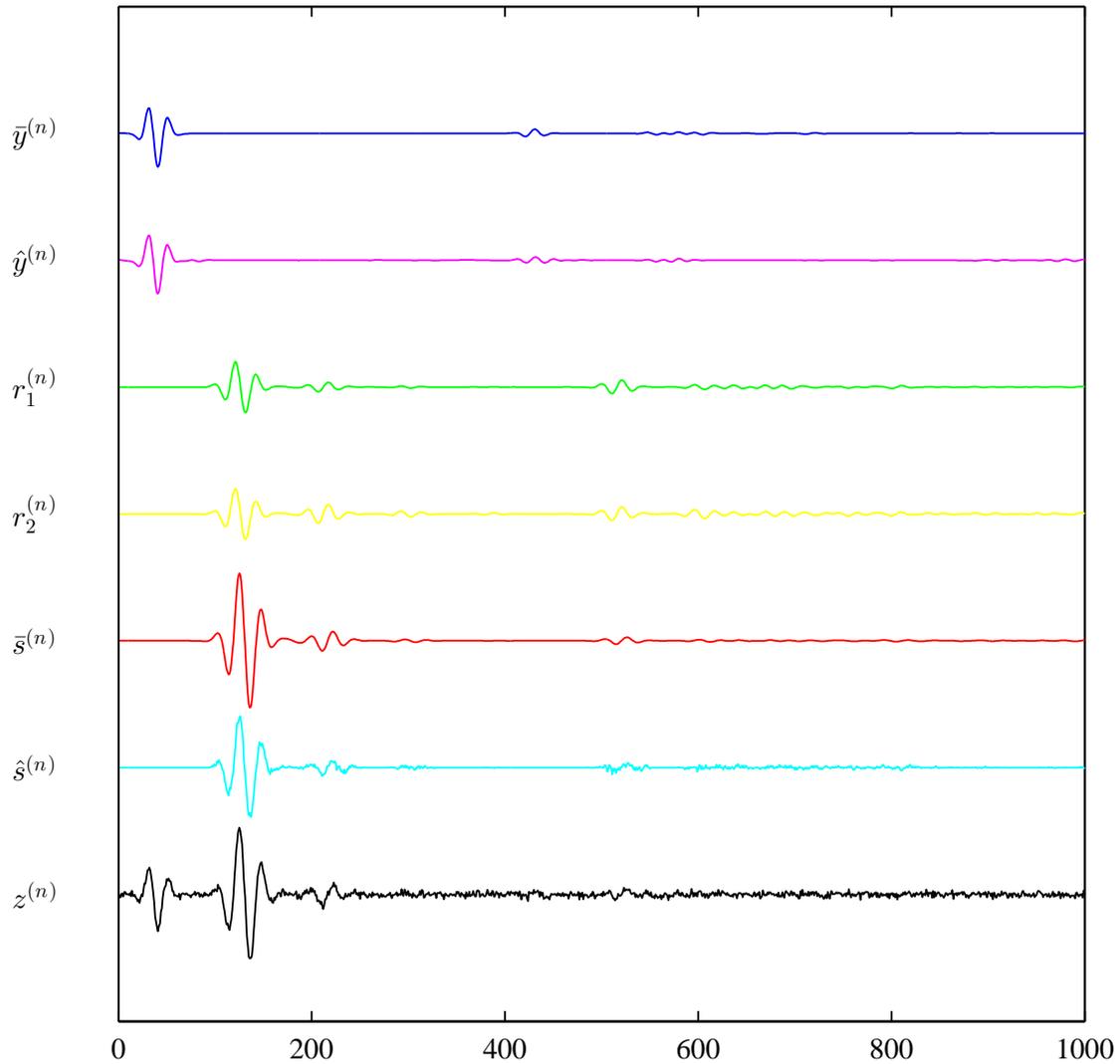


FIGURE 5.6 – Signaux sismiques considérés lorsque le bruit a un écart-type $\sigma = 0.08$. De haut en bas ; primaire (inconnu) \bar{y} , estimé \hat{y} , premier template r_0 , second template r_1 , multiple (inconnu) \bar{s} , estimé \hat{s} , et signal observé z .

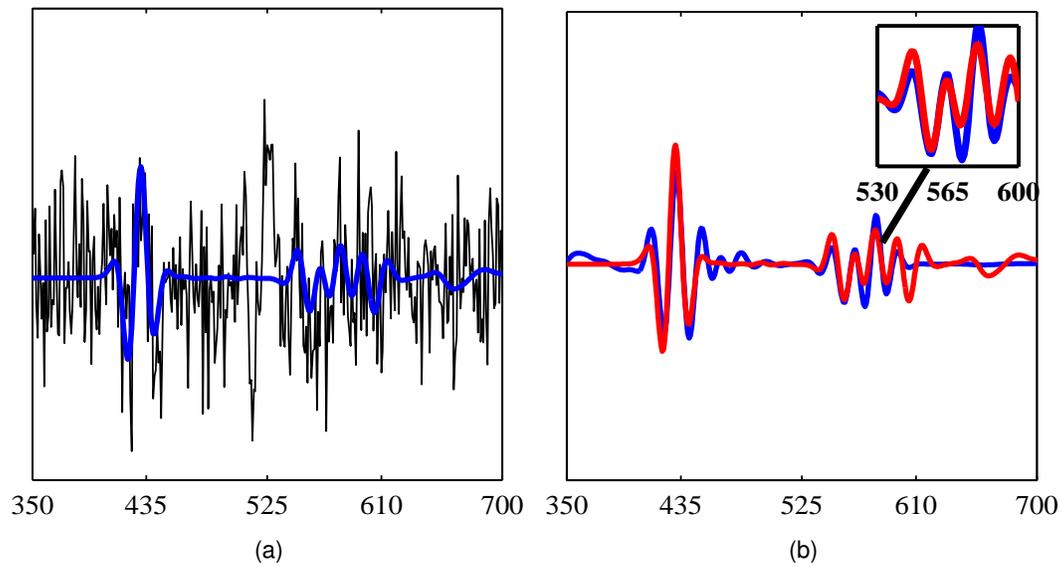


FIGURE 5.7 – Close-up when $\tilde{\rho}$ is the $\ell_{1,2}$ -norm; (a) observation z (en noir $\sigma = 0.08$) et primaire \bar{y} (en bleu), (b) primaire estimé \hat{y} (en rouge) et primaire original \bar{y} (en bleu).

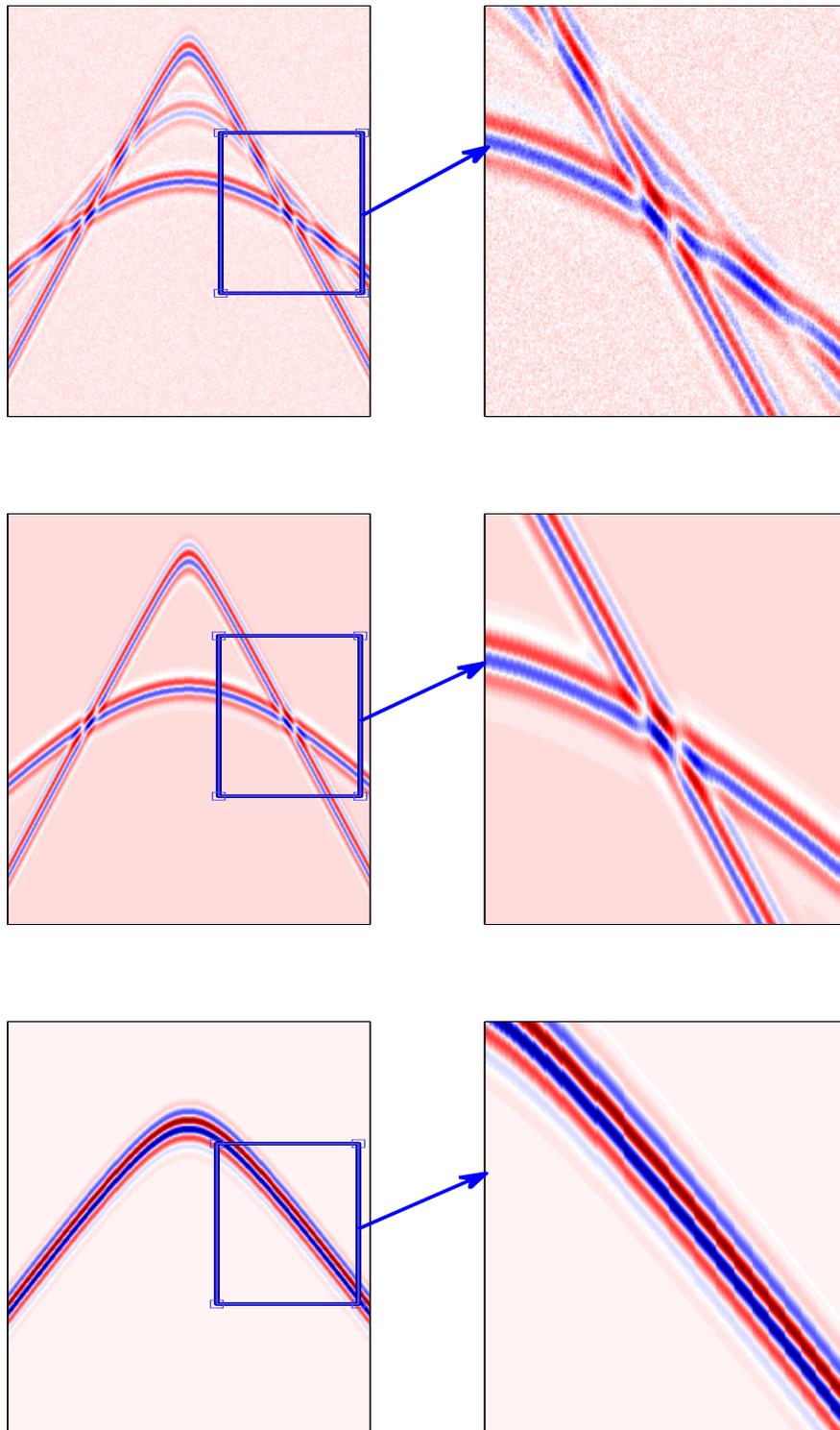


FIGURE 5.8 – De haut en bas : image observée z avec $\sigma = 0.04$, primaire \bar{y} et multiple \bar{s} .

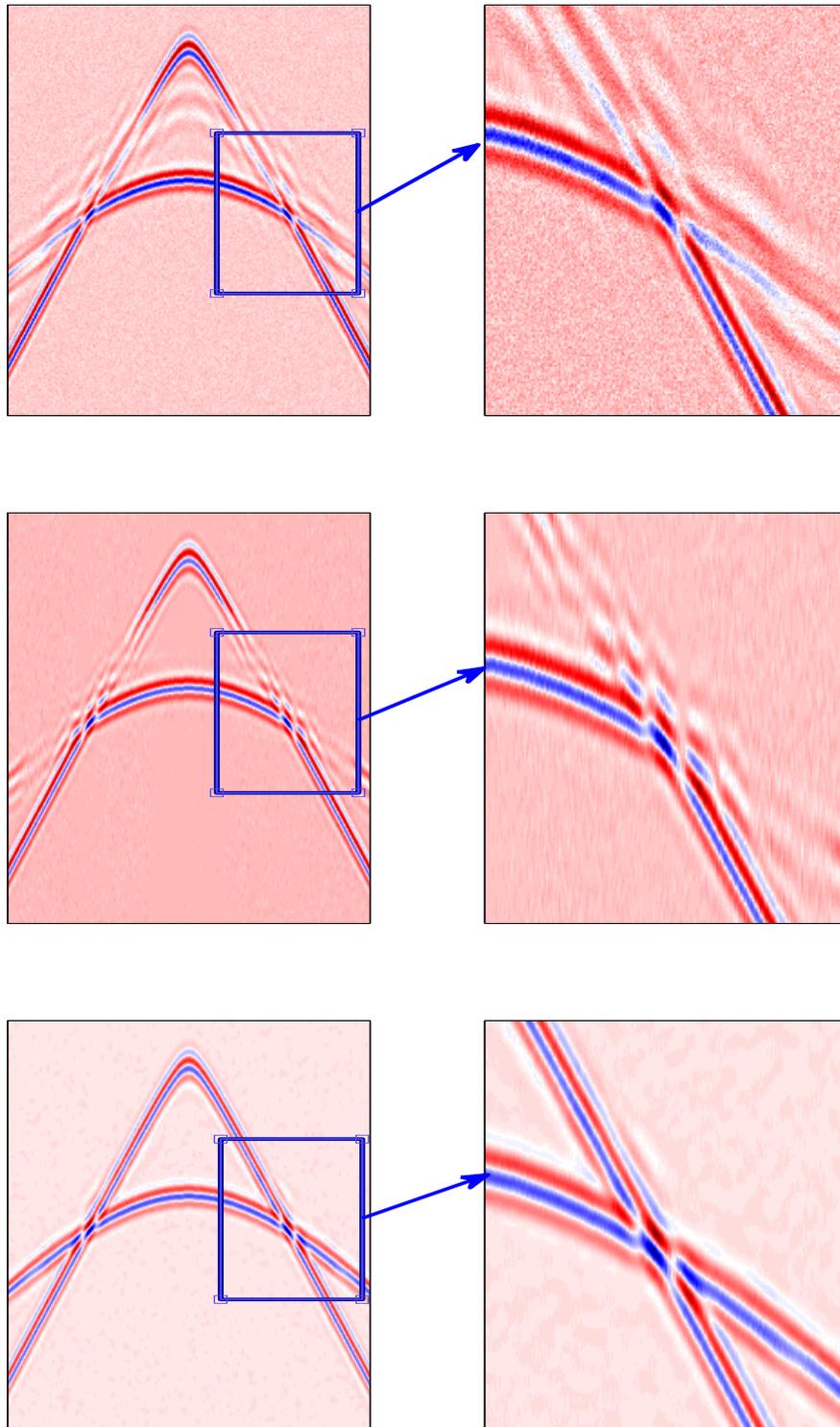


FIGURE 5.9 – De haut en bas : primaire estimé y avec [56], avec la version 1D et avec la version 2D.

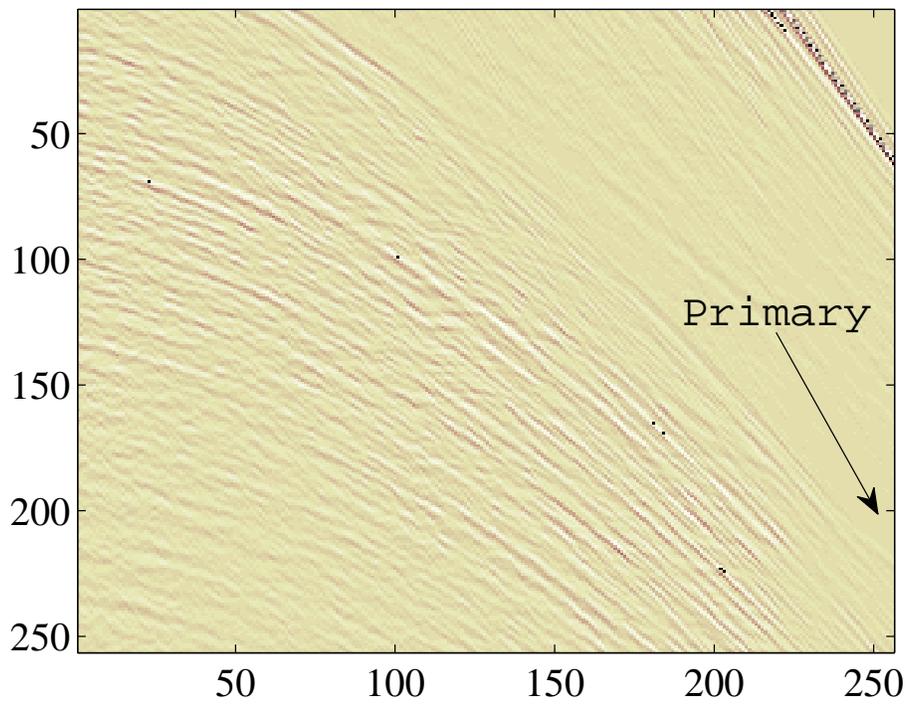
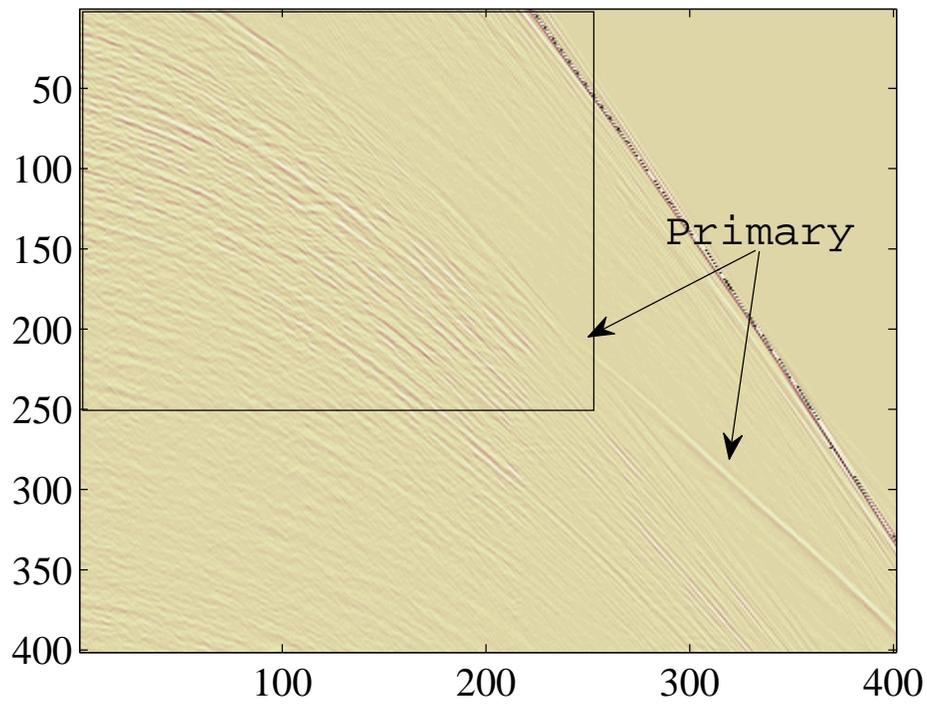


FIGURE 5.10 – De haut en bas : données sismiques réelles observées avec un primaire partiellement apparent et un zoom de ces données.

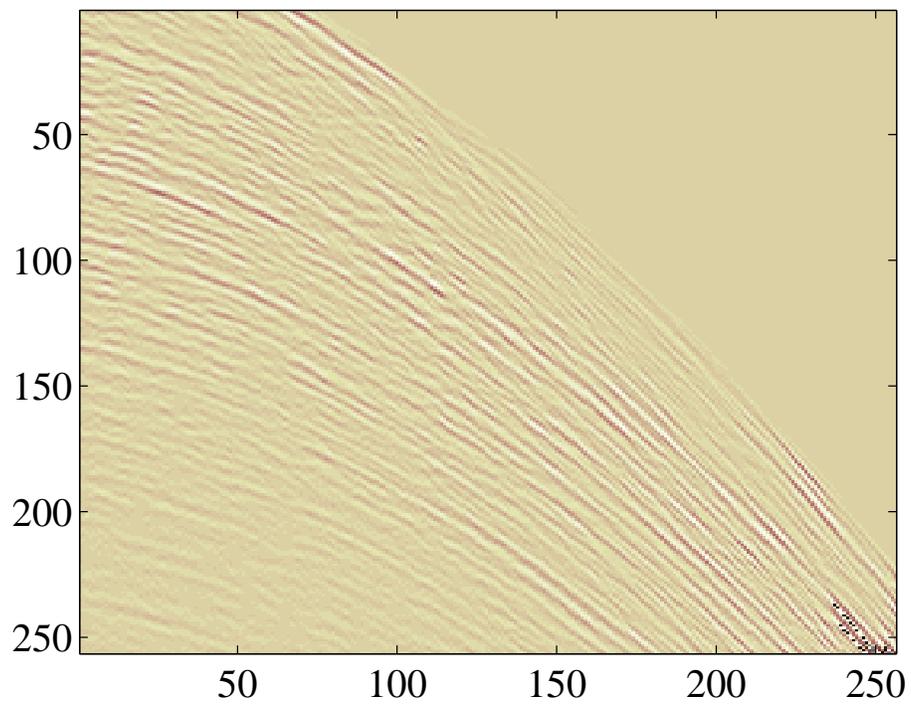
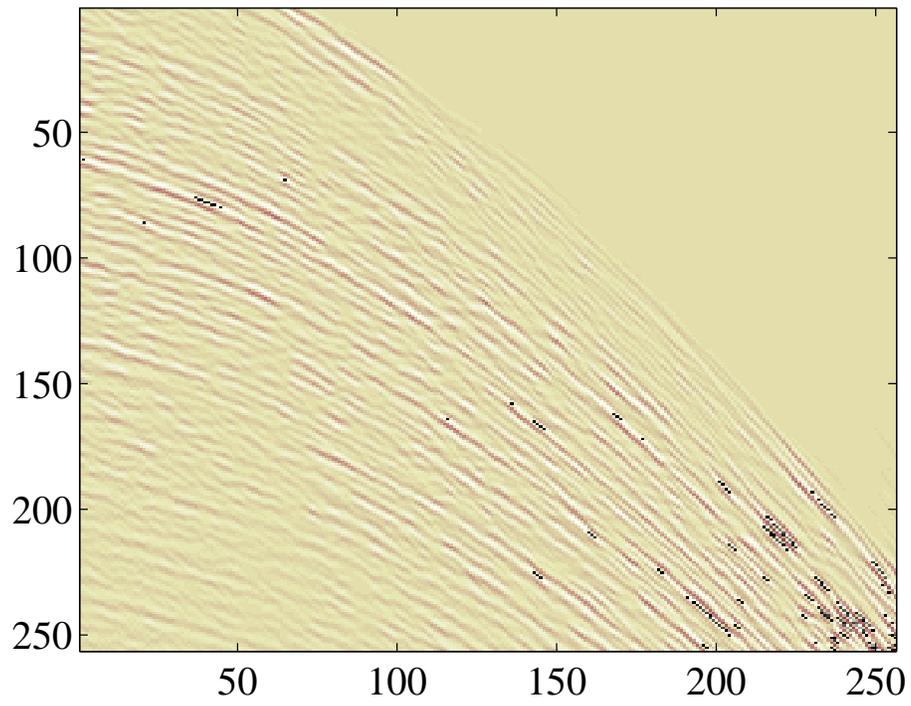


FIGURE 5.11 – Deux templates utilisés pour estimer les multiples.

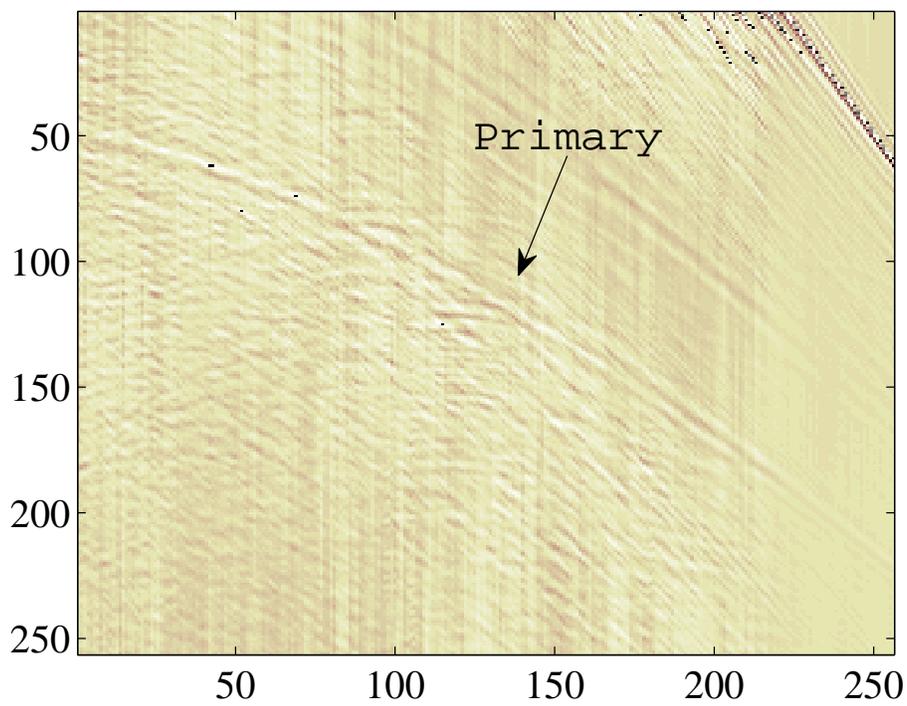
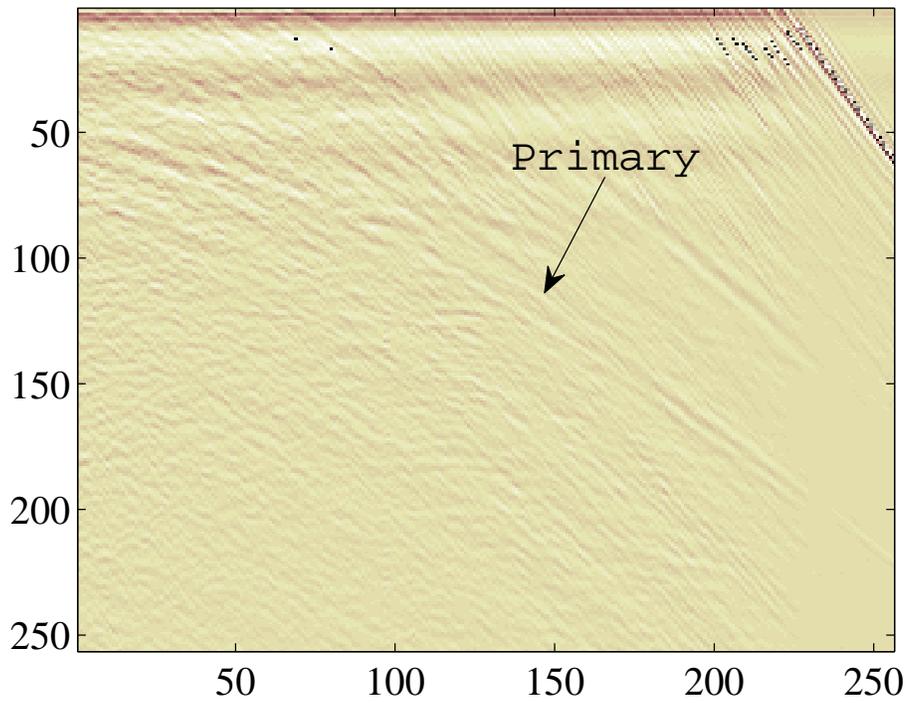


FIGURE 5.12 – De haut en bas : primaires estimés avec la méthode de l'état de l'art [56] et la version 1D de la méthode

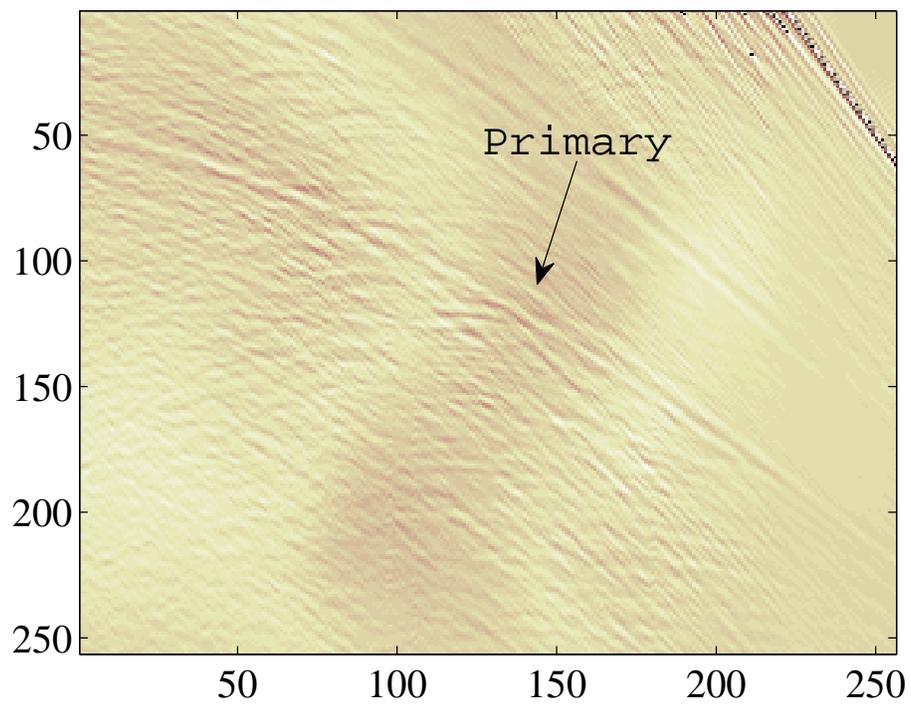


FIGURE 5.13 – Primaire estimé en utilisant la méthode 2D.

Chapitre 6

Optimisation en données complètes

Avant-propos

Ce dernier chapitre traite des cas où le modèle est parfaitement connu et que les paramètres de régularisation sont fixé empiriquement. On s'intéresse alors directement à estimer les variables d'intérêt. Je décrirai alors particulièrement deux applications : la restauration/reconstruction d'images et la factorisation non-négative de tenseurs.

Sommaire

6.1 Application : reconstruction d'images	81
6.2 Application : factorisation tensorielle non-négative	87
6.3 Quelques publications personnelles en lien avec ce chapitre	99

Nous nous plaçons ici dans le cas de données complètes (le modèle et ses paramètres sont supposés connus) et nous nous intéressons au problème d'optimisation à résoudre en supposant le modèle connu (paramètres inclus) et en déterminant les hyperparamètres de manière empirique.

6.1 Application : reconstruction d'images

6.1.1 Introduction

Nous nous intéressons ici à la restauration des données issues de la microscopie biphotonique [57, 58]. Cette modalité de microscopie par fluorescence, apparue dans les années 90, est fondée sur le principe de l'absorption à deux photons, par opposition à la microscopie confocale conventionnelle où l'absorption se fait à un photon. Elle possède de nombreux avantages par rapport à cette dernière, dont sa capacité à imager à de plus grandes profondeurs, à limiter la phototoxicité (ce qui permet d'effectuer des expériences *in vivo* [59]), ainsi qu'à réduire le bruit d'acquisition [60]. Le modèle de formation de l'image, ou problème direct, est le suivant

$$y = H\bar{x} + b, \tag{6.1}$$

où $\bar{x} \in \mathbb{R}^N$, avec $N = N_1 \times N_2 \times N_3$, représente les données 3D originales (inconnues), $\mathbf{y} \in \mathbb{R}^N$ est l'image mesurée (dégradée), \mathbf{H} est un opérateur linéaire de convolution 3D modélisant le système d'acquisition (lié à la PSF de l'instrument) et \mathbf{b} représente un bruit additif. L'objectif de ce travail¹ est de proposer des méthodes permettant de résoudre le problème inverse associé, qui consiste donc à restaurer l'image volumique \bar{x} à partir d'une version dégradée \mathbf{y} de celle-ci.

La méconnaissance de l'opérateur \mathbf{H} et la grande dimension du volume de données à traiter constituent deux verrous importants de ce problème. Pour adresser le premier point, nous avons mis en place une stratégie purement expérimentale, basée sur l'observation de micro-billes. Plus simple à mettre en place que des stratégies de restauration aveugle, cette approche, déjà utilisée en microscopie par fluorescence classique [61], nous a permis d'obtenir des mesures de PSF satisfaisantes. Nous proposons de traiter le problème inverse non aveugle résultant via la minimisation d'un critère pénalisé.

6.1.2 Méthodes proposées

Comme nous l'avons vu précédemment dans le chapitre 4 une approche classique de résolution est de définir une estimée de \bar{x} comme le minimiseur d'un critère composite $F(\mathbf{x})$ somme d'un terme d'attache aux données lié au modèle d'observation et d'un terme de régularisation renforçant certaines propriétés de l'objet recherché. Nous considérerons ici le critère de moindres carrés pénalisé

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad F(\mathbf{x}) = \frac{1}{2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 + R(\mathbf{x}). \quad (6.2)$$

Nous allons décrire à présent deux régularisations que nous avons considérées, et les algorithmes mis en place afin de résoudre le problème de minimisation ainsi formulé.

6.1.2-a) Pénalisation multirésolution non lisse

Ici, une pénalisation exprimée dans le domaine ondelette a tout d'abord été considérée. Cela conduit à un terme de régularisation de la forme $R = \lambda \|\mathcal{F}\mathbf{x}\|_{\ell_1}$ où $\lambda > 0$ est le paramètre de régularisation et $\mathcal{F} \in \mathbb{R}^{N \times N}$ représente l'opérateur de transformée en ondelettes orthogonale. L'anisotropie des données nous a conduit à considérer un même niveau de résolution sur les axes XY et un niveau de résolution différent suivant l'axe Z. La fonction de coût ainsi définie peut être minimisée grâce à l'algorithme Implicite-Explicite [5] qui est défini par les itérations suivantes :

$$(\forall k \in \mathbb{N}) \quad \mathbf{x}_{k+1} = \mathcal{F}^* (\text{Soft}_{\lambda/\gamma} (\mathcal{F} (\mathbf{x}_k - \gamma \mathbf{H}^* (\mathbf{H}\mathbf{x}_k - \mathbf{y}))))), \quad (6.3)$$

où $0 < \gamma < 2/\|\mathbf{H}\|^2$ est le pas de l'algorithme, $\mathcal{F}^* = \mathcal{F}^{-1}$ est l'opérateur de synthèse en ondelettes et Soft_α est l'opération de seuillage doux (seuil $\alpha > 0$).

On pourrait bien-sûr considérer un seuil différent pour chaque sous-bande de décomposition ; il faudrait néanmoins faire une estimation de ces hyperparamètres, ce qui est difficile.

6.1.2-b) Pénalisation hybride lisse

Une pénalisation hybride $R = R_1 + R_2 + R_3$ a ensuite été considérée. Tout d'abord, afin de contraindre l'intensité des pixels restaurés dans un certain intervalle $[x_{\min}, x_{\max}]$,

1. Projet jeune chercheur GdR ISIS - GRETSI / Projet Défi Imag'in OPTIMISME.

nous définissons

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad R_1(\mathbf{x}) = \eta \sum_{n=1}^N d_{[x_{\min}, x_{\max}]}^2(x_n), \quad (6.4)$$

où $\eta > 0$ et d_C est la distance au convexe C . La fonction R_1 est une approximation lisse de la fonction indicatrice $\iota_{[x_{\min}, x_{\max}]}^N$ (celle-ci est retrouvée quand η tend vers l'infini), et joue un rôle similaire à une pénalisation quadratique exacte.

En microscopie biphotonique, l'axe Z, correspondant à la profondeur d'acquisition, est souvent moins finement résolu que le plan XY. Par conséquent, deux termes distincts ont été considérés pour renforcer la régularité spatiale de l'objet estimé :

$$\begin{aligned} (\forall \mathbf{x} \in \mathbb{R}^N) \quad R_2(\mathbf{x}) &= \lambda \sum_{n=1}^N \sqrt{([\mathbf{V}^X \mathbf{x}]_n)^2 + ([\mathbf{V}^Y \mathbf{x}]_n)^2 + \delta^2}, \\ R_3(\mathbf{x}) &= \kappa \sum_{n=1}^N ([\mathbf{V}^Z \mathbf{x}]_n)^2, \end{aligned} \quad (6.5)$$

avec $(\lambda, \delta, \kappa) \in (0, +\infty)^3$, et $\mathbf{V}^X \in \mathbb{R}^{N \times N}$, $\mathbf{V}^Y \in \mathbb{R}^{N \times N}$, $\mathbf{V}^Z \in \mathbb{R}^{N \times N}$ des opérateurs linéaires calculant les gradients discrets de l'objet suivant les axes X, Y et Z. La fonction R_2 s'apparente à une version lissée de la variation totale (cette dernière est obtenue pour $\delta = 0$).

Nous proposons d'utiliser une stratégie de Majoration-Minimisation (MM) pour minimiser le critère convexe différentiable F défini par (6.2)-(6.5). Il s'agit de remplacer le problème de minimisation initial par une séquence de sous-problèmes plus simples, ici quadratiques.

En pratique, cela conduit à l'algorithme suivant :

$$(\forall k \in \mathbb{N}) \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{D}_k \mathbf{u}_k, \quad (6.6)$$

où \mathbf{D}_k définit le sous-espace de mémoire de gradient :

$$\mathbf{D}_k = \begin{cases} -\nabla F(\mathbf{x}_0) & \text{si } k = 0 \\ [-\nabla F(\mathbf{x}_k) \mid \mathbf{x}_k - \mathbf{x}_{k-1}] & \text{si } k > 0 \end{cases} \quad (6.7)$$

et \mathbf{u}_k est donné par

$$\mathbf{u}_k = - \left(\mathbf{D}_k^\top \mathbf{A}(\mathbf{x}_k) \mathbf{D}_k \right)^\dagger \mathbf{D}_k^\top \nabla F(\mathbf{x}_k), \quad (6.8)$$

où \dagger est l'opérateur de pseudo-inverse. La convergence de la suite $(\mathbf{x}_k)_{k \in \mathbb{N}}$ résultant de l'alternance des étapes (6.6) à (6.8) vers un minimiseur de F est établie dans [62]. En pratique, les itérations sont stoppées dès que le critère d'arrêt suivant est vérifié :

$$\|\nabla F(\mathbf{x}_k)\| \leq \epsilon, \quad (6.9)$$

où, typiquement, $\epsilon = 10^{-2}$. Notons que l'algorithme 3MG s'est distingué sur divers exemples pratiques par son faible coût calculatoire, et ses très bonnes performances en terme de vitesse de convergence [62, 63].

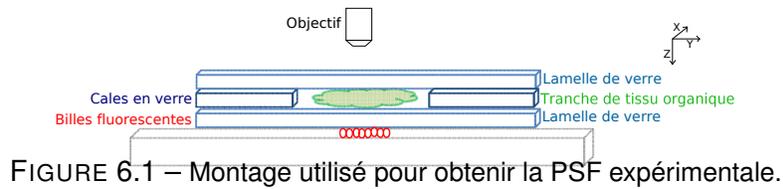


FIGURE 6.1 – Montage utilisé pour obtenir la PSF expérimentale.

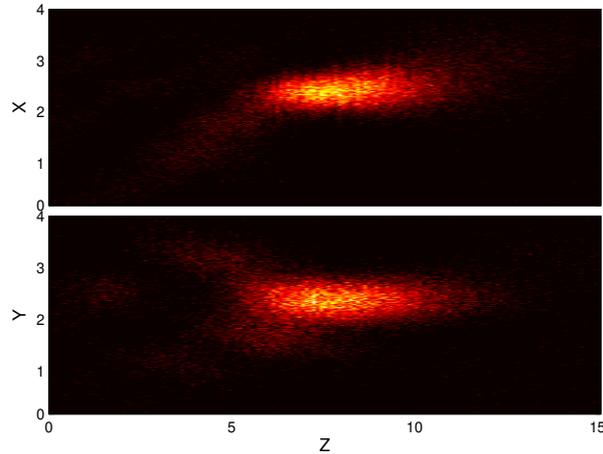


FIGURE 6.2 – Coupes centrales de la PSF expérimentale, issue de l'observation au microscope biphotonique de micro-billes fluorescentes de diamètre 0.5 microns placées sous une tranche de tissu cérébral d'épaisseur 100 microns.

6.1.3 Résultats de simulation

Dans cette section, nous présentons les résultats de restauration obtenus par l'algorithme 3MG appliqué à des données synthétiques et réelles. Les résultats obtenus avec l'algorithme Implicite-Explicite ne sont pas donnés ici par manque de place (les tests conduisent à des images de qualité similaire avec un temps de calcul légèrement supérieur). Les simulations sont effectuées sur un Intel(R) Xeon(R) CPU E5-2630v2 @ 2.6GHz avec Matlab R2013a. Tous les axes des figures sont gradués en microns. Pour tous les exemples présentés, nous considérerons que les données sont dégradées par une PSF expérimentale dont nous décrivons la stratégie d'obtention ci-dessous.

6.1.3-a) Détermination de la PSF expérimentale

Nous cherchons à obtenir une PSF expérimentale à profondeur 100 microns dans le cerveau à l'aide du microscope biphotonique LSM 780 (Zeiss), disponible au CERIMED², lorsque le contexte d'acquisition est le suivant : (i) le faisceau d'excitation est produit par un laser infrarouge cohérent pulsé à 150 femtosecondes et réglé à la longueur d'onde 840 nm, (ii) la fluorescence est acquise sur 5 détecteurs en cascade analysant chacun une bande spectrale distincte du spectre visible, (iii) seul le détecteur sensible dans la bande 555-605 nm est considéré dans notre étude. Nous mettons en place le montage illustré en figure 6.1. La tranche de tissu cérébral, d'une épaisseur de 100 microns, a été découpée au vibratome à partir d'un cerveau prélevé sur une souris. Plusieurs acquisitions (typiquement 5) de billes fluorescentes T7284 (produites par Molecular Probes Life technologies) de 0.5 microns de diamètre, placées sous cette tranche de cerveau, sont effectuées avec

2. <http://www.cerimed-web.eu/>

une résolution très fine ($0.02 \times 0.02 \times 0.10$ microns). Nous réalisons ensuite un recalage puis une moyenne des images obtenues, de façon à prendre en compte les éventuelles inhomogénéités structurelles du tissu cérébral. La figure 6.2 illustre la PSF résultante, de taille $191 \times 191 \times 151$ pixels. Nous pouvons remarquer la présence de bruit persistant malgré l'étape de moyennage, ainsi que le fort étalement suivant l'axe Z, caractéristique de la microscopie par fluorescence [64]. Dans la suite de cette section, H sera un opérateur de convolution 3D, dont le noyau est une version interpolée par splines cubiques de cette PSF expérimentale, de façon à s'adapter à la résolution des données à restaurer.

6.1.3-b) Restauration d'une image synthétique

Nous commençons avec la restauration de données synthétiques. L'objet original est l'image CT_aneurysm, issue du site <http://www.mathworks.com/matlabcentral/fileexchange/25987-showvol-isosurface-render/>, de taille $155 \times 154 \times 154$ pixels, retenue à cause de la présence de structures filaires dans celle-ci. Afin de générer y , nous avons appliqué à l'image originale la PSF expérimentale, en supposant une résolution de $0.08 \times 0.08 \times 0.1$ microns. Un bruit additif Gaussien a , de plus, été ajouté, afin d'avoir un SNR de 20 dB. Les paramètres de régularisation sont réglés manuellement afin de minimiser l'erreur entre l'image originale \bar{x} et l'image restaurée \hat{x} . Sur cet exemple, l'amélioration, en terme de SNR, entre y et \hat{x} est de 3.61 dB. La figure 6.3 illustre deux coupes, l'une suivant Z et l'autre suivant X, de l'image originale, dégradée, et restaurée. 360 itérations de l'algorithme sont nécessaires pour satisfaire le critère d'arrêt (6.9), ce qui correspond à un temps de reconstruction d'environ 8 minutes. La qualité visuelle des résultats, que nous avons confirmée sur d'autres jeux de données synthétiques, permet de valider la méthode proposée, dans le cas où l'opérateur H est connu de façon exacte.

6.1.3-c) Restauration de données réelles

Nous nous intéressons maintenant à la restauration d'images acquises par le microscope. Nous appliquons tout d'abord notre méthode à la restauration d'une image de bille de diamètre de 4 microns, de taille $186 \times 186 \times 351$ pixels, acquise par le montage décrit en figure 6.1, sous une résolution de $0.08 \times 0.08 \times 0.1$ microns. La figure 6.4 illustre deux coupes, suivant Z et X, de l'image observée et restaurée. Nous fixons les paramètres de régularisation afin d'obtenir une qualité visuelle satisfaisante de l'image restaurée. 25 minutes sont nécessaires pour la reconstruction. Nous pouvons noter une forte réduction du bruit de mesure. De plus, la forme sphérique et creuse de la bille est bien restaurée, en particulier dans le plan XY. La forte déformation sur l'axe Z liée à l'instrument est amoindrie mais reste néanmoins visible.

Nous nous intéressons ensuite à la restauration d'une image de taille $512 \times 512 \times 40$ pixels, correspondant à l'observation au microscope biphotonique d'un morceau de cerveau de souris transgénique clarifié par la méthode `scale` [65]. Contrairement aux tests précédents, cette fois-ci, une résolution moindre, de $0.8 \times 0.8 \times 2.5$ microns a été utilisée, afin de s'approcher des conditions réelles d'acquisition *in vivo*. La figure 6.5 illustre deux coupes, suivant Z et Y, de l'image observée, et restaurée après un temps de reconstruction d'environ 10 minutes. A nouveau, nous pouvons observer que la méthode proposée permet de réduire le bruit d'acquisition, et d'améliorer la qualité de l'image, en particulier dans le plan XY, ce qui permet une meilleure visualisation des corps cellulaires.

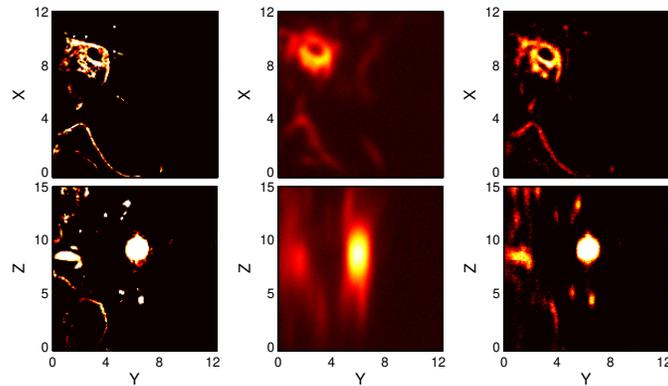


FIGURE 6.3 – Coupes originales (gauche), dégradées (centre) et restaurées (droite), à $Z = 13.6$ microns (haut) et $X = 5.36$ microns (bas), de l'image CT_aneurysm.

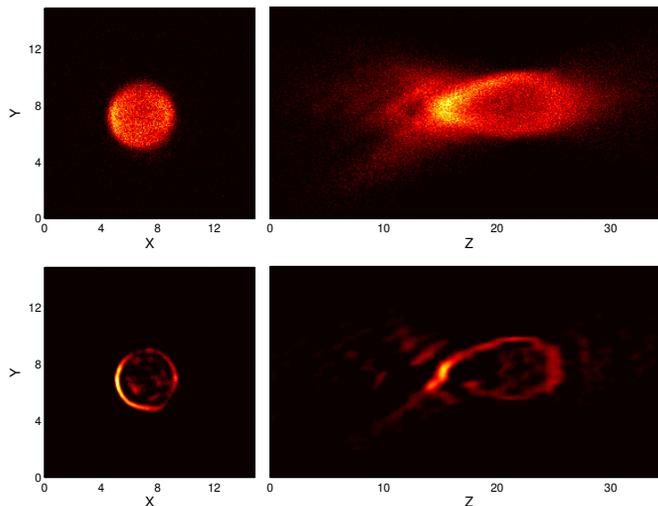


FIGURE 6.4 – Coupes observées au microscope biphotonique (haut) et restaurées (bas), à $Z = 22$ microns (gauche) et $X = 8$ microns (droite), d'une bille fluorescente de diamètre 4 microns.

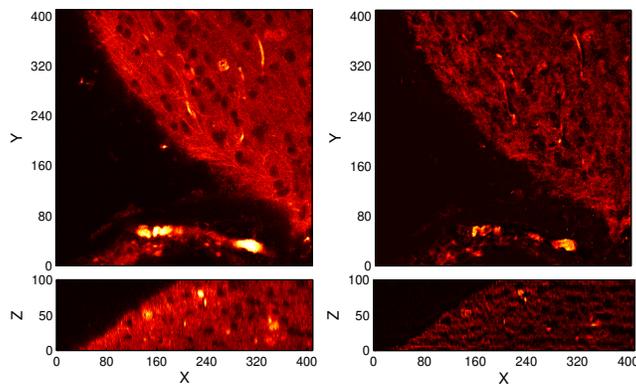


FIGURE 6.5 – Coupes observées au microscope biphotonique (gauche) et restaurées (droite), à $Z = 82.5$ microns (haut) et $Y = 369.6$ microns (bas), d'une tranche clarifiée de cerveau de souris.

6.1.4 Conclusion

Ce travail aborde le problème de déconvolution en microscopie biphotonique. Deux méthodes reposant sur une formulation variationnelle du problème et sur des algorithmes de type MM et proximal ont été proposées. Des résultats prometteurs ont été obtenus sur données réelles et simulées.

Ce projet ouvre les nombreuses perspectives de recherche suivantes : 1) l'obtention de modèles de PSF spatialement variantes ; en effet, des tests préliminaires montrent que la PSF varie suivant la profondeur (axe Z). L'algorithme Implicite-Explicite s'adapterait naturellement à cette modélisation [66] ; 2) l'élaboration d'une méthode automatique d'estimation des hyperparamètres, incluant une automatisation de l'évaluation de la qualité de la reconstruction ; 3) l'analyse de la nature du bruit. En effet, l'hypothèse d'un bruit Gaussien pour de telles données n'est peut-être pas adaptée (bruit de Poisson, mélange de bruit Poisson-Gauss ...); 4) l'exploitation du caractère multispectral des données acquises par le microscope.

6.2 Application : factorisation tensorielle non-négative

Nous nous intéressons ici à la factorisation de tenseurs non négatifs. Un tenseur est ici noté \mathcal{T} et il correspond à un tableau multi-entrées $(t_{i_1 \dots i_N})_{i_1, \dots, i_N} \in \mathbb{R}^{I_1 \times \dots \times I_N}$. Le nombre d'indices $N \in \mathbb{N}$ correspond à l'ordre du tenseur. Quand à l'indice i_n , il appartient à $\{1, \dots, I_n\}$ avec $I_n \in \mathbb{N}$ pour tout $n \in \{1, \dots, N\}$.

Considérons un tenseur $\overline{\mathcal{T}} = (\overline{t}_{i_1 \dots i_N})_{i_1, \dots, i_N} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ d'ordre N où $\{I_1, \dots, I_N\} \subset \mathbb{N}$, $N \in \mathbb{N}$. La décomposition canonique polyadique (CPD) du tenseur $\overline{\mathcal{T}}$ consiste à le décomposer en une somme minimale de tenseurs de rang 1. Plus précisément, le tenseur $\overline{\mathcal{T}}$ peut être décomposé comme suit

$$\overline{\mathcal{T}} = \sum_{r=1}^{\overline{R}} \overline{\mathbf{a}}_r^{(1)} \circ \overline{\mathbf{a}}_r^{(2)} \circ \dots \circ \overline{\mathbf{a}}_r^{(N)} = \llbracket \overline{\mathbf{A}}^{(1)}, \overline{\mathbf{A}}^{(2)}, \dots, \overline{\mathbf{A}}^{(N)} \rrbracket, \quad (6.10)$$

où \circ est le produit tensoriel. Dans le cas des tenseurs non-négatifs, l'existence d'une solution est assurée [67]. Quand à l'unicité de la solution, elle n'est pas garantie en générale [68, 69] et la solution est obtenue à une permutation et un facteur d'échelle prêt.

On peut par ailleurs introduire les quantités d'intérêt suivantes :

— le vecteur colonne

$$\overline{\mathbf{a}}_r^{(n)} = (\overline{a}_{1r}^{(n)}, \overline{a}_{2r}^{(n)}, \dots, \overline{a}_{I_n r}^{(n)})^\top \in \mathbb{R}^{I_n}$$

est appelé vecteur latent pour tout $n \in \{1, \dots, N\}, r \in \{1, \dots, \overline{R}\}$,

— la matrice

$$\overline{\mathbf{A}}^{(n)} = [\overline{\mathbf{a}}_1^{(n)}, \overline{\mathbf{a}}_2^{(n)}, \dots, \overline{\mathbf{a}}_{\overline{R}}^{(n)}] = (\overline{a}_{i_n r}^{(n)})_{i_n, r} \in \mathbb{R}^{I_n \times \overline{R}}$$

est appelée matrice latente pour tout $n \in \{1, \dots, N\}$,

— Le $\overline{R} \in \mathbb{N}$ minimal tel que l'eq. (6.10) est vérifiée est appelé le rang du tenseur.

La structure même des tenseurs d'ordre N nous incite à considérer N blocs distincts correspondant aux matrices facteurs $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}$ (Eq. 6.10). Afin de procéder à la

décomposition tensorielle, nous avons proposé dans le cadre de la thèse de Xuan Vu, de résoudre un problème d'optimisation régularisé de la forme

$$\underset{\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}, n \in \{1, \dots, N\}}{\text{minimiser}} \quad \mathcal{F}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) + \sum_{n=1}^N \mathcal{R}_n(\mathbf{A}^{(n)}). \quad (6.11)$$

Le terme de fidélité est choisi de manière très classique sous forme quadratique

$$\mathcal{F}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = \frac{1}{2} \|\mathcal{T} - \llbracket \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket\|_F^2 = \frac{1}{2} \|\mathbf{T}_{I_n, I-n}^{(n)} - \mathbf{A}^{(n)} \mathbf{Z}^{(-n)\top}\|_F^2, \quad (6.12)$$

et les termes de pénalisation $(\mathcal{R}_n(\mathbf{A}^{(n)}))_{n \in \{1, \dots, N\}}$ sont définis par

$$\mathcal{R}_n(\mathbf{A}^{(n)}) = \sum_{i_n=1}^{I_n} \sum_{r=1}^R \rho_n(a_{i_n r}^{(n)}) \quad \forall n \in \{1, \dots, N\}, \quad (6.13)$$

où les matrices facteurs sont définies par des éléments $\mathbf{A}^{(n)} = (a_{i_n r}^{(n)})_{(i_n, r) \in \{1, \dots, I_n\} \times \{1, \dots, R\}}$ et où

$$\rho_n(\omega) = \begin{cases} \alpha^{(n)} |\omega|^{\pi^{(n)}} & \text{si } \eta_{\min}^{(n)} \leq \omega \leq \eta_{\max}^{(n)} \\ +\infty & \text{sinon} \end{cases} \quad (6.14)$$

et $\alpha^{(n)} \in]0, +\infty[$, $\pi^{(n)} \in \mathbb{N}^*$, $\eta_{\min}^{(n)} \in [-\infty, +\infty[$ et $\eta_{\max}^{(n)} \in]\eta_{\min}^{(n)}, +\infty]$.

On peut noter que $\rho_n(\cdot)$ peut-être vu comme la combinaison d'une norme $\ell_{\pi^{(n)}}$ et d'une contrainte d'intervalle. Ce choix nous permet à la fois d'assurer la non-négativité des données en choisissant par exemple $\eta_{\min}^{(n)} = 10^{-10}$ et $\eta_{\max}^{(n)} = +\infty$, mais également de favoriser la parcimonie des données en choisissant $\pi^{(n)} = 1$ ou de promouvoir le caractère lisse des matrices en choisissant $\pi^{(n)} = 2$. Nous avons choisi ici des paramètres de régularisation dépendants du bloc mais constant au sein d'un bloc.

Il faut à présent trouver une stratégie d'optimisation pour résoudre le problème de minimisation ainsi défini. Nous avons choisi d'utiliser l'algorithme BC-VMFB [31, 70] où N blocs $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}$ sont considérés. Le schéma de l'algorithme est donné par l'Algorithme 7.

Algorithm 7 Algorithme BC-VMFB pour minimiser Eq. (6.11).

- 1: Soient $\mathbf{A}^{(n)} \in \text{dom } \mathcal{R}_n$, $n \in \{1, \dots, N\}$, $k \in \mathbb{N}$ et $\mu[k] \in]0, +\infty[$ // Initialisation
 - 2: **for** $k = 0, 1, \dots, k_{\max}$ **do** // k -ème itération
 - 3: Choisir un bloc $n \in \{1, \dots, N\}$ // règle quasi cyclique
 - 4: Calculer $\mathbf{P}^{(n)}[k] = \mathbf{P}^{(n)}(\mathbf{A}^{(1)}[k], \dots, \mathbf{A}^{(N)}[k])$ // Construction du préconditionnement
 - 5: Calcul de la matrice de gradient $\nabla_n[k]$ // Calcul du gradient
 - 6: $\tilde{\mathbf{A}}^{(n)}[k] = \mathbf{A}^{(n)}[k] - \mu[k] \nabla_n[k] \oslash \mathbf{P}^{(n)}[k]$ // Étape de gradient
 - 7: $\mathbf{A}^{(n)}[k+1] \in \text{prox}_{\mu[k]-1 \mathbf{P}^{(n)}[k], \mathcal{R}_n}(\tilde{\mathbf{A}}^{(n)}[k])$ // Étape proximale
 - 8: $\mathbf{A}^{(\bar{n})}[k+1] = \mathbf{A}^{(\bar{n})}[k]$ où $\bar{n} = \{1, \dots, N\} \setminus \{n\}$ // Les autres blocs restent inchangés
 - 9: **end for**
 - 10: $\forall n \in \{1, \dots, N\}$, $\hat{\mathbf{A}}^{(n)} = \mathbf{A}^{(n)}[k_{\max}]$ // Convergence atteint à l'itération k_{\max}
-

Lorsque le critère de convergence est atteint, (i.e. après k_{\max} itérations), l'algorithme fournit des estimations $\hat{\mathbf{A}}^{(n)}$ des matrices facteur $\mathbf{A}^{(n)}$ pour tout $n \in \{1, \dots, N\}$. Nous allons voir à présent en détails, chaque étape de l'algorithme.

6.2.1 Choix de la fonction coût et calcul des dérivées partielles

Comme nous pouvons le voir dans le schéma de l'Algorithme 7, la dérivées partielles de \mathcal{F} par rapport à chaque bloc, *i.e.* par rapport à chaque $\mathbf{A}^{(n)}$ pour tout $n = 1, \dots, N$ doivent être calculées.

Lorsque l'on choisi une forme quadratique pour \mathcal{F} , alors le gradient par rapport à chaque $\mathbf{A}^{(n)}$ pour tout $n = 1, \dots, N$, est donné par

$$\nabla_n \mathcal{F}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = -(\mathbf{T}_{I_n, I_n}^{(n)} - \mathbf{A}^{(n)} \mathbf{Z}^{(-n)\top}) \mathbf{Z}^{(-n)}, \quad (6.15)$$

où $\mathbf{Z}^{(-n)}$ est défini par

$$\mathbf{Z}^{(-n)} = \mathbf{A}^{(N)} \odot \dots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \dots \odot \mathbf{A}^{(1)} \in \mathbb{R}^{I_n \times \bar{R}}, \quad (6.16)$$

et \odot désigne le produit de Khatri-Rao.

Afin de simplifier les notations, on notera la matrice de gradient $\nabla_n \mathcal{F}(\mathbf{A}^{(1)}[k], \dots, \mathbf{A}^{(N)}[k])$ comme $\nabla_n[k]$ où k représente le numéro de l'itération.

6.2.2 Calcul du préconditionnement et des opérateurs proximaux associés

On voit également dans le schéma de l'algorithme que les opérateur proximaux par rapport à chaque terme de régularisation $(\mathcal{R}_n(\mathbf{A}^{(n)}))_{n \in \{1, \dots, N\}}$ associés à la métrique $\mathbf{P}^{(n)}$ doivent être calculés. On voit ici qu'un autre point clé de l'algorithme est le design de bon préconditionneurs permettant d'accélérer la vitesse de convergence. Dans ce travail, nous nous sommes inspirés des travaux [71] et [33], et avons généraliser aux tenseurs le préconditionnement proposer dans ce dernier pour les matrices. Cela a été possible en procédant à des dépliements de tenseur suivant le mode n .

Ces calculs nous ont amené au résultats suivant : la matrice de préconditionnement \mathbf{P} associée au bloc n peut-être définie par

$$\mathbf{P}^{(n)}(\mathbf{A}^{(1)}[k], \dots, \mathbf{A}^{(N)}[k]) = \mathbf{A}^{(n)} (\mathbf{Z}^{(-n)\top} \mathbf{Z}^{(-n)}) \oslash \mathbf{A}^{(n)}, \quad \forall n \in \{1, \dots, N\}, \quad (6.17)$$

où \oslash représente la division de Hadamard (terme à terme).

On peut à présent calculer les opérateurs proximaux associés à chaque terme de régularisation $(\mathcal{R}_n)_{n \in \{1, \dots, N\}}$ et correspondant donc au calcul de $\text{prox}_{\mu[k]-1\mathbf{P}^{(n)}[k], \mathcal{R}_n}(\tilde{\mathbf{A}}^{(n)}[k])$. Leurs calculs demandent une vectorisation des données³ dont nous définissons les éléments clés ici

- le n -th bloc estimé à l'itération k correspondant donc à la n -ème matrice facteur $\tilde{\mathbf{A}}^{(n)}[k]$ est stocké dans un vecteur noté $\tilde{\mathbf{a}}^{(n)}[k] \in \mathbb{R}^{RI_n \times 1}$.
- chaque élément de la matrice de préconditionnement $\mathbf{P}^{(n)}[k]$ in (6.17) est noté $(p_i^{(n)}[k])_{i \in \{1, \dots, RI_n\}}$ et l'on construit alors une matrice diagonale de taille $RI_n \times RI_n$ $\tilde{\mathbf{P}}^{(n)}[k]$ as $\tilde{\mathbf{P}}^{(n)}[k] = \text{Diag}(p_1^{(n)}[k], \dots, p_{RI_n}^{(n)}[k])$.

Sous ces hypothèses, calculer $\text{prox}_{\mu[k]-1\mathbf{P}^{(n)}[k], \mathcal{R}_n}(\tilde{\mathbf{A}}^{(n)}[k])$ consiste à calculer $\text{prox}_{\mu[k]-1\tilde{\mathbf{P}}^{(n)}[k], \mathcal{R}_n}(\tilde{\mathbf{a}}^{(n)}[k])$ dont la définition est donnée dans (4.16).

3. Non effectué en pratique car trop coûteux et sous-optimal par rapport à des calculs terme à terme

Ensuite, la régularisation étant séparable et les matrices de préconditionnement $\tilde{\mathbf{P}}^{(n)}[k]$ diagonales, l'opérateur proximal s'écrit simplement

$$\text{prox}_{\mu[k]^{-1}\mathbf{P}^{(n)}[k], \mathcal{R}_n}(y) = \left(\text{prox}_{\mu[k]^{-1}p_i^{(n)}[k], \rho_n}(y^{(i)}) \right)_{i \in \{1, \dots, RI_n\}} \quad (6.18)$$

$$(\forall y = (y^{(i)})_{i \in \{1, \dots, RI_n\}} \in \mathbb{R}^{RI_n}).$$

Pour tout $i \in \{1, \dots, RI_n\}$, nous avons [10] ($\forall v \in \mathbb{R}$)

$$\text{prox}_{\mu[k]^{-1}p_i^{(n)}, \rho_n}(v) = \min \left\{ \eta_{\max}^{(n)}, \max \left\{ \eta_{\min}^{(n)}, \text{prox}_{\mu[k]\alpha^{(n)}(p_i^{(n)}[k])^{-1} \cdot |\cdot|^{\pi(n)}}(v) \right\} \right\}. \quad (6.19)$$

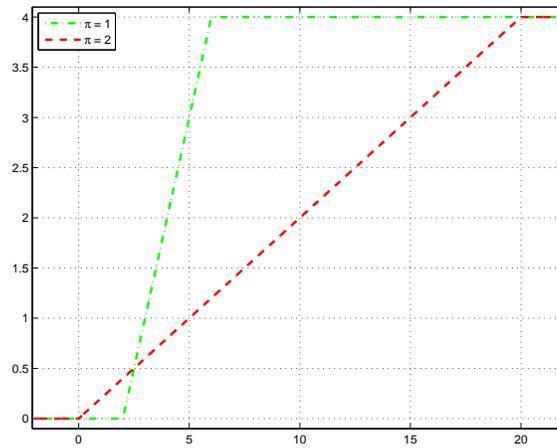


FIGURE 6.6 – Courbes de $\text{prox}_{\rho_n}(v)$ en fonction de $v \in [-2, 22]$ lorsque $\alpha^{(n)} = 2$ et $[\eta_{\min}^{(n)}, \eta_{\max}^{(n)}] = [0, 4]$.

L'opérateur proximal de ρ_n est illustré Fig. 6.6 pour des paramètres valant $\alpha^{(n)} = 2$, et $[\eta_{\min}^{(n)}, \eta_{\max}^{(n)}] = [0, 4]$. Comme désiré, les valeurs négatives sont projetées à zéro et lorsque l'on considère une norme ℓ_1 , on reconnaît l'effet de seuillage.

6.2.3 Simulations numériques

Nous nous sommes dans un premier temps intéressés à des données 3D mais aussi 4D (réalistes, afin de reproduire au mieux les signaux observés dans le cadre de la spectroscopie de fluorescence) simulées afin d'analyser le comportement de l'algorithme proposé. Bien que de nombreuses configurations aient été étudiées, observation sans bruit (modèle exacte), observation bruitée (bruit additif gaussien) et surestimation du rang du tenseur, je ne vais montrer ici que quelques résultats : données 3D avec surestimation du rang et données 4D, bruitée et surestimation du rang.

Les performances de l'algorithme ont été évaluées visuellement mais aussi selon plusieurs critères numériques :

— Erreur d'estimation E1

$$E1(\sigma) = \frac{\sum_{n=1}^N \|\widehat{\mathbf{A}}^{(n)\sigma}(1 : \bar{R}) - \mathbf{A}^{(n)}\|_1}{\sum_{n=1}^N \|\mathbf{A}^{(n)}\|_1}$$

$$\Rightarrow \begin{cases} E1 = \min_{\sigma} E1(\sigma) & \text{ou} & E1 = 10 \log_{10}(E1) \\ \sigma_{opt} = \arg \min_{\sigma} E1(\sigma) \end{cases} \quad (6.20)$$

— Erreur de surestimation E2

$$E2 = \left\| \sum_{r=\bar{R}+1}^R \mathbf{a}_{\sigma_{opt},r}^{(1)} \circ \dots \circ \mathbf{a}_{\sigma_{opt},r}^{(N)} \right\|_1 \quad \text{ou} \quad E2 = 10 \log_{10}(E2). \quad (6.21)$$

— Rapport Signal à Bruit (SNR)

$$\text{SNR} = 20 \log_{10} \frac{\|\bar{\mathcal{T}}\|_F}{\|\bar{\mathcal{T}} - \widehat{\mathcal{T}}\|_F} \quad (6.22)$$

où $\widehat{\mathcal{T}}$ représente le tenseur estimé.

6.2.3-a) Simulations sur données synthétiques 3D

Sont donc considérées ici des données synthétiques 3D (i.e. $N = 3$) imitant les données de spectroscopie de fluorescence. Les données ainsi générées $\bar{\mathcal{T}}$ sont construites comme suit : $\bar{\mathcal{T}}$ est de dimension $100 \times 100 \times 100$ (i.e. $I_1 = I_2 = I_3 = 100$). Les spectres d'excitation et d'émission sont créés en utilisant des sommes de densité (décalées) obtenues à partir d'une distribution normale généralisée tronquée de manière à assurer la positivité des données. Les concentrations sont générées aléatoirement selon une loi uniforme entre 0 et 10. Le rang du tenseur \bar{R} est fixé égal à 5. Lorsque l'on simule une surestimation du rang, cela veut dire que l'on considère que le rang est égale à 6 (i.e. $R = 6$).

Nous comparons les résultats de notre algorithme avec deux algorithmes de l'état de l'art : le N -way de Bro [72] et le fast Hierarchical Alternating Least Squares (HALS) de [73].

Voici les paramètres que nous avons utilisés dans nos simulations : $\eta_{\min}^{(n)} \equiv \eta_{\min} = 2e^{-16}$ (non négativité) et $\eta_{\max}^{(n)} \equiv \eta_{\max} = 1000$. La valeur de l'exposant $\pi^{(n)}$ est choisi par maximum de vraisemblance [74, p. 225] pour chaque bloc ce qui revient ici à le considérer égal à 1 pour $n = 1, 2$ et 2 pour $n = 3$. Le pas $\mu[k] \equiv \mu$ a été fixé à 0.99. Le paramètre de régularisation $\alpha^{(n)} \equiv \alpha$ a été fixé empiriquement à $\alpha = 0.05$ pour tous les blocs.

Nous considérons dans un premier temps le cas 3D sans bruit, c'est à dire $\mathcal{T} = \bar{\mathcal{T}}$ mais en considérant que $R = 6$. Les résultats (MEEF estimés) sont donnés Fig. 6.7.

Ce que l'on constate est que la version régularisée de l'algorithme permet de calculer correctement les matrices facteurs et détecte très bien l'absence de la 6ème composante. Ce n'est pas le cas lorsque la régularisation est omise ou pour l'algorithme N -way.

6.2.3-b) Simulations sur données synthétiques 4D

La quatrième dimension (venant donc s'ajouter aux dimensions spectres (excitation + émission) et concentrations) est obtenue en ajoutant un temps de vie aux composés chimique.

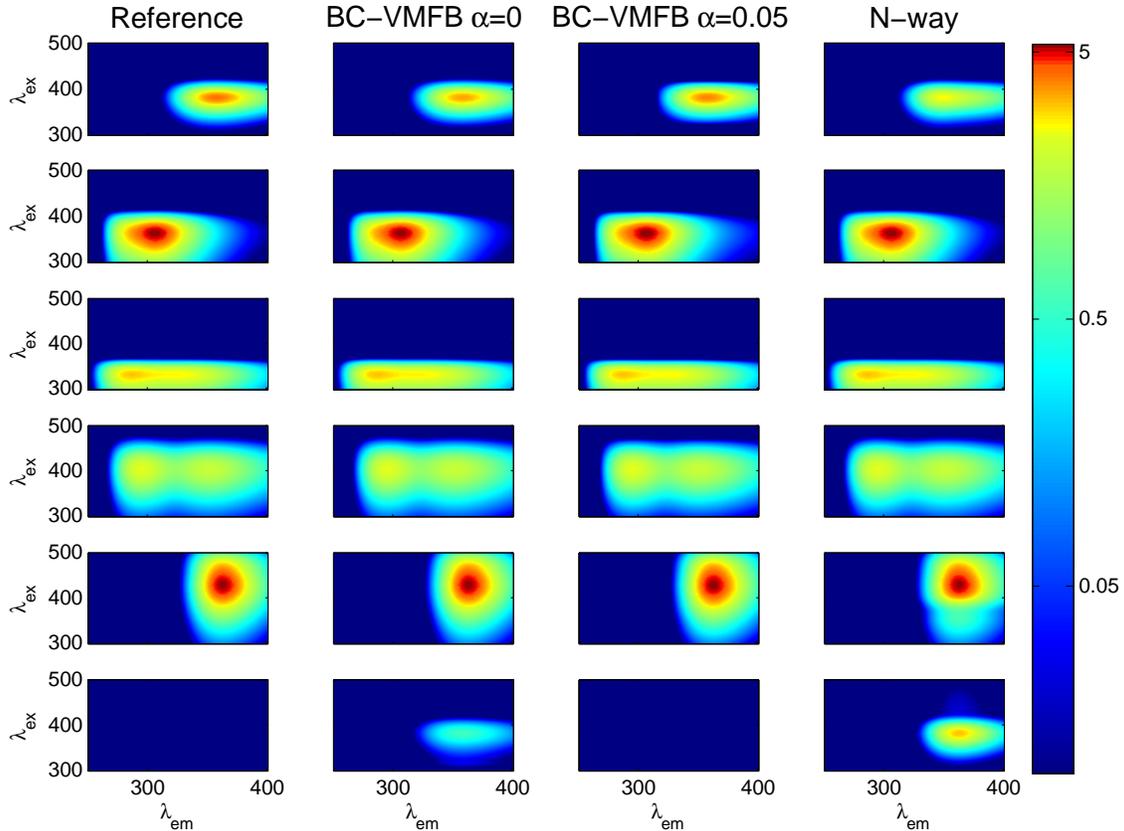


FIGURE 6.7 – MEEF de référence et MEEF reconstruits avec N -way ($E_1 = -8.7$ dB, $E_2 = 31.7$ dB) et l’algorithme proposé BC-VMFB sans régularisation $\alpha = 0$ ($E_1 = -12.3$ dB, $E_2 = 25.6$ dB) et avec régularisation $\alpha = 0.05$ ($E_1 = -15$ dB, $E_2 = -409$ dB).

Comme pour les données 3D, le tenseur 4D est construit en simulant des spectres d’émission et d’excitation de type uni ou bi-modal, des concentrations linéaires ou uni-modales et enfin, un temps de vie, décroissant au cours du temps [75]. Le rang du tenseur est fixé à $\bar{R} = 5$. Le tenseur résultant $\bar{\mathcal{T}}$ est de taille $I_1 = I_2 = I_3 = I_4 = 100$ et $\bar{\mathcal{T}} \in \mathbb{R}_+^{100 \times 100 \times 100 \times 100}$.

On considère ici que le tenseur observé est une version perturbée du tenseur original \mathcal{T} et que donc $\mathcal{T} = \bar{\mathcal{T}} + \mathcal{B}$ où \mathcal{B} est un bruit additif blanc gaussien avec $\sigma = 0.001$ conduisant à un SNR initial de 18.46 dB. De plus, le rang du tenseur est supposé être égal à 7 ($R = 7$, surestimation d’un facteur 2).

De nouveau, nous comparons notre approche (avec $\pi^{(n)} \equiv \pi = 1$ et $\alpha^{(n)} \equiv \alpha = 0.05$) aux deux méthodes de référence : 1) l’algorithme fast HALS [73] et 2) l’algorithme N -way [72]. Toutes les approches sont itératives et donc des critères d’arrêt doivent être considérés. Ce sont les suivant : le nombre maximum d’itérations pouvant être produites est fixé à $k_{\max} = 10^5$ et l’algorithme peut être stoppé avant si $\frac{\|\mathcal{F}^{[+K]} - \mathcal{F}^{[.]}\|}{\mathcal{F}^{[.]}} < 10^{-8}$, où $K = 1$ pour les algorithmes fast HALS et N -way et $K = 500$ pour le BC-VMFB.

Les spectres estimés sont représentés Fig. 6.8 et les MEEF estimés Fig. 6.9. On constate que malgré le bruit et la surestimation, la méthode proposée permet de retrouver correctement les données originales ce qui n’est pas le cas des approches fast HALS ou N -way. Tout ceci est confirmé numériquement par les calculs d’erreurs reportées dans le

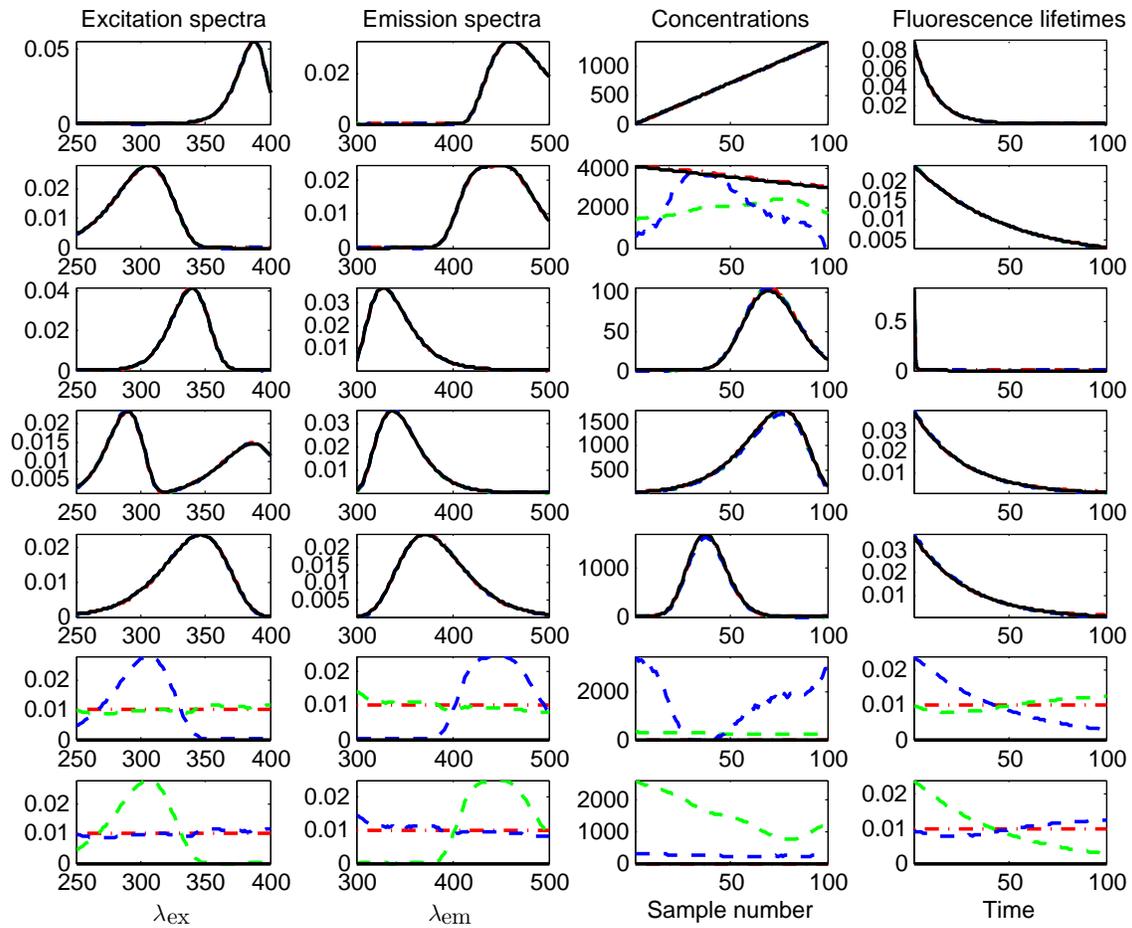


FIGURE 6.8 – Spectres (normalisés) estimés par *N*-way (bleu), fast HALS (vert) et BC-VMFB (rouge). La vérité terrain est en noir.

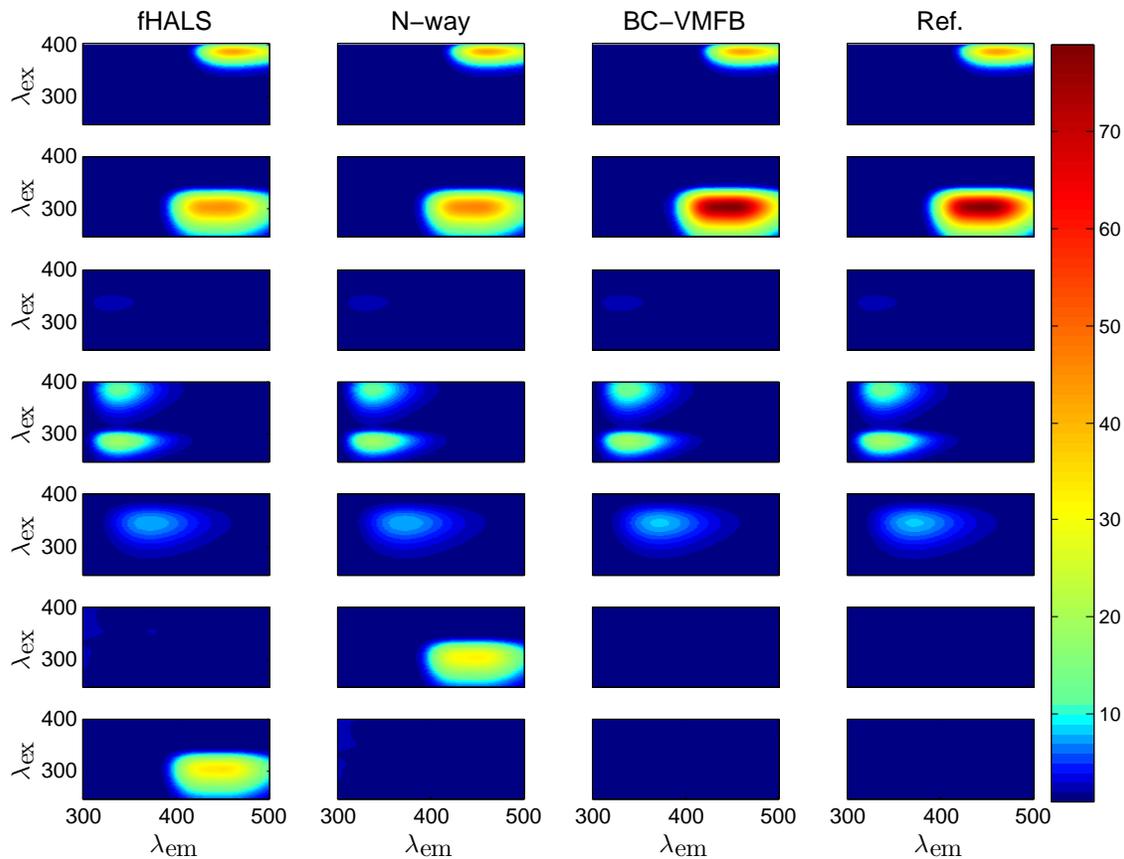


FIGURE 6.9 – MEEF estimés avec N -way, fast HALS, BC-VMFB et les MEEF de référence.

tableau 6.1 où l'on constate que l'erreur d'estimation est plus faible pour notre approche tout comme l'erreur de surestimation. Concernant le temps de calcul, l'algorithme reste compétitif malgré le fait que plus d'itérations soient nécessaires pour atteindre le critère d'arrêt.

Une question que l'on peut se poser est : qu'en est-il de l'influence de l'initialisation des algorithmes ? C'est ce que nous avons cherché à illustrer dans la figure 6.10, où les erreurs (E_1 et E_2) obtenues pour notre approche ainsi que pour le N -way et le fast HALS, pour 30 initialisations différentes. Notre approche mène aux meilleures performances dans 90% des cas que ce soit pour l'erreur d'estimation (mesurée au travers de E_1) ou de la détection des composants non existants (mesurée au travers de E_2). Dans les 10% restant, notre approche mène à des résultats équivalents à ceux obtenus pour le N -way

	fast HALS	N -way	BC-VMFB
Temps (en s) pour 50 itérations	15.4	251	9.7
Temps (en s) pour atteindre le critère d'arrêt	139	554	1689
Nb d'itérations pour atteindre le critère d'arrêt	506	176	17000
(SNR, E_1 , E_2) en dB	(29.61, 0.21, 47.74)	(29.61, 0.11, 46.61)	(34.15, 0.11, -543)

TABLE 6.1 – Temps de calculs et performances numériques des algorithmes fast HALS, N -way et BC-VMFB en utilisant la même initialisation.

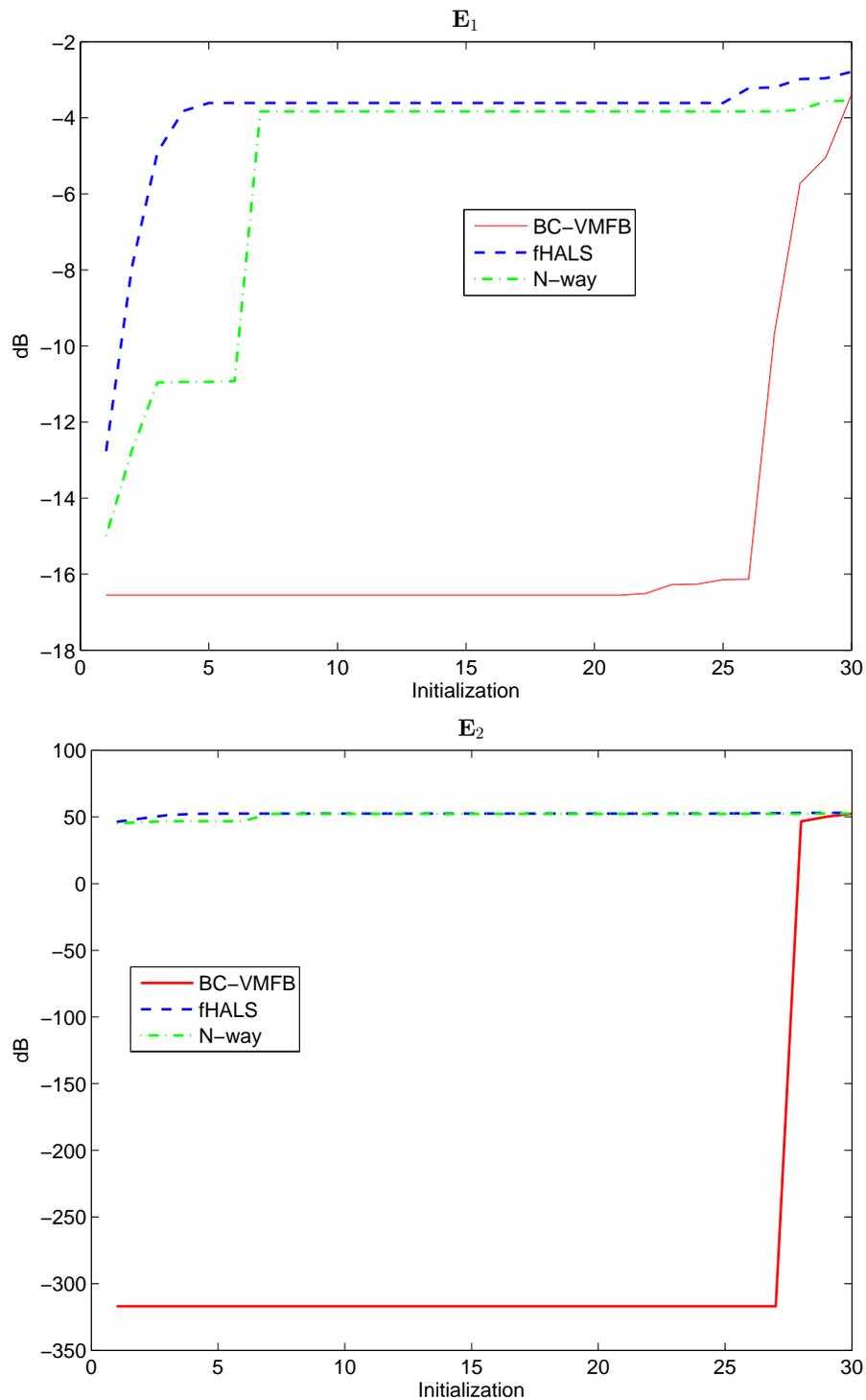


FIGURE 6.10 – Erreur d'estimation E_1 (haut) et erreur de surestimation E_2 (bas) en utilisant l'algorithme proposé (ligne rouge pointillé), le N -way (ligne verte point-pointillé), et le fast HALS (ligne bleu tirets).

et le fast HALS.

6.2.3-c) Application au monitoring de l'eau

L'objectif de cette dernière partie est de tester notre algorithme sur données réelles. Ces données sont souvent perturbées par des bruits aléatoires mais aussi des bruits liés à l'acquisition : problèmes de calibration, saturation, données manquantes, ce qui éloigne les données du modèle considéré. De plus, Rayleigh and Raman's scattering requiert que les données soient prétraitées. Bien sur, la vérité terrain est inconnue, et donc, le nombre de composés présents dans la solution est lui aussi inconnu.

Avant de présenter les résultats obtenus par la méthode proposées ici, nous allons décrire brièvement les données que nous avons traitées.

Les données réelles considérées ici sont celles introduites dans [76]. Les données ont été acquises automatiquement toutes les 3 minutes, durant les 10 jours d'une campagne d'étude de la qualité de l'eau d'une rivière. La taille des données ainsi mesurées est $36 \times 111 \times 2594$. Les longueurs d'onde d'excitation vont de 225nm à 400nm avec un pas de 5nm, alors que les longueurs d'ondes d'émission s'étalent de 280nm à 500nm avec un pas de 2nm. On sait qu'une pollution survient au cours du 7ème jour de campagne.

Les MEEF ont été prétraités par la méthode de Zepp [77] afin d'éliminer les effets de Rayleigh et Raman. Toutes les valeurs négatives (soit 0.34% des données) ont été mises à 0.

Afin de rendre les résultats plus lisibles, nous ne prenons pas en compte les 1200 premiers MEEF où il n'y a pas de changement notable. La taille des données finalement traitées est $36 \times 105 \times 1392$.

Comme le nombre de composés chimiques présents est inconnu, plusieurs valeurs du rang ont été testées : $R = 4/6$.

Concernant maintenant les paramètres de la méthode, une régularisation de type ℓ_1 ($\pi^{(n)} \equiv \pi = 1$) a été appliquée sur chaque matrice facteur. Lorsque $R = 4$ (resp. $R = 6$), on a choisi un paramètre de régularisation $\alpha^{(n)} \equiv \alpha = 2 \times 10^4$ (resp. $\alpha = 8 \times 10^3$). L'algorithme N -way incluant des contraintes de non-négativité est testé dans les mêmes conditions que la méthode proposée.

Les MEEF reconstruits par les méthodes sont données dans les figures Fig. 6.11 (resp. 6.13) considérant que $R = 4$ (resp. $R = 6$). Les concentrations normalisées sont affichées Fig. 6.12 pour $R = 4$, et dans Fig. 6.14 pour $R = 6$.

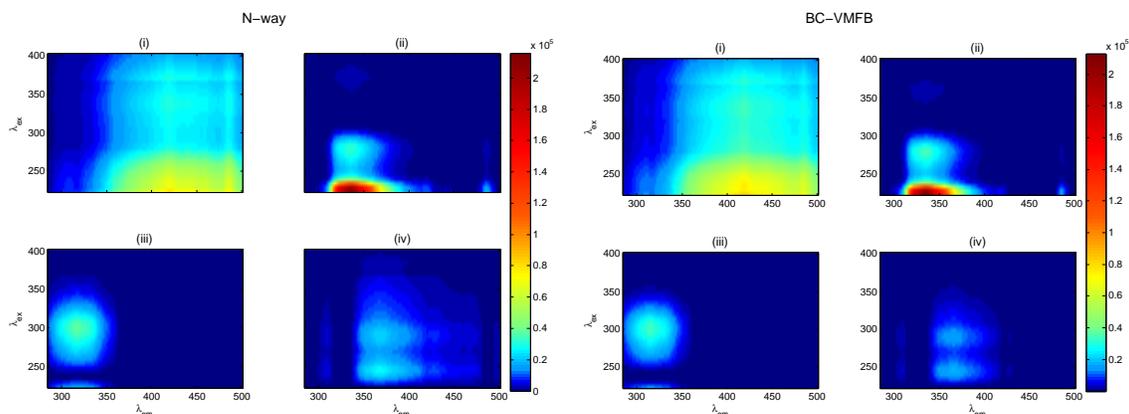


FIGURE 6.11 – MEEF estimés avec l'algorithme N -way (gauche), et l'algorithme BC-VMFB (droite), cas $R = 4$.

Les résultats deviennent assez différents lorsque $R = 6$, spécifiquement pour les composés (i), (v) et (vi) (voir Fig. 6.13, 6.14). On voit aussi que les composés obtenus à $R = 4$ avec le BC-VMFB sont plus ou moins préservés pour $R = 6$. La méthode non pénalisée tend à diviser le composé (i) obtenus pour $R = 4$ en les composés (i), (v) et (vi) dans le cas $R = 6$; les MEEF et les concentrations des autres composés (e.g. (iii) et (iv)) sont également un peu déformés. Ces phénomènes arrivent également lorsque les paramètres de régularisation sont mal choisis.

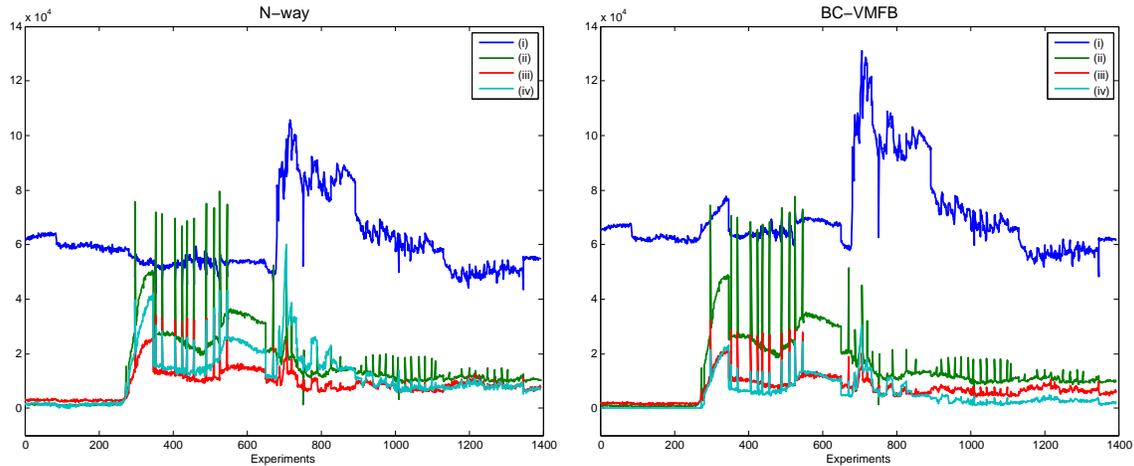


FIGURE 6.12 – Concentrations normalisées estimées avec le N-way (gauche) et le BC-VMFB (droite), cas $R = 4$.

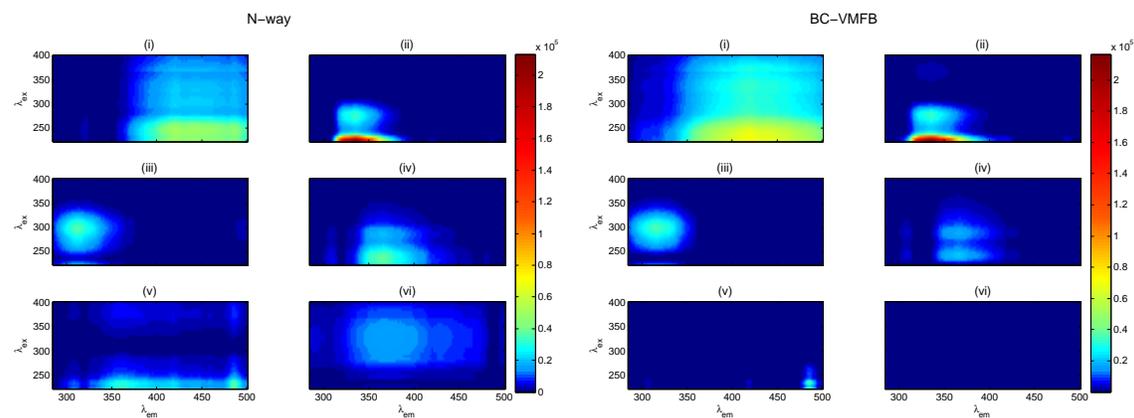


FIGURE 6.13 – MEEF estimés avec le N-way (gauche) et le BC-VMFB (droite), cas $R = 6$.

On peut noter par ailleurs des différences significatives des concentrations estimées pour les composés (i) et (iv) par l'algorithme BC-VMFB ou par l'algorithme N-way.

En conclusion, en regardant les résultats obtenus pour $R = 4$ et $R = 6$, on peut conclure que les estimations obtenues avec l'approche régularisée paraissent plus fiables car plus stables. Ceci étant dit, les paramètres de régularisation ont été vraiment difficile à fixer parfois. L'algorithme BC-VMFB reste toutefois intéressant car rapide et donc apte à tester plusieurs paramètres de régularisation. On aurait tendance à conclure que le rang du tenseur vaut 4 (les MEEF estimés des composés (v) et (vi) dans le cas $R = 6$

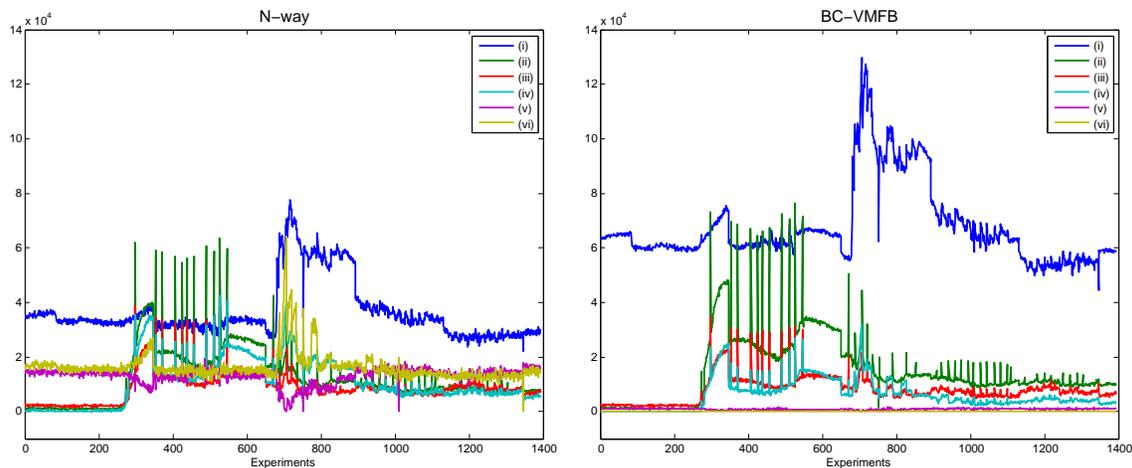


FIGURE 6.14 – Concentrations normalisées estimées avec le N-way (gauche) et le BC-VMFB (droite), cas $R = 6$.

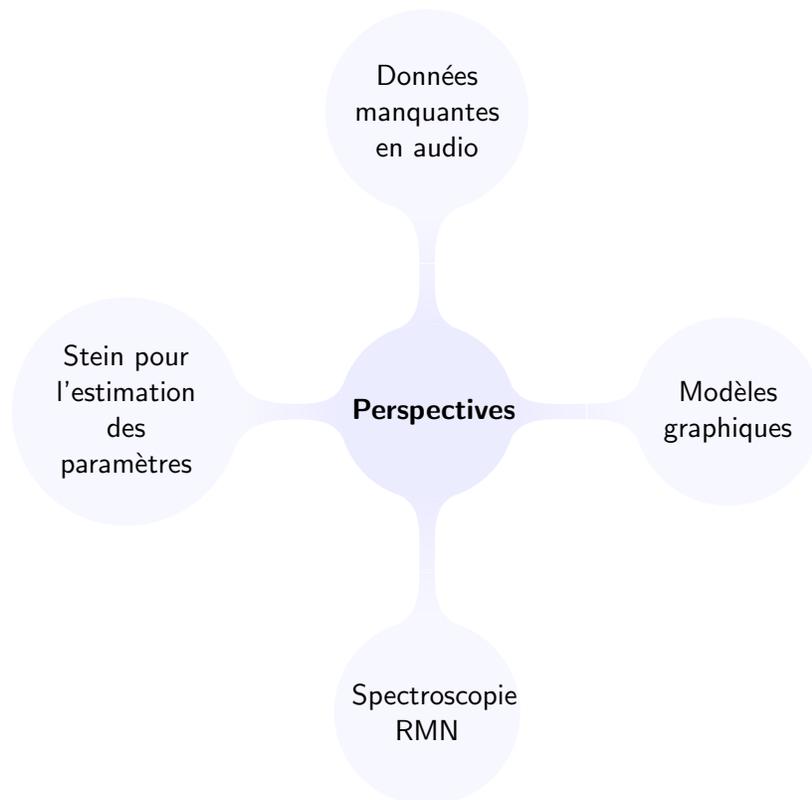
tout comme leurs concentrations sont proches de zéro, voir partie droite de la Fig. 6.13 et la partie basse de la Fig. 6.14). On peut aussi affirmer qu'une pollution a bien lieu au 7ème jour de campagne si l'on regarde attentivement les concentrations (e.g. Fig. 6.11). En effet, un hydrocarbure [78] a été clairement identifié. Avant l'échantillon 1480, un seul composé chimique est principalement présent (composé (i) que ce soit pour $R = 4$ ou $R = 6$) alors que deux autres sont présents en trace (composés (ii) et (iii)). Après ce 1480-ème échantillon, un quatrième composé fait irruption (composé (iv) que ce soit pour $R = 4$ ou $R = 6$) et l'on observe une augmentation importante des concentrations des deux composés précédemment présentés. Les 3 courbes de concentrations semblent suivre le même comportement.

6.3 Quelques publications personnelles en lien avec ce chapitre

- [CCo1] E. Chouzenoux, L. Lamassé, S. Anthoine, C. Chaux, A. Jaouen, I. Vanzetta, and F. Debarbieux. Approche variationnelle pour la déconvolution rapide de données 3D en microscopie biphotonique. In *Proc. GRETSI*, Lyon, France, 8-11 Sep. 2015.
- [CCo2] S. Cadoni, E. Chouzenoux, Pesquet J.-C., and C. Chaux. A block parallel majorize-minimize memory gradient algorithm. In *Proc. Int. Conf. Image Process.*, Phoenix, USA, 25-28 Sep. 2016.
- [CCo3] X. T. Vu, C. Chaux, N. D. Thirion-Moreau, S. Maire, and E. M. Carstea. A new penalized nonnegative third order tensor decomposition using a block coordinate proximal gradient approach : application to 3d fluorescence spectroscopy. *J. Chemometrics*, 31(4), Apr. 2017.

Partie III

Conclusions et perspectives



Chapitre 7

Conclusion

Avant-propos

Nous dressons ici un rapide bilan des travaux accomplis, source de nouvelles recherches, de collaborations, d'encadrements.

7.1 Conclusions

Durant ces dix années de recherche, la quête d'informations d'intérêt à partir d'acquisitions physiques de données de natures diverses, a constitué le fil conducteur de mes travaux. C'est le domaine des problèmes inverses qui englobe un nombre considérable d'applications dans des domaines variés comme l'imagerie médicale, la microscopie, l'imagerie satellitaire, la chimie séparative, l'audio, ... Cette quête m'a amenée à l'étude puis à la résolution de problèmes inverses, qui, comme nous l'avons vu, requiert de nombreuses attentions quant à sa formalisation (modèle physique de formation des données, parfois complexe), sa paramétrisation (paramètres régissant le modèle ou les méthodes mises en œuvre) et sa résolution (que ce soit les méthodes ou les outils employés et leurs applications aux données réelles). J'ai à cœur de développer des projets du début à la fin, depuis la modélisation à la résolution du problème et à l'application des méthodes aux données réelles. Même si, mes projets ont souvent commencé par une étude théorique, indépendante d'une application réelle, nombreux d'entre eux ont mené à des applications souvent conduites en collaboration avec des industriels. Cela a été le cas avec Thalès Alénia Space, pour lequel un capteur spécifique (le push broom) et une technique particulière d'acquisition (le flutter-shutter) ont été étudiés. Cela a également été le cas lors de développement de méthodes de déconvolution pour la macroscopie confocale en collaboration notamment avec les biologistes de l'INRA et les physiciens/opticiens de l'université de Mulhouse mais également lors de développement de méthodes de reconstruction en tomographie par émission de positrons avec le SHFJ d'Orsay. Il y a également tous les travaux entrepris avec l'IFP Energies Nouvelles, pour lesquels de nombreux outils ont été développés pour l'analyse des données de sismique. Une continuité s'est naturellement opérée à mon arrivée à Marseille où j'ai rapidement été contactée par des biologistes de Luminy afin de les éclairer sur les méthodes de déconvolution en microscopie (e.g. feuille

de lumière, biphoton). Cela a été également l'occasion d'élargir ce champ de recherche à d'autres domaines requérant l'appréhension et la conception de nouvelles méthodologies comme la décomposition tensorielle pour laquelle de récentes méthodes proximales ont été introduites et appliquées avec succès au monitoring de l'eau (détection de pollution), ou l'étude des modèles graphiques afin de modéliser les connexions cérébrales et d'effectuer des tâches de classification de patients atteints de différentes maladies neurodégénératives (appliqués à la fibromyalgie), ou encore la séparation de sources dans le domaine de la chimie analytique (appliquée à la métabolomique et aux biocarburants).

La théorie des ondelettes et l'optimisation convexe sont pour moi les deux domaines de recherche dans lesquels j'évolue. Les algorithmes proximaux font preuves d'une grande flexibilité et permettent de résoudre une large classe de problèmes d'optimisation convexes ou non. Le choix de l'algorithme à utiliser est important tant sur le point du problème d'optimisation à résoudre que sur le point de l'efficacité numérique de l'algorithme (temps de calculs, vitesse de convergence, ...). Les transformées en ondelettes ou plus généralement les trames fournissent de précieux outils d'analyse permettant souvent de mieux modéliser les données dans un domaine transformé, domaine dans lequel un a priori est plus facilement formulé. Les deux s'allient parfaitement et conduisent à divers problèmes de recherche d'intérêts comme la formulation du problème à l'analyse ou à la synthèse, le choix d'une bonne représentation pour un problème donné, le choix d'un a priori adapté et la résolution du problème associé, ...

Mais cela n'est pas tout. Avant même de mettre en place ces méthodes, aussi faut-il formuler le problème dans sa globalité en commençant simplement par la formulation du problème direct qui n'est pas toujours immédiate ni même simple à formuler. C'est à ce moment là que les échanges entre les physiciens, médecins, biologistes, chimistes et nous prennent tout leur sens. Ce dialogue est essentiel et s'avère souvent compliqué car il requiert l'élaboration d'un vocabulaire commun mais il est la source même de la formalisation du problème. Que ce soit le modèle direct ou le problème d'optimisation, les deux introduisent des paramètres de natures différentes (paramètres de bruit, paramètres de régularisation, paramètres d'algorithmes (i.e. pas)) qui sont primordiaux à déterminer, ceci dans le but de non seulement conduire à de bonnes solutions mais aussi d'emprunter le meilleur chemin qui soit.

Cette chaîne complète de traitement m'a permis de considérer différents outils, d'explorer différents domaines de recherche et m'a également offert l'opportunité de nombreuses collaborations, qui ont enrichi mes travaux. Cela a été également l'opportunité d'encadrer des doctorants, une riche expérience également.

Mais tout ne s'arrête pas là, beaucoup reste encore à étudier. Je propose de dresser dans le chapitre suivant une feuille de route, dans un premier temps assez générale et portant sur le long terme, puis dans un second temps, j'en profiterais pour détailler quelques pistes de recherche en cours d'exploration ou à explorer dans un futur proche.

Chapitre 8

Perspectives

Avant-propos

Ce dernier chapitre donne une ouverture à ces travaux et dresse quelques pistes plus précises pour de futurs travaux de recherche. Certaines ont déjà parcouru un certain chemin, d'autres apparaissent plus prospectives, offrant de nouvelles collaborations, de nouveaux champs d'applications et de nouveaux outils à explorer.

8.1 La suite

L'écriture de ce manuscrit a été l'occasion de faire le point sur mon parcours. Il marque une étape dans mon cheminement et me permet à présent de faire un tour d'horizon. Le paysage marseillais change, les laboratoires fusionnent, les parcours universitaires s'offrent un nouvelle jeunesse, des idex se créent, des instituts convergences ont vu le jour dont l'Institut Langage Communication Cerveau (ILCB). Tous ces changements offrent de nouvelles opportunités dans le monde de la recherche proposant, entre autre, de nouvelles collaborations, de nouveaux appels à projet. Il n'est pas toujours aisé de trouver sa place dans ce monde en perpétuel mouvement mais je pense qu'en faisant preuve d'ouverture et de flexibilité, beaucoup de choses peuvent être faites. Car non seulement l'environnement change mais de manière plus générale, la société et les technologies évoluent, apportant elles aussi leur lot de questions et leur lot de changements.

Je suis de plus en plus sensibilisée à la recherche reproductible et souhaite mettre en œuvre autant que faire se peut, cette reproductibilité. Cela demande un gros travail de développement, mais je trouve maintenant cet aspect de la recherche essentiel. Des ingénieurs de recherche en développement sont là pour nous accompagner dans cette démarche et nous aider à améliorer nos développements au quotidien. Cela concerne également les projets pluridisciplinaires pour lesquels les développements doivent être accessibles et utilisables par tous, demandant parfois quelques ajustements.

Je n'imagine pas l'avenir sans inversion, qui n'a jamais rêvé d'inverser le temps ? Je souhaite continuer à explorer d'autres domaines de recherches mais je souhaite également revenir à des recherches plus fondamentales, notamment en ce qui concerne les ondelettes. En effet, d'une manière plus générale, je souhaiterais étudier de nouvelles

représentations de données car elles jouent souvent un rôle important. Elles permettent par exemple de prendre en compte plus facilement certaines propriétés des données ou encore permettent d'introduire de la parcimonie. Je pense en effet à la transformée en ondelettes mais aussi aux transformées temps-fréquence, qui ont sans cesse été adaptées ou étendues à diverses problématiques. Prendre le temps d'étudier des outils d'analyse fait partie des axes de recherche que je souhaite privilégier. Un sujet connexe qui m'intéresse également est l'apprentissage de dictionnaire, qui allie à la fois, cet aspect analyse mais aussi l'aspect apprentissage qui apparaît incontournable et transversal à notre discipline. Par ailleurs, je souhaiterais également accorder plus d'importance à l'adéquation algorithme/architecture, côté que j'ai délaissé ces dernières années, mais qui conduit à de nouveaux formalismes (e.g. stochastique, asynchrone) et/ou à de nouveaux algorithmes (e.g. par blocs, parallèles). Certes les volumes de données augmentent, leur nature change (e.g. graphes) mais les moyens de calculs se modernisent eux aussi et je pense qu'il est opportun d'en tirer parti. Je pense qu'il y a des champs disciplinaires où tous ces développements seraient opportuns notamment ceux portés par l'ILCB par exemple comme l'imagerie médicale (nouveaux imageurs) ou plus généralement les neurosciences (magnétoencéphalographie (MEEG), ...).

J'ai démarré de nouveaux travaux il y a peu et je donne ci-après quelques développements nouveaux actuels.

8.2 Quelques directions en cours d'étude ou de développement

8.2.1 Données manquantes en audio

Dans le cadre de l'ANR MAD portée par Valentin Emiya et traitant de reconstruction de données audio manquantes, j'ai été confrontée à de nouveaux problèmes d'optimisation dans un contexte (audio) tout à fait nouveau pour moi. Un des objectifs était de faire un pont entre les techniques d'inpainting d'images et les futures techniques d'inpainting audio. Comme évoqué précédemment, j'ai beaucoup travaillé avec les outils multirésolution mais jamais avec les outils temps-fréquences tels que la transformée de Fourier à fenêtre (TFCT). Or le problème d'optimisation que j'ai étudié dans un premier temps était le suivant : soit $x \in \mathbb{R}^N$ un signal audio 1D et $\mathbf{X} = \Phi x \in \mathbb{C}^{F \times T}$ sa représentation via une TFCT,

$$\underset{\mathbf{X} \in \mathbb{C}^{F \times T}}{\text{minimiser}} \frac{1}{2} \|\mathbf{M} \odot (\mathbf{X} - \mathbf{Y})\|^2 + \lambda_1 \|\mathbf{X}\|_* + \frac{\lambda_2}{2} \|\mathbf{A}\mathbf{X}\|_2^2 \quad (8.1)$$

où \mathbf{M} est un masque binaire, $\mathbf{A} = \mathbf{I} - \Phi\Phi^*$ et \mathbf{Y} sont les observations. On suppose ici que Φ représente une trame ajustée avec une constante de trame égale à 1 (i.e. telle que $\Phi^*\Phi = \mathbf{I}$).

Pourquoi ce critère ? Le premier terme représente le terme de fidélité au données. Le deuxième terme représente un terme d'a priori sur les coefficients de TFCT et le dernier terme est un terme de consistance [79] qui assure que les \mathbf{X} recherchés correspondent effectivement à des coefficients de TFCT. Car oui, on peut à ce point remarquer que l'inpainting ne se fait pas dans le domaine temporel mais bien dans le domaine temps-fréquence.

Un algorithme de type Explicite-Implicite peut être mis en œuvre pour résoudre le problème 8.1 et les premières expériences ont montré des résultats partiellement satisfaisants. Comme nous avons pu le voir au cours de ce manuscrit, les choix des a

priori sont cruciaux pour obtenir de bonnes estimations. Or, l'a priori formulé ici est une contrainte de rang faible sur les coefficients de TFCT, a priori effectué dans de nombreux travaux sur le spectrogramme (c'est à dire sur le module ou module carré de X).

La question qui suit est naturellement : les coefficients de TFCT sont ils de rang faible ? Des travaux préliminaires [80] ont montré que

1. le rang de X dépend de la convention choisie pour effectuer la TFCT (passe-bas, passe-bande) ;
2. si l'on considère une classe de signaux tels que les signaux sinusoïdaux, la matrice X représentant la TFCT de tels signaux est de rang faible en convention passe-bande ;
3. à l'inverse, si l'on considère une classe de signaux tels que les signaux de Dirac, la matrice X représentant la TFCT de tels signaux est de rang faible en convention passe-bas.

S'est alors posée la question : peut-on essayer de caractériser les signaux dont la TFCT est de faible rang ? Nous avons des éléments de réponse dans des publications très récentes [81, 82] mais nous travaillons toujours sur ce sujet et cherchons à exploiter les résultats déjà obtenus.

D'autre part, si l'on regarde de nouveau le critère 8.1, on voit que la minimisation se fait sur la matrice des coefficients de TFCT qui sont complexes. Que se passe-t'il alors si l'on suppose que le module est connu et que la phase est partiellement inconnue (le masque M devient donc un masque sur les phases uniquement) ? Ce problème est alors un problème de reconstruction de phases dans le domaine de la TFCT. Dans le cadre de la thèse de Marina Krémé, co encadrée par V. Emiya, B. Torrèsani et moi-même, nous travaillons sur la reconstruction de données manquantes dans le domaine temps-fréquence. Des premiers résultats dans le contexte de reconstruction de phases et utilisant des méthodes d'optimisation SDP ont été publiés dans [83].

8.2.2 Modèles graphiques

Dans le cadre de la thèse de Khuyen Le, co-encadrée avec Frédéric Richard, nous nous sommes intéressés aux modèles graphiques. En effet, notre objectif est d'étudier l'estimation de matrices de covariance et/ou de précision (son inverse). Cela a par conséquent conduit à l'étude du modèle Graphical LASSO (GLASSO)[84] et ses extensions (Group Glasso, fused Glasso) [85] ainsi qu'aux méthodes d'optimisation associées. Ce domaine a été beaucoup étudié ces dernières années que ce soit par la communauté optimisation [86, 87] ou statistiques [88], dans le but de proposer des algorithmes plus performants, de fournir des tests associés et d'étudier des modèles plus large.

Ce qui a d'autant plus motivé ce travail sont les entrevues que nous avons eu avec Prof. Eric Guedj, médecin de physique nucléaire à la Timone, avec lequel nous avons échangé sur des problématiques de classification. Si la classification de patients sains versus atteints de la maladie d'Alzheimer est de nos jours un problème quasiment résolu, d'autres pathologies restent plus difficiles à classifier comme la fibromyalgie versus la dépression versus les douleurs chroniques. Or, la modélisation des connections cérébrales peut aider à cette classification. C'est pourquoi, nous avons proposé une méthode de classification basée sur une analyse discriminante linéaire (LDA) que nous avons adaptée, le nombre de variables p à estimer étant plus grand que le nombre n d'observations. La méthode proposée inclut l'estimation de la matrice de précision en résolvant

un problème GLASSO nous permettant d'estimer directement la matrice de précision. Nous avons appliqué cette approche à la classification de sujets sains ($n = 56$) versus dépressifs ($n = 32$) versus atteints de fibromyalgie ($n = 52$). Cette classification est effectuée sur des données TEP (Tomographie par Emission de Positons) au 18F-FDG (18-fluorodésoxyglucose) fournies par l'Hôpital de la Timone, mesurant la consommation cérébrale métabolique de glucose dans $p = 116$ régions cérébrales. Les premiers résultats obtenus (Tab. 8.1) sont encourageants et montrent que prendre en compte les connexions cérébrales aide à la classification. Nous continuons donc à travailler sur ce sujet.

Paire	Groupe	NB	SVM	LDA
1	FM	69.23	86.02	84.62
	HC	58.93	85.54	98.21
2	DEP	90.62	91.18	93.75
	HC	87.50	90.99	96.43
3	FM	90.38	95.34	98.08
	DEP	81.25	94.79	93.75

TABLE 8.1 – Performances de classification (%) pour 3 méthodes : classification Naïve Bayésienne (NB), SVM et proposé (LDA).

8.2.3 Spectroscopie RMN

Dans le cadre du projet Amidex interdisciplinaire BIFROST¹, nous (S. Anthoine, B. Torrèsani, F. Richard et moi-même) travaillons en collaboration avec des chimistes de l'ISM2 à Marseille et aussi avec l'IFP Énergies Nouvelles. L'objectif est de procéder à l'analyse et au démelange de mélanges complexes, c'est à dire de déterminer les molécules présentes dans un mélange et leur quantité (proportion/concentration). Ce projet vient de démarrer au 1er mars 2018.

Après une prise en main de données 1D classiques (Fig. 8.1), pour lesquelles nous avons testé quelques algorithmes de Nonnegative Matrix Factorization (NMF) avec succès (méthodes des moindres carrés alternés, méthodes des moindres carrés alternés projetés, moindres carrés avec a priori de parcimonie), nous traitons aujourd'hui des spectres 2D Heteronuclear Single Quantum Correlation (HSQC). Les données utilisées sont des données de terpènes (huiles essentielles) pour lesquelles nous disposons non seulement des spectres des mélanges mais aussi les spectres des composés purs et nous connaissons également la matrice de mélange. Ces données complètes nous permettent de mieux appréhender ces données et de les traiter progressivement afin de mieux les comprendre.

L'idée pour les données à deux dimensions est d'essayer d'exploiter la parcimonie des données en 2D (Lorentzienne 2D) en utilisant, par exemple, des ondelettes bidimensionnelles.

1. <http://www.i2m.univ-amu.fr/project/bifrost/>

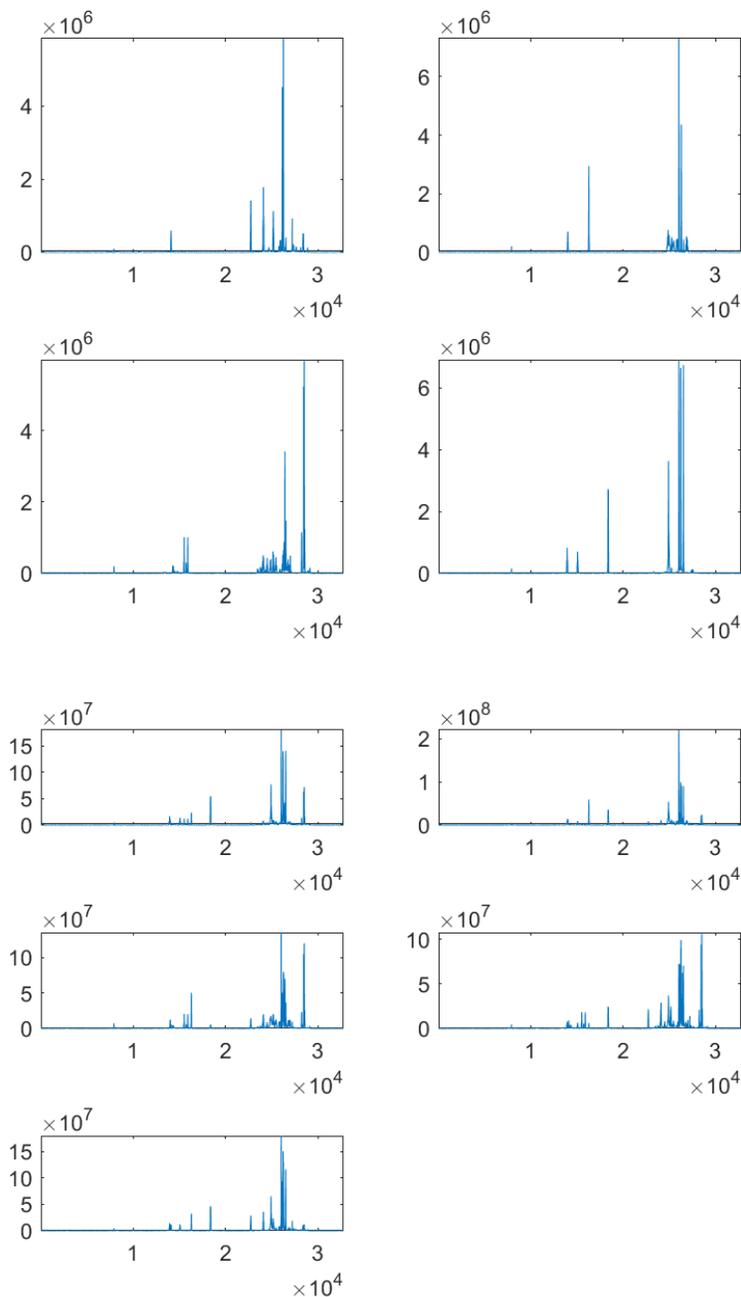


FIGURE 8.1 – 4 spectres purs (haut) et 5 mélanges associés (bas). Données de terpènes 1D RMN.

8.2.4 Stein pour l'estimation des paramètres

Nous avons un travail inachevé avec Jean-Christophe Pesquet qui porte sur l'estimation des paramètres de régularisation. Comme vous l'avez compris, ces paramètres jouent un rôle important et ne sont pas toujours facile à fixer. Une méthode, basée sur

des échantillonnages, est proposée dans le chapitre 5 dans le cas où la régularisation est effectuée dans le domaine des ondelettes.

Ce travail en cours vise à utiliser le principe de Stein [89] pour construire des ensembles de contraintes, les paramètres desquels sont estimés basé sur le principe de Stein. Ce travail s'inspire à la fois sur un des mes travaux antérieurs [90] mais aussi des travaux concernant le principe de Stein et les estimateurs Linear Expansion of Threshold (LET) [91]. Ce qui est intéressant est que le contexte du travail est assez large, permettant ainsi d'appréhender des problèmes de débruitage ou de déconvolution (et même reconstruction), associé à des opérateurs de trame (base d'ondelettes ou plus généralement trames).

Table des figures

4.1	Images de macroscopie - bruit et phénomène de photoblanchiment. . . .	35
4.2	Variation de la PSF (h) (haut) et son impact sur la scène observée (©CNES) (bas)	35
4.3	Images TEP. Sinogrammes observés (gauche) et images associées (droite).	36
4.4	Illustration d'une acquisition sismique : les signaux primaires y sont représentés en bleu et les réflexions multiples s en rouge.	37
4.5	Données de spectroscopie de fluorescence 3D.	38
5.1	Résultats sur données réelles.	51
5.2	Résultats d'identification du bruit à partir d'une seule image	52
5.3	Valeur de $\lambda_{\mathbf{m}}$ ($\mathbf{m} = (1, 1, 0)$) au cours des itérations	59
5.4	Valeur de $\lambda_{\mathbf{m}}$ ($\mathbf{m} = (1, 1, 1)$) au cours des itérations	59
5.5	Projection sur C_1/C_2 des points H_1, H_2 et H_3 de \mathbb{R}^2	64
5.6	Signaux sismiques considérés lorsque le bruit a un écart-type $\sigma = 0.08$. De haut en bas ; primaire (inconnu) \bar{y} , estimé \hat{y} , premier template r_0 , second template r_1 , multiple (inconnu) \bar{s} , estimé \hat{s} , et signal observé z	72
5.7	Close-up when $\tilde{\rho}$ is the $\ell_{1,2}$ -norm ; (a) observation z (en noir $\sigma = 0.08$) et primaire \bar{y} (en bleu), (b) primaire estimé \hat{y} (en rouge) et primaire original \bar{y} (en bleu).	73
5.8	De haut en bas : image observée z avec $\sigma = 0.04$, primaire \bar{y} et multiple \bar{s}	74
5.9	De haut en bas : primaire estimé y avec (Ventosa et al., 2012), avec la version 1D et avec la version 2D.	75
5.10	De haut en bas : données sismiques réelles observées avec un primaire partiellement apparent et un zoom de ces données.	76
5.11	Deux templates utilisés pour estimer les multiples.	77
5.12	From top to bottom : estimated primary obtained with (Ventosa et al., 2012) and with 1D version in Chapter 4.	78
5.13	Primaire estimé en utilisant la méthode 2D.	79
6.1	Montage utilisé pour obtenir la PSF expérimentale.	84
6.2	Coupes centrales de la PSF expérimentale, issue de l'observation au microscope biphotonique de micro-billes fluorescentes de diamètre 0.5 microns placées sous une tranche de tissu cérébral d'épaisseur 100 microns.	84
6.3	Coupes originales (gauche), dégradées (centre) et restaurées (droite), à $Z = 13.6$ microns (haut) et $X = 5.36$ microns (bas), de l'image CT_aneurysm.	86

6.4	Coupes observées au microscope biphotonique (haut) et restaurées (bas), à $Z = 22$ microns (gauche) et $X = 8$ microns (droite), d'une bille fluorescente de diamètre 4 microns.	86
6.5	Coupes observées au microscope biphotonique (gauche) et restaurées (droite), à $Z = 82.5$ microns (haut) et $Y = 369.6$ microns (bas), d'une tranche clarifiée de cerveau de souris.	86
6.6	Courbes de $\text{prox}_{\rho_n}(v)$ en fonction de $v \in [-2, 22]$ lorsque $\alpha^{(n)} = 2$ et $[\eta_{\min}^{(n)}, \eta_{\max}^{(n)}] = [0, 4]$	90
6.7	MEEF de référence et MEEF reconstruits avec N -way ($E1 = -8.7$ dB, $E2 = 31.7$ dB) et l'algorithme proposé BC-VMFB sans régularisation $\alpha = 0$ ($E1 = -12.3$ dB, $E2 = 25.6$ dB) et avec régularisation $\alpha = 0.05$ ($E1 = -15$ dB, $E2 = -409$ dB).	92
6.8	Spectres (normalisés) estimés par N -way (bleu), fast HALS (vert) et BC-VMFB (rouge). La vérité terrain est en noir.	93
6.9	MEEF estimés avec N -way, fast HALS, BC-VMFB et les MEEF de référence.	94
6.10	Erreur d'estimation $E1$ (haut) et erreur de surestimation $E2$ (bas) en utilisant l'algorithme proposé (ligne rouge pointillé), le N -way (ligne verte pointillé), et le fast HALS (ligne bleu tirets).	95
6.11	MEEF estimés avec l'algorithme N -way (gauche), et l'algorithme BC-VMFB (droite), cas $R = 4$	96
6.12	Concentrations normalisées estimées avec le N -way (gauche) et le BC-VMFB (droite), cas $R = 4$	97
6.13	MEEF estimés avec le N -way (gauche) et le BC-VMFB (droite), cas $R = 6$	97
6.14	Concentrations normalisées estimées avec le N -way (gauche) et le BC-VMFB (droite), cas $R = 6$	98
8.1	4 spectres purs (haut) et 5 mélanges associés (bas). Données de terpènes 1D RMN.	109

Liste des tableaux

5.1	Performance de l'algorithme EM pour des cas défavorables pour la méthode basée sur les cumulants.	50
5.2	Données simulées	59
5.3	Données réelles	60
6.1	Temps de calculs et performances numériques des algorithmes fast HALS, <i>N</i> -way et BC-VMFB en utilisant la même initialisation.	94
8.1	Performances de classification (%) pour 3 méthodes : classification Naïve Bayésienne (NB), SVM et proposé (LDA).	108

Liste des acronymes

- BC-VMFB** Explicite-Implicite à métrique variable par bloc de coordonnées (*Block Coordinate Variable Metric Forward-Backward*). 41, 42, 88, 92–94, 96–98, 112, 113
- CPD** Canonical Polyadic Decomposition. 37, 87
- EM** espérance-maximisation (*expectation-maximization*). 31, 46, 48–50, 52, 113
- GLASSO** Graphical LASSO. 107, 108
- HALS** Hierarchical Alternating Least Squares. 91–95, 112, 113
- HSQC** Heteronuclear Single Quantum Correlation. 108
- IRM** Imagerie par Résonance Magnétique. 14, 20, 36
- LDA** analyse discriminante linéaire. 107, 108, 113
- LET** Linear Expansion of Threshold. 110
- M+L FBF** Monotone + Lipschitz Forward-Backward-Forward. 41, 42, 63, 64, 69
- M+S FBF** Monotone + Skew Forward-Backward-Forward. 41
- MAP** maximum a posteriori. 62
- MEEF** Matrice d'Excitation et d'Emission de Fluorescence. 91, 92, 94, 96, 97, 112
- MEEG** magnetoencephalographie. 106
- MM** Majoration-Minimisation. 83, 87
- MV** maximum de vraisemblance. 46
- NMF** Nonnegative Matrix Factorization. 108
- PSD** symétrique définie positive. 40, 42, 43
- PSF** Point Spread Function. 14, 35, 82, 84, 85, 87, 111
- RMN** résonance magnétique nucléaire. 7, 14, 21, 32, 108, 109, 112
- SDP** semi-définie positive. 19, 107
- SNR** rapport signal à bruit. 58, 85, 91, 92, 94
- SVM** support vector machine. 108, 113
- TEP** tomographie par émission de positrons. 15, 19, 20, 32, 36, 108, 111
- TFCT** transformée de Fourier à fenêtre. 106, 107

Bibliographie

- [1] L. Jacques, L. Duval, C. Chaux, and G. Peyré. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Process.*, 91(12) :2699–2730, Dec. 2011.
- [2] A. Jezierska, C. Chaux, J.-C. Pesquet, and H. Talbot. An EM approach for Poisson-Gaussian noise modeling. In *European Signal Processing Conference (EUSIPCO)*, Barcelone, Espagne, 29 Aug. - 2 Sep. 2011.
- [3] A. Jezierska, C. Chaux, J.-C. Pesquet, H. Talbot, and G. Engler. An EM approach for time-variant Poisson-Gaussian model parameter estimation. *IEEE Trans. Signal Process.*, 62(1) :17–30, Jan. 2014.
- [4] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton Univ. Bull*, 13 :49–52, 1902.
- [5] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame based inverse problems. *Inverse Probl.*, 23(4) :1495–1518, Aug. 2007.
- [6] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS books in mathematics. Springer, 2011.
- [7] Ole Christensen. *An Introduction to Frames and Riesz Bases*. Birkhäuser Boston, 2002.
- [8] S. Mallat. *A wavelet tour of signal processing : the sparse way*. Academic Press, San Diego, CA, USA, 3rd edition, 2009.
- [9] N. Pustelnik. *Proximal methods for the resolution of inverse problems. Application to Positron Emission Tomography*. PhD thesis, Univ. Paris-Est, 2010.
- [10] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer Verlag, 2010.
- [11] N. Komodakis and J.-C. Pesquet. Playing with duality : An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Process. Mag.*, 2015.
- [12] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93 :273–299, 1965.

- [13] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4) :1168–1200, Nov. 2005.
- [14] P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Probl.*, 24(6) :27, Dec. 2008.
- [15] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Springer-Verlag, New York, 1993.
- [16] M. A. T. Figueiredo and R. D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.*, 12(8) :906–916, Aug. 2003.
- [17] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle. A l^1 -unified variational framework for image restoration. In T. Pajdla and J. Matas, editors, *Proc. Eur. Conf. Comput. Vis.*, volume LNCS 3024, pages 1–13, Prague, Czech Republic, May 2004. Springer.
- [18] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Applied Math.*, 57 :1413–1457, 2004.
- [19] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, 18(11) :2419–2434, Nov 2009.
- [20] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6) :964–979, 1979.
- [21] P. L. Combettes and J.-C. Pesquet. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE J. Sel. Top. Signal Process.*, 1(4) :564–574, Dec. 2007.
- [22] N. Pustelnik, C. Chaux, and J.-C. Pesquet. Parallel proximal algorithm for image restoration using hybrid regularization. *IEEE Trans. Image Process.*, 20(9) :2450–2462, Sep. 2011.
- [23] J.-C. Pesquet and N. Pustelnik. A parallel inertial proximal optimization method. *Pacific J. Opt.*, 8(2) :273–305, Apr. 2012.
- [24] H. Raguét, J. Fadili, and G. Peyré. A generalized forward-backward splitting. *SIAM J. Imaging Sci.*, 6(3) :1199–1226, 2013.
- [25] L. M. Briceños Arias and P. L. Combettes. A monotone + skew splitting model for composite monotone inclusions in duality. *SIAM J. Optim.*, 21(4) :1230–1250, Oct. 2011.
- [26] P. L. Combettes and J.-C. Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.*, 20(2) :307–330, Jun. 2012.
- [27] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40(1) :120–145, 2011.

- [28] L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158(2) :460–479, Aug. 2013.
- [29] Bang Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3) :667–681, Apr 2013.
- [30] Z. Q. Luo and P. Tseng. On the convergence of the coordinate descent method for convex differentiable minimization. *J. Optim. Theory Appl.*, 72(1) :7–35, Jan 1992.
- [31] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. A block coordinate variable metric forward-backward algorithm. *J. Global Optim.*, pages 1–29, 2016.
- [32] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Programm.*, 146(1) :459–494, Aug 2014.
- [33] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. A preconditioned forward-backward approach with application to large-scale nonconvex spectral unmixing problems. In *39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, Florence, Italie, May 2014.
- [34] C. Chaux, A. Jezierska, J.-C. Pesquet, and H. Talbot. A spatial regularization approach for vector quantization. *Journal of Mathematical Imaging and Vision*, 41 :23–38, 2011.
- [35] F. X. Dupé. Deconvolution under Poisson noise using proximal thresholding iteration, 2008. <http://fxdupe.free.fr/software.html>.
- [36] F. Luisier, T. Blu, and M. Unser. Image denoising in mixed poisson-gaussian noise. *IEEE Trans. Image Process.*, PP(99) :1 –1, 2010.
- [37] A. Jezierska, J.-C. Pesquet, H. Talbot, and C. Chaux. Iterative poisson-gaussian noise parametric estimation for blind image denoising. In *Proc. Int. Conf. Image Process.*, Paris, France, Oct. 2014.
- [38] C. Vonesch, S. Ramani, and M. Unser. Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint. In *Proc. Int. Conf. Image Process.*, pages 665–668, San Diego CA, USA, Oct. 12-15, 2008.
- [39] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, CA, USA, 2nd edition, 1998.
- [40] L. Younes. Parametric inference for imperfectly observed Gibbsian fields. *Probab. Theory Relat. Fields*, 82 :625–645, 1989. Springer-Verlag.
- [41] A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia. Hyperparameter estimation for satellite image restoration using a MCMC maximum likelihood method. *Pattern Recogn.*, 35(2) :341–352, 2002.
- [42] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8(1) :141–148, 1988.

- [43] B. Zhou, L. Gao, and Y. H. Dai. Gradient Methods with Adaptive Step-Sizes. *Comput. Optim. Appl.*, 35(1) :69–86, 2006.
- [44] G. Frassoldati, G. Zanghirati, and L. Zanni. New Adaptive Step-size Selections in Gradient Methods. *J. Ind. Manag. Optim.*, 4(2) :299–312, 2008.
- [45] S Bonettini, R Zanella, and L Zanni. A scaled gradient projection method for constrained image deblurring. *Inverse Probl.*, 25(1) :015002 (23pp), 2009.
- [46] Wanyou Cheng and Zixin Chen. Nonmonotone spectral method for large-scale symmetric nonlinear equations. *Numer. Algor.*, April 2012.
- [47] I. Daubechies. *Ten Lectures on Wavelets*. SIAM Lecture Series. CBMS-NSF, Philadelphia, PA, USA, 1992.
- [48] O. Christensen. *Frames and bases. An introductory course*. Birkhäuser, 2008.
- [49] D. Gragnaniello, C. Chaux, J.-C. Pesquet, and L. Duval. A convex variational approach for multiple removal in seismic data. In *European Signal Processing Conference (EUSIPCO)*, Bucarest, Roumanie, Aug. 27 - 31 2012.
- [50] M. Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet. Seismic multiple removal with a primal-dual proximal algorithm. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, page 5 pp., Vancouver, Canada, May, 26-31 2013.
- [51] M. Kowalski. Sparse regression using mixed norms. *Appl. Comp. Harm. Analysis*, 27(3) :303–324, Nov. 2009.
- [52] L. Jacques, L. Duval, C. Chaux, and G. Peyré. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Process.*, 91(12) :2699–2730, Dec. 2011.
- [53] L. M. Briceño-Arias and P. L. Combettes. A monotone + skew splitting model for composite monotone inclusions in duality. *SIAM J. Optim.*, 21(4) :1230–1250, Oct. 2011.
- [54] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2) :890–912, Nov. 2008.
- [55] L. Chaari, N. Pustelnik, C. Chaux, and J.-C. Pesquet. Solving inverse problems with overcomplete transforms and convex optimization techniques. In *Proc. SPIE, Wavelets*, San Diego, California, USA, Aug. 2-6 2009.
- [56] S. Ventosa, S. Le Roy, I. Huard, A. Pica, H. Rabeson, P. Ricarte, and L. Duval. Adaptive multiple subtraction with wavelet-based complex unary Wiener filters. *Geophysics*, 77(6) :V183–V192, Nov.-Dec. 2012.
- [57] P. T. C. So, C. Y. Dong, B. R. Masters, and K. M. Berland. Two-photon excitation fluorescence microscopy. *Annu. Rev. Biomed. Eng.*, 2 :399–429, 2000.
- [58] F. Helmchen and W. Denk. Deep tissue two-photon microscopy. *Nat. Methods*, 2(12), 2005.

- [59] F. Debarbieux. Towards a dynamic anatomopathology of glioblastoma in mice thanks to in vivo biphotonic microscopy. *Ann. Pathol.*, 30(5S1) :53–55, Nov. 2010.
- [60] P. Dufour, S. Dufour, A. Castonguay, N. McCarthy, and Y. De Koninck. Microscopie à deux photons pour l'imagerie cellulaire fonctionnelle : avantages et enjeux. *Médecine/Sciences*, 22 :837–844, Oct. 2006.
- [61] R. W. Cole, T. Jinadasa, and C. M. Brown. Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control. *Nature Protocols*, 6 :1929–1941, 2011.
- [62] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot. A majorize-minimize subspace approach for ℓ_2 - ℓ_0 image regularization. *SIAM J. Imaging Science*, 6(1) :563–591, 2013.
- [63] A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu, and S. Ciochina. A majorize-minimize memory gradient method for complex-valued inverse problem. *Signal Process.*, 103 :285–295, Oct. 2014. Special issue on Image Restoration and Enhancement : Recent Advances and Applications.
- [64] J. Pawley. *Handbook of Biological Confocal Microscopy*. Language of science. Springer, 2006.
- [65] H. Hama *et al.* Scale : a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nature Neuroscience*, 14 :1481–1488, Oct. 2011.
- [66] N. Hajlaoui, C. Chaux, G. Perrin, F. Falzon, and A. Benazza-Benyahia. Satellite image restoration in the context of a spatially varying point spread function. *J. Opt. Soc. Am. A*, 27(6) :1473–1481, 2010.
- [67] L.-H. Lim and P. Comon. Nonnegative approximations of nonnegative tensors. *J. Chemometrics*, 23 :432–441, August 2009.
- [68] J. B. Kruskal. Three-way arrays : rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Lin. Algebra Appl.*, 18(2) :95 – 138, 1977.
- [69] N. D. Sidiropoulos and R. Bro. On the uniqueness of multilinear decomposition of N-way arrays. *J. Chemometrics*, 14(3) :229–239, 2000.
- [70] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. *Journal of Optimization Theory and Applications*, 162(1) :107–132, July 2014.
- [71] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems 13*, pages 556–562. MIT Press, 2001.
- [72] R. Bro. Parafac : tutorial and applications. *Chemom. Intell. Lab. Syst.*, 38(2) :149–171, October 1997.
- [73] A. H. Phan, P. Tichavský, and A. Cichocki. Fast alternating ls algorithms for high order candecomp/parafac tensor factorizations. *IEEE Trans. on Sig. Proc.*, 61(19) :4834–4846, June 2013.

- [74] G. H. Paul. *Introduction to Mathematical Statistics*. Wiley, 5th edition, 1984.
- [75] J. R. Lakowicz, H. Szmazinski, K. Nowaczyk, K. W. Berndt, and M. Johnson. Fluorescence lifetime imaging. *Anal. Biochem.*, 202(2) :316–330, May 1992.
- [76] E. M. Carstea, S. Mounier, R. Redon, X. Luciani, C. Gadio, and A. Baker. On-line parafac analysis of fluorescence spectra. In *International Workshop on Organic Matter Spectroscopy (WOMS'13)*, Toulon, France, July 2013.
- [77] R. G. Zepp, W. M. Sheldon, and M. A. Moran. Dissolved organic fluorophores in southeastern us coastal waters : correction method for eliminating rayleigh and raman scattering peaks in excitation-emission matrices. *Marine Chemistry*, 89(1-4) :15–36, 2004.
- [78] Z. Fan, O. Schroder, and J. Krahl. Analysis of diesel fuels/biodiesel blends and identification of biodiesel using time-resolved laser-induced fluorescence spectroscopy (trlfs). *Appl Agric Forestry Res*, 65 :1–14, 2015.
- [79] P. Magron, R. Badeau, and B. David. Model-based STFT phase recovery for audio source separation. *IEEE Trans. Audio Speech Lang. Process.*, 2018.
- [80] R. Hamon, C. Chaux, and V. Emiya. De la faiblesse de rang en temps-fréquence. In *Proc. GRETSI*, Sep. 5-8 2017.
- [81] V. Emiya, R. Hamon, and C. Chaux. Being low-rank in the time-frequency plane. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, Calgary, Canada, Apr. 15-20 2018.
- [82] K. Usevich, V. Emiya, D. Brie, and C. Chaux. Characterization of finite signals with low-rank STFT. In *Proc. IEEE Workshop Stat. Signal Process.*, Freiburg, Germany, June 2018.
- [83] A. M. Krémé, V. Emiya, and C. Chaux. Phase reconstruction for time-frequency inpainting. In *14th International Conference on Latent Variable Analysis and Signal Separation*, Guildford, UK, Jul. 2-6 2018.
- [84] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3) :432–441, 2008.
- [85] P. Danaher, P. Wang, and D. M. Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *J. Roy. Stat. Soc. B Stat. Meth.*, 76(2) :373–397, March 2014.
- [86] G. Marjanovic and A. O. Hero. l0 sparse inverse covariance estimation. *IEEE Trans. Signal Process.*, 2015.
- [87] A. Benfenati, E. Chouzenoux, and J.-C. Pesquet. A proximal approach for a class of matrix optimization problems. <https://hal.archives-ouvertes.fr/hal-01673027/>, 2017.
- [88] M. Grazier G'Sell, J. Taylor, and R. Tibshirani. Adaptive testing for the graphical lasso. <https://arxiv.org/abs/1307.4765>, 2013.
- [89] C. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Stat.*, 9(6) :1135–1151, Nov. 1981.

-
- [90] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A SURE approach for digital signal/image deconvolution problems. *IEEE Trans. Signal Process.*, 57(12) :4616–4632, Dec. 2009.
- [91] T. Blu and F. Luisier. The SURE-LET approach to image denoising. *IEEE Trans. Image Process.*, 16(11) :2778–2786, Nov. 2007.