



HAL
open science

Algorithmes de majoration-minimisation. Application aux problèmes inverses de grande taille en signal/image

Emilie Chouzenoux

► **To cite this version:**

Emilie Chouzenoux. Algorithmes de majoration-minimisation. Application aux problèmes inverses de grande taille en signal/image. Traitement du signal et de l'image [eess.SP]. Université Paris Est - Marne-la-Vallée, 2017. tel-01661236

HAL Id: tel-01661236

<https://hal.science/tel-01661236v1>

Submitted on 13 Dec 2017

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

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

Université Paris-Est Marne-la-Vallée

Lab. Informatique Gaspard Monge - UMR CNRS 8049

Mémoire

soutenu par

Emilie Chouzenoux

le 1er décembre 2017

pour l'obtention de

L'HABILITATION

à

DIRIGER DES RECHERCHES

ALGORITHMES DE MAJORATION-MINIMISATION.
APPLICATION AUX PROBLÈMES INVERSES
DE GRANDE TAILLE EN SIGNAL/IMAGE.

Président	JUTTEN Christian	Professeur à l'Université Grenoble Alpes
Rapporteur	DE MOL Christine	Professeure à l'Université Libre de Bruxelles
Rapporteur	FIGUEIREDO Mario	Professeur à l'Instituto Superior Técnico, Université de Lisbonne
Rapporteur	MALICK Jérôme	Directeur de Recherche à l'Université Grenoble Alpes
Examineur	MALGOUYRES François	Professeur à l'Université Paul Sabatier, Toulouse
Examineur	STEIDL Gabriele	Professeure à l'Université de Kaiserslautern, Allemagne

Table des matières

I	Parcours académique et scientifique	1
1	Curriculum vitæ	3
1.1	Identification	3
1.2	Biographie	3
2	Activités scientifiques	5
2.1	Responsabilités d’encadrement	5
2.2	Projets collaboratifs	8
2.3	Diffusion de logiciels libres	9
2.4	Séminaires invités	10
2.5	Autres activités	13
2.5.1	Activités de relecture	13
2.5.2	Comités scientifiques	13
2.5.3	Animation scientifique	13
2.5.4	Implication dans la communauté	14
2.6	Liste de publications	14
3	Activités d’enseignement	21
3.1	Heures enseignées	21
3.1.1	Ecole CentraleSupélec	21
3.1.2	Université Paris Est Créteil	22
3.1.3	IUT de Marne-la-Vallée	22
3.1.4	Université Paris-Est Marne-la-Vallée	22
3.1.5	Ecole Centrale de Nantes	23
3.2	Responsabilités administratives	23
3.2.1	IUT de Marne la Vallée	23
3.2.2	Tutorat de stages/projets	24

II	Synthèse des activités de recherche	27
4	Introduction générale	29
4.1	Thématique des recherches	29
4.2	Résumé des travaux de thèse	31
4.2.1	Contexte	31
4.2.2	Résultats obtenus	31
4.3	Problématique	32
5	Approximation majorante quadratique	35
5.1	Introduction	35
5.2	Propriété de Lipschitz différentiabilité	36
5.3	Potentiels renforçant la parcimonie	37
5.4	Approximation majorante séparable	41
6	Algorithmes de majoration-minimisation pour l'optimisation non convexe	45
6.1	Introduction	45
6.2	Algorithme explicite-implicite à métrique variable	46
6.2.1	Introduction	46
6.2.2	Algorithme VMFB	47
6.2.3	Résultats de convergence	48
6.2.4	Application à la restauration d'images dégradées par un bruit Gaussien non homogène	49
6.3	Algorithme MM sous-espace	51
6.3.1	Introduction	51
6.3.2	Algorithme 3MG dans le cas déterministe	51
6.3.3	Algorithme 3MG dans le cas stochastique	54
6.3.4	Application à l'identification en ligne de flou 2D	58
6.4	Conclusion	59
7	Algorithmes de majoration-minimisation alternés par blocs	61
7.1	Introduction	61
7.2	Algorithme VMFB alterné par blocs	63
7.2.1	Analyse de convergence	64
7.2.2	Application à la déconvolution aveugle de signaux sismiques	64
7.3	Algorithme dual VMFB par blocs	65
7.3.1	Introduction	65
7.3.2	Algorithme proposé et analyse de convergence	67
7.3.3	Application à la restauration d'archives vidéo	69
7.4	Algorithme MM parallèle alterné par blocs	71
7.4.1	Introduction	71
7.4.2	Algorithme proposé	71

7.4.3	Application à la déconvolution d'images 3D	73
7.5	Conclusion	74
8	Approches semi-quadratiques	77
8.1	Introduction	77
8.2	Algorithmes semi-quadratiques	77
8.2.1	Formulation du problème	77
8.2.2	Interprétation comme approche MM quadratique	78
8.2.3	Interprétation comme minimisation alternée d'un critère augmenté	79
8.3	Approche Bayes variationnelle pour les modèles non Gaussiens	81
8.3.1	Position du problème	81
8.3.2	Approximation Bayes variationnelle	82
8.3.3	Approche VBA par majoration	83
8.3.4	Résultats expérimentaux	85
8.4	Approches MCMC accélérées pour la simulation en grande dimension	87
8.4.1	Motivations	87
8.4.2	Algorithme de MALA préconditionné	88
8.4.3	Stratégie d'augmentation de données	89
8.4.4	Application à la restauration d'images multicomposantes	91
8.5	Conclusion	91
9	Conclusion et perspectives	93
	Bibliographie	97

Première partie

PARCOURS ACADÉMIQUE ET SCIENTIFIQUE

- Chapitre 1 -

Curriculum vitæ

1.1 IDENTIFICATION

Nom :	CHOUZENOUX	Prénom :	Emilie
Naissance :	le 22/04/1985	Lieu :	Vannes, France
Affectation :	Université Paris-Est Marne-la-Vallée	Date :	1er septembre 2010
Corps :	Maître de conférences	Section CNU :	61ème
Grade :	5ème échelon depuis le 01/04/2017	Classe :	normale

1.2 BIOGRAPHIE

Depuis sept. 2016 **Chercheuse associée** (délégation INRIA) au Centre de Vision Numérique (CVN), CentraleSupélec, Université Paris-Saclay, INRIA Saclay.

Mars 2015 - Août 2015 Congé maternité.

Depuis sept. 2011 **Maître de conférences** à l'Université Paris-Est Marne-La-Vallée. Laboratoire d'Informatique Gaspard Monge, UMR CNRS 8049 Équipe « Signal et Communications ».

2010 – 2011 **ATER** à l'Université Paris-Est Marne-La-Vallée.

- 2007 – 2010 **Doctorat** à l'École Centrale de Nantes (soutenu le 8 déc. 2010).
Institut de Recherche en Communications et Cybernétique de Nantes, UMR CNRS 6597.
Équipe « Analyse et Décision en Traitement du Signal et de l'Image ».
- Sujet :* Recherche de pas par Majoration-Minoration.
Application à la résolution de problèmes inverses.
- Encadrants :* Jérôme Idier (Directeur de thèse) et Saïd Moussaoui (Co-encadrant de thèse).
- Jury :* Christian Jutten (Président), Gilles Aubert (Rapporteur), Jean-Christophe Pesquet (Rapporteur), Henri Lantéri (Examineur).
- 2007 – 2010 **Monitorat** à l'École Centrale de Nantes. Département Informatique et Mathématiques.
- 2006 – 2007 **Master** à l'École Centrale de Nantes.
Spécialité « Traitement du Signal » (diplômée en oct. 2007, mention Bien).
- 2007 **Stage de Master** - 6 mois - École Polytechnique de Montréal.
- Sujet :* Modélisation du problème direct en tomographie axiale.
- Soutenance :* le 8 décembre 2010 à l'École Centrale de Nantes
- Encadrant :* Yves Goussard.
- 2004 – 2007 **Elève Ingénieur de l'École Centrale de Nantes**.
Filière « Traitement du Signal » (diplômée en oct. 2007, mention Bien).
- 2002 – 2004 **Classes préparatoires aux grandes écoles**,
MPSI-MP, Lycée Alain-René Lesage, Vannes (56).
- 2002 **Baccalauréat série S**,
Spécialités Mathématiques - Sciences de l'ingénieur (mention Bien).
Lycée Alain-René Lesage, Vannes (56).

- Chapitre 2 -

Activités scientifiques

2.1 RESPONSABILITÉS D'ENCADREMENT

Depuis mon recrutement à l'UPEM, j'ai encadré plusieurs stagiaires de Master 2, ainsi que des doctorants (deux de mes doctorants ont soutenus et une autre soutient en octobre 2017). Depuis septembre 2014, je bénéficie de la **Prime d'encadrement doctoral et de recherche** (rang B). Je détaille ci-dessous mes responsabilités d'encadrement.

Encadrements de thèses :

1. Depuis juillet 2017, je co-encadre à 30% la thèse de Maissa Sghaier (co-directeurs de thèse : Giovanni Palma, GE Healthcare, et Jean-Christophe Pesquet, CentraleSupélec, INRIA Saclay). Il s'agit d'une thèse CIFRE en collaboration avec GE Healthcare, sur le sujet « Modélisation de la tâche clinique en reconstruction pour la tomosynthèse du sein ».
2. Depuis octobre 2016, je co-encadre à 50% la thèse de Marie-Caroline Corbiveau (directeur de thèse : Jean-Christophe Pesquet, CentraleSupélec, INRIA Saclay), sur le sujet « Algorithmes rapides d'optimisation en ligne pour l'apprentissage et l'imagerie médicale ».
3. Depuis septembre 2015, je co-encadre à 50% la thèse de Afef Cherni (directeur de thèse : Marc-André Delsuc, Université de Strasbourg, IGBMC), sur le sujet « Méthodes modernes d'analyse de données en biophysique analytique ». Cette thèse a donné lieu aux publications [47, 46, 48].
4. Depuis décembre 2013, je co-encadre à 30% la thèse de Feriel Abboud (directeur de thèse : Jean-Christophe Pesquet, UPEM, LIGM). Il s'agit d'une thèse CIFRE en collaboration avec l'Institut National de l'Audiovisuel. La thèse

porte sur « la restauration super-résolution de séquences d'images, application aux documents d'archives télévisuelles ». La soutenance de cette thèse est prévue pour octobre 2017. Cette thèse a donné lieu aux publications [4, 1, 3, 2]. Feriel Abboud est ingénieur recherche et développement dans l'entreprise Witbe depuis juillet 2017.

5. Entre décembre 2013 et avril 2017, j'ai co-encadré à 50% la thèse de Yosra Marnissi (directeur de thèse : Jean-Christophe Pesquet, UPEM, LIGM). La thèse portait sur « la restauration d'images multicomposantes par des approches mixtes variationnelles et stochastiques ». Elle a été soutenue le 25 avril 2017 à l'Université Paris Est Marne-la-Vallée, devant le jury composé de Gersende Fort (présidente du jury), Cédric Févotte (rapporteur), Philippe Ciuciu (rapporteur), Amel Benazza-Benyahia (examinatrice) et Nelly Pustelnik (examinatrice). Cette thèse a donné lieu aux publications [142, 141, 139, 140, 138]. Yosra Marnissi est ingénieur recherche et développement chez SafranTech depuis juillet 2017.
6. Entre octobre 2011 et juin 2015, j'ai co-encadré à 50% la thèse de Audrey Repetti (directeur de thèse : Jean-Christophe Pesquet, UPEM, LIGM). La thèse avait pour sujet « Algorithmes d'optimisation en grande dimension : Applications à la résolution de problèmes inverses ». Elle a été soutenue le 29 juin 2015, devant le jury composé de Stéphane Mallat (président du jury), Jérôme Bolte (rapporteur), Saïd Moussaoui (rapporteur), Patrick-Louis Combettes (examinateur) et Gabriele Steidl (examinatrice). Cette thèse a donné lieu aux publications [169, 168, 163, 164, 167, 165, 67, 66, 166, 162]. Audrey Repetti est chercheuse post-doctorante à Heriot-Watt University depuis fin 2015.

Encadrements de post-doctorats :

Dans le cadre du projet ANR GRAPHSIP, deux étudiants post-doctorants ont été recrutés au LIGM, avec qui j'ai étroitement collaboré :

1. Yuling Zheng (Novembre 2015 - Septembre 2016), dont le travail de post-doctorat a porté sur le développement d'approches de type Bayes variationnelles, pour la restauration d'images. Notre collaboration a donné lieu à la publication de l'article de journal [142].
2. Alessandro Benfenati (Mars 2016 - Avril 2017), dont le travail de post-doctorat a porté sur la proposition de nouvelles approches d'optimisation pour la résolution de problèmes d'estimation de matrices de covariance associées à des modèles graphiques gaussiens. Notre collaboration a mené à la publication de l'article de conférence [20]. Un article de journal est en préparation.

Encadrements de stages :

1. Entre avril et juin 2017, j'ai encadré le stage de Tsz Kit Lau, étudiant de 3ème année de CentraleSupélec, option Mathématiques Appliquées. Ce stage portait sur l'analyse théorique et l'implémentation pratique d'une nouvelle approche variationnelle pour l'estimation des paramètres de formes gaussiennes multidimensionnelles. Un article portant sur l'application de ces travaux à la microscopie biphotonique a été soumis récemment à la conférence ISBI 2018.
2. Entre mai 2016 et novembre 2016, j'ai encadré le stage de M2 de Azar Louzi. Ce stage portait sur le développement de nouvelles approches de majoration-minimisation stochastiques dans le cas de fonctions de coût non quadratiques.
3. Entre octobre 2015 et mars 2016, j'ai encadré le stage de M2 de Sara Cadoni. Le stage s'insérait dans le cadre du projet Mastodons CNRS OPTMISIME, et portait sur les algorithmes parallèles par majoration-minimisation pour la déconvolution 3D en microscopie biphotonique. Ce travail de stage a fait l'objet d'un article à la conférence IEEE ICIP 2016 [174], sélectionné parmi les 7 meilleurs articles de la conférence (sur un millier d'articles acceptés).
4. Entre février 2013 et juin 2013, j'ai encadré le stage de M2 de Yosra Marnissi, dans le cadre d'une collaboration avec A. Benazza Benyahia (Supcom Tunis). Le stage portait sur la loi a priori multivariée pour la restauration d'images multispectrales. Ce travail de stage a fait l'objet d'une publication dans la conférence IEEE ICIP 2013 [138].
5. Entre avril 2012 et septembre 2012, j'ai encadré le stage de M2 de Ferial Abboud, dans le cadre d'une collaboration avec J.-H. Chenot et L. Laborelli (INA). Le stage portait sur la déconvolution aveugle d'images par approches proximales alternées.
6. Entre avril 2012 et septembre 2012, j'ai encadré le stage de M2 de Fiona Zolyniak, dans le cadre d'une collaboration avec H. Talbot (LIGM) et E. Gouillard (Saint-Gobain Recherche). Le stage portait sur le développement de méthodes de reconstruction d'images 3D d'échantillons de verre biphasé, à partir de données tomographiques. Ce travail de stage a fait l'objet d'une publication dans la conférence IEEE ICIP 2013 [88].

2.2 PROJETS COLLABORATIFS

J'ai été impliquée dans plusieurs projets collaboratifs, académiques ou industriels, soit en tant que simple membre, ou en tant que porteuse de projet. Je les liste ci-dessous.

Coordination de projets collaboratifs :

- ◇ J'ai soumis en 2017 un projet ANR JCJC intitulé MajIC (Majoration-Minimization algorithms for Image Computing), qui a été accepté. Ce projet démarra en octobre 2017. Ce projet implique des chercheurs du LIGM (Giovanni Chierchia, Hugues Talbot et moi-même), ainsi que Philippe Ciuciu (CEA Neurospin), Laurent Duval (IFPEN), et Jean-Christophe Pesquet (CentraleSupélec). Financé à hauteur de 190k€, il vise à développer de nouvelles approches de majoration-minimisation pour la résolution des problèmes de grande dimension en traitement d'image.
- ◇ Depuis juin 2017, je suis porteuse du projet MAJIMED, financé par l'Université Paris Saclay. Ce projet, financé à hauteur de 13k€, a pour objectif le développement de nouvelles méthodes d'optimisation par majoration-minimisation, pour l'imagerie médicale.
- ◇ Depuis avril 2017, je suis porteuse du projet NextGenBP, financé par CNRS-CEFIPRA, en collaboration avec Angshul Majumdar, maître de conférences à l'Indraprastha Institute of Information Technology, New Delhi (porteur du projet, pour la partie Inde). Ce projet collaboratif franco-indien, financé à hauteur de 7k€ par an, a pour objectif le développement de nouvelles méthodes d'optimisation pour l'entraînement des réseaux profonds.
- ◇ Depuis janvier 2016, je suis porteuse du projet Mastodons CNRS TABASCO. Ce projet collaboratif rassemble des chercheurs du LIGM (UMR CNRS 8049 - Marne-la-Vallée), IGBMC (UMR CNRS 7104 - Strasbourg), MASP (USR CNRS 3290 - Lille), LJLL (UMR CNRS 7598 - Paris). Ce projet, financé à hauteur de 35k€ a pour objectif l'analyse et la réduction du bruit non gaussien dans les données de spectroscopie.
- ◇ Depuis octobre 2014, je suis porteuse locale du projet ANR GRAPHSIP. Ce projet collaboratif, porté par Olivier Lezoray (PR, GREYC, Caen) rassemble des chercheurs de l'ENS Lyon, du GREYC, du LIGM et du GIPSA-LAB. Financé à hauteur de 500k€, il a pour objectif le développement de méthodes de traitement de signaux sur des graphes.
- ◇ En 2013-2014, j'ai été porteuse d'un projet "jeunes chercheurs GDR ISIS", auquel participaient Sandrine Anthoine, Caroline Chaux (LATP, Aix-Marseille Université) et Franck Debarbieux (Institut de Neurosciences de la Timone). Ce projet, financé à hauteur de 6500€, avait pour objectif le développement

de méthodes de déconvolution pour les données 3D acquises par microscopie biphotonique.

Participation à des projets collaboratifs :

- ◇ En 2015 et 2016, j'ai été membre du projet Défi Imag'In CNRS OPTIMISME (porteur, Jean-Christophe Pesquet, LIGM), intégrant également Sandrine Anthoine, Caroline Chaux (LATP, Aix-Marseille Université), Franck Debarbieux (Institut de Neurosciences de la Timone), Jean-Yves Tourneret (IRIT, Toulouse) et Patrick-Louis Combettes (LJLL, Paris). Ce projet, financé à hauteur de 45k€ avait pour objectif le développement de méthodes d'optimisation stochastique pour les problèmes de grande taille d'imagerie hyperspectrale.
- ◇ Entre 2014 et 2016, j'ai été impliquée dans une collaboration avec Laurent Duval (Ingénieur de recherche, chef de projet Traitement du Signal et des Images, à l'Institut Français du Pétrole et des Energies Nouvelles), dans le cadre d'un contrat CIFRE sur la déconvolution aveugle de signaux sismiques parcimonieux (thèse de Mai Pham Quam).
- ◇ Entre 2013 et 2015, j'ai participé au projet Mastodons CNRS MeasureHD, porté par Marc-André Delsuc (IGBMC, Strasbourg). L'objectif de ce projet interdisciplinaire, financé à hauteur de 70k€, était le développement de nouveaux protocoles d'analyse et de traitement de données pour la biologie.
- ◇ En 2013, j'ai collaboré avec Emmanuelle Goullart (Directrice du laboratoire Surface du Verre et Interfaces, Saint-Gobain Recherche), dans le cadre du stage de M2 de Fiona Zolyniak. Le travail portait sur la reconstruction d'images de matériaux biphasiques en tomographie à rayons X.
- ◇ En 2011-2012, j'ai été impliquée, avec Caroline Chaux (CR, CNRS) dans un contrat industriel financé par Thalès Alenia Space. Ce contrat, d'un montant de 10k€, avait pour objectif l'estimation et la correction du flou de bougé dans les observations satellitaires en haute résolution.

2.3 DIFFUSION DE LOGICIELS LIBRES

Depuis le début de ma carrière de chercheuse, je me suis attachée à diffuser les codes / logiciels que j'ai produit dans le cadre de mes activités de recherche, dans un esprit de recherche reproductible et valorisée. Je détaille ci-dessous les codes disponibles à ce jour :

- ◇ **PALMA** : Plateforme en ligne, en Python, pour la reconstruction de données réelles de spectroscopie DOSY en résonance magnétique nucléaire par un algorithme proximal. Juillet 2016. Licence Cecill-B.
- ◇ **ToolboxPMRI** : toolbox Matlab de reconstruction d'images à partir de

données d'IRM parallèle, par l'algorithme 3MG complexe. Date de la version actuelle : Novembre 2014. Licence Cecill-B.

- ◇ **IPLS** : toolbox Matlab de démixage linéaire de données hyperspectrales. Date de la version actuelle : Octobre 2013. Licence Cecill-B
- ◇ **RestoVMFB-Lab** : toolbox Matlab de restauration d'images par l'algorithme Explicite-Implicite à métrique variable. Date de la version actuelle : Avril 2013. Licence Cecill-B.
- ◇ **LSMM** : boîte à outils Matlab de recherche de pas pour la minimisation d'un critère à barrière logarithmique. Date de la version actuelle : Mars 2013. Licence Cecill-B.
- ◇ **RestoMMMG-Lab** : boîte à outils Matlab pour la restauration d'images dans le cas d'un bruit gaussien. Date de la version actuelle : Août 2011. Licence Cecill-B.

Je suis par ailleurs impliquée dans deux demandes de valorisation logicielle. Tout d'abord, un dossier a été déposé en 2016 à la SATT Alsace (demande effectuée en collaboration avec M.-A. Delsuc, A. Cherni, IGBMC, Strasbourg) pour le logiciel **PALMA** évoqué ci-dessus. En outre, un dossier a été déposé en 2017 à la DVT IRS-TEA pour le logiciel **EMILIO**, toolbox Matlab réalisée (en partie) lors de ma thèse pour l'inversion de la transformée de Laplace multidimensionnelle par maximisation de l'entropie, pour la restauration des données de relaxométrie 2D acquises par résonance magnétique nucléaire (collaboration avec S. Moussaoui, J. Idier, IRCCyN, Nantes et F. Mariette, CEMAGREF, Rennes).

Enfin, j'ai réalisé en collaboration avec G. Chierchia, J.-C. Pesquet et P. L. Combettes le site internet **The Proximity Operator Repository**, qui sera mis en ligne en novembre 2017. Ce site, à visée participative, se propose de fournir les opérateurs proximaux (et leurs codes Matlab/Python/Julia associés) d'une quantité importante de fonctions, afin d'aider les utilisateurs dans leur mise en oeuvre pratique des algorithmes proximaux. Un tutoriel, présentant un exemple d'application de ces outils en restauration d'image, est également fourni.

2.4 SÉMINAIRES INVITÉS

Je suis invitée très régulièrement pour donner des exposés dans le cadre de workshop nationaux/internationaux, d'écoles d'été ou de visites dans des laboratoires. Je donne ci-dessous une liste des exposés (titre/lieu/date) que j'ai pu donner. J'inclus également dans cette liste les conférences sans actes auxquelles j'ai participé. Les transparents associés à ces présentations sont, pour la grande majorité, disponibles sur ma page web.

1. A variational Bayesian approach for image restoration with Poisson-Gaussian noise. Journées annuelles 2017 des GdR MOA et MIA, Bordeaux, 18 octobre 2017.

2. Proximity operator computation for video restoration. Journées Franco-Chiliennes d'Optimisation, Toulouse, 6 juillet 2017.
3. Majorization-Minimization Subspace Algorithms for Large Scale Data Processing. Ecole d'été "Structured Regularization for High-Dimensional Data Analysis", IHP, 21 juin 2017.
4. Accelerated dual forward-backward algorithms. Application to video restoration. Université de Modène, 3 avril 2017.
5. A Block Parallel Majorize-Minimize Memory Gradient Algorithm. Journée GDR ISIS, Inversion et problèmes multi-*, IAP, Paris, 20 mars 2017.
6. A Block Parallel Majorize-Minimize Memory Gradient Algorithm. Séminaire Parisien des Mathématiques Appliquées à l'Imagerie, IHP, Paris, 3 novembre 2016.
7. A stochastic Majorize-Minimize subspace algorithm with application to filter identification problems. Cavalieri workshop on Optimization and Optimal Transport for Imaging, INRIA Paris, 12 octobre 2016.
8. Convergence analysis of a stochastic Majorize-Minimize memory gradient algorithm. 28th European Conference on Operational Research (EURO 2016), Poznan, Polan, 4 juillet 2016.
9. Proximal methods : tools for solving large-scale, inverse problems of image and video processing. GT Scientifique Images TOTAL, Annecy, 15 juin 2016.
10. A stochastic Majorize-Minimize subspace algorithm with application to filter identification problems. Workshop Smart Energy and Stochastic Optimization (SESO 2016), ENPC, Champs/Marne, 3 mai 2016.
11. A stochastic Majorize-Minimize subspace algorithm with application to filter identification problems. Journées SMAI-MODE, Toulouse, 25 mars 2016.
12. Accelerated dual forward-backward algorithms : Application to video restoration. Mathematical Image Analysis (MIA) conference, 19 janvier 2016.
13. An overview of stochastic methods for solving optimization problems. Journée GDR ISIS, Sur les interactions Méthodes de Monte Carlo et Algorithmes d'Optimisation, Paris, France, 26 novembre 2015.
14. Some recent results on proximal primal dual methods for convex optimization. Seminar OGRE, LJK, Grenoble, 16 octobre 2015.
15. Some recent results on proximal primal-dual methods for convex optimization. PGMO Days, 2015, Massy-Palaiseau, France, 27-28 octobre 2015.
16. Application d'outils d'optimisation modernes au traitement rapide de données spectrométriques. Rencontre scientifique avec l'Agence Nationale de Sécurité Sanitaire, journée du Labex Bézout, Champs-sur-Marne, France, 3 mars 2015.
17. Proximal methods : tools for solving inverse problems in a large scale. Application to biophysics measurements processing. Chemical Complexity and Biology Symposium, Strasbourg, France, 19-20 janvier 2015.

18. A random block-coordinate primal-dual proximal algorithm with application to 3D mesh denoising. Horizon Maths 2014, rencontres scientifiques IFPEN, Rueil-Malmaison, France, 15-16 décembre 2014.
19. A stochastic 3MG algorithm with application to 2D estimation problems. RICAM workshop : Variational Methods in Imaging, Linz, Autriche, 27-31 octobre 2014.
20. Proximal methods : tools for solving inverse problems on a large scale. Data Science and Massive Data Analysis, journée du Labex Bézout, Champs-sur-Marne, France, 12 juin 2014.
21. Une nouvelle approche régularisée pour la reconstruction de phase. Séminaire Parisien d'Optimisation, Paris, France, 24 mars 2014.
22. Algorithme explicite-implicite préconditionné. Application à la résolution de problèmes inverses de grande taille. Journée "Images et Signaux" du pôle Math-STIC de l'Université Paris 13, 18 mars 2014.
23. Algorithme explicite-implicite préconditionné. Application à la résolution de problèmes inverses de grande taille. Séminaire "Signal Image" co-organisé par l'IMS, l'IMB et le LaBRI, Bordeaux, 15 mars 2014.
24. Algorithme de majoration-minimisation pour les fonctions à valeurs complexes. Journée GDR ISIS, "Traitement du signal de données à valeurs complexes" , Marseille, France, 12 juin 2013.
25. A convex variational approach for restoring data corrupted with Poisson-Gaussian noise. 26th European Conference on Operational Research (EURO 2013), Rome, Italie, 1-4 juillet 2013.
26. A majorize-minimize memory gradient algorithm applied to X-Ray tomography. Workshop on tomography reconstruction, Saint-Gobain Recherche, ENS ULM, Paris, 11 décembre 2012.
27. A memory gradient algorithm for non-convex regularization with applications to image restoration. Journées GDR MOA-MSPC, "Optimisation et traitement d'images", Saint-François, Guadeloupe, 4-7 juin 2012.
28. A memory gradient algorithm for non-convex regularization with applications to image restoration. SIAM Conference on Imaging Science, Philadelphie, Pennsylvanie, USA, 20-22 mai 2012.
29. A memory gradient algorithm for l2-l0 regularization with applications to image restoration. Journées GDR MOA-MSPC, "Optimisation et traitement d'images" , La Londe les Maures, France, 6-10 juin 2011.
30. Algorithmes de majoration-minimisation pour l'optimisation sous contraintes de positivité. Journée GDR ISIS, "Contraintes de nonnégalivité en traitement du signal et des images" , Paris, France, 1er février 2011.

2.5 AUTRES ACTIVITÉS

2.5.1 Activités de relecture

J'ai une activité soutenue de relectrice pour plusieurs journaux internationaux dans le domaine du traitement du signal/image (Applied Optics, IEEE Transactions on Signal Processing, IEEE Transactions on Image Processing, Journal of Mathematical Imaging and Vision, SIAM Journal on Imaging Sciences, Signal Processing, Signal, Image and Video Processing by Springer) et des mathématiques appliquées (Inverse Problems, Journal of Optimization Theory and Applications, Statistics and Computing, Computational Optimization and Applications, Mathematical Problems in Engineering, SIAM Journal of Optimization).

Depuis 2012, je rédige régulièrement (3 – 4 par an) des “Mathematical Reviews” pour l’American Mathematical Society. Je suis enfin fréquemment sollicitée pour effectuer des relectures pour les conférences GRETSI, EUSIPCO, ICIP, ICASSP, ISBI, SSP. En 2014, j’étais membre du Technical Program Committee de la conférence EUSIPCO.

J’ai réalisé deux expertises, pour des dossiers de demande de financement CIFRE. J’ai également expertisé une demande de financement pour un projet GdR ISIS.

2.5.2 Comités scientifiques

J’ai participé à 4 jurys de thèse, en tant qu’examinatrice (Cécile Chenot, 29 septembre 2017, CEA Saclay ; Simone Rebegoldi, 4 avril 2017, Université de Parme, Italie ; Rahul Mourya, 1er février 2016, Université de Saint-Etienne) ou rapportrice (Anupriya Gogna, 26 avril 2017, IIT New Delhi, Inde).

J’ai participé à 8 comités de sélection pour le recrutement de maître de conférences en section CNU 61 (2013 : INP Toulouse, INSA Lyon ; 2014 : Université Paris Sud, INP Toulouse ; 2016 : Université Paris 6, Ecole CentraleSupélec ; 2017 : IUT de Lyon) ou de PRAG (2017 : IUT de Meaux). Depuis 2012, je suis membre de la commission de spécialistes de l’Université Paris Est Marne-la-Vallée, pour la section CNU 61, et participe dans ce cadre au recrutement des ATER dans cette section.

2.5.3 Animation scientifique

J’ai organisé une session spéciale “Intéractions entre méthodes d’optimisation et algorithmes de simulation stochastique” à la conférence GRETSI 2017, en collaboration avec François Septier. De plus, j’ai organisé la session “Advanced optimization methods for solving inverse problems at a large scale”, au workshop international Biomedical and Astronomical Signal Processing (BASP) Frontiers en janvier 2017, à Villars-sur-Ollon (Suisse).

2.5.4 Implication dans la communauté

Je suis membre de plusieurs groupes et associations de chercheurs de la communauté signal/image et mathématiques appliquées. Je suis membre active des GDRs ISIS (correspondante locale), Madics (corres. locale), MOA (corres. locale), Calcul, MIA, et du club EEA. En particulier, j'ai été membre du comité d'organisation du prix de thèse en Traitement du Signal et des Images du club EEA-GRETSI-GDR ISIS, en 2014-2016, et je suis membre du jury du prix de thèse depuis 2017.

Je suis également membre du groupe de recherche MAORI, faisant partie du projet PGM0 (Gaspard Monge Program for Optimization and operations research) initié par EDF et la fondation pour les mathématiques Jacques Hadarmard.

2.6 LISTE DE PUBLICATIONS

J'ai publié 18 articles (dont 13 sans mes encadrants de thèses) de journaux internationaux et 40 articles de congrès (dont 14 invités et 32 sans mes encadrants de thèse). Ces articles ont été cités, au total, 639 fois, et mon h-index est de 13. Je présente ci-après la liste détaillée de ces publications, en indiquant par le symbole (*) les articles cités plus de 50 fois.

Prix :

J'ai obtenu plusieurs distinctions pour mes articles de journaux/congrès :

- ◇ L'article de conférence de A. Benfenati *et al.*, "A Proximal Approach for Solving Matrix Optimization Problems Involving a Bregman Divergence" a reçu le prix du meilleur poster au workshop BASP 2017, Villars-sur-Ollon, Suisse, 29 janvier - 3 février 2017.
- ◇ L'article de conférence de S. Cadoni *et al.*, "A Block Parallel Majorize-Minimize Memory Gradient Algorithm" a été sélectionné parmi les 7 finalistes du "Best paper award" parmi environ 1000 papiers acceptés à la conférence IEEE International Conference on Image Processing (ICIP 2016), Phoenix, Arizona, USA, 25-28 septembre 2016.
- ◇ L'article de journal de A. Florescu *et al.*, "A Majorize-Minimize Memory Gradient Method for Complex-Valued Inverse Problems", Signal Processing, Vol. 103, pages 285-295, 2014, a obtenu le "Romanian Executive Agency for Higher Education, Research Development and Innovation Award".
- ◇ L'article de conférence de S. Moussaoui *et al.*, "Primal-Dual Interior Point Optimization for Penalized Least Squares Estimation of Abundance Maps in Hyperspectral Imaging" a obtenu le "Best paper award" à la conférence Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS 2012), Shangai, Chine, 4-7 juin 2012.

Articles de journaux :

1. Y. Marnissi, Y. Zheng, **E. Chouzenoux** and J.-C. Pesquet : A Variational Bayesian Approach for Image Restoration. Application to Image Deblurring with Poisson-Gaussian Noise. A paraître dans IEEE Transactions on Image Computing, 2017.
2. 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. A paraître dans Journal of Mathematical Imaging and Vision, 2017.
3. **E. Chouzenoux** and J.-C. Pesquet : A Stochastic Majorize-Minimize Subspace Algorithm for Online Penalized Least Squares Estimation. IEEE Transactions on Signal Processing, Vol. 65, No. 18, pages 4770-4783, 2017.
4. A. Cherni, **E. Chouzenoux** and M.-A. Delsuc : PALMA, an Improved Algorithm for DOSY Signal Processing. Analyst, Vol. 142, No. 5, pages 772-779, 2016.
5. **E. Chouzenoux** and J.-C. Pesquet : Convergence Rate Analysis of the Majorize-Minimize Subspace Algorithm. IEEE Signal Processing Letters, Vol. 23, No. 9, pages 1284-1288, Sept. 2016.
6. (*) **E. Chouzenoux**, J.-C. Pesquet and A. Repetti : A Block Coordinate Variable Metric Forward-Backward Algorithm. Journal of Global Optimization, Vol. 66, No. 3, pages 457-485, 2016 .
7. M. Pereyra, P. Schniter, **E. Chouzenoux**, J.-C. Pesquet, J.-Y. Tourneret, A. Hero and S. McLaughlin : A Survey of Stochastic Simulation and Optimization Methods in Signal Processing. IEEE Journal of Selected Topics in Signal Processing, Vol. 10, No. 2, pages 224-241, Mar. 2016.
8. **E. Chouzenoux**, A. Jeziarska, J.-C. Pesquet and H. Talbot : A Convex Approach for Image Restoration with Exact Poisson-Gaussian Likelihood. SIAM Journal on Imaging Sciences, Vol. 8, No. 4, pages 2662-2682, 2015.
9. (*) A. Repetti, M. Q. Pham, L. Duval, **E. Chouzenoux** and J.-C. Pesquet : Euclid in a Taxicab : Sparse Blind Deconvolution with Smoothed $l_{1/2}$ Regularization. IEEE Signal Processing Letters, Vol. 22, No. 5, pages 539-543, Mai 2015.
10. (*) **E. Chouzenoux**, J.-C. Pesquet and A. Repetti : Variable Metric Forward-Backward Algorithm for Minimizing the Sum of a Differentiable Function and a Convex Function. Journal of Optimization Theory and Applications, Vol. 162, No. 1, pages 107-132, Jul. 2014.
11. A. Florescu, **E. Chouzenoux**, J.-C. Pesquet, P. Ciuciu and S. Ciochina : A Majorize-Minimize Memory Gradient Method for Complex-Valued Inverse Problems. Signal Processing, Vol. 103, pages 285-295, 2014. **Awarded by**

the Romanian Executive Agency for Higher Education, Research Development and Innovation.

12. **E. Chouzenoux**, M. Legendre, S. Moussaoui and J. Idier : Fast Constrained Least Squares Spectral Unmixing using Primal-Dual Interior Point Optimization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7, No. 1, pages 59-69, janvier 2014.
13. **E. Chouzenoux**, S. Moussaoui, M. Legendre and J. Idier : Algorithme primal-dual de points intérieurs pour l'estimation pénalisée des cartes d'abondances en imagerie hyperspectrale. *Traitement du Signal*, Vol. 30, No. 1-2, pages 35-59, 2013.
14. (*) **E. Chouzenoux**, A. Jezierska, J.-C. Pesquet and H. Talbot : A Majorize-Minimize Subspace Approach for l2-l0 Image Regularization. *SIAM Journal on Imaging Science*, Vol. 6, No. 1, pages 563-591, 2013.
15. **E. Chouzenoux**, S. Moussaoui and J. Idier : Majorize-Minimize Linesearch for Inversion Methods Involving Barrier Function Optimization. *Inverse Problems*, Vol. 28, No. 6, 2012.
16. L. Chaari, **E. Chouzenoux**, N. Pustelnik, C. Chaux and S. Moussaoui : OPTIMED : optimisation itérative pour la résolution de problèmes inverses de grande taille. *Traitement du Signal*, Vol. 28, No. 3-4, pages 329-374, 2011.
17. **E. Chouzenoux**, J. Idier and S. Moussaoui : A Majorize-Minimize Strategy for Subspace Optimization Applied to Image Restoration. *IEEE Transactions on Image Processing*, Vol. 20, No. 18, pages 1517-1528, juin 2011.
18. **E. Chouzenoux**, S. Moussaoui, J. Idier and F. Mariette : Efficient Maximum Entropy Reconstruction of Nuclear Magnetic Resonance T1-T2 Spectra. *IEEE Transactions on Signal Processing*, Vol. 58, No. 12, pages 6040-6051, décembre 2010.

Articles de conférences invités :

1. G. Chierchia, A. Cherni, **E. Chouzenoux**, J.-C. Pesquet. Approche de Douglas-Rachford aléatoire par blocs appliquée à la régression logistique parcimonieuse. In Actes du 26e colloque GRETSI (GRETSI 2017), Juan-les-Pins, 5-8 septembre 2017.
2. A. Benfenati, **E. Chouzenoux** and J.-C. Pesquet. A Proximal Approach for Solving Matrix Optimization Problems Involving a Bregman Divergence. In Proceedings of the International Biomedical and Astronomical Signal Processing Frontiers workshop (BASP 2017), Villars-sur-Ollon, Suisse, 29 janvier - 3 février 2017.
3. S. Cadoni, **E. Chouzenoux**, J.-C. Pesquet and C. Chaux. A Block Parallel Majorize-Minimize Memory Gradient Algorithm. In Proceedings of the International Biomedical and Astronomical Signal Processing Frontiers workshop (BASP 2017), Villars-sur-Ollon, Suisse, 29 janvier - 3 février 2017.

4. F. Abboud, **E. Chouzenoux**, J.-C. Pesquet, J.-H. Chenot and L. Laborelli : A Distributed Strategy for Computing Proximity Operators. In Proceedings of 50th Asilomar Conference on Signals, Systems, and Computers, Asilomar, Pacific Grove, CA, 6-9 novembre 2015.
5. A. Repetti, **E. Chouzenoux** and J.-C. Pesquet : Un petit tutoriel sur les méthodes primales-duales proximales pour l'optimisation convexe. In Actes du 25e colloque GRETSI, Lyon, France, 8-11 septembre 2015.
6. A. Repetti, **E. Chouzenoux** and J.-C. Pesquet : A Parallel Block-Coordinate Approach for Primal-Dual Splitting with Arbitrary Random Block Selection. In Proceedings of the 23rd European Signal Processing Conference, pages 235-239, Nice, France, 1-5 septembre 2015.
7. A. Repetti, **E. Chouzenoux** and J.-C. Pesquet : Proximal Primal-Dual Optimization Methods. In Proceedings of the International Biomedical and Astronomical Signal Processing Frontiers workshop (BASP 2014), page 25, Villars-sur-Ollon, Suisse, 25-30 janvier 2015.
8. M. Legendre, S. Moussaoui, **E. Chouzenoux** and J. Idier : Primal-Dual Interior-Point Optimization Based on Majorization-Minimization for Edge Preserving Spectral Unmixing. In Proceedings of the 21st IEEE International Conference on Image Processing (ICIP 2014), pages 4161-4165, Paris, France, 27-30 octobre 2014.
9. **E. Chouzenoux**, J.-C. Pesquet and A. Florescu : A Stochastic 3MG Algorithm with Application to 2D Filter Identification. In Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014), pages 1587-1591, Lisbonne, Portugal, 1-5 septembre 2014.
10. **E. Chouzenoux**, J.-C. Pesquet and A. Florescu : A Multi-Parameter Optimization Approach for Complex Continuous Sparse Modelling. In Proceedings of the 19th International Conference on Digital Signal Processing (DSP 2014), Hong-Kong, Chine, 20-23 août 2014.
11. A. Florescu, **E. Chouzenoux**, J.-C. Pesquet, P. Ciuciu and S. Ciochina : Cramer-Rao Bound for a Sparse Complex Model. In Proceedings of the 10th International Conference on Communications (COMM 2014), 4 p., Bucharest, Romania, 29-31 mai 2014.
12. A. Repetti, **E. Chouzenoux** and J.-C. Pesquet : A Preconditioned Forward-Backward Approach with Application to Large-Scale Nonconvex Spectral Unmixing Problems. In Proceedings of the 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), pages 1498-1502, Florence, Italie, 4-9 mai 2014.
13. **E. Chouzenoux**, S. Moussaoui, J. Idier and F. Mariette : Primal-Dual Interior Point Optimization for a Regularized Reconstruction of NMR Relaxation Time Distributions. In Proceedings of the 38th IEEE International Conference on

Acoustics, Speech, and Signal Processing (ICASSP 2013), pages 8747-8750, Vancouver, Canada, 26-31 mars 2013.

14. **E. Chouzenoux**, S. Moussaoui and J. Idier : A Majorize-Minimize Line Search Algorithm for Barrier Function Optimization. In Proceedings of the 17th European Signal Processing Conference (EUSIPCO 2009), pages 1379-1383, Glasgow, UK, 24-28 août 2009.

Articles de conférences internationales :

1. V. Dудар, G. Chierchia, **E. Chouzenoux**, J.-C. Pesquet, V. Semenov. A Two-Stage Subspace Trust Region Approach for Deep Neural Network Training. In Proceedings of the 25th European Signal Processing Conference (EUSIPCO 2017), Kos Island, Greece, 28 August - 2 September 2017.
2. A. Cherni, **E. Chouzenoux** and M.-A. Delsuc : Proximity Operators for a Class of Hybrid Sparsity+Entropy Priors. Application to DOSY NMR Signal Reconstruction. In Proceedings of the International Symposium on Signal, Image, Video and Communications (ISIVC 2016), 2016.
3. S. Cadoni, **E. Chouzenoux**, J.-C. Pesquet and C. Chaux : A Block Parallel Majorize-Minimize Memory Gradient Algorithm. In Proceedings of the 23rd IEEE International Conference on Image Processing (ICIP 2016), pages 3194–3198, Phoenix, Arizona, USA, 25-28 Sep. 2016. **Finalist of the Best Paper Award.**
4. Y. Marnissi, **E. Chouzenoux**, J.-C. Pesquet and A. Benazza-Benyahia : An Auxiliary Variable Method for Langevin based MCMC algorithms. In Proceedings of the IEEE Workshop on Statistical Signal Processing (SSP 2016), pages 297-301, Palma de Mallorca, Espagne, 26-29 juin 2016.
5. F. Abboud, **E. Chouzenoux**, J.-C. Pesquet, J.-H. Chenot and L. Laborelli : A Dual Block Coordinate Proximal Algorithm with Application to Deconvolution of Interlaced Video Sequences. In Proceedings of the 22nd IEEE International Conference on Image Processing (ICIP 2015), 5 p., Quebec City, Canada, 27-30 septembre 2015.
6. A. Repetti, **E. Chouzenoux** and J.-C. Pesquet : A Random Block-Coordinate Primal-Dual Proximal Algorithm with Application to 3D Mesh Denoising. In Proceedings of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015), pages 3561-3565, Brisbane, Australie, 19-24 avril 2015.
7. A. Repetti, **E. Chouzenoux** and J.-C. Pesquet : A Nonconvex Regularized Approach for Phase Retrieval. In Proceedings of the 21st IEEE International Conference on Image Processing (ICIP 2014), pages 1753-1757, Paris, France, 27-30 octobre 2014.

8. F. Abboud, **E. Chouzenoux**, J.-C. Pesquet, J.-H. Chenot and L. Laborelli : A Hybrid Alternating Proximal Method for Blind Video Restoration. In Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014), pages 1811-1815, Lisbonne, Portugal, 1-5 septembre 2014.
9. Y. Marnissi, A. Benazza-Benyahia, **E. Chouzenoux** and J.-C. Pesquet. Majorize-Minimize Adapted Metropolis Hastings Algorithm. Application to Multichannel Image Recovery. In Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014), pages 1332-1336, Lisbonne, Portugal, 1-5 septembre 2014.
10. A. Florescu, **E. Chouzenoux**, J.-C. Pesquet, P. Ciuciu and S. Ciochina : A Complex-Valued Majorize-Minimize Memory Gradient Method with Application to Parallel MRI. In Proceedings of the 21st European Signal Processing Conference (EUSIPCO 2013), pages 1-5, Marrakech, Maroc, 9-13 septembre 2013.
11. **E. Chouzenoux**, F. Zolyniak, E. Guillard and H. Talbot : A Majorize-Minimize Memory Gradient Algorithm Applied to X-Ray Tomography. In Proceedings of the 20th IEEE International Conference on Image Processing (ICIP 2013), pages 1011-1015, Melbourne, Australie, 15-18 septembre 2013.
12. Y. Marnissi, A. Benazza-Benyahia, **E. Chouzenoux** and J.-C. Pesquet : Generalized Multivariate Exponential Power Prior for Wavelet-Based Multichannel Image Restoration. In Proceedings of the 20th IEEE International Conference on Image Processing (ICIP 2013), pages 2402-2406, Melbourne, Australie, 15-18 septembre 2013.
13. A. Florescu, **E. Chouzenoux**, J.-C. Pesquet and S. Ciochina : A Constrained Optimization Approach for Complex Sparse Perturbed Models. In Proceedings of the Signal Processing with Adaptive Sparse Structured Representations (SPARS 2013), Lausanne, Suisse, 8-11 juillet 2013.
14. **E. Chouzenoux**, F. Zolyniak, E. Guillard and H. Talbot. A Majorize-Minimize Memory Gradient Algorithm Applied to X-Ray Tomography. In Proceedings of ICTMS, Ghent, Belgique, 2 juillet 2013.
15. A. Repetti, **E. Chouzenoux** and J.-C. Pesquet : A Penalized Weighted Least Squares Approach for Restoring Data Corrupted with Signal-Dependent Noise. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012), pages 1553-1557, Bucarest, Roumanie, 27-31 août 2012.
16. S. Moussaoui, **E. Chouzenoux** and J. Idier : Primal-Dual Interior Point Optimization for Penalized Least Squares Estimation of Abundance Maps in Hyperspectral Imaging. In Proceedings of the 4th Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS 2012), Shanghai, Chine, 4-7 juin 2012. **Best Paper Award**.
17. A. Jezierska, **E. Chouzenoux**, J.-C. Pesquet and H. Talbot : A Primal-Dual Proximal Splitting Approach for Restoring Data Corrupted with Poisson-

- Gaussian Noise. In Proceedings of the 37th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), pages 1085-1088, Kyoto, Japon, 25-30 mars 2012.
18. **E. Chouzenoux**, S. Moussaoui and J. Idier : Efficiency of Line Search Strategies in Interior Point Methods for Linearly Constrained Optimization. In Proceedings of the IEEE Workshop on Statistical Signal Processing (SSP 2011), pages 101-104, Nice, France, 28-30 juin 2011.
 19. **E. Chouzenoux**, J.-C. Pesquet, H. Talbot and A. Jezierska : A Memory Gradient Algorithm for l2-l0 Regularization with Applications to Image Restoration. In Proceedings of the 18th IEEE International Conference on Image Processing (ICIP 2011), pages 2717-2720, Bruxelles, Belgique, 11-14 septembre 2011.
 20. **E. Chouzenoux**, S. Moussaoui, J. Idier and F. Mariette : Optimization Algorithm for Maximum Entropy 2d Nuclear Resonance Reconstruction. In Proceedings of the 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010), pages 4154-4157, Dallas, Texas, USA, 14-19 mars 2010.

Articles de conférences nationales :

1. Y. Marnissi, **E. Chouzenoux**, A. Benazza-Benyahia, J.-C. Pesquet and L. Duval : Reconstruction de signaux parcimonieux à l'aide d'un algorithme rapide d'échantillonnage stochastique. In Actes du 25e colloque GRETSI, Lyon, France, 8-11 septembre 2015.
2. **E. Chouzenoux**, L. Lamassé, S. Anthoine, C. Chaux, A. Jaouen, I. Vanzetta and F. Debarbieux : Approche variationnelle pour la déconvolution rapide de données 3D en microscopie biphotonique. In Actes du 25e colloque GRETSI, Lyon, France, 8-11 septembre 2015.
3. A. Repetti, **E. Chouzenoux** and J.-C. Pesquet : Reconstruction d'image en présence de bruit gaussien dépendant par un algorithme Explicite-Implicite à métrique variable. In Actes du 24e colloque GRETSI, Brest, France, 3-6 septembre 2013.
4. A. Florescu, **E. Chouzenoux**, J.-C. Pesquet and S. Ciochina : Estimation d'un signal complexe à partir d'un modèle parcimonieux perturbé. In Actes du 24e colloque GRETSI, Brest, France, 3-6 septembre 2013.
5. **E. Chouzenoux**, S. Moussaoui and J. Idier : Algorithme primal-dual de points intérieurs pour l'estimation pénalisée des cartes d'abondances en imagerie hyperspectrale. In Actes du 23e colloque GRETSI, Bordeaux, France, 5-8 septembre 2011.
6. **E. Chouzenoux**, S. Moussaoui, J. Idier and F. Mariette : Reconstruction d'un spectre RMN 2D par maximum d'entropie. In Actes du 22e colloque GRETSI, Dijon, France, 8-11 septembre 2009

- Chapitre 3 -

Activités d’enseignement

3.1 HEURES ENSEIGNÉES

Je décris dans cette section les matières que j’ai enseigné depuis le début de ma thèse. Je fournis également un descriptif des cours dont j’ai été responsable. Mon activité d’enseignement est très réduite depuis septembre 2016, du fait de mon statut de chercheuse en délégation INRIA, me faisant bénéficier d’une décharge complète d’enseignement. La liste détaillée des heures enseignées est présentée dans le Tableau 3.1.

3.1.1 Ecole CentraleSupélec

Depuis 2014, je suis co-responsable du cours “Foundations of Distributed and Large Scale Computing Optimization” (avec Jean-Christophe Pesquet), s’insérant dans le cadre du master “Data Sciences” de l’Ecole CentraleSupélec (également inclus dans le cursus MVA ENS Cachan depuis 2016). L’objectif de ce cours est d’introduire les fondements théoriques nécessaires au développement d’algorithmes efficaces pour la résolution de problèmes d’optimisation de grande dimension. Le cours se focalise principalement sur les méthodes d’optimisation non linéaire pour la minimisation de fonctions convexes. Les approches proximales, les stratégies primales-duales et les méthodes de majoration-minimisation sont présentées. Les étudiants mettent en application les notions méthodologiques introduites en cours sur des exemples pratiques, issus du traitement du signal/image et de l’apprentissage, étudiés lors des séances de TP sous Python. Je participe à la rédaction des supports de cours, et des sujets de TP, et d’examen.

3.1.2 Université Paris Est Créteil

Depuis 2013, je suis responsable du cours “Restauration d’images” du Master 2 Signaux et images en médecine de l’UPEC. Ce cours vise à fournir un ensemble d’outils mathématiques permettant le traitement numérique des images médicales. Après avoir introduit le concept de problème inverse en traitement d’image, le cours décrit la construction des termes d’attache aux données et de régularisation d’un critère pénalisé et plusieurs algorithmes d’optimisation pour le minimiser. Une séance de TP sous Matlab permet aux étudiants de tester les algorithmes vus en cours sur un problème de déconvolution d’image, et un problème de reconstruction en tomographie. J’ai rédigé les supports de ce cours, les sujets de TP et le sujet d’examen.

3.1.3 IUT de Marne-la-Vallée

Depuis 2011, je suis responsable du module “Culture scientifique” du DUT Métiers du Multimédia et de l’Internet, du site de Meaux de l’IUT de Marne-la-Vallée. Au premier semestre de ce cours, les notions de signal sinusoïdal, série de Fourier et filtrage analogique sont introduites et illustrées sur un exemple concret de filtrage de signaux sonores, à travers des TPs sous Matlab. Au deuxième semestre, le cours présente le principe d’échantillonnage de Shannon et introduit les notions de transmission de signaux numériques par codage bande de base et bande transposée. Enfin, au troisième semestre, le cours a pour objectif la maîtrise de l’utilisation des codecs audio-vidéo et la connaissance des différents supports et standards de stockage et de diffusion. On s’intéresse en particulier aux techniques de compression JPEG, MPEG et MP3. J’ai rédigé les supports de cours/TD/TP, et les sujets d’examen.

3.1.4 Université Paris-Est Marne-la-Vallée

Lors de mon ATER en 2010-2011, j’ai enseigné dans plusieurs formations de l’UPEM, telles que l’IMAC et l’ESIPE, pour une charge d’enseignement totale de 192 heures équivalent TD. Depuis mon recrutement en tant que maître de conférences à l’UPEM en septembre 2011, j’enseigne principalement à l’IUT (voir plus haut). J’ai pris en complément la responsabilité de deux cours à l’UPEM sur le site de Champs sur Marne, que je décris ci-dessous.

Entre 2010 et 2013, j’étais responsable du cours “Traitement du signal aléatoire” enseigné aux étudiants de deuxième année de l’école d’ingénieur ESIPE de l’UPEM. L’objectif de ce cours est la mise en évidence à l’aide de multiples exemples, des concepts, des outils, des applications et des perspectives du traitement du signal aléatoire. A l’issue de ce cours, l’étudiant sait définir un processus aléatoire, énumérer les grandes classes de signaux aléatoires et manipuler les outils du domaine (valeur moyenne, fonction de corrélation, DSP). Les travaux pratiques permettent d’illustrer les points évoqués en cours à travers l’utilisation du logiciel de simulation

SCILAB. J'ai rédigé un polycopié, des transparents, des textes de TD et d'examen pour ce cours donné à la formation ESIPÉ.

Depuis 2012, je suis responsable du cours "Représentation et filtrage numérique" du Master 2 Informatique, filière Sciences de l'Image, de l'UPEM. Cet enseignement permet aux élèves, à partir de l'étude de signaux réels et de systèmes physiques, d'acquérir les notions de bases sur l'analyse temporelle ou fréquentielle des signaux, la modélisation des systèmes dynamiques, l'échantillonnage des signaux et systèmes, la synthèse de filtres. Les séances de travaux dirigés se décomposent en deux parties. Tout d'abord, les notions étudiées en cours sont mises en pratique à travers l'utilisation du logiciel de simulation SCILAB. La deuxième partie consiste en l'implémentation en langage C/C++ d'un compresseur d'image de type JPEG2000. Cette partie permet d'aborder les notions suivantes : Transformée en ondelettes monodimensionnelle et multidimensionnelle, schémas de bancs de filtres et lifting, compression par allocation de débit. J'ai participé à la rédaction du polycopié, des sujets de TD, et des sujets d'examens de cette matière.

3.1.5 Ecole Centrale de Nantes

Je suis intervenue pendant 3 ans (2007 – 2010) au département d'enseignement "Informatique et Mathématiques" de l'Ecole Centrale de Nantes en qualité de monitrice pour une charge d'enseignement totale de 200 heures équivalent TD. J'ai participé à la rédaction d'un sujet de TD sur l'optimisation contrainte, pour le module "optimisation". De plus, j'ai conçu les supports du cours/TD pour le groupe que j'avais en charge pour le module "algorithmique et programmation".

3.2 RESPONSABILITÉS ADMINISTRATIVES

3.2.1 IUT de Marne la Vallée

Depuis mon recrutement comme maître de conférence à l'UPEM, j'ai été très impliquée dans la gestion de la formation DUT Métiers du Multimédia et de l'Internet (MMI) de l'IUT de Marne la Vallée, site de Meaux. Entre 2012 et 2013, j'étais responsable de la gestion des stages et des projets tutorés du DUT, et depuis Mars 2012, je suis **Chef du département** de la formation. Cette activité regroupe les tâches suivantes : Gestion du budget du DUT, gestion des ressources humaines, gestion du recrutement des étudiants via le site APB, organisation et animation des conseils pédagogiques, organisation de la journée portes ouvertes et de la journée d'accueil des nouveaux entrants, participation aux jurys et aux conseils de direction de l'IUT, organisation de l'évaluation quadriennale du DUT, mise en place des nouveaux programmes pédagogiques nationaux, organisation de l'assemblée des chefs de département et du festival MMI Meaux 2017. Afin de mener à bien ces tâches, je

bénéficie d'une décharge d'enseignement de 96h (équivalents TD).

3.2.2 Tutorat de stages/projets

Je liste ci-dessous mes activités de tutorat de projets et de stages étudiants :

- ◇ Depuis juin 2017, je suis la tutrice du stage d'un étudiant du Master MVA de l'ENS Cachan.
- ◇ En 2017, j'ai suivi trois groupes d'étudiants de 3ème année de CentraleSupélec, option Mathématiques Appliquées, dans le cadre de leur séminaires d'initiation à la recherche.
- ◇ Entre 2011 et 2015, j'ai suivi deux groupes de projets tutorés du DUT MMI par an.
- ◇ Entre 2011 et 2015, j'ai suivi trois stagiaires de DUT MMI 2^e année par an.
- ◇ En mai-juin 2011, j'ai encadré deux stagiaires du M1 Electronique de l'UPEM, sur un projet portant sur le démélange d'images hyperspectrales.
- ◇ En 2010, j'ai encadré un groupe de projet pour le module "Recherche et développement" de l'Ecole Centrale de Nantes.
- ◇ Entre 2007 et 2010, j'ai encadré plusieurs groupes d'étudiants dans le cadre de leurs projets de programmation C (module "Algorithmique et Programmation") et java (module "Méthodes Logicielles") de l'Ecole Centrale de Nantes.
- ◇ En 2008 et en 2009, j'ai suivi des stagiaires de 2^e année de l'Ecole Centrale de Nantes.

Année	Formation	Intitulé	Volume total		
			CM	TD	TP
2014 – 2017	CentraleSupélec 3 ^e année / MSc in Data Sciences and Business Analytics / M2 MVA Cachan	Foundations of Distributed and Large Scale Computing Optimization	36	36	
2013 – 2016	Université Paris Est Créteil, M2 Signaux et images en médecine	Restauration d'images	9		9
2011 – 2016	IUT de Marne la Vallée, DUT MMI Meaux	Culture scientifique S1	30	60	64
		Culture scientifique S2	44	84	
		Culture scientifique S3	32	12	
2011 – 2016	Université Paris Est Marne la Vallée, M2 Informatique - Sciences de l'Image	Représentation et filtrage numérique	40	20	
2010 – 2013	Ingénieurs ESIPÉ 2 ^{ème} année	Traitement du signal aléatoire	54		
2010 – 2012	Ecole d'ingénieurs IMAC	Probabilités IMAC2		34	
2010 – 2011	Université Paris Est Marne la Vallée, L3 EEA-Meca	Traitement du signal		24	
		Traitement d'image	3	3	
	Université Paris Est Marne la Vallée, M1 ETG	Traitement du signal aléatoire		9	
		Traitement du signal IMAC1		22	
	Ecole d'ingénieurs IMAC	Compression d'image IMAC3		28	
		Traitement d'image IMAC3		18	
2007-2009	Ecole Centrale de Nantes	Optimisation Ei1		31	
		Automatismes Ei1			44
		Electronique Ei2			40
		Algorithmique Ei1		23	21.5
		Méthodes Logicielles Ei2			37.5

Table 3.1 – *Récapitulatif des heures enseignées*

Deuxième partie

SYNTHÈSE DES ACTIVITÉS DE RECHERCHE

- Chapitre 4 -

Introduction générale

PRÉAMBULE

Dans cette partie du document, je présente une synthèse de mes thématiques de recherche, et des résultats obtenus, depuis ma soutenance de thèse en décembre 2010. Ces travaux ont été réalisés, pour leur grande majorité, dans le cadre des thèses de AUDREY REPETTI, YOSRA MARNISSI et FERIEL ABOUD. Ils ont été consacrés au développement de nouvelles approches algorithmiques efficaces pour les problèmes de minimisation en grande dimension qui se posent lors de la résolution des problèmes inverses en traitement du signal et des images. Dans un tel contexte, la difficulté réside à la fois dans la forme complexe des fonctionnelles à minimiser (ex : fonctions non convexes, fonctions non lisses), et dans le très grand nombre de variables à gérer. Les contributions présentées dans ce manuscrit ont pour objectif de répondre à ces problématiques, en utilisant le cadre unificateur des approches par majoration-minimisation (MM).

4.1 THÉMATIQUE DES RECHERCHES

Considérons le problème inverse de l'estimation d'un signal vectoriel $x \in \mathbb{R}^N$ à partir d'un vecteur de données $y \in \mathbb{R}^M$. La résolution du problème inverse passe tout d'abord par l'élaboration d'un modèle direct liant x et y à partir d'équations de la physique. Mes travaux de recherche se placent le plus souvent dans le cadre usuel d'un modèle linéaire $y = Hx + w$, où $H \in \mathbb{R}^{N \times M}$ est la matrice d'observation décrivant le processus de mesure et $w \in \mathbb{R}^M$ un vecteur de perturbations représentant à la fois le bruit de mesure et les erreurs de modélisation. Les problèmes inverses rencontrés en traitement du signal et de l'image sont intrinsèquement instables du fait du mauvais conditionnement de la matrice H . Par conséquent, une

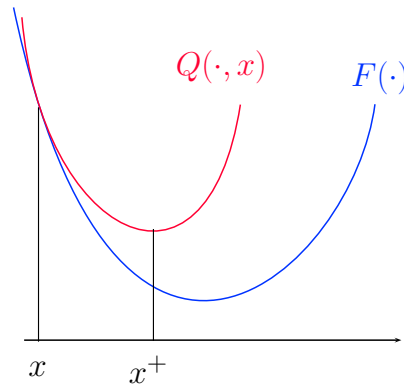


Figure 4.1 – *Principe MM* : $Q(\cdot, x)$ est une fonction tangente majorante de F en x . L'itérée MM x^+ est définie comme le minimiseur de $y \mapsto Q(y, x)$, ce qui conduit à $F(x^+) \leq F(x)$.

stratégie de résolution naïve à partir de la seule connaissance de l'observation y est vouée à l'échec. Une stratégie efficace consiste à renforcer certaines propriétés souhaitables de la solution sur la base de connaissances *a priori*. Cette approche, appelée approche pénalisée, définit la solution du problème comme le minimiseur d'un critère composite qui prend en compte conjointement les données observées et les informations préalables sur la solution. Résoudre le problème inverse relève donc formellement d'un problème d'optimisation de grande taille de la forme :

$$\underset{x \in \mathbb{R}^N}{\text{minimiser}} F(x), \quad (4.1)$$

où $F : \mathbb{R}^N \mapsto]-\infty, +\infty]$ est une fonction non nécessairement lisse et non nécessairement convexe. La fonction F s'exprime généralement comme une fonction composite, sommant un terme d'attache aux données décrivant les relations existant entre les données observées et les grandeurs à estimer, et des termes de pénalisation (ou régularisation) modélisant des connaissances *a priori* sur les inconnues du problème. Nous en donnerons plusieurs exemples concrets tout au long de ce manuscrit.

Le minimiseur du critère pénalisé F ne peut généralement pas s'exprimer sous une forme analytique et un algorithme de minimisation doit être utilisé pour en fournir une estimée. Parmi les différentes stratégies possibles pour construire cet algorithme, les méthodes MM reposent sur l'idée naturelle qu'il est fondamental, pour minimiser efficacement F , de tenir compte du maximum d'informations disponibles sur sa structure, plutôt que d'utiliser un solveur générique. Dans les algorithmes MM, la structure de F est exploitée à travers le principe d'approximation *tangente majorante*. A chaque itération d'un algorithme MM, une fonction tangente majorante i.e., qui majore F sur \mathbb{R}^N et est égale à celle-ci à l'itérée courante, est construite. L'itérée suivante est ensuite déduite, en minimisant la fonction tangente majorante (Fig. 4.1). Il en résulte une suite d'itérées qui font décroître, par construction, la fonction objectif à chaque itération de l'algorithme.

La principale difficulté à surmonter est le volume croissant des données qui entraîne un coût de calcul rédhibitoire lors de la résolution numérique du problème d'optimisation. Le traitement d'images a pour spécificité d'être presque systématiquement confronté à des problèmes de grande tailles. En effet, le nombre d'inconnues à estimer correspond au nombre de pixels et est par conséquent très important (par exemple, $N \geq 10^7$ variables, en imagerie 3D). Bien que la puissance des calculateurs augmente régulièrement, traiter de tels volumes de données en un temps raisonnable reste un défi majeur. Mes travaux de recherche portent sur le développement de méthodes d'optimisation basées sur des schémas MM permettant de résoudre efficacement des problèmes inverses de grande taille tout en offrant des garanties de convergence.

4.2 RÉSUMÉ DES TRAVAUX DE THÈSE

4.2.1 Contexte

Mon travail de thèse s'est déroulé entre septembre 2007 et décembre 2010, au sein de l'équipe Analyse et Décision en Traitement du Signal et de l'Image de l'IRCCyN (Nantes), sous la direction de Jérôme Idier et Saïd Moussaoui. Ma thèse s'intitule "Recherche de pas par Majoration-Minoration. Application à la résolution de problèmes inverses". Elle était financée par une bourse ministérielle, et était également intégrée dans le projet ANR OPTIMED (2006-2009), dont l'objectif était le développement d'algorithmes d'optimisation efficaces pour les problèmes d'imagerie médicale de grande taille [39].

4.2.2 Résultats obtenus

Mon travail de thèse a porté sur les approches itératives de résolution de problèmes d'optimisation différentiables issus de problèmes inverses en signal/image. Je me suis plus précisément focalisée sur la mise en œuvre algorithmique de l'étape de recherche de pas dans les algorithmes de descente itérative. L'objectif de mon travail de thèse était de proposer et d'étudier de nouvelles méthodes de recherche de pas par approche MM.

Dans la première partie de ma thèse, je me suis intéressée à la minimisation de critères contenant une barrière, c'est-à-dire une fonction dont le gradient est non borné, intervenant par exemple lors de la restauration d'images dégradées par un bruit de Poisson. Dans ce contexte, les procédures de recherche de pas MM quadratiques ne sont pas applicables. Ma contribution a été de développer une nouvelle stratégie de recherche de pas tenant compte de la singularité de la barrière à travers la proposition d'approximations majorantes quadratiques augmentées d'un terme logarithmique. La recherche de pas résultante est simple à mettre en œuvre et j'ai

démontré qu'elle garantit la convergence dans le cas convexe de plusieurs algorithmes standards de descente itérative tels que le gradient conjugué non linéaire. J'ai aussi montré sur plusieurs exemples que la méthode proposée présente de bonnes performances numériques par rapport aux stratégies concurrentes issues de la littérature des méthodes de points intérieurs. Ce travail a mené aux publications [56, 55, 53]. L'approche proposée a de plus conduit à l'élaboration d'une stratégie de reconstruction de spectres de relaxométrie issus de l'analyse RMN multidimensionnelle, à travers un travail en collaboration avec l'équipe *IRM Food* du CEMAGREF. Ces travaux ont fait l'objet des publications [58, 59, 57]. Enfin, la recherche de pas proposée dans le contexte des algorithmes de points intérieurs a été appliquée, à la suite de ma thèse, au problème d'estimation des cartes d'abondance en imagerie hyperspectrale, ce qui a conduit aux publications [52, 61, 60, 54].

Dans la deuxième partie de ma thèse, je me suis intéressée aux algorithmes de minimisation par sous-espace, qui consistent à explorer le critère non plus le long de directions de descente mais à l'intérieur de sous-espaces engendrés par plusieurs directions. Ma contribution a consisté à étendre la recherche de pas MM quadratique scalaire issue de [128] afin de pouvoir l'implanter au sein d'un algorithme de sous-espace. La nouvelle méthode est basée sur des approximations majorantes quadratiques du critère à l'intérieur du sous-espace. Elle donne lieu à un algorithme d'optimisation itératif élégant, avec une règle de mise à jour simple qui contraste avec les algorithmes de minimisation différentiable classiques (ex : gradient conjugué non linéaire [106], quasi-Newton L-BFGS [134]) nécessitant généralement une boucle interne de recherche de pas. J'ai montré que cette méthode bénéficiait en outre de garanties théoriques dans le cas convexe et montrait de très bonnes performances pratiques sur plusieurs problèmes de restauration/reconstruction d'image. Ce travail a été publié dans [49].

4.3 PROBLÉMATIQUE

Partant des résultats obtenus dans le cadre de ma thèse, et en particulier de la deuxième partie de celle-ci, de nombreuses questions méthodologiques se posaient, auxquelles j'ai cherché à répondre à travers mes travaux de recherche :

- ↪ comment construire des approximations majorantes quadratiques fiables ?
- ↪ quelle est la vitesse théorique de convergence des méthodes MM quadratiques ?
- ↪ quelles sont leurs garanties de convergence dans le cadre non convexe ?
- ↪ ces méthodes peuvent-elles résoudre des problèmes d'optimisation stochastique ?
- ↪ peut-on établir des points entre le cadre MM et l'optimisation non lisse, où les algorithmes proximaux prédominent ?

- ↪ comment gérer un nombre très important de variables et les problèmes de mémoire associés ?
- ↪ les approches MM peuvent-elles s'intégrer dans des méthodes de simulation stochastique en grande dimension ?

L'étude de ces aspects m'a mené à des contributions théoriques et aussi à la proposition de nouveaux algorithmes efficaces et rapides pour plusieurs problèmes applicatifs dans le domaine des problèmes inverses en signal/image, tels que la reconstruction d'images en imagerie par résonance magnétique parallèle, la déconvolution d'images microscopiques, et la restauration d'archives vidéos. Cette partie du manuscrit fournit une synthèse de ces différents résultats, et se structure ainsi : Le chapitre 5 introduit le principe de fonctions majorantes quadratiques, décrit trois techniques pour les construire, et illustre leur applicabilité sur des travaux que j'ai réalisés dans le contexte de la reconstruction d'images. Le chapitre 6 présente les résultats de convergence que j'ai établis dans le cadre de la minimisation non convexe pour deux familles d'algorithmes. Tout d'abord, je m'intéresse à la convergence d'algorithmes proximaux de type explicites-implicites, intégrant une métrique variable de type MM. Puis les algorithmes MM quadratiques par sous-espace sont considérés. Une extension stochastique de cette deuxième approche que j'ai récemment proposée est également présentée. Le chapitre 7 est consacré à mes contributions dans le contexte des approches de minimisation alternées par blocs. Je présente l'algorithme explicite-implicite à métrique variable alterné par blocs, ses propriétés de convergence et illustre son efficacité dans le contexte de la déconvolution aveugle. Une version duale de cette approche est également décrite, et son applicabilité est illustrée sur un problème de restauration de vidéos. Enfin, je termine ce chapitre en présentant un travail récent que j'ai effectué, portant sur la proposition d'une version parallèle de l'algorithme MM sous-espace, et son application à un exemple de déconvolution d'images microscopiques. Le chapitre 8 s'intéresse aux algorithmes semi-quadratiques. Après quelques rappels théoriques sur ces méthodes, je présente deux applications de leur principe sous-jacent d'augmentation de données issus de mes travaux dans le contexte de la restauration d'images dans un cadre bayésien. Enfin, dans le chapitre 9, je présente plusieurs perspectives, méthodologiques et applicatives, de ces différents travaux, et conclus ce manuscrit. Le lecteur est amené à se rediriger vers mes publications pour plus de détails techniques (ex : preuves mathématiques, mise en œuvre algorithmique) sur les travaux présentés.

- Chapitre 5 -

Approximation majorante quadratique

5.1 INTRODUCTION

Dans une approche d'optimisation de type MM, se succèdent des minimisations d'approximations majorantes de la fonction à minimiser [111]. Il est alors clair que l'efficacité d'une telle approche réside en deux aspects fondamentaux, à savoir (i) la possibilité de réaliser ces minimisations d'une façon peu coûteuse en complexité de calcul, (ii) la qualité d'approximation des fonctions majorantes employées. Un choix simple permettant un compromis idéal entre ces deux aspects est d'adopter une stratégie MM quadratique, c'est à dire de se baser sur des fonctions majorantes quadratiques. La condition de tangence (généralement) imposée dans les algorithmes MM fournit la forme d'une telle approximation quadratique de F en $x' \in \mathbb{R}^N$, lorsque F est supposée différentiable :

$$(\forall x \in \mathbb{R}^N) \quad Q(x, x') = F(x') + \nabla F(x')^\top (x - x') + \frac{1}{2}(x - x')^\top A(x')(x - x'), \quad (5.1)$$

où $A(x') \in \mathbb{R}^{N \times N}$ est une matrice symétrique, définie, positive (ainsi le minimiseur de $Q(\cdot, x')$ est bien défini et unique) telle que la propriété de majoration est vérifiée, c'est à dire

$$(\forall x \in \mathbb{R}^N) \quad Q(x, x') \geq F(x). \quad (5.2)$$

Il est intéressant de noter que (5.1) s'apparente à un développement de type Taylor au second ordre de F au voisinage de x' , avec un terme de second ordre modifié afin de vérifier la propriété de majoration (5.2). Nous verrons dans les chapitres suivants plusieurs approches se basant sur cette idée de fonction majorante quadratique. Leur mise en œuvre repose sur la capacité à construire des fonctions majorantes quadratiques, c'est à dire à savoir choisir, étant donnée une fonction F avec une structure spécifique, les matrices $A(x)$, $x \in \mathbb{R}^N$. Je présente dans ce chapitre les trois

techniques principales de construction que j’ai utilisées, et j’illustre leur applicabilité sur des exemples issus de mes travaux dans le contexte de la restauration d’images issus de mes travaux de recherche.

5.2 PROPRIÉTÉ DE LIPSCHITZ DIFFÉRENTIABILITÉ

Une façon simple de construire une fonction majorante quadratique est d’utiliser le “lemme de descente” [71], qui permet de majorer une fonction β -Lipschitz différentiable sur un domaine convexe par une fonction quadratique de courbure égale à $\mu \geq \beta$. Il est aisé de voir que l’application du schéma itératif MM à une telle fonction majorante conduit à un algorithme de descente de gradient à pas fixe μ^{-1} . Cet algorithme est connu pour sa vitesse de convergence très lente, ce qui se comprend bien puisqu’il est basé sur une approximation majorante assez rudimentaire du critère. Néanmoins, la stratégie de majoration par le lemme de descente peut s’avérer très utile, en particulier pour majorer des fonctions très complexes (Ex. 5.2.1) ou pour construire des algorithmes de type explicite-implicite pour l’optimisation non lisse (Sec. 6.2).

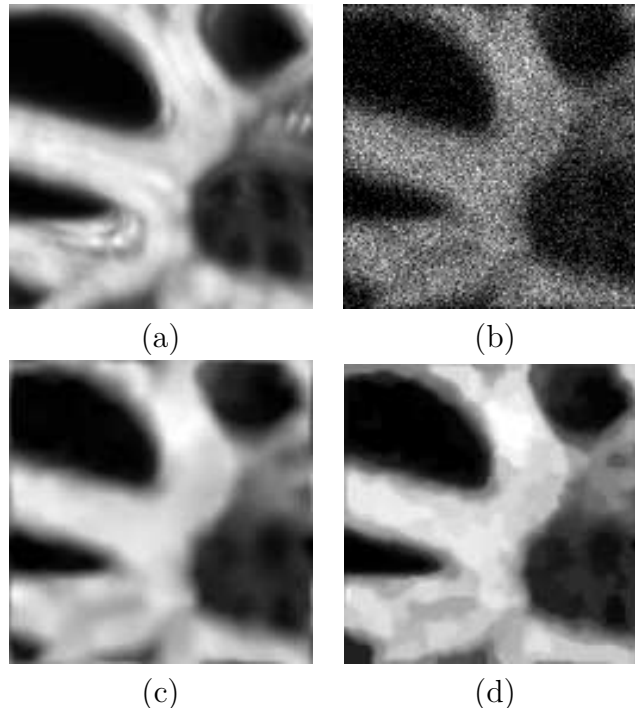


Figure 5.1 – (a) Image originale 128×128 ; (b) Image dégradée par un flou gaussien et un bruit Poisson-Gaussien; (c) Image restaurée par l’approche [51], avec une pénalisation hybride Hessien + TV; (d) Image restaurée par l’approche [51], avec une pénalisation TV non locale.

Exemple 5.2.1 Je présente ici le travail réalisé dans le cadre du projet ANR DIAMOND, s'intéressant à la restauration d'images en microscopie confocale, que nous avons publié dans [51, 117]. Nous considérons la restauration d'une image, dégradée par un opérateur linéaire et un bruit mixte Poisson-Gaussien. Ce modèle de bruit illustre la conjonction d'un bruit électronique de capteur (gaussien) et d'un bruit de comptage de photons (poissonien). Il se retrouve dans de nombreuses applications pratiques, incluant la microscopie confocale mais aussi l'imagerie satellitaire et la photographie numérique [21, 83, 149]. Dans un tel modèle, l'image dégradée notée $y \in \mathbb{R}^N$ est liée à l'image originale $\bar{x} \in \mathbb{R}^N$ par le modèle $y = z + w$ où $z \in \mathbb{R}^N$ et $w \in \mathbb{R}^N$ sont des vecteurs aléatoires mutuellement indépendants, tels que

$$z \sim \mathcal{P}(H\bar{x}), \quad w \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_M), \quad (5.3)$$

où \mathcal{P} est la distribution de Poisson, $H \in [0, +\infty[^{M \times N}$ est un opérateur linéaire de dégradation à entrées positives (ici, un flou), \mathcal{N} est la distribution normale, $\sigma > 0$ et \mathbf{I}_M la matrice identité de \mathbb{R}^M . La fonction de neg-log-vraisemblance associée à ce modèle de bruit, employée par exemple pour calculer l'estimateur par maximum *a posteriori* (MAP), s'écrit, pour tout $x \in [0, +\infty[^N$:

$$F(x) = - \sum_{m=1}^M \left(\sum_{i=1}^{+\infty} \frac{e^{-[Hx]_m} ([Hx]_m)^i e^{-\frac{1}{2\sigma^2}(y_m-i)^2}}{i! \sqrt{2\pi\sigma^2}} \right). \quad (5.4)$$

La fonction (5.4) est différentiable sur l'orthant positif. Nous avons établi dans [51] sa convexité et son caractère gradient β -Lipschitz différentiable, avec :

$$\beta = \|H\|^2 (1 - e^{-\frac{1}{\sigma^2}}) \max_{m \in \{1, \dots, M\}} e^{\frac{2y_m - 1}{\sigma^2}}. \quad (5.5)$$

Ce résultat nous a conduit à la proposition d'une approche efficace pour la restauration d'images dégradées par un bruit Poisson-Gaussien [51]. Cette méthode est une approche variationnelle basée sur l'estimateur MAP, dans laquelle la minimisation est effectuée par une méthode proximale primale-duale flexible au choix du terme de pénalisation et robuste aux erreurs numériques provenant de l'évaluation du gradient de (5.4). Je donne en Figure 5.1 quelques résultats obtenus dans le cadre de ce travail. Je présente de plus dans le Tableau 5.1 extrait de notre article [142] d'autres exemples de termes d'attaches aux données F de gradient Lipschitz qui peuvent être utilisés pour la restauration d'images en présence de bruit Poisson-Gaussien, Gaussien pur, Poisson pur, ou Cauchy.

5.3 POTENTIELS RENFORÇANT LA PARCIMONIE

Dans le cadre de la résolution des problèmes inverses mal posés, il est nécessaire d'introduire des informations *a priori* sur la solution recherchée afin d'obtenir une

Nom	$f_m(v; y_m)$	$\dot{f}_m(v; y_m)$	$\beta_m(y_m)$	Domaine de validité	Modèle de bruit
Gaussien	$\frac{1}{2\sigma^2}(v - y_m)^2$	$\frac{1}{\sigma^2}(v - y_m)$	$\frac{1}{\sigma^2}$	$y_m \in \mathbb{R}, \sigma > 0$	Gaussien
Cauchy	$\ln\left(1 + \frac{(v - y_m)^2}{\sigma^2}\right)$	$\frac{2(v - y_m)}{\sigma^2 + (v - y_m)^2}$	$\frac{2}{\sigma^2}$	$y_m \in \mathbb{R}, \sigma > 0$	Cauchy
Transformée d'Anscombe	$2\left(\sqrt{y_m + \frac{3}{8}} - \sqrt{v + \frac{3}{8}}\right)^2$	$2 - \frac{2\sqrt{y_m + \frac{3}{8}}}{\sqrt{v + \frac{3}{8}}}$	$\left(\frac{3}{8}\right)^{-3/2} \sqrt{y_m + \frac{3}{8}}$	$y_m \geq -\frac{3}{8}$	Poisson
Transformée d'Anscombe généralisée	$2\left(\sqrt{y_m + \sigma^2 + \frac{3}{8}} - \sqrt{v + \sigma^2 + \frac{3}{8}}\right)^2$	$2 - \frac{2\sqrt{y_m + \frac{3}{8} + \sigma^2}}{\sqrt{v + \frac{3}{8} + \sigma^2}}$	$\left(\frac{3}{8} + \sigma^2\right)^{-3/2} \sqrt{y_m + \frac{3}{8} + \sigma^2}$	$y_m \geq -\frac{3}{8} - \sigma^2$	Poisson-Gaussien
Distance de Kullback décalée	$(v + \sigma^2) - (y_m + \sigma^2) \ln(v + \sigma^2)$	$1 - \frac{y_m + \sigma^2}{v + \sigma^2}$	$\frac{y_m + \sigma^2}{\sigma^4}$	$y_m \geq -\sigma^2, \sigma > 0$	Poisson-Gaussien
Moindres carrés pondérés	$\frac{(y_m - v)^2}{2(\sigma^2 + v)} + \frac{1}{2} \ln(\sigma^2 + v)$	$\frac{1}{2} - \frac{(y_m + \sigma^2)^2}{2(v + \sigma^2)^2} + \frac{1}{2(\sigma^2 + v)}$	$\max\left\{\frac{(y_m + \sigma^2)^2}{\sigma^6} - \frac{1}{2\sigma^4}, \frac{1}{54(y_m + \sigma^2)^4}\right\}$	$y_m \in \mathbb{R} \setminus \{-\sigma^2\}, \sigma > 0$	Poisson-Gaussien

Table 5.1 – Exemples de fonctions d'attache aux données $F(x) = \sum_{m=1}^M f_m([Hx]_m; y_m)$ avec f_m à gradient Lipschitz. \dot{f}_m est la dérivée première de f_m et $\beta_m(y_m)$ est la constante de Lipschitz de \dot{f}_m (pour les fonctions en lignes 3-6, f_m peut être remplacée sur \mathbb{R}_- par son développement de Talor au 2ème ordre en zéro, afin de garantir la propriété Lipschitz sur \mathbb{R} .) La constante de Lipschitz β de ∇F se déduit de $(\beta_m(y_m))_{1 \leq m \leq M}$ par $\beta = \|H\|^2 \max_{m \in \{1, \dots, M\}} \beta_m(y_m)$.

méthode de reconstruction robuste au bruit. Supposons que l'on souhaite renforcer une propriété de parcimonie de la solution recherchée dans un domaine transformé déduit d'un opérateur $V \in \mathbb{R}^{P \times N}$, $P \geq 1$ (par exemple, V peut être un opérateur d'ondelettes à l'analyse [157]). On prendra alors une fonction de pénalisation de la forme :

$$(\forall x \in \mathbb{R}^N) \quad F(x) = \sum_{p=1}^P f([Vx]_p), \quad (5.6)$$

où $f : \mathbb{R} \mapsto \mathbb{R}$ est choisie de telle façon à favoriser la propriété de parcimonie souhaitée. En particulier, la (pseudo) norme ℓ_0 est obtenue en posant $f(u)$ égale à 1 pour $u = 0$, zéro sinon [89]. Ce choix renforce de façon explicite la parcimonie de la solution dans le domaine transformé induit de V . Il donne lieu à un problème d'optimisation complexe, car non convexe et non continu. Dans certaines applications, l'usage de la pénalité ℓ_0 est nécessaire à l'obtention de solutions satisfaisantes. Nous l'avons illustré par exemple dans nos travaux [94, 64], proposant une approche proximale non convexe pour l'analyse spectrale en haute résolution. Des choix alternatifs pour la fonction de pénalité f sont aussi possibles, la stratégie la plus utilisée étant probablement de prendre f égale à la fonction valeur absolue, ce qui conduit à une pénalisation de type norme ℓ_1 . Intéressons nous à une famille particulière de fonctions f paires et différentiables, renforçant la parcimonie. Ces pénalités peuvent être vues comme des approximations lisses et continues, non nécessairement convexes, de la norme ℓ_0 et bénéficient de la propriété de majoration suivante [113] :

Propriété 5.3.1 Soit $f : \mathbb{R} \mapsto \mathbb{R}$. f est continue sur \mathbb{R} , et pour tout $t \in \mathbb{R}$, on a $f(t) = \psi(|t|)$, avec :

- ψ différentiable sur $]0, +\infty[$,
- $\psi(\sqrt{\cdot})$ concave sur $]0, +\infty[$,
- ψ croissante sur $]0, +\infty[$,

- $\lim_{\substack{t \rightarrow 0 \\ t > 0}} \left(\omega(t) := \frac{\psi(t)}{t} \right) \in \mathbb{R}$.

La fonction quadratique (5.1) est une fonction majorante de (5.6) en $x' \in \mathbb{R}^N$ si :

$$(\forall x \in \mathbb{R}^N) \quad A(x) = V^\top \text{Diag}(\omega(|Vx|))V. \quad (5.7)$$

Le Tableau 5.2 issu de [63] répertorie plusieurs fonctions f vérifiant les hypothèses de la Propriété 5.3.1. La famille de pénalités ainsi définies est très intéressante, car elle inclut des potentiels convexes de forme $\ell_2 - \ell_1$ (comportement quadratique en 0, et linéaire à l'infini) approchant le potentiel ℓ_1 , et aussi non convexes de forme $\ell_2 - \ell_0$ (comportement quadratique en 0, et constant à l'infini), approchant le potentiel ℓ_0 [148, 50]. L'intérêt applicatif de ces pénalisations dans le contexte de la reconstruction d'images est illustré dans l'Exemple 5.3.1. La Propriété 5.3.1 est au cœur des algorithmes de "moindres carrés pondérés" [98, 104, 198, 37, 160] et "semi-quadratiques Geman-Reynolds" [6, 43, 113, 150] utilisés en traitement du signal/image pour la minimisation de critères de moindres carrés pénalisés. Nous reviendrons sur ces derniers dans le Chapitre 8.

	$f(t)$	Type	Nom
Convexe	$ t - \delta \log(t /\delta + 1)$	$\ell_2 - \ell_1$	
	$\begin{cases} t^2 & \text{si } t < \delta \\ 2\delta t - \delta^2 & \text{sinon} \end{cases}$	$\ell_2 - \ell_1$	Huber
	$\log(\cosh(t))$	$\ell_2 - \ell_1$	Green
	$\sqrt{1 + t^2/\delta^2} - 1$	$\ell_2 - \ell_1$	Hyperbolique
	$(1 + t^2/\delta^2)^{\kappa/2} - 1$	$\ell_2 - \ell_\kappa$	
Non convexe	$1 - \exp(-t^2/(2\delta^2))$	$\ell_2 - \ell_0$	Welsch
	$t^2/(2\delta^2 + t^2)$	$\ell_2 - \ell_0$	Geman -McClure
	$\begin{cases} 1 - (1 - t^2/(6\delta^2))^3 & \text{si } t \leq \sqrt{6}\delta \\ 1 & \text{sinon} \end{cases}$	$\ell_2 - \ell_0$	Tukey biweight
	$\tanh(t^2/(2\delta^2))$	$\ell_2 - \ell_0$	Tangente hyperbolique
	$\log(1 + t^2/\delta^2)$	$\ell_2 - \log$	Cauchy
	$1 - \exp(1 - (1 + t^2/(2\delta^2))^{\kappa/2})$	$\ell_2 - \ell_\kappa - \ell_0$	Chouzenoux

Table 5.2 – Exemples de fonctions f vérifiant la Propriété 5.3.1. $\kappa \in [1, 2]$

Exemple 5.3.1 Je fournis ici une illustration applicative de la Propriété 5.3.1 dans le cadre d'un travail mené en collaboration avec Saint Gobain Recherche, sur la

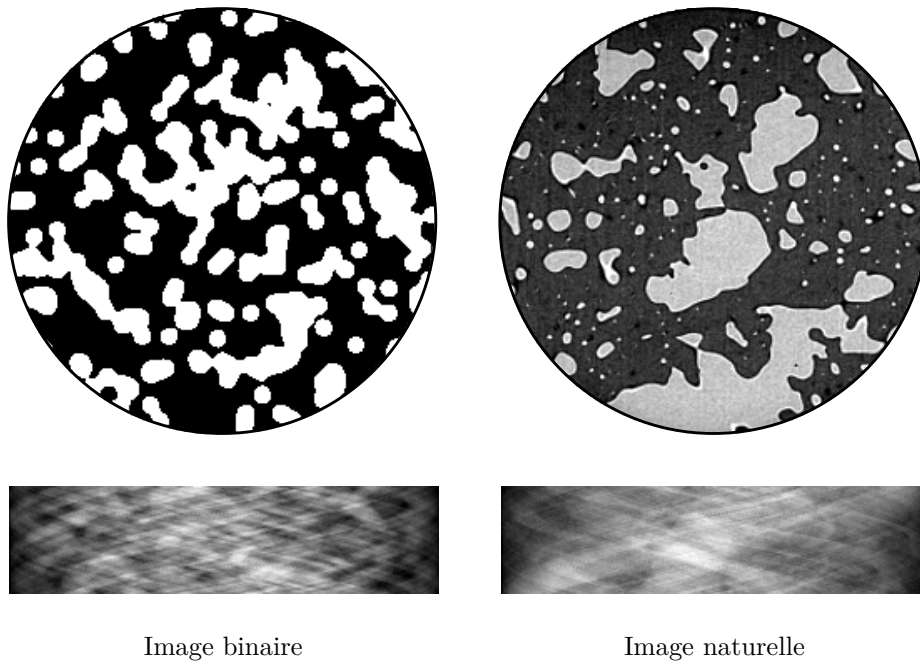


Figure 5.2 – Haut : Images originales. L'image naturelle est une image de séparation de phase de verre barium borosilicate acquise à l'ESRF synchrotron (courtoisie de David Bouttes). Bas : Sinogrammes bruités ($\sigma = 2$).

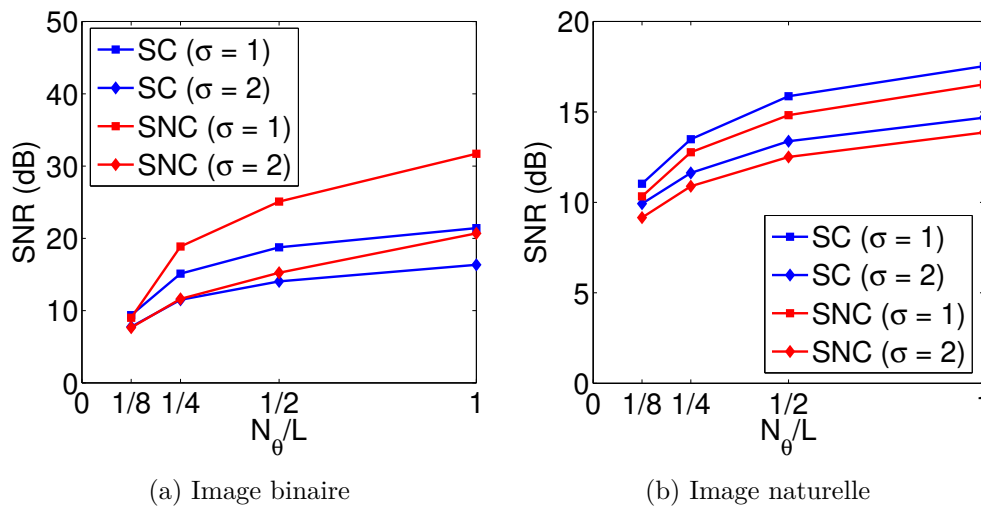


Figure 5.3 – Qualité de reconstruction, en terme de SNR, pour les pénalisations SC et SNC.

tomographie à rayons X du verre avec intégration angulaire continue publié dans

[88]. Nous avons considéré le problème de reconstruction d’une image originale \bar{x} à partir de projections discrètes bruitées $y = H\bar{x} + \omega \in \mathbb{R}^M$, où $H \in \mathbb{R}^{M \times N}$ est une matrice de projection de Radon, et $\omega \in \mathbb{R}^M$ un vecteur modélisant un bruit d’acquisition blanc gaussien à moyenne nulle. Ce type de problème inverse apparaît dans la modalité d’imagerie de tomographie à rayons X [119, 90]. Une approche efficace pour restaurer l’image est de minimiser un critère de moindres carrés pénalisé par un terme de régularisation introduisant des propriétés de régularité sur l’image, de la forme (5.6), où $V \in \mathbb{R}^{2N \times N}$ est la concaténation des opérateurs de gradient discret horizontal et vertical. Deux fonctions de pénalités du Tableau 5.2, à savoir la pénalité hyperbolique (*Smooth Convex*, SC) et la pénalité de Welsch (*Smooth Non Convex*, SNC), ont été comparées, pour restaurer deux images synthétiques mimant un échantillon biphasique de verre. L’approche d’optimisation utilisée était l’algorithme 3MG [50] décrit dans le Chapitre 6, associé à l’approximation majorante quadratique issue de la Propriété 5.3.1. Les résultats, présentés en Figures 5.2 et 5.3 illustrent l’influence du choix de la pénalisation sur la qualité des résultats, pour différentes valeurs de l’écart type du bruit σ , et du rapport du nombre d’angles d’acquisition N_θ sur le nombre de détecteurs du tomographe L . L’image binaire restaurée présente un meilleur rapport signal sur bruit (*Signal to Noise Ratio*, SNR) lorsque la pénalisation non convexe est employée alors que l’image naturelle est mieux restaurée par la pénalisation convexe. La flexibilité de la méthode de MM quadratique nous a permis d’effectuer simplement ces comparaisons, sans changer la structure du schéma d’optimisation.

5.4 APPROXIMATION MAJORANTE SÉPARABLE

Dans certains algorithmes basés sur les approximations majorantes quadratiques, il est nécessaire de choisir des matrices $A(x)$, $x \in \mathbb{R}^N$, avec une structure particulière, afin de limiter la complexité de calcul. C’est en particulier fondamental dans les approches proximales à métrique variable (Sec. 6.2, 7.2) qui utilisent la courbure de l’approximation majorante quadratique du critère comme métrique (ou préconditionneur) dans leur règle de mise à jour. Dans ce cas, le choix de fonctions majorantes séparables, basées sur des préconditionneurs diagonaux, permet bien souvent de préserver la forme explicite du schéma d’optimisation. Nous verrons par ailleurs dans la Section 7.4, que l’emploi de fonctions majorantes séparables permet de construire des algorithmes alternés par blocs propices à une implémentation efficace sur une architecture parallèle multi-processeur [174]. Je présente ici une propriété issue de [111] permettant de construire des fonctions majorantes quadratiques avec une structure séparable, à partir d’une fonction majorante quadratique sans structure particulière.

Propriété 5.4.1 *Soit $M \geq 1$ et $H \in \mathbb{R}^{M \times N}$. Alors :*

(i) $H^\top H \preceq \text{Diag}(b)$, avec $b \in \mathbb{R}^N$ défini par :

$$(\forall n \in \{1, \dots, N\}) \quad b_n = [|H|^\top |H| \mathbf{1}_N]_n \quad (5.8)$$

où $\mathbf{1}_N \in \mathbb{R}^N$ est un vecteur dont les entrées sont toutes égales à un, et $|H|$ est une matrice de même taille que H , dont les entrées sont égales à la valeur absolue de celles de H .

(ii) Supposons que $H \in [0, +\infty[^{M \times N}$. Alors, pour tout $x \in]0, +\infty[^N$, $H^\top H \preceq \text{Diag}(b(x))$ avec $b(x) \in \mathbb{R}^N$ défini par :

$$(\forall n \in \{1, \dots, N\}) \quad b_n(x) = [H^\top H x]_n / x_n. \quad (5.9)$$

La Propriété 5.4.1(i) est utilisée par exemple dans l'algorithme SART (*Simultaneous Algebraic Reconstruction Technique*) proposé par [118] pour la reconstruction d'images en tomographie, ainsi que dans le problème de restauration d'image que je présente dans la Section 6.2.4. Quant à la Propriété 5.4.1(ii), elle constitue la base de l'algorithme ISRA (*Iterative Image Space Reconstruction Algorithm*) proposé dans [7] pour la restauration d'images astronomiques, et de l'algorithme de factorisation non négative de [130]. Je présente dans l'Exemple 5.4.1 une version étendue de ce dernier algorithme pour la résolution d'un problème de démixage d'images multispectrales, issu de notre article [164]. Notons que ce travail a par la suite été appliqué dans [196] à la résolution d'un problème de décomposition canonique de tenseurs.

Exemple 5.4.1 Nous considérons dans [164] le problème de démixage d'un cube hyperspectral, noté $Y \in \mathbb{R}^{S \times M}$, modélisant un ensemble d'images de \mathbb{R}^M acquises dans S bandes spectrales. On suppose que le modèle d'observation suit l'équation de mélange linéaire $Y = UV + E$ où les colonnes de $U \in \mathbb{R}^{S \times P}$ représentent les spectres de P composantes distinctes présentes dans le cube, $V \in \mathbb{R}^{P \times M}$ leurs proportions respectives (abondances) en chaque pixel, et $E \in \mathbb{R}^{S \times M}$ le bruit d'acquisition. Ce type de modèle se rencontre dans diverses applications telles que la télédétection, la planétologie ou la spectromicroscopie [8, 186, 85]. Une illustration schématique de ce modèle est donnée dans la Figure 5.4. Le démixage spectral consiste à estimer conjointement U et V à partir des données observées Y [120]. Dans ce travail, nous nous sommes concentrés sur le cas où chaque spectre composant les colonnes de U est une combinaison linéaire de quelques spectres de référence issus d'un grand dictionnaire de taille $Q > P$, noté $\Omega \in \mathbb{R}^{S \times Q}$. Ainsi, le modèle d'observation devient $Y = \Omega TV + E$ où $T \in \mathbb{R}^{Q \times P}$ est une matrice supposée parcimonieuse donnant les coefficients de la combinaison linéaire permettant d'obtenir $U = \Omega T$. Nous avons proposé de définir des estimées \hat{T} et \hat{V} de T et V comme étant des minimiseurs d'un problème de moindres carrés pénalisé :

$$(\forall T \in \mathbb{R}^{Q \times P})(\forall V \in \mathbb{R}^{P \times M}) \quad F(T, V) = \frac{1}{2} \|Y - \Omega TV\|_F^2 + \Psi_1(T) + \Psi_2(V), \quad (5.10)$$

où les fonctions $\Psi_1 : \mathbb{R}^{Q \times P} \mapsto]-\infty, +\infty]$ et $\Psi_2 : \mathbb{R}^{P \times M} \mapsto]-\infty, +\infty]$ sont des fonctions de régularisation incorporant des informations a priori sur la positivité des entrées de T et V , la parcimonie de T , et la sommation à un des colonnes de V . La stratégie de minimisation alternée BC-VMFB (Sec. 7.2) est employée. Chaque étape de mise à jour est basée sur une stratégie explicite implicite préconditionnée par majoration quadratique, appliquée successivement sur les variables T puis V . Afin de limiter la complexité de l'algorithme, nous avons proposé d'utiliser des fonctions majorantes séparables obtenues par la Propriété 5.4.1(ii), similaires à celles utilisées dans les approches de factorisation de matrices non-négative (*Nonnegative Matrix Factorization*, NMF).

Présentons maintenant les résultats obtenus, pour une matrice U composée de $P = 5$ spectres distincts résultant d'une combinaison pondérée de 3 spectres purs sélectionnés parmi les colonnes d'un dictionnaire de $Q = 62$ spectres de $S = 224$ bandes spectrales, issu de la bibliothèque U.S. Geological Survey library [69]. Chaque ligne de la matrice d'abondance V est construite comme étant la superposition de gaussiennes 2D composées de $M = 128 \times 128$ pixels, d'espérances et de variances aléatoires, normalisées afin d'assurer la contrainte de somme à un. Finalement, on définit la matrice E comme une réalisation d'un vecteur aléatoire gaussien centré et dont la variance est choisie de manière à ce que le SNR initial soit égal à 20 dB. Sur la Figure 5.5(a), je représente les spectres exacts et ceux reconstruits par notre algorithme. Enfin, la Figure 5.5(b) présente l'évolution de l'erreur de reconstruction sur la variable T , pour notre méthode, et pour une méthode concurrente, basée sur une majoration pls simple utilisant la constante de Lipschitz du critère par rapport à chacune des variables. Nous pouvons constater l'amélioration conséquente, en terme de vitesse de convergence, apportée par l'approche de majoration séparable.

CONCLUSION

J'ai présenté dans ce chapitre trois familles d'approches utilisées dans mes travaux pour construire des approximations majorantes quadratiques menant à des algorithmes de minimisation efficaces. Ces approches sont les plus employées de par leur grande flexibilité, mais d'autres stratégies sont bien sûr possibles, pour des structures particulières de la fonction de coût (voire les exemples traités dans [67, 66]). Le choix d'une approche plutôt qu'une autre est étroitement lié à l'algorithme utilisé pour la minimisation. Par exemple, les algorithmes de type MM sous espace (Sec. 6.3) sont peu contraignants sur la forme retenue pour les métriques $A(\cdot)$, tandis que les méthodes proximales à métrique variable (Sec. 6.2, Sec. 7.2) vont avoir tendance à être plus efficaces pour des métriques diagonales. Dans chaque exemple d'application, il est toujours intéressant de comparer plusieurs choix possibles de fonctions majorantes, afin de sélectionner celui qui assure le meilleur compromis entre qualité d'approximation de la fonction de coût, et simplicité de mise en œuvre.

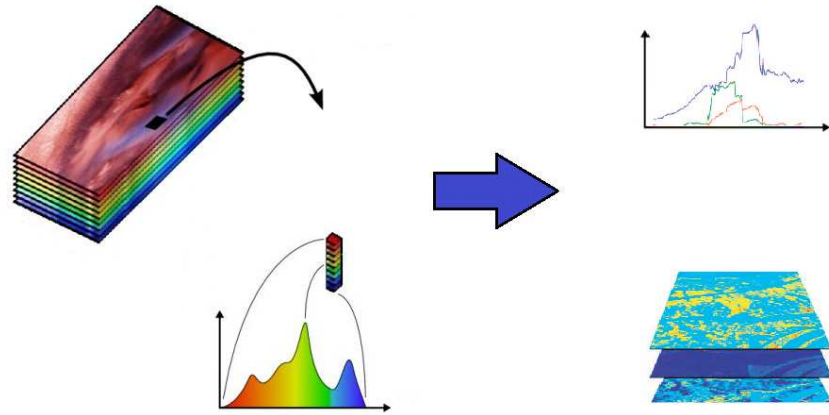


Figure 5.4 – Illustration du problème de démixage spectral. L'objectif est d'estimer U et V à partir d'un cube de données hyperspectrales Y .

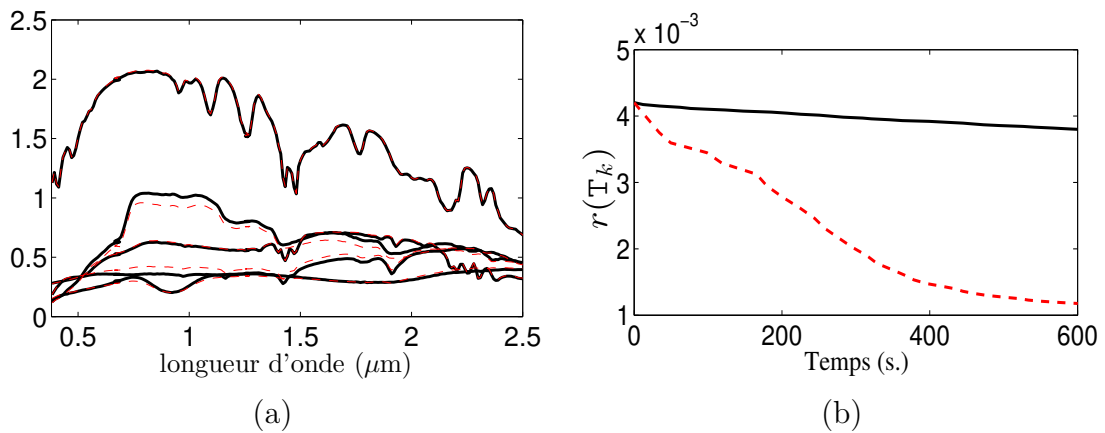


Figure 5.5 – (a) Spectres exacts (lignes continues noires) et reconstruits (lignes discontinues rouges); (b) Evolution de l'erreur relative de reconstruction sur T , pour l'algorithme de [164] basé sur une majorante séparable (ligne discontinue rouge) et l'algorithme PALM de [29] (ligne continue noire).

- Chapitre 6 -

Algorithmes de majoration-minimisation pour l'optimisation non convexe

6.1 INTRODUCTION

Les approches par MM connaissent une popularité croissante dans les domaines du signal/image et de l'apprentissage [183, 110, 202]. Le cadre MM fournit en effet une méthodologie simple, élégante et flexible pour construire des algorithmes efficaces en pratique. L'analyse de convergence de ces méthodes se fonde sur une propriété de convergence (monotone) du critère, inhérente au schéma MM. Autant, dans le cas fortement convexe, il est facile de déduire de la convergence de la fonctionnelle des garanties de convergence sur les itérées de l'algorithme, autant dans un cadre plus général, non nécessairement convexe, les résultats existants se limitent bien souvent à des résultats partiels sur les sous-suites engendrées par l'algorithme [161, 114]. Dans mes travaux de recherche, je me suis posée la question de la convergence des itérées des schémas MM pour une fonction de coût non convexe. Il s'agit d'une question fondamentale dans le cadre des problèmes inverses où la valeur limite des itérées minimisant la fonction de coût pénalisée conduit à une solution du problème inverse, tandis que la valeur finale de cette fonction de coût a peu d'importance.

En optimisation non convexe, un outil important qui permet de démontrer la convergence de suites minimisantes est *l'inégalité de Kurdyka-Łojasiewicz* (KL). Proposée initialement dans [135] dans le cadre des fonctions analytiques réelles, elle a été généralisée dans [126] pour les fonctions différentiables définissables dans une structure o-minimale (pour les structures o-minimales, voir par exemple [77, 194]). L'inégalité de KL peut être généralisée pour des fonctions non nécessairement différentiables en utilisant la notion de sous-différentiel [26, 27, 28]. Cette inégalité, s'écrit de la façon suivante :

Définition 6.1.1 La fonction $F : \mathbb{R}^N \mapsto]-\infty, +\infty]$ satisfait l'inégalité de Kurdyka-Łojasiewicz en $\hat{x} \in \text{dom } F$ s'il existe $\xi \in]0, +\infty[$, un voisinage $\mathcal{V}(\hat{x})$ de \hat{x} et une fonction $\varphi : [0, \xi[\rightarrow [0, +\infty[$ concave tels que

- (i) $\varphi(0) = 0$,
- (ii) φ est continue en 0 et continument différentiable sur $]0, \xi[$,
- (iii) pour tout $u \in]0, \xi[$, $\dot{\varphi}(u) > 0$,
- (iv) $\dot{\varphi}(F(x) - F(\hat{x}))\text{dist}(0, \partial F(x)) \geq 1$, pour tout $x \in E$ tel que $F(\hat{x}) \leq F(x) \leq F(\hat{x} + \xi)$, où $\text{dist}(0, \partial F(x)) = \inf_{u \in \partial F(x)} \|u\|$.

La convergence d'algorithmes proximaux pour des fonctions non nécessairement convexes vérifiant l'inégalité de KL ci-dessus a récemment été démontrée récemment dans [10, 11]. Cette inégalité est satisfaite par les fonctions définissables dans une structure o-minimale, et plus particulièrement par les fonctions semi-algébriques réelles et analytiques réelles. Pour ces dernières classes de fonctions, la fonction φ de l'inégalité de KL est de la forme $\varphi : u \in [0, \xi[\rightarrow u^{1-\theta}$, pour $\theta \in [0, 1[$. Cela donne lieu à une forme simplifiée de la définition [10], que nous considérerons dans toute la suite de ce manuscrit. Un intérêt majeur de cette forme particulière est qu'elle permet d'obtenir des résultats de vitesse de convergence [11, 96]. Dans ce chapitre, je présente les résultats de convergence issus des articles [66, 50], visant à établir la convergence des itérées générées par les algorithmes explicite-implicite à métrique variable (Sec. 6.2) et MM sous espace (Sec. 6.3) dans le cadre de fonctions non nécessairement convexes vérifiant l'inégalité de KL. Enfin, la dernière section présente l'extension de la méthode décrite en Section 6.3 dans le cadre de l'optimisation stochastique que j'ai récemment publié dans [63]. En particulier, je présente les résultats de convergence de cette approche établis dans le cas de fonctions de pénalisation non nécessairement convexes.

6.2 ALGORITHME EXPLICITE-IMPLICITE À MÉTRIQUE VARIABLE

6.2.1 Introduction

Considérons un problème de minimisation faisant intervenir une somme de deux fonctions dont l'une est différentiable de gradient Lipschitz, tandis que l'autre est convexe mais éventuellement non lisse. Dans ce cadre, une approche usuelle consiste à utiliser l'algorithme proximal explicite-implicite (*Forward-Backward*, FB) qui alterne, à chaque itération, une étape de gradient sur la fonction différentiable et une étape proximale sur la seconde fonction [44, 191]. Dans le cas convexe, il a été démontré dans [75] que la suite des itérées générée par un tel algorithme converge vers une solution du problème. Ce résultat a été étendu récemment au cas où les fonctions à minimiser ne sont pas nécessairement convexes mais satisfont l'inégalité

de KL [10, 12]. L'algorithme explicite-implicite est caractérisé par un faible coût de calcul à chaque itération. Cependant, comme la plupart des méthodes de minimisation de premier ordre, il peut avoir une convergence lente [44]. Une stratégie efficace d'accélération est basée sur l'introduction d'une métrique variant au cours des itérations [18, 34, 44, 36, 74, 136]. Cette métrique est obtenue grâce à une matrice de préconditionnement. On obtient alors l'algorithme explicite-implicite à métrique variable (*Variable Metric Forward-Backward*, VMFB). La convergence de cet algorithme a été démontrée dans [74, 131, 190], sous des hypothèses de convexité et des conditions spécifiques sur les métriques considérées. Dans le cadre de la thèse de AUDREY REPETI, nous nous sommes posés la question de la convergence de cet algorithme dans le cas où le terme lisse est non convexe. Je présente ici un résumé de nos résultats qui reposent d'une part sur l'inégalité de KL, et d'autre part sur un choix original des matrices de préconditionnement provenant d'une interprétation MM de la méthode VMFB. L'intérêt applicatif de la méthode est illustrée sur un exemple de restauration d'images en présence de bruit Gaussien d'écart type spatialement variable dépendant de l'intensité des pixels. Ces travaux ont été publiés dans [66, 162].

6.2.2 Algorithme VMFB

Intéressons-nous au problème de minimisation (4.1) où $F : \mathbb{R}^N \mapsto]-\infty, +\infty]$ est une fonction coercive, s'écrivant $F = f_1 + f_2$, avec $f_1 : \mathbb{R}^N \mapsto \mathbb{R}$ β -Lipschitz différentiable et f_2 convexe, propre, semi-continue inférieurement et dont la restriction à son domaine est continue. Supposons, de plus, que, pour tout x' dans le domaine de f_2 , la fonction lisse f_1 puisse être majorée par une fonction quadratique notée $q_1(\cdot, x')$ de courbure $A(x') \in \mathbb{R}^{N \times N}$ symétrique définie positive. Une approche MM pour la minimisation de F s'écrit alors :

$$\boxed{\begin{array}{l} \text{Pour } k = 0, 1, \dots \\ \left[\begin{array}{l} x_{k+1} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \quad q_1(x, x_k) + f_2(x). \end{array} \right. \end{array}} \quad (6.1)$$

Rappelons par ailleurs la définition suivante [109, Sec. XV.4] :

Définition 6.2.1 Soient $f : \mathbb{R}^N \mapsto]\infty, +\infty]$ une fonction propre et semi-continue inférieurement, $A \in \mathbb{R}^{N \times N}$ symétrique définie positive, et $x \in \mathbb{R}^N$. L'opérateur proximal de f en x relativement à la métrique induite par A est l'ensemble défini par

$$\text{prox}_{A,f}(x) = \underset{y \in \mathbb{R}^N}{\text{Argmin}} \quad f(y) + \frac{1}{2} \|y - x\|_A^2 \quad (6.2)$$

Par abus de notation, lorsque cet ensemble est égal à un singleton, on note aussi $\text{prox}_{A,f}(x)$ son unique élément.

La Définition 6.2.1 nous permet alors de montrer qu'une version relaxée de l'Algorithme (6.1) s'écrit comme un algorithme de type proximal suivant un schéma VMFB :

$$\begin{array}{l}
 x_0 \in \text{dom} f_2, (\lambda_k, \theta_k)_{k \in \mathbb{N}} \in]0, +\infty[^2. \\
 \text{Pour } k = 0, 1, \dots \\
 \left[\begin{array}{l}
 \tilde{x}_k \in \text{prox}_{\theta_k^{-1} A(x_k), f_2}(x_k - \theta_k A(x_k)^{-1} \nabla f_1(x_k)), \\
 x_{k+1} = x_k + \lambda_k (\tilde{x}_k - x_k).
 \end{array} \right.
 \end{array} \tag{6.3}$$

Lorsque $\theta_k \equiv 1$ et $\lambda_k \equiv 1$, les Algorithmes (6.1) et (6.3) sont identiques. Par ailleurs, comme souligné dans [93, 17], lorsque $A(x_k) \equiv I_N$, l'Algorithme (6.3) est équivalent à l'algorithme proximal explicite-implicite (i.e., FB) usuel de [44, 191]. D'autre part, lorsque $f_2 \equiv 0$, l'Algorithme (6.3) correspond à un algorithme de gradient préconditionné par les matrices $(A(x_k))_{k \in \mathbb{N}}$.

6.2.3 Résultats de convergence

La convergence de l'Algorithme (6.3) a été étudiée dans divers travaux, sous des hypothèses variées sur les fonctions mises en jeu, et le choix des métriques $(A(x_k))_{k \in \mathbb{N}}$. Le résultat le plus général est donné dans [74] : les auteurs ont démontré la convergence de la suite des itérés $(x_k)_{k \in \mathbb{N}}$ vers un minimiseur de $f_1 + f_2$ lorsque f_1 et f_2 sont convexes et qu'il existe une suite $(\eta_k)_{k \in \mathbb{N}}$ positive bornée telle que, pour tout $k \in \mathbb{N}$,

$$(1 + \eta_k)A(x_{k+1}) \succeq A(x_k). \tag{6.4}$$

D'autres résultats de convergence existent dans la littérature dans le cas particulier où f_2 est une fonction indicatrice sur un ensemble convexe [22, 25, 33] ou lorsque f_1 est localement fortement convexe avec un hessien Lipschitz continu et f_2 est convexe [131]. Plus récemment la convergence de l'Algorithme (6.3) a aussi été établie dans [190] lorsque f_2 est convexe, f_1 est une fonction lisse auto-concordante, et, pour tout $k \in \mathbb{N}$, $A(x_k)$ est soit le hessien de f_1 en x_k (ou une version approchée de celui-ci), soit une matrice diagonale. Cependant, dans aucun de ces travaux, l'étude de la convergence de l'Algorithme (6.3) n'avait été menée dans le cas d'une fonction f_1 non-convexe. Je présente ici le résultat de convergence que nous avons établi dans [66], dont la preuve est intrinsèquement liée à l'interprétation MM de l'algorithme VMFB décrite plus haut.

Théorème 6.2.1 *Supposons que :*

- Il existe $(\underline{\nu}, \bar{\nu}) \in]0, +\infty[^2$ tel que, pour tout $k \in \mathbb{N}$, $\underline{\nu}I_N \preceq A(x_k) \preceq \bar{\nu}I_N$.
- Il existe $\underline{\lambda} \in]0, +\infty[$ tel que, pour tout $k \in \mathbb{N}$, $\underline{\lambda} \leq \lambda_k \leq 1$,
- Il existe $(\eta, \bar{\eta}) \in]0, +\infty[^2$ tel que pour tout $k \in \mathbb{N}$, $\eta \leq \lambda_k \theta_k \leq 2 - \bar{\eta}$.
- Il existe $\alpha \in]0, 1]$ tel que, pour tout $k \in \mathbb{N}$, $F(x_{k+1}) \leq (1 - \alpha)F(x_k) + \alpha F(\tilde{x}_k)$.

Supposons de plus que la fonction F satisfasse l'inégalité de Kurdyka-Łojasiewicz. Les assertions suivantes sont alors satisfaites :

- (i) Les suites $(F(x_k))_{k \in \mathbb{N}}$ et $(F(\tilde{x}_k))_{k \in \mathbb{N}}$ convergent vers $F(\hat{x})$, où \hat{x} est un point critique de F . De plus, $(F(x_k))_{k \in \mathbb{N}}$ est décroissante.
- (ii) Les suites $(x_k)_{k \in \mathbb{N}}$ et $(\tilde{x}_k)_{k \in \mathbb{N}}$ générées par l'Algorithme (6.3) convergent toutes les deux vers un point critique \hat{x} de F .

Notons que nous avons présenté dans [66] une version étendue du théorème ci-dessus, pour une version inexacte de l'algorithme (6.3), s'inspirant de l'algorithme inexact explicite-implicite de [12], dans laquelle le calcul de l'opérateur proximal relatif à la métrique majorante peut être effectué de façon approchée. Notre résultat de convergence de la version inexacte de l'Algorithme (6.3) a par la suite été étendu dans [32, 42] sous des hypothèses relâchées sur les métriques utilisées.

6.2.4 Application à la restauration d'images dégradées par un bruit Gaussien non homogène

Illustrons les performances de l'algorithme VMFB sur le problème de reconstruction d'image en présence de bruit gaussien non homogène issu de notre article [66]. Ce modèle de bruit se rencontre dans de nombreux systèmes d'imagerie [108, 116, 187] dans lesquels les acquisitions sont dégradées par un bruit gaussien de moyenne nulle et d'écart type variant spatialement, proportionnellement à l'intensité des pixels, lié au comptage de photons, et un bruit électronique indépendant. Le bruit gaussien non homogène peut être également vu comme une approximation du bruit Poisson-Gaussien évoqué dans l'Exemple 5.2.1. Les observations sont de la forme :

$$(\forall m \in \{1, \dots, M\}) \quad y_m = [H\bar{x}]_m + \sqrt{a_m[H\bar{x}]_m + b_m} w_m, \quad (6.5)$$

où $H \in [0, +\infty[^{M \times N}$ est une matrice de projection, $[Hx]_m$ est la m -ème composante de Hx , $(a_m)_{1 \leq m \leq M} \in [0, +\infty[^M$, $(b_m)_{1 \leq m \leq M} \in [0, +\infty[^M$ et $(w_m)_{1 \leq m \leq M}$ est une réalisation d'un vecteur aléatoire gaussien de moyenne nulle et de matrice de covariance I_M . Le terme d'attache aux données est défini, à une constante additive près, pour tout $x \in [0, +\infty[^N$, comme étant la neg-log-vraisemblance des données :

$$(\forall x \in [0, +\infty[^N) \quad f_1(x) = \frac{1}{2} \sum_{m=1}^M \left(\frac{([Hx]_m - y_m)^2}{a_m[Hx]_m + b_m} + \log(a_m[Hx]_m + b_m) \right). \quad (6.6)$$

Les matrices $(A(x_k))_{k \in \mathbb{N}}$ associées sont construites suivant une approche similaire à [90]. D'autre part, une régularisation hybride est considérée : $f_2 = \iota_{[x_{\min}, x_{\max}]^N} + \sum_{p=1}^P \vartheta_p |[V \cdot]_p|$, où $\iota_{[x_{\min}, x_{\max}]^N}$ est la fonction indicatrice du pavé $[x_{\min}, x_{\max}]^N$, $V \in \mathbb{R}^{P \times N}$ et $(\vartheta_m)_{1 \leq p \leq P} \in [0, +\infty[^P$. Je présente ici les résultats obtenus lorsque l'image originale \bar{x} correspond à une coupe du fantôme Zubal de dimension $N = 128 \times 128$, où $x_{\min} = 0$ et $x_{\max} = 1$, et H est une matrice de Radon modélisant $M = 16384$ projections parallèles. Les Figures 6.1(a)-(b) présentent les projections bruitées et l'image restaurée par la méthode proposée. De plus, les figures 6.1(c)-(d) illustrent

les variations de $(F(x_k) - \hat{F})_k$ et de $(\|x_k - \hat{x}\|)_k$ en fonction du temps de calcul pour les algorithmes VMFB, FB [12] et FISTA [16]. Pour VMFB, le paramètre de relaxation est fixé à $\lambda_k \equiv 1$. Les résultats de FB et de VMFB sont présentés pour deux valeurs du pas : $\theta_k \equiv 1$ et $\theta_k \equiv 1.9$. Nous observons que le second choix conduit à de meilleurs résultats en terme de vitesse de convergence. De plus, nous remarquons que l'introduction de la métrique variable permet d'accélérer la vitesse de convergence du critère ainsi que celle des itérées.

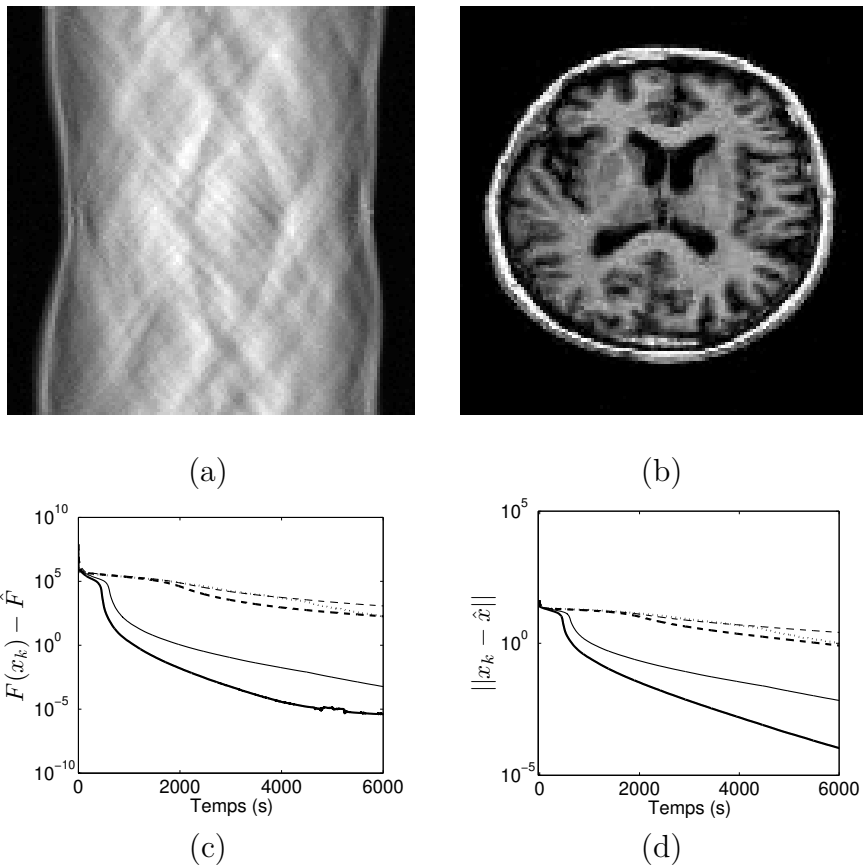


Figure 6.1 – (a) Projections bruitées ; (b) image restaurée, SNR=18.9 dB ; (c)-(d) Comparaisons entre l'algorithme VMFB avec $\theta_k \equiv 1$ (traits fins continus) et $\theta_k \equiv 1.9$ (traits épais continus), l'algorithme FB avec $\theta_k \equiv 1$ (traits fins discontinus) et $\theta_k \equiv 1.9$ (traits épais discontinus) et FISTA (traits pointillés) en terme de convergence du critère (c) et des itérées (d).

6.3 ALGORITHME MM SOUS-ESPACE

6.3.1 Introduction

Considérons le problème de la minimisation d'une fonction F différentiable, possédant des approximations majorantes quadratiques de la forme (5.1). L'application directe du schéma MM conduit à un algorithme simple, consistant à minimiser, à chaque itération $k \in \mathbb{N}$, la fonction quadratique $Q(\cdot, x_k)$, majorante de F en x_k . Les algorithmes semi-quadratiques [99, 100, 150, 113] en sont un cas particulier, nous y reviendrons dans le Chapitre 8. Néanmoins, dans le contexte de la grande dimension, la minimisation de la majorante quadratique à chaque itération peut s'avérer très coûteux, puisqu'il s'agit d'inverser un système linéaire de taille $N \times N$. Une possibilité est d'utiliser un solveur itératif, tel que l'algorithme de gradient conjugué [127]. Une solution alternative, que j'avais proposée dans le cadre de ma thèse, est d'adopter une stratégie de *sous-espace*. L'idée clé de cette approche est de restreindre l'espace de recherche du nouvel itéré à un sous-espace vectoriel engendré par un petit nombre de vecteur directionnels bien choisis, plutôt que de minimiser la fonction majorante sur tout \mathbb{R}^N . La complexité par itération est alors diminuée, et les exemples traités dans [49] illustraient les très bonnes performances de la méthode en terme de vitesse de convergence. Un résultat préliminaire de convergence avait été établi dans [49]. Cependant, la convergence des itérées n'était pas démontrée et, qui plus est, l'étude se limitait au cadre convexe. Je présente dans la première partie de cette section une extension de ces résultats, établissant la convergence des itérées de l'algorithme MM sous-espace pour une fonction F non convexe vérifiant l'inégalité de KL, ainsi qu'une illustration de l'intérêt de la méthode sur un exemple de restauration d'image. Dans une deuxième partie, j'introduis la version stochastique de l'algorithme MM sous-espace récemment publiée dans [63]. Je présente les propriétés de convergence obtenus pour cette méthode et illustre ses performances pratiques sur un problème d'identification de flou.

6.3.2 Algorithme 3MG dans le cas déterministe

Soit une fonction $F : \mathbb{R}^N \mapsto \mathbb{R}$ différentiable sur \mathbb{R} dont on souhaite déterminer itérativement un minimiseur. Supposons que pour tout $x' \in \mathbb{R}^N$, il existe $A(x') \in \mathbb{R}^{N \times N}$ symétrique définie positive, telle que la fonction quadratique $Q(\cdot, x')$ de courbure $A(x')$ est une approximation majorante de F en x' . L'algorithme MM par sous-espace de [49] s'écrit :

$$\begin{array}{l}
 x_0 \in \mathbb{R}^N \\
 \text{Pour } k = 0, 1, \dots \\
 \left[\begin{array}{l}
 \text{Choisir } D_k \in \mathbb{R}^{N \times M_k}, \\
 x_{k+1} \in \underset{x \in \text{ran} D_k}{\text{Argmin}} Q(x, x_k).
 \end{array} \right.
 \end{array} \tag{6.7}$$

Pour tout $k \in \mathbb{N}$, D_k est une matrice de $\mathbb{R}^{N \times M_k}$, $M_k \geq 1$, dont chaque colonne peut être vue comme une direction dans laquelle on va rechercher la future itérée, et $\text{ran}D_k$ est l'espace vectoriel engendré par ces colonnes. Notons que, même si $Q(\cdot, x_k)$ a un minimiseur unique sur \mathbb{R}^N , elle n'est pas forcément minimisée de façon unique sur $\text{ran}D_k$. En pratique, on choisit le minimiseur issu de l'inverse généralisé, c'est à dire :

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = x_k + D_k s_k, \quad s_k = - \left(D_k^\top A(x_k) D_k \right)^\dagger \left(D_k^\top \nabla F(x_k) \right). \quad (6.8)$$

Lorsque $\text{ran}D_k = \mathbb{R}^N$, l'Algorithme 6.7 est équivalent à un algorithme MM quadratique "classique". Cependant, on voit par l'Equation (6.8) que chaque itération $k \in \mathbb{N}$, nécessite dans ce cas d'inverser la matrice $A(x_k)$, de dimension $N \times N$, ce qui peut être très prohibitif dans le contexte de la grande dimension. Une meilleure solution est de prendre M_k assez petit, et des directions de recherche bien choisies, afin de réduire la complexité par itération tout en préservant une vitesse de convergence raisonnable (voir [49, Tab.I] pour une liste de choix possibles pour la matrice D_k).

La convergence de la suite $(\nabla F(x_k))_{k \in \mathbb{N}}$ vers 0 avait été montrée dans [49], pour une séquence de $\{A(x_k)\}_{k \in \mathbb{N}}$ à spectre borné, et sous l'hypothèse faible que $\text{ran}D_k$ contienne une direction "gradient reliée" [23] (ex : gradient, gradient préconditionné). Nous avons effectué de nombreux tests numériques dans [49], comparant, sur des exemples de restauration d'images, l'Algorithme 6.7 avec des méthodes d'optimisation différentiables usuelles tel que le gradient conjugué non linéaire [106], l'algorithme de Newton tronqué et l'algorithme L-BFGS [134]. Ces tests illustraient le fait que le choix du sous-espace avait une influence importante sur la vitesse de l'algorithme, le meilleur compromis étant réalisé pour le sous-espace de mémoire de gradient [145], engendré par la direction de plus grande pente, et la différence des itérés précédentes, donnant lieu à l'algorithme 3MG (*Majorize-Minimize Memory Gradient*). Ces travaux prometteurs nous ont amenés à effectuer une analyse plus poussée de la convergence de la méthode, dans un cadre non nécessairement convexe, et également une analyse théorique de sa vitesse. Les résultats obtenus ont été publiés dans [50, 62], et j'en présente ici un résumé.

Résultats de convergence :

Théorème 6.3.1 *Soit $(x_k)_{k \in \mathbb{N}}$ la suite générée par l'Algorithme 6.7. Supposons que :*

- $F : \mathbb{R} \rightarrow \mathbb{R}$ est coercive et différentiable,
- Il existe $(\underline{\nu}, \bar{\nu}) \in]0, +\infty[^2$ tel que $(\forall k \in \mathbb{N}) \quad \underline{\nu} I_N \preceq A(x_k) \preceq \bar{\nu} I_N$,
- Pour tout $k \in \mathbb{N}$, il existe $\delta_k \in \text{ran} D_k$, tel que $\delta_k \in \mathbb{R}^N$ est gradient-reliée i.e., il existe $(\gamma_0, \gamma_1) \in]0, +\infty[^2$ tel que pour tout $k \in \mathbb{N}$, $\langle \nabla F(x_k) | \delta_k \rangle \leq -\gamma_0 \|\nabla F(x_k)\|^2$ et $\|\delta_k\| \leq \gamma_1 \|\nabla F(x_k)\|$.

Supposons en outre que F satisfait l'inégalité de Kurdyka-Łojasiewicz. Les assertions suivantes sont alors satisfaites :

- (i) La suite $(\nabla F(x_k))_{k \in \mathbb{N}}$ converge vers 0. La suite $(F(x_k))_{k \in \mathbb{N}}$ est décroissante et converge vers $F(\hat{x})$ où \hat{x} est un point critique de F .
- (ii) La suite $(x_k)_{k \in \mathbb{N}}$ converge vers \hat{x} .

Notons que le résultat de convergence que nous avons établi dans [50] était un peu plus général, au sens où l'algorithme étudié faisait intervenir des sous-itérations dans la règle de mise à jour à l'intérieur du sous-espace. Cependant, nous avons remarqué qu'en pratique, ces sous-itérations n'apportaient pas d'accélération significative à l'algorithme. Par ailleurs, notons qu'une extension de ce résultat, pour le traitement de fonctions à variables complexes a été présentée dans notre article [95], dans le cadre de l'application de l'Algorithme 6.7 à la reconstruction d'images en Imagerie par Résonance Magnétique multi-canaux.

Vitesse de convergence : Présentons maintenant le résultat de vitesse de convergence que nous avons établi dans [62]. Considérons une fonction F fortement convexe de la forme :

$$(\forall x \in \mathbb{R}^N) \quad F(x) = \frac{1}{2}x^\top R x - r^\top x + \Psi(x), \quad (6.9)$$

où $R \in \mathbb{R}^{N \times N}$ est symétrique définie positive, $r \in \mathbb{R}^N$, et $\Psi : \mathbb{R}^N \mapsto \mathbb{R}$ est une fonction bornée inférieurement, deux fois continuellement différentiable et convexe sur \mathbb{R}^N . Dans ce cas, une forme générique pour la matrice de majoration $A(\cdot)$ peut être :

$$(\forall x \in \mathbb{R}^N) \quad A(x) = R + B(x), \quad (6.10)$$

telle qu'il existe une matrice $V \in \mathbb{R}^{N \times N}$ définie positive telle que, pour tout $k \in \mathbb{N}$, $\nabla^2 \Psi(x_k) \preceq B(x_k) \preceq V$.

Théorème 6.3.2 Soit F donnée par (6.9), et $(x_k)_{k \in \mathbb{N}}$ la suite générée par l'Algorithme 6.7. Soit $\epsilon \in]0, +\infty[$ tel que $\epsilon \mathbf{I}_N \prec R$ et $k_\epsilon \in \mathbb{N}$ tel que, pour tout $k \geq k_\epsilon$, $\nabla^2 F(x_k) \succeq R - \epsilon \mathbf{I}_N$. Alors pour tout $k \geq k_\epsilon$,

$$F(x_{k+1}) - \inf F \leq \theta_k (F(x_k) - \inf F), \quad (6.11)$$

avec $\theta_k = 1 - (1 + \epsilon)^{-1} \tilde{\theta}_k$, et

$$\tilde{\theta}_k = \frac{(\nabla F(x_k))^\top C(x_k) \nabla F(x_k)}{(\nabla F(x_k))^\top (\nabla^2 F(x_k))^{-1} \nabla F(x_k)}, \quad C(x_k) = D_k (D_k^\top A(x_k) D_k)^\dagger D_k^\top. \quad (6.12)$$

A partir du Théorème 6.3.2, on peut montrer (voir détails dans [62]) que θ_k appartient à $]0, 1[$ et est maximal (ce qui correspond au pire cas en terme de vitesse de convergence) lorsque $D_k = \nabla F(x_k)$. Dans ce cas, la vitesse est d'autant plus lente que la matrice Hessienne de F en x_k est mal conditionnée. A l'inverse, lorsque le sous-espace est tel que $\text{ran} D_k = \mathbb{R}^N$, θ_k est minimal (i.e. la convergence est rapide) et est

d'autant plus petit que la fonction quadratique majorante constitue une approximation fine de la fonction à minimiser. Un taux intermédiaire est obtenu pour le sous-espace de mémoire de gradient employé dans 3MG, avec l'avantage important d'une complexité réduite par itération. Cela confirme les résultats observés en pratique sur différentes applications [49, 50]. Enfin, soulignons qu'en conséquence du Théorème 6.3.2, l'algorithme MM sous-espace converge linéairement, avec un taux directement lié au spectre de $R+V$, et à la norme de la matrice $(A(x_k))^{\frac{1}{2}}(\nabla^2 F(x_k))^{-1}(A(x_k))^{\frac{1}{2}}$ quantifiant la qualité d'approximation de la fonction majorante. Outre son interprétation intéressante, le résultat que nous avons obtenu est très novateur, les seules analyses de la vitesse de convergence des approches MM quadratiques existant dans la littérature étant limitées au cas des algorithmes de type semi-quadratiques (correspondant au cas d'un sous-espace engendrant l'espace \mathbb{R}^N complet) [6, 150].

6.3.3 Algorithme 3MG dans le cas stochastique

Position du problème : De nombreux problèmes rencontrés en traitement du signal et des images et en apprentissage, reposent sur la résolution d'un problème d'optimisation stochastique de la forme :

$$\underset{x \in \mathbb{R}^N}{\text{minimiser}} \left(F(x) = \frac{1}{2} \mathbb{E}(\|y_k - H_k^\top x\|^2) + \Psi(x) \right), \quad (6.13)$$

où $(H_k)_{k \geq 1}$ sont des matrices aléatoires de $\mathbb{R}^{N \times Q}$ et $(y_k)_{k \geq 1}$ des vecteurs aléatoires de \mathbb{R}^Q vérifiant les propriétés de stationnarité au sens large :

$$(\forall n \in \mathbb{N}^*) \quad \mathbb{E}(\|y_k\|^2) = \varrho, \quad \mathbb{E}(H_k y_k) = r, \quad \mathbb{E}(H_k H_k^\top) = R, \quad (6.14)$$

$\varrho \in \mathbb{R}$, $r \in \mathbb{R}^N$, $R \in \mathbb{R}^{N \times N}$, et $\Psi: \mathbb{R}^N \rightarrow \mathbb{R}$ est une fonction de régularisation. Le Problème (6.13) se rencontre dans de nombreuses applications telles que l'identification de systèmes, l'égalisation de canal, la prédiction linéaire, l'annulation d'échos et la classification supervisée [129, 107, 73]. Dans cette dernière application, $Q = 1$ et $(H_k, y_k)_{k \geq 1}$ est un ensemble de paires entrée-sortie constituant la base d'entraînement, avec $(H_k)_{k \geq 1}$ les fonctions de bases ("features") et $(y_k)_{k \geq 1}$ les indices de classe correspondants, prenant des valeurs discrètes. Même si des mesures de risque non quadratiques, telles que la fonction de coût logistique, sont souvent utilisées dans ce contexte, l'emploi d'un critère de moindres carrés reste un choix compétitif, de par sa simplicité de mise en œuvre [184]. Dans ce contexte, le terme de régularisation permet de limiter le phénomène de sur-apprentissage qui peut advenir lorsque le nombre de fonctions de bases est élevé. La restauration de signaux/images constitue un autre domaine d'application pour la résolution du Problème (6.13). Dans ce cas, $(y_k)_{k \geq 1}$ sont des observations liées à un signal original inconnu \bar{x} via des produits avec des matrices d'observations $(H_k)_{k \geq 1}$ et dégradées par un bruit d'acquisition. Chaque matrice H_k^\top , avec $k \in \mathbb{N}^*$ correspond en fait à Q lignes d'une matrice d'acquisition complète H , et sont considérées comme aléatoires. Sous des hypothèses

adaptées de stationnarité, le terme d'attache aux données associées peut s'écrire $\mathbb{E}(\|y_k - H_k^\top x\|^2)/2$, tandis que le terme de régularisation Ψ permet de contraindre le caractère mal posé du problème inverse, en introduisant des connaissances *a priori* sur la solution recherchée. Par exemple, cette fonction peut intégrer des hypothèses de parcimonie, possiblement dans un domaine transformé, sur x .

Une difficulté majeure se pose quand le Problème (6.13) ne peut pas être résolu de façon directe, car les moments statistiques du second ordre ρ , r et R ne sont pas accessibles, et doivent être estimés par une stratégie en ligne. C'est fréquemment le cas dans le contexte de l'apprentissage ou du traitement du signal adaptatif. Dans le cas simple où $\Psi \equiv 0$, les algorithmes de type moindres carrés récursif (*Recursive Least Squares*, RLS) permettent de résoudre efficacement le Problème (6.13) [107]. Quand Ψ est non nulle, des algorithmes d'approximation stochastique peuvent être utilisés, souvent fondés sur un schéma itératif de type descente de gradient stochastique (*Stochastic Gradient Descent*, SGD) [171, 91, 24]. Une étude bibliographique des approches de type SGD, incluant leurs récentes extensions proximales, peut être trouvée dans notre article [153, Sec.IV]. Les algorithmes MM ont été peu étudiés dans un contexte stochastique. Nous pouvons mentionner les travaux récents [199, 137, 38] où sont introduits des approches MM stochastiques pour la minimisation de fonctions de coûts non nécessairement quadratiques. Néanmoins, la convergence de ces algorithmes a été étudiée sous des hypothèses spécifiques (par exemple, l'indépendance des variables aléatoires impliquées [199, 137]). Dans cette section, je présente un travail que j'ai récemment publié dans [63], proposant un nouvel algorithme d'optimisation MM stochastique intégrant une stratégie d'accélération par sous-espace.

Algorithme proposé : La stratégie proposée consiste à appliquer, à chaque $k \in \mathbb{N}^*$, une itération de l'approche MM par sous espace décrite en Section 6.3, sur l'approximation stochastique F_k de F définie par :

$$(\forall x \in \mathbb{R}^N) \quad F_k(x) = \frac{1}{2k} \sum_{i=1}^k \|y_i - H_i^\top x\|^2 + \Psi(x) \quad (6.15)$$

$$= \frac{1}{2} \rho_k - r_k^\top x + \frac{1}{2} x^\top R_k x + \Psi(x) \quad (6.16)$$

avec

$$\rho_k = \frac{1}{k} \sum_{i=1}^k \|y_i\|^2, \quad r_k = \frac{1}{k} \sum_{i=1}^k H_i y_i, \quad R_k = \frac{1}{k} \sum_{i=1}^k H_i H_i^\top. \quad (6.17)$$

Afin d'appliquer le schéma MM, il nous faut construire, pour tout $k \in \mathbb{N}^*$, une fonction quadratique majorante de F_k en x_k . Pour cela, précisons la forme de la fonction de régularisation employée. Par la suite, on supposera que Ψ s'écrit :

$$(\forall x \in \mathbb{R}^N) \quad \Psi(x) = \frac{1}{2} x^\top V_0 x - v_0^\top x + \sum_{s=1}^S \psi_s(\|V_s x - v_s\|), \quad (6.18)$$

Nom du sous-espace	Matrice des directions D_k
Gradient	$[\mathbf{x}_k, -\nabla F_k(\mathbf{x}_k)]$
Mémoire de gradient	$[\mathbf{x}_k, -\nabla F_k(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}_{k-1}]$
Super mémoire de gradient	$[\mathbf{x}_k, -\nabla F_k(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}_{n-1}, \dots$ $\dots, \mathbf{x}_{k-M_k+3} - \mathbf{x}_{k-M_k+2}]$
Sous-espace de gradient	$[\mathbf{x}_k, -\nabla F_k(\mathbf{x}_k), -\nabla F_{k-1}(\mathbf{x}_{k-1}), \dots$ $\dots, -\nabla F_{k-M_k+2}(\mathbf{x}_{k-M_k+2})]$
Sous-espace de Nesterov	$[\mathbf{x}_k, -\nabla F_k(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}_{k-1},$ $\nabla F_k(\mathbf{x}_k) - \nabla F_{k-1}(\mathbf{x}_{k-1})]$

Table 6.1 – Exemple de sous-espaces stochastiques. Le nombre de colonnes M_k de D_k est strictement supérieure à 2.

où $v_0 \in \mathbb{R}^N$, $V_0 \in \mathbb{R}^{N \times N}$ est une matrice symétrique semi-définie positive, et pour tout $s \in \{1, \dots, S\}$, $v_s \in \mathbb{R}^{P_s}$, $V_s \in \mathbb{R}^{P_s \times N}$, et $\psi_s: \mathbb{R} \rightarrow \mathbb{R}$ est une fonction différentiable vérifiant les hypothèses de la Propriété 5.3.1. Cette dernière propriété nous permet de construire, pour tout $k \in \mathbb{N}^*$, une fonction quadratique majorante de F_k en \mathbf{x}_k de courbure $A_k(\mathbf{x}_k)$. L'utilisation d'une règle de mise à jour récursive pour les paramètres $(\rho_k, r_k, R_k)_{k \geq 1}$ nous permet alors de déduire l'algorithme MM par sous-espace stochastique :

$$\begin{array}{l}
 r_0 = \mathbf{0}, R_0 = \mathbf{0} \\
 \text{Initialiser } \mathbf{x}_1 \in \mathbb{R}^N \\
 \text{Pour } k = 1, 2, \dots \\
 \left[\begin{array}{l}
 r_k = r_{k-1} + \frac{1}{k}(\mathbf{H}_k \mathbf{y}_k - r_{k-1}) \\
 R_k = R_{k-1} + \frac{1}{k}(\mathbf{H}_k \mathbf{H}_k^\top - R_{k-1}) \\
 \nabla F_k(\mathbf{x}_k) = R_k \mathbf{x}_k - r_k + \nabla \Psi(\mathbf{x}_k) \\
 \text{Choisir } D_k \text{ en utilisant } \nabla F_k(\mathbf{x}_k) \\
 B_k = D_k^\top A_k(\mathbf{x}_k) D_k \\
 \mathbf{u}_k = B_k^\dagger D_k^\top (A_k(\mathbf{x}_k) \mathbf{x}_k - \nabla F_k(\mathbf{x}_k)) \\
 \mathbf{x}_{k+1} = D_k \mathbf{u}_k
 \end{array} \right. \quad (6.19)
 \end{array}$$

Lorsque le sous-espace de mémoire de gradient stochastique

$$(\forall k \in \mathbb{N}^*) \quad D_k = \begin{cases} [-\nabla F_k(\mathbf{x}_k), \mathbf{x}_k, \mathbf{x}_k - \mathbf{x}_{k-1}] & \text{si } k > 1 \\ [-\nabla F_k(\mathbf{x}_1), \mathbf{x}_1] & \text{si } k = 1 \end{cases} \quad (6.20)$$

est sélectionné, on obtient un algorithme analogue à une version stochastique de la méthode 3MG, appelé S3MG (*Stochastic 3MG*), dont les bonnes performances pratiques seront illustrées dans la Section 6.3.4. D'autres choix possibles de sous-espace, garantissant la convergence de l'algorithme, sont présentés dans le Tableau 6.1.

Résultat de convergence : Je présente ici le résultat que nous avons établi dans [63] concernant les propriétés de convergence de l’Algorithme 6.19.

Théorème 6.3.3 *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ l’espace de probabilité sous-jacent. Pour tout $k \in \mathbb{N}^*$, notons $\mathcal{X}_k = \sigma((H_i, y_i)_{1 \leq i \leq k})$ la sous-sigma algèbre de \mathcal{F} engendrée par $(H_i, y_i)_{1 \leq i \leq k}$. Supposons que :*

- $R + V_0^\top V_0$ est une matrice définie positive.
- $((H_k, y_k))_{k \geq 1}$ est une suite stationnaire ergodique et pour tout $k \in \mathbb{N}^*$, les éléments de H_k et les entrées de y_k ont des moments d’ordre quatre finis.
- $(\forall k \in \mathbb{N}^*) \mathbb{E}(\|y_{k+1}\|^2 | \mathcal{X}_k) = \varrho$, $\mathbb{E}(H_{k+1}y_{k+1} | \mathcal{X}_k) = r$ et $\mathbb{E}(H_{k+1}H_{k+1}^\top | \mathcal{X}_k) = R$.
- Pour tout $k \in \mathbb{N}^*$, $\{\nabla F_k(x_k), x_k\} \subset \text{ran } D_k$.
- x_1 est \mathcal{X}_1 -mesurable et, pour tout $k \in \mathbb{N}^*$, D_k est \mathcal{X}_k -mesurable.

Soit $(x_k)_{k \geq 1}$ la suite générée par l’Algorithme 6.19. Alors les assertions suivantes sont vérifiées :

- (i) L’ensemble des points d’accumulation de $(x_k)_{k \geq 1}$ est presque sûrement un ensemble compact connexe non vide.
- (ii) Tout élément de cet ensemble est presque sûrement un point critique de F .
- (iii) Si les fonctions $(\psi_s)_{1 \leq s \leq S}$ sont convexes, alors la suite $(x_k)_{k \geq 1}$ converge presque sûrement vers le minimiseur unique (et global) de F .

Faisons quelques commentaires sur ces résultats. Tout d’abord, soulignons que l’hypothèse de stationnarité ergodique que nous faisons est moins restrictive que l’hypothèse i.i.d. usuellement employée dans les études de convergence des algorithmes de type gradient stochastique accéléré (voir, par exemple [137]). Le résultat de convergence que nous obtenons sur la suite $(x_k)_{k \geq 1}$ se démarque également des études existantes, au sens où il s’applique à la suite des itérées, et non à une version moyennée de celle-ci, et qu’il est obtenu dans un sens “presque sûr” (et non au sens de la moyenne quadratique). De plus, soulignons que l’Algorithme 6.19 traite vraiment le problème stochastique initial (6.13), et non une approximation par somme finie comme le font les algorithmes incrémentaux proposés dans la littérature de l’apprentissage [82, 176, 199, 81]. En particulier, il ne requiert pas de passer plusieurs fois sur les données pour parvenir à la convergence.

Vitesse de convergence : Nous avons présenté dans [62], une analyse de la vitesse de convergence de l’Algorithme 6.19, qui peut être vue comme une généralisation du Théorème 6.3.2 au cas stochastique. Il est intéressant de souligner que, comme dans le cas déterministe, la moins bonne vitesse théorique est obtenue pour le sous-espace de gradient, défini, pour tout $k \in \mathbb{N}$, par $D_k = [\nabla F_k(x_k), x_k]$. A l’opposé, le meilleur taux est obtenu lorsque les colonnes de D_k engendrent tout l’espace \mathbb{R}^N .

Version adaptative : Une version adaptative de l’algorithme S3MG est aussi disponible dans notre travail [63], permettant de gérer de façon automatique de

potentielles variations des statistiques du flux entrant. Un facteur d'oubli est intégré dans l'approximation stochastique de la fonction de coût, s'inspirant des travaux existant dans la littérature du filtrage adaptatif [84]. Cela représente une rupture par rapport aux approches d'apprentissage en ligne classiques qui ne possèdent pas cette faculté d'adaptivité. Les bonnes performances numériques de ce schéma, en comparaison avec plusieurs techniques de filtrage adaptatif de l'état de l'art, ont été illustrées dans [63] sur un exemple d'estimation de filtre parcimonieux.

6.3.4 Application à l'identification en ligne de flou 2D

Terminons cette section par une illustration des performances de l'Algorithme 6.19 sur l'exemple d'identification de flou 2D dans un contexte en ligne, issu de nos travaux [63, 65]. L'objectif est d'estimer un flou 2D $h \in \mathbb{R}^M$, à partir d'un couple d'images $(x, y) \in (\mathbb{R}^N)^2$, où y est liée à x par le modèle $y = S(\bar{h})x + w$. L'opérateur $S(h) \in \mathbb{R}^{N \times N}$ est la matrice bloc-circulante associée à un noyau h , telle que $S(h)x$ corresponde à la convolution 2D entre x et h et $w \in \mathbb{R}^N$ représente un bruit additif gaussien de moyenne nulle, et de variance σ^2 . Plaçons-nous dans le contexte où N est très grand, de telle sorte que l'estimation de h par une méthode d'inversion standard s'avère trop coûteuse en mémoire. Dans ce contexte, une approche de résolution consiste à estimer le flou 2D par une stratégie en ligne, en balayant successivement des blocs de pixels extraits des images x et y . Le problème d'optimisation considéré s'écrit alors :

$$\underset{h \in \mathbb{R}^M}{\text{minimiser}} \left(F(h) = \frac{1}{2} \mathbb{E}(\|y_k - X_k^\top h\|^2) + \Psi(h) \right), \quad (6.21)$$

avec $y_k \in \mathbb{R}^Q$ un groupe de Q pixels de y , et $X_k \in \mathbb{R}^{Q \times N}$, les Q lignes correspondantes de la matrice $X \in \mathbb{R}^{N \times N}$ (avec $S(h)x = Xh$). La Figure 6.2 montre les images (x, y) , ainsi que le noyau original et son estimation obtenue par l'Algorithme 6.19, lorsque Ψ est choisie comme une version lissée de la variation totale de h . La Figure 6.3(a) compare l'évolution de l'erreur quadratique relative de reconstruction au cours du temps, pour la méthode S3MG, et trois méthodes concurrentes, à savoir SGD [171], SAGA [81] et RDA [199]. Pour ces trois algorithmes, le réglage du pas est effectué manuellement de façon à optimiser les performances pratiques. Nous pouvons observer les très bonnes performances du S3GM. Elles s'expliquent tout d'abord par le fait qu'il n'y a pas de réglage de ce type à effectuer dans S3MG, le pas multi-dimensionnel associé au sous-espace étant automatiquement déterminé par la formule MM. De plus, la forme quadratique de la fonction de coût, et la stratégie MM appliquée sur la pénalisation, permettent d'introduire des informations de second ordre sur le critère, ce qui n'est pas possible dans les 3 autres stratégies testées. Je montre enfin en Figure 6.3(b) l'évolution de l'erreur pour S3MG, et pour d'autres variantes de l'Algorithme 6.19, associé aux sous-espaces de gradient (S2M GD), et super-mémoire de gradient (S2M SMG), pour différentes valeurs de la

mémoire M_k . Nous pouvons observer que S2M GD est le plus lent, ce qui était prédictible par les résultats obtenus dans l'étude théorique de vitesse de convergence. Par ailleurs, l'augmentation de la mémoire n'amène pas d'accélération significative de la convergence, et S3MG réalise la meilleure performance sur cet exemple. Il est intéressant de souligner que ces observations sont similaires à celles effectuées dans les exemples de restauration d'images présentés dans [50, 49] dans le contexte déterministe.

6.4 CONCLUSION

Dans ce chapitre, j'ai présenté deux familles d'algorithmes d'optimisation basés sur le principe MM. La première approche permet de minimiser une somme d'un terme lisse et d'un terme non lisse en utilisant une version préconditionnée de l'algorithme proximal explicite-implicite. L'utilisation d'une stratégie MM pour le choix du préconditionneur nous a conduit non seulement à de très bonnes performances pratiques, mais aussi à de nouvelles propriétés de convergence sur les itérées générées, vérifiées pour un critère non nécessairement convexe vérifiant l'inégalité de KL. La deuxième approche, traitant le cas de fonctions différentiables, se base sur un schéma MM par sous-espace. Là encore, nous avons démontré la convergence des itérées sous hypothèse KL. La méthode obtenue est très efficace en pratique, en particulier lorsqu'elle est associée au sous-espace de mémoire de gradient. Une extension de cette approche au cas de problèmes d'optimisation stochastique avec coût quadratique a de plus été proposée, avec des garanties de convergence établies, et de très bonnes performances pratiques. Dans le chapitre suivant, je m'intéresserai à l'intégration de stratégies de minimisation alternée au sein de ces deux types d'algorithmes, dans le but de les rendre plus flexibles et d'en diminuer la complexité dans le contexte de la grande dimension.

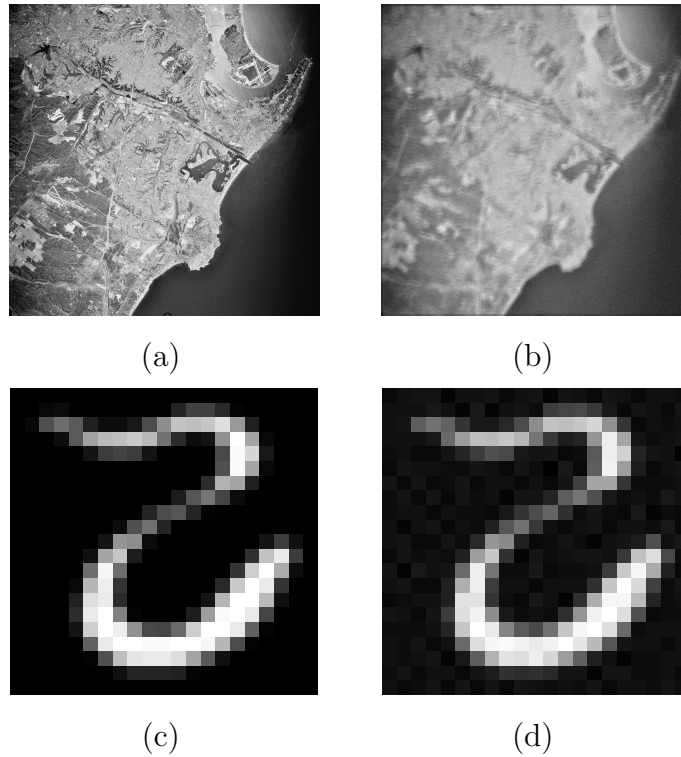


Figure 6.2 – (a) Image originale x de taille 4096×4096 ; (b) Image dégradée y , $\sigma = 0.03$, $BSNR = 25.7$ dB; (c) Flou original \bar{h} de taille 21×21 ; (d) Flou estimé, erreur relative 0.064.

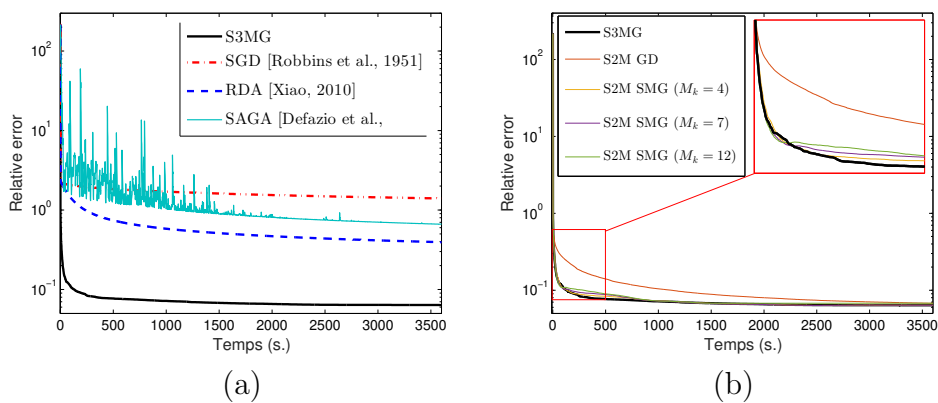


Figure 6.3 – (a) Comparaison des algorithmes $S3MG$, SGD [171], $SAGA$ [81] et RDA [199]; (b) Influence du sous-espace sur la vitesse de convergence.

- Chapitre 7 -

Algorithmes de majoration-minimisation alternés par blocs

7.1 INTRODUCTION

Dans de nombreuses applications, telles que la résolution de problèmes de déconvolution aveugle, ou lorsque les données considérées sont de très grande dimension, il est intéressant de considérer des méthodes de minimisation alternée pour la résolution du Problème (4.1). Définissons $(\mathbb{J}_j)_{1 \leq j \leq J}$ une partition de $\{1, \dots, N\}$ formée de $J > 2$ sous-ensembles, et, pour tout $j \in \{1, \dots, J\}$, $N_j \neq 0$ désignera le cardinal de \mathbb{J}_j . Tout vecteur $x = (x_n)_{1 \leq n \leq N} \in \mathbb{R}^N$ peut alors se décomposer par blocs :

$$x = (x^{(j)})_{1 \leq j \leq J} \in \mathbb{R}^{N_1} \times \dots \times \mathbb{R}^{N_J}, \quad (7.1)$$

avec, pour tout $j \in \{1, \dots, J\}$, $x^{(j)} = (x_n)_{n \in \mathbb{J}_j} \in \mathbb{R}^{N_j}$. Dans le reste de ce chapitre, nous noterons, pour tout $j \in \{1, \dots, J\}$, \bar{j} l'ensemble complémentaire de $\{j\}$ dans $\{1, \dots, J\}$, i.e., $\bar{j} = \{1, \dots, J\} \setminus \{j\}$ et, pour tout $x \in \mathbb{R}^N$,

$$x^{(\bar{j})} = (x^{(1)}, \dots, x^{(j-1)}, x^{(j+1)}, \dots, x^{(J)}). \quad (7.2)$$

Dans les approches de minimisation alternées, à chaque itération $k \in \mathbb{N}$, un indice $j_k \in \{1, \dots, J\}$ est sélectionné, et les entrées de x_k correspondantes sont mises à jour tandis que les autres (i.e., associées à \bar{j}_k) reste inchangées. Ce type de stratégie permet de réduire le coût de calcul et de mémoire par itération, et apporte aussi une flexibilité dans la résolution du problème. La plus simple d'entre elles est la minimisation alternée par blocs qui minimise, à chaque itération, de façon exacte la fonction de coût par rapport à un bloc donné [23, 152, 200]. Cette méthode peut être vue comme une généralisation de l'algorithme de Gauss-Seidel pour la résolution des systèmes linéaires [103]. Elle est d'ailleurs désignée dans certains ouvrages comme

algorithme non-linéaire de Gauss-Seidel [23, 152]. Dans la plupart des ouvrages de référence tels que [23, 152, 200], la méthode est présentée avec une mise à jour des blocs suivant une règle cyclique. Sa convergence peut néanmoins être garantie sous l'hypothèse plus faible d'une règle quasi cyclique [192]. Les résultats de convergence présentés dans [192] semblent de plus être les plus généraux concernant cet algorithme. Ils assurent la convergence de la suite générée dès lors que la fonction F est quasi-convexe et hemivariée en chaque bloc, et soit elle est pseudo-convexe sur chaque paire de blocs, soit elle admet au moins un minimum par rapport à chacun des blocs. Comme il est souligné dans [192], cette deuxième hypothèse est nécessaire au sens où l'algorithme peut ne pas converger si l'on suppose uniquement la convexité de F par rapport à chaque bloc (voir [156] pour une illustration). Pour pallier les problèmes de convergence rencontrés avec l'algorithme de minimisation alterné par blocs, une version proximale de ce dernier a été introduite par [13], où une étape proximale sur la fonction F relativement à un bloc donné est réalisée à chaque itération. La convergence est alors garantie dans le cas convexe, sous l'hypothèse que les termes non lisses éventuellement présents dans le critère soient séparables selon les blocs considérés. Des travaux récents [12] montrent que ce résultat de convergence reste valide dans le cas non convexe, si F vérifie l'inégalité de KL. L'inconvénient de ces approches de minimisation alternées est qu'elles ne permettent pas de tirer partie de façon explicite de propriétés de Lipschitz différentiabilité (ou d'existence de majorantes quadratiques) pouvant être vérifiées par certains termes de F . Une alternative est donc de remplacer la minimisation à chaque itération de la restriction de F (ou de sa régularisée de Moreau, dans le cas de [13]) au bloc courant, par la minimisation d'une approximation majorante de F par rapport à ce bloc. Cette idée a donné lieu à plusieurs algorithmes de type MM alternés [90, 114, 161, 110]. Les propriétés de convergence de ces approches ont été étudiées dans [110], qui établit la convergence des sous-suites vers un point stationnaire de F , pour une règle de mise à jour cyclique, sous une hypothèse de quasi-convexité et unicité du minimiseur de F par rapport à chaque blocs. Mentionnons également les travaux récents [72, 155, 170, 185] s'intéressant aux schémas proximaux alternés aléatoirement par blocs. Les algorithmes explicites-implicites alternés peuvent en effet être interprétés comme des schémas MM alternés, sous une hypothèse de majoration sur la métrique sous-jacente aux opérateurs proximaux employés. L'étude menée dans [72] permet d'établir la convergence presque sûre des itérées de ces schémas proximaux alternés par blocs, sous une hypothèse de convexité sur les fonctions, et de mise à jour aléatoire des blocs. Nous avons illustré dans [167, 165] l'applicabilité de tels résultats dans le contexte du débruitage de maillages 3D.

L'objectif de ce chapitre est de présenter plusieurs avancées méthodologiques réalisées dans le cadre des thèses de AUDREY REPETTI et FÉRIEL ABOUD, sur la problématique de la convergence des approches MM alternées par blocs dans le cadre non convexe. La Section 7.2 présente une version alternée par blocs de l'approche VMFB introduite dans le Chapitre 6, ses garanties de convergence dans

le cadre non convexe, et une application de cette méthode à la déconvolution aveugle de signaux sismiques. Dans la Section 7.3, une version duale de cette approche est décrite, ce qui conduit à une méthode rapide pour le calcul d'opérateurs proximaux de fonctions composites. Les performances de l'algorithme obtenu sont illustrées sur un problème de restauration de vidéos. Enfin, dans la Section 7.4, je m'intéresse à la construction d'une version alternée par blocs de l'algorithme 3MG introduit au Chapitre 6, dans le but de pouvoir en mettre en œuvre une implémentation parallèle efficace. Une application du schéma obtenu est donnée sur un problème de déconvolution d'images microscopiques.

7.2 ALGORITHME VMFB ALTERNÉ PAR BLOCS

Considérons le problème (4.1) avec $F = f_1 + f_2$, F coercive, $f_1 : \mathbb{R}^N \mapsto \mathbb{R}$ différentiable de gradient Lipschitz, et $f_2 : \mathbb{R}^N \mapsto]-\infty, +\infty]$ propre, semi-continue inférieurement, bornée inférieurement par une fonction affine et dont la restriction à son domaine de définition $\text{dom } f_2$ est séparable par blocs, i.e.,

$$(\forall x \in \mathbb{R}^N) \quad f_2(x) = \sum_{j=1}^J f_2^{(j)}(x^{(j)}). \quad (7.3)$$

Définissons, pour tout $x^{(\bar{j})} \in \prod_{i \in \bar{j}} \mathbb{R}^{N_i}$, la fonction $f_1^{(j)}(\cdot, x^{(\bar{j})})$ correspondant à la restriction de la fonction f_1 au j -ème bloc, en $x^{(\bar{j})}$:

$$(\forall v \in \mathbb{R}^{N_j}) \quad f_1^{(j)}(v, x^{(\bar{j})}) = f_1^{(j)}(x^{(1)}, \dots, x^{(j-1)}, v, x^{(j+1)}, \dots, x^{(J)}). \quad (7.4)$$

Nous avons proposé dans [67] une approche alternée par blocs pour la minimisation de $f_1 + f_2$, consistant à appliquer, à chaque itération $k \in \mathbb{N}$, une étape de l'Algorithme 6.3, à la fonction $f_1^{(j_k)}(\cdot, x^{(\bar{j}_k)}) + f_2^{(j_k)}$, où $j_k \in \{1, \dots, J\}$ est l'indice du bloc sélectionné. Cela donne lieu à l'Algorithme 7.5, nommé BC-VMFB (*Block Coordinate Variable Metric Forward Backward*) :

$$\boxed{\begin{array}{l} x_0 \in \text{dom } f_2, (\theta_k)_{k \in \mathbb{N}} \in]0, +\infty[. \\ \text{Pour } k = 0, 1, \dots \\ \left[\begin{array}{l} \text{Soit } j_k \in \{1, \dots, J\} \\ x_{k+1}^{(j_k)} \in \text{prox}_{\theta_k^{-1} A_{j_k}(x_k), f_2^{(j_k)}} \left(x_k^{(j_k)} - \theta_k A_{j_k}(x_k)^{-1} \nabla_{j_k} f_1(x_k) \right), \\ x_{k+1}^{(\bar{j}_k)} = x_k^{(\bar{j}_k)} \end{array} \right. \end{array}} \quad (7.5)$$

où, pour tout $j \in \{1, \dots, J\}$, pour tout $x \in \mathbb{R}^N$, $\nabla_j f_1(x) \in \mathbb{R}^{N_j}$ est le gradient partiel de f_1 par rapport à $x^{(j)}$ calculé en x , et $A_j(x) \in \mathbb{R}^{N_j \times N_j}$ est une matrice symétrique définie positive.

7.2.1 Analyse de convergence

Tout comme son analogue non alterné 6.3, l'Algorithme 7.5 s'interprète comme un algorithme MM lorsque $\theta_k \equiv 1$, à condition de choisir des métriques $(A_{j_k}(x_k))_{k \in \mathbb{N}}$ conduisant à des majorantes quadratiques de la restriction de f_1 aux blocs $(j_k)_{k \in \mathbb{N}}$. Cette observation permet de déduire certaines propriétés de convergence du BC-VMFB, à partir des travaux de [114, 161] sur les algorithmes MM alternés par blocs. Cependant, dans ces deux articles, la convergence des itérées $(x_k)_{k \in \mathbb{N}}$ vers une solution du problème est démontrée seulement sous des hypothèses très spécifiques. En particulier, il faut que la solution du problème (4.1) soit unique, et que chaque fonction $f_1^{(j)}(\cdot, x^{(\bar{j})}) + f_2^{(j)}$, $j \in \{1, \dots, J\}$ admette un unique minimiseur, à $x^{(\bar{j})}$ fixé. Je présente ci-après un résultat de convergence plus général, que nous avons établi dans [67], assurant la convergence des itérés de l'Algorithme 7.5 sous des hypothèses faibles. En particulier, nous ne faisons pas d'hypothèse de convexité sur les fonctions mises en jeu, et nous ne supposons pas une règle de mise à jour cyclique des blocs.

Théorème 7.2.1 *Supposons que :*

- *Les indices $(j_k)_{k \in \mathbb{N}}$ des blocs sont mis à jour selon une règle quasi cyclique, i.e., il existe une constante $K \geq J$ telle que :*

$$(\forall k \in \mathbb{N}) \quad \{1, \dots, J\} \subset \{j_k, \dots, j_{k+K-1}\}. \quad (7.6)$$

- *Il existe $(\underline{\nu}, \bar{\nu}) \in]0, +\infty[^2$ tels que*

$$(\forall k \in \mathbb{N}) \quad \underline{\nu} I_{N_{j_k}} \preceq A_{j_k}(x_k) \preceq \bar{\nu} I_{N_{j_k}}. \quad (7.7)$$

- *Il existe $(\underline{\theta}, \bar{\theta}) \in]0, +\infty[^2$ tel que, pour tout $k \in \mathbb{N}$, $\underline{\theta} \leq \theta_k \leq \tau - \bar{\theta}$, avec $\tau = 2$ si pour tout $j \in \{1, \dots, J\}$, $f_2^{(j)}$ est une fonction convexe, sinon $\tau = 1$.*

Supposons de plus que la fonction F satisfasse l'inégalité de Kurdyka-Łojasiewicz. Les assertions suivantes sont alors satisfaites :

- (i) *La suite $(F(x_k))_{k \in \mathbb{N}}$ converge vers $F(\hat{x})$, avec \hat{x} un point critique de F . De plus, $(F(x_k))_{k \in \mathbb{N}}$ est décroissante.*
- (ii) *La suite $(x_k)_{k \in \mathbb{N}}$ converge vers \hat{x} .*

Le résultat présenté dans [67] est en fait une version étendue du Théorème 7.2.1, s'appliquant à une version inexacte de l'Algorithme 7.5, dans laquelle le calcul de l'opérateur proximal relatif à la métrique majorante peut être effectué de façon approchée. Nous avons de plus présenté dans [67] un résultat de vitesse de convergence de l'algorithme, en fonction de l'exposant θ apparaissant dans l'inégalité de KL.

7.2.2 Application à la déconvolution aveugle de signaux sismiques

Je présente dans cette section une illustration de l'applicabilité de l'Algorithme 7.5 sur un problème de déconvolution aveugle, issu de notre article [169]. Deux autres

exemples d'application du BC-VMFB, dans le contexte de l'imagerie hyperspectrale, de la vidéo et de la reconstruction de phase, peuvent être trouvés dans [164], [1] et [163, 67], respectivement. Le problème considéré ici est celui de la restauration d'un signal sismique inconnu $\bar{x} \in \mathbb{R}^N$ à partir d'observations $y \in \mathbb{R}^N$, obtenues par le modèle $y = \bar{h} * \bar{x} + w$, où $\bar{h} \in \mathbb{R}^S$ représente un noyau de convolution inconnu, et $w \in \mathbb{R}^N$ est une réalisation d'un vecteur aléatoire modélisant un bruit additif (Fig. 7.1(a)). Nous définissons des estimées $(\hat{x}, \hat{h}) \in \mathbb{R}^N \times \mathbb{R}^S$ de (\bar{x}, \bar{h}) comme les minimiseurs de la somme d'une fonction d'attache aux données de type moindres carrés et de termes de régularisation agissant sur le signal et sur le noyau de convolution. Dans ce travail, nous avons proposé une nouvelle fonction de régularisation sur le signal, pouvant être vue comme une forme lissée de la fonction non convexe ℓ_1/ℓ_2 , dont l'intérêt est de rendre la résolution du problème aveugle moins sensible aux ambiguïtés d'échelles. Nous considérons en outre dans le critère pénalisé des fonctions indicatrices de domaine permettant de contraindre l'intensité du signal et du noyau estimé, ainsi que la norme de ce dernier. La fonction à minimiser s'écrit alors comme la somme d'un terme différentiable non convexe, et de termes non différentiables, séparables par rapport aux deux blocs, constitués de x et h , et convexes. Nous appliquons l'algorithme BC-VMFB pour résoudre le problème, en alternant K itérations sur la variable x suivie d'une itération sur la variable h . Le résultat obtenu par notre méthode pour l'estimation du noyau est représenté en Figure 7.1(b). Je donne également le résultat obtenu par la méthode concurrente de [125], également basée sur une régularisation de type ℓ_1/ℓ_2 . La Figure 7.1(c) donne le temps de reconstruction obtenu, en secondes, en fonction du nombre d'itérations internes K . On peut observer que le meilleur compromis en terme de vitesse de convergence est obtenu pour une valeur intermédiaire d'itérations internes égale à $K = 71$. Ce résultat illustre l'intérêt d'une flexibilité dans le choix des blocs dans la méthode d'optimisation alternée (la règle cyclique classiquement employée correspondant à $K = 1$).

7.3 ALGORITHME DUAL VMFB PAR BLOCS

7.3.1 Introduction

Les algorithmes proximaux 6.3 et 7.5 requièrent le calcul, à chaque itération, de l'opérateur proximal d'une fonction non lisse. De nombreuses propriétés et formules sont disponibles dans la littérature pour effectuer ce calcul de façon explicite [75, 71], incluant mes travaux [20, 47, 94]. Cependant, lorsque la fonction impliquée s'écrit comme une somme d'au moins deux termes, possiblement composés avec des opérateurs linéaires, et/ou lorsque l'opérateur proximal doit être calculé dans une métrique non triviale, la solution du problème de minimisation inhérent à la mise à jour proximale devient non explicite, et doit être obtenue par une stratégie itérative.

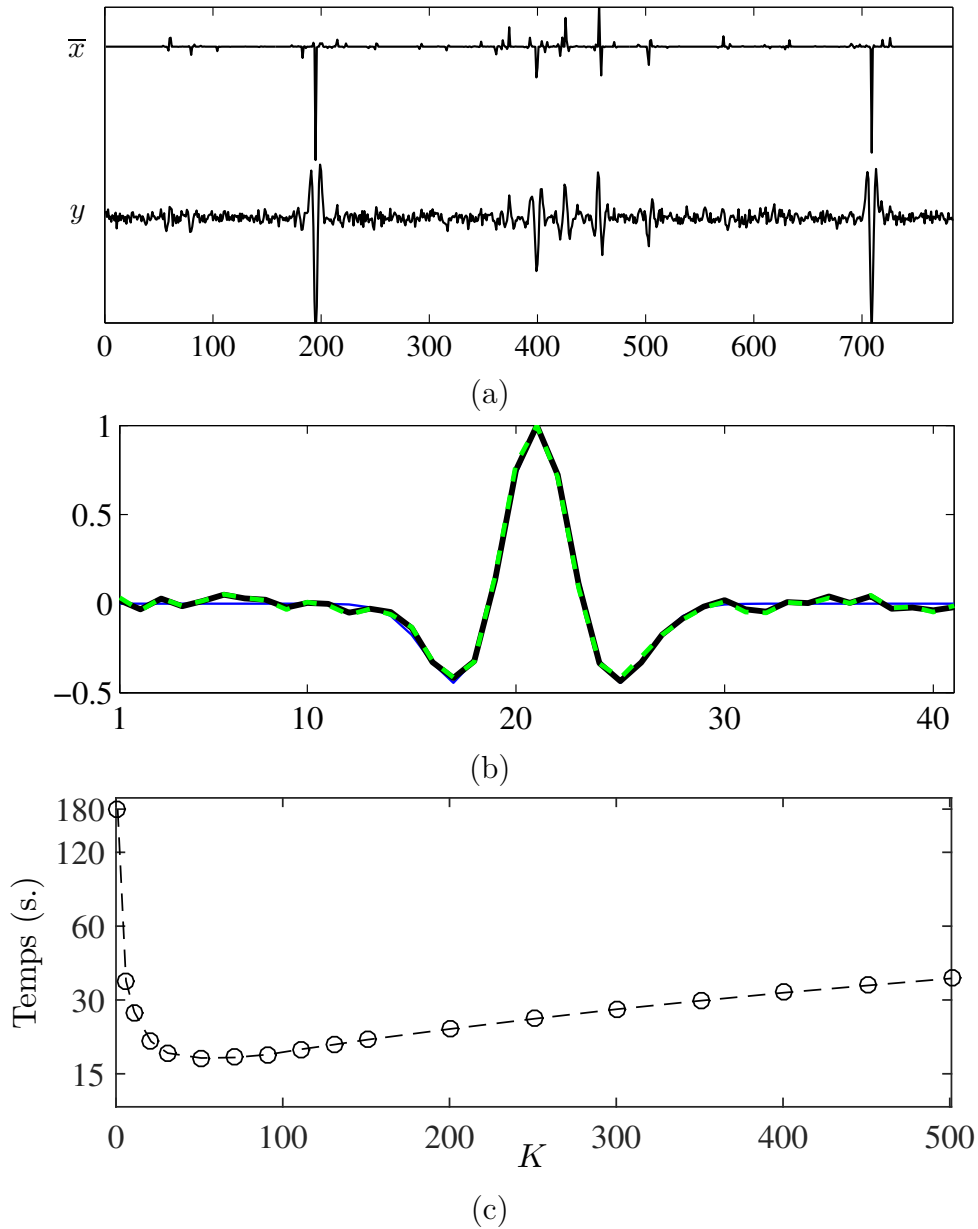


Figure 7.1 – (a) Signal original \bar{x} et dégradé y ; (b) Noyau original \bar{h} (ligne fine continue bleue), estimé par l’algorithme BC-VMFB (ligne épaisse continue noire) et par [125] (ligne discontinue épaisse verte); (c) Temps de reconstruction en fonction du nombre de sous-itérations internes K sur la variable x effectuées dans l’algorithme BC-VMFB (moyennes sur 30 réalisations de bruit).

Dans le cas convexe, l'approche la plus couramment utilisée est basée sur l'algorithme explicite-implicite dual (*Dual Forward-Backward*, DFB), qui consiste à appliquer l'algorithme FB au dual du problème, au sens de Fenchel-Rockafellar [76, 70, 71]. La convergence des itérées issues de l'algorithme DFB vers une solution primale-duale du problème a été établie dans [76], et son applicabilité pour le calcul d'opérateurs proximaux de fonctions complexes est illustrée par exemple dans [66, 31]. Dans la Section 7.3.2, je présente un travail issu de [4], dans lequel nous avons proposé une version alternée par blocs de l'Algorithme DFB permettant de gérer de façon efficace et flexible le cas où une structure de séparabilité est présente dans un des termes de la fonction impliquée.

7.3.2 Algorithme proposé et analyse de convergence

Considérons le problème du calcul de l'opérateur proximal de la fonction $F = f_1 + f_2 \circ H$, avec $f_1 : \mathbb{R}^N \rightarrow \mathbb{R}$, convexe, propre, semi-continue inférieurement, $H \in \mathbb{R}^{M \times N}$, et $f_2 : \mathbb{R}^N \rightarrow \mathbb{R}$, convexe, propre, semi-continue inférieurement. Supposons que $f_2 \circ H$ présente une structure séparable de la forme :

$$(\forall x \in \mathbb{R}^N) \quad f_2(Hx) = \sum_{j=1}^J f_2^{(j)}(H_j x), \quad (7.8)$$

avec, pour tout $j \in \{1, \dots, J\}$, $H_j \in \mathbb{R}^{M_j \times N}$ non identiquement nulle, $\sum_{j=1}^J M_j = M$, $f_2^{(j)} : \mathbb{R}^{N_j} \rightarrow \mathbb{R}^N$ propre, convexe et semi-continue inférieurement et $H = [H_1^\top | \dots | H_J^\top]^\top$. D'après la Définition 6.2.1, l'opérateur proximal de F en \tilde{x} relativement à la métrique induite par A , noté $\text{prox}_{A,F}(\tilde{x})$, est défini par :

$$\hat{x} = \underset{x \in \mathbb{R}^N}{\text{argmin}} \quad \frac{1}{2} \|x - \tilde{x}\|_A^2 + f_1(x) + \sum_{j=1}^J f_2^{(j)}(H_j x). \quad (7.9)$$

Introduisons les définitions suivantes :

Définition 7.3.1 Soit ψ et φ des fonctions de \mathbb{R}^N dans $] -\infty, +\infty]$.

(i) L'*infimale convolution* entre ψ et φ est

$$\psi \square \varphi : \mathbb{R}^N \rightarrow [-\infty, +\infty] : x \rightarrow \inf_{y \in \mathbb{R}^N} (\psi(y) + \varphi(x - y)). \quad (7.10)$$

(ii) La conjuguée de ψ est notée ψ^* et est définie par :

$$\psi^* : \mathbb{R}^N \rightarrow [-\infty, +\infty] : x \rightarrow \sup_{\nu \in \mathbb{R}^N} (\langle \nu | x \rangle - \psi(\nu)). \quad (7.11)$$

Alors, sous la contrainte de qualification ri ($A(\text{dom } f_1) \cap \text{ri}(\text{dom } f_2) \neq \emptyset$, où ri définit l'intérieur relatif d'un domaine, le problème dual associé à (7.9) est ([71]) :

$$\hat{y} = \underset{y=(y^{(j)})_{1 \leq j \leq J} \in \mathbb{R}^M}{\text{argmin}} \quad \varphi \left(-\sum_{j=1}^J A^{-1/2} H_j^\top y + A^{1/2} \tilde{x} \right) + \sum_{j=1}^J (f_2^{(j)})^*(y^{(j)}). \quad (7.12)$$

où $\varphi = ((f_1 \circ A^{-1/2})^* \square_{\frac{1}{2}} \|\cdot\|^2)$. Il est intéressant de noter que la variable duale y est décomposée en J blocs $(y^{(j)})_{1 \leq j \leq J}$, et que le deuxième terme de la fonction duale intervenant dans (7.12) est séparable par blocs. Pour cette raison, plusieurs approches de minimisation alternées par blocs ont été proposées pour résoudre (7.9)- (7.12). Citons en particulier, les algorithmes stochastiques introduits par [179, 159, 115] dans le domaine de l'apprentissage. Cependant, dans ces travaux, les méthodes proposées sont limitées à des blocs scalaires (i.e., $M_j \equiv 1$) et les garanties de convergence se limitent à des propriétés de décroissance sur la fonction duale. Mentionnons également l'approche de type FISTA alternée proposée dans [40], qui permet de traiter le cas où $M_j \equiv 2$, et assure la convergence des itérées primales. L'approche que nous avons proposé dans [4] permet de traiter le cas général (7.8), tout en bénéficiant de garanties de convergence sur les itérées primales et duales générées. Elle consiste à appliquer l'algorithme BC-VMFB introduit dans la Section 7.2 au problème dual (7.12), en utilisant le fait que la fonction φ est différentiable de gradient 1-Lipschitz. Après simplification, nous obtenons l'Algorithme 7.13 :

$$\begin{array}{l}
 \text{Initialiser } (y_0^{(j)})_{1 \leq j \leq J} \in \mathbb{R}^M, z_0 = -A^{-1} \sum_{j=1}^J H_j^\top y_0^{(j)}. \\
 \text{Poser } (\forall j \in \{1, \dots, J\}) B_j \succeq H_j A^{-1} H_j^\top. \\
 \text{Pour } k = 0, 1, \dots \\
 \left| \begin{array}{l}
 x_k = \text{prox}_{A, f_1}(\tilde{x} + z_k) \\
 j_k \in \{1, \dots, J\} \\
 \tilde{y}_k^{(j_k)} = y_k^{(j_k)} + \theta_k B_{j_k}^{-1} H_{j_k} x_k \\
 y_{k+1}^{(j_k)} = \tilde{y}_k^{(j_k)} - \theta_k B_{j_k}^{-1} \text{prox}_{\gamma_n B_{j_k}^{-1}, f_2^{(j_k)}}(\theta_k^{-1} B_{j_k} \tilde{y}_k^{(j_k)}) \\
 y_{k+1}^{(\bar{j}_k)} = y_k^{(\bar{j}_k)} \\
 z_{k+1} = z_k - A^{-1} H_{j_k}^\top (y_{k+1}^{(j_k)} - y_k^{(j_k)})
 \end{array} \right. \tag{7.13}
 \end{array}$$

A chaque itération $k \in \mathbb{N}$, un indice $j_k \in \{1, \dots, J\}$ est sélectionné, et la composante correspondante de variable duale $y^{(j_k)}$ est mise à jour par une étape explicite-implicite dans une métrique variable notée B_{j_k} . Les autres composantes de la variable duale, associées aux indices \bar{j}_k , restent inchangées.

Propriétés de convergence. Les propriétés de convergence de l'Algorithme 7.13, que nous avons établies dans [4] se déduisent du Théorème 7.2.1 :

Théorème 7.3.1 *Supposons que :*

- Pour tout $j \in \{1, \dots, J\}$, la restriction de $(f_2^{(j)})^*$ à son domaine est continue.
- Pour tout $j \in \{1, \dots, J\}$, B_j est définie positive.
- La suite $(j_k)_{k \in \mathbb{N}}$ suit une règle quasi cyclique.

Supposons en outre que F satisfait l'inégalité de Kurdyka-Łojasiewicz. Soit $(x_k)_{k \in \mathbb{N}}$ et $(y_k = (y_k^{(j)})_{1 \leq j \leq J})_{k \geq 1}$ des suites générées par l'Algorithme 7.13. Si $(y_k)_{k \geq 1}$ est bornée, alors $(x_k)_{k \in \mathbb{N}}$ converge vers la solution du problème primal (7.9) et $(y_k)_{k \geq 1}$ converge vers la solution du problème dual (7.12).

7.3.3 Application à la restauration d'archives vidéo

Je présente dans cette section une illustration de l'applicabilité de l'Algorithme 7.13, dans le cadre de la résolution du problème de restauration de vidéos que nous avons considéré dans [4, 3]. Le modèle d'observation est décrit en Figure 7.2. $(\bar{x}_t)_{1 \leq t \leq T}$ est une vidéo originale issue du site <http://media.xiph.org/video/derf/>, composée de $T = 50$ trames, de dimension $N = 288 \times 352$ pixels. La vidéo observée $(y_t)_{1 \leq t \leq T}$ correspond à une version convoluée par le noyau h , et bruitée de la vidéo originale. Nous considérons de plus que la vidéo observée est disponible dans un mode entrelacé, ce que nous modélisons par les opérateurs de décimation $(S_t)_{1 \leq t \leq T}$, extrayant les lignes paires (resp. impaires) de la vidéo pour t pair (resp. impair), de telle façon que chaque image de $(y_t)_{1 \leq t \leq T}$ est de dimension réduite $M = 144 \times 352$ pixels. Afin de restaurer la vidéo originale, nous proposons de minimiser un critère pénalisé $F = f_1 + f_2$ constitué d'un terme d'attaches aux données f_1 de moindre carrés, et de plusieurs termes de régularisation, renforçant des propriétés de régularité spatiale et temporelle sur la vidéo, regroupés dans f_2 . La minimisation de F est effectuée par le schéma alterné PALM (*Proximal Alternating Linearized Minimization*) [29], qui peut être vu comme un cas particulier de l'Algorithme 7.5, avec des métriques scalaires. Les blocs considérés sont les trames de la vidéo, ce qui fait qu'à chaque itération, il faut effectuer une étape de gradient sur la restriction de f_1 par rapport à x_t suivie d'une étape proximale sur la restriction de f_2 à x_t . Le calcul de cet opérateur proximal n'étant pas réalisable de façon explicite, nous proposons d'utiliser l'Algorithme 7.13, qui nous permet de traiter séparément tous les termes de la fonction de régularisation f_2 . Nous présentons dans la Figure 7.3 nos résultats de restauration, pour trois images extraites de la vidéo considérée. Les performances de l'Algorithme 7.13 sont illustrées en Figure 7.4. Nous avons comparé notre méthode à l'algorithme de [76], qui peut être vu comme une version parallèle de notre approche, dans laquelle tous les termes du critère sont traités à chaque itération. Sur la Figure 7.4(a), nous voyons que le temps moyen nécessaire au calcul de l'opérateur proximal par notre méthode est plus faible que si l'on utilise l'algorithme de [76], ce qui montre l'apport d'une stratégie de minimisation alternée. La Figure 7.4(b) illustre l'intérêt de pouvoir intégrer un préconditionneur diagonal dans l'Algorithme 7.13, par rapport à un préconditionneur scalaire. Pour ce dernier, il est nécessaire de calculer la norme de l'opérateur de mouvement impliqué dans le terme de régularisation temporel, ce qui engendre une étape de pré-traitement coûteuse et peu réaliste dans le contexte de traitement de vidéos. Nous montrons les résultats obtenus lorsque cette norme est calculée de façon exacte, ou par une approximation ne nécessitant pas de pré-traitement.

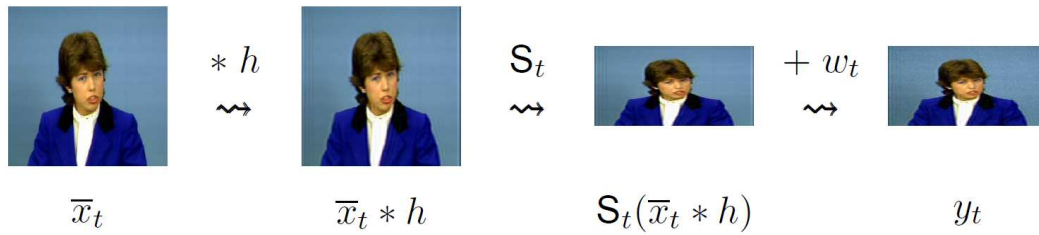


Figure 7.2 – *Modèle direct considéré, pour tout $t \in \{1, \dots, T\}$.*

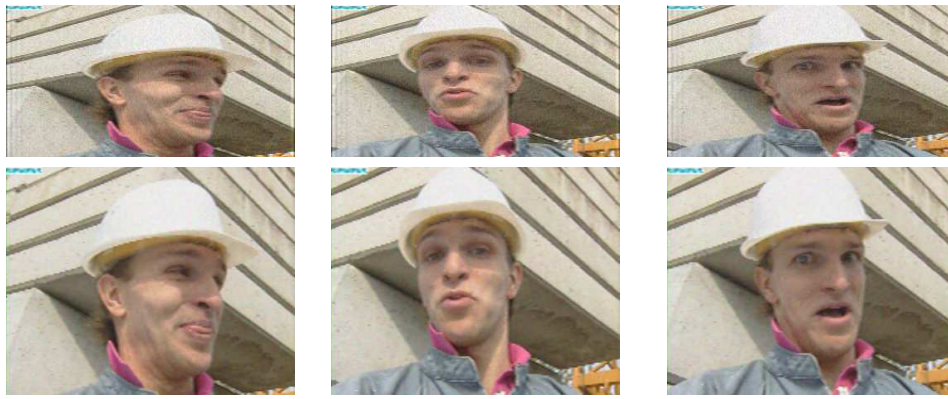


Figure 7.3 – *(haut) Images extraites de la vidéo dégradée en mode entrelacé, SNR = 25.54dB; (bas) Images de la vidéo restaurée en mode progressif, SNR = 28.95dB.*

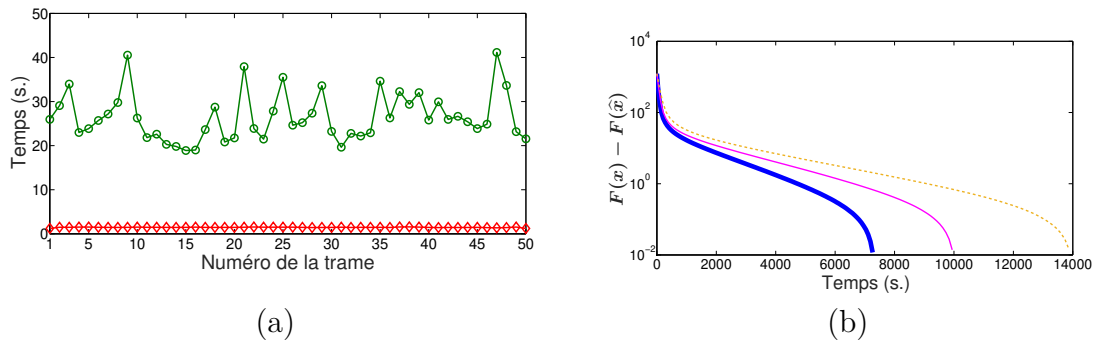


Figure 7.4 – *(a) Temps de calcul moyen par image (en s.), pour l'étape proximale de PALM, avec l'algorithme explicite-implicite dual parallèle de [76] (vert) ou l'Algorithme 7.13 (rouge); (b) Evolution du critère en fonction du temps. L'étape proximale est réalisée par l'Algorithme 7.13 avec préconditionneur diagonal (bleu), scalaire exact (magenta) ou scalaire approché (jaune).*

7.4 ALGORITHME MM PARALLÈLE ALTERNÉ PAR BLOCS

7.4.1 Introduction

Considérons la résolution du Problème (4.1), lorsque F est différentiable sur \mathbb{R}^N . Nous avons vu dans la Section 6.3 qu’une approche efficace pour minimiser F était d’adopter l’algorithme MM par sous-espace 6.7. Néanmoins, lorsque le problème est de très grande dimension (par exemple, lors du traitement de données multidimensionnelles), la mise en œuvre pratique de l’algorithme peut poser des difficultés. En particulier, des produits avec des matrices de grande taille interviennent dans le calcul du gradient de F , ainsi que dans la règle de calcul de pas (6.8), pouvant engendrer une saturation de l’espace mémoire de la machine de calcul utilisée. Un moyen efficace pour surmonter les difficultés liées aux exigences de mémoire d’un algorithme d’optimisation consiste à adopter une approche de minimisation parallèle alternée par blocs [151, 165, 124]. Comme pour les approches alternées par blocs classiques, le vecteur de variables est divisé en blocs de dimension réduite, mis à jour de façon itérative. La différence est qu’à chaque itération de l’algorithme, plusieurs blocs (et non un seul) sont sélectionnés et traités de façon parallèle selon un schéma d’optimisation donné. Cette stratégie permet de contrôler la mémoire requise et la répartition des tâches de calcul en fonction des caractéristiques de la machine utilisée. Par exemple, le nombre de blocs sélectionné à chaque itération peut être défini par l’utilisateur comme étant égal au nombre de processeurs disponibles.

Plusieurs travaux dans le contexte de la restauration d’image ont montré que les algorithmes MM peuvent s’implémenter efficacement sur une architecture parallèle [181, 132] à condition d’utiliser des fonctions majorantes ayant une structure séparable par blocs. L’utilisation d’approches alternées par blocs dans les stratégies MM est également possible [110, 114, 90], mais à notre connaissance, les méthodes résultantes ne présentent pas de structure parallèle, puisqu’un seul bloc doit être mis à jour par itération, de sorte que le gain en terme d’occupation mémoire est généralement obtenu au détriment d’une diminution substantielle de la vitesse de convergence. Afin de répondre à ces limitations, nous avons proposé dans [174] une nouvelle approche MM présentant une structure parallèle alternée par blocs, que je présente dans la Section 7.4.2. Une illustration des performances de l’algorithme que nous avons obtenues sur un exemple de déconvolution de d’images microscopiques est présentée dans la Section 7.4.3.

7.4.2 Algorithme proposé

Considérons que la fonction F à minimiser dans le Problème (4.1) soit de la forme suivante :

$$(\forall x \in \mathbb{R}^N) \quad F(x) = \sum_{s=1}^L f_s(L_s x), \quad (7.14)$$

où, pour tout $s \in \{1, \dots, S\}$, $L_s \in \mathbb{R}^{P_s \times N}$, $P_s \in \mathbb{N}^*$, et f_s est une fonction différentiable de \mathbb{R}^{P_s} dans \mathbb{R} . Dans le contexte de la résolution d'un problème inverse par estimateur MAP, L_1 va représenter la matrice d'observation liée au processus d'acquisition et la forme de f_1 est dictée par le modèle de bruit considéré. Les termes $(f_s)_{2 \leq s \leq S}$ sont des termes de régularisation intégrant des informations *a priori* sur la solution, ou plus généralement, des transformées linéaires de celle-ci, obtenues par les matrices $(L_s)_{2 \leq s \leq S}$.

Décomposons la variable x en blocs de dimension $N_j \neq 0$, $j \in \{1, \dots, J\}$, $\sum_{j=1}^J N_j = N$, comme décrit en (7.1). Nous proposons de minimiser F par un schéma MM sous-espace alterné par blocs. A chaque itération $k \in \mathbb{N}$, un sous-ensemble de blocs d'indices $\mathcal{S}_k \subset \{1, \dots, J\}$ est choisi. Les variables associées, regroupées dans le vecteur $x^{(\mathcal{S}_k)} = \left(x^{(j)}\right)_{j \in \mathcal{S}_k}$, sont mises à jour par une itération de l'Algorithme 6.7 appliqué à la restriction de F à \mathcal{S}_k , tandis que les autres variables, associées à l'ensemble d'indices complémentaire de \mathcal{S}_k noté $\bar{\mathcal{S}}_k$, restent fixées. Pour tout $x' \in \mathbb{R}^N$, on suppose que F possède une fonction majorante $Q(\cdot, x')$ de courbure $A(x') \in \mathbb{R}^{N \times N}$ symétrique définie positive. Alors, on peut en déduire aisément une majorante de la restriction de F aux variables contenues dans le bloc \mathcal{S}_k , en définissant la fonction quadratique de courbure :

$$(\forall k \in \mathbb{N}) \quad A^{(\mathcal{S}_k)}(x_k) = ([A(x_k)]_{p,p})_{p \in \mathcal{S}_k}, \quad (7.15)$$

où $\mathbb{S}_k \subset \{1, \dots, N\}$ contient les indices du bloc \mathcal{S}_k . Cette formule permet de déduire une version alternée par blocs de l'Algorithme 6.7. Cependant, cet algorithme ne présente pas de structure parallèle. En effet, la fonction majorante issue de (7.15) n'est pas séparable par rapport aux blocs $j \in \mathcal{S}_k$, et sa minimisation dans le sous-espace choisi ne peut pas être effectuée en parallèle sur ces blocs de variables. Nous avons donc proposé dans [174] de la remplacer par une approximation moins précise de F mais présentant une matrice de courbure bloc diagonale, de la forme :

$$(\forall x \in \mathbb{R}^N) \quad B^{(\mathcal{S})}(x) = \text{BDiag} \left\{ (B^{(j)}(x))_{j \in \mathcal{S}} \right\}, \quad (7.16)$$

où, pour tout $j \in \mathcal{S}$, $B^{(j)}(x) \in \mathbb{R}^{N_j \times N_j}$ est définie de telle sorte qu'on ait, pour tout $x \in \mathbb{R}^N$, la propriété de majoration $A^{(\mathcal{S})}(x) \preceq B^{(\mathcal{S})}(x)$.

La stratégie proposée dans [174] pour construire de telles matrices repose sur l'utilisation de l'inégalité de Jensen, et emprunte des techniques similaires à [80, Sec. III], [181, Sec. III-A]. Nous remplaçons donc, pour tout $k \in \mathbb{N}$, la fonction majorante initiale par une autre fonction majorante de courbure $B^{(\mathcal{S}_k)}(x)$, $x \in \mathbb{R}^N$. La structure bloc diagonale (7.16) de cette matrice permet alors d'obtenir l'Algorithme 7.17 qui présente une forme parallèle alternée par blocs :

$$\begin{array}{l}
\text{Initialiser } x_0 \in \mathbb{R}^N \\
\text{Pour } k = 0, 1, 2, \dots \\
\quad \left| \begin{array}{l}
\text{Choisir } \mathcal{S}_k \subset \{1, \dots, J\} \\
\text{Pour } j \in \mathcal{S}_k \text{ (en parallèle)} \\
\quad \left| \begin{array}{l}
\text{Construire } B_k^{(j)}(x_k) \\
\text{Choisir } D_k^{(j)} \\
s_k^{(j)} = - \left((D_k^{(j)})^\top B_k^{(j)}(x_k) D_k^{(j)} \right)^\dagger (D_k^{(j)})^\top \nabla F^{(j)}(x_k) \\
x_{k+1}^{(j)} = x_k^{(j)} + D_k^{(j)} s_k^{(j)} \\
x_{k+1}^{(\overline{\mathcal{S}}_k)} = x_k^{(\overline{\mathcal{S}}_k)}
\end{array} \right. \\
\end{array} \right.
\end{array} \quad (7.17)$$

Pour tout $k \in \mathbb{N}$, les colonnes de la matrice $D_k^{(\mathcal{S}_k)} = (D_k^{(j)})_{j \in \mathcal{S}_k}$ vont engendrer des sous-espaces restreints aux ensembles de blocs $(\mathcal{S}_k)_{k \in \mathbb{N}}$ sélectionnés. On peut par exemple définir le sous-espace de mémoire de gradient par blocs par :

$$(\forall k \in \mathbb{N})(\forall j \in \mathcal{S}_k) \quad D_k^{(j)} = \begin{cases} -\nabla F^{(j)}(x_k) & \text{si } j \notin \bigcup_{\ell=0}^{k-1} \mathcal{S}_\ell, \\ \left[-\nabla F^{(j)}(x_k) \middle| x_k^{(j)} - x_{k-1}^{(j)} \right] & \text{sinon,} \end{cases} \quad (7.18)$$

où, pour tout $j \in \{1, \dots, J\}$, $\nabla F^{(j)}$ est le gradient de F par rapport au bloc j . Il est important de remarquer qu'à chaque itération de l'Algorithme 7.17, la mise à jour d'un bloc $\in \{1, \dots, J\}$ nécessite d'avoir accès à la valeur de toutes les composantes du vecteur x_k . Du point de vue de l'implémentation sur une architecture multi-cœurs, cela signifie qu'à la fin de chaque itération, tous les coeurs de la machine doivent échanger entre eux afin d'avoir à disposition les informations utiles pour pouvoir continuer l'algorithme. Nous avons montré dans [174] que le contenu de ces échanges, et par conséquent le temps d'exécution global, était en fait directement liée à la structure des opérateurs linéaires $(L_s)_{1 \leq s \leq S}$ intervenant dans le critère (7.14), et plus précisément à leurs propriétés de parcimonie.

7.4.3 Application à la déconvolution d'images 3D

Illustrons maintenant les performances de l'Algorithme 7.17 sur le problème que nous avons traité dans [174], dans le cadre du projet collaboratif CNRS OPTIMISM visant à proposer des outils de restauration d'image en microscopie biphotonique. L'objectif est de restaurer un volume original $\bar{x} \in \mathbb{R}^N$, $N = N_1 \times N_2 \times N_3$, à partir d'une version dégradée de celui-ci, décrite par $y = H\bar{x} + w$, $y \in \mathbb{R}^N$, $H \in \mathbb{R}^{N \times N}$ un opérateur de flou 3D, et $w \in \mathbb{R}^N$ un bruit additif supposé Gaussien, de moyenne nulle et de variance σ^2 . Nous considérons un modèle de flou 3D variant avec l'indice $\ell \in \{1, \dots, N_3\}$ de l'axe Oz, modélisant les phénomènes optiques intervenant en microscopie biphotonique. La difficulté majeure avec ce type de modèle est que l'application de H est très coûteuse en temps de calcul et occupation mémoire, ce qui nuit aux performances pratiques des algorithmes de restauration. Nous avons

proposé d'appliquer l'Algorithme 7.17, en définissant les blocs comme les $J = N_3$ tranches 2D du volume 3D. Ainsi, l'application de H à un bloc est peu coûteuse, le flou 3D étant supposé peu étalé suivant Oz, et stationnaire pour $\ell \in \{1, \dots, N_3\}$ donné. L'algorithme est implémenté en Matlab, en utilisant les commandes MPI (*Message Passing Interface*) de la toolbox Parallel Computing. Nous répartissons les \overline{C} cœurs actifs de la machine multi-processeurs utilisée en un cœur maître et $C = \overline{C} - 1$ cœurs esclaves. Le cœur maître gère la boucle globale de l'Algorithme 7.17. A chaque itération $k \in \mathbb{N}$, il choisit C indices de blocs, stockés dans \mathcal{S}_k , et envoie à chaque cœur esclave les informations dont il a besoin pour mettre à jour un bloc j parmi ceux qui ont été sélectionnés. Les cœurs esclaves accomplissent leurs tâches de mises à jour de façon parallèle, puis renvoient les résultats au cœur maître.

Je présente les résultats obtenus pour deux exemples d'images microscopiques 3D FlyBrain¹ de taille $N = 256 \times 256 \times 48$ et Tube² de taille $N = 284 \times 280 \times 48$. Pour ces deux exemples, H modélise un flou 3D gaussien, de taille $5 \times 5 \times 11$, dont la matrice de covariance varie avec la profondeur. Des coupes extraites des volumes originaux, dégradés et restaurés sont illustrées en Figure 7.5. Nous montrons en Figure 7.6 l'accélération en terme de temps de calcul obtenue par une augmentation du nombre de cœurs actifs \overline{C} . Ces résultats montrent l'efficacité de l'implémentation parallèle utilisée, qui conduit à une accélération linéaire en fonction de \overline{C} , malgré la structure non séparable par blocs du problème traité.

7.5 CONCLUSION

Dans ce chapitre, j'ai présenté plusieurs méthodes de type MM, intégrant une stratégie alternée par blocs. Une telle stratégie apporte une grande flexibilité dans la mise en œuvre des algorithmes puisqu'elle permet de contrôler le temps de calcul et l'allocation de mémoire par itération. Les algorithmes MM alternés par blocs utilisent des fonctions majorantes de la restriction du critère par rapport aux blocs sélectionnés. Ces majorantes sont souvent plus simples à construire, et de courbure moins forte que des fonctions majorant le critère dans sa globalité, ce qui peut entraîner en pratique une accélération significative de la vitesse de convergence [67, 1]. Notons en outre que les résultats de convergence que nous avons obtenus sont établis pour une mise à jour quasi cyclique des blocs, ce qui apporte une flexibilité supplémentaire à l'algorithme de minimisation, et peut présenter un réel avantage pratique [169]. Les stratégies par blocs proposées ouvrent enfin la voie à un traitement efficace sur des architectures de calcul modernes des problèmes inverses de grande dimension, comme nous l'avons illustré dans la Section 7.4. Notons que nous avons également récemment proposé, dans le cadre de la thèse de FERIEL ABBOUD, une nouvelle forme de l'algorithme dual explicite implicite alterné par blocs 7.13,

1. <http://imagej.nih.gov/ij/plugins/3d-viewer/>
 2. <http://adm.irbbarcelona.org/image-j-fiji>

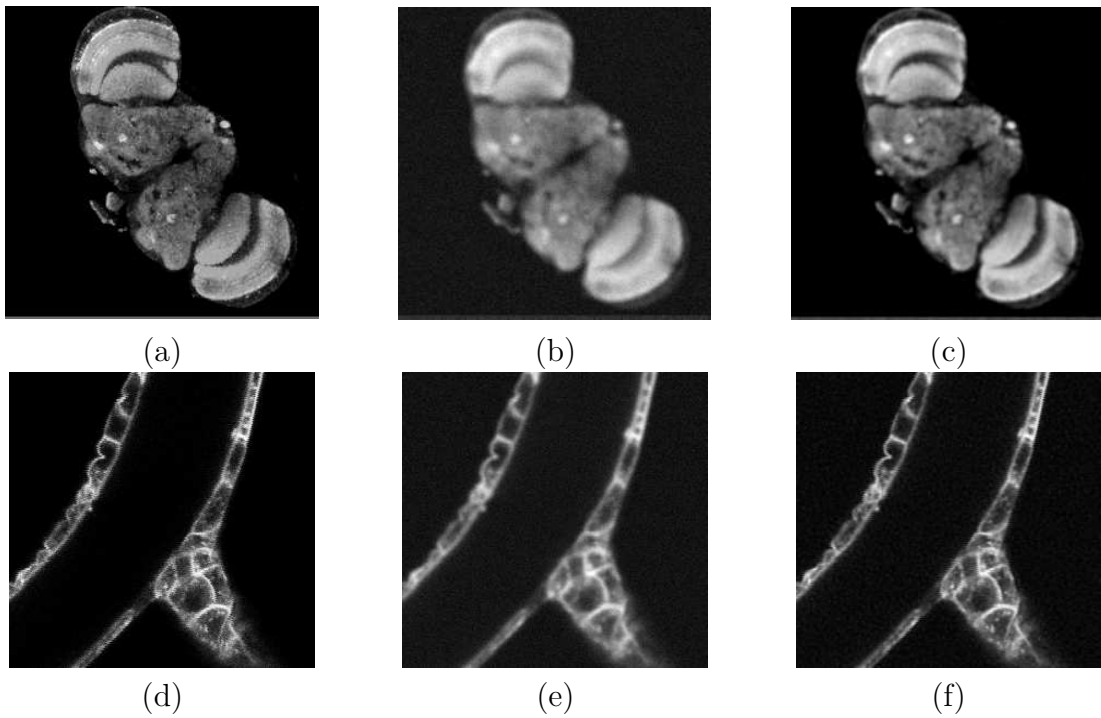


Figure 7.5 – (a)-(d) Coupe extraite du volume 3D original *FlyBrain* (resp. *Tube*) ; (b)-(e) Coupe du volume dégradé, $SNR = 13.42$ dB (resp. 11.53 dB) ; (c)-(f) Coupe du volume restauré par l'algorithme *BP3MG*, $SNR = 16.98$ dB (resp. 14.47 dB).

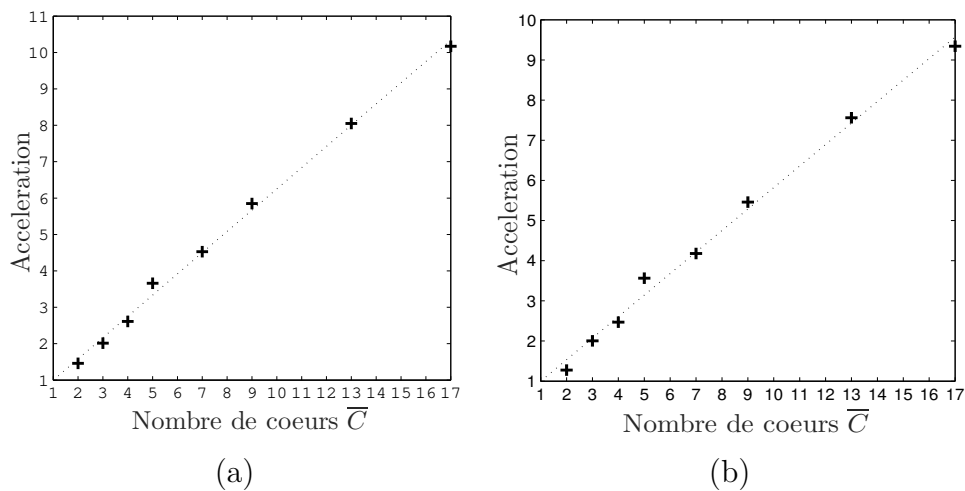


Figure 7.6 – Ratio entre le temps de calcul obtenu pour un cœur et le temps de calcul pour \bar{C} cœur (croix) et régression linéaire (pointillés), pour la restauration de *FlyBrain* (a) et *Tube* (b).

adaptée à une architecture de calcul distribué [2], avec l'objectif d'un traitement accéléré du problème de restauration de vidéos introduit en Section 7.3.3.

- Chapitre 8 -

Approches semi-quadratiques et applications en restauration d'images.

8.1 INTRODUCTION

Dans le cadre de la résolution des problèmes inverses en restauration d'image, une classe importante de méthode s'est imposée dans la littérature des années 1990-2000, à savoir les algorithmes semi-quadratiques [43, 41, 68, 113, 150, 178]. Ces algorithmes d'optimisation remontent à des idées assez anciennes, qu'on retrouvait déjà, par exemple dans l'approche de Weiszfeld [197]. Ils sont remarquables par la double interprétation qu'on peut en faire. En effet, on peut aussi bien les voir comme des algorithmes MM quadratiques, issus des règles de majoration définies dans les Sections 5.2 ou 5.3, que comme des algorithmes de minimisation alternés, se rapprochant ainsi plus des méthodes par blocs du Chapitre 7. Dans ce chapitre, je me propose de rappeler le principe de ces algorithmes. Ensuite, je présenterai deux travaux réalisés dans le cadre de la thèse de YOSRA MARNISSI, dans lesquels la formulation semi-quadratique conduit à des algorithmes rapides pour la restauration d'images dans un cadre bayésien.

8.2 ALGORITHMES SEMI-QUADRATIQUES

8.2.1 Formulation du problème

Les méthodes semi-quadratiques permettent de minimiser une classe de fonctions d'optimisation de la forme :

$$(\forall x \in \mathbb{R}^N) \quad F(x) = \frac{1}{2} \|Hx - y\|^2 + \sum_{p=1}^P f([Vx]_p) \quad (8.1)$$

avec $H \in \mathbb{R}^{M \times N}$, $y \in \mathbb{R}^M$, $V \in \mathbb{R}^{N \times P}$ et $f : \mathbb{R} \rightarrow \mathbb{R}$ différentiable. Dans le cadre de l'estimation par MAP de la solution du problème inverse $y = Hx + \text{bruit}$, le deuxième terme de (8.1) vise à renforcer des propriétés de régularité connues *a priori* sur le minimiseur \hat{x} de F recherché. Comme dit précédemment, les méthodes semi-quadratiques peuvent s'interpréter comme des algorithmes MM quadratiques pour la minimisation de (8.1). Deux classes d'approches semi-quadratiques ont été proposées dans la littérature, parfois identifiées par les appellations "Geman-Yang" et "Geman-Reynolds", du nom des auteurs des articles de référence associés. Chacune des classes correspond à une technique distincte pour la construction de la fonction majorante quadratique du critère à minimiser.

8.2.2 Interprétation comme approche MM quadratique

Stratégie de Geman-Yang : La stratégie de "Geman-Yang" [100], se fonde sur l'hypothèse de l'existence d'un $\mu > 0$ tel que $x \rightarrow \frac{x^2}{2} - \frac{f(x)}{\mu}$ est convexe sur \mathbb{R} . Remarquons que cette hypothèse est vérifiée pour tout $\mu \geq \beta$ dès lors que f est β -Lipschitz différentiable. Par ailleurs, si f est elle-même convexe, alors on a équivalence entre l'hypothèse de Geman-Yang et f Lipschitz différentiable. Sous cette hypothèse, on peut montrer que pour tout $x' \in \mathbb{R}^N$, la fonction quadratique (8.2) est une fonction majorante de (8.1) en x' :

$$(\forall x \in \mathbb{R}^N) \quad Q^{\text{GY}}(x, x') = F(x') + (x - x')^\top \nabla F(x') + \frac{1}{2}(x - x')^\top (H^\top H + \mu V^\top V)(x - x'). \quad (8.2)$$

Stratégie de Geman-Reynolds : La stratégie de "Geman-Reynolds" [99] suppose que f vérifie la Propriété 5.3.1. Dans ce cas, on peut déduire que pour tout $x' \in \mathbb{R}^N$, la fonction quadratique (8.3) est une fonction majorante de (8.1) en x' :

$$(\forall x \in \mathbb{R}^N) \quad Q^{\text{GR}}(x, x') = F(x') + (x - x')^\top \nabla F(x') + \frac{1}{2}(x - x')^\top (H^\top H + V^\top \text{Diag}(\omega(Vx'))V)(x - x'), \quad (8.3)$$

où $\omega : \mathbb{R}^+ \rightarrow \mathbb{R}$ est définie dans l'énoncé de la Propriété 5.3.1.

La forme des fonctions majorantes ayant été définie par (8.2) ou (8.3), l'algorithme semi-quadratique s'écrit simplement comme un algorithme MM quadratique avec relaxation ([6]) :

$$\boxed{\begin{array}{l} \text{Pour } k = 0, 1, \dots \\ \left[\begin{array}{l} \bar{x}_k = \underset{x \in \mathbb{R}^N}{\text{argmin}} \quad Q(x, x_k) = x_k - A(x_k)^{-1} \nabla F(x_k) \\ x_{k+1} = \theta_k \bar{x}_k + (1 - \theta_k)x_k, \quad \theta_k \in (0, 2). \end{array} \right. \end{array}} \quad (8.4)$$

avec A définie par

$$A^{\text{GY}} = H^\top H + \mu V^\top V, \quad \mu \geq \beta, \quad (8.5)$$

ou

$$(\forall x \in \mathbb{R}^N) \quad A^{\text{GR}}(x) = H^\top H + V^\top \text{Diag}(\omega(|Vx|))V, \quad (8.6)$$

respectivement. Dans l'équation de mise à jour (8.4), le paramètre $(\theta_k)_{k \in \mathbb{N}}$ agit comme un pas de relaxation et pour $\theta_k \equiv 1$, l'Algorithme (8.4) correspond au schéma MM classique. L'Algorithme (8.4) peut être vu comme un cas particulier (et relaxé) de l'Algorithme 6.7 décrit dans la Section 6.3, et bénéficie donc de propriétés de convergence similaires.

8.2.3 Interprétation comme minimisation alternée d'un critère augmenté

Le nom "semi-quadratique" vient du fait que l'application de l'algorithme (8.4) à un critère de la forme (8.1), en utilisant la matrice de majoration (8.5) ou (8.6), donne lieu à l'algorithme de minimisation alterné suivant [113, 150] :

$$\boxed{\begin{array}{l} \text{Pour } k = 0, 1, \dots \\ \left[\begin{array}{l} b_k = \underset{b \in \mathbb{R}^P}{\text{argmin}} \Phi(x_k, b) \\ \tilde{x}_k = \underset{x \in \mathbb{R}^N}{\text{argmin}} \Phi(x, b_k) \\ x_{k+1} = \theta_k \tilde{x}_k + (1 - \theta_k)x_k, \quad \theta_k \in (0, 2). \end{array} \right. \end{array}} \quad (8.7)$$

La fonction augmentée $\Phi : \mathbb{R}^N \times \mathbb{R}^P \rightarrow]-\infty, +\infty]$ est telle que pour tout $b \in \mathbb{R}^P$, $x \mapsto \Phi(x, b)$ est une fonction *quadratique* liée à la majorante Q par :

$$(\forall x \in \mathbb{R}^N)(\forall x' \in \mathbb{R}^N) \quad Q(x, x') = \Phi(x, \hat{b}(x')), \quad (8.8)$$

avec

$$(\forall x \in \mathbb{R}^N) \quad \hat{b}(x) = \underset{b \in \mathbb{R}^P}{\text{argmin}} \Phi(x, b). \quad (8.9)$$

De plus,

$$(\forall x \in \mathbb{R}^N) \quad F(x) = \inf_{b \in \mathbb{R}^P} \Phi(x, b). \quad (8.10)$$

Je donne ci-après l'expression de Φ et de la règle de mise à jour de la variable b , pour les deux approches Geman-Yang et Geman-Reynolds. La mise à jour de la variable x est identique à celle décrite dans (8.4), du fait des relations (8.8) liant Φ et Q .

Stratégie de Geman-Yang : Si l'approche de Geman-Yang (8.5) est retenue, le critère augmenté s'écrit :

$$(\forall x \in \mathbb{R}^N)(\forall b \in \mathbb{R}^P) \quad \Phi^{\text{GY}}(x, b) = \frac{1}{2} \|Hx - y\|^2 + \mu \sum_{p=1}^P \left(\frac{1}{2} ([Vx]_p - b_p)^2 + \sum_{p=1}^P \zeta^{\text{GY}}(b_p) \right) \quad (8.11)$$

avec

$$(\forall u \in \mathbb{R}) \quad \zeta^{\text{GY}}(u) = \sup_{x \in \mathbb{R}} \left(-\frac{1}{2}(x-u)^2 + \mu^{-1}f(x) \right). \quad (8.12)$$

L'expression du critère augmenté (8.11) est obtenue en remarquant que (8.12) implique :

$$(\forall x \in \mathbb{R}) \quad f(x) = \inf_{u \in \mathbb{R}} \mu \left(\frac{1}{2}(x-u)^2 + \zeta^{\text{GY}}(u) \right). \quad (8.13)$$

De plus, pour tout $x \in \mathbb{R}$, l'infimum de la fonction ci-dessus est atteint, de façon unique, en :

$$(\forall x \in \mathbb{R}) \quad \hat{u}(x) = x - \mu^{-1}\dot{f}(x). \quad (8.14)$$

La preuve de ces résultats, que nous avons donnée dans [142, Prop.III.1], est basée sur le caractère convexe, propre et continue, de la fonction $g : x \rightarrow x^2/2 - \mu^{-1}f(x)$, et sur la propriété de dualité $g = g^{**}$. Il est intéressant de remarquer que $x \rightarrow \Phi^{\text{GY}}(x, b)$ engendre, lorsque b varie dans \mathbb{R}^P , une famille de fonctions quadratiques de courbure constante A^{GY} qui sont simplement des versions décalées de la fonction quadratique $Q^{\text{GY}}(\cdot, x')$, pour $x' \in \mathbb{R}^N$. L'égalité (8.8) est obtenue pour $\hat{b}(x')$ décrit par la règle de mise à jour déduite de (8.14) :

$$\hat{b}(x') = x' - \mu^{-1}\nabla F(x'). \quad (8.15)$$

Stratégie de Geman-Reynolds : Si l'approche de Geman-Reynolds (8.6) est retenue, alors

$$(\forall x \in \mathbb{R}^N)(\forall b \in \mathbb{R}^P) \quad \Phi^{\text{GR}}(x, b) = \frac{1}{2}\|Hx - y\|^2 + \sum_{p=1}^P \frac{b_p}{2}([Vx]_p)^2 + \zeta^{\text{GR}}(b_p) \quad (8.16)$$

avec

$$(\forall u \in \mathbb{R}) \quad \zeta^{\text{GR}}(u) = \sup_{x \in \mathbb{R}} \left(-\frac{x^2 u}{2} + f(x) \right). \quad (8.17)$$

L'expression du critère augmenté (8.16) est obtenue en remarquant que (8.17) implique :

$$(\forall x \in \mathbb{R}) \quad f(x) = \inf_{u \in \mathbb{R}} \left(\frac{x^2 u}{2} + \zeta^{\text{GR}}(u) \right). \quad (8.18)$$

De plus, pour tout $x \in \mathbb{R}$, l'infimum de la fonction ci-dessus est atteint, de façon unique, en :

$$(\forall x \in \mathbb{R}) \quad \hat{u}(x) = \omega(|x|), \quad (8.19)$$

où $\omega : \mathbb{R}^+ \rightarrow \mathbb{R}$ est définie dans l'énoncé de la Propriété 5.3.1. La preuve de ces résultats, donnée par exemple dans [150], est basée sur le caractère convexe, propre et continue de la fonction $g : x \rightarrow -\psi(\sqrt{x})$, et sur la propriété de dualité $g = g^{**}$. Lorsque b varie dans \mathbb{R}^P , $x \rightarrow \Phi^{\text{GR}}(x, b)$ définit une famille de fonctions quadratiques.

Parmi celles-ci, la seule qui soit tangente majorante de F en $x' \in \mathbb{R}^N$ est obtenue pour b égal à :

$$\widehat{b}(x') = \omega(|Vx'|). \quad (8.20)$$

La propriété remarquable (8.10) permet d'interpréter la formulation (8.7) comme la minimisation alternée d'une version *augmentée* du critère F . Cette vision de l'algorithme semi-quadratique est au cœur de l'analyse de convergence menée dans [113]. Elle a par ailleurs été à l'origine de deux de mes travaux récents que je présente ci-après, où la reformulation du problème par ajout de variables auxiliaires mène à des approches rapides pour la restauration d'images dans le cadre bayésien.

8.3 APPROCHE BAYES VARIATIONNELLE POUR LES MODÈLES NON GAUSSIENS

8.3.1 Position du problème

Considérons le problème inverse consistant à retrouver une estimée de l'image d'intérêt $\bar{x} \in \mathbb{R}^N$ à partir d'observations dégradées et bruitées $y \in \mathbb{R}^M$ de celle-ci, à travers le modèle linéaire :

$$(\forall m \in \{1, \dots, M\}) \quad y_m = \mathcal{D}([Hx]_m), \quad (8.21)$$

avec $H \in \mathbb{R}^{M \times N}$ est un opérateur linéaire modélisant typiquement un flou ou une projection, et $\mathcal{D}(\cdot)$ modélise le bruit, non nécessairement additif, et dépendant éventuellement des données. Dans ce contexte, la neg-log-vraisemblance Φ des données s'écrit :

$$(\forall x \in \mathbb{R}^N) \quad \Phi(x; y) = -\log \mathbf{P}(y | x) = \sum_{m=1}^M f_m([Hx]_m; y_m), \quad (8.22)$$

où, l'expression de f_m , pour $m \in \{1, \dots, M\}$, dépend du modèle de bruit considéré (voire Table 5.1). Afin d'obtenir une solution régularisée du problème inverse, une distribution *a priori* $\mathbf{P}(x | \gamma)$, paramétrée par une constante de régularisation $\gamma \in]0, +\infty[$, est associée à x . L'objectif est alors d'estimer conjointement $\Theta = (x, \gamma)$. Pour cela, une loi *a priori* $\mathbf{P}(\gamma)$ est fixée (typiquement, une loi Gamma de paramètres (α, β) choisis assez petits, de façon à avoir une loi peu informative). D'après la règle de Bayes, la loi *a posteriori* de Θ sachant les données observées y est

$$\mathbf{P}(\Theta | y) \propto \mathbf{P}(y | x)\mathbf{P}(x | \gamma)\mathbf{P}(\gamma). \quad (8.23)$$

De cette loi, on peut déduire une solution du problème inverse, au sens de l'estimateur de l'espérance *a posteriori* (*Minimum Mean Square Estimator*, MMSE). Néanmoins, le calcul de (8.23) ou plus précisément son exploration, est un défi majeur. Même si la loi *a priori* et celle des observations sont simples, la loi *a posteriori* est

généralement compliquée dans le sens où elle ne peut être connue qu'à une constante multiplicative près ou elle possède une forme non usuelle ou son traitement nécessite des ressources informatiques massives. Une stratégie efficace pour pallier ces difficultés est d'approcher, de façon itérative, la loi (8.23), et d'en déduire le MMSE. A cet égard, plusieurs méthodes d'approximation ont été proposées. Nous nous intéressons dans cette section aux méthodes d'approximation basées sur le principe Bayes variationnel.

8.3.2 Approximation Bayes variationnelle

Les approches par approximation Bayes variationnelle (*Variational Bayes Approximation*, VBA) visent à approcher la loi *a posteriori* par une autre loi plus simple, notée $q(\Theta)$ à partir de laquelle la moyenne *a posteriori* peut être facilement calculée. Ces méthodes peuvent généralement conduire à une complexité de calcul relativement faible par rapport aux algorithmes basés sur l'échantillonnage, dans lesquels on cherche à simuler par rapport à (8.23). La loi approchante est choisie aussi proche que possible de la vraie loi au sens de la divergence de Kullback-Leibler :

$$q^{\text{opt}} = \operatorname{argmin}_q \mathcal{KL}(q(\Theta) \| \mathbb{P}(\Theta | y)), \quad (8.24)$$

où

$$\mathcal{KL}(q(\Theta) \| \mathbb{P}(\Theta | y)) = \int q(\Theta) \log \frac{q(\Theta)}{\mathbb{P}(\Theta | y)} d\Theta. \quad (8.25)$$

Une solution de (8.24) peut être déterminée si l'on fait une hypothèse de séparabilité sur $q(\Theta)$. Supposons que $q(\Theta) = \prod_{r=1}^R q(\Theta_r)$. Cette forme séparable permet de calculer la loi approchante optimale $q^{\text{opt}}(\Theta_r)$ pour chaque variable Θ_r , en résolvant (8.24) par rapport à cette variable, les autres variables restant fixées. On obtient alors :

$$(\forall r \in \{1, \dots, R\}) \quad q_r^{\text{opt}}(\Theta_r) \propto \exp \left(\langle \log \mathbb{P}(y, \Theta) \rangle_{\prod_{i \neq r} q_i^{\text{opt}}(\Theta_i)} \right) \quad (8.26)$$

avec $\langle \cdot \rangle_{\prod_{i \neq r} q(\Theta_i)} = \int \cdot \prod_{i \neq r} q(\Theta_i) d\Theta_i$. Néanmoins, nous pouvons voir que chaque composante séparable q_r^{opt} définie par (8.26) dépend de toutes les autres composantes q_i^{opt} avec i différent de r . Par conséquent, on ne peut pas obtenir une expression explicite pour $q^{\text{opt}}(\Theta)$ sauf dans des cas très simples. En pratique, ce problème est généralement surmonté en utilisant un algorithme de minimisation alternée, qui met à jour une composante séparable en fixant les autres à chaque étape, afin d'approcher la solution optimale de façon itérative [180].

Dans notre contexte, une approximation séparable naturelle de $q(\Theta)$ est :

$$q(\Theta) = q_X(x)q_\Gamma(\gamma), \quad (8.27)$$

et le schéma VBA classique consiste donc à alterner deux étapes de minimisation, dans le but d'approcher au mieux la loi $\mathbf{P}(\Theta \mid y)$:

1. Minimiser $\mathcal{KL}(q_\Theta(\Theta) \parallel \mathbf{P}(\Theta \mid y))$ par rapport à $q_X(x)$;
2. Minimiser $\mathcal{KL}(q_\Theta(\Theta) \parallel \mathbf{P}(\Theta \mid y))$ par rapport à $q_\Gamma(\gamma)$.

Des exemples d'applications d'approches VBA classiques peuvent être trouvés dans [86, 14, 45, 189] tandis que des versions accélérées ont été proposées dans [97, 203]. Notons que les estimateurs au sens MMSE, des grandeurs d'intérêt x et γ se déduisent des moyennes des lois approchantes $q_X(x)$ et $q_\Gamma(\gamma)$ obtenues à la convergence des itérations VBA.

8.3.3 Approche VBA par majoration

Dans le cas d'un modèle de bruit non gaussien, il est néanmoins difficile d'obtenir une solution simple à la première étape du schéma VBA (i.e., le calcul de $q_X(x)$), du fait de l'imbrication, dans $\mathbf{P}(\Theta \mid y)$ de la loi *a priori* sur x et de la vraisemblance non gaussienne de x sachant les données, sans faire une hypothèse très contraignante de conjugaison de ces lois. Une solution initialement proposée dans [177] consiste à remplacer les termes d'attache aux données $\Phi(x; y) = -\log \mathbf{P}(y \mid x)$ et de régularisation $\Psi(x; \gamma) = -\log \mathbf{P}(x \mid \gamma)$ par des approximations majorantes quadratiques de ceux-ci. Nous avons donc proposé dans [142] d'employer les techniques semi-quadratiques décrites dans la Section 8.2, pour construire ces approximations majorantes quadratiques. Ces majorantes conduisent à l'obtention d'une borne supérieure de la divergence de Kullback-Leibler (8.24). Les lois approchantes q_X et q_Γ sont alors obtenues en minimisant cette borne, par un schéma alterné.

Attache aux données : Supposons que Φ prend la forme (8.22), où, pour tout $m \in \{1, \dots, M\}$, $f_m(\cdot; y_m)$ vérifie l'hypothèse de Geman-Yang pour un paramètre $\mu_m > 0$. Alors, la stratégie de Geman-Yang permet de déduire :

$$(\forall x \in \mathbb{R}^N) \quad \Phi(x; y) = \inf_{w \in \mathbb{R}^M} T(x, w; y). \quad (8.28)$$

La fonction augmentée T est telle que, pour tout $w \in \mathbb{R}^M$, $x \mapsto T(x, w; y)$ est une fonction quadratique. Plus précisément,

$$(\forall x \in \mathbb{R}^N)(\forall w \in \mathbb{R}^M) \quad T(x, w; y) = \sum_{m=1}^M T_m([Hx]_m, w_m; y_m), \quad (8.29)$$

avec, pour tout $m \in \{1, \dots, M\}$

$$(\forall v \in \mathbb{R}^M)(\forall u \in \mathbb{R}) \quad T_m(v, u; y_m) = \mu_m(y_m) \left(\frac{1}{2}(v - u)^2 + \zeta_m(v; y_m) \right), \quad (8.30)$$

où

$$(\forall v \in \mathbb{R}) \quad \zeta_m(v; y_m) = \sup_{u \in \mathbb{R}} \left(-\frac{1}{2}(v - u)^2 + \frac{f_m(u; y_m)}{\mu_m(y_m)} \right). \quad (8.31)$$

Pénalisation : Considérons une loi *a priori* sur x de la forme générale :

$$\mathbf{P}(x \mid \gamma) = \tau \gamma^{\frac{N}{2\kappa}} \exp \left(-\gamma \sum_{s=1}^S \|V_s x\|^{2\kappa} \right) \quad (8.32)$$

avec $\kappa \in]0, 1]$, $(V_s)_{1 \leq s \leq S} \in (\mathbb{R}^{P \times N})^S$, tel que $V = [V_1^\top \dots V_S^\top]^\top$ est un opérateur linéaire. Par exemple, V peut calculer les différences horizontales et verticales entre les pixels voisins de l'image, et dans ce cas, $S = N$ et $P = 2$. De plus, $\tau \in [0, +\infty[$, et $\gamma \in]0, +\infty[$ est la constante de régularisation. Alors, la version généralisée [68, 41] de la stratégie de Geman-Reynolds nous permet de déduire que, pour tout $s \in \{1, \dots, S\}$,

$$(\forall x \in \mathbb{R}^N) \quad \|V_s x\|^{2\kappa} = \inf_{b \in \mathbb{R}} \left(\frac{b}{2} \|V_s x\|^2 + \xi(b) \right), \quad (8.33)$$

avec

$$(\forall b \in \mathbb{R}) \quad \xi(b) = \sup_{u \in \mathbb{R}} \left(-\frac{u^2 b}{2} + u^{2\kappa} \right). \quad (8.34)$$

La résolution de (8.34) et la substitution $b \leftrightarrow 2\kappa\lambda^{\kappa-1}$ conduit à :

$$(\forall x \in \mathbb{R}^N) \quad \Psi(x; \gamma) = \gamma \sum_{s=1}^S \|V_s x\|^{2\kappa} = \inf_{\lambda \in \mathbb{R}^S} R(x, \lambda; \gamma), \quad (8.35)$$

avec

$$(\forall x \in \mathbb{R}^N)(\forall \lambda \in \mathbb{R}^S) \quad R(x, \lambda; \gamma) = \sum_{s=1}^S \gamma \frac{\kappa \|V_s x\|^2 + (1 - \kappa)\lambda_s}{\lambda_s^{1-\kappa}}. \quad (8.36)$$

Algorithme proposé Des deux approximations majorantes précédentes, nous pouvons déduire :

$$\mathbf{P}(\Theta \mid y) \geq L(\Theta \mid y; w, \lambda), \quad (8.37)$$

où la fonction L est donnée par

$$L(\Theta \mid y; w, \lambda) = C(y) \exp [-T(x, w; y) - R(x, \lambda; \gamma)] \mathbf{P}(\gamma),$$

avec la constante de normalisation $C(y) = \mathbf{P}(y)^{-1} (2\pi)^{-M/2} \tau \gamma^{\frac{N}{2\kappa}}$. Nous en déduisons une majoration de la divergence de Kullback-Leibler :

$$\mathcal{KL}(q(\Theta) \parallel \mathbf{P}(\Theta \mid y)) \leq \mathcal{KL}(q(\Theta) \parallel L(\Theta \mid y; w, \lambda)). \quad (8.38)$$

En minimisant le terme de droite de l'inégalité (8.38) par rapport à w et λ , on rend cette majoration la plus précise possible. Notons que, pour tout $m \in \{1, \dots, M\}$ and $s \in \{1, \dots, S\}$, $\lambda_s \mapsto \mathcal{KL}(q(\Theta) \parallel L(\Theta \mid y; w, \lambda))$ et $w_m \mapsto \mathcal{KL}(q(\Theta) \parallel L(\Theta \mid y; w, \lambda))$ peuvent être minimisées de façon séparée. Le problème (8.24) peut donc maintenant être résolu de façon alternée, par le schéma suivant :

1. Minimiser la majorante $\mathcal{KL}(q(\Theta)||L(\Theta | y; w, \lambda))$ par rapport à $q_X(x)$;
2. Pour tout $m \in \{1, \dots, M\}$, mettre à jour les variables auxiliaires w_m de façon à minimiser $\mathcal{KL}(q(\Theta)||L(\Theta | y; w, \lambda))$;
3. Pour tout $s \in \{1, \dots, S\}$, mettre à jour les variables auxiliaires λ_s de façon à minimiser $\mathcal{KL}(q(\Theta)||L(\Theta | y; w, \lambda))$;
4. Minimiser la majorante $\mathcal{KL}(q(\Theta)||L(\Theta | y; w, \lambda))$ par rapport à $q_\Gamma(\gamma)$.

Supposons une distribution *a priori* Gamma sur le paramètre de régularisation γ . L'expression de l'approximation majorante $L(\Theta|y; w, \lambda)$ permet alors de déduire (voir [142] pour le détail des calculs) des expressions simples pour les lois de x et γ , de la forme :

$$q_X(x) \sim \mathcal{N}(\nu, \Sigma), \quad q_\Gamma(\gamma) \sim \mathcal{G}(a, b). \quad (8.39)$$

L'algorithme alterné va construire par un schéma itératif alterné des séquences $(\nu_k)_{k \in \mathbb{N}}$, $(\Sigma_k)_{k \in \mathbb{N}}$, $(a_k)_{k \in \mathbb{N}}$, et $(b_k)_{k \in \mathbb{N}}$ dans le but de converger vers les valeurs des paramètres ν , Σ , a , et b de ces deux lois. Il sera ensuite direct d'obtenir l'estimateur MMSE de x en regardant la valeur limite de $(\nu_k)_{k \in \mathbb{N}}$, et celui de γ par la limite de $(a_k/b_k)_{k \in \mathbb{N}}$.

8.3.4 Résultats expérimentaux

Illustrons l'efficacité de l'approche proposée sur un exemple extrait de notre article [142], portant sur la restauration d'une image dégradée par un flou, et un bruit mixte Poisson-Gaussien. L'image originale \bar{x} (Fig. 8.1(a)), d'intensité comprise entre 0 et $x^+ = 20$, est une image de microscopie confocale de taille 350×350 , issue du site <http://meyerinst.com/confocals/tcs-spe/index.htm>. L'observation y (Fig. 8.1(b)) est obtenue par le modèle (8.21), avec H modélisant un flou uniforme de taille 5×5 , et \mathcal{D} un bruit Poisson-Gaussien (Ex. 5.2.1), de paramètre $\sigma^2 = 9$. Nous présentons les résultats obtenus par l'approche VBA proposée. Le terme d'attache aux données Φ est construit à partir de la distance de Kullback décalée (Tab. 5.1), qui réalise le meilleur compromis entre complexité de calcul et approximation de la loi exacte Poisson-Gauss (5.4). Concernant le choix de la régularisation Ψ , la flexibilité de la technique VBA nous a permis de comparer plusieurs critères, nous présentons les résultats obtenus pour une régularisation de type variation totale (*total variation*, TV) (Fig. 8.1(c)) ou variation totale non locale (*non-local total variation*, NLTV) (Fig. 8.1(d)). Nous comparons enfin ces résultats à ceux obtenus par l'approche de type MAP, décrite dans [51]. Dans les approches de type MAP, le choix du paramètre de régularisation γ est critique, et a une influence importante sur la qualité de l'image restaurée. Nous présentons les images restaurées, lorsque γ est choisi en suivant le principe de "discrepancy" décrit dans [15] (Fig. 8.1(e)), ou bien en suivant une technique empirique consistant à minimiser l'erreur quadratique entre l'image restaurée et l'image originale (supposée connue) (Fig. 8.1(f)).

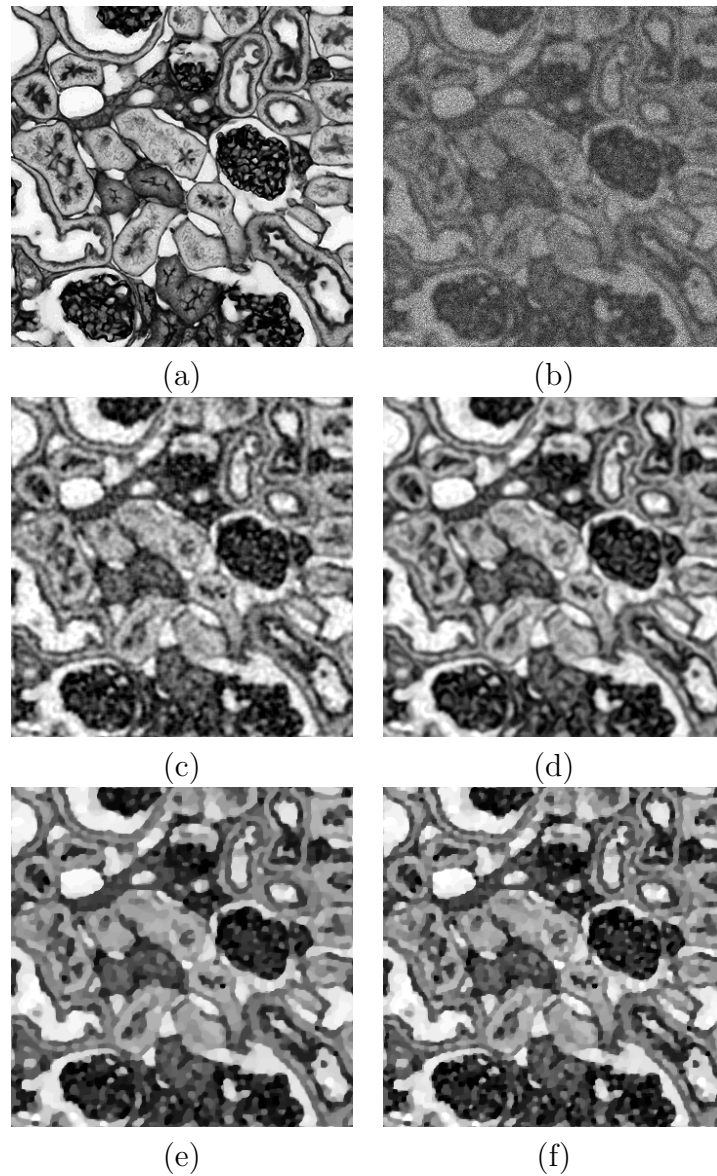


Figure 8.1 – (a) Image originale \bar{x} ; (b) Image observée y , $SNR=7.64$ dB; (c) Résultat de l’approche VBA proposée, régularisation TV, $SNR=13.90$ dB, 34 sec.; (d) Résultat de l’approche VBA proposée, régul. NLTV, $SNR=14.09$ dB, 8595 sec.; (e) Résultat de l’approche MAP [51], régul. TV, γ fixé par le principe de “discrepancy”, $SNR=13.60$ dB, 769 sec.; (f) Résultat de [51], régul. TV, γ optimisé par règle empirique, $SNR=13.71$ dB, 8477 sec.

Nous pouvons remarquer que, pour un même choix de régularisation, la méthode proposée, basée sur le MMSE, conduit à des résultats de qualité supérieure au MAP en terme de SNR. La régularisation NLTV permet d’encore améliorer la qualité de ces résultats. En ce qui concerne les temps de calculs, il est intéressant de souligner

que l’approche VBA avec régularisation TV est très rapide sur cet exemple. Ceci s’explique par la proposition d’une technique d’accélération par approximation diagonale (voir [142, Sec.IV.F]) qui permet de réduire considérablement la complexité de l’étape de mise à jour de la variable Σ de la loi approchante $q_X(x)$. A l’opposée, l’approche de type MAP est lente, notamment parce que le temps de calcul inclut l’optimisation du paramètre γ (en pratique, une quinzaine de valeurs doivent être testées). Enfin, soulignons que bien que la régularisation NLTV entraîne une amélioration significative de l’image restaurée, ceci est aux dépends de la complexité de calcul, et un compromis est donc à réaliser entre qualité d’image et temps passé à la restauration.

8.4 APPROCHES MCMC ACCÉLÉRÉES POUR LA SIMULATION EN GRANDE DIMENSION

8.4.1 Motivations

Une autre stratégie pour déterminer un estimateur de \bar{x} à partir de y au sens du MMSE dans le problème inverse (8.21) est d’utiliser un algorithme d’échantillonnage de type Monte Carlo par chaînes de Markov (MCMC), pour simuler la loi *a posteriori* $P(x | y)$. Les algorithmes MCMC sont en effet des outils bayésiens puissants pour explorer des lois compliquées. Dans le but d’échantillonner suivant $P(x | y)$, l’idée sous-jacente aux algorithmes MCMC est de produire, à partir d’une certaine loi de transition donnée, une chaîne de Markov qui converge vers la loi cible. Une classe importante de méthodes MCMC est inspirée de l’algorithme de Metropolis-Hastings (MH) où la loi de transition est définie par une loi de proposition $G(\tilde{x}_k | x_k)$ (dépendant de l’échantillon courant x_k) dont l’échantillonnage est simple et une règle d’acceptation-rejet assurant la réversibilité de la chaîne :

<pre> Initialiser $x_0 \in \mathbb{R}^N$ Pour $k = 0, 1, \dots$ Générer \tilde{x}_k selon la loi de proposition de densité $G(\cdot, x_k)$ Acceptation-Rejet : Générer $u \sim \mathcal{U}(0, 1)$ Calculer $\alpha(x_k, \tilde{x}_k) = \min\left(1, \frac{P(\tilde{x}_k y)G(x_k \tilde{x}_k)}{P(x_k y)G(\tilde{x}_k x_k)}\right)$ Si $u < \alpha(x_k, \tilde{x}_k)$ alors Accepter : $x_{k+1} = \tilde{x}_k$ Sinon Rejeter : $x_{k+1} = x_k$ </pre>	(8.40)
---	--------

Une tâche difficile lors de la mise en œuvre de cette méthode est le choix de la densité de proposition. Elle doit idéalement fournir une bonne approximation locale de la

densité cible avec un faible coût de calcul. L'algorithme de MH à marche aléatoire est associé au cas particulier d'une loi de proposition normale centrée sur l'état courant. Cet algorithme a une faible complexité, mais sa convergence peut devenir très lente lorsque la dimension du problème augmente [193]. Une alternative consiste à utiliser une approche de type MALA (*Metropolis-Adjusted Langevin Algorithm*) [172]. Dans le cadre de la thèse de YOSRA MARNISSI, nous avons proposé une stratégie d'accélération pour MALA basée sur une technique de préconditionnement MM, que je présente dans la Section 8.4.2 .

8.4.2 Algorithme de MALA préconditionné

Inspirées des outils d'optimisation par descente de gradient, les méthodes de type MALA [173, 172, 182] exploitent des informations sur la géométrie de la loi cible pour construire la densité de proposition dans le but de guider la chaîne vers l'espace cible où la plupart des échantillons doivent être concentrés. Supposons $x \rightarrow F(x) = -\log \mathbb{P}(x | y)$ différentiable sur \mathbb{R}^N . Alors, la composante directionnelle de la loi de proposition est choisie comme une itération d'un algorithme de descente de gradient préconditionné :

$$(\forall k \in \mathbb{N}) \quad \tilde{x}_k \sim \mathcal{N} \left(x_k - \frac{\epsilon^2}{2} A(x_k)^{-1} \nabla F(x_k), \epsilon^2 A(x_k)^{-1} \right), \quad (8.41)$$

où $\epsilon > 0$ est un pas déterminant l'avancée le long de la direction proposée, et $(A(x_k))_{k \in \mathbb{N}}$ sont des matrices symétriques définies positives $\mathbb{R}^{N \times N}$ jouant le rôle de préconditionneurs, l'algorithme MALA original de [173] correspondant à $A(x_k) \equiv I_N$. Afin d'obtenir de bonnes performances pratique, il est crucial de choisir des matrices $(A(x_k))_{k \in \mathbb{N}}$ réalisant un bon compromis entre faible coût de calcul par itération et préservation des propriétés de mélange de la chaîne. Plusieurs techniques ont été proposées dans la littérature [182, 143, 9, 201, 102, 193]. Nous avons introduit dans [139], l'approche 3MH (*Majorize-Minimize Metropolis Hastings*) qui consiste à choisir les matrices de préconditionnement par une stratégie MM, i.e., pour tout $x' \in \mathbb{R}^N$, $A(x')$ est construite de telle façon à assurer que la fonction quadratique :

$$(\forall x \in \mathbb{R}^N) \quad Q(x, x') = F(x) + (x - x')^\top \nabla F(x') + \frac{1}{2} (x - x')^\top A(x') (x - x') \quad (8.42)$$

soit une fonction majorante de F en x' . Nous avons établi dans le cadre de la thèse de YOSRA MARNISSI l'ergodicité géométrique de l'algorithme d'échantillonnage associé à cette nouvelle méthode, en nous basant sur les résultats théoriques de [9] concernant les algorithmes de MH classiques. De plus, les bonnes performances numériques de l'approche 3MH ont été illustrées sur des exemples de restauration de signaux parcimonieux [140], et d'images multicomposantes [139].

8.4.3 Stratégie d'augmentation de données

Comme nous l'avons déjà dit, une tâche délicate lors de la mise en œuvre des méthodes de type MALA préconditionné est la sélection de matrices de préconditionnement qui utilisent des informations sur la géométrie locale de la densité cible afin d'accélérer la convergence et améliorer les propriétés de mélange dans l'espace des paramètres, sans être trop coûteuses en termes de complexité de calcul. Ces deux problèmes sont accrus par la présence de deux sources de dépendances hétérogènes provenant soit du terme d'attache aux données $P(y | x)$, soit de la loi *a priori* $P(x)$, dans le sens où les matrices de préconditionnement associées peuvent rarement être diagonalisées dans le même domaine. Pour pallier ces difficultés, une stratégie d'augmentation de données (*Data Augmentation*, DA) [195] peut être adoptée afin de dissocier les deux sources de dépendances du modèle. Dans une telle approche, une variable auxiliaire $u \in \mathbb{R}^J$, $J \geq 1$ est ajoutée à l'ensemble des paramètres à estimer. Cette variable est caractérisée par la loi conditionnelle $P(u | x, y) = P(u | x)$ qui doit assurer que la densité jointe $P(x, u | y)$ satisfasse :

$$\int_{\mathbb{R}^J} P(x, u | y) du = P(x | y) \quad (8.43)$$

$$\int_{\mathbb{R}^N} P(x, u | y) dx = P(u | y) \quad (8.44)$$

où $P(u | y)$ définit une densité de probabilité valide (à valeurs positives, et d'intégrale égale à un). Au lieu de simuler directement selon $P(x | y)$, on va simuler de façon alternée suivant les lois conditionnelles des variables x et u , de densités respectives $P(x | u, y)$ et $P(u | x, y)$, ce qui permet de converger, sous certaines conditions techniques [101] vers une chaîne de distribution stationnaire $P(x, u | y)$. L'efficacité des stratégies de type DA est liée principalement à un choix approprié de $P(u | x, y)$ qui permette de simuler plus simplement suivant les deux lois conditionnelles impliquées que suivant la loi initiale $P(x | y)$.

Plusieurs algorithmes d'échantillonnage basés sur une technique de DA ont été proposés dans la littérature [78, 112, 87, 146, 102, 141]. Je présente ici la stratégie proposée dans notre article [141], qui repose sur la combinaison entre une formulation de type DA et l'approche 3MH présentée dans la section précédente. Considérons un problème inverse de la forme (8.21) où $\mathcal{D}(\cdot)$ modélise un bruit additif Gaussien, de moyenne nulle et de variance σ^2 , pour lequel nous souhaitons calculer l'estimateur au sens du MMSE, par une approche MCMC de type 3MH. La difficulté réside dans le choix des matrices de préconditionnement $(A(x_k))_{k \in \mathbb{N}}$. En effet, un choix naturel consiste à les choisir de la forme $H^\top H + V^\top \Omega V$ avec V la matrice de transformation apparaissant dans le terme de régularisation (typiquement, une transformée en ondelettes), et Ω une matrice diagonale liée à la forme de la loi *a priori*. Néanmoins, une telle matrice de préconditionnement est rarement inversible de façon simple, car H et V ne vont pas être diagonalisables dans le même domaine. Pour pallier ce problème, nous avons proposé de définir la variable auxiliaire $u \in \mathbb{R}^N$ de façon

à séparer les contributions des opérateurs H et V . Tout d'abord, la forme Gaussienne du terme d'attaches aux données nous incite à fixer une loi conditionnelle de u sachant x de la forme :

$$\mathbf{P}(u \mid x) = \mathcal{N}\left((\mathbf{I}_N + C)^{-1}x, \sigma^2\mu^{-1}(\mathbf{I}_N + C)^{-1}\right), \quad (8.45)$$

où $C = B(\mathbf{I}_N - B)^{-1}$, $B = \mu^{-1}H^\top H$ et $\mu > 0$ est telle que $\|H\|^2 < \mu$. Alors, pour tout $x \in \mathbb{R}^N$ et $u \in \mathbb{R}^N$, $\Phi(x, u) = -\log \mathbf{P}(x, u \mid y)$ s'écrit sous la forme d'un critère augmenté :

$$(\forall x \in \mathbb{R}^N)(\forall u \in \mathbb{R}^N) \quad \Phi(x, u) = \frac{\mu}{2\sigma^2} \left(\|u - x\|^2 + u^\top C^\top u - 2\mu^{-1} y^\top Hx \right) - \log \mathbf{P}(x), \quad (8.46)$$

tel que l'infimum de (8.46) par rapport à la variable u s'exprime :

$$-\log \mathbf{P}(x \mid y) = \frac{1}{2\sigma^2} \|Hx - y\|^2 - \log \mathbf{P}(x). \quad (8.47)$$

Cette propriété initialement établie dans [19] dans le contexte déterministe s'obtient en appliquant le résultat de Geman-Yang au terme de moindres carrés dans (8.47). Remarquons néanmoins que pour que la matrice C soit bien définie, il est nécessaire ici de choisir la constante μ strictement supérieure à la constante de Lipschitz $\|H\|^2$. D'après (8.46), on peut déduire (à une constante additive près) l'expression de $\Theta(x; u) = -\log \mathbf{P}(x \mid u, y)$ pour tout $x \in \mathbb{R}^N$ et $u \in \mathbb{R}^N$:

$$\Theta(x; u) = \frac{\mu}{2\sigma^2} \left(\|u - x\|^2 - 2\mu^{-1} y^\top Hx \right) - \log \mathbf{P}(x). \quad (8.48)$$

L'algorithme 3MH avec variable auxiliaire que nous proposons consiste à alterner une étape 3MH sur la loi conditionnelle $\mathbf{P}(x \mid u, y)$ et une étape d'échantillonnage de la variable auxiliaire u suivant (8.45) :

<p>Pour $k = 0, 1, \dots$</p> <p> Générer</p> <p> $\tilde{x}_k \sim \mathcal{N}\left(x_k - \frac{\varepsilon^2}{2}A(x_k)^{-1}\nabla\Theta(x_k; u_k), \varepsilon^2 A(x_k)^{-1}\right)$</p> <p> Accepter $(x_{k+1} = \tilde{x}_k)$ avec probabilité</p> <p> $\alpha(x_k, \tilde{x}_k) = \min\left(1, \frac{\mathbf{P}(\tilde{x}_k \mid y)\mathbf{G}(\tilde{x}_k, x_k)}{\mathbf{P}(x_k \mid y)\mathbf{G}(x_k, \tilde{x}_k)}\right)$</p> <p> Générer</p> <p> $u_{k+1} \sim \mathcal{N}\left((\mathbf{I}_N - \mu^{-1}H^\top H)x_{k+1}, \sigma^2\mu(\mathbf{I}_N - \mu^{-1}H^\top H)\right).$</p>	(8.49)
---	--------

Il est intéressant de noter que, si H est diagonalisable dans un domaine transformé (par exemple Fourier si H modélise un flou), alors l'étape d'échantillonnage suivant (8.45) peut être réalisée dans ce domaine. De plus, il est maintenant plus simple de déterminer des matrices de préconditionnement efficaces $(A(x_k))_{k \in \mathbb{N}}$ car l'opérateur linéaire H n'apparaît plus explicitement devant x dans la nouvelle loi de proposition (8.48) et il suffira de choisir une métrique adaptée au terme *a priori*.

8.4.4 Application à la restauration d'images multicomposantes

Je termine cette section par l'illustration des performances de l'approche 3MH, et de sa variante avec DA, sur un exemple de restauration d'images multicomposantes issus de notre article [141]. Le cube multispectral Hydice de taille $256 \times 256 \times 6$, extrait de <https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html>, est dégradé par un flou appliqué sur chacune de ses 6 composantes, et un bruit additif gaussien est ajouté, de telle sorte que le SNR de l'image dégradée est 12.47dB. Nous cherchons à estimer, au sens du MMSE, les coefficients en ondelettes des composantes du cube original, en supposant que ceux-ci suivent la loi GMEP (*Generalized Multivariate Exponential Power Prior*) que nous avons introduit dans un travail antérieur [138]. Nous comparons pour cela trois approches de type MH, à savoir la marche aléatoire (*Random Walk, RW*), MALA et 3MH, pour résoudre la formulation initiale du problème ou sa formulation après étape de DA (suivant un algorithme de la forme 8.49). Trois composantes du cube initial, dégradé et restauré (SNR = 16.68dB), sont illustrées en Figure 8.2. La vitesse des algorithmes d'échantillonnage est illustrée sur la Figure 8.3. Nous pouvons voir que l'approche 3MH est plus rapide que RW et MALA. De plus, l'utilisation de la stratégie d'augmentation de données entraîne une amélioration significative des performances des trois algorithmes d'échantillonnage considérés. La comparaison de l'indice MSJ (*Mean Square Jump in stationarity*) défini dans [9], pour les différents algorithmes, montre que, sur cet exemple, 3MH est supérieur à MALA, lui-même supérieur à RW en terme de qualité de mélange.

8.5 CONCLUSION

Dans ce chapitre, j'ai présenté deux nouvelles approches bayésiennes pour la restauration d'images en grande dimension. Les deux algorithmes proposés sont basés sur des schémas connus de la littérature bayésienne, à savoir la stratégie d'approximation Bayes variationnelle, ou le schéma Métropolis-Hastings avec diffusion de Langevin, respectivement. Ma contribution a été d'en améliorer l'efficacité pratique en y incorporant des techniques semi-quadratiques issues de l'optimisation déterministe. Ces dernières, qui constituent un sous-ensemble des approches MM, permettent une reformulation plus simple du problème initial, par l'ajout de variables auxiliaires. Les algorithmes obtenus permettent d'obtenir la solution au sens du MMSE du problème inverse à résoudre, en un temps comparable à celui d'une approche variationnelle basée sur le MAP, tout en présentant l'avantage de gérer de façon conjointe l'estimation des paramètres de régularisation.

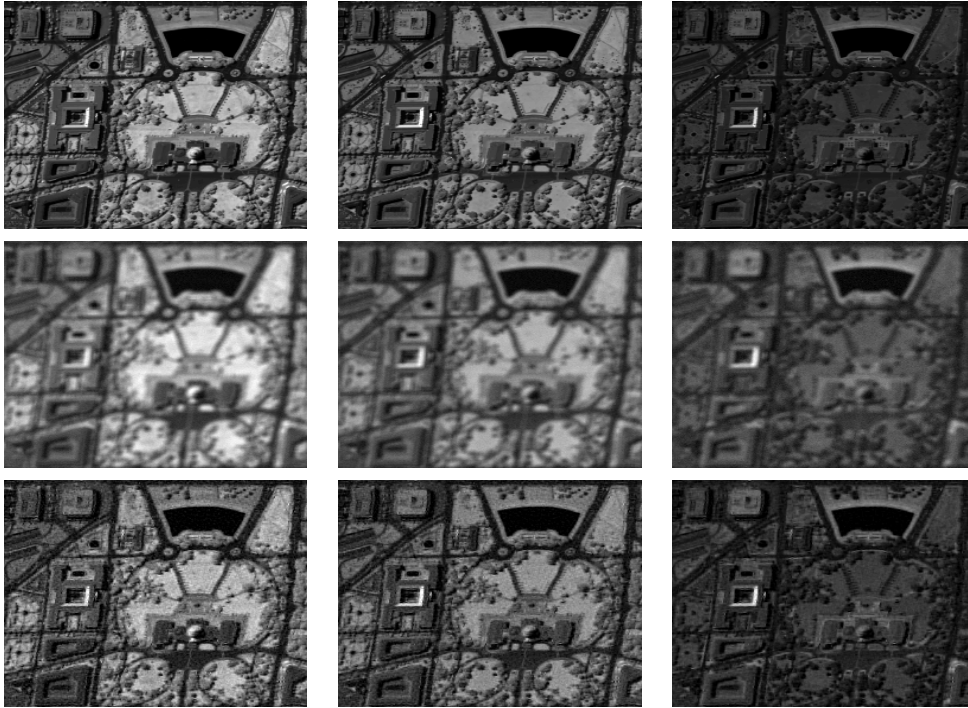


Figure 8.2 – De haut en bas : Images originales, images dégradées et images restaurées, pour les canaux n° 2, 4 et 6 du cube multispectral *Hydice*.

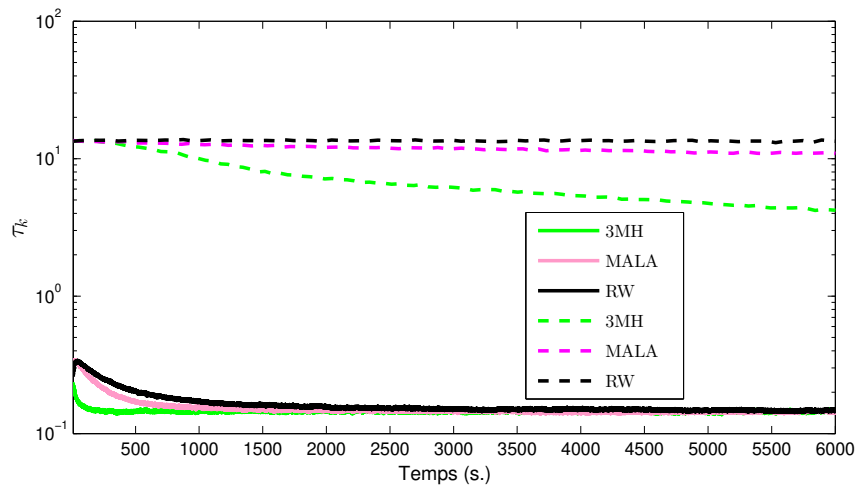


Figure 8.3 – Evolution du paramètre d'échelle de la 1ère sous-bande d'ondelettes au cours du temps (après étape de chauffage) pour les algorithmes d'échantillonnage RW, MALA et 3MH, en utilisant la formulation originale (traits discontinus) ou la formulation avec variable auxiliaire (traits continus).

- Chapitre 9 -

Conclusion et perspectives

J'ai présenté dans cette deuxième partie du manuscrit les contributions méthodologiques et applicatives que j'ai pu apporter depuis le début de ma carrière de chercheuse. Mes travaux se situent à la frontière entre le traitement du signal et des images, et les mathématiques appliquées. Ils visent à proposer des solutions algorithmiques efficaces pour l'optimisation des fonctions non nécessairement convexes ni différentiables, intervenant dans la résolution des problèmes inverses de grande taille en signal/image. Je m'intéresse particulièrement à la classe des algorithmes d'optimisation basés sur le principe de majoration-minimisation, qui se distinguent selon moi par leur simplicité d'interprétation, leur grande flexibilité, leur fiabilité théorique et leur efficacité pratique. Les méthodes que j'ai développées répondent à des besoins applicatifs réels dans des domaines variés tels que la microscopie biologique, l'imagerie médicale, l'imagerie des matériaux, l'analyse de données sismiques, ou le traitement d'archives vidéo. J'ai proposé des stratégies de résolution performantes et ayant une complexité de calcul limitée, malgré la grande dimensionalité des données mises en jeu dans ces domaines. De plus, je me suis attachée à établir la convergence théorique des algorithmes proposés en utilisant des outils récents d'analyse non lisse.

La recherche d'outils efficaces pour la résolution des problèmes inverses en signal/image n'est pas une problématique récente et de grandes avancées ont été réalisées depuis les travaux pionniers de Tikhonov. Le développement lors de ces quinze dernières années de l'optimisation non lisse, en conjonction avec la proposition de nouvelles stratégies de régularisation efficaces a conduit à des stratégies d'inversion rapides menant à des solutions de bonne qualité. Les travaux récents, dans le domaine des mathématiques appliquées, ont de plus permis d'apporter des solutions théoriques à l'analyse de convergence dans le cadre de l'optimisation non convexe basées sur la propriété de KL. Ces résultats ouvrent la voie à la résolution de formulations inverses plus complexes issues par exemple de problèmes de restau-

ration aveugle, ou de l'usage de pénalisations non convexes. De nombreux verrous restent néanmoins à lever, qui feront l'objet de mes recherches futures.

Modèles non linéaires et non-convexité : Une difficulté majeure se pose lorsque le modèle d'observation intervenant dans le problème inverse à résoudre est non linéaire. Cette non linéarité peut provenir par exemple d'une perte de la phase des données, donnant lieu au problème dit de reconstruction de phase fréquent en imagerie [92]. Elle peut aussi provenir d'un modèle linéaire associé à une méconnaissance partielle ou totale de la matrice d'observation dépendant de paramètres inconnus, qu'il faut estimer conjointement au signal d'intérêt [133]. Dans ces applications, la formulation d'un critère pénalisé même simpliste (par exemple, quadratique) donne lieu à un problème d'optimisation non convexe difficile à résoudre, pour lequel les approches de résolution standards sont souvent peu efficaces. Nous avons montré dans [67, 163] que les approches par majoration-minimisation pouvaient apporter une réponse à la non linéarité intervenant dans la fonction de coût en reconstruction de phase. Dans le cadre de mon projet ANR JCJC, je me poserai la question de l'extension de cette approche à un modèle plus complexe, tel que celui rencontré en tomographie par contraste de phase [79, 147]. Mes futurs travaux viseront aussi à proposer des stratégies MM plus flexibles, basées sur des conditions de majorations plus localisées, dans le but de pouvoir traiter plus efficacement ces modèles non linéaires.

Optimisation sous contraintes : Comme nous l'avons vu dans les chapitres 6 et 7 de ce manuscrit, les algorithmes proximaux permettent de gérer des problèmes d'optimisation non nécessairement différentiables. En particulier, cela inclut les problèmes d'optimisation sous contraintes, traités en incorporant dans la fonction de coût des fonctions indicatrices associées aux domaines des contraintes attendues sur la solution. Si les contraintes sont simples, on peut les faire intervenir via des opérations de projection, en utilisant des approches proximales primales telles que les algorithmes de type explicite implicite avec métrique MM que j'ai présenté dans ce manuscrit. Lorsqu'elles sont plus complexes (ex : non linéaires, ou faisant intervenir un opérateur linéaire coûteux), il est souvent plus efficace d'employer des stratégies proximales primales-duales [123], comme nous l'avons fait par exemple dans nos travaux [168, 167]. Ces approches ont néanmoins l'inconvénient de générer une suite d'itérées qui ne sont pas assurées de vérifier les contraintes qu'à la convergence asymptotique de l'algorithme. Or, dans certaines applications (ex : imagerie médicale, mécanique), la faisabilité de la solution est un point essentiel. Une réponse possible à cette problématique consiste à utiliser une stratégie de type points intérieurs, reposant sur l'idée de fonction barrière, pour garder les itérées dans l'intérieur strict du domaine contraint. Les algorithmes de points intérieurs ont deux inconvénients majeurs, à savoir leur difficulté à traiter des problèmes de grande dimension, et leur emploi limité à la minimisation de critères deux fois différentiables. J'ai montré dans les travaux [56, 52] que l'on pouvait accélérer ces algorithmes grâce à des outils issus de la méthodologie MM. J'aimerais poursuivre dans cette direction, en

cherchant à étendre ces résultats pour la minimisation de critères non lisses sous contraintes. Cette étude fait partie des sujets qui seront étudiés dans le cadre de la thèse de MARIE-CAROLINE CORBINEAU, que je co-encadre depuis octobre 2016 à CentraleSupélec.

Problèmes de très grande taille : Lorsque la taille de la donnée et/ou des variables à estimer devient très grande, comme cela se produit dans le traitement d'image 3D, le traitement de vidéos, ou la classification des images, l'application d'algorithmes d'optimisation standards devient impossible, en raison de problèmes de limitation de mémoire. Ainsi, un défi majeur consiste en la proposition de nouveaux algorithmes d'optimisation capables d'effectuer les traitements de façon distribuée, en utilisant des infrastructures pouvant être distantes de calcul parallèle et de stockage. En particulier, il devient nécessaire d'introduire des techniques d'optimisation combinant des approches parallèles par blocs agissant sur des sous-ensemble des variables à optimiser et des stratégies incrémentales (i.e. en ligne), traitant les données au fur et à mesure de leur chargement sur la mémoire locale de la machine. Dans le cadre de mes travaux de recherche, je chercherai à répondre à ces problématiques par la proposition de nouvelles méthodes bénéficiant du potentiel de parallélisation des stratégies développées dans [2, 174], tout en présentant la forme récursive de l'algorithme [63].

Interactions simulation/optimisation : Les approches basées sur la simulation de la loi *a posteriori* présentent plusieurs aspects très intéressants pour la résolution des problèmes inverses. Tout d'abord, elles fournissent l'estimateur au sens du MMSE, qui est reconnu théoriquement pour être supérieur à l'estimateur MAP en terme de qualité de solution (au sens de l'erreur quadratique). D'autre part, elles permettent de s'affranchir d'une recherche parfois fastidieuse des paramètres de régularisation, ceux-ci pouvant être estimés de façon conjointe à la solution recherchée. Enfin, elles traitent efficacement des modèles pour lesquels le problème d'optimisation associé serait non convexe et donc potentiellement sujet aux minimas locaux. Cependant, il est souvent reproché à ces approches d'être difficiles à mettre en œuvre, et de présenter un temps de convergence trop important, ce qui limite leur utilisation pour des problèmes de grande taille. Plusieurs travaux récents, dont ceux que j'ai présentés dans le Chapitre 8, ont illustré l'intérêt qu'il pouvait y avoir à intégrer dans ces approches des outils modernes d'optimisation [153]. Je me propose de poursuivre dans cette direction en étudiant notamment la possibilité d'accélérer par des stratégies MM le schéma proximal MALA [154]. Je m'intéresserai aussi à l'utilisation de fonctions majorantes non quadratiques dans les approches que nous avons introduites dans [142, 139], afin de pouvoir prendre en compte des termes d'attaches aux données de gradient non Lipschitz. Je me baserai pour cela sur les publications récentes [30, 175], dans le domaine de l'optimisation déterministe. Mon objectif serait de proposer des méthodes compétitives en terme de temps de calcul avec les stratégies de type MAP, capables de traiter des modèles avancées (ex : bruit non Gaussien, régularisations non lisses), tout en bénéficiant des avantages mentionnés

plus tôt des techniques de simulation (réglage automatique des paramètres de régularisation, notamment).

Approches par réseaux de neurones : Les approches d'apprentissage profond par réseaux de neurones ont suscité récemment un très fort engouement, dans la communauté de la vision par ordinateur, de par leur capacité à résoudre très efficacement des problèmes d'analyse d'images (ex : reconnaissance d'objets, segmentation, annotation automatique), pour peu qu'une base d'apprentissage suffisamment conséquente, et des outils de programmation parallèle intensifs (du type GPU) soient disponibles. Ces approches sont à l'heure actuelle difficiles à employer pour la résolution de problèmes inverses faisant intervenir un modèle d'observation sophistiqué (ex : transformée de Radon, flou variant spatialement) et il me semble intéressant de se poser la question de la proposition d'approches de résolution hybrides, mêlant apprentissage profond et estimation MAP par minimisation d'un critère pénalisé, dans la lignée des travaux récents [105, 5, 122, 144]. Par ailleurs, il faut souligner que l'étape d'apprentissage des poids d'un réseau de neurones constitue en soi un problème d'optimisation très intéressant de par la structure complexe, non convexe, et éventuellement non lisse, de la fonction de coût, le grand nombre de variables à estimer, liées entre elles par une structure multi-couches très spécifique, et la très grande dimension de la base d'apprentissage nécessitant l'usage d'algorithmes stochastiques [35]. Les réseaux de neurones sont souvent entraînés par des algorithmes de rétropropagation de type gradient stochastique. Malgré la proposition récente de variantes accélérées [121, 188, 158], le problème de la lenteur de convergence reste crucial. De plus, peu d'algorithmes permettent une gestion efficace des pénalisations non lisses, pourtant nécessaires pour limiter les problèmes de sur-entraînement. Il me semble intéressant de tirer partie des récents progrès en optimisation pour introduire de nouveaux algorithmes intégrant par exemple des approches de sous-espace, des stratégies d'optimisation parallèles/distribuées et l'application des outils proximaux, pour permettre la prise en compte de termes non lisses. Cette thématique fait l'objet de deux projets collaboratifs que j'ai démarré cette année (NextGenBP CEFIPRA-CNRS, MajiMED Université Paris Saclay) impliquant notamment GIOVANNI CHIERCHIA (ESIEE, Université Paris Est), ANGSUL MAJUMDAR (IIIT Delhi) et VYACHESLAV DUDAR (Université de Kiev).

Bibliographie

- [1] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot, and L. Laborelli. A hybrid alternating proximal method for blind video restoration. In *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)*, pages 1811–1815, 1-5 Sep. 2014. 6, 65, 74
- [2] F. Abboud, E. Chouzenoux, J. C. Pesquet, J. H. Chenot, and L. Laborelli. A distributed strategy for computing proximity operators. In *Proceedings of the 49th Asilomar Conference on Signals, Systems and Computers (ASILOMAR 2015)*, pages 396–400, 8-11 Nov. 2015. 6, 76, 95
- [3] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot, and L. Laborelli. A dual block coordinate proximal algorithm with application to deconvolution of interlaced video sequences. In *Proceedings of the 22nd IEEE International Conference on Image Processing (ICIP 2015)*, pages 4917–4921, Quebec City, Canada, 27-30 Sept. 2015. 6, 69
- [4] 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. *Journal of Mathematical Imaging and Vision*, 2017. (to appear). 6, 67, 68, 69
- [5] J. Adler and O. Oktom. Learned primal-dual reconstruction. Technical report, 2017. <https://arxiv.org/abs/1707.06474>. 96
- [6] M. Allain, J. Idier, and Y. Goussard. On global and local convergence of half-quadratic algorithms. *IEEE Transactions on Image Processing*, 15(5) :1130–1142, 2006. 39, 54, 78
- [7] G. E. B. Archer and D. M. Titterington. The iterative image space reconstruction algorithm (isra) as an alternative to the em algorithm for solving positive linear inverse problems. *Statistica Sinica*, 5(1) :77–96, 1995. 42
- [8] G. P. Asner and K. B. Heidebrecht. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions : comparing multispectral and hyperspectral observations. *International Journal of Remote Sensing*, 23(19) :3939–3958, Oct. 2002. 42

-
- [9] Y. F. Atchadé. An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift. *Methodology and Computing in applied Probability*, 8(2) :235–254, 2006. [88](#), [91](#)
- [10] H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming*, 116 :5–16, Jun. 2009. [46](#), [47](#)
- [11] 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. [46](#)
- [12] 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. *Mathematical Programming*, 137 :91–129, Feb. 2011. [47](#), [49](#), [50](#), [62](#)
- [13] A. Auslender. Asymptotic properties of the Fenchel dual functional and applications to decomposition problems. *Journal of Optimization Theory and Applications*, 73(3) :427–449, Jun. 1992. [62](#)
- [14] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational Bayesian super resolution. *IEEE Transactions on Image Processing*, 20(4) :984–999, 2011. [83](#)
- [15] J. M. Bardsley and J. Goldes. Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation. *Inverse Problems*, 25(9) :095005, 2009. [85](#)
- [16] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1) :183–202, 2009. [50](#)
- [17] A. Beck and M. Teboulle. *Gradient-based algorithms with applications to signal-recovery problems*, pages 42–88. Cambridge University Press, 2009. [48](#)
- [18] S. Becker and J. Fadili. A quasi-Newton proximal splitting method. In P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 25, pages 2618 – 2626, 2012. [47](#)
- [19] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle. *A ℓ_1 -Unified Variational Framework for Image Restoration*, pages 1–13. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. [90](#)
- [20] A. Benfenati, E. Chouzenoux, and J.-C. Pesquet. A proximal approach for solving matrix optimization problems involving a bregman divergence. In *Proceedings of the International Biomedical and Astronomical Signal Processing Frontiers workshop (BASP 2017)*, Villars-sur-Ollon, Switzerland, 29 Jan. - 3 Feb. 2017. [6](#), [65](#)

-
- [21] F. Benvenuto, A. La Camera, C. Theys, A. Ferrari, H. Lantéri, and M. Bertero. The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise. *Inverse Problems*, 24(3), 2008. 20 pp. [37](#)
- [22] D. P. Bertsekas. Projected Newton methods for optimization problems with simple constraints. *SIAM Journal on Control and Optimization*, 20 :221–246, 1982. [48](#)
- [23] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 2nd edition, 1999. [52](#), [61](#), [62](#)
- [24] D. P. Bertsekas and J. N. Tsitsiklis. Gradient convergence in gradient methods with errors. *SIAM Journal on Optimization*, 10(3) :627–642, 2000. [55](#)
- [25] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4) :1196–1211, 2000. [48](#)
- [26] J. Bolte, A. Daniilidis, and A. Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17 :1205–1223, 2006. [45](#)
- [27] J. Bolte, A. Daniilidis, and A. Lewis. A nonsmooth morse-sard theorem for subanalytic functions. *Journal of Mathematical Analysis and Applications*, 321 :729–740, 2006. [45](#)
- [28] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18(2) :556–572, 2007. [45](#)
- [29] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming, Series A*, Jul. 2013. [44](#), [69](#)
- [30] J. Bolte, S. Sabach, M. Teboulle, and Y. Vaisbourd. First order methods beyond convexity and Lipschitz gradient continuity with applications to quadratic inverse problems. Technical report, June 2017. <https://arxiv.org/abs/1706.06461>. [95](#)
- [31] S. Bonettini, I. Loris, F. Porta, and M. Prato. Variable metric inexact linesearch-based methods for nonsmooth optimization. *SIAM Journal on Optimization*, 26(2) :891–921, 2016. [67](#)
- [32] S. Bonettini, I. Loris, F. Porta, M. Prato, and S. Rebegoldi. On the convergence of a linesearch based proximal-gradient method for nonconvex optimization. *Inverse Problems*, 33(5) :055005, 2017. [49](#)
- [33] S. Bonettini, R. Zanella, and L. Zanni. A scaled gradient projection method for constrained image deblurring. *Inverse Problems*, 25(1) :015002+, 2009. [48](#)
- [34] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. A family of variable metric proximal methods. *Math. Program.*, 68 :15–47, 1995. [47](#)

- [35] L. Bottou, F. E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. Technical report, 2016. <http://arxiv.org/pdf/1606.04838v1.pdf>. 96
- [36] J. V. Burke and M. Qian. A variable metric proximal point algorithm for monotone operators. *SIAM Journal on Control and Optimization*, 37 :353–375, 1999. 47
- [37] R. H. Byrd and D. A. Payne. Convergence of the iteratively reweighted least squares algorithm for robust regression. Technical Report 131, The Johns Hopkins University, Baltimore, MD, June 1979. 39
- [38] D. Carlson, Y.-P. Hsieh, E. Collins, L. Carin, and V. Cevher. Stochastic spectral descent for discrete graphical models. *IEEE J. Sel. Top. Signal Process.*, 10(2) :296–311, Mar. 2016. 55
- [39] L. Chaari, E. Chouzenoux, N. Pustelnik, C. Chaux, and C. Moussaoui. OPTIMED : optimisation it erative pour la r solution de probl emes inverses de grande taille. *Traitement du Signal*, 28(3–4) :329–374, 2011. 31
- [40] A. Chambolle and T. Pock. A remark on accelerated block coordinate descent for computing the proximity operators of a sum of convex functions. *SMAI Journal of Computational Mathematics*, 1 :29–54, 2015. 68
- [41] T. F. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM Journal on Numerical Analysis*, 36(2) :354–367, 1999. 77, 84
- [42] J.-P. Chancelier. Auxiliary problem principle and inexact variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. Technical report, 2015. <https://arxiv.org/abs/1508.02994>. 49
- [43] P. Charbonnier, L. Blanc-F raud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6 :298–311, 1997. 39, 77
- [44] G. H.-G. Chen and R. T. Rockafellar. Convergence rates in forward-backward splitting. *SIAM Journal on Optimization*, 7 :421–444, 1997. 46, 47, 48
- [45] Z. Chen, S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational Bayesian methods for multimedia problems. *IEEE Transactions on Multimedia*, 16(4) :1000–1017, 2014. 83
- [46] A. Cherni, E. Chouzenoux, and M.-A. Delsuc. PALMA, an improved algorithm for DOSY signal processing. *Analyst*, 142(5) :772–779, 2016. 5
- [47] A. Cherni, E. Chouzenoux, and M.-A. Delsuc. Proximity operators for a class of hybrid sparsity+entropy priors. application to DOSY NMR signal reconstruction. In *Proceedings of the 8th International Symposium on Signal, Image, Video and Communications (ISIVC 2016)*, Tunis, Tunisia, 21–23 Nov. 2016. 5, 65

- [48] G. Chierchia, A. Cherni, E. Chouzenoux, and J.-C. Pesquet. Approche de Douglas-Rachford aléatoire par blocs appliquée à la régression logistique parcimonieuse. In *Actes du 26e colloque GRETSI (GRETSI 2017)*, 5-8 Sep. 2017. [5](#)
- [49] E. Chouzenoux, J. Idier, and S. Moussaoui. A Majorize-Minimize strategy for subspace optimization applied to image restoration. *IEEE Transactions Image Processing*, 20(18) :1517–1528, Jun. 2011. [32](#), [51](#), [52](#), [54](#), [59](#)
- [50] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot. A majorize-minimize subspace approach for ℓ_2 - ℓ_0 image regularization. *SIAM Journal Imaging Science*, 6(1) :563–591, 2013. [39](#), [41](#), [46](#), [52](#), [53](#), [54](#), [59](#)
- [51] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot. A convex approach for image restoration with exact Poisson-Gaussian likelihood. *SIAM Journal on Imaging Science*, 8(4) :2662–2682, 2015. [36](#), [37](#), [85](#), [86](#)
- [52] E. Chouzenoux, M. Legendre, S. Moussaoui, and J. Idier. Fast constrained least squares spectral unmixing using primal-dual interior-point optimization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1) :59–69, Jan. 2012. [32](#), [94](#)
- [53] E. Chouzenoux, S. Moussaoui, and J. Idier. A majorize-minimize line search algorithm for barrier function optimization. In *Proceedings of the 17th European Signal Processing Conference (EUSIPCO 2009)*, pages 1379–1393, Glasgow, UK, 24-28 Aug. 2009. [32](#)
- [54] E. Chouzenoux, S. Moussaoui, and J. Idier. Algorithme primal-dual de points intérieurs pour l’estimation pénalisée des cartes d’abondances en imagerie hyperspectrale. In *Actes du 23e colloque GRETSI (GRETSI 2011)*, Bordeaux, France, 5–8 Sep. 2011. [32](#)
- [55] E. Chouzenoux, S. Moussaoui, and J. Idier. Efficiency of line search strategies in interior point methods for linearly constrained optimization. In *Proceedings of IEEE Workshop on Statistical Signal Processing (SSP 2011)*, pages 101–104, Nice, France, 28-30 June 2011. [32](#)
- [56] E. Chouzenoux, S. Moussaoui, and J. Idier. Majorize-minimize linesearch for inversion methods involving barrier function optimization. *Inverse Problems*, 28(6) :065011, 2012. [32](#), [94](#)
- [57] E. Chouzenoux, S. Moussaoui, J. Idier, and F. Mariette. Reconstruction d’un spectre RMN 2D par maximum d’entropie. In *Actes 22e colloque GRETSI (GRETSI 2009)*, Dijon, France, 8-11 Sep. 2009. [32](#)
- [58] E. Chouzenoux, S. Moussaoui, J. Idier, and F. Mariette. Efficient maximum entropy reconstruction of nuclear magnetic resonance T1-T2 spectra. *IEEE Transactions on Signal Processing*, 58(12) :6040–6051, 2010. [32](#)
- [59] E. Chouzenoux, S. Moussaoui, J. Idier, and F. Mariette. Optimization of a maximum entropy criterion for 2D nuclear magnetic resonance reconstruction. In *Proceedings of the 35th IEEE International Conference on Acoustics*,

- Speech, and Signal Processing (ICASSP 2010)*, Dallas, TX, USA, 14-19 Mar. 2010. [32](#)
- [60] E. Chouzenoux, S. Moussaoui, J. Idier, and F. Mariette. Primal-dual interior point optimization for a regularized reconstruction of NMR relaxation time distributions. In *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, pages 8747–8750, Vancouver, Canada, 26-31 Mar. 2013. [32](#)
- [61] E. Chouzenoux, S. Moussaoui, M. Legendre, and J. Idier. Algorithme primal-dual de points intérieurs pour l'estimation pénalisée des cartes d'abondances en imagerie hyperspectrale. *Traitement du Signal*, 30(1-2) :35–59, 2013. [32](#)
- [62] E. Chouzenoux and J.-C. Pesquet. Convergence rate analysis of the majorize-minimize subspace algorithm. *IEEE Signal Processing Letters*, 23(9) :1284–1288, Sep. 2016. [52](#), [53](#), [57](#)
- [63] E. Chouzenoux and J.-C. Pesquet. A stochastic majorize-minimize subspace algorithm for online penalized least squares estimation. *IEEE Transactions Signal Processing*, 2017. (to appear). [39](#), [46](#), [51](#), [55](#), [57](#), [58](#), [95](#)
- [64] E. Chouzenoux, J.-C. Pesquet, and A. Florescu. A multi-parameter optimization approach for complex continuous sparse modeling. In *Proceedings of the 19th International Conference on Digital Signal Processing (DSP 2014)*, 20-23 Aug. 2014. [38](#)
- [65] E. Chouzenoux, J.-C. Pesquet, and A. Florescu. A stochastic 3MG algorithm with application to 2D filter identification. In *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)*, pages 1587–1591, Lisboa, Portugal, 1-5 Sep. 2014. [58](#)
- [66] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. *Journal of Optimization Theory and Applications*, 162(1) :107–132, Jul. 2014. [6](#), [43](#), [46](#), [47](#), [48](#), [49](#), [67](#)
- [67] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. A block coordinate variable metric forward-backward algorithm. *Journal on Global Optimization*, 66(3) :457–485, 2016. [6](#), [43](#), [63](#), [64](#), [65](#), [74](#), [94](#)
- [68] P. Ciuciu and J. Idier. A half-quadratic block-coordinate descent method for spectral estimation. *Signal Processing*, 82(7) :941–959, 2002. [77](#), [84](#)
- [69] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin. The U.S. geological survey digital spectral library : version 1 : 0.2 to 3.0 μm . *U.S. Geological Survey, Denver, CO, Open File Rep. 93-592*, 1993. [43](#)
- [70] P. L. Combettes, D. Dung, and B. C. Vũ. Dualization of signal recovery problems. *Set-Valued and Variational Analysis*, 18(3) :373–404, Dec. 2010. [67](#)
- [71] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R.

- Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York, 2010. [36](#), [65](#), [67](#)
- [72] P. L. Combettes and J.-C. Pesquet. Stochastic quasi-fejèr block-coordinate fixed point iterations with random sweeping. *SIAM Journal on Optimization*, 25(2) :1221–1248, 2015. [62](#)
- [73] P. L. Combettes and J.-C. Pesquet. Stochastic forward-backward and primal-dual approximation algorithms with application to online image restoration. In *Proceedings of the 24th European Signal Processing Conference (EUSIPCO 2016)*, Budapest, Hungary, 29 Aug.-3 Sep. 2016. [54](#)
- [74] P. L. Combettes and B. C. Vũ. Variable metric forward-backward splitting with applications to monotone inclusions in duality. *Optimization : A Journal of Mathematical Programming and Operations Research*, 63(9) :1289–1318, Sept. 2014. [47](#), [48](#)
- [75] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4) :1168–1200, Nov. 2005. [46](#), [65](#)
- [76] P.L. Combettes, D. Dung, and B.C. Vũ. Proximity for sums of composite functions. *Journal of Mathematical Analysis and Applications*, 380(2) :680–688, 2011. [67](#), [69](#), [70](#)
- [77] M. Coste. An introduction to o-minimal geometry. *RAAG Notes*, 81 pages, Institut de Recherche Mathématiques de Rennes, Nov. 1999. [45](#)
- [78] H. M. David. Auxiliary variable methods for Markov Chain Monte Carlo with applications. *Journal of the American Statistical Association*, 93 :585–595, Feb. 1997. [89](#)
- [79] V. Davidoiu, B. Sixou, M. Langer, and F. Peyrin. Nonlinear phase retrieval using projection operator and iterative wavelet thresholding. *IEEE Signal Processing Letters*, 19(9) :579–582, Sep. 2012. [94](#)
- [80] A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Transactions on Medical Imaging*, 14(1) :132–137, 1995. [72](#)
- [81] A. Defazio, F. Bach, and S. Lacoste. SAGA : A fast incremental gradient method with support for non-strongly convex composite objectives. In *Proceedings of the Neural Information Processing Systems Conference (NIPS 2014)*, pages 1646–1654, Montreal, Canada, Dec. 8-11 2014. [57](#), [58](#), [60](#)
- [82] A. Defazio, J. Domke, and T. Cartano. Finito : A faster, permutable incremental gradient method for big data problems. In *Proceedings 31st International Conference Machine Learning (ICML 2014)*, pages 1125–1133, Beijing, China, Jun. 21-26 2014. [57](#)

- [83] S. Delpretti, F. Luisier, S. Ramani, T. Blu, and M. Unser. Multiframe suretlet denoising of timelapse fluorescence microscopy images. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI 2008)*, pages 149–152, Paris, France, 14–17 May 2008. 37
- [84] P. S. R. Diniz. *Adaptive Filtering. Algorithms and Practical Implementation*. Springer Science & Business Media, New York, 4th edition, 2012. 58
- [85] N. Dobigeon and N. Brun. Spectral mixture analysis of eels spectrum-images. *Ultramicroscopy*, 120 :25–34, Sept. 2012. 42
- [86] A. Drémeau, C. Herzet, and L. Daudet. Boltzmann machine and mean-field approximation for structured sparse decompositions. *IEEE Transactions on Signal Processing*, 60(7) :3425–3438, 2012. 83
- [87] S. Duane, A.D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physical Letters B*, 195(2) :216 – 222, Sep. 1987. 89
- [88] E. Guillard E. Chouzenoux, F. Zolyniak and H. Talbot. A majorize-minimize memory gradient algorithm applied to x-ray tomography. In *Proceedings of 20th IEEE International Conference on Image Processing (ICIP 2013)*, pages 1011–1015, 15–18 Sep. 2013. 7, 41
- [89] M. Elad. *Sparse and Redundant Representations : From Theory to Applications in Signal and Image Processing*. Springer New York, 2010. 38
- [90] H. Erdogan and J. A. Fessler. Monotonic algorithms for transmission tomography. *IEEE Transactions on Medical Imaging*, 18(9) :801–814, Sept. 1999. 41, 49, 62, 71
- [91] J. M. Ermoliev and Z. V. Nekrylova. The method of stochastic gradients and its application. In *Seminar : Theory of Optimal Solutions. No. 1 (Russian)*, pages 24–47. Akad. Nauk Ukrain. SSR, Kiev, 1967. 55
- [92] J. R. Fienup. Phase retrieval algorithms : a personal tour. *Applied Optics*, 52(1) :45–56, Jan. 2013. 94
- [93] M. Figueiredo, J. Bioucas-Dias, and R. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 16 :2980–2991, 2007. 48
- [94] A. Florescu, E. Chouzenoux, J.-C. Pesquet, and S. Ciochina. A constrained optimization approach for complex sparse perturbed models. In *Proceedings of the Signal Processing with Adaptive Sparse Structured Representations (SPARS 2013)*, Lausanne, Switzerland, 8–11 July 2013. 38, 65
- [95] A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu, and S. Ciochina. A majorize-minimize memory gradient method for complex-valued inverse problem. *Signal Processing*, 103 :285–295, Oct. 2014. Special issue on Image Restoration and Enhancement : Recent Advances and Applications. 53

-
- [96] P. Frankel, G. Garrigos, and J. Peypouquet. Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates. *Journal of Optimization Theory and Applications*, 165(3) :874–900, 2015. 46
- [97] A. Fraysse and T. Rodet. A measure-theoretic variational Bayesian algorithm for large dimensional problems. *SIAM Journal on Imaging Sciences*, 7(4) :2591–2622, 2014. 83
- [98] J.-J. Fuchs. Convergence of a sparse representations algorithm applicable to real or complex data. *IEEE Transactions on Signal Processing*, 1(4) :598–605, December 2007. 39
- [99] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3) :367–383, Mar. 1992. 51, 78
- [100] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4(7) :932–946, July 1995. 51, 78
- [101] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6) :721–741, 1984. 89
- [102] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 73(91) :123–214, Mar. 2011. 88, 89
- [103] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 3rd edition, 1996. 61
- [104] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using focuss : a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3) :600–616, Mar. 1997. 39
- [105] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser. CNN-based projected gradient descent for consistent image reconstruction. Technical report, 2017. <https://arxiv.org/abs/1709.01809>. 96
- [106] W. W. Hager and H. Zhang. A survey of nonlinear conjugate gradient methods. *Pacific Journal on Optimization*, 2(1) :35–58, Jan. 2006. 32, 52
- [107] S. O. Haykin. *Adaptive Filter Theory*. Prentice Hall, Englewood Cliffs, NJ, 4th edition, 2002. 54, 55
- [108] G. E. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3) :267–276, Mar. 1994. 49
- [109] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Springer-Verlag, New York, 1993. 47

- [110] M. Hong, M. Razaviyayn, Z. Q. Luo, and J. S. Pang. A unified algorithmic framework for block-structured optimization involving big data : With applications in machine learning and signal processing. *IEEE Signal Processing Magazine*, 33(1) :57–77, Jan. 2016. [45](#), [62](#), [71](#)
- [111] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *American Statistical Association*, 58(1) :30–37, Feb. 2004. [35](#), [41](#)
- [112] M. Hurn. Difficulties in the use of auxiliary variables in Markov chain Monte Carlo methods. *Statistics and Computing*, 7 :35–44, Mar. 1997. [89](#)
- [113] J. Idier. Convex half-quadratic criteria and interacting auxiliary variables for image restoration. *IEEE Transactions on Image Processing*, 10(7) :1001–1009, July 2001. [38](#), [39](#), [51](#), [77](#), [79](#), [81](#)
- [114] M. W. Jacobson and J. A. Fessler. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Transactions on Image Processing*, 16(10) :2411–2422, Oct. 2007. [45](#), [62](#), [64](#), [71](#)
- [115] M. Jaggi, V. Smith, M. Takac, J. Terhorst, S. Krishnan, T. Hofmann, and M. I. Jordan. Communication-efficient distributed dual coordinate ascent. In Z. Ghahramani, M. Welling, C. Cortes, N.d. Lawrence, and K.q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3068–3076. Curran Associates, Inc., 2014. [68](#)
- [116] J. R. Janesick. *Photon Transfer*, volume PM170. SPIE Press Book, Bellingham, WA, 2007. [49](#)
- [117] A. Jeziarska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot. A primal-dual proximal splitting approach for restoring data corrupted with poisson-gaussian noise. In *Proceedings of the 37th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)*, pages 1085–1088, Kyoto, Japan, 25–30 Mar. 2012. [37](#)
- [118] M. Jiang and W. Wang. Convergence of the simultaneous algebraic reconstruction technique (SART). *IEEE Transactions on Image Processing*, 12(8) :957–961, Aug. 2003. [42](#)
- [119] A. Kak and M. Slaney. *Principles of computerized tomographic imaging*. Classics in Applied Mathematics. SIAM, Philadelphia, 1988. [41](#)
- [120] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1) :44–57, January 2002. [42](#)
- [121] D. P. Kingma and J. Ba. Adam : A method for stochastic optimization. In *Int. Conf. Learn. Representations*, Banff, Canada, 14-16 April 2014. [96](#)
- [122] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock. *Variational Networks : Connecting Variational Methods and Deep Learning*, pages 281–293. Springer International Publishing, Cham, 2017. [96](#)

-
- [123] N. Komodakis and J.-C. Pesquet. Playing with duality : An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Processing Magazine*, 32(6) :31–54, Nov. 2014. 94
- [124] J. Konečný, H. B. McMahan, and D. Ramage. Federated optimization : Distributed optimization beyond the datacenter. Technical report, 2015. <http://arxiv.org/pdf/1511.03575v1.pdf>. 71
- [125] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, pages 233–240, Colorado Springs, CO, USA, Jun. 21-25, 2011. 65, 66
- [126] K. Kurdyka. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier*, 48(3) :769–783, 1998. 45
- [127] C. Labat and J. Idier. Convergence of truncated half-quadratic and newton algorithms, with application to image restoration. Technical report, IRCCyN, Nantes, France, June 2007. 51
- [128] C. Labat and J. Idier. Convergence of conjugate gradient methods with a closed-form stepsize formula. *Journal of Optimization Theory and Applications*, 136(1) :43–60, Jan. 2008. 32
- [129] Y. LeCun, F. J. Huang, and L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, volume 2, pages II–97, 2004. 54
- [130] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural and Information Processing Systems*, volume 13, pages 556–562, 2001. 42
- [131] J. D. Lee, Y. Sun, and M. A. Saunders. Proximal Newton-type methods for convex optimization. In P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25, pages 827 – 835, 2012. 47, 48
- [132] M. Legendre, S. Moussaoui, F. Schmidt, and J. Idier. Parallel implementation of a primal-dual interior-point optimization method for fast abundance maps estimation. In *Proceedings of the 5th Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS 2013)*, Gainesville, FL, USA, June 2013. 71
- [133] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pages 1964–1971, Miami, Florida, USA, 20-25 Jun. 2009. 94
- [134] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(3) :503–528, 1989. 32, 52

- [135] S. Łojasiewicz. *Une propriété topologique des sous-ensembles analytiques réels*, pages 87–89. Editions du centre National de la Recherche Scientifique, 1963. [45](#)
- [136] P. A. Lotito, L. A. Parente, and M. V. Solodov. A class of variable metric decomposition methods for monotone variational inclusions. *Journal of Convex Analysis*, 16 :857–880, 2009. [47](#)
- [137] J. Mairal. Stochastic Majorization-Minimization algorithms for large-scale optimization. In *Proceedings of the Conference on Neural Information Processing Systems (NIPS 2013)*, pages x–x+8, Lake Tahoe, Nevada, Dec. 5-8 2013. [55](#), [57](#)
- [138] Y. Marnissi, A. Benazza-Benyahia, E. Chouzenoux, and J.-C. Pesquet. Generalized multivariate exponential power prior for wavelet-based multichannel image restoration. In *Proceedings of the 20th IEEE International Conference on Image Processing (ICIP 2013)*, pages 2402–2406, 15-18 Sept. 2013. [6](#), [7](#), [91](#)
- [139] Y. Marnissi, A. Benazza-Benyahia, E. Chouzenoux, and J.-C. Pesquet. Majorize-minimize adapted metropolis hastings algorithm. application to multichannel image recovery. In *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)*, pages 1332–1336, 1-5 Sept. 2014. [6](#), [88](#), [95](#)
- [140] Y. Marnissi, E. Chouzenoux, A. Benazza-Benyahia, J.-C. Pesquet, and L. Duval. Reconstruction de signaux parcimonieux à l’aide d’un algorithme rapide d’échantillonnage stochastique. In *Actes du 25e colloque GRETSI (GRETSI 2015)*, 8-11 Sept. 2015. [6](#), [88](#)
- [141] Y. Marnissi, E. Chouzenoux, J.-C. Pesquet, and A. Benazza-Benyahia. An auxiliary variable method for langevin based MCMC algorithms. In *Proceedings of the IEEE Workshop on Statistical Signal Processing (SSP 2016)*, pages 297–301, 26-29 June 2016. [6](#), [89](#), [91](#)
- [142] Y. Marnissi, Y. Zheng, E. Chouzenoux, and J.-C. Pesquet. A variational bayesian approach for image restoration. application to image deblurring with poisson-gaussian noise. *IEEE Transactions on Computational Imaging*, 2017. (to appear). [6](#), [37](#), [80](#), [83](#), [85](#), [87](#), [95](#)
- [143] J. Martin, C. L. Wilcox, C. Burstedde, and O. Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3) :1460–1487, Jan. 2012. [88](#)
- [144] M. T. McCann, K. H. Jin, and M. Unser. A review of convolutional neural networks for inverse problems in imaging. Technical report, 2017. <https://arxiv.org/abs/1710.04011>. [96](#)
- [145] A. Miele and J. W. Cantrell. Study on a memory gradient method for the minimization of functions. *Journal on Optimization Theory and Applications*, 3(6) :459–470, 1969. [52](#)

-
- [146] A. Mira and L. Tierney. On the use of auxiliary variables in Markov chain Monte Carlo sampling. Technical report, 1997. <http://cite-seerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.7814>. 89
- [147] K. A. Mohan, X. Xiao, and C. A. Bouman. Direct model-based tomographic reconstruction of the complex refractive index. In *Proceedings of the 23rd IEEE International Conference on Image Processing (ICIP 2016)*, pages 1754–1758, Phoenix, AZ, 25-28 Sep. 2016. 94
- [148] H. Mohimani, M. Babaie-Zadeh, and C. Jutten. A fast approach for overcomplete sparse decomposition based on smoothed ell^0 norm. *IEEE Transactions on Signal Processing*, 57(1) :289–301, Jan. 2009. 39
- [149] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy. Spatiotemporal reconstruction of list-mode PET data. *IEEE Transactions Medical Imaging*, 21(4) :396–404, Apr. 2002. 37
- [150] M. Nikolova and M. K. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing*, 27 :937–966, 2005. 39, 51, 54, 77, 79, 80
- [151] A. Onose, R. E. Carrillo, A. Repetti, J. D. McEwen, J.-T. Thiran, J.-C. Pesquet, and Y. Wiaux. Scalable splitting algorithms for big-data interferometric imaging in the SKA era. *Monthly Notices of the Royal Astronomical Society*, 462(4) :4314–4335, 2016. 71
- [152] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, NY, 1970. 61, 62
- [153] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. Hero, and S. McLaughlin. A survey of stochastic simulation and optimization methods in signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2) :224–241, Mar. 2016. 55, 95
- [154] Marcelo Pereyra. Proximal markov chain monte carlo algorithms. *Statistics and Computing*, 26(4) :745–760, Jul 2016. 95
- [155] J.-C. Pesquet and A. Repetti. A class of randomized primal-dual algorithms for distributed optimization. *Journal of nonlinear and convex analysis*, 16(12), June 2014. 62
- [156] M. J. D. Powell. On search directions for minimization algorithms. *Mathematical Programming*, 4 :193–201, 1973. 62
- [157] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, and J.-C. Pesquet. *Wavelet-based Image Deconvolution and Reconstruction*. John Wiley & Son, Feb. 2016. 38
- [158] N. Qian. On the momentum term in gradient descent learning algorithms. *Neural Network*, 12(1) :145–151, January 1999. 96

- [159] Z. Qu, P. Richtárik, and T. Zhang. Randomized dual coordinate ascent with arbitrary sampling. In *Proceedings of the Conference on Neural Information Processing Systems (NIPS 2015)*, pages 865–873, Montréal, Canada, 7-12 Dec. 2015. [68](#)
- [160] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado. Subset selection in noise based on diversity measure minimization. *IEEE Transactions on Signal Processing*, 51(3) :760–770, Mar. 2003. [39](#)
- [161] M. Razaviyayn, M. Hong, and Z. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2) :1126–1153, 2013. [45](#), [62](#), [64](#)
- [162] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. Reconstruction d’image en présence de bruit gaussien dépendant par un algorithme explicite-implicite à métrique variable. In *Actes du 24e colloque GRETSI (GRETSI 2013)*, Brest, France, 3-6 Sep. 2013. [6](#), [47](#)
- [163] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. A nonconvex regularized approach for phase retrieval. In *Proceedings of the 21st IEEE International Conference on Image Processing (ICIP 2014)*, pages 1753–1757, 27-30 Oct. 2014. [6](#), [65](#), [94](#)
- [164] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. A preconditioned forward-backward approach with application to large-scale nonconvex spectral unmixing problems. In *Proceedings of the 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, pages 1498–1502, Firenze, Italy, 4-9 May 2014. [6](#), [42](#), [44](#), [65](#)
- [165] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. A parallel block-coordinate approach for primal-dual splitting with arbitrary random block selection. In *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO 2015)*, pages 235–239, Nice, France, 1-5 Sep. 2015. [6](#), [62](#), [71](#)
- [166] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. Proximal primal-dual optimization methods. In *Proceedings of International Biomedical and Astronomical Signal Processing (BASP) Frontiers workshop*, Villars-sur-Ollon, Switzerland, 25-30 Jan. 2015. [6](#)
- [167] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. A random block-coordinate primal-dual proximal algorithm with application to 3d mesh denoising. In *Proceedings of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015)*, pages 3561–3565, Brisbane, Australia, 19-24 Apr. 2015. [6](#), [62](#), [94](#)
- [168] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. Un petit tutoriel sur les méthodes primales-duales pour l’optimisation convexe. In *Actes du 25e colloque GRETSI (GRETSI 2015)*, Lyon, France, 8-11 Sep. 2015. [6](#), [94](#)

-
- [169] A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet. Euclid in a taxicab : Sparse blind deconvolution with smoothed l1/l2 regularization. *IEEE Signal Processing Letters*, 22(5) :539–543, May 2015. 6, 64, 74
- [170] P. Richtárik and M. Talác. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming*, 144(1-2) :1–38, Apr. 2014. 62
- [171] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3) :400–407, 1951. 55, 58, 60
- [172] G. O. Roberts and O. Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Methodology and computing in applied probability*, 4(4) :337–357, 2002. 88
- [173] G. O. Roberts and L. R. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4) :341–363, Dec. 1996. 88
- [174] J.-C. Pesquet S. Cadoni, E. Chouzenoux and C. Chaux. A block parallel majorize-minimize memory gradient algorithm. In *23rd IEEE International Conference on Image Processing (ICIP 2016)*, pages 3194–3198, 25-28 Sep. 2016. 7, 41, 71, 72, 73, 95
- [175] S. Salzo. The variable metric forward-backward splitting algorithm under mild differentiability assumptions. *SIAM Journal on Optimization*, 27(4) :2153–2181, May 2017. 95
- [176] M. Schmidt, N. Le Roux, and F. Bach. Minimizing finite sums with the stochastic average gradient. 2013. <http://arxiv.org/abs/1309.2388>. 57
- [177] M. Seeger and G. Bouchard. Fast variational Bayesian inference for non-conjugate matrix factorization models. In *Proceedings of the Artificial Intelligence and Statistics Conference (AISTATS 2012)*, pages 1012–1018, La Palma, Canary Islands, 21-23 Apr. 2012. 83
- [178] S. Setzer, G. Steidl, and T. Teuber. On vector and matrix median computation. *Journal of Computational and Applied Mathematics*, 236(8) :2200–2222, Feb. 2012. 77
- [179] S. Shalev-Shwartz and T. Zhang. Stochastic dual coordinate ascent methods for regularized loss minimization. *Journal on Machine Learning Research*, 14(1) :567–599, February 2013. 68
- [180] V. Šmídl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer-Verlag, New York, USA, 2005. 82
- [181] S. Sotthivirat and J. A. Fessler. Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms. *IEEE Transactions on Signal Processing*, 11(3) :306–317, 2002. 71, 72

-
- [182] M. A. Stuart, J. Voss, and P. Wiberg. Conditional path sampling of SDEs and the Langevin MCMC method. *Communications in Mathematical Sciences*, 2(4) :685–697, 2004. [88](#)
- [183] Y. Sun, P. Babu, and D. P. Palomar. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, 65(3) :794–816, Aug. 2016. [45](#)
- [184] A. Tacchetti, P. Mallapragada, M. Santoro, and L. Rosasco. GURLS : a least squares library for supervised learning. *Journal on Machine Learning Research*, 14 :3201–3205, 2013. [54](#)
- [185] R. Tappenden, P. Richtárik, and J. Gondzio. Inexact coordinate descent : Complexity and preconditioning. Technical report, 2013. <http://arxiv.org/abs/1304.5530>. [62](#)
- [186] K. E. Themelis, F. Schmidt, O. Sykioti, A. A. Rontogiannis, K. D. Koutroumbas, and I. A. Daglis. On the unmixing of MEX/OMEGA hyperspectral data. *Planetary and Space Science*, 68(1) :34–41, 2012. [42](#)
- [187] H. Tian, B. Fowler, and A. E. Gamal. Analysis of temporal noise in CMOS photodiode active pixel sensor. *IEEE Solid State Circuits Magazine*, 36(1) :92–101, Jan. 2001. [49](#)
- [188] T. Tieleman and G. Hinton. Lecture 6.5 – RMSProp : Divide the gradient by a running average of its recent magnitude. COURSERA : Neural Networks for Machine Learning, 2012. [96](#)
- [189] E. W. Tramel, A. Drémeau, and F. Krzakala. Approximate message passing with restricted Boltzmann machine priors. *J. Stat. Mech. : Theory and Experiment*, 2016(7) :073401, 2016. [83](#)
- [190] Q. Tran-Dinh, A. Kyrillidis, and V. Cevher. Composite self-concordant minimization. *Journal of Machine Learning Research*, 16(1) :371–416, Jan. 2015. [47](#), [48](#)
- [191] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2) :431–446, 1998. [46](#), [48](#)
- [192] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3) :475–494, Jun. 2001. [62](#)
- [193] C. Vacar, J.-F. Giovannelli, and Y. Berthoumieu. Langevin and Hessian with Fisher approximation stochastic sampling for parameter estimation of structured covariance. In *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, pages 3964–3967, Prague, Czech Republic, 22-27 May 2011. [88](#)
- [194] L. Van den Dries and C. Miller. Geometries categories and o-minimal structures. *Duke Mathematical Journal*, 84 :497–540, 1996. [45](#)

-
- [195] D. A. Van Dyk and X-L Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 2012. [89](#)
- [196] X. Vu, C. Chaux, N. Thirion-Moreau, S. Maire, and E. Carstea. A new penalized nonnegative third-order tensor decomposition using a block coordinate proximal gradient approach : Application to 3d fluorescence spectroscopy. *Journal of Chemometrics*, 2016. [42](#)
- [197] E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tôhoku Mathematical Journal (first series)*, (43) :355–386, 1937. [77](#)
- [198] E. Weiszfeld and F. Plastria. On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research*, 167 :7–41, 2009. [39](#)
- [199] L. Xiao. Dual averaging methods for regularized stochastic learning and online optimization. *Journal on Machine Learning Research*, 11 :2543–2596, Oct. 2010. [55](#), [57](#), [58](#), [60](#)
- [200] W. I. Zangwill. *Nonlinear Programming*. Prentice-Hall, Englewood Cliffs, New Jersey, 1969. [61](#), [62](#)
- [201] Y. Zhang and C. A. Sutton. Quasi-Newton methods for Markov chain Monte Carlo. In *Proceedings of the conference on Neural Information Processing Systems (NIPS 2011)*, number 24, pages 2393–2401, Granada, Spain, 12-17 Dec. 2011. [88](#)
- [202] Z. Zhang, J. T. Kwok, and D.-Y. Yeung. Surrogate maximization/minimization algorithms and extensions. *Machine Learning*, 69 :1–33, 2007. [45](#)
- [203] Y. Zheng, A. Fraysse, and T. Rodet. Efficient variational Bayesian approximation method based on subspace optimization. *IEEE Transactions on Image Processing*, 24(2) :681–693, 2015. [83](#)