



Habilitation à Diriger les Recherches

présentée par

Nelly PUSTELNIK

le 30 septembre 2021

Approches variationnelles régularisées pour la résolution de problèmes inverses et pour l'apprentissage machine : de la modélisation aux algorithmes à grande échelle

Jury

Mme Laure BLANC FÉRAUD,	DR CNRS, Université de Nice-Sophia Antipolis	Rapportrice
M Gabriel PEYRÉ,	DR CNRS, Ecole Normale Supérieure	Rapporteur
M Bruno TORRESANI,	Professeur, I2M, Aix Marseille Université	Rapporteur
Mme Christine DE MOL,	Professeur, ULB, Belgique	Examinatrice
M Jérôme IDIER,	DR CNRS, Ecole Centrale Nantes	Examineur
M Simon MASNOU,	Professeur, Université Claude Bernard Lyon 1	Examineur
M Patrice ABRY,	DR CNRS, ENS de Lyon	Garant

Table des matières

1	Parcours académique et scientifique	1
1.1	Curriculum vitae	2
1.2	Responsabilités d'encadrement	3
1.3	Expertise scientifique	6
1.4	Animation de la recherche	8
1.5	Administration de la recherche	9
1.6	Contrats et projets de recherche	11
1.7	Enseignement	12
1.8	Responsabilités administratives	13
1.9	Vulgarisation scientifique	13
1.10	Publications	14
2	Synthèse des activités de recherches	19
2.1	Du modèle direct à la formulation régularisée	20
2.2	La projection épigraphique, un outil clé dans la gestion de contraintes non linéaires	35
2.3	Algorithmes proximaux et grand volume de données	58
2.4	Sélection des hyperparamètres	80
2.5	Applications variées des algorithmes proximaux sur des données réelles.	96
2.6	Conclusions et perspectives	99
2.7	Références	102

Remerciements

Mes remerciements s'adressent tout d'abord à ma rapportrice Laure Blanc Féraud et mes rapporteurs Gabriel Peyré et Bruno Torresani. Je les remercie pour leur lecture attentive et leurs retours très positifs. Je remercie également Christine De Mol, Jérôme Idier et Simon Masnou d'avoir accepté de faire partie de mon jury.

Je tiens ensuite à remercier l'équipe SiSyPhe qui m'a accueillie il y a presque 10 ans et où j'ai pu mener mes travaux de recherche dans un cadre extrêmement dynamique et bienveillant. Je remercie tout d'abord Patrick Flandrin et Pierre Borgnat, mes premiers collaborateurs au sein de l'équipe, qui m'ont permis de découvrir une première facette des travaux de SiSyPhe qu'était le temps fréquence et la décomposition modale empirique. Je remercie ensuite Patrice, mon principal collaborateur dans l'équipe, pour nos nombreux travaux en détection de changement, segmentation de texture et apprentissage. Je le remercie de m'avoir intégré à ses collaborations et de m'avoir ainsi enseigné les clés d'un travail multidisciplinaire réussi. Je le remercie également pour son dynamisme sans faille, sa disponibilité, ses nombreux et précieux conseils. Merci également à Pablo Jensen, Stéphane Roux et Benjamin Audit tout autant passionnés que passionnants.

Mes remerciements s'adressent ensuite à mes doctorantes et doctorants Giovanni Chierchia, Jordan Frécon, Barbara Pascal et Laurence Denneulin avec qui j'ai été ravie de partager ces années de travail commun. Je remercie également Trieu Vy Le actuellement en thèse et collaborateur précieux pour mener à bien mon ANR JCJC. Je remercie également Jérémy Schmitt, Mingyuan Jiu et Marion Foare anciennement chercheurs post-doctoraux et maintenant collaborateurs talentueux. Je remercie également Gabriel Michau avec qui j'ai eu l'opportunité de travailler durant sa thèse pour son fabuleux livre jeunesse intitulé *Opérateur dérivé va à la montagne* offert lors de la naissance de ma première fille.

Je remercie ensuite mes nombreux collaborateurs du Laboratoire de Physique, Valérie Vidal, Jean-Christophe Géminard, Sylvain Joubaud, Philippe Odier et également des chercheurs d'autres laboratoires en France ou à l'étranger Jérôme Boulanger, Luis Briceño-Arias, Laurent Condat, Nicolas Dobigeon, Maud Langlois, Jean-Christophe Pesquet, Béatrice Pesquet-Popescu, Eric Thiebaut, Samuel Vaiter, Herwig Wendt, pour nos collaborations extrêmement enrichissantes.

Je remercie enfin l'ensemble des membres du Laboratoire de Physique pour cette ambiance si conviviale qui y règne. Quelques remerciements particuliers s'adressent à Erika, Fatiha, Laure, Laurence, pour leur efficacité et leur bonne humeur communicative. Un merci particulier à Nadine qui a toujours été présente dans les démarches administratives les plus complexes. Je remercie également Artyom Petrossian pour son accueil à mon arrivée au laboratoire et ces quelques années où nous partagions le même bureau. Je tiens également à remercier Thierry Dauxois pour son soutien et conseils pendant ces 9 années en tant que directeur du laboratoire et merci également à Jean-Christophe Géminard pour son aide et son soutien dans les démarches de mission à l'UCLouvain.

Je remercie également Laurent Jacques pour son accueil à l'UCLouvain.

Mes derniers remerciements vont enfin à mon petit cercle familial dans lequel je puise l'énergie et l'envie d'aller plus loin chaque jour. Merci plus particulièrement à mes parents, pour leur dynamisme, leurs encouragements, les nombreuses attentions de ma mère et les discussions passionnées avec mon père. Un merci plus particulier s'adresse bien évidemment à mon conjoint Benoît, qui me supporte depuis tant d'années. Je le remercie pour son ouverture d'esprit, son positivisme sans faille, ses conseils précieux et son soutien quotidien. Enfin, je tiens à remercier mes filles Ambre et Camille pour ces moments de bonheur quotidien qui me nourrissent et me portent.

Chapitre 1

Parcours académique et scientifique

Sommaire

1.1 Curriculum vitae	2
1.1.1 Identification	2
1.1.2 Parcours académique et professionnel	2
1.2 Responsabilités d'encadrement	3
1.2.1 Doctorants	3
1.2.2 Encadrement post-doctoral et Agrégé Préparateur	4
1.2.3 Encadrements de stages et Projets de fin d'étude	4
1.3 Expertise scientifique	6
1.3.1 Séminaires invités	6
1.3.2 Ecoles doctorales	7
1.3.3 Activités éditoriales	7
1.3.4 Activités de relecture	7
1.3.5 Jurys de thèse	8
1.4 Animation de la recherche	8
1.4.1 En France	8
1.4.2 A l'étranger	9
1.5 Administration de la recherche	9
1.5.1 En France	10
1.5.2 A l'étranger	10
1.6 Contrats et projets de recherche	11
1.6.1 PI	11
1.6.2 Membre	11
1.7 Enseignement	12
1.8 Responsabilités administratives	13
1.9 Vulgarisation scientifique	13
1.10 Publications	14

1.1 Curriculum vitae

1.1.1 Identification

Nom	: Pustelnik
Prénom	: Nelly
Date et lieu de Naissance	: 12-11-1984 à Sarlat-la-canéda
Nationalité	: Française
Statut actuel	: Chargée de recherche classe normale au CNRS.
Laboratoire de rattachement	: Laboratoire de Physique à l'Ens de Lyon (LPENSL), 46, allée d'Italie, 68007 Lyon, FRANCE.
Section CNRS	: 7

1.1.2 Parcours académique et professionnel

Nov. 2011 – (en cours)	Chargée de recherche classe normale au CNRS Laboratoire de Physique, UMR CNRS 5672, Ens de Lyon, France Equipe SiSyPhe : Signaux, Systèmes et Physique
Oct. 2019 – (en cours)	Professeure visiteuse SST/ICT/INMA UCLouvain, Belgique
Sept. 2019	Professeure invitée SST/ICT/INMA UCLouvain, Belgique
2010 – 2011	Chercheure post-doctorale Laboratoire IMS, UMR CNRS 5218, Université Bordeaux, France Equipe : Traitement du signal et des images
2007 – 2010	Doctorat LIGM, UMR CNRS 8049, Université Paris-Est, France Equipe : Signal et Communication Directeurs : J.-C. Pesquet et C. Chaux Sujet : Méthodes proximales pour la résolution de problèmes inverses. Application à la tomographie par émission de positrons. Jury : Mohamed Najim (Président), Laure Blanc-Féraud (Rapportrice), Jalal Fadili (Rapporteur), Claude Comtat (Examineur), Michel Defrise (Examineur).

1.2 Responsabilités d'encadrement

Depuis mon recrutement au CNRS, j'ai eu la possibilité d'encadrer et continue d'encadrer des étudiant.e.s en thèse, chercheur.e.s en séjour post-doctoral, agrégé préparateur et également des stagiaires de divers niveaux.

1.2.1 Doctorants

- 2012-2015 : **G. Chierchia**, Télécom ParisTech.

Co-encadrement 40% en collaboration avec Béatrice Pesquet-Popescu (Télécom ParisTech) et Jean-Christophe Pesquet (LIGM, Univ. Paris-Est).

Sa thèse s'intitule "*Epigraphical decomposition to deal with non-linear convex constraints. Application to image reconstruction, supervised classification, and anomaly detection*" et fut financée par la fondation Mines-Telecom. Ses travaux ont conduit à 2 publications dans des journaux internationaux (IEEE TIP, SIVP), 4 conférences internationales (IEEE ICASSP (x3), SPARS), 1 toolbox MATLAB.

Giovanni Chierchia a reçu le Prix de thèse Fondation Télécom.

Giovanni Chierchia est actuellement Maître de Conférence à l'ESIEE à Marne-la-Vallée.

- 2013-2016 : **Jordan Frécon**, LPENSL.

Co-encadrement 50% en collaboration avec P. Abry (LPENSL).

Sa thèse s'intitule "*Apport des méthodes d'optimisation dans l'estimation des paramètres de régularité locale pour des signaux multivariés*" et fut financée par l'Ecole doctorale PHAST de l'Université de Lyon. Ses travaux ont conduit à 4 publications dans des journaux internationaux (3 IEEE TSP et 1 Biomedical Physics & Engineering Express), 5 conférences internationales (IEEE ICIP, IEEE ICASSP, IEEE EBM(x2), EUSIPCO), 1 conférence nationale (GRETSI) et plusieurs toolbox MATLAB.

Jordan Frécon est actuellement Chercheur postdoctoral dans le groupe LITOS à l'INSA Rouen.

- 2017 - 2020 : **Barbara Pascal**, LPENSL.

Co-encadrement 50% en collaboration avec P. Abry (LPENSL).

Sa thèse s'intitule "Estimation régularisée d'attributs fractals par minimisation convexe pour la segmentation de textures" et fut financée sur un Contrat Doctoral Spécifique Normalien (CDSN). Ses travaux ont conduit à 3 publications dans des journaux internationaux (JMIV, ACHA, Annals. Telecom), 5 conférences internationales dont 3 avec actes (EUSIPCO, IEEE ICIP, IEEE ICASSP, SIAM-IS, Curves and Surfaces), 1 conférence nationale (GRETSI) ainsi que deux toolbox Matlab

Barbara Pascal est actuellement Chercheure postdoctorale au Laboratoire CRIStAL à Lille.

- 2017 - 2020 : **Laurence Denneulin**, CRAL, Observatoire de Lyon/LPENSL.

Co-encadrement 20% en collaboration avec M. Langlois and E. Thiébaut (CRAL, Observatoire de Lyon).

Sa thèse s'intitule "*Approche inverse pour la reconstruction des environnements circumstellaires en polarimétrie avec l'instrument d'imagerie directe ESO/VLT SPHERE IRDIS*" et fut financée par l'école doctorale ED PHAST de l'Université de Lyon. Ses travaux ont conduit à 1 publication dans un journal international (A&A), 1 conférence internationale avec acte (SPIE), 1 conférence nationale (GRETSI).

Laurence Denneulin est actuellement Chercheure postdoctorale au Laboratoire Hubert Curien à Saint-Etienne.

- Sept. 2020 - (en cours) : **Hoang Trieu Ve LE**, LPENSL.

Co-encadrement 50% en collaboration avec M. Foare (LPENSL).

Sa thèse est financée par l'ANR JCJC Multisc-In dont je suis PI. Le sujet de sa thèse est "Analyse multi-échelle et détection d'interfaces dans des images de très grandes tailles". Les premiers résultats obtenus ont été présentés au workshop DGDVC.

1.2.2 Encadrement post-doctoral et Agrégé Préparateur

- **Agrégé préparateur** (2012-2015) : **Jérémy Schmitt**, LPENSL.

Jérémy Schmitt a travaillé sur une formulation de la décomposition modale empirique en 2D et son application à l'analyse des ondes internes. Ses travaux ont conduit à 2 articles de journaux (IEEE TIP et Experiments in Fluids), 2 conférences internationales (ICASSP 2014, EUSIPCO 2015) et 1 Toolbox MATLAB.

Jérémy Schmitt est actuellement Professeur agrégé de physique en BTS systèmes numériques au Lycée Marcel Sembat à Vénissieux.

- **Chercheur post-doctoral** (Fév. 2016 - Fév. 2017) : **Mingyuan Jiu**, LPENSL / TOTAL.

Mingyuan Jiu a été recruté dans le cadre d'un financement TOTAL SA afin de travailler sur des problèmes de classification parcimonieuse dans le contexte de fabrication de carburants spéciaux. Ses travaux ont conduit à un article de journal (IEEE J. Signal Proc. Systems), 2 conférences internationales (IEEE MLSP) et 1 prototype MATLAB pour TOTAL.

Mingyuan Jiu est actuellement *Assistant professor* à Zhengzhou University en Chine.

- **Chercheuse post-doctorale** (Sept. 2017 - Août. 2018) : **M. Foare**, LPENSL.

Marion Foare a été recrutée dans le cadre d'un financement du LPENSL. Son sujet de recherche était l'implémentation numérique du modèle de Mumford-Shah discret par des algorithmes proximaux. Ses travaux ont conduit à un article de journal (IEEE TIP), 2 conférences internationales (IEEE ICIP, SIAM IS) et 1 Toolbox MATLAB.

Marion Foare est actuellement Maître de Conférence à CPE Lyon et rattachée au LIP à l'ENS de Lyon.

1.2.3 Encadrements de stages et Projets de fin d'étude

- 2009 – **Adji Néné DIAO, Baptiste MAZIN** (Supélec) ont effectué leur **stage de M1**, d'une durée de 4 mois, au LIGM, sur le sujet "Implémentation parallèle d'algorithmes proximaux pour la restauration d'image". Il s'agissait d'un co-encadrement avec Caroline Chaux et Jean-Christophe Pesquet.
- 2011 – **Daniela-Ecaterina CRISTEA** (ENSEIRB-MATMECA) a effectué son **stage de M2**, d'une durée de 4 mois, à l'IMS sur le sujet "Algorithmes pour la résolution de problèmes de reconstruction tomographique à faible nombre de vues. Implantation et comparaison". Il s'agissait d'un co-encadrement avec Yannick Berthoumieu (IMS).
- 2012 **Lahlou Lamiae, Liebert Lise, Lotigier Rémi, Tajouri Slim** ont effectué leur **Projet d'Ingénierie** (Télécom St-Etienne) sur le sujet "Détection de personnes dans une foule". Il s'agissait d'un co-encadrement avec Denis Bartolo.

- 2015 – **Noura Baouan** (Supélec) a effectué son **stage de M1**, d’une durée de 4 mois, chez TOTAL SA sur le sujet “Création d’une base de données avec la répartition moléculaire et les propriétés physico-chimiques”. Il s’agissait d’un co-encadrement avec Mériam Chebre (TOTAL SA).
- 2017 – **Barbara Pascal** (M2 Sciences de la matière, ENS de Lyon) a effectué son **stage de M2**, d’une durée de 5 mois, au LPENSL sur le sujet “Segmentation de textures par méthodes proximales”. Il s’agissait d’un co-encadrement avec Patrice Abry. Son stage a donné suite à un article de conférence (IEEE ICASSP). Barbara Pascal a poursuivi son stage par une thèse sur le même sujet.
- 2017 – **Jules Colas** (L3 Sciences de la matière, ENS de Lyon) a effectué son **stage de L3**, d’une durée de 4 mois, au LPENSL sur le sujet “Débruitage de signaux de stick-slip”. Il s’agissait d’un co-encadrement avec Valérie Vidal. Son stage a donné suite à un article de journal (Physical Review E).
- 2018 – **Yacouba Kaloga** (M2 Systèmes Complexes, ENS de Lyon) a effectué son **stage de M2**, d’une durée de 5 mois, au LPENSL sur le sujet “Modèle de Mumford-Shah discret sur graphe. Application à l’estimation des matrices de transfert entre deux élections”. Il s’agissait d’un co-encadrement avec Marion Foare et Pablo Jensen. Son stage a donné suite à un article de journal (IEEE SPL).
- 2019 – **Vincent Paparlado** (M2 Maths en action, Univ. Lyon 1) a effectué son **stage de M2**, d’une durée de 5 mois, au LPENSL sur le sujet “Estimer le transfert des votes entre deux élections grâce à l’optimisation sur graphe”. Il s’agissait d’un co-encadrement avec Pablo Jensen.
- 2019 – **Loris Helmlinger** (Ecole d’Ingénieur Denis Diderot, Paris) a effectué son **stage de L3**, d’une durée de 3 mois au LPENSL sur le sujet “Apprentissage profond pour la segmentation de textures. Application à l’étude des écoulements multiphasiques.” Il s’agissait d’un co-encadrement avec Barbara Pascal et Valérie Vidal.
- 2020 – **Mamadou Ngom** (M2 Modélisation mathématique pour l’image et le signal, Bordeaux) a effectué son **stage de M2**, d’une durée de 6 mois, dans le département SST/ICT/INMA de l’UCLouvain (hybride présentiel/distanciel) sur le sujet “Minimisation du modèle d’Ambrosio-Tortorelli discret avec les outils proximaux”. Il s’agissait d’un co-encadrement avec Marion Foare (LIP).
- 2020 - 2021 – **Baptiste Pirlot** (UCLouvain, Belgique) a effectué son **Projet de Fin d’Etudes** entre septembre 2020 et juin 2021 (en distanciel) sur le sujet “Quantification d’incertitudes pour le traitement d’images par problèmes inverses”. Il s’agissait sur co-supervision avec Laurent Jacques.
- 2021 – **Hugo Artigas** (Ecole Polytechnique) effectue son **stage de L3**, d’une durée de 4 mois à IMT (hybride présentiel/distanciel) sur le sujet “Intervalle de confiance à partir de l’échantillonnage de trajectoire pour le facteur de reproduction Covid-19”. Il s’agit d’un co-encadrement avec Gersende Fort, Barbara Pascal et Patrice Abry.
- 2021 – **Benjamin Tibi** (ENSTA) effectue son **stage de L3**, d’une durée 3 mois, au LPENSL (hybride présentiel/distanciel) sur le sujet “Analyser le transport de charriage fluvial par minimisation ℓ_1 ”. Il s’agit d’un co-encadrement avec Pierre Borgnat et Mathieu Cassel (EVS, Lyon).
- 2021 – **Assia Chahidi** (M2 International Master of Biometrics and Intelligent Vision, Paris Est Creteil Univ.) effectue son **stage de M2**, d’une durée de 6 mois au CRAL- Observatoire de Lyon (hybride présentiel/distanciel) sur le sujet “Apprentissage profond pour la reconstruction d’environnements circumstellaires”. Il s’agit d’un co-encadrement avec Maud Langlois (CRAL- Observatoire de Lyon).
- 2021 – **Baptiste Desnos** (Ecole Centrale Nantes) effectue son **stage de M2**, d’une durée de 5 mois au LPENSL (distanciel) sur le sujet “Algorithmes déroulés pour la segmentation de texture”. Il s’agit d’un co-encadrement avec Barbara Pascal et Patrice Abry.

1.3 Expertise scientifique

Cette section présente l'ensemble des séminaires nationaux et internationaux dans lesquels j'ai eu l'occasion de présenter mes travaux, les écoles doctorales dans lesquelles je présente les principaux travaux de l'état-de-l'art en optimisation pour la restauration d'images et plus récemment en apprentissage machine. J'indique également mes responsabilités en tant qu'Editrice associée et en tant que relectrice. Enfin j'indique les jurys de thèse dans lesquels j'ai participé en tant qu'examinatrice.

1.3.1 Séminaires invités

- 2009 : **GdR MOA** Mathématiques de l'Optimisation et Applications, (Paris, France).
- 2009 : **SCAM** Séminaire Cristolien d'Analyse Multifractale, (Univ. de Créteil Paris-Est, France).
- 2010 : **CEA – Service Hospitalier Frédéric Joliot**, (Orsay, France).
- 2012 : **MIA**, (Paris, France).
- 2012 : **Séminaire Signal-Apprentissage du L2M**, (Marseille, France)
- 2012 : **Journées Bordelaises d'Analyse Mathématique des Images**, (Bordeaux, France).
- 2013 : **Worshop at Intitute for Pure & Applied Mathematics**, (Los Angeles, U.S.A).
- 2013 : **Groupe de travail image de l' Institut Mathématiques de Toulouse**, (France).
- 2013 : **Mathias 2013, Workshop TOTAL**, (La Défense, France).
- 2013 : **Colloque GDR Multifractale**, (Porquerolles, France).
- 2014 : **Séminaire d'équipe du LM2S**, (Troyes, France).
- 2014 : **Convegno Italo-Francese**, (Sestri-Levante, Italie).
- 2014 : **Mathias 2014, Workshop TOTAL**, (La Défense, France).
- 2014 : **CNES workshop**, Institut Aéronautique et Spatial, (Toulouse, France).
- 2015 : **SMILE workshop** (Paris, France).
- 2016 : **Mathias 2016, Workshop TOTAL**, (Paris, France).
- 2016 : **Séminaire d'équipe** au LBBE (Lyon, France),
- 2017 : **Mathias 2017, Workshop TOTAL**, (Paris, France).
- 2017 : **GdR Phenix/LIA MSD** (LPENSL, France)
- 2018 : **GdR MASCOT-NUM** (Nantes, France).
- 2018 : **Séminaire Parisien de Mathématiques Appliquées à l'Imagerie - IHP** (Paris, France)
- 2018 : **GdR MIA, Sparsity meeting** (Bordeaux, France)
- 2019 : **Statlearn** (Grenoble, France)
- 2019 : **Workshop "Graph signals : learning and optimization"** (Montpellier, France).
- 2019 : **Séminaire à l'Atelier MLima** (Nice, France).
- 2019 : **Séminaire au Dept. of Mathematical Engineering** (UCLouvain, Belgique).
- May 2021 : **One World IMAGIng and INvErse problems (IMAGINE)** (virtuel).
- May 2021 : **Séminaire Français d'Optimisation** (virtuel).
- July 2021 : **17ème colloque de la Société Française de Microscopie**, (virtuel).

1.3.2 Ecoles doctorales

En 2013, nous avons construit un cours avec Jean-Christophe Pesquet (Centrale-Supelec) pour l'école d'été Peyresq sur le thème "Opérateurs maximaux monotones et algorithmes de recherche de zéro, applications en optimisation convexe". Nous avons continué de faire évoluer ce cours dans le cadre d'un Tutoriel IEEE ICASSP 2014. Dans les dernières versions de ce cours, initialement orienté problèmes inverses, j'y intègre désormais les critères rencontrés en apprentissage machine supervisé, les algorithmes stochastiques, et encore plus récemment le principe des algorithmes déroulés.

- **Cours Peyresq 2013** – Il s'agissait d'une école d'été organisée par le GRETSI et le GdR ISIS sur le thème « Les nouvelles techniques d'optimisation : applications en signal, images et télécommunications ». Nous avons partagé 6h de cours avec Jean-Christophe Pesquet (Centrale-Supelec) sur le thème "Opérateurs maximaux monotones et algorithmes de recherche de zéro, applications en optimisation convexe".
- **Tutoriel IEEE ICASSP 2014**, Florence, Italie – Ce tutoriel d'une durée de 3h a été effectué en collaboration avec Jean-Christophe Pesquet (Centrale-Supelec) sur le thème "Monotone operators theory for signal and image processing".
- **Ecole d'été Big Optim**, Grenoble, France, Juin 2015 – [Site web](#) – Ce cours s'inscrivait dans le cadre d'une école intitulée "Optimisation en grande dimension". J'y ai donné 3h de cours sur le thème "Opérateurs maximaux monotones et algorithmes proximaux".
- **Ecole d'été GRAPHSIP**, Aussois, France, Juin 2018 – Ce cours s'inscrivait dans le cadre d'une école intitulée "Signaux sur graphes.". J'y ai donné 6h de cours sur le thème "Optimisation et réseaux".
- **Ecole doctorale MUSICS**, UCLouvain, Belgique, Sept. 2019 – [Site web](#) – Cette école doctorale organise tout au long de l'année des cours à destination des étudiants en Master, des étudiants en thèse, chercheurs et industriels. j'ai eu la possibilité de donner 10h de cours sur le thème "Optimisation en grande dimension en imagerie. Des méthodes de régularisation à l'apprentissage."
- **Ecole doctorale CIRM**, (hybride Marseille/distanciel), Jan. 2021 – [Site web](#) – Cette école avait pour thème : Mathématiques, traitement du signal et apprentissage. J'y ai donné 6h de cours/TP sur le thème "Optimisation".
- **Ecole doctorale Cimpa**, Essaouira, Maroc, Oct. 2021 – [Site web](#) – Cette école a pour thème "Transport optimal, EDP et optimisation : applications en sciences des données". Je suis en charge de la partie "Introduction au traitement d'image, données et graphes". Dans ce cours il sera très peu question d'optimisation, cette partie étant assurée par Jalal Fadili.
- **Mini-cours SMAI -MODE / GdR MOA**, Limoges, France, Juin. 2022 – Nous partagerons avec Audrey Repetti un cours sur le thème "Optimisation pour le traitement d'images".

1.3.3 Activités éditoriales

Entre 2016 et 2020, je fus Editrice associée (AE) pour IEEE Signal Processing Letters. Depuis je suis Editrice associée senior (SAE) qui consiste à identifier des articles devant être rejetés immédiatement et à motiver ce choix ou bien le soumettre au processus de review en indiquant un AE dont les thèmes sont en adéquation avec l'article.

Depuis début 2021, je suis également Editrice associée (AE) pour IEEE TIP.

1.3.4 Activités de relecture

Je participe régulièrement à des expertises, principalement pour les **journaux internationaux** : IEEE Trans. on Image processing, IEEE Trans. on Signal Processing, SIAM Journal on Imaging Sciences, Inverse problems, de l'ordre de 5 reviews par an.

Je participe également à des expertises pour des **conférences nationales et internationales** : GRETSI (depuis 2012), EUSIPCO (depuis 2012), SPARS (2013, 2019), IEEE ICASSP (depuis 2013), IEEE ICIP (depuis 2014), IEEE MLSP (depuis 2015), NIPS (2016-2018). Ceci représente une charge d'environ 20 articles par an.

Je suis également sollicitée par l'ANR pour l'évaluation de **projets JCJC** (4 projets expertisés en 2016, 2017, 2019, 2021).

J'ai également été sollicitée pour l'évaluation d'une **Bourse de thèse** de l'Université de Caen.

1.3.5 Jurys de thèse

Depuis 2016, je suis régulièrement invitée pour faire partie de jurys de thèse en tant qu'examinatrice.

- 2016 – **Odyssée Merveille** (ESIEE, Paris), RORPO : A morphological framework for curvilinear structure analysis. Application to the filtering and segmentation of blood vessels. Directeur de thèse : Hugues Talbot (Centrale-Supelec).
- 2017 – **Yosra Marnissi** (LIGM, Paris), Bayesian methods for inverse problems in signal and image. Directeurs de thèse : Jean-Christophe Pesquet (Centrale-Supelec) et Emilie Chouzenoux (INRIA, Centrale-Supelec).
- 2017 – **Federico Pierucci** (LJK, Grenoble), Nonsmooth Optimization for Statistical Learning with Structured Matrix Regularization. Directeurs de thèse : Anatoli Iouditski (LJK), Zaid Harchaoui (Univ. Washington), Jérôme Malick (CNRS, LJK)
- 2018 – **Eric Woillez** (ENS Lyon), Stochastic description of rare events for complex dynamics in the solar system. Directeur de thèse : Freddy Bouchet (CNRS, LPENSL)
- 2020 – **Daniel Martins Antunes** (Univ. Savoie), Contraintes géométriques et approches variationnelles pour l'analyse d'image. Directeur de thèse : Jacques-Olivier Lachaud (Univ. Savoie)
- 2021 – **Stéphanie Guérit** (UCLouvain), Regularized inverse problems in biomedical imaging.
- 2021 – **Quentin Klopfenstein** (IMB, Dijon), Non-smooth optimization for the estimation of cellular immune components in a tumoral environment. Directeur de thèse : Samuel Vaïter (CNRS, Institut de Mathématiques de Bourgogne).
- 2021 – **I. El Bouchairi** (GREYC, Caen). Directeur de thèse : Jalal Fadili.

1.4 Animation de la recherche

Je participe à l'animation de la recherche au niveau régional, national et international. Je détaille ci-dessous chacune d'entre elles.

1.4.1 En France

- 2011-2013 : **SIERRA** (Signal et Images En Région Rhône Alpes) :
 - Montage du dossier auprès de l'ARC6 (financement régional) pour organiser pendant deux ans des séminaires itinérants en région Rhône-Alpes.
 - Webmaster du [site web de ces journées](#).
 - Organisation d'une journée à l'ENS de Lyon en Mars 2013 sur la thématique « Avancées récentes autour de l'extraction de modes ».
- 2013-2014 : **GdR ISIS** : Co-responsable de l'*Action Optimisation* avec Saïd Moussaoui.
Journées organisées :
 - Sep 2013 : *Optimisation convexe sous contraintes*, Paris
 - Mar. 2014 : *Optimisation non-convexe*, Paris

- Oct. 2014 : *Optimisation non-convexe*, Paris
- Nov. 2014 : *Optimisation géométrique sur les variétés*, Paris (journée conjointe GdR MIA).
- 2015 : **GRETSI** : Comité d'organisation. Application mobile.
- 2017 : **Journées SMAI MODE** : Comité d'organisation. Sponsors industriels.

1.4.2 A l'étranger

- **IEEE ICIP 2014** : J'ai pris part au comité d'organisation en tant que « IEEE Student Activities Chair » :
 - Sélection des étudiants ayants droit à une bourse IEEE pour leur déplacement à ICIP. Expertise de 40 dossiers dont 20 acceptés.
 - Sélection des 46 volontaires s'occupant de l'accueil des participants, de l'orientation, de l'installation et la désinstallation du matériel dans les salles et du bon déroulement de chaque session. Création des emplois du temps pour les 46 participants. Création du document décrivant précisément chaque tâche.
 - Organisation de la rencontre entre professionnels et étudiants (« student-get-together meeting»). Industriels intervenants : Microsoft research, Google, NASA, Technicolor, DxO.
- **STATOS 2016** : Je faisais partie du Comité de programme. Ma tâche a principalement consisté à identifier des intervenants et à inviter certains d'entre eux.
- **IEEE IVMSPP 2016** : J'ai pris part au comité d'organisation en tant que "Publication Chair". J'étais en charge de proposer des templates compatibles au consignes IEEE puis de vérifier leur viabilité après soumission.
- Présidente d'une session sur l'apprentissage à IEEE ICASSP 2017.
- **IEEE MLSP 2018** (Aalborg, Denmark) : J'ai pris part au comité d'organisation en tant que "Program chair" en collaboration avec Zhanyu Ma (BUPT-Beijing). Nos responsabilités étaient les suivantes :
 - Répartition des articles soumis aux reviewers.
 - Sélectionner les articles acceptés sur base des expertises des reviewers et devant tenir compte de contraintes IEEE sur le nombre d'articles acceptés.
 - Identifier les thèmes des sessions et proposer le planning associé.
- **AIP 2019** (Grenoble, France) : Minisymposium sur le thème "Variational methods for inverse problems in image processing" co-organisé avec L. Condat (Gipsa-Lab).
- **LearnGraph 2022** (CIRM, Marseille, France) : Participation à l'organisation du Workshop sur le thème "Machine Learning and Signal Processing on Graphs" porté par Samuel Vaiter (CNRS, IMB, Dijon).
- **JDS 2022** (Lyon, France) : Participation au comité scientifique. Proposition d'orateurs pléniérs et relecture d'une vingtaine de communications.
- **IEEE ICIP 2022** (Bordeaux, France) : Organizing committee as "Tutorial chair".

1.5 Administration de la recherche

Cette section fait état de mes participations dans des comités de sélection mais également ma participation active au sein du comité de direction du GdR ISIS et d'autres comités nationaux ou internationaux.

1.5.1 En France

- **Comités de sélection** poste de Maître de Conférence et Agrégé Préparateur
 - 2013 McF : LJK Grenoble, profil Optimisation (26).
 - 2014 McF : ENSEEIHT Toulouse, profil Traitement du signal et des images (61).
 - 2014 McF : L2S Orsay, profil Traitement d’images astrophysique (61).
 - 2015 McF : INSA de Rouen, profil Apprentissage statistique et science des données (61).
 - 2016 McF : Université de Bourgogne, profil Contrôle, Optimisation et Transport (26).
 - 2018 Agrégé préparateur : LPENSL.
 - 2018 McF : Université de Bordeaux, profil Optimisation et traitement d’images (26).
 - 2019 McF : Ensimag, profil Traitement d’images (61)
 - 2020 McF : FIMI-CREATIS, profil Problèmes inverses (26, 27).
 - 2020 McF en tenure track : CentraleSupélec, profil Problèmes inverses (26, 61).
 - 2021 McF : SeaTech, profil Signal, image, vision, data, apprentissage (61).
 - 2021 McF : Université de Savoie Mont-Blanc, profil Traitement et analyse d’images (26).
- Depuis 2015, je suis membre du **Comité de direction du GdR ISIS** et responsable des relations Inter-GdR, internationales, et ouverture vers l’extérieur. Mes missions sont :
 - d’échanger avec les autres GdRs qui souhaiteraient organiser des journées communes.
 - la participation à l’évaluation des projets du GdR ISIS avec les autres membres du Comité de direction.
 - la gestion de la gazette du GdR ISIS. Lorsque Laure Blanc-Féraud (CNRS, I3S) était à la direction nous avons fait revivre la Gazette du GdR ISIS, sous forme électronique. Celle-ci est envoyée tous les 4-5 mois. Elle contient un Edito et annonce les actualités du GdR ISIS, des sociétés savantes et associations qui l’entourent (Club EEA, AMIES, GRETSI), de l’Académie des sciences, du club des partenaires. Elles récapitule les informations contenues dans le kiosque. Cette gazette est créée en collaboration avec Michel Jordan (ENSEA) et la direction du GdR ISIS.
 - la cartographie des compétences du GdR ISIS. Lors de l’AG de 2017, une cartographie des compétences avait été effectuée grâce à une analyse par graphe sur la base de données des membres du GdR ISIS. Nous avons regardé comment avaient évolué les compétences au cours du temps entre 1988 et 2016. Les affiches créées sont disponibles [sur ce lien](#). En 2021, suite à une demande du CNRS, en collaboration avec J. Idier (CNRS, Ecole Centrale Nantes), nous avons identifié un nombre de mots-clés discutés puis validés auprès du comité de direction, du club des partenaires et de l’INSII. La prochaine étape concerne la représentation des statistiques associées par exemple sous forme de carte sur le site du GdR ISIS.
- Depuis Juin 2021 : Secrétaire du **Comité de liaison du groupe SMAI-MODE**.

1.5.2 A l’étranger

- Depuis 2016 : Membre du comité technique **IEEE MLSP**. J’ai participé au sous-comité “Visibilité” puis “Award”.
- 2015-2018 : Membre du comité **EURASIP SAT Signal and Data Analytics for Machine Learning**.

1.6 Contrats et projets de recherche

1.6.1 PI

- 2019-2024 – **Projet ANR JCJC Multisc-In** portant sur l'estimation multi-échelle et la détection d'interfaces. Le financement est de 182k€. Les chercheurs partenaires à ce projet sont M. Foare (CPE Lyon, ENS de Lyon), Luis Briceño-Arias (Chili), L. Condat (GIPSA-lab puis KAUST). Hoang Trieu Vy Le (doctorant) a rejoint le projet dans le cadre d'une thèse.
- 2017-2019 – **Défi CNRS Imag'In SIROCCO** portant sur la segmentation de série d'images d'un financement de 15k€. Les chercheurs partenaires à ce projet étaient L. Condat (KAUST), H. Wendt (IRIT), V. Vidal, P. Abry, J.-C. Gémard.
- Avril-Juillet 2018 – **Equipe conseil pour la société AZOTH SYSTEMS** sur le thème Apprentissage supervisé au service de la prévention des accidents de décompression en plongée. Budget : 6 keuros. En collaboration avec Patrick Flandrin.
- 2016-2017 – **Financement industriel TOTAL** portant sur l'Optimisation pour l'apprentissage supervisé parcimonieux. Budget : 100k€ (financement d'un an post-doctorat + environnement). Prototypage transmis dans l'antenne TACS de TOTAL.
- 2014-2015 – **JCJC GdR ISIS GALILEO** sur le thème Segmentation de signaux et images par analyse d'invariance d'échelle, régularité locale et optimisation convexe non-lisse. Il s'agissait d'un financement de 5k€. Les chercheurs partenaires à ce projet étaient H. Wendt (IRIT), N. Dobigeon (IRIT) et P. Abry (LPENSL).

1.6.2 Membre

- 2007-2010 – **ANR OPTIMED** dont le sujet est "Algorithmes d'optimisation décomposés pour les problèmes d'imagerie médicale de grande taille." Les laboratoires impliqués étaient le LIGM (PI : J.-C. Pesquet), le LS2N, le CEA SHFJ, et le CEA Neurospin.
- 2012-2013 – **PEPS Bio-Maths-Info PROMIS** dont le sujet est "Proximal algorithms for Structured Illumination Microscopy". Les participants étaient J. Boulanger –PI– (Institut Curie, Paris) et L. Condat (Gipsa-lab, Grenoble).
- 2014-2017 – **ANR ASTRES** dont le sujet est "Analyse, Synthèse et Transformations par Réallocation, EMD et Synchrosqueezing". Les laboratoires impliqués étaient le LPENSL (PI : P. Flandrin), le LJK, le Gipsa-lab et l'IREENA.
- 2014 - 2017 – **ANR GRAPHSIP** dont le sujet est "Traitement de signaux sur graphes". Les laboratoires impliqués étaient UNICAEN (PI : O. Lazoray), le GIPSA-lab, le LIGM, et le LPENSL.
- 2015 - 2016 – **Défi Imag'In OPTIMISTE** dont le sujet est "Optimisation stochastique en imagerie hyperspectrale". Les laboratoires impliqués étaient le LIGM (PI : J.C. Pesquet), le LJLL, le LPENSL, l'INT, l'I2M, et l'IRIT.
- 2015 - 2016 – **Défi InFinITI CNRS OptMatGeo** dont le sujet est "Optimisation non lisse des matrices de transfert pour la géographie électorale". Les laboratoires impliqués sont le LPENSL (PI : P. Jensen) et Sciences Po Paris.
- 2017-2019 – **Projet Mastondons AGADIR** dont le sujet est "Adaptation et Graphes dynamiques pour l'Apprentissage Distribué sur les grands Réseaux". Les laboratoires impliqués étaient les Laboratoire Joseph-Louis Lagrange (PI : C. Richard), CVN, GIPSA-lab, LPENSL, L2S.
- 2017-2021 – **ANR MULTIFRACS** dont le sujet est "Théorie et méthodes multifractales pour les systèmes multivariés de grande taille - Applications à l'analyse libre à l'échelle de l'analyse de la dynamique cérébrale". Les laboratoires impliqués sont le LPENSL (PI : P. Abry), le CEA/DRF/NEUROSPIN, l'IRIT et le LAMA.

- 2018-2021 – **SBI ACADEMICS** dont le sujet est “Apprentissage machine et science des données pour les modèles complexes et dynamiques”. Les laboratoires impliqués sont le LPENSL (PI : P. Borgnat), l’ICJ, le LIP, le LIRIS et le Laboratoire Hubert Curien.
- 2019-2021 – **Pack Ambition Recherche GREENDAM** dont le sujet est “Évaluation de la transparence sédimentaire des ouvrages hydro-électriques, éléments opérationnels pour promouvoir une énergie hydraulique vertueuse”. Les laboratoires impliqués sont EVS (PI : H. Piégay) et le LPENSL.
- 2019-2023 – **ANR NanoPoRep** dont le sujet est “Profilage génomique haute performance de la réplication de l’ADN par séquençage nanopore”. Les laboratoires impliqués sont le LPENSL (PI : B. Audit) et l’IBENS.

1.7 Enseignement

Je mentionne ici uniquement les enseignements effectués après ma thèse. Entre 2012 et 2015, j’ai régulièrement été sollicitée pour participer à des enseignements en traitement du signal, problèmes inverses et optimisation : cours de traitement du signal à l’ENS de Lyon au niveau M1 Science de la Matière (18h de cours), cours d’optimisation dans le M2 Maths en Action de l’Univ. Lyon1 (6h de cours), cours d’optimisation de 12h à CPE Lyon, cours d’optimisation sur graphes aux étudiants de l’ENS de Lyon dans le Master Systèmes Complexes (10h). Je rapporte régulièrement des rapports de stage dans les différentes formations au sein de l’ENS de Lyon. Je participe également à des jurys de soutenance de stage. Plus d’informations pour chacun de ces cours sont données ci-dessous :

- **Traitement du signal** dans le M1 Science de la Matière à l’ENSL (Cours partagé avec P. Borgnat). Il s’agit de 12h de cours et 6h de TP. J’ai effectué ce cours tous les ans entre 2013 et 2016 au 2nd semestre. Le contenu du cours était le suivant : Filtrage, estimation spectrale non-paramétrique/paramétrique, temps-fréquence, ondelettes. TP sur le signal analytique, TP sur l’estimation spectrale, TP sur les ondelettes.
- **Optimisation pour la statistique en grande dimension appliquée à la génomique** (Cours partagé avec F. Picard (CNRS, LBBE), V. Viallon (CNRS, ICJ), L. Jacob (CNRS, LBBE)) dans le M2 Maths en Action de l’Université Lyon 1. Il s’agissait initialement de 4h de cours et de 2h de TP mais depuis 2 ans il est composé de 6h de cours et 4h de TP. J’enseigne ce cours tous les ans depuis 2014, au 2ème semestre. Le contenu du cours est le suivant : Optimisation convexe différentiable, sous-différentielle, conjuguée, opérateur proximal, algorithme forward-backward. TP sur l’implémentation des méthodes proximales pour un problème de détection de changements et de sélection de variables en génomique. Plus récemment j’ai ajouté à ce cours les algorithmes stochastiques et les algorithmes déroulés.
- **Optimisation sur graphes** (Cours partagé avec P. Borgnat) dans le M2 Systèmes Complexes de l’ENSL. Il s’agissait de 6h de cours, 4h de TP. J’ai enseigné ce cours du 1er semestre entre 2014 et 2019. Son intitulé et contenu a évolué chaque année. Le contenu du cours est le suivant : Optimisation convexe différentiable, sous-différentielle, conjuguée, opérateur proximal, algorithme forward-backward, algorithme Douglas-Rachford, algorithme primal-dual dont ADMM. TP sur l’implémentation des méthodes proximales pour des problèmes de débruitage de maillage 3D.
- **Introduction à l’optimisation convexe** pour les 4ème année ETI de CPE Lyon. Il s’agissait de 4h de cours sur la période 2015-2018. Le contenu du cours était le suivant : optimisation différentiable, optimisation non-différentiable. TP de restauration d’images (inversion directe, Tikhonov, restauration TV).
- **Problèmes inverses** (Cours partagé avec L. Condat) dans le M2 SIGMA, Grenoble-INP. Il s’agissait de 4h de cours sur la période 2015-2018. Le contenu du cours était le suivant : introduction aux problèmes inverses, exemples, critères rencontrés, algorithmes proximaux (forward-backward, algorithme primal-dual dont ADMM). TP de restauration d’images.

- **Apprentissage machine** (Cours partagé avec P. Borgnat, P. Abry et F. Picard) dans le M2 Systèmes Complexes de l'ENSL. Il s'agit de 6h de cours. Il s'agit d'un cours du 1er semestre débuté en 2020. Le contenu du cours est le suivant : Introduction aux critères rencontrés en machine learning, optimisation convexe différentiable, sous-différentielle, conjuguée, opérateur proximal, algorithme forward-backward, algorithme primal-dual dont ADMM, algorithmes stochastiques, algorithmes déroulés.
- **Fondements mathématiques de l'apprentissage profond** (Cours partagé avec R. Gribonval (INRIA, LIP) et A. Garivier(UMPA)) dans le M2R Maths de l'ENSL. Il s'agit de 6h de cours. Il s'agit d'un cours du 2nd semestre débuté en 2020. Le contenu du cours est le suivant : Introduction générale à l'optimisation et aux schémas proximaux, classe des fonctions convexes s.c.i. et propres, coercivité, existence et unicité, sous-différentielle, conjuguée, opérateur proximal, convergence faible/forte, opérateur Lipschitz continu, Théorème de Banach-Picard, opérateurs moyennés, Théorème convergence faible pour opérateurs moyennés. Convergence faible de gradient et du point proximal, puis de forward-backward par composition d'opérateurs alpha-moyennés, taux de convergence pour FB dans le cas fortement convexe, worst-case complexity, descente de gradient/forward-backward stochastique, algorithmes déroulés : forward-backward déroulé, LISTA, lien entre prox et fonctions d'activations, robustesse de réseau.

1.8 Responsabilités administratives

- 2021 : Recensement des thèmes de recherche en IA au sein du LPENSL.
- 2021 : Création du site web l'ANR JCJC MULTISC-IN.
- Depuis 2020 : Gestion du site web des séminaires MLSP de l'ENSL.
- Depuis 2020 : Gestion du site web de l'Equipe SiSyPhe du LPENSL.
- 2020 : Présentation orale devant le comité d'évaluation HCERES 2020 en collaboration avec V. Vidal sur le projet SIROCCO (collaboration physique/traitement du signal).
- 2015-2019 : Membre nommée du Conseil de Laboratoire du LPENSL.
- 2013-2019 : Correspondante bibliothèque (gestion d'un budget de 8 k€ pour les ouvrages et 20 k€ pour les périodiques)
- 2015 : Présentation orale devant le comité d'évaluation HCERES 2015.
- 2015 : En charge de la rédaction du chapitre "Instrumentation et imagerie" du rapport HCERES 2015 en collaboration avec Romain Volk (LabPhy ENSL).
- 2012-2015 : Création et gestion du site web l'ANR ASTRES.
- 2008-2010 : Responsable du site web de l'ANR OPTIMED.

1.9 Vulgarisation scientifique

- Article dans HIGHLIGHTS 2015, LPENSL.
- Article dans Wiley Encyclopedia of Electrical and Electronics Engineering, Feb. 2016, Wavelet-based Image Deconvolution and Reconstruction.
- Couverture HIGHLIGHTS 2020, LPENSL avec une image obtenue dans la collaboration du Defi Imag'In SIROCCO.
- Accueil de collégiens.

1.10 Publications

Journaux internationaux acceptés :

- [R1] L. Denneulin, M. Langlois, E. Thiebaut, and N. Pustelnik “*RHAPSODIE : Reconstruction of High-contrast Polarized Sources and Deconvolution for Circumstellar Environments*,” accepted to **A&A**, 2021.
- [R2] B. Pascal, S. Vaiteer, N. Pustelnik, and P. Abry, “*Automated data-driven selection of the hyperparameters for Total-Variation based texture segmentation*,” accepted to **Journal of Mathematical Imaging and Vision**, 2021.
- [R3] M. Jiu and N. Pustelnik, “*A deep primal-dual proximal network for image restoration*,” **IEEE JSTSP Special Issue on Deep Learning for Image/Video Restoration and Compression**, vol. 5, no. 2, pp. 190–203, 2021.
- [R4] E. Horne, J. Schmitt, N. Pustelnik, S. Joubaud, P. Odier “*Variational Mode Decomposition for estimating critical reflected internal wave in stratified fluid*,” **Experiments in Fluids** vol. 62, no. 5, pp. 1–13, 2021.
- [R5] B. Pascal, N. Pustelnik, and P. Abry, “*Strongly Convex Optimization for Joint Fractal Feature Estimation and Texture Segmentation*,” **Applied and Computational Harmonic Analysis**, vol. 54, pp 303-322, 2021.
- [R6] P. Abry, N. Pustelnik, S. Roux, P. Jensen, P. Flandrin, R. Gribonval, C.-G. Lucas, E. Guichard, P. Borgnat, N. Garnier, B. Audit, “*Spatial and temporal regularization to estimate COVID-19 Reproduction Number $R(t)$: Promoting piecewise smoothness via convex optimization*,” **PLoS One**, vol. 15, no. 8, Aug. 2020.
- [R7] B. Pascal, N. Pustelnik, P. Abry, J.-C. G  minard, V. Vidal, “*Parameter-free and fast non-linear piecewise filtering. Application to experimental physics*,” **Annals of Telecommunications**, vol. 75, pp. 655-671, Oct. 2020.
- [R8] M. Foare, N. Pustelnik, and L. Condat, “*Semi-linearized proximal alternating minimization for a discrete Mumford-Shah model*,” **IEEE Trans. on Image Processing**, vol. 29, pp 2176-2189, Oct. 2019.
- [R9] Y. Kaloga, M. Foare, N. Pustelnik, and P. Jensen, “*Discrete Mumford-Shah on graph for mixing matrix estimation*,” **IEEE Signal Processing Letters**, vol. 26, no. 9, pp. 1275-1279, Sept. 2019.
- [R10] J. Colas, N. Pustelnik, C. Oliver, J.-C. G  minard, V. Vidal, “*Nonlinear denoising for solid friction dynamics characterization*,” **Physical Review E**, 100, 032803, Sept. 2019.
- [R11] M. Jiu, N. Pustelnik, S. Janaqi, M. Chebre, P. Ricoux, “*Sparse hierarchical interaction learning with epigraphical projection*,” **Journal of Signal Processing Systems**, vol. 92, pp. 637–654, 2019.
- [R12] J. Boulanger, N. Pustelnik, L. Condat, T. Piolot, L. Sengmanivong, “*Nonsmooth convex optimization for Structured Illumination Microscopy image reconstruction*,” **Inverse problems**, vol. 34, no. 9, 22 pp., July 2018.
- [R13] P. Abry, J. Spilka, R. Leonarduzzi, V. Chudacek, N. Pustelnik, M. Doret, “*Sparse learning for Intrapartum fetal heart rate analysis*,” **Biomedical Physics & Engineering Express**, vol. 4, no. 3, pp. 034002, 2018.
- [R14] N. Pustelnik, L. Condat, “*Proximity Operator of a Sum of Functions; Application to Depth Map Estimation*,” **IEEE Signal Processing Letters**, vol. 24, no. 12, pp. 1827 - 1831, Dec. 2017.
- [R15] J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, and P. Abry, “*Bayesian selection for the regularization parameter in TVl0 denoising problems*,” **IEEE Trans. on Signal Processing**, vol. 65, no. 25, pp. 5215 - 5224, Jun. 2017.
- [R16] J. Spilka, J. Frecon, R. Leonarduzzi, N. Pustelnik, P. Abry, M. Doret, “*Sparse Support Vector Machine for Intrapartum Fetal Heart Rate Classification*,” **IEEE Journal of Biomedical and Health Informatics**, vol. 21, no. 3, pp. 664 - 671, May 2017.
- [R17] G. Michau, N. Pustelnik, P. Borgnat, P. Abry, A. Nantes, A. Bhaskar, E. Chung, “*A primal-dual algorithm for link dependent origin destination matrix estimation*,” **IEEE Transactions on Signal and Information Processing over Networks**, vol. 3, no. 1, pp. 104 - 113, Mar. 2017.
- [R18] N. Pustelnik, H. Wendt, P. Abry, N. Dobigeon, “*Local regularity, wavelet leaders and total va-*

- riation based procedures for texture segmentation,”* accepted to **IEEE Trans. on Computational Imaging**, vol. 2, no. 4, pp. 468–479, Dec. 2016.
- [R19] J. Frecon, G. Didier, N. Pustelnik, and P. Abry, “*Non-Linear Wavelet Regression and Branch and Bound Optimization for the Full Identification of Bivariate Operator Fractional Brownian Motion,*” accepted to **IEEE Trans. on Signal Processing**, vol. 64, no. 15, pp. 4040–4049, Aug. 2016.
- [R20] J. Frecon, N. Pustelnik, P. Abry, and L. Condat, “*Fast and On-the-fly Approximation of Multivariate Total Variation Minimization,*” **IEEE Trans. on Signal Processing**, vol. 64, no. 9, pp. 2355–2364, May 2016.
- [R21] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, J.-C. Pesquet, “*Wavelet-based Image Deconvolution and Reconstruction,*” accepted to **Wiley Encyclopedia of Electrical and Electronics Engineering**, 2016. Tutorial paper.
- [R22] G. Chierchia, N. Pustelnik, J.-C. Pesquet, B. Pesquet-Popescu, “*Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools,*” **Signal, Image and Video Processing**, vol.9, no. 8, pp.1737–1749, Nov. 2015.
- [R23] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, J.-C. Pesquet, “*A Non-Local Structure Tensor Based Approach for Multicomponent Image Recovery Problems,*” **IEEE Trans. Image processing**, Vol. 23, no. 12, pp. 5233–5248, Oct. 2014.
- [R24] J. Schmitt, N. Pustelnik, P. Borgnat, P. Flandrin, and L. Condat, “*A 2-D Prony-Huang Transform : A New Tool for 2-D Spectral Analysis,*” **IEEE Trans. Image processing**, Vol. 23, no. 12, pp. 5531–5544, Oct. 2014.
- [R25] C.R. Johnson, P. Messier, W.A. Sethares, A.G. Klein, C. Brown, A.H. Do, P. Klausmeyer, P. Abry, S. Jaffard, H. Wendt, S. Roux, N. Pustelnik, N. van Noord, L. van der Maaten, E. Potsma, J. Coddington, L.A. Daffner, H. Murata, H. Wilhelm, S. Wood, M. Messier, “*Pursuing automated classification of historic photographic papers from raking light photomicrographs,*” **Journal of the American Institute for Conservation**, vol. 53, no. 3, pp. 159–170, 2014.
- [R26] N. Pustelnik, P. Borgnat, P. Flandrin “*Empirical Mode Decomposition revisited by multicomponent non smooth convex optimization,*” **Signal Processing**, vol. 102, pp. 313–331, Sept. 2014.
- [R27] Y. Berthoumieu, Ch. Dossal, N. Pustelnik, F. Turcu et Ph. Ricoux, “*An evaluation of the sparsity degree for sparse recovery with deterministic measurement matrices,*” **Journal of Mathematical Imaging and Vision**, vol. 48, pp. 266–278, 2013.
- [R28] J.-C. Pesquet and N. Pustelnik, “*A Parallel Inertial Proximal Optimization Method,*” **Pacific Journal of Optimization**, vol. 8, no. 2, pp. 273–305, Apr. 2012.
- [R29] N. Pustelnik, J.-C. Pesquet, and C. Chaux, “*Relaxing Tight Frame Condition in Parallel Proximal Methods for Signal Restoration,*” **IEEE Transactions on Signal Processing**, vol. 60, no. 2, pp. 968–973, Feb. 2012.
- [R30] N. Pustelnik, C. Chaux, and J.-C. Pesquet, “*Parallel ProXimal Algorithm for image restoration using hybrid regularization,*” **IEEE Transactions on Image Processing**, vol. 20, no. 9, pp. 2450–2462, Sep. 2011.
- [R31] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, “*Proximal algorithms for multicomponent image processing,*” **Journal of Mathematical Imaging and Vision**, vol. 41, no. 1, pp. 3–22, Sep. 2011.

Journaux nationaux :

- [R1] L. Chaari, E. Chouzenoux, N. Pustelnik, C. Chaux et S. Moussaoui, “*OPTIMED : Optimisation itérative pour la résolution de problèmes inverses de grande taille,*” **Traitement du signal**, Vol. 28, No. 3-4, pp. 329–374, 2011.

Conférences internationales :

- [C1] B. Pascal, V. Mauduit, N. Pustelnik, P. Abry “*Scale-free Texture Segmentation : Expert Feature-based versus Deep Learning strategies,*” **EUSIPCO 2020**, Amsterdam, Pays-Bas, Jan. 18-21,

- 2021.
- [C2] L. Denneulin, M. Langlois, N. Pustelnik, I. Loris, E. Thibaut “*Primal-dual splitting scheme with backtracking for handling with epigraphic constraint and sparse analysis regularization*,” **iTwist**, Nantes, France, Dec. 2-4, 2020.
 - [C3] B. Pascal, N. Pustelnik, P. Abry, M. Serres, V. Vidal “*Joint estimation of local variance and local regularity for texture segmentation. Application to multiphase flow characterization*,” **IEEE ICIP**, Athens, Greece, Oct. 7-10, 2018.
 - [C4] M. Jiu, N. Pustelnik, L. Qi “*Multiclass SVM With Hierarchical Interaction : Application To Face Classification*,” **IEEE MLSP**, Aalborg, Danemark, Sept. 17-20, 2018.
 - [C5] N. Pustelnik, L. Condat, “*Proximity operator of a sum of functions : Application to image segmentation*,” **SIAM Conference on Imaging Science**, Bologna, Italy June, 5-8 2018.
 - [C6] B. Pascal, N. Pustelnik, P. Abry, “*Combining Local Regularity Estimation and Total Variation Optimization for Scale-Free Texture Segmentation*,” **SIAM Conference on Imaging Science**, Bologna, Italy June, 5-8 2018.
 - [C7] M. Foare, N. Pustelnik, L. Condat, “*Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford-Shah Model*,” **SIAM Conference on Imaging Science**, Bologna, Italy June, 5-8 2018.
 - [C8] M. Foare, N. Pustelnik, L. Condat “*A new proximal method for joint image restoration and edge detection with the Mumford-Shah model*,” **IEEE ICASSP**, Calgary, Alberta, Canada, Apr. 15-20, 2018.
 - [C9] B. Pascal, N. Pustelnik, P. Abry, J.-C. Pesquet “*Block-coordinate proximal algorithms for scale-free texture segmentation*,” **IEEE ICASSP**, Calgary, Alberta, Canada, Apr. 15-20, 2018.
 - [C10] J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, and P. Abry “*Bayesian-driven criterion to automatically select the regularization parameter in the l1-Potts model*,” **IEEE ICASSP**, New-Orleans, USA, May, 5-9 2017.
 - [C11] G. Chierchia, N. Pustelnik, J.-C. Pesquet “*Random primal-dual proximal iterations for sparse multiclass SVM*,” **IEEE MLSP**, Vietri sul Mare, Salerno, Italy, Sep. 13-16, 2016
 - [C12] M. Jiu, N. Pustelnik, M. Chebre, S. Janaqi, P. Ricoux “*Multiclass SVM with graph path coding regularization for face classification*,” **IEEE MLSP**, Vietri sul Mare, Salerno, Italy, Sep. 13-16, 2016.
 - [C13] J. Frecon, N. Pustelnik, H. Wendt, L. Condat, and P. Abry “*Multifractal-based texture segmentation using variational procedure*,” **IEEE IVMS**, Bordeaux, France, Jul., 11-12 2016.
 - [C14] J. Frecon, R. Fontugne, G. Didier, N. Pustelnik, K. Fukuda, and P. Abry “*Non-linear regression for bivariate self-similarity identification - Application to anomaly detection in Internet traffic based on a joint scaling analysis of packet and byte counts*,” **IEEE ICASSP**, Shanghai, China, Mar, 20-25 2016.
 - [C15] P. Flandrin, N. Pustelnik, P. Borgnat, “*On Wigner-based sparse time-frequency distributions*,” **IEEE CAMSAP**, Cancun, Mexico, Dec. 13-16 2015.
 - [C16] J. Frecon, N. Pustelnik, H. Wendt, and P. Abry “*Multivariate optimization for multifractal-based texture segmentation*,” **IEEE ICIP**, Quebec City, Canada, Sept, 27-30 2015.
 - [C17] J. Schmitt, E. Horne, N. Pustelnik, S. Joubaud, P. Odier “*An improved variational mode decomposition method for internal waves separation*,” **EUSIPCO**, Nice, France, Aug. 31- Sep. 4 2015.
 - [C18] J. Spilka, J. Frecon, R.F. Leonarduzzi, N. Pustelnik, P. Abry, M. Doret “*Intrapartum Fetal Heart Rate classification from Trajectory in Sparse SVM feature space*,” **IEEE EMBC**, Milan, Italy, Aug. 25-29 2015.
 - [C19] R.F. Leonarduzzi, J. Spilka, J. Frecon, H. Wendt, N. Pustelnik, S. Jaffard, P. Abry, M. Doret “*P-leader multifractal analysis and sparse SVM for intrapartum fetal acidosis detection*,” **IEEE EMBC**, Milan, Italy, Aug. 25-29 2015.
 - [C20] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, “*An Epigraphic Splitting Technique for Sparse Multiclass SVM*,” **SPARS**, Cambridge, UK, July 6-9, 2015.
 - [C21] G. Michau, P. Borgnat, N. Pustelnik, P. Abry, A. Nantes, E. Chung, “*Estimating link-dependen*

- origin-destination matrices from sample trajectories and traffic counts*," **IEEE ICASSP**, Brisbane, Australia, April 19-24, 2015
- [C22] N. Pustelnik, P. Abry, H. Wendt, and N. Dobigeon, "Inverse problem formulation for regularity estimation in images," **IEEE ICIP**, La Défense, Paris, France, October 27-30, 2014. Top 10% papers.
- [C23] J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, and P. Abry, "Hybrid Bayesian variational scheme to handle parameter selection in total variation signal denoising," **EUSIPCO**, Lisbon, Portugal, Sept, 1-5 2014.
- [C24] J. Schmitt, N. Pustelnik, P. Borgnat, and P. Flandrin, "2D Hilbert-Huang Transform," **IEEE ICASSP**, Florence, Italy, May 4-9, 2014
- [C25] G. Cherchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphic proximal projection for sparse Multiclass SVM," **IEEE ICASSP**, Florence, Italy, May 4-9, 2014.
- [C26] J. Boulanger, N. Pustelnik, L. Condat, "Non-smooth convex optimization for an efficient reconstruction in structured illumination microscopy," **IEEE ISBI**, Beijing, China, April 28-May 2, 2014.
- [C27] L. Condat, J. Boulanger, N. Pustelnik, S. Sahnoun, L. Sengmanivong "A 2-D spectral analysis method to estimate the modulation parameters in structured illumination microscopy," **IEEE ISBI**, Beijing, China, April 28-May 2, 2014.
- [C28] N. Saulig, N. Pustelnik, P. Borgnat, P. Flandrin, and V. Sucic, "Instantaneous counting of components in nonstationary signals," **EUSIPCO**, Marrakech, Morocco, Sept. 9-13, 2013. Invited paper.
- [C29] N. Pustelnik, H. Wendt, and P. Abry, "Local regularity for texture segmentation : combining wavelet leaders and proximal minimization," **IEEE ICASSP**, Vancouver, Canada, May 26-31, 2013.
- [C30] G. Cherchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "An epigraphical convex optimization approach for multicomponent image restoration using non-local structure tensor," **IEEE ICASSP**, Vancouver, Canada, May 26-31, 2013
- [C31] N. Pustelnik, P. Borgnat, and P. Flandrin, "A multicomponent proximal algorithm for Empirical Mode Decomposition," **EUSIPCO**, Bucharest, Romania, August, 27-31, 2012.
- [C32] N. Pustelnik, C. Dossal, F. Turcu, Y. Berthoumieu, and Ph. Ricoux, "A greedy algorithm to extract sparsity degree for l_1/l_0 -equivalence in a deterministic context," **EUSIPCO**, Bucharest, Romania, August, 27-31, 2012.
- [C33] G. Cherchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "A proximal approach for constrained cosparsity modelling," **IEEE ICASSP**, Kyoto, Japan, March 25-30, 2012.
- [C34] C. Chaux, C. Comtat, J.-C. Pesquet, and N. Pustelnik, "Dynamic PET Reconstruction using Parallel ProXimal Algorithm," **SIAM Conference on Optimization**, Darmstadt, Germany, May 16-19, 2011. Invited paper.
- [C35] N. Pustelnik, C. Chaux, J.-C. Pesquet, and C. Comtat, "Parallel Algorithm and Hybrid Regularization for Dynamic PET Reconstruction," **IEEE MIC**, Knoxville, Tennessee, Oct. 30 - Nov. 6 2010.
- [C36] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, "Proximal method for geometry and texture image decomposition," **IEEE ICIP**, Honk Kong, 26-29 Septembre 2010.
- [C37] N. Pustelnik, J.-C. Pesquet, and C. Chaux, "Proximal methods for image restoration using a class of non-tight frame representations," **EUSIPCO**, Aalborg, Denmark, 23-27 August 2010.
- [C38] C. Chaux, J.-C. Pesquet, and N. Pustelnik, "Frame-based proximal algorithms for Poisson data recovery," **SIAM Conference on Imaging Science**, Chicago, Illinois, April 12-14 2010. Invited paper.
- [C39] 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," **Fully3D**, Beijing, China, September 5-10, 2009.
- [C40] L. Chaari, N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Solving inverse problems with over-complete transforms and convex optimization techniques," **SPIE**, San Diego, California, USA

- , August 2-6, 2009. (Invited)
- [C41] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "A wavelet-based quadratic extension method for image deconvolution in the presence of Poisson noise," **IEEE ICASSP**, Taipei, Taiwan, April 19-24, 2009.
- [C42] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "A constrained forward-backward algorithm for image recovery problems," **EUSIPCO**, Lausanne, Switzerland, August 25-29, 2008.

Conférences nationales :

- [C1] L. Denneulin, M. Langlois, N. Pustelnik, E. Thiébaud "Reconstruction polarimétrique d'environnements circumstellaires à partir des données ESO/VLT-SPHERE IRDIS," **GRETSI**, Lille, France, Aug. 26-29, 2019.
- [C2] B. Pascal, T. Busser, N. Pustelnik, P. Abry and V. Vidal "Segmentation d'images texturées en grande dimension. Application à l'analyse d'écoulements," **GRETSI**, Lille, France, Aug. 26-29, 2019.
- [C3] J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, and P. Abry, "Sélection du paramètre de régularisation dans le problème l_2 -Potts," **GRETSI**, Juan-les-Pins, France, September 5-8, 2017.
- [C4] L. Condat, N. Pustelnik, "Segmentation d'image par optimisation proximale," **GRETSI**, Lyon, France, September 8-11, 2015.
- [C5] J. Frecon, N. Pustelnik, H. Wendt, and P. Abry, "Variation totale multivariée pour la détection de changement du spectre multifractal," **GRETSI**, Lyon, France, September 8-11, 2015.
- [C6] G. Michau, P. Borgnat, N. Pustelnik, P. Abry, A. Nantes, E. Chung "Estimating link-dependent origin-destination matrices from sample trajectories and traffic counts," **GRETSI**, Lyon, France, September 8-11, 2015.
- [C7] N. Pustelnik, H. Wendt, and P. Abry, "Régularité locale pour l'analyse de texture : le mariage des coefficients dominants et de la minimisation proximale," **GRETSI**, Brest, France, September 3-6, 2013.
- [C8] N. Pustelnik, J.-C. Pesquet, and C. Chaux, "Bancs de filtres et méthodes proximales pour la restauration d'images," **GRETSI**, Bordeaux, France, September 5-8, 2011.
- [C9] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Extension des algorithmes imbriqués pour la résolution de problèmes d'optimisation convexe en imagerie," **GRETSI**, Dijon, France, September 8-11, 2009.

Chapitre 2

Synthèse des activités de recherches

Sommaire

2.1 Du modèle direct à la formulation régularisée	20
2.1.1 Reconstruction d'images	21
2.1.2 Détection de changements/segmentation	24
2.1.3 Apprentissage machine	26
2.1.4 Algorithmes proximaux	28
2.1.5 Résumé des travaux de thèse	34
2.2 La projection épigraphique, un outil clé dans la gestion de contraintes non linéaires	35
2.2.1 Motivations	35
2.2.2 Nouvelles formes explicites de projections épigraphiques	35
2.2.3 Contraintes épigraphiques pour la gestion de contraintes non-linéaires en traitement du signal et des images	38
2.2.4 Contraintes épigraphiques dans l'étude des environnements circumstellaires	47
2.2.5 Contraintes épigraphiques et apprentissage machine	52
2.3 Algorithmes proximaux et grand volume de données	58
2.3.1 Motivations et enjeux	58
2.3.2 <i>Splitting or not splitting</i>	59
2.3.3 Forte convexité en traitement du signal et des images. Le cas de la segmentation de texture par attributs invariants d'échelles	67
2.3.4 Opérateur proximal versus gradient, que choisir? Quelques éléments de réponse.	74
2.4 Sélection des hyperparamètres	80
2.4.1 Motivations et état-de-l'art	80
2.4.2 Approche bayésienne hiérarchique	81
2.4.3 SURE	86
2.4.4 Algorithmes déroulés	93
2.5 Applications variées des algorithmes proximaux sur des données réelles.	96
2.6 Conclusions et perspectives	99
2.7 Références	102

2.1 Du modèle direct à la formulation régularisée

Mes travaux visent à développer des outils numériques pour le traitement et l'analyse de données structurées (signaux, images, graphes). L'ensemble de mes travaux s'articule autour d'un objet mathématique simple qui se formule comme la minimisation d'une énergie de la forme :

$$\hat{x} \in \underset{x \in \mathcal{H}}{\operatorname{Argmin}} \psi(x; z) + \lambda \varphi(x) \quad (2.1)$$

où $z \in \mathbb{R}^{\bar{N}}$ modélise les données structurées de taille $\bar{N}$ que l'on souhaite analyser, $\hat{x} \in \mathcal{H}$ modélise la quantité que l'on souhaite estimer où $\mathcal{H}$ est un espace d'Hilbert réel de dimension finie, $\psi: \mathcal{H} \rightarrow]-\infty, +\infty]$ est une fonction d'attache aux données qui permet d'assurer une adéquation entre z et l'information d'intérêt, $\varphi: \mathcal{H} \rightarrow]-\infty, +\infty]$ est une fonction qui intègre des *a priori* relatifs à l'information recherchée et $\lambda > 0$ est un paramètre de régularisation permettant d'ajuster le poids du terme d'attache aux données et le terme d'*a priori*.

Parmi les problèmes communément modélisés par (2.1) on peut citer la résolution de problèmes inverses par pénalisation parcimonieuse [DDDM04; PBBZP16], la segmentation de Chan-Vese [CV01], le débruitage par variation totale [ROF92], le codage parcimonieux [OF97; JMOB11], ou encore les problèmes d'apprentissage machine supervisés comme la classification par SVM [CV95], ou la régression ridge/lasso [HK70; Tib96].

Les avancées sur les modalités d'acquisitions de données combinées à celles relatives aux méthodes d'optimisation et également aux techniques d'extraction d'information (e.g. échantillonnage comprimé, algorithmes proximaux, deep-learning) ont conduit à de nouveaux cas particuliers de (2.1). Malgré la simplicité apparente de (2.1), cette formulation présente toujours de nombreuses questions fondamentales auxquelles les spécialistes en analyse de données doivent être en mesure de répondre.

Il est par exemple indispensable d'identifier l'information $\hat{x}$ souhaitée par le spécialiste (physicien, biologiste, sociologue, écologiste, ...) et d'identifier comment choisir ψ et φ de façon à construire un problème de minimisation qui permettra d'extraire au mieux la quantité d'intérêt mais qui aura également de bonnes propriétés en terme d'unicité et d'algorithmie. De plus, il faut noter que de nombreuses analyses reposent sur une succession de traitements (débruitage, déconvolution, extraction de descripteurs, segmentation, ...) qui peuvent prendre la forme (2.1), chacun d'entre eux nécessitant des choix ad hoc de paramètres et où une partie de l'information se perd d'une étape à l'autre. Il est donc important, lorsque cela est possible, de réduire le nombre d'étapes en les combinant comme cela est maintenant usuel entre débruitage et déconvolution.

D'autre part, la littérature en optimisation est prolifique et pour un problème de minimisation donné de nombreuses stratégies algorithmiques peuvent être employées. Il est donc nécessaire d'identifier la plus adaptée pour un problème donné en fonction des garanties souhaitées (convergence des itérées, convergence linéaire, temps de calcul de chaque itération, facilité d'implémentation, ...).

Enfin, une question extrêmement importante est le choix du paramètre λ , en effet celui-ci est souvent laissé au libre arbitre de l'utilisateur mais sa valeur impacte considérablement la qualité de l'estimation et les conclusions ou le diagnostic qui sera fait. Il est donc indispensable de savoir estimer précisément ce paramètre.

Toutes ces questions sont au cœur de mes travaux de recherche avec des contributions en modélisation dans des problèmes très variés, sur l'obtention de nouvelles formes explicites d'opérateurs proximaux indispensables au design d'algorithmes proximaux rapides, sur la construction et l'analyse d'algorithmes proximaux permettant de gérer de gros volumes de données, et sur le développement de méthodes permettant le réglage de l'hyperparamètre λ .

Comme cela apparaît déjà dans cette introduction, la majorité des travaux mentionnés ne parlent pas d'apprentissage profond. Il s'agit cependant d'un champ d'étude qui ne peut être ignoré au vu des performances obtenues dans des nombreux domaines tels que la classification

ou la segmentation, et même récemment en restauration d'image. J'ai donc commencé à m'intéresser à la construction d'algorithmes déroulés qui permettent de mettre à profit tout le savoir faire acquis depuis des années en résolution de problèmes inverses dans le design de réseaux profonds. Cette contribution sera présentée dans la sous-section 2.4.4 dédiée à l'estimation des paramètres.

Dans ce document, j'ai choisi de regrouper les contributions qui ont suivi ma thèse en quatre parties :

- la section 2.2 sera dédiée aux contraintes épigraphiques qui jouent un rôle clé dans de nombreux problèmes de traitement de signaux et d'images mais également en apprentissage machine. Cette section sera l'occasion de présenter les nouveaux résultats de projections épigraphiques que nous avons obtenus dans le cadre de la thèse de Giovanni Chierchia et également leur intérêt dans des problèmes de restauration d'images formulés sous forme contrainte, pour la décomposition modale de signaux, pour l'analyse des environnements circumstellaires dont les développements proviennent de la thèse de Laurence Denneulin et enfin dans des problèmes d'apprentissage (régression ou classification) avec *a priori* de parcimonie de diverses formes.
- la section 2.3 propose ensuite d'analyser différentes pistes d'accélération dans les algorithmes proximaux : d'une part à travers de nouvelles formes explicites d'opérateurs proximaux associés à une somme de deux fonctions permettant d'éviter des stratégies d'éclatement (*splitting*), puis à travers la construction d'une fonction fortement convexe dans un contexte de segmentation de texture permettant de combiner les étapes d'estimation de descripteurs locaux et de segmentation (résultats obtenus dans le cadre de la thèse de Barbara Pascal). Je conclurai cette section par une analyse théorique de plusieurs algorithmes proximaux de premier ordre autour de la question faut-il préférer des pas proximaux ou des pas de descente dans un contexte où les deux stratégies sont envisageables.
- la section 2.4 se concentre sur la question du choix de/des paramètre(s) de régularisation à travers une première contribution dans le contexte d'une formulation bayésienne hiérarchique menée dans le cadre de la thèse de Jordan Frécon, puis grâce au principe de SURE (*Stein Unbiased riesk estimator*) étudié dans le cadre de la thèse de Barbara Pascal et enfin à travers la construction d'un schéma d'apprentissage profond reposant sur le principe des algorithmes déroulés.
- la section 2.5 présentera brièvement l'intérêt d'une formulation (2.1) combinée à l'usage d'algorithmes proximaux pour diverses analyses de données réelles : en microscopie par illumination structurée, pour l'analyse du rythme cardiaque fœtal, pour l'estimation des matrices origine-destination dans l'étude du trafic, pour l'analyse des ondes internes et enfin dans l'estimation du facteur de reproduction de l'épidémie Covid19.

L'ensemble de ces résultats repose sur des notions de base en reconstruction d'image, en détection de changement/d'interfaces et segmentation, en apprentissage machine et sur les algorithmes proximaux que je propose de rappeler brièvement dans les sous sections 2.1.1, 2.1.2, 2.1.3 et 2.1.4. Je ferai également un bref rappel de mes travaux de thèse dans la section 2.1.5.

2.1.1 Reconstruction d'images

Mon cœur d'expertise étant la résolution de problèmes inverses, je vais naturellement commencer par introduire cette problématique.

Pour de nombreuses modalités en imagerie le modèle d'observation peut s'écrire sous la forme :

$$z = \mathcal{D}(A\tilde{u}) \quad (2.2)$$

où l'objet d'étude $\tilde{u} = (\tilde{u}_n)_{n \in \Omega}$ que l'on souhaite estimer est défini dans $\mathbb{R}^N$. Par exemple, un signal 1-D de taille N ou un graphe composé de N nœuds correspondent à $\Omega = \{1, \dots, N\}$. Les images 2-D

de taille $N_1 \times N_2$ correspondent à $\Omega = \{1, \dots, N_1\} \times \{1, \dots, N_2\}$, avec $n = (n_1, n_2)$ un indice 2-D. Notons que les images seront toujours représentées sous forme vectorisée. La dégradation linéaire $A: \mathbb{R}^N \rightarrow \mathbb{R}^{\bar{N}}$ relie la quantité à estimer et les données et peut prendre la forme d'un opérateur de flou permettant de modéliser un mouvement entre la scène et la caméra, la défocalisation d'un système d'imagerie optique, les imperfections de la lentille ou des turbulences atmosphériques. Le flou peut être invariant spatialement, dans ce cas le produit $A\tilde{u}$ peut être exprimé comme une convolution entre $\tilde{u}$ et la réponse impulsionnelle spatiale (PSF) [CNR03; PBBZP16] ou, comme dans de nombreux problèmes réels, le flou peut varier spatialement à cause de distortions optiques des miroirs [BW99], de turbulences atmosphériques [NPPT97] ou encore l'utilisation d'optiques adaptatives [NPPT05]. Parmi les opérateurs fréquemment rencontrés on peut également citer les opérateurs de projection de tomographie produisant des données qui sont représentées en 2D sous forme d'un sinogramme et également les opérateurs linéaires associés à un processus d'acquisition IRM [CPCBB10; BCC⁺16]. Dans de nombreuses modalités interviennent les opérateurs permettant de modéliser une information manquante, par exemple provenant d'une acquisition à faible résolution (i.e. super-résolution), ou d'un processus d'occultation (i.e. *inpainting*). Depuis les travaux fondateurs de Candès and Tao [CRT06] en échantillonnage comprimé, il existe également un axe de recherche qui s'attèle à construire des techniques d'acquisitions où A modélise une matrice aléatoire avec $\bar{N} \ll N$. On peut se référer à [BCC⁺16; MW11; GSL⁺21] pour des exemples d'intégration de ces concepts en IRM, en radio-astronomie et en endoscopie sans lentille.

Concernant les dégradations aléatoires $\mathcal{D}$, la plus communément rencontrée est la dégradation par un bruit additif. Le modèle de dégradation se formule $z = A\tilde{u} + \epsilon$ où ϵ est alors un vecteur de réalisations d'un bruit de moyenne nulle et de variance σ^2 ou covariance Σ , qui suit souvent une loi gaussienne. Dans les techniques d'imagerie à faible nombre de photons, il est standard de modéliser la dégradation par un processus de Poisson puisque ce dernier est un procédé de comptage apte à exprimer la distribution du nombre de photons reçus par le capteur dans un interval de temps donné. Dans ce cas, chaque observation scalaire z_j avec $j \in \{1, \dots, \bar{N}\}$ peut être vue comme une réalisation d'une variable aléatoire de Poisson d'intensité $\alpha(A\tilde{u})_j$ où A et $\tilde{u}$ sont supposées avoir des composantes positives. La variance de z_j est alors égale à $\alpha(A\tilde{u})_j$ et par conséquent dépend de l'intensité de $(A\tilde{u})_j$. On peut également avoir recours à des modèles de bruit plus réalistes comme les modèles probabilistes Poisson-Gauss [LT05; PKDH08; CJPT15; Son09] qui reflètent à la fois le comptage des photons et la modélisation du bruit thermique induit par l'électronique du système d'imagerie.

Formulation régularisée En reconstruction d'image, l'objectif est de trouver une estimée $\hat{u}$ de l'image originale $\tilde{u}$ à partir des mesures z . La solution est considérée comme supervisée quand A et la statistique du bruit sont connues, sinon on parle de mode non-supervisé ou aveugle. Dans une telle situation, l'identification des caractéristiques de dégradation doit être effectuée avant ou simultanément à la procédure de reconstruction [LBB90; KH96]. L'ensemble de mes travaux se concentre sur une approche supervisée.

Le modèle de dégradation fait généralement intervenir un opérateur pouvant être non-inversible ou mal conditionné. Par conséquent, des estimateurs de la forme $\hat{u} = A^{-1}z$ ou $\hat{u} \in \text{Argmin}_{\tilde{u}} \|A\tilde{u} - z\|_2^2 = (A^*A)^{-1}A^*z$ ne sont pas envisageables ou conduisent à un estimateur irrégulier provenant de l'amplification du bruit de mesure. Il est donc nécessaire de proposer une régularisation de la solution.

Afin d'intuiter la formulation générique (2.1), il peut être opportun de faire appel à la formulation bayésienne. Pour cela, nous supposons que u et z sont des réalisations de vecteurs aléatoires U et Z et l'estimation peut alors être atteinte par une stratégie du Maximum A Posteriori (MAP). L'objectif est alors d'estimer $\hat{u}$ qui maximise la distribution a posteriori $\mu_{U|Z=z}$ qui peut être refor-

mulée par the théorème de Bayes puis par monotonie du logarithme comme :

$$\hat{u} \in \underset{u \in \mathbb{R}^N}{\text{Argmin}} \left(-\log \underbrace{\mu_{Z|U=u}(z)}_{\text{vraisemblance}} - \log \underbrace{\mu_U(u)}_{\text{a priori}} \right). \quad (2.3)$$

Le premier terme de (2.3) quantifie la fidélité aux observations z et dépend du modèle de formation des données (2.2). A titre d'exemple, lorsque $\mathcal{D}$ modélise un bruit blanc Gaussien de variance σ^2 , la vraisemblance prend la forme :

$$\mu_{Z|U=u}(z) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{\|Au-z\|^2}{2\sigma^2}}$$

conduisant à

$$\Psi(Au; z) = \frac{1}{2\sigma^2} \|Au - z\|^2.$$

Ce même raisonnement permet de motiver une attache aux données de type divergence de Kullback-Leibler lorsque le modèle d'acquisition des données est un processus de Poisson (cf. Section 2.1.5 relative à mes travaux de thèse) ou une attache aux données Huber pour des données qui suivent une loi gaussienne pour des intensité au dessous d'un certain seuil et une loi laplacienne pour des intensités supérieures à ce seuil [PCJ21].

Le second terme est relié à l'information *a priori* sur l'image originale conduisant à

$$\Phi(u) = -\log \mu_U(u).$$

Si la littérature en restauration d'image s'appuie fortement sur le modèle bayésien pour justifier le choix de l'attache aux données, les modèles de pénalisation ont quant à eux plutôt émergé du côté de la littérature dédiée aux ondelettes, à la parcimonie ou aux modèles variationnels continus. La régularisation la plus connue est certainement la pénalisation par variation totale, proposée par Rudin, Osher, et Fatemi (ROF) [ROF92]. dont la définition originale fut donnée dans une formulation continue. En dimension finie, plusieurs définitions de la variation totale sont possibles et dépendent essentiellement du choix de l'opérateur de discrétisation. La plus utilisée est la variation totale isotrope proposée par Chambolle [Cha04]. Pour la définir, on s'appuie sur l'opérateur $D: \mathbb{R}^N \rightarrow \mathbb{R}^{|\mathbb{E}|}$ avec $\mathbb{E}$ un sous-ensemble de Ω^2 , constitué de paires d'indices, qui peuvent être vues comme les arêtes d'un graphe. La formulation usuelle de la variation totale isotrope s'exprime :

$$(\forall u \in \mathbb{R}^N), \quad \Phi(u) = \text{TV}(u) = \|Du\|_{2,1} = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \sqrt{((D_1 u)_{n_1, n_2})^2 + ((D_2 u)_{n_1, n_2})^2} \quad (2.4)$$

avec

$$(D_1 u)_{n_1, n_2} = \begin{cases} u_{n_1+1, n_2} - u_{n_1, n_2} & \text{si } n_1 < N_1, \\ 0 & \text{si } n_1 = N_1, \end{cases} \quad (2.5)$$

et

$$(D_2 u)_{n_1, n_2} = \begin{cases} u_{n_1, n_2+1} - u_{n_1, n_2} & \text{si } n_2 < N_2, \\ 0 & \text{si } n_2 = N_2. \end{cases} \quad (2.6)$$

La pénalisation par variation totale permet d'obtenir de très bonnes performances en débruitage, surtout si l'image contient de larges régions uniformes. Cependant, pour des images naturelles, cette méthode laisse apparaître un effet de *staircasing* [Lou08] que les utilisateurs préfèrent éviter, par exemple en astronomie où une version lissée de la variation totale est souvent utilisée. Il s'agit de la pénalisation par variation totale hyperbolique [CBFAB97] qui s'exprime :

$$\Phi(u) = \sum_n \sqrt{\|(Du)_n\|_2^2 + v} \quad (2.7)$$

faisant apparaître un paramètre supplémentaire $v > 0$ à ajuster. Elle possède cependant l'avantage de lisser l'effet constant par morceaux et de fournir une transition plus lisse qui peut apparaître

plus réaliste dans certaines modalités d'imagerie. D'autres pénalisations avancées ont été proposées dans la littérature. Par exemple les pénalisations favorisant la parcimonie dans une base ou une trame d'ondelettes [JDCP11; PBBZP16] et dont la pénalisation prend une forme similaire à la pénalisation par variation totale :

$$\varphi(u) = \|Du\|.$$

où D modélise dans ce cas l'opérateur associé à la base ou trame d'ondelettes et où $\|\cdot\|$ est une pénalisation favorisant la parcimonie pouvant prendre la forme d'une norme ℓ_1 ou d'une norme mixte $\ell_{1,2}$ [KT09] comme définie dans (2.4) offrant une flexibilité dans le choix de D et des groupes de coefficients considérés. On peut se référer à [JMOB11; BJMO12] pour des choix de groupes judicieux. Il existe également des formes avancées de pénalisation par variation totale comme la variation totale généralisée (TGV) [BKP10]. On peut également noter l'utilisation de pénalisations non-convexes [Nik05; NNZC08; CPR14], qui cependant introduisent des difficultés associées à la minimisation.

La modélisation de φ comme la composition d'une fonction avec un opérateur linéaire nous permet de modéliser la plupart des pénalisations standards convexes lisses telles que Tikhonov [Tik63] ou la pénalisation TV hyperbolique [CBFAB97], mais également la variation totale [ROF92], ou ses généralisations basées sur le principe de non-localité [BCM05; GO07; GO09], les pénalisations reposant sur des ondelettes ou trames dans leur formulation à l'analyse [EMR07; CPCP09], les pénalisations basées sur des modèles de mélanges gaussiens (GMM) telles que EPLL [ZW11] ou BM3D [DFKE07; DKE12]. On pourra également se référer à [PBBZP16; CP16] pour un aperçu détaillé et des comparaisons de ces pénalisations. Le choix de la pénalisation est très dépendant de l'application considérée, du temps de traitement attendu et de la structure des données qui peut varier considérablement d'une application à l'autre.

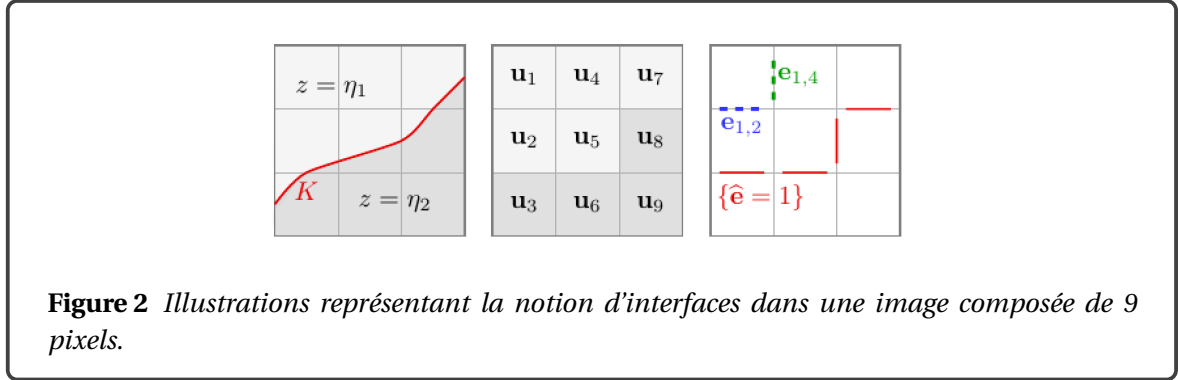
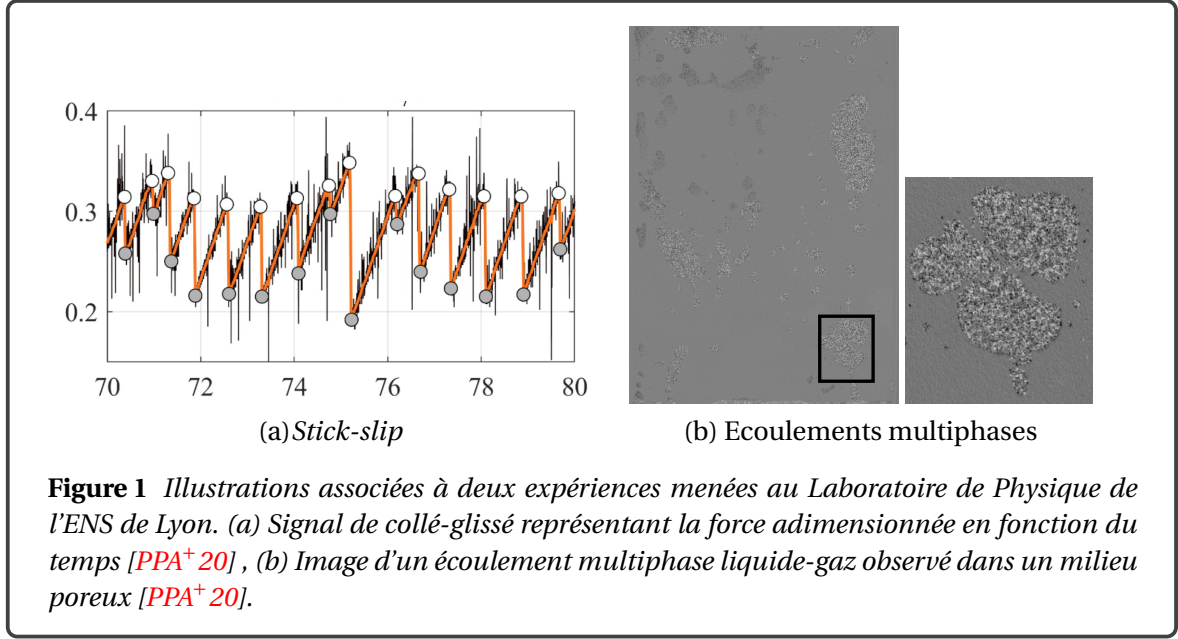
2.1.2 Détection de changements/segmentation

Lors de mon arrivée au Laboratoire de Physique, il me tenait à coeur de m'intégrer aux thématiques de l'équipe et plus généralement à celles du laboratoire. Mes travaux menés jusqu'alors (cf. section 2.1.5) étaient essentiellement centrés sur la restauration d'images qui en tant que tels ont trouvé assez peu d'écho au sein du laboratoire. Cependant, à travers des échanges avec Patrice Abry, Valérie Vidal et Jean-Christophe Geminard, il est apparu que des méthodes de traitement de données bien plus utiles dans leurs expériences étaient des méthodes de détection de changement dans des signaux, de débruitage de signaux linéaires par morceaux, de segmentation, ou encore de détection d'interfaces afin d'étudier par exemple des signaux de collé-glissé (*Stick-slip*) ou des données d'écoulements multiphases (cf. Figure 1).

La détection de changements dans des signaux est étroitement liée à la question de débruitage de signaux constants par morceaux. Cette question a été largement étudiée dans la littérature et des solutions diverses existent telles que les procédures Cusum [BN93], le cadre d'inférence bayésienne hiérarchique [DTS05; DTS07] ou les formulations d'optimisation fonctionnelle, impliquant soit la norme ℓ_1 ou une pénalisation ℓ_0 sur les premières différences du signal (ℓ_2 -Potts) [Pot52; SWD14]. Cette dernière classe est celle qui nous intéresse ici car elle nous permet de formuler dans un cadre unifié la question de détection de changement des signaux et la détection de contours dans des images, tous deux se ramenant à un problème de minimisation de la forme (2.1).

Formellement, dans cette section on s'intéresse à la question d'estimer le signal (ou image/graphes) $\tilde{u}$ et ses points de changements (resp. interfaces) $\tilde{e}$ à partir de données bruitées (cf. Figure 2) :

$$\begin{cases} z = \tilde{u} + \epsilon \\ \tilde{e} = \text{Interfaces}(\tilde{u}). \end{cases}$$



Entre les années 1980 et 2000, un effort de recherche important fut dédié à la construction de modèles permettant d'effectuer des tâches de traitement d'images et en particulier des modèles permettant de faire de la détection d'interfaces (ou des modèles de segmentation) conjointement à une tâche de débruitage. Une façon intuitive d'introduire ces différents modèles s'appuie sur le modèle de Mumford-Shah (MS) [MS89] qui consiste (dans une formulation continue) à estimer conjointement $\hat{u} \in W^{1,2}(\Omega)$ ¹, une approximation lisse par morceaux des observations $z \in L^\infty(\Omega)$, et l'ensemble des discontinuités $K \subset \Omega$, telles que la paire $(\hat{u}, K)$ est une solution optimale de :

$$\underset{u, K}{\text{minimiser}} \int_{\Omega} (z - u)^2 dy_1 dy_2 + \beta \int_{\Omega \setminus K} |\nabla u|^2 dy_1 dy_2 + \chi |K|. \quad (2.8)$$

Le premier terme agit comme un terme d'attache aux données et force l'approximation u à être proche de z , le second terme autorise des petites variations sur u , excepté à la position des discontinuités K , et $|K|$ correspond à la longueur totale des arcs de discontinuités K . Par conséquent, la minimisation de ce terme permet de contrôler la longueur de $|K|$. Les paramètres $\beta > 0$ et $\chi > 0$ correspondent aux paramètres de régularisation qui contrôlent le caractère lisse de l'estimée et la longueur de K . On retrouve un modèle similaire formulé sous forme bayésienne et dans un cadre discret dans Geman et Geman [GG84] prenant la forme du problème de minimisation suivant :

$$\underset{u \in \mathbb{R}^N, e \in \{0,1\}^{|E|}}{\text{minimiser}} \frac{1}{2} \|u - z\|_2^2 + \beta \sum_{i=1}^{|E|} |(Du)_i|^2 (1 - e_i) + \chi \sum_{i=1}^{|E|} e_i. \quad (2.9)$$

Un autre modèle permettant de favoriser ce comportement lisse par morceaux est celui de l'énergie à membrane lâche (*weak membrane*) suggérée par Blake et Zisserman [BZ87]. Ce modèle se ramène au problème standard (2.1) d'une somme d'une terme d'attache aux données et d'un terme

1. $W^{1,2}(\Omega) = \{u \in L^2(\Omega) : \partial u \in L^2(\Omega)\}$ où ∂ correspond à l'opérateur de dérivée faible)

de pénalisation prenant la forme d'une quadratique tronquée :

$$\underset{u \in \mathbb{R}^N}{\text{minimiser}} \frac{1}{2} \|u - z\|_2^2 + \lambda \sum_i \min(|(Du)_i|^2, \tilde{\chi}) \quad (2.10)$$

avec $\lambda > 0$ et $\tilde{\chi} > 0$. Les interfaces sont ensuite obtenues par seuillage. Plus de détails sur les liens entre (2.13) et le modèle de Mumford-Shah et (2.9) sont présentés dans [GY95; CBFAB97].

La restriction aux fonctions constantes par morceaux du problème de Mumford-Shah conduit au problème de segmentation pouvant être formulé via la fonctionnelle de Chan-Vese ou celle de Potts. Dans la littérature [CEN06; CCP12], il apparaît qu'une relaxation convexe et discrète efficace de la fonctionnelle de Chan-Vese continue, dont l'objectif est de segmenter en Q régions une image z , se formule :

$$\hat{\Theta} \in \underset{\Theta \in \mathbb{R}^{(Q+1) \times N}}{\text{Argmin}} \sum_{q=1}^{Q-1} \langle \alpha^{(q)}, \theta^{(q)} \rangle + \lambda \sum_{q=1}^Q \|DT_q \Theta\|_{2,1} + t_C(\Theta), \quad (2.11)$$

où

$$\begin{cases} \alpha^{(q)} = \eta^{(q+1)} - \eta^{(q)}, \\ T_q: \mathbb{R}^{(Q+1) \times N} \rightarrow \mathbb{R}^N: \Theta \mapsto \theta^{(q-1)} - \theta^{(q)}, \\ C = \{\Theta \in \mathbb{R}^{(Q+1) \times N} : 1 \equiv \theta^{(0)} \geq \theta^{(1)} \geq \dots \geq \theta^{(Q-1)} \geq \theta^{(Q)} \equiv 0\}. \end{cases}$$

où $\eta_n^{(q)}$ correspond au coût d'assigner le pixel n dans la région $\Omega^{(q)}$ telle que $\Omega = \cup_q \Omega^{(q)}$ et $\cap_q \Omega^{(q)} = \emptyset$. Dans un contexte de segmentation, le choix le plus standard est $\eta^{(q)} = \frac{1}{2\sigma_q^2} (z - c_q)^2$ où c_q et σ_q^2 désignent la moyenne et la variance au sein de la région $\Omega^{(q)}$. Le paramètre $\lambda > 0$ permet d'agir sur la régularité des interfaces entre les régions. On obtient une estimée composée de Q régions via $\hat{u} = \sum_{q=1}^Q c_q (\hat{\theta}_{q-1} - \hat{\theta}_q)$ et les interfaces sont obtenues par les points de discontinuités de $D\hat{u}$.

La fonctionnelle de Potts quant à elle s'écrit simplement à l'aide de la fonction de comptage $\|\cdot\|_0$:

$$\underset{u \in \mathbb{R}^N}{\text{minimiser}} \frac{1}{2} \|u - z\|_2^2 + \lambda \sum_i \|(Du)_i\|_0 \quad (2.12)$$

et dont une approximation convexe conduit à la minimisation par variation totale :

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \frac{1}{2} \|u - z\|_2^2 + \lambda \sum_i \|(Du)_i\|_1 \quad (2.13)$$

pour lesquels des liens avec la fonctionnelle de Mumford-Shah sont faits dans [CS13] à l'aide d'une stratégie de seuillage itératif, la méthode résultante est nommée T-ROF.

Parmi les contributions récentes sur ce sujet, on peut se référer au travail de Strelakovski et al. [SC14] ainsi qu'aux travaux de Storath, Weimann et collaborateurs proposant des algorithmes efficaces pour résoudre la formulation Blake-Zisserman (2.13) [SWD14; SWFU15; KSW20] mais également le problème de Potts (2.12). On peut également citer les travaux de Foare et collaborateurs qui proposent une nouvelle formulation de la fonctionnelle d'Ambrosio-Tortorelli dans le cadre du Calcul Discret [FLT16] que nous redétaillerons dans la section 2.3.2.

2.1.3 Apprentissage machine

L'apprentissage machine est au cœur de nombreuses analyses de données. Il peut être supervisé comme c'est le cas dans des tâches de régression/classification ou non supervisé permettant alors d'identifier des structures sous-jacentes des données. Depuis 2015 j'ai commencé à m'intéresser à ces questions d'apprentissage supervisé à travers tout d'abord le concept de projection épigraphique, puis par la mise en œuvre sur divers problèmes réels (Contrat TOTAL, contrat AZOTH SYSTEMS, classification de signaux FETUSES en collaboration avec Patrice Abry) et encore plus récemment dans le contexte des algorithmes déroulés permettant la construction d'architectures de type réseaux profonds dans un contexte de restauration d'images. L'apprentissage

machine supervisé prend également la forme d'un problème de minimisation (2.1) que nous proposons de détailler dans le contexte de la classification SVM.

Je présente ici la construction du modèle de classification SVM multiclasse sous forme (2.1) sur lequel plusieurs contributions s'appuieront.

Un classifieur multiclasse repose sur Q fonctions de discrimination $d_x^{(q)} : \mathbb{R}^{\bar{N}} \mapsto \mathbb{R}$, pour tout $q \in \{1, \dots, Q\}$, construites à partir d'un ensemble de paires d'entrée-sortie² :

$$\mathcal{S} = \{(z_s, \tilde{u}_s) \in \mathbb{R}^{\bar{N}} \times \{1, \dots, Q\} \mid s = \{1, \dots, S\}\}.$$

Dans un contexte d'apprentissage standard, ces fonctions sont supposées linéaires dans un certain espace de représentation des entrées et sont obtenues par une transformation $\mathcal{F} : \mathbb{R}^{\bar{N}} \rightarrow \mathbb{R}^{\bar{M}}$ [Cov65], modélisant par exemple des coefficients temps-fréquence [Fla99], des coefficients de scattering [BM13] ou des coefficients de CNN [LBH15], avec $\bar{N} \ll \bar{M}$. conduisant à la forme suivante des fonctions de discrimination :

$$d_x^{(q)}(z) = \mathcal{F}(z)^\top \theta^{(q)} + b^{(q)}, \quad (2.14)$$

où, pour tout $q \in \{1, \dots, Q\}$, $(\theta^{(q)}, b^{(q)}) \in \mathbb{R}^{\bar{M}} \times \mathbb{R}$ correspondent aux paramètres à estimer. Dans la suite nous concaténons ces paramètres dans un unique vecteur $x \in \mathbb{R}^{(\bar{M}+1)Q}$

$$x = \begin{bmatrix} \theta^{(1)} \\ b^{(1)} \\ \vdots \\ \theta^{(Q)} \\ b^{(Q)} \end{bmatrix} = \left\{ \begin{array}{l} \text{ } \\ \text{ } \end{array} \right\} x^{(1)} \quad \left\{ \begin{array}{l} \text{ } \\ \text{ } \end{array} \right\} x^{(Q)}$$

et nous définissons la fonction $\widetilde{\mathcal{F}} : \mathbb{R}^{\bar{N}} \rightarrow \mathbb{R}^{\bar{M}+1}$ comme

$$\widetilde{\mathcal{F}}(z) = [\mathcal{F}(z)^\top \ 1]^\top,$$

Par conséquent (2.14) peut-être simplifié sous la forme $d_x^{(q)}(z) = \widetilde{\mathcal{F}}(z)^\top x^{(q)}$.

L'objectif de l'apprentissage est de trouver un vecteur x qui propose un compromis entre complexité du modèle et attache aux données, et qui consiste à prédire correctement les paires $(z_s, \tilde{u}_s) \in \mathcal{S}$ pour tout $s \in \{1, \dots, S\}$, c'est à dire :

$$\tilde{u}_s = \arg \max_{q \in \{1, \dots, Q\}} \widetilde{\mathcal{F}}(z_s)^\top x^{(q)}.$$

Par définition de l'argmax,

$$(\forall s \in \{1, \dots, S\}) \quad \max_{q \neq \tilde{u}_s} \widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\tilde{u}_s)}) < 0,$$

et de manière équivalente

$$(\forall s \in \{1, \dots, S\}) \quad \max_{q \neq \tilde{u}_s} \widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\tilde{u}_s)}) \leq -v_s, \quad (2.15)$$

où, pour tout $s \in \{1, \dots, S\}$, v_s est un scalaire positif. Malheureusement cette contrainte n'a pas d'intérêt pratique étant donné qu'elle devient non faisable quand $\mathcal{F}$ ne conduit pas à une séparation linéaire de l'ensemble d'apprentissage. Les SVMs multiclassés surmontent ce problème en introduisant la notion de marges douces (*soft margins*), qui consiste à ajouter un vecteur de variable de marges (*slack variables*) $\zeta = (\zeta_s)_{1 \leq s \leq S}$ dans (2.15) :

$$\begin{cases} (\forall s \in \{1, \dots, S\}) \quad \max_{q \neq \tilde{u}_s} \widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\tilde{u}_s)}) \leq \zeta_s - v_s, \\ (\forall s \in \{1, \dots, S\}) \quad \zeta_s \geq 0, \end{cases} \quad (2.16)$$

2. Dans un contexte de régression, $\mathcal{S} = \{(z_s, \tilde{u}_s) \in \mathbb{R}^{\bar{N}} \times \mathbb{R}^N \mid s = \{1, \dots, S\}\}$

Le problème d'apprentissage SVM multiclass est obtenu en ajoutant une régularisation quadratique [CS01], conduisant à³

$$\underset{(x, \zeta) \in \mathbb{R}^{(M+1)Q} \times \mathbb{R}^S}{\text{minimiser}} \sum_{q=1}^Q \|x^{(q)}\|_2^2 + \alpha \sum_{s=1}^S \zeta_s \quad \text{s.c.} \begin{cases} (\forall s \in \{1, \dots, S\}) & \max_{q \neq \tilde{u}_s} \widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\tilde{u}_s)}) \leq \zeta_s - v_s, \\ (\forall s \in \{1, \dots, S\}) & \zeta_s \geq 0, \end{cases} \quad (2.17)$$

où $\alpha \in]0, +\infty[$. La pénalisation linéaire sur les variables de marge permet de minimiser la violation à la contrainte (2.15). En utilisant des arguments d'analyse convexe [BV04], le problème de minimisation ci-dessus peut alors être reformulé sans variable de marge comme :

$$\underset{x \in \mathbb{R}^{(M+1)Q}}{\text{minimiser}} \sum_{s=1}^S \max \{0, v_s + \max_{q \neq \tilde{u}_s} \widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\tilde{u}_s)})\} + \lambda \|x\|_2^2. \quad (2.18)$$

où $\lambda = \alpha^{-1}$. On reconnaît bien ici un problème de la forme (2.1). Dans le cas où $v_s \equiv 1$, la fonction de coût est la fonction *hinge*, souvent approchée par la fonction logistique qui est une fonction différentiable et de gradient Lipschitz [BSU13]. Les variations autour de ce critère seront détaillées dans la section 2.2.5.

2.1.4 Algorithmes proximaux

Dans les sections précédentes, j'ai introduit les principaux critères de la forme (2.1) sur lesquels s'articulent mes contributions. Cependant une étape clé dans la résolution repose sur l'algorithme mis en oeuvre pour estimer un (voire le) minimiseur de cette fonction.

Dans la majorité de mes travaux, ce critère, bien que convexe, implique soit une pénalisation non-lisse, soit des contraintes non triviales (cf. contraintes épigraphiques présentées dans la section 2.2), soit des fonctions différentiables mais n'étant pas de gradient Lipschitz, et dans ces nombreuses configurations les méthodes standards de descente de gradient ne peuvent être mises en oeuvre. Il est alors usuel d'avoir recours aux méthodes d'optimisation dites méthodes proximales [BC17; CP11b; BJMO12; PB14]. Ces méthodes sont présentées comme une alternative aux méthodes de gradient dans le cas de la minimisation de fonctions convexes non-lisses. Rappelons que si le problème considéré est de la forme $\min_{x \in \mathcal{H}} f(x)$ avec f une fonction convexe non-lisse, les itérations de l'algorithme de sous-gradient [Sho85] sont les suivantes :

$$x^{[k+1]} = x^{[k]} - \tau_k g^{[k]} \quad \text{avec} \quad g^{[k]} \in \partial f(x^{[k]})$$

où $\tau_k > 0$ et $\partial f(x) = \{w \in \mathcal{H} \mid (\forall y \in \mathcal{H}) \langle y - x, w \rangle + f(x) \leq f(y)\}$ désigne la sous-différentielle de Moreau [Mor65] qui se réduit au gradient lorsque la fonction est différentiable (i.e. $\partial f = \{\nabla f\}$). Cette stratégie est cependant rarement utilisée à cause de conditions restrictives liées au choix du pas $\tau_k > 0$ devant vérifier $\sum_{k=0}^{+\infty} \tau_k^2 < +\infty$ et $\sum_{k=0}^{+\infty} \tau_k = +\infty$ afin de garantir la convergence vers un minimiseur de f , noté $\hat{x}$.

Une solution proposée dans les années 80 pour améliorer ce comportement numérique consiste à procéder à des itérations implicites prenant la forme :

$$x^{[k+1]} = x^{[k]} - \tau_k g^{[k]} \quad \text{avec} \quad g^{[k]} \in \partial f(x^{[k+1]}) \quad (2.19)$$

$$\Leftrightarrow x^{[k+1]} = \text{prox}_{\tau_k f} x^{[k]} \quad (2.20)$$

où, pour tout $\tau_k > 0$, $\text{prox}_{\tau_k f}$ désigne l'opérateur proximal associé à une fonction f [Mor65] tel que

$$(\forall x \in \mathcal{H}) \quad \text{prox}_{\tau_k f} x = \underset{y \in \mathcal{H}}{\text{argmin}} \|y - x\|_2^2 + \tau_k f(y). \quad (2.21)$$

La convergence est assurée lorsque $\sum_{k=0}^{+\infty} \tau_k = +\infty$. L'opérateur proximal est une généralisation de la projection sur un ensemble convexe fermé non vide C de $\mathcal{H}$, notée P_C , en remarquant que

3. Il est à noter que la régularisation n'implique pas $(b^{(q)})_{1 \leq q \leq Q}$.

$\text{prox}_{\iota_C} = P_C$ avec ι_C la fonction indicatrice associée à l'ensemble C telle que, pour tout $x \in \mathcal{H}$, $\iota_C(x) = 0$ si $x \in C$ et $+\infty$ sinon.

L'efficacité des itérations proximales est conditionnée à la connaissance sous forme explicite de l'opérateur proximal. De nombreuses formes explicites ont été proposées dans la littérature et sont répertoriées pour la plupart sur le site web [Prox-Repository](#). Une difficulté supplémentaire intervient lorsqu'il s'agit de calculer l'opérateur proximal associé à une somme de deux fonctions (i.e. $f = f_1 + f_2$) comme c'est le cas dans (2.1). Dans ce cas, très peu de formes explicites existent (cf. section 2.3.2.1 pour plus de détails). Il est alors nécessaire d'avoir recours à des algorithmes proximaux d'éclatement (*splitting*) qui sont des variations autour de ces itérations de façon à minimiser une somme de fonctions convexes pouvant être non-lisses, typiquement pour estimer la solution du Problème (2.1).

2.1.4.1 Algorithmes proximaux du premier ordre

Nous commençons par rappeler les grandes classes d'algorithmes adaptées à un problème de minimisation de la forme :

$$\underset{x \in \mathcal{H}}{\text{minimiser}} f_1(x) + f_2(x), \quad (2.22)$$

où f_1 et f_2 sont des fonctions propres, semi-continues inférieurement et convexes de $\mathcal{H}$ dans $]-\infty, +\infty]$ (i.e dans $\Gamma_0(\mathcal{H})$). Les algorithmes présentés génèrent une suite récursive de la forme :

$$(\forall k \in \mathbb{N}) \quad x^{[k+1]} = \Phi x^{[k]}, \quad (2.23)$$

où $x^{[0]} \in \mathcal{H}$ et $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ est un opérateur contractant reposant soit sur le gradient soit sur l'opérateur proximal des fonctions f_1 et f_2 et tel que nous pouvons retrouver une solution de (2.22) à partir de ses points fixes. Les principaux algorithmes rencontrés dans la littérature sont :

- l'**algorithme explicite (EA)** ou algorithme de descente de gradient [Cau47; Cur44] qui peut-être mis en oeuvre lorsque f_1 et f_2 sont différentiables

$$\Phi = G_{\tau(\nabla f_1 + \nabla f_2)} = \text{Id} - \tau(\nabla f_1 + \nabla f_2), \quad (2.24)$$

pour $\tau > 0$. De plus, il est direct que $\text{Fix } G_{\tau(\nabla f_1 + \nabla f_2)} = \text{zer}(\nabla f_1 + \nabla f_2)$.

- l'**algorithme du point proximal (PPA)** repose sur un choix d'opérateur de la forme

$$\Phi = J_{\tau(\partial f_1 + \partial f_2)} = \text{prox}_{\tau(f_1 + f_2)} \quad (2.25)$$

pour $\tau > 0$. Cependant, le calcul de $\text{prox}_{\tau(f_1 + f_2)}$ peut être difficile, et d'autres méthodes de fractionnement comme FBS, PRS et DRS (décrite ci-dessous) peuvent être considérées afin de réduire le temps de calcul par itération. De plus, il est direct que $\text{Fix } J_{\tau(\partial f_1 + \partial f_2)} = \text{zer}(\partial f_1 + \partial f_2)$.

- l'**algorithme explicite-implicite (FBS pour *Forward-backward splitting*)** aussi connu sous le nom d'algorithme du gradient proximal requiert que la fonction f_1 soit différentiable. Ce schéma alterne une étape de descente de gradient sur la fonction différentiable et une étape proximale sur la fonction non différentiable :

$$\Phi = T_{\tau \partial f_2, \tau \nabla f_1} = \text{prox}_{\tau f_2} \circ (\text{Id} - \tau \nabla f_1), \quad (2.26)$$

pour $\tau > 0$. Cette approche est une version généralisée de la méthode du gradient projeté [LP66] (qui correspond au choix $f_2 = \iota_C$ pour un ensemble convexe C). De plus, en suivant [BC17, Proposition 26.1(iv)(a)], on montre que $\text{Fix } T_{\tau \partial f_2, \tau \nabla f_1} = \text{zer}(\nabla f_1 + \partial f_2)$.

- l'**algorithme de Peaceman-Rachford (PRS)** – Cet algorithme est obtenu est choisissant :

$$\Phi = R_{\tau \partial f_2, \tau \partial f_1} = (2\text{prox}_{\tau f_2} - \text{Id}) \circ (2\text{prox}_{\tau f_1} - \text{Id}), \quad (2.27)$$

pour $\tau > 0$. De plus, $\text{prox}_{\tau f_1}(\text{Fix } R_{\tau \partial f_2, \tau \partial f_1}) = \text{zer}(\partial f_1 + \partial f_2)$ en suivant [BC17, Proposition 26.1(iii)(b)].

— l'**algorithme de Douglas-Rachford (DRS)** – Il s'agit d'un schéma moyenné entre Id et $R_{\tau\partial f_2, \tau\partial f_1}$ conduisant à :

$$\Phi_{\text{DR}} = S_{\tau\partial f_2, \tau\partial f_1} = \frac{\text{Id} + R_{\tau\partial f_2, \tau\partial f_1}}{2} = \text{prox}_{\tau f_2}(2\text{prox}_{\tau f_1} - \text{Id}) + \text{Id} - \text{prox}_{\tau f_1}, \quad (2.28)$$

pour $\tau > 0$. De plus, $\text{prox}_{\tau f_1}(\text{Fix } R_{\tau\partial f_2, \tau\partial f_1}) = \text{zer}(\partial f_1 + \partial f_2)$ en suivant [BC17, Proposition 26.1 (iii) (b)].

Convergence – Les principaux résultats de convergence des suites générées par (2.23) dans les configurations mentionnées précédemment reposent sur deux résultats généraux de la littérature, à savoir le théorème de convergence de Banach-Picard dans le cas d'opérateurs Lipschitz continus et le théorème invoquant des opérateur moyennés.

Commençons par rappeler qu'un opérateur $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ est ω -Lipschitz continu avec $\omega \in [0, +\infty[$ si

$$(\forall x_1 \in \mathcal{H})(\forall x_2 \in \mathcal{H}) \quad \|\Phi x_1 - \Phi x_2\| \leq \omega \|x_1 - x_2\|, \quad (2.29)$$

Si l'opérateur est 1-Lipschitz continu alors on parle d'opérateur contractant. Le résultat de convergence suivant, extrait de [BC17, Théorème 1.50], est connu sous le nom du Théorème de Banach-Picard et prouve la convergence forte⁴ et linéaire des itérations générées en appliquant à plusieurs reprises un opérateur continu ω -Lipschitz quand $\omega \in [0, 1[$.

Proposition 1 Soit $\omega \in [0, 1[$, soit $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ un opérateur ω -Lipschitz continu, et soit $x^{[0]} \in \mathcal{H}$. Si

$$(\forall k \in \mathbb{N}) \quad x^{[k+1]} = \Phi x^{[k]}. \quad (2.30)$$

Alors, $\text{Fix } \Phi = \{\hat{x}\}$ avec $\hat{x} \in \mathcal{H}$ et

$$(\forall k \in \mathbb{N}) \quad \|x^{[k]} - \hat{x}\| \leq \omega^k \|x^{[0]} - \hat{x}\|. \quad (2.31)$$

De plus, $(x^{[k]})_{k \in \mathbb{N}}$ converge fortement et linéairement vers $\hat{x}$ avec un taux de convergence ω .

Dans le cas où $\omega = 1$, la Proposition 1 n'est plus valide. On peut alors avoir recours à un théorème de convergence reposant sur la propriété d'opérateur moyenné. Rappelons qu'un opérateur $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ est α -moyenné avec $\alpha \in]0, 1]$ si, pour tout $x_1 \in \mathcal{H}$ and $x_2 \in \mathcal{H}$,

$$\|\Phi x_1 - \Phi x_2\|^2 \leq \|x_1 - x_2\|^2 - \left(\frac{1-\alpha}{\alpha}\right) \|(\text{Id} - \Phi)x_1 - (\text{Id} - \Phi)x_2\|^2, \quad (2.32)$$

et Φ est fermement contractant s'il est $1/2$ -moyenné. De plus, nous observons que Φ est contractant si et seulement si Φ est 1-moyenné. Le résultat présenté ci-dessous est extrait de [BC17, Proposition 5.16] et garantit la convergence faible⁵ dans un contexte α -moyenné.

Proposition 2 Soit $\alpha \in]0, 1[$, soit $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ un opérateur α -moyenné tel que $\text{Fix } \Phi \neq \emptyset$, et soit $x^{[0]} \in \mathcal{H}$. Soit

$$(\forall k \in \mathbb{N}) \quad x^{[k+1]} = \Phi x^{[k]}. \quad (2.33)$$

Alors $(x^{[k]})_{k \in \mathbb{N}}$ converge faiblement vers $\text{Fix } \Phi$.

Afin de montrer la convergence des algorithmes mentionnés précédemment (EA, FBS, PRS, DRS), il sera parfois nécessaire d'ajouter une hypothèse sur f_1 et/ou f_2 , qui est l'appartenance

4. Soit $(x^{[k]})_{k \in \mathbb{N}}$ une suite d'éléments de $\mathcal{H}$ et soit $\hat{x} \in \mathcal{H}$. $(x^{[k]})_{k \in \mathbb{N}}$ converge fortement vers $\hat{x}$ si $\lim_{k \rightarrow +\infty} \|x^{[k]} - \hat{x}\| = 0$.
 5. Soit $(x^{[k]})_{k \in \mathbb{N}}$ une suite d'éléments de $\mathcal{H}$ et soit $\hat{x} \in \mathcal{H}$. $(x^{[k]})_{k \in \mathbb{N}}$ converge faiblement vers $\hat{x}$ si pour tout $y \in \mathcal{H}$, $\lim_{k \rightarrow +\infty} \langle y | x^{[k]} - \hat{x} \rangle = 0$.

à la classe $C_{\beta}^{1,1}(\mathcal{H})$ avec $\beta > 0$ qui correspond à la classe des fonctions Gâteaux différentiables dans $\mathcal{H}$ ⁶ avec $\nabla f: \mathcal{H} \rightarrow \mathcal{H}$ est β -Lipschitz continu. De plus, les résultats de convergence liés à la Proposition 1 nécessitent une hypothèse de forte convexité. On rappelle qu'une fonction f est ρ -fortement convexe, pour tout $\rho \in]0, +\infty[$, si $f - \frac{\rho}{2} \|\cdot\|_2^2$ est convexe.

Les principaux résultats de la littérature sont les suivants :

- **EA** – Si $f_1 + f_2 \in C_{\beta}^{1,1}(\mathcal{H})$ et $\tau \in]0, \frac{2}{\beta}[$ alors $G_{\tau(\nabla f_1 + \nabla f_2)}$ est $\frac{\tau\beta}{2}$ -moyenné [BC17, Proposition 4.39] et, par conséquent, EA converge faiblement vers $\hat{x}$ en vue de la Proposition 2. Si de plus $f_1 + f_2$ est fortement convexe, $G_{\tau(\nabla f_1 + \nabla f_2)}$ est Lipschitz continu avec constante dans $]0, 1[$ [RHY19, Fact 7] et EA atteint une convergence linéaire en vue de la Proposition 1.
- **PPA** – Pour tout $\tau > 0$, $\text{prox}_{\tau(f_1 + f_2)}$ est fermement non expansif [BC17, Proposition 23.8], ce qui permet de prouver la convergence faible de PPA vers $\hat{x}$. Si nous supposons en plus la forte convexité de $f = f_1 + f_2$, on obtient que $\text{prox}_{\tau f}$ est Lipschitz continu avec une constante dans $]0, 1[$ et PPA converge linéairement [BC17, Proposition 23.13].
- **FBS** – La convergence faible des itérations générées par (2.23) avec $\Phi = T_{\tau\partial f_2, \tau\nabla f_1}$ est garantie si $f_1 \in C_{\beta}^{1,1}(\mathcal{H})$ et $\tau \in]0, \frac{2}{\beta}[$ [BC17, Théorème 26.14]. Ceci est une conséquence de la composition d'un opérateur $\frac{\tau\beta}{2}$ -moyenné et d'un opérateur fermement contractant [BC17, Proposition 26.1(iv)(d)]. Si de plus on suppose la forte convexité de f_1 ou f_2 , la convergence linéaire de FBS est garantie [BC17, Théorème 26.16].
- **PRS** et **DRS** – Sans hypothèse supplémentaire sur f_1 et f_2 autre que l'appartenance à la classe $\Gamma_0(\mathcal{H})$ définie après l'équation (2.22), $2\text{prox}_{f_1} - \text{Id}$ et $2\text{prox}_{f_2} - \text{Id}$ sont simplement contractants, et la convergence de PRS n'est pas garantie. Ce constat a motivé le schéma moyenné DRS qui conduit à un opérateur moyenné et ainsi à la convergence faible vers une $\hat{x}$. Avec l'hypothèse supplémentaire $f_1 \in C_{\beta}^{1,1}(\mathcal{H})$, la convergence faible de PRS est garantie [LM79, Corollaire 1 & Remarque 2(2)]. Si de plus on suppose la forte convexité de f_1 , la réflexion $2\text{prox}_{f_1} - \text{Id}$ est Lipschitz continue avec une constante dans $]0, 1[$ [Gis17] et, par conséquent, PRS converge linéairement et fortement vers une solution. Cette propriété est également valable pour DRS, mais avec un taux de convergence moins bon [RHY19; DY17; THG18; GB17a].

En résumé, sans aucune hypothèse de différentiabilité sur le problème (2.22) la seule méthode convergente disponible est DRS, si les opérateurs proximaux sont faciles à calculer.

L'ensemble de ces résultats sont généralement présentés dans la littérature dans le contexte des opérateurs maximaux montones. L'hypothèse $f \in C_{\beta}^{1,1}(\mathcal{H})$ est quant à elle formulée dans le cadre d'opérateurs cocoercive (en notant le lien entre $f \in C_{\beta}^{1,1}(\mathcal{H})$ et ∇f qui est β^{-1} -cocoercif) et la notion de forte convexité est formulée dans le contexte de forte monotonie.

Vers la dualité Comme mentionné dans les sections précédentes, il est commun d'être confronté à une fonction objectif où $f_2 = h_2(H\cdot)$ conduisant à un problème de minimisation de la forme :

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f_1(x) + h_2(Hx), \quad (2.34)$$

avec $f_1 \in \Gamma_0(\mathcal{H})$, $H: \mathcal{H} \rightarrow \mathcal{G}$ et $h_2 \in \Gamma_0(\mathcal{G})$. Or dans ce cas, si h_2 n'est pas différentiable il est nécessaire de calculer $\text{prox}_{h_2 \circ H}$. Les principaux résultats de la littérature conduisant à une forme explicite de cet opérateur proximal sont les suivants :

- **Norme euclidienne composée avec un opérateur linéaire** – Pour tout $w \in \mathcal{G}$, $h_2(w) = \|w - z\|_2^2$ avec $z \in \mathcal{G}$, alors

$$(\forall x \in \mathcal{H}) \quad \text{prox}_{h_2 \circ H}(x) = (H^*H + \text{Id})^{-1}(H^*z + x).$$

6. pour tout $x \in \mathcal{H}$ il existe un opérateur linéaire borné $Dh(x): \mathcal{H} \rightarrow \mathbb{R}$ tel que, pour tout $d \in \mathcal{H}$, $Dh(x)d = \lim_{t \downarrow 0} \frac{h(x+td) - h(x)}{t} = \langle \nabla h(x) | d \rangle$, où nous notons $\nabla h(x) \in \mathcal{H}$ le représentant de Riesz-Fréchet.

- **Composition d'une fonction convexe avec un opérateur linéaire** – Ce résultat extrait de [CP07a, Proposition 11] propose une forme explicite de $\text{prox}_{h_2 \circ H}$ lorsque $HH^* = v\text{Id}$ avec $v > 0$ prenant la forme :

$$(\forall x \in \mathcal{H}) \quad \text{prox}_{h_2 \circ H} x = x + v^{-1} H^* (\text{prox}_{v h_2} (Hx) - Hx). \quad (2.35)$$

- **Composition d'une fonction convexe avec un opérateur linéaire** (fonction séparable) – Le résultat précédent est étendu dans [PCP11] dans le contexte de fonctions séparables avec $\mathcal{H} = \mathbb{R}^X$ et $\mathcal{G} = \mathbb{R}^Y$ où $X \in \mathbb{N}^*$, $Y \in \mathbb{N}^*$, et soit $(o_j)_{1 \leq j \leq Y}$ une base orthonormale de $\mathbb{R}^Y$. Soit h_2 une fonction telle que

$$(\forall w \in \mathbb{R}^Y) \quad h_2(w) = \sum_{j=1}^Y h_{2,j}(\langle w, o_j \rangle) \quad (2.36)$$

où $(h_{2,j})_{1 \leq j \leq Y}$ sont des fonctions de $\Gamma_0(\mathbb{R})$. Soit H une matrice dans $\mathbb{R}^{Y \times X}$ telle que

$$\underbrace{HH^T}_{\Lambda} = \sum_{j=1}^Y v_j o_j o_j^T \quad (2.37)$$

où $(v_j)_{1 \leq j \leq Y}$ est une suite de réels positifs. Alors $h_2 \circ H \in \Gamma_0(\mathbb{R}^X)$ et, pour tout $x \in \mathbb{R}^X$

$$(\forall x \in \mathbb{R}^X) \quad \text{prox}_{h_2 \circ H} x = x + H^T \Lambda^{-1} (\text{prox}_{\Lambda h_2} (Hx) - Hx) \quad (2.38)$$

où Λh_2 est la fonction définie par

$$(\forall w \in \mathbb{R}^Y) \quad \Lambda h_2(w) = \sum_{j=1}^Y v_j h_{2,j}(\langle w, o_j \rangle). \quad (2.39)$$

En dehors de ces hypothèses sur h_2 et H , nous ne pouvons obtenir une forme explicite $\text{prox}_{h_2 \circ H}$ et devons avoir recours à des sous-itérations. Bien que les algorithmes proximaux soient robustes aux erreurs numériques dans les calculs de gradient et d'opérateurs proximaux, effectuer des sous-itérations conduit à des problèmes comme identifier le nombre de sous itérations permettant d'assurer que les erreurs soient sommables et un coût numérique plus important de chaque itérations. Lorsque cela est possible il est donc préférable de les éviter. Une alternative peut être de reformuler le problème de minimisation (2.34) dans le dual conduisant à :

$$\underset{w \in \mathcal{G}}{\text{minimiser}} f_1^*(-H^* w) + h_2^*(w), \quad (2.40)$$

faisant intervenir les fonctions conjuguées f_1^* et f_2^* de f_1 et f_2 , puis d'exploiter les relations existantes entre solution primale et solution duale (i.e. $H^* \hat{w} \in \partial f_1(\hat{x})$ et $H \hat{x} \in \partial h_2^*(\hat{w})$). Cette formulation duale est souvent considérée dans le contexte d'une minimisation où f_1 est un terme quadratique, sa conjuguée étant également quadratique, on déplace alors la gestion de l'opérateur linéaire au niveau de la fonction différentiable permettant alors d'employer une technique de type FBS sur le dual [ACP⁺17; PPAP18].

Algorithmes proximaux primaux duaux La résolution par FBS sur le dual est fréquemment employée dans des configurations simples. Cependant, dans un contexte plus général, par exemple lorsque le critère implique plus de deux fonctions, ou si f_1 ne s'exprime pas facilement dans le dual ou n'a pas de bonnes propriétés dans le dual, il est nécessaire d'avoir recours à d'autres schémas algorithmiques.

Parmi les plus connus on peut citer GFB (Generalized Forward-Backward) basé sur des itérations proximales FBS [RFP13] et PPXA [CP11b] reposant sur DRS. Ces deux schémas permettent de minimiser une somme finie de fonctions de $\Gamma_0(\mathcal{H})$.

Dans un contexte où des opérateurs linéaires sont impliqués il est usuel de s'orienter vers les méthodes du type ADMM (Alternating Direction methods of Multipliers) ou SDMM (sa version parallélisée [BSS17]), pouvant être interprétées comme une procédure Douglas-Rachford dans le dual

[Set09], et PPXA+ [PP12] une version étendue de PPXA dont les itérations se rapprochent de celles d'ADMM et les algorithmes proximaux primaux-duaux [CKCH19]. Ces schémas ont pour objectif de réduire la complexité du calcul de l'opérateur proximal lié à la composition d'une fonction non lisse avec un opérateur linéaire en séparant leur activation dans les étapes algorithmiques. La première nécessite de calculer $(\text{Id} + H^*H)^{-1}$, tandis que la seconde approche ne nécessite que l'activation de H et H^* . L'inversion nécessite une procédure itérative qui est peut-être coûteuse en temps de calcul. Je me contente ici de présenter les itérations de l'algorithme Condat-Vũ [Con13b; Vû13] (ou PDHG pour Primal dual hybrid gradient) qui est le schéma le plus utilisé dans mes différents travaux. On se place dans le contexte de minimisation suivant :

$$\hat{x} \in \underset{x \in \mathcal{H}}{\text{Argmin}} f_1(x) + f_2(x) + h_2(Hx), \quad (2.41)$$

où $H: \mathcal{H} \rightarrow \mathcal{G}$ désigne un opérateur linéaire borné, $\mathcal{H}$ et $\mathcal{G}$ étant deux espaces de Hilbert de dimension finie, $f_1: \mathcal{H} \rightarrow]-\infty, +\infty]$, $f_2: \mathcal{H} \rightarrow]-\infty, +\infty]$ et $h_2: \mathcal{G} \rightarrow]-\infty, +\infty]$ désignent les fonctions convexes, semi-continues inférieurement et propres. De plus $f_1 \in C_{\beta}^{1,1}(\mathcal{H})$. La suite $(x^{[k]})_{k \in \mathbb{N}}$ générée par l'algorithme de Condat-Vũ est présentée dans l'Algorithme 1 et converge vers $\hat{x}$ définie dans (2.41).

Algorithm 1 Algorithme de Condat-Vũ

Choix des paramètres : Soit $\tau > 0$, $\gamma > 0$, tel que $\frac{1}{\tau} - \gamma \|H\|^2 > \frac{\beta}{2}$

Initialisation : $(w^{[0]}, x^{[0]}) \in \mathcal{H} \times \mathcal{G}$

Pour $k = 0, 1, \dots$

$$\begin{cases} w^{[k+1]} = \text{prox}_{\tau f_2}(w^{[k]} - \tau \nabla f_1(w^{[k]}) - \tau H^* x^{[k]}) \\ x^{[k+1]} = \text{prox}_{\gamma h_2^*}(x^{[k]} + \gamma H(2w^{[k+1]} - w^{[k]})) \end{cases}$$

Ce schéma est particulièrement intéressant car pour certains choix de f_1 , f_2 et h_2 , on peut retrouver les itérations de l'algorithme Chambolle-Pock [CP11a] (avec $f_1 = 0$), un algorithme primal-dual reposant uniquement sur des activations proximales, mais également FBS ou DRS. Il faut également noter que dans le cas spécifique de l'algorithme Chambolle-Pock, les pas τ et γ peuvent être mis à jour de façon à prendre en compte la constante de forte convexité du critère est ainsi conduire à un algorithme avec un meilleur taux de convergence.

Ce schéma algorithmique peut être étendu au contexte d'une somme de plus de trois fonctions en définissant $h_2(Hx) = \sum_{\ell} h_{2,\ell}(H_{\ell}x)$ impliquant le calcul de $\text{prox}_{\gamma h_{2,\ell}^*}$.

Les preuves de convergence des algorithmes primaux-duaux reposent principalement sur les résultats associés au FBS et DRS dans le contexte des opérateurs maximaux monotones.

2.1.5 Résumé des travaux de thèse

Ma thèse était intitulée “Méthodes proximales pour la résolution de problèmes inverses. Application à la Tomographie par Emission de Positrons.” L’objectif était de proposer des méthodes fiables, efficaces et rapides pour minimiser des critères convexes apparaissant dans la résolution de problèmes inverses en imagerie. Plus précisément, je m’étais intéressée à des problèmes de restauration/reconstruction lorsque les données sont dégradées par un opérateur linéaire et un bruit de Poisson.

En s’appuyant sur le raisonnement mené dans la section 2.1.1 dédiée à la construction du terme d’attache aux données, on remarque que dans le contexte d’un modèle de bruit de Poisson de paramètre d’échelle α , la vraisemblance est donnée par

$$\mu_{Z|U=u}(u) = \prod_{j=1}^{\bar{N}} \frac{\exp(-\alpha(Au)_j)}{z_j!} (\alpha(Au)_j)^{z_j}. \quad (2.42)$$

où $z = (z_j)_{1 \leq j \leq \bar{N}}$ sont des valeurs entières non négatives. Cette vraisemblance conduit à un terme d’attache aux données relié à la divergence généralisée discrète de Kullback-Leibler [Tit87; CP07a] notée D_{GKL} :

$$-\log \mu_{Z|U=u}(u) \propto \sum_{j=1}^{\bar{N}} -z_j \log(\alpha(Au)_j) + \alpha(Au)_j \propto D_{\text{GKL}}(z, \alpha Au). \quad (2.43)$$

Ce terme d’attache aux données s’avérait beaucoup plus difficile à prendre en compte que celui du contexte de statistiques gaussiennes présenté dans la section 2.1.1 car d’une part il n’est pas fini pour toutes valeurs de u , mais seulement lorsque $(Au)_j > 0$. De plus, ce terme est différentiable mais n’est pas de gradient Lipschitz, ce qui limitait le choix des méthodes permettant de minimiser le critère (2.1) car lors de ma thèse (soutenue en 2010) les algorithmes proximaux primaux-duaux n’avaient pas encore été proposés.

La fiabilité de la méthode fut assurée par l’utilisation d’algorithmes proximaux dont la convergence est garantie lorsqu’il s’agit de minimiser des critères convexes. Les solutions algorithmes développées reposaient essentiellement sur l’algorithme PPXA qui permettait d’avoir recours uniquement à des activations proximales et qui nécessitait un jeu subtil de découpe de l’opérateur linéaire impliqué afin de pouvoir appliquer la propriété (2.38). Nous avons également exploré des approximations quadratiques du terme d’attache aux données afin de retrouver le caractère Lipschitz du gradient.

La quête d’efficacité a impliqué le choix d’un critère adapté aux caractéristiques du bruit, à l’opérateur linéaire et au type d’image à reconstruire. En particulier, nous avons utilisé des termes de régularisation basés sur la variation totale et/ou favorisant la parcimonie des coefficients du signal recherché dans une trame. L’utilisation de trames nous a amené à considérer deux approches : une formulation du critère à l’analyse et une formulation du critère à la synthèse. De plus, nous avons étendu les algorithmes proximaux et leurs preuves de convergence aux cas de problèmes inverses multicomposantes.

La recherche de la rapidité de traitement s’est traduite par l’utilisation d’algorithmes proximaux parallélisables.

Les résultats théoriques obtenus ont été illustrés sur différents types de problèmes inverses de grandes tailles comme la restauration d’images mais aussi la stéréoscopie, l’imagerie multispectrale, la décomposition en composantes de texture et de géométrie. Une application avait plus particulièrement attiré notre attention ; il s’agissait de la reconstruction de l’activité dynamique en Tomographie par Emission de Positrons (TEP) qui constitue un problème inverse difficile mettant en jeu un opérateur de projection et un bruit de Poisson dégradant fortement les données observées. Pour optimiser la qualité de reconstruction, nous avons exploité les caractéristiques spatio-temporelles de l’activité dans les tissus.

2.2 La projection épigraphique, un outil clé dans la gestion de contraintes non linéaires

2.2.1 Motivations

Pour analyser des données, il est commun de gérer des contraintes sur la dynamique ou des contraintes de support. Il existe cependant une classe de contraintes moins connues, les contraintes épigraphiques, qui jouent pourtant un rôle clé dans de nombreux problèmes de traitement du signal et des images mais également en apprentissage machine. On peut par exemple faire face à des contraintes épigraphiques par construction afin de gérer des contraintes non-linéaires de la forme $f(x) \leq \eta$ avec $\eta > 0$ comme cela sera introduit dans la section 2.2.3. On peut également rencontrer des contraintes épigraphiques par la formulation du problème comme cela apparaît en astronomie (cf. section 2.2.4) ou en apprentissage machine (cf. section 2.2.5).

Dans cette section, je présente l'ensemble de mes travaux relatifs à des problèmes d'optimisation impliquant une contrainte épigraphique, c'est à dire pouvant se formuler comme :

$$E = \text{epi } f = \{(x, \zeta) \in \mathcal{H} \times \mathbb{R} \mid f(x) \leq \zeta\}.$$

Dans la section suivante, nous présentons plusieurs nouvelles formes explicites de projection $P_{\text{epi } f}$ dont les résultats sont le cœur du travail de thèse de Giovanni Chierchia et qui seront ensuite utilisés tout au long de cette section 2.2.

2.2.2 Nouvelles formes explicites de projections épigraphiques

Formellement, l'épigraphie d'une fonction f convexe, s.c.i., propre de $\mathcal{H}$ dans $]-\infty, +\infty]$ est l'ensemble convexe fermé non-vide de $\mathcal{H} \times \mathbb{R}$ défini par $\text{epi } f = \{(x, \zeta) \in \mathcal{H} \times \mathbb{R} \mid f(x) \leq \zeta\}$.

Dans un premier temps nous présentons un résultat obtenu dans [CPPPP15] qui montre que l'opérateur proximal, pour un fonction f bien choisie, la fonction $\zeta \rightarrow \frac{1}{2}(\max\{f - \zeta, 0\})^2$ joue un rôle important dans le calcul de la projection sur $\text{epi } f$.

Proposition 3 [CPPPP15, Proposition 3.2] Soit $\mathcal{H}$ un espace d'Hilbert réel et soit $\mathcal{H} \times \mathbb{R}$ muni de la norme d'espace produit standard. Soit f une fonction dans $\Gamma_0(\mathcal{H})$ telle que $\text{dom } f$ est ouvert. La projection épigraphique est donnée par : pour tout $(x, \zeta) \in \mathcal{H} \times \mathbb{R}$,

$$P_{\text{epi } f}(x, \zeta) = (p, \theta), \quad (2.44)$$

où

$$\begin{cases} p = \text{prox}_{\frac{1}{2}(\max\{f - \zeta, 0\})^2}(x), \\ \theta = \max\{f(p), \zeta\}. \end{cases} \quad (2.45)$$

Nous fournissons maintenant plusieurs exemples de fonctions f pour lesquelles $P_{\text{epi } f}$ admet une forme simple.

Proposition 4 [CPPPP15, Proposition 3.3] Soit $\gamma > 0$. On suppose que

$$(\forall x \in \mathbb{R}) \quad f(x) = \gamma|x|. \quad (2.46)$$

Pour tout $(x, \zeta) \in \mathbb{R} \times \mathbb{R}$, $P_{\text{epi } f}(x, \zeta) = (p, \theta)$ est donnée par

$$p = \begin{cases} x, & \text{si } \gamma|x| \leq \zeta, \\ \frac{\text{sign}(x)}{1 + \gamma^2} \max\{|x| + \gamma\zeta, 0\}, & \text{sinon.} \end{cases} \quad (2.47)$$

et $\theta = \max\{\gamma|p|, \zeta\}$.

Le résultat ci-dessus découle du fait que $\frac{1}{2}(\max\{\gamma|x| - \zeta, 0\})^2 = (\gamma^2/2)x^2 - \gamma\zeta|x| + \zeta^2/2$ si $\gamma|x| > \zeta$ et 0 sinon, pour lequel l'opérateur proximal est connu [CCPW07, Exemple 4.6]. Cet exemple montre clairement que la Proposition 3 permet de valoriser les résultats sur les opérateurs proximaux déjà connus dans la littérature. La proposition suivante considère le cas où la valeur absolue est élevée par une puissance positive.

Proposition 5 [CPPPP15, Proposition 3.3] Soit $r \in]1, +\infty[$ and $\gamma > 0$. On suppose que

$$(\forall x \in \mathbb{R}) \quad f(x) = \gamma|x|^r. \quad (2.48)$$

Pour tout $(x, \zeta) \in \mathbb{R} \times \mathbb{R}$, $P_{\text{epi } f}(x, \zeta) = (p, \theta)$ est donnée par

$$p = \begin{cases} \text{sign}(x)\chi_0, & \text{si } \zeta \leq 0, \\ y, & \text{si } \zeta > 0 \text{ et } \gamma|x|^r \leq \zeta, \\ \text{sign}(x)\chi_\zeta, & \text{sinon,} \end{cases} \quad (2.49)$$

et $\theta = \max\{\gamma|p|^r, \zeta\}$, où χ est l'unique solution sur $[(\cdot/\gamma)^{1/r}, +\infty[$ de l'équation suivante :

$$r\gamma^2\chi^{2r-1} - r\gamma\zeta\chi^{r-1} + \chi = |x|. \quad (2.50)$$

Il est à noter que lorsque r est un nombre rationnel, (2.50) équivaut à une équation polynomiale pour laquelle soit des solutions avec une forme explicite existent soit des solutions numériques standards peuvent être mises en œuvre.

Les propositions précédentes nous permettent d'établir un résultat concernant la fonction de distance à un ensemble convexe.

Proposition 6 [CPPPP15, Proposition 3.4] Soit C un sous-ensemble convexe non vide de $\mathcal{H}$. Soit $r \in [1, +\infty[$, $\gamma > 0$, et $\zeta \in \mathbb{R}$. Supposons que

$$(\forall x \in \mathcal{H}) \quad f(x) = \gamma d_C^r(x) = \gamma \|x - P_C(x)\|^r. \quad (2.51)$$

Pour tout $(x, \zeta) \in \mathcal{H} \times \mathbb{R}$, $P_{\text{epi } f}(x, \zeta) = (p, \theta)$ est donnée par

$$p = \begin{cases} x, & \text{si } x \in C, \\ \alpha x + (1 - \alpha)P_C(x), & \text{sinon,} \end{cases} \quad (2.52)$$

et $\theta = \max\{\gamma d_C^r(p), \zeta\}$, où

$$\alpha = \frac{\text{prox}_{\frac{1}{2}(\max\{\gamma|\cdot|^r - \zeta, 0\})^2}(d_C(x))}{d_C(x)} \quad (2.53)$$

et l'expression ci-dessus est donnée par les Propositions 4–5.

Dans le cas où $r = 1$ et $C = \{v\}$ pour $v \in \mathcal{H}$, le précédent résultat peut être particularisé par $d_C(x) = \|x - v\|$ and $P_C(x) = v$. Ce faisant, nous obtenons un corollaire sur la norme euclidienne, pour laquelle l'épigraphe correspondant est appelé le *cône de Lorentz* et la projection épigraphique associée est connue dans la littérature (cf. e.g. [PB14, Section 6.3.2]).

Corollaire 1 [CPTPP15, Corollaire 3.5] Soit $\gamma > 0$, $\zeta \in \mathbb{R}$ et $v \in \mathcal{H}$. Nous supposons que

$$(\forall x \in \mathcal{H}) \quad f(x) = \gamma \|x - v\|. \quad (2.54)$$

Pour tout $(x, \zeta) \in \mathcal{H} \times \mathbb{R}$, $P_{\text{epi} f}(x, \zeta) = (p, \theta)$ est donné par

$$p = \begin{cases} v, & \text{si } x = v, \\ x, & \text{si } \gamma \|x - v\| \leq \zeta, \\ \alpha x + (1 - \alpha)v, & \text{sinon,} \end{cases} \quad (2.55)$$

et $\theta = \max\{\gamma \|p - v\|, \zeta\}$, où

$$\alpha = \frac{1}{1 + \gamma^2} \max\left\{1 + \frac{\gamma \zeta}{\|x - v\|}, 0\right\}. \quad (2.56)$$

Nous concluons la section avec un résultat sur le maximum pondéré des valeurs absolues. Lorsque les poids sont tous égaux, cette fonction se ramène à la norme ℓ_∞ standard, pour laquelle l'expression de la projection épigraphique est donnée dans [DST12]. La Proposition suivante fournit un résultat légèrement plus général⁷.

Proposition 7 [CPTPP15, Proposition 3.6] Soit $(\gamma_m)_{1 \leq m \leq M} \in]0, +\infty[^M$. Nous supposons que

$$(\forall x = (x_m)_{1 \leq m \leq M} \in \mathbb{R}^M) \quad f(x) = \max_{1 \leq m \leq M} \gamma_m |x_m|, \quad (2.57)$$

où les valeurs $(v_m = \gamma_m |x_m|)_{1 \leq m \leq M}$ sont triées par ordre croissant. Alors, pour tout $(x, \zeta) \in \mathbb{R}^M \times \mathbb{R}$, $P_{\text{epi} f}(x, \zeta) = (p, \theta)$ est donnée par

$$p = \left(\text{sign}(x_m) \min\{x_m, \theta / \gamma_m\} \right)_{1 \leq m \leq M}$$

et

$$\theta = \max \left\{ \frac{\zeta + \sum_{m=\bar{m}}^M v_m (\gamma_m)^{-2}}{1 + \sum_{m=\bar{m}}^M (\gamma_m)^{-2}}, 0 \right\}, \quad (2.58)$$

où $\bar{m}$ est l'unique entier dans $\{1, \dots, M+1\}$ tel que

$$v_{\bar{m}-1} < \frac{\zeta + \sum_{m=\bar{m}}^M v_m (\gamma_m)^{-2}}{1 + \sum_{m=\bar{m}}^M (\gamma_m)^{-2}} \leq v_{\bar{m}}$$

(avec les conventions $\sum_{m=1}^0 \cdot = \sum_{m=M+1}^M \cdot = 0$, $v_0 = -\infty$ et $v_{M+1} = +\infty$).

7. Il est à noter que ces résultats font intervenir une étape de tri. L'opération de tri coûteuse peut être évitée en utilisant une structure de données en *tas* [CLR90], qui conserve une séquence partiellement triée de sorte que le premier élément soit le plus grand. Cette approche a été utilisée, par exemple, par van den Berg et al. [VDBF08, Algorithm 2] afin d'implémenter la projection sur la boule ℓ_1 .

2.2.3 Contraintes épigraphiques pour la gestion de contraintes non-linéaires en traitement du signal et des images

Si la formulation régularisée (2.1) est la plus commune en problèmes inverses, il est souvent opportun de préférer une approche contrainte pouvant se formuler comme :

$$\hat{u} \in \underset{u \in \mathcal{H}}{\text{Argmin}} \Psi(u; z) \quad \text{s.t.} \quad \Phi(u) \leq \eta_\Phi \quad (2.59)$$

ou

$$\hat{u} \in \underset{u \in \mathcal{H}}{\text{Argmin}} \Phi(u) \quad \text{s.t.} \quad \Psi(u; z) \leq \eta_\Psi. \quad (2.60)$$

Ces formulations sont généralement motivées par un choix plus facile des paramètres $\eta_\Phi > 0$ ou $\eta_\Psi > 0$. En effet, η_Φ dans (2.59) peut plus aisément être relié à des propriétés physiques de l'objet d'étude. Quant à la borne η_Ψ dans (2.60), par exemple dans le cas d'un bruit additif gaussien de moyenne nulle et de variance σ^2 , elle peut être calculée analytiquement en s'appuyant sur des arguments de concentration de la mesure et des propriétés de la distribution χ^2 [RPW19]. Des arguments similaires existent dans le cas d'un bruit de Poisson [BPBT⁺10].

2.2.3.1 Fractionnement épigraphique

Si le choix du paramètre apparaît plus aisé, la difficulté est alors déplacée dans la résolution du problème variationnel. En effet, on remarque que (2.59) et (2.60) sont des cas particulier de problèmes de minimisation plus génériques tels que (2.22) ou (2.41) pour lesquels des solutions algorithmiques existent sous la condition d'être en mesure de calculer en forme explicite la projection sur les contraintes $C = \{u \in \mathcal{H} \mid \Phi(u) \leq \eta_\Phi\}$ ou $C = \{u \in \mathcal{H} \mid \Psi(u; z) \leq \eta_\Psi\}$. Hors, rares sont les formes explicites. Une exception réside pour une contrainte de la forme $C = \{u \in \mathcal{H} \mid \|u - z\| \leq \eta\}$ [TSY11].

Dans le cadre de la thèse de Giovanni Chierchia, nous nous sommes intéressés aux contraintes non linéaires décomposables par blocs de la forme :

$$C = \left\{ (\forall x \in \mathbb{R}^L), \quad f(x) \leq \eta \right\} \quad (2.61)$$

$$= \left\{ (\forall x = [\underbrace{x^{(1)}}_{\text{Taille : } L_1}, \dots, \underbrace{x^{(B)}}_{\text{Taille : } L_B}] \in \mathbb{R}^L), \quad \sum_{b=1}^B f_b(x^{(b)}) \leq \eta \right\} \quad (2.62)$$

avec $L = \sum_{b=1}^B L_b$, $\eta > 0$ et où les fonctions $(f_b)_{1 \leq b \leq B}$ vérifient les hypothèses classiques d'optimisation convexe à savoir convexes, semi-continues inférieurement et propres de $\mathbb{R}^{L_b}$ dans $] -\infty, +\infty]$.

Dans le cas où l'opérateur de projection sur l'ensemble convexe C ne possède pas une forme explicite, nous proposons de déplacer la difficulté en fractionnant cette contrainte en autant d'épigraphes que de fonctions impliquées dans la somme. Une contrainte de demi-espace fermé est également appliquée, afin de limiter la somme des variables épigraphiques introduites à la limite supérieure η . La contrainte se reformule :

$$\sum_{b=1}^B f_b(x^{(b)}) \leq \eta \quad \Leftrightarrow \quad \begin{cases} \sum_{b=1}^B \zeta_b \leq \eta, \\ (\forall b \in \{1, \dots, B\}) \quad (x^{(b)}, \zeta_b) \in \text{epi } f_b. \end{cases} \quad (2.63)$$

2.2.3.2 Formes explicites utiles en restauration d'images

Le point clé du découpage que nous proposons est l'introduction d'épigraphes dans la procédure de minimisation. Les formes explicites de $P_{\text{epi } f_b}$ pour divers exemples de f_b rencontrés en traitement du signal et des images sont listées ci-dessous.

• La **norme** ℓ_r à la **puissance** r dont les cas particuliers $\|\cdot\|_1$ et $\|\cdot\|_2$ sont largement utilisés tant pour modéliser une attache aux données qu'une régularisation en combinaison avec un opérateur linéaire. Formellement, soit $r \in [1, +\infty[$, $(\gamma_b)_{1 \leq b \leq B} \in]0, +\infty[^B$, et

$$(\forall x \in \mathbb{R}^B) \quad f(x) = \sum_{b=1}^B \gamma_b |x_b|^r. \quad (2.64)$$

Alors, f peut être reformulée comme (2.63), avec pour tout $b \in \{1, \dots, B\}$,

$$(\forall x_b \in \mathbb{R}) \quad f_b(x_b) = \gamma_b |x_b|^r. \quad (2.65)$$

La projection épigraphique $P_{\text{epi } f_b}$ est donnée par la Proposition 4 pour $r = 1$ et la Proposition 5 pour $r > 1$.

• La **fonction de distance** est particulièrement intéressante lorsqu'il s'agit de relâcher une contrainte de support ou de dynamique [CP08]. Formellement, soit $B \in \mathbb{N}^*$, $(\gamma_b)_{1 \leq b \leq B} \in]0, +\infty[^B$, $(r_b)_{1 \leq b \leq B} \in [1, +\infty[^B$ et, pour tout $b \in \{1, \dots, B\}$, soit C_b un sous-ensemble convexe fermé non vide de $\mathbb{R}^{L_b}$, et

$$(\forall x \in \mathbb{R}^L) \quad f(x) = \sum_{b=1}^B \gamma_b d_{C_b}^{r_b}(x^{(b)}). \quad (2.66)$$

Alors, f peut être reformulée comme (2.63) avec, pour tout $b \in \{1, \dots, B\}$,

$$(\forall x^{(b)} \in \mathbb{R}^{L_b}) \quad f_b(x^{(b)}) = \gamma_b d_{C_b}^{r_b}(x^{(b)}). \quad (2.67)$$

La projection épigraphique $P_{\text{epi } f_b}$ associée est donnée par la Proposition 6.

• La **norme** $\ell_{1,2}$ est largement utilisée dans de nombreux problèmes favorisant une parcimonie multivariée [WLJ⁺11; ROF92]. Soit $B \in \mathbb{N}^*$, $(\gamma_b)_{1 \leq b \leq B} \in]0, +\infty[^B$, et

$$(\forall x \in \mathbb{R}^L) \quad f(x) = \sum_{b=1}^B \gamma_b \|x^{(b)}\|_2. \quad (2.68)$$

Alors, f peut être reformulée comme (2.63) avec, pour tout $b \in \{1, \dots, B\}$,

$$(\forall x^{(b)} \in \mathbb{R}^{L_b}) \quad f_b(x^{(b)}) = \gamma_b \|x^{(b)}\|_2. \quad (2.69)$$

La projection épigraphique $P_{\text{epi } f_b}$ associée est donnée par le Corollaire 1.

• La **norme** $\ell_{1,\infty}$ a également suscité beaucoup d'intérêt à des fins de régularisation [TVW05; QCCD09; CH12; SGYB15]. Soit $B \in \mathbb{N}^*$ et, pour tout $b \in \{1, \dots, B\}$, soit $(\gamma_{b,m})_{1 \leq m \leq L_b} \in]0, +\infty[^{L_b}$. Nous supposons que :

$$(\forall x \in \mathbb{R}^L) \quad f(x) = \sum_{b=1}^B \max_{1 \leq m \leq L_b} \gamma_{b,m} |x_m^{(b)}|. \quad (2.70)$$

Alors, la fonction ci-dessus peut être modélisée comme dans (2.63) avec, pour tout $b \in \{1, \dots, B\}$,

$$(\forall x^{(b)} \in \mathbb{R}^{L_b}) \quad f_b(x^{(b)}) = \max_{1 \leq m \leq L_b} \gamma_{b,m} |x_m^{(b)}|. \quad (2.71)$$

La projection épigraphique $P_{\text{epi } f_b}$ associée est donnée par la Proposition 7. Quand $\gamma_{b,m} \equiv 1$, la fonction impliquée dans (2.71) se réduit à la norme ℓ_∞ standard $\|\cdot\|_\infty$, dont l'intérêt en terme de régularisation est discuté dans [TVW05; QCCD09; CH12; SGYB15].

2.2.3.3 Application à la restauration d'images

Modèle direct Dans [CPPPP15], nous avons proposé d'évaluer les performances d'une stratégie de découpe épigraphique dans un cadre de restauration d'image, où $z \in \mathbb{R}^{\bar{N}}$ est un vecteur d'observation tel que :

$$z = TA\tilde{u} + \epsilon. \quad (2.72)$$

où $\tilde{u} \in \mathbb{R}^N$ représente l'image originale, $A : \mathbb{R}^N \rightarrow \mathbb{R}^{\bar{N}}$ est un opérateur de flou, $T : \mathbb{R}^{\bar{N}} \rightarrow \mathbb{R}^{\bar{N}}$ est un opérateur de décimation, et $\epsilon \in \mathbb{R}^{\bar{N}}$ est une réalisation d'un bruit blanc gaussien de moyenne nulle.

Approche variationnelle Dans cette étude nous proposons de résoudre le problème de minimisation suivant :

$$\underset{u \in [0,255]^N}{\text{minimiser}} \quad \|TAu - z\|_2^2 \quad h(Du) \leq \eta, \quad (2.73)$$

où $\eta \geq 0$ et $D : \mathbb{R}^N \rightarrow \mathbb{R}^{|\mathbb{E}|}$ est un opérateur linéaire associé à une transformation d'analyse par exemple l'opérateur TV-*non-local*, défini par

$$Du = \begin{bmatrix} D_1 u \\ \vdots \\ D_N u \end{bmatrix} = \begin{bmatrix} (\omega_{1,n}(u_1 - u_n))_{n \in \mathcal{N}_1} \\ \vdots \\ (\omega_{N,n}(u_N - u_n))_{n \in \mathcal{N}_N} \end{bmatrix}, \quad (2.74)$$

et $h = \|\cdot\|_q^r$ une norme $\ell_{q,r}$ où, pour tout $i \in \{1, \dots, N\}$, $\mathcal{N}_i \subset \{1, \dots, N\} \setminus \{i\}$ modélise le voisinage du pixel i et $(\omega_{i,n})_{n \in \mathcal{N}_i}$ sont des poids positifs. Dans le cas où $\mathcal{N}_i$ contient uniquement les plus proches gradients horizontaux et verticaux de i et $\omega_{i,n} \equiv 1$, (2.74) se réduit à l'opérateur classique de gradient (cf. (2.4)). Des méthodes pour construire $\mathcal{N}_i$ et choisir les poids sont décrites dans [BCM05; GO07; FB12]. Dans nos simulations, nous choisissons une régularisation $\ell_{2,1}$ -TV pour obtenir une première estimée de $\hat{u}$, à partir de laquelle nous calculons les poids par une mesure d'auto-similarité proposée dans [FB12]. Nous considérons

$$(\forall w \in \mathbb{R}^{|\mathbb{E}|}) \quad h(w) = \sum_{b=1}^B \|w^{(b)}\|_q^r, \quad (2.75)$$

où $q \geq 1$, $r \geq 1$ et w est décomposé en blocs comme présenté dans (2.62).

Reformulation épigraphique Comme mentionné dans la section 2.1.4, les algorithmes proximaux peuvent efficacement résoudre le Problème (2.73) tant que les opérateurs proximaux des fonctions non-lisses impliquées peuvent être calculés efficacement. Malheureusement, la projection sur $h(w) \leq \eta$ avec h défini par (2.75) n'a pas de forme explicite. Il est alors possible de contourner ce problème en utilisant le fractionnement épigraphique proposé dans la Section 2.2.3.1. En effet, le problème (2.73) peut être écrit comme :

$$\underset{u \in [0,255]^N, \zeta \in \mathbb{R}^L}{\text{minimiser}} \quad \|TAu - z\|_2^2 \quad \begin{cases} (Du, \zeta) \in E, \\ \zeta \in V, \end{cases} \quad (2.76)$$

où

$$E = \{(x, \zeta) \in \mathbb{R}^{|\mathbb{E}|} \times \mathbb{R}^B \mid (\forall b \in \{1, \dots, B\}) \quad (x^{(b)}, \zeta_b) \in \text{epi}_{\|\cdot\|_q^r}\}, \quad (2.77)$$

$$V = \{\zeta \in \mathbb{R}^B \mid 1_B^\top \zeta \leq \eta\}, \quad (2.78)$$

avec $1_B = (1, \dots, 1)^\top \in \mathbb{R}^B$.

Algorithmes Afin de minimiser (2.76), nous avons considéré deux algorithmes proximaux : l'algorithme proximal primal-dual *Monotone+Lipschitz Forward Backward Forward* (M+LFBF) [CP12], et l'algorithme proximal *Simultaneous-Direction Method of Multipliers* (SDMM) [CP11b]. Notons que M+LFBF et l'algorithme Condat-Vũ sont tous deux deux algorithmes proximaux-duaux et reposent sur des itérations proches. En s'appuyant sur les résultats généraux de convergence de ces algorithmes (cf. [CP12, Theorem 4.2] et [BC17]), les suites $(x^{[k]})_{k \in \mathbb{N}}$ générées par M+LFBF et SDMM convergent vers un minimiseur du Problème (2.76).

Résultats expérimentaux Plusieurs types de normes $\ell_{q,r}$ peuvent être employées pour pénaliser des gradients non-locaux. Le cas $q = r = 2$ peut être vu comme une extension de *NonLocal Means* [BCM05], alors que le cas $q = r = 1$ conduit à *NonLocal Medians* [LO09]. Le cas $r = 1$ et $q \in \{2, +\infty\}$ (i.e. la norme $\ell_{q,1}$) est préférée en imagerie couleur [BC08b; MS12].

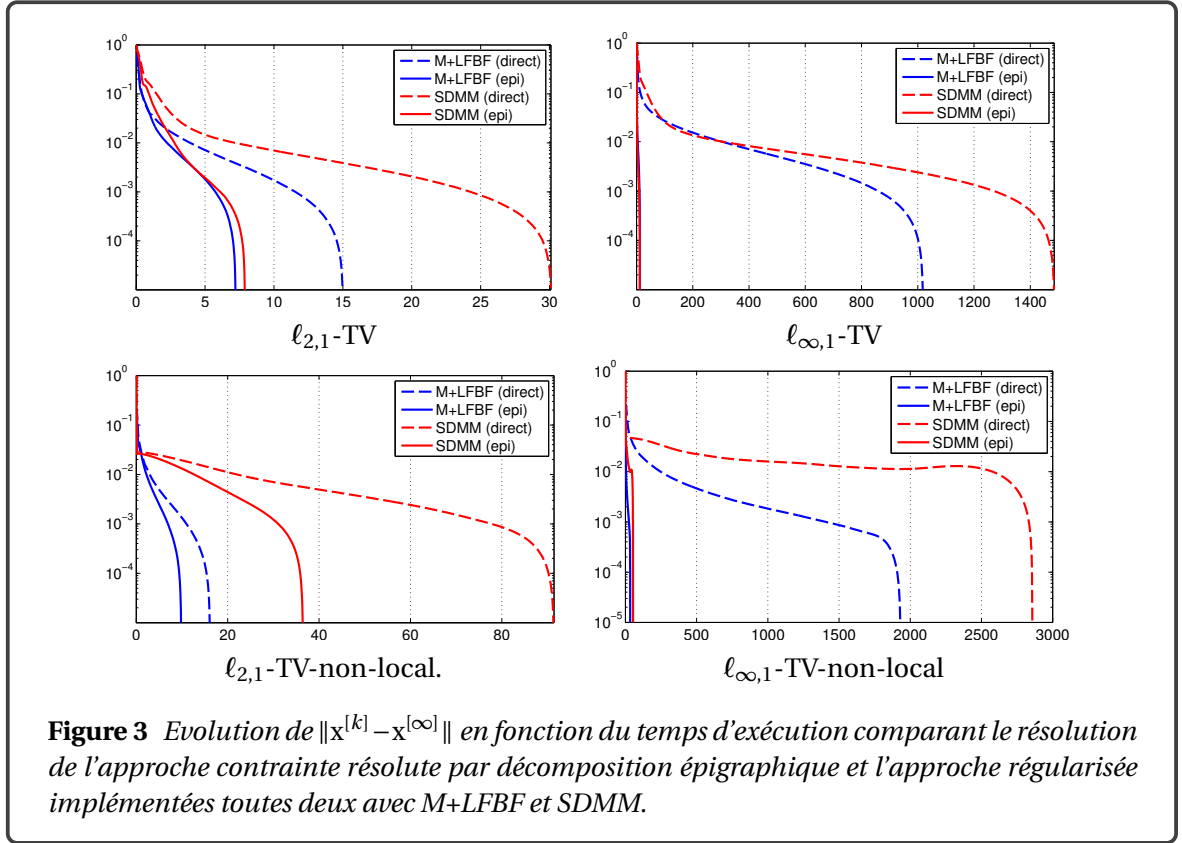
Dans la Figure 3, extraite de [CPPPP15], nous affichons $\|x^{[k]} - x^{[\infty]}\|$ pour M+LFBF et SDMM, lorsque ces algorithmes sont employés pour minimiser (2.76) ou la version équivalente régularisée

$$\underset{u \in [0,255]^N}{\text{minimiser}} \|\text{TA}u - z\|_2^2 + \lambda h(Du). \quad (2.79)$$

Cette figure est obtenue pour une réalisation de l'image couleur dégradée présentée Figure 5 et dont les paramètres de dégradation sont précisés dans la légende.

Le fractionnement épigraphique est plus efficace numériquement que la résolution par une approche régularisée, et cela pour différents choix de q , r , et poids $\omega_{j,n}$. Plus de résultats, en particulier, la sensibilité au choix des paramètres η et λ sont présentés dans [CPPPP15].

Dans la Figure 5 (2ème et 3ème ligne) extraite de [CPPPP14b], nous présentons les résultats visuels pour différents choix de TV non-locale $\ell_{q,r}$ pour différentes valeurs de q et r .



2.2.3.4 Projection épigraphique et pénalisation tensorielle

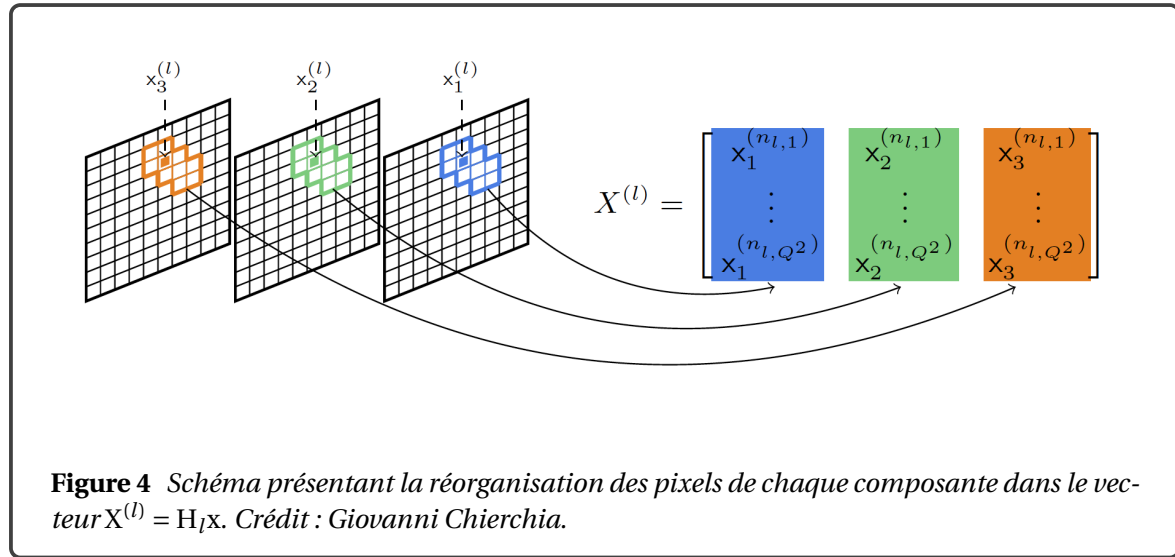
Dans [CPPPP14b], nous avons étendu les travaux présentés dans la section précédente au contexte de la restauration d'images multicomposantes pour lequel il est d'une importance primordiale d'exploiter les corrélations intrinsèques le long des dimensions spatiales et spectrales. Pour cela, nous adoptons une approche variationnelle basée sur l'introduction d'une régularisation par *tenseur de structure de variation totale non locale* (ST-NLTV) et nous montrons comment la mettre en oeuvre en pratique avec des techniques d'optimisation convexe et de découpe épigraphique afin de bénéficier des gains en temps de calculs observés précédemment. Ce travail est une extension des résultats présentés dans [CPPPP15] dans le contexte de régularisations basées sur les normes nucléaires, Frobenius et spectrales. De plus nous proposons une comparaison aux deux méthodes de pointe en imagerie spectroscopique [YZS12; CSZL14].

Approche variationnelle Le modèle de dégradation est de la forme (2.72) dans un contexte où $\tilde{u} \in (\mathbb{R}^N)^M$ et $z \in (\mathbb{R}^{\tilde{N}})^M$. Nous proposons de résoudre la formulation contrainte suivante

$$\underset{u \in (\mathbb{R}^N)^M}{\text{minimiser}} \|TAu - z\|_2^2 \quad \text{s.t.} \quad \sum_{\ell=1}^N \|D_\ell H_\ell u\|_q \leq \eta \quad (2.80)$$

où TA désigne la matrice système associée à une convolution décimée. $\|\cdot\|_q$ modélise une norme Schatten, c'est à dire la norme usuelle appliquée sur les valeurs singulières de $D_\ell H_\ell u$ avec $H_\ell: (\mathbb{R}^N)^M \rightarrow \mathbb{R}^{\tilde{Q}^2 \times M}$ qui sélectionne $\tilde{Q} \times \tilde{Q}$ blocs de chaque composante (incluant le pixel ℓ) et qui le ré-organise dans une matrice de taille $\tilde{Q}^2 \times M$, et où $D_\ell: \mathbb{R}^{\tilde{Q}^2 \times M} \rightarrow \mathbb{R}^{|\mathbb{E}_\ell| \times M}$ est une matrice qui rend parcimonieux $H_\ell u$ avec $|\mathbb{E}_\ell| \leq \tilde{Q}^2$. Une illustration pour $M = 3$ canaux est présentée Figure 4.

Cette régularisation apparaît comme une extension de la variation totale non-locale aux images multicomposantes en se basant sur la notion de tenseur de structure. Le terme de régularisation développé permet alors de pénaliser des variations non-locales de façon conjointes à travers les différentes composantes. La solution algorithmique mise en œuvre fait appel à la projection épigraphique présentée dans la Section 2.2.



Reformulation épigraphique – Le Problème (2.80) peut s'écrire de façon équivalente sous la forme :

$$\underset{u \in (\mathbb{R}^N)^M}{\text{minimiser}} \|TAu - z\|_2^2 \quad \text{s.t.} \quad \Phi u \in C \quad (2.81)$$

où

$$\Phi: u \mapsto X = \begin{bmatrix} D_1 H_1 u \\ \vdots \\ D_N H_N u \end{bmatrix} \begin{matrix} \} X^{(1)} \\ \\ \} X^{(N)} \end{matrix} \quad (2.82)$$

avec $X \in \mathbb{R}^{M \times R}$ et $M = M_1 + \dots + M_N$, alors que C est l'ensemble convexe fermé défini par

$$C = \{X \in \mathbb{R}^{M \times R} \mid \sum_{\ell=1}^N \gamma_\ell \|X^{(\ell)}\|_q \leq \eta\}. \quad (2.83)$$

La projection sur C n'ayant pas de forme explicite, nous avons proposé une technique de fractionnement épigraphique similaire à celle présentée dans la section précédente. Différentes solutions de fractionnement doivent être proposées en fonction de la norme q -Schatten impliquée :

— dans le cas $q = 1$,

$$X \in C \Leftrightarrow \sum_{\ell=1}^N \sum_{m=1}^{\tilde{M}_\ell} \gamma_\ell \left| \sigma_{X^{(\ell)}}^{(m)} \right| \leq \eta, \quad (2.84)$$

— dans le cas $q > 1$,

$$X \in C \Leftrightarrow \sum_{\ell=1}^N \gamma_{\ell} \|\sigma_{X^{(\ell)}}\|_q \leq \eta, \quad (2.85)$$

Le résultat suivant permet de relier la projection sur l'épigraphe associée à une norme matricielle à partir de la projection épigraphique sur la norme vectorielle.

Proposition 8 Pour tout $\ell \in \{1, \dots, N\}$, soit

$$X^{(\ell)} = U^{(\ell)} S^{(\ell)} (V^{(\ell)})^{\top} \quad (2.86)$$

la décomposition en valeurs singulières de $X^{(\ell)} \in \mathbb{R}^{M_{\ell} \times R}$, où

- $(U^{(\ell)})^{\top} U^{(\ell)} = \text{Id}_{M_{\ell}}$,
- $(V^{(\ell)})^{\top} V^{(\ell)} = \text{Id}_{M_{\ell}}$,
- $S^{(\ell)} = \text{diag}(s^{(\ell)})$, avec $s^{(\ell)} = (\sigma_{X^{(\ell)}}^{(m)})_{1 \leq m \leq M_{\ell}}$.

Alors, la projection sur $E = \{(X, \zeta) \in \mathbb{R}^{M \times R} \times \mathbb{R}^N \mid (\forall \ell \in \{1, \dots, N\}) (\sigma_{X^{(\ell)}}, \zeta^{(\ell)}) \in \text{epi} \|\cdot\|_q\}$, est :

$$P_E(X, \zeta) = \left(U^{(\ell)} T^{(\ell)} (V^{(\ell)})^{\top}, \theta^{(\ell)} \right)_{1 \leq \ell \leq N}, \quad (2.87)$$

où $T^{(\ell)} = (t^{(\ell)})$ et,

- dans le cas $q = 1$, pour tout $m \in \{1, \dots, M_{\ell}\}$

$$(t^{(\ell, m)}, \theta^{(\ell, m)}) = P_{\text{epi}|\cdot|}(s^{(\ell)_m}, \zeta^{(\ell)_m}), \quad (2.88)$$

- dans le cas $q > 1$,

$$(t^{(\ell)}, \theta^{(\ell)}) = P_{\text{epi} \|\cdot\|_q}(s^{(\ell)}, \zeta^{(\ell)}). \quad (2.89)$$

Imagerie couleur Dans cette section, nous évaluons les performances de la pénalisation par régularisation ST-NLTV proposée et décrite précédemment et nous la comparons avec une pénalisation standard *canal par canal* (CC) présentée dans la section 2.2.3.3.

La première expérience est centrée sur l'imagerie couleur, c'est-à-dire le cas $R = S = 3$. Les observations bruitées sont obtenues avec le modèle de dégradation dans (2.72).

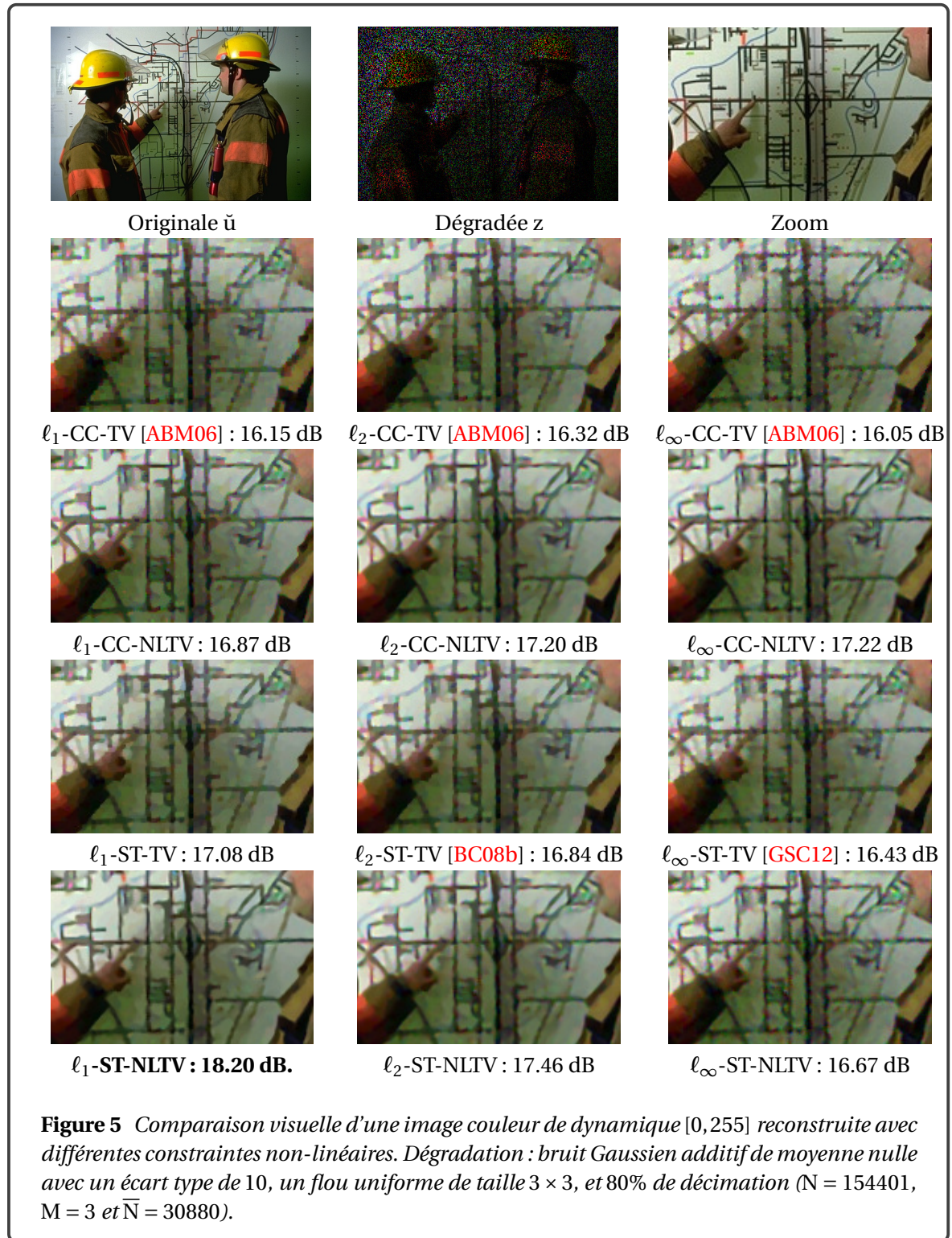
Dans l'exemple de la Figure. 5, nous regroupons les résultats obtenues avec les pénalisations ℓ_q -CC-TV [ABM06], ℓ_q -CC-NLTV (cf. Section 2.2.3.3), ℓ_q -ST-TV [BC08b; GSC12] et la pénalisation ℓ_q -ST-NLTV for $q \in \{1, 2, +\infty\}$ décrite dans cette section. Toutes ces méthodes ont été implémentées avec l'algorithme M+LFBE.

Les résultats présentés Figure. 5 démontrent d'une part l'intérêt de la variation total non locale par rapport à la variation totale standard (cf. lignes 2 et 3 sans couplage entre les canaux et lignes 4 et 5 avec couplage). D'autre part, l'intérêt de considérer des mesures de tenseur de structures non locales est visible sur la comparaison entre les lignes 1 et 3 (TV) et 2 et 4 (NLTV). ℓ_1 -ST-NLTV étant la régularisation la plus efficace. En effet, la régularisation ST-NLTV combine les avantages de ST et NLTV, comme une meilleure préservation des détails et une réduction des traînées de couleur.

Imagerie hyperspectrale Dans cette section, nous comparons la méthode proposée ℓ_1 -ST-NLTV avec des méthodes de l'état-de-l'art en imagerie spectrale : Hyperspectral TV (H-TV) [YZS12] et Multichannel NLTV (M-NLTV) [CSZL14]. Nous présentons Figure 6, les résultats obtenus dans le contexte d'un scénario d'échantillonnage comprimé. À des fins de reproductibilité, nous avons utilisé des images multispectrales et hyperspectrales accessibles au public.⁸ Des tests approfon-

8. <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>

dis ont été réalisés sur plusieurs images de différentes tailles et sont disponibles dans [CPPPP14b].



Les résultats présentés Figure 6 confirment l'intérêt de combiner le principe de non-localité avec des mesures de lissage structure-tenseur. En effet, ℓ_1 -ST-NLTV s'avère ici encore être la régularisation la plus efficace avec des gains de SNR (jusqu'à 1.4 dB) par rapport à M-NLTV, qui à son tour est comparable à ℓ_1 -ST-TV (approche composante par composante).

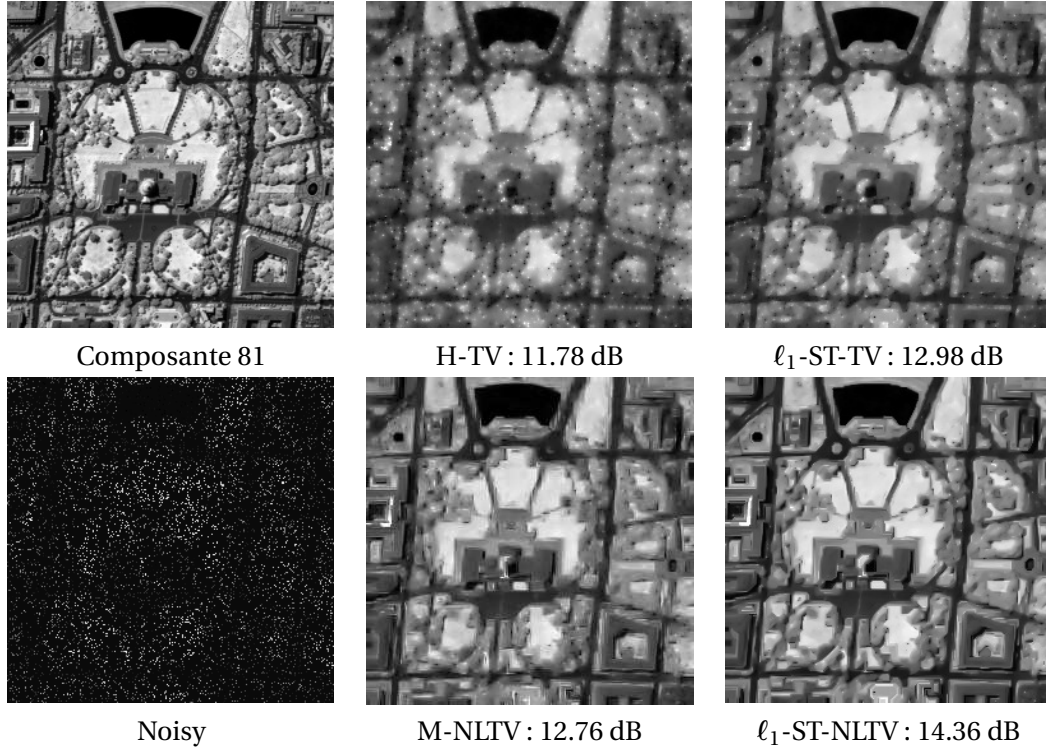


Figure 6 Comparaisons visuelles et quantitatives de reconstructions de l'image hydrique de dynamique $[0, 255]$ avec H-TV [YZS12], ℓ_1 -ST-TV (cf. Section 2.2.3.3), M-NLTV [CSZL14] et ℓ_1 -ST-NLTV. Dégradation : scénario d'échantillonnage comprimé et un bruit additif gaussien d'écart type 5 et une décimation de 90% ($N = 65536$, $R = 191$, $\bar{N} = 6553$).

2.2.3.5 Forme approchée en traitement d'images

Dans [CPTPP12], nous avons considéré le cas où la projection sur $\text{epi } f_b$ défini dans (2.62) n'a pas de forme explicite. Nous avons proposé d'approcher inférieurement f_b par une fonction affine par morceaux de la forme

$$(\forall x^{(b)} \in \mathbb{R}^{L_b}) \quad \hat{f}_b(x^{(b)}) = \max \left\{ f_b(a_m^{(b)}) + (\delta_m^{(b)})^\top (x^{(b)} - a_m^{(b)}) \mid 1 \leq m \leq M_b \right\}, \quad (2.90)$$

où, pour tout $m \in \{1, \dots, M_b\}$, $a_m^{(b)} \in \mathbb{R}^{L_b}$ et $\delta_m^{(b)} \in \partial f_b(a_m^{(b)})$. Une illustration de la fonction approchée est présentée Figure 7. La résolution passe ensuite par le calcul de la projection épigraphique associée à la fonction maximum, mentionnée précédemment, pour laquelle nous avons proposé une forme explicite. Cette question a été regardée de façon plus spécifique dans [CPTPP12] lorsque f_b modélise la divergence de Kullback-Leibler.

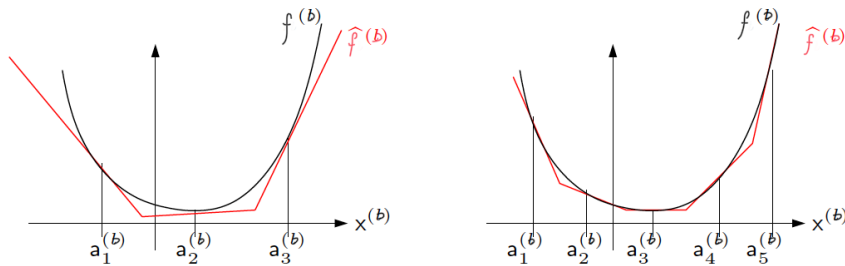


Figure 7 Approximation inférieure affine par morceaux pour $M_b \equiv 3$ (gauche) et $M_b \equiv 5$ (droite)

2.2.3.6 Projection épigraphique et décomposition modale empirique

Dans [PBF14], nous avons appliqué le principe de la découpe épigraphique au contexte de l'analyse de signaux par décomposition modale empirique.

De nombreux signaux qui proviennent de phénomènes naturels ou de systèmes fabriqués par l'homme, tels que la météorologie, l'océanographie, la biologie, font apparaître des comportements non-linéaires et/ou non-stationnaires. Une manière de modéliser une telle classe de signaux est l'écriture sous forme d'une somme de signaux AM-FM

$$(\forall t \in \mathbb{R}) \quad z(t) = \sum_{m=1}^M \alpha_m(t) \cos \theta_m(t) \quad (2.91)$$

où $\alpha_m(t)$ est supposé positif et plus lisse que $\cos \theta_m(t)$. Le challenge associé à ce modèle est double : d'une part, il faut extraire chaque composante $v_m(t) = \alpha_m(t) \cos \theta_m(t)$ à partir de $u(t)$ et, d'autre part, évaluer la fréquence instantanée $\omega_m(t)$ (déduite de $\theta_m(t)$) et l'amplitude instantanée $\alpha_m(t)$. Si $M = 1$, l'amplitude et la fréquence instantanée peuvent être facilement obtenues grâce au concept de signal analytique mais si $M > 1$ cela n'est plus possible. Plusieurs solutions ont été proposées telles que les méthodes de réallocation, du synchrosqueezing ou de la décomposition modale empirique. Ces méthodes peuvent se découper en deux classes. Une première classe regroupe les techniques de réallocation et de synchrosqueezing [DLW11; MOM12; TW11], qui permettent de rendre plus localisée l'information contenue dans une représentation temps-fréquence. La difficulté de mise en œuvre de ces méthodes réside dans une étape nécessitant une supervision pour extraire chacune des composantes. La *décomposition modale empirique* (EMD) [RFG07] appartient quant à elle à la deuxième classe de méthodes. Elle permet d'extraire chaque composante de façon non supervisée mais fait face à la difficulté de ne pas avoir de définition mathématique claire.

Notre première contribution fut de proposer une alternative à l'EMD basée sur une approche variationnelle permettant d'avoir des garanties de convergence, d'extraire des tendances pouvant avoir une forme constante par morceaux, de gérer des signaux ayant des portions nulles, et qui est robuste aux effets d'échantillonnage [PBF14]. Dans ce travail, on suppose que le signal à analyser peut s'écrire sous la forme

$$z = \sum_{m=1}^M v_m + a_M \quad (2.92)$$

où v_m sont des modes oscillants autour de zéro ayant des enveloppes supérieures et inférieures symétriques et où a_M modélise une tendance. La solution proposée consiste à extraire successivement le mode et la tendance d'ordre m à partir de la tendance d'ordre $m - 1$, sachant que $a_0 = z$. Le critère considéré est le suivant :

$$(a_m, v_m) \in \underset{a \in \mathbb{R}^N, v \in \mathbb{R}^N}{\text{Argmin}} \|a_{m-1} - v - a\|_2^2 \quad \text{s.c.} \quad \begin{cases} \|Ha\|_r^r \leq \eta_m, \\ \|D_m v\|_q^q \leq \epsilon_m, \\ (\forall j \in \{0, \dots, m-1\}), |\langle v, v_j \rangle| \leq \eta_{m,j}, \end{cases} \quad (2.93)$$

où le terme quadratique permet d'assurer que la somme des a_m et v_m estimés est proche de a_{m-1} . La première contrainte $\|Ha\|_r^r$ permet d'imposer à la tendance un comportement lisse ou constant par morceaux suivant la valeur de r choisie. On considère généralement un opérateur linéaire de type gradient. La contrainte $\|D_m v\|_q^q$ permet de modéliser le comportement d'un mode, c'est-à-dire oscillant autour de zéro. La dernière contrainte permet d'assurer la quasi-orthogonalité entre les modes. Un exemple de décomposition sur des données réelles est illustré Figure 2.1. La gestion des contraintes non-linéaires dans le cas $r = 1$ repose sur la technique de projection épigraphique présentée dans les sections précédentes.

Une seconde contribution fut d'étendre ce formalisme à la décomposition d'images [SPB⁺14]. Cette version 2-D se rapproche fortement des méthodes de décomposition en texture et géométrie avec une part d'adaptativité supplémentaire. Dans cette étude, nous avons également proposé

une nouvelle méthode d'estimation de l'amplitude, de la fréquence et de l'orientation instantanée basée sur une version améliorée de l'algorithme de Cadzow. Nous avons conduits de nombreux tests pour vérifier la stabilité de notre méthode et le gain obtenu par rapport aux méthodes de l'état-de-l'art que ce soit au niveau de l'étape de séparation ou d'estimation spectrale.

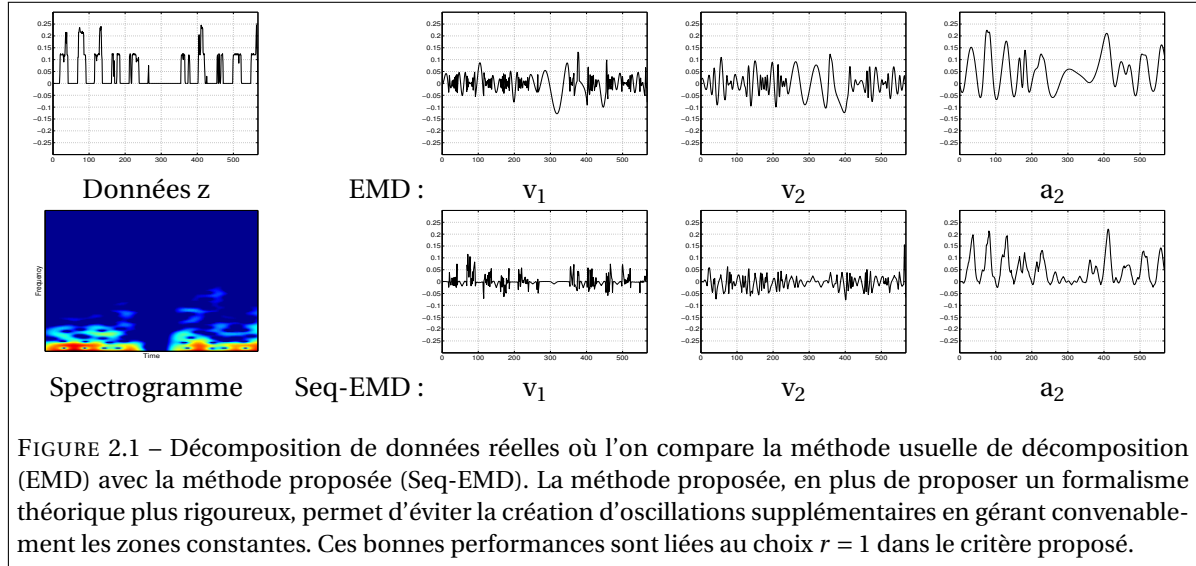


FIGURE 2.1 – Décomposition de données réelles où l'on compare la méthode usuelle de décomposition (EMD) avec la méthode proposée (Seq-EMD). La méthode proposée, en plus de proposer un formalisme théorique plus rigoureux, permet d'éviter la création d'oscillations supplémentaires en gérant convenablement les zones constantes. Ces bonnes performances sont liées au choix $r = 1$ dans le critère proposé.

2.2.3.7 Conclusions

Cette section a montré l'intérêt des contraintes épigraphiques comme outil de gestion de contraintes non-linéaires. De nouvelles formes explicites de projections épigraphiques associées à des fonctions usuellement rencontrées en traitement du signal et des images ont été présentées. Ces formes explicites sont indispensables à une mise en œuvre efficace dans les algorithmes proximaux. Ces nouvelles formes explicites ont été considérées en restauration d'images, potentiellement multi-composantes, où des pénalisations avancées telles que la TV non locale avec structure de tenseur (ST-NLTV) ont été proposées et dont une mise en œuvre sur de grands volumes de données a été rendue possible grâce à l'efficacité d'une approche inverse formulée sous forme contrainte et gérée par la combinaison d'algorithmes proximaux et de découpe épigraphique. Plusieurs algorithmes ont été implémentés et des comparaisons entre l'approche contrainte proposée et son équivalent sous forme régularisée illustrent le gain systématique en temps de calcul d'une formulation contrainte.

Une approximation permettant de gérer des contraintes non linéaires dont la projection épigraphique ne possède pas de forme explicite, par exemple lorsque f correspond à la divergence de Kullback-Leibler, a été proposée. Cette approximation repose sur l'évaluation du sous-gradient de f_b en un nombre fini de points dont une étude approfondie du nombre et de la position mériterait d'être menée. Il serait par exemple intéressant d'exploiter le formalisme des méthodes à faisceaux [HUL01] pour guider ces choix.

Enfin, cette découpe épigraphique a été mis en œuvre pour proposer une formulation variationnelle de la décomposition modale empirique (EMD) robuste aux effets d'échantillonnage et permettant de gérer des signaux contenant des informations manquantes. Ces travaux étaient un premier pas vers une formulation variationnelle de l'EMD mais mériteraient d'être repensés par des approches de type Prony locales [SPB⁺14] ou par décomposition tensorielles.

2.2.4 Contraintes épigraphiques dans l'étude des environnements circumstellaires

L'étude des environnements circumstellaires fut le cœur de la thèse de Laurence Denneulin. Ses principales contributions furent de modéliser le problème direct à la fois via le formalisme de Jones puis via formalisme de Stokes conduisant alors à un modèle linéaire. Plusieurs critères de la forme (2.1) ont été proposés dans sa thèse. Dans cette section, nous concentrons notre attention sur l'une des formulations variationnelles convexes proposées, qui par la physique du problème, conduit à une contrainte épigraphique.

L'imagerie polarimétrique est l'une des techniques les plus efficaces pour l'imagerie à haut contraste et la caractérisation des environnements circumstellaires. Ces environnements peuvent être caractérisés par polarimétrie en imagerie directe aux longueurs d'ondes proches infrarouge. L'instrument Spectro-Polarimetric High Contrast Exoplanet REsearch (SPHERE)/IRDIS installé sur le VLT (Very Large Telescope) dans son mode d'imagerie polarimétrique à double faisceau (DPI), offre la possibilité d'acquérir des images polarimétriques à haut contraste et à haute résolution angulaire. Cependant, un traitement d'image dédié est nécessaire pour éliminer les dégradations par la lumière stellaire, les effets de polarisation instrumentale, et les dégradations optiques.

Linéarisation : Du formalisme de Jones au paramètres de Stokes Les principaux paramètres d'intérêt sont l'intensité de la lumière polarisée linéairement I^p et l'angle de polarisation θ qui est causé par la réflexion de la lumière stellaire sur la poussière circumstellaire. En modulant l'orientation de la polarisation instrumentale, ces paramètres peuvent être dissociés de l'intensité I^u de la lumière non polarisée reçue de l'étoile et de son environnement. Le modèle qui relie les paramètres d'intérêts à l'intensité à l'entrée du détecteur s'exprime par la relation non-linéaire suivante :

$$(\forall j \in \{1, 2\})(\forall k \in \{1, \dots, K\}) \quad I_{j,k}^{\text{det}} = \frac{1}{2} \underbrace{(I_{\text{étoile}}^u + I_{\text{disque}}^u)}_{I^u} + I^p \cos^2(\theta - 2\alpha_k - \psi_j) \quad (2.94)$$

où ψ_j désigne l'angle d'orientation du polariseur gauche/droit et α_k correspond à l'angle de plaque demi-onde (HWP) pour les K acquisitions. Lors d'une séquence d'observations avec SPHERE / IRDIS en mode DPI, le HWP est tourné plusieurs fois le long d'un cycle donné d'angles $\alpha \in \{0^\circ, 45^\circ, 22.5^\circ, 67.5^\circ\}$. L'ensemble de données consiste en K trames composées de deux, gauche et droite, sous-images, chacune avec une position différente du HWP. L'ordre de grandeur typique de K varie de 32 à plus de 512 selon la cible observée. Les valeurs typiques de α_k et ψ_j sont listées dans le Tableau 2.1 (1ère et 2ème colonnes).

Afin de simplifier la résolution du problème, nous avons proposé de reformuler l'intensité du détecteur à l'aide des paramètres de Stokes qui représentent la lumière totale I , la lumière polarisée linéairement dans deux directions à 45° l'une de l'autre, notée Q et U :

$$\begin{cases} I = I^u + I^p \\ Q = I^p \cos(2\theta) \\ U = I^p \sin(2\theta) \end{cases} \quad (2.95)$$

Le quatrième paramètre de Stokes caractérisant la lumière polarisée circulairement ne peut pas être mesuré avec un instrument tel que SPHERE / IRDIS (une modulation par une plaque quart d'onde aurait été nécessaire). Pour la plupart des cibles astronomiques d'intérêt, la lumière polarisée circulairement est négligeable, donc cette limitation n'est pas un problème en pratique. Dans la suite, on néglige donc la lumière polarisée circulairement.

En combinant (2.94) et (2.95), puis par relations usuelles de trigonométrie nous obtenons une expression linéaire de l'intensité $I_{j,k}^{\text{det}}$ en fonction des paramètres de Stokes :

$$I_{j,k}^{\text{det}} = \frac{1}{2} I + \frac{1}{2} \cos(4\alpha_k + 2\psi_j) Q + \frac{1}{2} \sin(4\alpha_k + 2\psi_j) U, \quad (2.96)$$

ou de façon équivalente :

$$I_{j,k}^{\text{det}} = v_{j,k,1} I + v_{j,k,2} Q + v_{j,k,3} U \quad (2.97)$$

où la Table 2.1 liste les valeurs des coefficients $(v_{j,k,m})_{1 \leq m \leq 3}$ pour des choix usuels d'angle d'orientation du polariseur et d'angle de plaque demi-onde.

α_k (HWP)	ψ_j (Analyzer)	$v_{j,k,1}$	$v_{j,k,2}$	$v_{j,k,3}$
0°	0° (left : $j = 1$)	0.5	0.5	0
0°	90° (right : $j = 2$)	0.5	-0.5	0
45°	0° (left : $j = 1$)	0.5	-0.5	0
45°	90° (right : $j = 2$)	0.5	0.5	0
22.5°	0° (left : $j = 1$)	0.5	0	0.5
22.5°	90° (right : $j = 2$)	0.5	0	-0.5
77.5°	0° (left : $j = 1$)	0.5	0	-0.5
77.5°	90° (right : $j = 2$)	0.5	0	0.5

TABLEAU 2.1 – Ce tableau liste les positions du HWP α_k et les orientations du polariseur ψ_j ainsi que les valeurs correspondantes des coefficients $v_{j,k,\ell}$ calculés par Eq. (2.97) en supposant qu'il n'y a pas de polarisation instrumentale et en ignorant le champ de rotation.

Il est ensuite possible de reconstruire les paramètres d'intérêt I^u , I^p , et θ à partir d'une combinaison des paramètres Stokes grâce aux relations suivantes :

$$\begin{cases} I^p = \sqrt{Q^2 + U^2} \\ \theta = (1/2) \arctan(U/Q) \mod \pi \\ I^u = I - \sqrt{Q^2 + U^2}. \end{cases}$$

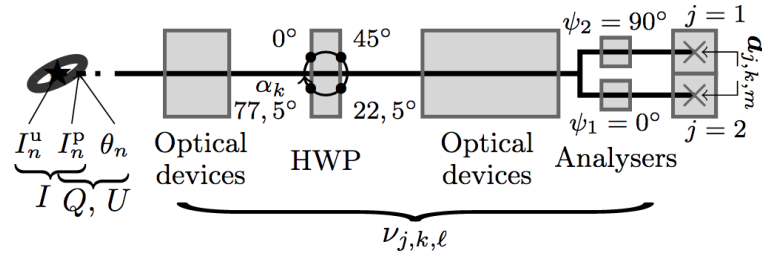


Figure 8 Représentation schématique de l'instrument ESO/VLT-SPHERE IRDIS montrant les différentes parties optiques pouvant introduire des effets de polarisation.

Modèle de dégradation En imagerie polarimétrique, chaque paramètre polarimétrique est une fonction du champ de vue bidimensionnel. On considère que les paramètres de Stokes sont représentés par des images de N pixels chacun et nous notons $S_{m,n}$ la valeur du n -ième pixel pour la m -ième composante de Stokes (i.e. $S_1 = I$, $S_2 = Q$ et $S_3 = U$).

Sous condition que les effets de polarisation s'appliquent uniformément sur l'ensemble du champ de vue et que la réponse impulsionnelle spatiale (PSF) ne dépend pas de la polarisation de la lumière, tous les paramètres de Stokes d'une source spatialement incohérente sont indépendamment et identiquement affectés par la PSF spatiale [BRW18]. Comme les effets de la PSF spatiale sont linéaires, nous pouvons écrire :

$$z_{j,k} = \sum_{m=1}^3 v_{j,k,m} T_{j,k} A S_m + \epsilon_{j,k}, \quad (2.98)$$

où $A: \mathbb{R}^N \rightarrow \mathbb{R}^N$ modélise un flou invariant par translation et $T_{j,k}: \mathbb{R}^N \rightarrow \mathbb{R}^{\bar{N}_j}$ modélise une transformation géométrique du j -ième polariseur de l'analyseur durant l'acquisition k . $\varepsilon_{j,k}$ modélise une perturbation aléatoire inconnue due au bruit. Ce bruit peut être supposé centré et indépendant entre pixels ou images et de variance $\Sigma_{j,k} = \text{Var}(z_{j,k})$ estimé lors de l'étape d'étalonnage. Il est à noter que puisque le modèle proposé n'est pas valide dans le voisinage du masque coronographique, les pixels de cette région ne peuvent être pris en compte. De plus, le détecteur contient des pixels défectueux (e.g. des pixels morts, des pixels saturés) qui ne doivent également pas être pris en compte. Ces pixels peuvent être gérés en supposant que leur variance est infinie.

La Figure 8 fournie une représentation de l'instrument.

RHAPSODIE : approche inverse et contrainte épigraphique Imposer la positivité des intensités restaurées a montré son efficacité en imagerie astronomique étant donné qu'une grande partie de l'image correspond à des pixels du fond où les valeurs devraient être nulles. Quelque soit la paramétrisation (I^u, I^p, θ) ou (I, Q, U) , les intensités I , I^u , et I^p doivent toutes être positives. Dans le domaine des paramètres de Stokes = (I, Q, U) , la positivité conduit à une contrainte épigraphique :

$$E = \left\{ (I, Q, U) \in \mathbb{R}^{N \times 3} \mid \forall n \in \{1, \dots, N\}, I_n \geq \sqrt{Q_n^2 + U_n^2} \right\}. \quad (2.99)$$

Une telle contrainte a précédemment été employée dans un contexte d'interférométrie dans [BRW18], mais n'avait encore jamais été mise en œuvre en imagerie polarimétrique haut contraste. Cette contrainte est intégrée à un formalisme usuel de la forme (2.1) où le terme d'attache aux données est motivé par [Tar05] et prend la forme d'une distance au carré de Mahalanobis $\|\cdot\|_{\Sigma^{-1}}^2 = \langle \cdot, \Sigma^{-1} \cdot \rangle$. Afin de régulariser les estimées nous proposons de regarder l'impact de différentes pénalisations φ (TV, TGV, Schatten, TV-l1). Le critère que nous proposons de minimiser prend la forme :

$$\hat{u} \in \underset{u=(I,Q,U) \in E}{\text{Argmin}} \sum_{j=1}^2 \sum_{k=1}^K \left\| z_{j,k} - T_{j,k} A(v_{j,k,1} I + v_{j,k,2} Q + v_{j,k,3} U) \right\|_{\Sigma_{j,k}^{-1}}^2 + \lambda \varphi(u), \quad (2.100)$$

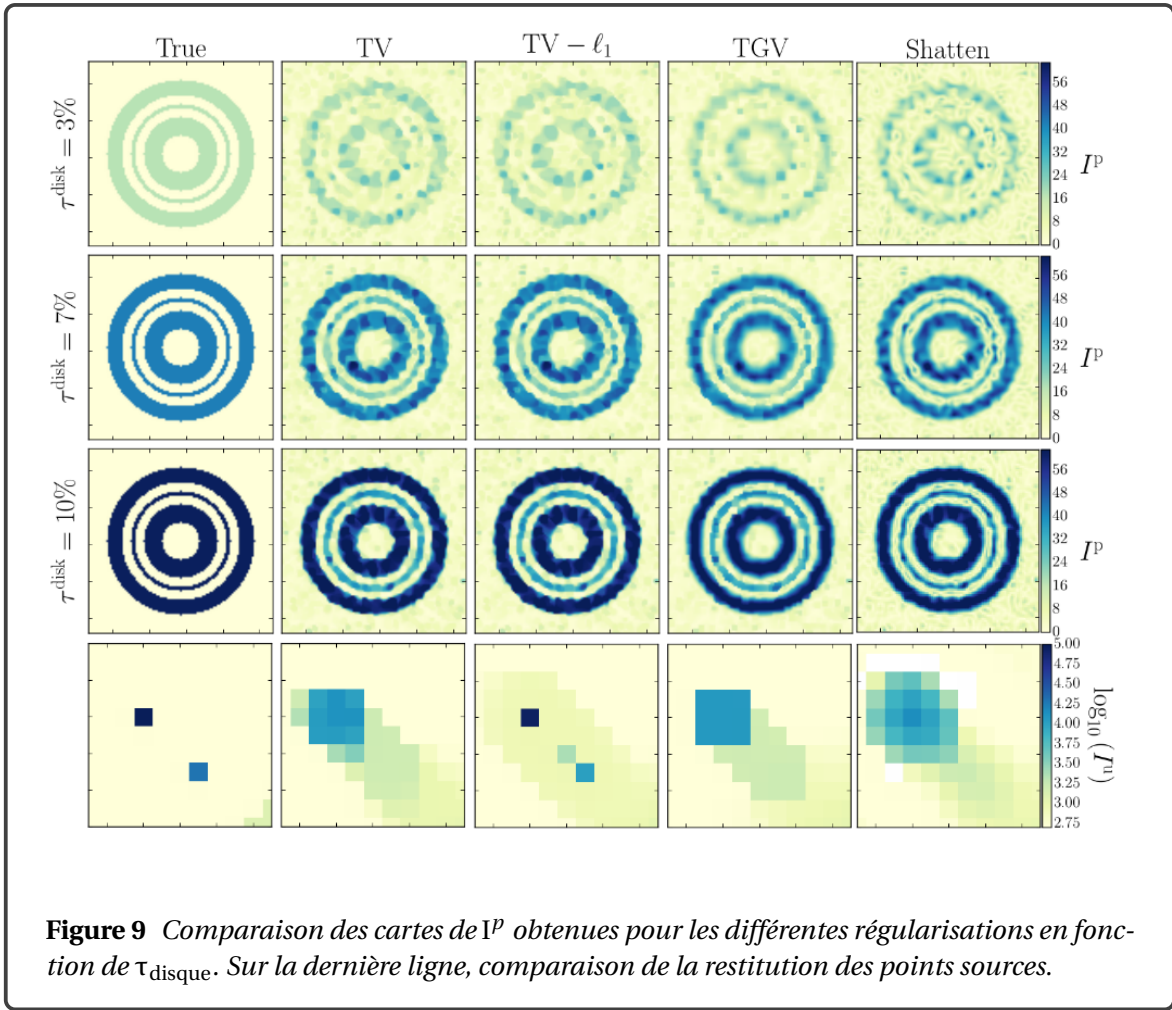
Algorithme La minimisation de (2.100) est effectuée par l'algorithme primal-dual Condat-Vũ et la projection sur la contrainte épigraphique E est donnée par le Corollaire 1.

Résultats numériques Nous avons simulé des données associées à un taux de polarisation $\tau_{\text{disque}} = I^p / (I_{\text{disque}}^u + I^p)$ donné. Ce taux n'est pas estimable en pratique car I_{disque}^u est intégrée dans I^u . Les jeux de données sont simulés avec $\tau_{\text{disque}} = \{10\%, 7\%, 3\%\}$, correspondant respectivement à un cas simple, intermédiaire et difficile, une intensité polarisée faible étant plus facilement noyée dans le bruit de photons. Les données sont composées de trois anneaux notés A1, A2, et A3, de mêmes intensités mais de tailles différentes. Deux points sources sont ajoutés dans I^u . Les images simulées sont ensuite combinées d'après (2.98) et altérées par l'ajout de 10% de pixels non valides pris aléatoirement. Chaque jeu de données est constitué de 64 réalisations du bruit.

La Figure 9, montre que les cartes de I_p estimées en régularisant par Schatten apparaissent plus précises pour les cas difficiles. Cette figure montre également l'efficacité de TV+ ℓ_1 par rapport aux autres régularisations pour la reconstruction des points sources. Le Tableau 2.2 permet de confirmer cela. On voit également que les méthodes de l'état-de-l'art ont tendance à surestimer le taux de polarisation dans les cas les plus simples.

τ_{disk}		SNR	TV	TGV	Shatten	TV- ℓ_1	Non-Lin.	Dbl. Ratio	Dbl. Diff.
3 %	A ₁	1,69	0,904	0,890	0,986	0,903	0,886	0,956	0,956
	A ₂	3,23	0,602	0,633	0,825	0,602	0,492	0,499	0,493
	A ₃	4,08	0,967	0,957	0,985	0,967	0,860	0,889	0,859
7 %	A ₁	4,11	0,955	0,924	0,968	0,954	0,941	1,086	1,086
	A ₂	7,80	0,665	0,725	0,839	0,662	0,540	0,558	0,539
	A ₃	9,68	0,962	0,950	0,957	0,962	0,855	0,899	0,853
10 %	A ₁	6,05	0,950	0,914	0,948	0,950	1,293	1,570	1,570
	A ₂	11,45	0,618	0,685	0,770	0,618	0,697	0,720	0,688
	A ₃	14,17	0,980	0,959	0,964	0,980	0,912	1,037	0,908

TABLEAU 2.2 – SNR moyens de chaque anneaux entre les quantités $I^p/(I^p + I^u)$ estimées et réelles avec les différentes régularisations, la méthode non linéaire et les méthodes de l'état de l'art.



2.2.4.1 Conclusions

La travail de thèse de Laurence Denneulin fut dédié à la modélisation du système d'acquisition des données issues de l'instrument Spectro-Polarimetric High Contrast Exoplanet REsearch (SPHERE)/IRDIS et a permis de proposer un modèle direct linéaire permettant de prendre en compte l'ensemble des transformations subies par le signal d'intérêt (i.e. I^u, θ) lors de l'acquisition des données telles que les transformations géométriques, les dégradations optiques, le bruit de mesure, les pixels morts et une paramétrisation adéquate. Muni de ce modèle direct, une formulation variationnelle de la forme (2.1) et sa mise en œuvre sur données synthétiques et sur données réelles a pu être envisagée.

Sur des données synthétiques, nous avons pu illustrer l'intérêt de pénalisations avancées de type Schatten. Cependant, à ce stade, la résolution sous forme régularisée non-lisse ne nous a pas

permis d'obtenir des algorithmes suffisamment rapides, malgré une implémentation avec stratégie de préconditionnement ou *backtracking*, pour être envisagés pour la gestion de données réelles.

Pour la gestion de données réelles, nous nous sommes limités à une pénalisation de type TV hyperbolique (cf. (2.7)) conduisant à un critère différentiable pouvant être minimisé par des algorithmes de type VMLMB (BFGS à mémoire limité [Noc80]) dont une forme étendue permet de gérer des contraintes simples séparables de type positivité [Thi02].

Au vu de ces conclusions et des résultats présentés dans la section précédente 2.2.3, un axe prometteur afin d'accélérer les algorithmes et de simplifier le choix des hyperparamètres, serait d'envisager une résolution sous forme contrainte puis d'avoir recours à une découpe épigraphique. Une telle formulation conduirait à des contraintes épigraphiques présentes à la fois au niveau de la gestion du terme de pénalisation ou d'attache aux données et également au niveau de la contrainte permettant d'assurer la positivité des intensités (cf. (2.99)).

2.2.5 Contraintes épigraphiques et apprentissage machine

Une troisième classe de problèmes sur lesquels j'ai travaillé et dans lesquels les contraintes épigraphiques jouent un rôle important sont les problèmes d'apprentissage machine qui ont conduit aux contributions [CPPPP16; JPJ⁺19].

2.2.5.1 SVM multiclass

Dans la section introductive aux SVM (cf. section 2.1.3) une contrainte épigraphique a été introduite sans la nommer dans l'eq. (2.17).

Rappelons qu'un problème de classification permet de prédire la classe $q \in \{1, \dots, Q\}$ à laquelle appartient une observation $z \in \mathbb{R}^{\bar{N}}$ (e.g. un signal, une image ou un graphe). Ce problème consiste à estimer des vecteurs de discriminations $(x^{(q)})_{1 \leq q \leq Q}$ à partir d'un ensemble de S paires entrées-sorties

$$\mathcal{S} = \{(z_s, \tilde{u}_s) \in \mathbb{R}^{\bar{N}} \times \{1, \dots, Q\} \mid s = \{1, \dots, S\}\}.$$

L'estimation de $x = (x^{(q)})_{1 \leq q \leq Q}$ passe par la résolution du problème de minimisation de la forme (2.17), généralisé dans cette section afin de prendre en compte une plus grande classe de pénalisation $\boldsymbol{\varphi}$ sur x (cf. [BJMO12]) :

$$\underset{(x, \zeta) \in \mathbb{R}^{(\bar{M}+1)Q} \times [0, +\infty]^S}{\text{minimiser}} \quad \boldsymbol{\varphi}(x) + \alpha \sum_{s=1}^S \zeta_s \quad \text{s.t.} \quad \max\{0, v_s + \max_{q \neq \tilde{u}_s} \widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\tilde{u}_s)})\} \leq \zeta_s, \quad (2.101)$$

et dont la formulation régularisée prend la forme :

$$\underset{x \in \mathbb{R}^{(\bar{M}+1)Q}}{\text{minimiser}} \quad \boldsymbol{\varphi}(x) + \alpha \sum_{s=1}^S \max\{0, v_s + \max_{q \neq \tilde{u}_s} \varphi(z_s)^\top (x^{(k)} - x^{(\tilde{u}_s)})\}, \quad (2.102)$$

Ces deux problèmes peuvent être reformulés sous forme synthétique comme :

$$\underset{(x, \zeta) \in \mathbb{R}^{(\bar{M}+1)Q} \times [0, +\infty]^S}{\text{minimiser}} \quad \boldsymbol{\varphi}(x) + \alpha \sum_{s=1}^S \zeta_s \quad \text{s.t.} \quad (\forall s \in \{1, \dots, S\}) \quad \boldsymbol{\Psi}_s(T_s x) \leq \zeta_s, \quad (2.103)$$

et

$$\underset{x \in \mathbb{R}^{(\bar{M}+1)Q}}{\text{minimiser}} \quad \boldsymbol{\varphi}(x) + \alpha \sum_{s=1}^S \boldsymbol{\Psi}_s(T_s x), \quad (2.104)$$

où pour tout $s \in \{1, \dots, S\}$, l'opérateur linéaire $T_s: \mathbb{R}^{(\bar{M}+1)Q} \mapsto \mathbb{R}^Q$ est défini comme :

$$(\forall x \in \mathbb{R}^{(\bar{M}+1)Q}) \quad T_s x = [\widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\tilde{u}_s)})]_{1 \leq q \leq Q},$$

et le vecteur $c_s = (c_s^{(k)})_{1 \leq k \leq Q} \in \mathbb{R}^Q$ est défini par :

$$(\forall q \in \{1, \dots, Q\}) \quad c_s^{(q)} = \begin{cases} 0, & \text{si } q = \check{u}_s, \\ v_s, & \text{sinon,} \end{cases}$$

et la fonction $\Psi_s: \mathbb{R}^Q \mapsto \mathbb{R}$ est définie, pour tout $w = (w_q)_{1 \leq q \leq Q} \in \mathbb{R}^Q$, par

$$\Psi_s(w) = \max_{1 \leq q \leq Q} w_q + c_s^{(q)}, \quad (2.105)$$

Algorithme Les problèmes de minimisation (2.103) et (2.104) font tous deux intervenir un terme d'attache aux données (potentiellement sous forme contrainte) et une pénalisation non-lisse. Par conséquent, nous proposons une résolution basée sur l'algorithme Condat-Vũ.

La projection sur epi_{Ψ_s} possède une forme explicite dérivée de la Proposition 7 alors que prox_{Ψ_s} peut être calculé via la projection sur le simplexe.

La formulation sous forme de contrainte épigraphique (2.103) conduit à deux avantages potentiels. Premièrement, la projection $P_{\text{epi}_{\Psi_s}}$ est calculée sur un sous ensemble convexe de plus faible dimension epi_{Ψ_s} dans $\mathbb{R}^Q \times \mathbb{R}$, dont la dimension est fixée par le nombre Q de classes. Deuxièmement, ces projections peuvent être calculées en parallèles, puisqu'elles sont définies sur des blocs disjoints dont le nombre est donné par la cardinalité S de l'entraînement (cf. [GCPP12] pour un exemple d'implémentation en parallèle sur GP-GPUs).

Dans les simulations présentées dans le paragraphe suivant **hinge (Alg.1)** (resp. **hinge (Alg.2)**) fera référence à la minimisation de (2.104) (resp. (2.103)) par l'algorithme Condat-Vũ décrit précédemment.

Résultats numériques Dans [CPCPP14a; CPCPP16], nous avons évalué les erreurs de classification obtenues avec les SVM multiclassés et *a priori* de parcimonie formulés dans (2.103) et (2.104) et nous les avons comparé à des approches de l'état-de l'art :

- **Square** – SVM multiclassé proposé dans [BSU13] résolue avec FISTA.

$$\underset{x \in \mathbb{R}^{(\bar{M}+1)Q}}{\text{minimiser}} \quad \Phi(x) + \alpha \sum_{s=1}^S \sum_{q \neq \check{u}_s} \left(\max \{0, v_s + \widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\check{u}_s)})\} \right)^2, \quad (2.106)$$

- **Logit** – la régression logistique multinomiale (e.g., [KCFH05]) résolue avec FBS.

$$\underset{x \in \mathbb{R}^{(\bar{M}+1)Q}}{\text{minimiser}} \quad \Phi(x) + \alpha \sum_{s=1}^S \log \left(1 + \sum_{q \neq \check{u}_s} \exp \left\{ v_s + \widetilde{\mathcal{F}}(z_s)^\top (x^{(q)} - x^{(\check{u}_s)}) \right\} \right). \quad (2.107)$$

- **One-vs-all** – le SVM binaire [LFC⁺14] basé sur une stratégie *one-vs-all*, qui pour tout $q \in \{1, \dots, Q\}$, se formule

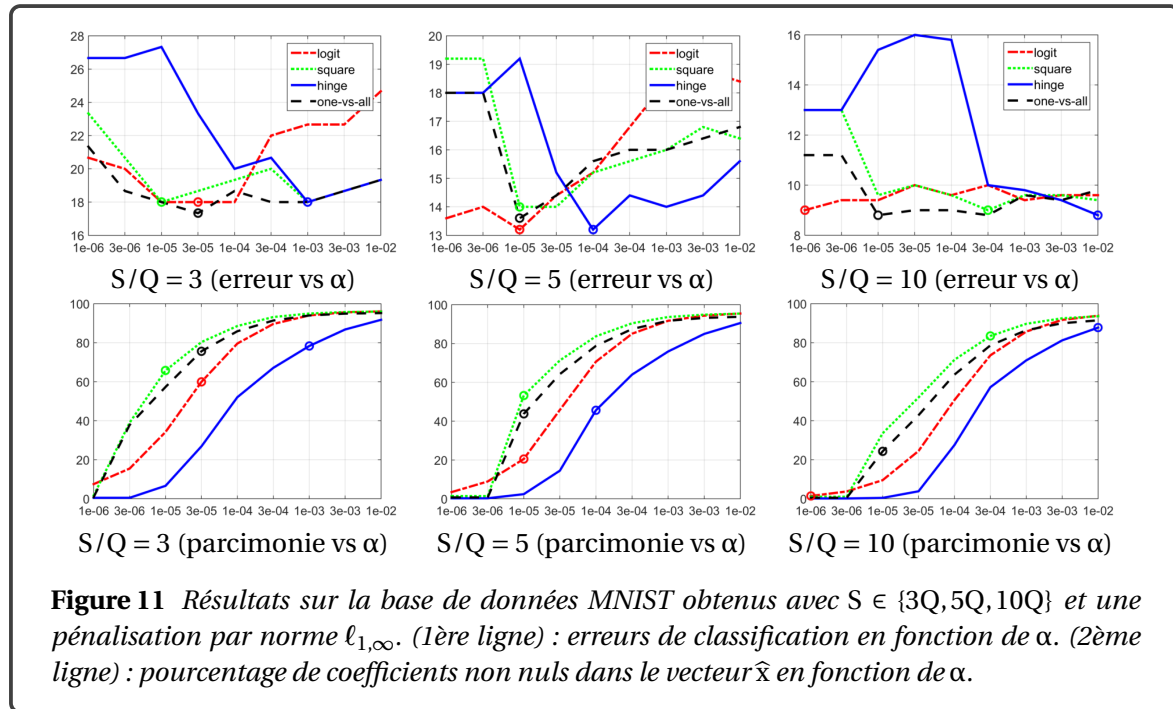
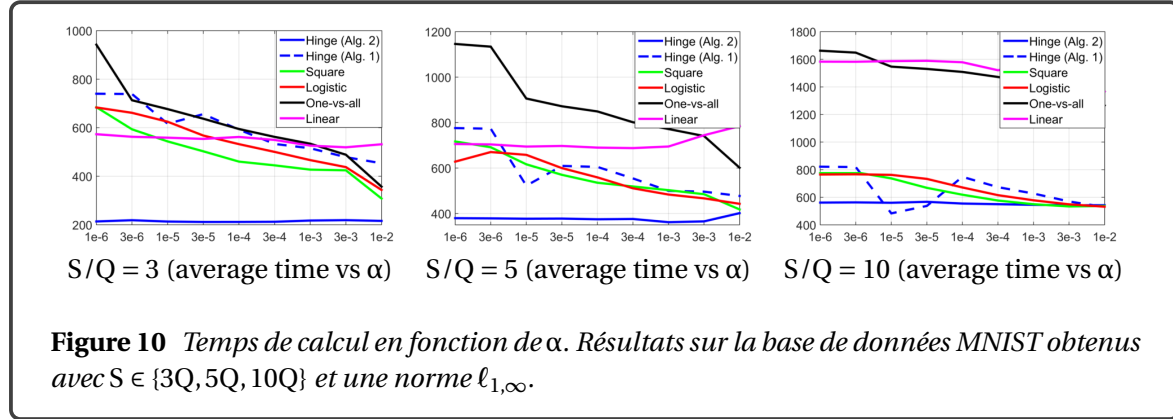
$$\underset{x^{(q)} \in \mathbb{R}^{(\bar{M}+1)Q}}{\text{minimiser}} \quad \Phi(x) + \alpha \sum_{s=1}^S \left(\max \{0, v_s + \widetilde{z}_\ell \widetilde{\mathcal{F}}(z_s)^\top x^{(s)}\} \right)^2, \quad (2.108)$$

avec $\widetilde{z}_s$ égal à 1 si $\check{u}_s = q$, et -1 sinon. On peut noter que (2.108) peut être vu comme un cas particulier de (2.106).

La Figure 10 compare les temps d'exécution (moyennés entre 10 différents ensembles d'apprentissage) obtenus avec les algorithmes mentionnés précédemment pour différentes valeurs de α et η sur la base de données MNIST dans le cas où $S \in \{3Q, 5Q, 10Q\}$. Une implémentation avec l'algorithme Condat-Vũ du Problème (2.103) reformulé sous forme de contraintes linéaires est également proposée. Cette approche est conceptuellement similaire aux méthodes de programmation linéaire proposées par [WS07] et [ZLWZ08] pour des SVM régularisés ℓ_1 - ou $\ell_{+\infty,1}$. Ces résultats démontrent que l'algorithme proposé reposant sur la minimisation de (2.104) est

comparable en termes de temps d'exécution aux autres approches, alors que l'algorithme proposé reposant sur la découpe épigraphique (2.103) est plus rapide.

La Figure 11 (1ère ligne) affiche les erreurs de classification en fonction de l'hyperparamètre de régularisation. Ces résultats ont été obtenus avec une régularisation $\ell_{1,\infty}$, car elle conduisait aux meilleurs résultats dans toutes nos expériences sur cette base de données. Les erreurs de classification indiquent que l'approche *hinge* est comparable aux autres. D'autre part, la Figure 11 (2ème ligne) reporte le pourcentage de coefficients nuls dans les vecteurs $\hat{x}$ en fonction de α . Les graphiques montrent que l'approche *hinge* donne des solutions aussi parcimonieuses que les autres. Seul le résultat obtenu avec l'approche contrainte (i.e **hinge (Alg2)**) est présenté.



2.2.5.2 Interactions quadratiques pour la régression et la classification

Les travaux précédents ont attiré l'attention de TOTAL SA qui a souhaité initier une collaboration autour de cette thématique et plus exactement sur l'estimation conjointe des caractéristiques discriminantes parmi les caractéristiques $\mathcal{F}(z)$ et l'identification des interactions significatives entre ces caractéristiques dans le contexte de la régression et de la classification. Dans le cas de la régression, de nombreux travaux avaient déjà abordé ce sujet. Nous avons proposé un cadre unificateur que nous détaillons dans cette section dans le contexte de la régression. L'ensemble d'apprentissage de régression est noté $\mathcal{S} = \{(z_s, \tilde{u}_s) \in \mathbb{R}^N \times \mathbb{R}^N \mid s = \{1, \dots, S\}\}$.

Le problème qui nous intéresse est de la forme :

$$(\hat{x}, \hat{\Theta}) \in \underset{\substack{x \in \mathbb{R}^{\bar{M}} \\ \Theta \in \mathbb{R}^{\bar{M} \times \bar{M}}}}{\text{Argmin}} \frac{1}{2} \sum_{s=1}^S (\tilde{y}_s - \mathcal{F}(z_s)^\top x - \mathcal{F}(z_s)^\top \Theta \mathcal{F}(z_s))^2 + \lambda \boldsymbol{\varphi}(x, \Theta) \quad (2.109)$$

où la pénalisation $\boldsymbol{\varphi}$ est définie pour tout vecteur de poids de régression $x = (x_i)_{1 \leq i \leq \bar{M}}$ et toute matrice d'interaction $\Theta = (\Theta_{i,j})_{1 \leq i \leq \bar{M}, 1 \leq j \leq \bar{M}}$,

$$\boldsymbol{\varphi}(x, \Theta) = \frac{1}{2} \|\Theta\|_1 + \sum_{i=1}^{\bar{M}} \|(x_i, g(\Theta_{i,\cdot}))\|_q + \iota_C(\Theta) \quad (2.110)$$

avec $g: \mathbb{R}^{\bar{M}} \rightarrow \mathbb{R}$, $1 \leq q \leq +\infty$, et ι_C correspond à la fonction indicatrice d'un ensemble convexe fermé $C \subset \mathbb{R}^{\bar{M} \times \bar{M}}$.

Choix de la contrainte C : Hiérarchie faible et forte La pénalisation $\boldsymbol{\varphi}$ crée une relation entre x et Θ , comme illustré sur la Fig. 12. Deux types de contraintes de hiérarchie sont définies dans "hierNet" [BTT13] : (i) la *hiérarchie faible* où les interactions se produisent (i.e., $\Theta_{i,j} \neq 0$) seulement si l'un des descripteurs associé dans x est non-nul (i.e., $x_i \neq 0$ ou $x_j \neq 0$) et (ii) la *hiérarchie forte* lorsque les deux caractéristiques associées sont différentes de zéro. Ces deux types de hiérarchies sont imposées au moyen de l'ensemble convexe fermé $C \subset \mathbb{R}^{\bar{M} \times \bar{M}}$. La hiérarchie faible est obtenue avec $C = \mathbb{R}^{\bar{M} \times \bar{M}}$ et une hiérarchie forte en imposant une structure symétrique pour la matrice d'interactions utilisant

$$C = F = \{\Theta \in \mathbb{R}^{\bar{M} \times \bar{M}} \mid \Theta = \Theta^\top\}. \quad (2.111)$$

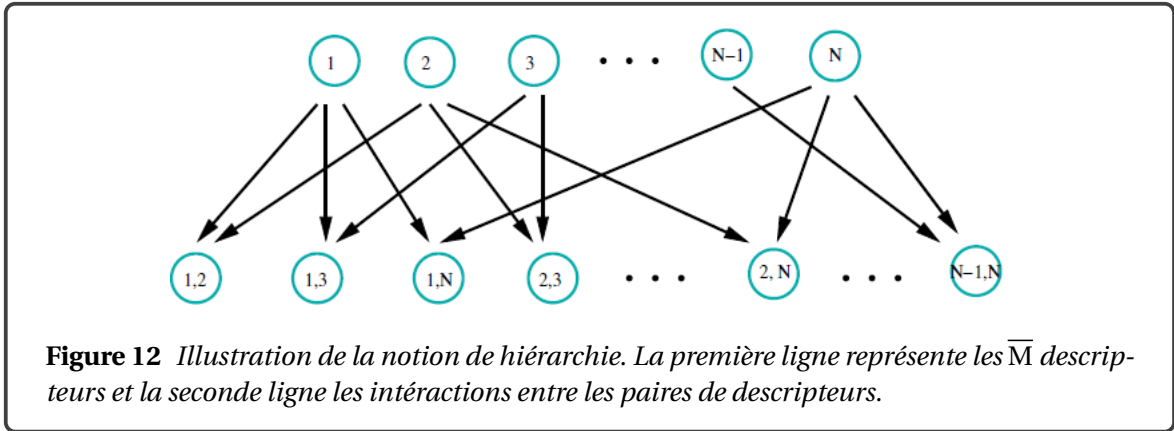


Figure 12 Illustration de la notion de hiérarchie. La première ligne représente les $\bar{M}$ descripteurs et la seconde ligne les interactions entre les paires de descripteurs.

Différentes formes de pénalisation (2.110) rencontrées dans la littérature La plupart des pénalisations de l'état-de-l'art peuvent être interprétées comme un cas spécifique du problème 2.109 avec $\boldsymbol{\varphi}$ défini par (2.110), par exemple :

- [JMOB11] : $g(\cdot) = (\cdot)^\top$, $q = \{2, +\infty\}$,
- [BTT13; HWS14] : $g(\cdot) = \|\cdot\|_1$ and $q = +\infty$,
- [SJ16] : $g(\cdot) = (\cdot)^\top$ and $q = \{2\}$.

Nous pouvons remarquer que [JMOB11] se concentre sur une hiérarchie faible en utilisant $C = \mathbb{R}^{\bar{M} \times \bar{M}}$. [BTT13] se rencontre sous le nom de «hierNet», alors que [HWS14] est appelée "FAMILY" et [SJ16] est nommée "Type-A GRESH". La formulation du lasso latent de groupe superposé de [BTT13] est connue sous le nom de "glinternet" [LH15].

La résolution du problème (2.109) avec $\boldsymbol{\varphi}$ définie par (2.110) est difficile, car elle implique des fonctions convexes non lisses et des contraintes de symétrie. Dans [JMOB11], les auteurs proposent un schéma algorithmique reposant sur FISTA et sont limités au contexte de la hiérarchie

faible. Dans un contexte de hiérarchie forte, leur schéma nécessite des sous-itérations basées sur l'algorithme de Dykstra qui ralentissent considérablement la convergence [CPP09]. Dans [BTT13], les auteurs reformulent le problème (2.109) lorsque $g(\cdot) = \|\cdot\|_1$ sous une formulation épigraphique qui représente le problème comme un ensemble de contraintes dures. Cette reformulation aide à l'interprétation ainsi qu'à la conception d'un schéma algorithmique reposant sur ADMM. Dans la suite nous généralisons cette reformulation épigraphique.

Reformulation épigraphique Notre contribution fut de fournir un cadre unifié et une stratégie algorithmique efficace basée sur une reformulation épigraphique et l'usage d'algorithmes proximaux primaux-duaux pour résoudre une formulation relaxée de (2.110) quand $C = \mathbb{R}^{\bar{M} \times \bar{M}}$ ou $C = F$ (cf (2.111)), $g(\cdot) = \|\cdot\|_r$ avec $r = \{1, +\infty\}$ et $q = +\infty$.

Le problème (2.109) avec $\boldsymbol{\varphi}$ définie par (2.110) à l'aide de contraintes épigraphiques se formule :

$$\begin{aligned} \underset{(x^+, x^-, \Theta) \in O \times O \times C}{\text{minimiser}} \quad & \frac{1}{2} \sum_{s=1}^S \left(u_s - \mathcal{F}(z_s)^\top (x^+ - x^-) - \mathcal{F}(z_s)^\top \Theta \mathcal{F}(z_s) \right)^2 \\ & + \frac{\lambda}{2} \|\Theta\|_1 + \lambda 1^\top (x^+ + x^-) + \sum_{i=1}^{\bar{M}} \iota_{E_r}(x_i^+, x_i^-, \Theta_{i,\cdot}) \end{aligned} \quad (2.112)$$

où O désigne l'orthant positif, $x = x^+ - x^-$ et $E_r = \{(\omega^+, \omega^-, \mathbf{u}) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{\bar{M}} \mid \|\mathbf{u}\|_r \leq \omega^+ + \omega^-\}$ la contrainte épigraphique avec $r > 1$. Nous avons pu obtenir une forme explicite de la projection sur E_r lorsque $r = \{1, +\infty\}$. La projection sur E_∞ est obtenue à partir d'arguments similaires à ceux permettant de dériver le résultat de la Proposition 7. Quant à la projection sur E_1 , elle se déduit de la projection sur $E_1^+ = \{(\omega^+, \omega^-, \mathbf{u}) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{\bar{M}} \mid \|\mathbf{u}\|_1 = \omega^+ + \omega^-, \mathbf{u} \geq 0\}$ proposée dans [DS09][Lemma 3]. L'algorithme de Condat-Vũ peut alors être considéré pour minimiser (2.112). Une suite $((x^+)^{[k]}, (x^-)^{[k]}, \Theta^{[k]})_{k \in \mathbb{N}}$ est alors construite avec garantie de convergence vers un minimiseur de (2.112).

Résultats numériques Je présente ici des résultats extraits de [JPJ⁺19] dans lesquels nous comparons les résultats obtenus par combinaison de l'algorithme Condat-Vu et formulation épigraphique dans le cas *Weak-PD- ℓ_1* ($r = 1$ et $C = \mathbb{R}^{\bar{M} \times \bar{M}}$), *Strong-PD- ℓ_1* ($r = 1$ et $C = F$), *Weak-PD- ℓ_∞* ($r = +\infty$ et $C = \mathbb{R}^{\bar{M} \times \bar{M}}$), et *Strong-PD- ℓ_∞* ($r = +\infty$ et $C = F$) avec deux méthodes de l'état de l'art : “hierNet” [BTT13] et le travail de Jenatton et ses collaborateurs [JMOB11]. Les méthodes sont comparées dans un contexte de régression, nous présentons dans ce manuscrit les résultats sur des données simulées mais des résultats sur données réelles sont également disponibles dans [JPJ⁺19].

Les données synthétiques sont générées selon la procédure décrite dans [HWS14]. Il est composé de $\bar{M}$ caractéristiques principales et de $\bar{M}(\bar{M} - 1)/2$ interactions (dûes à la symétrie). La *vérité terrain* $(\bar{x}, \bar{\Theta}) \in \mathbb{R}^{\bar{M}} \times \mathbb{R}^{\bar{M} \times \bar{M}}$ est générée afin de vérifier une hiérarchie forte.

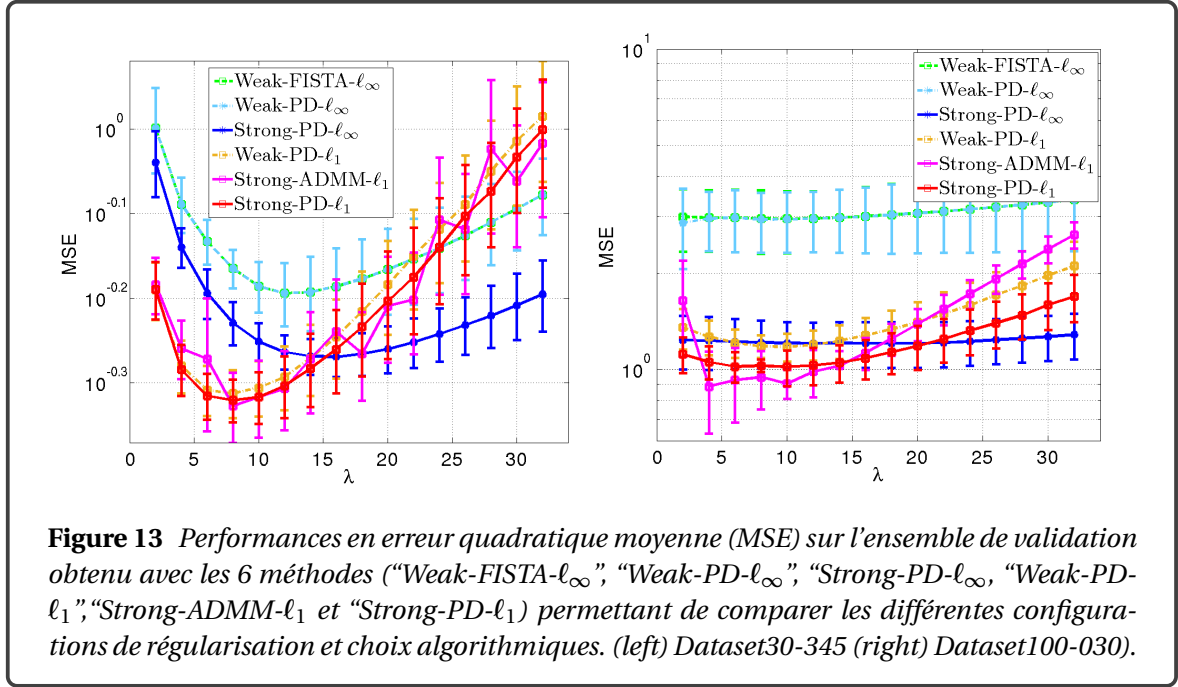
Les performances algorithmiques sont évaluées sur deux jeux de données simulés. Le premier est composé de caractéristiques $N = 30$, les 10 premières caractéristiques principales sont non nulles et le rapport d'interaction est $\rho = 3.45\%$ (Dataset30-345). Le deuxième jeu de données est de taille $N = 100$, les 30 premières entités ne sont pas des zéros et le rapport d'interaction est $\rho = 0.30\%$ (Dataset100-030).⁹

Les performances moyennes en EQM pour différents modèles d'interaction avec différentes valeurs de λ sur l'ensemble de validation sont présentées dans la Figure 13. On observe que *Weak-PD- ℓ_∞* et *Weak-FISTA- ℓ_∞* ont des performances quasiment identiques pour différentes valeurs de λ sur les deux jeux de données de simulation. Pour le Dataset30-345, *Strong-ADMM- ℓ_1* a des performances similaires à *Strong-PD- ℓ_1* . Pour le Dataset100-030, il existe des écarts pour différents

9. Le rapport d'interaction est calculé comme le nombre d'interactions non nul divisé par le nombre d'interactions possibles dans $\bar{\Theta}$ (c'est-à-dire $N(N - 1)/2$).

λ entre $Strong-ADMM-\ell_1$ et $Strong-PD-\ell_1$. La méthode basée sur ADMM devient sensible pour de grandes valeurs de λ . Le bon comportement de la contrainte de hiérarchie forte apparaît clairement pour les deux ensembles de données. En effet, nous rappelons que les données ont été créées avec une structure hiérarchique forte et nous pouvons clairement observer que pour les deux jeux de données les performances avec la contrainte de hiérarchie forte conduisaient toujours aux meilleurs résultats que ce soit pour $r = 1$ ou $r = +\infty$. Dans l'ensemble de nos simulations, la régularisation avec $r = 1$ conduit toujours à de meilleures performances comparé à $r = +\infty$.

Formulation similaire pour la classification Des travaux similaires ont été menés dans le contexte de la classification multiclasse [JPJ⁺19].



2.2.5.3 Conclusions

Dans cette section nous avons proposé de considérer l'algorithme Condat-Vũ combiné à de nouvelles formes explicites de projections épigraphiques afin de construire, d'une part, des algorithmes plus rapides que les algorithmes de l'état-de-l'art (basés sur FBS ou FISTA) dans le contexte de classification multiclassées avec a priori de parcimonie et, d'autre part, permettant une plus grande flexibilité et robustesse des solutions proposées dans le contexte de la régression ou classification multiclassée avec a priori de parcimonie sur les descripteurs et prise en compte d'interactions quadratiques entre les descripteurs.

A l'ère de l'apprentissage profond, ce genre d'approches peut paraître obsolète cependant elles peuvent être un outil d'analyse préliminaire ou complémentaire à une étude reposant sur des stratégies d'apprentissage profond.

2.3 Algorithmes proximaux et grand volume de données

2.3.1 Motivations et enjeux

Dans de nombreuses applications sur données réelles sur lesquelles j'ai eu l'occasion de travailler, allant de la restauration d'images satellitaires, à l'analyse des écoulements multiphasiques ou encore la prédiction de matrice de transfert entre deux élections, le volume de données à traiter est de l'ordre du milliard de pixels dans une image (ou de noeuds d'un graphe). De plus, lors de la mise en œuvre d'une approche variationnelle de la forme (2.1), une recherche exhaustive du paramètre optimal λ est faite et nécessite de faire tourner plusieurs fois un même algorithme pour différentes valeurs de λ . Il est donc indispensable d'avoir recours à des algorithmes rapides.

Une première piste d'accélération a été évoquée dans la section précédente à travers la formulation sous contrainte et par décomposition épigraphique. Dans cette section nous considérons des solutions alternatives dans le contexte régularisé (2.1).

Les développements autour des algorithmes proximaux introduits dans la Section 2.1.4 ont permis d'envisager la minimisation de critères convexes, potentiellement non-lisses, de plus en plus complexes. Cette flexibilité provient du concept d'éclatement (e.g. *splitting*) permettant d'activer séparément les fonctions à travers un pas de gradient ou un pas proximal. En effet, si l'on considère le problème général suivant :

$$\underset{x \in \mathcal{H}}{\text{minimiser}} \quad f_1(x) + f_2(x) + f_3(x) \quad (2.113)$$

où les fonctions f_1 , f_2 et f_3 appartiennent à $\Gamma_0(\mathcal{H})$ et sont potentiellement non-lisses, sachant que l'opérateur proximal d'une somme de fonctions a rarement une forme explicite, il est préférable d'activer les opérateurs proximaux séparément plutôt que d'avoir recours à des sous-itérations. La flexibilité des algorithmes de *splitting* permet d'envisager une telle stratégie. Cependant, l'intuition vise à penser que plus le problème est éclaté, plus la convergence sera lente. La section 2.3.2.1, vise à donner quelques éléments de réponses à cette intuition. Dans un premier temps, nous rappellerons les principaux résultats de la littérature dans lesquels

$$\text{prox}_{f_1+f_2} = \text{prox}_{f_2} \circ \text{prox}_{f_1}, \quad (2.114)$$

où $\circ$ modélise l'opération de composition. Puis nous illustrerons l'intérêt d'éviter l'éclatement lorsque cela est possible sur un exemple d'estimation de disparité et un exemple de détection d'interfaces. Dans chacun de ces problèmes, de nouvelles formes explicites d'opérateurs proximaux d'une somme de deux fonctions seront données.

Une source d'accélération importante provient de la formulation sous forme d'un critère fortement convexe. Cependant peu d'algorithmes proximaux permettent d'exploiter une telle propriété. Une des raisons pourrait être que le développement des algorithmes proximaux fut fortement lié aux développements en restauration d'images, contexte dans lequel la forte convexité du critère est rarement envisageable. Notons qu'à ma connaissance seul l'algorithme Chambolle-Pock peut intégrer cette hypothèse. La section 2.3.3 est dédiée à la question de la segmentation de texture. Dans ce contexte, un critère fortement convexe est proposé. Dans cette section, nous commencerons par discuter l'impact de la forte convexité par rapport à des approches par blocs dans le cas spécifique de la résolution du modèle ROF. Nous présenterons ensuite comment construire un critère fortement convexe dans le cadre de la segmentation de texture par des attributs multi-échelles. Enfin nous mettrons en avant sur cet exemple l'intérêt de la prise en compte de la forte convexité dans les algorithmes proximaux.

Les résultats numériques effectués en segmentation de texture mais également en détection d'interfaces nous amèneront à nous poser la question suivante : faut-il préférer faire un pas proximal ou un pas de gradient ? Il s'agit d'une question bien connue mais pour laquelle nous avons quand même proposé dans [BAP20] quelques pistes de réflexions supplémentaires permettant de comparer théoriquement et numériquement les algorithmes de premier ordre en utilisant et développant de nouveaux résultats fins sur les taux de convergence (cf. section 2.3.4).

2.3.2 Splitting or not splitting

2.3.2.1 Opérateur proximal d'une somme de deux fonctions

Plusieurs travaux ont été consacrés au calcul de l'opérateur proximal d'une somme de deux fonctions. Par exemple, nous pouvons nous référer à [Cha04; BC08a; CDuCVu09] pour des solutions itératives et à [CPP09; MAB12] pour les garanties de convergence lorsque l'opérateur proximal est calculé de façon inexact par exemple lorsque des sous-itérations sont nécessaires.

Nous rappelons maintenant les principaux résultats de l'état-de-l'art pour lesquels (2.114) est vérifiée. Par exemple, si C est réduit à un singleton ou un sous-ensemble non vide de $\mathbb{R}$, la propriété (2.114) est vérifiée, avec $f_2 = \iota_C$ et tout $f_1 \in \Gamma_0(\mathbb{R})$. Ce résultat se généralise lorsque C et f_1 sont séparables dans la même base de $\mathcal{H}$ [CPP09]. Lorsque $\mathcal{H} = \mathbb{R}$, avec f_1 la fonction de support σ_C d'un sous-ensemble fermé, convexe, non vide C de $\mathcal{H}$ [BC17], et f_2 différentiable en 0 avec $f_2'(0) = 0$, l'égalité (2.114) est également valable [CP07b, Proposition 3.6]. Une condition suffisante plus générale est $\partial f_1(x) \subset \partial f_1(\text{prox}_{f_2}(x))$, pour tout $x \in \mathcal{H}$ [Yu13].

2.3.2.2 Nouvelle forme explicite et son intérêt pour l'estimation de disparité et la segmentation.

Dans cette section, nous considérons $x = (x_n)_{n \in \Omega} \in \mathbb{R}^N$ comme un graphe avec des valeurs x_n sur ses sommets et soit $\mathbb{E}$ un sous-ensemble de Ω^2 , constitué de paires d'indices, qui peuvent être vues comme les arêtes du graphe.

Dans [PC17], nous considérons des hypothèses sur f_1 et f_2 différentes de celles mentionnées dans la section 2.3.2.1 et permettant d'obtenir une expression de la forme (2.114). Ce résultat est rappelé dans la Proposition 9.

Proposition 9 [PC17] Soient f_1 et f_2 deux fonctions de $\Gamma_0(\mathbb{R}^N)$, qui pénalisent les arêtes et les sommets d'un graphe, telles que :

(i) f_2 est séparable avec

$$(\forall x = (x_n)_{n \in \Omega} \in \mathbb{R}^N) \quad f_2(x) = \sum_{n \in \Omega} h(x_n), \quad (2.115)$$

pour une fonction $h \in \Gamma_0(\mathbb{R})$.

(ii) f_1 est de la forme :

$$(\forall x = (x_n)_{n \in \Omega} \in \mathbb{R}^N) \quad f_1(x) = \sum_{(m,m') \in \mathbb{E}} \sigma_{C_{m,m'}}(x_{m'} - x_m), \quad (2.116)$$

où $\sigma_{C_{m,m'}} : t \in \mathbb{R} \mapsto \sup \{tp, p \in C_{m,m'}\}$ est la fonction support d'un intervalle réel fermé $C_{m,m'}$, tel que $\inf C_{m,m'} = a_{m,m'}$ et $\sup C_{m,m'} = b_{m,m'}$, pour $a_{m,m'} \in \mathbb{R} \cup \{-\infty\}$ et $b_{m,m'} \in \mathbb{R} \cup \{+\infty\}$, avec $a_{m,m'} \leq b_{m,m'}$. Alors nous avons

$$(\forall t \in \mathbb{R}) \quad \sigma_{C_{m,m'}}(t) = \begin{cases} a_{m,m'} t & \text{si } t < 0, \\ 0 & \text{si } t = 0, \\ b_{m,m'} t & \text{si } t > 0, \end{cases} \quad (2.117)$$

avec la convention $(-\infty)t = +\infty$ et $(+\infty)t = +\infty$.

Alors,

$$\text{prox}_{f_1+f_2} = \text{prox}_{f_2} \circ \text{prox}_{f_1}. \quad (2.118)$$

Voici quelques exemples où cette proposition possède un intérêt pratique :

- Dans le cas 1D, où $\Omega = \{1, \dots, N\}$ et $\mathbb{E} = \{(1, 2), (2, 3), \dots, (N-1, N)\}$, la pénalisation par variation totale pondérée de x coïncide avec notre formulation lorsque $b_{n,n+1} = -a_{n,n+1} = \omega_n \geq 0$ tel

que :

$$(\forall x \in \mathbb{R}^N) \quad f_1(x) = \sum_{n=1}^{N-1} \omega_n |x_{n+1} - x_n|. \quad (2.119)$$

Il existe des algorithmes très efficaces pour calculer l'opérateur proximal de cette fonction spécifique [DK01; Con13a]. Son extension aux images 2D est connue sous le nom de variation totale anisotrope et son opérateur proximal peut être calculé efficacement en utilisant un algorithme de coupure de graphes [CD12]. Le Lasso fusionné (*Fused Lasso*) [TSR⁺05] correspond au cas où f_1 est définie par (2.119) et f_2 est une norme ℓ_1 ; c'est-à-dire dans (2.115), $h = \lambda |\cdot|$, avec $\lambda > 0$. L'opérateur proximal du Lasso fusionné est obtenu en appliquant l'opérateur proximal de la variation totale, suivi d'un seuillage doux, qui est l'opérateur proximal de la norme ℓ_1 . Ce cas particulier de la Proposition 9 était déjà connu [FHHT07].

- Dans le cas 1D, avec $\Omega = \{1, \dots, N\}$ et $\mathbb{E} = \{(1, 2), (2, 3), \dots, (N-1, N)\}$, la régression isotonique correspond à

$$f_1(x) = \iota_C(x) \quad \text{avec} \quad C = \{x \in \mathbb{R}^N : x_1 \geq \dots \geq x_N\}. \quad (2.120)$$

Il s'agit également d'un cas particulier du contexte d'étude de la Proposition 9 avec $a_{n,n+1} = 0$ et $b_{n,n+1} = +\infty$. L'algorithme PAVA (Pool Adjacent Violators Algorithm) [ABE⁺95] permet de calculer l'opérateur proximal de f_1 avec une complexité $O(N)$. De plus, quand $h = \iota_{[0,1]}$, le résultat de la Proposition 9 était déjà connu dans [GW84]. L'extension de ce résultat dans un contexte où la structure des données n'est plus une chaîne mais un arbre est présentée dans [CS15].

Dans [PC17], nous avons proposé d'évaluer l'intérêt de la Proposition 9 dans un contexte d'estimation de disparité en stéréovision et de segmentation d'images. Ces deux questions prennent la forme d'un problème de minimisation décrit par (2.11) et que nous rappelons ici :

$$\hat{\Theta} \in \underset{\Theta}{\text{Argmin}} \sum_{q=1}^{Q-1} \langle \alpha^{(q)}, \theta^{(q)} \rangle + \lambda \sum_{q=1}^Q \|DT_q \Theta\|_{2,1} + \iota_{[0,1]^{(Q-1) \times N}}(\Theta) + \iota_C(\Theta), \quad (2.121)$$

où

$$\begin{cases} \alpha^{(q)} = \eta^{(q+1)} - \eta^{(q)}, \\ T_q: \mathbb{R}^{(Q+1) \times N} \rightarrow \mathbb{R}^N: \Theta \mapsto \theta^{(q-1)} - \theta^{(q)}, \\ C = \{\Theta \in \mathbb{R}^{(Q-1) \times N} : \theta^{(1)} \geq \dots \geq \theta^{(Q-1)}\}. \end{cases}$$

avec $\theta^{(Q)} \equiv 0$ et $\theta^{(0)} \equiv 1$ et où $(\eta^{(q)})_{1 \leq q \leq Q}$ sont choisis différemment pour un problème d'estimation de disparité ou de segmentation.

Estimation de la disparité Dans le cas de l'estimation de la carte de disparité en stéréovision [MPP09], les données sont une paire d'images $z^L \in \mathbb{R}^{N \times M}$ et $z^R \in \mathbb{R}^{N \times M}$ qui correspondent aux vues de la même scène par les yeux gauche et droit. $M = 3$ dans le cas des images couleurs. Le principe utilisé pour l'estimation de la profondeur est que les valeurs de pixels $z_{(n_1, n_2)}^L$ et $z_{(n_1, n'_2)}^R$ ont tendance à être similaires. La profondeur de l'objet imagé étant inversement proportionnelle à $n_2 - n'_2$. Par conséquent, nous fixons l'ensemble des valeurs possibles de $n_2 - n'_2$ comme la plage d'entiers $q = \{1, \dots, Q\}$, conduisant à :

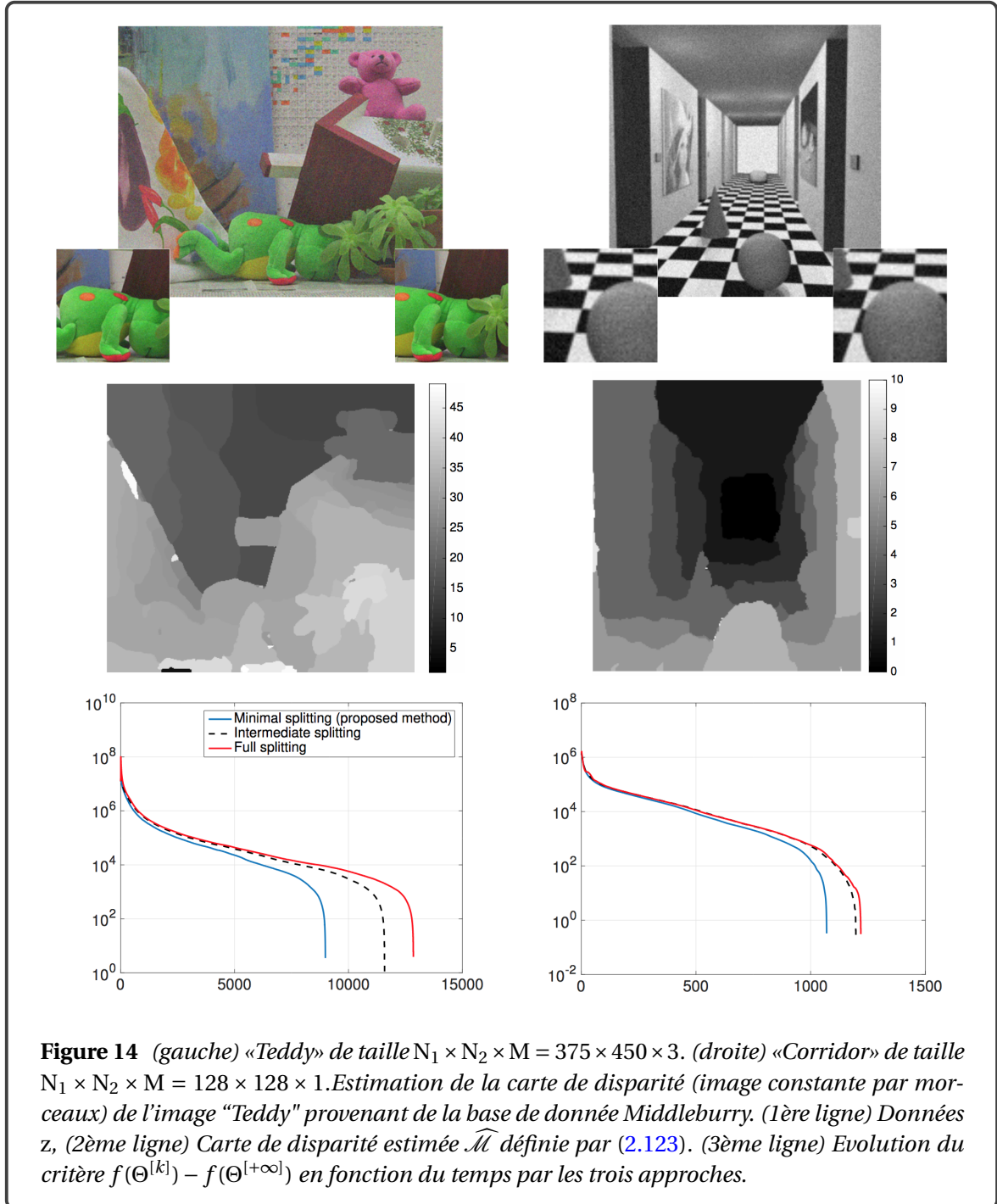
$$(\forall n = (n_1, n_2) \in \Omega) \quad \eta_n^{(q)} = \sum_{m=1}^M |z_{(n_1, n_2), m}^L - z_{(n_1, n_2 - q + 1), m}^R|^r, \quad (2.122)$$

où le choix usuel de r est 1 ou 2. Enfin, la carte de disparité estimée est obtenue à partir de $\hat{\Theta} = (\hat{\theta}^{(q)})_{1 \leq q \leq Q}$ avec :

$$\widehat{\mathcal{M}} = \sum_{q=1}^Q (q-1) (\hat{\theta}^{(q-1)} - \hat{\theta}^{(q)}). \quad (2.123)$$

Pour la minimisation de (2.11), nous proposons dans [PC17] trois solutions algorithmes, reposant toutes sur le même algorithme à savoir l'algorithme de Condat-Vũ mais pour différents niveaux de spitting :

- *Full splitting* – La contrainte C est décomposée en deux contraintes C_1 et C_2 qui sont définies afin d'éviter le chevauchement des variables : $C_1 = \{\Theta \in \mathbb{R}^{(Q-1) \times N} : \theta^{(1)} \geq \theta^{(2)}, \theta^{(3)} \geq \theta^{(4)}, \dots\}$ et $C_2 = \{\Theta \in \mathbb{R}^{(Q-1) \times N} : \theta^{(2)} \geq \theta^{(3)}, \theta^{(4)} \geq \theta^{(5)}, \dots\}$. Notons que $C = C_1 \cap C_2$ et que $\iota_C = \iota_{C_1} + \iota_{C_2}$. P_{C_1} et P_{C_2} ont alors des formes explicites.
- *Intermediate splitting* – La projection sur C est calculée par l'algorithme PAVA mentionné précédemment. P_C et $P_{[0,1]^{(Q-1) \times N}}$ sont calculées séparément.
- *Minimal splitting* – Nous considérons le résultat de la Proposition 9 avec $f_2 = \iota_{[0,1]^{(Q-1) \times N}}$ et $f_1 = \iota_C$ nous permettant de calculer $P_{[0,1]^{(Q-1) \times N}} \circ P_C$.



Dans [PC17], nous avons appliqué la procédure d'estimation décrite précédemment à deux paires d'images «Teddy» et «Corridor» de la base de données classique de 2003 Middlebury. Les images originales ont été corrompues par un bruit blanc additif Gaussien d'écart type 10. Les images gauche et droite sont montrées dans la Figure 14 (1ère ligne). Nous fixons $Q = 50$ (resp. $Q =$

20), $\lambda = 150$ (resp. $\lambda = 50$), car ces valeurs conduisent aux meilleurs résultats. Notez que la valeur de Q peut être déduite de la traduction des éléments de premier plan sur les données bruitées. Les résultats sont présentés Figure 14 (2ème ligne) ainsi que l'évolution du critère $\Psi(\Theta^{[k]}) - \Psi(\Theta^{[+\infty]})$ en fonction du temps. La comparaison en temps de calcul des trois différentes stratégies illustre que plus le problème est découplé, plus la procédure d'estimation est lente. Nous avons donc tout intérêt quand cela est possible de calculer l'opérateur proximal d'une somme de fonctions plutôt que de les gérer séparément.

Segmentation d'images Des conclusions similaires ont été obtenues dans [CP15] dans le contexte de segmentation d'images où $\eta^{(q)} = \frac{1}{2\sigma_q^2}(z - c_q)^2$ où c_q et σ_q^2 désignent la moyenne et la variance au sein de la région $\Omega^{(q)} = \theta^{(q)} - \theta^{(q)}$.

2.3.2.3 Nouvelle forme explicite et son intérêt pour la détection d'interfaces par le modèle de Mumford-Shah discret

Cette section, associée aux résultats présentés dans [FPC19; FPC18], présente une nouvelle forme explicite de l'opérateur proximal d'une somme de deux fonctions et nous illustrons son intérêt dans le contexte de la minimisation de la fonctionnelle de Mumford-Shah discrète. Notons qu'une première contribution de [FPC19] fut de proposer une reformulation discrète des fonctionnelles de Mumford-Shah (cf. (2.8)) et de Geman et Geman (cf. (2.9)). Cette fonctionnelle sera désignée par D-MS. Une seconde contribution fut de proposer un nouveau schéma proximal alterné. Ce dernier repose sur le calcul de l'opérateur proximal d'une somme de deux fonctions qui possède une forme explicite dans ce cadre d'étude et qui conduit à des gains significatifs en temps de calculs.

Modèle D-MS La généralisation de la formulation discrète de la fonctionnelle de Mumford-Shah que nous avons proposée dans [FPC19] prend la forme :

$$\underset{u \in \mathbb{R}^{NM}, e \in \mathbb{R}^{|E|}}{\text{minimiser}} \underbrace{\Psi(u; z) + \beta \sum_m \|(1 - e) \odot Du_m\|_2^2}_{\varphi_1(u, e)} + \chi \varphi_2(e). \quad (2.124)$$

où $z \in \mathbb{R}^{NM}$ sont des données dégradées par le modèle (2.2) et où D est défini par (2.4). On considère un terme d'attache aux données générique $\Psi(\cdot, z) : \mathbb{R}^{NM} \rightarrow]-\infty, +\infty]$ pouvant par exemple modéliser une distance euclidienne ou une divergence de Kullback-Leibler (cf. Section 2.1.1) et $\varphi_2 : \mathbb{R}^{|E|} \rightarrow]-\infty, +\infty]$ est un terme de régularisation qui favorise la parcimonie et agit comme un terme de longueur de contours, toutes deux sont supposées propres et semi-continues inférieurement. $\varphi_1 : \mathbb{R}^{NM} \times \mathbb{R}^{|E|} \rightarrow \mathbb{R}$ est un terme de couplage qui pénalise les fortes variations excepté au niveau des contours.

Cette fonctionnelle permet :

- de donner plus de flexibilité dans le terme d'attache aux données, permettant de gérer des termes d'attache aux données n'étant pas limités à la norme euclidienne comme dans la fonctionnelle originale de Mumford-Shah et permettant de gérer des opérateurs linéaires.
- de considérer un large choix de termes de régularisation φ_2 . L'un des choix le plus rencontré dans la littérature fut proposé par Ambrosio et Tortorelli (AT) [AT90; AT92] et se formule $\varphi_2 = \epsilon \|\tilde{D}e\|_2^2 + \frac{1}{4\epsilon} \|e\|_2^2$, avec $\epsilon > 0$ et $\tilde{D}$ étant un opérateur de différence sur les arêtes. Une telle pénalisation de contour Γ -converge vers la fonctionnelle de Mumford-Shah quand ϵ tend vers 0 [FLT16]. Notons que de grandes valeurs de ϵ conduisent à des contours épais mais aident à détecter l'ensemble des discontinuités et lorsque ϵ tend vers 0, la pénalisation de $\|e\|_2^2$ augmente et permet ainsi à e de devenir de plus en plus parcimonieux, et donc à des contours de plus en plus fins. Le choix $\varphi_2 = \|\cdot\|_1$ permet de se rapprocher de (2.9) mais possède de moins bonnes propriétés théoriques que la pénalisation AT.

Le choix du terme de couplage dans (2.124) est motivé par (2.8) et [FLT16]. Cependant, dans [Sha96], Shah propose de remplacer la norme ℓ_2 par un terme de couplage impliquant la norme ℓ_1 tel que $\boldsymbol{\varphi}_1(\mathbf{u}, \mathbf{e}) = \sum_m \|(1 - \mathbf{e}) \odot (1 - \mathbf{e}) \odot \mathbf{D}\mathbf{u}_m\|_1$, combiné avec une régularisation AT. Alicandro et al. [ABS99] ont prouvé la Γ -convergence de cette fonctionnelle particulière vers une variante de la fonctionnelle de Mumford-Shah. Des expériences montrent que ce terme de couplage de type ROF est plus robuste aux gradients d'image, mais élimine le contenu haute fréquence. Dans le critère de Geman et Geman introduit (2.9), le terme de couplage prend la forme $\sum_i (\mathbf{D}\mathbf{u})_i^2 (1 - e_i)$ qui nécessite l'ajout d'une contrainte sur la dynamique de $\mathbf{e}$ afin d'assurer que e_i appartienne à $[0, 1]$. Plus récemment, Li et al. [LMZ17] ont suggéré de choisir $e_i = \{\mathbf{e}_{i,j}\}_{j \in \mathcal{N}_i}$, où $\mathcal{N}_i$ modélise un effet de voisinage autour du pixel i . La fonction de régularisation est donc une fonctionnelle ROF non locale (NL-ROF) de la forme $\boldsymbol{\varphi}_1(\mathbf{u}, \mathbf{e}) = \sum_m \sum_{i=1}^N \sqrt{\sum_{j \in \mathcal{N}_i} \mathbf{e}_{i,j} (\mathbf{u}_{m,i} - \mathbf{u}_{m,j})^2}$. Dans cette approche, les contours ne sont pas obtenus à partir de $\mathbf{e}$ mais par un seuillage de $\mathbf{u}$, conduisant à une estimation moins précise.

On peut remarquer que, lorsqu'il s'agit d'images multivariées, les contours peuvent être définis soit comme des arêtes similaires à travers toutes les composantes, soit comme des arêtes distinctes, conduisant à des contours qui peuvent être différents pour toutes les composantes. Afin de faciliter la compréhension et la lecture, nous avons choisi de formuler le Problème (2.124) dans le contexte d'arêtes similaires mais l'analyse qui suit peut être dérivée de manière similaire pour des arêtes distinctes en considérant $\mathbf{e} \in \mathbb{R}^{|\mathcal{E}| \times M}$. Quand $M = 1$, les deux formalismes sont équivalents.

Algorithme Afin de résoudre le Problème (2.124), nous considérons deux algorithmes proximaux alternés. Le premier est l'algorithme PALM (*Proximal Alternating Linearized Minimization*) [ABRS10], particularisé dans l'Algorithme 2 pour la résolution spécifique du Problème (2.124). Le second est l'algorithme SL-PAM (*Semi-Linearized Proximal Alternating Minimization*), que nous avons proposé dans [FPC19], qui correspond à une version partiellement linéarisée de PALM mais qui requiert le calcul de l'opérateur proximal de $\boldsymbol{\varphi}_1(\mathbf{u}, \cdot) + \boldsymbol{\varphi}_2$. Cet algorithme spécifié à la résolution du Problème (2.124) est présenté dans l'Algorithme 3.

Algorithm 2 PALM pour résoudre D-MS (2.124)

Soit $\mathbf{u}^{[0]} \in \mathbb{R}^{\text{NM}}$ et $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathcal{E}|}$.
 Pour $k \in \mathbb{N}$
 Soit $\gamma > 1$ et $c_k = \gamma \beta_{\nabla \boldsymbol{\varphi}_1(\cdot, \mathbf{e}^{[k]})}$.
 $\mathbf{u}^{[k+1]} \in \text{prox}_{\frac{1}{c_k} \boldsymbol{\Psi}(\cdot, \mathbf{z})} \left(\mathbf{u}^{[k]} - \frac{1}{c_k} \nabla_{\mathbf{u}} \boldsymbol{\varphi}_1(\mathbf{u}^{[k]}, \mathbf{e}^{[k]}) \right)$
 Soit $\delta > 1$ et $d_k = \delta \beta_{\nabla \boldsymbol{\varphi}_1(\mathbf{u}^{[k+1]}, \cdot)}$.
 $\mathbf{e}^{[k+1]} \in \text{prox}_{\frac{1}{d_k} \chi \boldsymbol{\varphi}_2} \left(\mathbf{e}^{[k]} - \frac{1}{d_k} \nabla_{\mathbf{e}} \boldsymbol{\varphi}_1(\mathbf{u}^{[k+1]}, \mathbf{e}^{[k]}) \right)$

Algorithm 3 SL-PAM pour résoudre D-MS (2.124)

Soit $\mathbf{u}^{[0]} \in \mathbb{R}^{\text{NM}}$ et $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathcal{E}|}$.
 Pour $k \in \mathbb{N}$
 Soit $\gamma > 1$ et $c_k = \gamma \beta_{\nabla \boldsymbol{\varphi}_1(\cdot, \mathbf{e}^{[k]})}$.
 $\mathbf{u}^{[k+1]} \in \text{prox}_{\frac{1}{c_k} \boldsymbol{\Psi}(\cdot, \mathbf{z})} \left(\mathbf{u}^{[k]} - \frac{1}{c_k} \nabla_{\mathbf{u}} \boldsymbol{\varphi}_1(\mathbf{u}^{[k]}, \mathbf{e}^{[k]}) \right)$
 Soit $d_k > 0$.
 $\mathbf{e}^{[k+1]} \in \text{prox}_{\frac{1}{d_k} (\beta \boldsymbol{\varphi}_1(\mathbf{u}^{[k+1]}, \cdot) + \chi \boldsymbol{\varphi}_2)} (\mathbf{e}^{[k]})$

La convergence de ces deux algorithmes est assurée par un ensemble d'hypothèses telles que les hypothèses KL (Kurdyka-Łojasiewicz) de $\boldsymbol{\Psi}$ et $\boldsymbol{\varphi}_2$, les fonctions doivent être bornées inférieu-

rement, $\nabla_{\mathbf{u}} \boldsymbol{\varphi}_1$ doit être Lipschitz continue par rapport à chaque variable pour PALM et seulement par rapport à $\mathbf{u}$ pour SL-PAM et enfin les étapes de mises à jour de $\mathbf{u}^{[k+1]}$ et $\mathbf{e}^{[k+1]}$ doivent avoir une forme explicite. Si la suite $\{\mathbf{u}^{[k]}, \mathbf{e}^{[k]}\}_{k \in \mathbb{N}}$ est bornée, alors la suite converge vers un point critique de la fonction que l'on minimise dans (2.124). La convergence de PALM a été démontrée dans [ABRS10] et nous avons proposé celle de SL-PAM dans [FPC19] en suivant un raisonnement similaire. Il est cependant important de noter que les pas de mise à jour de la variable $\mathbf{e}^{[k]}$ dans l'algorithme SL-PAM ne sont pas bornés, permettant ainsi une plus grande flexibilité dans le choix des pas de descente proximale.

Afin d'assurer la convergence, il est important que $\text{prox}_{\frac{1}{d_k}(\beta \boldsymbol{\varphi}_1(\mathbf{u}^{[k+1]}, \cdot) + \chi \boldsymbol{\varphi}_2)}$ possède une forme explicite dans SL-PAM. Dans [FPC19], nous remarquons que sous l'hypothèse d'une pénalisation séparable, une telle forme explicite existe et est donnée par la Proposition 10.

Proposition 10 Soit $D: \mathbb{R}^{|\mathbb{E}| \times \mathbb{N}}$. Pour tout $(\mathbf{u}, \mathbf{e}) \in \mathbb{R}^{\text{NM}} \times \mathbb{R}^{|\mathbb{E}|}$, nous supposons que $\boldsymbol{\varphi}_1(\mathbf{u}, \mathbf{e}) = \|(1 - \mathbf{e}) \odot D\mathbf{u}\|_2^2$ et que $\boldsymbol{\varphi}_2$ est une fonction séparable telle que

$$(\forall \mathbf{e} = (e_i)_{1 \leq i \leq |\mathbb{E}|}) \quad \boldsymbol{\varphi}_2(\mathbf{e}) = \sum_{i=1}^{|\mathbb{E}|} g_i(e_i), \quad (2.125)$$

où $g_i: \mathbb{R}^{|\mathbb{E}|} \rightarrow]-\infty; +\infty]$, et possède un opérateur proximal s'exprimant sous forme explicite. A chaque itération $k \in \mathbb{N}$, avec $d_k > 0$, $\beta > 0$ and $\chi > 0$, l'étape de mise à jour de $\mathbf{e}^{[k+1]}$ dans l'Algorithme 3 s'exprime : pour tout $i \in \{1, \dots, |\mathbb{E}|\}$,

$$\mathbf{e}_i^{[k+1]} \in \text{prox}_{\frac{\chi}{2\beta(D\mathbf{u}^{[k+1]})_i^2 + d_k}} g_i \left(\frac{\beta(D\mathbf{u}^{[k+1]})_i^2 + \frac{d_k \mathbf{e}_i^{[k]}}{2}}{\beta(D\mathbf{u}^{[k+1]})_i^2 + \frac{d_k}{2}} \right). \quad (2.126)$$

Résultats numériques Dans [FPC19], nous avons proposé une série d'expérience pour évaluer les performances de SL-PAM pour la résolution de D-MS (2.124). Nous avons comparé les performances pour différents choix de $\boldsymbol{\varphi}_2$ séparables : norme ℓ_1 , pénalisation ℓ_0 , et fonction de BerHu (Huber inversée) définie par

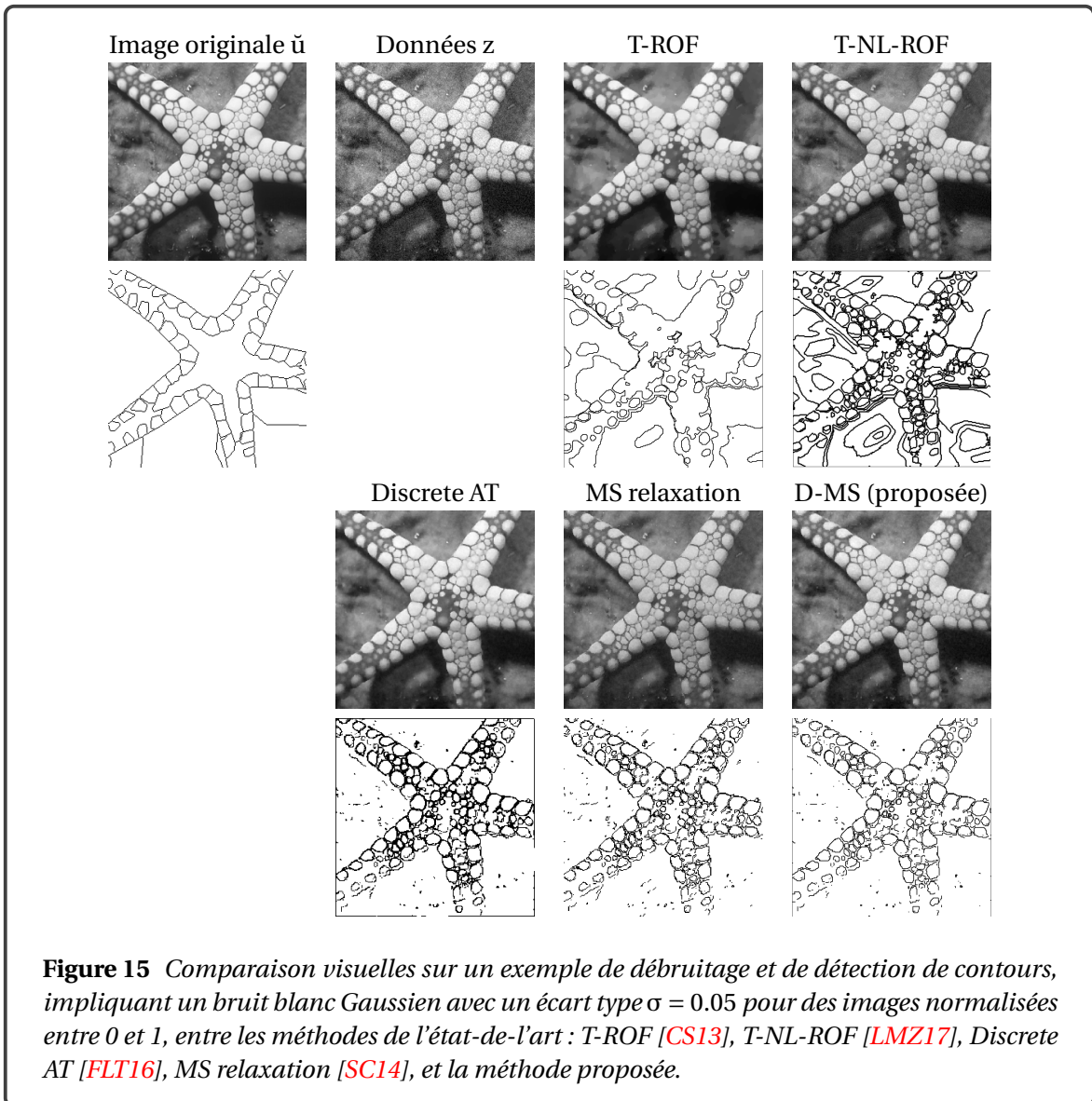
$$\boldsymbol{\varphi}_2(\mathbf{e}) = \sum_{i=1}^{|\mathbb{E}|} \max \left\{ |e_i|, \frac{|e_i|^2}{4\epsilon} \right\}, \quad (2.127)$$

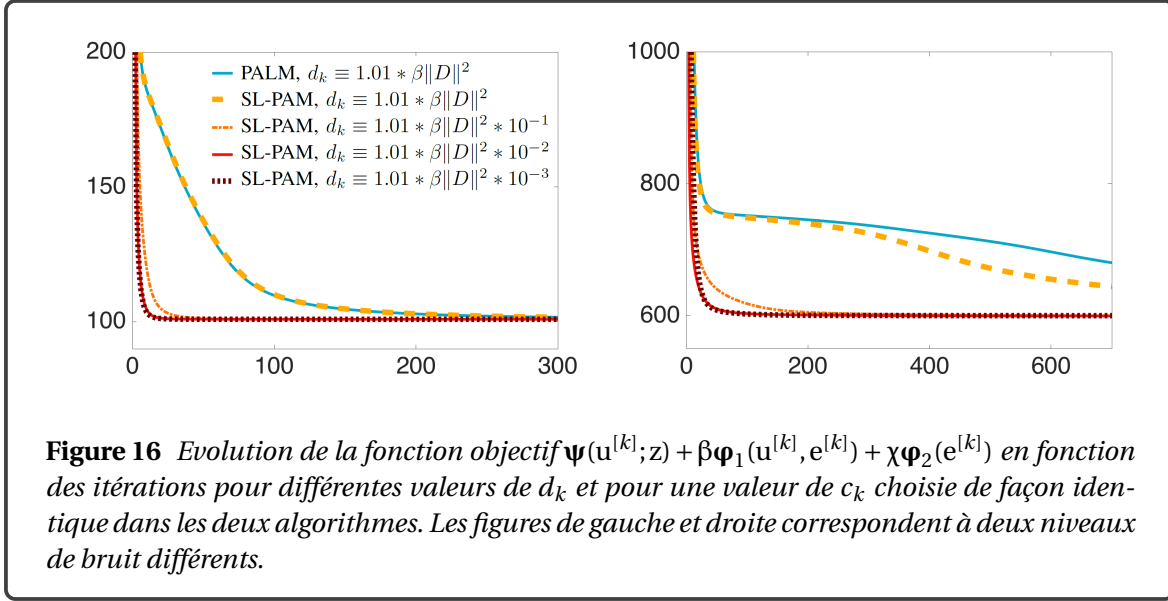
qui sont toutes des fonctions séparables, KL, bornées inférieurement et dont l'opérateur proximal possède une forme explicite. La pénalisation de BerHu a pour objectif de reproduire le comportement quadratique de la pénalisation AT : $\frac{1}{4\epsilon} \|\cdot\|_2^2$ pour de faibles valeurs de ϵ et d'assurer la parcimonie. Son opérateur proximal s'exprime pour tout $\eta \in \mathbb{R}$,

$$\text{prox}_{\tau \max \left\{ |\cdot|, \frac{|\cdot|^2}{4\epsilon} \right\}}(\eta) = \text{sign}(\eta) \max \left\{ 0, \min \left[|\eta| - \tau, \max \left(4\epsilon, \frac{|\eta|}{\frac{\tau}{2\epsilon} + 1} \right) \right] \right\}. \quad (2.128)$$

La première série d'expériences numériques a été menée dans le cas du débruitage gaussien et il apparaît que les pénalisation ℓ_1 et BerHu conduisent aux meilleurs résultats numériques tant en SNR, SSIM et indice de Jaccard mais qu'une pénalisation ℓ_0 conduit à des algorithmes plus rapides. D'autre part, comme illustré Figure 16, SL-PAM est systématiquement plus rapide que PALM. Ce résultat s'explique par une plus grande flexibilité dans le choix du pas sur la mise à jour de la variable de contours. Remarquons que pour un même choix de d_k , le comportement algorithmique est équivalent. Cependant, en bénéficiant de la flexibilité offerte par SL-PAM dans le choix de d_k , on obtient un gain d'au minimum un facteur 100 suivant les configurations considérées. Nous avons également mis en avant la robustesse à l'initialisation et avons montré de bons résultats dans le contexte de débruitage Poisson (attache aux données DKL). Nous présentons Figure 15,

les résultats obtenus avec différentes méthodes de l'état-de-l'art et notre méthode pour une pénalisation de BerHu dans le contexte d'une image dégradée par un bruit additif gaussien. Il apparaît que la stratégie proposée permet d'obtenir des performances au moins équivalentes voire meilleures en termes de reconstruction. Le même constat est fait concernant la qualité d'estimation des interfaces de changement. Enfin, comparé aux méthodes de l'état-de-l'art permettant d'extraire conjointement l'image et ses contours, notre méthode possède le double intérêt d'avoir des garanties de convergences et d'avoir des temps de calculs très faibles. De nombreux résultats numériques complémentaires sont disponibles dans [FPC19; FPC18].





2.3.2.4 Conclusions

Dans cette section, nous avons illustré à travers deux exemples (estimation de disparité et détection de contours), le bénéfice de calculer l'opérateur proximal d'une somme de deux fonctions plutôt que d'activer l'opérateur proximal de chaque fonction indépendamment. Ce résultat pourrait apparaître contradictoire avec les conclusions de la découpe épigraphique présentée dans la section 2.2.3 mais il faut noter qu'une différence majeure réside entre ces deux approches. Dans le premier cas, il est nécessaire de doubler l'espace d'étude (généralement $\bar{N}$) sur lequel on travaille pour gérer une minimisation impliquant une somme de plus de deux fonctions alors que dans le contexte des projections épigraphiques, on augmente seulement la dimension via l'introduction du vecteur ζ qui est de la taille du nombre de blocs $B \ll N$.

La fonctionnelle D-MS combinée à SL-PAM ouvre quant à elle de nombreuses pistes de recherches au cœur de mon ANR JCJC Multisc'In, telles que la prise en charge d'une pénalisation de type AT, non séparable et par conséquent qui ne s'inscrit pas de le cadre présenté ci-dessus, et également des développements théoriques supplémentaires dans le cas où $A \neq \text{Id}$ afin d'avoir des garanties de convergence. Il est également important d'être en mesure de caractériser les minimiseurs de D-MS obtenus avec SL-PAM, par exemple en s'appuyant sur des techniques de lifting [ZH17].

2.3.3 Forte convexité en traitement du signal et des images. Le cas de la segmentation de texture par attributs invariants d'échelles

Depuis 2013 via le projet JCJC GALILEO du GdR ISIS puis le Defi Imag'IN SIROCCO du CNRS, une part importante de mes travaux de recherche fut dédiée à la segmentation de texture. Cette thématique a été renforcée par la thèse de Barbara Pascal (2017-2020).

Segmenter des textures est un problème bien connu en traitement d'images [CRD07]. De nombreux outils particulièrement efficaces ont été proposés dans le contexte de textures géométriques. Pour les textures stochastiques, la segmentation reste un défi car la géométrie est plus difficile à capturer, s'appuyant essentiellement sur les statistiques des caractéristiques de texture. Il est commun d'avoir recours à des approches en deux étapes : extraction de descripteurs puis segmentation. Cependant, les approches d'apprentissage profond ont fortement contribué à revisiter cette approche en deux étapes en combinant l'estimation des descripteurs et la segmentation de ces derniers (cf. e.g. [AW17]). Cependant, leur utilisation reste principalement conçue pour le contexte spécifique de la segmentation supervisée : un ensemble de données d'entraînement (généralement volumineux) doit être disponible, ainsi que de grandes puissances de calcul et de stockage. Souvent, ces bases de données ne sont pas disponibles car les annotations d'experts peuvent être trop coûteuses (en temps et en argent) à mettre en œuvre ou techniquement hors de portée. -On peut également regretter le manque d'interprétabilité voire d'identification des caractéristiques sur lesquelles reposent les décisions : Les médecins par exemple pourraient légitimement rester réticents à fonder le diagnostic sur des caractéristiques non identifiées. Par conséquent, malgré le potentiel de l'apprentissage en profondeur, dans des contextes d'absence de base de données documentée, d'importance de la précision dans les limites estimées et d'un besoin de compréhension, les stratégies de segmentation non supervisées restent d'une importance critique.

Dans [PWAD16; PPA21], nous avons proposé de revisiter les techniques de segmentation de textures stochastiques en proposant une approche non-supervisée combinant estimation des descripteurs et segmentation formulé comme un cas particulier de (2.1). Notre étude se concentre sur des modèles de textures autosimilaires (ou fractales). Ces modèles sont largement utilisés pour décrire des textures du monde réel de natures très différentes allant de la classification de photos historiques aux diagnostics dans les mammographies [ADR99; BPL⁺01; AJW13; GAR⁺14].

Modèle fractal et régularité locale L'autosimilarité est caractérisée par la notion de *régularité locale*, quantité mathématique estimée par régression sur des quantités multiéchelles (e.g. logarithme des coefficients d'ondelettes). Formellement, si on note $z = (z_n)_{1 \leq n \leq N} \in \mathbb{R}^N$ l'image que l'on cherche à analyser. La régularité locale autour de la position n est mesurée par l'exposant dit de Hölder h_n tel que si h_n est proche de 0, l'image est localement très irrégulière et inversement, une grande valeur de h_n conduit à un champ localement lisse. Si on note $\mathcal{L}_{j,n}$ une quantité multi-échelle associée à l'image z définie pour l'échelle $j \in \{1, \dots, J\}$ et la position n , la régularité locale est reliée à cette quantité par la relation suivante :

$$\mathcal{L}_{j,n} \simeq \eta_n 2^{jh_n}, \quad (2.129)$$

quand $2^j \rightarrow 0$. L'estimation locale de h_n peut être obtenue par régression linéaire à travers les échelles :

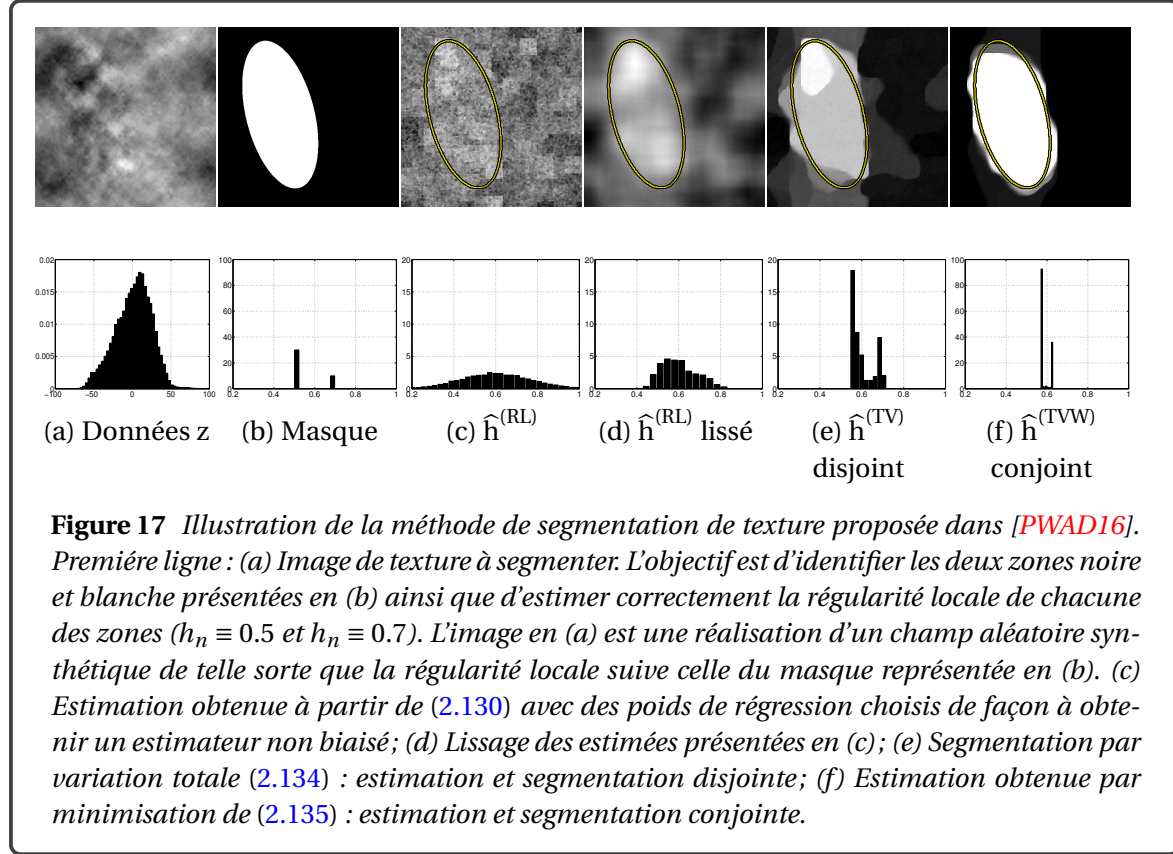
$$\hat{h}_n^{(RL)} = \sum_j w_{j,n} \log_2 \mathcal{L}_{j,n} \quad (2.130)$$

où $(w_{j,n})_{j,n}$ modélisent les poids de régression à l'échelle j et position n . On parle d'estimation non-biaisée lorsque

$$(\forall n \in \{1, \dots, N\}) \quad \sum_j w_{j,n} \equiv 0 \quad \text{et} \quad \sum_j j w_{j,n} \equiv 1, \quad (2.131)$$

Dans les études dédiées à l'estimation de la régularité locale, il est usuel de considérer un champ homogène, par conséquent, s'il s'agit de caractériser une texture uniforme sur une image,

l'estimation se fait en moyennant les valeurs des estimées obtenues en chaque position n de l'image. La question se complique lorsqu'il s'agit de segmenter l'image de texture à partir de ces estimées locales, c'est-à-dire d'identifier des changements de *régularité locale* sur une image (cf. Figure 17-a et b). En effet, l'estimation de la régularité locale montre une grande variabilité qui nuit à la détection des changements (cf. Figure 17-c), tandis que lisser les estimées (même sur un petit voisinage) empêche de localiser les changements correctement (cf. Figure 17-d).



Dans [FPWA15], nous avons étudié les performances de l'estimation de la régularité locale pour différentes quantités multi-échelles, à savoir les coefficients en ondelettes

$$\mathcal{L}_{j,n} = \frac{1}{3} \sum_{m=1}^3 |(Dz)_{j,n}^{(m)}| \quad (2.132)$$

où D représente une transformée en ondelettes, m désigne les coefficients horizontaux, verticaux, diagonaux, et les coefficients dominants,

$$\mathcal{L}_{j,n} = \sup_{\substack{m=\{1,2,3\} \\ (j',n') \in \mathcal{N}(j,n)}} |(Dz)_{j',n'}^{(m)}| \quad (2.133)$$

où $\mathcal{N}(j,n)$ représente un voisinage à travers les échelles les plus fines et les positions voisines comme illustré Figure 18. Quelque soit le choix de $\mathcal{L}_{j,n}$, un compromis biais-variance relatif à la taille des fenêtres d'analyse apparaît clairement.

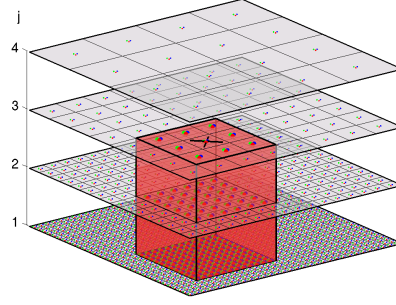


Figure 18 Représentation des coefficients dominants. Crédit : Herwig Wendt.

Dans un premier temps, nous avons proposé d'extraire les zones de régularité homogène en effectuant une segmentation des estimées de régularité locale par segmentation Chan-Vese (cf. (2.11)) [FPWA15]. Ce travail préliminaire a montré l'efficacité d'une telle approche pour identifier des zones ayant une régularité locale qui varie spatialement. Nous avons également considéré une segmentation via T-ROF [CS13] reposant sur le seuillage de $\hat{h}$ obtenu par :

$$\hat{h}^{(TV)} = \arg \min \frac{1}{2} \left\| h - \hat{h}^{(RL)} \right\|_2^2 + \lambda TV(h) \quad (2.134)$$

dont le résultat est illustré Figure 17-e. Notons que des schémas de minimisation par blocs avaient été envisagés dans [PPAP18]. Il s'agit d'une implémentation par blocs de FBS sur le dual pour minimiser (2.134). Nous y avons remarqué que la construction des blocs avait une importance cruciale dans ces approches dans un but de réduire la constante de Lipschitz par bloc. Aux résultats extraits de [PPAP18], nous avons ajouté une implémentation de l'algorithme de Chambolle-Pock prenant en compte la forte convexité [CP11a, Algorithme 2]. Les comparaisons sont présentées Figure 19. Il y apparaît clairement le bénéfice que l'on peut retirer d'une stratégie algorithmique pouvant prendre en compte une telle hypothèse de forte convexité.

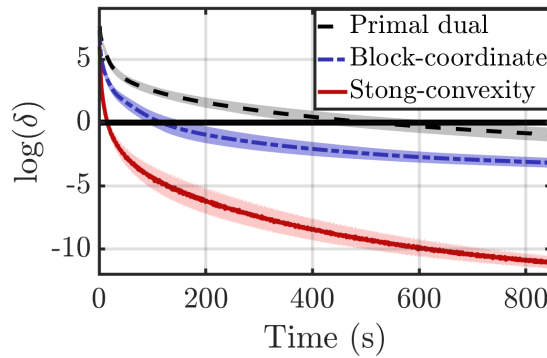


Figure 19 Gap de dualité en fonction du temps de calcul. Comparaison entre des itérations classiques de l'algorithme de Chambolle-Pock [CP11a, Algorithme 1] (en noir), une stratégie par blocs FBS sur le dual [PPAP18] (en bleu), des itérations de l'algorithme de Chambolle-Pock intégrant la forte convexité [CP11a, Algorithme 2] (en rouge). Crédit : Barbara Pascal.

Estimation et segmentation jointe : 1ère formulation Afin d'améliorer les performances d'estimation, nous avons proposé une solution algorithmique permettant de considérer conjointement estimation de la régularité locale et segmentation plutôt que de découpler ces deux étapes comme mentionné précédemment et comme il est usuellement proposé en segmentation de textures (cf. Figure 17-e). Nous avons montré qu'il était possible de formuler cette question comme un problème d'optimisation convexe non lisse pour lequel un algorithme efficace d'optimisation a été

développé [PWAD16]. Le critère prend la forme

$$(\hat{h}^{(TV)}, \hat{w}) \in \underset{h, w}{\operatorname{Argmin}} \sum_{j=j_1}^{j_2} \sum_{n=1}^N (w_{j,n} \log_2 \mathcal{L}_{j,n} - h_n)^2 + \lambda \operatorname{TV}(h) + \chi_1 \sum_{n=1}^N d_{C_1}(w_n) + \chi_2 \sum_{n=1}^N d_{C_2}(w_n) \quad (2.135)$$

où $w_n = (w_{j_1,n}, \dots, w_{j_2,n})$ avec $1 \leq j_1 < j_2 \leq J$ et où λ , η_1 et η_2 sont des quantités positives. Le premier terme assure que la solution respecte (2.131). d_{C_1} et d_{C_2} modélisent les distances aux ensembles convexes $C_1 = \{(\omega_{j_1}, \dots, \omega_{j_2}) \in \mathbb{R} \times \dots \times \mathbb{R} \mid \sum_{j=j_1}^{j_2} \omega_j = 0\}$ et $C_2 = \{(\omega_{j_1}, \dots, \omega_{j_2}) \in \mathbb{R} \times \dots \times \mathbb{R} \mid \sum_{j=j_1}^{j_2} j \omega_j = 1\}$. Ces distances permettent d'assurer une certaine flexibilité, principalement au niveau des zones de changement de régularité pour lesquelles le choix d'un estimateur non-biaisé est moins pertinent. Enfin le terme de pénalisation sur h , qui correspond à la pénalisation TV, permet de détecter les changements de régularité locale en imposant une solution de variation totale minimale. Une illustration des résultats obtenus est présentée Figure 17-f. Plus de résultats sont disponibles dans [PWAD16] incluant des comparaisons avec des méthodes de l'état-de-l'art et des résultats sur données réelles.

Cette étude a permis de mesurer le bénéfice d'utiliser des critères combinant estimation et segmentation comme illustré Figure 17. La principale limitation de cette approche variationnelle est le coût de calcul élevé lorsqu'il s'agit d'analyser de grands volumes de données, indispensable dans ce cadre de travail multiéchelle. La contribution suivante est dédiée à cette question.

Estimation et segmentation jointe : 2ème formulation Nous poursuivons ici notre effort d'estimation et segmentation jointe de textures fractales par morceaux en proposant un critère convexe alternatif à (2.135), visant à favoriser le caractère constant par morceaux de la régularité locale (comme dans la contribution précédente) mais également d'une quantité liée à la variance locale (notée η_n dans (2.129)), qui sont les deux paramètres qui caractérisent les textures fractales. L'énergie que nous proposons de minimiser dans [PPA21] s'écrit cette fois-ci encore comme un cas particulier de (2.1) qui prend la forme :

$$(\hat{v}, \hat{h}) = \underset{h, v}{\operatorname{argmin}} \frac{1}{2} \sum_j \|v + jh - \log_2 \mathcal{L}_j\|_{\text{Fro}}^2 + \lambda \| [Dv; \alpha Dh]^\top \|_{2,1}, \quad (2.136)$$

où $v = \log_2(\eta)$.

Ce critère est particulièrement intéressant car l'attache aux données est fortement convexe, son opérateur proximal et sa fonction conjuguée possèdent une forme explicite. Ces éléments apparaissent clairement lorsque que l'on ré-écrit le terme d'attache aux données sous cette forme :

$$\Psi(v, h; \mathcal{L}) = \frac{1}{2} \sum_n \sum_j (v_n + j h_n - \log_2 \mathcal{L}_{j,n})^2 \quad (2.137)$$

$$= \frac{1}{2} \sum_n \left\| A \begin{pmatrix} v_n \\ h_n \end{pmatrix} - \log_2 \mathcal{L}_n \right\|_2^2 \quad \text{où } A = \begin{pmatrix} 1 & j_1 \\ \vdots & \vdots \\ 1 & j_2 \end{pmatrix}. \quad (2.138)$$

La forme explicite de l'opérateur proximal provient de calculs standards des opérateurs proximaux :

$$\begin{pmatrix} p_n \\ q_n \end{pmatrix} = \operatorname{prox}_{\frac{1}{2} \|A \cdot - \log_2 \mathcal{L}_n\|_2^2} \begin{pmatrix} v_n \\ h_n \end{pmatrix} = (A^* A + \operatorname{Id})^{-1} \left(A^* \log_2 \mathcal{L}_n + \begin{pmatrix} v_n \\ h_n \end{pmatrix} \right)$$

où

$$\begin{cases} A^* A = \begin{pmatrix} R_0 & R_1 \\ R_1 & R_2 \end{pmatrix} & \text{avec } R_i = \sum_j j^i \\ A^* \log_2 \mathcal{L}_n = \begin{pmatrix} \mathcal{L}_n \\ \mathcal{T}_n \end{pmatrix} = \begin{pmatrix} \sum_j \log_2 \mathcal{L}_{j,n} \\ \sum_j j \log_2 \mathcal{L}_{j,n} \end{pmatrix} \end{cases}$$

conduisant à la forme explicite explicitée dans la Proposition 11.

Proposition 11 [PPA21] Pour tout $(v, h) \in \mathbb{R}^N \times \mathbb{R}^N$, soit $(p, q) = \text{prox}_\Psi(v, h) \in \mathbb{R}^N \times \mathbb{R}^N$ alors

$$\begin{cases} p = \frac{(1+R_2)(\mathcal{S}+v) - R_1(\mathcal{T}+h)}{(1+R_0)(1+R_2) - R_1^2}, \\ q = \frac{(1+R_0)(\mathcal{T}+h) - R_1(\mathcal{S}+v)}{(1+R_0)(1+R_2) - R_1^2}. \end{cases}$$

où $R_i = \sum_j j^i$, $\mathcal{S}_n = \sum_j \log_2 \mathcal{L}_{j,n}$, et $\mathcal{T}_n = \sum_j j \log_2 \mathcal{L}_{j,n}$.

Le second résultat extrait de [PPA21], concerne la forte convexité de Ψ dont la constante de forte convexité dépend des échelles sur lesquelles est effectuée la régression. Les valeurs de constantes de forte convexité sont présentées Figure 20. On voit en particulier que la constante de forte convexité croît avec le nombre d'échelles considérées. Ce constat est particulièrement intéressant lorsqu'il s'agit de considérer des images de très grande taille car malgré le plus grand volume de données, la structure du problème de minimisation permet d'assurer une plus grande accélération.

Proposition 12 [PPA21] La fonction $\Psi(v, h; \mathcal{L})$ est ρ -fortement convexe par rapport aux variables (v, h) , où $\rho > 0$ est la plus petite valeur propre de la matrice définie symétrique et positive $A^*A = \begin{pmatrix} R_0 & R_1 \\ R_1 & R_2 \end{pmatrix}$ où $R_i = \sum_j j^i$.

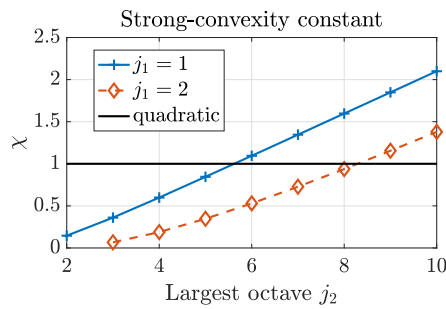


Figure 20 Constante de forte convexité en fonction de la dynamique des échelles sur laquelle est effectuée la régression.

Dans [PPA21], nous avons également proposé une forme explicite de la fonction conjuguée de Ψ dont l'expression est détaillée dans la Proposition 13. L'intérêt de calculer la fonction conjuguée est de pouvoir mettre en oeuvre un FBS sur le dual pour minimiser (2.136) ou une version accélérer FISTA sur le dual.

Proposition 13 [PPA21] Soit Ψ définie par (2.137), sa fonction conjuguée Ψ^* prend la forme :

$$\Psi^*(v, h; \mathcal{L}) = \frac{1}{2} \langle (v, h)^\top, (A^* A)^{-1} (v, h)^\top \rangle + \langle (\mathcal{S}, \mathcal{T})^\top, (A^* A)^{-1} (v, h)^\top \rangle + \mathcal{C},$$

où

$$\begin{cases} \mathcal{C} &= \frac{1}{2} \langle (\mathcal{S}, \mathcal{T})^\top, \mathbf{J}^{-1}(\mathcal{S}, \mathcal{T})^\top \rangle - \frac{1}{2} \sum_j (\log_2 \mathcal{L}_j)^2. \\ \mathcal{S} &= \sum_j \log_2 \mathcal{L}_j \\ \mathcal{T} &= \sum_j j \log_2 \mathcal{L}_j \\ A^* A &= \begin{pmatrix} R_0 & R_1 \\ R_1 & R_2 \end{pmatrix} \quad \text{et} \quad R_m = \sum_j j^m, \end{cases}$$

Deux catégories d'algorithmes proximaux sont mis en oeuvre pour minimiser (2.136). Nous considérons d'une part une version de l'algorithme FBS sur le dual [PPAP18] ainsi qu'une version accélérée de la forme FISTA également sur le dual. Cette formulation est rendue possible grâce à la forme explicite de la fonction conjuguée de Ψ^* présentée dans la Proposition 13. La seconde classe d'algorithmes repose sur l'algorithme de Chambolle-Pock pour lequel il existe une version accélérée dans le cas fortement convexe. Une description précise de ces algorithmes (DFB, FISTA, PD et AcPD) est disponible dans [PPA21]. Les différentes stratégies algorithmiques sont comparées dans un contexte de textures simulées de taille 256×256 . Les résultats sont présentés dans le Tableau 2.3. Il apparaît clairement que la prise en compte d'algorithmes inertiels de type FISTA ou reposant sur la forte convexité est nécessaire pour envisager d'employer ces approches sur de grands volumes de données.

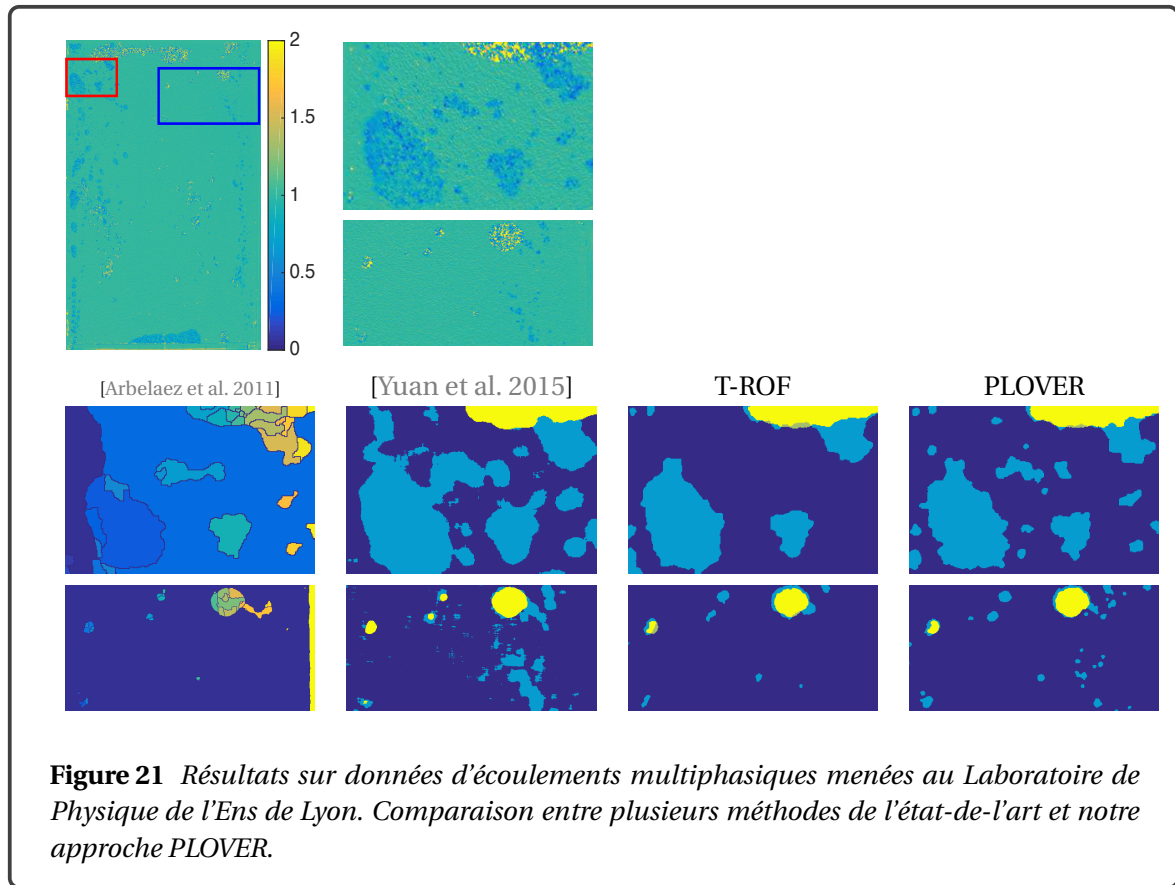
Dans [PPA21], nous comparons les résultats obtenu par estimation et segmentation disjointe reposant sur (2.134) avec les résultats obtenus par la minimisation de (2.136) pour différents choix de pénalisation de type TV permettant soit de pénaliser simultanément les contours de h et v soit de les pénaliser séparément. Il apparaît qu'une approche couplée améliore les performances au prix d'un coût de calcul plus important comme illustré dans le Tableau 2.3. Plus de résultats sont disponibles dans [PPA21].

		Configuration I			Configuration III		
		T-ROF	T-joint	T-couplé	T-ROF	T-joint	T-couplé
Itérations (10^3 it.)	DFB	96 ± 48	> 250	> 250	241 ± 18	> 250	> 250
	FISTA	1.7 ± 0.4	50.2 ± 21.0	231 ± 37	3.7 ± 0.7	48.1 ± 3.4	> 250
	PD	31.8 ± 17.0	> 250	> 250	201 ± 69	> 250	> 250
	AcPD	1.5 ± 0.4	31.4 ± 4.6	125 ± 67	45.2 ± 43	40.5 ± 2.8	121 ± 42
Temps (s)	DFB	1,090 ± 520	4,840 ± 15	4,210 ± 76	2,010 ± 73	4,810 ± 215	4,200 ± 76
	FISTA	16 ± 4	1,030 ± 410	4,800 ± 560	30 ± 5	989 ± 64	5,110 ± 340
	PD	297 ± 150	4,180 ± 69	4,110 ± 43	1,580 ± 490	4,150 ± 18	4,100 ± 15
	AcPD	15 ± 4	619 ± 96	2,420 ± 1,300	349 ± 330	785 ± 59	2,320 ± 790

TABLEAU 2.3 – Nombre d'itérations et temps de calcul nécessaire pour atteindre une condition d'arrêt basée sur le gap de dualité pour les différents algorithmes proposés, dans le cadre de deux configurations ($\Delta H = 0.2$, $\Delta \Sigma^2 = 0.1$) et III ($\Delta H = 0.1$, $\Delta \Sigma^2 = 0.1$). **DFB** : FBS sur le dual, **FISTA** : accélération inertielle de FBS sur le dual, **PD** : Chambolle-Pock, **AcPD** : accélération de Chambolle-Pock reposant sur la forte convexité.

Muni d'un critère fortement convexe permettant d'effectuer estimation et segmentation jointe puis d'un algorithme exploitant la forte convexité, il devient alors possible d'analyser un grand volume de données tel que les données d'écoulements multiphasés acquises au Laboratoire de Physique. Les résultats obtenus avec notre procédure PLOVER sont présentés sur la Figure 21, extraite de [PPA⁺18], et comparés à des méthodes de l'état-de-l'art en segmentation de texture. La méthode proposée permet d'extraire bien plus précisément les structures à petite échelle. Remarquons que le passage à l'échelle des algorithmes fut un point crucial pour l'analyse de ces données. Sans la forte convexité, la mise en oeuvre sur les données d'écoulements n'aurait été pos-

sible. En effet, dans le contexte des données d'écoulements chaque image à analyser est composée de plusieurs millions de pixels. Le traitement doit ensuite être effectué pour plusieurs dynamiques d'écoulements afin de construire un diagramme de phase. Sans une accélération notable, ces analyses avancées de traitement d'images n'auraient pu être mis en oeuvre.



2.3.3.1 Conclusions

La segmentation de textures est une question largement étudiée dans la littérature de traitement d'images que nous avons proposé de revisiter à l'aide de descripteurs fractals. Notre contribution principale fut de montrer que la combinaison de l'estimation des descripteurs et leur segmentation pouvait être formulée comme un critère intelligible de la forme (2.1) fortement convexe. L'intelligibilité de notre formulation est à comparer aux approches usuelles échaînant les étapes de prétraitement des données, extraction de descripteurs et segmentation, chacune nécessitant des réglages de paramètres ad hoc. D'autre part, en exploitant la forte convexité, nous proposons un algorithme rapide permettant de gérer un grand volume de données contrairement à certaines méthodes de l'état-de-l'art où il est parfois nécessaire d'effectuer le traitement sur des patches de l'image plutôt que globalement.

Dans des travaux futurs, nous souhaiterions intégrer des descripteurs d'orientation, de couleurs et d'analyse multifractale, afin d'obtenir une méthode efficace sur une plus grande gamme de données texturées.

Les réseaux de neurones profonds étant particulièrement adaptés pour une tâche de segmentation, une analyse comparative a été menée dans [PMPPA21] et motive nos travaux actuels sur le développements d'algorithmes déroulés dans le contexte de la segmentation de texture sur lesquels je reviendrai en conclusion générale.

2.3.4 Opérateur proximal versus gradient, que choisir? Quelques éléments de réponse.

Dans les sections 2.3.2.3 et 2.3.3, les accélérations sont respectivement dûes à l'expression d'une forme explicite de l'opérateur proximal d'une somme de deux fonctions et à la mise en œuvre de stratégies d'accélération (inertie ou forte convexité). Cependant, dans ces deux approches, il existe une autre interrogation, à savoir, faut-il mieux considérer un pas proximal ou un pas de gradient lorsque les deux sont envisageables. Par exemple, dans la minimisation du modèle discret de Mumford-Shah, nous pouvons soit considérer une linéarisation impliquant une descente de gradient suivi d'un pas proximal (PALM) ou effectuer un pas proximal d'une somme de deux fonctions (SL-PAM). Dans ce premier exemple, la solution reposant sur un pas proximal apparaît numériquement la plus efficace et nous l'avons argumenté par le fait que nous évitions un *splitting* mais la différence majeure réside aussi dans le choix du pas qui n'est alors plus contraint par une constante de Lipschitz locale. Dans le contexte de la segmentation de texture, nous avons également comparé une méthode combinant étape de gradient et une étape proximal (FBS ou sa version accélérée FISTA) et une méthode reposant uniquement sur des pas proximaux (l'algorithme de Chambolle-Pock et sa version accélérée avec forte convexité). Dans le Tableau 2.3, il apparaît que l'approche reposant uniquement sur des pas proximaux est la plus rapide. Des remarques similaires sont faites dans [CG19] où il a été montré numériquement qu'activer des fonctions par des pas proximaux plutôt que par leur gradient pouvait être avantageux numériquement.

Ces constats nous laissent penser qu'une méthode basée sur des pas proximaux est toujours plus rapide qu'une méthode basée gradient. Dans cette section, nous nous posons deux questions : faut-il systématiquement activer une fonction par son opérateur proximal lorsqu'une forme explicite existe? et d'autre part, au vu d'un problème de minimisation donné, est-il possible d'identifier parmi un ensemble d'algorithmes envisageables celui qui sera le plus efficace en fonctions des propriétés du critère et de ses paramètres tels que le paramètre de régularisation, sa constante de forte convexité, la constante de Lipschitz de son gradient?

Dans cette section nous proposons quelques éléments de réponses dans le cas où le problème à minimiser prend la forme spécifique suivante :

Problème 1 Soit $f_1 \in C_{1/\alpha}^{1,1}(\mathcal{H})$ et $f_2 \in C_{1/\beta}^{1,1}(\mathcal{H})$, avec $\alpha \in]0, +\infty[$ et $\beta \in]0, +\infty[$. De plus, on suppose que f_1 est ρ -fortement convexe avec $\rho \in]0, 1/\alpha[$. Le problème consiste à estimer :

$$\hat{x} = \arg \min_{x \in \mathcal{H}} f_1(x) + f_2(x). \quad (2.139)$$

Etat-de-l'art Du point de vue de l'utilisateur, le choix de l'algorithme le plus efficace pour résoudre un problème de minimisation donné de la forme (2.1) est une tâche compliquée. Le taux de convergence est un outil utile pour fournir une comparaison théorique entre les algorithmes. Cependant, le comportement théorique d'un schéma algorithmique peut différer considérablement de son efficacité numérique. Il est donc important d'obtenir des taux de convergence précis exploitant les propriétés de f_1 et f_2 . Dans ce contexte, des taux de convergence linéaires précis peuvent être obtenus pour plusieurs algorithmes proximaux sous hypothèse de forte convexité de f_1 et/ou f_2 [THG18; GB17b; DY17; RHY19; RHY19], et peuvent être étendus lorsque la forte convexité est satisfaite sur des variétés particulières, par exemple dans le cas de fonctions partiellement lisses [Lew02; LFP17]. De plus, les taux de convergence sous-linéaire de quelques méthodes de premier ordre en fonction de l'exposant Kurdyka-Łojasiewicz (KL) sont obtenus dans [ABS13] lorsque $f_1 + f_2$ est une fonction KL (voir [BDL06]). Les exposants KL étant généralement difficiles à calculer [BNPS17], nous concentrons notre analyse sur des résultats liés à l'hypothèse de forte convexité et aux taux de convergence linéaires associés.

Contributions Dans le cas où f_1 est fortement convexe, nous explicitons les constantes de Lipschitz continues des opérateurs régissant la méthode du gradient (EA), FBS, PRS et DRS (cf. section 2.1.4), permettant ainsi une comparaison de leur taux de convergence. Ces résultats sont détaillés dans la Proposition 14.

Proposition 14 [BAP20, Proposition 9] Dans le contexte du Problème 1, on suppose que f_1 est ρ -fortement convexe, avec $\rho \in]0, \alpha^{-1}[$. Soit $G_{\bullet}, T_{\bullet}, R_{\bullet}, T_{\bullet}$ et $S_{\bullet}$ définis dans le paragraphe 2.1.4.1. Soit $\tau > 0$. Alors,

[1] On suppose que $\tau \in]0, 2\beta\alpha/(\beta + \alpha)[$. Alors, $G_{\tau(\nabla f_1 + \nabla f_2)}$ est $r_G(\tau)$ -Lipschitz continue, où

$$r_G(\tau) := \max\{|1 - \tau\rho|, |1 - \tau(\beta^{-1} + \alpha^{-1})|\} \in]0, 1[. \quad (2.140)$$

En particulier, le minimum de (2.140) est atteint pour

$$\tau^* = \frac{2}{\rho + \alpha^{-1} + \beta^{-1}} \quad \text{et} \quad r_G(\tau^*) = \frac{\alpha^{-1} + \beta^{-1} - \rho}{\alpha^{-1} + \beta^{-1} + \rho}. \quad (2.141)$$

[2] Supposons que $\tau \in]0, 2\alpha[$. Alors $T_{\tau\nabla f_2, \tau\nabla f_1}$ est $r_{T_1}(\tau)$ -Lipschitz continue, avec

$$r_{T_1}(\tau) := \max\{|1 - \tau\rho|, |1 - \tau\alpha^{-1}|\} \in]0, 1[. \quad (2.142)$$

En particulier, le minimum de (2.142) est atteint pour

$$\tau^* = \frac{2}{\rho + \alpha^{-1}} \quad \text{et} \quad r_{T_1}(\tau^*) = \frac{\alpha^{-1} - \rho}{\alpha^{-1} + \rho}. \quad (2.143)$$

[3] Supposons que $\tau \in]0, 2\beta[$. Alors $T_{\tau\nabla f_1, \tau\nabla f_2}$ est $r_{T_2}(\tau)$ -Lipschitz continue, où

$$r_{T_2}(\tau) := \frac{1}{1 + \tau\rho} \in]0, 1[. \quad (2.144)$$

En particulier, le minimum de (2.144) est atteint pour

$$\tau^* = 2\beta \quad \text{et} \quad r_{T_2}(\tau^*) = \frac{1}{1 + 2\beta\rho}. \quad (2.145)$$

[4] $R_{\tau\nabla f_2, \tau\nabla f_1}$ et $R_{\tau\nabla f_1, \tau\nabla f_2}$ sont $r_R(\tau)$ -Lipschitz continue, où

$$r_R(\tau) = \max\left\{\frac{1 - \tau\rho}{1 + \tau\rho}, \frac{\tau\alpha^{-1} - 1}{\tau\alpha^{-1} + 1}\right\} \in]0, 1[. \quad (2.146)$$

En particulier, le minimum de (2.146) est atteint

$$\tau^* = \sqrt{\frac{\alpha}{\rho}} \quad \text{et} \quad r_R(\tau^*) = \frac{1 - \sqrt{\alpha\rho}}{1 + \sqrt{\alpha\rho}}. \quad (2.147)$$

[5] $S_{\tau\nabla f_2, \tau\nabla f_1}$ et $S_{\tau\nabla f_1, \tau\nabla f_2}$ sont $r_S(\tau)$ -Lipschitz continue, où

$$r_S(\tau) = \min\left\{\frac{1 + r_R(\tau)}{2}, \frac{\beta + \tau^2\rho}{\beta + \tau\beta\rho + \tau^2\rho}\right\} \in]0, 1[\quad (2.148)$$

et r_R est définie par (2.146). En particulier, le pas optimal et le minimum de (2.148) sont

$$(\tau^*, r_S(\tau^*)) = \begin{cases} \left(\sqrt{\frac{\alpha}{\rho}}, \frac{1}{1 + \sqrt{\alpha\rho}}\right), & \text{si } \beta \leq 4\alpha; \\ \left(\sqrt{\frac{\beta}{\rho}}, \frac{2}{2 + \sqrt{\beta\rho}}\right), & \text{sinon.} \end{cases} \quad (2.149)$$

Ceci donne un support théorique aux résultats obtenus dans [CG19] pour le cas fortement convexe. Dans [BAP20], nous proposons également les taux de convergence linéaire des quatre algorithmes mentionnés précédemment dans le contexte des opérateurs cocoercifs¹⁰ fortement monotones, qui conduisent à de moins bonnes valeurs de taux de convergence que celles obtenues dans le contexte d'optimisation. Nous proposons en particulier une amélioration du taux de convergence pour DRS inspiré de [Gis17; GB17b], qui exploite le contexte entièrement lisse, qui est répliqué dans le cadre cocoercif. De plus, nous obtenons une convergence améliorée pour la méthode du gradient dans le cadre fortement monotone et coercif inspiré de [Gis17]. Dans ce cadre, nous améliorons plusieurs taux de convergence de la littérature dont [Mer80; CR97; Tse91; LM79; RTBG20].

Plusieurs remarques peuvent être faites sur cette proposition :

- Quand $\rho \sim 0$ le choix de (2.143) justifie le choix classique $\tau^* \approx 2\alpha$ où α est l'inverse de la constante de Lipschitz. Ce choix apparaît naturellement dans plusieurs problèmes inverses.
- Les Figures 22 et 23 représentent les taux r de la Proposition 14 pour les différentes méthodes. Sur la Figure 22 (première ligne), on peut observer que pour une valeur fixée de α et β (inverses des constantes de Lipschitz de f_1 et f_2), plus grande est la constante de forte convexité ρ , meilleur est le taux de convergence. De plus, pour α et β fixés, les itérations de PRS $R_{\tau\nabla f_2, \tau\nabla f_1}$ et les itérations de FBS $T_{\tau\nabla f_2, \tau\nabla f_1}$ et $T_{\tau\nabla f_1, \tau\nabla f_2}$ conduisent aux meilleurs taux de convergence. Sur la deuxième ligne de la Figure 22, nous concluons que plus la constante de Lipschitz du gradient de la fonction fortement convexe est petite (i.e. plus grande est α), meilleur est le taux de convergence à l'exception de $T_{\tau\nabla f_1, \tau\nabla f_2}$. Dans ce cas, la constante de Lipschitz ne dépend que du paramètre de forte convexité ρ . Enfin, pour ρ et β fixes, nous observons que les itérations PRS $R_{\tau\nabla f_2, \tau\nabla f_1}$ et les itérations FBS $T_{\tau\nabla f_2, \tau\nabla f_1}$ conduisent aux meilleurs taux de convergence. Sur la troisième ligne de la Figure 22, où ρ et α sont fixes, nous observons qu'une constante de Lipschitz plus petite de ∇f_2 (i.e. plus grand β) affecte positivement $G_{\tau\nabla f_2, \tau\nabla f_1}$, $T_{\tau\nabla f_1, \tau\nabla f_2}$, et $S_{\tau\nabla f_2, \tau\nabla f_1}$. Le dernier taux de convergence profite du contexte de Problème 1.
- Sur la Figure 23, nous observons l'avantage du raffinement des taux de convergence dans un cadre d'optimisation (lignes continues) par rapport au cadre des opérateurs cocoercif (lignes discontinues) (cf. [BAP20, Proposition 7]) dans toutes les méthodes à l'exception de $T_{\tau\nabla f_1, \tau\nabla f_2}$, dont le taux est le même. Nous avons aussi observé qu'en général les itérations PRS $R_{\tau\nabla f_2, \tau\nabla f_1}$ conduisent au meilleur taux de convergence pour plusieurs configurations de (α, β, ρ) .

Résultats expérimentaux Dans [BAP20], nous avons souhaité confronter les résultats théoriques à des résultats numériques. Pour cela nous avons considéré le problème jouet de débruitage de signaux constants par morceaux pour lequel il existe d'autre part de nombreux algorithmes très efficaces dans la littérature. Le problème de minimisation considéré se formule :

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \quad \frac{1}{2} \|u - z\|_2^2 + \lambda h(Dz), \quad (2.150)$$

où $D \in \mathbb{R}^{N-1 \times N}$ correspond à l'opérateur de différences finies de premier ordre :

$$(\forall n \in \{1, \dots, N-1\}) \quad (Du)_n = \frac{1}{2}(u_n - u_{n-1})$$

et h correspond à la pénalisation de Huber, une approximation lisse de la norme ℓ_1 paramétrée par $v > 0$, définie par (voir, e.g., [CG19, Exemple 2.5])

$$h : \mathbb{R}^{N-1} \rightarrow \mathbb{R} : (w_n)_{1 \leq n \leq N-1} \mapsto \sum_{n=1}^{N-1} h_n(w_n) \quad \text{et} \quad h_n : w \mapsto \begin{cases} |w| - \frac{v}{2}, & \text{si } |w| > v; \\ \frac{|w|^2}{2v}, & \text{si } |w| \leq v. \end{cases} \quad (2.151)$$

10. Soit D un sous-ensemble non vide d'un espace de Hilbert $\mathcal{H}$, soit $T : D \rightarrow \mathcal{H}$ et soit $\beta > 0$. Alors T est β^{-1} -cocoercive si $\beta^{-1}T$ est fermement contractant, i.e., $(\forall x \in D)(\forall y \in D) \langle x - y | Tx - Ty \rangle \geq \beta^{-1} \|Tx - Ty\|^2$

La fonction $h \in C_{1/\nu}^{1,1}(\mathbb{R}^{N-1})$.

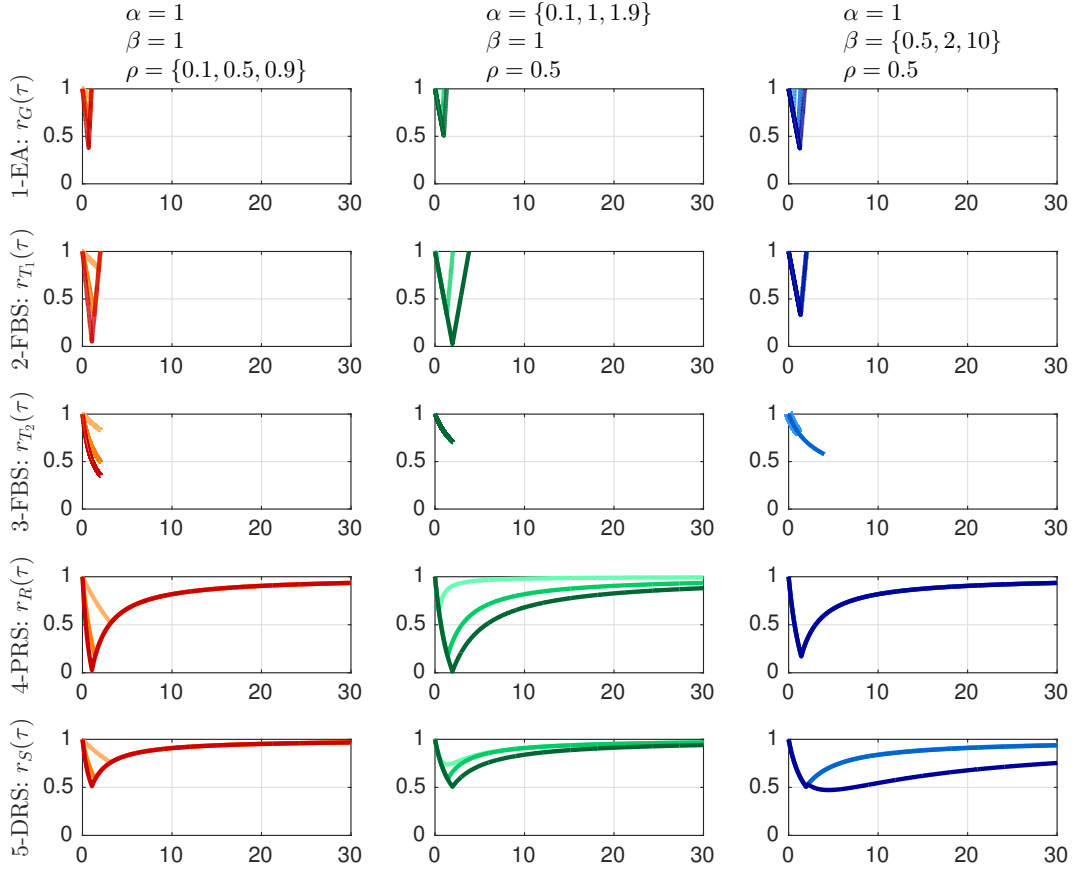


Figure 22 Taux de convergence (constante de Lipschitz continue) donnée par la Proposition 14 en fonction du pas τ pour différents choix de α, β , et ρ .

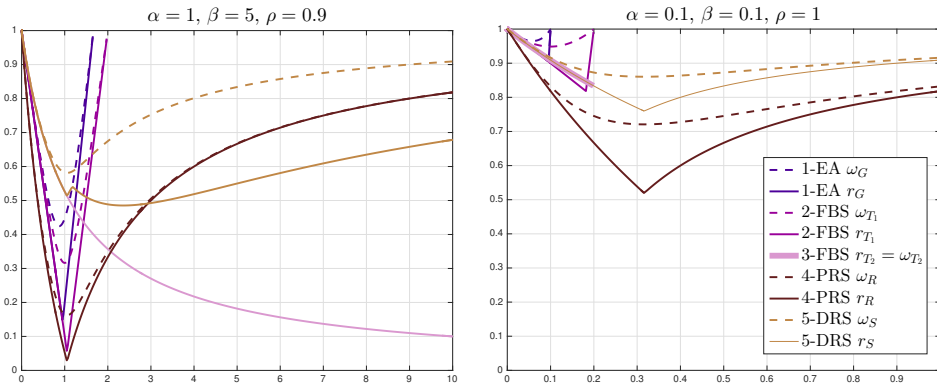


Figure 23 Comparaison des taux de convergence de EA, FBS, PRS, DRS obtenus dans la Proposition 14 (lignes continues) et dans [BAP20, Proposition 7] (lignes discontinues) pour deux choix de α, β et ρ . On peut noter que les taux d'optimisation sont meilleurs que ceux du contexte général cocoercif.

En choisissant $f_1 = \frac{1}{2} \|\cdot - z\|_2^2$ et $f_2 = \lambda h(D\cdot)$, (2.150) est alors un cas particulier du Problème 1, où f est $\rho = 1$ fortement convexe, $\alpha = 1$, et $\beta = \frac{\nu}{\lambda \|D\|^2}$. Ce problème de minimisation peut être directement résolu par

- 1- **EA** : En considérant $G_{\tau(\nabla f_1 + \nabla f_2)}$ avec le pas τ^* donné par (2.141).
- 2- **FBS** : En considérant $T_{\tau \nabla f_1, \tau \nabla f_2}$ avec le pas τ^* donné par (2.145).

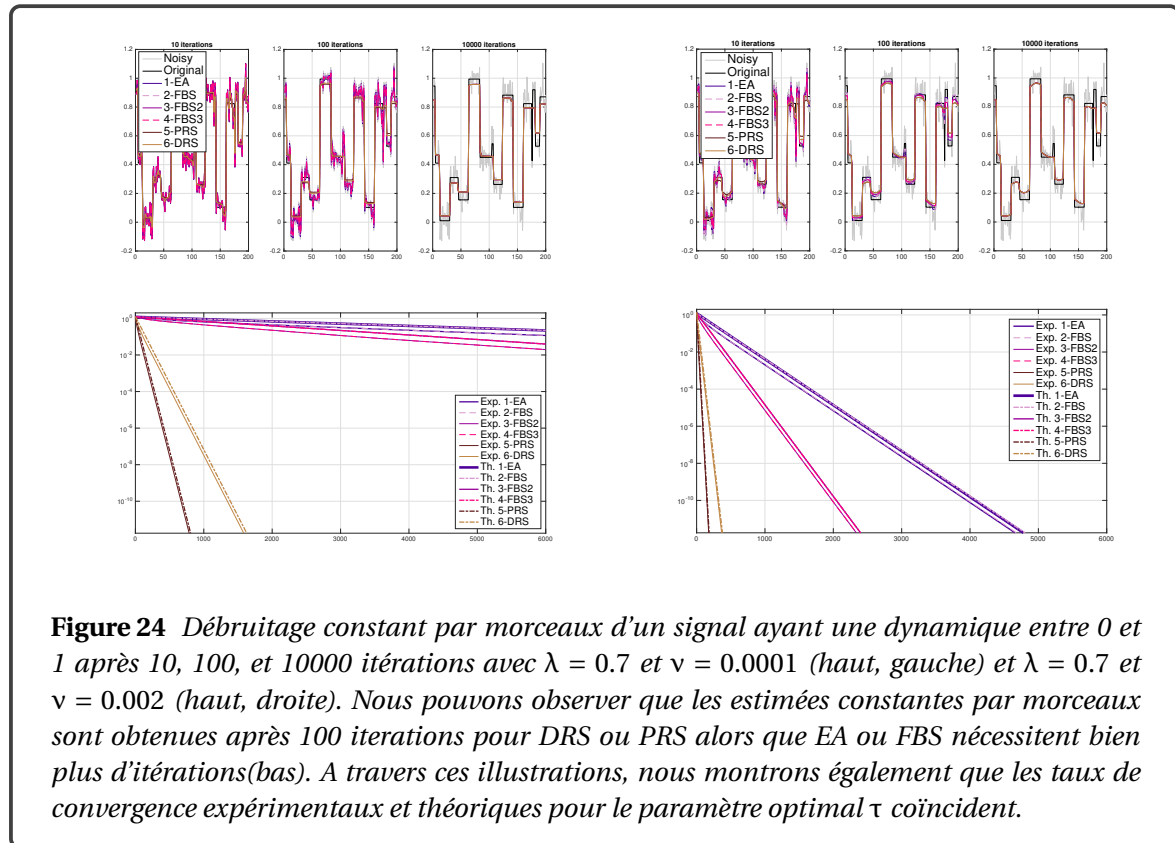
L'opérateur proximal de h possède une forme explicite [CG19] mais l'opérateur proximal de $h \circ D$ n'est pas explicite à cause de l'influence de D (cf. section 2.1.4). En exploitant la structure séparable de h , on obtient une formulation équivalente de (2.150) qui se formule :

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \quad \frac{1}{2} \|u - z\|_2^2 + \lambda h_{\mathbb{I}_1}(D_{\mathbb{I}_1} x) + \lambda h_{\mathbb{I}_2}(D_{\mathbb{I}_2} x), \quad (2.152)$$

où $\mathbb{I}_1 = \{1, 3, \dots\}$ et $\mathbb{I}_2 = \{2, 4, \dots\}$ sont les ensembles d'indices impairs et pairs pour $k \in \{1, 2\}$, $h_{\mathbb{I}_k}(y_{\mathbb{I}_k}) = \sum_{i \in \mathbb{I}_k} h_i(y_i)$, et $D_{\mathbb{I}_k} \in \mathbb{R}^{\mathbb{I}_k \times N}$ correspond à la sous matrice de D associée regroupant les $\mathbb{I}_k$ lignes. Etant donné que $L_{\mathbb{I}_1} L_{\mathbb{I}_1}^\top = \text{Id}/2$ et $L_{\mathbb{I}_2} L_{\mathbb{I}_2}^\top = \text{Id}/2$, la ré-écriture (2.152) permet d'obtenir une forme explicite de l'opérateur proximal de $h_{\mathbb{I}_k} \circ L_{\mathbb{I}_k}$, pour $k \in \{1, 2\}$. En choisissant $\tilde{f}_1 = \frac{1}{2} \|\cdot - z\|_2^2 + \chi h_{\mathbb{I}_2}(L_{\mathbb{I}_2} \cdot)$ and $\tilde{f}_2 = \chi h_{\mathbb{I}_1}(L_{\mathbb{I}_1} \cdot)$, nous écrivons (2.152) comme le Problème 1, où $\tilde{f}_1$ est $\rho = 1$ fortement convexe, $\alpha = \frac{\nu}{\nu + \lambda \|L_{\mathbb{I}_2}\|^2}$, et $\beta = \frac{\nu}{\lambda \|L_{\mathbb{I}_1}\|^2}$. Cette approche donne lieu à 4 méthodes alternatives pour résoudre (2.152).

- 3- **FBS 2** : En considérant $T_{\tau \nabla \tilde{f}_2, \tau \nabla \tilde{f}_1}$ avec le pas τ^* dans (2.143).
- 4- **FBS 3** : En considérant $T_{\tau \nabla \tilde{f}_1, \tau \nabla \tilde{f}_2}$ avec le pas τ^* in (2.145).
- 5- **PRS** : En considérant $R_{\tau \nabla \tilde{f}_1, \tau \nabla \tilde{f}_2}$ avec le pas τ^* dans (2.147).
- 6- **DRS** : En considérant $S_{\tau \nabla \tilde{f}_1, \tau \nabla \tilde{f}_2}$ avec le pas τ^* dans (2.149).

Les taux de convergence théoriques et numériques de ces différentes approches sont présentés Figure 24 pour différents choix de ν et λ conduisant à des estimées différentes. Pour une faible valeur de ν les performances sont proches du résultat obtenu par minimisation ℓ_1 .



Sur la Figure 24 (bas), nous observons que les itérations PRS offrent les meilleurs taux théoriques et expérimentaux lorsque la taille de pas optimal est sélectionnée. Les itérations DRS fournissent également un bon comportement, tandis que les stratégies EA et FBS reposant sur le fractionnement $f_1 = \frac{1}{2} \|\cdot - z\|_2^2$ et $f_2 = \lambda h(L \cdot)$ apparaissent moins efficaces que celles impliquant le fractionnement $\tilde{f}_1 = \frac{1}{2} \|\cdot - z\|_2^2 + \lambda h_{\mathbb{I}_2}(L_{\mathbb{I}_2} \cdot)$ et $\tilde{f}_2 = \lambda h_{\mathbb{I}_1}(L_{\mathbb{I}_1} \cdot)$. Des conclusions similaires peuvent être observées à partir de la Figure 24 (en haut), où la solution optimale est atteinte après 100 itérations pour DRS (marron clair) et PRS (marron foncé) tandis que les procédures basées sur le

gradient nécessitent beaucoup plus d'itérations. Cela est particulièrement vrai lorsque v est petit, ce qui conduit à une grande constante de Lipschitz.

2.3.4.1 Conclusions

Cette étude préliminaire a montré qu'il était possible de proposer des comparaisons théoriques fiables entre divers algorithmes proximaux en amont d'une implémentation. Le cadre d'étude est ici très spécifique (fonctions différentiables de gradient Lipschitz) cependant des analyses similaires sont en cours sous des hypothèses moins restrictives. Cette analyse mériterait également d'être étendue dans le contexte des algorithmes proximaux primaux-duaux et dans le cadre d'accélérationnelles inertielles.

2.4 Sélection des hyperparamètres

2.4.1 Motivations et état-de-l'art

Le choix de l'hyperparamètre λ dans un problème de minimisation de la forme (2.1) en traitement du signal et des images est d'une importance cruciale car il permet de contrôler la qualité d'estimation. Dans cette section, nous concentrons notre attention sur un problème de minimisation de la forme :

$$\hat{u}(z; \Lambda) \in \underset{u \in \mathcal{H}}{\operatorname{Argmin}} \frac{1}{2} \|Au - z\|_2^2 + \|Du\|_\Lambda. \quad (2.153)$$

où $\|\cdot\|_\Lambda$ peut modéliser une pénalisation ℓ_0 (i.e $\lambda \|\cdot\|_0$), une pénalisation ℓ_1 (i.e $\lambda \|\cdot\|_1$) ou une pénalisation mixte $\ell_{2,1}$ (i.e $\lambda \|\cdot\|_{2,1}$). Ce formalisme permet également de considérer des formes plus complexes de pénalisation comme (2.136) pouvant impliquer plusieurs paramètres (e.g. $\Lambda = (\lambda, \alpha)$).

Parmi les grandes classes de méthodes permettant de régler les hyperparamètres, on peut d'abord citer les approches hiérarchiques, les principe de SURE, l'approche par courbe en L [HO93], l'approche par validation croisée ou encore des règles empiriques [SB10], où le paramètre de régularisation λ est sélectionné selon une règle heuristique, à savoir $\lambda = 0.25\sqrt{N}\sigma$.

Le cadre d'inférence hiérarchique bayésienne a suscité un intérêt considérable pour résoudre les problèmes de point de changement ou de débruitage par morceaux [LL01; DTS07] mais également des problèmes plus complexes de restauration d'images [MNCM01; BMK09; CBFe09; CPTC11]. Cela résulte principalement de leur capacité à inclure les hyperparamètres dans la modélisation bayésienne et à les estimer conjointement avec les paramètres d'intérêt. En contrepartie de cette flexibilité intrinsèque, l'approximation des estimateurs bayésiens associés à ce modèle hiérarchique nécessite généralement l'utilisation d'algorithmes de Monte Carlo à chaînes de Markov (MCMC), souvent réputés gourmands en temps de calcul. Dans la section 2.4.2, nous proposons de revisiter cette approche hiérarchique dans le contexte spécifique de la fonctionnelle L2- Potts 1-D (cf. (2.12)) avec une approche hybride hiérarchique-variationnelle.

Une seconde classe de méthodes repose sur l'estimateur SURE [Ste81] qui vise à produire un estimateur non biaisé de l'erreur quadratique moyenne. La mise en oeuvre d'une approche SURE nécessite la connaissance préalable de la variance de l'erreur résiduelle, souvent indisponible *a priori* mais pour laquelle des estimateurs empiriques (cf. [Mal97]) peuvent être mis en oeuvre et permettent d'obtenir de bonnes performances. On peut se référer à [BBP05; PBBC09] pour le débruitage hyperspectral ou la déconvolution d'images impliquant des trames où la variance est estimée. Dans la section 2.4.3, nous proposons de mettre en oeuvre l'estimateur SURE dans le cas où A est une matrice de rang plein et lorsque le bruit est corrélé. Afin d'analyser de telles données nous avons dû proposer dans [PVPA21] une extension des résultats existants liés à SURE et à son gradient SUGAR [DVPF14], permettant de mettre en oeuvre des solutions de recherche efficace des hyperparamètres. Cette approche est d'abord validée sur des données synthétiques puis sur des données réelles d'écoulement multiphasés. Cette approche est également mis en oeuvre dans le cadre plus simple (bruit non-corrélé) des signaux stick-slip.

Enfin, dans la section 2.4.4, je présenterai une technique d'estimation reposant sur le formalisme de l'apprentissage profond pour apprendre le paramètre de régularisation mais également l'opérateur D et les pas de l'algorithme. Plus précisément, nous proposons de reformuler l'algorithme de Condat-Vũ pour la minimisation de (2.153) avec $\|\cdot\|_\Lambda = \lambda \|\cdot\|_1$ dans un contexte d'algorithme déroulé.

2.4.2 Approche bayésienne hiérarchique

Cette section est dédiée au problème de détection de changement par la fonctionnelle L2-Potts (cf. (2.12)) qui correspond à (2.153) quand $A = \text{Id}$ et $\|\cdot\|_\Lambda = \lambda \|\cdot\|_0$ avec $\lambda > 0$. Partant de l'interprétation bayésienne de (2.12) et capitalisant sur des algorithmes de programmation dynamique efficaces [WL02] pour ce contexte précis de minimisation, permettant de récupérer $\hat{u}(z; \lambda)$ pour une valeur prédéfinie de λ , nous proposons une procédure d'estimation conjointe du signal débruité et de sélection de l'hyperparamètre optimal λ , sans supposer de connaissance préalable supplémentaire sur la variance résiduelle du bruit de mesure σ^2 . Nous allons considérer un critère étendu de (2.12) de la forme :

$$\underset{u \in \mathbb{R}^N, \lambda > 0, \sigma^2 > 0}{\text{minimiser}} \frac{1}{2\sigma^2} \|u - z\|_2^2 + \frac{\lambda}{\sigma^2} \|Du\|_0 + \phi(\lambda, \sigma^2) \quad (2.154)$$

où $\phi: \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ est une fonction pénalisant λ et σ^2 .

Paramétrisation En suivant les formulations [Lav98; LL01; DTD07], les signaux constants par morceaux $u \in \mathbb{R}^N$ peuvent être explicitement paramétrés par les emplacements des points de changement e et les amplitudes des segments constants par morceaux c .

Pour localiser les instants temporels des points de changement dans un signal constant par morceaux, un vecteur $e = (e_i)_{1 \leq i \leq N} \in \{0, 1\}^N$ est introduit comme suit :

$$e_i = \begin{cases} 1, & \text{si il y a un point de changement à l'instant } i, \\ 0, & \text{sinon.} \end{cases} \quad (2.155)$$

Par convention, $e_i = 1$ signifie que u_i est le dernier échantillon appartenant au segment courant, et donc que u_{i+1} appartient au segment suivant. De plus, nous choisissons $e_N = 1$ afin de garantir que le nombre Q de segments soit égal au nombre de points de changement, c'est-à-dire, $Q = \sum_{i=1}^N e_i$.

Pour chaque indice de segment $q \in \{1, \dots, Q\}$, l'ensemble $\mathcal{E}_q \subset \{1, \dots, N\}$ est utilisé pour désigner l'ensemble d'indices temporels associés au q -ième segment. En particulier, il est intéressant de noter que $\mathcal{E}_q \cap \mathcal{E}_{q'} = \emptyset$ pour $q \neq q'$ et $\bigcup_{q=1}^Q \mathcal{E}_q = \{1, \dots, N\}$. Ci-après, la notation Q_e sera adoptée pour souligner la dépendance du nombre Q de segments sur le vecteur indicateur e , tel que $Q = \|e\|_0$.

Les amplitudes de chaque segment du signal constant par morceaux peuvent être codées en introduisant le vecteur $c = (c_q)_{1 \leq q \leq Q_e}$ tel que

$$(\forall q \in \{1, \dots, Q_e\})(\forall i \in \mathcal{E}_q) \quad u_i = c_q. \quad (2.156)$$

Un problème important intrinsèque à la fonctionnelle L2-Potts, mais également à cette formulation, concerne le domaine de définition de c , qui appartient à $\mathbb{R}^{Q_e}$ avec $Q_e = \{1, \dots, N-1\}$ dont la dimension est *a priori* inconnue, car cela dépend du nombre Q de points de changements.

La paramétrisation présentée précédemment conduit au lemme suivant qui implique que l'estimation du signal constant par morceaux u peut être formulée de manière équivalente via l'estimation du vecteur de paramètres $\{e, c\}$.

Lemme 1 Soit $z \in \mathbb{R}^N$ et $\phi: \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$. Le problème (2.12) est équivalent à

$$\underset{e \in \{0,1\}^N, c \in \mathbb{R}^{Q_e}, \lambda > 0, \sigma^2 > 0}{\text{minimize}} \left\{ \frac{1}{2\sigma^2} \sum_{q=1}^{Q_e} \sum_{i \in \mathcal{E}_q} (z_i - c_q)^2 + \frac{\lambda}{\sigma^2} (Q_e - 1) + \phi(\lambda, \sigma^2) \right\}. \quad (2.157)$$

Formalisme bayésien hiérarchique Dans [LL01; DTD07], le problème de la détection des points de changement dans une séquence stationnaire a été abordé suivant une procédure d'inférence bayésienne qui construit une distribution *a posteriori* du vecteur de paramètres $\{e, c\}$ à partir de

la fonction de vraisemblance associée au modèle d'observation et des distributions *a priori* choisies pour les paramètres inconnus. Nous avons proposé une approche similaire pour produire un modèle bayésien hiérarchique qui peut être étroitement lié au problème (2.157). Les hypothèses formulées sont les suivantes :

- Nous supposons que $z = \tilde{u} + \epsilon$ avec $(\epsilon_i)_{1 \leq i \leq N}$ des variables aléatoires gaussiennes de moyenne nulle indépendantes et identiquement distribuées de variance inconnue σ^2 . La fonction de vraisemblance conjointe résultante des observations z conduisant à un modèle constant par morceaux paramétré par $\{e, c\}$ se formule :

$$\mu(z|e, c, \sigma^2) = \prod_{q=1}^{Q_e} \prod_{i \in \mathcal{E}_q} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(c_q - z_i)^2}{2\sigma^2}\right). \quad (2.158)$$

Sous l'hypothèse que (e, c) et σ^2 sont indépendants et que les $(c_q)_{1 \leq q \leq Q_e}$ définissant le vecteur c , conditionnellement à e , sont indépendants, on construit un modèle *a priori* pour (e, c, σ^2) de la forme :

$$\mu(e, c, \sigma^2|p) = \mu(c, e|p)\mu(\sigma^2) = \mu(c|e)\mu(e|p)\mu(\sigma^2).$$

où l'hyperparamètre p représente la probabilité *a priori* d'occurrence d'un changement.

- Un choix naturel pour $\mu(c|e)$ consiste à choisir des distributions *a priori* gaussiennes conjuguées indépendantes $\mathcal{N}(c_0, \sigma_0^2)$ pour les amplitudes des segments c_q , i.e.,

$$\mu(c|e) = \prod_{q=1}^{Q_e} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(c_q - c_0)^2}{2\sigma_0^2}}. \quad (2.159)$$

En effet, cet ensemble de priors conjugués garantit que les distributions postérieures conditionnelles des amplitudes des segments soient toujours des distributions gaussiennes.

- Suite à des choix bien admis tels que ceux de [Lav98; LL01; PADF02; DTD07], les composantes e_i du vecteur e sont supposées distribuées selon une distribution de Bernoulli d'hyperparamètre p :

$$\mu(e|p) = \prod_{i=1}^{N-1} p^{e_i} (1-p)^{1-e_i} = \left(\frac{p}{1-p}\right)^{(Q_e-1)} (1-p)^{(N-1)}. \quad (2.160)$$

- L'indépendance *a priori* entre les composantes de e suppose implicitement que l'occurrence d'un changement à un indice de temps donné ne dépend pas de l'occurrence de changement à un autre indice de temps. De plus, l'hyperparamètre p représente la probabilité *a priori* d'occurrence d'un changement, qui est supposé indépendant de la localisation. Évidemment, pour des applications particulières, des choix alternatifs et plus spécifiques peuvent être adoptés en s'appuyant par exemple sur des contraintes dures [KTHD12] ou des modèles markoviens [DTS06], par exemple pour éviter des solutions composées de segments trop courts.
- Dans un formalisme bayésien hiérarchique, les hyperparamètres tels que la variance du bruit et autres paramètres définissant les distributions *a priori* peuvent être inclus dans le modèle à estimer conjointement avec $\{e, c\}$. En particulier, pour tenir compte de l'absence de connaissance *a priori* sur la variance du bruit σ^2 , un *a priori* de Jeffreys non informatif est considéré :

$$\mu(\sigma^2) \propto \frac{1}{\sigma^2}. \quad (2.161)$$

Proposée dans [Jef46], l'utilisation de cette distribution impropre a été largement préconisée dans la littérature bayésienne pour son invariance sous reparamétrisation [Rob07, Chap. 3].

Enfin, comme dans [DTS06; DTD07], une distribution Beta conjuguée $\mathcal{B}(\alpha_0, \alpha_1)$ est affectée à l'hyperparamètre inconnu p , qui est un choix naturel pour modéliser un paramètre contraint dans $]0, 1[$

$$\mu(p) = \frac{\Gamma(\alpha_0 + \alpha_1)}{\Gamma(\alpha_0)\Gamma(\alpha_1)} p^{\alpha_1-1} (1-p)^{\alpha_0-1}. \quad (2.162)$$

Une grande variété de formes de distribution peut être obtenue en ajustant les deux hyperparamètres α_0 et α_1 , tout en s'assurant que le paramètre p appartienne à l'ensemble $]0, 1[$, comme requis étant donné que p représente une probabilité [Rob07, App. A]. En particulier, lorsque les hyperparamètres sont sélectionnés comme $\alpha_0 = \alpha_1 = 1$, le prior dans (2.162) se réduit à la distribution uniforme.

A partir de la fonction de vraisemblance (2.158) et des distributions *a priori* et hyper-prior (2.160)–(2.162) introduites ci-dessus, la distribution *a posteriori* jointe s'écrit :

$$\mu(\Theta|z) \propto \mu(z|e, c, \sigma^2) \mu(c|e) \mu(e|p) \mu(p) \mu(\sigma^2) \quad (2.163)$$

avec $\Theta = \{e, c, \sigma^2, p\}$. Une approche MAP consiste à maximiser la distribution *a posteriori* jointe (2.163) ou de façon équivalent à résoudre le problème de minimisation défini dans le Problème 2.

Problème 2 Soit $z = (z_i)_{1 \leq i \leq N} \in \mathbb{R}^N$ et soit $\{\alpha_0, \alpha_1, \sigma_0^2\}$ un ensemble d'hyperparamètres. Le problème de minimisation est le suivant :

$$\begin{aligned} \underset{\Theta = \{e, c, \sigma^2, p\}}{\text{minimiser}} \quad & \frac{1}{2\sigma^2} \sum_{q=1}^{Q_e} \sum_{i \in \mathcal{C}_q} (z_i - c_q)^2 + (Q_e - 1) \left(\log\left(\frac{1-p}{p}\right) + \frac{1}{2} \log(2\pi\sigma_0^2) \right) \\ & + \frac{N}{2} \log(2\pi\sigma^2) - (N-1) \log(1-p) + \log \sigma^2 \\ & - (\alpha_1 - 1) \log p - (\alpha_0 - 1) \log(1-p) + \frac{1}{2\sigma_0^2} \sum_{q=1}^{Q_e} (c_q - c_0)^2 + \frac{1}{2} \log(2\pi\sigma_0^2). \end{aligned}$$

Malgré des différences apparentes de paramétrage entre les problème (2.154) et le Problème 2, nous démontrons dans [FPD⁺17] que les deux sont équivalents pour des choix spécifiques de λ and ϕ rappelés dans la Proposition ci-dessous.

Proposition 15 Pour σ_0^2 suffisamment grand, le problème (2.154) avec

$$\lambda = \sigma^2 \left(\log\left(\frac{1-p}{p}\right) + \frac{1}{2} \log(2\pi\sigma_0^2) \right) \quad (2.164)$$

et

$$\begin{aligned} \phi(\lambda, \sigma^2) = & \frac{N}{2} \log(2\pi\sigma^2) + \log(\sigma^2) \\ & - \frac{\lambda}{\sigma^2} (N + \alpha_0 - 2) + \frac{N + \alpha_0 - 1}{2} \log(2\pi\sigma_0^2) \\ & + (N + \alpha_0 + \alpha_1 - 3) \log \left(1 + \exp \left(\frac{\lambda}{\sigma^2} - \frac{1}{2} \log(2\pi\sigma_0^2) \right) \right) \end{aligned} \quad (2.165)$$

coïncide avec le Problème 2.

L'identification est rendue possible sous la condition que le terme $\frac{1}{2\sigma_0^2} \sum_{q=1}^{Q_e} (c_q - c_0)^2$ qui dépend explicitement de e à travers Q_e puisse être négligé. Par conséquent, choisir σ_0^2 suffisamment grand permet d'écrire

$$\frac{1}{2\sigma_0^2} \sum_{q=1}^{Q_e} (c_q - c_0)^2 \ll \frac{1}{2} \log(2\pi\sigma_0^2), \quad (2.166)$$

Une condition suffisante pour vérifier (2.166) est :

$$\frac{pN}{2} \ll \frac{1}{2} \log(2\pi\sigma_0^2) \leq \log \frac{(1-p)}{p} + \log \zeta \quad (2.167)$$

où $\zeta > 0$ est une constante arbitraire dont la valeur est discutée dans [FPD⁺17]. Le choix de $\log(2\pi\sigma_0^2)$ étant d'une importance cruciale, une discussion approfondie est également menée dans [FPD⁺17].

Algorithme La procédure algorithmique consiste ensuite, pour différentes valeurs de λ ,

- à évaluer la solution résultant du problème L2-Potts : $\hat{u}(z; \lambda) = \underset{u}{\operatorname{argmin}} \frac{1}{2} \|u - z\|_2^2 + \lambda \|Du\|_0$ en utilisant un algorithme rapide reposant sur de la programmation dynamique [WL02];
- à estimer $\hat{\sigma}(\lambda) = \|\hat{u}(z; \lambda) - z\|^2 / (N - 1)$;
- à évaluer le critère dans (2.154) combiné à la Proposition 15 pour $(\hat{u}(z; \lambda), \lambda, \hat{\sigma}(\lambda))$ pour différentes valeurs de λ .

La solution du MAP coïncide alors avec la solution $\hat{u}(z; \lambda)$ qui minimise le critère (2.154) où ϕ est défini par la Proposition 15.

Résultats numériques Les comportements et les performances de la méthode proposée pour débruiter les signaux 1D sont évalués qualitativement et quantitativement. Ils se comparent favorablement à ceux d'une procédure hiérarchique entièrement bayésienne et à d'autres procédures plus standards, à la fois en termes de précision et de temps de calcul. Une figure extraite de [FPD⁺17] résume ces résultats.

2.4.2.1 Conclusions

Nous avons proposé une méthode hybride variationnelle et bayésienne hiérarchique permettant de tirer partie du meilleur de chaque approche : la rapidité d'estimation reposant sur une minimisation variationnelle et le cadre hiérarchique pour le choix des hyperparamètres. La difficulté majeure réside dans l'extension de ce travail au cadre 2-D en particulier l'extension de la notion de point de changements. Une piste envisageable reposerait sur la formulation D-MS présentée dans la section 2.3.2.3 combinée à SLPAM.

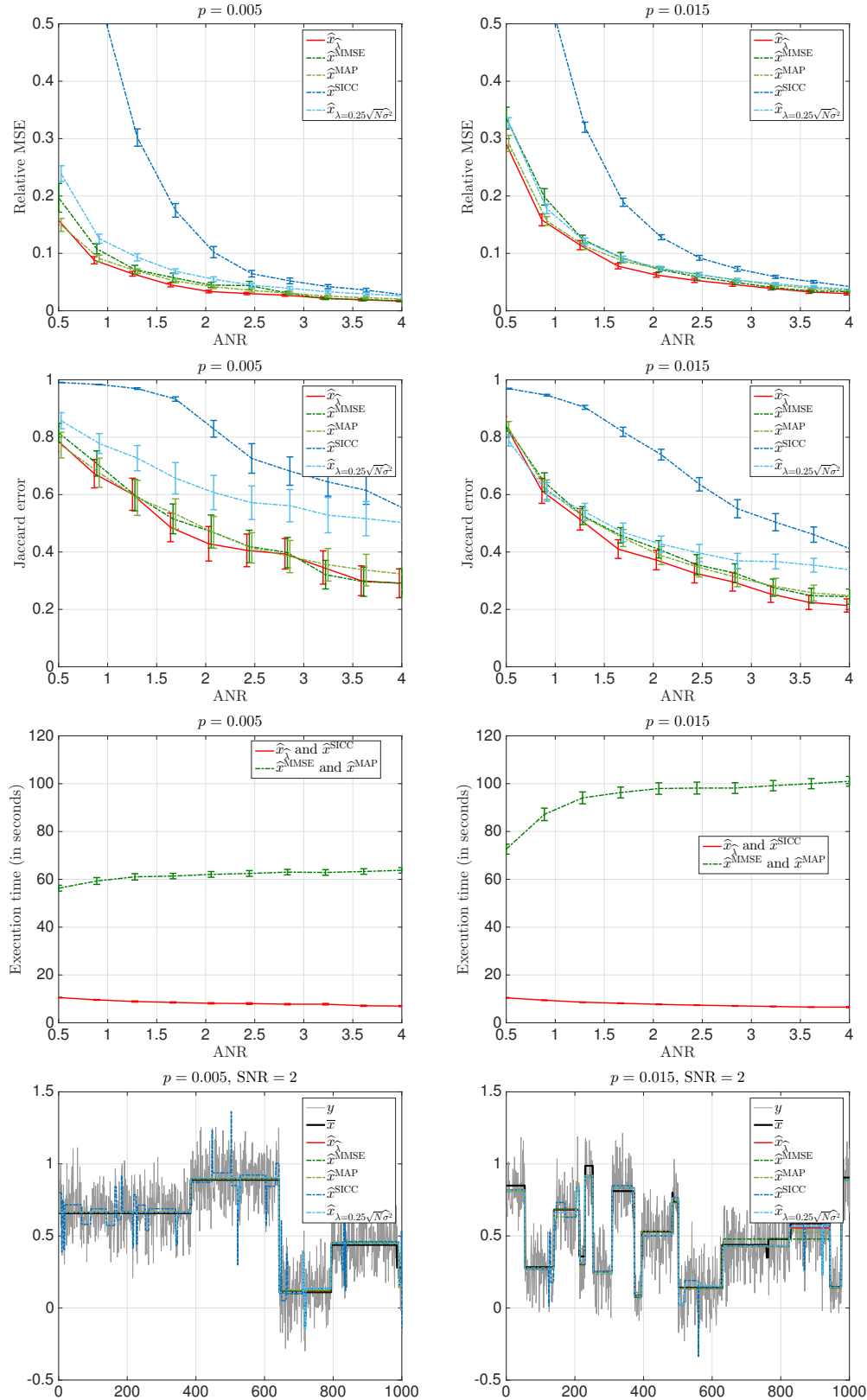


Figure 25 Comparaisons avec les méthodes de l'état-de-l'art. Sur chacune des colonnes sont représentés les résultats obtenus pour différentes valeurs de : $p = 0.005$ et 0.015 (valeur impactant le nombre de discontinuités). (1ère ligne) EQM relative, (2ème ligne) erreur de Jaccard, (3ème ligne) temps d'exécution, (4ème ligne) solutions débruitées. Différentes méthodes sont comparées : L'estimateur proposé (rouge) conduit à des performances comparables à celles obtenues avec des estimateurs bayésien (vert) mais avec des temps de calcul significativement réduits. De plus, notre estimateur conduit à de meilleures performances que l'estimateur SICC (bleu) pour un même temps de calcul. Les résultats obtenus avec $\lambda = 0.25\sqrt{N}\sigma^2$ sont affichés en bleu clair.

2.4.3 SURE

Dans cette section, nous considérons le principe de l'estimateur non-biaisé de Stein (SURE) pour sélectionner le(s) hyperparamètre(s) du problème de minimisation variationnel régularisé (2.153).

Dans [PVP21], nous proposons une généralisation de plusieurs résultats rencontrés dans la littérature dans le contexte de données z corrélées. Nous proposons également une expression de son gradient (i.e. SUGAR [DVPF14]) dans ce contexte afin d'envisager une procédure efficace de recherche des hyperparamètres. Cette approche est ensuite validée sur des données synthétiques et mise à l'épreuve de l'estimation de données réelles sur des données issues d'expériences menées au LPENSL [PPA⁺20] (expérience de collé-glissé et d'écoulement multiphase). L'ensemble de ces contributions a été obtenu dans le cadre de la thèse de Barbara Pascal.

2.4.3.1 Etat-de-l'art

Dans cette section, nous proposons d'estimer l'ensemble de paramètres Λ qui minimise le risque quadratique, c'est à dire :

$$\underset{\Lambda}{\text{minimiser}} \mathbb{E} \|T\hat{u}(z; \Lambda) - T\tilde{u}\|^2. \quad (2.168)$$

avec T un opérateur linéaire, qui permet d'envisager différents types de risques. Par exemple, $T = \Pi$ permet de modéliser un projecteur sur un sous-ensemble de $\mathcal{H}$ [Eld09]. Le *risque quadratique projeté* mesure l'erreur d'estimation sur la quantité projetée $\Pi\tilde{u}$. On retrouve également le risque quadratique habituel lorsque $T = \text{Id}$. Quand $T = A$, on parle de *risque de prédiction*. Ce dernier quantifie la qualité de prédiction $A\hat{u}(z; \Lambda)$ par rapport à l'observation sans bruit $A\tilde{u}$ dans $\mathbb{R}^{\bar{N}}$.

La principale difficulté dans la résolution de (2.168) est que nous n'avons pas accès à la vérité terrain $\tilde{u}$. Dans [Ste81; TW15], Stein a proposé un estimateur de $\mathbb{E} \|A\hat{u}(z; \Lambda) - A\tilde{u}\|^2$ dans le cas d'un modèle linéaire gaussien indépendant, identiquement distribué, (i.i.d.), c'est-à-dire pour des données de la forme $z = A\tilde{u} + \epsilon$ avec $\epsilon \sim \mathcal{N}(0, \Sigma)$ où $\Sigma = \sigma^2 \text{Id}$. Son résultat s'appuie sur une intégration par partie intelligente, qui conduit à l'estimateur du *Risque non biaisé de Stein* (SURE : Stein's Unbiased Risk Estimate) ne faisant plus intervenir $\tilde{u}$:

$$\|A\hat{u}(z; \Lambda) - z\|^2 + 2\sigma^2 \text{Tr}(\partial_z \hat{u}(z; \Lambda)) - \bar{N}\sigma^2, \quad (2.169)$$

dont l'espérance est égale au risque quadratique projeté.

Au cours de ces 15 dernières années, le principe de SURE a été largement utilisé en débruitage voire en restauration de signaux et d'images [BL07; PBBC09]. Son principe a été étendu au cas d'un bruit indépendant mais non distribué de manière identique [CDBBP08; XKB12], correspondant à la matrice de covariance diagonale $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_{\bar{N}}^2)$, et au cas où le bruit est gaussien avec une matrice de covariance générale Σ . Peu d'évaluations numériques sont cependant disponibles pour un bruit gaussien avec une matrice de covariance non scalaire. Une exception notable est [CDBBP08], dans laquelle des expériences numériques sont proposées sur des données multi-composantes non corrélées mais où chaque composante est dégradée par un niveau de bruit différent. Des formulations de SURE dans le cas d'un bruit de Poisson ont été proposées dans [LB08; LMAOM14] et généralisées (G-SURE) pour les familles exponentielles dans [Eld09].

En ce qui concerne l'évaluation pratique de SURE, des outils plus sophistiqués peuvent être nécessaires pour évaluer le second terme de (2.169), notamment lorsque $\hat{u}(z; \Lambda)$ est obtenu à partir d'un algorithme itératif (cf. section 2.1.4) résolvant le problème (2.153). En effet, l'estimateur de Stein fait intervenir la matrice jacobienne de $\hat{u}(z; \Lambda)$ par rapport aux observations z , qui peut ne pas être directement accessible. Afin de gérer ce problème, Vonesch *et al* ont proposé dans [VU08]

d'effectuer une différentiation récursive explicite. Cette approche, bien que partiellement heuristique, s'est avérée efficace pour de nombreux problèmes [DVPF14].

La méthode de résolution la plus directe de (2.168) consiste à calculer SURE (2.169) sur une grille de paramètres, et de sélectionner le(s) paramètre(s) de la grille pour lequel SURE est minimal. Or, les méthodes de recherche par grille souffrent d'un coût de calcul élevé car d'une part la taille de la grille augmente exponentiellement avec le nombre de paramètres de régularisation à estimer et d'autre part chaque point de la grille repose sur un algorithme itératif souvent coûteux en temps de calcul. On peut avoir recours à des stratégies aléatoires pour améliorer l'efficacité de la recherche sur grille [BB12]. Sous l'hypothèse $\Sigma = \sigma^2 \text{Id}$, Deledalle *et al.* [DVPF14] ont élaboré un estimateur du gradient du risque (SUGAR) dans le but de remplacer une recherche exhaustive sur une grille par une stratégie de recherche de minimum reposant sur BFGS.

2.4.3.2 Méthode proposée

Hypothèses Dans [PVP21], nous considérons un modèle spécifique de dégradation $z = A\tilde{u} + \epsilon$ où l'opérateur de dégradation $A \in \mathbb{R}^{\bar{N} \times N}$ est supposé être de rang plein (i.e. A^*A est inversible) et où $\epsilon \sim \mathcal{N}(0, \Sigma)$, où $\Sigma \in \mathbb{R}^{\bar{N} \times \bar{N}}$ est la matrice de covariance du bruit.

Soit $\hat{u}(z; \Lambda)$ un estimateur paramétrique de la vérité terrain $\tilde{u}$, défini de manière unique à partir des observations z et des hyperparamètres Λ , par exemple provenant de (2.153). On suppose que la quantité d'intérêt peut être une projection de u sur le sous-espace $\mathcal{I}$ de $\mathcal{H} = \mathbb{R}^N$ (cf. (2.136) où seule la vérité terrain relative à la régularité locale h est accessible). L'opérateur linéaire $\Pi : \mathcal{H} \rightarrow \mathcal{I}$ effectue la projection orthogonale sur le sous-espace $\mathcal{I}$ capturant les informations pertinentes sur u . De plus, nous définissons l'opérateur linéaire $H : \mathbb{R}^{\bar{N}} \rightarrow \mathcal{I}$ comme

$$H = \Pi (A^*A)^{-1} A^*. \quad (2.170)$$

Formulation du risque Dans notre étude, le risque est défini comme l'erreur d'estimation projetée :

$$R[\hat{u}](\Lambda) = \mathbb{E}_\epsilon \|\Pi\hat{u}(z; \Lambda) - \Pi\tilde{u}\|_2^2. \quad (2.171)$$

Le Théorème 1 proposé dans [PVP21] fournit une généralisation de l'estimateur original de Stein (2.169) et constitue le point de départ de notre travail.

Théorème 1 [PVP21] *Considérons le modèle de dégradation $z = A\tilde{u} + \epsilon$ avec l'hypothèse de gaussianité du bruit, A de rang plein et l'opérateur H défini par (2.170). De plus, on suppose que l'estimateur $\hat{u}(z; \Lambda)$ est continu et faiblement dérivable par rapport aux observations z et que les quantités $\langle H^* \Pi\hat{u}(z; \Lambda), \epsilon \rangle$ et $\partial_z \hat{u}(z; \Lambda)$ sont intégrables par rapport à la densité gaussienne :*

$$\frac{1}{\sqrt{(2\pi)^{\bar{N}} |\det \Sigma|}} \exp\left(-\frac{\|\epsilon\|_{\Sigma^{-1}}^2}{2}\right) d\epsilon.$$

Le lemme de Stein généralisé s'applique et conduit à

$$R[\hat{u}](\Lambda) = \mathbb{E}_\epsilon \left[\|H(A\hat{u}(z; \Lambda) - z)\|_2^2 + 2\text{Tr}(\Sigma H^* \Pi \partial_z \hat{u}(z; \Lambda)) - \text{Tr}(H \Sigma H^*) \right]. \quad (2.172)$$

Notons qu'une autre définition usuelle du risque implique l'inverse de la matrice de covariance [Eld09] via la distance de Mahalanobis :

$$R_M[\hat{u}](\Lambda) \triangleq \mathbb{E}_\epsilon \|\Pi\hat{u}(z; \Lambda) - \Pi\tilde{u}\|_{\Sigma^{-1}}. \quad (2.173)$$

nécessitant la connaissance de Σ^{-1} , qui peut être non triviale. Dans ce cas, la reformulation a la même structure globale, mais au lieu d'impliquer la matrice de covariance Σ elle requiert son inverse :

$$R_M[\hat{u}](\Lambda) = E_\epsilon \left[\|H(A\hat{u}(z; \Lambda) - z)\|_{\Sigma^{-1}}^2 + 2\text{Tr}(H^* \Pi \partial_z \hat{u}(z; \Lambda)) - \text{Tr}(HH^*) \right]. \quad (2.174)$$

FDMC SURE Dans le Théorème 1, la quantité $\text{Tr}(\Sigma H^* \Pi \partial_z \hat{u}(z; \Lambda))$ fait apparaître des difficultés majeures de mise en oeuvre, comme en témoigne la littérature prolifique traitant de cette question dans le cas d'un bruit non corrélé [Kat09; TT12; DKF⁺13; VDF⁺17]. En effet, il s'agit du produit d'une matrice $\Sigma H^* \Pi$ avec la matrice jacobienne $\partial_z \hat{u}(z; \Lambda)$. Le calcul de $\partial_z \hat{u}(z; \Lambda)$ est effectué par différences finies et une stratégie de Monte Carlo est mise en oeuvre pour proposer une estimation de la trace. Sous un certain nombre d'hypothèses, nous pouvons construire un estimateur du risque non biaisé de Stein utilisable en pratique, noté $\hat{R}_{v,\epsilon}(z; \Lambda|\Sigma)$, défini dans le Théorème suivant.

Théorème 2 [PVP21] *Considérant les mêmes hypothèses que celles énoncées dans le Théorème 1 combinées à une hypothèse de Lipschitzianité par rapport à z . Finite Difference Monte Carlo (FDMC) GSURE s'exprime :*

$$\hat{R}_{v,\epsilon}(z; \Lambda|\Sigma) \triangleq \|H(A\hat{u}(z; \Lambda) - z)\|_2^2 + \frac{2}{v} \langle H^* \Pi (\hat{u}(z + v\epsilon; \Lambda) - \hat{u}(z; \Lambda)), \Sigma \epsilon \rangle - \text{Tr}(H\Sigma H^*),$$

qui est un estimateur asymptotique sans biais du risque $R[\hat{u}](\Lambda)$ quand $v \rightarrow 0$, i.e.

$$\lim_{v \rightarrow 0} E_{\epsilon,\epsilon} \hat{R}_{v,\epsilon}(z; \Lambda|\Sigma) = R[\hat{u}](\Lambda). \quad (2.175)$$

Estimateur du gradient du risque sans biais de Stein (SUGAR) De l'estimateur du risque $\hat{R}_{v,\epsilon}(z; \Lambda|\Sigma)$ fourni dans le Théorème 2, une approche par recherche sur une grille pourrait être effectuée, afin d'estimer $\Lambda^\dagger$, tel que

$$\Lambda^\dagger \in \underset{\Lambda}{\text{Argmin}} \hat{R}_{v,\epsilon}(z; \Lambda|\Sigma). \quad (2.176)$$

Comme discuté précédemment, cette stratégie est rarement efficace. En nous inspirant de [DVPF14], nous avons proposé dans [PVP21] un estimateur généralisé du gradient du risque par rapport aux hyperparamètre i.e. $\partial_\Lambda \hat{R}_{v,\epsilon}(z; \Lambda|\Sigma)$. De plus, nous avons démontré que l'estimateur $\partial_\Lambda \hat{R}_{v,\epsilon}(z; \Lambda|\Sigma)$ est un estimateur asymptotiquement sans biais du gradient du risque $R[\hat{u}](\Lambda)$ (cf. [PVP21]). L'algorithme complet fournissant à la fois $\hat{u}(z; \Lambda)$ et sa jacobienne $\partial_\Lambda \hat{u}(z; \Lambda)$ par rapport à Λ est détaillé dans [PVP21] permettant ainsi d'effectuer une estimation de $\Lambda^\dagger$ reposant sur une méthode de descente de gradient, par exemple BFGS [PVP21].

2.4.3.3 Evaluation numérique

Nous avons mis en oeuvre la stratégie d'estimation proposée dans la section précédente reposant sur SURE combiné à SUGAR dans divers problèmes de traitement du signal et des images : débruitage linéaire par morceaux et application à l'analyse de signaux de collé-glissé (*stick-slip*), la segmentation de texture via l'approche présentée dans la section 2.3.3 sur des données synthétiques puis réelles venant de données d'écoulements multiphasés.

Débruitage de signaux de *stick-slip* Les expériences de frottement visent à sonder non seulement les caractéristiques des matériaux, mais aussi la dynamique des systèmes impliquant des surfaces en contact. En particulier, ces approches permettent de modéliser et de tenter de prédire la dynamique sismique [Mar98]. L'expérience classique de frottement solide consiste à tracter une masse m sur un substrat via un ressort de raideur k tiré à la vitesse V (voir par exemple Figure 2 dans [CPO⁺19]). Le signal représentatif de la dynamique de ce système est la force mesurée au

point de contact entre le ressort et le curseur. Parmi les différents régimes décrits dans le frottement solide, on peut distinguer le stick-slip, caractérisé par un signal en forme de dent alternant montée lente, linéaire et descente rapide, le régime inertiel, dans lequel le signal devient périodique et ressemble à une courbe sinusoïdale, et le régime de glissement continu, caractérisé par un signal presque constant superposé de bruit [BC06]. L'apparition d'un fluage peut également modifier la forme du signal. Le défi de telles études est d'établir un *diagramme de phase* décrivant (et donc prédisant) la dynamique du système en fonction des paramètres (m, k, V) . Si ce diagramme est aisé à construire pour les grandes masses m , les expériences avec une faible pression de confinement, nécessaires pour éviter l'usure de surface, sont difficiles car elles conduisent à des signaux fortement dégradés [CPO⁺19].

Dans ce contexte, de nouveaux outils de traitement du signal sont nécessaires. Dans [CPO⁺19], nous avons proposé d'effectuer le débruitage du signal en modélisant les signaux stick-slip comme des signaux linéaires par morceaux conduisant au problème de minimisation suivant :

$$\hat{u}(z; \Lambda) \in \underset{u \in \mathcal{H}}{\text{Argmin}} \frac{1}{2} \|u - z\|_2^2 + \lambda \|Du\|_1 \quad (2.177)$$

où z correspond aux signaux de force mesurés et

$$(Du)_n = \lambda (u_{n+1} - 2u_n + u_{n-1}). \quad (2.178)$$

qui est un cas particulier de (2.153). La minimisation de ce critère est effectuée par des itérations de l'algorithme Chambolle-Pock exploitant la forte convexité.

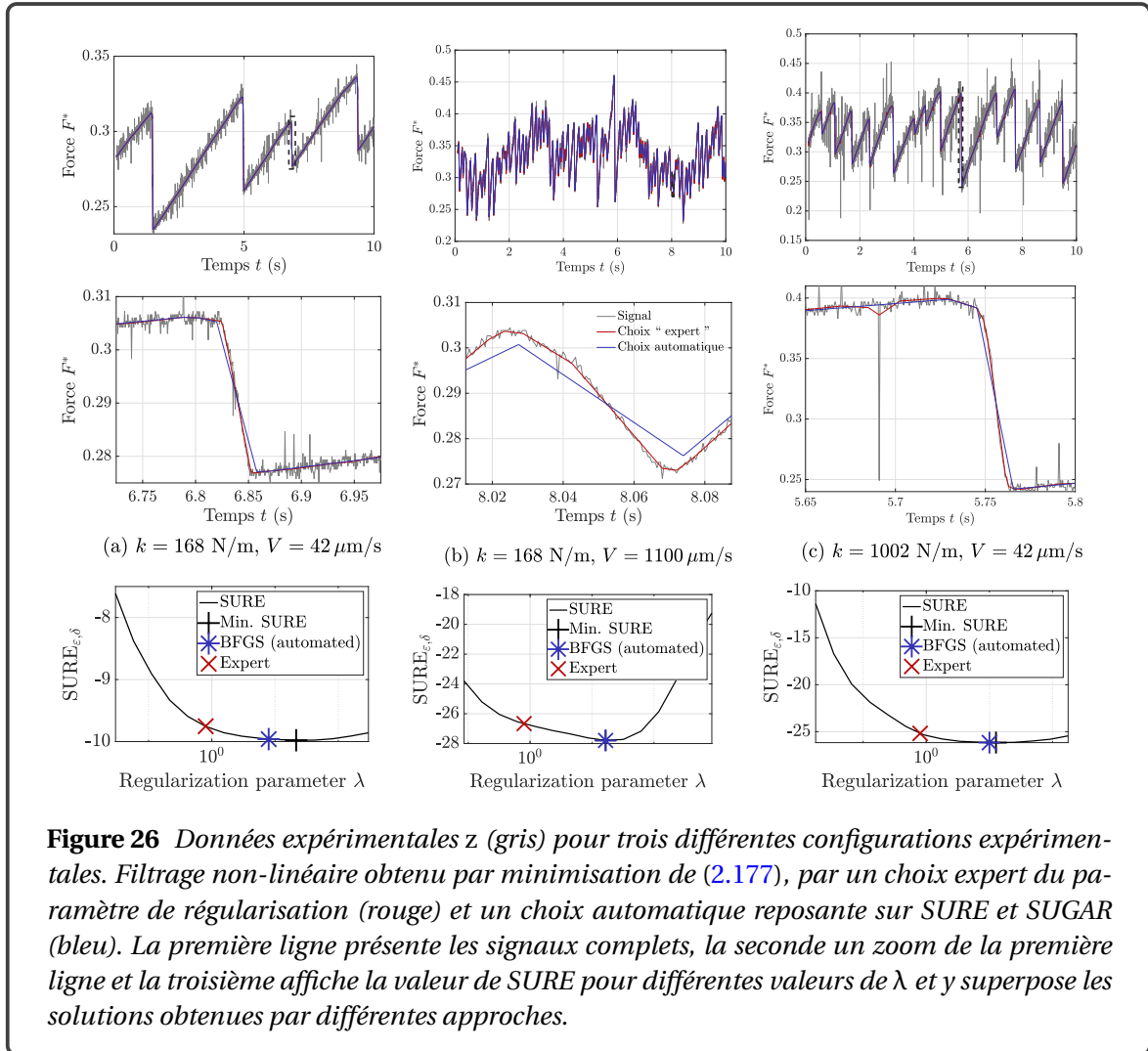


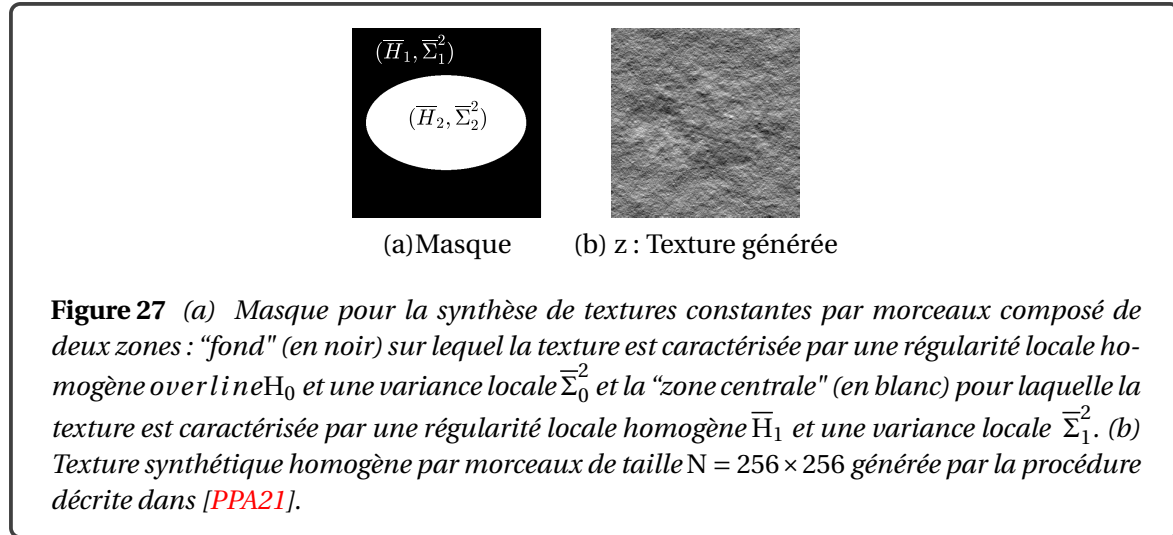
Figure 26 Données expérimentales z (gris) pour trois différentes configurations expérimentales. Filtrage non-linéaire obtenu par minimisation de (2.177), par un choix expert du paramètre de régularisation (rouge) et un choix automatique reposant sur SURE et SUGAR (bleu). La première ligne présente les signaux complets, la seconde un zoom de la première ligne et la troisième affiche la valeur de SURE pour différentes valeurs de λ et y superpose les solutions obtenues par différentes approches.

La tâche fastidieuse de réglage du paramètre de régularisation λ a d'abord été réalisée par inspection visuelle d'experts et a conduit à un choix $\lambda_{\text{expert}} = 0,8$ appliqué uniformément à tous les signaux, quels que soient les différents paramètres d'expérimentation. Des exemples d'observations bruitées sont montrés dans la Figure 26 (gris), avec des estimations linéaires par morceaux associées obtenues avec λ_{expert} affiché en rouge. Les signaux régularisés semblent bien saisir la transition entre les régimes de “collé” et “glissé”

Nous proposons également d'illustrer l'utilisation de la stratégie de réglage automatique des paramètres de régularisation présentée dans la section précédente dans ce contexte de débruitage où $\hat{R}_{v,\epsilon}(z; \lambda | \Sigma)$ est utilisé comme critère de qualité et minimisé par rapport à λ . Notons que ce cas simple, se ramène à [DVPF14]. La recherche sur une grille et le réglage automatisé sont implémentés et comparés. Les résultats obtenus par choix automatique sont présentés sur la Figure 26 (bleu). Ce résultat est satisfaisant car le réglage automatique permet d'atteindre un réglage proche de celui fait par l'utilisateur mais sans aucune information *a priori*.

Segmentation de texture sur données synthétiques Le formalisme de sélection automatique présenté précédemment a été mis en oeuvre dans le contexte de la segmentation de texture présentée Section 2.3.3. Dans ce contexte, l'opérateur A défini par (2.137), est bien un opérateur de rang plein et l'estimateur $\hat{u}(z, \Lambda) = \{\hat{h}, \hat{v}\}(\mathcal{L}, (\lambda, \alpha))$ est obtenu par minimisation de (2.136) qui n'est qu'un cas particulier de (2.153). Nous montrons dans [PVP21], que cet estimateur vérifie l'hypothèse de Lipschitzianité par rapport aux observations. La question de la Lipschitzianité par rapport à Λ reste quant à elle une question ouverte.

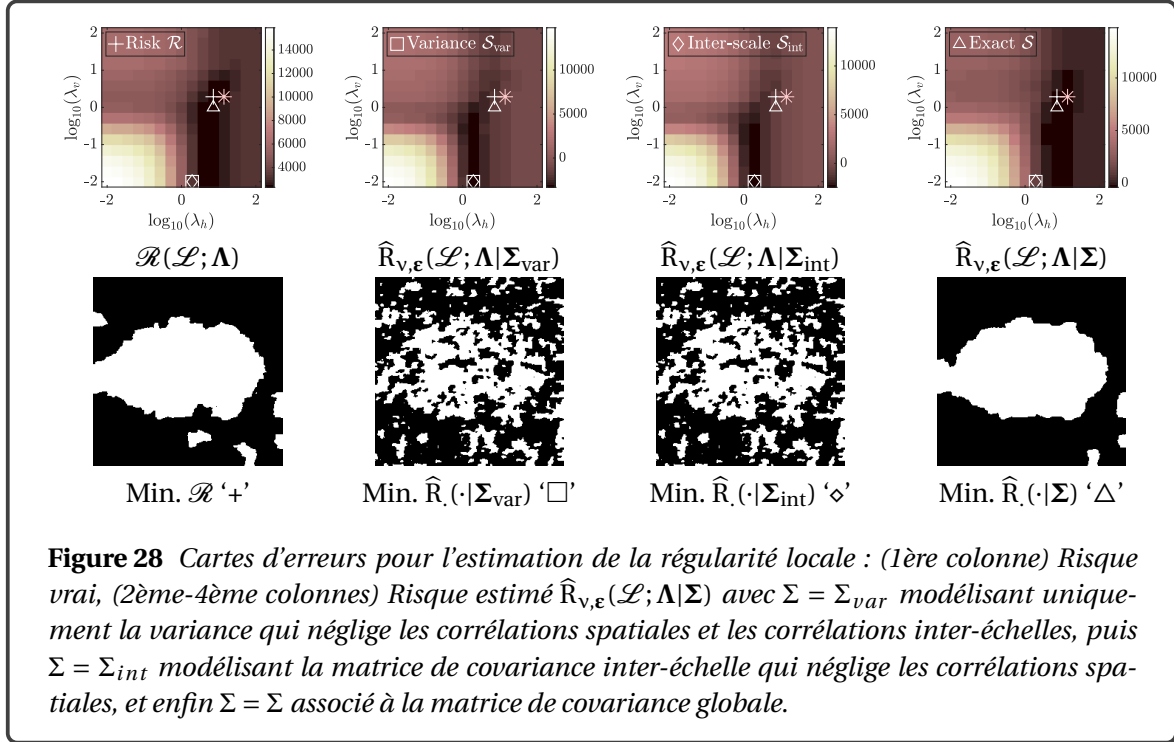
Dans [PVP21], nous proposons d'évaluer les performances d'estimation pour deux textures “facile” et “difficile”. Dans ce manuscrit, je me contente de reporter les résultats obtenus dans le contexte “difficile”. L'image synthétiques est composée de deux zones : un fond de régularité locale $H_1 \equiv 0.5$ et de variance locale $\Sigma_1^2 = 0.6\text{Id}$ et d'une ellipse centrale de régularité locale $H_2 = 0.75$ et de variance locale $\Sigma_2^2 = 0.7\text{Id}$. L'image synthétique est générée grâce à un processus proposé dans le cadre de la thèse de Barbara Pascal détaillé dans [PPA21] et elle est affichée dans la Figure 27.



La procédure de sélection automatique de (λ, α) est mise en oeuvre et repose sur une procédure d'estimation de la matrice de covariance des log-leaders. Nous comparons sur la Figure 28, les résultats lorsque le risque est calculé avec Σ_{var} modélisant uniquement la variance qui néglige les corrélations spatiales et les corrélations inter-échelles, puis Σ_{int} modélisant la matrice de covariance inter-échelle qui néglige les corrélations spatiales, et enfin Σ associée à la matrice de covariance globale. On observe que la prise en compte de la matrice de covariance globale est nécessaire pour fournir une estimée précise.

Dans [PVP21], des résultats supplémentaires sont présentés. Nous y comparons les résultats obtenus avec la vraie matrice de covariance Σ et une matrice Σ estimée et également les perfor-

mances d'une recherche automatique reposant sur BFGS et donc l'utilisation de SUGAR versus une recherche sur une grille. Nous observons que les résultats sont très similaires entre une estimation reposante sur Σ vrai ou Σ estimé. Les performances peuvent également être similaires entre une recherche sur une grille et une recherche par BFGS mais au prix d'un temps de calcul plus important. En effet, une recherche du couple $\{\lambda, \alpha\}$ optimal reposant sur BFGS ne demande que 40 appel alors que pour obtenir des performances équivalentes avec une recherche sur une grille 225 appels sont nécessaires.



Segmentation de texture sur données issues d'expériences d'écoulements multiphasés Comprendre et prédire la dynamique des écoulements multiphasés est un enjeu majeur en géosciences (décontamination des sols, séquestration de CO_2) et dans l'industrie (récupération assistée du pétrole, catalyse hétérogène). Parmi ces procédés, beaucoup impliquent un écoulement conjoint de gaz et de liquide à travers un milieu poreux. La quantification des zones de contact entre les différentes phases, où se déroulent les réactions chimiques, est d'une importance capitale pour analyser et prédire l'efficacité de tels processus. Cependant, même lorsque la visualisation directe est possible, le milieu poreux génère une texture globale et multi-échelle sur les images qui rend difficile l'extraction des interfaces gaz-liquide. Les techniques de segmentation basées sur des outils morphologiques utilisés jusqu'à présent pour différencier les phases dans les écoulements multiphasiques présentent de sérieuses limitations : réglage arbitraire de seuils, contours de bulles irréguliers et non physique, mauvaise détection de petites bulles. De plus, les développements récents en imagerie haute résolution et haute vitesse produisent des images de grande taille et des ensembles de données volumineux, amenant ainsi à des problèmes de mémoire et de coûts de calcul.

Des expériences d'écoulements gaz-liquide à travers un milieu poreux ont été réalisées dans une cellule de Hele-Shaw verticale quasi-2D. Plus de détails sont disponibles dans [MSV16]. Un exemple est fourni dans la Figure 29(a), montrant que la phase gazeuse (structures gris foncé ou blanches) est texturée en raison de la présence de la mousse. La phase liquide (en gris) est également texturée mais à des échelles plus petites, comme on peut l'observer sur la Figure 29(a).

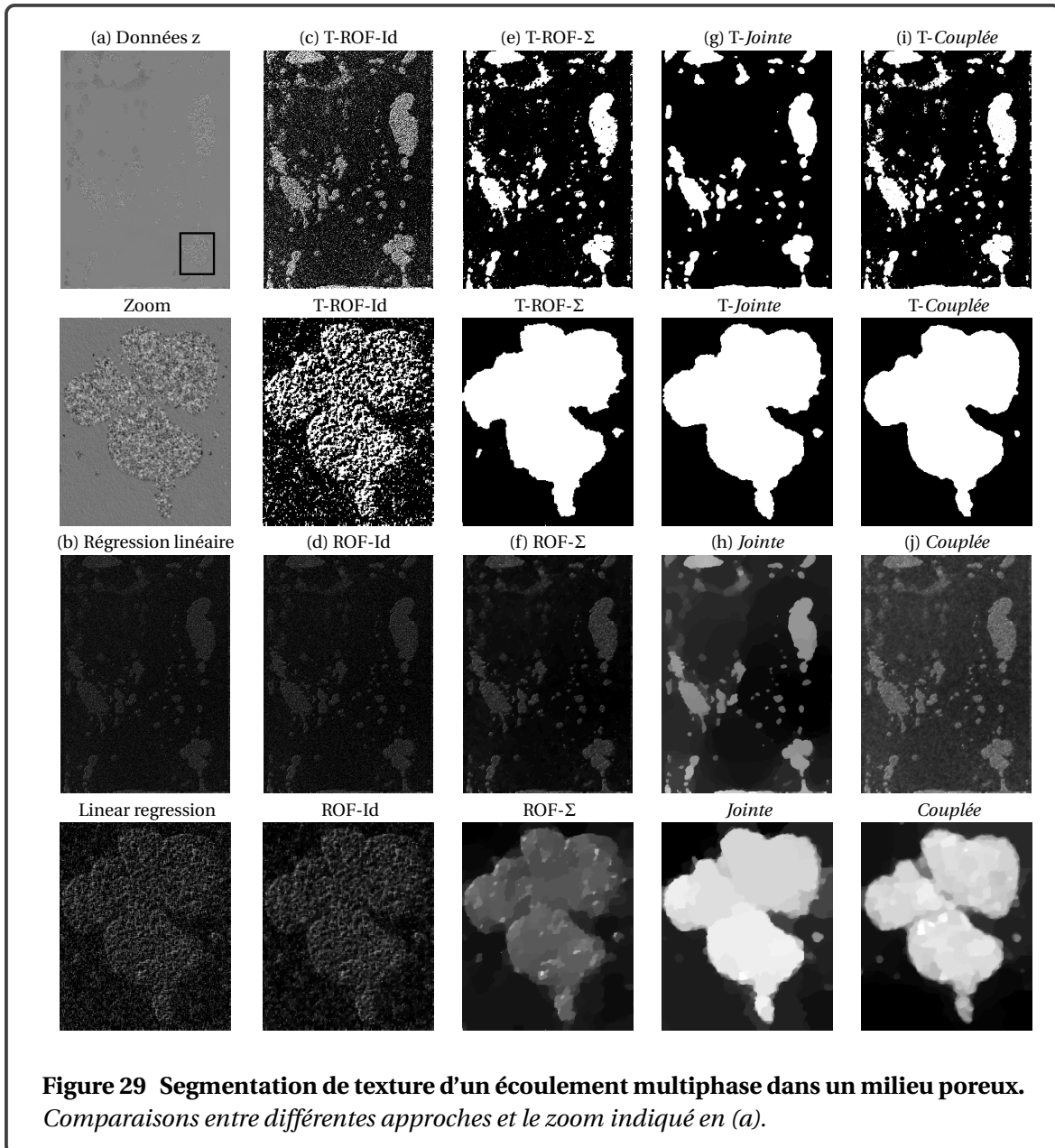


Figure 29 Segmentation de texture d'un écoulement multiphase dans un milieu poreux.
Comparaisons entre différentes approches et le zoom indiqué en (a).

Pour tous les ensembles de données expérimentales, 50 à 3000 images sont enregistrées. Un grand nombre de jeux de données associés à différents paramètres (ici les débits de gaz et de liquide) sont investigués, pour analyser les différents régimes hydrodynamiques.

Quatre procédures relevant du modèle (2.153) ont été comparées dans [PPA⁺20]. Il est à noter que ces modèles diffèrent à la fois par le critère minimisé et le modèle de bruit, qui est d'une importance cruciale dans les procédures de Stein.

La première, notée ROF-Id repose sur la résolution de (2.134) où l'on applique la procédure SURE-SUGAR en supposant que $\Sigma = \sigma^2 \text{Id}$. Les procédures ROF- Σ , reposent sur la stratégie de segmentation de texture (2.136) ("couplée") ou une variation appliquant TV sur h et v séparément (cf. *Jointe*) et prennent en compte la structure de covariance des coefficients *log-leaders*, mettant en évidence à la fois des corrélations inter-échelles et spatiales encapsulées dans une matrice de covariance non diagonale Σ .

On observe que les procédures *Jointe* et *Couplée*, prenant en compte à la fois la régularité locale et la variance locale, donnent des contours plus réguliers. De plus, les méthodes *Jointe* et *Couplée* détectent moins d'artefacts (voir Figures 29(g) et 29(i)). La procédure *Couplée* s'avère réaliser un compromis satisfaisant, évitant les artefacts, tout en détectant de petites bulles de gaz, comme illustré sur les Figures 29(i) et 29(j).

2.4.3.4 Conclusions

SURE combiné à SUGAR offre une stratégie d'estimation d'hyperparamètres très efficace dont la formulation étendue, permettant de prendre en compte une matrice de covariance, a montré son efficacité dans le contexte de segmentation de texture qui reste un contexte très spécifique où l'opérateur A est de rang plein.

Dans un contexte plus général, comme pour l'analyse des environnements circumstellaires (cf. section 2.2.4), où $A \neq \text{Id}$ et donc une erreur en prédiction doit être formulée, les performances ne sont pas aussi bonnes. Cette stratégie permet néanmoins d'obtenir un ordre de grandeur raisonnable pour le choix des hyperparamètres. Une étude est actuellement en cours pour estimer les paramètres (β, χ) dans D-MS (cf. (2.124)).

2.4.4 Algorithmes déroulés

2.4.4.1 Motivation

Une alternative récente à la formulation variationnelle et à l'optimisation non lisse repose sur un apprentissage supervisé basé sur des réseaux de neurones (voir [LIMK18; RYF20] pour des articles de synthèse). De nombreuses contributions sont dédiées à ce sujet mais le point commun est d'apprendre une fonction de prédiction d_Θ reposant sur un ensemble de paramètres Θ à partir de l'ensemble de données d'apprentissage $\mathcal{S} = \{(z_s, \tilde{u}_s) \in \mathbb{R}^{\bar{N}} \times \mathbb{R}^N | s = 1, \dots, S\}$ composée de S couples d'images originales $\tilde{u}_s$ et images dégradées z_s en minimisant une fonction de perte empirique de la forme :

$$\mathcal{C}(\Theta) := \frac{1}{S} \sum_{s=1}^S \|d_\Theta(g_z(z_s)) - g_u(\tilde{u}_s)\|^2. \quad (2.179)$$

L'approche la plus simple consiste à considérer d_Θ comme un CNN [MSY16; UVL20; XRLJ14], conduisant à $g_z(z) = z$ et $g_u(u) = u$. Des performances améliorées peuvent être obtenues lorsque l'apprentissage est effectué sur des coefficients d'ondelettes ou de trame [KMY17; YHC18; BKL⁺19] (conduisant à g_z et g_x associé à la transformation de trame), et sur les données rétroprojetées, c'est-à-dire $g_z(z) = A^\dagger z$ (ou aussi une alternative reposant sur $g_z(z) = A^* z$) [JMFU17]. Plus récemment, la conception de d_Θ s'appuie sur les connaissances construites depuis de nombreuses années dans les problèmes inverses, par exemple en tronquant une série de Neumann [GOW19], ou en déployant des méthodes itératives telles que le schéma de descente de gradient dans [RZZ⁺19], des itérations ISTA proposées dans les travaux pionniers de Gregor et LeCun [GLC10], des itérations de point intérieur proximales comme dans [BCC⁺20], et plus récemment en déployant un algorithme proximal primal-dual [AO18] où les opérateurs proximaux sont remplacés par des CNN. La dernière classe d'approches, notamment [BCC⁺20], offre un cadre particulièrement adapté à l'analyse de stabilité [CP20].

2.4.4.2 Contributions

Dans [JP21], nous avons proposé un réseau d'apprentissage profond dédié à la restauration d'images lorsque la base de données provient du modèle de dégradation suivant $z = A\tilde{u} + \epsilon$. Les couches de notre réseau reposent sur itérations déroulées d'une instance spécifique de l'algorithme Condat-Vũ dans un contexte de minimisation de (2.153) avec $\|\cdot\|_\Lambda = \lambda h(\cdot)$ où h est une fonction convexe, semi-continue inférieurement et propre. Cette reformulation nous permet d'apprendre à la fois les paramètres de l'algorithme σ, γ ainsi que le paramètre de régularisation λ et l'opérateur linéaire D (en tant qu'entité unique $H = \lambda D$) pour un nombre fixe de couches K , conduisant au réseau DeepPDNet. Une procédure de rétropropagation par rapport aux paramètres d'intérêt combiné à un gradient stochastique sont mis en oeuvre.

Algorithme Condat-Vũ déroulé Le design de notre réseau s'appuie sur une reformulation de (2.153) qui prend la forme :

$$\hat{u}(z; \Lambda) \in \underset{u \in \mathbb{R}^N}{\operatorname{Argmin}} \frac{1}{2} \|Au - z\|_2^2 + h(Hu). \quad (2.180)$$

Afin d'estimer $\hat{u}(z; \Lambda)$ sans hypothèses spécifiques sur A et H , l'algorithme le plus flexible et intelligible de la littérature est certainement l'algorithme Condat-Vũ [Con13b], dont les itérations spécifiques à la résolution du Problème (2.180) sont données dans Algorithme 4.

Algorithme 4 Algorithme Condat-Vũ pour résoudre le Problème (2.180).

Soit $\tau > 0$, $\gamma > 0$, tel que $\frac{1}{\tau} - \gamma \|H\|^2 > \frac{\|A\|^2}{2}$

Initialisation : $(u^{[0]}, w^{[0]}) \in \mathbb{R}^N \times \mathbb{R}^{|\mathcal{E}|}$ Pour $k = 0, 1, \dots$

$$\begin{cases} u^{[k+1]} = u^{[k]} - \tau A^* (Au^{[k]} - z) - \tau H^* w^{[k]} \\ w^{[k+1]} = \operatorname{prox}_{\gamma h^*}(w^{[k]} + \gamma H(2u^{[k+1]} - u^{[k]})) \end{cases}$$

La convergence de la suite $(u^{[k+1]})_{k \in \mathbb{N}}$, générée par l'Algorithme 4, vers un minimiseur de (2.180) est assurée sous une hypothèse technique faisant intervenir les paramètres de pas τ , γ , mais également $\|A\|^2$ et $\|H\|^2$ donnés dans l'algorithme.

DeepPDNet Les réseaux profonds sont composés d'un empilement de couches. Chaque couche est composée d'une transformation linéaire (par exemple, des convolutions) et une fonction d'activation non linéaire. Formellement, un réseau composé de K couches peut être écrit comme

$$u^{[K]} = \eta^{[K]}(T^{[K]} \dots \eta^{[1]}(T^{[1]} u^{[1]} + b^{[1]}) \dots + b^{[K]}) \quad (2.181)$$

où, pour chaque $k \in \{1, \dots, K\}$, $T^{[k]}$ est la matrice de poids, $b^{[k]}$ est le biais, et $\eta^{[k]}$ est la fonction d'activation non linéaire. Dans le cadre d'apprentissage en profondeur classique (par exemple CNNs), $T^{[k]}$ est considéré comme l'opération de convolution avec une collection de filtres de petite taille et chaque filtre est associé à un biais $b^{[k]}$, $\eta^{[k]}$ est une fonction d'activation non linéaire telle que tanh, sigmoid ou la fonction ReLU, suivie d'une couche de pooling (par exemple max pooling, average pooling, etc.) qui a pour effet d'augmenter les champs récepteurs locaux et de diminuer également le nombre de paramètres du réseau.

La proposition 16 présente comment déplier l'instance spécifique de l'algorithme primal-dual Condat-Vũ décrit dans Algorithme 4 dans un réseau à plusieurs couches. Ce réseau effectue une restauration d'image non supervisée lorsque le nombre de couches $K \rightarrow +\infty$ et il converge vers la solution de l'Eq. (2.180). En pratique, K doit être défini suffisamment grand et sa valeur est soit définie manuellement (par exemple $K > 10^5$) soit basée sur un critère d'arrêt (par exemple basé sur le résidu des itérées ou sur le gap de dualité).

Proposition 16 [JP21] L'algorithme 4 coïncide avec le réseau (2.181) lorsque, pour tout $k \in \{1, \dots, K\}$, $T^{[k]} \in \mathbb{R}^{(N+|\mathcal{E}|) \times (\bar{N}+|\mathcal{E}|)}$, $b^{[k]} \in \mathbb{R}^{\bar{N}+|\mathcal{E}|}$ and $\eta^{[k]} : \mathbb{R}^{\bar{N}+|\mathcal{E}|} \rightarrow \mathbb{R}^{\bar{N}+|\mathcal{E}|}$ tel que

$$\begin{cases} T^{[k]} = \begin{pmatrix} \operatorname{Id} - \tau A^* A & -\tau H^* \\ \gamma H(\operatorname{Id} - 2\tau A^* A) & \operatorname{Id} - 2\tau \gamma H H^* \end{pmatrix} \\ b^{[k]} = \begin{pmatrix} \tau A^* z \\ 2\tau \gamma H A^* z \end{pmatrix} \\ \eta^{[k]} = \begin{pmatrix} \operatorname{Id} \\ \operatorname{prox}_{\gamma h^*} \end{pmatrix} \end{cases} \quad (2.182)$$

Lorsqu'il est reformulé dans un cadre d'apprentissage supervisé, un nombre extrêmement important de K devient peu pratique et un réseau profond avec une profondeur moyenne (par

exemple dix ou cent couches), qui pourrait être interprété comme un arrêt précoce, ne fournirait pas une si bonne estimation. Nous apportons donc plusieurs modifications au réseau présenté dans la Proposition 16 afin de l'adapter à un formalisme d'apprentissage profond.

Réseau DeepPDNet Dans un contexte de restauration où A est connu, les degrés de liberté dans le réseau proposé (cf. Proposition 16) ne peuvent provenir que de γ , τ , et H . La stratégie d'apprentissage supervisé proposée, nommée DeepPDNet, nous permet d'apprendre les paramètres $\hat{\Theta} = \{\hat{\gamma}^{[k]}, \hat{\tau}^{[k]}, \hat{H}^{[k]}\}_{1 \leq k \leq K}$ solutions de :

$$\underset{\Theta}{\text{minimize}} \mathcal{C}(\Theta) := \frac{1}{S} \sum_{s=1}^S \|\bar{x}_s - d_{\Theta}(z_s)\|_2^2 \quad (2.183)$$

où la fonction de prédiction est

$$d_{\Theta}(z_s) = \eta^{[K]} \left(T^{[K]} (\dots \eta^{[1]} (T^{[1]} u_s^{[1]} + b^{[1]}) \dots + b^{[K]} \right)$$

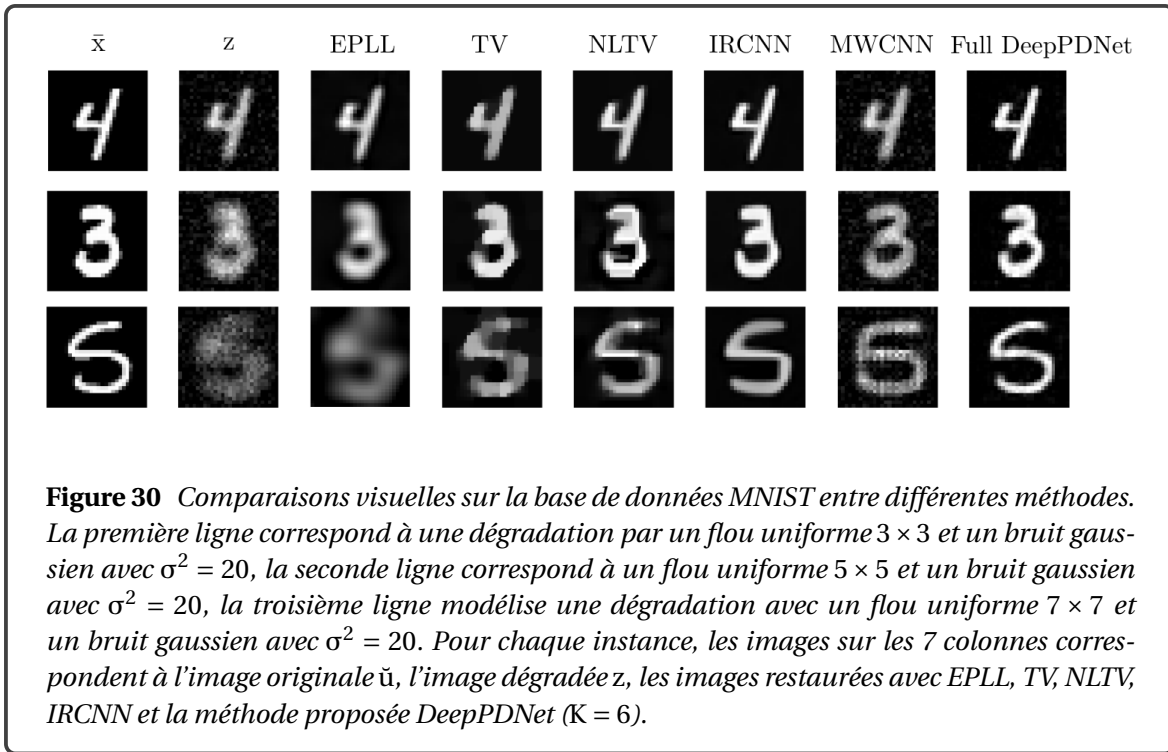
avec

$$\left\{ \begin{array}{l} u_s^{[1]} = A^* z_s \\ T^{[1]} = \begin{pmatrix} \text{Id} - \tau^{[1]} A^* A \\ \sigma^{[1]} H^{[1]} (\text{Id} - 2\tau^{[1]} A^* A) \end{pmatrix} \\ T^{[k]} = \begin{pmatrix} \text{Id} - \tau^{[k]} A^* A & -\tau^{[k]} (H^{[k]})^* \\ \sigma^{[k]} H^{[k]} (\text{Id} - 2\tau^{[k]} A^* A) & \text{Id} - 2\tau^{[k]} \gamma^{[k]} H^{[k]} (H^{[k]})^* \end{pmatrix} \\ b^{[k]} = \begin{pmatrix} \tau^{[k]} A^* z \\ 2\tau^{[k]} \gamma^{[k]} H^{[k]} A^* z_s \end{pmatrix} \\ \eta^{[k]} = \begin{pmatrix} \text{Id} \\ \text{prox}_{\gamma^{[k]} h^*} \end{pmatrix} \\ T^{[K]} = \begin{pmatrix} \text{Id} - \tau^{[K]} A^* A & \tau^{[K]} (H^{[K]})^* \end{pmatrix} \\ b^{[K]} = \tau^{[K]} A^* z_s \\ \eta^{[K]} = \text{Id}. \end{array} \right. \quad (2.184)$$

2.4.4.3 Evaluation expérimentale

Afin d'évaluer les performances du réseau proposé, nous avons tout d'abord considéré le problème de déconvolution sur la base de données MNIST puis sur la base de données BSD68 dans un contexte de restauration et de super-résolution. Nous y avons présenté :

- l'impact du nombre de couche. Il apparaît que plus les images sont complexes plus le nombre de couches doit être grand. A titre d'exemple 6 couches sont suffisantes pour MNIST alors que 20 couches sont nécessaires pour la base de données BSD69.
- l'impact du choix de $H \in \mathbb{R}^{|\mathbb{E}| \times N}$ comme étant une matrice dense ou parcimonieuse (reproduisant l'idée de filtrage). Le choix parcimonieux permet d'atteindre les meilleures performances pour une valeur de $|\mathbb{E}|$ identique.
- des comparaisons avec les méthodes de l'état-de-l'art : classiques comme EPLL, TV, NLTV (reposant sur une minimisation de la forme (2.180) avec $H = \lambda D$ et où le paramètre λ est appris par validation croisée et également des méthodes reposant sur des réseaux de neurones comme IRCNN [ZZGZ17] et MWCNN [LZZ⁺18]. Nous observons que les résultats sont identiques voire meilleurs avec l'approche proposée comme illustré Figure 30. Plus de résultats sont disponibles dans [JP21].



2.4.4.4 Conclusions

Ce travail préliminaire offre des perspectives prometteuses sur la combinaison des approches variationnelles présentées dans l'ensemble de ce manuscrit et des méthodes d'apprentissage profond. Il n'en reste pas moins que de nombreuses questions ouvertes restent en suspens comme la gestion d'un plus grand volume de données et la stabilité. Plusieurs pistes de recherche seront précisées dans la conclusion générale.

2.5 Applications variées des algorithmes proximaux sur des données réelles.

La formulation variationnelle (2.1) combinée à une minimisation par des algorithmes proximaux a été mise en œuvre pour le traitement de diverses données réelles. Dans cette partie, je résume brièvement chacune de ces contributions.

Microscopie par illumination structurée Les développements récents en microscopie ont permis de révéler le fonctionnement de divers mécanismes moléculaires. Parmi les méthodes développées on remarque une importance croissante de la combinaison de la stratégie d'acquisition et d'une étape computationnelle de reconstruction adaptée. C'est dans ce contexte que s'est inscrit ce travail. En effet, la microscopie par illumination structurée (SIM) permet de doubler la résolution des microscopes à épifluorescence sans contrainte spécifique sur le marquage des protéines d'intérêts. Cette approche est basée sur l'acquisition d'une série d'images correspondant à différentes modulations, suivie d'une reconstruction permettant l'estimation d'une image haute résolution. Pour cette étape de reconstruction, deux contributions principales ont été proposées. La première concerne l'estimation des paramètres de modulation à partir des images acquises et la seconde traite du problème de la reconstruction d'images haute résolution à partir des images modulées.

L'estimation des paramètres de modulation des images SIM est primordiale pour leur reconstruction. Ainsi nous avons proposé une méthode basée sur la propriété d'annihilation pour estimer ces paramètres. Dans ce projet nous avons adapté le formalisme de filtre annihilant et d'al-

gorithme de Cadzow pour traiter le problème d'estimation spectrale 2-D. La méthode proposée a permis d'estimer de façon précise les paramètres de modulation des images en SIM [CBP⁺14].

Récemment les méthodes de déconvolution basées sur la minimisation de critères convexes non-lisses ont permis un gain significatif en terme de résolution en microscopie à champ large. Plus précisément, les performances obtenues permettent de se rapprocher des résolutions atteintes en microscopie par illumination structurée. Des travaux récents se sont alors intéressés à la reconstruction de données en SIM et c'est dans ce cadre que se place notre seconde contribution [BPC⁺18]. La reconstruction SIM mentionnée précédemment se base sur la minimisation d'un critère convexe quadratique conduisant à une forme explicite de la solution. Pour améliorer les performances de reconstruction, nous nous sommes inspirés de travaux récents en optimisation convexe non-lisse de façon à étendre la classe des critères pouvant être considérés. La procédure de reconstruction proposée permet par exemple de gérer un terme d'attache aux données quadratique mais également une antilog-vraisemblance poissonienne et une régularisation de type Tikhonov (quadratique) mais aussi pouvant gérer des pénalisation plus complexes. Munis de cette méthode de reconstruction, nous avons montré que la microscopie SIM permettait de rester compétitive vis à vis de la microscopie à champ large déconvoluée. Nous avons validé ces approches sur ces données simulées ainsi que sur des données réelles.

Afin de tester la robustesse des algorithmes développés, nous avons choisi des échantillons aux propriétés connues. Nous nous sommes ainsi orientés vers des cellules HeLa transfectées et exprimant la protéine fluorescente LifeAct. Pour cela nous avons travaillé avec les ingénieurs de la plateforme de microscopie du Nikon Imaging Center de l'Institut Curie. Nous avons aussi caractérisé la réponse impulsionnelle du microscope en collectant des images de billes fluorescentes afin d'intégrer cette information lors de la déconvolution des images acquises.

Apprentissage supervisé parcimonieux. Application à l'analyse du rythme cardiaque fœtal. La surveillance de la fréquence cardiaque fœtale (FHR) est utilisée pendant l'accouchement pour évaluer le bien-être du fœtus. Classiquement basée sur l'évaluation visuelle des critères FIGO, la caractérisation de la FHR reste une tâche difficile qui fait l'objet d'efforts de recherche continus et intensifs. L'analyse de la FHR intrapartum est encore compliquée pour deux différentes étapes du travail (dilatation et poussée active). Les travaux de recherche visant à mettre au point des procédures automatisées de prévision de l'acidose reposent soit sur la conception de nouvelles analyses avancées de traitement du signal, soit sur la combinaison efficace d'un grand nombre de caractéristiques proposées dans la littérature. Ces procédures s'appuient soit sur une étape préalable de sélection des caractéristiques, soit sur des règles de décision impliquant de longues listes de caractéristiques mais qui ne permet pas d'interpréter facilement la décision et n'est donc pas adaptée à la pratique clinique. L'évaluation des règles de décision basée sur l'apprentissage machine est souvent entravée par l'utilisation de bases de données différentes et de petite taille, ce qui empêche de comparer de manière exhaustive les résultats rapportés dans la littérature.

Dans [SFL⁺17], nous avons considéré l'apprentissage parcimonieux (cf. Section 2.2.5) comme moyen d'effectuer conjointement la sélection des caractéristiques et la prédiction de l'acidose, produisant ainsi à une règle de décision optimale basée sur le moins de caractéristiques possibles. En utilisant un ensemble de 20 caractéristiques (rassemblant des caractéristiques de type FIGO, des caractéristiques spectrales classiques et des caractéristiques fractales), appliquées à deux bases de données de grande taille (respectivement 1800 et 500 sujets), bien documentées, collectées indépendamment dans des hôpitaux français et tchèques, les avantages de l'apprentissage parcimonieux sont quantifiés en termes de i) prise en compte du déséquilibre des classes (peu de sujets acidotiques), ii) de production de règles de décisions simples et interprétables, iii) de mise en évidence des différences entre les dynamiques temporelles des étapes de poussée active et de dilatation, et iv) de validité/généralisation des règles de décision apprises sur une base de données et appliquées à l'autre.

Estimation des matrices origine-destination. Application à l'analyse du trafic à Brisbane L'estimation de la matrice origine-destination (ODM) est un problème classique en ingénierie des transports visant à récupérer les flux de chaque origine vers chaque destination à partir des comptages de trafic mesurés et des informations de modèle a priori. En plus des comptages de trafic, dans [MPB⁺17] nous avons tiré parti des trajectoires issues de données Bluetooth, dont la capture est rendue possible par les nouvelles technologies de mesure. Il étend le concept d'ODM à celui d'LODM dépendant de la liaison (LODM), en conservant les informations sur la distribution du flux sur les liaisons et contenant de manière inhérente l'affectation ODM. De plus, une formulation originale de l'estimation LODM, à partir des comptages de trafic et des trajectoires issues de données Bluetooth est présentée comme un problème d'optimisation, où la fonctionnelle à minimiser est composée de cinq fonctions convexes, chacune modélisant une contrainte ou une propriété du problème de transport : cohérence avec les comptages de trafic, cohérence avec les trajectoires des données Bluetooth échantillonnées, cohérence avec la conservation du trafic (loi de Kirchhoff), similitude des flux ayant des origines et des destinations proches, positivité des flux de trafic. Un algorithme Condat-Vũ est conçu pour minimiser le critère construit, car les fonctions objectives correspondantes ne sont pas nécessairement dérivables. Une étude de cas, sur un réseau et un trafic simulés, valide la faisabilité de la procédure et détaille ses bénéfices pour l'estimation d'une LODM. L'étude de cas de Brisbane City, où plus de 600 centaines de détecteurs Bluetooth ont été installés, illustre également l'intérêt que ces matrices de liaison-origine-destination créent pour l'analyse du trafic.

Prédiction du $R(t)$ dans l'épidémie Covid19 Parmi les différents indicateurs qui quantifient la propagation d'une épidémie, telle que la COVID-19 en cours, figure d'abord le facteur de reproduction qui mesure combien de personnes peuvent être contaminées par une personne infectée. Afin de permettre le suivi de l'évolution de ce nombre, nous avons proposé [APR⁺20] une nouvelle procédure d'estimation en supposant un modèle bien accepté pour les données d'incidence actuelles, basé sur des observations passées. La nouveauté de l'approche proposée est double : 1) l'estimation du nombre de reproduction est réalisée par optimisation convexe proximale pour minimiser un critère favorisant le comportement linéaire par morceaux; 2) l'approche est développée dans un cadre multivarié, permettant le traitement simultané de plusieurs séries temporelles rattachées à différentes régions géographiques, ainsi qu'une régularisation spatiale (graphique) de leurs évolutions dans le temps. L'efficacité de l'approche est d'abord étayée par des simulations, et deux applications principales à des données COVID-19 réelles sont ensuite discutées. La première se réfère à l'évolution comparée du nombre de reproduction pour plusieurs pays, tandis que la seconde se concentre sur les départements français et leur analyse conjointe, conduisant à des cartes dynamiques révélant la co-évolution temporelle de leur facteur de reproduction.

Décomposition modale pour estimer l'onde interne réfléchie critique dans un fluide stratifié

Le cisaillement résultant des réflexions internes des ondes peut jouer un rôle crucial dans le transport et la remise en suspension des sédiments dans des conditions océaniques. En particulier, lorsque ces ondes subissent un phénomène critique de réflexion, l'onde réfléchie peut produire un cisaillement très important. Séparer l'onde réfléchie de l'onde incidente est un défi technique puisque les deux ondes partagent la même fréquence temporelle. Dans [HSP⁺21], nous présentons une série de mesures expérimentales d'ondes internes en configuration de réflexion critique et nous les analysons en utilisant la méthode de décomposition 2D-VMD-prox. Pouvoir isoler avec confiance l'onde réfléchie nous a permis de comparer nos résultats à un modèle visqueux et non linéaire de réflexion critique qui décrit correctement la dépendance du taux de cisaillement en fonction des paramètres expérimentaux.

2.6 Conclusions et perspectives

A travers ce manuscrit, j'ai pu présenter les principales contributions obtenues pendant les dix années qui ont suivi ma thèse. Les contributions et réflexions proposées à travers ce document sont le fruit de l'encadrement de plusieurs doctorant.e.s, chercheur.e.s en séjours post-doctoral, agrégés préparateurs et stagiaires nommés au début de ce document et également de nombreuses collaborations avec des collègues du LPENSL (P. Abry, P. Borgnat, P. Flandrin, V. Vidal, J.C. Gémignard, S. Joubaud, P. Odier, P. Jensen) et également avec des chercheur.e.s d'autres laboratoires dont les principales collaborations ont été menées avec L. Briceño-Arias, L. Condat, J. Boulanger, J.C. Pesquet, B. Pesquet-Popescu, N. Dobigeon, H. Wendt, M. Langlois, E. Thiebaut.

J'y ai présenté des contributions en modélisation à travers une généralisation du modèle de Mumford-Shah discret offrant une flexibilité dans le choix du terme d'attache aux données et de pénalisation des contours, une nouvelle formulation variationnelle fortement convexe pour la segmentation de texture qui repose sur des attributs fractals et enfin une formulation variationnelle de la décomposition modale empirique.

J'ai également proposé des formulations variationnelles reposant sur (2.1) pour diverses analyses de données réelles telles que l'étude des environnements circumstellaires, l'analyse d'écoulements multiphasés en milieu poreux, l'analyse d'images acquises par microscopie par illumination structurée, l'estimation du facteur de reproduction dans l'épidémie Covid19, l'analyse des ondes internes dans un fluide stratifié, l'estimation des matrices origine-destination pour l'analyse du trafic routier.

Les choix des approches variationnelles mentionnées précédemment sont guidés par le cadre flexible des algorithmes proximaux, principalement des algorithmes proximaux primaux-duaux, qui reposent sur des activations de fonctions par un pas proximal. Dans ce manuscrit j'ai présenté de nouvelles formes explicites d'opérateurs proximaux à savoir des nouvelles formes explicites pour une somme de deux fonctions ainsi que de nouvelles formes explicites de projections épigraphiques. J'ai montré l'importance de calculer l'opérateur proximal d'une somme de deux fonctions plutôt que de les activer séparément. J'ai présenté une technique de découpe épigraphique permettant de gérer efficacement des contraintes non-linéaires. J'ai également montré l'importance de l'activation d'une fonction par un pas proximal plutôt qu'un pas de descente de gradient à travers des comparaisons théoriques et numériques. Enfin, si la plupart de mes travaux reposent sur l'utilisation et la comparaison numérique d'algorithmes proximaux existants, dans le contexte du modèle de Mumford-Shah discret, un nouvel algorithme, l'algorithme SLPAM, a été proposé.

Dans l'ensemble des approches variationnelles considérées, le choix de l'hyperparamètre de régularisation est crucial tant en terme de performances d'estimation que de temps de calcul. Cette question a été regardée par différentes stratégies : d'une part, par la reformulation du modèle pénalisé sous une formulation contrainte combiné à un découpage épigraphique, mais également par la formulation d'une approche hybride variationnelle et bayésienne hiérarchique dans le cas spécifique du débruitage de signaux constants par morceaux. Je me suis également intéressée au principe de SURE (combiné à une recherche rapide reposant sur SUGAR) où une extension permettant de gérer un bruit corrélé a été proposée. Enfin à travers un formalisme d'apprentissage profond reposant sur le principe des algorithmes déroulés nous a permis d'estimer conjointement les pas de l'algorithme, l'hyperparamètre de régularisation λ et la matrice d'analyse D .

Les travaux mentionnés précédemment conduisent pour la plupart à des questions ouvertes qui ont été précisées dans les conclusions partielles de chaque section. Ici, j'indique les axes de recherche sur lesquels je souhaite plus précisément travailler dans les années à venir.

- **Estimation des interfaces combinée à la reconstruction d'images ou l'estimation de descripteurs de texture**

La fonctionnelle de Mumford-Shah discrète que nous avons proposé dans [FPC19] combiné à un schéma algorithmique SL-PAM a été largement validé dans un contexte de détection

d’interfaces dans des images bruitées par un bruit additif gaussien. Cependant, une étude approfondie théorique et numérique doit être menée dans d’autres contextes d’études, par exemple lorsque les données sont dégradées par un bruit de Poisson ou avec un bruit corrélié, mais également lorsqu’un opérateur linéaire intervient dans le terme d’attache aux données. On pourra en particulier s’intéresser à l’opérateur décrit dans la section 2.3.3 nous permettant alors de détecter des interfaces dans des données texturées. Cette étude est extrêmement importante car la stabilité de la solution impacte directement l’interface estimée, c’est donc un enjeu majeur d’avoir des garanties théoriques sur le schéma numérique proposé. Pour aborder cette question, il serait dans un premier temps raisonnable de limiter l’étude à une formulation bi-convexe de D-MS, voire bi-fortement-convexe. D’autre part, l’optimum obtenu avec SL-PAM pourrait être comparé avec une stratégie reposant sur un schéma de lifting consistant à reformuler le problème D-MS non convexe sous une forme équivalente convexe en dimension supérieure ayant alors des garanties d’optimalité globale. Une étude approfondie du choix du schéma de discrétisation (i.e. D) qui fournit une estimation précise du champ de gradient, et donc des contours, est également envisagée par exemple en suivant [FLT16; Con17; ALT21]. Une attention particulière sera portée au design de l’opérateur discret afin d’obtenir un estimateur multiniveaux convergent [CLL14].

Dans le contexte spécifique d’estimation d’interfaces dans des données texturées, l’intégration d’estimation des descripteurs de régularité locale et variance locale est une étape préliminaire à l’intégration d’autres descripteurs tels que l’orientation locale, la couleur ou des descripteurs multifractals.

— Accélération : Algorithmes multiniveaux

En optimisation lisse, des méthodes efficaces pour faire face à des problèmes à grande échelle sont les approches multiniveaux [Nas00; LN15; WG19]. Ces approches exploitent la structure des données pour les représenter à différents niveaux de résolution. Des approches multiniveaux combinées aux algorithmes proximaux ont été proposées par Parpas dans [Par17] reposant sur des itérations FBS. Cependant, comme cela a été illustré à travers ce manuscrit, FBS est rarement suffisant pour des modèles avancés de restauration d’images pour lesquels des schémas proximaux primaux-duaux sont souvent plus adaptés. L’extension des travaux de Parpas à d’autres schémas numériques tels que les algorithmes primaux-duaux ou les algorithmes de minimisation alternée est un enjeu majeur pour envisager le traitement de gros volumes de données par des approches variationnelles reposant sur des a priori complexes. On peut aisément imaginer que l’élaboration d’algorithmes efficaces reposera sur l’intégration de propriétés multiéchelles, conduisant éventuellement à une forte convexité des sous-problèmes. La littérature spécifique aux EDP et approches multiniveaux nous aidera à aborder cette question, en particulier en remarquant que les algorithmes proximaux et les EDP reposent tous deux sur la théorie des opérateurs monotones [BC17]. Une autre direction à explorer est d’établir des connexions avec des approches par blocs avec un choix intelligent de blocs en s’appuyant sur la structure des transformées multirésolution.

— Algorithmes déroulés et apprentissage profond

Les résultats préliminaires obtenus par le déroulement de l’algorithme Condat-Vũ dans le contexte de restauration d’image (section 2.4.4) sont très encourageants. Ce travail s’inscrit dans la continuité du travail pionnier de Gregor et Le Cun [GLC10], et d’autres qui ont suivi, montrant qu’il était possible de construire un réseau de neurones en s’appuyant sur les connaissances contruites depuis des années en approches variationnelles et algorithmes proximaux. Ce travail nécessite cependant d’être approfondi tant sur le plan de l’implémentation numérique que sur son analyse théorique. En effet, l’implémentation actuelle ne repose ni sur Keras ni sur Tensorflow mais seulement sur une implémentation interne en Matlab qui nécessite d’être améliorée afin d’être en mesure de traiter de plus grands volumes de données et une plus large gamme d’opérateurs linéaires. D’autre part, à l’heure actuelle, il est encore complexe de savoir comment construire un bon réseau reposant sur des itérations proximales car nous avons pu observer que deux schémas distincts construits

à partir d'un même critère et d'un même algorithme pouvait conduire à des résultats différents. Dans le cas de l'analyse théorique des algorithmes proximaux, on voit qu'il est déjà complexe d'identifier en amont d'une implémentation le schéma le plus efficace mais des bornes fines sur les taux de convergence sont une façon de répondre à cette question. L'équivalent dans le contexte des réseaux de neurones reste à établir, par exemple en considérant les analyses théoriques présentées dans [CP20].

La mise en œuvre de ces stratégies d'apprentissage dans le contexte de l'analyse des environnements cricumstellaires (section 2.2.4) et la segmentation de texture (section 2.3.3) sont en cours et sera également considérée pour la détection d'interfaces (section 2.3.2.3).

- **Quantification d'incertitudes** La quantification d'incertitudes est un problème extrêmement important en analyse de signaux et d'images car il permet de valider la présence d'une structure dans les données. Les méthodes traditionnelles pour quantifier l'incertitude sont les méthodes de Monte-Carlo par chaînes de Markov (MCMC) malheureusement souvent trop coûteuses pour être employées sur de gros volumes de données. Afin d'aborder cette question nous explorerons d'une part les méthodes de type Prox-MALA [PSC⁺16] raisonnant globalement et d'autre part la méthode BUQO [RPW19] permettant d'analyser l'incertitude autour d'une structure d'intérêt à l'aide de la notion de région de crédibilité et d'un test d'hypothèse bayésien. L'analyse par la méthode BUQO, pourra bénéficier des développements menés sur le fractionnement épigraphique présentés dans la section 2.2.3.1.
- **Discret versus continu** Au sein du LPENS, plusieurs signaux que nous souhaitons analyser peuvent être représentés de façon parcimonieuse dans un dictionnaire construit par la physique ou la biologie du problème [CPO⁺19; HAL⁺20]. A ce jour, dans [CPO⁺19], nous nous sommes contentés d'analyser les données stick-slip par minimisation d'un critère de la forme (2.177) (favorisant les solutions linéaires par morceaux) permettant d'analyser une grande gamme de signaux et de construire un diagramme de phase dans un contexte expérimental non considéré jusqu'alors (faible masse). Cependant, une analyse approfondie de ces signaux par une approche de type codage parcimonieux [EA06], où la construction du dictionnaire serait effectuée par la connaissance du spécialiste, est une piste prometteuse. Une première analyse consisterait à construire un dictionnaire discret mais l'extension reposant sur l'usage de dictionnaires continus [ECHG20; DDUF20] serait encore plus pertinente. La limitation proviendrait peut-être de coût de calculs plus importants avec une telle approche [DCGHL16].
- **Applications** Dans les années à venir, je souhaite d'une part poursuivre mes collaborations existantes au sein du LPENSL et d'autre part je souhaiterais renforcer les collaborations initiées récemment dans le contexte applicatif des données astronomiques (collaboration avec le CRAL, Observatoire de Lyon) et également la collaboration avec EVS (H. Piégay et collaborateurs) pour l'analyse des rivières et plus généralement l'analyse de données environnementales. Le point commun de ces deux domaines applicatifs est le grand volume de données. Dans la première collaboration, il sera principalement question de résolution de problèmes inverses alors que dans la seconde, les méthodes mises en œuvre seront des techniques de segmentation, détection d'interfaces et apprentissage.

2.7 Références

- [ABE⁺95] M. Ayer, H.D. Brunk, G.M. Ewing, W.T. Reid, and E. Silverman. An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.*, 26(4) :641–647, 1995. [60](#)
- [ABM06] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV Spaces : Applications to PDEs and Optimization*. MPS-SIAM Series on Optimization, 2006. [43](#), [44](#)
- [ABRS10] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems : An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2) :438–457, 2010. [63](#), [64](#)
- [ABS99] R. Alicandro, A. Braides, and J. Shah. Free-discontinuity problems via functionals involving the L1-norm of the gradient and their approximations. *Interfaces and free boundaries*, 1(1) :17–37, 1999. [63](#)
- [ABS13] H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems : proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math. Program.*, 137(1-2, Ser. A) :91–129, 2013. [74](#)
- [ACP⁺17] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot, and L. Laborelli. Dual block-coordinate forward-backward algorithm with application to deconvolution and deinterlacing of video sequences. *J. Math. Imag. Vis.*, 59 :415–431, 2017. [32](#)
- [ADR99] A. Arneodo, N. Decoster, and S.G. Roux. Intermittency, log-normal statistics, and multifractal cascade process in high-resolution satellite images of cloud structure. *Physical Review Letters*, 83(6) :1255–1258, 1999. [67](#)
- [AJW13] P. Abry, S. Jaffard, and H. Wendt. When Van Gogh meets Mandelbrot : Multifractal classification of painting’s texture. *Signal Process.*, 93(3) :554–572, 2013. [67](#)
- [ALT21] D. Antunes, J.-O. Lachaud, and H. Talbot. An elastica-driven digital curve evolution model for image segmentation. *J. Math. Imag. Vis.*, 63 :1–17, 2021. [100](#)
- [AO18] J. Adler and O. Oktem. Learned primal-dual reconstruction. *IEEE Trans. Med. Imag.*, 37(6) :1322–1332, 2018. [93](#)
- [APR⁺20] P. Abry, N. Pustelnik, S. Roux, P. Jensen, P. Flandrin, R. Gribonval, C.-G. Lucas, E. Guichard, P. Borgnat, N. Garnier, and B. Audit. Spatial and temporal regularization to estimate COVID-19 reproduction number $R(t)$: Promoting piecewise smoothness via convex optimization. *PLoS One*, 15(8), August 2020. [98](#)
- [AT90] L. Ambrosio and V.M. Tortorelli. Approximation of functional depending on jumps by elliptic functional via t-convergence. *Comm. Pure Applied Math.*, 43(8) :999–1036, 1990. [62](#)
- [AT92] L. Ambrosio and V.M. Tortorelli. On the approximation of free discontinuity problems. *Boll. Un. Mat. Ital.*, 6(B) :105–123, 1992. [62](#)
- [AW17] V. Andrearczyk and P.F. Whelan. Texture segmentation with Fully Convolutional Networks. *arXiv :1703.05230*, 2017. [67](#)

- [BAP20] L. M. Briceño-Arias and N. Pustelnik. Proximal or gradient steps for cocoercive operators. *arXiv2101.06152*, 2020. [58](#), [75](#), [76](#), [77](#)
- [BB12] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13 :281–305, February 2012. [87](#)
- [BBP05] A. Benazza-Benyahia and J.-C Pesquet. Building robust wavelet estimators for multicomponent images using Stein’s principles. *IEEE Trans. Image Process.*, 14(11) :1814–1830, Nov. 2005. [80](#)
- [BC06] T. Baumberger and C. Caroli. Solid friction from stick–slip down to pinning and aging. *Advances in Physics*, pages 279–348, 2006. [89](#)
- [BC08a] H. H. Bauschke and P. L. Combettes. A Dykstra-like algorithm for two monotone operators. *Pac. J. Optim.*, 4 :383–391, Sept. 2008. [59](#)
- [BC08b] X. Bresson and T. F. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processingfast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Prob. Imaging*, 2(4) :455–484, Nov. 2008. [41](#), [43](#), [44](#)
- [BC17] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2017. [28](#), [29](#), [30](#), [31](#), [40](#), [59](#), [100](#)
- [BCC⁺16] C. Boyer, N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss. On the generation of sampling schemes for magnetic resonance imaging. *SIAM J. Imaging Sci.*, 9(4) :2039–2072, 2016. [22](#)
- [BCC⁺20] C. Bertocchi, E. Chouzenoux, M.-C. Corbineau, J.-C. Pesquet, and M. Prato. Deep unfolding of a proximal interior point method for image restoration. *Inverse Problems*, 36(3) :034005, February 2020. [93](#)
- [BCM05] A. Buades, B. Coll, and J. Morel. A review of image denoising algorithms, with a new one. *Multiscale Model. and Simul.*, 4(2) :490–530, 2005. [24](#), [40](#), [41](#)
- [BDL06] J. Bolte, A. Daniilidis, and A. Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.*, 17(4) :1205–1223, 2006. [74](#)
- [BJMO12] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning*, 4(1) :1–106, 2012. [24](#), [28](#), [52](#)
- [BKL⁺19] T. A. Bubba, G. Kutyniok, M. Lassas, M. M’arz, W. Samek, S. Siltanen, and V. Srinivasan. Learning the invisible : A hybrid deep learning-shearlet framework for limited angle computed tomography. *Inverse Problems*, 35(6), 2019. [93](#)
- [BKP10] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3) :492–526, Sep. 2010. [24](#)
- [BL07] T. Blu and F. Luisier. The SURE-LET approach to image denoising. *IEEE Trans. Image Process.*, 16(11) :2778–2786, 2007. [86](#)
- [BM13] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Match. Int.*, 35(8) :1872–1886, August 2013. [27](#)
- [BMK09] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational Bayesian blind deconvolution using a total variation prior. *IEEE Trans. Image Process.*, 18(1) :12–26, Jan. 2009. [80](#)

- [BN93] M. Basseville and I.V. Nikiforov. Detection of Abrupt Changes : Theory and Application. *Prentice-Hall, Inc., Upper Saddle River, NJ, USA*, 1993. 24
- [BNPS17] J. Bolte, Trong P. Nguyen, J. Peypouquet, and B. W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Math. Program.*, 165(2, Ser. A) :471–507, 2017. 74
- [BPBT⁺10] M. Bertero, P. P. Boccacci, G. Talenti, R. Zanella, and L. Zanni. A discrepancy principle for poisson data. *Inverse Problems*, 26(10) :105004, 2010. 38
- [BPC⁺18] J. Boulanger, N. Pustelnik, L. Condat, T. Piolot, and L. Sengmanivong. Nonsmooth convex optimization for Structured Illumination Microscopy image reconstruction. *Inverse problems*, 34(9) :22pp, 2018. 97
- [BPL⁺01] C.L. Benhamou, S. Poupon, E. Lespessailles, S. Loiseau, R. Jennane, V. Siroux, W. J. Ohley, and L. Pothuau. Fractal analysis of radiographic trabecular bone texture and bone mineral density : two complementary parameters related to osteoporotic fractures. *Journal of Bone and Mineral Research*, 16(4) :697–704, 2001. 67
- [BRW18] J. Birdi, A. Repetti, and Y. Wiaux. Sparse interferometric stokes imaging under the polarization constraint (Polarized SARA). *Monthly Notices of the Royal Astronomical Society*, 478(4) :4442–4463, 2018. 49, 50
- [BSS17] M. Burger, A. Sawatzky, and G. Steidl. First order algorithms in variational image processing. (eds) *Splitting Methods in Communication, Imaging, Science, and Engineering*, Springer, 2017. 32
- [BSU13] M. Blondel, K. Seki, and K. Uehara. Block coordinate descent algorithms for large-scale sparse multiclass classification. *Machine Learning*, 93(1) :31–52, October 2013. 28, 53
- [BTT13] J. Bien, J. Taylor, and R. Tibshirani. A lasso for hierarchical interactions. *Annals of Statistics*, 41(3) :1111–1141, 2013. 55, 56
- [BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004. 28
- [BW99] M. Born and E. Wolf. *Principles of Optics : Electromagnetic theory of propagation interference and diffraction of light*. Cambridge University Press, 1999. 22
- [BZ87] A. Blake and A. Zisserman. *Visual reconstruction*. MIT press, 1987. 25
- [Cau47] A. Cauchy. Méthode générale pour la résolution des systèmes d’équations simultanées. *C. R. Acad. Sci. Paris*, 22 :536–538, 1847. 29
- [CBFAB97] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.*, 6(2) :298–311, 1997. 23, 24, 26
- [CBFe09] C. Chaux and L. Blanc-Féraud. Estimation d’hyperparamètres pour la résolution de problèmes inverses à l’aide d’ondelettes. In *Proc. GRETSI*, Dijon, 8-11 sept. 2009. 80
- [CBP⁺14] L. Condat, J. Boulanger, N. Pustelnik, S. Sahnoun, and L. Sengmanivong. A 2-D spectral analysis method to estimate the modulation parameters in structured illumination microscopy. In *IEEE International Symposium on Biomedical Imaging*, Beijing, China, April 28-May 2 2014. 4p. 97

- [CCP12] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *SIAM J. Imaging Sci.*, 5(4) :1113–1158, 2012. [26](#)
- [CCPW07] C. Chaux, P.L. Combettes, J.-C. Pesquet, and V.R. Wajs. A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4) :1495–1518, Jun. 2007. [36](#)
- [CD12] A. Chambolle and J. Darbon. Image processing and analysis with graphs : Theory and practice. *R. Lukac, Ed., Digital Imaging and Computer Vision.,CRC Press / Taylor and Francis*, 2012. [60](#)
- [CDBBP08] 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, August 2008. [86](#)
- [CDuCVu09] P. L. Combettes, D. Dũng, and B. Công Vũ. Dualization of signal recovery problems. *Set-Valued and Variational Analysis*, 18 :373–404, 2009. [59](#)
- [CEN06] T. Chan, S. Esedoglu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. App. Math.*, 66(5) :1632–1648, 2006. [26](#)
- [CG19] P. L. Combettes and L. E. Glaudin. Proximal activation of smooth functions in splitting algorithms for convex image recovery. *SIAM J. Imaging Sci.*, 12(4) :1905–1935, 2019. [74](#), [76](#), [78](#)
- [CH12] Y. Chen and A. O. Hero. Recursive $\ell_{1,\infty}$ lasso. *IEEE Trans. Signal Process.*, 60(8) :3978–3987, Aug. 2012. [39](#)
- [Cha04] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imag. Vis*, 20(1-2) :89–97, Jan. 2004. [23](#), [59](#)
- [CJPT15] E. Chouzenoux, A. Jezierska, J.C. Pesquet, and H. Talbot. A convex approach for image restoration with exact Poisson-Gaussian likelihood. *SIAM J. Imaging Sci.*, 8(4) :2662–2682, 2015. [22](#)
- [CKCH19] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi. Proximal splitting algorithms : A tour of recent advances, with new twists. *arXiv :1912.00137*, 2019. [33](#)
- [CLL14] D. Coeurjolly, J.-O. Lachaud, and J. Levallois. Multigrid convergent principal curvature estimators in digital geometry. *Computer Vision and Image Understanding*, 129 :27–41, 2014. Publisher : Elsevier. [100](#)
- [CLR90] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. MIT Press, 1990. [37](#)
- [CNR03] B. R. Corner, R. M. Narayanan, and S. E. Reichenbach. Noise estimation in remote sensing imagery using data masking. *Internat. J. Remote Sensing*, 24(4) :689–702, 2003. [22](#)
- [Con13a] L. Condat. A direct algorithm for 1D total variation denoising. *Signal Process. Lett.*, 20(11) :1054–1057, 2013. [60](#)
- [Con13b] L. Condat. A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Opt. Theory and Applications*, 158(2) :460–479, 2013. [33](#), [94](#)

- [Con17] L. Condat. Discrete total variation : New definition and minimization. *SIAM J. Imaging Sci.*, 10(3) :1258–1290, 2017. [100](#)
- [Cov65] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. on Electronic Computers*, EC-14(3) :326–334, June 1965. [27](#)
- [CP07a] P. L. Combettes and J.-C. Pesquet. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE J. Selected Topics Signal Process.*, 1(4) :564–574, Dec. 2007. [32](#), [34](#)
- [CP07b] P. L. Combettes and J.-C. Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Opt.*, 18(4) :1351–1376, Nov. 2007. [59](#)
- [CP08] P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6), Dec. 2008. [39](#)
- [CP11a] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.*, 40(1) :120–145, 2011. [33](#), [69](#)
- [CP11b] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. 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, New York, 2011. [28](#), [32](#), [40](#)
- [CP12] 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 and Variational Analysis*, 20(2) :307–330, June 2012. [40](#)
- [CP15] L. Condat and N. Pustelnik. Segmentation d’image par optimisation proximale. *GRETSI, Lyon, France*, Sept. 8-11, 2015. [62](#)
- [CP16] A. Chambolle and T. Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25 :161–319, 2016. [24](#)
- [CP20] P. Combettes and J.-C. Pesquet. Deep neural network structures solving variational inequalities. *Set-Valued Var. Anal.*, 28 :491—518, 2020. [93](#), [101](#)
- [CPCBB10] L. Chaari, J.-C. Pesquet, P. Ciuciu, and A. Benazza-Benyahia. An iterative method for parallel MRI SENSE-based reconstruction in the wavelet domain. *Medical Image Analysis*, 15(2) :185–201, November 2010. In press. [22](#)
- [CPCP09] L. Chaari, N. Pustelnik, C. Chaux, and J.-C. Pesquet. Solving inverse problems with over-complete transforms and convex optimization techniques. In *Proc. SPIE, Wavelets XIII*, volume 7446, page 14, San Diego, California, USA, Aug. 2-8 2009. [24](#)
- [CPO⁺19] J. Colas, N. Pustelnik, C. Oliver, J.-C. Geminard, and V. Vidal. Nonlinear denoising for solid friction dynamics characterization. *Physical Review E*, 100 :032803, September 2019. [88](#), [89](#), [101](#)
- [CPP09] C. Chaux, J.-C. Pesquet, and N. Pustelnik. Nested iterative algorithms for convex constrained image recovery problems. *SIAM J. Imaging Sci.*, 2(2) :730–762, June 2009. [56](#), [59](#)

- [CPPPP12] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. A proximal approach for constrained cospase modelling. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, March 25-30 2012. 4p. [45](#)
- [CPPPP14a] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. Epigraphical proximal projection for sparse multiclass SVM. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, 4-9 May 2014. [53](#)
- [CPPPP14b] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet. A nonlocal structure tensor-based approach for multicomponent image recovery problems. *IEEE Trans. Image Process.*, 23(12) :5531–5544, Dec. 2014. [41](#), [44](#)
- [CPPPP15] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. Epigraphical projection and proximal tools for solving constrained convex optimization problems. *Signal, Image and Video Processing*, 9 :1737–1749, November 2015. [35](#), [36](#), [37](#), [39](#), [41](#)
- [CPPPP16] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. A proximal approach for sparse multiclass SVM. *arXiv :1501.03669*, 23(12) :5531–5544, 2016. [52](#), [53](#)
- [CPR14] 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 Opt. Theory and Applications*, 162(1) :107–132, 2014. [24](#)
- [CPTC11] L. Chaari, J.-C. Pesquet, J.-Y. Tourneret, and P. Ciuciu. Parameter estimation for hybrid wavelet-total variation regularization. In *Proc. IEEE Workshop Stat. Sign. Proc.*, Nice, France, Jun. 28-30 2011. [80](#)
- [CR97] G. H.-G. Chen and R. T. Rockafellar. Convergence rates in forward-backward splitting. *SIAM J. Optim.*, 7(2) :421–444, 1997. [76](#)
- [CRD07] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation : integrating color, texture, motion and shape. *Int. J. Comp. Vis.*, 72(2) :195–215, 2007. [67](#)
- [CRT06] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2) :489–509, Feb. 2006. [22](#)
- [CS01] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *J. Mach. Learn. Res.*, 2 :265–392, 2001. [28](#)
- [CS13] X. Cai and G. Steidl. Multiclass segmentation by iterated ROF thresholding. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 237–250. Springer, 2013. [26](#), [65](#), [69](#)
- [CS15] D. Cremers and J. Stühmer. A fast projection method for connectivity constraints in image segmentation. *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 8932 :183–196, January 2015. [60](#)
- [CSZL14] Q. Cheng, H. Shen, L. Zhang, and P. Li. Inpainting for remotely sensed images with a multichannel nonlocal total variation model. *IEEE Trans. on Geoscience and Remote Sensing*, 52(1) :175–187, January 2014. [41](#), [43](#), [45](#)

- [Cur44] H. B. Curry. The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2 :258–261, 1944. [29](#)
- [CV95] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3) :273–297, September 1995. [20](#)
- [CV01] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Process.*, 10(2) :266–277, 2001. [20](#)
- [DCGHL16] Y. De Castro, F. Gamboa, D. Henrion, and J.B. Lasserre. Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Trans. Inform. Theory*, 63(1) :621–630, 2016. [101](#)
- [DDDM04] 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(11) :1413–1457, Nov. 2004. [20](#)
- [DDUF20] T. Debarre, Q. Denoyelle, M. Unser, and J. Fageot. Sparsest continuous piecewise-linear representation of data. *arXiv :2003.10112*, 2020. [101](#)
- [DFKE07] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8) :2080 – 2095, Aug. 2007. [24](#)
- [DK01] P.L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Ann. Stat.*, 29(1) :1–65, 2001. [60](#)
- [DKE12] A. Danielyan, V. Katkovnik, and K. Egiazarian. BM3D frames and variational image deblurring. *IEEE Trans. Image Process.*, 21(4) :1715–1728, 2012. [24](#)
- [DKF⁺13] C. Dossal, M. Kachour, J. Fadili, G. Peyré, and C. Chesneau. The degrees of freedom of the Lasso for general design matrix. *Statistica Sinica*, pages 1–39, 2013. [88](#)
- [DLW11] I. Daubechies, J. Lu, and H.-T. Wu. Synchrosqueezed wavelet transforms : An empirical mode decomposition-like tool. 30(2) :243–261, Mar. 2011. [46](#)
- [DS09] J. Duchi and Y. Singer. Boosting with structural sparsity. In *International Conference on Machine Learning*, pages 297–304, Montreal, Canada, 14-18 June 2009. [56](#)
- [DST12] C. Ding, D. F. Sun, and K.-C. Toh. An introduction to a class of matrix cone programming. *Math. Programm.*, pages 1–39, Dec. 2012. [37](#)
- [DTD07] N. Dobigeon, J.-Y. Tournet, and M. Davy. Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach. *IEEE Trans. Signal Process.*, 55(4) :1251–1263, April 2007. [81](#), [82](#), [83](#)
- [DTS05] N. Dobigeon, J.-Y. Tournet, and J. D. Scargle. Change-point detection in astronomical data by using a hierarchical model and a bayesian sampling approach. In *Proc. IEEE Workshop Stat. Sign. Proc.*, pages 369–374, Bordeaux, France, July 2005. [24](#)
- [DTS06] N. Dobigeon, J.-Y. Tournet, and J. D. Scargle. Joint segmentation of multivariate poissonian time series applications to burst and transient source experiments. In *Proc. European Signal Processing Conf. (EUSIPCO)*, Florence, Italy, Sept. 2006. [82](#), [83](#)

- [DTS07] N. Dobigeon, J.-Y. Tournet, and J.D. Scargle. Joint segmentation of multivariate astronomical time series : Bayesian sampling with a hierarchical models. *IEEE Trans. Signal Process.*, 55(2) :414–423, Feb. 2007. [24](#), [80](#)
- [DVPF14] C. Deledalle, S. Vaite, G. Peyré, and J. Fadili. Stein Unbiased GrAdient estimator of the Risk (SUGAR) for multiple parameter selection. *SIAM J. Imaging Sci.*, 7(4) :2448–2487, 2014. [80](#), [86](#), [87](#), [88](#), [90](#)
- [DY17] D. Davis and W. Yin. Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions. *Math. Oper. Res.*, 42(3) :783–805, 2017. [31](#), [74](#)
- [EA06] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. 15(12) :3736–3745, Dec. 2006. [101](#)
- [ECHG20] C. Elvira, J. E. Cohen, C. Herzet, and R. Gribonval. Continuous dictionaries meet low-rank tensor approximations. *arXiv :2009.06340*, 2020. [101](#)
- [Eld09] Y. C. Eldar. Generalized SURE for exponential families : Applications to regularization. *IEEE Trans. Signal Process.*, 57 :471–481, 2009. [86](#), [87](#)
- [EMR07] M. Elad, P. Milanfar, and R. Ron. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3) :947–968, June 2007. [24](#)
- [FB12] A. Foi and G. Boracchi. Foveated self-similarity in nonlocal image filtering. In *Proc. SPIE Electronic Imaging 2012, Human Vision and Electronic Imaging XVII*, volume 8291, Burlingame (CA), USA, Jan. 2012. [40](#)
- [FHHT07] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani. Pathwise coordinate optimization. *Ann. Appl. Statist.*, 1(2) :302–332, 2007. [60](#)
- [Fla99] P. Flandrin. Time-frequency/time-scale analysis. *Academic Press*, 10, 1999. [27](#)
- [FLT16] M. Foare, J.-O. Lachaud, and H. Talbot. Image restoration and segmentation using the Ambrosio–Tortorelli functional and discrete calculus. In *Proc. Int. Conf. Patt. Rec.*, Cancun, Mexico, December 2016. [26](#), [62](#), [63](#), [65](#), [100](#)
- [FPC18] M. Foare, N. Pustelnik, and L. Condat. A new proximal method for joint image restoration and edge detection with the Mumford-Shah model. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, Calgary, Canada, April 2018. :hal
- [FPC19] M. Foare, N. Pustelnik, and L. Condat. Semi-linearized proximal alternating minimization for a discrete Mumford-Shah model. *IEEE Trans. on Image Process.*, 29(1) :2176–2189, 2019. [62](#), [63](#), [64](#), [65](#), [99](#)
- [FPD⁺17] J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, and P. Abry. Bayesian selection for the L2-Potts model regularization parameter : 1D piecewise constant signal denoising. *IEEE Trans. Signal Process.*, 65(19) :5215–5224, 2017. [83](#), [84](#)
- [FPWA15] J. Frecon, N. Pustelnik, H. Wendt, and P. Abry. Multivariate optimization for multifractal-based texture segmentation. In *Proc. IEEE Int. Conf. Image Process.*, pages 4957–4961, Quebec City, Canada, September 2015. [68](#), [69](#)
- [GAR⁺14] E. Gerasimova, B. Audit, S. G. Roux, A. Khalil, O. Gileva, F. Argoul, O. Naimark, and A. Arneodo. Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis. *Frontiers in Physiology*, 5, May 2014. [67](#)

- [GB17a] P. Giselsson and S. Boyd. Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, 62(2) :532–544, 2017. [31](#)
- [GB17b] P. Giselsson and S. Boyd. Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. *IEEE Trans. Automat. Control*, 62(2) :532–544, 2017. [74](#), [76](#)
- [GCPP12] R. Gaetano, G. Chierchia, and B. Pesquet-Popescu. Parallel implementations of a disparity estimation algorithm based on a proximal splitting method. In *Visual Communication and Image Processing*, pages 1–6, San Diego, USA, 27–30 Nov. 2012. [53](#)
- [GG84] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Match. Int.*, 6(6) :721 – 741, November 1984. [25](#)
- [Gis17] P. Giselsson. Tight global linear convergence rate bounds for Douglas-Rachford splitting. *Journal of Fixed Point Theory and Applications*, 19(4) :2241–2270, 2017. [31](#), [76](#)
- [GLC10] K. Gregor and Y. Le Cun. Learning fast approximations of sparse coding. In *Proc. International Conference on Machine Learning*, pages 399–406, Haifa, Israel, Jun. 21–24 2010. [93](#), [100](#)
- [GO07] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *Multiscale Model. and Simul.*, 6(2) :595–630, 2007. [24](#), [40](#)
- [GO09] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Model. and Simul.*, 7(3) :1005, 2009. [24](#)
- [GOW19] D. Gilton, G. Ongie, and R. Willett. Neumann networks for linear inverse problems in imaging. *IEEE Trans. Comput. Imaging.*, 9 :328–343, 2019. [93](#)
- [GSC12] B. Goldluecke, E. Strekalovskiy, and D. Cremers. The natural vectorial total variation which arises from geometric measure theory. *SIAM J. Imaging Sci.*, 5(2) :537–563, 2012. [43](#), [44](#)
- [GSL⁺21] S. Guérit, S. Sivankutty, J.A. Lee, H. Rigneault, and L. Jacques. Compressive lensless endoscopy with partial speckle scanning. *arXiv :2104.10959*2021, 2021. [22](#)
- [GW84] J. Grotzinger and C. Witzgall. Projections onto order simplexes. *Applied Mathematics and Optimization*, 12(1) :247–270, June 1984. [60](#)
- [GY95] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Process.*, 4(7) :932–945, 1995. [26](#)
- [HAL⁺20] M. Hennion, J.-M. Arbona, L. Lacroix, C. Cruaud, B. Theulot, B. Le Tallec, F. Proux, X. Wu, E. Novikova, S. Engelen, A. Lemainque, B. Audit, and O. Hyrien. FORK-seq : replication landscape of the *saccharomyces cerevisiae* genome by nanopore sequencing. *Genome Biology*, 21(125), 2020. [101](#)
- [HK70] A.E. Hoerl and R.W. Kennard. Ridge Regression : Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1) :55–67, 1970. [20](#)
- [HO93] P. C. Hansen and D. P. O’Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14(6) :1487–1503, 1993. [80](#)

- [HSP⁺21] E. Horne, J. Schmitt, N. Pustelnik, S. Joubaud, and P. Odier. Variational mode decomposition for estimating critical reflected internal wave in stratified fluid. *Experiments in Fluids*, 2021. to appear. [98](#)
- [HUL01] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms, Part II : Advanced theory and bundle methods*, volume 305 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, Heidelberg, N.Y., 2nd edition, 2001. [47](#)
- [HWS14] A. Haris, D. Witten, and N. Simon. Convex modeling of interactions with strong heredity. *J. Comput. Graph. Statist.*, pages 1–31, 2014. [55](#), [56](#)
- [JDCP11] L. Jacques, L. Duval, C. Chaux, and G. Peyré. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Processing*, 91(12) :2699–2730, Dec. 2011. [24](#)
- [Jef46] Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proc. of the Royal Society of London. Series A*, 186(1007) :453–461, 1946. [82](#)
- [JMFU17] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep Convolutional Neural Network for inverse problems in imaging. *IEEE Trans. Image Process.*, 26(9) :4509–4522, Jun. 2017. [93](#)
- [JMOB11] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for hierarchical sparse coding. *J. Mach. Learn. Res.*, 12 :2297–2334, 2011. [20](#), [24](#), [55](#), [56](#)
- [JP21] M. Jiu and N. Pustelnik. A deep primal-dual proximal network for image restoration. *IEEE J. Selected Topics Signal Process.*, 15(2) :190–203, January 2021. [93](#), [94](#), [95](#)
- [JPJ⁺19] M. Jiu, N. Pustelnik, S. Janaqi, M. Chebre, and P. Ricoux. Sparse hierarchical interaction learning with epigraphical projection. *Journal of Signal Processing Systems*, 92 :637–654, 2019. [52](#), [56](#), [57](#)
- [Kat09] K. Kato. On the degrees of freedom in shrinkage estimation. *J. Multivariate Anal.*, 100(7) :1338–1352, 2009. [88](#)
- [KCFH05] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink. Sparse multinomial logistic regression : Fast algorithms and generalization bounds. *IEEE Trans. Pattern Anal. Match. Int.*, 27(6) :957 – 968, June 2005. [53](#)
- [KH96] D. Kundur and D. Hatzinakos. Blind image deconvolution. *Signal Processing Magazine*, 13(3) :43–64, May 1996. [22](#)
- [KMY17] E. Kang, J. Min, and J.C. Ye. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys.*, 44(10) :360–375, Oct. 2017. [93](#)
- [KSW20] L. Kiefer, M. Storath, and A. Weinmann. PALMS image partitioning-a new parallel algorithm for the piecewise affine-linear mumford-shah model. *Image Processing On Line*, 10 :124–149, 2020. [26](#)
- [KT09] M. Kowalski and B. Torrèsani. Sparsity and persistence : mixed norms provide simple signal models with dependent coefficients. *Signal, image and video processing*, 3(3) :251–264, Sep. 2009. [24](#)

- [KTHD12] G. Kail, J.-Y. Tourneret, F. Hlawatsch, and N. Dobigeon. Blind deconvolution of sparse pulse sequences under a minimum distance constraint : a partially collapsed Gibbs sampler method. *IEEE Trans. Signal Processing*, 60(6) :2727–2743, June 2012. [82](#)
- [Lav98] M. Lavielle. Optimal segmentation of random processes. *IEEE Trans. Signal Process.*, 46(5) :1365–1373, 1998. [81](#), [82](#)
- [LB08] F. Luisier and T. Blu. SURE-LET multichannel image denoising : Interscale orthonormal wavelet thresholding. *IEEE Trans. Image Process.*, 17(4) :482–492, 2008. [86](#)
- [LBB90] R. L. Lagendijk, J. Biemond, and D. E. Boeke. Identification and restoration of noisy blurred images using the expectation maximization algorithm. *IEEE Trans. Acous., Speech Signal Process.*, 38 :1180–1191, July 1990. [22](#)
- [LBH15] Y. LeCun, Y. Bengio, and G. Hinto. Deep learning. *Nature*, 521(7553) :436–444, 2015. [27](#)
- [Lew02] A. S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.*, 13(3) :702–725 (2003), 2002. [74](#)
- [LFC⁺14] L. Laporte, R. Flamary, S. Canu, S. Déjean, and J. Mothe. Non-convex regularizations for feature selection in ranking with sparse SVM. *IEEE Trans. on Neural Networks and Learning Systems*, 25(6) :1118 – 1130, June 2014. [53](#)
- [LFP17] J. Liang, J. Fadili, and G. Peyré. Activity identification and local linear convergence of forward-backward-type methods. *SIAM J. Optim.*, 27(1) :408–437, 2017. [74](#)
- [LH15] M. Lim and T. Hastie. Learning interactions through hierarchical group-lasso regularization. *J. Comput. Graph. Statist.*, 24 :627–654, 2015. [55](#)
- [LIMK18] A. Lucas, M. Iliadis, R. Molina, and A.K. Katsaggelos. Using deep neural networks for inverse problems in imaging : Beyond analytical methods. *IEEE Signal Process. Mag.*, 35(1) :20–36, Jan. 2018. [93](#)
- [LL01] M. Lavielle and E. Lebarbier. An application of MCMC methods for the multiple change-points problem. *Signal processing*, 81(1) :39–53, January 2001. [80](#), [81](#), [82](#)
- [LM79] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *IMA J. Numer. Anal.*, 16(6) :964–979, Dec. 1979. [31](#), [76](#)
- [LMAOM14] Y. Le Montagner, E. D. Angelini, and J.-C. Olivo-Marin. An unbiased risk estimator for image denoising in the presence of mixed poisson–gaussian noise. *IEEE Trans. Image Process.*, 23(3) :1255–1268, 2014. [86](#)
- [LMZ17] Z. Li, F. Malgouyres, and T. Zeng. Regularized non-local total variation and application in image restoration. *J. Math. Imag. Vis.*, 59(2) :296–317, 2017. [63](#), [65](#)
- [LN15] R.M. Lewis and S.G. Nash. Model problems for the multigrid optimization of systems governed by differential equations. *J. Sci. Comput.*, 26(6) :1811–1837, 2015. [100](#)
- [LO09] Y. Li and S. Osher. A new median formula with applications to pde based denoising. *Commun. Math. Sci.*, 7(3) :741–753, Sep. 2009. [41](#)

- [Lou08] C. Louchet. *Modèles variationnels et bayésiens pour le débruitage d'images : de la variation totale vers les moyennes non-locales*. Mathématiques appliquées, Université Paris Descartes, 2008. [23](#)
- [LP66] E. S. Levitin and B. T. Polyak. Constrained minimization methods. *USSR Computational mathematics and mathematical physics*, 6(5) :1–50, 1966. Publisher : No longer published by Elsevier. [29](#)
- [LT05] H. Lantéri and C. Theys. Restoration of astrophysical images – The case of Poisson data with additive Gaussian noise. *EURASIP EURASIP Journal on Advances in Signal Processing*, 2005(15) :2500–2513, 2005. [22](#)
- [LZZ⁺18] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo. Multi-level wavelet-CNN for image restoration. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 886–88609, Los Alamitos, CA, USA, jun 2018. IEEE Computer Society. [95](#)
- [MAB12] P. Machart, S. Anthoine, and L. Baldassarre. Optimal computational trade-off of inexact proximal methods. *arXiv :1210.5034*, 2012. [59](#)
- [Mal97] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, USA, 1997. [80](#)
- [Mar98] C. Marone. Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences*, 1998. [88](#)
- [Mer80] B. Mercier. *Inéquations variationnelles de la mécanique*, volume 1 of *Publications Mathématiques d'Orsay 80 [Mathematical Publications of Orsay 80]*. Université de Paris-Sud, Département de Mathématique, Orsay, 1980. [76](#)
- [MNCM01] R. Molina, J. Nunez, F. J. Cortijo, and J. Mateos. Image restoration in astronomy : A Bayesian perspective. *IEEE Signal Proc. Mag.*, 18(2) :11–29, March 2001. [80](#)
- [MOM12] S. Meignen, T. Oberlin, and S. McLaughlin. On the mode synthesis in the synchroqueezing method. In *Proc. Eur. Sig. and Image Proc. Conference*, pages 1865–1869, Bucharest, Romania, August 2012. [46](#)
- [Mor65] J. J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93 :273–299, 1965. [28](#)
- [MPB⁺17] G. Michau, N. Pustelnik, P. Borgnat, P. Abry, A. Nantes, A. Bhaskar, and E. Chung. A primal-dual algorithm for link dependent origin destination matrix estimation. *IEEE Transactions on Signal and Information Processing over Networks*, 3(1) :104–113, March 2017. [98](#)
- [MPP09] W. Miled, J.-C. Pesquet, and M. Parent. A convex optimization approach for depth estimation under illumination variation. *IEEE Trans. Image Process.*, 18(4) :813–830, 2009. [60](#)
- [MS89] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied Math.*, 42(5) :577–685, 1989. [25](#)
- [MS12] T. Miyata and Y. Sakai. Vectorized total variation defined by weighted L-infinity norm for utilizing inter channel dependency. In *Proc. IEEE Int. Conf. Image Process.*, pages 3057–3060, Orlando, FL, USA, Sep. 30-Oct. 3 2012. [41](#)

- [MSV16] R. Philippe M. Serres, M.-L. Zanota and V. Vidal. On the stability of taylor bubbles inside a confined highly porous medium. *International Journal of Multiphase Flow*, 85 :157–163, 2016. [91](#)
- [MSY16] X.-J. Mao, C. Shen, and Y.-B. Yang. Using deep neural networks for inverse problems in imaging : Beyond analytical methods. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, volume 29, Barcelona, Spain, Dec. 5-10 2016. [93](#)
- [MW11] J. D. McEwen and Y. Wiaux. Compressed sensing for wide-field radio interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 413(2) :1318–1332, May 2011. [22](#)
- [Nas00] S.G. Nash. A multigrid approach to discretized optimization problems. *Optimization Methods and Software*, 14(1-2) :99–116, 2000. [100](#)
- [Nik05] M. Nikolova. Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *Multiscale Model. and Simul.*, 405(3) :960–991, 2005. [24](#)
- [NNZC08] M. Nikolova, M.K. Ng, S. Zhang, and W.-K. Ching. Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. *SIAM J. Imaging Sci.*, 1(1) :2–25, 2008. [24](#)
- [Noc80] J. Nocedal. Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, 35(151) :773–782, July 1980. [52](#)
- [NPPT97] J. G. Nagy, V. P. Pauca, R. J. Plemons, and T. C. Torgersen. Space-varying restoration of optical images. *Journal of the Optical Society of America A*, 14 :3162–3174, 1997. [22](#)
- [NPPT05] J. G. Nagy, V. P. Pauca, R. J. Plemmons, and T. C.J. Torgersen. Anisoplanatic deconvolution of adaptive optics images. *Opt. Soc. Amer. A*, 22 :504—513, March 2005. [22](#)
- [OF97] A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set : A strategy employed by V1? *Vision Research*, 37 :3311–3325, 1997. [20](#)
- [PADF02] E. Punsakaya, C. Andrieu, A. Doucet, and W.J. Fitzgerald. Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Trans. Signal Process.*, 50 :747–758, 2002. [82](#)
- [Par17] P. Parpas. A multilevel proximal gradient algorithm for a class of composite optimization problems. *SIAM J. Sci. Comput.*, 39(5) :S681–S701, 2017. [100](#)
- [PB14] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3) :123–231, 2014. [28](#), [36](#)
- [PBBC09] 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, 2009. [80](#), [86](#)
- [PBBZP16] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, and J.-C. Pesquet. Wavelet-based image deconvolution and reconstruction. *Wiley Encyclopedia of EEE*, 2016. [20](#), [22](#), [24](#)
- [PBF14] N. Pustelnik, P. Borgnat, and P. Flandrin. Empirical mode decomposition revisited by multicomponent non smooth convex optimization. *Signal processing*, 102 :313–331, Sep. 2014. [46](#)

- [PC17] N. Pustelnik and L. Condat. Proximity operator of a sum of functions; application to depth map estimation. *IEEE Signal Process. Lett.*, 24(12) :1827–1831, December 2017. [59](#), [60](#), [61](#)
- [PCJ21] B. Pairet, F. Cantalloube, and L. Jacques. MAYONNAISE : a morphological components analysis pipeline for circumstellar discs and exoplanets imaging in the near-infrared. *Monthly Notices of the Royal Astronomical Society*, 503(3) :3724–3742, May 2021. [23](#)
- [PCP11] N. Pustelnik, C. Chaux, and J.-C. Pesquet. Parallel proXimal algorithm for image restoration using hybrid regularization. *IEEE Transactions on Image Processing*, 20(9) :2450–2462, Sep. 2011. [32](#)
- [PKDH08] P. Paul, D. Kalamatianos, V. Duesman, and H. Huber. Automatic quality assessment for fluorescence microscopy images. In *Proc. IEEE International Conference on BioInformatics and BioEngineering*, pages 1–6, Athens, Greece, October 2008. [22](#)
- [PMPPA21] B. Pascal, V. Mauduit, N. Pustelnik, and P. P. Abry. Scale-free texture segmentation : Expert feature-based versus deep learning strategies. In *Proc. Eur. Sig. and Image Proc. Conference*, Amsterdam, The Netherlands, Jan. 18-22 2021. [73](#)
- [Pot52] R. B. Potts. Some generalized order-disorder transformations. *Proc. Camb. Phil. Soc.*, 48 :106–109, 1952. [24](#)
- [PP12] J.-C. Pesquet and N. Pustelnik. A parallel inertial proximal optimization method. *Pac. J. Optim.*, 8(2) :273–305, Apr. 2012. [33](#)
- [PPA⁺18] B. Pascal, N. Pustelnik, P. Abry, M. Serres, and V. Vidal. Joint estimation of local variance and local regularity for texture segmentation. Application to multiphase flow characterization. In *Proc. IEEE Int. Conf. Image Process.*, pages 2092–2096, Athens, Greece, 2018. [72](#)
- [PPA⁺20] B. Pascal, N. Pustelnik, P. Abry, J.-C. Géminard, and V. Vidal. Parameter-free and fast nonlinear piecewise filtering. Application to experimental physics. *Annals of Telecommunications*, 75 :655–671, 2020. [25](#), [86](#), [92](#)
- [PPA21] B. Pascal, N. Pustelnik, and P. Abry. Strongly convex optimization for joint fractal feature estimation and texture segmentation. *Appl. Comp. Harm. Analysis*, 54 :303–322, 2021. [67](#), [70](#), [71](#), [72](#), [90](#)
- [PPAP18] B. Pascal, N. Pustelnik, P. Abry, and J.-C. Pesquet. Block-coordinate proximal algorithms for scale-free texture segmentation. In *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Calgary, Alberta, Canada, Apr. 15-20 2018. [32](#), [69](#), [72](#)
- [PSC⁺16] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tournier, A. Hero, and S. McLaughlin. Tutorial on stochastic simulation and optimization methods in signal processing. *IEEE J. Selected Topics Signal Process.*, 10(2) :224 – 241, March 2016. [101](#)
- [PVPA21] B. Pascal, S. Vaiter, N. Pustelnik, and P. Abry. Automated data-driven selection of the hyperparameters for total-variation based texture segmentation. *J. Math. Imag. Vis.*, 2021. to appear. [80](#), [86](#), [87](#), [88](#), [90](#)
- [PWAD16] N. Pustelnik, H. Wendt, P. Abry, and N. Dobigeon. Combining local regularity estimation and total variation optimization for scale-free texture segmentation. *IEEE Trans. Comput. Imaging*, 2(4) :468–479, 2016. [67](#), [68](#), [70](#)

- [QCCD09] A. Quattoni, X. Carreras, M. Collins, and T. Darrell. An efficient projection for $\ell_{1,\infty}$ regularization. In *International Conference on Machine Learning*, Montreal, Quebec, Jun., 14-18 2009. 39
- [RFGL07] G. Rilling, P. Flandrin, P. Goncalves, and J. M. Lilly. Bivariate Empirical Mode Decomposition. *IEEE Signal Process. Lett.*, 14 :936–939, 2007. 46
- [RFP13] H. Raguét, M.J. Fadili, and G. Peyré. A generalized forward-backward splitting. *SIAM J. Imaging Sci.*, 6(3) :1199–1226, 2013. 32
- [RHY19] E. K. Ryu, R. Hannah, and W. Yin. Scaled relative graph : Nonexpansive operators via 2D euclidean geometry. *arXiv :1902.09788*, 2019. 31, 74
- [Rob07] Christian P. Robert. *The Bayesian Choice : from Decision-Theoretic Motivations to Computational Implementation*. Springer Texts in Statistics. Springer-Verlag, New York, 2 edition, 2007. 82, 83
- [ROF92] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D : Nonlinear Phenomena*, 60(1-4) :259–268, 1992. 20, 23, 24, 39
- [RPW19] A. Repetti, M. Pereyra, and Y. Wiaux. Scalable bayesian uncertainty quantification in imaging inverse problems via convex optimization. *SIAM J. Imaging Sci.*, 12(1) :87–118, 2019. 38, 101
- [RTBG20] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson. Operator splitting performance estimation : Tight contraction factors and optimal parameter selection. *SIAM J. Opt.*, 30(3) :2251–2271, 2020. 76
- [RYF20] S. Ravishankar, J. C. Ye, and J. A. Fessler. Image reconstruction : From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1) :86–109, 2020. 93
- [RZZ⁺19] D. Ren, W. Zuo, D. Zhang, L. Zhang, and M. Yang. Simultaneous fidelity and regularization learning for image restoration. *IEEE Trans. Pattern Anal. Match. Int.*, pages 1–1, 2019. 93
- [SB10] I. W. Selesnick and I. Bayram. Total variation filtering. Technical report, 2010. 80
- [SC14] E. Strekalovskiy and D. Cremers. Real-time minimization of the piecewise smooth Mumford–Shah functional. In *Proc. European Conference on Computer Vision*, pages 127–141. Zurich, Germany, 2014. 26, 65
- [Set09] S. Setzer. *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage*, volume 5567, chapter Scale Space and Variational Methods in Computer Vision. SSVM 2009. Lecture Notes in Computer Science, pages 464–476. 2009. 33
- [SFL⁺17] J. Spilka, J. Frecon, R. Leonarduzzi, N. Pustelnik, P. Abry, and M. Doret. Sparse support vector machine for intrapartum fetal heart rate classification. *IEEE Journal of Biomedical and Health Informatics*, 21(3) :664–671, May 2017. 97
- [SGYB15] C. Studer, T. Goldstein, W. Yin, and R. G. Baraniuk. Democratic representations. *arXiv :1401.3420*, 2015. [http ://arxiv.org/abs/1401.3420](http://arxiv.org/abs/1401.3420). 39
- [Sha96] J. Shah. A common framework for curve evolution, segmentation and anisotropic diffusion. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, June 1996. IEEE. 63

- [Sho85] N. Z. Shor. *Minimization Methods for Non-differentiable Functions*. Springer-Verlag, New York, 1985. [28](#)
- [SJ16] Y. She and H. Jiang. Group regularized estimation under structural hierarchy. *Journal of the American Statistical Association*, pages 445–454, 2016. [55](#)
- [Son09] M. Sonka. *Medical Image Processing and Analysis*. SPIE, 2009. [22](#)
- [SPB⁺14] J. Schmitt, N. Pustelnik, P. Borgnat, P. Flandrin, and L. Condat. A 2-D Prony-Huang transform : A new tool for 2-D spectral analysis. *IEEE Trans. Image Process.*, 23(12) :5233 – 5248, December 2014. [46](#), [47](#)
- [Ste81] C. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Stat.*, 9(6) :1135–1151, 1981. [80](#), [86](#)
- [SWD14] M. Storath, A. Weinmann, and L. Demaret. Jump-sparse and sparse recovery using Potts functionals. *IEEE Trans. Signal Process.*, 62(14) :3654–3666, July 2014. [24](#), [26](#)
- [SWFU15] M. Storath, A. Weinmann, J. Friel, and M. Unser. Joint image reconstruction and segmentation using the Potts model. *Inverse Problems*, 31(2) :025003, 2015. [26](#)
- [Tar05] A. Tarantola. Inverse problem theory and methods for modelparameter estimation. *SIAM*, 2005. [50](#)
- [THG18] A. B. Taylor, J. M. Hendrickx, and F. Glineur. Exact worst-case convergence rates of the proximal gradient method for composite convex minimization. *J. Optim. Theory Appl.*, 178(2) :455–476, 2018. [31](#), [74](#)
- [Thi02] Eric Thiébaud. Optimization issues in blind deconvolution algorithms. In *Astronomical Data Analysis II*, volume 4847, pages 174–183. International Society for Optics and Photonics, 2002. [52](#)
- [Tib96] R. Tibshirani. Regression shrinkage and selection via the Lasso. *J. R. Statist. Soc. B*, 58 :267–288, 1996. [20](#)
- [Tik63] A. Tikhonov. Tikhonov regularization of incorrectly posed problems. *Soviet Mathematics Doklady*, 4 :1624–1627, 1963. [24](#)
- [Tit87] D. Titterton. On the iterative image space reconstruction algorithm for ECT. *IEEE Trans. Med. Imag.*, 6(1) :52–56, Mar. 1987. [34](#)
- [Tse91] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim.*, 29(1) :119–138, 1991. [76](#)
- [TSR⁺05] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *J. Royal Stat. Soc. B*, 67(1) :91–108, June 2005. [60](#)
- [TSY11] S. Theodoridis, K. Slavakis, and I. Yamada. Adaptive learning in a world of projections. *IEEE Signal Process. Mag.*, 28(1) :97–123, Jan. 2011. [38](#)
- [TT12] R. J. Tibshirani and J. Taylor. Degrees of freedom in lasso problems. *Ann. Stat.*, 40(2) :1198–1232, 2012. [88](#)
- [TVW05] B. A. Turlach, W. N. Venables, and S. J. Wright. Simultaneous variable selection. *Technometrics*, 47(3) :349–363, Aug. 2005. [39](#)

- [TW11] G. Thakur and H.-T. Wu. Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples. *SIAM J. Math. Anal.*, 43 :2078–2095, 2011. [46](#)
- [TW15] R. Tibshirani and L. Wasserman. Stein unbiased risk estimate. *Course notes from “Statistical Machine Learning”*, pages 1–12, 2015. [86](#)
- [UVL20] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. *Int. J. Comp. Vis.*, 128 :1867–1888, Oct. 2020. [93](#)
- [VDBF08] 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. [37](#)
- [VDF⁺17] S. Vaiteer, C. Deledalle, J. Fadili, G. Peyré, and C. Dossal. The degrees of freedom of partly smooth regularizers. *Ann. Inst. Stat. Math.*, 69(4) :791–832, 2017. [88](#)
- [VU08] C. Vonesch and M. Unser. A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution. *IEEE Trans. Image Process.*, 17(4) :539–549, Apr. 2008. [86](#)
- [Vũ13] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3) :667–681, April 2013. [33](#)
- [WG19] Z. Wen and D. Goldfarb. A line search multigrid method for large-scale nonlinear optimization. *SIAM J. Opt.*, 2026(36) :1478–1503, December 2019. [100](#)
- [WL02] G. Winkler and V. Liebscher. Smoothers for discontinuous signals. *Journal of Nonparametric Statistics*, 14(1-2) :203–222, 2002. [81](#), [84](#)
- [WLJ⁺11] J. Wu, F. Liu, L. C. Jiao, X. Wang, and B. Hou. Multivariate compressive sensing for image reconstruction in the wavelet domain. *IEEE Trans. Image Process.*, 20(12) :3483–3494, Dec. 2011. [39](#)
- [WS07] L. Wang and X. Shen. On l_1 -norm multi-class support vector machines : methodology and theory. *Journal of the American Statistical Association*, 102 :583–594, 2007. [53](#)
- [XKB12] X. Xie, S.C. Kou, and L. D. Brown. SURE estimates for a heteroscedastic hierarchical model. *J. American Statist. Ass.*, 107(500) :1465–1479, 2012. [86](#)
- [XRLJ14] L. Xu, J.S.J. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, volume 27, Montréal, Canada, Dec. 8-13 2014. [93](#)
- [YHC18] J.C. Ye, Y. Han, and E. Cha. Deep convolutional framelets : A general deep learning framework for inverse problems. *SIAM J. Imaging Sci.*, 11(2) :991–1048, 2018. [93](#)
- [Yu13] Y.-L. Yu. On decomposing the proximal map. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, pages 91–99, Oaxaca, Mexico, 2013. [59](#)
- [YZS12] Q. Yuan, L. Zhang, and H. Shen. Hyperspectral image denoising employing a spectral-spatial adaptive total variation model. *IEEE Trans. on Geoscience and Remote Sensing*, 50(10) :3660 – 3677, October 2012. [41](#), [43](#), [45](#)
- [ZH17] C. Zach and C. Häne. Discretized convex relaxations for the piecewise smooth Mumford-Shah model. In *Proc. Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 548–563, 2017. [66](#)

- [ZLWZ08] H.H. Zhang, Y. Liu, Y. Wu, and J. Zhu. Variable selection for multcategory SVM via sup-norm regularization. *Electronic Journal of Statistics*, 2 :149–167, 2008. [53](#)
- [ZW11] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *Proc. IEEE Int. Conf. Comput. Vis*, pages 479–486, Barcelona, Spain, Nov. 6-13 2011. [24](#)
- [ZZGZ17] K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep CNN denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, Honolulu, HI, USA, Jul. 21-26, 2017. [95](#)