Inversion Bayésienne : illustration sur des problèmes tomographiques et astrophysiques

Thomas Rodet

Thomas Rodet. Inversion Bayésienne : illustration sur des problèmes tomographiques et astrophysiques. Traitement du signal et de l'image [eess.SP]. Université Paris Sud - Paris XI, 2012. tel-00819179

HAL Id: tel-00819179
https://theses.hal.science/tel-00819179

Submitted on 30 Apr 2013

N° d'ordre :

Université Paris-Sud 11
Faculté des sciences d'Orsay

Mémoire

pour obtenir

L'habilitation à diriger des recherches

préparé au laboratoire des signaux et systèmes
dans le Groupe des Problèmes Inverses

soutenue publiquement
par

Thomas RODET

le 20 novembre 2012

Inversion Bayésienne : illustration sur des problèmes tomographiques et astrophysiques

JURY

Mme. Laure BLANC-FERRAUD	Rapporteur
M. Christophe COLLET	Rapporteur
Mme. Sylvie ROQUES	Rapporteur
Mme. Irène BUVAT	Présidente
M. Hichem SNOUSSI	Examinateur
M. Eric WALTER	Examinateur

Remerciements

Avant toute chose, à l'ocassion de cette HdR, je souhaite ici exprimer mes plus sincères remerciements :

à Laure BLANC-FERAUD, Directrice de Recherche CNRS à I3S, Sylvie ROQUES, Directrice de Recherche CNRS au LATT et Christophe COLLET, Professeur à l'Université de Strasbourg au LSIIT pour avoir accepté la lourde tâche d'être rapporteur de ce manuscrit. Leurs conseils et leurs remarques m'ont été très utiles.

À Irène BUVAT, Directrice de Recherche CNRS à IMNC, présidente du jury de cette HdR, pour son examen minutieux et ses remarques pertinentes.

À Hichem SNOUSSI, Professeur à l'Université Technologique de Troyes au LM2S, qui m'a fait l'honneur d'être examinateur de ce travail.

À Eric WALTER, Directeur de Recherche CNRS à double titre, pour avoir accepté d'être examinateur et pour m'avoir acceuilli très chaleureusement lors de mon arrivé au L2S. J'ai apprécié pendant ses deux mandats de directeur de laboratoire, ses qualités humaines hors du commun.

À Silviu NICULESCU, Directeur de Recherche CNRS, actuel directeur du L2S, pour son travail constant pour le rayonnement de notre laboratoire, et pour l'aide qu'il m'a apportée pour développer le groupe des problèmes inverses.

Ce retour sur le travail que j'ai mené ces dix dernières années m'amène à penser que les qualités personnelles passent en grande partie par des recontres. De ce point de vue, j'ai eu la chance de croiser Michel DEFRISE et Guy DEMOMENT, qui sont des chercheurs execptionnels autant d'un point de vue scientifique que d'un point de vue humain. Ils restent pour moi des références en terme de déontologie.

Je tiens aussi à exprimer ma reconnaissance à Alain ABERGEL, qui depuis mon recrutement n'a pas cessé de me soutenir. Je le remercie aussi pour les rudiments d'astrophysique qu'il m'a appris avec patience.

Merci également, à Jean–François GIOVANNELLI pour m'avoir accueilli au Groupe Problèmes Inverses qu'il dirigeait à l'époque, et pour m'avoir converti à l'inférence Bayésienne à force de discussions philosophiques au café.

Aux membres actuels du Groupe Problèmes Inverses, en particulier aux permanents Aurélia FRAYSSE, Nicolas GAC, Matthieu KOWALSKI, Ali MOHAMMAD-DJAFARI, Hélène PAPADOPOULOS avec lesquels j'ai de nombreux échanges, en particulier pour les choix pour l'équipe.

Aux thésards que j'ai (eu) la chance d'encadrer : Nicolas BARBEY, François ORIEUX, Caifang CAI, Yuling ZHENG et Long CHEN. Beaucoup de leur travail est présent dans ce document mais surtout ils m'ont appris à diriger des recherches.

A tous les membres du GPI qui sont passés dans la salle des machines avec moi Olivier, Mehdi, Aurélien, Boris, Fabrice, Patrice, Sofia, Nadia, Hachem, Diarra, Doriano, Sha, Ning, Thomas, Leila, Mircea. Merci de toutes ces discussions à travers lesquelles j'ai pu découvrir une partie des cultures des sept nationalités dont vous provenez.

Enfin, je remercie ma femme Cécile pour son soutien et ses relectures minutieuses de ce manuscrit.

Table des matières

Table des figures

Avant propos

Dans ce manuscrit, je vais décrire mes activités de recherche de ces neuf dernières années au sein du Groupe Problèmes Inverses.

Tout d'abord, j'exposerai un Curriculum Vitae assez succinct où mes encadrements et mes publications seront mise en avant.

Puis, je présenterai de manière non exhaustive mes travaux de recherche. Je commencerai par préciser ma démarche lorsque je suis confronté à la résolution d'un problème inverse (Chap. 2). Dans ce chapitre, je traiterai de l'inversion tomographique et j'illustrerai comment utiliser les principaux leviers qui permettent d'améliorer la qualité de l'estimation : l'amélioration du modèle direct, le choix pertinent de la paramétrisation des inconnues et la détermination de l'information *a priori* à prendre en compte. Dans le chapitre 3, j'exposerai ma contribution à la méthodologie bayésienne variationnelle, approche qui permet de résoudre efficacement des problèmes inverses comportant un grand nombre d'inconnues. J'illustrerai l'efficacité de l'approche en résolvant un problème estimation-séparation de composantes astrophysiques. J'exposerai la résolution de ce problème de manière détaillée afin d'appréhender le travail nécessaire à sa mise en œuvre. Cette méthode est le point de départ de la majorité de mes perspectives de recherche exposées dans le chapitre 4.

Dans la dernière partie, j'ai sélectionné quelques uns de mes articles afin de représenter de manière plus exhaustive que dans la partie II mes travaux de recherche. En effet, cette deuxième partie n'exposera ni mes travaux sur la déconvolution myope **[101]** [1] (reproduit à la page 137), ni ceux sur l'inversion sur résolue issue de données d'un spectro-imageur **[118]** (reproduit à la page 154), ni ceux sur la reconstruction sur-résolue des données issues de détecteurs infrarouge bolométriques **[102]** (reproduit à la page 119).

---

[1] Les références sur des travaux auxquels j'ai contribué sont en gras dans le texte

Première partie

Parcours Professionnel

Chapitre 1

Curriculum Vitae

1.1 État Civil

Thomas RODET
Laboratoire des Signaux et Systèmes
Supélec, Plateau de Moulon, 91192 Gif–sur–Yvette, Cedex
Tél. : 01 69 85 17 47
Mél. : Thomas.Rodet@lss.supelec.fr
Web : http ://thomas.rodet.lss.supelec.fr/

Né le 3 juin 1976
Nationalité française
Marié
Trois enfants

1.1.1 Situation actuelle

Maître de conférences à l'Université Paris-Sud 11, section CNU 61, affecté au L2S depuis septembre 2003.

1.2 Titres Universitaires

***1999 - 2002* :** Thèse de Doctorat en traitement du signal de l'Institut National Polytechnique de Grenoble (INPG), intitulée « Algorithmes rapides de Reconstruction en tomographie par compression des calculs. Application à la tomofluoroscopie », soutenue la 18 octobre 2002, devant le jury composé de :
  - Pierre Yves COULON, Professeur, INPG (Président)
  - Michel BARLAUD, Professeur, Université de Nice-Sophia Antipolis (Rapporteur)
  - Michel DEFRISE, Directeur de Recherche, Univ. Libre de Belgique (Rapporteur)
  - Laurent DESBAT, Professeur, Université Joseph Fourier (UJF), (Directeur de thèse)
  - Pierre GRANGEAT, Expert senior, Commissariat à l'Énergie Atomique, (CEA), (Co-directeur de thèse)
  - Françoise PEYRIN, Directrice de Recherche, CNRS, CREATIS, (Examinatrice)

***1999* :** Obtention du DEA «image de» l'Université Jean Monnet de Saint Étienne.

***1999* :** Ingénieur de l'École Supérieure de Chimie Physique Électronique de Lyon.

1.3 Parcours

***2011* :** Obtention d'un demi CRCT, séjour de 6 mois au Laboratoire Informatique Gaspard Monge (LIGM UMR 8049, Marne la Vallée).

***2002 - 2003* :** Étude post doctorale à la Vriej Universiteit Brussel.

1.4 Activités d'enseignement

- Coordinateur de l'équipe d'enseignants (2 Pr, 4 MdC, 1 ATER, 4 moniteurs) de traitement du signal depuis 2009.
- Création de deux nouveaux enseignements de traitement du signal pour des astrophysiciens (M2P et M2R).
- Création de projets d'informatique pluridisciplinaires (chimie des matériaux, traitement du signal, traitement de l'image).
- Responsable de quatre enseignements (gestion d'une équipe pédagogique de cinq personnes).
- Enseignement de traitement du signal , de probabilité, d'informatique (langages de programmation C et Java).
- Production de polycopiés :
  - Transformation d'un signal, théorie de l'échantillonnage, analyse temps-fréquence, transformée en ondelettes, M2 P IST Capteur (18 pages)
  - Outils de traitement du signal, transformée en ondelettes et problèmes inverses, M2R Astronomie et Astrophysique (52 pages).
  - Inversion de données en imagerie, M1 IST EU 453 (40 pages)
  - Initiation à la programmation : Scilab, L2 Phys 230 (21 pages)
  - C dans la poche, L3 PIST 306 (31 pages)
- J'enseigne dans des formations très variées en terme de public :
  - Deuxième année de Licence de Physique,
  - Troisième année de Licence de Physique parcours Information, Systèmes et Technologie (IST),
  - Première et deuxième année du Master IST,
  - Deuxième année du Master de Recherche Astronomie et Astrophysique,
  - Troisième et quatrième années à Polytech Paris-Sud,
  - École doctorale d'Astronomie et d'Astrophysique d'Ile de France.
  - Formation continue et dernière année formation initiale à Supélec.

1.5 Activités liées à l'administration

- Co-responsable du parcours L3 Physique IST, refonte de l'emploi du temps du premier semestre (2011-2012).
- Co-responsable du Parcours Réseau et Telecom du Master IST (2012-).
- Membre élu du conseil de perfectionnement de l'IFIPS de 2006 à 2010.
- Membre élu au conseil du laboratoire depuis janvier 2010.
- Membre élu à la CCSU 60/61/62 depuis 2010.
- Membre du bureau de la sous-CCSU 61 depuis 2010.
- Membre élu des CS-61 à l'UPS (*2004 - 2007*).
- Membre extérieur du comité de sélection 61 sur le poste MCF 566 à l'école centrale de Nantes (2011)
- Membre interne dans quatre comités de sélection 61 (sur les postes 209, 417, 2197, 4058).
- Membre du conseil scientifique du Groupement d'Intérêt Scientifique (GIS) « Signaux, Images et Commande optimale des procédés(SIC) » entre EDF et Supélec.
- Forte implication dans la définition de la politique informatique du L2S (définition d'un système commun pour les postes de travail, achat d'un cluster de 64 coeurs de calcul).

1.6 Activités liées à la recherche

- Responsable du Groupe Problèmes Inverses depuis juin 2008, le groupe est constitué de 6 permanents (1 DR CNRS, 1 CR CNRS, 4 MdC), 9 thésards, 1 ATER. Gestion du budget, animation scientifique, organisation de séminaires internes, entretien du parc informatique, accueil des nouveaux entrants, rédaction des rapports d'activité, demande de nouveaux postes, réponse aux appels à projets etc...
- Quatre jours d'échanges avec un parlementaire (le sénateur maire Michel HOUEL), et un membre de l'Académie des Sciences (Odile MACCHI). Le but était de sensibiliser les parlementaires au métier de chercheur et au rôle de la recherche en France.
- Président du Jury de la thèse de J. SCHMITT soutenue le 7/12/2011 devant le jury composé d'A. BIJAOUI (Rapporteur), G. FAY, I. GRENIER (Co-directeur), M. NGUYEN-VERGER (Rapporteur), T. RODET (Président) et J.-L. STARCK (Directeur de thèse).
- Titulaire de la Prime d'Excellence Scientifique (PES) depuis le 1^{er} octobre 2009.
- Organisation des séminaires d'audition informels pour les concours de Maître de Conférences et Professeur des Universités en mai 2009.
- Premier inventeur de deux brevets dont un étendu à l'international.
- Expert ANR : blanc, jeune chercheur et Astrid.
- Relecteur auprès de : IEEE Trans. on Medical Imaging, IEEE Trans. on Signal Processing, IEEE Journal of Selected Topics in Signal Processing, Optical Engineering, Applied Optics, Optics Express, The Journal of Electronic Imaging et The Journal of X-Ray Science and Technology.
- Relecteur pour les conférences : Gresti et ICIP.

1.7 Collaborations

- Très forte collaboration avec l'Institut d'Astrophysique Spatiale (IAS) (*depuis 2003*).
  Collaborateur privilégié de l'IAS dans les projets suivants :
  - Consultant en traitement du signal pour le Projet EUCLID.
  - Consultant en traitement du signal pour le consortium SPIRE de la mission Herschel ESA.
  - Expert en traitement de données pour le projet prospectif de détection d'eau dans l'atmosphère des planètes extra-solaires, projet Darwin.
  - Collaboration avec le groupe de physique solaire de l'IAS, le Centre National d'Études Spatiales (CNES) et la société Collecte Localisation Satellites (CLS) sur le projet STEREO.
- Collaboration avec le LIGM, dépôt d'une ANR blanche 2011.
- Collaboration avec l'ONERA Palaiseau dépôt d'une ANR blanche 2011.
- Collaboration avec Arcellor Mittal, dépôt d'une ANR industrielle 2011.
- Collaboration avec le LGEP projet Carnot C3S MillTER.
- Collaboration internationale avec la Vriej Universiteit Brussel (*depuis 2002*).
- Projet industriel avec Trophy (Kodak Dental System) (2011-).
- Participation au projet européen DynCT (1999-2002).

1.8 Encadrements

1.8.1 Post-doctorat et ATER.

1. Nom : Boujemaa AIT EL FQUIH
   Thème : Approches bayésiennes variationnelles récursives.
   Encadrement : T. RODET
   Cadre : ATER temps plein à l'Université Paris-Sud 11

Période : 01 septembre 2009, durée 12 mois
Situation actuelle : Post doctorant au CEA
Publications : [C4]

2. Nom : Hacheme AYASSO
   Thème : Inversion des données SPIRE/Herschel.
   Encadrement : A. Abergel (1/2) et T. RODET (1/2)
   Cadre : Post doctorant consortium SPIRE
   Période : 01 décembre 2010, durée 21 mois
   Situation actuelle : Maître de conférences à Grenoble (UJF)
   Publications : [A1,S2]

1.8.2 Doctorants

1. Nom : Nicolas BARBEY
   Titre : Détermination de la structure tridimensionnelle de la couronne solaire à partir d'images des missions spatiales SoHO et STEREO.
   Encadrement : F. Auchère (2/5), J.-C. Vial (1/5) et T. Rodet (2/5)
   Cadre : Thèse Cifre CNES- CLS
   Soutenance : le 19 novembre 2008 devant le jury composé de :
   – Alain ABERGEL, Professeur, Université Paris-Sud 11, (Président)
   – Jérôme IDIER, Directeur de Recherche, CNRS IRCCYN, (Rapporteur)
   – Craig DEFOREST, Chercheur, Southwest Research Institute, (Rapporteur)
   – Jean-Claude VIAL, Directeur de Recherche, CNRS (Directeur de thèse)
   – Frédéric AUCHÈRE, Astronome adjoint, Univ. Paris Sud 11 (Co-encadrant)
   – Thomas RODET, Maître de conférences, Univ. Paris Sud 11, (Co-encadrant)
   Situation actuelle : CDI dans une SSII
   Publications : [A4, C14, CN5]
2. Nom : François ORIEUX
   Titre : Inversion bayésienne myope et non-supervisée pour l'imagerie sur-résolue. Application à l'instrument SPIRE de l'observatoire spatial Herschel.
   Encadrement : A. Abergel (15%), J.-F. Giovannelli (50%) et T. Rodet (35%)
   Cadre : Allocation ministérielle
   Soutenance : le 16 novembre 2009 devant le jury composé de :
   – Alain ABERGEL, Professeur, Université Paris-Sud 11, (Co-encadrant)
   – Jean–François GIOVANNELLI, Professeur, Université Bordeaux 1 (Directeur de thèse)
   – Jérôme IDIER, Directeur de Recherche, CNRS IRCCYN, (Rapporteur)
   – Guy LE BESNERAIS, Chercheur, ONÉRA (Examinateur)
   – Thomas RODET, Maître de conférences, Université Paris-Sud 11, (Co-encadrant)
   – Sylvie ROCQUES, Directeur de Recherche, CNRS LATT, (Rapporteur)
   – Eric THIÉBAUT, Astronome adjoint, Observatoire de Lyon, (Examinateur)
   Situation actuelle : Ingénieur de Recherche à l'Institut d'Astrophysique de Paris
   Publications :[S4, A1, A2, C8, C10, C11, C12, CN4]
3. Nom : Caifang CAI
   Titre : Approches bayésiennes pour la reconstruction des fractions d'os et d'eau à partir de données tomographiques polychromatiques.
   Encadrement : A. Djafari (3/10), S. Legoupil (2/10) et T. Rodet (1/2)
   Cadre : thèse CEA-DGA
   Soutenance : prévue en novembre 2012
   Publications : [C2, C6]

4. Nom : Yuling ZHENG
   Titre : Nouvelles méthodes bayésiennes variationnelles.
   Encadrement : A. Fraysse (2/5) et T. Rodet (3/5)
   Cadre : Allocation ministérielle
   Soutenance : prévue en octobre 2014
   Publications : Néant
5. Nom : Long CHENG
   Titre : Approches statistiques dédiées à la reconstruction d'images de la mâchoire.
   Encadrement : N. Gac (1/5), C. Maury (1/5) et T. Rodet (3/5)
   Cadre : Cifre avec la société Trophy Dental System
   Soutenance : prévue en décembre 2014
   Publications : Néant

1.8.3 Stagiaires

1. Nom : Alex GILBERT
   Niveau : deuxième année de l'ESIEE (Marne la Vallée) (bac +4)
   Sujet : La reconstruction de cartes infrarouges étendues (mips/Spitzer).
   Encadrement : T. Rodet
   Période : mai 2004, 3 mois
   Publications : Néant
2. Nom : Guillaume BITTOUN
   Niveau : deuxième année de FIUPSO (Université Paris-Sud 11)
   Sujet : La reconstruction tomographique de plumes solaires.
   Encadrement : T. Rodet
   Période : mai 2005, 3 mois
   Publications : Néant
3. Nom : François ORIEUX
   Niveau : troisième année ESEO Angers
   Sujet : L'inversion de données issues du spectro-imageur (IRS/Spitzer).
   Encadrement : J.-F. Giovannelli (1/10) et T. Rodet (9/10)
   Période : mars 2006, 5 mois
   Publications : [A3]
4. Nom : Anne-Aelle DRILLIEN
   Niveau : M1 Physique et application (Université Paris-Sud 11)
   Sujet : Évaluation de la résolution d'un algorithme de sur résolution.
   Encadrement : T. Rodet
   Période : mai 2007, 3 mois
   Publications : Néant
5. Nom : Yuling ZHENG
   Niveau : L3 Physique parcours IST (Université Paris-Sud 11)
   Sujet : Approche bayésienne variationnelle appliquée à la déconvolution impulsionnelle.
   Encadrement : T. Rodet
   Période : septembre 2008, 3 mois
   Publications : [CN3]
6. Nom : Belkacem KHERRAB
   Niveau : M2R IST spécialité Automatique et Traitement du Signal et des Images (Université Paris-Sud 11)

Sujet : La résolution d'un problème de séparation de sources étendues, appliquée à la reconstruction 3D dynamique de l'atmosphère solaire à partir des données STEREO.
Encadrement : T. Rodet
Période : avril 2009, 5 mois
Publications : [C9]

7. Nom : Cyril DELESTRE
   Niveau : M1 IST (Université Paris-Sud 11)
   Sujet : Simulation d'une caméra teraHetz (infra-rouge) pour développer des systèmes de sécurité dans les aéroports.
   Encadrement : T. Rodet
   Période : mai 2010, 3 mois
   Publications : Néant
8. Nom : Yuling ZHENG
   Niveau : M1 IST (Université Paris-Sud 11)
   Sujet : Approches statistiques pour la réduction des artefacts métalliques.
   Encadrement : T. Rodet
   Période : mai 2010, 3 mois
   Publications : [C6]
9. Nom : Long CHEN
   Niveau : M2R IST Spécialité Automatique et Traitement du Signal et des Images
   Sujet : La reconstruction tomographique 3D en imagerie dentaire.
   Encadrement : N. Gac (3/10) et T. Rodet (7/10)
   Période : mars 2011, 6 mois
   Publications : Néant
10. Nom : Leila GHARSALLI
    Niveau : M2R Mathématiques / Vision / Apprentissage (ENS Cachan)
    Sujet : Les nouveaux algorithmes bayésiens variationnels accélérés en utilisant des outils d'analyse fonctionnelle.
    Encadrement : A. Fraysse (4/5) et T. Rodet (1/5)
    Période : avril 2011, 6 mois
    Publications : Néant

1.9 Liste de publications

Les références surmontées d'une étoile * concernent mon travail de thèse.

1.9.1 Articles de revues internationales avec comité de lecture

[A1] F. Orieux, J.-F. Giovannelli, **T. Rodet**, A. Abergel, H. Ayasso et M. Husson, « Superresolution in map-making based on physical instrument model and regularized inversion. Application to SPIRE/Herschel. », *Astronomy & Astrophysics*, vol. 539, n°A38, pp. 16, mars 2012.

[A2] F. Orieux, J.-F. Giovannelli et **T. Rodet**, « Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution », *Journal of the Optical Society of America (A)*, vol. 27, n°7, pp. 1593–1607, 2010.

[A3] **T. Rodet**, F. Orieux, J.-F. Giovannelli et A. Abergel, « Data inversion for over-resolved spectral imaging in astronomy », *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, n°5, pp. 802–811, octobre 2008.

[A4] N. Barbey, F. Auchère, **T. Rodet** et J.-C. Vial, « A time-evolving 3D method dedicated to the reconstruction of solar plumes and results using extreme ultra-violet data », *Solar Physics*, vol. 248, n°2, pp. 409–423, avril 2008.

[A5] G. Lagache, N. Bavouzet, N. Fernandez-Conde, N. Ponthieu, **T. Rodet**, H. Dole, M.-A. Miville-Deschênes et J.-L. Puget, « Correlated anisotropies in the cosmic far-infrared background detected by the multiband imaging photometer for spitzer : First constraint on the bias », *The Astrophysical Journal Letters*, vol. 665, n°2, pp. 89–92, août 2007.

[A6] **T. Rodet**, F. Noo et M. Defrise, « The cone–beam algorithm of Feldkamp, Davis, and Kress preserves oblique line integral », *Medical Physics*, vol. 31, n°7, pp. 1972–1975, juillet 2004.

[A7] H. Kudo, **T. Rodet**, F. Noo et M. Defrise, « Exact and approximate algorithms for helical cone-beam CT », *Physics in Medicine and Biology*, vol. 49, n°13, pp. 2913–2931, juillet 2004.

[A8]* **T. Rodet**, P. Grangeat et L. Desbat, « Multichannel algorithm for fast reconstruction », *Physics in Medicine and Biology*, vol. 47, n°15, pp. 2659–2671, August 2002.

[A9]* P. Grangeat, A. Koenig, **T. Rodet** et S. Bonnet, « Theoretical framework for a dynamic cone–beam reconstruction algorithm based on a dynamic particle model », *Physics in Medicine and Biology*, vol. 47, n°15, pp. 2611–2625, August 2002.

1.9.2 Articles de revues internationales à paraître ou soumis

[S1] A. Fraysse et **T. Rodet**, « A measure-theoretic variational bayesian algorithm for large dimensional problems. », Soumis à *SIAM Imaging Sciences*, 36 pages, 2012.

[S2] H. Ayasso, **T. Rodet** et A. Abergel, « A variational bayesian approach for unsupervised super-resolution using mixture models of point and smooth sources applied to astrophysical mapmaking », en révision dans *Inverse Problems*, 33 pages, 2012.

[S3] C. Cai, **T. Rodet**, S. Legoupil et A. Mohammad-Djafari, « Bayesian reconstruction of water and bone fractions in polychromatic multi-energy computed tomography », Soumis à *Medical Physics*, 2012.

[S4] F. Orieux, J.-F. Giovannelli, **T. Rodet** et A. Abergel, « Estimation of hyperparameters and instrument parameters in regularized inversion. illustration for SPIRE/Herschel map making. », Soumis à *Astronomy & Astrophysics*, 15 pages, 2012.

1.9.3 Communications internationales dans des congrès avec comité de lecture et actes

[C1] E. Dupraz, F. Bassi, **T. Rodet** et M. Kieffer, « Distributed coding of sources with bursty correlation », in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, Kyoto, Japan, 2012, pp. 2973 – 2976.

[C2] C. Cai, A. Mohammad-Djafari, S. Legoupil et **T. Rodet**, « Bayesian data fusion and inversion in X–ray multi–energy computed tomography », in *Proceedings of the International Conference on Image Processing*, septembre 2011, pp. 1377 – 1380.

[C3] A. Fraysse et **T. Rodet**, « A gradient-like variational bayesian algorithm », in *Proceedings of IEEE Statistical Signal Processing Workshop (SSP)*, Nice, France, juin 2011, n°S17.5, pp. 605 – 608.

[C4] B. Ait el Fquih et **T. Rodet**, « Variational bayesian Kalman filtering in dynamical tomography », in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, Prague, mai 2011, n°Id 2548, pp. 4004 - 4007.

[C5] M. Kowalski et **T. Rodet**, « An unsupervised algorithm for hybrid/morphological signal decomposition », in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, Prague, mai 2011, n°Id 2155, pp. 4112 – 4115.

[C6] Y. Zheng, C. Cai et **T. Rodet**, « Joint reduce of metal and beam hardening artifacts using multi-energy map approach in X-ray computed tomography », in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, mai 2011, n°Id 3037, pp. 737–740.

[C7] P. Guillard, **T. Rodet**, S. Ronayette, J. Amiaux, A. Abergel, V. Moreau, J.-L. Auguères, A. Bensalem, T. Orduna, C. Nehmé, A. R. Belu, E. Pantin, P.-O. Lagage, Y. Longval, A. C. H. Glasse, P. Bouchet, C. Cavarroc, D. Dubreuil et S. Kendrew, « Optical performance of the JWST/MIRI flight model : characterization of the point spread function at high resolution », in *Space Telescopes and Instrumentation 2010 : Optical, Infrared, and Millimeter Wave (Proceedings Volume)*, San Diego, California, USA, juillet 2010, vol. 7731.

[C8] F. Orieux, **T. Rodet** et J.-F. Giovannelli, « Instrument parameter estimation in bayesian convex deconvolution », in *Proceedings of the International Conference on Image Processing*, Hong Kong, Hong-Kong, décembre 2010, pp. 1161 – 1164.

[C9] B. Kherrab, **T. Rodet** et J. Idier, « Solving a problem of sources separation stemming from a linear combination : Applied to the 3D reconstruction of the solar atmosphere », in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, Dallas, USA, 2010, pp. 1338 – 1341.

[C10] F. Orieux, J.-F. Giovannelli et **T. Rodet**, « Deconvolution with gaussian blur parameter and hyperparameters estimation », in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, Dallas, USA, 2010, pp. 1350 – 1353.

[C11] **T. Rodet**, F. Orieux, J.-F. Giovannelli et A. Abergel, « Data inversion for hyperspectral objects in astronomy », in *Proc. of IEEE International Workshop on Hyperspectral Image and Signal Processing (WHISPERS 2009)*, Grenoble France, août 2009.

[C12] F. Orieux, **T. Rodet** et J.-F. Giovannelli, « Super–resolution with continuous scan shift », in *Proceedings of the International Conference on Image Processing*, Cairo, Egypt, novembre 2009, pp. 1169 – 1172.

[C13] Y. Grondin, L. Desbat, M. Defrise, **T. Rodet**, N. Gac, M. Desvignes et S. Mancini, « Data sampling in multislice mode PET for multi-ring scanner », in *Proceedings of IEEE Medical Imaging Conference*, San Diego, CA, USA, 2006, vol. 4, pp. 2180–2184.

[C14] N. Barbey, F. Auchère, **T. Rodet**, K. Bocchialini et J.-C. Vial, « Rotational Tomography of the Solar Corona-Calculation of the Electron Density and Temperature », in *SOHO 17 - 10 Years of SOHO and Beyond*, H. Lacoste, Ed., Noordwijk, juillet 2006, ESA Special Publications, pp. 66–68, ESA Publications Division.

[C15] A. Coulais, J. Malaizé, J.-F. Giovannelli, **T. Rodet**, A. Abergel, B. Wells, P. Patrashin, H. Kaneda et B. Fouks, « Non-linear transient models and transient corrections methods for IR low-background photo-detectors », in *Astronomical Data Analysis Software & Systems 13*, Strasbourg, octobre 2003.

[C16] H. Kudo, F. Noo, M. Defrise et **T. Rodet**, « New approximate filtered backprojection algorithm for helical cone-beam CT with redundant data », in *Proceedings of IEEE Medical Imaging Conference*, octobre 2003, vol. 5, pp. 3211 – 3215.

[C17]⋆ **T. Rodet**, L. Desbat et P. Grangeat, « Parallel algorithm based on a frequential decomposition for dynamic 3D computed tomography », in *International Parallel and Distributed Processing Symposium (IPDPS 2003)*, Nice, France, 2003, n°237, p. 7.

[C18]⋆ **T. Rodet**, P. Grangeat et L. Desbat, « A fast 3D reconstruction algorithm based on the computation compression by prediction of zerotree wavelet », in *Fully 3D Reconstruction in Radiology and Nuclear Medicine*, Saint Malo, France, juin 2003, vol. 7, pp. 135–138.

[C19] **T. Rodet**, F. Noo, M. Defrise et H. Kudo, « Exact and approximate algorithms for helical cone–beam CT », in *Fully 3D Reconstruction in Radiology and Nuclear Medicine*, Saint Malo, France, juin 2003, n°Mo-AM1-3, pp. 17–20.

[C20] **T. Rodet**, J. Nuyts, M. Defrise et C. Michel, « A study of data sampling in PET with planar detectors », in *Proceedings of IEEE Medical Imaging Conference*, Portland, Oregon, USA, novembre 2003, n°M8-8.

[C21]* **T. Rodet**, P. Grangeat et L. Desbat, « Multifrequential algorithm for fast reconstruction », in *Fully 3D Reconstruction in Radiology and Nuclear Medicine*, Pacific Grove CA ,USA, 2001, vol. 6, pp. 17–20.

[C22]* P. Grangeat, A. Koenig, **T. Rodet** et S. Bonnet, « Theoretical framework for a dynamic cone–beam reconstruction algorithm based on a dynamic particle model », in *Fully 3D Reconstruction in Radiology and Nuclear Medicine*, 2001, pp. 171–174.

[C23]* **T. Rodet**, P. Grangeat et L. Desbat, « A new computation compression scheme based on a multifrequential approach », in *Proceedings of IEEE Medical Imaging Conference*, 2000, vol. 15, pp. 267–271.

1.9.4 Communications nationales dans des congrès avec comité de lecture et actes

[CN1] M. Kowalski et **T. Rodet**, « Un algorithme de seuillage itératif non supervisé pour la décomposition parcimonieuse de signaux dans des unions de dictionnaires », in *Actes du 23^{e} colloque GRETSI*, Bordeaux, France, septembre 2011, p. ID46.

[CN2] A. Fraysse et **T. Rodet**, « Sur un nouvel algorithme bayésien variationnel », in *Actes du 23^{e} colloque GRETSI*, Bordeaux, France, septembre 2011, p. ID91.

[CN3] **T. Rodet** et Y. Zheng, « Approche bayésienne variationnelle : application à la déconvolution conjointe d'une source ponctuelle dans une source étendue. », in *Actes du 22^{e} colloque GRETSI*, septembre 2009, p. ID432.

[CN4] F. Orieux, **T. Rodet**, J.-F. Giovannelli et A. Abergel, « Inversion de données pour l'imagerie spectrale sur-résolus en astronomie », in *Actes du 21^{e} colloque GRETSI*, Troyes, septembre 2007, pp. 717–720.

[CN5] N. Barbey, F. Auchère, **T. Rodet** et . J.-C. Vial, « Reconstruction tomographique de séquences d'images 3D. Application aux données SOHO/STEREO », in *Actes du 21^{e} colloque GRETSI*, septembre 2007, pp. 709–712.

[CN6] **T. Rodet**, A. Abergel, H. Dole et A. Coulais, « Inversion de données infrarouges issues du télescope Spitzer », in *Actes du 20^{e} colloque GRETSI*, septembre 2005, vol. 1, pp. 133–136.

[CN7]* **T. Rodet**, P. Grangeat et L. Desbat, « Algorithme rapide de reconstruction tomographique basé sur la compression des calculs par ondelettes », in *Actes du 19^{e} colloque GRETSI*, Paris, France, 2003, p. ID297.

[CN8]* **T. Rodet**, P. Grangeat et L. Desbat, « Reconstruction accélérée par compression de calculs utilisant une approche fréquentielle », in *Actes du 18^{e} colloque GRETSI*, Toulouse, France, 2001, p. ID168.

1.9.5 Brevets

[B1] **T. Rodet**, P. Grangeat et L. Desbat, « Procédé de reconstruction accéléré d'une image tridimensionnelle », Brevet Français n° 00 07278 0, CEA, Juin 2000.

[B2] **T. Rodet**, P. Grangeat et L. Desbat, « Procédé de reconstruction accéléré d'une image à partir d'un jeu de projections par application d'une transformée en ondelettes », Brevet Français n° 02 12925 2, CEA, octobre 2002.

1.9.6 Thèse

[Th1]* **T. Rodet**. *Algorithmes rapides de reconstruction en tomographie par compression des calculs. Application à la tomofluoroscopie 3D*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, octobre 2002.

Deuxième partie

Synthèse des travaux de recherches

Chapitre 2

Tomographie

2.1 Introduction

Un nom qui recouvre des problématiques variées :

La tomographie est un terme très général qui regroupe un ensemble très varié de problématiques. Il n'est donc pas facile de trouver une définition satisfaisante. Une définition possible serait la suivante : la tomographie consiste à estimer certaines propriétés physiques d'une partie d'un objet en utilisant toutes sortes d'ondes pénétrantes. Faire un panel de toutes les problématiques tomographiques est un travail que nous n'aborderons pas dans ce document. Nous traiterons uniquement les problèmes de tomographie où la propriété mesurée est intégrée suivant des droites, et plus particulièrement l'inversion de la transformée en rayons X et de la transformée en rayons divergents. Nous commencerons par les définitions de ces deux transformées dans $\mathbb{R}^n$. Puis nous présenterons leurs inversions analytiques, leurs inversions dans un cadre probabiliste et le lien entre ces deux cadres de résolution.

2.1.1 Cadre général et définitions

Transformée de Radon : La transformée de Radon Rf d'une fonction intégrable $f \in L^1(\mathbb{R}^n)$ s'écrit comme l'intégrale selon des hyperplans de $\mathbb{R}^n$, $\forall \boldsymbol{\theta} \in S^{n-1}$ et $\forall s \in \mathbb{R},\ \Pi(\boldsymbol{\theta}, s) = \{\boldsymbol{r} \in \mathbb{R}^n,\ \boldsymbol{r} \cdot \boldsymbol{\theta} = s\ \}$, où S^{n-1} est l'hypersphère unité de $\mathbb{R}^n$.

$$Rf(\boldsymbol{\theta}, s) = \int_{\boldsymbol{u} \in \Pi(\boldsymbol{\theta}, s)} f(s\boldsymbol{\theta} + \boldsymbol{u}) \mathrm{d}\boldsymbol{u} \tag{2.1}$$

Transformée en rayons X : Définissons la transformée en rayons X, $\mathcal{X}f$, d'une fonction $f \in L^1(\mathbb{R}^n)$ comme un ensemble d'intégrales sur des droites.

$$\mathcal{X}f(\boldsymbol{\theta}, \boldsymbol{u}) = \int_{-\infty}^{+\infty} f(\boldsymbol{u} + t\boldsymbol{\theta}) \mathrm{d}t \tag{2.2}$$

$\mathcal{X}f(\boldsymbol{\theta}, \boldsymbol{u})$ est l'intégrale de f suivant la droite de vecteur directeur $\boldsymbol{\theta} \in S^{n-1}$ passant par $\boldsymbol{u} \in \boldsymbol{\theta}^\perp$, où $\boldsymbol{\theta}^\perp$ est l'hyperplan de $\mathbb{R}^n$ de vecteur normal $\boldsymbol{\theta}$.

Transformée en rayons divergents : Définissons la transformée en rayons divergents, $\mathcal{D}f$, d'une fonction $f \in L^1(\mathbb{R}^n)$ comme un ensemble d'intégrales sur des demi-droites. L'intégrale de f suivant la demi-droite issue de $\boldsymbol{a}$ et ayant pour vecteur directeur $\boldsymbol{\theta} \in S^{n-1}$ s'écrit,

$$\mathcal{D}f(\boldsymbol{a}, \boldsymbol{\theta}) = \int_{0}^{+\infty} f(\boldsymbol{a} + t\boldsymbol{\theta}) dt. \tag{2.3}$$

On s'intéressera particulièrement à cette transformée car c'est elle qui décrit le plus de systèmes médicaux. En effet, dans ces systèmes, la source de rayonnement peut souvent être considérée comme ponctuelle. Elle est représentée par $\boldsymbol{a}$ dans l'Eq. (2.3). De plus, un système d'acquisition permet de mesurer simultanément l'intégrale de l'objet f sur un ensemble de demi-droites. La transformée en rayons divergents dépend de la trajectoire de $\boldsymbol{a}$, nous allons considérer dans la suite que cette trajectoire est $\boldsymbol{a}(\cdot)$ une courbe paramétrée par le scalaire λ.

Cas particulier d'un objet bi-dimensionnel (2D) $n = 2$ **:**

Dans ce cas le vecteur unitaire $\boldsymbol{\theta}$ se réduit à un angle $\theta \in [0, 2\pi]$ et la transformée en rayons X est confondue avec la transformée de Radon. Pour reconstruire exactement f, il faut mesurer Rf pour tous les couples (θ, s) appartenant au domaine $[0, \pi] \times \mathbb{R}$, [112].

Pour ce qui est de la transformée en rayons divergents, la reconstruction dépend de la trajectoire. Lorsque $\boldsymbol{a}(\lambda)$ suit une trajectoire circulaire autour de l'objet (trajectoire la plus utilisée en 2D), les conditions pour une reconstruction exacte consiste à faire un peu plus d'un tour « Short Scan » (pour plus de détails voir [106] **[115]** [1]). Lorsque l'on veut reconstruire uniquement une partie d'un objet ou un objet allongé, on peut utiliser moins d'un demi tour c'est ce qu'ont montré F. NOO *et al* dans [97].

Cas particulier d'un objet tridimensionnel (3D) $n = 3$ **:**

Ce cas est celui que nous développerons principalement dans la suite de ce manuscrit. En effet, les objets que l'on veut imager en pratique, des patients, des pièces pour le contrôle non destructif, l'atmosphère du soleil, sont tous des objets 3D.

À ma connaissance, il n'y a pas de système qui permette de mesurer directement la transformée de Radon pour $n = 3$, car l'acquisition des intégrales de l'objet 3D selon des plans, posent de nombreux problèmes. On s'intéressera uniquement à la transformée en rayons X et à la transformée en rayons divergents. On rencontre la transformée en rayons X lorsque l'on résout des problèmes de Tomographie à Émission de Positron (TEP). Dans ce cas, la reconstruction exacte est mathématiquement possible si les données vérifient les conditions d'Orlov [103], c'est à dire si le vecteur $\boldsymbol{\theta}$ décrit un grand cercle [2].

Comme dans le cas 2D, c'est la transformée en rayons divergents qui correspond à la plus grande diversité d'applications : scanners médicaux, imagerie dentaire, imagerie de l'atmosphère solaire, scanner pour des petits animaux, appareil de contrôle non destructif etc... Pour cette transformée, l'obtention d'une reconstruction théoriquement exacte est possible si les données respectent la condition de Tuy [136]. Pour vérifier cette condition il faut que pour chaque point de l'objet à reconstruire tous les plans qui passent par ce point coupent la trajectoire de la source en au moins un point. On parle alors de trajectoire complète.

Au delà des conditions de reconstruction exacte se pose aussi le problème du choix des fréquences d'échantillonnage de la transformée de Radon. Ce problème est très bien décrit dans l'habilitation à diriger des recherches de L. DESBAT [43]. Nous y avons contribué **[117][67]** dans le cas de l'imagerie TEP en faisant une étude sur l'impact de l'utilisation de schémas d'échantillonnage efficace introduits dans [42], et en particulier l'étude de la robustesse au bruit de mesure de ces approches.

2.1.2 Inversion tomographique analytique

Inversion de la transformée en rayons X

Si l'on a acquis des données sur un grand cercle (vérifiant les conditions Orlov), le problème d'inversion de la transformée en rayons X 3D $(n = 3)$, peut se factoriser comme N problèmes d'inversion

---

[1] Les références sur des travaux auxquels j'ai contribué sont en gras dans le texte

[2] cercle dont le plan générateur contient le centre de la sphère

de la transformée en rayons X 2D ($n = 2$). De plus, nous avons vu dans le § 2.1.1 que cette transformé en rayons X 2D était équivalente à la transformée de Radon (TR) 2D. Son inversion a été introduite par J. RADON en 1917 [112].

$$f(r_1, r_2) = \frac{1}{2}(2\pi)^{-1} \int_0^{2\pi} \int_{\mathbb{R}} e^{2j\pi\sigma(r_1 \cos\theta + r_2 \sin\theta)} |\sigma| \widetilde{Rf}(\theta, \sigma) \mathrm{d}\sigma \mathrm{d}\theta \tag{2.4}$$

avec $\widetilde{Rf}$ la transformée de Fourier de la transformée de Radon par rapport au deuxième paramètre $(Rf(\theta, \cdot))$.

Nous exposons ci-dessus un algorithme de reconstruction de type filtrage rétroprojection. Dans l'Eq. (2.4), l'intégrale en σ et le produit par $|\sigma|$ correspondent à une opération de filtrage par un filtre « passe haut » (la fréquence nulle est mise à zéro et les hautes fréquences sont amplifiées). L'intégrale en θ correspond à l'étape de rétroprojection. Il existe d'autres types d'algorithmes analytiques de reconstruction comme l'approche rétroprojection puis filtrage ou la reconstruction dans le domaine de Fourier [17]. Ces autres approches sont moins répandues car elles sont plus coûteuses en temps de calcul pour une sensibilité au bruit souvent plus importante.

Pour implanter les approches sur des calculateurs, on doit discrétiser l'Eq. (2.4). On échantillonne alors uniformément la dimension angulaire θ et les dimension spatiales s, r_1 et r_2. Les intégrales sont remplacées par des sommes. L'opération de filtrage est aussi discrétisée. Elle est très souvent réalisée dans le domaine de Fourier car le noyau de convolution est de grande taille. Enfin, on discrétise l'opération de rétroprojection. De nombreux travaux portent sur l'accélération et la discrétisation des opérateurs de projection [125] [75] et de rétroprojection [59][120][35], nous évoquerons ceux-ci dans la partie 2.1.3.

Inversion de la transformée en rayons divergents pour $n = 3$ (3D)

La trajectoire circulaire : La trajectoire la plus naturelle est bien sûr la trajectoire circulaire. Mais cette trajectoire ne vérifie pas les conditions de Tuy. Les reconstructions que l'on peut faire seront donc mathématiquement approchées. L'approche la plus répandue pour cette trajectoire est l'algorithme FDK introduit par FELDKAMP DAVIS et KRESS [48]. Bien que les reconstructions soient approchées nous avons montré **[116]** (reproduit à la page 184) que les intégrales sur les lignes obliques les plus proches de l'axe perpendiculaire à la trajectoire étaient exactes.

La trajectoire hélicoïdale : Dans ce cas la courbe paramétrée est définie de la façon suivante

$$\boldsymbol{a}(\lambda) = \begin{pmatrix} R\cos\lambda \\ R\sin\lambda \\ h\frac{\lambda}{2\pi} \end{pmatrix},$$

avec R le rayon de l'hélice, λ un angle en radian et h son pas («pitch» en anglais). Cette trajectoire permet une reconstruction exacte, et elle est mise en œuvre facilement comme la composition d'une trajectoire circulaire avec une translation uniforme. De ce fait, c'est la trajectoire complète au sens de Tuy la plus utilisée. Déterminer une formule exacte de type filtrage rétroprojection est quelque chose de très difficile. En effet, certains plans passant par un point $\boldsymbol{r}$ fixé coupent trois fois la trajectoire[3] alors que d'autres ne la coupent qu'une fois. Cela conduit indirectement à ce que la transformée de Radon 3D est mesurée une ou trois fois selon les plans considérés. C'est seulement en 2002 que A. KATSEVICH a trouvé un algorithme de reconstruction exacte [78]. Son principe est de trouver des directions de filtrage particulières qui permettent de pondérer les plans par -1 ou 1. Cette pondération des plans permet de résoudre la dissymétrie de redondance entre les différents plans et ainsi d'avoir une formule d'inversion exacte. La formule d'inversion associée à l'algorithme de Katsevich est la suivante en utilisant les

---

[3] uniquement la portion la plus faible permettant de faire la reconstruction appelée le PI-segment[33]

notations de l'article de H. KUDO **[83]** :

$$f(\boldsymbol{r}) = \int_{\Lambda(\boldsymbol{r})} \frac{1}{R - r_1 \cos\lambda - r_2 \sin\lambda} g^F(\boldsymbol{r}, \lambda, U(\boldsymbol{r}, \lambda), V(\boldsymbol{r}, \lambda)) \mathrm{d}\lambda \tag{2.5}$$

où , $\boldsymbol{r} = (r_1, r_2, r_3)^{\mathrm{t}}$, $\Lambda(\boldsymbol{r}) = [\lambda_b(\boldsymbol{r}), \lambda_t(\boldsymbol{r})]$ est le PI-segment [33] correspondant au point $\boldsymbol{r}$. C'est l'unique segment sur un tour d'hélice où les deux vecteurs $(\overrightarrow{\boldsymbol{a}(\lambda_b), \boldsymbol{r}})$ et $(\overrightarrow{\boldsymbol{r}, \boldsymbol{a}(\lambda_t)})$ sont colinéaires. De plus $U(\boldsymbol{r}, \lambda)$ et $V(\boldsymbol{r}, \lambda)$ sont les coordonnées détecteur de l'intersection entre la droite issue de $\boldsymbol{a}(\lambda)$ et passant par $\boldsymbol{r}$ et le plan détecteur. L'intégrale de l'Eq. (2.5) correspond à l'opération de rétroprojection. g^F correspond aux données filtrées

$$g^F(\boldsymbol{r}, \lambda, u', v') = \frac{1}{2\pi} \int_{\mathbb{R}} h(u' - u) g^D(\lambda, u, v' - (u' - u) \tan\eta(\lambda, \boldsymbol{r})) \mathrm{d}u \tag{2.6}$$

avec $h(t)$ le filtre associé à la transformée de Hilbert, et $v' - (u' - u) \tan\eta(\lambda, \boldsymbol{r})$ définissant la ligne de filtrage permettant de résoudre le problème de redondance associé à l'acquisition des mesures dans la géométrie hélicoïdale. Enfin g^D correspond à la dérivée des projections,

$$g^D(\lambda, u, v) = \frac{R}{\sqrt{R^2 + u^2 + v^2}} \left(\frac{\partial}{\partial\lambda} + \frac{R^2 + u^2}{R^2} \frac{\partial}{\partial u} + \frac{uv}{R} \frac{\partial}{\partial v} \right) \mathcal{D}f(\boldsymbol{a}(\lambda), u, v) \tag{2.7}$$

Ce travail a été une grande avancée pour la communauté. Par contre toutes les données acquises ne sont pas utilisées. C'est pourquoi, il y a eu beaucoup de travaux sur les méthodes approchées [82][77][132] même après 2002 pour concevoir des méthodes qui utilisent toutes les données acquises et qui sont donc moins sensibles au bruit [16]**[83]** (reproduit à la page 190).

Autres trajectoires : Il existe aussi des travaux sur d'autres trajectoires qui sont complètes au sens de Tuy, comme par exemple la trajectoire cercle + droite [135]. Mais cette dernière, du fait de sa dissymétrie, rend les reconstructions instables dans certaines zones.

En revanche la trajectoire «saddle» qui a une forme de selle de cheval ou de chips a de bonnes propriétés en particulier pour l'imagerie cardiaque [104].

2.1.3 Inversion statistique

On considère ici que les données collectées, $\boldsymbol{y} \in \mathbb{R}^M$, sont discrètes et que l'objet à reconstruire, bien que continu, est représenté par un nombre fini de coefficients de base.

$$f(\boldsymbol{r}) = \sum_i x_i \phi_i(\boldsymbol{r}),$$

où x_i représente le i^{e} coefficient de décomposition associé à la fonction de base ϕ_i. L'ensemble des ϕ_i forme une base ou une famille génératrice de $L^2(\mathbb{R})$. Dans la plupart des cas, les fonctions ϕ_i sont des indicatrices de pixels ou de voxels, mais on peut aussi utiliser des fonctions Kaiser Bessel [85][89] ou des Gaussiennes **[118]**. On obtient dans tous les cas une relation entre des grandeurs discrètes, les données $\boldsymbol{y}$ et les coefficients de décomposition $\boldsymbol{x} = (x_1, \ldots, x_i, \ldots, x_N)^{\mathrm{t}}$,

$$\boldsymbol{y} = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{b}. \tag{2.8}$$

On désigne par $\mathcal{H}$ le modèle direct, $\boldsymbol{x} \in \mathbb{R}^N$ les coefficients et par $\boldsymbol{b}$ le bruit. Le modèle direct peut être linéaire ou non en fonction des cas. On étudiera des modèles directs non linéaires dans le cadre de l'imagerie médicale dans le § 2.2. Le modèle direct ne représente pas strictement la réalité physique du mode d'acquisition. On utilise souvent des modèles simplifiés afin de faciliter la mise en œuvre et pour que le calcul de la sortie modèle $\mathcal{H}(\boldsymbol{x})$ soit suffisamment rapide. Par conséquent, le bruit $\boldsymbol{b}$ modélise l'incertitude sur les mesures due à certaines propriétés physiques des détecteurs mais aussi aux erreurs causées par l'utilisation d'un modèle simplifié.

Modèle direct linéaire : résolution algébrique du problème d'inversion

Rappelons tout d'abord que $\boldsymbol{y} \in \mathbb{R}^M$ et que $\boldsymbol{x} \in \mathbb{R}^N$. De plus, on suppose que le modèle direct qui relie les inconnues $\boldsymbol{x}$ au données $\boldsymbol{y}$ est linéaire. Dans l'Eq. (2.8) l'opérateur $\mathcal{H}$ peut donc être remplacé par une matrice que nous noterons $\boldsymbol{H}$ la résolution du problème inverse dépend de la structure de cette dernière.

- si p est le rang de la matrice $\boldsymbol{H}$ et si $p = N = M$, alors l'opérateur $\mathcal{H}$ est bijectif. Alors, la solution du problème inverse s'écrit,
  $$\hat{\boldsymbol{x}} = (\boldsymbol{H})^{-1}\boldsymbol{y}. \tag{2.9}$$
  Cette configuration n'arrive pratiquement jamais dans le cas de la tomographie, car le nombre d'inconnues ne correspond pratiquement jamais au nombre de données.
- si $p > M$, alors il n'y a pas existence de la solution quelles que soient les données. On détermine la solution qui minimise la distance quadratique entre $\boldsymbol{Hx}$ et $\boldsymbol{y}$. La solution minimise le critère des moindres carrés,
  $$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{Hx}\|^2 \tag{2.10}$$
  Simplement en annulant le gradient du critère ci-dessus, on détermine le minimum du critère,
  $$\hat{\boldsymbol{x}} = (\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{t}}\boldsymbol{y}. \tag{2.11}$$
- si $p < M$, alors il n'y a pas unicité de la solution, afin de déterminer une solution parmi toutes celles qui vérifient $\boldsymbol{y} = \boldsymbol{Hx}$, on choisit la solution de norme minimale. Cette solution s'obtient en résolvant le problème d'optimisation sous contrainte suivant,
  $$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|^2 \qquad s.c \quad \boldsymbol{y} = \boldsymbol{Hx}. \tag{2.12}$$
  La solution s'obtient facilement en construisant un critère étendu en utilisant des multiplicateurs de Lagrange afin de tenir compte de la contrainte. Le calcul mène à la formule d'inversion
  $$\hat{\boldsymbol{x}} = \boldsymbol{H}^{\mathrm{t}}(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}\boldsymbol{y}. \tag{2.13}$$

Ces approches qui étudient la matrice associée au problème direct sont très utiles, mais il y a une part d'arbitraire lorsque l'on détermine l'estimateur. Par exemple dans le cas où $p > M$ pourquoi choisit -on la distance quadratique ? et dans le cas où $p < M$ pourquoi choisit on la solution de norme minimale ? Pour répondre à ces questions naturelles nous allons utiliser l'inférence bayésienne.

Inférence bayésienne :

La formalisation bayésienne du problème d'inférence (qui consiste à construire un raisonnement à partir d'hypothèses plausibles) décrit toute notre connaissance du processus de formation des données et notre connaissance *a priori* sur l'objet recherché par des probabilités [40][41][71][72]. Le problème d'inférence sur les grandeurs qui nous intéressent est équivalent à déterminer la fonction densité de probabilité *a posteriori*. Cette fonction est déterminée en utilisant la règle de Bayes.

Soit C toutes les informations (hypothèses) utilisées pour arriver à notre modélisation du problème. La règle de Bayes s'écrit :
$$p(\boldsymbol{x}|\boldsymbol{y}, C) = \frac{p(\boldsymbol{y}|\boldsymbol{x}, C)p(\boldsymbol{x}|C)}{p(\boldsymbol{y}|C)} \tag{2.14}$$

Comme nous ne connaissons pas de manière exacte le processus de formation de la mesure $\boldsymbol{y}$ nous intégrons toute l'information que l'on possède sur ce processus de formation à l'aide de la probabilité conditionnelle des mesures sachant la grandeur d'intérêt $p(\boldsymbol{y}|\boldsymbol{x}, C)$. Cette grandeur est aussi appelée la vraisemblance du $\boldsymbol{x}$. Notre problème inverse comme son nom l'indique consiste à extraire l'information que l'on possède sur la grandeur physique connaissant les mesures, c'est à dire la fonction densité de probabilité *a posteriori* $p(\boldsymbol{x}|\boldsymbol{y}, C)$. Cette fonction résume toute l'information que l'on possède sur l'objet que l'on cherche. Par contre, on est très souvent intéressé par un estimateur ponctuel, permettant de résumer cette fonction par une valeur particulière. L'estimation ponctuelle est très réductrice par rapport à la richesse de l'information contenue dans la loi *a posteriori*, c'est pourquoi on préfère souvent déterminer la valeur de l'estimateur ponctuel ainsi que la valeur de variance *a posteriori*. Dans ce chapitre nous nous intéresserons principalement à l'estimateur de Maximum *a posteriori* (MAP), pour des raisons de commodité de calcul.

2.1.4 Étude empirique sur les similitudes entre les approches analytiques et les approches statistiques

Pour mener à bien cette étude nous nous sommes placés dans la configuration la plus simple possible. Nous résolvons un problème tomographique 2D où les fonctions de base ϕ_i sont des indicatrices de pixel. L'opération de projection de l'objet est en géométrie parallèle, et le modèle direct considéré est linéaire. Enfin, nous allons traiter des problèmes de tailles réduites afin de faciliter les calculs.

Lien entre les approches rétroprojection puis filtrage et la solution des moindres carrés

Analysons tout d'abord les opérations effectuées lorsque l'on calcule la solution des moindres carrés à l'aide de l'Eq. (2.11). L'opération $\boldsymbol{H}\boldsymbol{x}$ correspond à l'opération de projection de l'objet 2D dans une géométrie projective parallèle[4] (transformée de Radon). L'opération rétroprojection correspond à l'application de la matrice $\boldsymbol{H}^{\mathrm{t}}$. Le calcul de l'estimateur des moindres carrés peut être vu comme une opération de rétroprojection puis une opération linéaire (multiplication par la matrice $(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H})^{-1}$). Dans le cadre des reconstructions analytiques, la majorité des approches consistent à filtrer les données puis à les rétroprojeter, mais on peut très bien commencer par l'opération de rétroprojection puis appliquer l'opération de filtrage. Dans ce cas, le filtre est bi-dimensionnel et à symétrie de révolution. Il faut maintenant analyser la structure de la matrice $(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H})^{-1}$ afin de voir si elle correspond à une opération de filtrage.

Afin de calculer facilement l'inverse de la matrice $\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H}$, nous avons choisi de reconstruire une image de petite taille (32×32 pixels). De plus, pour que cette dernière soit bien conditionnée nous avons un nombre important de données (151 projections composées de 139 valeurs), la matrice $\boldsymbol{H}$ est donc rectangulaire et elle comporte 20989 lignes et 1024 colonnes. La figure 2.1 représente les coefficients de la matrice $(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H})^{-1}$ (1024×1024), les points blancs représentent les valeurs positives, les points noirs les valeurs négatives et le gris correspond aux valeurs proches de zéro. On remarque que la structure de cette matrice est Toeplitz. On peut en conclure qu'on peut approcher l'application de cette matrice par une opération de filtrage. Nous allons maintenant extraire une ligne de la matrice $(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H})^{-1}$ afin d'étudier la réponse impulsionnelle du filtre de reconstruction (voir Fig. 2.2 (a)). Cette réponse impulsionnelle est bi-dimensionnelle et à première vue sa moyenne est quasiment nulle. En prenant le module de la transformée de Fourier de cette réponse impulsionnelle (voir Fig. 2.2 (b)), on remarque qu'elle est à symétrie de révolution et qu'elle correspond à un filtre « passe haut ». Ces propriétés sont identiques au filtre de reconstruction $h(\nu_1, \nu_2) = \sqrt{\nu_1^2 + \nu_2^2}$ de l'approche rétroprojection puis filtrage. On remarque sur la Fig. 2.2 (b) que la valeur du filtre pour la fréquence nulle est différente de zéro ce qui permet de conserver la moyenne de l'objet à reconstruire. On peut en conclure qu'il y a pratiquement équivalence dans ce cas entre l'approche analytique et la solution algébrique.

---

[4] Nous avons utilisé dans l'expérimentation numérique le projecteur parallèle codé sous matlab par la fonction `radon`.

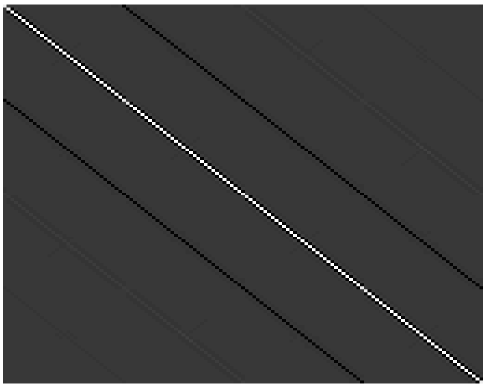
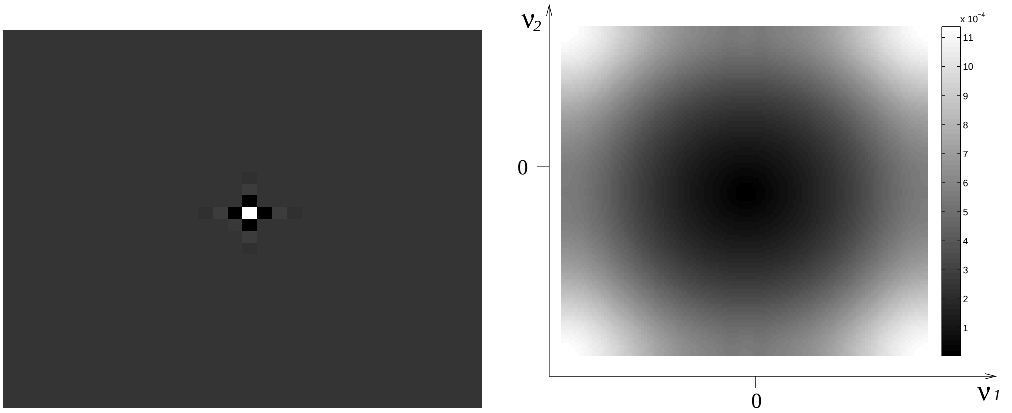


FIGURE 2.1 – Visualisation des coefficients de la matrice $(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H})^{-1}$



(a) (b)

FIGURE 2.2 – (a) Réponse impulsionnelle du filtre de reconstruction (b) Transformée de Fourier de la réponse impulsionnelle

Lien entre les approches filtrage puis rétroprojection et la solution de norme minimale

La solution de norme minimale consiste à faire une opération linéaire sur les données (application de la matrice $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$) puis de rétroprojeter le résultat. Nous allons maintenant décrire l'expérience menée afin d'identifier l'opération effectuée lorsque nous appliquons la matrice $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$: comme précédemment, nous avons choisi une configuration où la matrice $\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}}$ est de petite taille pour être facilement inversible. Nous avons un nombre de mesures réduit (529), celles-ci sont composées de 23 angles de projections uniformément répartis sur $[0, \pi[$. Afin que la matrice $\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}}$ soit bien conditionnée nous avons défini une image relativement grande ($151 \times 151 = 22801$ pixels[5]). La figure 2.3 représente les coefficients de la matrice $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$. On remarque visuellement que cette matrice peut être approchée par une matrice Toeplitz. L'obtention de la solution de norme minimale correspond à un algorithme de type filtrage (multiplication par la matrice $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$ puis rétroprojection (multiplication par la matrice $\boldsymbol{H}^{\mathrm{t}}$). Nous avons extrait une ligne de la matrice $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$ afin de visualiser la réponse impulsionnelle du filtre (voir Fig. 2.4 (a)) et le module de sa transformée de Fourier (voir Fig. 2.4 (b)). Dans ce cas la réponse impulsionnelle est très différente des approches filtrage rétroprojection. En effet, dans ces approches on applique le même filtre rampe (dont la transformée de Fourier est égale à $|\sigma|$) sur toutes les projections de manière identique. De manière schématique, on corrige l'effet passe bas dû à l'intégration de la transformée de Radon par un filtrage « passe haut » (filtre rampe) sur la variable spatiale s de la

---

[5] Un nombre impaire de pixels dans l'images permet de limiter les symétrie du problème de projection permet d'avoir une matrice $\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}}$ ayant un meilleur conditionnement.

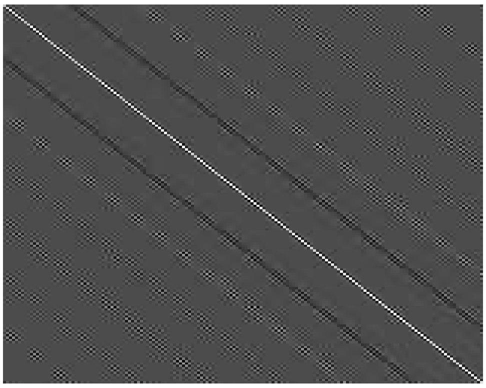
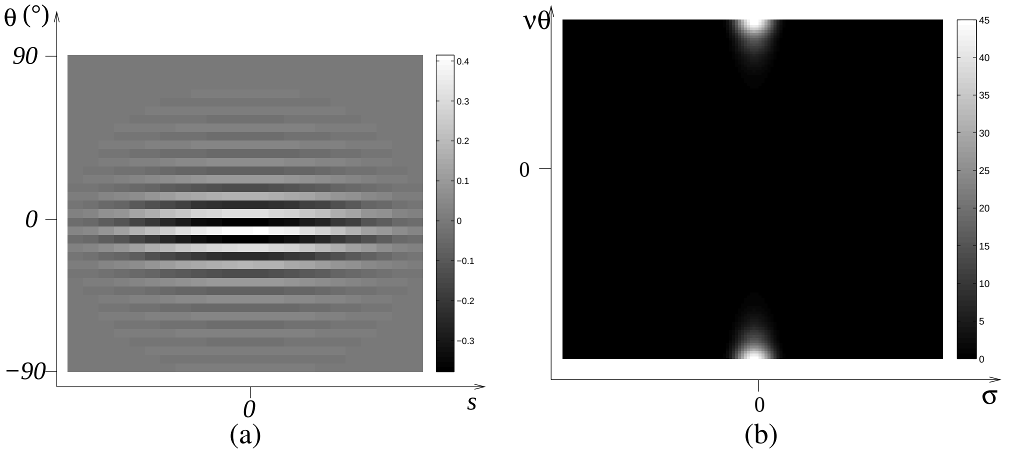


FIGURE 2.3 – Visualisation des coefficients de la matrice $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$



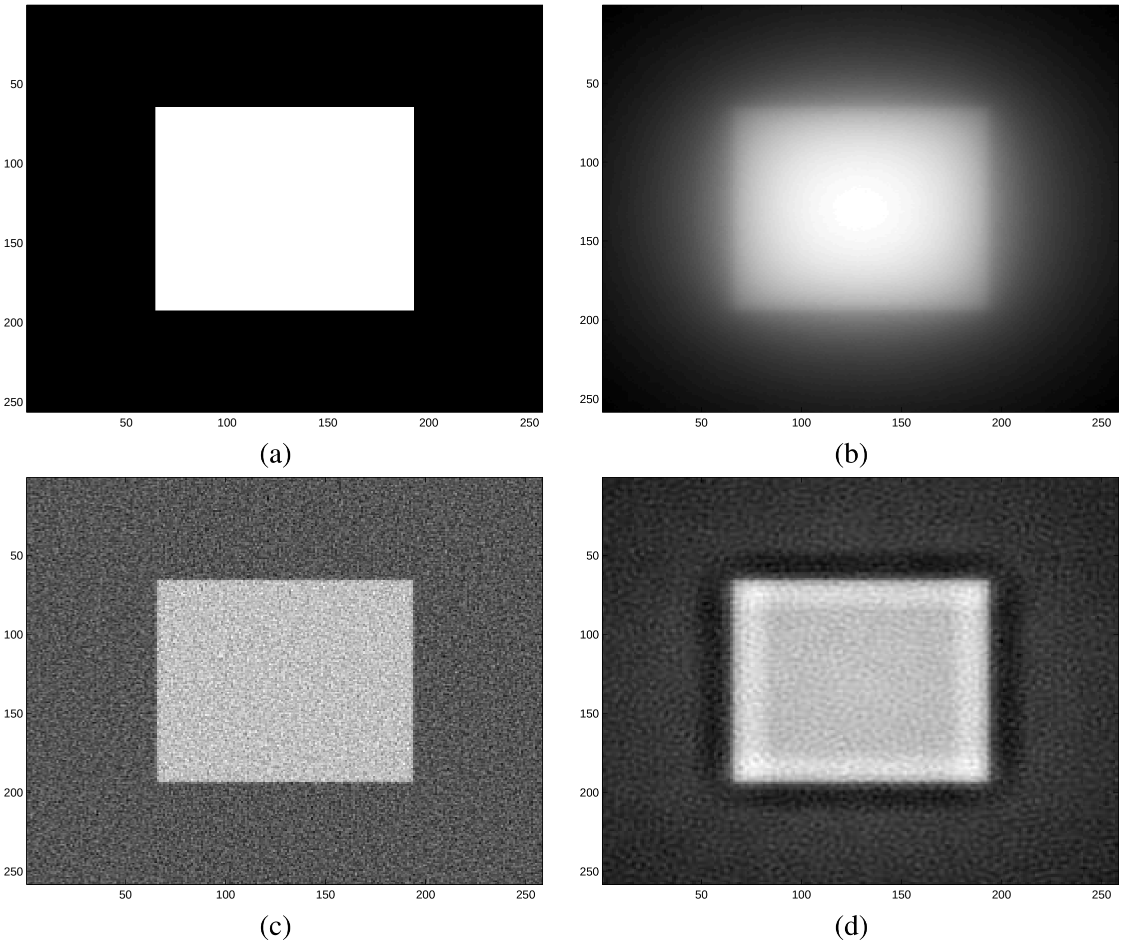
(a) (b)

FIGURE 2.4 – (a) Réponse impulsionnelle du filtre appliquée aux données. On utilise les mêmes notations que dans la définition de la transformée de Radon (TR) Eq. (2.1) : l'axe des abscisses représente le paramètre de position s et l'axe des ordonnées représente l'angle d'incidence θ (b) Transformée de Fourier de la réponse impulsionnelle

transformée de Radon (TR). Comme on peut le voir sur les Fig. 2.4 (a) et (b) l'opération de filtrage effectuée par l'application de la matrice $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$ est totalement différente. Le filtre est bi-dimensionnel, c'est un « passe bas » selon la direction spatiale de la TR et c'est un « passe haut » selon la direction angulaire de la TR. En conclusion, on retrouve bien une opération de filtrage comme dans la reconstruction classique de type FBP, mais le filtrage passe haut se fait sur l'axe angulaire et non sur la dimension spatiale des projections.

Étude de l'impact des phénomènes de bord sur la structure des matrices

Le filtre de la Fig. 2.4 a été obtenu sur un très faible nombre de projections afin de pouvoir facilement inverser la matrice. On pourrait penser que le filtre obtenu est le fruit d'effets de bord vu le nombre réduit de données utilisées pour le générer. C'est pourquoi nous avons utilisé ce filtre pour reconstruire des données de taille plus importante. On fait l'expérience numérique suivante : on simule des projections suivant 256 angles avec un détecteur comprenant 362 cellules détectrices, en projetant une image composée de 256×256 pixels. L'image qui nous a servi pour simuler les données est représentée sur la Fig. 2.5 (a). Les projections en géométrie parallèle ont été obtenues grâce à la fonction matlab `radon` et nous avons ajouté un bruit Gaussien i.i.d. de variance égale à 20 (SNR 24,3 dB).



(a) (b)

(c) (d)

FIGURE 2.5 – (a) Image vraie (b) rétroprojection uniquement (c) inversion de la transformée de Radon (filtre rampe) (d) rétroprojection filtrée en utilisant le filtre identifié à partir d'un petit problème.

On compare les approches suivantes : uniquement l'opération de rétroprojection (Fig. 2.5 (b)), la reconstruction avec l'algorithme classique FBP (fonction matlab `iradon`) en utilisant uniquement le filtre rampe (Fig. 2.5 (c)) et une approche filtrage rétroprojection en utilisant le filtre de la Fig. 2.4 (a) (Fig. 2.5 (d)). En comparant les Figs 2.5 (b) et (d), on voit clairement que le filtre garde des propriétés de reconstruction et que l'image de la Fig. 2.5 (d) possède des structures plus hautes fréquences que uniquement l'opération de rétroprojection. De plus lorsqu'on compare les résultats de reconstruction de l'approche FBP avec le nouveau filtre on observe un effet régularisant qui est dû à l'opération passe bas sur les projections. Il est normal d'avoir un effet régularisant car on a déduit le filtre à partir d'une solution de norme minimale.

Conclusion :

Cette petite étude nous a permis de voir qu'il y avait un lien assez fort dans le cas de la géométrie parallèle entre les approches analytiques et les approches algébriques. Cette étude est plus une illustration pédagogique qu'un véritable travail de recherche, mais elle permet d'ouvrir des perspectives de recherches intéressantes. Il serait très utile d'avoir des approches intermédiaires entre les approches analytiques et les approches algébriques afin d'avoir des approches plus précises que les approches analytiques et moins coûteuses en temps de calcul que les approches algébriques. Pour cela, il faudrait étudier les matrices $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$ lorsque la géométrie est divergente, lorsque la source ne peut pas être considérée comme ponctuelle, *etc*.... À partir de ces études on déterminerait de nouveaux filtres de reconstruction qui pourraient tenir compte partiellement d'une source de rayons X non ponctuelle.

2.1.5 Discussion

Nous évoquerons ici les avantages et les inconvénients des approches analytiques et statistiques à travers leurs applications à l'imagerie médicale en rayons X, car ce domaine couvre la majorité des problèmes rencontrés dans l'inversion tomographique.

Apport des approches analytiques

Les approches analytiques restent les méthodes de référence dans pratiquement tous les systèmes médicaux. Elles sont relativement faciles à mettre en œuvre. De plus, elles sont très efficaces en terme de temps de calcul, et assez économes en ressource mémoire. En effet, les approches analytiques nécessitent une étape de filtrage puis une étape de rétroprojection. La complexité dans le cas d'une reconstruction 2D d'une image $N \times N$ à partir de mesures $M \times N$ est en $O(MN \log_2 N)$ opérations élémentaires et la complexité de la rétroprojection est en $O(MN^2)$ opérations élémentaires. De plus, ces algorithmes sont relativement peu coûteux en terme de ressources mémoires car on peut traiter les données séquentiellement. La mémoire vive utilisée correspond seulement à la taille de l'image ou du volume que l'on veut reconstruire et les données uniquement à un ou deux angles différents. Avec l'évolution des détecteurs, l'imagerie est principalement 3D et le nombre de mesures est croissant. De plus, le temps de reconstruction doit être le plus court possible pour recevoir un plus grand nombre de patients. Ces approches sont très adaptées lorsque le modèle direct est une intégration d'une propriété physique (modèle linéaire), que l'échantillonnage angulaire est régulier, et que la trajectoire de la source de rayons X est complète au sens de Tuy.

Problèmes de discrétisation des formules analytiques

Des artefacts apparaissent sur les reconstructions lorsque les conditions exposées ci-dessus ne sont pas remplies. Si le modèle direct reste linéaire mais que la matrice $(\boldsymbol{H}\boldsymbol{H}^{\mathrm{t}})^{-1}$ n'a plus une structure invariante, on peut voir apparaître des artefacts pour les raisons suivantes :

- Les données ont une couverture angulaire limitée (moins de 180°) : par exemple la tomosynthèse [13].
- L'échantillonnage angulaire est non uniforme.
- Il y a des cellules détectrices défectueuses.
- Le vieillissement des détecteurs conduit à une modification des gains détecteurs, qui provoque des artefacts en anneau.

Problèmes lorsque le processus de formation des données n'est plus linéaire

Des artefacts apparaissent aussi lorsque le modèle direct n'est plus linéaire. Dans le cas de l'imagerie à rayons X, la loi de Beer Lambert (2.15) n'est pas linéaire à cause de l'exponentielle.

$$I(\lambda, p) = I_s e^{-\int_{\mathcal{L}(\lambda,p)} \mu(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}} \tag{2.15}$$

avec $\mathcal{L}(\lambda, p)$ le segment de droite issu de la source $\boldsymbol{a}(\lambda)$ et passant par la cellule détectrice p, μ l'atténuation des matériaux traversés et I_s l'énergie émise par la source de rayons X. Afin de linéariser le problème, on prend le logarithme des mesures, mais cette opération modifie les propriétés statistiques du bruit, ce qui amplifie très fortement ce dernier lorsque l'on est en présence d'objets de forte atténuation comme nous le verrons dans le § 2.2.1.

Une autre cause de non linéarité provient du fait que les sources de rayons X ne sont pas monochromatiques sauf cas particulier (rayonnement synchrotron). L'hypothèse d'une source monochromatique faite par les approches analytiques provoque des artefacts de durcissement de spectre. Nous verrons plus précisément dans le § 2.2.2 comment les approches statistiques peuvent partiellement s'affranchir de ce problème.

Les approches analytiques supposent aussi que l'objet est immobile pendant l'acquisition. Le mouvement de l'objet pendant l'acquisition des projections provoque des artefacts importants. Nous verrons dans le § 2.3 que dans le cadre de l'inversion tomographique de données issues de la couronne solaire, on peut résoudre ce problème lorsque « le mouvement » a une structure particulière.

Difficulté pour intégrer des informations *a priori* aux approches analytiques

Les problèmes tomographiques sont mal posés au sens d'Hadamard. L'inversion a tendance à amplifier le bruit présent dans les données. Pour palier ce problème, il faut introduire de l'information *a priori*. Malheureusement dans les approches analytiques, la seule information *a priori* exploitable, est la douceur (ou la régularité spatiale) de la solution. Cette régularité est imposée à l'aide d'opérations de filtrage sur la reconstruction (post filtrage) ou directement sur les données avant d'appliquer le filtre de reconstruction (pré-traitement).

À l'inverse les approches statistiques peuvent tenir compte d'une plus grande variété d'information *a priori*. Par exemple, on peut favoriser les solutions régulières par morceaux en utilisant un *a priori* introduisant de la parcimonie sur le gradient de l'objet (Variation Totale), ou favoriser les objets composés d'un nombre connu de matériaux différents à l'aide d'une loi *a priori* formée d'un mélange de Gaussiennes. Ces *a priori* permettent de mieux préserver les contours présents dans les images qu'un *a priori* de douceur tout en réduisant le bruit dans les reconstructions.

Limitations des approches statistiques

Il existe bien des facteurs qui limitent la qualité des images reconstruites, mais dans la majorité des problèmes tomographiques difficiles (forte non linéarité, données avec un mauvais SNR, couverture angulaire limitée) les approches statistiques fournissent des images de meilleure qualité que celles analytiques.

Le principale verrou qui limite l'utilisation de ces approches est le temps de calcul qui est très important. Comme l'estimateur est calculé de manière itérative, que chaque itération nécessite deux à trois fois plus d'opérations élémentaires qu'une inversion analytique et qu'il faut entre 150 et 3000 itérations pour arriver à la convergence, le temps de calcul est plus important de 2 à 3 ordres de grandeur. Le temps de calcul est vraiment le talon d'Achile de ces approches.

Les raisons de la démocratisation des approches statistiques dans les systèmes d'imagerie médicale

Il y a principalement deux raisons au développement de ces approches :

- La démocratisation dans les pays développés du recours à l'imagerie médicale, entraîne une hausse du nombre d'examens par patient. L'irradiation de la population est en constante augmentation. Si cette augmentation persiste dans les années futures, on pourrait être confronté à un problème de santé publique. La réduction drastique des doses délivrées aux patients est donc un sujet de première importance pour l'imagerie à rayons X. Les approches statistiques permettent d'avoir une meilleur robustesse au bruit de mesures que les approches analytiques.
- Comme nous venons de le voir le temps de calcul des approches statistiques est trop important pour que les systèmes d'imagerie médicale soient économiquement viables.
  Le développement de la puissance de calcul des ordinateurs associés à l'utilisation de FPGA pour la reconstruction tomographique [110] permettent d'obtenir des temps de calcul plus raisonnables.

De plus, le développement des cartes graphiques conçues pour générer des images de synthèse à haut débit afin de permettre aux joueurs d'être plongés dans un monde tridimensionnel, peuvent être utilisés désormais pour faire du calcul scientifique. Ces cartes sont de réels calculateurs possédant jusqu'à 1500 cœurs de calcul. Le problème du rendu d'une scène 3D est similaire à un problème de projection tomographique : on a une ou plusieurs sources ponctuelle de lumière et on veut former une image 2D à partir d'une scène 3D. Comme la géométrie projective est très semblable entre les deux problèmes, on peut utiliser de manière très efficace les cartes graphiques afin de calculer la projection et la rétroprojection tomographique [59, 120]. Ces cartes graphiques sont des calculateurs ayant une architecture de type pipeline permettant de traiter un flot très important de données. Á l'inverse d'un super calculateur la structure particulière des cartes graphiques ne permet pas d'avoir de bonnes performances sur tous les problèmes de calculs scientifiques. Par contre, l'énorme avantage des cartes graphiques est qu'elles sont bon marché et que le coût de fonctionnement est relativement faible (pas besoin de salle climatisée et insonorisée).

2.2 Deux exemples illustrant l'apport des approches statistiques lorsque le modèle de formation des données est non linéaire

2.2.1 Artefacts causés par la linéarisation de la loi de Beer Lambert

On se place dans cette partie dans le cas où la source de rayons X est monochromatique et les photons possèdent tous l'énergie ϵ_0. Dans ce cas l'Eq. (2.15) devient :

$$I(\lambda, s) = I_s(\epsilon_0) e^{-\int_{\mathcal{L}(\lambda,s)} \mu(\epsilon_0, \boldsymbol{r}) \mathrm{d}\boldsymbol{r}} \tag{2.16}$$

De plus les mesures sont entachées de bruit :

$$I_m(\lambda, s) = I(\lambda, s) + \eta(\lambda, s). \tag{2.17}$$

Le bruit $\eta(\lambda, p)$ modélise un ensemble de phénomènes variés. Tout d'abord les photons X ont une certaine probabilité d'être absorbés par le scintillateur[6], ce phénomène conduit à ce que les données suivent une loi de Poisson. De plus, l'électronique de lecture et la quantification conduisent à commettre des erreurs. De manière générale ces sources d'erreurs sont souvent modélisées par un bruit Gaussien. Il faut aussi ajouter toutes les erreurs de modélisation dues aux hypothèses simplificatrices : la non prise en compte de la diffusion et la connaissance imparfaite de I_s l'intensité de la source de rayons X. Il y a deux causes à la mauvaise connaissance de I_s : elle est mesurée avec un détecteur imparfait et elle peut varier au cours de l'acquisition. La complexité et la diversité des phénomènes perturbateurs nous conduisent à penser que la modélisation du bruit devrait se faire au prix d'une loi très complexe. En pratique, on utilise des modèles avec des lois simples pour réduire la complexité des approches.

On voit dans l'Eq. (2.16) que la partie intégrale est une opération linéaire que nous allons modéliser par une matrice après discrétisation. La non linéarité de la loi Beer Lambert vient de l'exponentielle. L'approche classique consiste à linéariser le problème. Pour cela, un pré-traitement détermine l'atténuation du faisceau de rayons X à partir de l'énergie déposée par les photons détectés :

$$y(\lambda, s) = \log\left(\frac{I_s}{I_m(\lambda, s)}\right). \tag{2.18}$$

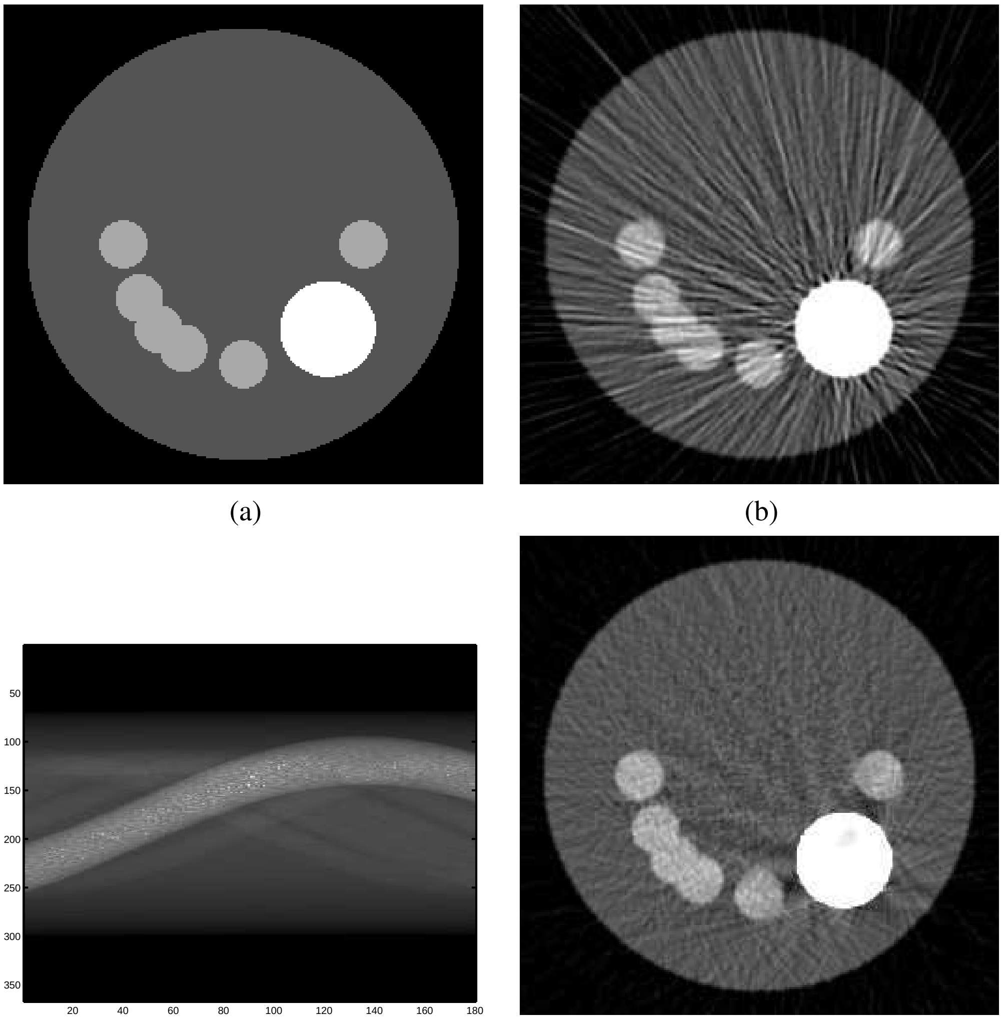
Ce pré-traitement entraîne des artefacts en étoile lorsqu'il y a une zone dans l'image possédant une forte atténuation (voir figure 2.6 (b)). Ensuite, on considère le modèle suivant :

$$\begin{aligned} y(\lambda, s) &= \int_{\mathcal{L}(\lambda,s)} \mu(\epsilon_0, \boldsymbol{r}) \mathrm{d}\boldsymbol{r} + b(\lambda, s) && (2.19)\\ &= \mathcal{D}f(\boldsymbol{a}(\lambda), s) + b(\lambda, s), && (2.20) \end{aligned}$$

---

[6] Il est souvent composé d'un cristal qui absorbe les photons X et qui réémet des photons visibles

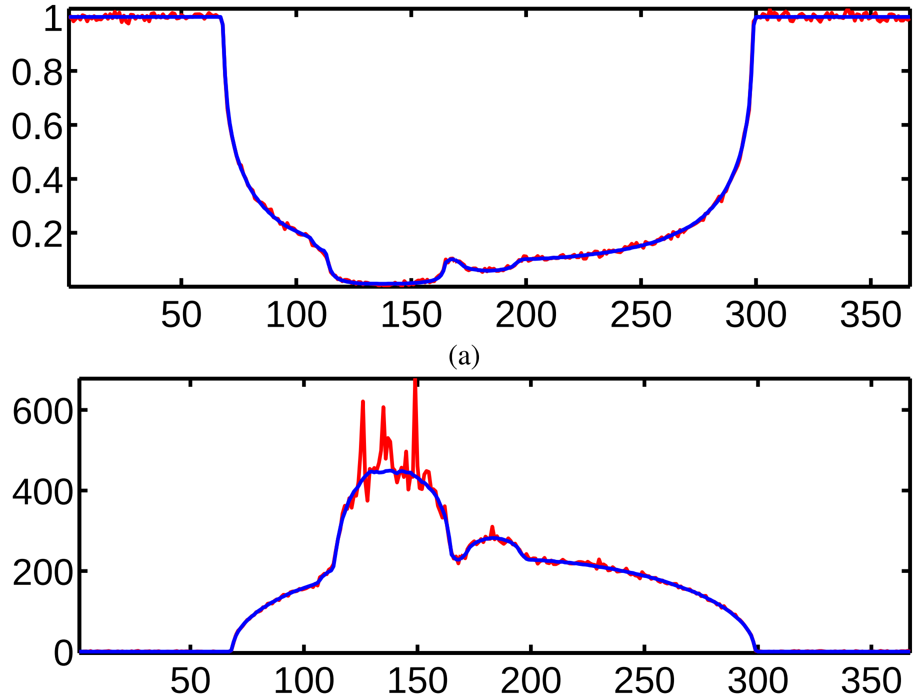
Dans les approches analytiques toutes les données ont le même poids ce qui correspond à un bruit $b(\lambda, s)$ Gaussien indépendant identiquement distribué (i.i.d.) dans une interprétation statistique. D'autres approches statistiques considèrent que $y(\lambda, s)$ suit une loi de Poisson. Dans les deux cas, on néglige l'effet du pré-traitement sur la statistique du bruit. Le problème apparaît (voir Fig. 2.6 (b) et (c)) lorsque que les flux de photons $I(\lambda, s)$ sont faibles. Dans ce cas, d'une part l'énergie du bruit peut être supérieure au signal d'intérêt et d'autre part les faibles énergies mesurées correspondent à des fortes atténuations. Il y a donc une amplification importante du bruit causée par le pré-traitement non linéaire. Ce phénomène est illustré sur la Fig. 2.7. Nous avons fait l'expérience numérique suivante : on a simulé des projections parallèles de l'objet 2D de la Fig. 2.6 (a) auxquelles nous avons ajouté du bruit Poissonnien et du bruit Gaussien d'un niveau relativement faible (Fig. 2.6 (c)). Le résultat du pré-traitement est visualisé sur la Fig. 2.7 (b). On voit très clairement qu'autour du pixel 150 le bruit a été très fortement amplifié et on remarque sur le tracé d'une projection (Fig. 2.7 (a)) que cela correspond à un niveau de flux faible. À l'inverse autour du pixel 25 (abscisse de la Fig. 2.7 (b)) le niveau du bruit est inchangé voir atténué par le pré-traitement et on remarque que cela correspond à une zone de fort flux sur la Fig. 2.7 (a).



(a) (b)

(c) (d)

FIGURE 2.6 – (a) Image vraie utilisée pour faire les simulations (b) Reconstruction de données pré-traitées avec l'algorithme FBP (c) Représentation sous forme d'un sinogramme de $\boldsymbol{y}$ (données pré-traitées) (d) Minimum du critère des moindres carrés régularisés où les données pré-traitées les plus bruitées ont été supprimées.



(a)

(b)

FIGURE 2.7 – Tracés d'une projection pour un angle arbitrairement fixé. En bleu simulation sans bruit. En rouge simulation avec bruit (a) mesure de l'énergie déposée sur le détecteur (b) données après pré-traitement Eq. (2.18).

Réduction de ces artefacts

La réduction de ces artefacts passe par la prise en compte de la loi Beer Lambert Eq. (2.16) et par une modélisation assez fine du bruit η.

Détermination du modèle d'incertitude sur les mesures : On remarque dans la littérature que la modélisation retenue pour le bruit est Poissonnienne [98][99][46][45] dans une grande majorité des approches. Ce choix semble légitime car le processus de détection des photons est Poissonnien. Mais on peut opposer deux arguments à ce choix : premièrement dans la tomographie à rayons X il y a beaucoup de photons qui frappent les détecteurs et le processus Poissonnien peut être approché très précisément par une Gaussienne ; deuxièmement, comme nous l'avons déjà exposé dans le § 2.2.1, les sources d'incertitude sur les mesures sont multiples, la loi résultante n'est donc ni Gaussienne, ni Poissonnienne. C'est pourquoi, dans le § 2.2.2 nous avons choisi de modéliser l'incertitude sur les mesures par un bruit blanc Gaussien donc la variance dépend du flux mesuré par le détecteur. De ce fait on tire parti de la simplicité de la loi Gaussienne tout en ayant une modélisation plus fine que les approches utilisant une loi de Poissonnien. Notons qu'il existe un faible nombre de travaux utilisant un modèle composite associant un bruit Poissonnien et un bruit Gaussien [54].

Différentes stratégies d'inversion : Schématiquement on peut diviser l'état de l'art en deux grandes familles : les approches qui cherchent à inverser le problème non-linéaire et les approches qui gardent le pré-traitement mais qui vont essayer de tenir compte de l'amplification du bruit en pondérant ou en éliminant des données extrêmement bruitées.

L'inversion du problème non-linéaire avec un bruit Poissonnien est un problème d'optimisation complexe. L'optimisation directe n'est pas réalisable. C'est pourquoi, dans [45][46] les auteurs utilisent une famille de fonctions majorantes plus facilement minimisables (les "surogate") pour approcher itérativement le critère cible. Dans le même esprit, J. NUYTS [99][98][36][37] fait une approximation quadratique séparable du critère puis il utilise une approche de descente de gradient pour minimiser ce nouveau critère. Il est bien évident que, comme dans l'approche précédente, le critère quadratique approchant est modifié à chaque itération. Nous avons contribué dans **[144]** en inversant le problème non-linéaire avec un bruit Gaussien. Dans ce cas nous avons une expression explicite du gradient du critère. Nous avons développé un algorithme de descente de type gradient conjugué. Par contre, nous avons été obligés de faire une approximation quadratique du critère afin de déterminer un pas de descente efficace. Enfin, dans le cadre de la thèse de Caifang CAI nous avons résolu le même problème avec un bruit Gaussien dont la variance est différente pour chaque donnée et nous estimons ces variances au cours de l'algorithme itératif **[20]**.

La deuxième famille d'approches consiste à pré-traiter les données en tenant compte de l'amplification du bruit. Une première approche consiste à éliminer les données qui sont trop corrompues par le bruit. De manière générale ces données correspondent à des objets très fortement atténuants dont la reconstruction avec précision n'est pas très utile. Pour ce faire on applique un seuil sur les données pré-traitées. Nous avons appliqué cette approche sur les données simulées de la Fig. 2.6. Dans ce cas, on obtient des résultats assez satisfaisants (voir Fig. 2.6 (d)). Mais cette approche dépend d'un seuil qui n'est pas facile à régler quand les images ont une structure complexe (images réelles), de plus il reste des artefacts résiduels (voir Fig. 2.6 (d)). Dans [53][134] les auteurs affectent une confiance différente à chaque mesure pré-traitée en fonction de leur amplitude. Pour cela, on construit une vraisemblance Gaussienne qui possède une variance différente pour chaque mesure y. Cette vraisemblance correspond à un critère des moindres carrés pondérés [53] ("Weighted Least Square"). Cette approche est plus souple que la précédente car elle ne nécessite pas de décision brutale. La qualité de reconstruction de cette approche est un petit peu plus faible que les approches visant à inverser le problème non linéaire, par contre, comme le critère à optimiser est quadratique, les algorithmes utilisés sont plus efficaces et plus facilement parallèlisables.

2.2.2 Résolution du problème de durcissement de spectre de la source de rayons X

Les artefacts de durcissement de spectre sont particulièrement difficiles à éliminer par des approches classiques de traitement d'image car ils sont aux mêmes fréquences que les signaux d'intérêt. Toutes les approches de filtrage sont donc inefficaces. Un des artefacts symptomatique du durcissement de spectre est l'artefact « coupe ». Il se manifeste de la façon suivante : au lieu de reconstruire un objet ayant un niveau uniforme, la reconstruction présente une atténuation plus faible au centre de l'objet et plus forte sur les bords. Ce phénomène provient du fait que les générateurs de rayons X produisent des spectres d'émission continus de zéro jusqu'à la tension d'accélération des électrons. Les approches classiques de reconstruction supposent que l'on reconstruit l'image à l'énergie moyenne du spectre d'émission. On sait tous que les photons de faible énergie sont beaucoup plus atténués que ceux possédant une énergie importante, lorsqu'un rayon traverse le bord de l'objet, son spectre d'émission est pratiquement inchangé et l'atténuation est correctement estimée. Par contre lorsqu'un rayon passe au centre de l'objet, l'atténuation est tellement importante que pratiquement aucun photon de faible énergie ne le traverse, la proportion de photons haute énergie devient donc majoritaire. On dit que le spectre est durci car son énergie moyenne a augmenté. Comme l'atténuation des matériaux peut être considérée comme inversement proportionnelle à l'énergie moyenne du spectre, il y a une sous estimation de l'atténuation au centre de l'objet. On peut voir cet effet sur la gauche du profil de la Fig. 2.8 (h). Afin de bien comprendre la problématique du durcissement de spectre, nous allons présenter les principes de base de l'interaction rayonnement-matière. Nous exposerons ensuite différentes paramétrisations de l'objet inconnu ainsi que les hypothèses associées. Puis, nous présenterons quelques résultats de reconstruction sur des données simulées et sur des données réelles.

Interaction rayonnement-matière

Comme nous l'avons vu dans l'Eq. (2.16) la loi Beer Lambert relie l'atténuation aux mesures, mais elle correspond uniquement a une loi empirique qui exprime le ratio entre les photons émis par la source et ceux détectés par le capteur. Cette atténuation des photons provient de différents phénomènes de la physique de l'interaction entre le rayonnement et la matière.

Dans la gamme de rayonnements qui nous intéresse (photons possédant une énergie entre 5 et 500 keV), il y a principalement trois types d'interaction rayonnement-matière[7] :

- l'effet photoélectrique : le photon X frappe un atome, ce dernier est absorbé et un électron est éjecté.
- la diffusion Rayleigh (ou diffusion élastique) : le photon X rebondit sur l'atome sans perdre d'énergie.
- la diffusion Compton (ou diffusion inélastique) : le photon X donne de l'énergie à un électron faiblement lié à l'atome. Il y a donc transfert d'énergie entre le photon et l'électron. Le photon X perd une partie de son énergie. En moyenne, l'angle de diffusion est proportionnel à l'énergie transmise à l'électron.

La probabilité d'une interaction rayonnement-matière dépend bien évidemment de la densité des atomes, mais aussi du type d'atomes. En première approximation, on peut utiliser un modèle qui dépend uniquement du numéro atomique des atomes considérés.

Loi Beer-Lambert généralisée

Si on fait l'hypothèse que tous les rayons diffusés ne sont pas détectés, alors, pour des photons d'une énergie ϵ, on peut modéliser toutes les interactions rayonnement-matière par la loi Beer-Lambert généralisée où l'atténuation $\mu(\epsilon, \boldsymbol{r})$ dépend de l'énergie du faisceau incident :

$$I(\lambda, s) = \int_{\mathbb{R}} S(\epsilon) I_s(\epsilon) e^{-\int_{\mathcal{L}(\lambda,s)} \mu(\epsilon,\boldsymbol{r}) \mathrm{d}\boldsymbol{r}} \mathrm{d}\epsilon \tag{2.21}$$

où $I_s(\epsilon)$ correspond au spectre émis par la source et $S(\epsilon)$ correspond à la sensibilité du détecteur en fonction de l'énergie ϵ des photons incidents. La difficulté ici est de reconstruire $\mu(\epsilon, \boldsymbol{r})$ car il y a beaucoup de degrés de liberté vis à vis des mesures acquises.

Paramétrisations de l'inconnue

Le but de ces paramétrisations est de limiter le nombre de degrés de liberté du problème. Ces paramétrisations impliquent certaines hypothèses sur les objets et la physique de détection. Tout d'abord, on considère dans toutes les approches qu'aucun photon diffusé n'arrive sur le capteur (nous verrons à la fin de cette partie, bien que cette hypothèse ne soit pas très réaliste, qu'il est très difficile de trouver une paramétrisation permettant de s'en affranchir). Dans la littérature, il y a principalement deux types de paramétrisation :

- une approche proposée par B. DE MAN [34] qui suppose que l'atténuation peut être vue comme une combinaison linéaire de deux fonctions $\phi(\epsilon)$ et $\psi(\epsilon)$

  $$\mu(\epsilon, \boldsymbol{r}) = a(\boldsymbol{r})\phi(\epsilon) + b(\boldsymbol{r})\psi(\epsilon) \tag{2.22}$$

  où $\phi(\epsilon)$ représente le comportement photoélectrique et $\psi(\epsilon)$ la diffusion Compton. Dans les gammes d'énergie qui nous intéressent, on a des modélisations analytiques pour déterminer $\phi(\epsilon)$ et $\psi(\epsilon)$ relativement précisément. Il ne reste plus qu'à déterminer les coefficients $a(\boldsymbol{r})$ et $b(\boldsymbol{r})$ en chaque point $\boldsymbol{r}$. Ici, les différentes valeurs des coefficients correspondent à des matériaux différents. B. DE MAN est allé plus loin pour contraindre encore plus le problème. Il a considéré que

---

[7] Nous ne nous intéresserons pas à la production paire qui survient dans des gammes d'énergie supérieures à 1 000 keV

l'on pouvait identifier les matériaux présents, uniquement à partir de leurs atténuations à l'énergie à 70 keV :

$$\mu(\epsilon, \boldsymbol{r}) = a(\mu_{70}(\boldsymbol{r}))\phi(\epsilon) + b(\mu_{70}(\boldsymbol{r}))\psi(\epsilon) \tag{2.23}$$

où $a(\mu_{70}(\boldsymbol{r}))$ et $b(\mu_{70}(\boldsymbol{r}))$ sont des fonctions analytiques de $\mu_{70}(\boldsymbol{r})$. De cette manière il suffit de déterminer l'atténuation à 70 keV.

- Une deuxième approche consiste à considérer que l'objet est une combinaison linéaire de matériaux référence. Cette approche est très utilisée dans la reconstruction à partir de deux énergies [73][126][65]. Récemment, une approche de factorisation en matrice creuse de faible rang à permis d'avoir une plus grande diversité de matériaux [60]. Dans le contexte d'une application médicale, les matériaux référence les plus naturels sont l'eau et l'os :

$$\mu(\epsilon, \boldsymbol{r}) = a(\boldsymbol{r})\mu_{eau}(\epsilon) + b(\boldsymbol{r})\mu_{os}(\epsilon) \tag{2.24}$$

L'atténuation des matériaux référence en fonction de l'énergie est connue il suffit donc de déterminer les coefficients linéaires $a(\boldsymbol{r})$ et $b(\boldsymbol{r})$ qui correspondent à l'abondance des différents matériaux. D'un côté cette approche permet de décrire une moins grande variété de matériaux que la précédente ; par contre, elle permet d'avoir un modèle d'atténuation plus fidèle pour les matériaux référence, par exemple elle peut tenir compte de la diffusion Rayleigh. De plus, il permet de connaître l'abondance de deux matériaux référence et de réduire des artefacts dus au durcissement de spectre. La connaissance de l'abondance d'os peut se révéler très intéressante pour le suivi de l'ostéoporose. C'est pourquoi dans le cadre de la thèse de Caifang CAI nous avons choisi ce deuxième modèle **[19][20]**.

Remarque 1 *Comme nous l'avons vu précédemment, les deux modèles ci-dessus font l'hypothèse qu'aucun photon diffusé n'est détecté. Cette hypothèse est vérifiée s'il y a uniquement des diffusions simples*[8] *et que le détecteur est à une distance infinie de l'objet. En pratique il y a forcément des diffusions multiples*[9]*, par contre ils ont une plus grande probabilité d'être atténués car à chaque diffusion Compton les photons perdent de l'énergie. Finalement, l'énergie détectée issue des diffusions multiples est dans la plupart des cas négligeable. Dans les systèmes d'imagerie médicale 3D, le détecteur est très proche du patient. En effet, comme la paire source-détecteur tourne autour du patient, il faut que le détecteur soit proche du patient pour des considérations mécaniques et d'encombrement. En conséquence, pour les systèmes possédant des détecteurs plats grand champ, l'énergie des photons diffusés représente jusqu'à 25 % de l'énergie mesurée.*

Cette fraction de photons détectés est très dure à déterminer car elle dépend de la géométrie d'acquisition mais aussi de la composition de l'objet (position relative des matériaux très diffusants par rapport aux matériaux très absorbants). De ce fait, produire un modèle précis qui tient compte uniquement de la fraction non détectée des photons diffusés est un problème ouvert.

Utilisation d'un système de tomographie à rayons X pour obtenir de la diversité d'information

Nous voulons donc reconstruire deux cartes correspondant à l'abondance d'os et l'abondance d'eau. Pour faire cette séparation en deux matériaux nous devons obtenir des données pouvant les discriminer. Nous avons donc acquis des mesures en faisant varier la tension de la source de rayons X pour que le spectre d'émission $I_s(\epsilon)$ soit modifié. Nous avons choisi les tensions 70 keV et 150 keV, car les spectres générés sont suffisamment différents. De plus, ces tensions sont facilement applicables sur des générateurs médicaux standards. En intégrant le modèle de matériaux Eq. (2.24) dans l'Eq. (2.21) et en

---

[8] Diffusion simple : un photon qui diffuse qu'une fois

[9] Diffusion multiple : un photon qui diffuse au moins deux fois

faisant varier le spectre de la source, on obtient le couple d'équations suivant :

$$I_{70}(\lambda, s) = \int_{\mathbb{R}} S(\epsilon) I_{s70}(\epsilon) e^{-\mu_{eau}(\epsilon)\int_{\mathcal{L}(\lambda,s)} a(\boldsymbol{r})\mathrm{d}\boldsymbol{r}} e^{-\mu_{os}(\epsilon)\int_{\mathcal{L}(\lambda,s)} b(\boldsymbol{r})\mathrm{d}\boldsymbol{r}} \mathrm{d}\epsilon \tag{2.25}$$

$$I_{150}(\lambda, s) = \int_{\mathbb{R}} S(\epsilon) I_{s150} e^{-\mu_{eau}(\epsilon)\int_{\mathcal{L}(\lambda,s)} a(\boldsymbol{r})\mathrm{d}\boldsymbol{r}} e^{-\mu_{os}(\epsilon)\int_{\mathcal{L}(\lambda,s)} b(\boldsymbol{r})\mathrm{d}\boldsymbol{r}} \mathrm{d}\epsilon \tag{2.26}$$

avec I_{s70} (resp. I_{s150}) l'énergie délivrée par la source de rayons X pour une tension à l'anode de 70 keV (resp. 150 keV). On voit clairement que ce modèle est non linéaire à cause de la présence du terme exponentiel. Contrairement au cas du § 2.2.1, les approches de type moindres carrés pondérés sont moins faciles à mettre en œuvre du fait de la présence de l'intégrale suivant l'énergie ϵ. Nous allons donc inverser ce modèle non linéaire. Suite à la discussion du § 2.2.1, on considère que les données mesurées sont entachées d'un bruit blanc Gaussien de variance proportionnelle à sa moyenne. On choisit un *a priori* favorisant la parcimonie dans l'image des gradients des abondances. Afin de préserver la différentiabilité du critère, la norme L^1 est remplacée par la fonction potentielle de Huber.

Une fois que la loi *a posteriori* est déterminée en utilisant la règle de Bayes, Eq. (2.14), nous avons choisi de calculer le Maximum *a posteriori* (MAP). Nous devons résoudre un problème d'optimisation qui est en fait non convexe à cause de la non-linéarité du problème direct et de l'indétermination partielle pour la séparation des deux composantes. Néanmoins, nous avons résolu ce problème d'optimisation en utilisant une approche de descente à pas adaptatif **[20]**. Nous n'avons donc aucune garantie de convergence de cette approche vers le minimum global du critère, mais de manière pratique nous avons initialisé les inconnues par des zéros et les résultats lorsque l'algorithme a convergé, ont toujours été conformes à nos attentes.

Résultats

Afin de valider cette approche décrite dans **[20]**, nous avons fait une expérience numérique (Fig. 2.8) et une expérimentation sur des données réelles (Fig. 2.9). Lors de ces deux expériences, nous avons cherché à être dans les mêmes conditions avec des fantômes les plus proches possibles. Nous avons fabriqué un fantôme composé d'un grand cylindre d'eau (Fig. 2.8 (a)). Ce dernier est percé de 4 quatre cavités :

- à droite, nous avons de l'air : Fig. 2.8 (a)
- en bas, nous avons de l'eau salée : Fig. 2.8 (a)
- en haut, nous avons de l'«os» [10] : Fig. 2.8 (e)
- à gauche, nous avons un anneau d'aluminium : Fig. 2.8 (e).

Afin de comparer les performances de notre approche, nous avons développé une approche similaire aux travaux de Granton [65]. Cette approche que nous nommerons Moindres Carrés Pondérés avec Séparation (MCPS), consiste à reconstruire de manière indépendante les deux jeux de données pour une tension au générateur de rayons X de 70 keV (resp. 150 keV), puis de déterminer de l'abondance de chaque matériau en résolvant un problème de séparation de sources. Le problème de séparation est résolu en calculant la solution d'un critère des moindres carrés pondérés **[19]**.

Comme nous pouvons le voir, en particulier sur les Figs. 2.8 (f) et (h), la méthode MCPS ne résout pas le problème de durcissement de spectre : il y a des artefacts symptomatiques du durcissement de spectre. On voit un effet « coupe » (l'atténuation au centre est sous évaluée par rapport au bord de l'objet) sur le profil passant dans l'aluminium. De plus, la séparation entre les matériaux n'est pas parfaite, une partie importante de l'atténuation de l'os a été expliquée dans l'image d'eau voir Figs. 2.8 (b) et (d). Ces artefacts n'apparaissent pas dans notre nouvelle approche, les niveaux sont respectés et la séparation est bien meilleure que pour l'approche MCPS (voir Figs. 2.8 (c), (d), (g) et (h)).

---

[10] Pour des raisons d'expérimentation nous avons pris de Hydroxyapatite, matériau dont les propriétés sont très proches de l'os.

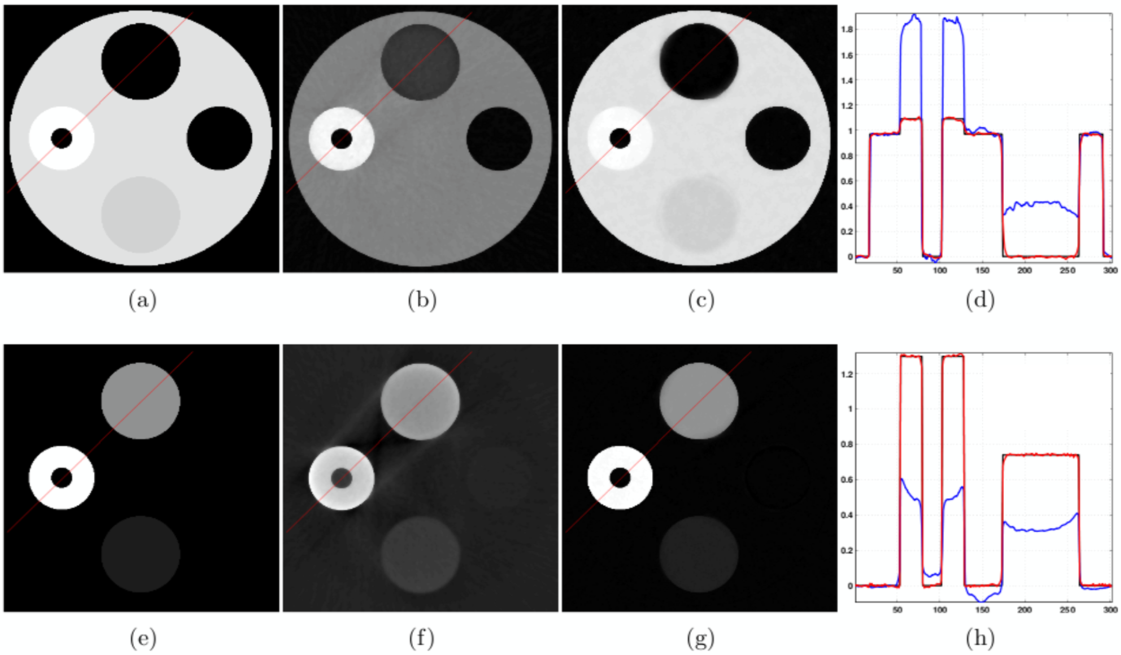
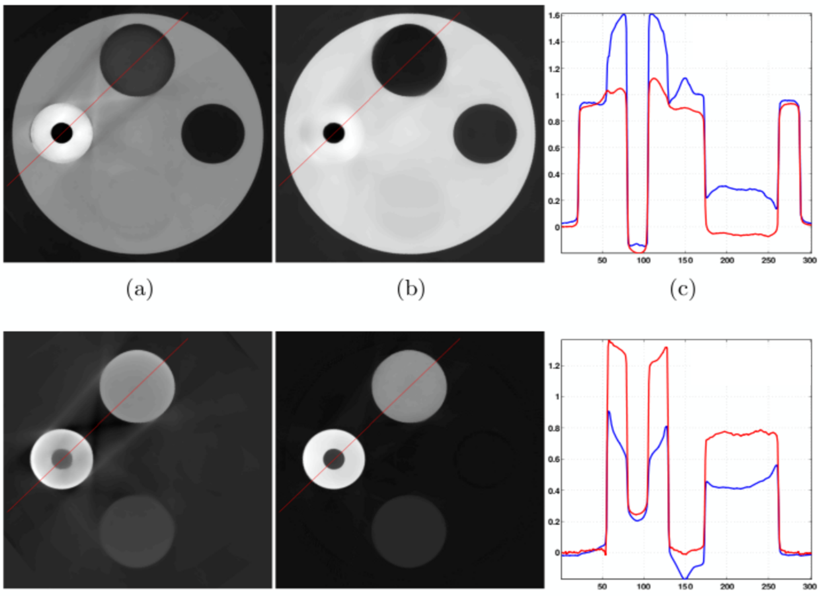


FIGURE 2.8 – Reconstruction polychromatique à partir de données simulées pour une tension de générateur de 70 et 150 keV. La première ligne correspond à l'abondance d'eau et la deuxième à l'abondance d'os. (a) (e) sont les abondances de référence utilisées pour les simulations (b) (f) sont les résultats obtenus avec la méthode MCPS (c) (g) sont les résultats obtenus en inversant le problème décrit par les équations (2.25) et (2.26) (d) (h) correspondent aux profils extraits des images selon la ligne rouge ; les traits noirs correspondent à la référence, les traits bleus à la méthode de reconstruction puis séparation et les traits rouges à l'inversion du modèle non linéaire.

Nous avons aussi évalué l'approche sur des données réelles acquises avec un système que nous ne détaillerons pas ici mais qui est équivalent aux systèmes tomographiques médicaux ; en particulier, le générateur de rayons X fonctionne dans les mêmes gammes d'énergie. Les résultats sont nettement moins bons que sur les simulations. Dans les reconstructions de l'approche MCPS, les artefacts sont largement amplifiés et la séparation reste relativement mauvaise (voir Figs. 2.9 (a), (d), (c) et (f)). Notre approche améliore de façon significative les images tant en terme de réduction d'artefacts qu'en terme de séparation, mais des imperfections demeurent comme au centre de l'anneau d'aluminium où le niveau est anormalement élevé (voir Figs. 2.9 (e) et (f)). Il est évident que dans le contexte des données réelles, la physique est beaucoup plus compliquée que la modélisation que nous avons utilisée. En particulier, l'hypothèse selon laquelle le rayonnement diffusé ne frappe pas le détecteur n'est pas vérifiée en pratique (voir remarque précédente). De plus, nous ne connaissons pas précisément le spectre d'émission du générateur et la sensibilité du détecteur en fonction de l'énergie, malgré un effort conséquent de simulation. En effet, nous n'avons pas pu mesurer de manière précise le spectre réellement émis par notre générateur, car nous ne possédons pas de système de mesure dédié à cette tâche. Malgré ces problèmes, notre approche fait preuve d'une certaine robustesse par rapport aux approches concurrentes.

2.3 Problème de la reconstruction 3D de la couronne solaire

L'étude de la couronne solaire est un enjeu scientifique assez important. En effet, l'activité solaire a des répercussions sur l'activité humaine. En particulier, elle peut fortement endommager les satellites de télécommunication. L'objectif à terme est de prédire l'activité solaire afin de se prémunir de ces effets.



(a) (b) (c)

(d) (e) (f)

FIGURE 2.9 – Reconstruction polychromatique à partir de données réelles sur fantôme physique très proche des simulations pour des tensions de générateur de 70 et 150 keV. La première ligne correspond à l'abondance d'eau et la deuxième à l'abondance d'os. (a) (d) approche MCPS (b) (e) inversion du modèle non-linéaire (d) (h) correspondent aux profils extraits des images selon la ligne rouge ; les traits bleus correspondent à la méthode de reconstruction puis séparation et les traits rouges à l'inversion du modèle non linéaire.

Pour cela, il faut comprendre les mécanismes physiques sous-jacents. On peut citer deux phénomènes dont on ne connaît pas l'origine. Premièrement, la température de la couronne solaire est d'environ un million de kelvins alors que la surface est uniquement de 5800 kelvins. Plusieurs théories [3][22] ont été développées mais aucune n'a encore pu être validée par des observations. Deuxièmement, la vitesse du vent solaire rapide est encore inexpliquée. Le vent solaire est un flux continu de particules chargées qui émane du soleil. Il y a un vent lent (400 km/s) issu des régions équatoriales et un vent rapide (750 km/s) issu des pôles [49].

Mission STEREO

Afin de mieux comprendre le fonctionnement du soleil, des missions d'observation de son activité ont été lancées ces dernières années : Solar and Heliospheric Observatory (SoHO) en 1995 et STEREO en 2006. Le but de la mission STEREO est de reconstruire la structure 3D de la couronne solaire. Pour cela, on souhaite avoir deux mesures simultanées sous deux angles de vue différents du soleil. Cette mission a consisté à envoyer deux satellites quasiment identiques (STEREO A et STEREO B). Ces satellites s'écartent l'un de l'autre au court de la mission afin d'avoir des angles de vue de plus en plus différents et donc une plus grande diversité d'information (voir Fig. 2.10). A l'intérieur de chaque satellite sont embarqués plusieurs instruments : deux coronographes champ proche COR1 et COR2, deux coronographes grand champ H1 et H2 qui peuvent couvrir toute la zone entre le Soleil et la Terre et les EUVI (Extreme UltraViolet Imager) qui sont quatre imageurs qui observent dans l'extrême ultraviolet. Les quatre imageurs EUVI observent aux longueurs d'ondes 17,1 nm, 19,5 nm, 28,4 nm et 30,4 nm. Chaque longueur d'onde correspond à une raie de désexcitation d'un ion particulier qui est présent uniquement si le plasma atteint une température particulière (17,1 nm, ions Fe IX et Fe X, $T_e \approx 10^6 K$; 19,5 nm, ion Fe XII $T_e \approx 1,6 \times 10^6 K$; 28,4 nm ion Fe XV,$T_e \approx 2 \times 10^6 K$; 30,4 nm, ion He II, $T_e \approx 8 \times 10^4 K$). On

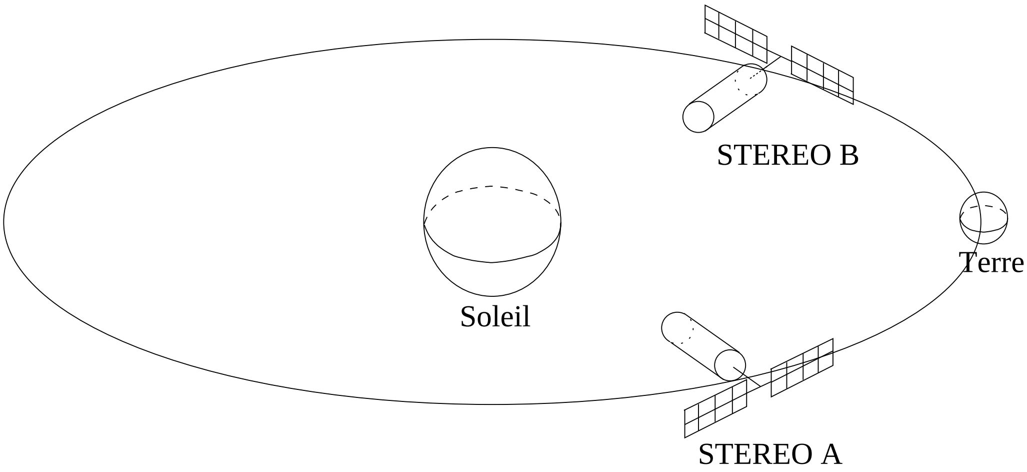


FIGURE 2.10 – Position des satellites STEREO A et STEREO B par rapport au soleil

peut donc accéder à la densité du plasma à différentes altitudes dans la couronne proche du soleil (température entre 1 et 2 millions de degrés Kelvin). Nous nous intéresserons ici uniquement aux données EUVI bien que l'approche que nous allons présenter pourrait facilement être applicable sur les données des coronographes (voir thèse de Nicolas BARBEY [11]). L'image formée (voir Fig. 2.11) sur le plan focal du détecteur est acquise à l'aide d'un capteur CCD. Chaque pixel détecteur intègre tous les photons arrivant dans le télescope pour une direction donnée. Cette mesure est donc l'intégrale selon une droite de la propriété d'émissivité qui nous intéresse. Enfin, l'atténuation à ces longueurs d'ondes des photons ultraviolets par l'atmosphère solaire est négligeable, on peut donc considérer chaque image Fig. 2.11 comme une projection tomographique de l'émissivité de la couronne solaire.

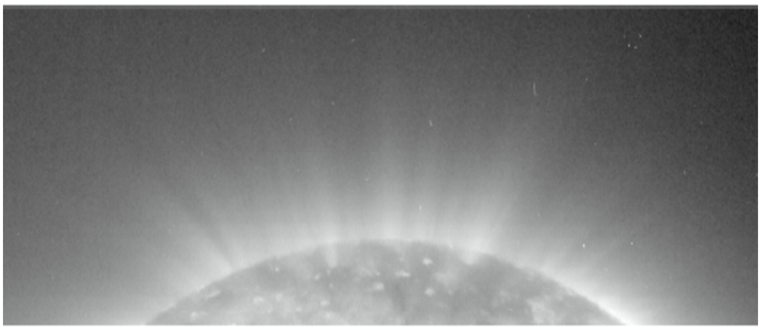


FIGURE 2.11 – Image à 17,1 nm du pôle nord du soleil à l'aide de l'instrument EUVI. On voit clairement des structures verticales lumineuses correspondant aux plumes polaires

2.3.1 Problématiques en traitement du signal

L'objectif astrophysique est la connaissance de la structure 3D de la couronne solaire, et en particulier la forme des plumes polaires. Ce sont des structures fines qui sont sur les pôles du soleil Fig. 2.11. Leurs morphologies ainsi que leurs paramètres physiques (densité et température du plasma) sont des éléments très importants pour expliquer la présence du vent rapide.

Tout d'abord, nous connaissons simultanément deux projections de notre objet d'intérêt grâce à STEREO A et STEREO B, mais cette information est trop lacunaire pour faire une reconstruction précise de l'objet. On utilise aussi la rotation du soleil sur lui même pour acquérir d'autres angles de projection. Sa période de rotation au pôle est d'environ 27 jours [140]. L'un des problèmes que l'on n' abordera pas ici est que la rotation du Soleil n'est pas solide mais différentielle, c'est à dire que la période de rotation à l'équateur est plus rapide que celle aux pôles [140]. Pour les structures du soleil qui nous intéressent ici, nous avons une difficulté supplémentaire : la durée d'apparition d'une plume polaire peut être très courte (2 jours) au regard de la période de rotation du soleil. L'hypothèse implicite de la reconstruction tomographique est que l'objet n'évolue pas au court de l'acquisition. Cette évolution rapide déforme les structures et crée des artefacts très importants dans les images. Afin de pouvoir tirer des conclusions astrophysiques des reconstructions tomographiques, nous devons résoudre un problème de tomographie dynamique.

La littérature dans ce domaine est assez abondante pour le domaine de l'imagerie médicale [14, 121, 143, 138, 69]. Dans ce contexte la tomographie dynamique est utilisée pour s'affranchir des mouvements du patient. Le mouvement le plus étudié est le mouvement cardiaque [69, 1]. La difficulté de ce mouvement est sa rapidité, mais il a l'avantage d'être périodique. Pratiquement l'ensemble des approches utilise ce caractère périodique du mouvement. On acquiert un grand nombre de périodes et on sélectionne les données associées à la même phase cardiaque [69, 1]. Cette approche est très efficace et permet de supprimer la majeure partie des artefacts. Cette hypothèse est malheureusement inutilisable dans notre cas. La tomographie dynamique est aussi utilisée dans le cas du mouvement respiratoire. Ce dernier n'est pas forcément périodique, mais on peut soit mesurer le mouvement à partir d'une mesure extérieure [138], soit faire une estimation du mouvement **[64]**[14][143], car il est relativement lent.

Dans le domaine de l'astrophysique solaire, bien que la reconstruction tomographique soit connue depuis plus de 60 ans [137], le nombre de travaux est relativement faible. Cela s'explique facilement par la difficulté du problème associé à la communauté scientifique nettement plus réduite qu'en imagerie médicale. R. FRAZIN s'est intéressé au problème de la reconstruction dynamique de la couronne solaire [57, 58]. Il a cherché à reconstruire des structures présentes dans la haute atmosphère qui ont une évolution plus lente que nos plumes. Sa résolution repose sur l'introduction d'une information *a priori* de douceur selon l'axe temporel de l'objet 4D à reconstruire. D'autres travaux ont consisté à définir un *a priori* selon des directions en accord avec un modèle de MagnétoHydroDynamique (MHD) de la couronne solaire [141].

2.3.2 Première approche : paramétrisation des plumes polaires

Le problème de reconstruction dynamique de la couronne solaire est extrêmement mal posé car on a dans les meilleures configurations 50 fois plus d'inconnues que de données. Ces travaux se sont fait dans le cadre de la thèse de N. BARBEY, et sont détaillés dans **[12]** (reproduit à la page 166). L'inversion de ce problème passe donc par l'introduction d'informations *a priori* très contraignantes. Pour ce faire, nous faisons l'hypothèse que la morphologie des plumes ne varie pas sur 13,5 jours. Cette modélisation de l'évolution des plumes vient de la supposition que les plumes sont tubulaires et suivent les lignes de champs magnétiques du Soleil [86, 39]. Par contre, nous allons faire évoluer l'intensité de chaque plume au cours du temps. Notre objet 3D + T que l'on note $\boldsymbol{x}$ va être paramétré à l'aide d'une composante morphologique 3D notée $\boldsymbol{e}$ et des intensités variant dans le temps $\boldsymbol{\gamma}$. La nouvelle paramétrisation du problème est décrite par l'équation

$$\boldsymbol{x}(\vec{r},t) = \boldsymbol{e}(\vec{r}). \times \boldsymbol{L}\boldsymbol{\gamma}(t). \tag{2.27}$$

Dans notre paramétrisation nous avons besoin de connaître une segmentation de chaque plume (Fig. 2.12 (b)). Cette segmentation est contenue dans la matrice $\boldsymbol{L}$. De plus, l'opération $.\times$ fait le produit terme à terme des vecteurs. Afin de contraindre suffisamment le problème nous avons considéré au maximum 4 ou 5 plumes actives.

Inversion :

On considère, un modèle direct linéaire correspondant à la transformée en rayons divergents. Nous modéliserons une projection à un instant t donné par la matrice $\boldsymbol{H}_t$.

$$\boldsymbol{y}(t) = \boldsymbol{H}_t\left(\boldsymbol{e}.\times\boldsymbol{L}\boldsymbol{\gamma}(t)\right) + \boldsymbol{b}(t) \tag{2.28}$$

avec $\boldsymbol{b}(t)$ un bruit Gaussien i.i.d. de variance σ_b^2. On déduit que la vraisemblance $p(\boldsymbol{y}|\boldsymbol{e},\boldsymbol{\gamma},\sigma_b)$ suit une loi Gaussienne. Afin d'avoir une optimisation efficace dans un contexte où la taille des reconstructions est importante (problème 3D + T), nous avons choisi des *a priori* conjugués avec la vraisemblance.

Pour ce qui est de la composante morphologique, nous avons choisi d'introduire une information de douceur à travers un *a priori* Gaussien corrélé,

$$p(\boldsymbol{e}|\sigma_e) \propto \exp\left(-\frac{\|\boldsymbol{D}_e\boldsymbol{e}\|^2}{2\sigma_e^2}\right) \tag{2.29}$$

où la matrice de corrélation est égale à $(\boldsymbol{D}_e^{\mathrm{t}}\boldsymbol{D}_e)^{-1}$ et où $\boldsymbol{D}$ est la matrice des différences finies entre pixels voisins.

Pour ce qui est de la composante temporelle, on considère que la fluctuation est relativement douce. De plus, pour lever une indétermination du problème, on ajoute un terme de rappel à la valeur 1. On favorisera donc les solutions dont la valeur multiplicative est proche de 1, en utilisant la loi *a priori* suivante :

$$p(\boldsymbol{\gamma}|\sigma_\gamma,\sigma_1) \propto \exp\left(-\frac{\|\boldsymbol{D}_t\boldsymbol{\gamma}\|^2}{2\sigma_\gamma^2} - \frac{\|\boldsymbol{\gamma}-\boldsymbol{1}\|^2}{2\sigma_1^2}\right), \tag{2.30}$$

où la matrice $\boldsymbol{D}_t$ correspond aux différences finies selon la dimension temporelle.

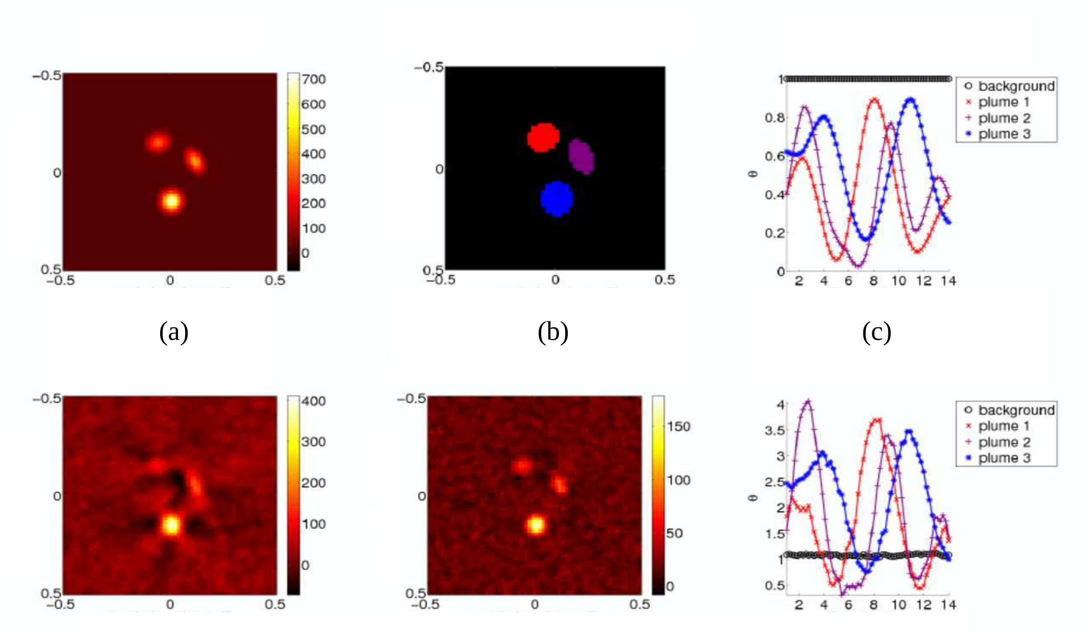
En utilisant la règle de Bayes (2.14), on obtient la loi *a posteriori* jointe suivante :

$$p(\boldsymbol{e},\boldsymbol{\gamma}|\boldsymbol{y},\sigma_\gamma,\sigma_1,\sigma_e,\sigma_b) \propto \exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{e}.\times\boldsymbol{L}\boldsymbol{\gamma})\|}{2\sigma_b^2} - \frac{\|\boldsymbol{D}_e\boldsymbol{e}\|^2}{2\sigma_e^2} - \frac{\|\boldsymbol{D}_t\boldsymbol{\gamma}\|^2}{2\sigma_\gamma^2} - \frac{\|\boldsymbol{\gamma}-\boldsymbol{1}\|^2}{2\sigma_1^2}\right) \tag{2.31}$$

Le modèle étant bilinéaire, la loi jointe $p(\boldsymbol{e},\boldsymbol{\gamma}|\boldsymbol{y},\ldots)$ n'est pas Gaussienne. Par contre les lois conditionnelles $p(\boldsymbol{e}|\boldsymbol{\gamma},\boldsymbol{y},\ldots)$ et $p(\boldsymbol{\gamma}|\boldsymbol{e},\boldsymbol{y},\ldots)$ sont Gaussiennes. Nous allons, pour des questions d'efficacité de calculs, choisir l'estimateur du maximum *a posteriori*. Puis nous utiliserons une procédure de maximisation alternée afin de trouver un optimum local. En effet, cette approche ne nous garantit pas de déterminer l'optimum global du critère conjoint. Comme la dimension du vecteur $\boldsymbol{\gamma}$ est faible, sa minimisation connaissant $\boldsymbol{e}$ peut se faire en appliquant directement la formule analytique d'inversion (Eq. (2.11)). Pour ce qui est de l'optimisation suivant la composante morphologique la dimension du problème nous contraint à utiliser une approche de descente suivant une direction liée au gradient du critère. Nous procédons ensuite avec un schéma de minimisation alternée où le critère d'arrêt repose sur la norme de la mise à jour inférieure à une constante ϵ pendant au moins trois itérations.

Résultats sur des données simulées :

Nous avons simulé des données en utilisant notre modèle de l'Eq. (2.27) composé de trois plumes. La composante morphologique est représentée sur la Fig. 2.12 (a), et l'évolution temporelle de chaque plume sur la Fig. 2.12 (c). Nous avons simulé les données avec les mêmes configurations que pour un jeu de données réelles et nous avons ajouté un bruit Gaussien avec une variance équivalente à celle présente sur les données réelles. Nous avons comparé notre approche (Fig. 2.12 (e) et Fig. 2.12 (f)) avec la reconstruction de type filtrage rétroprojection Eq. (2.4) sur la Fig. 2.12 (d). Il apparaît clairement sur la Fig. 2.12 (d) que la non prise en compte de l'évolution temporelle crée des artefacts très importants et que les structures ne peuvent pas être reconstruites de manière précise. L'approche proposée permet bien de reconstruire de manière fidèle la composante morphologique (Fig. 2.12 (e)) et de manière assez fidèle l'allure des évolutions temporelles de chaque plume. Par contre, les valeurs des gains ne sont pas correctement estimées à cause de l'indétermination du modèle à une constante multiplicative près.



(a) (b) (c)

(d) (e) (f)

FIGURE 2.12 – Reconstruction en utilisant notre première approche : la première ligne correspond aux valeurs que nous avons utilisées pour faire les simulations, la deuxième ligne correspond à des estimations : (a) image morphologique vraie : elle est composée de trois tâches Gaussiennes, (b) segmentation en zones possédant la même dynamique d'évolution, (c) valeur du gain γ en fonction du temps d'acquisition : les couleurs courbes correspondent aux couleurs de la segmentation de la figure (b), (d) reconstruction statique en utilisant une approche filtrage-rétroprojection classique, (e) estimation de la morphologie avec notre première approche, (f) estimation de la dynamique temporelle.

Résultats sur des données réelles :

Nous avons appliqué notre approche à des données réelles issues d'un seul satellite car au moment de ces travaux la calibration photométrique entre les deux satellites n'était pas encore faite. Pour appliquer notre approche, on suppose que l'on connaît la segmentation des différentes plumes qui composent l'atmosphère. Pour déterminer cette segmentation (Fig. 2.13 (a)), on fait une première reconstruction avec une approche filtrage rétroprojection (Fig. 2.13 (c)). On identifie les structures les plus lumineuses puis on détermine manuellement le contour des plumes. L'estimation des évolutions temporelles (Fig. 2.13 (b)) est cohérente avec ce que nous attendions. L'estimation de la composante morphologique (Fig. 2.13 (d)) est plus contrastée que l'approche classique (Fig. 2.13 (c)). De plus, les valeurs négatives dans l'image sont beaucoup moins nombreuses que dans l'approche classique. Tous ces éléments nous laissent penser que notre estimation est plus proche de la réalité que celle de la méthode de filtrage rétroprojection.

Conclusions :

Au regard de ces premiers résultats, la paramétrisation qui sépare la composante morphologique de l'évolution temporelle paraît très pertinente. Les résultats sur des données simulées sont concluants et ceux sur des données réelles sont encourageants à bien des égards. Par contre, le point négatif de l'approche est de considérer que la segmentation des différentes plumes est connue au moment de l'estimation. Il est bien évident qu'une erreur d'estimation est très préjudiciable sur la qualité de la reconstruction finale. C'est pourquoi nous avons continué des travaux après la fin de la thèse de N. BARBEY afin de s'affranchir cette hypothèse.

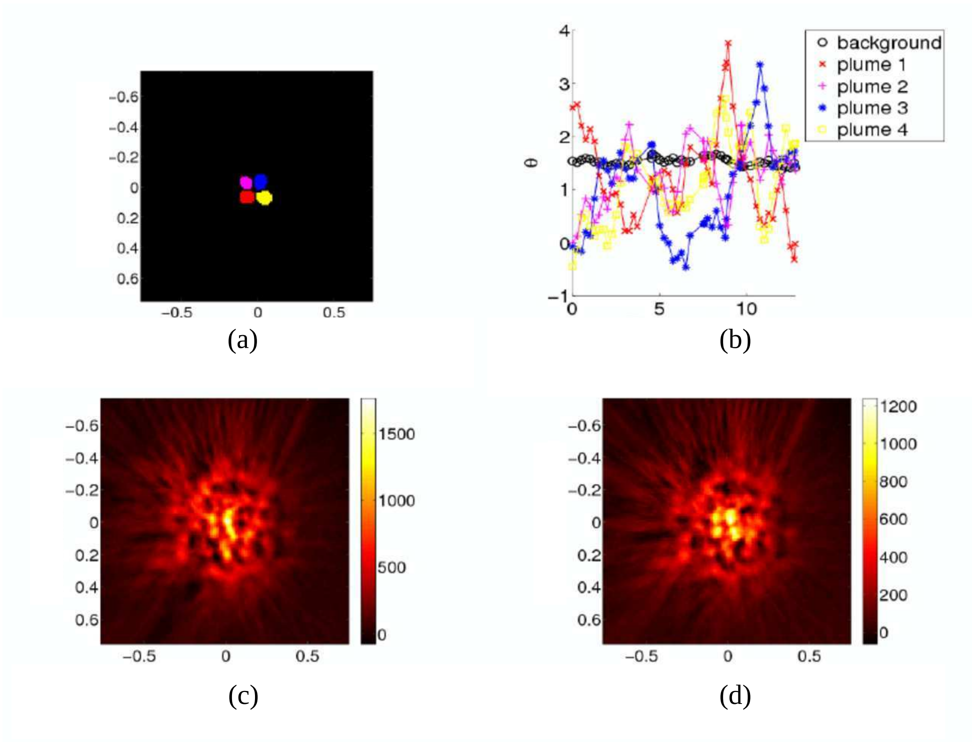


FIGURE 2.13 – Reconstruction de données réelles en utilisant un seul satellite. Visualisation d'une seule coupe au dessus du pôle nord : (a) segmentation des plumes définies manuellement, (b) évolution temporelle, (c) reconstruction par un algorithme de filtrage-rétroprojection, (d) reconstruction de la composante morphologique.

2.3.3 Deuxième approche : identification comme un problème de séparation de sources

Le problème de l'estimation de la segmentation dans le cadre précédent est un problème non linéaire, difficile à résoudre de manière efficace. On pourrait sans doute utiliser une méthode de recuit simulé qui serait très coûteuse en temps de calcul. Nous avons choisi une autre voie qui consiste à garder la structure linéaire du modèle direct. Par contre, cette paramétrisation entraîne une augmentation du nombre d'inconnues (3 à 5 fois plus). On suppose que l'objet est la somme d'une composante de fond $\boldsymbol{e}_1$ qui est invariante dans le temps plus des composantes associées à une ou plusieurs plumes partageant la même dynamique temporelle $\boldsymbol{e}_{2,\dots,I}$. Afin de simplifier les expressions, nous présenterons tous les résultats uniquement avec une composante statique $\boldsymbol{e}_1$ modélisant le fond et deux composantes $\boldsymbol{e}_2$ resp. $\boldsymbol{e}_3$ ayant des dynamiques différentes $\alpha(t)$ resp. $\beta(t)$,

$$\boldsymbol{x}(\vec{r},t) = \boldsymbol{e}_1(\vec{r}) + \alpha(t)\boldsymbol{e}_2(\vec{r}) + \beta(t)\boldsymbol{e}_3(\vec{r}). \tag{2.32}$$

On peut réécrire l'équation ci-dessus sous forme matricielle,

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{e}, \tag{2.33}$$

avec

$$\boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \boldsymbol{e}_3 \end{pmatrix} \quad \text{et} \quad \boldsymbol{A} = \begin{pmatrix} 1 & \boldsymbol{\alpha}_1 & \boldsymbol{\beta}_1 \\ \vdots & \vdots & \vdots \\ 1 & \boldsymbol{\alpha}_T & \boldsymbol{\beta}_T \end{pmatrix}.$$

On reconnaît un modèle de mélange instantané de sources. Ce modèle a fait l'objet de très nombreux travaux [76][32][21][31]. On trouve aussi des travaux sur les mélanges convolutifs de sources

[107][94][93], mais à notre connaissance il n'existe quasiment pas de travaux sur les mélanges linéaires de sources. C'est pourtant le problème auquel nous sommes confrontés car nous devons appliquer la matrice de projection $\boldsymbol{H}$ à l'Eq. (2.33),

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{e} + \boldsymbol{b}. \tag{2.34}$$

$\boldsymbol{b}$ est ici un bruit Gaussien i.i.d. de variance σ_b^2. On déduit que la vraisemblance $p(\boldsymbol{y}|\boldsymbol{e}, \boldsymbol{A}, \sigma_b)$ est Gaussienne.

Comme nous l'avons dit au début de cette partie, l'idée de cette approche est de garder un modèle linéaire, mais en contrepartie le nombre de paramètres inconnus est bien plus important (trois fois plus dans notre cas). Nous devons donc introduire une quantité d'information *a priori* importante. La paramétrisation ne permettant pas de réduire suffisamment les degrés de liberté, nous allons utiliser une information *a priori* de parcimonie sur la composante morphologique des plumes $\boldsymbol{e}_2$ et $\boldsymbol{e}_3$ au moyen d'une loi à queue lourde. Notre choix s'est porté sur la loi de Student T car elle peut être générée comme un mélange infini de Gaussienne. Cette propriété va nous permettre d'utiliser uniquement des *a priori* conjugués avec la vraisemblance. Cette propriété nous permet d'avoir des lois *a posteriori* conditionnelles dans des familles connues. Nous pourrons donc utiliser des approches bayésiennes variationelles décrites dans le chapitre 3 pour résoudre notre problème d'inférence.

Les différentes lois *a priori* utilisées :

- pour les composantes morphologiques des plumes, exprimons ici la loi de $\boldsymbol{e}_2$:

$$\begin{aligned} p(\boldsymbol{e}_2(i)|\lambda,\nu) &= St(\boldsymbol{e}_2(i)|\lambda,\nu) \\ &= \int_{\mathbb{R}} p(\boldsymbol{e}_2(i)|\boldsymbol{a}_2(i),\lambda)\, p(\boldsymbol{a}_2(i) \,|\, \nu)\, \mathrm{d}\boldsymbol{a}_2(i) \\ &= \int_{\mathbb{R}} \mathcal{N}\left(e_2(i)|0,(\boldsymbol{a}_2(i)*\lambda)^{-1}\right)\, \mathcal{G}\left(\boldsymbol{a}_2(i)|\frac{\nu}{2},\frac{\nu}{2}\right) \mathrm{d}\boldsymbol{a}_2(i) \end{aligned} \tag{2.35}$$

  où $St(\cdot|\lambda,\nu)$ [11] est une loi de Student centrée, $\mathcal{N}(\cdot|m,\sigma^2)$ est une loi Gaussienne de moyenne m et de variance σ^2 et $\mathcal{G}(\cdot|p_1,p_2)$ [12] est une loi Gamma centrée de paramètres p_1 et p_2. On choisit la même loi avec les mêmes valeurs de paramètres pour e_3. Nous allons estimer les précisions (l'inverse de la variance) des lois Gaussiennes $\boldsymbol{a}_2$ et $\boldsymbol{a}_3$ au lieu de les marginaliser pour conserver la conjugaison des *a priori* (pour plus de précision, cette approche est décrite dans le chap. 3, § 3.3).
- pour les composantes dynamiques des plumes $(\boldsymbol{\alpha},\boldsymbol{\beta})$ nous utilisons une information de douceur,

$$p(\boldsymbol{\alpha},\boldsymbol{\beta}|\sigma_t) \propto \exp(-\frac{\|\boldsymbol{D}_t\boldsymbol{\alpha}\|^2 + \|\boldsymbol{D}_t\boldsymbol{\beta}\|^2}{2\sigma_t^2}) \tag{2.36}$$

- pour la composante de fond $\boldsymbol{e}_1$, nous utiliserons une information de douceur spatiale,

$$p(\boldsymbol{e}_1|\sigma_e) \propto \exp(-\frac{\|\boldsymbol{D}_e\boldsymbol{e}_1\|^2}{2\sigma_e^2}) \tag{2.37}$$

Loi *a posteriori* :

$$p(\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_2,\boldsymbol{a}_2,\boldsymbol{a}_3,\boldsymbol{\alpha},\boldsymbol{\beta}|\boldsymbol{y},\sigma_b,\sigma_t,\sigma_e,\lambda,\nu) \tag{2.38}$$

$$\propto p(\boldsymbol{y}|\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3,\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_b)p(\boldsymbol{e}_1|\sigma_e)p(\boldsymbol{e}_2|\boldsymbol{a}_2,\lambda)p(\boldsymbol{a}_2|\nu) \tag{2.39}$$

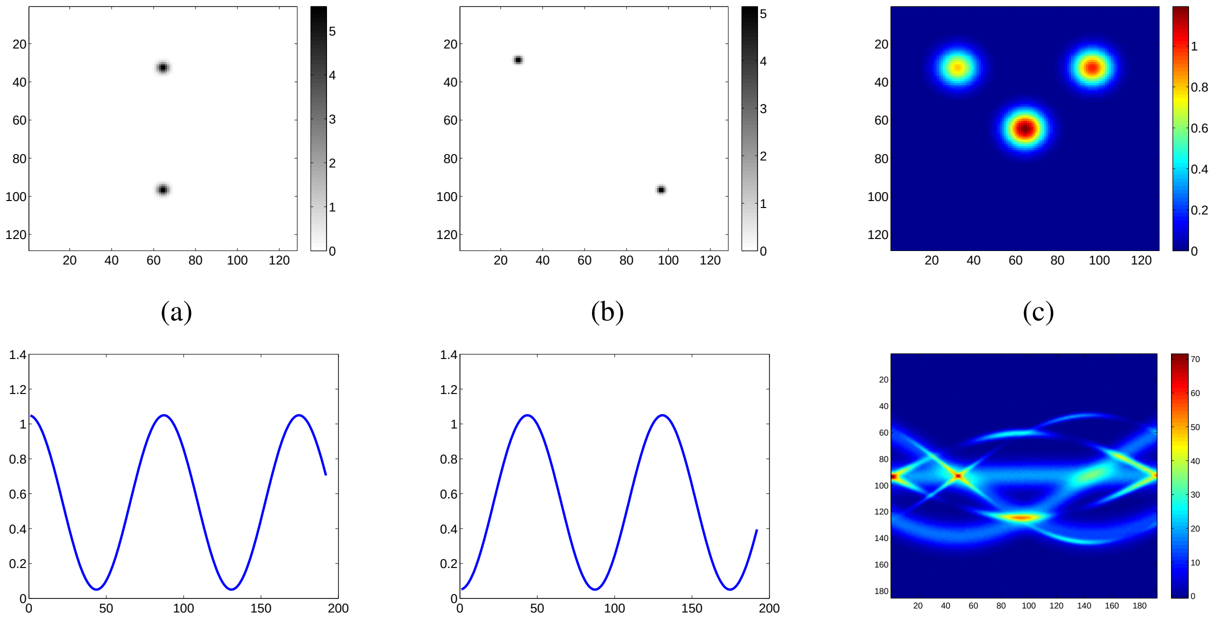
$$p(\boldsymbol{e}_3|\boldsymbol{a}_3,\lambda)p(\boldsymbol{a}_3|\nu)p(\boldsymbol{\alpha},\boldsymbol{\beta}|\sigma_t) \tag{2.40}$$

On utilise l'algorithme 2 décrit dans le Chap. 3, § 3.5.1. L'initialisation est faite par des zéros pour la composante de fond et par une approche filtrage rétroprojection pour les composantes morphologiques des plumes, enfin, les composantes dynamiques sont initialisées par des 1.

---

[11] $St(x|\lambda,\nu) \propto \left(1+\frac{\lambda}{\nu}x^2\right)^{-\frac{\nu+1}{2}}$

[12] $\mathcal{G}(x|p_1,p_2) \propto x^{p_1-1}\exp\left[-p_2 x\right]$



(a) (b) (c)

(d) (e) (f)

FIGURE 2.14 – Simulation des données : (a) composante morphologique $\boldsymbol{e}_2$, (b) composante morphologique $\boldsymbol{e}_3$, (c) composante statique $\boldsymbol{e}_1$, (d) évolution dynamique $\boldsymbol{\alpha}$ associée à $\boldsymbol{e}_2$, (e) évolution dynamique $\boldsymbol{\beta}$ associée à $\boldsymbol{e}_3$, (f) sinogramme légèrement bruité $\boldsymbol{y}$

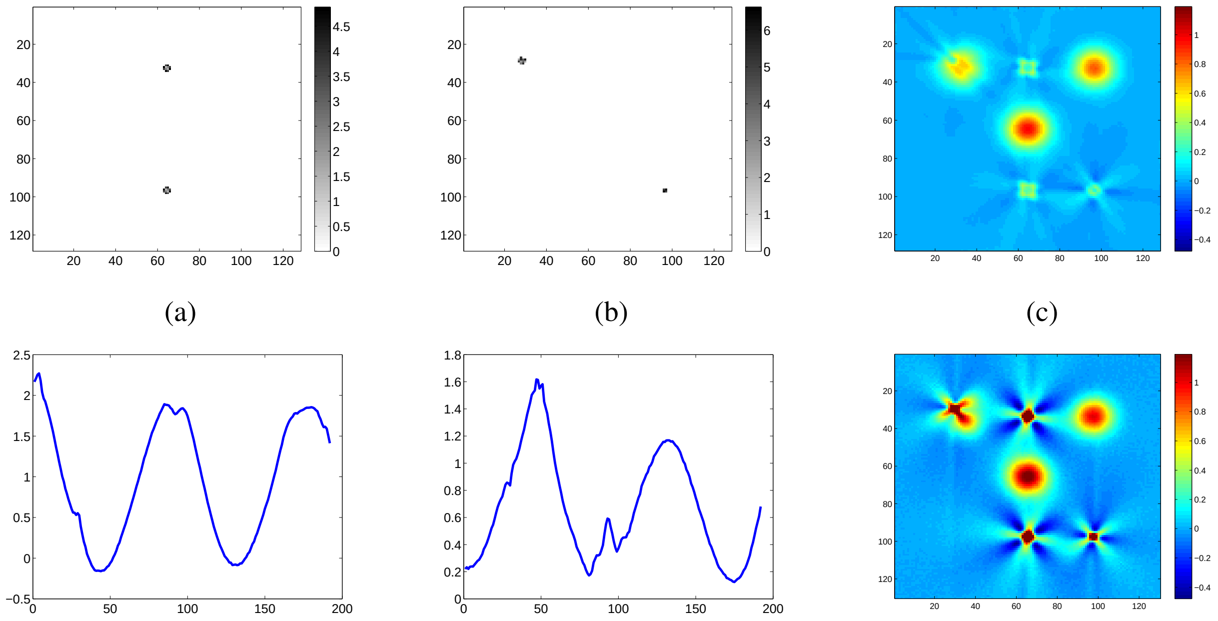
Premiers résultats sur des données simulées :

Ces travaux sont encore à un stade préliminaire, ils ont débouché uniquement sur une communication dans une conférence internationale **[79]**. Il manque un travail important de validation sur des données réelles pour qu'ils soient aboutis. Afin de montrer la faisabilité de l'approche, nous avons fait l'expérience numérique suivante : nous avons simulé des données en géométrie parallèle à partir d'un objet 2D + T 128 × 128 × 180. Cet objet comporte trois composantes :

- le fond qui est invariant dans le temps et qui est relativement régulier (Fig. 2.14 (c)),
- une première composante dynamique composée de deux plumes (Fig. 2.14 (a)) qui suit l'évolution dynamique représentée sur la Fig. 2.14 (d).
- une deuxième composante dynamique aussi composée de deux plumes dont l'une est superposée avec le fond (Fig. 2.14 (b) et (c)). L'évolution dynamique de cette composante est représentée sur la Fig. 2.14 (e), elle est en opposition de phase par rapport à l'autre composante dynamique.

Les données sont représentées sur la Fig. 2.14 (f). Il y a une projection tous les degrés. Nous avons ajouté du bruit Gaussien i.i.d. de variance égale à 10. On voit clairement des évanouissements des structures dynamiques dans le sinogramme (Fig. 2.14 (f)). La difficulté de ce problème inverse est directement visible sur la Fig. 2.14. On cherche à retrouver les images des Figs. 2.14 (a)-(c) et les signaux représentés sur les Figs. 2.14 (d) et (e) uniquement à partir du sinogramme de la Fig. 2.14 (f).

On commence par faire une première reconstruction avec une approche filtrage rétroprojection (Fig. 2.15 (f)). On observe que la composante statique est parfaitement reconstruite. Par contre, l'inconsistance dans les données due aux composantes dynamiques provoque de sévères artefacts en étoile autour des plumes. Suite à l'utilisation de l'algorithme 2 page 49 adapté à notre problème, nous obtenons des lois approchantes pour chaque paramètre. Nous représentons sur les Figs. 2.15 (a), (b), (c), (d) et (e) la moyenne de ces lois. On remarque que l'estimation du fond (Fig. 2.15 (c)) représentée avec la même échelle de couleurs est de bien meilleure qualité. Les artefacts sont très fortement réduits, les valeurs négatives sont quasiment inexistantes. Si on compare par rapport à l'image vraie (Fig. 2.14 (c)) on voit



(a) (b) (c)

(d) (e) (f)

FIGURE 2.15 – Comparaison entre une reconstruction de type filtrage rétroprojection (f) et notre nouvelle approche (a)-(e) : (a) estimation de la composante morphologique $\boldsymbol{e}_2$, (b) estimation de la composante morphologique $\boldsymbol{e}_3$, (c) estimation de composante statique $\boldsymbol{e}_1$, (d) estimation de l'évolution dynamique $\boldsymbol{\alpha}$ associée à $\boldsymbol{e}_2$, (e) estimation de l'évolution dynamique $\boldsymbol{\beta}$ associée à $\boldsymbol{e}_3$, (f) reconstruction par une approche classique de filtrage rétroprojection (fonction `iradon` de matlab).

que l'introduction de l'information *a priori* de douceur nous fait légèrement sous estimer le maximum de l'image. De plus, on voit en haut à gauche qu'à l'endroit où la plume est superposée au fond, il y a un trou dans l'estimation du fond. L'approche lors de la séparation a préféré mettre tout le signal dans la composante dynamique. En ce qui concerne les composantes associées aux plumes, l'approche a très bien séparé les deux composantes dynamiques (rappelons que l'initialisation de ces composantes est identique ; on utilise l'image de la Fig. 2.15 (f) pour la partie morphologique et 1 pour l'évolution temporelle). De plus les plumes sont correctement positionnées et l'allure des évolutions dynamiques Fig. 2.15 (d) (resp. Fig. 2.15 (e)) sont très proches des vraies Fig. 2.14 (d) (resp. Fig. 2.14 (e)). Il y a uniquement une erreur relativement importante autour de la projection 90 très visible sur la Fig. 2.15 (e). Cette projection correspond à une intégration suivant des lignes horizontales. Dans ce cas, il y a clairement une indétermination dans le problème de séparation que l'information *a priori* que nous avons introduite ne peut que partiellement lever. Enfin, les composantes morphologiques reconstruites ont un support plus petit que les images de départ. L'introduction de l'information de parcimonie a contraint les faibles valeurs des composantes associées aux plumes de se retrouver dans l'image du fond. On voit des artefacts résiduels autour de l'emplacement des plumes sur la Fig. 2.15 (c) causés par cette mauvaise estimation.

Chapitre 3

Bayesien Variationnel

3.1 Introduction

Problématique liée à résolution de problèmes inverses de grande dimension

Comme on a pu le voir dans le chapitre précédent, les problèmes inverses sont souvent de très grandes dimensions (plus d'un millions d'inconnues). On peut citer par exemple, les problèmes de reconstructions 3D, 3D + T ou 2D mettant en jeu de grand détecteurs (ex : les CCD 4096×4096 sont couramment dans l'imagerie satellitaire).

Dans beaucoup d'applications le besoin d'avoir des approches non-supervisées se fait sentir. Dans ces approches, les paramètres qui règlent le compromis entre l'information apportée par les données et l'information *a priori*, sont aussi estimés. Nous appellerons dans la suite ces paramètres, hyperparamètres, car ils ne correspondent pas à la grandeur d'intérêt du problème inverse. Dans un cadre bayésien, l'estimation des hyperparamètres passe par l'introduction d'une loi *a priori* sur ces hyperparamètres qui conduit à une loi *a posteriori* jointe. Ces approches s'appellent «bayésiennes hiérarchiques». Une fois la loi *a posteriori* obtenue, on cherche un estimateur ponctuel. On a principalement le choix entre le maximum ou la moyenne *a posteriori*. L'obtention du maximum *a posteriori* est équivalent à un problème d'optimisation du critère $Q(\boldsymbol{w}) = -\log\big(p(\boldsymbol{w}|\boldsymbol{y})\big)$. Le critère que l'on optimise est en général non convexe lorsqu'on estime conjointement les hyperparamètres. L'optimisation non-convexe étant délicate en pratique, beaucoup de travaux récents ont choisi de calculer l'estimateur de la moyenne *a posteriori* par une approche stochastique.

État de l'art

Les méthodes les plus utilisées pour obtenir l'estimateur de la moyenne *a posteriori*, sont les approches de calculs stochastiques de type Monte Carlo «Markov Chain» (MCMC) [113][114][90]. L'idée des MCMC est de générer une chaîne d'échantillons qui converge asymptotiquement vers des échantillons sous la loi cible (loi *a posteriori*). On conserve ensuite uniquement les échantillons sous la loi *a posteriori* obtenus après un certain temps, appelé temps de chauffe. En faisant la moyenne empirique d'un grand nombre de ces échantillons, on obtient un très bon estimateur de la moyenne *a posteriori*. Il y a principalement deux types d'algorithmes MCMC : les algorithmes de Hasting-Metropolis et l'échantillonnage de Gibbs. Le principe des méthodes de Hasting-Metropolis est de générer un échantillon sous une loi plus simple que l'on peut échantillonner facilement. Cet échantillon est alors accepté avec une certaine probabilité. La principale limitation de cette méthode est la dimension de l'espace à explorer qui peut devenir importante quand le nombre d'inconnues augmente. Il est donc quasiment impossible de trouver une loi de proposition suffisamment proche de la loi cible sur un problème de grande dimension. En ce qui concerne l'échantillonneur de Gibbs, son principe est de tirer des échantillons sous des lois conditionnelles. Lorsque la chaîne a convergé, l'algorithme génère des échantillons sous la loi *a posteriori* jointe. Dans un problème avec un grand nombre d'inconnues, cet algorithme converge extrêmement lentement si on échantillonne conditionnellement inconnue par inconnue. Il est par contre efficace lorsque l'on peut échantillonner conjointement un grand nombre d'inconnues, ce qui est le cas si la loi jointe est séparable ou si elle est Gaussienne avec une matrice de covariance facilement inversible. Par exemple, il y a de nombreuses approches bayésiennes développées dans le cas de la déconvolution [90][61]**[101]** [1], car la matrice de covariance est diagonalisable par transformée de Fourier (sous approximation circulante). Par contre, il est extrêmement difficile de générer un échantillon vectoriel de grande dimension efficacement sous une loi corrélée si la matrice de covariance n'a pas de structure particulière qui la rend facilement inversible ou si l'inverse de la matrice de covariance n'est pas creuse.

D. MACKAY a proposé en 1995 une alternative à ces méthodes d'échantillonnage : le bayésien variationnel [88]. Il s'est inspiré des approches en champ moyen «mean field» développées en physique statistique [105]. Cette approche a aussi été introduite sous une autre forme dans la communauté des informaticiens [74]. Dans le domaine du traitement du signal, cette méthodologie a été appliquée à différents sujets comme la séparation de sources [91][28], le choix de modèles pour décomposer un signal comme une somme de signaux AR [92], la réduction d'images hyper spectrales [9], les problèmes d'apprentissage dans des dictionnaires en utilisant la parcimonie [142][124], les problèmes de déconvolution [25][8] et le problème de déconvolution myope d'un flou dû au mouvement du photographe [50]. L'approche bayésienne variationnelle est moins générique que les approches MCMC, et en théorie plus approximative, mais elle est plus efficace en pratique surtout en grande dimension. Elle est limitée car elle ne permet de résoudre des problèmes inverses que lorsque la loi *a posteriori* appartient à une famille de lois conjuguées avec la vraisemblance. Cependant, comme les calculs sont analytiques, la vitesse de convergence est nettement plus rapide que les approches MCMC. L'idée de cette méthodologie est d'approcher la loi *a posteriori* par des lois séparables. On cherche donc la fonction séparable qui est la plus proche de la loi *a posteriori* au sens de la divergence de Kullback-Leiber. Il faut donc résoudre un problème d'optimisation dans un espace de dimension infinie, à savoir un espace fonctionnel. En pratique, la minimisation de la fonctionnelle se fait composante séparable par composante séparable, de manière alternée. Cette minimisation est une version fonctionnelle de l'algorithme de minimisation de Gauss-Seidel dans un espace de dimension finie [96]. Lorsque l'on a un très grand nombre de composantes séparables, l'algorithme devient très lent.

Contributions

Il y a une réelle difficulté à développer des approches non supervisées pour les problèmes inverses de très grande dimension. L'algorithme que nous avons proposé dans **[56]** (reproduit à la page 81) permet de s'affranchir de ce problème de dimension dans de très nombreux cas. L'idée est de transposer les approches classiques de descente de gradient à pas optimal, bien connue dans les espace de dimension finie [96], à l'espace des distribution de probabilité, qui est de dimension infinie. Nous avons aussi, établi une preuve de convergence de notre algorithme dans ces travaux. Nous exposerons brièvement dans ce chapitre l'approche bayésienne variationnelle classique § 3.2, puis notre nouvel algorithme dans le § 3.3. Nous appliquerons cette méthodologie à un problème conjoint de reconstruction sur-résolue associé à un problème de séparation de composante dans le contexte de l'imagerie astrophysique. Nous finirons par présenter des résultats de notre approche sur des données simulées et réelles.

3.2 Méthodologie bayésienne variationelle

Nous décrivons ici l'approche bayésienne variationnelle classique introduite par D. MACKAY [88]. Dans cette partie, afin d'avoir une présentation générale des approches nous noterons par $\boldsymbol{y} \in \mathbb{R}^M$ le vecteur de dimension M contenant les données et nous noterons par $\boldsymbol{w} \in \mathbb{R}^N$ le vecteur rassemblant

---

[1] Les références sur des travaux auxquels j'ai contribué sont en gras dans le texte

toutes les inconnues à estimer, qu'elles soient ou non cachées. La méthodologie bayésienne variationnelle vise à approcher la loi *a posteriori* par une loi approchante séparable que l'on notera $q(\boldsymbol{w})$

$$q(\boldsymbol{w}) = \prod_i q_i(w_i).$$

On choisit généralement le degré de séparabilité en fonction de la complexité des dépendances de l'*a posteriori*. Dans le cas qui nous intéresse ici, les problèmes de grandes dimensions, le choix d'une forte séparabilité est pratiquement imposé par la difficulté à calculer les statistiques à l'ordre deux d'une loi vectorielle de grande taille **[56]**.

Une fois le choix de la séparabilité fait, on détermine la fonction de densité de probabilité q la plus proche de l'*a posteriori* au sens de la divergence de Kullback-Leibler, sous contrainte d'obtenir une densité de probabilité (pour plus de détails voir [129][5]). Comme cette loi approchante est supposée séparable, nous pouvons mener les calculs analytiquement. On obtient ainsi la loi approchante optimale suivante :

$$\forall i \in \{1, \ldots, N\}, \quad q_i(w_i) = \frac{1}{K_i} \exp\left(< \log p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j \neq i} q_j(w_j)} \right), \tag{3.1}$$

où K_i est une constante de normalisation, et où $< \boldsymbol{w} >_q = \int_{\mathbb{R}^N} \boldsymbol{w} q(\boldsymbol{w}) \mathrm{d}\boldsymbol{w}$.

On remarque clairement que cette solution est analytique mais malheureusement qu'elle n'a pas une forme explicite. Pour obtenir l'optimum de manière pratique, il faut mettre en oeuvre un algorithme itératif de point fixe :

$$q_i^{k+1}(w_i) = \frac{1}{K_i} \exp\left(< \log p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j \neq i} q_j^k(w_j)} \right), \tag{3.2}$$

La procédure d'obtention de la loi approchante correspond donc à une optimisation alternée par groupe de coordonnées, résumée dans l'algorithme 1 ci-dessous :

Algorithm 1 Algorithme bayésien variationnel classique

```
INITIALISATION($q^0$)
repeat
    for $i \in \{1, \ldots, N_{groupe}\}$ do
        function ESTIMER $q_i^{k+1}(w_i)$(connaissant $q_j^{k+1}(w_j)$ pour $j < i$ et $q_l^k(w_l)$ pour $l \geq i$)
            Calcul de $q_i^{k+1}(w_i)$ en utilisant Eq. (3.2))
        end function
    end for
    Estimer l'énergie libre $F(q^k)$.
until Convergence
```

Il est bien connu dans le cas de l'optimisation en dimension finie que ces approches deviennent très lentes quand le nombre de groupes de coordonnées est élevé et donc pratiquement inapplicable lorsqu'on veut résoudre un problème de très grande dimension. C'est pourquoi nous nous intéresserons à une autre approche, basée sur le gradient exponentialisé, introduit par W. KIVINEN [80]. L'avantage de cette dernière approche est qu'elle se révèle bien adaptée à des problèmes de grande dimension.

3.3 Gradient exponentiel pour le bayésien variationnel

Le but du gradient exponentiel pour le bayésien variationnel est de mettre à jour toutes les lois approchantes simultanément. Pour cela, il s'inspire des approches de type gradient à pas optimal, bien connues dans la résolution des problèmes d'optimisation en dimension finie. Toutefois, dans le cas qui nous intéresse ici nous voulons trouver un loi approchante conjointe permettant de minimiser la divergence de Kullback-Leibler.

Nous commencerons par introduire la fonctionnelle à optimiser, puis nous définirons le type de différentielle utilisée ainsi que l'espace fonctionnel dans lequel nous nous placerons. Nous exposerons ensuite la direction de descente utilisée ainsi que la définition du pas de descente.

Nous pouvons facilement trouver la relation suivante entre l'évidence du modèle $\mathcal{M}$ et la distance de Kullback-Leibler que nous voulons minimiser [2] :

$$\log p(\boldsymbol{y}|\mathcal{M}) = F(q(\boldsymbol{w})) + \mathcal{KL}[q(\boldsymbol{w})||p(\boldsymbol{w}|\boldsymbol{y},\mathcal{M}))]. \tag{3.3}$$

où

$$\begin{aligned} F(q(\boldsymbol{w})) &= \int_{\mathbb{R}^N} \log(\frac{p(\boldsymbol{y},\boldsymbol{w}|\mathcal{M})}{q(\boldsymbol{w})}) q(\mathrm{d}\boldsymbol{w}) \\ &= \int_{\mathbb{R}^N} \log p(\boldsymbol{y},\boldsymbol{w}) q(\mathrm{d}\boldsymbol{w}) - \int_{\mathbb{R}^N} \log(q(\boldsymbol{w})) q(\mathrm{d}\boldsymbol{w}) \end{aligned} \tag{3.4}$$

est l'énergie libre négative. Le premier terme de gauche correspond à l'espérance du logarithme de la loi jointe par rapport à q et le terme de droite correspond à l'entropie de la loi approchante q. Pour un jeu de données $\boldsymbol{y}$, l'évidence du modèle est constante. Pour trouver la loi approchante q nous pouvons donc soit minimiser la divergence de Kullback-Leibler, soit maximiser l'énergie libre négative. D'un point de vue pratique, on préfère maximiser l'énergie libre négative car elle fait intervenir la loi jointe de $p(\boldsymbol{y},\boldsymbol{w}|\mathcal{M})$ et non la loi *a posteriori* $p(\boldsymbol{y}|\boldsymbol{w},\mathcal{M})$. Or la loi jointe est plus facile à manipuler car elle ne nécessite pas de connaître la fonction de partition de la loi *a posteriori* qui est souvent difficilement calculable.

Notre problème d'optimisation fonctionnelle est de maximiser l'énergie libre négative. Afin de résoudre ce problème, nous avons choisi de nous placer dans l'espace des mesures de Radon signées noté $\mathcal{R}$ [119]. Nous nous intéressons plus particulièrement à l'espace $\mathcal{R}(\mathbb{R})^N = \mathcal{R}(\mathbb{R}) \times \mathcal{R}(\mathbb{R}) \times \ldots \times \mathcal{R}(\mathbb{R})$ qui est le produit cartésien de N espaces de mesures de Radon signées sur $\mathbb{R}$. L'avantage d'utiliser les mesures signées est qu'elles définissent ainsi un espace de Banach, muni de la norme de la variation totale $\|\nu\|_{TV}$.

Dans cet espace fonctionnel, le problème d'optimisation se formalise comme suit. On cherche :

$$q(\boldsymbol{w})^{opt} = \underset{q\in\mathcal{R}(\mathbb{R})^N}{\arg\max}\, F(q(\boldsymbol{w})) \quad s.c. \;\; q \geq 0. \tag{3.5}$$

Pour déterminer cette mesure à densité optimale, nous utilisons la différentielle de Fréchet de la fonctionnelle $F(q(\boldsymbol{w}))$. Cette différentielle est définie par l'application linéaire $\mathrm{d}F(q(\boldsymbol{w});\,\cdot\,) : \mathcal{R}(\mathbb{R})^N \to \mathbb{R}$ vérifiant

$$\forall \nu \in \mathcal{R}(\mathbb{R})^N, \quad \lim_{t\to 0} \frac{\|F(q(\boldsymbol{w}) + t\nu(\boldsymbol{w})) - F(q(\boldsymbol{w})) - t\mathrm{d}F(q(\boldsymbol{w});\nu(\boldsymbol{w}))\|}{t} = 0. \tag{3.6}$$

De plus, dans le cas qui nous occupe, notre différentielle peut s'exprimer comme :

$$\begin{aligned} \forall \nu \in \mathcal{R}(\mathbb{R})^N, \quad \mathrm{d}F(q(\boldsymbol{w});\nu(\boldsymbol{w})) &= \int_{\mathbb{R}^N} \mathrm{d}f(q,\boldsymbol{w})\; \nu(\mathrm{d}\boldsymbol{w}) \\ &= \sum_{i=1}^{N} \int_{\mathbb{R}} \mathrm{d}_i f(q_i,w_i)\; \nu_i(\mathrm{d}w_i). \end{aligned} \tag{3.7}$$

Dans ce cas

$$\mathrm{d}f(q,\boldsymbol{w}) = \sum_i \mathrm{d}_i f(q_i,w_i).$$

---

[2] la démonstration se trouve dans [28] ou dans [129]

À l'aide des définitions précédentes nous pouvons calculer la différentielle de Fréchet de la fonctionnelle donnée par l'Eq. (3.4). Ces calculs sont détaillés dans **[55][56]**. Ainsi on a $\forall i \in \{1, \dots, N\}$, $\forall \boldsymbol{w} \in \mathbb{R}^N$,

$$\mathrm{d}_i f(q_i, w_i) = \langle \log p(\boldsymbol{y}, \boldsymbol{w}) \rangle_{\prod_{j \neq i} q_j(w_j)} - \log q_i(w_i) - 1. \tag{3.8}$$

Afin de garantir la positivité de la densité q, nous avons choisi une mise à jour à l'aide d'un gradient exponentiel par analogie avec l'approche de W. KIVINEN [80] en dimension finie.

$$q^{k+1}(\boldsymbol{w}) = K q^k(\boldsymbol{w}) \exp\left[\mathrm{d}f(q, \boldsymbol{w})\right]^{\alpha_k} \tag{3.9}$$

De plus nous avons ajouté la constante K afin de garantir que $\int_{\mathbb{R}^N} q^{k+1}(\mathrm{d}\boldsymbol{w}) = 1$. Nous avons aussi introduit un pas de descente α_k qui apparaît en puissance dans l'Eq. (3.9) car la mise à jour est multiplicative. Nous remarquons que si nous prenons le logarithme de l'expression de l'Eq. (3.9), nous retrouvons une structure de mise à jour additive avec un pas multiplicatif, comme dans la descente de gradient classique. Si le pas est suffisamment bien choisi il permet de garantir la convergence de l'algorithme vers le minimum global de la fonctionnelle, l'énergie libre étant une fonctionnelle concave au sens large. L'énoncé du théorème de convergence et sa preuve se trouvent dans **[56]**.

En remplaçant Eq. (3.8) dans Eq. (3.9) nous obtenons l'équation de mise à jour des lois approchantes,

$$\begin{aligned} q^{k+1}(\boldsymbol{w}) &= K q^k(\boldsymbol{w}) \left(\prod_i \frac{\exp\left(< \log p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j \neq i} q_j^k(w_j)} \right)}{q_i^k(w_i)} \right)^{\alpha_k} \\ &= K q^k(\boldsymbol{w}) \left(\prod_i \frac{q_i^r(w_i)}{q_i^k(w_i)} \right)^{\alpha_k}, \end{aligned} \tag{3.10}$$

avec

$$q_i^r(w_i) = \exp\left(< \log p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j \neq i} q_j^k(w_j)} \right). \tag{3.11}$$

À chaque itération, l'obtention du pas de descente α_k nécessite la résolution du problème d'optimisation scalaire suivant

$$\alpha_{k+1} = \underset{\alpha \in \mathbb{R}}{\arg\max}\, F(q^k(\boldsymbol{w}) \exp\left[\mathrm{d}f(q, \boldsymbol{w})\right]^{\alpha}). \tag{3.12}$$

En pratique, le critère est trop compliqué pour obtenir une solution optimale. Nous avons donc déterminé un pas sous-optimal en calculant le pas optimal du développement de Taylor à l'ordre 2 du critère.

Algorithm 2 Gradient exponentiel pour le bayésien variationnel

```
INITIALISATION( $q^0 \in \mathcal{R}(\mathbb{R})^N$ )
repeat
    Calcul des $q_i^r(w_i)$ (connaissant $\forall j, q_j^k(w_j)$)
    Calcul de $\alpha_k$
    mise à jour $q^{k+1}$ en utilisant l'Eq. (3.10)
until Convergence
```

3.4 Application à la reconstruction-séparation de composantes astrophysiques

Une question cruciale en astrophysique à l'heure actuelle est la détermination de la structure de l'univers juste après le big-bang. Cette connaissance passe par l'étude des fluctuations du fond diffus cosmologique afin d'en extraire l'information liée à l'énergie fossile. Les modèles cosmologiques utilisent le spectre du fond diffus. Cependant, la présence d'avant plans (étoiles, galaxies non résolues) biaise son estimation. L'objectif de notre approche est de pouvoir séparer les étoiles du fond.

Nous voulons donc construire une méthode permettant à la fois de faire une inversion sur-résolue et non supervisée, et de résoudre conjointement un problème de séparation entre une source étendue corrélée et une source impulsionnelle séparable. La méthode développée doit de plus être rapide et efficace sur des données de grandes tailles. Typiquement, pour les données réelles issues du satellite, il y a jusqu'à 100 millions de données et 50 millions d'inconnues à traiter.

3.4.1 Modèle direct :

Nous cherchons une image infrarouge composée de sources ponctuelles, étoiles et galaxies non résolues par l'instrument, et d'une composante étendue qui correspond au fond diffus cosmologique. L'idée est de reconstruire la grandeur observée comme la somme de deux composantes, cette idée était déjà présente dans [30][38][62]. Nous considérons donc que notre image d'intérêt $\boldsymbol{x} \in \mathbb{R}^N$ comprend deux composantes, une image lisse $\boldsymbol{s}$ et une image de pics $\boldsymbol{p}$. Le modèle du ciel s'écrit alors

$$\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{p}.$$

Nous l'intéressons ici aux données issues du télescope Herschel et plus particulièrement de l'imageur infrarouge lointain ($\lambda = 250\mu m$ ou $\lambda = 350\mu m$ ou $\lambda = 500\mu m$) SPIRE.

Les observations sont données par plusieurs balayages du ciel horizontaux ou/et verticaux. Cet imageur est composé de détecteurs bolométriques répartis suivant une grille hexagonale. De plus, il y a un sous échantillonnage dû à l'encombrement de chaque bolomètre. La mesure du ciel est faite grâce à une procédure de balayage du ciel sub-pixellique. Le modèle prend aussi en compte la réponse optique de l'instrument, donnée par la somme de la réponse des miroirs et des cornets qui fonctionnent comme des antennes permettant de concentrer l'onde électromagnétique. Nous considérons que la réponse du télescope correspond à la convolution du ciel par une réponse impulsionnelle supposée connue. Ce modèle direct a été développé dans le cadre de la thèse de François ORIEUX (pour plus de détails sur ce point voir [100]**[102]**). Les données sont aussi entachées d'une erreur $\boldsymbol{\epsilon}$. La relation qui lie les données, notées $\boldsymbol{y} \in \mathbb{R}^M$, au ciel $\boldsymbol{x}$ est donc la suivante

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\epsilon} = \boldsymbol{H}(\boldsymbol{s} + \boldsymbol{p}) + \boldsymbol{\epsilon} \tag{3.13}$$

Cette erreur $\boldsymbol{\epsilon}$ est décomposable en deux parties : une erreur de mesure reproductible (offset) $\boldsymbol{o}$ qui est estimable et une erreur que l'on considère comme un bruit blanc Gaussien centré $\boldsymbol{b}$ d'inverse variance (précision) γ_b. On peut supposer sans introduire trop d'erreur de modélisation que les offsets sont indépendants d'un bolomètre à l'autre et qu'ils sont identiques pour chaque bolomètre pendant une phase d'acquisition. La dimension du vecteur d'offset ($N_{balayage} \times N_{bolometres}$) est donc très inférieure au nombre de mesures. De plus, la stratégie de mesure rend les données redondantes, chaque pixel du ciel étant vu plusieurs fois par les capteurs (bolomètres) à des instants différents. Tout cela fait que le problème d'estimation des offsets est bien posé. Si on introduit une matrice de réplication des offsets, $\boldsymbol{R}$, qui duplique les offsets afin d'avoir une valeur par mesure, l'équation (3.13) devient

$$\boldsymbol{y} = \boldsymbol{H}(\boldsymbol{s} + \boldsymbol{p}) + \boldsymbol{R}\boldsymbol{o} + \boldsymbol{b}. \tag{3.14}$$

Comme le bruit est Gaussien, en utilisant l'Eq. (3.14) on détermine la vraisemblance de $\boldsymbol{s}$, $\boldsymbol{p}$ et $\boldsymbol{o}$:

$$\Pr(\boldsymbol{y}|\boldsymbol{p}, \boldsymbol{s}, \boldsymbol{o}, \gamma_b) \propto \exp[-\frac{1}{2}\gamma_b\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{s} - \boldsymbol{H}\boldsymbol{p} - \boldsymbol{R}\boldsymbol{o}\|^2] \tag{3.15}$$

3.4.2 Lois *a priori* :

Ces lois *a priori* vont modéliser l'information que l'on va introduire pour séparer les deux composantes (impulsionnelle et douce). En effet, lorsqu'on regarde l'équation (3.14) on s'aperçoit aisément qu'il y a indétermination entre $\boldsymbol{s}$ et $\boldsymbol{p}$. La séparation entre les deux composantes ne proviendra donc que de l'information *a priori*. C'est pourquoi, il faut qu'*a priori*, l'information soit suffisamment discriminante. On choisit ici une loi corrélée Gaussienne pour la composante douce $\boldsymbol{s}$ et une loi à lourde queue séparable pour la composante impulsionnelle $\boldsymbol{p}$. Ces lois sont par exemple, la loi de Laplace, la loi Bernoulli Gaussien [111][63], les mélanges finis de Gaussiennes [131][70], la loi de Cauchy, les lois α-stables et plus généralement, certaines lois de la famille des mélanges infinis de Gaussiennes (Gaussian Scale Mixture GSM) [139][23].

Une autre propriété importante pour le choix de nos *a priori* est qu'ils soient conjugués avec la vraisemblance. Cette propriété assure que l'*a posteriori* corresponde à des lois connues et donc que ses cumulants soient facilement calculables. Ces contraintes nous conduisent aux choix suivants :

- La loi *a priori* sur les offsets $\boldsymbol{o}$ est une loi Gaussienne iid de moyenne o_m et d'inverse variance γ_o :

$$\Pr(\boldsymbol{o}|o_m, \gamma_o) \propto \exp[-\frac{1}{2}\gamma_o\|\boldsymbol{o} - o_m\|^2]$$

- La loi *a priori* pour $\boldsymbol{s}$ est une loi Gaussienne corrélée d'inverse variance γ_s et de matrice de covariance $(\gamma_s\, \boldsymbol{D}^{\mathrm{t}}\boldsymbol{D})^{-1}$:

$$\Pr(\boldsymbol{s}|\gamma_s) \propto \exp[-\frac{1}{2}\gamma_s\|\boldsymbol{D}\boldsymbol{s}\|^2]$$

  où $\boldsymbol{D}$ est la matrice des différences finies.
- La loi *a priori* pour $\boldsymbol{p}$ est une loi de Student-t iid.

$$\Pr(\boldsymbol{p}|\nu, \lambda) \propto \prod_i \left(1 + \frac{\lambda}{\nu}p_i^2\right)^{-\frac{\nu+1}{2}}.$$

  Nous avons choisi cette loi à queue lourde plutôt qu'une loi de Laplace [27] ou qu'une loi α-stable car elle peut s'écrire comme un mélange infini de Gaussiennes (GSM) faisant intervenir une loi Gamma,

$$\Pr(p_i|\nu, \lambda) = \int_{\mathbb{R}} \Pr(p_i|a_i, \lambda) \Pr(a_i|\nu) \mathrm{d}a_i$$

  avec

$$\begin{aligned} \Pr(p_i|a_i, \lambda) &= \mathcal{N}(0, (\lambda a_i)^{-1}), \\ \Pr(a_i|\nu) &= Gamma(\frac{\nu}{2}, \frac{\nu}{2}). \end{aligned}$$

Remarque 2 *En utilisant ce problème étendu, on définit la loi de Student-t comme un mélange infini de Gaussiennes centrées. En remplaçant la loi Gamma par d'autres lois, on peut obtenir d'autres types de probabilités telles que les lois de Laplace, voir [139].*

L'idée introduite par [25] pour la résolution de problèmes inverses est d'introduire des variables cachées $\boldsymbol{a}$ correspondant à la précision des lois Gaussiennes. On va donc résoudre un problème étendu dans lequel l'on estimera la composante lisse $\boldsymbol{s}$, la composante impulsionnelle $\boldsymbol{p}$, et la précision $\boldsymbol{a}$ de cette composante impulsionnelle. Comme souvent, nous prenons le parti de résoudre un problème étendu possédant plus d'inconnues et dans lequel nous devons estimer aussi des variables cachées $\boldsymbol{a}$. Nous estimons maintenant trois fois plus d'inconnues que de données. La contrepartie est que l'on a un *a posteriori* Gaussien pour $\boldsymbol{s}, \boldsymbol{p}$ et Gamma pour $\boldsymbol{a}$.

$$\Pr(\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{a}|\gamma_s, \nu, \lambda) = \Pr(\boldsymbol{s}|\gamma_s) \Pr(\boldsymbol{p}, \boldsymbol{a}|\nu, \lambda) = \Pr(\boldsymbol{s}|\gamma_s) \Pr(\boldsymbol{p}|\boldsymbol{a}, \lambda) \Pr(\boldsymbol{a}|\nu) \quad (3.16)$$

Notons ici, que l'on aurait pu introduire des variables cachées discrètes, cela nous aurait conduit à une loi *a priori* mélange de Gaussiennes comme dans [70][6], ou Bernoulli Gaussien [111]**[81]**.

Maintenant que nous avons défini des lois *a priori* conjuguées avec la vraisemblance, nous sommes sûr que les lois *a posteriori* conditionnelles appartiendront à des familles connues. L'application de la méthodologie bayésienne variationnelle présentée dans la partie 3.2 est donc possible. Grâce à cela, l'optimisation des lois approchantes se fera à travers les paramètres de ces lois, ce qui facilite l'implémentation des méthodes.

3.4.3 Loi *a posteriori* :

Pour obtenir la loi *a posteriori*, nous utilisons la règle de Bayes. On déduit donc des équations (3.15) et (3.16) la loi *a posteriori* étendue pour $(\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{a}, \boldsymbol{o})$

$$\begin{aligned}\Pr(\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{a}, \boldsymbol{o} | \boldsymbol{y}, \gamma_b, \gamma_s, \nu, \lambda, o_m, \gamma_o, \mathcal{M}) \propto \exp -\frac{1}{2}\left[\gamma_b \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{s} - \boldsymbol{H}\boldsymbol{p} - \boldsymbol{R}\boldsymbol{o}\|^2\right. \\ \left. + \gamma_s \|\boldsymbol{D}\boldsymbol{s}\|^2 + \gamma_o \|\boldsymbol{o} - o_m\|^2\right] \prod_i \sqrt{a_i} \exp\left[-\frac{1}{2} p_i^2 a_i \lambda\right] a_i^{\frac{\nu}{2}-1} \exp\left[-\frac{1}{2}\nu a_i\right] \quad (3.17)\end{aligned}$$

où $\mathcal{M}$ est le modèle choisi.

3.5 Mise en œuvre de l'approche bayésienne variationnelle

3.5.1 Étude de la séparabilité

En étudiant la loi *a posteriori* Eq. (3.17), on remarque que la loi conditionnelle

$$\Pr(\boldsymbol{a} | \boldsymbol{s}, \boldsymbol{p}, \boldsymbol{y}, \boldsymbol{o}, \gamma_b, \gamma_s, \nu, \lambda, \mathcal{M})$$

est séparable, ce qui n'est pas le cas pour les autres lois conditionnelles :

$$\begin{aligned}&\Pr(\boldsymbol{s} | \boldsymbol{a}, \boldsymbol{p}, \boldsymbol{y}, \boldsymbol{o}, \gamma_b, \gamma_s, \nu, \lambda, \mathcal{M}), \\ &\Pr(\boldsymbol{p} | \boldsymbol{a}, \boldsymbol{s}, \boldsymbol{y}, \boldsymbol{o}, \gamma_b, \gamma_s, \nu, \lambda, \mathcal{M}), \\ &\Pr(\boldsymbol{o} | \boldsymbol{a}, \boldsymbol{s}, \boldsymbol{y}, \boldsymbol{p}, \gamma_b, \gamma_s, \nu, \lambda, \mathcal{M}).\end{aligned}$$

De plus, nous allons traiter un volume de données important (4 millions de données) et nous avons un nombre d'inconnues aussi très grand, à savoir 3×700000 inconnues. C'est pourquoi nous avons choisi de prendre des lois approchantes toutes séparables afin de réduire le coût de calcul et la taille mémoire nécessaire. En effet, le stockage de la loi vectorielle impose au minimum le stockage d'une matrice de covariance de taille N^2. Nous allons donc appliquer les méthodologies présentées dans les parties 3.2 et 3.3. En prenant $\boldsymbol{w} = (\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{a}, \boldsymbol{o})$, la loi approchante q s'écrit dans ce cas :

$$q(\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{a}, \boldsymbol{o}) = \prod_i q_{s_i}(s_i) \prod_j q_{p_j}(p_j) \prod_k q_{a_k}(a_k) \prod_l q_{o_l}(o_l) \quad (3.18)$$

Rappelons que l'objectif est de trouver la loi approchante q maximisant l'énergie libre négative de l'Eq. (3.4). Pour ce faire, nous allons utiliser une procédure alternée décrite dans l'algorithme ci-dessous :

Algorithm 3 Algorithme d'inversion du bi-modèle

```
INITIALISATION($q_{\boldsymbol{s}}^0, q_{\boldsymbol{p}}^0, q_{\boldsymbol{a}}^0, q_{\boldsymbol{o}}^0$)
repeat
    function MISE À JOUR DE LA LOI DE $\boldsymbol{a}(q_{\boldsymbol{p}}^k)$
        Approche bayésienne variationnelle classique algorithme 1
    end function
    function MISE À JOUR DE LA LOI DE $\boldsymbol{p}(q_{\boldsymbol{s}}^k, q_{\boldsymbol{o}}^k, q_{\boldsymbol{a}}^{k+1})$
        Approche gradient exponentiel algorithme 2
    end function
    function MISE À JOUR DE LA LOI DE $\boldsymbol{s}(q_{\boldsymbol{o}}^k, q_{\boldsymbol{p}}^{k+1})$
        Approche gradient exponentiel algorithme 2
    end function
    function MISE À JOUR DE LA LOI DE $\boldsymbol{o}(q_{\boldsymbol{s}}^{k+1}, q_{\boldsymbol{p}}^{k+1})$
        Approche gradient exponentiel algorithme 2
    end function
until Convergence
```

3.5.2 Mise à jour des lois approchantes :

Comme nous avons choisi des *a priori* conjugués, nous savons que la loi optimale appartient forcément à la même famille que la loi *a posteriori*. Nous choisissons donc des lois approchantes q dans la même famille que la loi *a posteriori*. On en déduit donc que les lois approchantes $q(\boldsymbol{s})$ et $q(\boldsymbol{p})$ seront des Gaussiennes tandis que la loi approchante $q(\boldsymbol{a})$ appartiendra à la famille des lois Gamma. Comme nous l'avons vu dans la partie précédente nous avons choisi le degré de séparabilité maximum afin de limiter le coût de calcul. Les lois approchantes à l'itération k sont donc définies comme suit :

$$\begin{array}{rcl} q_{o_r}^k(o_r) & = & \mathcal{N}(\boldsymbol{mo}_k(r), \boldsymbol{\Upsilon o}_k^2(r)), \\ q_{s_i}^k(s_i) & = & \mathcal{N}(\boldsymbol{ms}_k(i), \boldsymbol{\Upsilon s}_k^2(i)), \\ q_{p_j}^k(p_j) & = & \mathcal{N}(\boldsymbol{mp}_k(j), \boldsymbol{\Upsilon p}_k^2(j)), \\ q_{a_l}^k(a_l) & = & Gamma(\boldsymbol{\alpha}_k(l), \boldsymbol{\gamma}_k(l)). \end{array}$$

avec $\boldsymbol{mo}$, $\boldsymbol{ms}$ (resp. $\boldsymbol{mp}$) le vecteur de moyenne des lois Gaussiennes approchantes de $\boldsymbol{o}$, $\boldsymbol{s}$ (resp. $\boldsymbol{p}$) et $\boldsymbol{\Upsilon o}$, $\boldsymbol{\Upsilon s}$ (resp. $\boldsymbol{\Upsilon p}$) le vecteur des écarts types des lois Gaussiennes approchantes de $\boldsymbol{o}$, $\boldsymbol{s}$ resp. $\boldsymbol{p}$.

Mise à jour de la loi de a

Si on examine finement la loi *a posteriori* de l'Eq. (3.17), on remarque que la loi conditionnelle de $\boldsymbol{a}$ est déjà séparable. La mise à jour des a_i peut donc se faire de façon indépendante dans l'approche bayésienne variationnelle «classique». Dans ce cas l'algorithme 2 est inutile. C'est pourquoi, nous allons utiliser l'algorithme 1. Les équations (3.2) et (3.17) nous donnent $\forall i \in \{1, \ldots, N\}$,

$$\begin{aligned}
\tilde{q}_i^{k+1}(a_i) &= \frac{1}{K_i} \exp\left(< \log p(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{p}, \boldsymbol{a}) >_{\prod_{j \neq i} q_j^k(a_j) q^k(\boldsymbol{s}) q^k(\boldsymbol{p}) q^k(\boldsymbol{o})} \right) \\
&\propto \exp\Bigg((\frac{\nu}{2} - \frac{1}{2}) \log(a_i) \\
&- \int (\frac{p_i^2 a_i \lambda}{2} + a_i \frac{\nu}{2}) \prod_l q_l^k(p_l) \prod_{j \neq i} \tilde{q}_j^k(a_j) \mathrm{d}\boldsymbol{p} \mathrm{d}\boldsymbol{a} \Bigg) \\
&\propto \exp\left((\frac{\nu}{2} - \frac{1}{2}) \log(a_i) - a_i \frac{\nu}{2} - \int \frac{p_i^2 a_i \lambda}{2} q_i^k(p_i) \mathrm{d}p_i \right) \\
&\propto \exp\left((\frac{\nu}{2} - \frac{1}{2}) \log(a_i) - a_i \frac{\nu}{2} - \frac{(\boldsymbol{\Upsilon p}_k^2(i) + \boldsymbol{mp}_k^2(i)) a_i \lambda}{2} \right) \\
&\propto a_i^{\frac{\nu}{2} - \frac{1}{2}} \exp\left(-a_i \Big[\frac{\nu}{2} + \frac{(\boldsymbol{\Upsilon p}_k^2(i) + \boldsymbol{mp}_k^2(i)) \lambda}{2} \Big] \right)
\end{aligned} \tag{3.19}$$

On en déduit les équations de mise à jour des paramètres de la loi Gamma :

$$\forall i \in \{1, \ldots, N\}, \qquad \boldsymbol{\alpha}_{k+1}(i) = \frac{\nu}{2} + \frac{1}{2}, \tag{3.20}$$

$$\boldsymbol{\gamma}_{k+1}(i) = \frac{\nu}{2} + \frac{\lambda(\boldsymbol{\Upsilon p}_k^2(i) + \boldsymbol{mp}_k^2(i))}{2}. \tag{3.21}$$

Mise à jour de la loi de $\boldsymbol{p}$

Contrairement au cas précédent, la loi *a posteriori* conditionnelle de $\boldsymbol{p}$ n'est pas séparable, nous allons donc utiliser l'algorithme 2 afin de mettre à jour notre loi approchante. La première étape de cet algorithme est de calculer la densité intermédiaire $q_{\boldsymbol{p}}^r$ Eq. (3.11) qui appartient à la même famille que $q_{\boldsymbol{p}}$. On a pour tout i appartenant à $\{1, \ldots, N\}$:

$$\begin{aligned}
q_{p_i}^r(p_i) &= \exp\left(< \log p(\boldsymbol{y}, \boldsymbol{p}, \boldsymbol{s}, \boldsymbol{a}) >_{\prod_{j \neq i} q_{p_j}^k(p_j) q^k(\boldsymbol{s}) q^k(\boldsymbol{o}) q^{k+1}(\boldsymbol{a})} \right) \\
&\propto \exp\Bigg[- \int \left(\frac{\gamma_b \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{s} - \boldsymbol{H}\boldsymbol{p} - \boldsymbol{R}\boldsymbol{o}\|^2}{2} + \frac{\lambda p_i^2 a_i}{2} \right) \\
&\prod_{j \neq i} q_{p_j}^k(p_j) q^k(\boldsymbol{s}) q^k(\boldsymbol{o}) q^{k+1}(\boldsymbol{a}) \mathrm{d}\boldsymbol{p} \mathrm{d}\boldsymbol{s} \mathrm{d}\boldsymbol{o} \mathrm{d}\boldsymbol{a} \Bigg] \\
&\propto \exp\Bigg[- \int \Bigg(\frac{\gamma_b \boldsymbol{p}^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{p} - 2\gamma_b \boldsymbol{p}^T \boldsymbol{H}^T (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{s} - \boldsymbol{R}\boldsymbol{mo}_k)}{2} \\
&+ \frac{\lambda p_i^2 a_i}{2} \Bigg) \prod_{j \neq i} q_{p_j}^k(p_j) q^k(\boldsymbol{s}) q^{k+1}(\boldsymbol{a}) \mathrm{d}\boldsymbol{p} \mathrm{d}\boldsymbol{s} \mathrm{d}\boldsymbol{a} \Bigg] \\
&\propto \exp\Bigg[- \frac{\gamma_b}{2} \Bigg(p_i^2 \mathrm{diag}(\boldsymbol{H}^T \boldsymbol{H}) - 2p_i \Big(\boldsymbol{H}^T (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{ms}_k - \boldsymbol{R}\boldsymbol{mo}_k) \Big)_i \\
&+ 2p_i (\boldsymbol{H}^T \boldsymbol{H} - \mathrm{diag}(\boldsymbol{H}^T \boldsymbol{H})) \boldsymbol{mp}_k \Bigg) + \frac{\lambda p_i^2 \boldsymbol{\alpha}_{k+1}(i)}{2\boldsymbol{\gamma}_{k+1}(i)} \Bigg]
\end{aligned} \tag{3.22}$$

où $\mathrm{diag}(\boldsymbol{A})$ est un vecteur composé des éléments diagonaux de la matrice $\boldsymbol{A}$. On remarque que $q_{\boldsymbol{p}}^r$ reste Gaussienne de moyenne $\boldsymbol{mp}_r$ et de variance $\boldsymbol{\Upsilon p}_r^2$, ces paramètres prenant alors les valeurs suivantes :

$$\boldsymbol{\Upsilon\!\!p}_r^2(i) = \left(\gamma_b \mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H})_i + \frac{\lambda\, p_i^2 \boldsymbol{\alpha}_{k+1}(i)}{\boldsymbol{\gamma}_{k+1}(i)}\right)^{-1} \tag{3.23}$$

et

$$\boldsymbol{m\!p}_r(i) = \boldsymbol{\Upsilon\!\!p}_r^2(i) \times \gamma_b \left(\boldsymbol{H}^T(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{ms}_k - \boldsymbol{R}\boldsymbol{mo}_k) - (\boldsymbol{H}^T\boldsymbol{H} - \mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H}))\boldsymbol{m\!p}_k\right)_i \tag{3.24}$$

La deuxième étape consiste à déterminer la loi q^{k+1} en fonction du pas de la descente α_k. En injectant l'expression de $q_{\boldsymbol{p}}^r$ dans l'équation (3.10) et en utilisant les calculs détaillés dans **[56]**, on obtient les équations de mise à jour de $q_{\boldsymbol{p}}$ en fonction de α_k :

$$\begin{aligned} \boldsymbol{\Upsilon\!\!p}_{\alpha_k}^2(i) &= \frac{\boldsymbol{\Upsilon\!\!p}_r^2(i)\ \boldsymbol{\Upsilon\!\!p}_k^2(i)}{\boldsymbol{\Upsilon\!\!p}_r^2(i) + \alpha(\boldsymbol{\Upsilon\!\!p}_k^2(i) - \boldsymbol{\Upsilon\!\!p}_r^2(i))} \\ \boldsymbol{m\!p}_{\alpha_k}(i) &= \frac{\boldsymbol{m\!p}_k(i)\ \boldsymbol{\Upsilon\!\!p}_r^2(i) + \alpha_k\big(\boldsymbol{m\!p}_r(i)\ \boldsymbol{\Upsilon\!\!p}_k^2(i) - \boldsymbol{m\!p}_k(i)\ \boldsymbol{\Upsilon\!\!p}_r^2(i)\big)}{\boldsymbol{\Upsilon\!\!p}_r^2(i) + \alpha_k\big(\boldsymbol{\Upsilon\!\!p}_k^2(i) - \boldsymbol{\Upsilon\!\!p}_r^2(i)\big)}. \end{aligned} \tag{3.25}$$

La dernière étape avant la mise à jour de la loi consiste à déterminer un pas de descente α_k comme exposé dans la section 3.3.

Mise à jour de la loi de $\boldsymbol{s}$

Comme dans la partie précédente, la loi *a posteriori* conditionnelle de $\boldsymbol{s}$ n'est pas séparable et nous utilisons l'algorithme 2.

Commençons par calculer la densité intermédiaire $q_{\boldsymbol{s}}^r$, on a pour tout i appartenant à $\{1, \dots, N\}$,

$$\begin{aligned} q_{s_i}^r(s_i) &= \exp\left(< \log p(\boldsymbol{y}, \boldsymbol{p}, \boldsymbol{s}, \boldsymbol{a}) >_{\prod_{j\neq i} q_{s_j}^k(s_j) q^k(\boldsymbol{o}) q^{k+1}(\boldsymbol{p}) q^{k+1}(\boldsymbol{a})}\right) \\ &\propto \exp\Bigg[-\int\left(\frac{\gamma_b\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{s} - \boldsymbol{H}\boldsymbol{p} - \boldsymbol{R}\boldsymbol{o}\|^2}{2} + \frac{\gamma_s\|\boldsymbol{D}\boldsymbol{s}\|^2}{2}\right) \\ &\quad \prod_{j\neq i} q_{s_j}^k(s_j) q^k(\boldsymbol{o}) q^{k+1}(\boldsymbol{p}) q^{k+1}(\boldsymbol{a}) \mathrm{d}\boldsymbol{s}\mathrm{d}\boldsymbol{o}\mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{a}\Bigg] \\ &\propto \exp\Bigg[-\int\Bigg(\frac{\gamma_b \boldsymbol{s}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{s} - 2\gamma_b \boldsymbol{s}^T\boldsymbol{H}^T(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{p} - \boldsymbol{R}\boldsymbol{mo}_k)}{2} \\ &\quad + \frac{\gamma_s\|\boldsymbol{D}\boldsymbol{s}\|^2}{2}\Bigg) \prod_{j\neq i} q_{s_j}^k(s_j) q^{k+1}(\boldsymbol{p}) q^{k+1}(\boldsymbol{a}) \mathrm{d}\boldsymbol{s}\mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{a}\Bigg] \\ &\propto \exp\Bigg[-\frac{\gamma_b}{2}\Bigg(s_i^2 \mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H}) - 2s_i\Big(\boldsymbol{H}^T(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{m\!p}_{k+1} - \boldsymbol{R}\boldsymbol{mo}_k)\Big)_i \\ &\quad + 2s_i(\boldsymbol{H}^T\boldsymbol{H} - \mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H}))\boldsymbol{ms}_k\Bigg) \\ &\quad + \gamma_s(s_i^2 \mathrm{diag}(\boldsymbol{D}^T\boldsymbol{D}) + 2s_i\Big(\boldsymbol{D}^{\mathrm{t}}\boldsymbol{D}\boldsymbol{ms}_k - \mathrm{diag}(\boldsymbol{D}^{\mathrm{t}}\boldsymbol{D})\boldsymbol{ms}_k\Big)_i\Bigg] \end{aligned} \tag{3.26}$$

On remarque que $q_{\boldsymbol{s}}^r$ reste Gaussienne de moyenne $\boldsymbol{ms}_r$ et de variance $\boldsymbol{\Upsilon\!\!s}_r^2$ où

$$\boldsymbol{\Upsilon\!\!s}_r^2(i) = \left(\gamma_b \mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H})_i + \gamma_s \mathrm{diag}(\boldsymbol{D}^T\boldsymbol{D})_i\right)^{-1} \tag{3.27}$$

et

$$\begin{aligned}\boldsymbol{ms}_r(i) &= \boldsymbol{\Upsilon s}_r^2(i) \times \gamma_b \Bigg[\boldsymbol{H}^T(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{mp}_{k+1} - \boldsymbol{R}\boldsymbol{mo}_k) \\ &- (\boldsymbol{H}^T\boldsymbol{H} - \mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H}))\boldsymbol{ms}_k - \frac{\gamma_s}{\gamma_b}(\boldsymbol{D}^T\boldsymbol{D} - \mathrm{diag}(\boldsymbol{D}^T\boldsymbol{D}))\boldsymbol{ms}_k \Bigg]_i \end{aligned} \tag{3.28}$$

On remarque ici que la variance est constante ce qui facilite le calcul. Le reste des calculs est similaire à la partie 3.5.2.

Calcul de mise à jour de la loi de $\boldsymbol{o}$

En procédant de manière identique aux parties précédentes, on obtient les équations de mise à jour suivantes pour les paramètres de la loi intermédiaire :

$$\boldsymbol{\Upsilon o}_r^2(i) = \left(\gamma_b \mathrm{diag}(\boldsymbol{R}^T\boldsymbol{R})_i + \gamma_o\right)^{-1} \tag{3.29}$$

et

$$\boldsymbol{mo}_r(i) = \boldsymbol{\Upsilon o}_r^2(i) \times \gamma_b \left[\boldsymbol{R}^T(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{mp}_k - \boldsymbol{H}\boldsymbol{ms}_k) + \frac{\gamma_s}{\gamma_b} o_m \right]_i \tag{3.30}$$

3.6 Résultats

Cette partie vise à montrer, d'une part l'efficacité de l'approche gradient exponentiel pour résoudre de manière non-supervisée des problèmes de grande dimension et d'autre part le pouvoir de séparation entre les avant plans (composante impulsionnelle) et le fond diffus cosmologique (composante régulière). Pour ce faire, nous allons dans un premier temps simuler des données dans la configuration de données réelles, c'est-à-dire qu'il y a 4 millions d'observations pour estimer 700 000 inconnues. Nous avons simulé à partir d'une carte d'émission étendue mesurée par Herschel, à laquelle nous avons ajouté des sources ponctuelles. Après cette validation, nous avons appliqué notre approche à des données réelles. Nous verrons que l'apport de notre approche en terme de résolution et de correction d'artefacts est flagrant.

Tout d'abord, un des points que nous n'avons pas abordé dans ce chapitre, est le réglage des hyperparamètres. Ce réglage est particulièrement fastidieux dans le cas de la reconstruction et de la séparation conjointe de deux composantes car il y a alors quatre paramètres à régler, à savoir $\gamma_b, \gamma_o, \gamma_s$ et λ. On considère ici que le paramètre de forme de la loi de Student ν n'a pas besoin d'être réglé, il suffit de prendre une valeur permettant de favoriser les impulsions. Nous avons donc développé une approche où les quatre hyperparamètres cités ci-dessus sont estimés conjointement aux deux cartes $\boldsymbol{s}, \boldsymbol{p}$ et aux offsets $\boldsymbol{o}$.

3.6.1 Données simulées

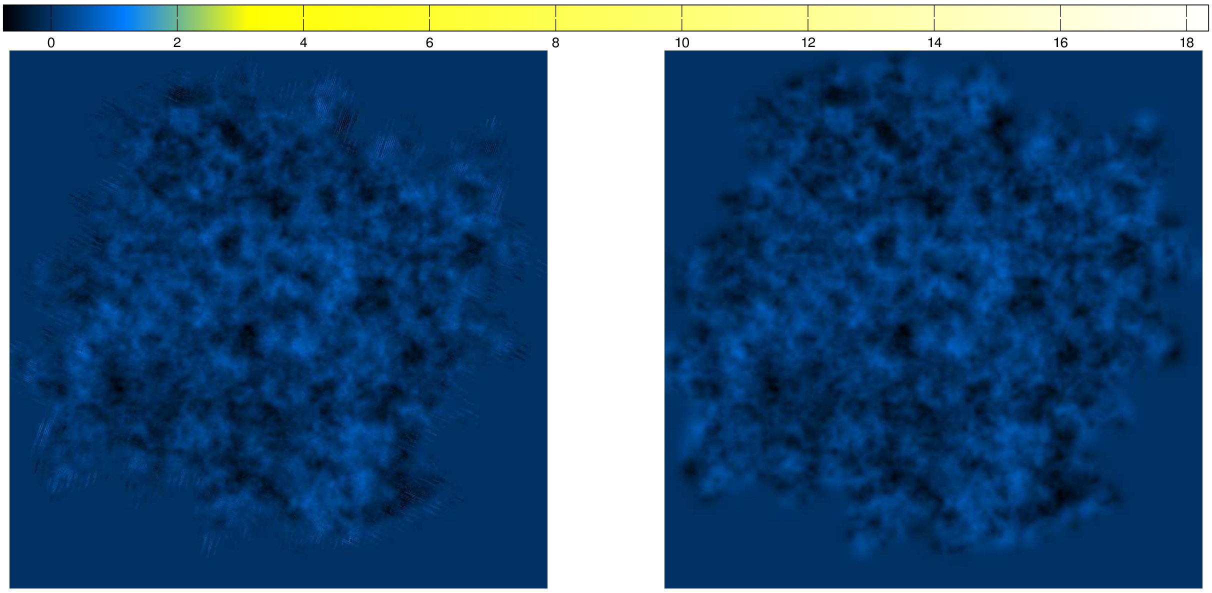
Nous présentons ici deux expérimentations numériques. Dans ces expériences, nous avons construit une image vraie à partir de deux composantes puis nous avons généré des données à l'aide de notre modèle direct en ajoutant un bruit blanc Gaussien identiquement distribué. Dans ces cas, nous n'avons pas ajouté d'offset $\boldsymbol{o}$ afin de rendre l'analyse des résultats plus facile.

La première expérimentation est une simulation réaliste. On a donc pris comme vraie carte d'émission étendue, une carte reconstruite via le code officiel du consortium Herschel à partir de données réelles (voir figure 3.1 (a)). Puis, nous avons ajouté des sources qui sont légèrement étendues, à savoir un disque de 5 pixels sur-résolu de diamètre, voir figure 3.2 (a)). Nous avons ensuite simulé des données grâce à notre modèle direct puis nous avons ajouté un bruit Gaussien iid d'écart type $0,035$, qui correspond au niveau de bruit observé sur les données réelles.

La deuxième expérience numérique a comme objectif de tester le fonctionnement de l'approche quand la composante étendue est nulle et de vérifier que la position des sources est bien retrouvée et que la photométrie, i.e. l'énergie de la source, est conservée.

Simulation réaliste

Un point particulièrement important pour les approches bayésiennes variationnelles est l'initialisation. Dans cet exemple, nous avons initialisé les moyennes $\boldsymbol{mp}$ à zéro et $\boldsymbol{ms}$ avec la carte issue de l'algorithme de ESA déconvoluée. Les variances ont été initialisées à partir d'une estimation empirique de la variance du bruit, σ_b, obtenu par le résidu issu de l'initialisation des moyennes. On prend alors $\boldsymbol{\Upsilon p}$ égal à σ_b et $\boldsymbol{\Upsilon s}$ égal à $\sigma_b/10$. Après convergence, lorsque l'évolution de l'énergie libre négative est inférieure à un seuil déterminé, nous obtenons les résultats des figures 3.1 (b) et 3.2 (b).



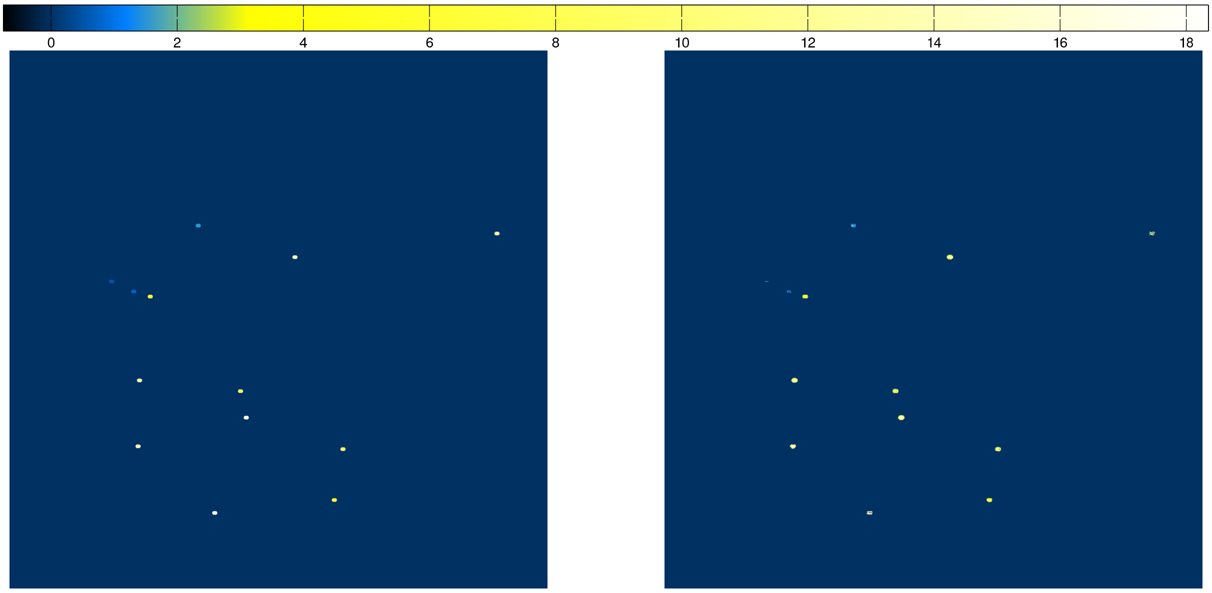
(a) (b)

FIGURE 3.1 – Composante étendue $\boldsymbol{s}$: (a) vraie carte utilisée pour simuler les données (b) carte reconstruite avec notre approche

On remarque que nous avons bien séparé la composante étendue de la composante impulsionnelle. Au delà de la séparation des sources, on arrive aussi à augmenter la qualité d'image reconstruite. Pour voir cela, nous sommons la composante étendue et la composante impulsionnelle et nous comparons le résultat obtenu, figure 3.3 (d), avec la méthode fournie par l'ESA, figure 3.3 (b), et une méthode où la vraisemblance est identique à notre approche mais où l'on utilise un *a priori* Gaussien sur les gradients de l'image, figure 3.3 (c). On voit clairement que l'*a priori* favorisant les impulsions permet d'exploiter au maximum le modèle sur-résolu. De plus, la composante impulsionnelle est très proche de la vraie, voir figure 3.3 (a).

Étude de la reconstruction des pics

Cette deuxième étude a deux objectifs. D'une part, nous voulons déterminer le comportement de la méthode en l'absence de composante étendue. D'autre part, nous souhaitons obtenir un résultat quantitatif sur la qualité de reconstruction des pics, leur position et le respect de la photométrie (énergie



(a) (b)

FIGURE 3.2 – Composante impulsionnelle : (a) vraie carte utilisée pour simuler les données (b) carte reconstruite avec notre approche.

conservée). Dans cet exemple, l'image des sources ponctuelles a été générée en choisissant les positions des sources aléatoirement ainsi que leurs amplitudes, voir figure 3.4 (a).

La composante étendue reconstruite est négligeable car son énergie est de 3 contre 190 pour la carte impulsionnelle. De plus, l'énergie de la carte des pics est concentrée sur quelques valeurs alors que celle de la carte étendue est répartie sur l'ensemble des pixels de l'image (15 contre 700 000). Aucune source n'a donc été expliquée par la composante étendue. On observe sur la figure 3.4 (a) que les positions des pics sont toutes bien estimées. La valeur des sources est elle aussi très bien estimée (voir figure 3.4 (b)). Ce dernier point est très important pour les astrophysiciens car la répartition des amplitudes des sources est un élément permettant de déterminer des paramètres cosmologiques.

Nous avons aussi étudié la qualité de reconstruction des impulsions en traçant une coupe passant par une impulsion (voir figure 3.5). On observe tout d'abord que nous avons pris les impulsions avec une certaine extension pour qu'elles soient plus visibles sur les images (voir courbe noir de la figure 3.5). Sur ce même graphique est tracé en bleu la méthode officielle de l'ESA, en vert la version de notre algorithme sur-résolu utilisant uniquement un *a priori* de douceur, et en rouge notre nouvelle approche. On remarque que l'approche sur-résolue avec *a priori* de douceur permet d'améliorer la résolution de l'image mais qu'elle introduit aussi des rebonds (effet de Gibbs), l'*a priori* de douceur modélisant mal les sources impulsionnelles. Notre nouvelle approche permet d'augmenter considérablement la résolution des impulsions tout en éliminant les effets de rebonds autour des pics.

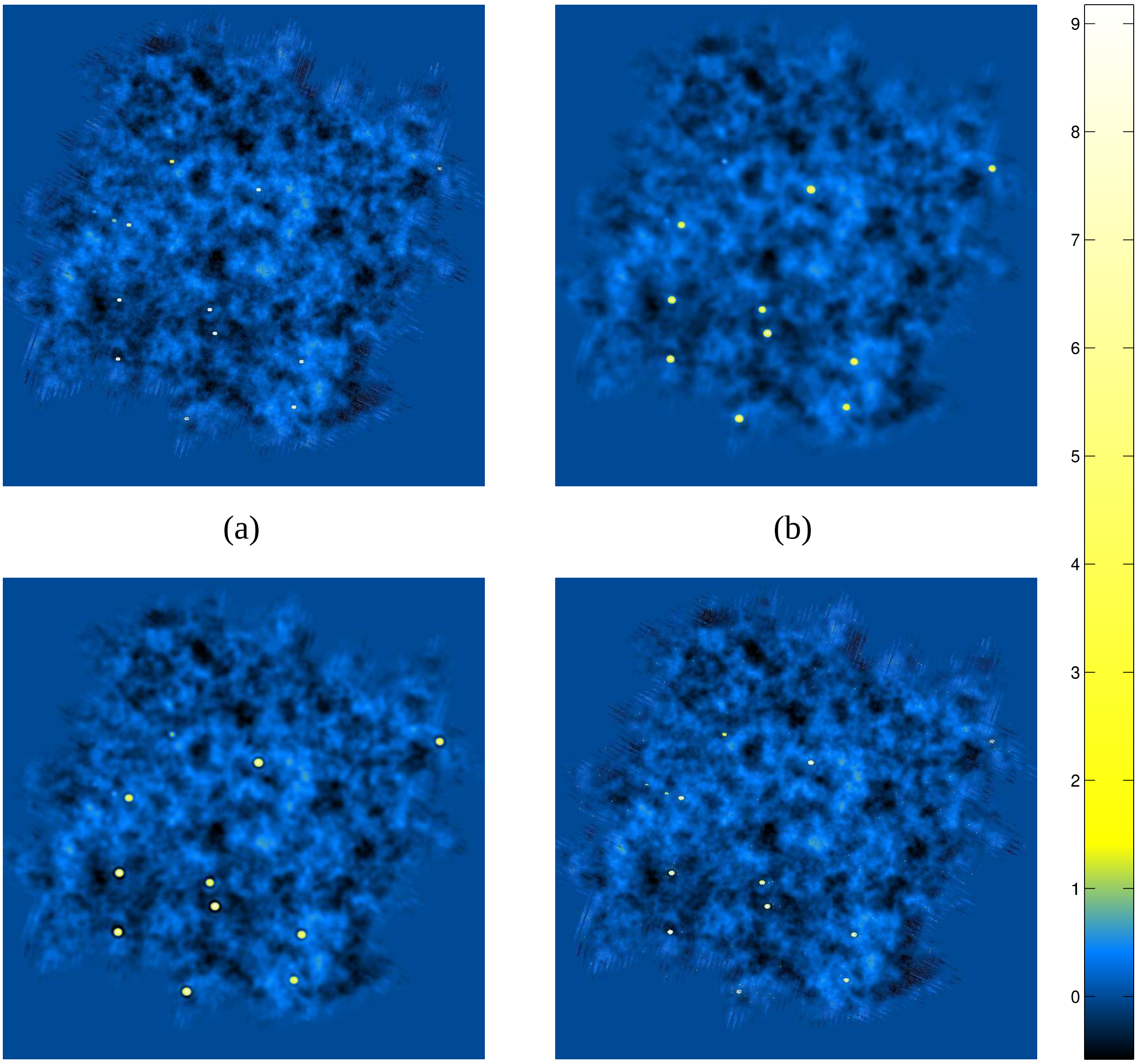
3.6.2 Données réelles

Les données utilisées sont issues du satellite Herschel de l'instrument SPIRE. Cet instrument est un imageur du rayonnement infra-rouge lointain (la longueur d'onde moyenne est de 250 μm). Les premiers résultats sont très encourageants. Cette approche est très efficace car nous estimons conjointement les offsets, la carte des impulsions, la carte étendue, la variance du bruit, la variance de la carte impulsionnelle et la variance de la carte douce. Chaque carte est composée de 700 000 inconnues et on traite en tout 4 millions de données. Le temps de calcul est seulement de 2000 secondes sur un ordinateur de bureau (Intel core 2 Duo 2.66 GHz). De plus, lorsqu'on compare le résultat obtenu par notre approche (voir figure 3.6 (b)) avec le résultat obtenu par l'algorithme de l'ESA (voir figure 3.6 (a)), on remarque que la résolution est nettement plus importante. Notre méthode, grâce à l'estimation conjointe des offsets, élimine aussi les effets systématiques (oscillations sur le bord des images).

En plus du gain en résolution, nous avons séparé les sources ponctuelles ou légèrement étendues d'un fond continu. Ces sources sont des galaxies lointaines non résolues par l'instrument. On remarque que certaines de ces sources non résolues sont expliquées par deux ou trois impulsions. Enfin, grâce à l'estimation conjointe des sources et de la carte étendue, on peut détecter des sources d'amplitude beaucoup plus faible qu'avec des approches de type “clean”. Comme ces approches se basent sur l'estimation d'un maximum dans une carte de résidus, les sources faibles sont détectées lorsque la carte étendue a un niveau important, ce qui n'est pas le cas lorsque le niveau de la carte étendue est faible. Dans ce cadre, notre méthode est donc particulièrement intéressante car elle s'affranchit de ce problème.

3.7 Conclusion

Dans ce chapitre, nous avons montré comment les approches bayésiennes, et plus particulièrement l'approche basée sur le gradient exponentiel adapté au bayésien variationnel, peuvent être utilisées dans le cadre de la reconstruction d'images issues de télescope spatiaux. Plus particulièrement, la méthode détaillée ici permet de résoudre des problèmes inverses de très grandes dimensions. En effet, on reconstruit ici des cartes de 700 000 pixels avec un ordinateur de bureau, voire 50 millions de pixels sur des super-calculateurs. Nous avons montré comment cette approche récente permet de résoudre un problème d'estimation sur-résolue conjointement à de la séparation de deux composantes : une composante lisse représentant le fond diffus cosmologique et une composante impulsionnelle, l'avant plan. Cette approche a été mise en œuvre dans le cadre d'un algorithme non supervisé, en exploitant le caractère très parcimonieux de la carte de sources impulsionnelles. Les résultats sur données simulées sont conformes à nos attentes et les résultats sur données réelles sont très bons car nous sommes arrivés à modéliser les effets prépondérants qui dégradent les reconstructions, en particulier la présence d'offsets sur les bolomètres. Cette approche bayésienne est très prometteuse car elle permet d'étendre les approches non supervisées exploitant de la parcimonie aux problèmes de très grande dimension.



(a) (b)

(c) (d)

FIGURE 3.3 – Comparaison de différents estimateurs sur la somme des deux composantes : (a) Vraie carte, (b) carte reconstruite avec le pipeline officiel de l'ESA, (c) carte reconstruite sans bi-modèle avec un *a priori* de douceur (d) carte reconstruite en utilisant un bi-modèle

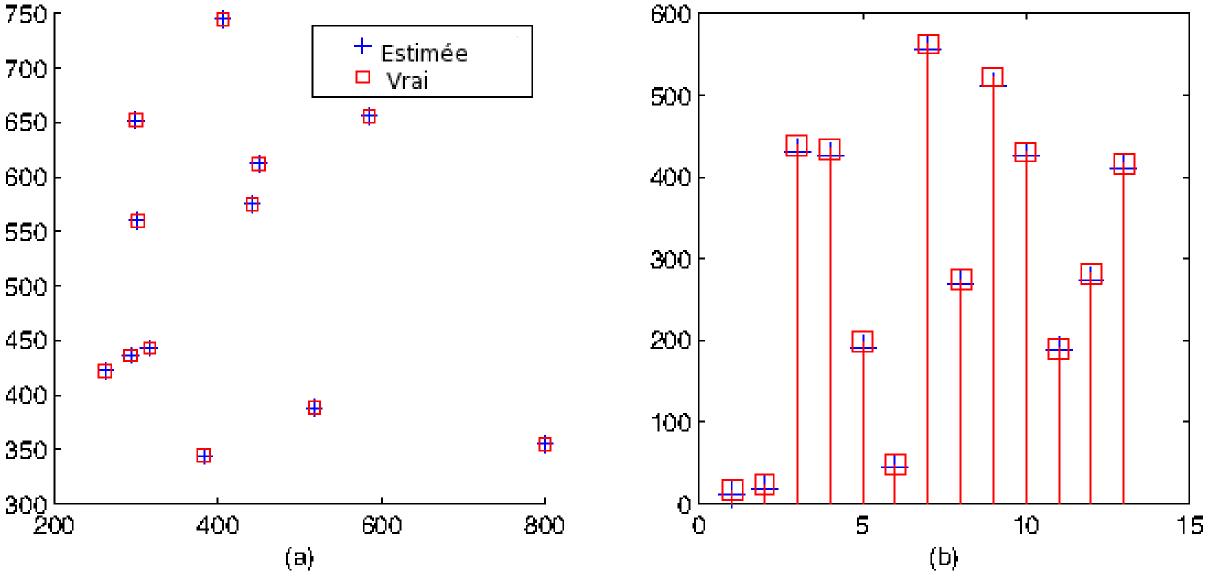


FIGURE 3.4 – Reconstruction d'une image de sources ponctuelles : (a) répartition spatiale des sources, vraie position (carré rouge), position estimée (croix bleue), (b) amplitude des treize sources dans un ordre arbitraire.

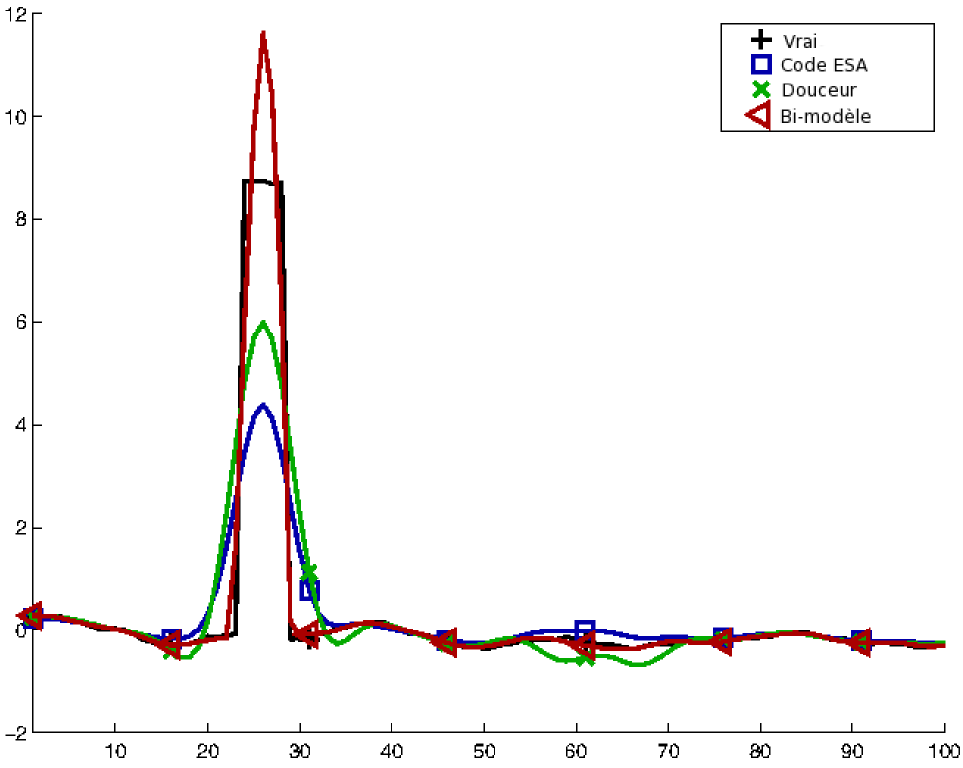
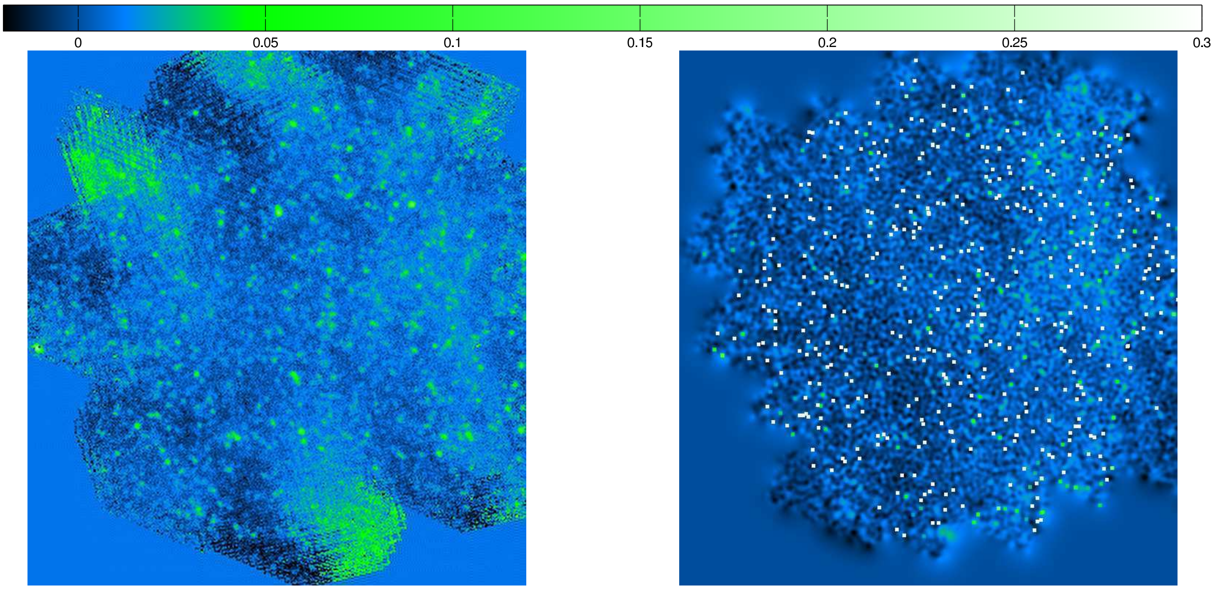
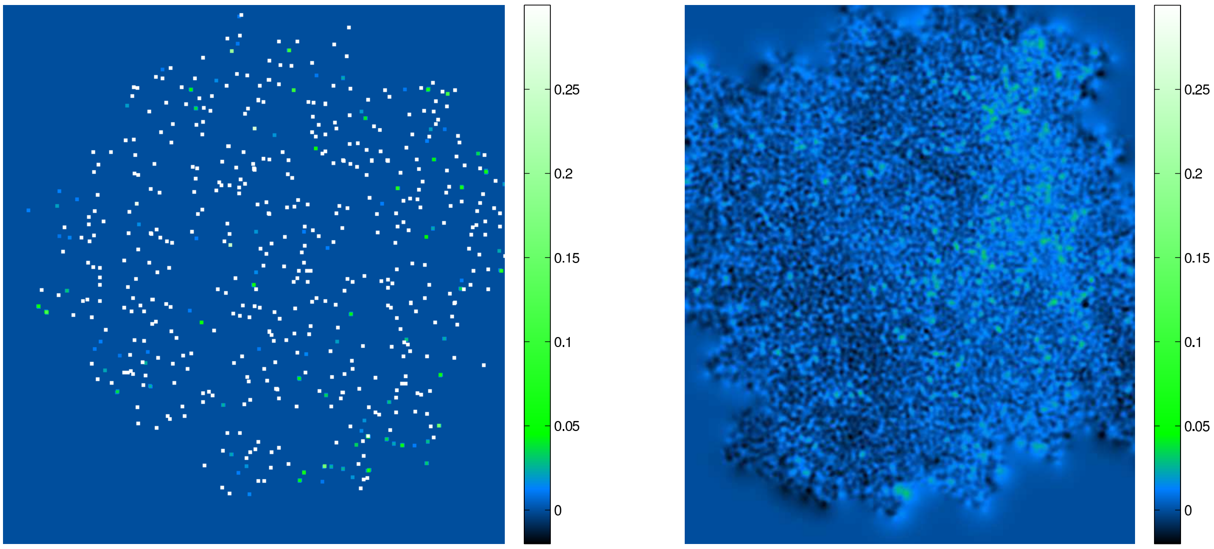


FIGURE 3.5 – Profil d'une source : (ligne continue) vrai ciel, (ligne plus carrés) algorithme de l'ESA, (ligne plus croix) reconstruction sans bi-modèle avec un *a priori* de douceur, (ligne plus triangle) reconstruction en utilisant un bi-modèle



(a) (b)

FIGURE 3.6 – Visualisation de la somme des deux composantes sur les données réelles reconstruites : (a) reconstruction à l'aide de l'algorithme de l'ESA (b) reconstruction bayésienne variationnelle avec bi-modèle



(a) (b)

FIGURE 3.7 – Reconstruction de données réelles avec notre approche : (a) composante impulsionnelle, (b) composante étendue

Chapitre 4

Perspectives

4.1 Inversion de données astrophysiques

Inversion sur-résolue de données issues d'un Spectro-imageur

Nous avons eu de très bons résultats sur l'inversion des données du spectro-imageur IRS/SPITZER. Ce travail a fait l'objet d'une publication dans *IEEE Journal of Selected Topics in Signal Processing* **[118]** (reproduit à la page 154). Dans ce travail, la composante spectrale des données est obtenue grâce à un réseau de diffraction placé sur le plan focal du détecteur. Dans les missions lancées récemment : HERCHEL (FTS) ou Mars Express (PFS), on utilise des spectro-imageurs à transformée de Fourier. Le principe de ces spectro-imageurs est comparable à celui d'un interféromètre de Michelson. Cette configuration leur confère une très bonne résolution et un faible encombrement. L'objectif serait de développer un modèle direct précis comme dans le cas d'IRS, mais aussi de développer une approche bayésienne variationnelle permettant d'avoir une solution non-supervisée en un temps réduit. Ces travaux permettront de tirer le maximum d'information des mesures réalisées. De plus, cette action participera à la diffusion des approches statistiques dans la communauté des astrophysiciens.

Échantillonnage adaptatif

Les avancées récentes en inférence bayésienne, ont permis de résoudre des nouveaux problèmes qui construisent en temps réel un schéma d'échantillonnage optimal vis à vis de l'objet que l'on est en train de reconstruire («Experimental Design»). Ce problème revient à répondre à la question suivante : Sachant que l'on a déjà fait N mesures de notre objet d'intérêt, où doit-on faire la $N + 1^e$ mesure pour que cette dernière soit la plus informative possible ? Ces approches ont déjà été introduites par M. Seeger dans le cadre des données IRM [123] et dans un cadre plus général de problèmes de grandes dimensions [122]. Les approches de M. Seeger sont conditionnées par l'utilisation d'un type d'*a priori* particulier favorisant la parcimonie. L'approche bayésienne variationnelle permet aussi de résoudre ce problème d'«Experimental Design», car elle fournit des lois simples appartenant à des familles connues donc facilement marginalisables. L'utilisation d'une plus grande variété de lois *a priori* et de vraisemblances serait donc rendue possible par cette approche. Enfin, il serait très pertinent d'appliquer cette méthodologie dans le contexte de fabrication de cartes astrophysiques. En effet, pour des raisons de protection contre le rayonnement solaire, il y a un laps de temps important entre deux séries de mesures d'une même carte. L'idée serait, après une première acquisition, de déterminer la configuration de mesure optimale permettant de mesurer uniquement les données nécessaires et ainsi gagner du temps d'acquisition qui est extrêmement coûteux sur ce type d'instrument (la minute de mesure sur le satellite Herschel revient à environ 7 000 euros).

4.2 Inversion de données tomographiques

Utilisation du rayonnement diffusé pour discriminer des matériaux

Nous avons présenté dans le § 2.2.2 une méthode permettant de discriminer différents matériaux en utilisant leurs différences d'atténuation en fonction de l'énergie de photon. Une des limitations de cette approche est que, pour certains matériaux, la différence d'atténuation est très faible. Par exemple, dans le contexte du contrôle des bagages dans les aéroports, les shampoings ont une atténuation très proche du plastique (l'explosif).

L'idée serait d'utiliser la différence de diffusion des matériaux pour les séparer. Il faut tout d'abord acquérir une partie du rayonnement diffusé pour avoir un espoir de pouvoir remonter à la fraction de rayonnement diffusé en chaque point de l'objet. Il existe des systèmes tomographiques où l'on peut facilement mesurer le rayonnement diffusé. Par exemple, les tomographes à faisceau d'électrons «Electron Beam Computed Tomography» (EBCT) sont constitués d'un anneau de détecteur fixe et le déplacement de la source de rayons X provient du déplacement d'un faisceau d'électrons sur une anode circulaire en périphérie de l'anneau de détection. Ces tomographes sont très utilisés en imagerie cardiaque [4]. Le temps d'acquisition d'un jeu de données complètes est inférieur à un cycle cardiaque. La présence de l'anneau complet de détection permet de mesurer simultanément le rayonnement direct sur une portion de l'anneau et une partie du rayonnement diffusé sur une autre portion. On peut bien sûr concevoir d'autres systèmes tomographiques possédant la même propriété.

Il faut ensuite pouvoir, à partir de ces mesures, connaître la fraction diffusé en chaque point de l'image. Pour ce faire, on peut appliquer une approche de type Monte Carlo comme ce qui a été proposé dans le cas de l'imagerie PET quantitative [84]. On peut aussi concevoir des méthodes d'inversion en s'inspirant des travaux de M. NGUYEN sur l'imagerie SPECT [95]. En réalité, la connaissance de la diffusion a un intérêt supplémentaire en plus de son pouvoir de discrimination : elle permet de déterminer précisément l'atténuation effective des matériaux qui composent l'objet, car, comme nous l'avons expliqué dans la remarque du § 2.2.2, il est important de connaître la fraction de rayonnement diffusé qui se mélange au rayonnement direct dans nos mesures. Dans ce contexte, le développement d'une méthode d'estimation de l'abondance de différents matériaux en utilisant conjointement l'information d'atténuation et de diffusion paraît très pertinente.

Traitement récursif des données tomographiques 3D + T

Comme nous l'avons vu dans le § 2.3, la reconstruction 3D + T de la couronne solaire nécessite un temps de calcul très important. De plus, les satellites STEREO A et STEREO B délivrent un flot quasiment continu de mesures. Dans le cadre de l'imagerie médicale nous avons affaire au même type de problème lorsque l'on veut développer un système de tomofluoroscopie 3D **[115][64]**[15]. Dans les deux exemples précédents, on aimerait appliquer une approche de type filtre de Kalman afin de réduire le coût de calcul et d'avoir une approche cohérente pour reconstruire un flot continu de données. Malheureusement le nombre d'inconnues est très important, donc la matrice de covariance des estimées n'est pas stockable en mémoire vive. En conséquence la mise à jour récursive de la matrice de covariance est impossible, car nous devons calculer son inverse à chaque itération. Les travaux de BUTULA *et al.* [18] résolvent partiellement ce problème de mise à jour en estimant de manière empirique la nouvelle matrice de covariance à partir d'un ensemble d'échantillons qui ont été mis à jour. Cependant, cette approche fonctionne uniquement sur des problèmes de dimension encore trop faible. De plus, contrairement à l'estimation d'une moyenne, il faut beaucoup d'échantillons pour estimer de manière fiable une matrice de covariance de grande dimension.

L'approche que l'on propose consiste à se placer dans un cadre bayésien variationnel. En effet, on peut y développer facilement des approches récursives [129][130]. Si l'on choisit un degré de séparabilité maximum associé à l'approche que nous avons exposée dans le chapitre 3 **[56]**, on s'affranchit de la matrice de covariance. La mise à jour de la matrice de covariance est remplacée par la mise à jour d'un vecteur de variance qui a la même taille que le vecteur des inconnues. Des résultats préliminaires de cette approche ont déjà été présentés sur un problème 2D + T à *«the International Conference Acoustics, Speech and Signal Processing»* **[2]**. Il serait intéressant d'appliquer cette approche sur des données réelles 3D + T, afin de comparer les performances de cette nouvelle approche par rapport à celles existantes. Dans ces premiers travaux, nous avons uniquement résolu un problème où la loi *a posteriori* était Gaussienne (ce qui correspond au filtre de Kalman), il serait très intéressant d'appliquer l'approche pour des lois *a posteriori* non Gaussiennes comme par exemple celles présentées dans le § 2.3.3 et le § 3.4. Cette approche pourrait palier les problèmes rencontré avec le filtrage particulaire [44] lorsque le nombre d'inconnues est trop important.

4.3 Méthodologie bayésienne variationnelle

Amélioration de l'approche gradient

Ce travail se fera dans le cadre de la thèse de Yuling Zheng. Il s'inscrit dans la continuité du travail déjà entamé sur l'accélération des approches bayésiennes variationnelles exposées dans le chapitre 3 et l'article **[56]**. Dans ces travaux, nous avons déterminé le gradient de notre fonctionelle puis un pas sous optimum de descente. Dans le cas de l'optimisation en dimension finie pour accélérer les algorithmes de descente, on peut soit appliquer un préconditionneur, soit déterminer un pas de descente optimum, soit déterminer une direction de descente plus efficace que l'opposé du gradient. Nous travaillerons ici sur le choix de la direction de descente. Une première idée serait d'utiliser des directions de descente conjuguées (gradient conjugué) car c'est l'approche la plus répandue en dimension finie. La transposition de cette approche dans notre problème d'optimisation fonctionnelle est beaucoup plus compliquée qu'il n'y parait. En effet, l'espace dans lequel nous cherchons à optimiser n'est pas un espace vectoriel, donc il n'existe pas de produit scalaire donc pas de conjugaison. L'idée est de transposer les approches de descente à mémoire qui sont clairement exposées dans la thèse d'E. CHOUZENOUX [29]. Ces algorithmes réduisent de manière importante le nombre d'itérations mais le coût de calcul d'une itération augmente aussi fortement. Il faudra donc déterminer si ces méthodes sont pertinentes pour notre problème d'optimisation.

Estimation robuste

On pourrait appliquer l'algorithme présenté dans le chapitre 3 à l'estimation robuste. En effet, dans beaucoup de systèmes nous voulons avoir une estimation qui est robuste aux données aberrantes. L'utilisation d'une vraisemblance gaussienne ne permet pas d'avoir une telle propriété. L'idée est de trouver une vraisemblance possédant des bonnes propriétés de conjugaison, ce qui est une condition nécessaire pour pouvoir appliquer l'approche bayésienne variationnelle. La vraisemblance doit être une loi à queue lourde afin de posséder la propriété de robustesse. Dans le cas de mesures positives nous pouvons sans doute obtenir de bons résultats en utilisant une vraisemblance Gamma.

Approches bayesiennes variationnelles et dictionnaire à base d'ondelettes

Nous pouvons aussi appliquer notre nouvelle méthodologie bayésienne variationnelle aux dictionnaires fortement redondants par exemple les ondelettes non décimées. Les ondelettes non décimées sont déjà très largement utilisées dans le domaine du débruitage d'images [24]. Il est évident qu'un schéma optimal de seuillage des coefficients d'ondelettes est de déterminer un seuil par type de coefficient (type, échelle). La difficulté de cette approche réside dans le fait qu'il faut déterminer un nombre important de seuils (environ 10 seuils pour seulement trois niveaux décomposition). Lorsque l'on résout un problème de débruitage, on peut utiliser les méthodes d'approximation de l'erreur quadratique moyenne en utilisant le lemme de Stein [87] qui donnent de très bon résultats. Ces approches restent applicables dans le cas de la déconvolution [26][109], par contre on ne peut pas les appliquer à des problèmes inverses de grande dimension.


(a) (b) (c) (d)

FIGURE 4.1 – Débruitage d'image : (a) Données bruitées SNR 20 db, (b) seuillage dur avec le seuil universel SNR 21,97 db, (c) meilleure image obtenue avec le seuillage dur 28,00 db, (d) notre approche qui détermine des «seuils» adaptatifs 27,35 db

Nous avons développé une approche permettant d'estimer conjointement les coefficients d'ondelettes et les paramètres pour chaque type d'ondelette. Pour l'instant, nous avons uniquement fait une expérimentation sur un problème de débruitage d'image. Les données bruitées de la Fig. 4.1 (a) ont été obtenues en ajoutant un bruit i.i.d. Gaussien à l'image Lena afin d'obtenir un SNR de 20 db. On utilise


(a) (b)

FIGURE 4.2 – Zoom correspondant au carré rouge dans la figure 4.1 : (a) meilleure image obtenue avec le seuillage dur, (b) notre approche

un dictionnaire d'ondelettes basé sur une transformée non décimée (Symlet 4).

On compare notre approche Fig. 4.1 (d) avec le seuillage dur sur le même dictionnaire utilisant le seuil universel Fig. 4.1(b) et en utilisant le seuil qui minimise le SNR Fig. 4.1(c) [1]. On observe que notre approche permet d'avoir une qualité d'image comparable (SNR 27,35 db contre SNR 28 db en utilisant le meilleur seuil). De plus, quand on regarde les zooms sur la Fig. 4.2 (a) et la Fig. 4.2 (b), on observe que les artefacts de seuillage dur sont moins visibles dans notre approche.

Le travail qui reste à mener, seraient dans un premier temps de comparer notre méthode avec les approches utilisant le lemme de Stein. Ensuite, on pourrait appliquer cette approche à des problèmes inverses linéaires (déconvolution, inpainting, sur-résolution, inversion de données tomographiques *etc.*...).

4.4 Sélection de modèles

Dans cette partie, nous aborderons le choix (ou la sélection) de modèles dans deux cas distincts. Le premier correspond au choix entre un petit nombre de modèles connus à l'avance. L'objectif est de déterminer le meilleur modèle vis à vis des données collectées. Dans le deuxième cas, le nombre de modèles concurrents est théoriquement infini. Il s'agit de déterminer la paramétrisation discrète (calculable et stockable sur un ordinateur) associée à une grandeur observée continue. Le choix du modèle dans ce cas revient à choisir le nombre de paramètres qui vont décrire la grandeur continue.

Choix entre un faible nombre de modèles

Le choix entre les différents modèles se fait en calculant l'évidence $p(\boldsymbol{y}|\mathcal{M}_i)$ de chaque modèle. Le modèle qui maximise l'*a posteriori* du modèle sera retenu. L'*a posteriori* du modèle se détermine à une constante près très simplement comme le produit de l'évidence par l'*a priori* sur le modèle. En pratique, il est très difficile de calculer l'évidence d'un modèle, car on connaît de manière explicite uniquement la loi

---

[1] il faut connaître la vraie image pour obtenir ce résultat

jointe $p(\boldsymbol{y}, \boldsymbol{x}|\mathcal{M}_i)$. Dans ce cas, le calcul de l'évidence consiste au calcul de l'intégrale $\int p(\boldsymbol{y}, \boldsymbol{x}|\mathcal{M}_i)\mathrm{d}\boldsymbol{x}$. Cette intégrale n'est pas calculable analytiquement sauf dans des cas très particuliers.

Il y deux méthodologies récentes qui permettent d'estimer simultanément l'évidence du modèle et la solution du problème inverse. Il s'agit des approches bayésiennes variationnelles [129] et du «nested sampling» introduit par J. SKILLING [127][128]. Pour ce qui est de l'approche bayésienne variationnelle, comme nous l'avons vu dans le § 3.3 l'Eq. (3.3) il y a égalité entre l'évidence du modèle et la somme de l'énergie libre négative et la distance de Kullback Leiber entre la loi *a posteriori* et la loi approchante q. Lorsque les algorithmes bayésiens variationels ont convergé, nous pouvons considérer que la distance de Kullback Leiber étant quasiment nulle, l'énergie libre négative donne une borne inférieur de l'évidence du modèle. Le principe du «nested sampling» est de définir une variable $\mathcal{X}(\lambda)$ qui correspond à l'intégrale de l'*a priori* pour toutes les valeurs de paramètre où la vraisemblance est supérieure à λ. De plus, si on note $\mathcal{V}(\mathcal{X}) = \lambda$ alors l'évidence du modèle est égale à l'intégrale de $\mathcal{V}$ par rapport à $\mathcal{X}$. Le calcul de l'évidence se réduit à une intégrale monodimensionnelle. Son calcul est fait de manière stochastique, en estimant l'intégrale des plus faibles valeurs de la vraisemblance jusqu'aux plus importantes. Cette approche est très efficace car les parties de l'espace où la vraisemblance est faible sont peu explorées et les parties où la vraisemblance est importante sont très explorées. Grâce à ces bonnes propriétés, cette approche a été très utilisée dans le choix de modèle pour estimer les paramètres cosmologiques à partir du fond diffus [51][52].

La difficulté que l'on rencontre dans l'approche «nested sampling» est qu'il faut déterminer à chaque étape le domaine où la vraisemblance est supérieure à une certaine valeur λ. L'idée serait d'utiliser l'approche bayésienne variationnelle pour déterminer une première approximation de l'évidence et une loi approchante de la loi *a posteriori*. La loi approchante permettra de déterminer plus efficacement (sans évaluation multiple de la vraisemblance) les domaines où la vraisemblance est supérieure à λ.

Paramétrisation d'une grandeur d'intérêt continue

Dans la majorité des problèmes inverses, la grandeur d'intérêt (notée $f(t)$) est une fonction continue, mais comme les instruments ont des résolutions limitées et que les algorithmes de traitement se font sur des ordinateurs, on la représente par un ensemble fini de coefficients discrets. L'approche la plus répandue est de considérer que la fonction est une combinaison linéaire de fonctions régulièrement réparties,

$$f(t) = \sum_i \langle f, \phi_i \rangle \phi_i(t) = \sum_i x_i \phi(t - \Delta i),$$

avec $x_i = \langle f, \phi_i \rangle$ et Δ le pas d'échantillonnage. Les fonctions de décomposition ϕ_i sont le plus souvent des indicatrices de pixels régulièrement espacées. On peut utiliser d'autres fonctions comme par exemple les fonctions de Kaiser Bessel [85] [89] ou des Gaussiennes tronquées **[118]**. L'utilisation de fonctions plus régulières que les indicatrices de pixels a un effet légèrement régularisant, mais apporte très peu d'information *a priori*.

D'autres approches plus récentes permettent de représenter la fonction continue comme une somme de fonctions de base dont la forme est variable et dont la répartition est libre (non régulièrement espacée). De plus, le nombre de composantes n'est pas fixé, il est aussi estimé. Il y a deux familles d'approches, les processus ponctuels marqués [68][133][108] et les approches bayésiennes non-paramétriques [10][47]. La première approche consiste à générer des points sur une image qui suivent un processus de Poisson ou de Strauss, puis à associer une marque à chaque point. Par exemple, le point peut définir le centre d'une Gaussienne et la marque est une fonction Gaussienne dont on estime la matrice de covariance. L'estimation de ces processus ponctuels marqués se fait à l'aide de méthodes Monte Carlo par chaînes de Markov à saut réversible (RJMCMC) [66].

Dans l'approche bayésienne non paramétrique on considère que l'inconnue est une densité de probabilité. De ce fait, l'*a priori* sur cette densité est un processus. On utilise souvent les processus de Dirichlet, ou de Pitman-Yor. L'obtention de la densité se fait grâce à des approches MCMC, dans lesquelles on utilise des échantillonneurs spécifiques («stick-breaking») afin de gérer les problèmes de troncature des lois.

La limitation de ces approches réside dans la complexité de calculs lorsque que le nombre de fonctions de base devient élevé (> 10 000). L'idée serait par exemple d'étendre le modèle de reconstruction et de séparation conjointe de sources ponctuelles et de sources étendues décrit dans **[7]** par un modèle variationnel pour la composante étendue et un modèle non paramétrique pour la composante impulsionnelle. Nous limiterions ainsi le nombre de fonctions de base et nous pourrions accéder comme cela à la position des étoiles avec une précision sub-pixelique. Des premiers travaux sur des modèles très simples (AR) ont déjà proposé de combiner non paramétrique et bayésien variationnel [92].

Bibliographie

[1] S. Achenbach, D. Ropers, A. Kuettner, T. Flohr, B. Ohnesorge, H. Bruder, H. Theessen, M. Karakaya, W. G. Daniel, W. Bautz, et al. Contrast-enhanced coronary artery visualization by dual-source computed tomography-initial experience. *European Journal of Radiology*, 57(3) :331–335, 2006.

[2] B. Ait el Fquih and T. Rodet. Variational bayesian Kalman filtering in dynamical tomography. In *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, pages 4004 – 4007, Prague, mai 2011.

[3] H. Alfvén. Existence of electromagnetic-hydrodynamic waves. *Nature*, 150(3805) :405–412, 1942.

[4] Y. Arad, L. A. Spadaro, K. Goodman, D. Newstein, and A. D. Guerci. Prediction of coronary events with electron beam computed tomography. *Journal of the American College of Cardiology*, 36(4) :1253–1260, 2000.

[5] H. Ayasso. *Une approche bayésienne de l'inversion. Application à l'imagerie de diffraction dans les domaines micro-onde et optique*. Thèse de Doctorat, Université Paris Sud, décembre 2010.

[6] H. Ayasso and A. Mohammad-Djafari. Joint NDT image restoration and segmentation using gauss-markov-potts prior models and variational bayesian computation. *IEEE Transactions on Image Processing*, 19(9) :2265 – 2277, septembre 2010.

[7] H. Ayasso, T. Rodet, and A. Abergel. A variational bayesian approach for unsupervised super-resolution using mixture models of point and smooth sources applied to astrophysical mapmaking. Technical report, LSS, 33 pages, 2012. Submitted to Inverses Problem.

[8] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variationnal Bayesian Blind Deconvolution Using a Total Variation Prior. *IEEE Transactions on Image Processing*, 18(1) :12–26, janvier 2009.

[9] N. Bali and A. Mohammad-Djafari. Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis. *IEEE Transactions on Image Processing*, 17(2) :217–225, Feb. 2008.

[10] E. Barat, T. Dautremer, and T. Montagu. Nonparametric bayesian inference in nuclear spectrometry. In *Proceedings of IEEE Medical Imaging Conference*, volume 1, pages 880–887. IEEE, 2007.

[11] N. Barbey. *Détermination de la structure tridimensionnelle de la couronne solaire à partir d'images des missions spatiales SoHO et STEREO*. Thèse de Doctorat, Université Paris–sud 11, Orsay, décembre 2008.

[12] N. Barbey, F. Auchère, T. Rodet, and J.-C. Vial. A time-evolving 3D method dedicated to the reconstruction of solar plumes and results using extreme ultra-violet data. *Solar Physics*, 248(2) :409–423, avril 2008.

[13] P. Bleuet. *Reconstruction 3D par tomosynthèse généralisée. Application à l'imagerie médicale par rayons X.* Thèse de Doctorat, Institut National des Sciences Appliquées de Lyon, 2002.

[14] C. Blondel, R. Vaillant, G. Malandain, and N. Ayache. 3D tomographic reconstruction of coronary arteries using a precomputed 4D motion field. *Physics in Medicine and Biology*, 49 :2197–2208, 2004.

[15] S. Bonnet, A. Koenig, S. Roux, P. Hugonnard, R. Guillemaud, and P. Grangeat. Dynamic X-ray computed tomography. *Proceedings of the IEEE*, 91(10) :1574–1587, 2003.

[16] C. Bontus, T. Kohler, and R. Proksa. EnPiT : filtered back-projection algorithm for helical CT using an n-Pi acquisition. *IEEE Transactions on Medical Imaging*, 24(8) :977–986, 2005.

[17] D. Brasse, P. E. Kinahan, R. Clackdoyle, M. Defrise, C. Comtat, and D. W. Townsend. Fast fully 3-D image reconstruction in PET using planograms. *IEEE Transactions on Medical Imaging*, 23(4) :413–425, avril 2004.

[18] M. D. Butala, F. Kamalabadi, R. A. Frazin, and Y. Chen. Dynamic tomographic imaging of the solar corona. *IEEE Journal of Selected Topics in Signal Processing*, 2(5) :755–766, 2008.

[19] C. Cai, A. Mohammad-Djafari, S. Legoupil, and T. Rodet. Bayesian data fusion and inversion in X–ray multi–energy computed tomography. In *Proceedings of the International Conference on Image Processing*, pages 1377 – 1380, septembre 2011.

[20] C. Cai, T. Rodet, S. Legoupil, and A. Mohammad-Djafari. Bayesian reconstruction of water and bone fractions in polychromatic multi-energy computed tomography. *Soumis à Medical Physics*, 2012.

[21] J.-F. Cardoso. Blind signal separation : statistical principles. *Proceedings of the IEEE*, 9 :2009–2025, 1998.

[22] P. J. Cargill and J. A. Klimchuk. Nanoflare heating of the corona revisited. *The Astrophysical Journal*, 605 :911–920, 2004.

[23] F. Champagnat and J. Idier. A connection between half-quadratic criteria and EM algorithm. *IEEE Signal Processing Letters*, 11(9) :709–712, septembre 2004.

[24] S. G. Chang, B. Yu, and M. Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9) :1532–1546, 2000.

[25] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders. Variational Bayesian image restoration based on a product of t-distributions image prior. *IEEE Transactions on Image Processing*, 17(10) :1795–1805, octobre 2008.

[26] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet. A nonlinear stein based estimator for multichannel image denoising. *IEEE Transactions on Signal Processing*, 56(8) :3855–3879, août 2008.

[27] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. Technical report, Stanford University, février 1996.

[28] R. A. Choudrey. *Variational Methods for Bayesian Independent Component Analysis*. Thèse de Doctorat, University of Oxford, 2002.

[29] E. Chouzenoux. *Recherche de pas par Majoration-Minoration. Application à la résolution de problèmes inverses.* Thèse de Doctorat, École Centrale de Nantes, 2010.

[30] P. Ciuciu, J. Idier, and J.-F. Giovannelli. Analyse spectrale non paramétrique haute résolution. In *Actes du 17$^{\text{e}}$ colloque GRETSI*, pages 721–724, Vannes, septembre 1999.

[31] P. Comon and C. Jutten. *Handbook of Blind Source Separation : Independent component analysis and applications*. Academic Press, 2010.

[32] P. Comon, C. Jutten, and J. Herault. Blind separation of sources, part II : Problems statement. *Signal processing*, 24(1) :11–20, 1991.

[33] P. E. Danielsson, P. Edholm, J. Eriksson, and M. Magnusson. Towards exact reconstruction for helical cone-beam scanning of long objects. a new detector arrangement and a new completeness condition. In DW Townsend and PE Kinahan, editors, *Proc. 1997 Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, pages 141–144, 1997.

[34] B. De Man. *Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography*. Thèse de Doctorat, Katholieke Universiteit Leuven (Belgium), 2001.

[35] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Physics in Medicine and Biology*, 49(11) :2463–75, juin 2004.

[36] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. Reduction of metal streak artifacts in x-ray computed tomography using a transmission maximum a posteriori algorithm. In *Proceedings of IEEE Medical Imaging Conference*, volume 2, pages 850–854. IEEE, 1999.

[37] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. An iterative maximum-likelihood polychromatic algorithm for CT. *IEEE Transactions on Medical Imaging*, 20(10) :999–1008, 2001.

[38] C. De Mol and M. Defrise. Inverse imaging with mixed penalties. In *URSI Symposium on Electromagnetic Theory*, pages 798–800, Pisa Italie, 2004. Ed. PLUS Univ. Pisa.

[39] G. Del Zanna, B. J. I. Bromage, and H. E. Mason. Spectroscopic characteristics of polar plumes. *Astronomy and Astrophysics*, 398(2) :743–761, 2003.

[40] G. Demoment. Image reconstruction and restoration : Overview of common estimation structure and problems. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-37(12) :2024–2036, décembre 1989.

[41] G. Demoment, J. Idier, J.-F. Giovannelli, and A. Mohammad-Djafari. Problèmes inverses en traitement du signal et de l'image. volume TE 5 235 of *Traité Télécoms*, pages 1–25. Techniques de l'Ingénieur, Paris, 2001.

[42] L. Desbat. Efficient sampling on coarse grids in tomography. *Inverse Problems*, 9 :1141–1148, 1993.

[43] L. Desbat. *Échantillonnage efficace en tomographie.* Habilitation à diriger des recherches, Université Joseph Fourier - Grenoble I, Octobre 1997.

[44] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3) :197–208, 2000.

[45] I. A. Elbakri and J. A. Fessler. Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation. *Physics in Medicine and Biology*, 48 :2453, 2003.

[46] H. Erdogan and J. A. Fessler. Monotonic algorithms for transmission tomography. *IEEE Transactions on Medical Imaging*, 18(9) :801–814, septembre 1999.

[47] M. D. Fall. *Modélisation stochastique de processus pharmaco-cinétiques, applications à la reconstruction tomographique par émission de positrons (TEP) spatio-temporelle*. Thèse de Doctorat, Université Paris-Sud 11, 2012.

[48] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America (A)*, 1(6) :612–619, 1984.

[49] U. Feldman, E. Landi, and N. A. Schwadron. On the sources of fast and slow solar wind. *Journal of Geophysical Research*, 110(A07109) :1–12, 2005.

[50] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *ACM Transactions on Graphics (TOG)*, 25(3) :787–794, 2006.

[51] F. Feroz and M. P. Hobson. Multimodal nested sampling : an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Monthly Notices of the Royal Astronomical Society*, 384(2) :449–463, 2008.

[52] F. Feroz, M. P. Hobson, and M. Bridges. MultiNest : an efficient and robust bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4) :1601–1614, 2009.

[53] J. A. Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Transactions on Medical Imaging*, 13(2) :290–300, 1994.

[54] J. A. Fessler, I. A. Elbakri, P. Sukovic, and N. H. Clinthorne. Maximum-likelihood dual-energy tomographic image reconstruction. In *SPIE Medical Imaging*. spie, 2002.

[55] A. Fraysse and T. Rodet. A gradient-like variational bayesian algorithm. In *Proceedings of IEEE Statistical Signal Processing Workshop (SSP)*, number S17.5, pages 605 – 608, Nice, France, juin 2011.

[56] A. Fraysse and T. Rodet. A measure-theoretic variational bayesian algorithm for large dimensional problems. Technical report, LSS UMR 8506, 36 pages, 2012. http ://hal.archives-ouvertes.fr/docs/00/70/22/59/PDF/var_bayV8.pdf.

[57] R. A. Frazin. Tomography of the Solar Corona. I. A Robust, Regularized, Positive Estimation Method. *Astrophysical Journal*, 530 :1026–1035, février 2000.

[58] R. A. Frazin and P. Janzen. Tomography of the Solar Corona. II. Robust, Regularized, Positive Estimation of the Three-dimensional Electron Density Distribution from LASCO-C2 Polarized White-Light Images. *Astrophysical Journal*, 570 :408–422, mai 2002.

[59] N. Gac, A. Mohammad-Djafari, A. Vabre, and F. Buyens. GPU implementation of a 3D bayesian CT algorithm and its application on real foam reconstruction. In *First international conference on image formation in X-ray computes tomography*, Salt lake city, USA, June 2010.

[60] H. Gao, H. Yu, and G. Wang. Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM). In *Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 124–7, Postdam, Germany, juillet 2011.

[61] J.-F. Giovannelli. Unsupervised bayesian convex deconvolution based on a field with an explicit partition function. *IEEE Transactions on Image Processing*, 17(1) :16–26, janvier 2008.

[62] J.-F. Giovannelli and A. Coulais. Positive deconvolution for superimposed extended source and point sources. *Astronomy and Astrophysics*, 439 :401–412, 2005.

[63] Y. Goussard, G. Demoment, and F. Monfront. *Maximum a posteriori dectection-estimation of Bernoulli-Gaussian processes*. Mathematics in Signal Processing II. Clarendon Press, Oxford, UK, 1990.

[64] P. Grangeat, A. Koenig, T. Rodet, and S. Bonnet. Theoretical framework for a dynamic cone–beam reconstruction algorithm based on a dynamic particle model. *Physics in Medicine and Biology*, 47(15) :2611–2625, August 2002.

[65] P. V. Granton, S. I. Pollmann, N. L. Ford, M. Drangova, and D. W. Holdsworth. Implementation of dual-and triple-energy cone-beam micro-CT for postreconstruction material decomposition. *Medical Physics*, 35 :5030–42, 2008.

[66] P. J. Green. Reversible jump MCMC computation and Bayesian model determinaion. *Biometrika*, 82(4) :711–732, 1995.

[67] Y. Grondin, L. Desbat, M. Defrise, T. Rodet, N. Gac, M. Desvignes, and S. Mancini. Data sampling in multislice mode PET for multi-ring scanner. In *Proceedings of IEEE Medical Imaging Conference*, volume 4, pages 2180–2184, San Diego, CA, USA, 2006.

[68] O. Häggström, M.-C. N. M. Van Lieshout, and J. Møller. Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli*, 5(4) :641–658, 1999.

[69] C. Hong, C. R. Becker, A. Huber, U. J. Schoepf, B. Ohnesorge, A. Knez, R. Brüning, and M. F. Reiser. ECG-gated reconstructed multi-detector row CT coronary angiography : Effect of varying trigger delay on image quality. *Radiology*, 220(3) :712–717, 2001.

[70] M. M. Ichir and A. Mohammad-Djafari. Hidden Markov models for wavelet-based blind source separation. *IEEE Transactions on Image Processing*, 15(7) :1887–1899, July 2006.

[71] J. Idier, editor. *Approche bayésienne pour les problèmes inverses*. Traité IC2, Série traitement du signal et de l'image, Hermès, Paris, 2001.

[72] J. Idier, editor. *Bayesian Approach to Inverse Problems*. ISTE Ltd and John Wiley & Sons Inc., London, 2008.

[73] S. M. Johnston, G. A. Johnson, and C. T. Badea. A material decomposition method for dual energy micro-CT. In *Proceedings of SPIE*, volume 725841, pages 1–12, 2009.

[74] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An Introduction to variational Methods for Graphical Models. *Machine Learning*, 37(2) :183–233, 1999.

[75] P. M. Joseph. An improved algorithm for reprojecting rays through pixel images. *IEEE Transactions on Medical Imaging*, MI-1(3) :192–196, mars 1982.

[76] C. Jutten and J. Herault. Blind separation of sources, part I : An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1) :1–10, 1991.

[77] M. Kachelrieß, S. Schaller, and W. A. Kalender. Advanced single-slice rebinning in cone-beam spiral CT. *Medical Physics*, 27 :754, 2000.

[78] A. Katsevich. Theoretically exact filtered backprojection-type inversion algorithm for spiral CT. *SIAM Journal on Applied Mathematics*, pages 2012–2026, 2002.

[79] B. Kherrab, T. Rodet, and J. Idier. Solving a problem of sources separation stemming from a linear combination : Applied to the 3D reconstruction of the solar atmosphere. In *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, pages 1338 – 1341, Dallas, USA, 2010. DOI : 10.1109/ICASSP.2010.5495439.

[80] J. Kivinen and M. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1) :1–63, 1997.

[81] M. Kowalski and T. Rodet. An unsupervised algorithm for hybrid/morphological signal decomposition. In *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, number Id 2155, pages 4112 – 4115, Prague, mai 2011.

[82] H. Kudo, F. Noo, and M. Defrise. Quasi-exact filtered backprojection algorithm for long-object problem in helical cone-beam tomography. *IEEE Transactions on Medical Imaging*, 19(9) :902–921, 2000.

[83] H. Kudo, T. Rodet, F. Noo, and M. Defrise. Exact and approximate algorithms for helical cone-beam CT. *Physics in Medicine and Biology*, 49(13) :2913–2931, juillet 2004.

[84] D. Lazaro, Z. El Bitar, V. Breton, D. Hill, and I. Buvat. Fully 3D Monte Carlo reconstruction in SPECT : a feasibility study. *Physics in Medicine and Biology*, 50 :3739–54, 2005.

[85] R. M. Lewitt. Alternative to voxels for image representation in iterative reconstruction algorithms. *Physics in Medicine and Biology*, 37 :705–716, 1992.

[86] A. Llebaria, P. Lamy, C. E. DeForest, and S. Koutchmy. Time domain analysis of polar plumes observed with LASCO-C2 and EIT. In *Solar Jets and Coronal Plumes*, volume 421, pages 87–92, 1998.

[87] F. Luisier, T. Blu, and M. Unser. Image denoising in mixed Poisson-Gaussian noise. *IEEE Transactions on Image Processing*, 20(3) :696 – 708, 2010.

[88] D. J. C. MacKay. Ensemble learning and evidence maximization. http ://citeseerx.ist.psu.edu/viewdoc/summary ?doi=10.1.1.54.4083, 1995.

[89] S. Matej and R. M. Lewitt. Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *IEEE Transactions on Medical Imaging*, 15 :68–78, janvier 1996.

[90] V. Mazet. *Développement de méthodes de traitement de signaux spectroscopiques : estimation de la ligne de base et du spectre de raies.* Thèse de Doctorat, Université Henri Poincaré, Nancy 1, 2005.

[91] J.W Miskin. *Ensemble Learning for Independent Component Analysis*. Thèse de Doctorat, University of Cambridge, `http://www.inference.phy.cam.ac.uk/jwm1003/`, 2000.

[92] K. Morton, P. Torrione, and L. Collins. Variational bayesian learning for mixture autoregressive models with uncertain-order. *IEEE Transactions on Signal Processing*, 59(6) :2614–2627, 2011.

[93] E. Moulines, J.-F. Cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, Munich, Allemagne, avril 1997.

[94] N. Murata, S. Ikeda, and A. Ziehe. An approach to blind source separation based on temporal structure of speech signals. *Neurocomputing*, 41(1) :1–24, 2001.

[95] M. K. Nguyen and T. T. Truong. On an integral transform and its inverse in nuclear imaging. *Inverse Problems*, 18 :265, 2002.

[96] J. Nocedal and S. J. Wright. *Numerical Optimization*. Series in Operations Research. Springer Verlag, New York, 2000.

[97] F. Noo, M. Defrise, R. Clackdoyle, and H. Kudo. Image reconstruction from fan-beam projections on less than a short scan. *Physics in Medicine and Biology*, 47 :2525–2546, 2002.

[98] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. *IEEE Transactions on Medical Imaging*, 18(5) :393–403, 1999.

[99] J. Nuyts, B. D. Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans. Iterative reconstruction for helical CT : a simulation study. *Physics in Medicine and Biology*, 43 :729, 1998.

[100] F. Orieux. *Inversion bayésienne myope et non-supervisée pour l'imagerie sur-résolue. Application à l'instrument SPIRE de l'observatoire spatial Herschel.* Thèse de Doctorat, Université Paris-Sud 11, novembre 2009.

[101] F. Orieux, J.-F. Giovannelli, and T. Rodet. Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution. *Journal of the Optical Society of America (A)*, 27(7) :1593–1607, 2010.

[102] F. Orieux, J.-F. Giovannelli, T. Rodet, A. Abergel, H. Ayasso, and M. Husson. Superresolution in map-making based on physical instrument model and regularized inversion. Application to SPIRE/Herschel. *Astronomy and Astrophysics*, 539(A38) :16, mars 2012.

[103] S. S. Orlov. Theory of three dimensional reconstruction. I. Conditions for a complete set of projections. *Soviet Physics Crystallography*, 20(3) :312–314, 1975.

[104] J. D. Pack, F. Noo, and H. Kudo. Investigation of saddle trajectories for cardiac CT imaging in cone-beam geometry. *Physics in Medicine and Biology*, 49 :2317, 2004.

[105] G. Parisi. *Statistical field theory*. Addison Wesley, Redwood City,CA, 1988.

[106] D. L. Parker. Optimal short scan convolution reconstruction for fan-beam CT. *Medical Physics*, 9(2) :245–257, 1982.

[107] L. Parra and C. Spence. Convolutive blind separation of non-stationary sources. *IEEE Transactions on Speech and Audio Processing*, 8(3) :320–327, 2000.

[108] G. Perrin, X. Descombes, and J. Zerubia. A marked point process model for tree crown extraction in plantations. In *Proceedings of the International Conference on Image Processing*, volume 1, pages I–661. IEEE, 2005.

[109] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A SURE approach for digital signal/image deconvolution problems. *IEEE Transactions on Signal Processing*, 57(12) :4616–4632, décembre 2009.

[110] F. Pfanner, M. Knaup, and M. Kachelriess. High performance parallel backprojection on FPGA. In *3rd Workshop on High Performance Image Reconstruction*, pages 23–26, Postdam, Germany, juillet 2011.

[111] O. Rabaste and T. Chonavel. Estimation of multipath channels with long impulse response at low SNR via an MCMC method. *IEEE Transactions on Signal Processing*, 55(4) :1312 – 1325, avril 2007.

[112] J. Radon. Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Berichte Sächsische Akademie der Wissenschaften, Math.-Phys.*, 69 :262–267, 1917.

[113] C. Robert. *Simulations par la méthode MCMC*. Economica, Paris, 1997.

[114] C. P. Robert and G. Casella. *Monte-Carlo Statistical Methods*. Springer Texts in Statistics. Springer, New York, NY, USA, 2000.

[115] T. Rodet. *Algorithmes rapides de reconstruction en tomographie par compression des calculs. Application à la tomofluoroscopie 3D*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, octobre 2002.

[116] T. Rodet, F. Noo, and M. Defrise. The cone–beam algorithm of Feldkamp, Davis, and Kress preserves oblique line integral. *Medical Physics*, 31(7) :1972–1975, juillet 2004.

[117] T. Rodet, J. Nuyts, M. Defrise, and C Michel. A study of data sampling in PET with planar detectors. In *Proceedings of IEEE Medical Imaging Conference*, number M8-8, Portland, Oregon, USA, novembre 2003.

[118] T. Rodet, F. Orieux, J.-F. Giovannelli, and A. Abergel. Data inversion for over-resolved spectral imaging in astronomy. *IEEE Journal of Selected Topics in Signal Processing*, 2(5) :802–811, octobre 2008.

[119] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, 1987.

[120] S. Sawall, L. Ritschl, M. Knaup, and M. Kachelriess. Performance comparison of openCL and CUDA by benchmarking an optimized perspective backprojection. In *3rd Workshop on high Performance Image Reconstruction*, pages 15–18, Postdam, Germany, juillet 2011.

[121] D. Schafer, J. Borgert, V. Rasche, and M. Grass. Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography. *IEEE Transactions on Medical Imaging*, 25(7) :898–906, 2006.

[122] M. Seeger and Nickisch H. Large scale bayesian inference and experimental design for sparse linear models. *SIAM Journal on Imaging Sciences*, 4(1) :166–199, 2011.

[123] M. Seeger, H. Nickisch, R. Pohmann, and B. Schoelkopf. Optimization of k-space trajectories for compressed sensing by bayesian experimental design. *Magnetic Resonance in Medicine*, 63(1) :116–126, 2009.

[124] M. Seeger and D. P. Wipf. Variational bayesian inference techniques. *IEEE Signal Processing Magazine*, 27(6) :81–91, november 2010.

[125] R. L. Siddon. Fast calculation of the exact radiological path for a three-dimensional CT array. *Medical Physics*, 12 :252–255, 1985.

[126] E. Y. Sidky, Y. Zou, and X. Pan. Impact of polychromatic X-ray sources on helical, cone-beam computed tomography and dual-energy methods. *Physics in Medicine and Biology*, 49 :2293, 2004.

[127] J. Skilling. Nested sampling. In *American Institute of Physics Conference Proceedings*, volume 735, page 395, 2004.

[128] J. Skilling. Nested sampling for general bayesian computation. *Bayesian Analysis*, 1(4) :833–860, 2006.

[129] V. Smidl and A. Quinn. *The variational Bayes Method in Signal Processing*. Springer, 2006.

[130] V. Smidl and A. Quinn. Variational bayesian filtering. *IEEE Transactions on Signal Processing*, 56(10) :5020–5030, octobre 2008.

[131] H. Snoussi and A. Mohammad-Djafari. Bayesian unsupervised learning for source separation with mixture of gaussians prior. *International Journal of VLSI Signal Processing Systems*, 37 :263–279, 2004.

[132] K. Sourbelle, H. Kudo, G. Lauritsch, K. C. Tam, M. Defrise, and F. Noo. Performance evaluation of exact cone-beam algorithms for the long-object problem in spiral computed tomography. In *Fully 3D Reconstruction in Radiology and Nuclear Medicine*, pages id 7–10, Pacific Grove CA USA, novembre 2001.

[133] R. Stoica, X. Descombes, and J. Zerubia. A Gibbs point process for road extraction from remotely sensed images. *International Journal of Computer Vision*, 57(2) :121–136, 2004.

[134] P. Sukovic and N. H. Clinthorne. Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography. *IEEE Transactions on Medical Imaging*, 19(11) :1075–1081, 2000.

[135] K. C. Tam, G. Lauritsch, K. Sourbelle, F. Sauer, and B. Ladendorf. Exact (spiral+ circles) scan region-of-interest cone beam reconstruction via backprojection. *IEEE Transactions on Medical Imaging*, 19(5) :376–383, 2000.

[136] H. Tuy. An inversion formula for cone-beam reconstruction. *SIAM Journal of Applied Mathematics*, 43 :546–552, 1983.

[137] H. C. Van de Hulst. The electron density of the solar corona. *Bulletin of the Astronomical Institutes of the Netherlands*, 11 :135–149, 1950.

[138] S. S. Vedam, P. J. Keall, V. R. Kini, H. Mostafavi, H. P. Shukla, and R. Mohan. Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Physics in Medicine and Biology*, 48 :45–62, 2003.

[139] M. J. Wainwright and E. P. Simoncelli. Scale mixtures of gaussians and the statistics of natural images. *Advances in Neural Information Processing Systems*, 12(1) :855–861, 2000.

[140] F. Ward. Determination of the solar-rotation rate from the motion of identifiable features. *The Astrophysical Journal*, 145 :416–425, 1966.

[141] T. Wiegelmann and B. Inhester. Magnetic modeling and tomography : First steps towards a consistent reconstruction of the solar corona. *Solar Physics*, 214 :287–312, juin 2003.

[142] D. P. Wipf and B. D. Rao. Sparse Bayesian learning for basis selection. *IEEE Transactions on Signal Processing*, 52(8) :2153–2164, 2004.

[143] H. Yu and G. Wang. Data consistency based rigid motion artifact reduction in fan-beam CT. *IEEE Transactions on Medical Imaging*, 26(2) :249–260, 2007.

[144] Y. Zheng, C. Cai, and T. Rodet. Joint reduce of metal and beam hardening artifacts using multi-energy map approach in X-ray computed tomography. In *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, number Id 3037, pages 737–740, mai 2011.

Troisième partie

Sélection d’articles

A. Fraysse et **T. Rodet**, « A measure-theoretic variational bayesian algorithm for large dimensional problems. », Soumis à *SIAM Imaging Sciences*, 36 pages, 2012.

A MEASURE-THEORETIC VARIATIONAL BAYESIAN ALGORITHM FOR LARGE DIMENSIONAL PROBLEMS.

A. FRAYSSE AND T. RODET*

Abstract. In this paper we provide an algorithm allowing to solve the variational Bayesian issue as a functional optimization problem. The main contribution of this paper is to transpose a classical iterative algorithm of optimization in the metric space of measures involved in the Bayesian methodology. Once given the convergence properties of this algorithm, we consider its application to large dimensional inverse problems, especially for unsupervised reconstruction. The interest of our algorithm is enhanced by its application to large dimensional linear inverse problems involving sparse objects. Finally, we provide simulation results. First we show the good numerical performances of our method comparatively with classical ones on a small example. On a second time we treat a large dimensional dictionary learning problem.

keywords: ill-posed inverse problems, variational bayesian methodology, sparse signal reconstruction, infinite dimensional convex optimization

1. Introduction. The recent development of information technologies has increased the expansion of inverse problems for very large dimensional datasets. Indeed whereas the 90's decade have seen the introduction of image reconstruction problems, the current main interest is on 3D sequences (3D + T), thus on large dimensional sets of data. There is therefore a significant growth in the number of measurements in the involved problems. One has frequently to treat the reconstruction of more than one million data. At the same time, the signal processing techniques have helped to overcome the limitations of measuring instruments as they supplied the design of systems involving indirect measures. These new equipments introduced in exchange novel signal processing challenges, such as super resolution deconvolution, source separation or tomographic reconstruction. All these problems are ill posed, the only information contained in the data and in the model of acquisition are not sufficient to obtain a good estimation of the unknown objects.

To solve these ill-posed problems, we introduce an *a priori* information on the data. The Bayesian approach appearing in this paper consists in a modelisation of source of information as probability density functions [7, 19, 12]. This approach allows the development of unsupervised methods, such that the parameters of probability distributions (mean, variance, etc. ..), also called hyperparameters, are adjusted automatically. These hyperparameters can tune the trade-off between information coming from data (likelihood) and *a priori* information. We call these methods "fully Bayesian" as they consist in a construction of a posterior distribution from the likelihood and from the prior information thanks to the Bayes rule. In general, this posterior distribution is known up to a constant of proportionality K. In order to determine an estimation of the unknown object, the posterior law is summed up by a point: generally, the maximum *a posteriori* (MAP) or the posterior mean (PM) are chosen. The maximization of the posterior law leads to an non convex optimization issue. In the posterior mean case, we must calculate an integral. When the constant K is unknown, we cannot determine analytically this solution. Therefore, a classical way to determine this posterior mean is to perform an empirical mean of sample under

---

*L2S, SUPELEC, CNRS, University Paris-Sud, 3 rue Joliot-Curie, 91190 Gif-Sur-Yvette, FRANCE. email:{firstname.lastname}@lss.supelec.fr

the posterior law thanks to stochastic Markov Chain Monte Carlos (MCMC) [30, 31]. In the MCMC principle a Markov chain is generated which converges asymptotically to a sample of the desired distribution. After a sufficient time, the so called burning time, one retains only the samples of the Markov chain close to the posterior distribution. There are two main types of MCMC methods, algorithms of Metropolis-Hastings and Gibbs sampler [31]. However for large dimensional problems involving complicated covariance matrix, MCMC methods fail to give an acceptable approximation. Furthermore these methods need a long time solely to explore the space of possibilities.

Therefore D. MacKay, inspired by statistical physics works, introduced the variational Bayesian inference as an alternative method to MCMC, see [20]. This methodology was ever since involved in computer science such as in [14] and in signal processing, for different applications such as: source separation using ICA [21, 6], Mixture Models estimation [24], hyperspectral imaging reduction [3], deconvolution [4, 2] recursive methods [35, 36]. More recently a sparse bayesian methodology using a Laplace prior was developed by Seeger [33, 34]. The main outline of the variational Bayesian methodology is to determine an approximation the posterior distribution by the closest separable law, in terms of Kullback-Leiber divergence. Thanks to this method, the initial inverse problem appears as a convex optimization problem in a function space. Classical variational approach gives then an analytical approximation of the posterior. However, this solution has no explicit forms. For large dimensional problems this turn out to be an important drawback.

The goal of this paper is to construct an iterative algorithm able to provide in a reduced computation time a close approximation of the solution of this variational problem. The main principle is to adapt a classical optimization algorithm, the gradient descent method, [27], to the space of probability distributions.

The second contribution consists in the application of the mentioned methodology to inverse problems corresponding to linear Gaussian white noise model. This model can be applied for instance for deconvolution, super-resolution, tomography, or source separation. Concerning the prior distribution, we emphasize on sparse information.

The sparse prior information is classically introduced by heavy-tailed distributions. These distributions can be for instance Laplace [40], Bernoulli Gaussian law [29, 10, 18], mixtures of Gaussian [37, 11], Cauchy distribution, or α-stable distributions. Recently, the Gaussian Scale Mixture class was introduced in [38] as a model of wavelet coefficients of natural images. This class of probability densities generalizes the previous ones. The main advantage of Gaussian Scale Mixture is that they can be easily written as Gaussian distribution, conditioned by an hidden variable and thus enable the use of Gaussian based tools. In our context, we consider that the sparsity information can be modeled by the family of Gaussian Scale Vector Mixture distribution introduced in [13]. In this case, such as Chantas *et al.* in [4], we define an extended problem where hidden variables have also to be estimated.

This article is divided as follows. In Section 2 we present the optimization algorithm in probability density space involved in the paper whereas Section 3 presents our algorithm, based on its application to the classical variational Bayesian approach. Thus, in Section 4 we apply our new method for inverse problems with a Gaussian likelihood and a prior in the Gaussian Scale Vector Mixture (GSVM) family and present our supervised and unsupervised algorithms. In section 5, we present simulation results, first on a tomographic example where the *a priori* information promotes pulses or solutions extremely sparse and secondly on an identification problem in a very large dictionary learning context. Finally, Section 6 concludes the paper.

2. Optimization algorithm in a measures space. The present section is devoted to the construction of a gradient based algorithm adapted to the optimization problem involved in the variational Bayesian methodology.

In this part, we assume that we stand in the measurable space $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$, where $\mathcal{B}(\mathbb{R}^N)$ is the σ-field of Borel sets of $\mathbb{R}^N$. Our purpose is to construct an iterative algorithm able to provide a close approximation of the posterior distribution involved in Bayesian estimation. This problem can be seen as the problem of maximization of a concave functional in a probability measures set.

Concerning the probability density functions, there are several possible representations of such objects. The first one is to consider that this space is a subset of $L^1(\mathbb{R}^N)$, that is the subset of positive integrable functions with a total mass equal to one. As $L^1(\mathbb{R}^N)$ is a Banach space, classical algorithms holds. However, one has to pay a particular attention to the fact that the positivity of the functions together with the fixed total mass imposes additional constraints. We will see in the following why this point of view is not adapted to the Bayesian variational methodology. Another point of view, see [22], is to consider this set as a subset of the space of signed Radon measures $\mathcal{M}(\mathbb{R}^N)$, that is measures that can be written $\mu = \mu^+ - \mu^-$, endowed with the norm of total variation. Once again this is a Banach space, see [22]. The classical gradient descent can also be adapted in this framework, as done in [23]. However in [23], measures obtained at each iteration have no longer densities, and cannot converge to a solution of our optimization problem.

In the following we consider as an important constraint the fact that the optimal measure has to be separable. Hence we rather stand in $\widetilde{\mathcal{M}}(\mathbb{R}^N) = \prod_{i=1}^N \mathcal{M}(\mathbb{R})$ endowed with norm:

$$\forall \mu \in \widetilde{\mathcal{M}}, \quad \|\mu\|_{TV} = \sum_{i=1}^{N} \sup_{A_i \in \mathcal{B}(\mathbb{R})} \int_{A_i} \mathrm{d}\mu_i^+(x_i) + \int_{A_i} \mathrm{d}\mu_i^-(x_i). \tag{2.1}$$

Note that when μ is a density measure, i.e. $\mathrm{d}\mu(\boldsymbol{x}) = q(\boldsymbol{x})\mathrm{d}\mathcal{L}(\boldsymbol{x}) = q(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$, $\mathcal{L}$ standing for the Lebesgue measure, its total variation norm coincides with the L^1 norm of its density function q.

Furthermore, a probability density is considered as an element of the closed convex set Ω, defined by

$$\Omega = \{\mu \in \widetilde{\mathcal{M}};\ \mathrm{d}\mu(\boldsymbol{x}) = \prod_{i=1}^{N} q_i(x_i)\mathrm{d}x_i, \text{ where } q_i \in L^1(\mathbb{R}) \text{ is a function}$$
$$\text{such that } q_i \geq 0 \text{ a.e. and } \int_{\mathbb{R}} \mathrm{d}\mu_i(x_i) = 1\}.$$

Note that this set can also be written as the cartesian product of the $(\Omega_i)_{i=1,\dots,N}$ where

$$\Omega_i = \{\mu_i \in \mathcal{M}(\mathbb{R});\ \mathrm{d}\mu_i(\boldsymbol{x}) = q_i(x_i)\mathrm{d}x_i, \text{ where } q_i \in L^1(\mathbb{R}) \text{ is a function}$$
$$\text{such that } q_i \geq 0 \text{ a.e. and } \int_{\mathbb{R}} \mathrm{d}\mu_i(x_i) = 1\}.$$

Hence our purpose is, given a concave functional $F : \widetilde{\mathcal{M}} \to \mathbb{R}$, to define an algorithm which approximates a probability measure μ^{opt} solution of

$$\mu^{opt} = \arg\max_{\mu \in \Omega} F(\mu). \tag{2.2}$$

This problem can be seen as a constrained convex optimization problem in the infinite dimensional Banach space $(\widetilde{\mathcal{M}}, \|.\|_{TV})$.

In this framework, most results of optimization are based on duality results, see for instance [16, 5]. In the present paper we consider a gradient-like descent algorithm defined directly on the probability measures set.

Let us introduce some notations from [22]. Let $F : \widetilde{\mathcal{M}} \to \mathbb{R}$. As $\widetilde{\mathcal{M}}$ is a Banach space, one can compute its Fréchet derivative at $\mu \in \widetilde{\mathcal{M}}$ as the bounded linear functional $\mathrm{d}F_\mu(.) : \widetilde{\mathcal{M}} \to \mathbb{R}$ satisfying

$$F(\mu+\nu) - F(\mu) - \mathrm{d}F_\mu(\nu) = o(\|\nu\|_{TV}), \quad \text{when } \|\nu\| \to 0.$$

In the following we also consider the Gateaux derivative of F:

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

In some cases, see [22], one can find a function $\mathrm{d}f : \mathcal{M} \times \mathbb{R}^N \to \mathbb{R}$ continuous, such that

$$\forall \nu \in \widetilde{\mathcal{M}}, \quad \partial F_\mu(\nu) = \int_{\mathbb{R}^N} \mathrm{d}f(\mu, \boldsymbol{x})\mathrm{d}\nu(\boldsymbol{x}). \tag{2.3}$$

Consider an auxiliary concave functional $G : \widetilde{\mathcal{M}} \to \mathbb{R}$. An important property appearing in the following is that the Fréchet differential of G is L-Lipschitz on Ω, i.e.

$$\forall (\mu_1, \mu_2) \in \Omega^2, \;\; \forall \nu \in \widetilde{\mathcal{M}} \quad |\mathrm{d}G_{\mu_1}(\nu) - \mathrm{d}G_{\mu_2}(\nu)| \leq L\|\nu\|_{TV}\|\mu_1 - \mu_2\|_{TV}. \tag{2.4}$$

The Lipschitz differential condition of G together with its concavity implies that, see [26] for instance,

$$\forall (\mu_1, \mu_2) \in \Omega^2, \quad 0 \geq G(\mu_1) - G(\mu_2) - \mathrm{d}G_{\mu_2}(\mu_1 - \mu_2) \geq -L\|\mu_1 - \mu_2\|^2_{TV}. \tag{2.5}$$

Furthermore we say that a function F is twice differentiable at $\mu \in \widetilde{\mathcal{M}}$, in the sense of Gateaux, if

$$\forall (\nu_1, \nu_2) \in \widetilde{\mathcal{M}}^2, \quad \partial^2 F_\mu(\nu_1, \nu_2) = \lim_{t \to 0} \frac{\mathrm{d}F_{\mu + t\nu_1}(\nu_2) - \mathrm{d}F_\mu(\nu_2)}{t}.$$

If it exists, $\partial^2 F$ is a bilinear application from $\widetilde{\mathcal{M}} \times \widetilde{\mathcal{M}}$ to $\mathbb{R}$. If F is a concave functional, this second-order derivative must satisfy for every $\mu \in \widetilde{\mathcal{M}}$,

$$\forall \nu \in \widetilde{\mathcal{M}}, \quad \partial^2 F_\mu(\nu, \nu) \leq 0. \tag{2.6}$$

Our purpose is to construct an iterative algorithm providing a density at each iteration and approximating the solution of (2.2) for a certain class of functions F. The key principle of our method is given by the Radon-Nikodym theorem, see [32] for instance. Let $k \geq 0$ be an integer and assume that $\mu^k \in \widetilde{\mathcal{M}}$ is a probability measure absolutely continuous respectively to the Lebesgue measure. In order to keep a measure absolutely continuous with respect to the Lebesgue measure we construct $\mu^{k+1} \in \widetilde{\mathcal{M}}$ as a measure absolutely continuous with respect to μ^k. In this case, the Radon-Nikodym theorem ensures that this measure should be written as

$$\mathrm{d}\mu^{k+1}(\boldsymbol{x}) = h_k(\boldsymbol{x})\mathrm{d}\mu^k(\boldsymbol{x}), \tag{2.7}$$

where $h_k \in L^1(\mu^k)$ is a positive function. Our aim is thus to determine a function $h_k \in L^1(\mu^k)$ which ensures that $F(\mu^{k+1}) \geq F(\mu^k)$. Following the classical scheme given by the gradient descent method, h_k is given as a function of the derivative of F at μ^k and, according to our structure, by

$$h_k(\boldsymbol{x}) = K_k(\alpha_k)\exp(\alpha_k \mathrm{d}f(\mu^k, \boldsymbol{x})), \tag{2.8}$$

where $\mathrm{d}f$ is defined by (2.3) whereas $\alpha_k > 0$ is the algorithm step-size at iteration k and $K_k(\alpha_k)$ is the normalization constant such that $\int_{\mathbb{R}^N} \mathrm{d}\mu^{k+1}(\boldsymbol{x}) = 1$. We also impose the convention that $h_k(\boldsymbol{x}) = \infty$ when $\exp(\alpha_k \mathrm{d}f(\mu^k, \boldsymbol{x}))$ is not integrable. One can see that as soon as μ^0 can be written as $\mu^0 = q^0 \mathrm{d}\boldsymbol{x}$ where q^0 a positive density, so can be each μ^k. We thus can consider in the following that $\mathrm{d}\mu^k = q^k \mathrm{d}\boldsymbol{x}$. This choice of h is motivated by the positive, integrable assumption together with, as mentioned earlier, its coherence with the structure of the gradient descent method. This descent algorithm is defined as the "exponentiated gradient" descent in [17]. Since [17] it has been widely studied in the context of machine learning even in the Bayesian framework, see [9] for instance.

The optimization algorithm involved in this paper is the following:

Algorithm 1 Exponentiated Gradient algorithm

```
INITIALIZE( $\mu^0 \in \Omega$)
repeat
Compute $\mathrm{d}f(\mu^k, \boldsymbol{x})$
Compute $\alpha_k = \arg\max_\alpha K_k(\alpha)\exp(\alpha \mathrm{d}f(\mu^k, .))\mu^k$
Compute $\mu^{k+1} = K_k(\alpha_k)\exp(\alpha_k \mathrm{d}f(\mu^k, .))\mu^k$
until Convergence
```

In order to determine the convergence properties of this exponentiated gradient method in our context, let us define the hypotheses that we impose on the functional F.

ASSUMPTION 1. *Let $F : \widetilde{M} \to \mathbb{R}$ be a concave functional. We say that F satisfies Assumption 1 if:*

(i) *F can be written as $F = G + H$ where G is a concave L-Lipschitz Fréchet-differentiable functional whereas H corresponds to the entropy of the probability measure, that is the Kullback-Leibler divergence from this measure to the Lebesgue measure.*

(ii) *F is twice differentiable in the sense of Gateaux and its first order derivative satisfies Equation (2.3).*

(iii) $\lim_{\|\mu\| \to \infty} F(\mu) = -\infty$.

REMARK 1. *The two points (ii), (iii) of the definition of hypothesis 1 just ensure that the optimal stepsize α_k defined in our algorithm indeed exists. Concerning hypothesis (i), it can be replaced by the more restrictive hypothesis that F is L-Lipschitz*

Fréchet differentiable. Note that as our purpose is to construct measures with a density, the term H also corresponds to the entropy of the density function.

Let us now state the convergence result.

THEOREM 2.1. *Let F be a concave functional satisfying Assumption 1 of Definition 1. Let α_k be the optimal stepsize of Algorithm 1, for every $k \geq 0$. Then the sequence $(\mu^k)_{k\geq 0}$ of elements of $\widetilde{\mathcal{M}}$ given by $\mu^{k+1} = K_k(\alpha_k)\exp(\alpha_k \mathrm{d}f(\mu^k, .))\mu^k$ converges to a maximizer of F on Ω.*

Proof. The proof of this theorem involves two main steps. In a first step we prove that the sequence $(F(\mu^k))_{k\in\mathbb{N}}$ is an increasing sequence. This ensures in a second time to infer the convergence of the sequence $(\mu^k)_{k\geq 0}$ to a solution of (2.2).

LEMMA 2.2. *Let F be a functional satisfying Hypothesis 1. Let also $(\mu^k)_{k\in\mathbb{N}}$ be the sequence provided by Algorithm 1. Then*

$$\exists \alpha_0 > 0,\ \forall \alpha \in (0, \alpha_0) \quad F(\mu^\alpha) - F(\mu^k) \geq 0. \tag{2.9}$$

Proof.

Let $k > 0$ be fixed and μ^k be given. For every $\alpha \geq 0$ we define μ^α as the measure defined for all $\boldsymbol{x} \in \mathbb{R}^N$ by $\mathrm{d}\mu^\alpha(\boldsymbol{x}) = K_k(\alpha)\exp(\alpha \mathrm{d}f(\mu^k, \boldsymbol{x}))\mathrm{d}\mu^k(\boldsymbol{x}) := h_\alpha(\mu^k, \boldsymbol{x})\mathrm{d}\mu^k(\boldsymbol{x})$.

We define furthermore $g_k(\alpha) := F(\mu^\alpha)$. Thus g_k is a function from $\mathbb{R}^+$ to $\mathbb{R}$ twice differentiable and α_{opt} is an optimal step-size if $g_k(\alpha_{opt}) = \max g_k(\alpha)$, i.e.

$$\alpha_{opt} = \arg\max_\alpha g_k(\alpha). \tag{2.10}$$

The fact that $F(\mu) \to -\infty$ when $\|\mu\| \to \infty$ ensures that we can find an α_{opt}, not necessarily unique, such that

$$\forall \alpha > 0, \quad F(\mu^\alpha) \leq F(\mu^{\alpha_{opt}}). \tag{2.11}$$

Let $\alpha > 0$ be given and consider the decomposition given by point (i) of Definition 1. Thanks to Equation (2.5) one has

$$G(\mu^\alpha) \geq G(\mu^k) + \mathrm{d}G_{\mu^k}(\mu^\alpha - \mu^k) - L\|\mu^\alpha - \mu^k\|^2_{TV}. \tag{2.12}$$

Furthermore, as $\mu^\alpha = h_\alpha(\mu^k, .)\mu^k$ and μ^k is a probability measure one can notice that

$$-L\|\mu^\alpha - \mu^k\|^2_{TV} = -L\|h_\alpha(\mu^k, .) - 1\|^2_{L^1(\mu^k)} \geq -L\|h_\alpha(\mu^k, .) - 1\|^2_{L^2(\mu^k)}. \tag{2.13}$$

Furthermore,

$$\begin{aligned} H(\mu^\alpha) &= -\int_{\mathbb{R}^N} \ln\left(\frac{\mathrm{d}\mu^\alpha}{\mathrm{d}\mathcal{L}}\right)\mathrm{d}\mu^\alpha(\boldsymbol{x}) \\ &= H(\mu^k) - \int_{\mathbb{R}^N} \ln(h_\alpha(\mu^k, \boldsymbol{x}))\mathrm{d}\mu^\alpha(\boldsymbol{x}) - \int_{\mathbb{R}^N} \ln\left(\frac{\mathrm{d}\mu^k}{\mathrm{d}\mathcal{L}}\right)(h_\alpha(\mu^k, \boldsymbol{x}) - 1)\mathrm{d}\mu^k(\boldsymbol{x}). \end{aligned}$$

Following the development made in [8] for simple functions, the Gateaux derivative at q^k of the entropy $\mathcal{H} : L^1 \to \mathbb{R}^+$ in direction $h \in L^1$ is given by

$$\partial\mathcal{H}_{q^k}(h) = -\int (\ln(q^k(\boldsymbol{x})) + 1)h(\boldsymbol{x})\mathrm{d}(\boldsymbol{x}),$$

as soon as $q^k + h$ cannot be negative. A similar approach in $\widetilde{\mathcal{M}}$ ensures that

$$\partial H_{\mu^k}(\mu^\alpha - \mu^k) = \int_{\mathbb{R}^N} (-\ln\left(\frac{\mathrm{d}\mu^k}{\mathrm{d}\mathcal{L}}\right) - 1)(h_\alpha(\mu^k, \boldsymbol{x}) - 1)\mathrm{d}\mu^k(\boldsymbol{x}), \tag{2.14}$$

as $\mu^k + (\mu^\alpha - \mu^k)$ is obviously always positive.

This entails

$$H(\mu^\alpha) \geq H(\mu^k) - \int_{\mathbb{R}^N} \ln(h_\alpha(\mu^k, \boldsymbol{x}))\mathrm{d}\mu^\alpha(\boldsymbol{x}) + \partial H_{\mu^k}(\mu^\alpha - \mu^k). \tag{2.15}$$

Finally, from (2.12), (2.13) and (2.15) one has

$$F(\mu^\alpha) - F(\mu^k) \geq \partial F_{\mu^k}(\mu^\alpha - \mu^k) - L\|h_\alpha(\mu^k, .) - 1\|^2_{L^2(\mu^k)} - \int_{\mathbb{R}^N} \ln(h_\alpha(\mu^k, \boldsymbol{x}))\mathrm{d}\mu^\alpha(\boldsymbol{x}). \tag{2.16}$$

Finally, $F(\mu^\alpha) - F(\mu^k)$ is positive if the right side of Equation (2.16) is positive. This part being quite technical it is reported to Appendix 7.1.

□

Lemma 2.2 ensures that for $\alpha > 0$ small enough, $F(\mu^\alpha) \geq F(\mu^k)$. As we choose $\mu^{k+1} = \mu^{\alpha_{opt}}$, where α_{opt} is defined by (2.10), we obviously have $F(\mu^{k+1}) \geq F(\mu^k)$.

Finally the sequence $(F(\mu^k))_{k\in\mathbb{N}}$ is increasing and upper bounded in $\mathbb{R}$, so convergent, it thus also satisfies that $F(\mu^{k+1}) - F(\mu^k) \to 0$.

In order to conclude we have to show that $(\mu^k)_{k\in\mathbb{N}}$ indeed converges to a maximum of F on Ω. But, for every $k \geq 0$, $\mu^k \in \Omega$, which is a bounded set in $\widetilde{\mathcal{M}}$ and thus in $\mathcal{M}(\mathbb{R}^N)$. Furthermore, this last one is the dual of C_0, the space of continuous functions that tend to zero at infinity, which is a separable Banach space. As a consequence of the Banach-Alaoglu Theorem, see [32] for instance, there exists $\mu^{lim} \in \widetilde{\mathcal{M}}$ and a subsequence $(\mu^{k_n})_{n\in\mathbb{N}}$ such that for every continuous function that goes to zero at infinity f,

$$\int f(\boldsymbol{x})\mu^{k_n}(\mathrm{d}\boldsymbol{x}) \to \int f(\boldsymbol{x})\mu^{lim}(\mathrm{d}\boldsymbol{x}).$$

i.e. when $k \to \infty$, we have $\mu^{k_n} \rightharpoonup^* \mu^{lim} \in \Omega$.

From Lemma 2.2 we know that there exists $\alpha_0 > 0$ such that

$$F(\mu^{k+1}) = g_k(\alpha_{opt}) \geq g_k(\alpha), \quad \forall \alpha \in (0, \alpha_0).$$

However the analytic form of α_{opt} is not attainable. We thus approximate it by a calculable α_{subopt}, not necessarily smaller than α_0. In order to determine this α_{subopt} we can notice that the Taylor-Young formula gives, for α small enough,

$$g_k(\alpha) = g_k(0) + \alpha g_k'(0) + \frac{\alpha^2}{2} g_k''(0) + \alpha^2 \varepsilon(\alpha) := \varphi_k(\alpha) + \alpha^2\varepsilon(\alpha), \tag{2.17}$$

where $\varepsilon(\alpha) \to 0$ when $\alpha \to 0$.

Let us determine the derivative of g. For this purpose, we have to determine the derivative of the function $\widetilde{g} : \alpha \mapsto h_\alpha(\mu^k, .)$. As $h_\alpha(\mu^k, \boldsymbol{x}) = K_k(\alpha)e^{\alpha \mathrm{d}f(\mu^k, \boldsymbol{x})}$, its derivative is given by

$$\forall \boldsymbol{x} \in \mathbb{R}^N, \quad \widetilde{g}'(\alpha)(\boldsymbol{x}) = K_k'(\alpha) e^{\alpha \mathrm{d}f(\mu^k, \boldsymbol{x})} + \mathrm{d}f(\mu^k, \boldsymbol{x}) K_k(\alpha) e^{\alpha \mathrm{d}f(\mu^k, \boldsymbol{x})}.$$

As α is supposed to be close to zero, one can assume that it is not greater than one and the Lebesgue's Theorem thus yields

$$K_k'(\alpha) = -\frac{\int \mathrm{d}f(\mu^k, \boldsymbol{x}) K_k(\alpha) e^{\alpha \mathrm{d}f(\mu^k, \boldsymbol{x})} \mathrm{d}\mu^k(\boldsymbol{x})}{\left(\int e^{\alpha \mathrm{d}f(\mu^k, \boldsymbol{x})} \mathrm{d}\mu^k(\boldsymbol{x})\right)^2} = -K_k(\alpha) \int \mathrm{d}f(\mu^k, \boldsymbol{x}) h_\alpha(\mu^k, \boldsymbol{x}).$$

Finally one obtains that

$$\forall \boldsymbol{x} \in \mathbb{R}^N, \quad \widetilde{g}'(\alpha)(\boldsymbol{x}) \mathrm{d}\mu^k(\boldsymbol{x}) = \mathrm{d}\mu^\alpha(\boldsymbol{x}) \left(\mathrm{d}f(\mu^k, \boldsymbol{x}) - \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k, \boldsymbol{y}) \mathrm{d}\mu^\alpha(\boldsymbol{y}) \right). \quad (2.18)$$

And

$$g_k'(\alpha) = \partial F_{\mu^\alpha}(\widetilde{g}'(\alpha)\mu^k) = \int_{\mathbb{R}^N} \mathrm{d}f(\mu^\alpha, \boldsymbol{x}) \left(\mathrm{d}f(\mu^k, \boldsymbol{x}) - \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k, \boldsymbol{y}) \mu^\alpha(\mathrm{d}\boldsymbol{y}) \right) \mathrm{d}\mu^\alpha(\boldsymbol{x}),$$

and

$$\begin{aligned} g_k'(0) = \partial F_{\mu^k}(\widetilde{g}'(0)\mu^k) &= \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k, \boldsymbol{x}) \left(\mathrm{d}f(\mu^k, \boldsymbol{x}) - \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k, \boldsymbol{y}) \mu^k(\mathrm{d}\boldsymbol{y}) \right) \mathrm{d}\mu^k(\boldsymbol{x}) \\ &= \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k, \boldsymbol{x})^2 \mu^k(\mathrm{d}\boldsymbol{x}) - \left(\int_{\mathbb{R}^N} \mathrm{d}f(\mu^k, \boldsymbol{y}) \mathrm{d}\mu^k(\boldsymbol{y}) \right)^2 \end{aligned} \quad (2.19)$$

The critical point of $\varphi_k(\alpha)$ is $\alpha_{subopt} = -\frac{g_k'(0)}{g_k''(0)}$, as soon as $g_k''(0) \neq 0$, which gives in (2.17):

$$g_k(\alpha_{subopt}) = g_k(0) - \frac{g_k'(0)^2}{2g_k''(0)} + \alpha^2 \varepsilon(\alpha), \quad (2.20)$$

and by construction of $F(\mu^{k+1})$,

$$F(\mu^{k+1}) \geq g_k(\alpha_{subopt}) = g_k(0) - \frac{g_k'(0)^2}{2g_k''(0)} + \alpha^2 \varepsilon(\alpha) = F(\mu^k) - \frac{g_k'(0)^2}{2g_k''(0)} + \alpha^2 \varepsilon(\alpha).$$

As $F(\mu^{k+1}) - F(\mu^k) \to 0$, obviously $\lim\limits_{k \to \infty} \frac{g_k'(0)^2}{2g_k''(0)} = 0$. Let us consider the convergent subsequence $(k_n)_{n \in \mathbb{N}}$ and denote by $(\alpha_{k_n})_{n \in \mathbb{N}}$ the sequence defined $\forall n \in \mathbb{N}$ by $\gamma_{k_n} = \alpha_{subopt}$. Hence, $-\frac{g_{k_n}'(0)^2}{g_{k_n}''(0)} = g_{k_n}'(0)\gamma_{k_n} \to 0$. As $\mathrm{d}f$ is supposed to be continuous, the sequence $(g_{k_n}'(0))_{n \in \mathbb{N}}$ is convergent and either $\gamma_{k_n} \to 0$ or $g_{k_n}'(0) \to 0$. Let us assume that $\gamma_{k_n} \to 0$ and that $g_{k_n}'(0) \to l \neq 0$. Let $\beta > 0$ be given. As $\gamma_{k_n} \to 0$ we have, for n large enough that

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

Hence,

$$\frac{g_{k_n}(\frac{\gamma_{k_n}}{\beta}) - g_{k_n}(0)}{\gamma_{k_n}/\beta} = g'_{k_n}(0)\left(1 - \frac{1}{2\beta}\right) + \gamma_{k_n}\varepsilon(\gamma_{k_n}), \tag{2.21}$$

and when n tends to infinity, γ_{k_n} tends to zero, and taking limits in (2.21) one obtains

$$l = l\left(1 - \frac{1}{2\beta}\right),$$

which is impossible as soon as $\beta \neq \frac{1}{2}$. Hence, $g'_{k_n}(0) \to 0$ when $n \to \infty$.

Furthermore,

$$g'_k(0) = \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k, \boldsymbol{x})^2 \mu^k(\mathrm{d}\boldsymbol{x}) - \left(\int_{\mathbb{R}^N} \mathrm{d}f(\mu^k, \boldsymbol{y})\mu^k(\mathrm{d}\boldsymbol{y})\right)^2.$$

Hence, for n large enough,

$$\|\mathrm{d}f(\mu^{k_n}, .)\|^2_{L^2(\mu^{k_n})} - \|\mathrm{d}f(\mu^{k_n}, .)\|^2_{L^1(\mu^{k_n})} \to 0,$$

and thus $\mathrm{d}f(\mu^{k_n}, .)$ converges to a constant λ, independent of $\boldsymbol{x} \in \mathbb{R}^N$.

Let ν be any element of Ω. Applying the Main Value Theorem to the real valued function $f(\theta) = F(\theta\mu^{lim} + (1-\theta)\nu)$ we see that

$$F(\nu) = F(\mu^{lim}) + \mathrm{d}F_{\mu^{lim}}(\nu - \mu^{lim}) + \frac{1}{2}\partial^2 F_{\tilde{\mu}}(\nu - \mu^{lim}, \nu - \mu^{lim}),$$

where $\tilde{\mu} = \theta_0\nu + (1-\theta_0)\mu^{lim}$ and $\theta_0 \in (0,1)$ is fixed.

As $\partial F_{\mu^{lim}}(\nu - \mu^k) = 0$, and from concavity of F we have

$$F(\nu) \leq F(\mu^{lim}) \qquad \forall \nu \in \Omega,$$

which means that $F(\mu^{lim})$ is a maximum of F over Ω.

□

In the present part we have presented an algorithm adapted to the space of probability measures. Our main interest is its application in the context of variational Bayesian methodology. For the sake of completeness, let us remind this methodology introduced in [20].

3. Our Variational Bayesian Algorithm. For the sake of simplicity, we consider in the following density functions q instead of the corresponding measures μ. We denote by $\boldsymbol{y} \in \mathbb{R}^M$ the M dimensional vector containing the data information whereas $\boldsymbol{w} \in \mathbb{R}^N$ represents the vector to be estimated, which is considered as a realization of a random vector $\boldsymbol{W}$. We also denote by p the prior probability density function (p.d.f.) of $\boldsymbol{W}$. The Bayes rule entails that this prior distribution is closely related to the posterior one, $p(\boldsymbol{w}|\boldsymbol{y})$, up to a normalization constant. Even in simple cases this posterior may not be separable. Hence, in the variational Bayesian framework, we approximate it by a separable probability density

$$q(\boldsymbol{w}) = \prod_i q_i(w_i). \tag{3.1}$$

Taking separable laws obviously simplify the problem even if it introduces some approximation errors.

Therefore, we want to determine a separable probability density function q close of the true posterior, in the sense defined by the Kullback-Leibler divergence, see [35] for instance. The optimal approximating density q is given by

$$\forall i \in \{1, \ldots, N\}, \quad q_i(w_i) = K_k^i \exp\left(< \ln p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j \neq i} q_j} \right), \tag{3.2}$$

where K_k^i is the normalization constant and $< \ln p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j\neq i} q_j} = \int_{\mathbb{R}^{N-1}} \ln p(\boldsymbol{y}, \boldsymbol{w}) \prod_{j \neq i} q_j(w_j)$ is the mean of $\ln p(\boldsymbol{y}, \boldsymbol{w})$ under the probability $\prod_{j\neq i} q_j$.

Although this solution is obtained analytically, Equation (3.2) clearly does not have an explicit form. In order to have implementable methods a first step is to consider conjugate priors. Hence the optimization turns out to an update of the distribution parameters. However even for conjugate priors, this solution is hardly tractable in practice, and is thus approximated by iterative fixed points methods. In Equation (3.2), the calculus of q_i imposes the knowledge of all q_j for j different from i, and this optimization is either performed alternatively or by group of coordinate. In both cases, the computation complexity can be important.

For large dimensional problems these methods are not tractable in practice. Our purpose is thus to solve the functional optimization problem given by the Bayesian variational method in an efficient way thanks to the algorithm defined in Section 2.

3.1. Variational Bayesian Exponentiated Gradient Algorithm. In this section we define an iterative method which allows to compute efficiently at each iteration each q_i independently of the others in order to decrease the computational cost of one iteration.

A first step in this part is to rewrite the minimization problem as a convex optimization problem independent of the posterior distribution to be approximated. Instead of minimizing the Kullback-Leibler divergence, we thus remark, as in [6], that

$$\ln p(\boldsymbol{y}) = \ln \frac{p(\boldsymbol{y}, \boldsymbol{w})}{p(\boldsymbol{w}|\boldsymbol{y})}, \tag{3.3}$$

where $\boldsymbol{w}$ is the vector of hidden variables and parameters.

As the log-likelihood $\ln p(\boldsymbol{y})$ in (3.3) does not depends on $\boldsymbol{w}$ one can write

$$\ln p(\boldsymbol{y}) = \mathcal{F}(q) + \mathcal{KL}[q||p(.|\boldsymbol{y}))].$$

In this case,

$$\mathcal{F}(q) = \int_{\mathbb{R}^N} q(\boldsymbol{w}) \ln(\frac{p(\boldsymbol{y}, \boldsymbol{w})}{q(\boldsymbol{w})}) \mathrm{d}\boldsymbol{w}, \tag{3.4}$$

is the negative free energy. Thus minimizing the Kullback-Leibler divergence is obviously equivalent to maximize this negative free entropy.

Therefore in the following we will consider the problem of maximizing

$$\begin{aligned}\mathcal{F}(q) &= \int_{\mathbb{R}^N} \ln p(\boldsymbol{y}, \boldsymbol{w}) q(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} - \int_{\mathbb{R}^N} \ln(q(\boldsymbol{w})) q(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \\ &= \langle \ln p(\boldsymbol{y}, .) \rangle_q + \mathcal{H}(q),\end{aligned} \tag{3.5}$$

where

$$\mathcal{H}(q) = -\int_{\mathbb{R}^N} \ln(q(\boldsymbol{w})) q(\boldsymbol{w}) \mathrm{d}\boldsymbol{w},$$

is the entropy of q. The main advantage of this approach is that the objective functional does not depend on the true posterior anymore but only on the joint distribution $p(\boldsymbol{y}, .)$, which is more easily tractable.

One can also notice that the problem of finding

$$q^{opt} = \arg\max_{q\ p.d.f.} \mathcal{F}(q) \tag{3.6}$$

is equivalent to the problem of finding

$$\mu^{opt} = \arg\max_{\mu \in \Omega} F(\mu). \tag{3.7}$$

Where the functional F is defined by $\forall \mu \in \Omega$, such that q is the density of μ by $F(\mu) = \mathcal{F}(q)$. Furthermore the corresponding function F satisfies hypothesis 1. Note that in the following we consider that . Hence we can apply Theorem 2.1 in this context.

Remark 2. *As mentioned, a classical method in our context is to consider each density function q as a $L^1(\mathbb{R}^N)$ function and to apply classical algorithms. In the present framework, taking the non-negativity and the total mass assumptions into account, the algorithm involved is given by the projected gradient method which gives:*

$$\forall \boldsymbol{w} \in \mathbb{R}^N \qquad q^{k+1}(\boldsymbol{w}) = P_\Theta(q^k(\boldsymbol{w}) + \rho^k f(q^k, \boldsymbol{w})), \tag{3.8}$$

where P_Θ is the projector operator on the subspace $\Theta = \{q \in L^1(\mathbb{R}^N);\ q(\boldsymbol{w}) \geq 0 \text{ and } \|q\|_{L^1} = 1\}$.

However, this algorithm requires that $f(q^k, \boldsymbol{w}) \in L^1(\mathbb{R}^N)$ which is not the case in general.

Hence, we apply the algorithm introduced in Section 2 to the Variational Bayesian framework of Section 3. We consider

$$\mathcal{F}(q) = \langle \ln p(\boldsymbol{y}, .) \rangle_q + \mathcal{H}(q).$$

In this case, the Gateaux differential of $F(\mu) = \mathcal{F}(q)$ at $\mu \in \Omega$ separable is given by $\mathrm{d}F_\mu(\nu) = \sum_i \int_{\mathbb{R}^N} f(q_i, x_i) \nu_i(\mathrm{d}x)$ where

$$\forall i \in \{1, \ldots, N\}, \quad \forall \boldsymbol{w} \in \mathbb{R}^N, \quad f(q_i, w_i) = \langle \ln p(\boldsymbol{y}, \boldsymbol{w}) \rangle_{\prod_{j \neq i} q_j} - \ln q_i(w_i) - 1.$$

Let $k \geq 0$ be given and q^k be constructed. Following the scheme defined by Algorithm 1 and Equation (2.8), at the following iteration we consider q^α given, for

$\alpha > 0$, by

$$\begin{aligned}
\forall \boldsymbol{w} \in \mathbb{R}^N, \quad q^{\alpha}(\boldsymbol{w}) &= K_k q^k \exp\left[\alpha_k f(q^k, \boldsymbol{w})\right] && (3.9)\\
&= \tilde{K}_k q^k(\boldsymbol{w}) \left(\prod_i \frac{\exp\left(< \ln p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j \neq i} q_j^k(w_j)}\right)}{q_i^k(w_i)} \right)^{\alpha_k} \\
&= \tilde{K}_k q^k(\boldsymbol{w}) \left(\prod_i \frac{q_i^r(w_i)}{q_i^k(w_i)} \right)^{\alpha_k} && (3.10)
\end{aligned}$$

where $\tilde{K}_k$ is the normalization constant and q^r is an intermediate measure given by

$$\forall i \in \{1, \ldots, N\}, \quad q_i^r(w_i) = \exp\left(< \ln p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j \neq i} q_j^k(w_j)}\right)$$

The main challenge is to determine the optimal value of $\alpha > 0$. This optimal value should satisfy $g_k'(\alpha_{opt}) = 0$. However, this quantity is hardly tractable in practice. Therefore, we consider instead the suboptimal value given by

$$\alpha_{subopt} = -\frac{g_k'(0)}{g_k''(0)}, \tag{3.11}$$

when $g_k''(0) \neq 0$. This leads to the main algorithm of this paper.

Algorithm 2 Variational Bayesian Exponentiated Gradient Like Algorithm

```
INITIALIZE( $q^0 \in \Theta$)
repeat
    function ITERATION( Compute $q^{k+1} = K_k q^k \exp\left[\alpha_k f(q^k, \boldsymbol{w})\right]$)
        Compute $q_i^r(w_i) = \exp\left(< \ln p(\boldsymbol{y}, \boldsymbol{w}) >_{\prod_{j \neq i} q_j^k(w_j)}\right)$ for every $i = 1, \ldots, N$
        Compute $\alpha_{subopt} = -\frac{g'(0)}{g''(0)}$
        Compute $q^{\alpha_{subopt}}(\boldsymbol{w}) = q^k(\boldsymbol{w}) \left(\frac{q^r(\boldsymbol{w})}{q^k(\boldsymbol{w})}\right)^{\alpha_{subopt}}$.
        Take $q^{k+1} = q^{\alpha_{subopt}}$.
    end function
until Convergence
```

4. Application to linear inverse problems.

4.1. Statement of the problem. The following of this paper presents the application of Algorithm 2 to linear inverse ill-posed problems. More precisely we consider its implementation for Bayesian study of heavy-tailed information. The model of observations chosen in the following is given by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{b}, \tag{4.1}$$

where $\boldsymbol{y} \in \mathbb{R}^M$ is the vector of observations given as a linear function of the unknowns vector $\boldsymbol{x} \in \mathbb{R}^N$ to be estimated. Here, $\boldsymbol{b} \in \mathbb{R}^M$ is the noise vector whereas $\boldsymbol{H}$ is a matrix in $M_{N \times M}$. We also suppose that $\boldsymbol{x}$ is a realization of a random vector $\mathbf{X}$.

In the following we stand in a white noise model which induces that the noise is supposed to be an iid Gaussian vector $\mathcal{N}(0,\sigma_b^2\mathbf{I})$. We can deduct easily the likelihood

$$p(\boldsymbol{y}|\boldsymbol{x}) = (2\pi\sigma_b^2)^{-M/2}\exp\left[-\frac{\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2}{2\sigma_b^2}\right]. \qquad (4.2)$$

Concerning the prior distribution we choose to take sparsity into account by considering $\mathbf{X}$ distributed following a separable heavy tailed distribution. The most general case is given by Gaussian Vector Scale Mixture (GVSM) defined in [13]. In this case, for $i=1,\dots,N$, we suppose that $X_i \sim U_i/\sqrt{Z_i}$ where $\mathbf{U}\sim\mathcal{N}(0,\sigma_s^2\mathbf{I})$, $\mathbf{Z}=\prod Z_i$ is a positive random vector of independent positive coordinates and $\mathbf{U}$ and $\mathbf{Z}$ are independents. As a consequence the density of $\mathbf{X}$ is given in an integral form as

$$\forall i\in\{1,\dots,N\},\qquad p(x_i)=\int_{\mathbb{R}}\frac{\sqrt{z_i}}{(2\pi)^{1/2}\sigma_s}e^{-\frac{z_ix_i^2}{2\sigma_s^2}}\phi_{z_i}(z_i)dz_i.$$

Note that in the definition, for the sake of simplicity, we consider $\mathbf{Z}$ as a precision parameter and not as a dilatation one. Gaussian Vector Scale Mixture form a large class of nongaussian random variables recently developed as a model of wavelet coefficients of natural images, see [38]. The main interest of this model is, by solving an extended problem due to the presence of a hidden random vector $\mathbf{Z}$, to allow the use of Bayesian hierarchic approaches.

The Gaussian Scale Mixture family offers a large class of random variables including Gaussian mixing, when $\mathbf{Z}\sim Z\mathbf{I}$ a discrete random vector or Student laws if all Z_i are Gamma random variables. With different hypothesis on the distribution of $\mathbf{Z}$ one can also define Generalized Gaussian distributions or α-stable ones, see [38]. Indeed GSM offer a simple representation of a large class of nongaussian probability distributions, which justify the increasing interest on this model.

In our context, we choose to consider $\mathbf{Z}$ as a independent Gamma random vector, i.e. for $i=1,\dots,N$, we have $Z_i\sim\mathcal{G}(\tilde{a}_i,\tilde{b}_i)$ and

$$\forall i\in\{1,\dots,N\},\qquad p(x_i)=\frac{\tilde{b}_i^{\tilde{a}_i}}{\Gamma(\tilde{a}_i)}\int_{\mathbb{R}}\frac{\sqrt{z_i}}{(2\pi)^{1/2}\sigma_s}e^{-\frac{z_ix_i^2}{2\sigma_s^2}}z_i^{\tilde{a}_i-1}e^{-z_i\tilde{b}_i}dz_i. \qquad (4.3)$$

For $\tilde{a}_i=\tilde{b}_i=\frac{\nu}{2}$, the p.d.f. of $\mathbf{X}$ corresponds to a Student-t distribution, as in the model used in [4]. This model of $\boldsymbol{Z}$ ensures that $\boldsymbol{X}$ satisfies the conjugate priors condition.

One can easily check that when the prior information is given by (4.3), Equation (4.2) gives the following posterior distribution

$$p(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{y})\propto\sigma_b^{-M}\exp\left[-\frac{\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2}{2\sigma_b^2}\right]\prod_{i=1}^N\frac{\sqrt{z_i}}{\sigma_s}\exp\left[-\frac{z_ix_i^2}{2\sigma_s^2}\right]\frac{\tilde{b}_i^{\tilde{a}_i}z_i^{\tilde{a}_i-1}e^{-z_i\tilde{b}_i}}{\Gamma(\tilde{a}_i)}. \qquad (4.4)$$

Considering that we do not know the involved constants and that the mixing matrix $\boldsymbol{H}$ is high dimensional, this posterior distribution cannot be evaluated directly.

4.2. Numerical implementation. The aim of variational Bayesian methodology and therefore of our method in the context established in part 4 is the approximation of the posterior p.d.f given by (4.4) by a separable one $q(\boldsymbol{x},\boldsymbol{z})=\prod_i q_i(x_i)\prod_j\tilde{q}_j(z_j)$.

As we have chosen conjugate prior for $\boldsymbol{X}$ and $\boldsymbol{Z}$, the optimum approximating distribution of $\boldsymbol{X}$ is known to belong to a Gaussian family, whereas the p.d.f. of $\boldsymbol{Z}$ belongs to a Gamma one.

$$\begin{aligned} q^k(\boldsymbol{x}) &= \prod_i \mathcal{N}(\boldsymbol{m}_k(i), \boldsymbol{\sigma}_k^2(i)) \\ \tilde{q}^k(\boldsymbol{z}) &= \prod_j \mathcal{G}(a_k(j), b_k(j)) \end{aligned}$$

Hence at the initialization stage, we consider

$$\begin{aligned} q^0(\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{m}_0, \mathrm{Diag}(\boldsymbol{\sigma}_0^2)) \\ \tilde{q}^0(\boldsymbol{z}) &= \prod_j \mathcal{G}(a_0(j), b_0(j)) \end{aligned}$$

where $\mathrm{Diag}(\boldsymbol{v})$ is a diagonal matrix with $\boldsymbol{v}$ on its diagonal, and $\boldsymbol{\sigma}_0^2 \in \mathbb{R}^N$ is the vector of initial variances.

Our minimization problem can be analyzed following the alternate iterative scheme:

$$\begin{aligned} \tilde{q}^{k+1}(\boldsymbol{z}) &= \arg\max_{\tilde{q}(\boldsymbol{z})} \mathcal{F}\big(q^k(\boldsymbol{x})\tilde{q}(\boldsymbol{z})\big) \\ q^{k+1}(\boldsymbol{x}) &= \arg\max_{q(\boldsymbol{x})} \mathcal{F}\big(q(\boldsymbol{x})\tilde{q}^{k+1}(\boldsymbol{z})\big) \end{aligned}$$

4.2.1. Approximation of $\tilde{q}$. One can see in Equation (4.4) that the conditional posterior $p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})$ is separable. In this case the classical Bayesian variational approach is efficient enough to be implemented directly. Actually for $i = 1, \dots, N$, $\tilde{q}_i^{k+1}(z_i)$ does not depend on other coordinates. Hence all the z_i can be computed simultaneously, knowing only $q(\boldsymbol{x})$. Thanks to the classical Bayesian variational approach [20] described in Section 3, we deduce $\tilde{q}^{k+1}$ thanks to Equation (3.2) and Equation (4.4), for every $i = 1, \dots, N$

$$\begin{aligned} \tilde{q}_i^{k+1}(z_i) &= \exp\Big(< \ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) >_{\prod_{j\neq i} \tilde{q}_j^k(z_j) q^k(\boldsymbol{x})}\Big) \qquad (4.5) \\ &\propto \exp\left((\tilde{a}_i - \frac{1}{2})\ln(z_i) - \int (\frac{x_i^2 z_i}{2\sigma_s^2} + z_i\tilde{b}_i) \prod_l q_l^k(x_l) \prod_{j\neq i} \tilde{q}_j^k(z_j) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z}\right) \\ &\propto \exp\left((\tilde{a}_i - \frac{1}{2})\ln(z_i) - z_i\tilde{b}_i - \int \frac{x_i^2 z_i}{2\sigma_s^2} q_i^k(x_i)\mathrm{d}x_i\right) \\ &\propto \exp\left((\tilde{a}_i - \frac{1}{2})\ln(z_i) - z_i\tilde{b}_i - \frac{(\boldsymbol{\sigma}_k^2(i) + \boldsymbol{m}_k^2(i))z_i}{2\sigma_s^2}\right) \\ &\propto z_i^{\tilde{a}_i - \frac{1}{2}} \exp\left(-z_i\Big[\tilde{b}_i + \frac{(\boldsymbol{\sigma}_k^2(i) + \boldsymbol{m}_k^2(i))}{2\sigma_s^2}\Big]\right) \qquad (4.6) \end{aligned}$$

This entails that $\tilde{q}_i^{k+1}(z_i)$ corresponds to a Gamma p.d.f. of parameters:

$$\forall i \in \{1, \dots, N\}, \qquad \boldsymbol{a}_{k+1}(i) = \tilde{a}_i + \frac{1}{2}, \qquad (4.7)$$

$$\boldsymbol{b}_{k+1}(i) = \frac{\boldsymbol{m}_k^2(i) + \boldsymbol{\sigma}_k^2(i)}{2\sigma_s^2} + \tilde{b}_i. \qquad (4.8)$$

4.2.2. Approximation of q by Algorithm 2. Let us assume that we start with a Gaussian p.d.f. $q^0(\boldsymbol{x})$ with mean $\boldsymbol{m}_0$ and covariance matrix $\mathrm{Diag}(\boldsymbol{\sigma}_0^2)$. At each iteration we determine the approximation of $q^k(\boldsymbol{x})$ thanks to our method. At each iteration $k{+}1$, we define an auxiliary measure $q^r(\boldsymbol{x})$ by $q_i^r(x_i) = \exp\left(\langle \ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})\rangle_{\prod_{j\neq i} q_j^k(x_j) q^{k+1}(\boldsymbol{z})}\right)$.

Hence $\forall i \in \{1,\dots,N\}$,

$$\begin{aligned}
q_i^r(x_i) &= \exp\left(< \ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) >_{\prod_{j\neq i} q_j^k(x_j)\tilde{q}^{k+1}(\boldsymbol{z})}\right) &(4.9)\\
&\propto \exp\left(-\int\left(\frac{\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2}{2\sigma_b^2} + \frac{x_i^2 z_i}{2\sigma_s^2}\right)\prod_{j\neq i} q_j^k(x_j)\tilde{q}^{k+1}(\boldsymbol{z})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z}\right)\\
&\propto \exp\left[-\int\left(\frac{\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{y}}{2\sigma_b^2} + \frac{x_i^2 z_i}{2\sigma_s^2}\right)\prod_{j\neq i} q_j^k(x_j)\tilde{q}^{k+1}(z_i)\mathrm{d}\boldsymbol{x}\mathrm{d}z_i\right]\\
&\propto \exp\left[-\frac{1}{2\sigma_b^2}\Big(x_i^2\mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H})_i - 2x_i(\boldsymbol{H}^T\boldsymbol{y})_i + 2x_i(\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{m}_k)_i\right.\\
&\qquad \left. -2x_i\mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H}))_i\boldsymbol{m}_k(i)\Big) + \frac{x_i^2\boldsymbol{a}_{k+1}(i)}{2\sigma_s^2\boldsymbol{b}_{k+1}(i)}\right] &(4.10)
\end{aligned}$$

where $\mathrm{diag}(A)$ is the vector composed by the diagonal entries of A. Note that $q^r(\boldsymbol{x})$ corresponds, up to the normalization term, to a Gaussian distribution with mean $\boldsymbol{m}_r$ and variance $\boldsymbol{\sigma}_r^2$, where, for every $i = 1,\dots,N$,

$$\boldsymbol{\sigma}_r^2(i) = \left(\frac{\mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H})_i}{\sigma_b^2} + \frac{\boldsymbol{a}_{k+1}(i)}{\boldsymbol{b}_{k+1}(i)\sigma_s^2}\right)^{-1} \tag{4.11}$$

and

$$\boldsymbol{m}_r(i) = \boldsymbol{\sigma}_r^2(i) \times \left(\frac{\boldsymbol{H}^T\boldsymbol{y} - (\boldsymbol{H}^T\boldsymbol{H} - \mathrm{diag}(\boldsymbol{H}^T\boldsymbol{H}))\boldsymbol{m}_k}{\sigma_b^2}\right)_i \tag{4.12}$$

Therefore, by Equation (3.10), we have for every $i = 1,\dots,N$,

$$\begin{aligned}
q_i^\alpha(x_i) &= K_k q_i^k(x_i)\left(\frac{q_i^r(x_i)}{q_i^k(x_i)}\right)^\alpha\\
&= \sqrt{\frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)}}^{\alpha} K_k \exp\left[-\frac{(x_i - \boldsymbol{m}_k(i))^2}{2\boldsymbol{\sigma}_k^2(i)}\right]\exp\left[-\alpha\frac{x_i^2(\boldsymbol{\sigma}_k^2(i) - \boldsymbol{\sigma}_r^2(i))}{2\boldsymbol{\sigma}_r^2(i)\boldsymbol{\sigma}_k^2(i)}\right]\\
&\times \exp\left[-\alpha\frac{-2x_i(\boldsymbol{m}_r(i)\boldsymbol{\sigma}_k^2(i) - \boldsymbol{m}_k(i)\boldsymbol{\sigma}_r^2(i)) + \boldsymbol{m}_r(i)^2\boldsymbol{\sigma}_k^2(i) - \boldsymbol{m}_k(i)^2\boldsymbol{\sigma}_r^2(i)}{2\boldsymbol{\sigma}_r^2(i)\boldsymbol{\sigma}_k^2(i)}\right]\\
&= \sqrt{\frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)}}^{\alpha} K_k \exp\left[-\frac{1}{2}\left(x_i^2\frac{\boldsymbol{\sigma}_r^2(i) + \alpha(\boldsymbol{\sigma}_k^2(i) - \boldsymbol{\sigma}_r^2(i))}{\boldsymbol{\sigma}_r^2(i)\boldsymbol{\sigma}_k^2(i)}\right)\right]\\
&\times \exp\left[-\frac{1}{2}\left(-2x_i\frac{\boldsymbol{m}_k(i)\boldsymbol{\sigma}_r^2(i) + \alpha(\boldsymbol{m}_r(i)\boldsymbol{\sigma}_k^2(i) - \boldsymbol{m}_k(i)\boldsymbol{\sigma}_r^2(i))}{\boldsymbol{\sigma}_r^2(i)\boldsymbol{\sigma}_k^2(i)} + t(\alpha)\right)\right]
\end{aligned}$$

where q^α is defined in Section 5.1.1, and $t(\alpha) = \alpha\frac{\boldsymbol{m}_r(i)^2\boldsymbol{\sigma}_k^2(i) - \boldsymbol{m}_k(i)^2\boldsymbol{\sigma}_r^2(i)}{2\boldsymbol{\sigma}_r^2(i)\boldsymbol{\sigma}_k^2(i)}$ is a constant. Finally, $q_i^\alpha(\boldsymbol{x})$ is still a Gaussian p.d.f. with parameters $\boldsymbol{m}_\alpha$ and $\mathrm{Diag}(\boldsymbol{\sigma}_\alpha^2)$

satisfying:

$$\boldsymbol{\sigma}_\alpha^2(i) = \frac{\boldsymbol{\sigma}_r^2(i)\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i) + \alpha(\boldsymbol{\sigma}_k^2(i) - \boldsymbol{\sigma}_r^2(i))} \tag{4.13}$$

$$\boldsymbol{m}_\alpha(i) = \frac{\boldsymbol{m}_k(i)\boldsymbol{\sigma}_r^2(i) + \alpha(\boldsymbol{m}_r(i)\boldsymbol{\sigma}_k^2(i) - \boldsymbol{m}_k(i)\boldsymbol{\sigma}_r^2(i))}{\boldsymbol{\sigma}_r^2(i) + \alpha(\boldsymbol{\sigma}_k^2(i) - \boldsymbol{\sigma}_r^2(i))}. \tag{4.14}$$

In order to construct q^{k+1} we choose in the previous equation $\alpha = \alpha_{subopt}$ defined in Equation (3.11).

Finally, we obtain the following algorithm.

Algorithm 3 Supervised Sparse Reconstruction algorithm (SSR)

```
INITIALIZE($q^0, \tilde{q}^0$)
repeat
    function ESTIMATE $\tilde{q}^{k+1}(\boldsymbol{z})(q^k(\boldsymbol{x}))$
        update $\boldsymbol{a}_{k+1}$ by Equation (4.7)
        update $\boldsymbol{b}_{k+1}$ by Equation (4.8)
    end function
    function ESTIMATE $q^{k+1}(\boldsymbol{x})(\tilde{q}^{k+1}(\boldsymbol{z}))$
        compute $q^r(\boldsymbol{x}) \leftarrow (\boldsymbol{m}_r, \boldsymbol{\sigma}_r^2)$ by Equation (4.12) and Equation (4.11)
        compute $\alpha_{subopt}$
        compute $q^\alpha(\boldsymbol{x}) \leftarrow (\boldsymbol{m}_\alpha, \boldsymbol{\sigma}_\alpha^2)$ by Equation (4.14) and Equation (4.13)
    end function
until Convergence
```

4.3. Unsupervised algorithm. The algorithm described in the previous part is not a fully Bayesian one as it still depends on some parameters, namely the parameters induced by the model (4.1) and (4.3). We see in the following how this method can be extended to an unsupervised one by estimating these parameters. The parameters of the underlying Gamma random variable are not estimated in the following as they define the sharpness of the prior distribution. We thus only estimate the variance parameter of this prior together with the trade off between the prior and the noise.

The main interest of the variational Bayesian approach introduced by D. MacKay [20] and involved in this paper is its flexibility. That is an unsupervised algorithm can be easily deduced from the method proposed in Section 4.2, by considering that the parameters are realizations of a random variable with a given Jeffrey's prior distribution.

In order to simplify the different expressions, we introduce in the following the notations $\gamma_b = 1/\sigma_b^2$ and $\gamma_s = 1/\sigma_s^2$. Hence, γ_b and γ_s are the precision parameters of the noise and of the prior distribution. From now on they are also assumed to be random variable with Gamma prior of parameters $(\tilde{a}_b, \tilde{b}_b)$ resp. $(\tilde{a}_s, \tilde{b}_s)$. As we do not have information on these precision parameters γ_b and γ_s, this prior is obtained by fixing $(\tilde{a}_b = 0, \tilde{b}_b = 0)$ resp. $(\tilde{a}_s = 0, \tilde{b}_s = 0)$.

With these assumptions, the posterior distribution from Equation (4.4) can be written as

$$p(\boldsymbol{x},\boldsymbol{z},\gamma_b,\gamma_s|\boldsymbol{y}) \propto \gamma_b^{\frac{M}{2}} \exp\left[-\frac{\gamma_b\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2}{2}\right] \gamma_s^{\frac{N}{2}} \prod_i \sqrt{z_i} \exp\left[-\frac{\gamma_s z_i x_i^2}{2}\right] \frac{\tilde{b}_i^{\tilde{a}_i} z_i^{\tilde{a}_i-1} e^{-z_i\tilde{b}_i}}{\Gamma(\tilde{a}_i)}$$
$$\times \frac{\tilde{b}_b^{\tilde{a}_b}\gamma_b^{\tilde{a}_b-1}e^{-\gamma_b\tilde{b}_b}}{\Gamma(\tilde{a}_b)} \frac{\tilde{b}_s^{\tilde{a}_s}\gamma_s^{\tilde{a}_s-1}e^{-\gamma_s\tilde{b}_s}}{\Gamma(\tilde{a}_s)}. \tag{4.15}$$

As in the previous section, the conditional posterior $p(\boldsymbol{z},\gamma_b,\gamma_s|\boldsymbol{x},\boldsymbol{y})$ is separable and can be approximated thanks to the classical variational Bayesian approach. Once again only the distribution of $\boldsymbol{X}$ needs the use of Algorithm 2. Here the alternate optimization scheme to carry out is:

$$\begin{aligned}
\tilde{q}^{k+1}(\boldsymbol{z}) &= \arg\max_{\tilde{q}(\boldsymbol{z})} \mathcal{F}\big(q^k(\boldsymbol{x})\tilde{q}(\boldsymbol{z})q_b^k(\gamma_b)q_s^k(\gamma_s)\big)\\
q^{k+1}(\boldsymbol{x}) &= \arg\max_{q(\boldsymbol{x})} \mathcal{F}\big(q(\boldsymbol{x})\tilde{q}^{k+1}(\boldsymbol{z})q_b^k(\gamma_b)q_s^k(\gamma_s)\big)\\
q_b^{k+1}(\gamma_b) &= \arg\max_{q(\gamma_b)} \mathcal{F}\big(q^{k+1}(\boldsymbol{x})\tilde{q}^{k+1}(\boldsymbol{z})q_b(\gamma_b)q_s^k(\gamma_s)\big)\\
q_s^{k+1}(\gamma_s) &= \arg\max_{q(\gamma_s)} \mathcal{F}\big(q^{k+1}(\boldsymbol{x})\tilde{q}^{k+1}(\boldsymbol{z})q_b^{k+1}(\gamma_b)q_s(\gamma_s)\big)
\end{aligned}$$

4.3.1. Optimization of the approximate p.d.f. q_b. Concerning the random vectors $\boldsymbol{Z}$ and $\boldsymbol{X}$, the updating process follows the same scheme than the supervised case, see Section 4.2, and are not recalled here. The main differences reside in the update of the parameters distributions.

As the distributions of γ_b and γ_s are supposed to be Gamma, which is coherent with the conjugate priors hypothesis, at each iteration we just adapt the parameters. Hence we initialize our algorithm by considering that

$$q_b^0(\gamma_b) = \mathcal{G}(a_b^0, b_b^0)$$

At iteration $k+1$ we consider the maximum of the free energy from Equation (3.2) which gives

$$\begin{aligned}
q_b^{k+1}(\gamma_b) &\propto \exp\left[< \ln \Pr(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\gamma_b,\gamma_s) >_{\prod_j q_j^{k+1}(x_j)\prod_j \tilde{q}_j^{k+1}(z_j) q_s^k(\gamma_s)}\right]\\
&\propto \exp\left[< (\frac{M}{2}+\tilde{a}_b-1)\ln(\gamma_b) - \gamma_b(\frac{\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2}{2}+\tilde{b}_b) >_{\prod_j q_j^{k+1}(x_j)}\right]\\
&\propto \exp\Bigg[(\frac{M}{2}+\tilde{a}_b-1)\ln(\gamma_b)\\
&\qquad -\gamma_b\Big(\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{m}_{k+1}\|^2 + \frac{1}{2}\sum_{i=1}^{N} \operatorname{diag}\left(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H}\right)_i \boldsymbol{\sigma}_{k+1}^2(i) + \tilde{b}_b\Big)\Bigg]
\end{aligned}$$

So $q_b^{k+1}(\gamma_b)$ is a Gamma p.d.f. of parameters:

$$a_b^{k+1} = \frac{M}{2} + \tilde{a}_b \tag{4.16}$$

$$b_b^{k+1} = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{m}_{k+1}\|^2 + \frac{1}{2}\sum_{i=1}^{N} \operatorname{diag}\left(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H}\right)_i \boldsymbol{\sigma}_{k+1}^2(i) + \tilde{b}_b \tag{4.17}$$

4.3.2. Optimization of the approximate p.d.f. q_s. As for γ_b, the approximation of q_s is performed in the family of Gamma distributions. Hence at the initialization step we assume that

$$q_s^0(\gamma_s) = \mathcal{G}(a_s^0, b_s^0)$$

and at iteration $k+1$, thanks again to Equation (3.2) we obtain

$$\begin{aligned} q_s^{k+1}(\gamma_s) &\propto \exp\left[< \ln \Pr(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \gamma_b, \gamma_s | I) >_{\prod_j q_j^{k+1}(x_j) \prod_j \tilde{q}_j^{k+1}(z_j) q_b^{k+1}(\gamma_b)}\right] \\ &\propto \exp\left[< (\frac{N}{2} + \tilde{a}_s - 1)\ln(\gamma_s) - \gamma_s(\frac{1}{2}\sum_{i=1}^{N} z_i x_i^2 + \tilde{b}_s) >_{\prod_j q_j^{k+1}(x_j) \tilde{q}_j^{k+1}(z_j)}\right] \\ &\propto \exp\left[< (\frac{N}{2} + \tilde{a}_s - 1)\ln(\gamma_s) - \gamma_s(\frac{1}{2}\sum_{i=1}^{N} \frac{\boldsymbol{a}_{k+1}(i)}{\boldsymbol{b}_{k+1}(i)}(\boldsymbol{m}_{k+1}^2(i) + \boldsymbol{\sigma}_{k+1}^2(i)) + \tilde{b}_s)\right] \end{aligned}$$

So $q_s^{k+1}(\gamma_s)$ is a Gamma p.d.f. and its parameters are deduced by identification.

$$a_s^{k+1} = \frac{N}{2} + \tilde{a}_s \tag{4.18}$$

$$b_s^{k+1} = \frac{1}{2}\sum_{i=1}^{N} \frac{\boldsymbol{a}_{k+1}(i)}{\boldsymbol{b}_{k+1}(i)}(\boldsymbol{m}_{k+1}^2(i) + \boldsymbol{\sigma}_{k+1}^2(i)) + \tilde{b}_s \tag{4.19}$$

Finally the algorithm performed can be summed up as follows.

Algorithm 4 UnSupervised Sparse Reconstruction algorithm (USSR)

```
INITIALIZE($q^0, \tilde{q}^0, q_b^0, q_s^0$)
repeat
    function ESTIMATE $\tilde{q}^{k+1}(\boldsymbol{z})(q^k(\boldsymbol{x}), q_b^k(\gamma_b), q_s^k(\gamma_s))$
        update $\boldsymbol{a}_{k+1}$ using Equation (4.7)
        update $\boldsymbol{b}_{k+1}$ using Equation (4.8)
    end function
    function ESTIMATE $q^{k+1}(\boldsymbol{x})(\tilde{q}^{k+1}(\boldsymbol{z}), q_b^k(\gamma_b), q_s^k(\gamma_s))$
        compute $q^r(\boldsymbol{x}) \leftarrow (\boldsymbol{m}_r, \boldsymbol{\sigma}_r^2)$ using Equation (4.12) and Equation (4.11)
        compute $\alpha_{subopt}$
        compute $q^\alpha(\boldsymbol{x}) \leftarrow (\boldsymbol{m}_\alpha, \boldsymbol{\sigma}_\alpha^2)$ using Equation (4.14) and Equation (4.13)
    end function
    function ESTIMATE $q_b^{k+1}(\gamma_b)(\tilde{q}^{k+1}(\boldsymbol{z}), q^{k+1}(\boldsymbol{x}))$
        update $a_b^{k+1}$ using Equation (4.16)
        update $b_b^{k+1}$ using Equation (4.17)
    end function
    function ESTIMATE $q_s^{k+1}(\gamma_s)(\tilde{q}^{k+1}(\boldsymbol{z}), q^{k+1}(\boldsymbol{x}))$
        update $a_s^{k+1}$ using Equation (4.18)
        update $b_s^{k+1}$ using Equation (4.19)
    end function
until Convergence
```

5. Simulations. This section is devoted to numerical validations of the method proposed in this paper. For the sake of completeness we will treat two cases. The first one is given by a noisy tomographic problem in a small dimensional case. This example allows a comparison of our method with classical reconstruction ones. In a second example, we will see a component identification problem in a large dimensional dataset. This second case ensures that the method proposed in this paper is valid for large dimensional cases.

The first inverse problem considered is given by a tomographic example. The goal is to enhance the accuracy and the effectiveness of our approach, by comparison with classical ones, such as classical Variational Bayesian methods or Monte Carlos Markov Chain (MCMC) methods. From the limitations of these concurrent approaches, we choose to consider only a small dimensional inverse problem (4096 unknowns), and thus to invert the Radon transform of a small sparse image (64×64 pixels).

The second experimental result is devoted to a relatively large inverse problem (≈ 300000 unknowns). In this case, the problem is to identify different components in a dictionary learning process. This problem is performed in a very noisy environment, such as the signal to noise ratio can take negative values. This signal processing problem can appear for instance in astrophysical context (detection of gravitational waves [28]) or in radar imaging [39, 1].

In both cases the sparsity information is introduced by an iid Student't prior. This prior is a particular case of GSM. In the following we fix $\tilde{a}_i = \frac{\nu}{2}$ and $\tilde{b}_i = \frac{\nu}{2}$ in Equation (4.4).

5.1. Tomographic example. For the sake of completeness, a short description of the comparative approaches is given, enhancing the main differences between them. In a second part, we describe the phantom together with the simulation parameters and the results.

5.1.1. Algorithms descriptions.

Filtered Back Projection (FBP). Filtered Back Projection is the classical approach to invert the Radon transform [25, 15]. This algorithm is obtained by sampling the continuous inversion formula. Each line of the sinogram (see Fig. 5.1) is filtered with a ramp filter. The filtered data are backprojected. The discrete version of the backprojection operator is given by $\boldsymbol{H}^{\mathrm{t}}$.

Monte Carlos Markov Chain. The MCMC method contains a large class of Bayesian algorithms [31]. In the following we consider the Gibbs algorithm for its efficiency when the size of the problem increases. The principle is to obtain samples of the posterior law given by Equation (4.4) by an alternate sampling with conditional laws. The algorithm is as follows:

(i) $\boldsymbol{z}^k$ sampled with $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{x}^{k-1})$
(ii) $\boldsymbol{x}^k$ sampled with $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}^k)$
(iii) go to i) until convergence of the Markov chain.

As the conditional law $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{x}^{k-1})$ is a separable Gamma distribution, the computation of the sample $\boldsymbol{z}^k$ is easy. Furthermore $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}^k)$ is a correlated Gaussian distribution with a covariance matrix $\boldsymbol{R}_k = \boldsymbol{M}_k^{\mathrm{t}} \boldsymbol{M}_k = [\frac{1}{\sigma_b^2} \boldsymbol{H}^{\mathrm{t}} \boldsymbol{H} + \frac{1}{\sigma_s^2} \mathrm{Diag}(\boldsymbol{z}^k)]^{-1}$ and a mean $\boldsymbol{m}_k = \frac{1}{\sigma_b^2} \boldsymbol{R}_k \boldsymbol{H}^{\mathrm{t}} \boldsymbol{y}$. The sampling under this correlated distribution is performed by sampling a vector of centered iid Gaussian random variables with variance 1. Afterward this vector is multiplied by the correlation matrix $\boldsymbol{M}_k$ and added to $\boldsymbol{m}_k$.

Remark 3. *At each sampling iteration the covariance matrix of size $N \times N$ have to be inverted.*

Classical Bayesian Variational approach. This method was already described in Section 3. In the alternate gradient descent algorithm, one can chose the degree of separability of the approximative distribution. In the following we consider two cases. In the first case, the so called VBBloc, we consider that the separation of the approximating law is only between $\boldsymbol{x}$ and $\boldsymbol{z}$. This leads to consider the approximating distribution as:

$$q(\boldsymbol{x},\boldsymbol{z}) = q(\boldsymbol{x})\tilde{q}(\boldsymbol{z})$$

and

$$\begin{aligned} q(\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{m},\boldsymbol{R}) \\ \tilde{q}(\boldsymbol{z}) &= \mathcal{G}(\boldsymbol{a},\boldsymbol{b}) \end{aligned}$$

Thus, with Equation (3.2), we obtain $\forall i \in \{1,\dots,N\}$ the following updating equations:

$$\begin{aligned} \boldsymbol{a}(i) &= \frac{\nu}{2} + \frac{1}{2}, \\ \boldsymbol{b}_{k+1}(i) &= \frac{\nu}{2} + \frac{\boldsymbol{m}_k^2(i) + \mathrm{diag}(\boldsymbol{R}_k)(i)}{2\sigma_s^2} \\ \boldsymbol{R}_{k+1} &= \left(\frac{1}{\sigma_s^2}\mathrm{Diag}(\boldsymbol{a}/\boldsymbol{b}_{k+1}) + \frac{1}{\sigma_b^2}\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H}\right)^{-1} \\ \boldsymbol{m}_{k+1} &= \frac{1}{\sigma_b^2}\boldsymbol{R}_{k+1}\boldsymbol{H}^{\mathrm{t}}\boldsymbol{y}. \end{aligned}$$

Remark 4. *At each step, the updating of the covariance matrix requires the inversion of a $N \times N$ matrix, but the convergence rate is better than for the MCMC approach.* To overcome the limit given by a matrix inversion in the classical variational Bayesian framework, we can construct an approximative distribution separable on $\boldsymbol{x}$. Hence, we estimate a vector of variance instead of the matrix of covariance. This approach is called VBComp in the following.

$$q(\boldsymbol{x},\boldsymbol{z}) = \Pi_i q_i(x_i)\tilde{q}(\boldsymbol{z})$$

In this case Equation (3.2) give the following updating equations, $\forall i \in \{1,\dots,N\}$:

$$\begin{aligned} \boldsymbol{a}(i) &= \frac{\nu}{2} + \frac{1}{2}, \\ \boldsymbol{b}_{k+1}(i) &= \frac{\nu}{2} + \frac{\boldsymbol{m}_k^2(i) + \boldsymbol{\sigma}_k^2(i)}{2\sigma_s^2} \end{aligned}$$

And, for every $i \in \{1,\dots,N\}$:

$$\begin{aligned} \boldsymbol{\sigma}_{k+1}^2(i) &= \left(\frac{1}{\sigma_s^2}\boldsymbol{a}(i)/\boldsymbol{b}_{k+1}(i) + \frac{1}{\sigma_b^2}(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H})_{(i,i)}\right)^{-1} \\ \boldsymbol{m}_{k+1}(i) &= \frac{\boldsymbol{\sigma}_{k+1}^2(i)}{\sigma_b^2}(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{y}(i) - (\boldsymbol{d}(i) - (\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H})_{(i,i)}\boldsymbol{m}_k(i))) \\ \boldsymbol{d} &= \boldsymbol{H}^{\mathrm{t}}\boldsymbol{H}\boldsymbol{m}_k \end{aligned}$$

Remark 5. *For each pixel x_i, the corresponding value of $\boldsymbol{d} = \boldsymbol{H}^{\mathrm{t}}\boldsymbol{H}\boldsymbol{m}_k$ must be determined.*

Coordinate	(28,28)	(25,28)	(28,25)	(40,28)	(32,38)	(48,48)	(8,52)
Value	1	1	1	0.5	0.7	0.8	0.6

Table 5.1
Peaks definition in the phantom

5.1.2. Simulation configuration. The test image is given by a sparse phantom, composed of 7 peaks on a grid 64 × 64 (see Tab. 5.1 and Fig. 5.2(a)). We have simulated data in a parallel beam geometry. These projections are collected from 32 angles θ, uniformly spaced over $[0, 180[$. Each projection is composed of 95 detector cells. We add a white Gaussian noise (iid) with standard deviation equal to 0.3 (see Fig. 5.1). Data have thus a relatively bad signal to noise ratio and the number of unknowns is larger than the number of data, which leads to an ill-posed inverse problem.

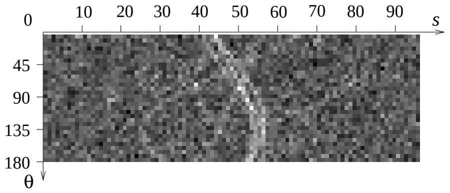


Fig. 5.1. *Data collected : sinogram composed of 32 angles and 95 detector cells.*

5.1.3. Results and discussion. In this section, we expose the inversion of a tomographic problem with the approaches described earlier. All the iterative approaches are initialized with a zero mean and a variance equal to one, and the hyperparameters σ_b^2, σ_s^2 and ν are respectively fixed to 1, 0.05 and 0.1. The original image and its different reconstructions are summed up on Fig. 5.2. A comparison of Fig. 5.2 (b) with 5.2 (c), 5.2 (d) and 5.2 (e) clearly shows that the analytical inversion of the Radon transform performed by Filtered Back Projection (FBP) is less robust to noise than Bayesian approaches. Asymptotically, in Bayesian cases, theoretical results are favorable to the MCMC approach, as it does not need any approximation. In practice, the number of samples is too small to fit with the asymptotic results of MCMC methods, which explains the bad reconstruction observed in Fig. 5.2(c). Finally, the Supervised Sparse Reconstruction (SSR) (see Fig. 5.2(f)) has the same reconstruction quality than the classical variational Bayesian approaches (see VBBloc Fig. 5.2(d) and VBComp Fig. 5.2(e)). However when we compare the execution time (see Tab. 5.2), we see that our approach is 10 time faster than the VBBloc approach, 40 time faster than the VBComp approach and 370 faster than the MCMC approach for this small inverse problem. Moreover this ratio increases with the size of the problem as both MCMC and classical variational Bayesian need the inversion of a covariance matrix at each iteration, which is not the case for our algorithm.

5.1.4. Hyperparameters estimation. As seen in the section 4.2, our approach is defined in a fully Bayesian framework. We thus estimate the values of hyperparameters in introducing a non informative Jeffrey's prior, as described in Part 4.3. We estimate thus the trade off between the likelihood and the prior through the estimation of σ_b^2 and σ_s^2. Hence, we apply the algorithm UnSupervised Sparse Reconstruction (USSR) (see Algorithm 4) in our tomographic dataset. As for the previous simulation,

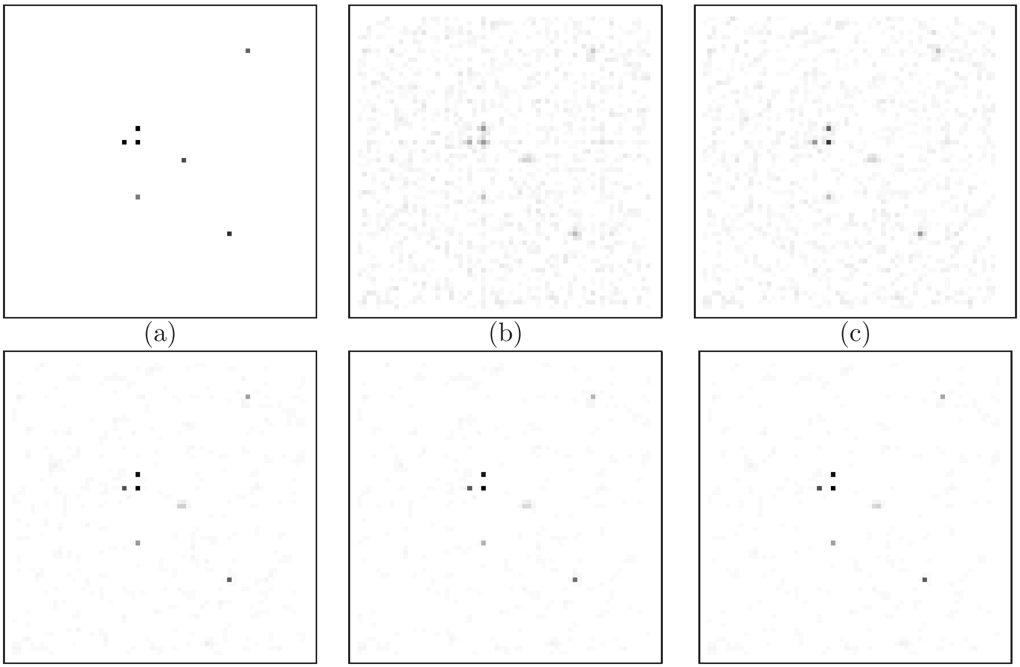

(a) (b) (c)
(d) (e) (f)

Fig. 5.2. *Images are presented with the same inversed grayscale: (a) true image of 7 peaks, (b) FBP with ramp filter, (c) MCMC Gibbs approach, (d) classical variational Bayesian (VBBloc) with bloc optimization, (e) classical variational Bayesian (VBComp) with component optimization, (f) SSR approach.*

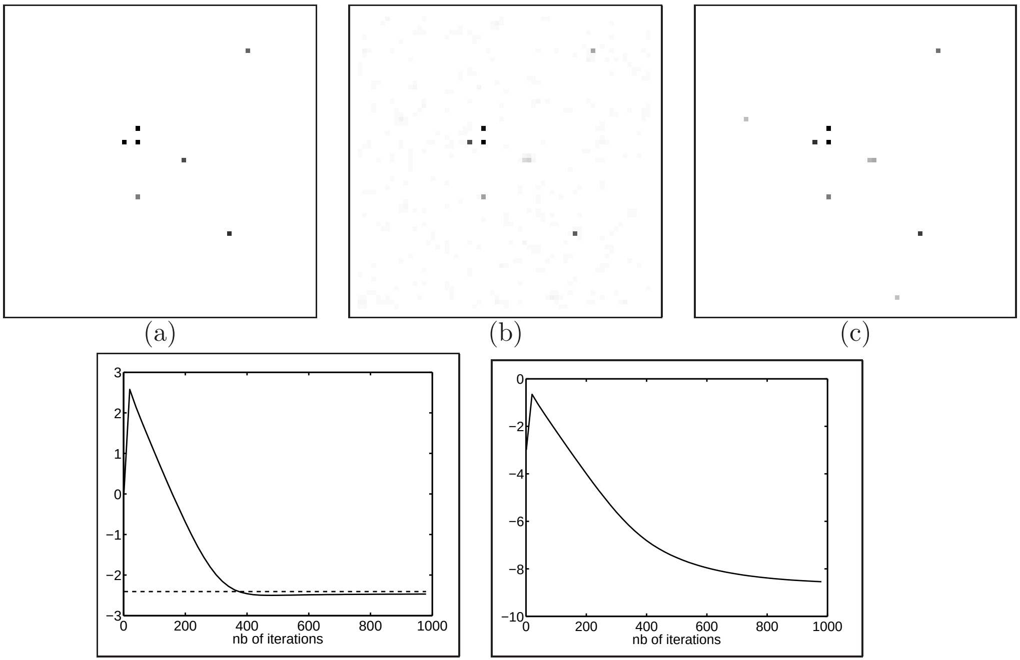
Table 5.2
*Comparison of the different approaches: computing time (second) and quality of estimation (*SNR*).*

Method	FBP	VBBloc	VBComp	VBGrad (SSR)	MCMC Gibbs
CPU time (s)	0.05	586.20	1759.1	44.41	37079.50
Nb of iterate	1	15	8(×4096)	500	1000
SNR	-2.04	5.87	5.76	6.00	-0.60

the initial values of the mean are fixed to zero and the variance are fixed to one. For the hyperparameters σ_b^2 and σ_s^2 the initial values are respectively fixed to 1 and 0.05 to begin with a prior information more important than the likelihood.

The results are summed up in Fig. 5.3. We observe that the hyperparameters estimation intensifies the sparsity of the reconstructed image together with the SNR, as it goes from 6.00 db in the previous case to 10.06 db. Estimating the true hyperparameters is in this case more relevant than arbitrarily chosen parameters. We observe on Fig. 5.3 (c) that the background is equal to zero even if some additional peaks appear in the reconstructed image

Finally we see on Fig. 5.3, (d) and (e), the behavior of the estimation of the hyperparameters respectively to the number of iterations. This plot is in a logarithm scale due to the dynamic of the fluctuations. We observe that for σ_b^2 this estimation converges to the true value (dashed line). For σ_s^2 we do not known the true value, but one can notice that this estimation converges also to some given value.



(a) (b) (c)

(d) (e)

FIG. 5.3. *Results with hyperparameters estimation: (a) True image, (b) reconstructed image with SSR algorithm (hyperparameters are fixed), (c) reconstructed image with USSR (image and hyperparameters are estimated jointly), (d) logarithm of σ_b^2 during the iterations the dashed line correspond to the true value, (e) logarithm of σ_s^2 during the iterations*

5.2. Component identification (denoising). As enhanced by the previous simulation, our method can be more efficient than classical ones on relatively small problem (4096 unknowns). But its main advantage is that it allows to address larger problems (≈ 300000 unknowns in the present case). In the present part, our purpose is to treat an identification problem on a dictionary decomposition. More precisely, we identify many chirps function $\psi_k(t)$ in a linear mixture. This identification issue appears for instance in astrophysics, in order to detect gravitational waves [28] or in radar imaging [39, 1]. Unfortunately, this mixture is corrupted by noise and spurious signal. To identify each chirps in the mixture and to remove the effect of the spurious signal, we develop a dictionary learning approach. To build our dictionary we make the following assumptions : all chirps have the same duration T_{Chirp}, the chirps rate ζ is digitalized on very few values (eight in the present case), the spurious signal can be represented on very few coefficients of Discrete Fourier Transform and the noise is white and Gaussian. However, we do not make any assumption on the variance of the noise or on the number of chirp functions in the mixture.

In the following section, we present the formalism of the dictionary learning approach. After that, we expose the simulation condition. Then, we illustrate the efficiency of our approach with three numerical experiments. In the first one, we consider a nominal case to study the quality of estimation. After what, we study a limit case with a very high level of noise (SNR at -5 db), in order to illustrate the robustness of our approach with regard to noise. In the last experiment, we study the behavior of the reconstruction quality when the number of chirp functions or the noise level increase.

5.2.1. Dictionary decomposition. In this section the signal considered, denoted by $s(t)$, is obtained by the following dictionary decomposition.

$$s(t) = \sum_{i=1}^{N_{freq}} (u_i + jv_i)\phi_i(t) + \sum_{l=1}^{N_{trans}} \sum_{k=1}^{N_{rate}} c_{l,k}\psi_k(t - l\Delta_t), \tag{5.1}$$

where

$$\phi_i(t) = \exp\left[2j\pi f_i t\right],$$

corresponds to a pure frequency f_i with $j^2 = -1$ whereas

$$\psi_k(t) = cos(2\pi(f_0 t + \frac{1}{2}t^2\zeta_k))\Pi_{0,T_{Chirp}}(t)$$

corresponds to the chirps components. Here, f_0 is the starting frequency (at time $t = 0$), ζ_k is the chirp rate, that is the increasing rate of the frequency, $\Pi(t)$ is a gate function, T_{Chirp} is the duration of the chirp, Δ_t is the shift between two chirps functions and $t_l = l\Delta_t$ is the first time where the $\psi_k(t-l\Delta_t)$ is not null. We merge all the dictionary's coefficients in a single vector $\boldsymbol{x} = (u_1, \ldots, u_{Nfreq}, v_1, \ldots, v_{Nfreq}, c_{1,1}, \ldots, c_{Ntrans,Nrate})^{\mathrm{t}}$. Where $Nfreq$ is the number of pure frequency functions contained on the dictionary whereas $Ntrans$ is the number of chirp shifts and $Nrate$ is the number of chirp rate. Moreover, we store sampled version of functions ϕ_i and $\psi_{l,k}$ into a matrix $\boldsymbol{H}$. Hence, the sampled version of Equation (5.1) is given by

$$\boldsymbol{s} = \boldsymbol{H}\boldsymbol{x}, \tag{5.2}$$

where $\boldsymbol{s} = (s(t_0), \ldots, s(t_N))^{\mathrm{t}}$.

The measurement of the signal $\boldsymbol{s}$ is also corrupted by an iid Gaussian noise $\boldsymbol{b}$ with a variance σ_b^2, such that the observations are given by

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{b} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{b}. \tag{5.3}$$

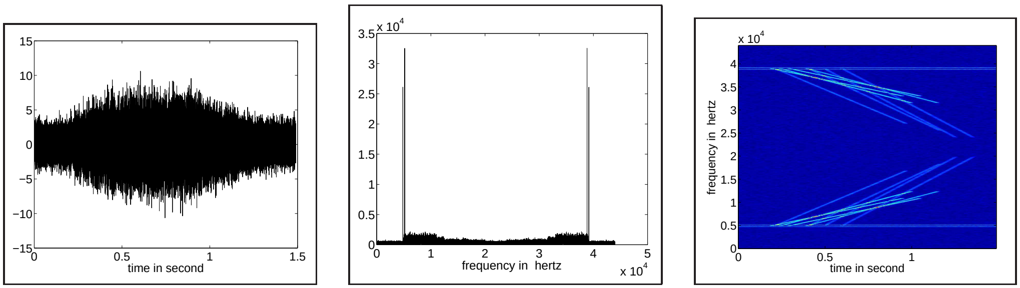
In the following, the dictionary is composed of chirp functions with only 8 different chirp rates ($N_{rate} = 8$). The frequency f_0 is equal to 5 000 hz, the chirp rates ζ_k are uniformly spaced between 6 000 and 20 000 hz, the shift parameters Δ_t is fixed up to a sampling period ($T_e = 1/F_e$). Finally, the duration of the chirp (T_{Chirp}) is equal to the half time of the measurements (T_{mes}).

REMARK 6. *Our dictionary is redundant as the number of coefficients is 4.5 times greater than the number of observations. Thus, the component identification is an ill-posed problem.*

5.2.2. Common simulation conditions. Simulated data are composed by the sum of N_{Cos} cosine functions with different frequencies and N_{Chirp} chirp functions taken in our dictionary. The simulated data are sampled at a frequency $F_e = 44$khz, and they are composed of $N = 2^{16}$ points, thus the duration of measurement T_{mes} is equal to 1.5 second.

TABLE 5.3
Parameters of the different components

type of function	Amplitude	frequency (hz)	rate (hz)	first time (s)
cosine	1	5169.7	-	-
cosine	0.8	4834	-	-
Chirp	1.4	-	8000	0.2
Chirp	1.4	-	10000	0.25
Chirp	1.0	-	16000	0.22
Chirp	1.0	-	20000	0.5
Chirp	1.2	-	10000	0.4
Chirp	1.0	-	18000	0.41
Chirp	1.0	-	20000	0.6
Chirp	1.4	-	8000	0.3



(a) (b) (c)

FIG. 5.4. *Data in different fields: (a) time, (b) frequency, (c) time-frequency*

5.2.3. First result. The simulated data are composed of two cosine functions and eight chirp functions, the parameters of all this functions are given in Tab. 5.3. Once again simulated data have a relatively bad SNR, at 5.68 db in this case. They are plotted in Fig. 5.4 (a), whereas their Fourier transform are given in Fig. 5.4 (b) and their time-frequency transform, computed with a STFT (Short Time Fourier Transform) are on Fig. 5.4(c). We can see in the time frequency representation that there are overlaps between the different components making the identification harder.

Moreover this inverse problem is treated with the unsupervised approach given by Algorithm 4. Our algorithm was launched with the shape parameter of the Student't equals to $\nu = 0.01$ in order to introduce a very sparse prior. The initialization parameters are:

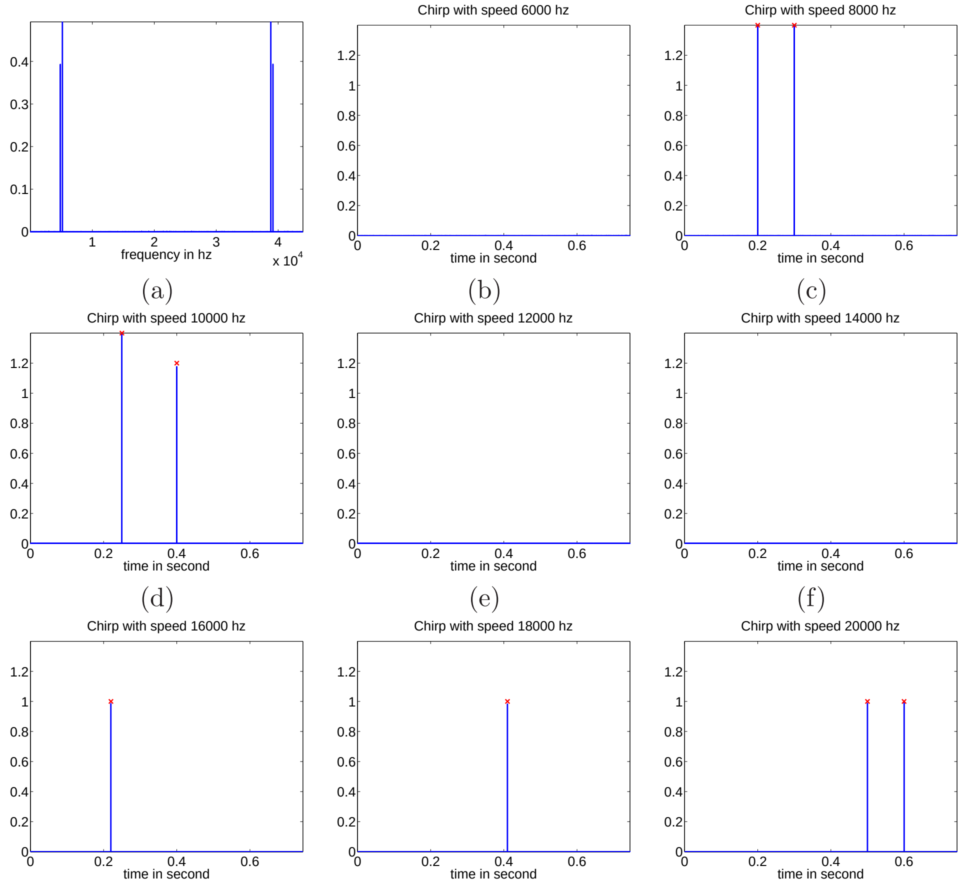
- The mean of $q^0(\boldsymbol{x})$, $\boldsymbol{m}_0 = 0$,
- the variance of $q^0(\boldsymbol{x})$, $\boldsymbol{\sigma}_0^2 = 1$,
- the mean of $q_b^0(\gamma_b)$, is equal to 10^{-5},
- the mean of $q_s^0(\gamma_s)$, is equal to 10^5.

After 400 iterations (316 s), the algorithm (USSR) converges to a stable solution. It gives parameters of different approximated laws. We consider here that the estimation $\hat{\boldsymbol{x}}$ is obtained by taking the posterior mean for the coefficients (see Fig. 5.5).

Fig. 5.5 (a) represents the real part of the Fourier coefficients (the vector $\boldsymbol{u}$ is the Equation (5.1)). We recognize the Fourier transform of the two cosine functions. Moreover, when we compare Fig. 5.5 (a) with the Fourier transform of the data, Fig.

5.4 (b), we observe that the algorithm USSR selected only the sparse components of the Fourier transform. We plot on Figs 5.5 (b-i) the chirp coefficients $c_{l,k}$ for eight chirp rates (6000 until 20000) with respect to the first time t_l where the chirp does not vanishes. In these figures the estimated values are plotted with a line and the true value of non null coefficients are marked by a cross. All chirp coefficients have been reconstructed in the right place, and their amplitudes are very close to the real ones.

Estimation $\hat{\boldsymbol{s}}$ of s is performed thanks to the estimation of the coefficients $\hat{\boldsymbol{x}}$ and Equation (5.2). The SNR of $\hat{\boldsymbol{s}}$ is equal to 22.6 db. It is a very good result knowing that the SNR of data $\boldsymbol{y}$ is equal to 5.68 db.



(a) (b) (c)
(d) (e) (f)
(g) (h) (i)

FIG. 5.5. *Dictionary decomposition: true values of coefficients are marked by a cross*

5.2.4. Limit case. This section illustrates the performance of the algorithm USSR when the signal s is hidden by the noise, the SNR of data being equal to -5 db (see Fig. 5.6 (a)). We generate a signal composed of four chirps Equation (5.1) which parameters are summed up in Tab. 5.4.

The estimation of coefficients is performed using our USSR algorithm with the same initialization as in the previous case. After 400 iterations we obtain the coefficients illustrated on Fig. 5.6, (c) and (d). The reconstructed coefficients are

Table 5.4
Parameters of the different components

type of function	Amplitude	Chirp rate (hz)	first time (s)
Chirp	0.9526	16000	0.1530
Chirp	1.1606	16000	0.1945
Chirp	0.7369	18000	0.2000
Chirp	1.1724	18000	0.1865

represented by a line and the true values of the coefficients are marked by a cross. We observe that all coefficients are in the right place and that the peaks amplitudes are systematically underestimated, but the estimated values are relatively close to the true ones. Fig. 5.6 (b) points out the estimator $\hat{\boldsymbol{s}}$. We see that the shape of the signal is close to the true one, and that when the signal is missing, between 1 and 1,45 s, the level of the residual noise is relatively low.

Table 5.5
*Signal to Noise Ratio (*SNR*) in db*

data	USSR approach	best Wiener filter	best wavelet soft threshold
-5.0	15.05	1.1941	1.8677

In Tab. 5.5 we compare the reconstructed signal $\hat{\boldsymbol{s}}$ with the reconstruction obtained with two classical denoising approaches, namely the Wiener filter and the soft wavelet shrinkage, with the four vanishing moments symmlet. In these two methods, we furthermore have to tune a parameter which is the correlation power for the Wiener filter and the threshold for the soft wavelet shrinkage. Hence we choose the value of this hyperparameter which minimizes the Signal to Noise Ratio (SNR). Unlike in the USSR approach, we thus have to we know the true signal in order to tune this parameter. Furthermore, our approach increases hugely the SNR (20 db), thus the noise level is divided by 100 whereas the classical methods reduce the noise only by a factor 4 or 5. This example enhances the efficiency of the dictionary decomposition in the denoising context.

5.2.5. Behavior of our method versus level of noise and number of components. In this part, we perform the study of the robustness of the USSR algorithm regarding Signal to Noise Ratio. This simulation allows a better understanding of the reconstruction properties of our method.

Simulations. In the following we consider simulated data with 6 different SNR $(-5, -2, 1, 2, 5, 10)$ and with low (4) and high (16) number of components. For each SNR and each number of components, we simulate 30 set of data, the components of the signal being randomly chosen between 1 and 9. If the number is equal to 1, we consider a cosine function with a frequency f_i is uniformly taken between 0 and the sampling frequency (F_e= 44 khz), and an amplitude randomly chosen between 0.6 and 1.4. If the number is equal to 2 resp. $3, \dots, 9$, we simulate a chirp function with a chirp rate ζ_k equal to 6000 resp. $8000, \dots, 20000$. The first time of the chirp t_l is uniformly taken between 0.1 and 0.6 second, and as previously, the amplitude $c_{l,k}$ is randomly chosen between 0.6 and 1.4.

Reconstruction. We reconstruct this 360 sets of data by the algorithm USSR taking the same configuration and the same initialization as in previous cases.

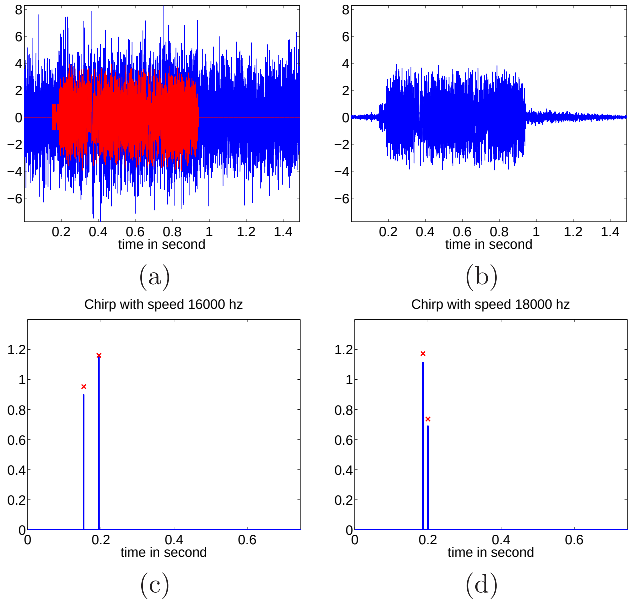


FIG. 5.6. *Limit case: (a) data and true signal, (b) reconstructed signal*

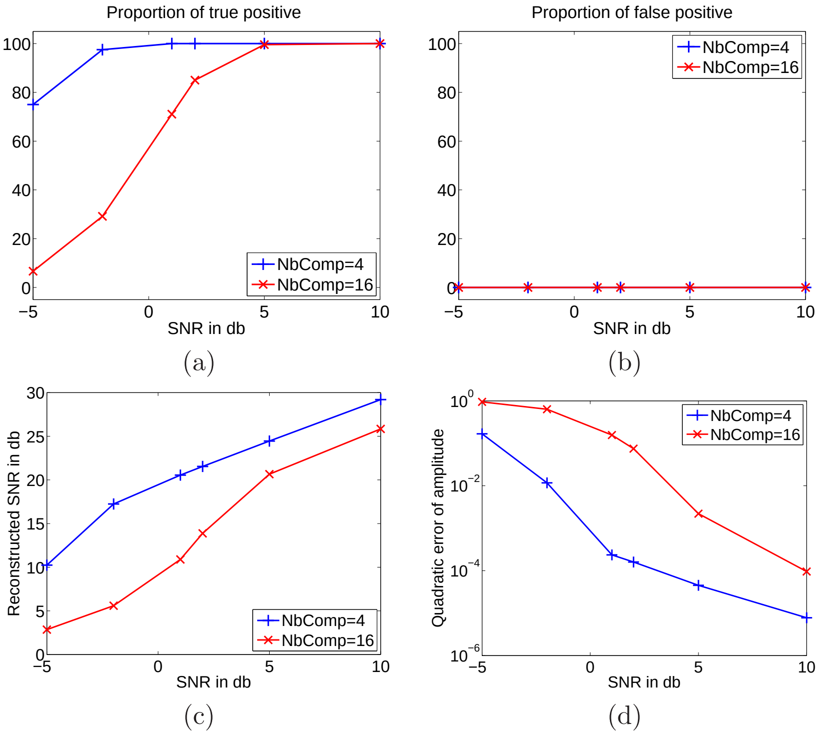
The results are summed up on Fig 5.7 (a-d). Each point of the plot is computed by averaging the result of 30 reconstructions. In Fig 5.7 (a), resp. (b), we plot the true positive proportion, resp. false positive proportion, of the significant reconstructed coefficients[1]. At first sight we see that there are no false positive with our approach. Indeed, as the approach is unsupervised with a sparsity prior, the coefficients with low energy are considered as noise. Moreover we can reconstruct 16 components without lost when the SNR is greater or equal to 5 db and resp. 4 components when the SNR is greater to 1 db. Fig. 5.7 shows that the reconstruction is more difficult when the number of components increases.

Fig. 5.7 (c) is obtained by calculating the SNR of the reconstructed signal $\hat{\boldsymbol{s}}$. We observe a quite linear behavior. For 4 components the gain is of 17 db whereas for 16 components we gain 11.5 db. Finally, Fig. 5.7 (d) exposes the quadratic error of the peaks amplitude. There are here two cases. When all the components are found this error is linear (see Fig. 5.7 (d) the bottom curve when SNR>1) but it increases more rapidly when some components are not found (see Fig. 5.7 (d) the bottom curve when SNR<1).

6. Conclusion. In this paper, we have defined an iterative algorithm based on the descent gradient principle and adapted to the context of variational Bayesian methods. The main interest of this algorithm is that it converges faster than the classical Bayesian methods and allows an use on large dimensional datasets. We have furthermore give its implementation in the case of white noise model when the prior information enhances some sparse behavior. A small tomographic application allows us to compare our method with classical ones. We see that even in small cases, our algorithm can be faster than classical ones. A second simulation part, corresponding

---

[1] the significant coefficients are obtained by applying a threshold equal to 0.2 on the coefficients vector. This threshold is equal to the third of the minimum value of the true non zero coefficients.



(a) (b)

(c) (d)

FIG. 5.7. *Study of Robustness versus noise, each point of the curve being is calculated by averaging 30 reconstructions with components randomly chosen: (a) True positive proportion of the significant reconstructed coefficients, (b) False positive proportion of the significant reconstructed coefficients, (c)* SNR *of the estimated signal* $\hat{\mathbf{s}}$*, (d) Quadratic error of the peaks amplitude*

to a dictionary identification allows to understand the behavior of our method for large dimensional problems. Once again this method has good reconstruction properties in this case.

7. Annexe.

7.1. Proof of Lemma 2.2. In order to ensure that for small values of α, we have $F(\mu^\alpha) > F(\mu^k)$, we show that the right part of Equation (2.16),

$$F(\mu^\alpha)-F(\mu^k) \geq \partial F_{\mu^k}(\mu^\alpha-\mu^k)-L\|h_\alpha(\mu^k,.)-1\|^2_{L^2(\mu^k)}-\int_{\mathbb{R}^N}(\alpha \mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))\mathrm{d}\mu^\alpha(\boldsymbol{x}),$$

is positive.

First, one can notice that

$$-\int_{\mathbb{R}^N}\ln(h_\alpha(\mu^k,\boldsymbol{x}))\mathrm{d}\mu^\alpha(\boldsymbol{x}) = -\int_{\mathbb{R}^N}(\alpha \mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))h_\alpha \mathrm{d}\mu^k(\boldsymbol{x}) = -\alpha\left(\int_{\mathbb{R}^N}\mathrm{d}f(\mu^k,\boldsymbol{x})h_\alpha \mathrm{d}\mu^k(\boldsymbol{x}) -\right.$$

But Jensen's inequality ensures that

$$\ln K_k(\alpha) \leq -\alpha\int_{\mathbb{R}^N}\mathrm{d}f(\mu^k,\boldsymbol{x})\mathrm{d}\mu^k(\boldsymbol{x}). \tag{7.1}$$

Which leads to

$$-\int_{\mathbb{R}^N}\ln(h_\alpha(\mu^k,\boldsymbol{x}))\mathrm{d}\mu^\alpha(\boldsymbol{x}) \geq -\alpha\int_{\mathbb{R}^N}\mathrm{d}f(\mu^k,\boldsymbol{x})(h_\alpha(\mu^k,\boldsymbol{x})-1)\mathrm{d}\mu^k(\boldsymbol{x}). \tag{7.2}$$

Hence,

$$\begin{aligned}\partial F_{\mu^k}(\mu^\alpha-\mu^k)-\int_{\mathbb{R}^N}\ln(h_\alpha(\mu^k,\boldsymbol{x}))\mathrm{d}\mu^\alpha(\boldsymbol{x}) &\geq \int_{\mathbb{R}^N}(1-\alpha)\mathrm{d}f(\mu^k,\boldsymbol{x})(h_\alpha(\mu^k,\boldsymbol{x})-1)\mathrm{d}\mu^k(\boldsymbol{x})\\ &= \frac{1-\alpha}{\alpha}\int_{\mathbb{R}^N}(\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))(h_\alpha(\mu^k,\boldsymbol{x})-1)\mathrm{d}\mu^k(\boldsymbol{x})-\int_{\mathbb{R}^N}\frac{(1-\alpha)\ln K_k(\alpha)}{\alpha}(h_\alpha(\mu\\ &= \frac{1-\alpha}{\alpha}\int_{\mathbb{R}^N}(\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))(h_\alpha(\mu^k,\boldsymbol{x})-1)\mathrm{d}\mu^k(\boldsymbol{x}),\end{aligned}$$

as $\frac{(1-\alpha)\ln K_k(\alpha)}{\alpha}$ is constant and $\int_{\mathbb{R}^N}(h_\alpha(\mu^k,\boldsymbol{x})-1)\mathrm{d}\mu^k(\boldsymbol{x})=0$.

Finally,

$$\begin{aligned}&\mathrm{d}F_{\mu^k}(\mu^\alpha-\mu^k)-L\|h_\alpha-1\|^2_{L^2(\mu^k)}-\int_{\mathbb{R}^N}(\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))\mathrm{d}\mu^\alpha(\boldsymbol{x})\\ &\geq\int_{\mathbb{R}^N}\left((\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))(\frac{1}{\alpha}-1)-L(h_\alpha(\mu^k,\boldsymbol{x})-1)\right)(h_\alpha(\mu^k,\boldsymbol{x})-1)\mathrm{d}\mu^k(\boldsymbol{x})\\ &=\int_{\{x:\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha)>0\}}\left((\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))(\frac{1}{\alpha}-1)-L(h_\alpha(\mu^k,\boldsymbol{x})-1)\right)(h_\alpha(\mu^k,\boldsymbol{x})-1)\mathrm{d}\mu^k(\boldsymbol{x}\\ &+\int_{\{x:\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha)\leq 0\}}\left((\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))(\frac{1}{\alpha}-1)-L(h_\alpha(\mu^k,\boldsymbol{x})-1)\right)(h_\alpha(\mu^k,\boldsymbol{x})-1)\mathrm{d}\mu^k(\boldsymbol{x}\end{aligned} \tag{7.3}$$

Let us consider each integrals appearing in Eq. (7.3) separately. First, let us notice that if $\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha)<0$ then so is $h_\alpha(\mu^k,.)-1$. Furthermore, for every $\alpha>0$ and $\boldsymbol{x}\in\mathbb{R}^N$, we have $h_\alpha(\mu^k,.)-1\geq\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha)$. Hence if $\boldsymbol{x}\in\mathbb{R}^N$ is such that $\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha)<0$ then

$$\left((\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))(\frac{1}{\alpha}-1)-L(h_\alpha(\mu^k,\boldsymbol{x})-1)\right)(h_\alpha(\mu^k,\boldsymbol{x})-1)\geq(h_\alpha(\mu^k,\boldsymbol{x})-1)^2\left(\frac{1}{\alpha}-1-L\right),$$

which is positive as soon as $\alpha\leq\frac{1}{1+L}$.

Consider now that $\boldsymbol{x}\in\mathbb{R}^N$ is such that $\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha)\geq 0$. The Mean Value Theorem applied to the function exponential ensures that one can find, for every $\boldsymbol{x}\in\mathbb{R}^N$ and $\alpha>0$ a $\theta(\boldsymbol{x},\alpha)\in(0,\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))$ such that

$$h_\alpha(\mu^k,\boldsymbol{x})=e^{\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha)}=1+(\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))e^{\theta(\boldsymbol{x},\alpha)}.$$

This entails that

$$\left((\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))(\frac{1}{\alpha}-1)-L(h_\alpha(\mu^k,\boldsymbol{x})-1)\right)=(\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha))(\frac{1}{\alpha}-1-Le^{\theta(\boldsymbol{x},\alpha)}).$$

Furthermore, Jensen's inequality ensures that

$$0\leq\alpha\mathrm{d}f(\mu^k,\boldsymbol{x})+\ln K_k(\alpha)\leq\alpha\left(\mathrm{d}f(\mu^k,\boldsymbol{x})-\int_{\mathbb{R}^N}\mathrm{d}f(\mu^k,\boldsymbol{x})\mathrm{d}\mu^k(\boldsymbol{x})\right).$$

Thus

$$\mathrm{d}f(\mu^k,\boldsymbol{x}) - \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k,\boldsymbol{x})\mathrm{d}\mu^k(\boldsymbol{x}) \geq 0.$$

And

$$1 \leq e^{\theta(\boldsymbol{x},\alpha)} \leq h_\alpha(\mu^k,\boldsymbol{x}) \leq e^{\alpha\left(\mathrm{d}f(\mu^k,\boldsymbol{x}) - \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k,\boldsymbol{x})\mathrm{d}\mu^k(\boldsymbol{x})\right)},$$

which leads to

$$\frac{1}{\alpha} - 1 - Le^{\alpha\left(\mathrm{d}f(\mu^k,\boldsymbol{x}) - \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k,\boldsymbol{x})\mathrm{d}\mu^k(\boldsymbol{x})\right)} \leq \frac{1}{\alpha} - 1 - Le^{\theta(\boldsymbol{x},\alpha)} \leq \frac{1}{\alpha} - 1 - L.$$

Concerning the left part of the previous equation one can notice that the function of α defined here is positive for small positive values of α. Hence there exists a value $\alpha_0 > 0$ such that for every $\alpha \leq \alpha_0$,

$$\frac{1}{\alpha} - 1 - Le^{\alpha\left(\mathrm{d}f(\mu^k,\boldsymbol{x}) - \int_{\mathbb{R}^N} \mathrm{d}f(\mu^k,\boldsymbol{x})\mathrm{d}\mu^k(\boldsymbol{x})\right)} \geq 0.$$

Finally one has that there exists $\alpha_0 \geq 0$ such that

$$\forall \alpha \leq \alpha_0,\ \ \mathrm{d}F_{\mu^k}(\mu^\alpha - \mu^k) - L\|h_\alpha - 1\|^2_{L^2(\mu^k)} - \int_{\mathbb{R}^N} (\alpha \mathrm{d}f(\mu^k,\boldsymbol{x}) + \ln K_k(\alpha))\mathrm{d}\mu^\alpha(\boldsymbol{x}) \geq 0. \tag{7.4}$$

7.2. Optimization of the parameter α. In order to obtain the optimal value of $\alpha > 0$, we have to optimize

$$g_k(\alpha) = F(q^\alpha) = \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \ln p(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) q^\alpha(\boldsymbol{x})\tilde{q}^k(\boldsymbol{z})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z} - \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \ln(q^\alpha(\boldsymbol{x})\tilde{q}^k(\boldsymbol{z}))q^\alpha(\boldsymbol{x})\tilde{q}^k(\boldsymbol{z})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z}.$$

Let us compute $F(q^\alpha)$. To achieve this, we first notice that

$$\ln p(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = \tilde{C} - \frac{\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{y}}{2\sigma_b^2} + \sum_i \frac{1}{2}\ln(z_i/\sigma_s^2) - \frac{z_i x_i^2}{2\sigma_s^2} + (\tilde{a}_i - 1)\ln(z_i) - z_i\tilde{b}_i,$$

where $\tilde{C}$ is a positive constant. Thus

$$\begin{aligned}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \ln p(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) q^\alpha(\boldsymbol{x})\tilde{q}^k(\boldsymbol{z})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z} = \tilde{C} &- \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{y}}{2\sigma_b^2} \prod q_j^\alpha(x_j)\tilde{q}^k(\boldsymbol{z})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z} \\ &+ \sum_i \int_{\mathbb{R}} \frac{1}{2}\ln(z_i/\sigma_s^2)\tilde{q}_i^k(z_i)\mathrm{d}z_i + \sum_i \int_{\mathbb{R}} \Big((\tilde{a}_i - 1)\ln(z_i) - z_i\tilde{b}_i\Big)\tilde{q}^k(z_i \\ &- \sum_i \int_{\mathbb{R}^2} \frac{z_i x_i^2}{2\sigma_s}\tilde{q}_i^k(z_i)\mathrm{d}x_i\mathrm{d}z_i\end{aligned}$$

We have thus five terms:

- $$\begin{aligned}A &= -\int_{\mathbb{R}^N\times\mathbb{R}^N} \frac{\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}}{2\sigma_b^2}\prod_j q_j^\alpha(x_j)\tilde{q}^k(\boldsymbol{z})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z}\\ &= -\frac{1}{2\sigma_b^2}\int_{\mathbb{R}^N} \frac{\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}}{2\sigma_b^2}\prod_j q_i j^\alpha(x_j)\mathrm{d}\boldsymbol{x}\\ &= -\frac{1}{2\sigma_b^2}\int_{\mathbb{R}^N}\sum_{l=1}^N\sum_{p=1}^N x_l x_p h_{lp}\prod_j q_j^\alpha(x_j)\mathrm{d}\boldsymbol{x}\\ &= -\frac{1}{2\sigma_b^2}\int_{\mathbb{R}^N}\sum_{l=1}^N\sum_{p\neq l}^N x_l x_p h_{lp}\prod_j q_j^\alpha(x_j)\mathrm{d}\boldsymbol{x} - \frac{1}{2\sigma_b^2}\int_{\mathbb{R}^N}\sum_{p=1}^N x_p^2 h_{pp}\prod_j q_j^\alpha(x_j)\mathrm{d}\boldsymbol{x}\\ &= -\frac{1}{2\sigma_b^2}\left(\sum_{l=1}^N\sum_{p\neq l}^N \boldsymbol{m}_\alpha(l)\boldsymbol{m}_\alpha(p)h_{lp} + \sum_{p=1}^N h_{pp}(\boldsymbol{\sigma}_\alpha^2(p)+\boldsymbol{m}_\alpha^2(p))\right)\\ &= -\frac{1}{2}\left(\boldsymbol{m}_\alpha^T\frac{\boldsymbol{H}^T\boldsymbol{H}}{\sigma_b^2}\boldsymbol{m}_\alpha + \sum_{p=1}^N h_{pp}\boldsymbol{\sigma}_\alpha^2(p)\right)\end{aligned}$$

  Where $(h_{lp})_{1\leq l\leq N, 1\leq p\leq N}$ are the coefficients of $\boldsymbol{H}^T\boldsymbol{H}$.

- $$\begin{aligned}B &= \int_{\mathbb{R}^N\times\mathbb{R}^N}\frac{\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{y}}{\sigma_b^2}\prod q_i^\alpha(x_i)\tilde{q}^k(\boldsymbol{z})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z}\\ &= \boldsymbol{m}_\alpha^T\frac{\boldsymbol{H}^T\boldsymbol{y}}{\sigma_b^2}.\end{aligned}$$

- $$\begin{aligned}C &= \sum_i\int_{\mathbb{R}}\frac{1}{2}\ln(z_i)\tilde{q}^k(z_i)\mathrm{d}z_i + \sum_i\int_{\mathbb{R}}\Big((\tilde{a}_i-1)\ln(z_i)\mathrm{d}z_i\\ &= \sum_{i=1}^N\int_{\mathbb{R}}(\tilde{a}_i-\frac{1}{2})\ln(z_i)\tilde{q}^k(z_i)\mathrm{d}z_i.\end{aligned}$$

- $$D = -\sum_i\frac{1}{2}\ln(\sigma_s^2) + \int_{\mathbb{R}} z_i\tilde{b}_i\tilde{q}^k(z_i)\mathrm{d}z_i = -\sum_{i=1}^N\tilde{b}_i\frac{\boldsymbol{a}(i)}{\boldsymbol{b}_k(i)} - \frac{N}{2}\ln(\sigma_s^2).$$

- $$\begin{aligned}E &= -\sum_i\int_{\mathbb{R}^2}\frac{z_i x_i^2}{2\sigma_s^2}q_i^\alpha(x_i)\tilde{q}^k(z_i)\mathrm{d}x_i\mathrm{d}z_i\\ &= -\sum_i\int_{\mathbb{R}}\frac{x_i^2}{2\sigma_s^2}q_i^\alpha(x_i)\mathrm{d}x_i\int_{\mathbb{R}} z_i\tilde{q}^k(z_i)\mathrm{d}z_i\\ &= -\frac{\boldsymbol{a}(i)}{2\boldsymbol{b}_k(i)\sigma_s^2}\int_{\mathbb{R}} x_i^2 q_i^\alpha(x_i)dx_i\\ &= -\sum_i\frac{\boldsymbol{a}(i)(\boldsymbol{\sigma}_\alpha^2(i)+\boldsymbol{m}_\alpha^2(i))}{2\boldsymbol{b}_k(i)\sigma_s^2}.\end{aligned}$$

Furthermore, concerning the entropy we have

$$\mathcal{H}(\mathbf{X},\mathbf{Z}) = \sum_{i=1}^{N} \mathcal{H}(q^{\alpha}(x_i)) + \sum_{i=1}^{N} \mathcal{H}(\tilde{q}^{k}(z_i)).$$

And,

$$\begin{aligned}
\mathcal{H}(q^{\alpha}(x_i)) &= \frac{1}{2}(1+\ln(2\pi\boldsymbol{\sigma}_{\alpha}^{2}(i))),\\
\mathcal{H}(\tilde{q}^{k}(z_i)) &= -\int_{\mathbb{R}} (\boldsymbol{a}(i)-1)\ln(z_i)\tilde{q}^{k}(z_i)\mathrm{d}z_i + Const.
\end{aligned}$$

From Equation (4.7) we have $\boldsymbol{a}(i) = \tilde{a}_i + \frac{1}{2}$ and

$$C+\sum_{i=1}^{N}\mathcal{H}(\tilde{q}^{k}(z_i)) = \sum_{i=1}^{N}\sum_{i=1}^{N}\int_{\mathbb{R}}(\tilde{a}_i-\frac{1}{2})\ln(z_i)\tilde{q}^{k}(z_i)\mathrm{d}z_i - \int_{\mathbb{R}}(\tilde{a}_i+\frac{1}{2}-1)\ln(z_i)\tilde{q}^{k}(z_i)\mathrm{d}z_i + Const = Const$$

Finally, the negative free entropy is

$$F(q^{\alpha}) = A+B+D+E+\sum_{i}\mathcal{H}(x_i)+Const, \tag{7.5}$$

Let us determine the critical values of $g_k(\alpha) = F(q^{\alpha})$ and their signs.

As one can see, the derivative of $g_k(\alpha)$ is closely related to the derivatives of $\boldsymbol{\sigma}_{\alpha}^{2}(i)$ and $\boldsymbol{m}_{\alpha}(i)$ which, for $i = 1,\ldots,N$, are given by

$$\frac{\mathrm{d}\boldsymbol{\sigma}_{\alpha}^{2}(i)}{\mathrm{d}\alpha} = \frac{-\boldsymbol{\sigma}_{k}^{2}(i)\boldsymbol{\sigma}_{r}^{2}(i)(\boldsymbol{\sigma}_{k}^{2}(i)-\boldsymbol{\sigma}_{r}^{2}(i))}{(\boldsymbol{\sigma}_{r}^{2}(i)+\alpha(\boldsymbol{\sigma}_{k}^{2}(i)-\boldsymbol{\sigma}_{r}^{2}(i)))^{2}}, \tag{7.6}$$

whereas

$$\frac{\mathrm{d}\boldsymbol{m}_{\alpha}(i)}{\mathrm{d}\alpha} = \frac{\boldsymbol{\sigma}_{r}^{2}(i)\boldsymbol{\sigma}_{k}^{2}(i)(\boldsymbol{m}_{r}(i)-\boldsymbol{m}_{k}(i))}{(\boldsymbol{\sigma}_{r}^{2}(i)+\alpha(\boldsymbol{\sigma}_{k}^{2}(i)-\boldsymbol{\sigma}_{r}^{2}(i)))^{2}}. \tag{7.7}$$

And the second order derivatives are

$$\frac{\mathrm{d}^{2}\boldsymbol{\sigma}_{\alpha}^{2}(i)}{\mathrm{d}\alpha^{2}} = \frac{2\boldsymbol{\sigma}_{k}^{2}(i)\boldsymbol{\sigma}_{r}^{2}(i)(\boldsymbol{\sigma}_{k}^{2}(i)-\boldsymbol{\sigma}_{r}^{2}(i))^{2}}{(\boldsymbol{\sigma}_{r}^{2}(i)+\alpha(\boldsymbol{\sigma}_{k}^{2}(i)-\boldsymbol{\sigma}_{r}^{2}(i)))^{3}} \tag{7.8}$$

and,

$$\frac{\mathrm{d}^{2}\boldsymbol{m}_{\alpha}(i)}{\mathrm{d}\alpha^{2}} = \frac{2\boldsymbol{\sigma}_{r}^{2}(i)\boldsymbol{\sigma}_{k}^{2}(i)(\boldsymbol{m}_{r}(i)-\boldsymbol{m}_{k}(i))(\boldsymbol{\sigma}_{r}^{2}(i)-\boldsymbol{\sigma}_{k}^{2}(i))}{(\boldsymbol{\sigma}_{r}^{2}(i)+\alpha(\boldsymbol{\sigma}_{k}^{2}(i)-\boldsymbol{\sigma}_{r}^{2}(i)))^{3}}. \tag{7.9}$$

This entails that the first order derivative of g satisfies

$$\begin{aligned}
g'(\alpha) = \sum_{i=1}^{N} -\frac{1}{\sigma_b^2}\sum_{p=1}^{N}\frac{\mathrm{d}\boldsymbol{m}_{\alpha}(i)}{\mathrm{d}\alpha}\boldsymbol{m}_{\alpha}(p)h_{ip}\\
-\frac{h_{ii}}{2\sigma_b^2}\frac{\mathrm{d}\boldsymbol{\sigma}_{\alpha}^{2}(i)}{\mathrm{d}\alpha} + \frac{\mathrm{d}\boldsymbol{m}_{\alpha}(i)}{\mathrm{d}\alpha}\left(\frac{\boldsymbol{H}^{T}\boldsymbol{y}}{\sigma_b^2}\right)_i - \frac{\boldsymbol{a}(i)}{2\boldsymbol{b}_k(i)\sigma_s^2}\frac{\mathrm{d}\boldsymbol{\sigma}_{\alpha}^{2}(i)}{\mathrm{d}\alpha} - \frac{\boldsymbol{a}(i)\boldsymbol{m}_{\alpha}(i)}{\boldsymbol{b}_k(i)\sigma_s^2}\frac{\mathrm{d}\boldsymbol{m}_{\alpha}(i)}{\mathrm{d}\alpha} + \frac{\mathrm{d}\boldsymbol{\sigma}_{\alpha}^{2}(i)/\mathrm{d}\alpha}{4\pi\boldsymbol{\sigma}_{\alpha}^{2}(i)}
\end{aligned} \tag{7.10}$$

Whereas,

$$g''(\alpha) = \sum_{i=1}^{N} -\frac{1}{\sigma_b^2} \sum_{p=1}^{N} \left(\frac{\mathrm{d}^2 \boldsymbol{m}_\alpha(i)}{\mathrm{d}\alpha^2} \boldsymbol{m}_\alpha(p) + \frac{\mathrm{d}\boldsymbol{m}_\alpha(i)}{\mathrm{d}\alpha} \frac{\mathrm{d}\boldsymbol{m}_\alpha(p)}{\mathrm{d}\alpha} \right) h_{ip} - \frac{h_{ii}}{2\sigma_b^2} \frac{\mathrm{d}^2 \boldsymbol{\sigma}_\alpha^2(i)}{\mathrm{d}\alpha^2} + \frac{\mathrm{d}^2 \boldsymbol{m}_\alpha(i)}{\mathrm{d}\alpha^2} \left(\frac{\boldsymbol{H}^T \boldsymbol{y}}{\sigma_b^2} \right)_i$$
$$- \frac{\boldsymbol{a}(i)}{2\boldsymbol{b}_k(i)\sigma_s^2} \frac{\mathrm{d}^2 \boldsymbol{\sigma}_\alpha^2(i)}{\mathrm{d}\alpha^2} - \frac{\boldsymbol{a}(i)\boldsymbol{m}_\alpha(i)}{\boldsymbol{b}_k(i)\sigma_s^2} \frac{\mathrm{d}^2 \boldsymbol{m}_\alpha(i)}{\mathrm{d}\alpha^2} - \frac{\boldsymbol{a}(i)}{\boldsymbol{b}_k(i)\sigma_s^2} \left(\frac{\mathrm{d}\boldsymbol{m}_\alpha(i)}{\mathrm{d}\alpha} \right)^2 + \frac{\mathrm{d}^2 \boldsymbol{\sigma}_\alpha^2(i)/\mathrm{d}\alpha^2 \boldsymbol{\sigma}_\alpha^2(i) - \mathrm{d}\boldsymbol{\sigma}_\alpha^2(i)/}{4\pi(\boldsymbol{\sigma}_\alpha^2(i))^2} \tag{7.11}$$

In this case, the approximated critical value of F is given by α such that

$$\alpha_{subopt} = -\frac{\left.\frac{dF(q^\alpha)}{d\alpha}\right|_{\alpha=0}}{\left.\frac{\mathrm{d}^2 F(q^\alpha)}{d\alpha^2}\right|_{\alpha=0}}.$$

Finally, we consider $\alpha_{subopt} = -\frac{g'(0)}{g''(0)}$, where

$$g'(0) = -\left(\left.\frac{\mathrm{d}\boldsymbol{m}_\alpha}{\mathrm{d}\alpha}\right|_{\alpha=0} \right)^T \left(\frac{\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{m}_k - \boldsymbol{H}^T\boldsymbol{y}}{\sigma_b^2} \right) - \sum_{i=1}^{N} \left.\frac{\mathrm{d}\boldsymbol{m}_\alpha(i)}{\mathrm{d}\alpha}\right|_{\alpha=0} \frac{\boldsymbol{a}(i)\boldsymbol{m}_k(i)}{\boldsymbol{b}_k(i)\sigma_s^2}$$
$$+ \frac{1}{2} \sum_{i=1}^{N} \frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)} \left(\boldsymbol{\sigma}_r^2(i) - \boldsymbol{\sigma}_k^2(i)\right) \left(-\frac{h_{ii}}{\sigma_b^2} - \frac{\boldsymbol{a}(i)}{\boldsymbol{b}_k(i)\sigma_s^2} + \frac{1}{2\pi\boldsymbol{\sigma}_k^2(i)} \right) \Bigg]. \tag{7.12}$$

$$g'(0) = \sum_{i=1}^{N} -\frac{1}{\sigma_b^2} \sum_{p=1}^{N} \frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)} (\boldsymbol{m}_r(i) - \boldsymbol{m}_k(i)) \boldsymbol{m}_0(p) h_{ip} - \frac{h_{ii}}{2\sigma_b^2} \frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)} (\boldsymbol{\sigma}_r^2(i) - \boldsymbol{\sigma}_k^2(i))$$
$$+ \frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)} (\boldsymbol{m}_r(i) - \boldsymbol{m}_k(i)) \left(\frac{\boldsymbol{H}^T\boldsymbol{y}}{\sigma_b^2} \right)_i - \frac{a_i^k b_i^k}{2\sigma_s^2} \frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)} (\boldsymbol{\sigma}_r^2(i) - \boldsymbol{\sigma}_k^2(i)) - \frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)} (\boldsymbol{m}_r(i) - \boldsymbol{m}_k(i)) \frac{\boldsymbol{m}_0(i)}{\sigma_s^2}$$
$$+ \frac{\boldsymbol{\sigma}_r^2(i) - \boldsymbol{\sigma}_k^2(i)}{4\pi\boldsymbol{\sigma}_r^2(i)}$$
$$= \sum_{i=1}^{N} \frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)} \Bigg[(\boldsymbol{m}_r(i) - \boldsymbol{m}_k(i)) \left(-\frac{(\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{m}_k - \boldsymbol{H}^T\boldsymbol{y})_i}{\sigma_b^2} - \frac{\boldsymbol{m}_k(i)}{\sigma_s^2} \right)$$
$$+ (\boldsymbol{\sigma}_r^2(i) - \boldsymbol{\sigma}_k^2(i)) \left(-\frac{h_{ii}}{2\sigma_b^2} - \frac{a_i^k b_i^k}{2\sigma_s^2} + \frac{1}{4\pi\boldsymbol{\sigma}_k^2(i)} \right) \Bigg]$$
$$= -\left(\left.\frac{d\boldsymbol{m}_\alpha}{d\alpha}\right|_{\alpha=0} \right)^T \left(\frac{\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{m}_k - \boldsymbol{H}^T\boldsymbol{y}}{\sigma_b^2} + \frac{\boldsymbol{m}_k}{\sigma_s^2} \right) + \frac{1}{2} \sum_{i=1}^{N} \frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)} (\boldsymbol{\sigma}_r^2(i) - \boldsymbol{\sigma}_k^2(i)) \left(-\frac{h_{ii}}{\sigma_b^2} - \frac{a_i^k b_i^k}{\sigma_s^2} + \right. \tag{7.13}$$

And,

$$
\begin{aligned}
g''(0) = &-\left(\frac{\mathrm{d}^2\boldsymbol{m}_\alpha}{\mathrm{d}\alpha^2}\Big|_{\alpha=0}\right)^T\left(\frac{\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{m}_k-\boldsymbol{H}^T\boldsymbol{y}}{\sigma_b^2}\right)-\sum_{i=1}^{N}\frac{\mathrm{d}^2\boldsymbol{m}_\alpha(i)}{\mathrm{d}\alpha^2}\Big|_{\alpha=0}\frac{\boldsymbol{a}(i)\boldsymbol{m}_k(i)}{\boldsymbol{b}_k(i)\sigma_s^2}\\
&-\left(\frac{\mathrm{d}\boldsymbol{m}_\alpha}{\mathrm{d}\alpha}\Big|_{\alpha=0}\right)^T\left(\frac{\boldsymbol{H}^T\boldsymbol{H}}{\sigma_b^2}-\frac{\mathrm{Diag}(\boldsymbol{a})(\mathrm{Diag}(\boldsymbol{b}))^{-1}}{\sigma_s^2}\right)\frac{\mathrm{d}\boldsymbol{m}_\alpha}{\mathrm{d}\alpha}\Big|_{\alpha=0}\\
&+\sum_{i=1}^{N}\frac{\boldsymbol{\sigma}_k^2(i)}{(\boldsymbol{\sigma}_r^2(i))^2}(\boldsymbol{\sigma}_r^2(i)-\boldsymbol{\sigma}_k^2(i))^2\left(-\frac{h_{ii}}{\sigma_b^2}-\frac{\boldsymbol{a}(i)}{\boldsymbol{b}_k(i)\sigma_s^2}+\frac{1}{2\pi\boldsymbol{\sigma}_k^2(i)}\right)-\sum_{i=1}^{N}\frac{\boldsymbol{\sigma}_k^2(i)}{\boldsymbol{\sigma}_r^2(i)}\frac{\boldsymbol{\sigma}_r^2(i)-\boldsymbol{\sigma}_k^2(i)}{4\pi\boldsymbol{\sigma}_k^2(i)}.
\end{aligned}
\tag{7.14}
$$

REFERENCES

[1] R. Amiri, M. Alaee, H. Rahmani, and M. Firoozmand, *Chirplet based denoising of reflected radar signals*, Asia International Conference on Modelling & Simulation **0** (2009), 304–308.

[2] S. D. Babacan, R. Molina, and A. K. Katsaggelos, *Variationnal Bayesian Blind Deconvolution Using a Total Variation Prior*, IEEE Trans. Image Processing **18** (2009), no. 1, 12–26.

[3] N. Bali and A. Mohammad-Djafari, *Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis*, IEEE Trans. on Image Processing **17** (2008), no. 2, 217–225.

[4] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders, *Variational Bayesian image restoration based on a product of t-distributions image prior*, IEEE Trans. Image Processing **17** (2008), no. 10, 1795–1805.

[5] R. Chartrand and W. Brendt, *A Gradient Descent Solution to the Monge-Kantorovich Problem*, Applied Math. Sciences **3** (2009), no. 22, 1071–1080.

[6] R. A. Choudrey, *Variational methods for bayesian independent component analysis*, Phd thesis, University of Oxford, 2002.

[7] G. Demoment, *Image reconstruction and restoration: Overview of common estimation structure and problems*, IEEE Trans. Acoust. Speech, Signal Processing **assp-37** (1989), no. 12, 2024–2036.

[8] B. A. Frigyik, S. Srivastava, and M. R. Gupta, *Functional bregman divergence and bayesian estimation of distributions*, IEEE Trans. Inf. Theory **54** (2008), no. 11, 5130–5139.

[9] A. Globerson, T. Koo, X. Carreras, and M. Collins, *Exponentiated gradient algorithms for log-linear structured prediction*, In Proc. ICML, 2007, pp. 305–312.

[10] Y. Goussard, G. Demoment, and F. Monfront, *Maximum a posteriori dectection-estimation of Bernoulli-Gaussian processes*, Mathematics in Signal Processing II, Clarendon Press, Oxford, UK, 1990.

[11] M. M. Ichir and A. Mohammad-Djafari, *Hidden Markov models for wavelet-based blind source separation*, IEEE Trans. Image Processing **15** (2006), no. 7, 1887–1899.

[12] J. Idier (ed.), *Bayesian approach to inverse problems*, ISTE Ltd and John Wiley & Sons Inc., London, 2008.

[13] S. Jana and P. Moulin, *Optimality of klt for high-rate transform coding of gaussian vector-scale mixtures : Application to reconstruction, estimation, and classification*, IEEE Trans. on Info. Theory **52** (2006), no. 9, 4049–4067.

[14] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, *An Introduction to variational Methods for Graphical Models*, Machine Learning **37** (1999), no. 2, 183–233.

[15] A C Kak and M Slaney, *Principles of computerized tomographic imaging*, IEEE Press, New York, NY, 1988.

[16] H. Kellerer, *Measure theoretic versions of linear programming*, Math. Z. **198** (1988), no. 3, 367–400.

[17] J. Kivinen and M. Warmuth, *Exponentiated gradient versus gradient descent for linear predictors*, Information and Computation **132** (1997), no. 1, 1–63.

[18] M. Kowalski and T. Rodet, *An unsupervised algorithm for hybrid/morphological signal decomposition*, Proc. IEEE ICASSP (Prague), no. Id 2155, May 2011, pp. 4112 – 4115.

[19] D. MacKay, *Information theory, inference, and learning algorithms*, Cambridge University Press, 2003.

[20] D. J. C. MacKay, *Ensemble learning and evidence maximization*, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.54.4083, 1995.

[21] J.W Miskin, *Ensemble learning for independent component analysis*, Phd thesis, University of Cambridge, `http://www.inference.phy.cam.ac.uk/jwm1003/`, 2000.

[22] I. Molchanov, *Tangent sets in the space of measures: with applications to variational analysis*, J. Math. Anal. Appl. **249** (2000), no. 2, 539–552.

[23] I. Molchanov and S. Zuyev, *Steepest descent algorithms in a space of measures*, Statistics and Computing **12** (2002), 115–123.

[24] K. Morton, P. Torrione, and L. Collins, *Variational bayesian learning for mixture autoregressive models with uncertain-order*, IEEE Trans. Signal Processing **59** (2011), no. 6, 2614–2627.

[25] F. Natterer, *Algorithms in tomography*, Clarendon Press, 1997.

[26] Y. Nesterov, *Smooth minimization of non-smooth functions*, Mathematical Programming **103** (2005), no. 1, 127–152.

[27] J. Nocedal and S. J. Wright, *Numerical optimization*, Series in Operations Research, Springer Verlag, New York, 2000.

[28] A. Pai, E. Chassande-Mottin, and O. Rabaste, *Best network chirplet chain: Near-optimal coherent detection of unmodeled gravitational wave chirps with a network of detectors*, Phys. Rev. D **77** (2008), no. 062005, 1–22, Also available at gr-qc/0708.3493.

[29] O. Rabaste and T. Chonavel, *Estimation of multipath channels with long impulse response at low SNR via an MCMC method*, IEEE Trans. Signal Processing **55** (2007), no. 4, 1312 – 1325.

[30] C. Robert, *Simulations par la méthode MCMC*, Economica, Paris, France, 1997.

[31] C. P. Robert and G. Casella, *Monte-Carlo statistical methods*, Springer Texts in Statistics, Springer, New York, NY, 2000.

[32] W. Rudin, *Real and complex analysis*, McGraw-Hill Book Co., New York, 1987.

[33] M. Seeger, *Bayesian inference and optimal design for sparse linear model*, Jour. of Machine Learning Research **9** (2008), 759–813.

[34] M. Seeger and D. P. Wipf, *Variational bayesian inference techniques*, IEEE Trans. Signal Processing Mag. **27** (2010), no. 6, 81–91.

[35] V. Smidl and A. Quinn, *The variational bayes method in signal processing*, Springer, 2006.

[36] ———, *Variational bayesian filtering*, IEEE Trans. Signal Processing **56** (2008), no. 10, 5020–5030.

[37] H. Snoussi and A. Mohammad-Djafari, *Bayesian unsupervised learning for source separation with mixture of gaussians prior*, International Journal of VLSI Signal Processing Systems **37** (2004), no. 2-3, 263–279.

[38] M. J. Wainwright and E. P. Simoncelli, *Scale mixtures of gaussians and the statistics of natural images*, Advances in Neural Information Processing Systems **12** (2000), no. 1, 855–861.

[39] G. Wang and Z. Bao, *Inverse synthetic aperture radar imaging of maneuvering targets based on chirplet decomposition*, Opt. Eng. **38** (1999), no. 1534.

[40] M. Zibulevsky and M. Elad, *L1-l2 optimization in signal and image processing*, IEEE Signal Processing Magazine (2010), 76–88.

F. Orieux, J.-F. Giovannelli, **T. Rodet**, A. Abergel, H. Ayasso et M. Husson, « Superresolution in map-making based on physical instrument model and regularized inversion. Application to SPIRE/Herschel. », *Astronomy & Astrophysics*, vol. 539, n°A38, pp. 16, mars 2012.

A&A 539, A38 (2012)
DOI: 10.1051/0004-6361/201116817

Astronomy & Astrophysics

Super-resolution in map-making based on a physical instrument model and regularized inversion

Application to SPIRE/*Herschel*

F. Orieux[1], J.-F. Giovannelli[1,2], T. Rodet[1], A. Abergel[3], H. Ayasso[3], and M. Husson[3]

[1] Laboratoire des Signaux et Systèmes (– Supélec – Univ. Paris-Sud 11), Plateau de Moulon, 91192 Gif-sur-Yvette, France
e-mail: `[orieux;rodet]@lss.supelec.fr`
[2] Univ. Bordeaux, IMS, UMR 5218, 33400 Talence, France
e-mail: `Giova@IMS-Bordeaux.fr`
[3] Institut d'Astrophysique Spatiale (– Univ. Paris-Sud 11), 91 405 Orsay, France
e-mail: `abergel@ias.u-psud.fr`

Received 3 March 2011 / Accepted 5 December 2011

ABSTRACT

We investigate super-resolution methods for image reconstruction from data provided by a family of scanning instruments like the *Herschel* observatory. To do this, we constructed a model of the instrument that faithfully reflects the physical reality, accurately taking the acquisition process into account to explain the data in a reliable manner. The inversion, i.e. the image reconstruction process, is based on a linear approach resulting from a quadratic regularized criterion and numerical optimization tools. The application concerns the reconstruction of maps for the SPIRE instrument of the *Herschel* observatory. The numerical evaluation uses simulated and real data to compare the standard tool (coaddition) and the proposed method. The inversion approach is capable to restore spatial frequencies over a bandwidth four times that possible with coaddition and thus to correctly show details invisible on standard maps. The approach is also applied to real data with significant improvement in spatial resolution.

Key words. methods: numerical – techniques: photometric – methods: data analysis – techniques: image processing – instrumentation: photometers – techniques: high angular resolution

1. Introduction

Map making is a critical step in the processing of astronomical data of various imaging instruments (interferometers, telescopes, spectro-imager, etc.), and two recent special issues have been published (Leshem et al. 2010, 2008) on the subject. Because the observed sky may contain structures of various scales, from extended emission to point sources, the challenge is to design reconstruction methods that deliver maps that are photometrically valid for the broadest range of spatial frequencies.

For long-wavelength instruments, be they ground based (SCUBA/JCMT, LABOCA/APEX, etc.), on-board balloons (Archeops, BLAST, etc.) or space borne (IRAS, ISO, *Spitzer*, WMAP, *Planck*, *Herschel*, etc.), the task is especially challenging for two reasons. First, the physical resolution is poor at these wavelengths. Second, the distance between the detectors of these instruments generally prevents a proper sampling of the focal plane, given the maximum spatial frequency allowed by the optical response. Therefore, specific scanning strategies have to be defined, which depend on the detector positions and need to be closely combined with a well-designed image reconstruction method.

The *Herschel* Space Observatory (Pilbratt et al. 2010) was launched in May 2009 together with the *Planck* satellite. It continuously covers the 55–672 μm spectral range with its very high spectral resolution spectrometer HIFI (de Graauw et al. 2010) and its two photometers / medium resolution spectrometers PACS (Poglitsch et al. 2010) and SPIRE (Griffin et al. 2010). With a 3.5 m primary mirror, *Herschel* is the largest space telescope launched to date. In order to take full advantage of the telescope size, the accurate representation and processing of the highest spatial frequencies presents a particular challenge. To this end, two step-by-step photometer pipelines have been developed by the instrument consortia by Griffin et al. (2008) for SPIRE and by Wieprecht et al. (2009) for PACS: they produce flux density timelines corrected for various effects, calibrated and associated with sky coordinates (level-1 products), then produce maps (level-2 products). An important step is the correction of the $1/f$ noise components, which can be correlated or uncorrelated between bolometers. For SPIRE, a significant fraction of the correlated component is processed using the signals delivered by blind bolometers. For PACS, it is currently processed using different kinds of filtering. The glitches caused by the deposit of thermal energy by ionizing cosmic radiation are flagged or corrected. Finally, the timeline outputs can be simply coadded on a spatial grid to produce "naive maps", with a rounded pointing approximation. Maximum likelihood approaches with the same coaddition algorithm, namely MADmap (Cantalupo et al. 2010) and SANEPIC (Patanchon et al. 2008) have also been developed to compute maps, using the spatial redundancy to correct for the $1/f$ noise.

There are several drawbacks to these pipelines. First, because they work on a step-by-step basis, the performance of the whole process is limited by the step with the worst performance. Second, the ultimate performance of one step is out of reach because only a reduced part of the available information is handed over from the previous step. This mean that better perfomances can be achieved by a more global approach. More important, the instrument and the telescope properties (mainly the diffraction) are not taken into account, which is why the maps are unavoidably smoothed by the Point Spread Function (PSF), whereas the scanning strategy allows higher spatial frequencies to be indirectly observed.

To overcome these limitations, we resorted to an inverse problem approach (Idier 2008) that is based on an instrument model and an inverson method.

- It requires an instrument model that faithfully reflects the physical reality to distinguish in the observations between what is caused by the instrument and what is due to the actual sky. To this end, an important contribution of our paper is an analytical instrument model based on a physical description of the phenomena as functions of continuous variables. Moreover, it includes scanning strategy, mirror, wavelength filter, feedhorns, bolometers and read-out electronics. The point for the resolution is the following. On the one hand, the field of view is covered by hexagonally packed feedhorn-coupled bolometers, the sampling period is twice the PSF width, which potentially lead to spectral aliasing for wide-band objects. On the other hand, the scanning strategy with a pointing increment lower than the bolometer spacing introduces an higher equivalent sampling frequency. Therefore, it is crucial to properly take into account the scanning strategy and the whole instrument including irregular sampling to obtain super-resolution (see also the analysis in Orieux et al. 2009). To the best of our knowledge, a physical model of the instrument this accurate has never been used in a map making method.
- The inversion of our instrument model constitutes an ill-posed problem (Idier 2008) because of the deficit of available information induced by convolution with the instrument PSF. Moreover, the ill-posedness becomes all the more marked as the resolution requirement increases. The inversion methods must therefore exploit other information by regularization to compensate for the deficits in the observations. Each reconstruction method is therefore specialized for a certain class of maps (point sources, diffuse emission, superposition of the two, etc.) according to the information that is included. From this standpoint, the present paper is essentially devoted to extended emission.
  The method is linear w.r.t. the data for the sake of simplicity and computational burden. From the methodological point of view, it is built within the framework of quadratic regularization (Tikhonov & Arsenin 1977; Andrews & Hunt 1977). It relies on a criterion involving an adequation measure (observed data vs model output) and a spatial smoothness measure. From a numerical standpoint, we resort to a gradient-based optimisation algorithm (Nocedal & Wright 2000) to compute the map.
  Moreover, in as much as it relies on two sources of information, the method is based on a trade-off tuned by means of an hyperparameter. It is empirically set in the present paper and work in progress, based on Robert & Casella (2000) and Orieux et al. (2010), is devoted to the question of the hyperparameter and instrument parameter auto-calibration (myopic and unsupervised inversion).

One of the most striking results of our research is the correct restoration of small-scale structures (wide-band), which are not detectable on naive maps. This result is reached thanks to the developed instrument model together with the used inversion: they jointly enable the proposed method to reduce instrument effects, overtake instrument limitations and restore high spatial frequencies.

In the image processing community, these capabilities are referred to as super-resolution (Park et al. 2003) and we were partly inspired by recent developments in this field. They are usually based on various (scene or camera) motion or scanning strategy. Some of them account for possible rotation (Elad & Feuer 1999) and/or a magnifying factor (Rochefort et al. 2006). Other approaches introduce an edge-preserving prior (Nguyen et al. 2001; Woods et al. 2006). These works rely on the description of the unknown object as a function of continuous variables that is decomposed on pixel indicator basis (Hardie et al. 1997; Patti et al. 1997), on a truncated discrete Fourier basis (Vandewalle et al. 2007), on a family of regularly shifted Gaussian functions (Rodet et al. 2008), or spline family (Rochefort et al. 2006). Other approaches have been proposed, based on shift-and-add step (Farsiu et al. 2004) followed by a deconvolution step (Molina & Ripley 1989). Finally, several contributions are devoted to the performance of super-resolution approaches (Champagnat et al. 2009; Orieux et al. 2009).

The paper is organized as follows. The instrument model describing the relationship between the measured data and the unknown sky is presented in Sect. 2. Section 3 details the method that we propose to inverse the data and compute high-resolution maps. Finally, Sect. 4 presents experimental results, first on simulated data (Sect. 4.1), then on real data (Sect. 4.2).

2. Instrument model

The prime objective of the instrument model is the reproduction of observed data taking into account the physics of the acquisition. In addition, the reconstruction algorithms use the instrument model many times, it is therefore necessary to adopt hypotheses and approximations to the reduce computational burden. This is one of the differences to a simulator (Sibthorpe et al. 2009; Sibthorpe & Griffin 2006), which is designed to be run once per data set.

2.1. Physical models

2.1.1. Mode of observation

The sky, $\mathcal{X}(\alpha,\beta,\lambda)$, is characterized by two spatial dimensions (α,β) and one spectral dimension λ. To model telescope translations, we used a frame of reference defined by the instrument. The map at the input is time-dependent and can be written

$$\mathcal{X}(\alpha,\beta,\lambda,t) = \mathcal{X}(\alpha - p_\alpha(t), \beta - p_\beta(t), \lambda), \tag{1}$$

where α and β define the central angular position of the observation and $(p_\alpha(t), p_\beta(t))$ the translations into the two directions as a function of time t.

Here, we present only the *"Large map"* protocol. Data were acquired over a complete observation sequence composed of two almost perpendicular directions and several scans back and forth for each of the two directions. The pointing acceleration and deceleration phases were not included in the zone of interest and there was no rotation during the observation sequence. The pointing functions are therefore written

$$p_\alpha(t) = v_\alpha t + c_\alpha \qquad \text{and} \qquad p_\beta(t) = v_\beta t + c_\beta \tag{2}$$

for scanning at a constant velocity (v_α, v_β). The pointing accuracy is of the order of a few seconds of arc. This protocol enables us to introduce spatial redundancy, which is an essential element for the reconstruction of a sky at a resolution greater than the detector spatial sampling period (Orieux et al. 2009; Champagnat et al. 2009).

2.1.2. Optics

The *Herschel* Telescope is a classical Cassegrain instrument with a 3.5 m diameter primary mirror and a 308 mm diameter secondary mirror. The SPIRE photometer has three channels for a single field of view. The light is split by a combination of dichroics and flat-folding mirrors. The spectral channels are defined by a sequence of metal mesh filters and the reflection/transmission edges of the dichroics. They are centred at approximately 250, 350 and 500 μm (noted as PSW, PMW and PLW respectively). We assumed the overall transmission curves of the wavelength filter $h_k(\lambda)$, for $k = 1, 2, 3$, as given by the SPIRE Observers' Manual (no analytical form is available).

The three detector arrays contain 139 (250 μm), 88 (350 μm) and 43 (500 μm) bolometers, each coupled to the telescope beam with hexagonally close-packed circular feedhorns. The beam solid angle is apodized by a bell-shaped weight whose width increases with λ. Efforts have been made to correctly integrate the feedhorns in the instrument model but the detailed coupling of feedhorns on incoming radiation is, to the best of our knowledge (Griffin et al. 2002), not fully understood at present.

Our final choice as an effective PSF for the telescope coupled with feedhorns was a Gaussian shape $h_o(\alpha, \beta, \lambda)$. This choice has two advantages: (i) it allows a closed equation for the instrument model (see Sect. 2.2), and (ii) it agrees with the response measured from observations of Neptune (Griffin et al. 2010). As a first approach, we assumed isotropic Gaussians with standard deviations $\sigma_o(\lambda) = c\lambda$ proportional to the wavelength since the width of the beam varies almost linearly with the wavelength. The widths obtained are close to the FWHM measured on the sky with 18.1″, 25.2″, and 36.9″ at 250 μm, 350 μm and 500 μm, respectively (Griffin et al. 2010). The feedhorn diameter is $2F\lambda$, which introduces a detector spatial sampling period of $2F\lambda$ (50″ for the 350 μm array, or equivalently with sampling frequency $f_s \approx 0.02\,\text{arcsec}^{-1}$).

The output after each feedhorn is then written as a 2D convolution of the input $\mathcal{X}(\alpha, \beta, \lambda, t)$ and the effective PSF h_o in addition to the h_k wavelength filter

$$\mathcal{X}_k^{lm}(\lambda, t) = h_k(\lambda) \iint \mathcal{X}(\alpha, \beta, \lambda, t) \times h_o(\alpha - \alpha_{lm}, \beta - \beta_{lm}, \lambda)\, \mathrm{d}\alpha \mathrm{d}\beta \quad (3)$$

where $(\alpha_{lm}, \beta_{lm})$ is the direction pointed at by the feedhorn (l, m), for $l = 1, \ldots L$ and $m = 1, \ldots M$. The k subscript can be safely removed from $\mathcal{X}_k^{lm}$ since each spectral band is processed separately. Finally, the optics was modelled as a linear invariant system w.r.t. continuous variable.

2.1.3. Bolometers

To set up the bolometer model, we took the thermal model of Sudiwala et al. (2002), which was also used in the simulator developed by Sibthorpe et al. (2009). Bolometers absorb the entire received radiation

$$P^{lm}(t) = \int_\lambda \mathcal{X}^{lm}(\lambda, t)\, \mathrm{d}\lambda, \quad (4)$$

and this power provides the system excitation. The temperature $T^{lm}(t)$ determines the system output. The link between the input $P(t)$ and the response $T(t)$ is described by the differential equation deduced from a thermal balance,

$$C\frac{\mathrm{d}T}{\mathrm{d}t} - \frac{R(T)V_\mathrm{p}^2}{R_\mathrm{c}^2} + \frac{G_0}{T_0^\nu(\nu + 1)}\left(T^{\nu+1} - T_0^{\nu+1}\right) = P,$$

where C is the heat capacity of the bolometer, $R(T)$ is its resistivity, T_0 is the temperature of the thermal bath, ν is a physical parameter that depends on the bolometer, G_0 is the thermal conductance (at temperature T_0) and V_p and R_c are the polarization voltage and charge. No explicit solution of this equation is available in the literature. Sudiwala's approach (Sudiwala et al. 2002), which we adopted here, is to linearize this equation around an operating point $(\bar{T}, \bar{P})$. In the following, we consider only the variable part of the flux and exclude the constant part that defines the operating point. All constants are defined with respect to the operating point.

For SPIRE, most of the observations should be carried out in the linear regime (Griffin 2006, 2007). We therefore considered that a development is sufficient to model the bolometer behaviour correctly. Then, knowing the variations of the resistivity $R(T)$ with temperature, it is possible to determine the tension at the terminals. This first-order development models the bolometer as a first-order, low-pass filter with an impulse response

$$h_\mathrm{b}(t) = S \exp\left[-t/\tau\right], \quad (5)$$

where the gain S and the time constant τ depend on the physical parameters in the differential equation (Sudiwala et al. 2002). The values of these parameters are defined with respect to the operating point and correspond to the official SPIRE characteristics (Griffin 2006, 2007). The output voltage around the operating point can then be written as a function of the incident flux,

$$y_{lm}(t) = \int_{t'} \int_\lambda \mathcal{X}^{lm}(\lambda, t') h_\mathrm{b}(t' - t)\mathrm{d}t'\, \mathrm{d}\lambda. \quad (6)$$

Finally, downstream, we have the read-out electronics, composed of several stages (modulation, amplification, low-pass filter, demodulation, quantification). However, except for the low-pass filters, they seem to have negligible effects to the other elements and are not included in our model. The equations are nevertheless available (Griffin 2007) and it is possible to integrate them into the model.

The low-pass filters introduce a delay on the data with respect to the telescope position along the scan. As a trade-off between model accuracy and computational burden, we have chosen to model the combination of the low-pass filter and the bolometer as a unique first-order filter. The time constant[1] value (0.2 s) is taken to be representative of the combination.

Finally, we accounted for regular time sampling that takes the values at times $t = nT_\mathrm{s}$ (with a sampling frequency $F_\mathrm{s} = 1/T_\mathrm{s} \approx 30\,\text{Hz}$) and then $y_{lm}^n = y_{lm}(nT_\mathrm{s})$, for $n = 1, \ldots N$. Given the scanning speed of 30″ s^{-1} this induces a spatial sampling period of 2″ between two succesive time samples for one bolometer, while the detector sampling period is 50″ for the 350 μm array.

---

[1] For the illustration on real data in Sect. 4.2, the correction of the low-pass filter was performed using the *Herschel* Interactive Processing Environment (Ott 2010), and the time constant of the first-order low-pass filter was set to the time constant for the bolometer alone (5.7 ms).

2.1.4. Complete model equation

Adding these elements yields the equation of the acquisition chain. For a spectral channel k, the time signal at the bolometer (l, m) at time n is

$$y^n_{lm} = \iint h_k(\lambda) \iint \mathcal{X}(\alpha - p_\alpha(t), \beta - p_\beta(t), \lambda) \times h_\mathrm{o}(\alpha - \alpha_{lm}, \beta - \beta_{lm}, \lambda)\,\mathrm{d}\alpha\,\mathrm{d}\beta\,h_\mathrm{b}(nT_\mathrm{s} - t)\,\mathrm{d}\lambda\,\mathrm{d}t. \quad (7)$$

This equation introduces four integrals: two from the optics (spatial convolution), one from the spectral integration, and one from the time convolution. This is the fundamental equation of the instrument model since it describes the data y^n_{lm} bolometer by bolometer at each instant as a function of the sky $\mathcal{X}(\alpha, \beta, \lambda)$. It should be noted that this model includes the discretization process (and possible aliasing) in the sense that the data y^n_{lm} is a discret set and $\mathcal{X}$ is a function of continuous variables.

2.1.5. Super-resolution sky model

The model of the sky is an important element for the reconstruction method. As stated in the introduction and presented in Sect. 2.1.1, the sub-pixel scanning strategy should allow for reverse aliasing and enable to estimate a super-resolved sky (Orieux et al. 2009). The model of $\mathcal{X}$ must therefore be suitable for super-resolved reconstruction and, in particular, allows a fine description of the physical reality and integration with the instrument model.

Firstly, unlike conventional models of SPIRE (Sibthorpe et al. 2009; Cantalupo et al. 2010), we considered the sky spectrum within each channel. The emission observed by SPIRE is mainly caused by thermal equilibrium (between emission and absorption of UV and visible photons from incident radiation), and the intensities can be written

$$I_\lambda = \tau_{\lambda_0} \times \left(\frac{\lambda}{\lambda_0}\right)^{-\beta} \times B_\lambda(T), \quad (8)$$

where τ_{λ_0} is the optical depth at wavelength λ_0, β is the spectral index, B_λ is the *Planck* function, and T the dust temperature. The SPIRE data alone do not allow the proper measurement of the dust temperature (the combination of SPIRE and PACS is mandatory, Abergel et al. 2010), consequently we decided to exclude the dust temperature in our sky model and work in the Rayleigh-Jeans approximation, so that $B_\lambda(T) \propto \lambda^{-4}$. Moreover, we assumed $\beta = 2$, which is the "standard" value of the diffuse ISM (e.g., Boulanger et al. 1996). Finally, we have

$$\mathcal{X}(\alpha, \beta, \lambda) = \lambda^{-\varrho} \mathcal{X}(\alpha, \beta) \quad (9)$$

with $\varrho = 6$. However, as we will see in Sect. 2.2, the wavelength integration of the acquisition model will be performed numerically. In other words, the spectrum profile can be set adequately with the available knowledge of the observed sky.

Secondly, $\mathcal{X}(\alpha, \beta)$ was generated onto a family of functions regularly shifted in space: $\psi_{ij}(\alpha, \beta) = \psi(\alpha - i\delta_\alpha, \beta - j\delta_\beta)$ where ψ is an elementary function and $(\delta_\alpha, \delta_\beta)$ are the shifts between the ψ_{ij} in (α, β). We then obtain

$$\mathcal{X}(\alpha, \beta) = \sum_{ij} x_{ij}\,\psi(\alpha - i\delta_\alpha, \beta - j\delta_\beta), \quad (10)$$

where ψ is the generating function and x_{ij} are the coefficients. In addition, the axis α is determined by the first scan of the observation.

One of the purposes of this model is to describe maps with arbitrary fine details, that is to say, arbitrary wide band. Within this model, the usual function ψ is the cardinal sine with shift and width adapted to the target band. However, cardinal sines require analytical calculations that cannot be made explicit. To lighten the computational burden, we chose Gaussian ψ functions. These functions are parametrized by their spatial shifts $(\delta_\alpha, \delta_\beta)$ and their standard deviations $(\sigma_\alpha, \sigma_\beta)$. The parameters $(\delta_\alpha, \delta_\beta)$ are chosen to be equal to the inverse of the target band width as for the cardinal sines. In the numerical processing of Sect. 4, the values of $(\delta_\alpha, \delta_\beta)$ are equal to the sampling period of $2''$ induced by the scanning scheme of the *"Large map"* protocol (Orieux 2009). For the Gaussian function width parameters $(\sigma_\alpha, \sigma_\beta)$, we determined the value that minimizes the difference between the width at half-maximum of the cardinal sine and the Gaussian: $\sigma_{\alpha/\beta} \approx 0.6\,\delta_{\alpha/\beta}$ in a similar manner in α and β.

2.2. Explicit calculation of the acquisition model

Given the linearity of the instrument model (7) and the sky model (9), (10), the instrument output for a given sky is

$$y^n_{lm} = \sum_{ij} x_{ij} \int \lambda^{-\varrho} h_k(\lambda) \iint h_\mathrm{o}(\alpha - \alpha_{lm}, \beta - \beta_{lm}, \lambda) \times \psi(\alpha - i\delta_\alpha - p_\alpha(t), \beta - j\delta_\beta - p_\beta(t))\,\mathrm{d}\alpha\mathrm{d}\beta\,h_\mathrm{b}(nT_\mathrm{s} - t)\,\mathrm{d}t\,\mathrm{d}\lambda. \quad (11)$$

Thus, to obtain the contribution of a sky coefficient x_{ij} to a data item y^n_{lm}, it is necessary to calculate four integrals, whose discretization by brute force would result in time-consuming numerical computations.

Concerning the optics, the convolution of the function ψ with the optical response h_o appears in Eq. (11) and, because these are Gaussians, the convolution can be made explicit

$$\iint \psi(\alpha - i\delta_\alpha - p_\alpha, \beta - j\delta_\beta - p_\beta) h_\mathrm{o}(\alpha - \alpha_{lm}, \beta - \beta_{lm})\,\mathrm{d}\alpha\mathrm{d}\beta \propto \exp\left[-\frac{(p_\alpha + i\delta_\alpha - \alpha_{lm})^2}{2\Sigma^2_\alpha} - \frac{(p_\beta + j\delta_\beta - \beta_{lm})^2}{2\Sigma^2_\beta}\right] \quad (12)$$

with, in a similar manner in α and β: $\Sigma^2_{\alpha/\beta} = \sigma^2_{\alpha/\beta} + \sigma^2_\mathrm{o}$.

For the integral over time, only the constant velocity phases can be explicitly described for the *"Large map"* protocol. To integrate over time in (11), we used the expressions of (2) for $p_\alpha(t)$ and $p_\beta(t)$, which gives

$$\sum_{ij} x_{ij} \int \lambda^{-\varrho} h_k(\lambda) \int \exp\left[-\frac{(v_\alpha t + c_\alpha + i\delta_\alpha - \alpha_{lm})^2}{2\Sigma^2_\alpha}\right] \times \exp\left[-\frac{(v_\beta t + c_\beta + j\delta_\beta - \beta_{lm})^2}{2\Sigma^2_\beta}\right] h_\mathrm{b}(nT_\mathrm{s} - t)\,\mathrm{d}t\,\mathrm{d}\lambda. \quad (13)$$

It can be shown that explicit integration can be performed by including the Gaussians and the bolometer response (see details of the calculations in Appendix B, and the model becomes

$$y_{lm}^{n} = \frac{S}{2\sqrt{2\pi}\Sigma_{v}} \sum_{ij} x_{ij} \int \lambda^{-\varrho} h_{k}(\lambda) \times \operatorname{erfcx}\left(\frac{\Sigma_\alpha \Sigma_\beta}{\sqrt{2}\tau\Sigma_v} - \frac{\Sigma_\beta v_\alpha (o_\alpha + nT_\mathrm{s} v_\alpha)}{\sqrt{2}\Sigma_\alpha \Sigma_v} - \frac{\Sigma_\alpha v_\beta (o_\beta + nT_\mathrm{s} v_\beta)}{\sqrt{2}\Sigma_\beta \Sigma_v} \right) \times \exp\left[-\frac{(o_\alpha + nT_\mathrm{s} v_\alpha)^2}{2\Sigma_\alpha^2} - \frac{\left(o_\beta + nT_\mathrm{s} v_\beta\right)^2}{2\Sigma_\beta^2} \right] \mathrm{d}\lambda. \quad (14)$$

In this equation, the angles o_α and o_β are defined by $o_\alpha = c_\alpha + i\delta_\alpha - \alpha_{lm}$ and $o_\beta = c_\beta + j\delta_\beta - \beta_{lm}$. Moreover, $\Sigma_v^2 = \Sigma_\beta^2 v_\alpha^2 + \Sigma_\alpha^2 v_\beta^2$.

The data point y_{lm}^n does not depend directly on the "scanning time" t because it is integrated. It depends on time through the sampling instant n occuring *only* in $nT_\mathrm{s}v_{\{\alpha,\beta\}}$, i.e. a distance. In practice, the sampling period along the scans $T_\mathrm{s}\sqrt{v_\alpha^2 + v_\beta^2}$ is much shorter than the sampling period of the detecteor array. Thus, this properly modelled sampling scheme is a key element for reconstruction with an higher resolution.

In addition, the time constant of the bolometer and the electronics τ appears only in the argument of the function erfcx. It is consequently through this function that the bolometers and the electronics influence the *spatial* response.

The dependence on the wavelength through $\sigma_\mathrm{o}(\lambda)$ precludes explicit integration with respect to λ. However, the integral depends neither on the data nor on the unknown object but only on the protocol. Accordingly, for a given protocol, these integrals can be calculated once and for all. Finally, the work described above allow three explicit intergration of the four introduced by the initial model.

Equation (14) models the acquisition of the data item y_{lm}^n at time n by bolometer (l, m) from the coefficients x_{ij}. These equations can be written

$$y_{lm}^{n} = \sum_{ij} x_{ij} \mathcal{H}_{lmn}(\psi_{ij}), \quad (15)$$

where $\mathcal{H}$ is calculated from Eq. (14). The model is linear and we can therefore write

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}, \quad (16)$$

where $\boldsymbol{y}$ and $\boldsymbol{x}$ are vectors of size LMN and IJ, and $\boldsymbol{H}$ is a $LMN \times IJ$ matrix, each row of which can be deduced from (14) by varying l, m, n for fixed i, j.

2.3. Invariant structure

Initially, the physical model (7) is based on convolutive (so invariant) transforms w.r.t. continuous variables. However, the discretization operation is inhomogeneous, consequently the invariance property does not hold anymore, which lead to long computational time. Nevertheless, the trace of this initial invariance can still be perceived because $\boldsymbol{H}$ is a *sum* of terms at different spatial positions of the Gaussians (cf. Eq. (14)). Because the problem is now discretized, we seek to bring out an invariance by *quantified* shifts in

$$c_\alpha + i\delta_\alpha + nT_\mathrm{s}v_\alpha - \alpha_{lm}$$

for the α direction, and similarly for β. With the approximation that the terms are multiples of a common factor Δ_α, the continuous shift is

$$o_\alpha + nT_\mathrm{s}v_\alpha = (n_0 + in_1 + nn_2 - ln_3 - mn_4)\Delta_\alpha.$$

The pointed directions are therefore rounded to match the grid of sky. The MADmap and SANEPIC methods use this idea but there is a notable difference: they perform the operation on a low-resolution grid, which limits the map resolution. In contrast, the developments proposed here exploit the idea of a high-resolution grid, enabling super-resolution reconstruction. By acting in the same way in the β direction, we have

$$y_{lm}^{n} = \sum_{ij} x_{ij} \mathcal{H}\Big((n_0 + in_1 + nn_2 - ln_3 - mn_4)\Delta_\alpha, (n_0' + in_1' + nn_2' - ln_3 - mn_4)\Delta_\beta\Big) \quad (17)$$

and by computing the discrete convolution, we obtain

$$\tilde{y}(i', j') = \sum_{ij} x_{ij} \mathcal{H}\Big((i - i')\Delta_\alpha, (j - j')\Delta_\beta\Big). \quad (18)$$

Therefore, $y_{lm}^n = \tilde{y}(i', j')$ if, and only if,

$$i - i' = in_1 + ln_3 + mn_4 - nn_2 - n_0 \quad (19)$$

$$j - j' = jn_1' + ln_3 + mn_4 - nn_2' - n_0'. \quad (20)$$

In these conditions, the data $\boldsymbol{y}$, for a given scanning direction are computed by discrete convolution (18) followed by (inhomogeneous) down-sampling defined by (19), (20), which is much more efficient than using a generic linear model (16). First of all, the decomposition by convolution then decimation is faster than the direct calculation and, what is more, the convolution can be computed by FFT. Finally, given that only the impulse response is necessary, there is no need to compute and store all elements of the matrix.

In this form, some computations may be made even though they are useless, because the convolution is performed for all indices, whereas only some of them are used. In practice, the excess computation is reduced because we chose shifts $(\delta_\alpha, \delta_\beta)$ close to the sampling period induced by the scanning scheme. Almost all convolution results are observed, from 1 to 7 times for PSW as illustrated in Fig. 1.

There is, however, the disadvantage that the bolometer positions are approximated. Yet these positions are important because they allow to best exploit the data and to properly manage the information needed to estimate high frequencies. We chose a step Δ that is close to the sampling period along the scan, i.e. $\Delta \approx 2''$. The error introduced is therefore small. This can be seen to be all the more valid when we consider the expected level of noise and telescope pointing errors, which are of the same order of magnitude, $2''$.

Finally, the initial model (16) is decomposed in the discrete convolution defined by (18) following the (inhomogeneous) down-sampling defined by (19), (20), that is to say, $\boldsymbol{H}$ is factorised and

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{H}_\mathrm{c}\boldsymbol{x}, \quad (21)$$

where $\boldsymbol{H}_\mathrm{c}$ is a convolution matrix and $\boldsymbol{P}$ a pointing matrix that takes the values observed after convolution. It has one, and only one, "1" per row because each data item can only come from one position. Some columns may be entirely zero because certain coefficients may not be observed. Conversely, some columns may contain several "1" because certain coefficients may be observed several times.

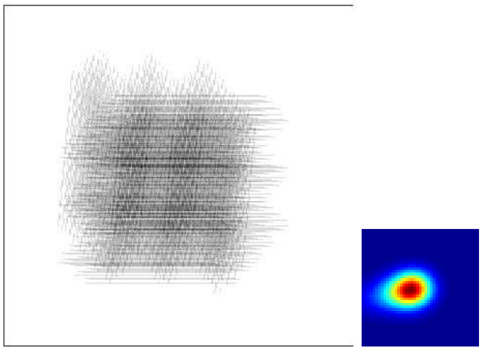


Fig. 1. Factorised physical model (PSW detector, velocity of 30″/s towards the left): map of spatial redundancies $\boldsymbol{P}$ (*left*) and spatial impulse response $\boldsymbol{H}_\mathrm{c}$ (*right*). The spatial scales are different for better visualisation of the impulse response.

To summarize, using an approximation of the pointed direction, we have separated the model $\boldsymbol{H}$ into two sub-models $\boldsymbol{H} = \boldsymbol{P}\boldsymbol{H}_\mathrm{c}$, where $\boldsymbol{H}_\mathrm{c}$ is invariant and $\boldsymbol{P}$ contains the non-invariant structure. This decomposition is broadly similar to the one generally found in super-resolution in the field of image processing (see references in the introduction).

Figure 1 presents this decomposition for the PSW detector with a velocity of 30″/s towards the left: spatial redundancy contained in $\boldsymbol{P}$ (the blacker the pixel, the more often it was observed) and spatial impulse response (the time response of the bolometer and the electronics is clearly visible as the spatial extent of the Gaussian lobe).

2.4. Conclusion

We have constructed a linear instrument model from the physical description of the phenomena involved during acquisition: scanning, optics, filters, bolometers, and electronics were taken into account, together with a description of the sky in continuous variables in the three dimensions. We next explicitly described certain calculations and approximated the model in a factorised form to lighten the numerical computational burden.

The proposed model differs from those currently used in SANEPIC (Patanchon et al. 2008) or MADmap (Cantalupo et al. 2010) in that it includes the physics of acquisition. Moreover, unlike monochromatic models (Sibthorpe et al. 2009), the sky model extends spectrally across the whole channel. Again, unlike (Sibthorpe et al. 2009), our bolometer model is linearized, which simplifies the developments and allows the bolometer time response to be made explicit.

Finally, the consistent, global definition of the acquisition allows the over-sampling to be directly exploited and a processing method to be designed that uses these properties to estimate the sky at higher resolution than the detector sampling period.

3. Data inversion for high-resolution maps

The previous section was dedicated to the instrument model and we deduced the relationship between the measured data $\boldsymbol{z}$ and the unknown sky $\mathcal{X}$ or its coefficients $\boldsymbol{x}$ through

$$\boldsymbol{z} = \mathcal{H}\mathcal{X} + \boldsymbol{o} + \boldsymbol{n} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{o} + \boldsymbol{n}. \tag{22}$$

The matrix $\boldsymbol{H}$ is relatively complex and high-dimensional, but the forward model (16) remains linear. The terms $\boldsymbol{o}$ and $\boldsymbol{n}$ account for measurement and modelling errors and quantify the data uncertainties. The term $\boldsymbol{o}$ is the noise mean (offset) and $\boldsymbol{n}$ is a zero-mean white and stationnary Gaussian noise with variance σ_n^2. We assumed that each bolometer denoted b is affected by an unknown offset o_b. Equation (16) can be rewritten for the bolometer b

$$\boldsymbol{z}_b = \boldsymbol{H}_b\boldsymbol{x} + o_b + \boldsymbol{n}_b, \tag{23}$$

where $\boldsymbol{z}_b$ contains data from bolometer b, $\boldsymbol{H}_b$ is the corresponding part of the instrument model and $(\boldsymbol{n}_b, o_b)$ accounts for errors of the bolometer b. This section presents the method to estimate the unknown $\boldsymbol{x}$ and the offsets $\boldsymbol{o}$ from the data $\boldsymbol{z}$.

We tackled the map-making question in an inverse problem framework. Abundant literature is available on the subject (Idier 2008; Demoment 1989; Tikhonov & Arsenin 1977; Twomey 1962). As presented in the previous section, the instrument model embeds convolutions and low-pass systems. The inverse problem is ill-posed (Park et al. 2003) and this is particularly true when super-resolution is intended. In this context, a naive inversion, such as a least-squares solution, would lead to an unacceptably noisy and unstable solution.

A usual class of solutions relies on regularization, i.e. the introduction of prior information on the unknown object $\boldsymbol{x}$ to compensate for the lack of information in the data. A consequence of regularization is that reconstruction methods are specific to a class of (sky) maps, according to the introduced information. From this standpoint, the present paper considers extended sources and relatively spatially regular maps.

Since it is defined as a function of continuous variables, the regularity can be measured by the squared energy[2] of derivatives of $\mathcal{X}$. For first derivatives in both directions, it can be shown (see Appendix A) that

$$\left\|\frac{\partial\mathcal{X}(\alpha,\beta)}{\partial\alpha}\right\|^2 + \left\|\frac{\partial\mathcal{X}(\alpha,\beta)}{\partial\beta}\right\|^2 = \boldsymbol{x}^t\left(\boldsymbol{D}_\alpha + \boldsymbol{D}_\beta\right)\boldsymbol{x}, \tag{24}$$

where $\boldsymbol{D} = \boldsymbol{D}_\alpha + \boldsymbol{D}_\beta$ is obtained from the sum of the autocorrelation of the derivative of ψ with respect to α and β and is similar to a discrete gradient operator. This relation illustrates the equivalence between the measure on the continuous function $\mathcal{X}$ and the measure on coefficient $\boldsymbol{x}$, thanks to the use of a Gaussian generating function.

With the regularity measure (24) and the white Gaussian model hypothesis for $\boldsymbol{n}$, the regularized least-squares criterion is

$$J_\mathcal{X}(\mathcal{X}, \boldsymbol{o}) = \|\boldsymbol{z} - \mathcal{H}\mathcal{X} - \boldsymbol{o}\|^2 + \mu\left(\left\|\frac{\partial\mathcal{X}(\alpha,\beta)}{\partial\alpha}\right\|^2 + \left\|\frac{\partial\mathcal{X}(\alpha,\beta)}{\partial\beta}\right\|^2\right)\cdot \tag{25}$$

Another consequence of ill-posedness and regularization is the need to tune the compromise between different sources of information. The hyperparameter μ tunes this trade-off. With Eqs. (22) and (24), we obtain a regularized least-squares criterion that depends only on the coefficients

$$J_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{o}) = \|\boldsymbol{z} - \boldsymbol{H}\boldsymbol{x} - \boldsymbol{o}\|^2 + \mu\boldsymbol{x}^t\boldsymbol{D}\boldsymbol{x}, \tag{26}$$

the desired map is defined as the minimizer

$$\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{o}} = \arg\min_{\boldsymbol{x},\boldsymbol{o}} J_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{o}).$$

---

[2] As an alternative, a non-quadratic norm of the derivative, e.g. convex penalty, could also be used. Its interest is less penalization of high gradients in the map. Unfortunately, the measure on coefficients is no more explicit.

As a consequence $\widehat{X}(\alpha,\beta) = \sum_{ij} \widehat{x_{ij}}\, \psi(\alpha - i\delta_\alpha, \beta - j\delta_\beta)$, is the optimum of the criterion Eq. (25).

Remark 1. A Bayesian interpretation of criterion (26) is a Gaussian *posterior* law with Gaussian iid likelihood, Gaussian correlated prior and flat prior law for $\boldsymbol{o}$. An advantage of the Bayesian interpretation is the ability to derive an uncertainty around the maximum through the variance (see Sect. 4) of the *posterior* law. Another important advantage of the Bayesian interpretation deals with the estimation of hyperparameter and instrument parameters (Orieux et al. 2010).

The proposed algorithm for the computation of $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{o}}$ is an alternating minimization algorithm: after an initialization, the following two steps are iterated

1. find $\widehat{\boldsymbol{x}}$ for fixed $\boldsymbol{o}$

$$\widehat{\boldsymbol{x}}^{k+1} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{z} - \boldsymbol{H}\boldsymbol{x} - \widehat{\boldsymbol{o}}^k\|^2 + \mu \boldsymbol{x}^t \boldsymbol{D}\boldsymbol{x}; \tag{27}$$

2. find $\widehat{\boldsymbol{o}}$ for fixed $\boldsymbol{x}$

$$\widehat{\boldsymbol{o}}^{k+1} = \arg\min_{\boldsymbol{o}} \|\boldsymbol{z} - \boldsymbol{H}\widehat{\boldsymbol{x}}^{k+1} - \boldsymbol{o}\|^2 \tag{28}$$

until a criterion is met. For fixed $\boldsymbol{x}$, the solution is straightforward and $\widehat{o_b}$ is the empirical mean of the residual $\boldsymbol{z}_b - \boldsymbol{H}_b\boldsymbol{x}$ for each bolometer separately. For fixed $\boldsymbol{o}$, the solution Eq. (27) is unique and explicit

$$\widehat{\boldsymbol{x}} = \left(\boldsymbol{H}^t\boldsymbol{H} + \mu\boldsymbol{D}\right)^{-1} \boldsymbol{H}^t(\boldsymbol{z} - \boldsymbol{o}). \tag{29}$$

The estimator is linear w.r.t. data $\boldsymbol{z}$. Unfortunately, since $\boldsymbol{H}$ is not circulant, $\widehat{\boldsymbol{x}}$ cannot be computed with a "brute force" algorithm: the practical inversion of the Hessian matrix $\boldsymbol{H}^t\boldsymbol{H} + \mu\boldsymbol{D}$ is impossible (the size of this matrix is the square of the number of coefficients $\boldsymbol{x}$). The proposed solution relies on an iterative conjugate gradient descent algorithm (Nocedal & Wright 2000; Shewchuk 1994). The most expensive part is the computation of the product between the matrix $\boldsymbol{H}^t\boldsymbol{H}$ and the current point $\boldsymbol{x}_k$ and it can be efficiently computed based on FFT, decimation, and zero-padding (see Appendix C).

4. Experimental results

This part illustrates the improvement that our approach can bring w.r.t. to the standard approach based on coaddition first using simulated data and then with actual data transmitted by *Herschel*.

4.1. Simulated data

4.1.1. Experimental protocol

We chose three $20' \times 20'$ maps used by the SPIRE consortium to assess reconstruction methods (Clements et al. 2006): a map of galactic cirrus (Fig. 3) complying with the a priori regularity model, a map of galactic cirrus superimposed on point sources (Fig. 6), and a galaxy map (Fig. 7).

We studied the PMW channel and the *"Large Map"* protocol with three scans in each direction and a velocity of $30''$/s. The data were generated using a simulated map of coefficients $\boldsymbol{x}$ and (Clements et al. 2006) the instrument model (16), considering for this simulation part that the bolometers are not affected by any offset. We used a flat spectrum ($\varrho = 0$ in Eq. (9)) for the simulations and the inversions. The noise is zero-mean white Gaussian with three levels characterized by their standard deviation σ_n ("standard noise" hereafter), $10\,\sigma_n$ ("high noise") and $0.1\,\sigma_n$ ("low noise"). The standard deviation is the same for all bolometers and, unless stated otherwise, all data sets were generated with the same noise realization.

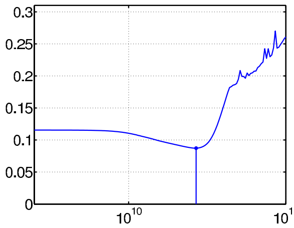


Fig. 2. Reconstruction error $\mathcal{E}$ vs. regularization parameter μ for of cirrus with "standard noise". The minimum error is $\mathcal{E}_{\min} = 0.08$ for the proposed method, while $\mathcal{E} = 0.12$ for $\mu = 0$.

The proposed reconstruction for the $20' \times 20'$ maps performed using $\delta_\alpha = \delta_\beta = 2''$, i.e. maps of 600×600 coefficients. We compare our results with the map obtained by coaddition, with $6''$ as pixel size.

In agreement with Sect. 3, the map was reconstructed as the minimizer of criterion (25), (26) and the minimization was performed by a conjugate gradient algorithm with optimal step size. The value of the criterion decreases at each iteration and a few tens of iterations appear to be sufficient to reach the minimum.

In the simulated cases, the original map (the "sky truth") is known, accordingly, we can quantitatively assess the reconstruction through an error measure defined by

$$\mathcal{E} = \sum_{i,j} |x^*_{ij} - \widehat{x_{ij}}| \;/\; \sum_{i,j} |x^*_{ij}|, \tag{30}$$

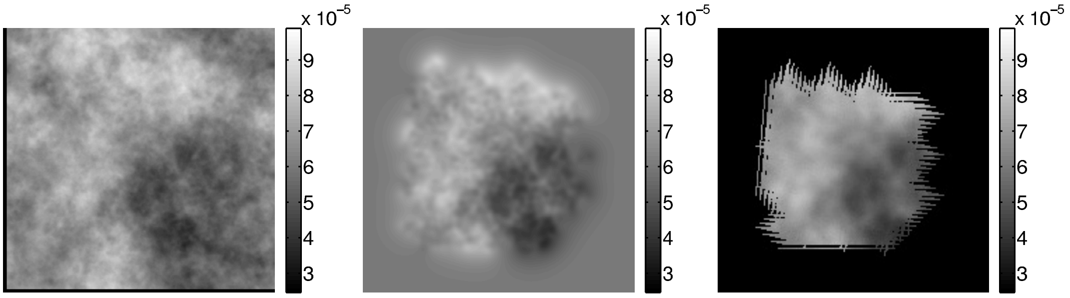
where $\boldsymbol{x}^*$ and $\widehat{\boldsymbol{x}}$ are the coefficients of the true map and the reconstructed map.

The estimate $\widehat{\boldsymbol{x}}$ depends on the regularization parameter, as illustrated in Fig. 2. A non-zero optimum value μ_{opt} appears (here $\sim 10^{12}$) for which $\mathcal{E}$ is a minimum, which confirm the interest of the regularization. A value lower than 10^{11} leads to an under-regularized map and a value greater than 10^{13} to an over-regularized one. In the following, it is, of course, the optimal value that is used to reconstruct the maps. Also, it appears empirically that μ needs to vary by a factor 2 around μ_{opt} to obtain a noticeable modification of the map. This result is confirmed in Fig. 2, where the minimum is not very marked compared to the horizontal scale.

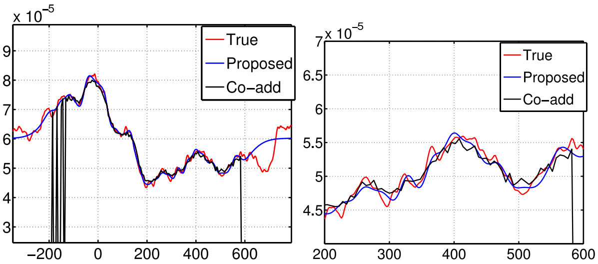
Figure 2 also illustrates the improvement provided by the regularization: the errors for the non-regularized and optimum-regularized maps are 0.12 and 0.08 respectively.

4.1.2. Restoration of galactic cirrus

Figure 3 summarises the results concerning the cirrus in the "standard noise" case. The proposed map is very close to the true one. Our method restores details of small spatial scales (with spectral extension from low to high frequency) that are invisible on the coaddition but present on the true map (see the profiles in Figs. 3d and e, especially the fluctuations around pixels 250 and 350). In addition, our method also correctly restores large-scale structures, which correspond to low-frequencies down to the null frequency (mean level of the map). We conclude that our method properly estimates the photometry.



(a) True map (b) Proposed map (c) Coaddition



(d) one horizontal profile (e) one horizontal profile (zoom)

Fig. 3. Comparison of results. **a)** Shows the true map; **b)** presents the proposed map and **c)** the coaddition. A horizontal profile is shown in **d)** and **e)** gives a zoom.

Remark 2. Moreover, the reconstruction method is linear with respect to the data (see Sect. 2), which means that the use of arbitrary units is valid.

To quantify the gain in correctly restored bandwidth, we considered the power spectra of the maps (Fig. 4) for the true sky, the sky convolved with the PSF h_o (see Sect. 2.1.2), the coaddition, and the proposed sky. As mentioned in Sect. 2.1.2, the sampling frequency of the detector is $f_s \approx 0.02$ arcsec^{-1}. Consequently the acquired data during one integration cannot correctly represent frequencies above $f_s/2 \approx 0.01$ arcsec^{-1}. We have also seen in Sect. 2.1.2 that the FWHM of the PSF is 25.2$''$ at 350 μm, i.e. a cutoff frequency of the optical transfer function of $\approx$0.04 arcsec^{-1}. The attenuation effect of the convolution by the PSF on the true map is visible the power spectra of the convolved and coaddition maps for all frequencies above $\approx$0.008 arcsec^{-1} (Fig. 4).

The power spectrum of the proposed map perfectly follows the power spectrum of the true map, from the null frequency up to a limit frequency that depends on the noise level. In the "standard noise" case (Fig. 4a) this limit is 0.025 arcsec^{-1}, that is to say, almost three times the limit frequency of each integration ($f_s/2 \approx 0.01$ arcsec^{-1}). It illustrates that our method also takes full advantage of the high-frequency temporal sampling. In any case and compared to the coaddition, we have multiplied the spectral bandwidth by a factor $\approx$4 (starting from the null frequency) where frequencies attenuated by the optical transfer function are accurately inverted.

Our method also yields the uncertainties through the Bayesian interpretation (see remark 1), from the standard deviation $\widehat{\sigma}$ of the a posteriori law (Figs. 5a and b). The uncertainties increase as we move away from the centre of the map because the data contain less information. Moreover, we see in Fig. 5c that the true map is inside a $\pm 3\widehat{\sigma}$ interval around the estimated map. In the Fourier space (Fig. 5d), up to the 0.03 arcsec^{-1}, the true power spectrum is inside a $\pm 3\widehat{\sigma}$ interval around the estimated power spectrum.

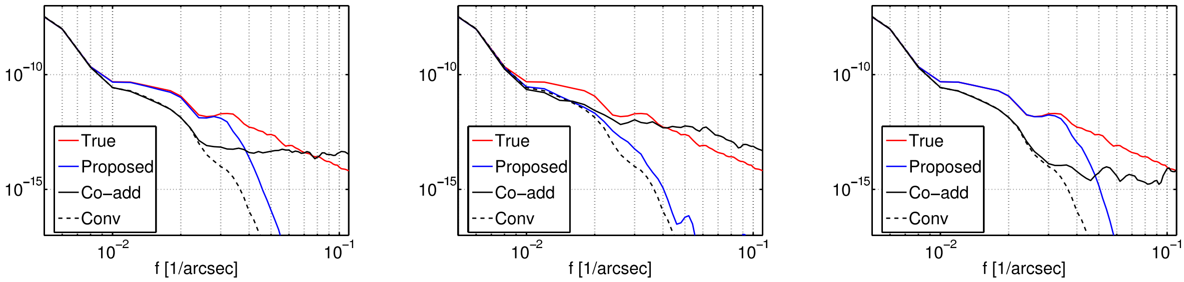
The possibilities of restoring frequencies obviously depend on the noise levels, as illustrated in the spectra shown in Fig. 4. When the noise level is lower, it is possible to restore slightly higher frequencies: up to 0.03 arcsec^{-1} for "low noise", compared to 0.025 arcsec^{-1} for "standard noise". Conversely, in the case of "high noise", our method no longer restores the frequencies attenuated by the optical transfer function Fig. 4b. The deconvolution effect is reduced and the essential effect is one of denoising. Nevertheless, the proposed method gives better (or equivalent) results than coaddition in all cases.

4.1.3. Other types of sky

Our method is based on spatial regularity information but to assess its robustness as it is, we tested it with two other types of sky in which the spatial regularity is less pronounced: galactic cirrus superimposed on point sources, and a galaxy image (Figs. 6 and 7).

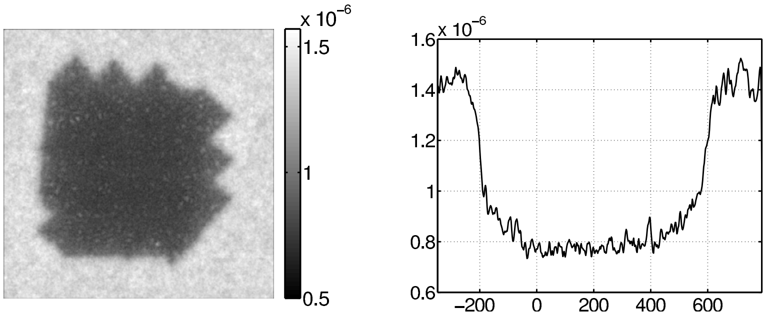
The coaddition map (Fig. 6c) is smoother than the proposed one (Fig. 6b), and several point sources are visible on the proposed map but not on the coaddition one. The amplitude of point sources is underestimated but markedly less so by the proposed method than by coaddition (Figs. 6d and e). Rebounds also appear around the point sources, a feature characteristic of linear deconvolution (resulting from a quadratic criterion).

The galaxy does not contain any point source but has spatial structures that are more complex than the galactic cirrus. These structures are considerably better restored by our method than by coaddition (Fig. 7) and it is particularly clear around pixels 250 and 300.

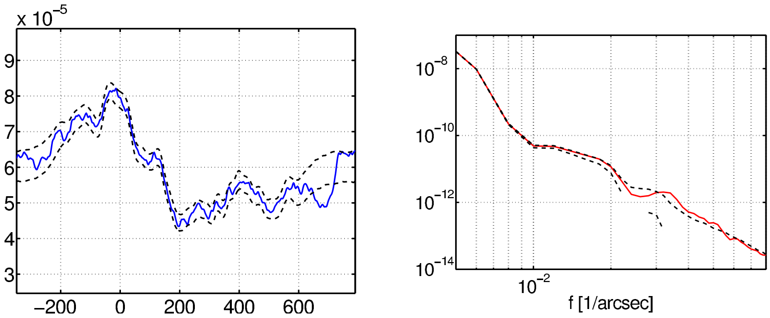


(a) Standard noise (b) High noise (c) Low noise

Fig. 4. Circular means of power spectra for the three levels of noise (standard deviations: σ_n, $10\,\sigma_n$ and $0.1\,\sigma_n$). The parameter μ is chosen to be optimal each time from the point of view of the error $\mathcal{E}$ (Eq. (30)).



(a) σ map (b) σ map profile



(c) Map profile $\pm 3\widehat{\sigma}$ (d) Power spectra $\pm 3\widehat{\sigma}$

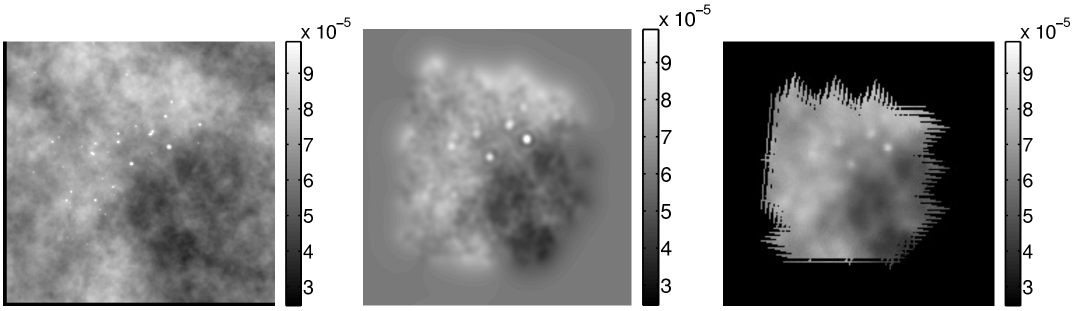
Fig. 5. Uncertainty provided by the a posteriori standard deviation $\widehat{\sigma}$. **a)** Shows the map of the standard deviation for each pixel and **b)** gives a profile. **c)** Shows a profile of the *true map* as a solid line and the two dashed lines give a $\pm 3\widehat{\sigma}$ interval around the *estimated map*. **d)** Shows the power spectrum of the *true map* as a solid red line and the two dashed lines give a $\pm 3\widehat{\sigma}$ interval around the *estimated power spectrum* in the "standard noise" case.

In conclusion, the proposed method is flexible and shows a good restoration capacity for various types of maps. In particular, it possesses a certain robustness compared to an input sky presenting characteristics that are poorly taken into account by the a priori model based on regularity information. It provides a sky that is closer to the real one than that obtained by coaddition, even in the least favourable cases.

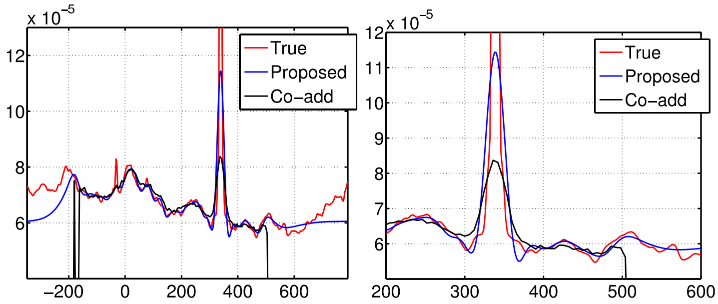
4.2. Processing real data

We conducted tests with real data of the reflection nebula NGC 7023 and of the Polaris flare (which is a high Galactic latitude cirrus cloud) performed during the science demonstration phase of *Herschel* and already presented in (Abergel et al. 2010) and (Miville-Deschênes et al. 2010), respectively. In order to run our algorithm, we took the level-1 files processed using HIPE. The true sky is not known, of course, so the value of the regularization parameter was fixed for each of the spectral channels by a trade-off between the gain in spectral resolution and the amplification of the noise.

Figures 8 to 12 illustrate the results for NGC 7023 and the Polaris flare. The gain in spatial resolution is spectacular in the three channels. It is interesting to note that the map of NGC 7023

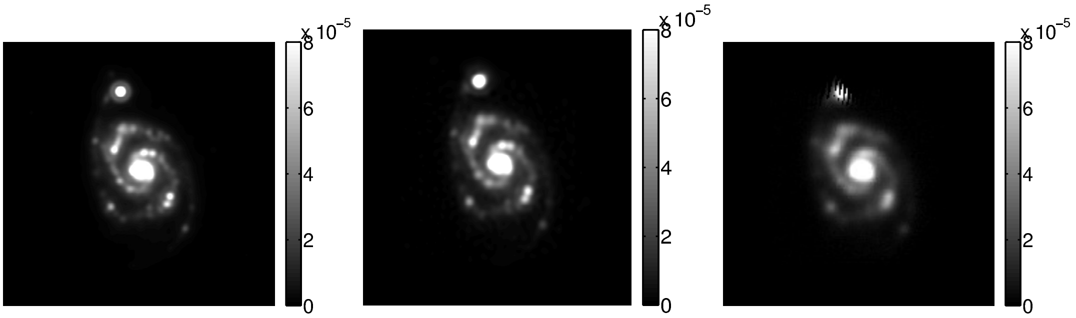


(a) True map (b) Proposed (c) Coaddition

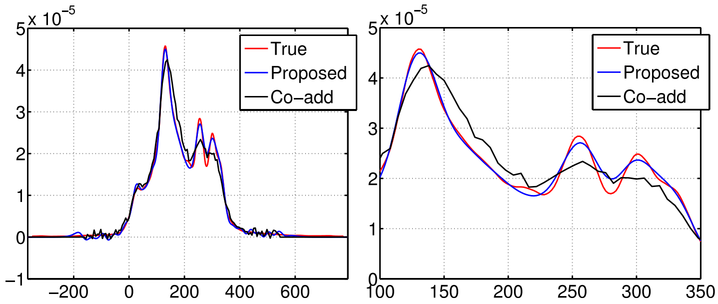


(d) one horizontal profile (e) one horizontal profile (zoom)

Fig. 6. Restoration of cirrus superimposed on point sources.



(a) True map (b) Proposed (c) Coaddition



(d) one horizontal profile (e) one horizontal profile (zoom)

Fig. 7. Restoration of galaxy.

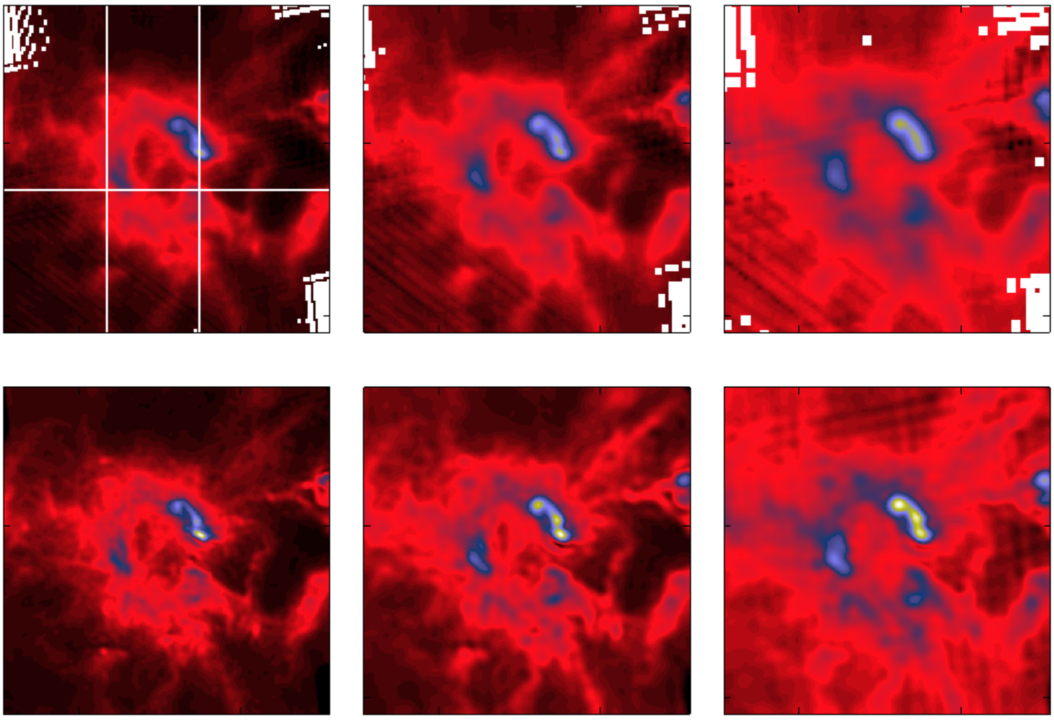


Fig. 8. Central part (23′× 23′) of NGC 7023 in the three channels PSW, PMW and PLW (left, middle and right, respectively). *Top panels*: coadded maps; *bottom panels*: proposed maps.

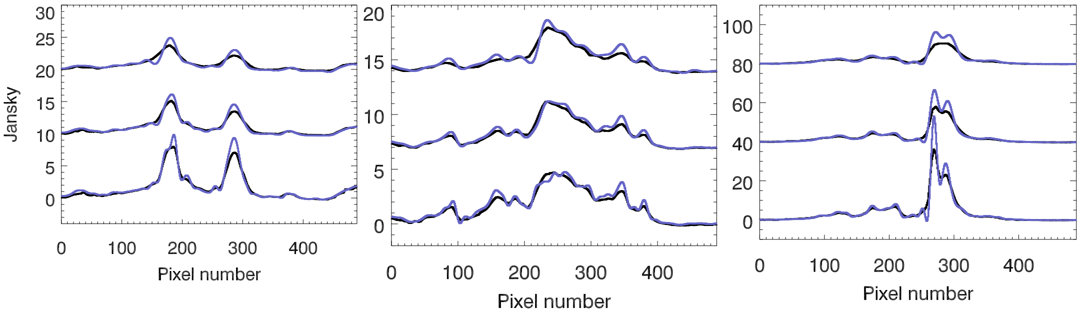


Fig. 9. Profiles along the three sections shown in the top left panel of Fig. 8. Each panel shows the profiles within the PSW, PMW and PLW channels, offset for clarity from bottom to top, respectively. *Left panel*: horizontal profile; *central and right panels*: vertical profiles. Black: coadded maps, blue: proposed maps.

obtained by our method in the longest wavelength channel (500 μm, PLW channel) shows spatial structures that are not visible in the coaddition but are real since they are visible at shorter wavelengths (250 μm, PSW channel), as illustrated for instance in the right panel of Fig. 9. The same panel also shows that negative rebounds appear on the sharpest side of the brightest filament of NGC 7023. This filament is the narrowest structure of the map and its width is comparable to the width of the PSF. Similar rebounds were also seen in our simulations with point sources (Fig. 6). The Polaris flare does not contain comparable bright and narrow filament, so the proposed map does not present this kind of artefact. The zoom on a 10′ × 10′ square (Fig. 11) illustrates the gain in angular resolution for faint structures.

Figure 12 shows that the power spectra of the proposed maps of the Polaris flare in the three channels follow the expected power law that is typical of the infrared emission of high Galactic cirrus $P(k) \propto k^{\gamma}$ with $\gamma = -2.7$ (e.g., Miville-Deschênes et al. 2010) on a frequency range from 10^{-3} to 3×10^{-2} arcsec^{-1}. For the simulated data of our Sect. 4.1, the attenuation effect of the convolution by the PSF is accurately inverted up to the frequency where the noise is dominant. Thanks to this correction, the contrast of the small-scale structures is enhanced (Figs. 10

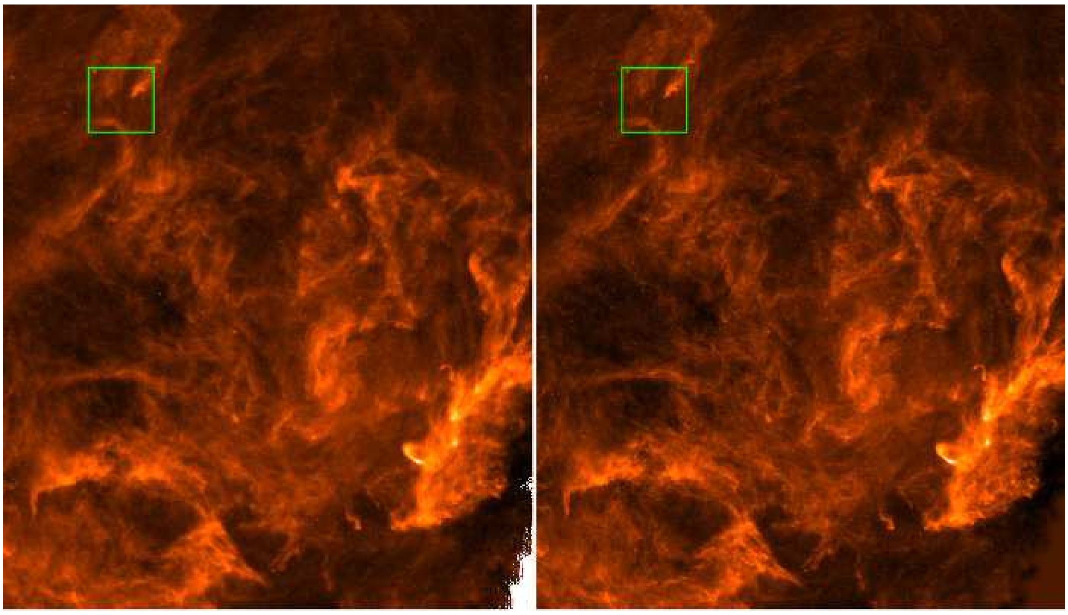


Fig. 10. 85′ × 110′ rectangle in the Polaris flare for the PSW channel (the total field observed during the science demonstration phase of *Herschel* is presented in Miville-Deschênes et al. 2010). *Left panel*: coadded map. *Right panel*: proposed result.

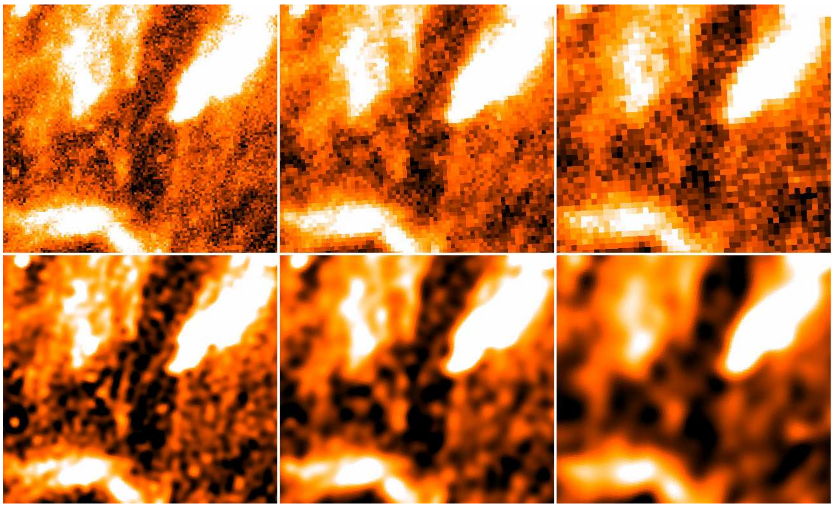


Fig. 11. Zoom on the 10′ × 10′ green square seen in Fig. 10. *Top panels* and from left to right: coadded maps in the PSW, PMW and PLW channels, respectively; *bottom panels*: proposed maps in the three channels.

and 12) w.r.t. the coaddition, since the energy of each structure is put in a shorter number of pixels than for the coaddition. At smaller frequencies, Miville-Deschênes et al. (2010) have shown that the SPIRE spectra are attenuated compared to IRAS, which is likely owing to the correction of $1/f$ noise attributed to thermal drifts in the preprocessing of the data.

5. Conclusion

We have proposed a new method for super-resolved image reconstruction for scanning instruments and its application to the SPIRE instrument of the *Herschel* observatory.

The first key element is an instrument model that describes the physical processes involved in the acquisition. To explain the data in a reliable way, our model combines the descriptions of three elements: (i) the sky as a function of continuous variables in the three dimensions (two spatial and one spectral), (ii) the optics and the rest of the instrumentation (bolometer, electronics, etc.) and (iii) the scanning strategy. We thus arrived at a linear model in integral form (Eq. (7)). We then wrote it in a matrix form (Eq. (16)) by making certain calculations explicit. Next, by coming close to the pointed directions (on a fine grid), we decomposed it into a convolution followed by inhomogeneous down-sampling (Eq. (21)). This model provides a faithful link

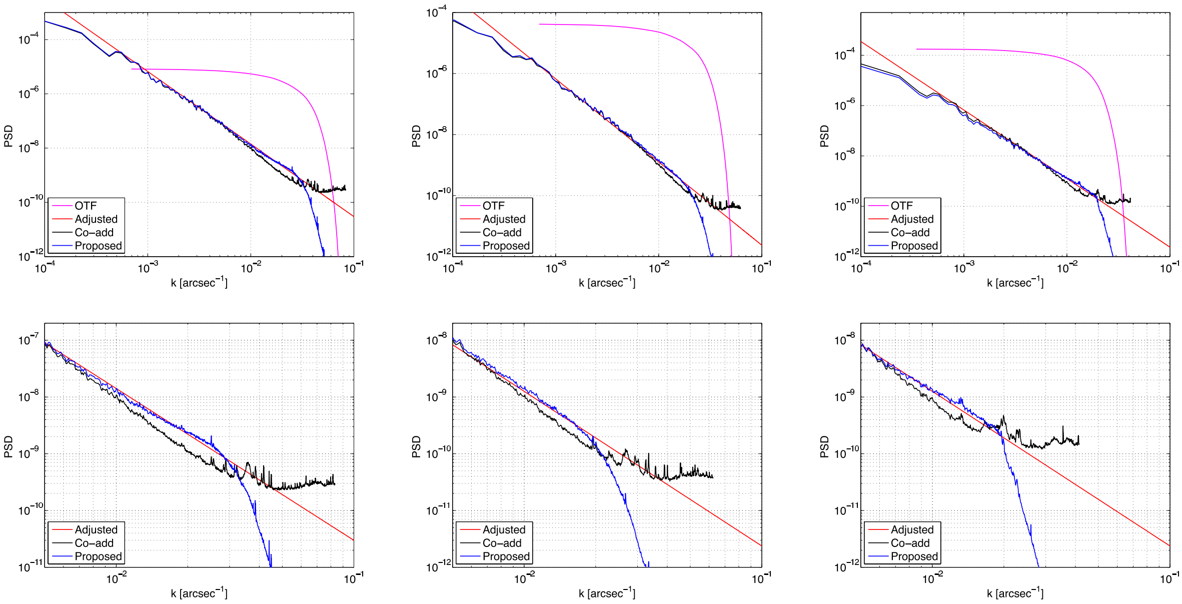


Fig. 12. Circular means of the power spectrum of the Polaris flare in the PSW (*left panels*), PMW (*middle panels*) and PSW (*right panels*) channels. The *bottom panels* present plots on the frequency range from 5×10^{-3} to 10^{-1} arcsec^{-1}. The red lines show the power law $P(k) \propto k^{\gamma}$ adjusted in a frequency range from 10^{-3} arcsec^{-1} to 3×10^{-2} arcsec^{-1}, with $\gamma = -2.7$. The pink solid lines show the optical transfer functions (OTF) for each band.

between the data, the sky actually observed, and the instrument effects.

On the sole basis of this instrument model and the data, the inversion is an ill-posed problem, especially if resolution enhancement is desired. The lack of information brought by the data, considering the limitations of the instrument, leads to instability of the inversion, which is all the more noticeable when the target resolution is high. This difficulty is overcome by a standard regularization method that constitutes the second key element. The method relies on spatial regularity information introduced by quadratic penalisation and on a quadratic data attachment term, the trade-off being governed by a regularization parameter. Thus, the inversion is based on a relatively standard linear approach and its implementation uses standard numerical optimization tools (conjugate gradient with optimal step).

The presented results for the SPIRE instrument illustrate, for simulated and real data, the potential of our method. Through the use of the accurate instrument model and a priori regularity information, we restored spatial frequencies over a bandwidth ~4 times that obtained with coaddition. In all channels, the attenuation by the optical transfer function is accurately inverted up to the frequency where the noise is dominant. The photometry is also properly restored.

A future work will focus on the question of hyperparameter and instrument parameter estimation, that is to say, unsupervised and myopic problems. We have a work in progess about this problem and it is developed in a Bayesian framework and resorts to an Markov Chain Monte-Carlo algorithm. Moreover, an estimation of the correlation matrix parameters (cutoff frequency, attenuation coefficients, spectral profile, etc.) could be achieved for the object or the noise (typically for the $1/f$ noise).

From another perspective, quadratic prior is known for possible excessive sharp edge penalisation in the restored object. The use of convex $L_2 - L_1$ penalisation (Künsch 1994; Charbonnier et al. 1997; Mugnier et al. 2004; Thiébaut 2008) can overcome this limitation, if needed. Moreover, the proposed method can be specialized to deal with construction/separation of two superimposed components: (i) an extended component together with (ii) a set of point sources (Giovannelli & Coulais 2005).

Finally, another relevant contribution could rely on the introduction of the spectral dependence between the different channels in the data inversion. The conjunction with a PACS direct model and the joint inversion of SPIRE and PACS data would greatly improve the map reconstruction.

Acknowledgements. The authors would like to thank M. Griffin (– Cardiff University), B. Sibthorpe (– Royal Observatory Edinburgh), G. Le Besnerais (–) and F. Champagnat (–), for fruitful discussions, and and the consortium for funding. The authors also thank K. Dassas for her help in the preprocessing of SPIRE data.

Appendix A: Energy spectral density

This appendix gives the details of the calculations concerning the regularity measure used in Sect. 3 and its frequency interpretation. Based on Eq. (10), the energy of the first derivative can be written

$$\begin{aligned}\left\|\frac{\partial X}{\partial \alpha}\right\|^2 &= \iint_{\mathbb{R}^2} \left(\frac{\partial X}{\partial \alpha}\right)^2 \mathrm{d}\alpha\, \mathrm{d}\beta \\ &= \sum_{i j i' j'} x_{ij}\, x_{i' j'} \iint_{\mathbb{R}^2} \left(\frac{\partial}{\partial \alpha}\psi_{i' j'}\right)\left(\frac{\partial}{\partial \alpha}\psi_{ij}\right) \mathrm{d}\alpha\, \mathrm{d}\beta.\end{aligned}$$

By noting the derivative $\psi'_\alpha = \partial\psi/\partial\alpha$, we obtain the autocorrelation $\Psi_\alpha = \psi'_\alpha \star \psi'_\alpha$ of the first derivative of the generating function and we have

$$\begin{aligned}\left\|\frac{\partial\mathcal{X}}{\partial\alpha}\right\|^2 &= \sum_{i j i' j'} x_{i j}\, x_{i' j'} \iint_{\mathbb{R}^2} \psi'_\alpha\Big(\alpha - i'\delta_\alpha, \beta - j'\delta_\beta\Big) \\ &\qquad \psi'_\alpha\Big(\alpha - i\delta_\alpha, \beta - j\delta_\beta\Big)\,\mathrm{d}\alpha\,\mathrm{d}\beta \\ &= \sum_{i j i' j'} x_{i j}\, x_{i' j'}\, [\psi'_\alpha \star \psi'_\alpha]\left\{(i' - i)\delta_\alpha, (j' - j)\delta_\beta\right\} \\ &= \sum_{i j i' j'} x_{i j}\, x_{i' j'} \Psi_\alpha \left\{(i' - i)\delta_\alpha, (j' - j)\delta_\beta\right\}. \qquad \text{(A.1)}\end{aligned}$$

As there is a finite number of coefficients x_{ij}, the measure can be put in the form of a quadratic norm

$$\left\|\frac{\partial\mathcal{X}}{\partial\alpha}\right\|^2 = \boldsymbol{x}^t \boldsymbol{D}_\alpha \boldsymbol{x}$$

where the matrix $\boldsymbol{D}_\alpha$ is obtained from Ψ_α. Considering the invariant structure of (A.1), the matrix $\boldsymbol{D}_\alpha$ has a Tœplitz structure. The calculation is performed by discrete convolution and can be computed by FFT.

By introducing the dimension β,

$$\left\|\frac{\partial\mathcal{X}}{\partial\alpha}\right\|^2 + \left\|\frac{\partial\mathcal{X}}{\partial\beta}\right\|^2 = \boldsymbol{x}^t \boldsymbol{D}_\alpha \boldsymbol{x} + \boldsymbol{x}^t \boldsymbol{D}_\beta \boldsymbol{x}.$$

The quadratic regularity measure on the function $\mathcal{X}$ with continuous variables is expressed through a quadratic regularity measure on the coefficients $\boldsymbol{x}$.

The autocorrelation Fourier transform (FT) is the energy spectral density, i.e. the squared modulus of the FT of ψ'_α

$$\begin{aligned}\mathring{\Psi}_\alpha(f_\alpha, f_\beta) &= \iint_{\mathbb{R}^2} \Psi_\alpha(\alpha,\beta)\mathrm{e}^{-2j\pi(\alpha f_\alpha + \beta f_\beta)}\,\mathrm{d}\alpha\,\mathrm{d}\beta \\ &= \left|\iint_{\mathbb{R}^2} \psi'_\alpha(\alpha,\beta)\mathrm{e}^{-2j\pi(\alpha f_\alpha + \beta f_\beta)}\,\mathrm{d}\alpha\,\mathrm{d}\beta\right|^2 \\ &= 4\pi^2 f_\alpha^2 \left|\mathring{\psi}(f_\alpha, f_\beta)\right|^2,\end{aligned}$$

where $\mathring{\psi}$ is the FT of ψ. When the dimension β is introduced, the a priori energy spectral density for the sky naturally has circular symmetry

$$\mathring{\Psi}(f_\alpha, f_\beta) = 4\pi^2 \left(f_\alpha^2 + f_\beta^2\right) \left|\mathring{\psi}(f_\alpha, f_\beta)\right|^2. \qquad \text{(A.2)}$$

This calculation brings out the frequency structure introduced a priori for the sky according to the chosen function ψ. This is a high-pass structure since the factor $f_\alpha^2 + f_\beta^2$ tends to cancel $\mathring{\Psi}$ around zero, which is consistent with a regularity measure.

Appendix B: Explicit calculation of the model

In order to integrate over time in (13), we use the expressions of (2) for $p_\alpha(t)$ and $p_\beta(t)$, which give

$$\begin{aligned}\frac{1}{2\pi}\frac{1}{\Sigma_\alpha\Sigma_\beta} \int_t \exp&\left[-\frac{1}{2}\frac{(v_\alpha t + c_\alpha + \alpha^{ij} - \alpha_{lm})^2}{\Sigma_\alpha^2}\right] \\ &\times \exp\left[-\frac{1}{2}\frac{(v_\beta t + c_\beta + \beta^{ij} - \beta_{lm})^2}{\Sigma_\beta^2}\right] h_\mathrm{b}(nT_\mathrm{s} - t)\,\mathrm{d}t.\end{aligned}$$

With the bolometer response

$$h_\mathrm{b}(nT_\mathrm{s} - t) = \mathbb{1}_{[0\ +\infty[}(nT_\mathrm{s} - t) S \exp\left[-\frac{nT_\mathrm{s} - t}{\tau}\right],$$

we have

$$\begin{aligned}\frac{1}{2\pi}\frac{S}{\Sigma_\alpha\Sigma_\beta} \exp\left[-\frac{nT_\mathrm{s}}{\tau}\right] \int_{-\infty}^{nT_\mathrm{s}} \exp&\left[-\frac{1}{2}\frac{(v_\alpha t + o_\alpha)^2}{\Sigma_\alpha^2}\right] \\ &\times \exp\left[-\frac{1}{2}\frac{(v_\beta t + o_\beta)^2}{\Sigma_\beta^2}\right] \exp\left[\frac{t}{\tau}\right]\mathrm{d}t \qquad \text{(B.1)}\end{aligned}$$

with $o_\alpha = c_\alpha + \alpha^{ij} - \alpha_{lm}$ and $o_\beta = c_\beta + \beta^{ij} - \beta_{lm}$. This is the integration of a truncated Gaussian since the argument of the exponential is a quadratic form w.r.t. t.

B.1. Calculation of the argument

Here, we express the quadratic form in question

$$n(t) = \tau\Sigma_\beta^2(v_\alpha t + o_\alpha)^2 + \tau\Sigma_\alpha^2(v_\beta t + o_\beta)^2 - 2\Sigma_\alpha^2\Sigma_\beta^2 t.$$

Expanding and factorizing the numerator $n(t)$ gives

$$\begin{aligned}n(t) &= \tau\Sigma_\beta^2\left(v_\alpha^2 t^2 + 2v_\alpha o_\alpha t + o_\alpha^2\right) \\ &\qquad + \tau\Sigma_\alpha^2\left(v_\beta^2 t^2 + 2v_\beta o_\beta t + o_\beta^2\right) - 2\Sigma_\alpha^2\Sigma_\beta^2 t \\ &= \left((t + a)^2 + b - a^2\right)/\Sigma^2\end{aligned}$$

with the constants $a = \Sigma^2\left(\tau\Sigma_\beta^2 v_\alpha o_\alpha + \tau\Sigma_\alpha^2 v_\beta o_\beta - \Sigma_\alpha^2\Sigma_\beta^2\right)$, $b = \tau\Sigma^2\left(\Sigma_\beta^2 o_\alpha^2 + \Sigma_\alpha^2 o_\beta^2\right)$ and $\Sigma^{-2} = \tau\left(\Sigma_\beta^2 v_\alpha^2 + \Sigma_\alpha^2 v_\beta^2\right)$. Putting this t-quadratic form into the integral, we obtain

$$\begin{aligned}\frac{1}{2\pi}S\frac{\sqrt{\pi\tau}\Sigma}{\sqrt{2}} \exp&\left[-\frac{nT_\mathrm{s}}{\tau} - \frac{1}{2}\frac{b - a^2}{\Sigma^2\Sigma_\alpha^2\Sigma_\beta^2\tau}\right] \\ &\times \left(1 + \mathrm{erf}\left(\frac{nT_\mathrm{s} + a}{\sqrt{2\tau}\Sigma\Sigma_\alpha\Sigma_\beta}\right)\right), \qquad \text{(B.2)}\end{aligned}$$

where the function erf is defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-\theta^2}\,\mathrm{d}\theta = -\mathrm{erf}(-x).$$

This expression can be simplified by using the function $\mathrm{erfcx}(x) = \exp(x^2)(1 - \mathrm{erf}(x))$.

B.2. Argument of the exponential

For the sake of notational simplcity, let us note $S = \Sigma^2\Sigma_\alpha^2\Sigma_\beta^2$. The argument of the function exp then is

$$\begin{aligned}\frac{nT_\mathrm{s}}{\tau} - \frac{b - a^2}{2\Sigma^2\Sigma_\alpha^2\Sigma_\beta^2\tau} &= -\frac{nT_\mathrm{s}}{\tau} - \frac{b - a^2}{2S\tau} \\ &\qquad + \left(\frac{n^2T_\mathrm{s}^2}{2S\tau} - \frac{n^2T_\mathrm{s}^2}{2S\tau}\right) + \left(\frac{2nT_\mathrm{s}a}{2S\tau} - \frac{2nT_\mathrm{s}a}{2S\tau}\right)\end{aligned}$$

and then

$$\frac{nT_\mathrm{s}}{\tau} - \frac{b - a^2}{2\Sigma^2\Sigma_\alpha^2\Sigma_\beta^2\tau} = -\frac{nT_\mathrm{s}}{\tau} - \frac{b + 2nT_\mathrm{s}a + n^2T_\mathrm{s}^2}{2S\tau} + \left(\frac{nT_\mathrm{s} + a}{\sqrt{2S\tau}}\right)^2.$$

So, by injecting this expression in (B.2), the function erfcx appears

$$\exp\left[-\frac{nT_\mathrm{s}}{\tau}-\frac{b-a^2}{2S\tau}\right]\left(1+\operatorname{erf}\left(\frac{nT_\mathrm{s}+a}{\sqrt{2S\tau}}\right)\right)=$$
$$\exp\left[-\frac{nT_\mathrm{s}}{\tau}-\frac{b+2nT_\mathrm{s}a+n^2T_\mathrm{s}^2}{2S\tau}\right]\operatorname{erfcx}\left(-\frac{nT_\mathrm{s}+a}{\sqrt{2S\tau}}\right).$$

The values of S, a and b can be replaced. First of all, the argument of the exponential is

$$\begin{aligned}
&-\frac{nT_\mathrm{s}}{\tau}-\frac{\tau\Sigma^2\left(\Sigma_\beta^2 o_\alpha^2+\Sigma_\alpha^2 o_\beta^2\right)}{2\Sigma^2\Sigma_\alpha^2\Sigma_\beta^2\tau}\\
&\quad-\frac{2nT_\mathrm{s}\Sigma^2\left(\tau\Sigma_\beta^2 v_\alpha o_\alpha+\tau\Sigma_\alpha^2 v_\beta o_\beta-\Sigma_\alpha^2\Sigma_\beta^2\right)}{2\Sigma^2\Sigma_\alpha^2\Sigma_\beta^2\tau}-\frac{n^2T_\mathrm{s}^2}{2S\tau}=\\
&\quad-\frac{nT_\mathrm{s}}{\tau}-\frac{o_\alpha^2}{2\Sigma_\alpha^2}-\frac{o_\beta^2}{2\Sigma_\beta^2}-\frac{2nT_\mathrm{s}v_\alpha o_\alpha}{2\Sigma_\alpha^2}-\frac{2nT_\mathrm{s}v_\beta o_\beta}{2\Sigma_\beta^2}\\
&\qquad+\frac{nT_\mathrm{s}}{\tau}-\frac{n^2T_\mathrm{s}^2}{2\Sigma^2\Sigma_\alpha^2\Sigma_\beta^2\tau},
\end{aligned}\tag{B.3}$$

and the terms nT_s/τ simplify. We then use the expression for Σ^2

$$\frac{n^2T_\mathrm{s}^2}{2\Sigma^2\Sigma_\alpha^2\Sigma_\beta^2\tau}=\frac{n^2T_\mathrm{s}^2v_\alpha^2}{2\Sigma_\alpha^2}+\frac{n^2T_\mathrm{s}^2v_\beta^2}{2\Sigma_\beta^2},$$

to obtain two perfect squares. Finally the argument of the exponential (B.3) in (B.2) is

$$-\frac{(o_\alpha+nT_\mathrm{s}v_\alpha)^2}{2\Sigma_\alpha^2}-\frac{\left(o_\beta+nT_\mathrm{s}v_\beta\right)^2}{2\Sigma_\beta^2},\tag{B.4}$$

which is exactly the argument of a bivariate Gaussian. We again find the same standard deviations Σ_α and Σ_β. However, the response of the optics, initially (o_α, o_β), is now shifted by $(nT_\mathrm{s}v_\alpha, nT_\mathrm{s}v_\beta)$, i.e. the pointing difference between two successive time samples.

B.3. Argument of the function erfcx and final expression

Another term is needed to know the global response. It comes from the function erfcx, which corresponds to the influence of the bolometer. The argument of the function erfcx is

$$\begin{aligned}
-\frac{nT_\mathrm{s}+a}{\sqrt{2\tau}\Sigma\Sigma_\alpha\Sigma_\beta}=\frac{\Sigma_\alpha\Sigma_\beta}{\sqrt{2\tau}\Sigma_v}-&\\
\frac{\Sigma_\beta v_\alpha(o_\alpha+nT_\mathrm{s}v_\alpha)}{\sqrt{2}\Sigma_\alpha\Sigma_v}-\frac{\Sigma_\alpha v_\beta(o_\beta+nT_\mathrm{s}v_\beta)}{\sqrt{2}\Sigma_\beta\Sigma_v}&,
\end{aligned}\tag{B.5}$$

where $\Sigma_v^2=\Sigma_\beta^2v_\alpha^2+\Sigma_\alpha^2v_\beta^2$, and what is of interest here is that the same factors are found in the argument of the exponential. To know the global response, we need to bring everything together. By injecting the expressions of the arguments (B.4) and (B.5), we obtain

$$\begin{aligned}
&\frac{1}{2\pi}S\frac{\sqrt{\pi\tau}\Sigma}{\sqrt{2}}\exp\left[\frac{nT_\mathrm{s}}{\tau}-\frac{1}{2}\frac{b-a^2}{\Sigma^2\Sigma_\alpha^2\Sigma_\beta^2\tau}\right]\\
&\quad\times\left(1+\operatorname{erf}\left(\frac{nT_\mathrm{s}+a}{\sqrt{2\tau}\Sigma\Sigma_\alpha\Sigma_\beta}\right)\right)=\\
&\quad\times\frac{S}{2\sqrt{2\pi}\Sigma_v}\exp\left[-\frac{(o_\alpha+nT_\mathrm{s}v_\alpha)^2}{2\Sigma_\alpha^2}-\frac{\left(o_\beta+nT_\mathrm{s}v_\beta\right)^2}{2\Sigma_\beta^2}\right]\\
&\quad\times\operatorname{erfcx}\left(\frac{\Sigma_\alpha\Sigma_\beta}{\sqrt{2\tau}\Sigma_v}-\frac{\Sigma_\beta v_\alpha(o_\alpha+nT_\mathrm{s}v_\alpha)}{\sqrt{2}\Sigma_\alpha\Sigma_v}-\frac{\Sigma_\alpha v_\beta(o_\beta+nT_\mathrm{s}v_\beta)}{\sqrt{2}\Sigma_\beta\Sigma_v}\right)
\end{aligned}$$

with, similarly for α and β: $\Sigma_{\alpha/\beta}^2=\sigma_{\alpha/\beta}^2+\sigma_\mathrm{o}^2$, which finishes the integration of (13) over time.

Appendix C: Direct model computation algorithm

This part gives some more details on the concrete computation of a model output $\boldsymbol{Hx}$ of Sect. 2.3. First of all, there are four different impulse responses whatever the number of scans. For scans in the same direction, the response is the same. Thus we can construct four different convolution matrices $\boldsymbol{H}_i$ for $i=1,2,3,4$ and apply four different discrete convolutions to the coefficients $\boldsymbol{x}$.

We can also deduce the structure of the transpose of the model $\boldsymbol{H}^t=\boldsymbol{H}_\mathrm{c}^t\boldsymbol{P}^t$. The matrix $\boldsymbol{P}^t$ is a data summation/zero padding matrix (addition of the data that possess the same pointing while setting the other coefficients to zero), and $\boldsymbol{H}_\mathrm{c}^t$ corresponds to a convolution with the space reversal impulse responses.

The product by $\boldsymbol{P}^t$ is very similar to the construction of a naive map except that the data are added rather than averaged. Also, the operation is done by velocity and not globally. Finally, the products by $\boldsymbol{H}_\mathrm{c}$ and $\boldsymbol{H}_\mathrm{c}^t$ are convolutions computed by FFT.

References

Abergel, A., Arab, H., Compiègne, M., et al. 2010, A&A, 518, L96

Andrews, H. C., & Hunt, B. R. 1977, Digital Image Restoration (Englewood Cliffs, : Prentice-Hall)

Boulanger, F., Abergel, A., Bernard, J., et al. 1996, A&A, 312, 256

Cantalupo, C. M., Borrill, J. D., Jaffe, A. H., Kisner, T. S., & Stompor, R. 2010, ApJS, 187, 212

Champagnat, F., Le Besnerais, G., & Kulcsár, C. 2009, J. Opt. Soc. Am. A, 26, 1730

Charbonnier, P., Blanc-Féraud, L., Aubert, G., & Barlaud, M. 1997, IEEE Trans. Image Processing, 6, 298

Clements, D., Chanial, P., Bendo, G., et al. 2006, Mapmaking Algorithm Review Report, Tech. Rep., Astrophysics group at Imperial College London

de Graauw, T., Helmich, F. P., Phillips, T. G., et al. 2010, A&A, 518, L6

Demoment, G. 1989, IEEE Trans. Acoust. Speech, Signal Processing, -37, 2024

Elad, M., & Feuer, A. 1999, IEEE Trans. Image Processing, 8, 387

Farsiu, S., Robinson, M., Elad, M., & Milanfar, P. 2004, Image Processing, IEEE Transactions on, 13, 1327

Giovannelli, J.-F., & Coulais, A. 2005, A&A, 439, 401

Griffin, M. J. 2006, Revised Photometer sensitivity model, working version after sensitivity review meeting

Griffin, M. J. 2007, The SPIRE Analogue Signal Chain and Photometer Detector Data Processing Pipeline, Tech. Rep., University of Wales Cardiff

Griffin, M. J., Bock, J. J., & Gear, W. K. 2002, Appl. Opt., 41, 6543

Griffin, M., Swinyard, B., Vigroux, L., et al. 2008, in SPIE Conf., 7010

Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, A&A, 518, L3

Hardie, R. C., Barnard, K. J., & Armstrong, E. E. 1997, IEEE Trans. Image Processing, 6, 1621

Idier, J., 2008, Bayesian Approach to Inverse Problems (London: ISTE Ltd and John Wiley & Sons Inc.)
Künsch, H. R. 1994, Ann. Inst. Stat. Math., 46, 1
Leshem, A., Christou, J., Jeffs, B. D., Kuruoglu, E., & van der Veen, A. J. 2008, IEEE Journal of Selected Topics in Signal Processing, 2
Leshem, A., Kamalabadi, F., Kuruoglu, E., & van der Veen, A.-J. 2010, Signal Processing Magazine, 27
Miville-Deschênes, M., Martin, P. G., Abergel, A., et al. 2010, A&A, 518, L104
Molina, R., & Ripley, B. D. 1989, J. Appl. Stat., 16, 193
Mugnier, L., Fusco, T., & Conan, J.-M. 2004, J. Opt. Soc. Amer., 21, 1841
Nguyen, N., Milanfar, P., & Golub, G. 2001, IEEE Trans. Image Processing, 10, 573
Nocedal, J., & Wright, S. J. 2000, Numerical Optimization, Series in Operations Research (New York: Springer Verlag)
Orieux, F. 2009, Ph.D. Thesis, Université Paris-Sud 11
Orieux, F., Rodet, T., & Giovannelli, J.-F. 2009, in Proc. of IEEE International Conference on Image Processing (ICIP 2009), Cairo, Egypt
Orieux, F., Giovannelli, J.-F., & Rodet, T. 2010, J. Opt. Soc. Am. A, 27, 1593
Ott, S. 2010, in Astronomical Data Analysis Software and Systems XIX, ed. Y. Mizumoto, K.-I. Morita, & M. Ohishi, ASP Conf. Ser., 434, 139
Park, S. C., Park, M. K., & Kang, M. G. 2003, IEEE Trans. Signal Processing Mag., 21
Patanchon, G., Ade, P. A. R., Bock, J. J., et al. 2008, ApJ, 681, 708
Patti, A. J., Sezan, M. I., & Tekalp, A. M. 1997, IEEE Trans. Image Processing, 6, 1064
Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1
Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, A&A, 518, L2
Robert, C. P., & Casella, G. 2000, Monte-Carlo Statistical Methods, Springer Texts in Statistics (New York, : Springer)
Rochefort, G., Champagnat, F., Le Besnerais, G., & Giovannelli, J.-F. 2006, IEEE Trans. Image Processing, 15, 3325
Rodet, T., Orieux, F., Giovannelli, J.-F., & Abergel, A. 2008, IEEE J. Selec. Topics in Signal Proc., 2, 802
Shewchuk, J. R. 1994, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Tech. Rep., Carnegie Mellon University
Sibthorpe, B., & Griffin, M. J. 2006, Spire Photometer Simulator, Tech. Rep., University of Wales Cardiff
Sibthorpe, B., Chanial, P., & Griffin, M. J. 2009, A&A, 503, 625
Sudiwala, R. V., Griffin, M. J., & Woodcraft, A. L. 2002, International Journal of Infrared and Millimeter Waves, 23, 545
Thiébaut, E. 2008, in Astronomical Telescopes and Instrumentation, Proc. SPIE, 7013, 70131–I
Tikhonov, A., & Arsenin, V. 1977, Solutions of Ill-Posed Problems (Washington, : Winston)
Twomey, S. 1962, J. Assoc. Comp. Mach., 10, 97
Vandewalle, P., Sbaiz, L., Vandewalle, J., & Vetterli, M. 2007, IEEE Trans. Signal Processing, 55, 3687
Wieprecht, E., Schreiber, J., de Jong, J., et al. 2009, in ASP Conf. Ser., ed. D. A. Bohlender, D. Durand, & P. Dowler, 411, 531
Woods, N. A., Galatsanos, N. P., & Katsaggelos, A. K. 2006, IEEE Trans. Image Processing, 15, 201

F. Orieux, J.-F. Giovannelli et **T. Rodet**, « Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution », *Journal of the Optical Society of America (A)*, vol. 27, n°7, pp. 1593–1607, 2010.

Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution

François Orieux,[1,*] Jean-François Giovannelli,[2] and Thomas Rodet[1]

[1]*Laboratoire des Signaux et Systèmes (CNRS–SUPELEC–Univ. Paris-Sud 11), SUPELEC, Plateau de Moulon,*
3 rue Joliot-Curie, 91 192 Gif-sur-Yvette, France
[2]*Laboratoire d'Intégration du Matériau au Système (CNRS–ENSEIRB–Univ. Bordeaux 1–ENSCPB),*
351 cours de la Liberation, 33405 Talence, France
**Corresponding author: orieux@lss.supelec.fr*

Received October 21, 2009; revised March 5, 2010; accepted April 15, 2010;
posted April 30, 2010 (Doc. ID 118863); published June 9, 2010

This paper tackles the problem of image deconvolution with joint estimation of point spread function (PSF) parameters and hyperparameters. Within a Bayesian framework, the solution is inferred via a global *a posteriori* law for unknown parameters and object. The estimate is chosen as the posterior mean, numerically calculated by means of a Monte Carlo Markov chain algorithm. The estimates are efficiently computed in the Fourier domain, and the effectiveness of the method is shown on simulated examples. Results show precise estimates for PSF parameters and hyperparameters as well as precise image estimates including restoration of high frequencies and spatial details, within a global and coherent approach. © 2010 Optical Society of America

OCIS codes: 100.1830, 100.3020, 100.3190, 150.1488.

1. INTRODUCTION

Image deconvolution has been an active research field for several decades, and recent contributions can be found in papers such as [1–3]. Examples of application are medical imaging, astronomy, nondestructive testing, and, more generally, imagery problems. In these applications, degradations induced by the observation instrument limit the data resolution, while the need of precise interpretation can be of major importance. For example, this is particularly critical for long-wavelength astronomy (see e.g., [4]). In addition, the development of a high-quality instrumentation system must be completed rationally at an equivalent level of quality in the development of data processing methods. Moreover, even for poor performance systems, the restoration method can be used to bypass instrument limitations.

When the deconvolution problem is ill-posed, a possible solution relies on regularization, i.e., introduction of information in addition to the data and the acquisition model [5,6]. As a consequence of regularization, deconvolution methods are specific to the class of image in accordance with the introduced information. From this standpoint, the present paper is dedicated to relatively smooth images encountered for numerous applications in imagery [4,7,8]. The second-order consequence of ill-posedness and regularization is the need to balance the compromise between different sources of information.

In the Bayesian approach [1,9], information about unknowns is introduced by means of probabilistic models. Once these models are designed, the next step is to build the *a posteriori* law, given the measured data. The solution is then defined as a representative point of this law; the two most representative points are (1) the maximizer and (2) the mean. From a computational standpoint, the first leads to a numerical optimization problem and the latter leads to a numerical integration problem. However, the resulting estimate depends on two sets of variables in addition to the data:

1. First, the estimate naturally depends on the response of the instrument at work, namely, the point spread function (PSF). The literature is devoted to deconvolution predominantly in the case of known PSF. In contrast, the present paper is devoted to the case of unknown or poorly known PSF, and there are two main strategies to tackle its estimation from the available data set (without extra measurements):

(i) In most practical cases, the instrument can be modeled using physical operating description. It is thus possible to find the equation for the PSF, at least in a first approximation. This equation is usually driven by a relatively small number of parameters. It is a common case in optical imaging, where a Gaussian-shaped PSF is often used [10]. It is also the case in other fields: interferometry [11], magnetic resonance force microscopy [12], fluorescence microscopy [13], etc. Nevertheless, in real experiments, the parameter values are unknown or imperfectly known and need to be estimated or adjusted in addition to the image of interest: the problem is called *myopic* deconvolution.

(ii) The second strategy forbids the use of the parametric PSF deduced from the physical analysis, and the PSF then naturally appears in a non-parametric form. Practically, the non-parametric PSF is unknown or imperfectly known and needs to be estimated in addition to the image of interest: the problem is referred to as *blind* deconvolution, for example in interferometry [14–17].

From an inference point of view, the difficulty of both myopic and blind problems lies in the possible lack of information, resulting in ambiguity between image and PSF, even in the noiseless case. In order to resolve the ambiguity, information must be added [3,18], and it is crucial to make inquiries based on any available source of information. To this end, the knowledge of the parametric PSF represents a valuable means to structure the problem and possibly resolve the degeneracies. Moreover, due to the instrument design process, a nominal value as well as an uncertainty are usually available for the PSF parameters. In addition, from a practical and algorithmic standpoint, the myopic case, i.e., the case of parametric PSF, is often more difficult due to the non-linear dependence of the observation model with respect to the PSF parameters. In contrast, the blind case, i.e., the case of non-parametric PSF, yields a simpler practical and algorithmic problem since the observation model remains linear w.r.t. the unknown elements given the object. Despite the superior technical difficulty, the present paper is devoted to the myopic format since it is expected to be more efficient than the blind format from an information standpoint. Moreover, the blind case has been extensively studied, and a large number of papers are available [19–21], while the myopic case has been less investigated, though it is of major importance.

2. Second, the solution depends on the probability law parameters called hyperparameters (means, variances, parameters of correlation matrix, etc.). These parameters adjust the shape of the laws, and at the same time they tune the compromise between the information provided by the *a priori* knowledge and the information provided by the data. In real experiments, their values are unknown and need to be estimated: the problem is called *unsupervised* deconvolution.

For both families of parameters (PSF parameters and hyperparameters), two approaches are available. In the first one, the parameter values are empirically tuned or estimated in a preliminary step (with maximum likelihood [7] or calibration [22] for example); then the values are used in a second step devoted to image restoration given the parameters. In the second step, the parameters and the object are jointly estimated [2,19].

For the myopic problem, Jalobeanu *et al.* [23] address the case of a symmetric Gaussian PSF. The width parameter and the noise variance are estimated in a preliminary step by maximum likelihood. A recent paper [24] addresses the estimation of a Gaussian blur parameter, as in our experiment, with an empirical method. They found the Gaussian blur parameter by minimizing the absolute value of the second derivative (Laplacian) of the restored images.

The present paper addresses the myopic and unsupervised deconvolution problem. We propose a new method that jointly estimates the PSF parameters, the hyperparameters, and the image of interest. It is built in a coherent and global framework based on an extended *a posteriori* law for all the unknown variables. The posterior law is obtained via the Bayes rule, founded on *a priori* laws: Gaussian for image and noise, uniform for PSF parameters, and gamma or Jeffreys for hyperparameters.

Regarding the image prior law, we have paid special attention to the parameterization of the covariance matrix in order to facilitate law manipulations such as integration, conditioning, and hyperparameter estimation. The possible degeneracy of the *a posteriori* law in some limit cases is also studied.

The estimate is chosen as the mean of the posterior law and is computed using Monte Carlo simulations. To this end, Monte Carlo Markov chain (MCMC) algorithms [25] enable one to draw samples from the posterior distribution despite its complexity and especially the non-linear dependence w.r.t. the PSF parameters.

The paper is structured in the following manner. Section 2 presents the notation and describes the problem. The three following sections describe our methodology: first the Bayesian probabilistic models are detailed in Section 3, then a proper posterior law is established in Section 4; and an MCMC algorithm to compute the estimate is described in Section 5. Numerical results are presented in Section 6. Finally, Section 7 is devoted to the conclusion and perspectives.

2. NOTATION AND CONVOLUTION MODEL

Consider N-pixel real square images represented in lexicographic order by vector $\boldsymbol{x} \in \mathbb{R}^N$, with generic elements x_n. The forward model is written

$$\boldsymbol{y} = \boldsymbol{H_w x} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the vector of data, $\boldsymbol{H_w}$ a convolution matrix, $\boldsymbol{x}$ the image of interest, and $\boldsymbol{\epsilon}$ the modelization errors or the noise. Vector $\boldsymbol{w} \in \mathbb{R}^P$ stands for the PSF parameters, such as width or orientation of a Gaussian PSF.

The matrix $\boldsymbol{H_w}$ is block-circulant with circulant-block (BCCB) for computational efficiency of the convolution in the Fourier space. The diagonalization [26] of $\boldsymbol{H_w}$ is written $\boldsymbol{\Lambda_H} = \boldsymbol{F H_w F}^\dagger$, where $\boldsymbol{F}$ is the unitary Fourier matrix and † is the transpose conjugate symbol. The convolution in the Fourier space, is then

$$\mathring{\boldsymbol{y}} = \boldsymbol{\Lambda_H} \mathring{\boldsymbol{x}} + \mathring{\boldsymbol{\epsilon}}, \tag{2}$$

where $\mathring{\boldsymbol{x}} = \boldsymbol{F x}$, $\mathring{\boldsymbol{y}} = \boldsymbol{F y}$ and $\mathring{\boldsymbol{\epsilon}} = \boldsymbol{F \epsilon}$ are the 2D discrete Fourier transform (DFT-2D) of image, data, and noise, respectively.

Since $\boldsymbol{\Lambda_H}$ is diagonal, the convolution is computed with a term-wise product in the Fourier space. There is a strict equivalence between a description in the spatial domain [Eq. (1)] and in the Fourier domain [Eq. (2)]. Consequently, for coherent description and computational efficiency, all the developments are equally done in the spatial space or in the Fourier space.

For notational convenience, let us introduce the component at null frequency $\mathring{x}_0 \in \mathbb{R}$ and the vector of component at non-null frequencies $\mathring{\boldsymbol{x}}_* \in \mathbb{C}^{N-1}$ so that the whole set of components is written $\mathring{\boldsymbol{x}} = [\mathring{x}_0, \mathring{\boldsymbol{x}}_*]$.

Let us denote $\mathbb{1}$ the vector of N components equal to $1/N$, so that $\mathbb{1}^{\mathrm{t}} \boldsymbol{x}$ is the empirical mean level of the image. The Fourier components are the $\mathring{\mathbb{1}}_n$, and we have $\mathring{\mathbb{1}}_0 = 1$ and

$\mathring{\mathbb{1}}_n = 0$ for $n \neq 0$. Moreover, $\boldsymbol{\Lambda}_{\mathbb{1}} = \boldsymbol{F}\mathbb{1}\mathbb{1}^{\mathrm{t}}\boldsymbol{F}^{\dagger}$ is a diagonal matrix with only one non-null coefficient at null frequency.

3. BAYESIAN PROBABILISTIC MODEL

This section presents the prior law for each set of parameters. Regarding the image of interest, in order to account for smoothness, the law introduces high-frequency penalization through a differential operator on the pixel. A conjugate law is proposed for the hyperparameters, and a uniform law is considered for the PSF parameters.

Moreover, we have paid special attention to the image prior law parameterization. In Subsection 3.A we present several parameterizations in order to facilitate law manipulations such as integration, conditioning, and hyperparameter estimation. Moreover, the correlation matrix of the image law may become singular in some limit cases, resulting in a degenerated prior law [when $p(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{R}^N$]. Based on this parameterization, in Section 4 we study the degeneracy of the posterior in relation to the parameters of the prior law.

A. Image Prior Law

The probability law for the image is a Gaussian field with a given precision matrix $\boldsymbol{P}$ parameterized by a vector γ. The pdf reads

$$p(\boldsymbol{x}|\boldsymbol{\gamma}) = (2\pi)^{-N/2} \det[\boldsymbol{P}]^{1/2} \exp\left[-\frac{1}{2}\boldsymbol{x}^{\mathrm{t}}\boldsymbol{P}\boldsymbol{x} \right]. \tag{3}$$

For computational efficiency, the precision matrix is designed (or approximated) in a toroidal manner, and it is diagonal in the Fourier domain $\boldsymbol{\Lambda_P} = \boldsymbol{FPF}^{\dagger}$. Thus, the law for $\boldsymbol{x}$ also is written

$$p(\boldsymbol{x}|\boldsymbol{\gamma}) = (2\pi)^{-N/2} \det[\boldsymbol{F}]\det[\boldsymbol{\Lambda_P}]^{1/2} \det[\boldsymbol{F}^{\dagger}] \times \exp\left[-\frac{1}{2}\boldsymbol{x}^{\mathrm{t}}\boldsymbol{F}^{\dagger}\boldsymbol{\Lambda_P}\boldsymbol{F}\boldsymbol{x} \right], \tag{4}$$

$$= (2\pi)^{-N/2} \det[\boldsymbol{\Lambda_P}]^{1/2} \exp\left[-\frac{1}{2}\mathring{\boldsymbol{x}}^{\dagger}\boldsymbol{\Lambda_P}\mathring{\boldsymbol{x}} \right], \tag{5}$$

and it is sometimes referred to [27] as a Whittle approximation (see also [28], p. 133) for the Gaussian law. The filter obtained for fixed hyperparameters is also the Wiener–Hunt filter [29], as described in Subsection 5.A, below.

This paper focuses on smooth images and thus on positive correlation between pixels. It is introduced by high-frequency penalty using any circulant differential operator: pth differences between pixels, Laplacian and Sobel, among others. The differential operator is denoted by $\boldsymbol{D}$ and its diagonalized form by $\boldsymbol{\Lambda_D} = \boldsymbol{FDF}^{\dagger}$. Then, the precision matrix is written $\boldsymbol{P} = \gamma_1 \boldsymbol{D}^{\mathrm{t}}\boldsymbol{D}$, and its Fourier counterpart is written

$$\boldsymbol{\Lambda_P} = \gamma_1 \boldsymbol{\Lambda}_{\boldsymbol{D}}^{\dagger}\boldsymbol{\Lambda_D} = \mathrm{diag}(0, \gamma_1|\mathring{d}_1|^2, \ldots, \gamma_1|\mathring{d}_{N-1}|^2), \tag{6}$$

where γ_1 is a positive scale factor, diag builds a diagonal matrix from elementary components, and $\mathring{d}_n$ is the nth DFT-2D coefficient of $\boldsymbol{D}$.

Under this parameterization of $\boldsymbol{P}$, the first eigenvalue is equal to zero, corresponding to the absence of penalty for the null frequency $\mathring{x}_0$, i.e., no information is accounted for about the empirical mean level of the image. As a consequence, the determinant vanishes, $\det[\boldsymbol{P}] = 0$, resulting in a degenerated prior. To manage this difficulty, several approaches have been proposed.

Some authors [2,30] still use this prior despite its degeneracy, and this approach can be analyzed in two ways:

1. On the one hand, this prior can be seen as a non-degenerated law for $\mathring{\boldsymbol{x}}_*$, the set of non-null frequency components only. In this format, the prior does not affect any probability for the null frequency component, and the Bayes rule does not apply to this component. Thus, this strategy yields an incomplete posterior law, since the null frequency is not embedded in the methodology.
2. On the other hand, this prior can be seen as a degenerated prior for the whole set of frequencies. The application of the Bayes rule is then somewhat confusing due to degeneracy. In this format, the posterior law cannot be guaranteed to remain non-degenerated.

In any case, neither of the two standpoints yields a posterior law that is both non-degenerated and addresses the whole set of frequencies.

An alternative parameterization relies on the energy of $\boldsymbol{x}$. An extra term $\gamma_0\boldsymbol{I}$, tuned by $\gamma_0 > 0$, in the precision matrix [31], introduces information for all the frequencies including $\mathring{x}_0$. The precision matrix is written

$$\boldsymbol{\Lambda_P} = \gamma_0\boldsymbol{I} + \gamma_1\boldsymbol{\Lambda}_{\boldsymbol{D}}^{\dagger}\boldsymbol{\Lambda_D} = \mathrm{diag}(\gamma_0, \gamma_0 + \gamma_1|\mathring{d}_1|^2, \ldots, \gamma_0 + \gamma_1|\mathring{d}_{N-1}|^2) \tag{7}$$

with a determinant

$$\det[\Lambda_{\boldsymbol{P}}] = \prod_{n=0}^{N-1} (\gamma_0 + \gamma_1|\mathring{d}_n|^2). \tag{8}$$

The obtained Gaussian prior is not degenerated and undoubtedly leads to a proper posterior. Nevertheless, the determinant Eq. (8) is not separable in γ_0 and γ_1. Consequently, the conditional posterior for these parameters is not a classical law, and future development will be more difficult. Moreover, the non-null frequencies $\mathring{\boldsymbol{x}}_*$ are controlled by two parameters, γ_0 and γ_1:

$$p(\mathring{\boldsymbol{x}}|\gamma_0,\gamma_1) = p(\mathring{x}_0|\gamma_0)p(\mathring{\boldsymbol{x}}_*|\gamma_0,\gamma_1). \tag{9}$$

The proposed approach to manage the degeneracy relies on the addition of a term for the null frequency only, $\boldsymbol{\Lambda}_{\mathbb{1}} = \mathrm{diag}(1, 0, \ldots, 0)$:

$$\boldsymbol{\Lambda_P} = \gamma_0\boldsymbol{\Lambda}_{\mathbb{1}}^{\dagger}\boldsymbol{\Lambda}_{\mathbb{1}} + \gamma_1\boldsymbol{\Lambda}_{\boldsymbol{D}}^{\dagger}\boldsymbol{\Lambda_D}. = \mathrm{diag}(\gamma_0, \gamma_1|\mathring{d}_1|^2, \ldots, \gamma_1|\mathring{d}_{N-1}|^2). \tag{10}$$

The determinant has a separable expression,

$$\det[\boldsymbol{\Lambda_P}] = \gamma_0\gamma_1^{N-1}\prod_{n=1}^{N-1} |\mathring{d}_n|^2, \tag{11}$$

i.e., the precision parameters have been factorized. In addition, each parameter controls a different set of frequencies:

$$p(\mathring{\boldsymbol{x}}|\gamma_0,\gamma_1) = p(\mathring{x}_0|\gamma_0)p(\mathring{\boldsymbol{x}}_*|\gamma_1),$$

where γ_0 drives the empirical mean level of the image $\mathring{x}_0$ and γ_1 drives the smoothness $\mathring{\boldsymbol{x}}_*$ of the image. With the Fourier precision structure of Eq. (10), we have the non-degenerated prior law for the image that addresses separately all the frequencies with a factorized partition function w.r.t. (γ_0,γ_1):

$$p(\boldsymbol{x}|\gamma_0,\gamma_1) = (2\pi)^{-N/2}\prod_{n=1}^{N-1}|\mathring{d}_n|\gamma_0^{1/2}\gamma_1^{(N-1)/2} \times \exp\left[-\frac{\gamma_0}{2}\|\mathring{x}_0\|^2 - \frac{\gamma_1}{2}\|\boldsymbol{\Lambda}_{\boldsymbol{D}^*}\mathring{\boldsymbol{x}}_*\|^2\right], \qquad (12)$$

where $\boldsymbol{\Lambda}_{\boldsymbol{D}^*}$ is obtained from $\boldsymbol{\Lambda}_{\boldsymbol{D}}$ without the first line and column. The next step is to write the *a priori* law for the noise in an explicit form and the other parameters, including the law parameters $\boldsymbol{\gamma}$ and the instrument parameters $\boldsymbol{w}$.

B. Noise and Data Laws

From a methodological standpoint, any statistic can be included for errors (measurement and model errors). It is possible to account for correlations in the error process or to account for a non-Gaussian law, e.g., Laplacian law, generalized Gaussian law, or other laws based on a robust norm. In the present paper, the noise is modeled as zero-mean white Gaussian vector with unknown precision parameter γ_ϵ:

$$p(\boldsymbol{\epsilon}|\gamma_\epsilon) = (2\pi)^{-N/2}\gamma_\epsilon^{N/2}\exp\left[-\frac{\gamma_\epsilon}{2}\|\boldsymbol{\epsilon}\|^2\right]. \qquad (13)$$

Consequently, the likelihood for the parameters given the observed data is written

$$p(\boldsymbol{y}|\boldsymbol{x},\gamma_\epsilon,\boldsymbol{w}) = (2\pi)^{-N/2}\gamma_\epsilon^{N/2}\exp\left[-\frac{\gamma_\epsilon}{2}\|\boldsymbol{y}-\boldsymbol{H}_{\boldsymbol{w}}\boldsymbol{x}\|^2\right]. \qquad (14)$$

It naturally depends on the image $\boldsymbol{x}$, on the noise parameter γ_ϵ, and the PSF parameters $\boldsymbol{w}$ embedded in $\boldsymbol{H}_{\boldsymbol{w}}$. It clearly involves a least-squares discrepancy that can be rewritten in the Fourier domain: $\|\boldsymbol{y}-\boldsymbol{H}_{\boldsymbol{w}}\boldsymbol{x}\|^2=\|\mathring{\boldsymbol{y}}-\boldsymbol{\Lambda}_{\boldsymbol{H}}\mathring{\boldsymbol{x}}\|^2$.

C. Hyperparameter Law

A classical choice for the hyperparameter law relies on the conjugate prior [32]: the conditional posterior for the hyperparameters is in the same family as its prior. It results in practical and algorithmic facilities: update of the laws amounts to update of a small number of parameters.

The three parameters γ_0, γ_1, and γ_ϵ are precision parameters of Gaussian laws Eq. (12) and (14); a conjugate law for these parameters is the gamma law (see Appendix B). Given parameters (α_i,β_i), for $i=0$, 1 or ϵ, the pdf reads

$$p(\gamma_i) = \frac{1}{\beta_i^{\alpha_i}\Gamma(\alpha_i)}\gamma_i^{\alpha_i-1}\exp(-\gamma_i/\beta_i), \qquad \forall\ \gamma_i \in [0,+\infty[. \qquad (15)$$

In addition to computational efficiency, the law allows for non-informative priors. With specific parameter values, one obtains two improper non-informative priori: the Jeffreys law $p(\gamma)=1/\gamma$ and the uniform law $p(\gamma)=\mathcal{U}_{[0,+\infty[}(\gamma)$ with (α_i,β_i) set to $(0,+\infty)$ and $(1,+\infty)$, respectively. The Jeffreys law is a classical law for the precisions and is considered non-informative [33]. This law is also invariant to power transformations: the law of γ^n [33,34] is also a Jeffreys law. For these reasons development is done using the Jeffreys law.

D. PSF Parameter Law

Regarding the PSF parameters $\boldsymbol{w}$, we consider that the instrument design process or a physical study provides a nominal value $\bar{\boldsymbol{w}}$ with uncertainty $\boldsymbol{\delta}$, that is, $\boldsymbol{w}\in[\bar{\boldsymbol{w}}-\boldsymbol{\delta},\bar{\boldsymbol{w}}+\boldsymbol{\delta}]$. The "Principle of Insufficient Reason" [33] leads to a uniform prior on this interval:

$$p(\boldsymbol{w}) = \mathcal{U}_{\bar{\boldsymbol{w}},\boldsymbol{\delta}}(\boldsymbol{w}), \qquad (16)$$

where $\boldsymbol{\mathcal{U}}_{\bar{\boldsymbol{w}},\boldsymbol{\delta}}$ is a uniform pdf on $[\bar{\boldsymbol{w}}-\boldsymbol{\delta},\bar{\boldsymbol{w}}+\boldsymbol{\delta}]$. Nevertheless, within the proposed framework, the choice is not limited and other laws, such as Gaussian, are possible. In any case, other choices do not allow easier computation because of the non-linear dependency of the observation model w.r.t. PSF parameters.

4. PROPER POSTERIOR LAW

At this point, the prior law of each parameter is available: the PSF parameters, the hyperparameters, and the image. Thus, the joint law for all the parameters is built by multiplying the likelihood Eq. (14) and the *a priori* laws Eq. (12), (15), and (16):

$$p(\mathring{\boldsymbol{x}},\gamma_\epsilon,\gamma_0,\gamma_1,\boldsymbol{w},\mathring{\boldsymbol{y}}) = p(\mathring{\boldsymbol{y}}|\mathring{\boldsymbol{x}},\gamma_\epsilon,\boldsymbol{w})p(\mathring{\boldsymbol{x}}|\gamma_0,\gamma_1)p(\gamma_\epsilon)p(\gamma_0)p(\gamma_1)p(\boldsymbol{w}) \qquad (17)$$

and explicitly

$$p(\mathring{\boldsymbol{x}},\gamma_\epsilon,\gamma_0,\gamma_1,\boldsymbol{w},\mathring{\boldsymbol{y}}) = \frac{(2\pi)^{-N}\prod_{n=1}^{N-1}|\mathring{d}_n|}{\beta_\epsilon^{\alpha_\epsilon}\Gamma(\alpha_\epsilon)\beta_0^{\alpha_0}\Gamma(\alpha_0)\beta_1^{\alpha_1}\Gamma(\alpha_1)}\gamma_\epsilon^{\alpha_\epsilon+N/2-1}\gamma_0^{\alpha_0-1/2}\gamma_1^{\alpha_1+(N-1)/2-1} \times\exp\left[-\frac{\gamma_\epsilon}{\beta_\epsilon}-\frac{\gamma_0}{\beta_0}-\frac{\gamma_1}{\beta_1}\right]\boldsymbol{\mathcal{U}}_{\bar{\boldsymbol{w}},\boldsymbol{\delta}}(\boldsymbol{w}) \times\exp\left[-\frac{\gamma_\epsilon}{2}\|\mathring{\boldsymbol{y}}-\boldsymbol{\Lambda}_{\boldsymbol{H}}\mathring{\boldsymbol{x}}\|^2-\frac{\gamma_0}{2}|\mathring{x}_0|^2-\frac{\gamma_1}{2}\|\boldsymbol{\Lambda}_{\boldsymbol{D}}\mathring{\boldsymbol{x}}\|^2\right]. \qquad (18)$$

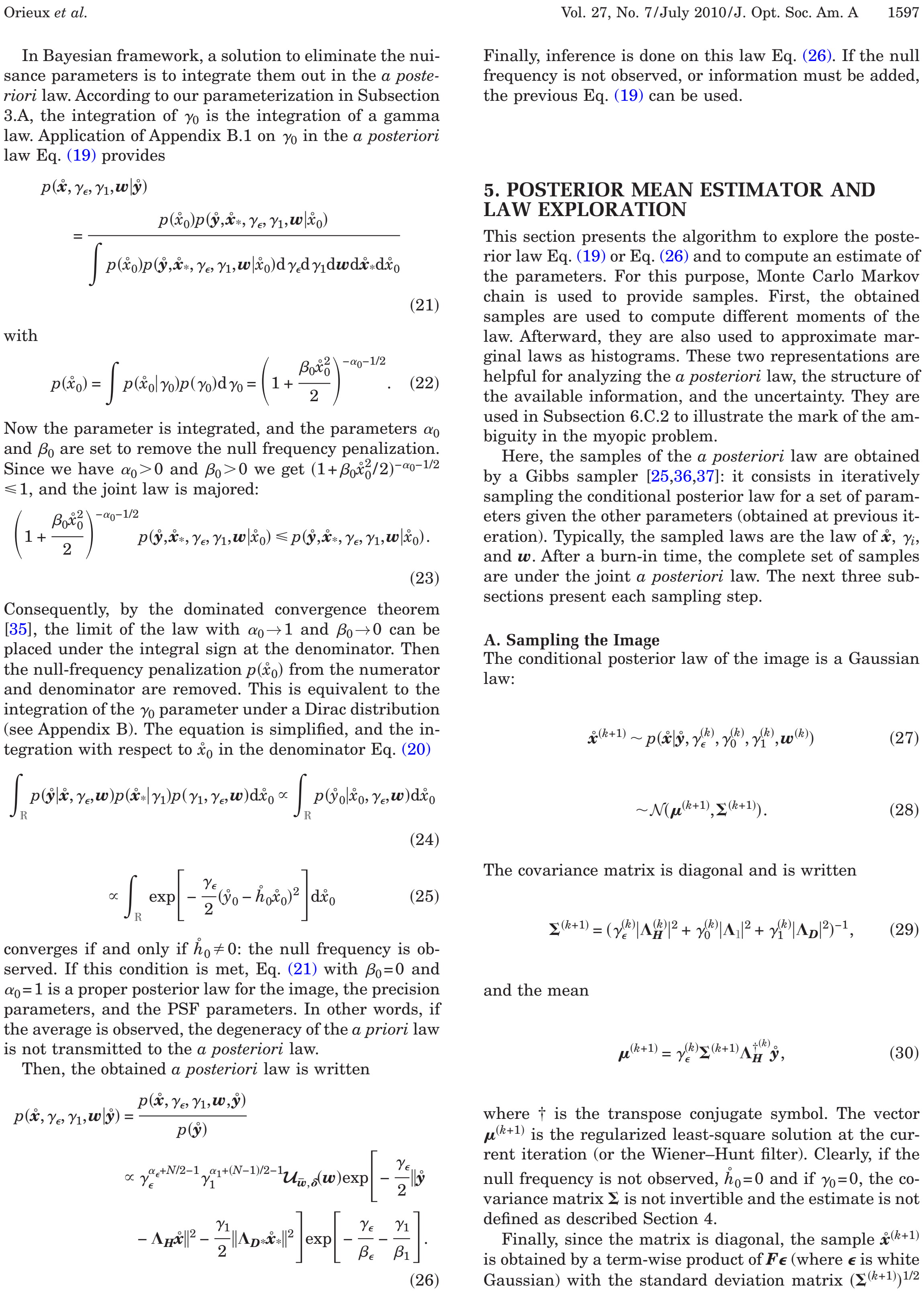
According to the Bayes rule, the *a posteriori* law reads

$$p(\mathring{\boldsymbol{x}},\gamma_\epsilon,\gamma_0,\gamma_1,\boldsymbol{w}|\mathring{\boldsymbol{y}}) = \frac{p(\mathring{\boldsymbol{x}},\gamma_\epsilon,\gamma_0,\gamma_1,\boldsymbol{w},\mathring{\boldsymbol{y}})}{p(\mathring{\boldsymbol{y}})}, \qquad (19)$$

where $p(\mathring{\boldsymbol{y}})$ is a normalization constant

$$p(\mathring{\boldsymbol{y}}) = \int p(\mathring{\boldsymbol{y}},\mathring{\boldsymbol{x}},\boldsymbol{\gamma},\boldsymbol{w})\mathrm{d}\mathring{\boldsymbol{x}}\mathrm{d}\boldsymbol{\gamma}\mathrm{d}\boldsymbol{w}. \qquad (20)$$

As described before, setting $\gamma_0=0$ leads to degenerated prior and joint laws. However, when the observation system preserves the null frequency, γ_0 can be considered a nuisance parameter. In addition, only prior information on the smoothness is available.



In Bayesian framework, a solution to eliminate the nuisance parameters is to integrate them out in the *a posteriori* law. According to our parameterization in Subsection 3.A, the integration of γ_0 is the integration of a gamma law. Application of Appendix B.1 on γ_0 in the *a posteriori* law Eq. (19) provides

$$p(\mathring{\boldsymbol{x}}, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{\boldsymbol{y}}) = \frac{p(\mathring{x}_0) p(\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}_*, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{x}_0)}{\int p(\mathring{x}_0) p(\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}_*, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{x}_0) \mathrm{d}\gamma_\epsilon \mathrm{d}\gamma_1 \mathrm{d}\boldsymbol{w} \mathrm{d}\mathring{\boldsymbol{x}}_* \mathrm{d}\mathring{x}_0} \tag{21}$$

with

$$p(\mathring{x}_0) = \int p(\mathring{x}_0 | \gamma_0) p(\gamma_0) \mathrm{d}\gamma_0 = \left(1 + \frac{\beta_0 \mathring{x}_0^2}{2}\right)^{-\alpha_0 - 1/2}. \tag{22}$$

Now the parameter is integrated, and the parameters α_0 and β_0 are set to remove the null frequency penalization. Since we have $\alpha_0 > 0$ and $\beta_0 > 0$ we get $(1 + \beta_0 \mathring{x}_0^2 / 2)^{-\alpha_0 - 1/2} \leqslant 1$, and the joint law is majored:

$$\left(1 + \frac{\beta_0 \mathring{x}_0^2}{2}\right)^{-\alpha_0 - 1/2} p(\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}_*, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{x}_0) \leqslant p(\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}_*, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{x}_0). \tag{23}$$

Consequently, by the dominated convergence theorem [35], the limit of the law with $\alpha_0 \to 1$ and $\beta_0 \to 0$ can be placed under the integral sign at the denominator. Then the null-frequency penalization $p(\mathring{x}_0)$ from the numerator and denominator are removed. This is equivalent to the integration of the γ_0 parameter under a Dirac distribution (see Appendix B). The equation is simplified, and the integration with respect to $\mathring{x}_0$ in the denominator Eq. (20)

$$\int_{\mathbb{R}} p(\mathring{\boldsymbol{y}} | \mathring{\boldsymbol{x}}, \gamma_\epsilon, \boldsymbol{w}) p(\mathring{\boldsymbol{x}}_* | \gamma_1) p(\gamma_1, \gamma_\epsilon, \boldsymbol{w}) \mathrm{d}\mathring{x}_0 \propto \int_{\mathbb{R}} p(\mathring{y}_0 | \mathring{x}_0, \gamma_\epsilon, \boldsymbol{w}) \mathrm{d}\mathring{x}_0 \tag{24}$$

$$\propto \int_{\mathbb{R}} \exp\left[-\frac{\gamma_\epsilon}{2} (\mathring{y}_0 - \mathring{h}_0 \mathring{x}_0)^2\right] \mathrm{d}\mathring{x}_0 \tag{25}$$

converges if and only if $\mathring{h}_0 \neq 0$: the null frequency is observed. If this condition is met, Eq. (21) with $\beta_0 = 0$ and $\alpha_0 = 1$ is a proper posterior law for the image, the precision parameters, and the PSF parameters. In other words, if the average is observed, the degeneracy of the *a priori* law is not transmitted to the *a posteriori* law.

Then, the obtained *a posteriori* law is written

$$p(\mathring{\boldsymbol{x}}, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{\boldsymbol{y}}) = \frac{p(\mathring{\boldsymbol{x}}, \gamma_\epsilon, \gamma_1, \boldsymbol{w}, \mathring{\boldsymbol{y}})}{p(\mathring{\boldsymbol{y}})}$$
$$\propto \gamma_\epsilon^{\alpha_\epsilon + N/2 - 1} \gamma_1^{\alpha_1 + (N-1)/2 - 1} \mathcal{U}_{\bar{\boldsymbol{w}}, \boldsymbol{\delta}}(\boldsymbol{w}) \exp\left[-\frac{\gamma_\epsilon}{2} \| \mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}} \mathring{\boldsymbol{x}} \|^2 - \frac{\gamma_1}{2} \| \boldsymbol{\Lambda}_{\boldsymbol{D}*} \mathring{\boldsymbol{x}}_* \|^2\right] \exp\left[-\frac{\gamma_\epsilon}{\beta_\epsilon} - \frac{\gamma_1}{\beta_1}\right]. \tag{26}$$

Finally, inference is done on this law Eq. (26). If the null frequency is not observed, or information must be added, the previous Eq. (19) can be used.

5. POSTERIOR MEAN ESTIMATOR AND LAW EXPLORATION

This section presents the algorithm to explore the posterior law Eq. (19) or Eq. (26) and to compute an estimate of the parameters. For this purpose, Monte Carlo Markov chain is used to provide samples. First, the obtained samples are used to compute different moments of the law. Afterward, they are also used to approximate marginal laws as histograms. These two representations are helpful for analyzing the *a posteriori* law, the structure of the available information, and the uncertainty. They are used in Subsection 6.C.2 to illustrate the mark of the ambiguity in the myopic problem.

Here, the samples of the *a posteriori* law are obtained by a Gibbs sampler [25,36,37]: it consists in iteratively sampling the conditional posterior law for a set of parameters given the other parameters (obtained at previous iteration). Typically, the sampled laws are the law of $\mathring{\boldsymbol{x}}$, γ_i, and $\boldsymbol{w}$. After a burn-in time, the complete set of samples are under the joint *a posteriori* law. The next three subsections present each sampling step.

A. Sampling the Image

The conditional posterior law of the image is a Gaussian law:

$$\mathring{\boldsymbol{x}}^{(k+1)} \sim p(\mathring{\boldsymbol{x}} | \mathring{\boldsymbol{y}}, \gamma_\epsilon^{(k)}, \gamma_0^{(k)}, \gamma_1^{(k)}, \boldsymbol{w}^{(k)}) \tag{27}$$

$$\sim \mathcal{N}(\boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}). \tag{28}$$

The covariance matrix is diagonal and is written

$$\boldsymbol{\Sigma}^{(k+1)} = (\gamma_\epsilon^{(k)} |\boldsymbol{\Lambda}_{\boldsymbol{H}}^{(k)}|^2 + \gamma_0^{(k)} |\boldsymbol{\Lambda}_{\mathbb{1}}|^2 + \gamma_1^{(k)} |\boldsymbol{\Lambda}_{\boldsymbol{D}}|^2)^{-1}, \tag{29}$$

and the mean

$$\boldsymbol{\mu}^{(k+1)} = \gamma_\epsilon^{(k)} \boldsymbol{\Sigma}^{(k+1)} \boldsymbol{\Lambda}_{\boldsymbol{H}}^{\dagger(k)} \mathring{\boldsymbol{y}}, \tag{30}$$

where † is the transpose conjugate symbol. The vector $\boldsymbol{\mu}^{(k+1)}$ is the regularized least-square solution at the current iteration (or the Wiener–Hunt filter). Clearly, if the null frequency is not observed, $\mathring{h}_0 = 0$ and if $\gamma_0 = 0$, the covariance matrix $\boldsymbol{\Sigma}$ is not invertible and the estimate is not defined as described Section 4.

Finally, since the matrix is diagonal, the sample $\mathring{\boldsymbol{x}}^{(k+1)}$ is obtained by a term-wise product of $\boldsymbol{F\epsilon}$ (where $\boldsymbol{\epsilon}$ is white Gaussian) with the standard deviation matrix $(\boldsymbol{\Sigma}^{(k+1)})^{1/2}$

followed by the addition of the mean $\boldsymbol{\mu}^{(k+1)}$, also computed with term-wise products Eq. (30). Consequently, the sampling of the image is effective even with a high-dimensional object.

B. Sampling Precision Parameters

The conditional posterior laws of the precisions are gamma corresponding to their prior law with parameters updated by the likelihood

$$\gamma_i^{(k+1)} \sim p(\gamma_i | \mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}^{(k+1)}, \boldsymbol{w}^{(k)}) \tag{31}$$

$$\sim \mathcal{G}(\gamma_i | \alpha_i^{(k+1)}, \beta_i^{(k+1)}). \tag{32}$$

For γ_ϵ, γ_0, and γ_1 the parameters law are, respectively,

$$\alpha_\epsilon^{(k+1)} = \alpha_\epsilon + N/2 \quad \text{and} \quad \beta_\epsilon^{(k+1)} = \left(\beta_\epsilon^{-1} + \frac{1}{2}\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}}^{(k)} \mathring{\boldsymbol{x}}^{(k+1)}\|^2\right)^{-1}, \tag{33}$$

$$\alpha_0^{(k+1)} = \alpha_0 + 1/2 \quad \text{and} \quad \beta_0^{(k+1)} = \left(\beta_0^{-1} + \frac{1}{2}(\mathring{x}_0^{(k+1)})^2\right)^{-1}, \tag{34}$$

$$\alpha_1^{(k+1)} = \alpha_1 + (N-1)/2 \quad \text{and} \quad \beta_1^{(k+1)} = \left(\beta_1^{-1} + \frac{1}{2}\|\boldsymbol{\Lambda}_{\boldsymbol{D}} \mathring{\boldsymbol{x}}^{(k+1)}\|^2\right)^{-1}. \tag{35}$$

In the case of the Jeffreys prior, the parameters are

$$\alpha_\epsilon^{(k+1)} = N/2 \quad \text{and} \quad \beta_\epsilon^{(k+1)} = 2/\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}}^{(k)} \mathring{\boldsymbol{x}}^{(k+1)}\|^2, \tag{36}$$

$$\alpha_0^{(k+1)} = 1/2 \quad \text{and} \quad \beta_0^{(k+1)} = 2/(\mathring{x}_0^{(k+1)})^2, \tag{37}$$

$$\alpha_1^{(k+1)} = (N-1)/2 \quad \text{and} \quad \beta_1^{(k+1)} = 2/\|\boldsymbol{\Lambda}_{\boldsymbol{D}} \mathring{\boldsymbol{x}}^{(k+1)}\|^2. \tag{38}$$

Remark 1: *If the a posteriori law Eq. (26) without γ_0 is considered, there is no need to sample this parameter [Eqs. (34) and (37) are not useful] and $\gamma_0^{(k)}=0$ in Eq. (29).*

C. Sample PSF Parameters

The conditional law for PSF parameters is written

$$\boldsymbol{w}^{(k+1)} \sim p(\boldsymbol{w} | \mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}^{(k+1)}, \gamma_\epsilon^{(k+1)}) \tag{39}$$

$$\propto \exp\left[-\frac{\gamma_\epsilon^{(k+1)}}{2}\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H},\boldsymbol{w}} \mathring{\boldsymbol{x}}^{(k+1)}\|^2\right], \tag{40}$$

where parameters $\boldsymbol{w}$ are embedded in the PSF $\boldsymbol{\Lambda}_{\boldsymbol{H}}$. This law is not standard and is intricate: no algorithm exists for direct sampling and we use the Metropolis–Hastings method to bypass this difficulty. In the Metropolis–Hastings algorithm, a sample $\boldsymbol{w}_{\text{p}}$ is proposed and accepted with a certain probability. This probability depends on the ratio between the likelihood of the proposed value and the likelihood of the current value $\boldsymbol{w}^{(k)}$. In practice, in the independent form described in appendix C, with prior law as proposition law, it is divided into several steps.

1. PROPOSITION: Sample a proposition:

$$\boldsymbol{w}_{\text{p}} \sim p(\boldsymbol{w}) = \mathcal{U}_{[\mathbf{a}\ \mathbf{b}]}(\boldsymbol{w}). \tag{41}$$

2. PROBABILITY OF ACCEPTANCE: Calculate the criterion:

$$J(\boldsymbol{w}^{(k)}, \boldsymbol{w}_{\text{p}}) = \frac{\gamma_\epsilon^{(k+1)}}{2}\left(\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H},\boldsymbol{w}^{(k)}} \mathring{\boldsymbol{x}}^{(k+1)}\|^2 - \|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H},\boldsymbol{w}_{\text{p}}} \mathring{\boldsymbol{x}}^{(k+1)}\|^2\right). \tag{42}$$

3. UPDATE: Sample $t \sim \mathcal{U}_{[0\ 1]}$ and take

$$\boldsymbol{w}^{(k+1)} = \begin{cases} \boldsymbol{w}_{\text{p}} & \text{if } \log t < J \\ \boldsymbol{w}^{(k)} & \text{otherwise} \end{cases}. \tag{43}$$

D. Empirical Mean

The sampling of $\mathring{\boldsymbol{x}}$, $\boldsymbol{\gamma}$, and $\boldsymbol{w}$ is repeated iteratively until the law has been sufficiently explored. These samples $[\mathring{\boldsymbol{x}}^{(k)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{w}^{(k)}]$ follow the global *a posteriori* law of Eq. (19). By the large-numbers law, the estimate, defined as the posterior mean, is approximated by

$$\hat{\boldsymbol{x}} = \boldsymbol{F}^\dagger \mathbb{E}[\mathring{\boldsymbol{x}}] \approx \boldsymbol{F}^\dagger \left[\frac{1}{K}\sum_{k=0}^{K-1} \mathring{\boldsymbol{x}}^{(k)}\right]. \tag{44}$$

As described by Eq. (44), to obtain an estimate of the image in the spatial space, all the computations are achieved recursively in the Fourier space with a single IFFT at the end. An implementation example in pseudo code is described Appendix D.

6. DECONVOLUTION RESULTS

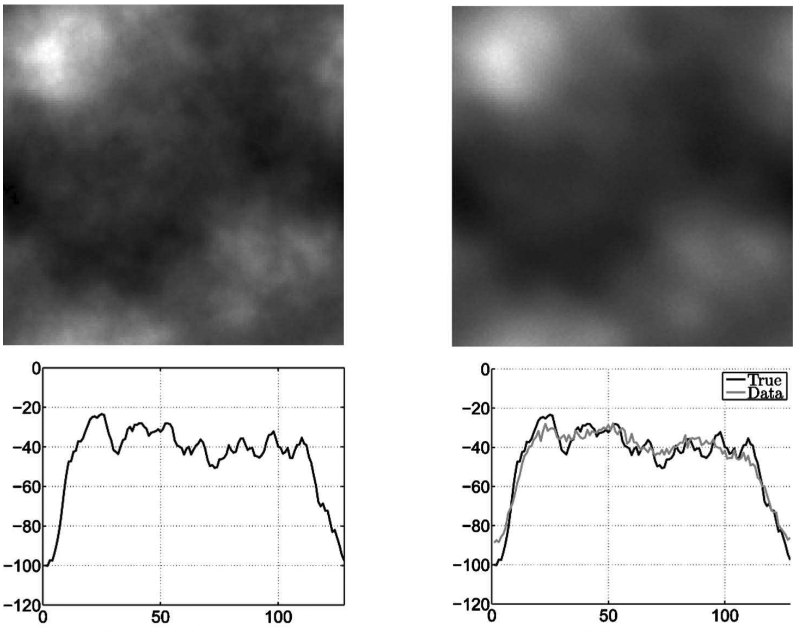
This section presents numerical results obtained by the proposed method. In order to completely evaluate the method, true value of all parameters $\boldsymbol{x}$, $\boldsymbol{w}$, γ_ϵ but also γ_1, γ_0 is needed. In order to achieve this, an entirely simulated case is studied: image and noise are simulated under their respective prior laws Eq. (12) and (13) with given values of γ_0, γ_1, and γ_ϵ. Thanks to this protocol, all experimental conditions are controlled and the estimation method is entirely evaluated.

The method has also been applied in different conditions (lower signal-to-noise ratio, broader PSF, different and realistic (non-simulated) images, and so on) and showed similar behavior. However, in the case of realistic images, since the true value of the hyperparameters γ_0 and γ_1 is unknown, the evaluation cannot be complete.

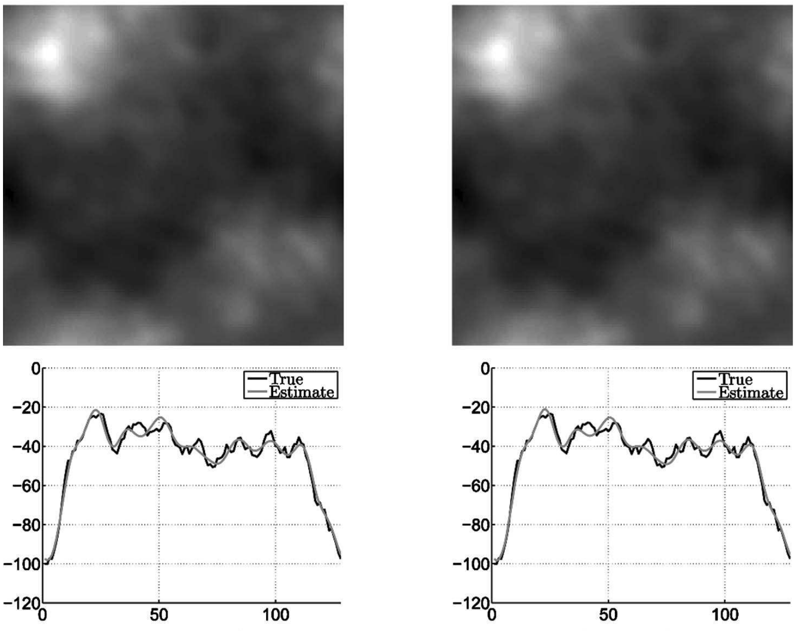
A. Practical Experimental Conditions

Concretely, a 128×128 image is generated in the Fourier space as the product of a complex white Gaussian noise and the *a priori* standard deviation matrix $\boldsymbol{\Sigma} = (\gamma_0 \boldsymbol{\Lambda}_1^\dagger \boldsymbol{\Lambda}_1 + \gamma_1 \boldsymbol{\Lambda}_{\boldsymbol{D}}^\dagger \boldsymbol{\Lambda}_{\boldsymbol{D}})^{-1/2}$, given by Eq. (10). The chosen matrix $\boldsymbol{\Lambda}_{\boldsymbol{D}}$ results from the FFT-2D of the Laplacian operator $[0\ 1\ 0;\ 1\ {-4}\ 1;\ 0\ 1\ 0]/8$, and the parameter values are $\gamma_0 = 1$ and $\gamma_1 = 2$.

These parameters provide the image shown in Fig. 1(a): it is an image with smooth features similar to a cloud. Pixels have numerical values between -100 and 150, and the profile line 68 shows fluctuations around a value of -40.



(a) True image (b) Data



(c) Non-myopic (d) Myopic

Fig. 1. (a) 128×128 sample of the *a priori* law for the object with $\gamma_0 = 1$ and $\gamma_1 = 2$. (b) Data computed with the PSF shown in Fig. 2. (c) and (d) Estimates with non-myopic and the myopic estimate, respectively. Profiles correspond to the 68th line of the image.

The *a priori* law for the hyperparameters are set to the non-informative Jeffreys law by fixing the (α_i, β_i) to $(0, +\infty)$, as explained in Subsection. 3.C. In addition, the PSF is obtained in the Fourier space by discretization of a normalized Gaussian shape,

$$\begin{aligned}\mathring{h}(\nu_\alpha, \nu_\beta) = \exp[-2\pi^2(&\nu_\alpha^2(w_\alpha \cos^2\varphi + w_\beta \sin^2\varphi) \\ &+ \nu_\beta^2(w_\alpha \sin^2\varphi + w_\beta \cos^2\varphi) \\ &+ 2\nu_\alpha\nu_\beta \sin\varphi \cos\varphi(w_\alpha - w_\beta))],\end{aligned} \quad (45)$$

with frequencies $(\nu_\alpha, \nu_\beta) \in [-0.5;0.5]^2$. This low-pass filter, illustrated in Fig. 2, is controlled by three parameters:

- Two width parameters w_α and w_β set to 20 and 7, respectively. Their *a priori* laws are uniform: $p(w_\alpha) = \mathcal{U}_{[19\ 21]}(w_\alpha)$ and $p(w_\beta)=\mathcal{U}_{[6\ 8]}(w_\alpha)$ corresponding to an uncertainty of about 5% and 15% around the nominal value (see Subsection 3.D).
- A rotation parameter φ set to $\pi/3$. The *a priori* law is also uniform: $p(\varphi)=\mathcal{U}_{[\pi/4\ \pi/2]}(\varphi)$, corresponding to 50% uncertainty.

Then, the convolution is computed in the Fourier space, and the data are obtained by adding a white Gaussian noise with precision $\gamma_\epsilon=0.5$. Data are shown Fig. 1(b): they are naturally smoother than the true image, and the small fluctuations are less visible and are corrupted by the noise. The empirical mean level of the image is correctly observed (the null frequency coefficient of $\boldsymbol{H_w}$ is $\mathring{h}_0=1$) so the parameter γ_0 is considered a nuisance parameter. Consequently it is integrated out under a Dirac distribution (see Section 4). This is equivalent to fix its value to 0 in the algorithm of Appendix D, line 4.

Finally, the method is evaluated on two different situations.

1. The unsupervised and non-myopic case: the parameters $\boldsymbol{w}$ are known. Consequently, there is no M.H. step (Subsection 5.C): lines 9–16 are ignored in the algorithm of Appendix D, and $\boldsymbol{w}$ is set to its true value. To obtain sufficient law exploration, the algorithm is run until the difference between two successive empirical means is less than 10^{-3}. In this case, 921 samples are necessary, and they are computed in approximately 12 s on a processor at 2.66 GHz with Matlab,
2. The unsupervised and myopic case: all the parameters are estimated. To obtain sufficient law exploration, the algorithm is run until the difference between two successive empirical means is less than 5×10^{-5}. In this case, 18, 715 samples are needed, and they are computed in approximately 7 min.

Remark: *The algorithm has also been run for up to 1,000,000 samples, in both cases, without perceptible qualitative changes.*

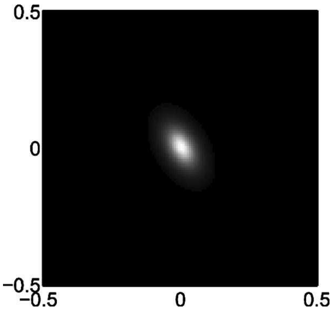


Fig. 2. PSF with $w_\alpha=20$, $w_\beta=7$ and $\varphi=\pi/3$. The x axis and y axis are reduced frequency.

B. Estimation Results

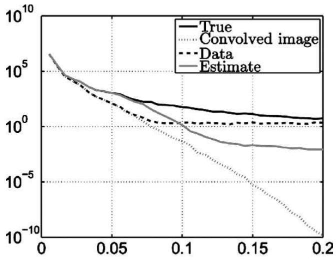
1. Images

The two results for the image are given Figs. 1(c) and 1(d) for the non-myopic and the myopic cases, respectively.

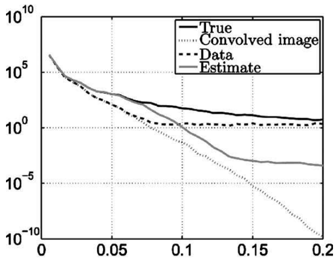
The effect of deconvolution is notable on the image, as well as on the shown profile. The object is correctly positioned, the orders of magnitude are respected, and the mean level is correctly reconstructed. The image is restored, more details are visible, and the profiles are closer matches to the true image than are the data. More precisely, pixels 20–25 of the 68th line of the image in Fig. 1 show the restoration of the original dynamic, whereas it is not visible in the data. Between pixels 70 and 110, fluctuations not visible in the data are also correctly restored.

In order to visualize and study the spectral contents of the images, the circular average of the empirical power spectral density is considered and is called "spectrum" hereafter. The subjacent spectral variable is a radial frequency f, such as $f^2=\nu_\alpha^2+\nu_\beta^2$. The spectra of the true object, data, and restored object are shown Figs. 3(a) and 3(b) in the non-myopic and myopic cases, respectively. It is clear that the spectrum of the true image is correctly retrieved, in both cases, up to the radial frequency $f\approx0.075$. Above this frequency, noise is clearly dominant and information about the image is almost lost. In other words, the method produces correct spectral equalization in the properly observed frequency band. The result is expected from the Wiener–Hunt method, but the achievement is the joint estimation of hyperparameter and instrument parameters in addition to the correct spectral equalization.

Concerning a comparison between non-myopic and myopic cases, there is no visual difference. The spectra in



(a) Non-Myopic



(b) Myopic

Fig. 3. Circular average of the empirical power spectral density of the image, the convolved image, the data (convolved image corrupted by noise), and the estimates, in radial frequency with the y axis in logarithmic scale. The x axis is the radial frequency.

Table 1. Error e [Eq. (46)] and Averaged Standard Deviation $\hat{\sigma}$ of the Posterior Image Law[a]

	Data	Non-myopic	Myopic	Best
Error (e)	11.092 %	6.241 %	6.253 %	6.235 %
$\hat{\sigma}$ of $\boldsymbol{x}$ law	–	3.16	3.25	–

[a] The "Best" error has been obtained with the knowledge of the true image.

Figs. 3(a) and 3(b) in the non-myopic and myopic cases, respectively, are visually indistinguishable. This is also the case when comparing Figs. 1(c) and 1(d) and especially, the 68th line. From a more precise quantitative evaluation, a slight difference is observed is and detailed below.

In order to quantify performance, a normalized Euclidean distance

$$e = \|\boldsymbol{x} - \boldsymbol{x}^*\| / \|\boldsymbol{x}^*\| \tag{46}$$

between an image $\boldsymbol{x}$ and the true image $\boldsymbol{x}^*$ is considered. It is computed between the true image and the estimated images as well as between the true image and the data. Results are reported in Table 1 and confirm that the deconvolution is effective with an error of approximately 6% in the myopic case compared with 11% for the data. Both non-myopic and myopic deconvolution reduce error by a factor of 1.7 with respect to the observed data.

Regarding a comparison between the non-myopic and myopic cases, the errors are almost the same, with a slightly lower value for the non-myopic case, as expected. This difference is consistent with the intuition: more information is injected in the non-myopic case through the true PSF parameter values.

2. Hyperparameters and Instrument Parameters

Concerning the other parameters, their estimates are close to the true values and are reported in Table 2. The γ_ϵ estimate is very close to the true value with $\hat{\gamma}_\epsilon = 0.49$ instead of 0.5 in the two cases. The error for the PSF parameters are 0.35%, 2.7%, and 1.9% for w_α, w_β, and ϕ, respectively. The value of γ_1 is underestimated in the two cases, with approximately 1.7 instead of 2. All the true values fall in the $\hat{\mu} \pm 3\hat{\sigma}$ interval.

In order to deepen the numerical study, the paper evaluates the capability of the method to accurately select the best values for hyperparameters and instrument parameters. To this end, we compute the estimation error Eq. (46) for a set of "exhaustive" values of the parameters $[\gamma_\epsilon, \gamma_1, w_\alpha, w_\beta, \phi]$. The protocol is the following: (1) choose a new value for a parameter (γ_ϵ, for example) and fix the other parameters to the value provided by our algorithm, (2) compute the Wiener–Hunt solution (Subsection 5.A), and (3) compute the error index.

Results are reported in Fig. 4. In each case, smooth variation of error is observed when varying hyperparameters and instrument parameters, and unique optimum is visible. In this way, one can find the values of the parameters that provide the best Wiener–Hunt solution when the true image $\boldsymbol{x}^*$ is known. This is reported in Table 1 and shows almost imperceptible improvement: optimization of the parameters (based on the true image $\boldsymbol{x}^*$) allows negligible improvement (smaller than 0.02% as reported in Table 1).

Thus the main conclusion is that the proposed unsupervised and myopic approach is a relevant tool for tuning parameters: it works (without knowledge of the true image) as well as being an optimal approach (based on the knowledge of the true image).

C. A Posteriori Law Characteristics

This section describes the *a posteriori* law using histograms, means and variances of the parameters. The sample histograms, Figs. 5 and 6, provide an approximation of the marginal posterior law for each parameter. Tables 1 and 2 report the variance for the image and law parameters respectively and thus allow to quantify the uncertainty.

1. Hyperparameter Characteristics

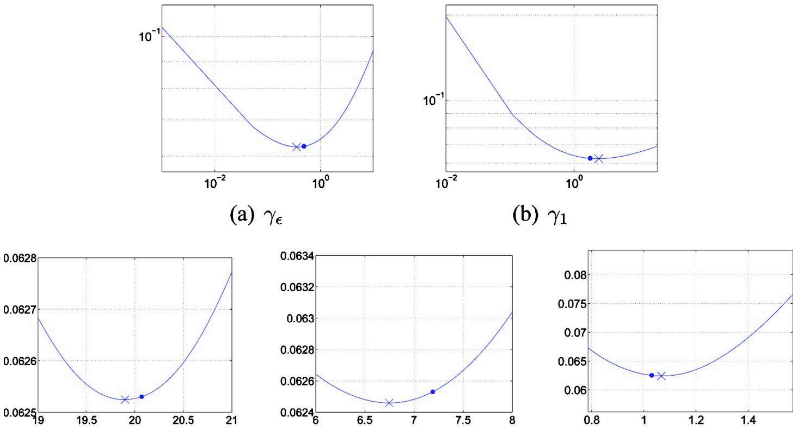
The histograms for γ_ϵ and γ_1, Fig. 5, are concentrated around a mean value in both non-myopic and myopic cases. The variance for γ_ϵ is lower than the one for γ_1, and it can be explained as follows.

The observed data are directly affected by noise (present at the system output), whereas they are indirectly affected by the object (present at the system input). The convolution system damages the object and not the noise: as a consequence, the parameter γ_ϵ (which drives noise law) is more reliably estimated than γ_1 (which drives object law).

A second observation is the smaller variance for γ_1 in the non-myopic case, Fig. 5(c), than in the myopic case, Fig. 5(d). This is the consequence of the addition of information in the non-myopic case w.r.t. the myopic one, through the value of the PSF parameters. In the myopic case, the estimates are based on the knowledge of an interval for the values of the instrument parameters, whereas in the non-myopic case, the estimates are based on the true values of the instrument parameters.

Table 2. Quantitative Evaluation: True and Estimated Values of Hyperparameters and PSF Parameters

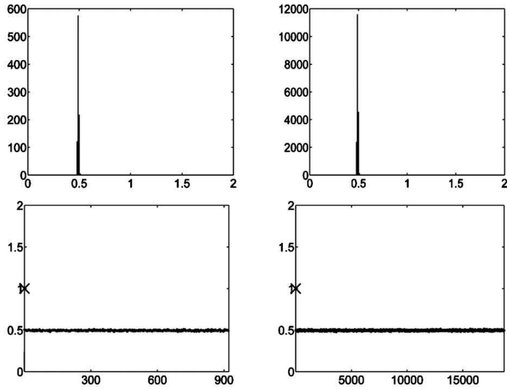
		$\hat{\gamma}_\epsilon \pm \hat{\sigma}$	$\hat{\gamma}_1 \pm \hat{\sigma}$	$\hat{w}_\alpha \pm \hat{\sigma}$	$\hat{w}_\beta \pm \hat{\sigma}$	$\hat{\varphi} \pm \hat{\sigma}$
Case	True Value	0.5	2	20	7	1.05 ($\pi/3$)
Non-myopic	Estimate	0.49±0.0056	1.78±0.14	–	–	–
	Error	2.0 %	11 %	–	–	–
Myopic	Estimate	0.49±0.0056	1.65±0.15	20.07±0.53	7.19±0.38	1.03±0.04
	Error	2.0 %	18 %	0.35 %	2.7 %	1.9 %



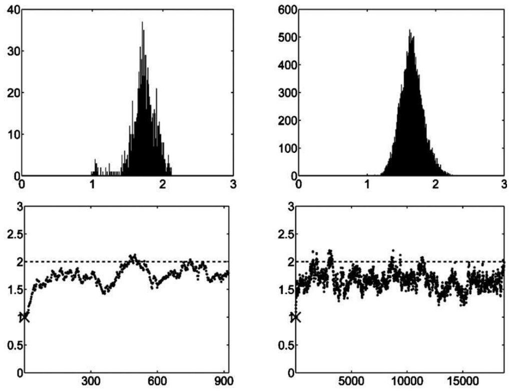
(a) γ_ϵ (b) γ_1

(c) w_α (d) w_β (e) φ

Fig. 4. (Color online) Computation of the best parameters in the sense e, Eq. (46). The symbol × is the minimum and the dot is the estimated value with our approach. The y axes of γ_ϵ and γ_1 are in logarithmic scale.



(a) γ_ϵ for non-myopic case (b) γ_ϵ for myopic case



(c) γ_1 for non-myopic case (d) γ_1 for myopic case

Fig. 5. Histograms and chains for the non-myopic case [(a) and (c)] and the myopic case [(b) and (d)] for γ_ϵ and γ_1, respectively. The symbol × on the y axes localizes the initial value, and the dashed line corresponds to the true value. The x axes are the iteration's index for the chains (bottom of figures) and the parameter value for the histograms (top of figures).

2. PSF Parameter Characteristics

Figure 6 gives histograms for the three PSF parameters and their appearance is quite different from the one for the hyperparameters. The histograms for w_α and w_β, Figs. 6(a) and 6(b), are not as concentrated as the one of Fig. 5 for hyperparameters. Their variances are quite large with respect to the interval of the prior law. In contrast, the histogram for the parameter φ, Fig. 6(c), has the smallest variance. It is analyzed as a consequence of a larger sensitivity of the data w.r.t. the parameter φ than w.r.t. the parameters w_α and w_β. In an equivalent manner, the observed data are more informative about the parameter φ than about the parameters w_α and w_β.

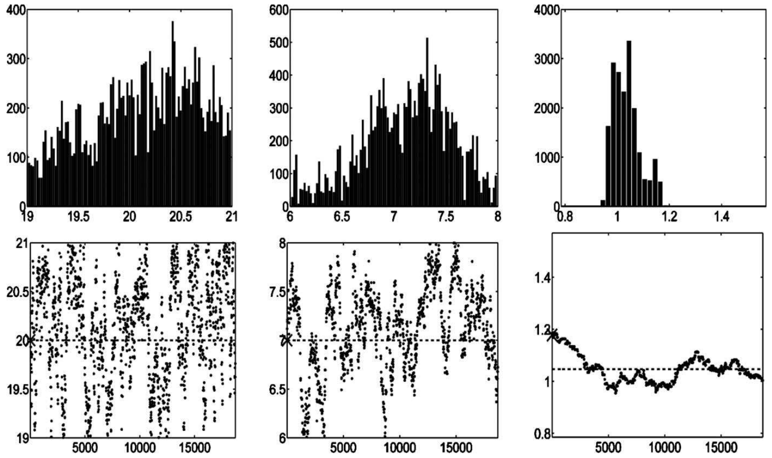
3. Mark of the Myopic Ambiguity

Finally, a correlation between parameters (γ_1, w_α) and (γ_1, w_β) is visible on their joint histograms, Fig. 7. It can be interpreted as a consequence of the ambiguity in the primitive myopic deconvolution problem, in the following manner: the parameters γ_1 and $\boldsymbol{w}$ both participate in the interpretation of the spectral content of data, γ_1, as a scale factor and $\boldsymbol{w}$ as a shape factor. An increase of w_α or w_β results in a decrease of the cutoff frequency of the observation system. In order to explain the spectral content of a given data set, the spectrum of the original image must contain more high frequencies, i.e., a smaller γ_1. This is also observed on the histogram illustrated, Fig. 7(a)

D. MCMC Algorithm Characteristics

Globally, the chains of Figs. 5 and 6, have a Markov feature (correlated) and explore the parameter space. They have a burn-in period followed by a stationary state. This characteristic has always been observed regardless of the initialization. For fixed experimental conditions, the stationary state of multiple runs was always around the same value. Considering different initializations, the only visible change is on the length of the burn-in period.

More precisely, the chain of γ_ϵ is concentrated in a small interval, the burn-in period is very short (fewer

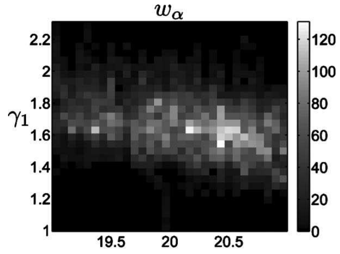


(a) w_α (b) w_β (c) φ

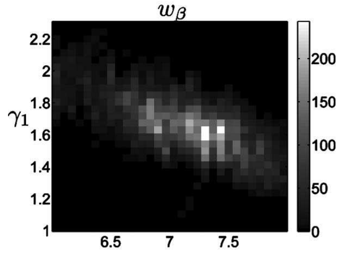
Fig. 6. Histogram and chain for the PSF parameters (a) w_α, (b) w_β, and (c) φ. The symbol × on the y axes localizes the initial value, and the dashed line corresponds to the true value. The x axis for the histograms and the y axis of the chain are limits of the *a priori* law.

than 10 samples), and its evolution seems independent of the other parameters. The chain of γ_1 has a larger exploration, the burn-in period is longer (approximately 200 samples), and the histogram is larger. This is in accordance with the analysis of Subsection 6.C.1.

Regarding PSF parameters, the behavior is different for (w_α, w_β) and φ. The chain of the two width parameters has a very good exploration of the parameter space with a quasi-instantaneous burn-in period. Conversely, the chain of φ is more concentrated, and its burn-in period is approximately 4,000 samples. This is also in accordance with previous analysis (Subsection 6.C.2).



(a) (γ_1, w_α)



(b) (γ_1, w_β)

Fig. 7. Joint histograms for the couple (a) (γ_1, w_α) and (b) (γ_1, w_β). The x and y axes are the parameter values.

Table 3. Acceptance Rate

	Parameter		
	w_α	w_β	φ
Acceptance rate	14.50 %	9.44 %	2.14 %

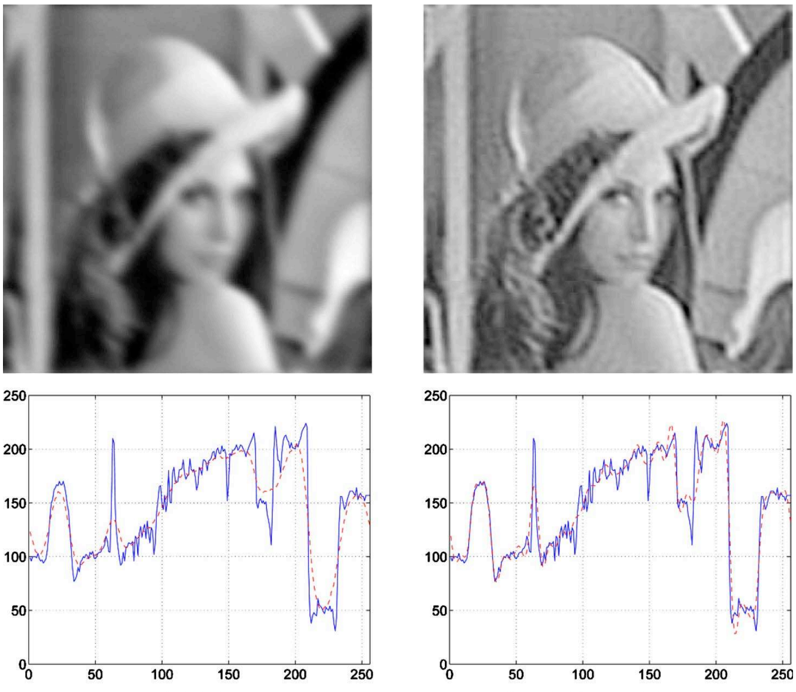
Acceptance rates in the Metropolis–Hastings algorithm are reported in Table 3: they are quite small, especially for the rotation parameter. This is due to the structure of the implemented algorithm: an independent Metropolis–Hastings algorithm with the prior law as a proposition law. The main advantage of this choice is its simplicity, but as a counterpart, a high rejection rate is observed due to a large *a priori* interval for the angle parameter. A future study will be devoted to the design of a more accurate proposition law.

E. Robustness of Prior Image Model

Figure 8 illustrates the proposed method on a more realistic image with heterogeneous spatial structures. The original is the Lena image, and the data have been obtained with the same Gaussian PSF and also with corruption by white Gaussian noise. Figure 8(b) shows that the restored image is closer to the true one than are the data. Smaller structures are visible and edges are sharper, for example, around pixel 200. The estimated parameters are $\hat{\gamma}_\epsilon = 1.98$ while the true value is $\gamma_\epsilon^* = 2$. Concerning the PSF parameters, the results are $\hat{w}_\alpha = 19.3$, $\hat{w}_\beta = 7.5$ and $\hat{\varphi} = 1.15$, while the true values are, respectively, $w_\alpha^* = 20$, $w_\beta^* = 7$ and $\varphi^* = 1.05$ as in Subsection 6.B. Here again, the estimated PSF parameters are close to the true values, giving an initial assessment of the capability of the method in a more realistic context.

7. CONCLUSION AND PERSPECTIVES

This paper presents a new global and coherent method for myopic and unsupervised deconvolution of relatively smooth images. It is built within a Bayesian framework



(a) Data

(b) Estimated image

Fig. 8. (Color online) (a) Observed image and (b) restored image. Profiles correspond to the 68th line. The solid curve is the true profile, and the dashed curve correspond to (a) data and (b) estimated profiles.

and a proper extended *a posteriori* law for the PSF parameters, the hyperparameters, and the image. The estimate, defined as the posterior mean, is computed by means of an MCMC algorithm in less than a few minutes.

Numerical assessment attests that the parameters of the PSF and the parameters of the prior laws are precisely estimated. In addition, results also demonstrate that the myopic and unsupervised deconvolved image is closer to the true image than are the data and shows true restored high frequencies as well as spatial details.

The paper focuses on a linear invariant model often encountered in astronomy, medical imaging, and nondestructive testing and especially in optical problems. Noninvariant linear models can also be considered in order to address other applications such as spectrometry [4] or fluorescence microscopy [13]. The loss of the invariance property precludes entirely Fourier-based computations, but the methodology remains valid and practicable. In particular, it is possible to draw samples of the image by means of an optimization algorithm [38].

Gaussian law, related to L_2 penalization, is known for possible excessive sharp edges penalization in the restored object. The use of convex L_2-L_1 penalization [39–41] or non-convex L_2-L_0 penalization [42] can overcome this limitation. In these cases a difficulty occurs in the development of myopic and unsupervised deconvolution: the partition function of the prior law for the image is in intricate or even unknown dependency w.r.t. the parameters [1,7,43]. However a recent paper [41] overcomes the difficulty, resulting in an efficient unsupervised deconvolution, and we plan to extend this work for the myopic case.

Regarding noise, Gaussian likelihood limits robustness to outliers or aberrant data, and it is possible to appeal to a robust law such as Huber penalization in order to bypass the limitation. Nevertheless, the partition function for the noise law is again difficult or impossible to manage, and it is possible to resort to the idea proposed in [41] to overcome the difficulty.

Finally, estimation of the parameters of the correlation matrix (cutoff frequency, attenuation coefficients, etc.) is possible within the same methodological framework. This could be achieved for the correlation matrix of the object or the noise. As for the PSF parameters, the approach could rely on an extended *a posteriori* law, including the new parameters and a Metropolis–Hastings sampler.

APPENDIX A: LAW IN FOURIER SPACE

For a Gaussian vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the law for $\mathring{\boldsymbol{x}} = \boldsymbol{F}\boldsymbol{x}$ (the FFT of $\boldsymbol{x}$) is also Gaussian, whose first two moments are the following:

- The mean is

$$\mathring{\boldsymbol{\mu}} = \mathbb{E}[\mathring{\boldsymbol{x}}] = \boldsymbol{F}\mathbb{E}[\mathring{\boldsymbol{x}}] = \boldsymbol{F}\boldsymbol{\mu}. \tag{A1}$$

- The covariance matrix is

$$\mathring{\boldsymbol{\Sigma}} = \mathbb{E}[(\mathring{\boldsymbol{x}} - \mathring{\boldsymbol{\mu}})(\mathring{\boldsymbol{x}} - \mathring{\boldsymbol{\mu}})^\dagger] = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^\dagger. \qquad \text{(A2)}$$

Moreover, if the covariance matrix $\boldsymbol{\Sigma}$ is circulant, it is written

$$\mathring{\boldsymbol{\Sigma}} = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^\dagger = \boldsymbol{\Lambda}_{\boldsymbol{\Sigma}}, \qquad \text{(A3)}$$

i.e., the covariance matrix $\mathring{\boldsymbol{\Sigma}}$ is diagonal.

APPENDIX B: GAMMA PROBABILITY DENSITY

1. Definition

The gamma pdf for $\gamma > 0$, with given parameter $\alpha > 0$ and $\beta > 0$, is written

$$\mathcal{G}(\gamma|\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \gamma^{\alpha-1} \exp(-\gamma/\beta). \qquad \text{(B1)}$$

Table 4 gives three limit cases for (α,β). The following properties hold:

- The mean is $\mathbb{E}_{\mathcal{G}}[\gamma] = \alpha\beta$.
- The variance is $\mathbb{V}_{\mathcal{G}}[\gamma] = \alpha\beta^2$.
- The maximizer is $\beta(\alpha-1)$ if and only if $\alpha > 1$.

2. Marginalization

First consider an N-dimensional zero-mean Gaussian vector with a given precision matrix $\gamma\boldsymbol{\Gamma}$ with $\gamma > 0$. The pdf reads

$$p(\boldsymbol{x}|\gamma) = (2\pi)^{-N/2} \gamma^{N/2} \det[\boldsymbol{\Gamma}]^{1/2} \exp[-\gamma \boldsymbol{x}^{\mathrm{t}} \boldsymbol{\Gamma} \boldsymbol{x}/2]. \qquad \text{(B2)}$$

So consider the conjugate pdf for γ as a gamma law with parameters (α,β) (see Appendix B.1). The joint law for $(\boldsymbol{x},\gamma)$ is the product of the pdf given by Eqs. (B1) and (B2): $p(\boldsymbol{x},\gamma) = p(\boldsymbol{x}|\gamma)p(\gamma)$. The marginalization of the joint law is known [44]:

$$\begin{aligned} p(\boldsymbol{x}) &= \int_{\mathbb{R}_+} p(\boldsymbol{x}|\gamma) p(\gamma) \mathrm{d}\gamma \\ &= \frac{\beta^{N/2} \det[\boldsymbol{\Gamma}]^{1/2} \Gamma(\alpha + N/2)}{(2\pi)^{N/2} \Gamma(\alpha)} \left(1 + \frac{\beta \boldsymbol{x}^{\mathrm{t}} \boldsymbol{\Gamma} \boldsymbol{x}}{2}\right)^{-\alpha-N/2}, \end{aligned} \qquad \text{(B3)}$$

which is an N-dimensional t-Student law of 2α degrees of freedom with a $\beta\Gamma$ precision matrix. Finally, the conditional law reads

$$p(\gamma|\boldsymbol{x}) = \frac{(2\pi)^{-N/2} \det[\boldsymbol{\Gamma}]^{1/2}}{\beta^\alpha \Gamma(\alpha)} \gamma^{\alpha+N/2-1} \exp[-\gamma \boldsymbol{x}^{\mathrm{t}} \boldsymbol{\Gamma} \boldsymbol{x}/2 + 1/\beta]. \qquad \text{(B4)}$$

Thanks to conjugacy, it is also a gamma pdf with parameters $\bar{\alpha}, \bar{\beta}$ given by $\bar{\alpha} = \alpha + N/2$ and $\bar{\beta}^{-1} = \beta^{-1} + 2/(\boldsymbol{x}^{\mathrm{t}} \boldsymbol{\Gamma} \boldsymbol{x})$.

Table 4. Specific Laws Obtained As Limit of the Gamma PDF

Law	α	β
Jeffreys	0	$+\infty$
Uniform	1	$+\infty$
Dirac	–	0

APPENDIX C: METROPOLIS–HASTINGS ALGORITHM

The Metropolis–Hastings algorithm provides samples of a target law $f(\boldsymbol{w})$ that cannot be directly sampled but can be evaluated, at least up to a multiplicative constant. Using the so called "instrument law" $q(\boldsymbol{w}_{\mathrm{p}}|\boldsymbol{w}^{(t)})$, samples of the target law are obtained by the following iterations:

1. Sample the proposition $\boldsymbol{w}_{\mathrm{p}} \sim q(\boldsymbol{w}_{\mathrm{p}}|\boldsymbol{w}^{(t)})$.
2. Compute the probability:

$$\rho = \min\left\{ \frac{f(\boldsymbol{w}_{\mathrm{p}})}{f(\boldsymbol{w}^{(t)})} \frac{q(\boldsymbol{w}^{(t)}|\boldsymbol{w}_{\mathrm{p}})}{q(\boldsymbol{w}_{\mathrm{p}}|\boldsymbol{w}^{(t)})}, 1 \right\}. \qquad \text{(C1)}$$

3. Take

$$\boldsymbol{w}^{(t+1)} = \begin{cases} \boldsymbol{w}_{\mathrm{p}} & \text{with } \rho \text{ probability} \\ \boldsymbol{w}^{(t)} & \text{with } 1-\rho \text{ probability} \end{cases}. \qquad \text{(C2)}$$

At convergence, the samples follow the target law $f(\boldsymbol{w})$ [25,36]. When $q(\boldsymbol{w}_{\mathrm{p}}|\boldsymbol{w}^{(t)}) = q(\boldsymbol{w}_{\mathrm{p}})$ the algorithm is named independent Metropolis–Hastings. In addition, if the instrument law is uniform, the acceptance probability becomes simpler:

$$\rho = \min\left\{ \frac{f(\boldsymbol{w}_{\mathrm{p}})}{f(\boldsymbol{w}^{(t)})}, 1 \right\}. \qquad \text{(C3)}$$

APPENDIX D

Pseudo-code algorithm. `gamrnd`, `rand` and `randn` draw samples of the gamma-variable, uniform-variable, and zero-mean unit-variance white complex Gaussian vector, respectively.

```
Initialization of $[\overset{o}{\boldsymbol{x}}^{(0)}, \gamma^{(0)}, \boldsymbol{w}^{(0)}, k=0]$
repeat
    % Sample of $\overset{o}{\boldsymbol{x}}$
    $\Sigma \leftarrow \gamma_{\epsilon}^{(k)}|\boldsymbol{\Lambda}_{\boldsymbol{H}}|^2 + \gamma_0^{(k)}|\boldsymbol{\Lambda}_{\mathbb{1}}|^2 + \gamma_1^{(k)}|\boldsymbol{\Lambda}_{\boldsymbol{D}}|^2$
    $\boldsymbol{\mu} \leftarrow \gamma_{\epsilon}^{(k)}\Sigma^{-1}\boldsymbol{\Lambda}_{\boldsymbol{H}}^{*}\overset{o}{\boldsymbol{y}}$
    $\overset{o}{\boldsymbol{x}}^{(k)} \leftarrow \boldsymbol{\mu} + \Sigma^{-1/2}.*\mathrm{randn}$
    % Sample of $\gamma$
    $\gamma_{\epsilon}^{(k)} \leftarrow \mathrm{gamrnd}(\alpha_{\epsilon}, \beta_{\epsilon})$
    $\gamma_1^{(k)} \leftarrow \mathrm{gamrnd}(\alpha_1, \beta_1)$
    $\gamma_0^{(k)} \leftarrow \mathrm{gamrnd}(\alpha_0, \beta_0)$
    % Sample of $\boldsymbol{w}$
    $\boldsymbol{w}_{\mathrm{p}} \leftarrow \mathrm{rand}*(\boldsymbol{a}-\boldsymbol{b})+\boldsymbol{a}$
    $J \leftarrow \gamma_{\epsilon}(\|\overset{o}{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}}\overset{o}{\boldsymbol{x}}\|^2 - \|\overset{o}{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H},w_{\mathrm{p}}}\overset{o}{\boldsymbol{x}}\|^2)/2$
    if $\log(\mathrm{rand}) < \min\{J, 0\}$ then
        $\boldsymbol{w}^{(k)} \leftarrow w_{\mathrm{p}}$
        $\boldsymbol{\Lambda}_{\boldsymbol{H}} \leftarrow \boldsymbol{\Lambda}_{\boldsymbol{H},w_{\mathrm{p}}}$
    else
        $\boldsymbol{w}^{(k)} \leftarrow \boldsymbol{w}^{(k-1)}$
    end if
    % Empirical mean
    $k \leftarrow k+1$
    $\overset{o}{\boldsymbol{x}}^{(k)} \leftarrow \Sigma_i \overset{o}{\boldsymbol{x}}^{(i)}/k$
until $|\bar{\boldsymbol{x}}^{(k)} - \bar{\boldsymbol{x}}^{(k-1)}|/|\bar{\boldsymbol{x}}^{(k)}| \leqslant \mathrm{criterion}$
```

ACKNOWLEDGMENTS

The authors would like to thank Professor Alain Abergel of Laboratoire Institute d'Astrophysique Spatial at Université Paris-Sud 11, France, for fruitful discussions and constructive suggestions. The authors are also grateful to Cornelia Vacar, Laboratoire d'Intégration du Matériau au Système (IMS), Univ. Bordeaux, for carefully reading the paper.

REFERENCES

1. J. Idier, ed., *Bayesian Approach to Inverse Problems* (Wiley, 2008).
2. R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," IEEE Trans. Image Process. **15**, 3715–3727 (2006).
3. P. Campisi and K. Egiazarian, eds., *Blind Image Deconvolution* (CRC Press, 2007).
4. T. Rodet, F. Orieux, J.-F. Giovannelli, and A. Abergel, "Data inversion for over-resolved spectral imaging in astronomy," IEEE J. Sel. Top. Signal Process. **2**, 802–811 (2008).
5. A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems* (Winston, 1977).
6. S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature," J. Assoc. Comput. Mach. **10**, 97–101 (1962).
7. A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia, "Hyperparameter estimation for satellite image restoration by a MCMC maximum likelihood method," Pattern Recogn. **35**, 341–352 (2002).
8. J. A. O'Sullivan, "Roughness penalties on finite domains," IEEE Trans. Image Process. **4**, 1258–1268 (1995).
9. G. Demoment, "Image reconstruction and restoration: overview of common estimation structure and problems," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-37**, 2024–2036 (1989).
10. P. Pankajakshani, B. Zhang, L. Blanc-Féraud, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, "Blind deconvolution for thin-layered confocal imaging," Appl. Opt. **48**, 4437–4448 (2009).
11. E. Thiébaut and J.-M. Conan, "Strict *a priori* constraints for maximum likelihood blind deconvolution," J. Opt. Soc. Am. A **12**, 485–492 (1995).
12. N. Dobigeon, A. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," IEEE Trans. Image Process. (2009).
13. B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models," Appl. Opt. **46**, 1819–1829 (2007).
14. L. Mugnier, T. Fusco, and J.-M. Conan, "MISTRAL: a myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images," J. Opt. Soc. Am. A **21**, 1841–1854 (2004).
15. E. Thiébaut, "MiRA: an effective imaging algorithm for optical interferometry," Proc. SPIE **7013**, 70131-I (2008).
16. T. Fusco, J.-P. Véran, J.-M. Conan, and L. M. Mugnier, "Myopic deconvolution method for adaptive optics images of stellar fields," Astron. Astrophys. Suppl. Ser. **134**, 193 (1999).
17. J.-M. Conan, L. Mugnier, T. Fusco, V. Michau, and G. Rousset, "Myopic deconvolution of adaptive optics images by use of object and point-spread function power spectra," Appl. Opt. **37**, 4614–4622 (1998).
18. A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," IEEE Trans. Image Process. **52**, 2222–2233 (2004).
19. T. Bishop, R. Molina, and J. Hopgood, "Blind restoration of blurred photographs via AR modelling and MCMC," in *Proceedings of 15th IEEE International Conference on Image Processing, 2008, ICIP 2008* (IEEE Signal Processing Society, 2008).
20. E. Y. Lam and J. W. Goodman, "Iterative statistical approach to blind image deconvolution," J. Opt. Soc. Am. A **17**, 1177–1184 (2000).
21. Z. Xu and E. Y. Lam, "Maximum *a posteriori* blind image deconvolution with Huber–Markov random-field regularization," Opt. Lett. **34**, 1453–1455 (2009).
22. M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," IEEE Trans. Acoust., Speech, Signal Process. **24**, 58–63 (1976).
23. A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, "Estimation of blur and noise parameters in remote sensing," in *Proceedings of 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002)* (IEEE Signal Processing Society, 2002), Vol. 4, pp. 3580–3583.
24. F. Chen and J. Ma, "An empirical identification method of Gaussian blur parameter for image deblurring," IEEE Trans. Signal Process. (2009).
25. C. P. Robert and G. Casella, *Monte-Carlo Statistical Methods*, Springer Texts in Statistics (Springer, 2000).
26. B. R. Hunt, "A matrix theory proof of the discrete convolution theorem," IEEE Trans. Autom. Control **AC-19**, 285–288 (1971).
27. M. Calder and R. A. Davis, "Introduction to Whittle (1953) 'The analysis of multiple stationary time series'," Breakthroughs in Statistics **3**, 141–148 (1997).
28. P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods* (Springer-Verlag, 1991).
29. B. R. Hunt, "Deconvolution of linear systems by constrained regression and its relationship to the Wiener theory," IEEE Trans. Autom. Control **AC-17**, 703–705 (1972).
30. K. Mardia, J. Kent, and J. Bibby, *Multivariate Analysis* (Academic, 1992), Chap. 2, pp. 36–43.
31. C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," IEEE Trans. Image Process. **2**, 296–310 (1993).
32. D. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ. Press, 2003).
33. R. E. Kass and L. Wasserman, "The selection of prior dis-

tributions by formal rules," J. Am. Stat. Assoc. **91**, 1343–1370 (1996).

34. E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge Univ. Press, 2003).
35. S. Lang, *Real and Functional Analysis* (Springer, 1993).
36. P. Brémaud, *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues*, Texts in Applied Mathematics 31 (Springer, 1999).
37. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).
38. P. Lalanne, D. Prévost, and P. Chavel, "Stochastic artificial retinas: algorithm, optoelectronic circuits, and implementation," Appl. Opt. **40**, 3861–3876 (2001)
39. H. R. Künsch, "Robust priors for smoothing and image restoration," Ann. Inst. Stat. Math. **46**, 1–19 (1994).
40. P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," IEEE Trans. Image Process. **6**, 298–311 (1997).
41. J.-F. Giovannelli, "Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function," IEEE Trans. Image Process. **17**, 16–26 (2008).
42. D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," IEEE Trans. Image Process. **4**, 932–946 (1995).
43. X. Descombes, R. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood," IEEE Trans. Image Process. **8**, 954–963 (1999).
44. G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (Addison-Wesley, 1972).

T. Rodet, F. Orieux, J.-F. Giovannelli et A. Abergel, « Data inversion for over-resolved spectral imaging in astronomy », *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, n°5, pp. 802–811, octobre 2008.

Data inversion for over-resolved spectral imaging in astronomy

T. Rodet, F. Orieux, J.-F. Giovannelli and A. Abergel

Abstract—**We present an original method for reconstructing a three-dimensional object having two spatial dimensions and one spectral dimension from data provided by the infrared slit spectrograph on board the Spitzer Space Telescope. During acquisition, the light flux is deformed by a complex process comprising four main elements (the telescope aperture, the slit, the diffraction grating and optical distortion) before it reaches the two-dimensional sensor.**

The originality of this work lies in the physical modelling, in integral form, of this process of data formation in *continuous variables*. The inversion is also approached with *continuous variables* in a semi-parametric format decomposing the object into a family of Gaussian functions. The estimate is built in a deterministic regularization framework as the minimizer of a quadratic criterion.

These specificities give our method the power to over-resolve. Its performance is illustrated using real and simulated data. We also present a study of the resolution showing a 1.5-fold improvement relative to conventional methods.

Index Terms—**inverse problems, bayesian estipmation, over-resolved imaging, spectral imaging, irregular sampled, interpolation, IRS Spitzer.**

I. INTRODUCTION

Since the end of the 1970's, infra-red to millimetric observations of the sky from space have brought about a revolution in practically all fields of astrophysics. It has become possible to observe distant galaxies, perform detailed physicochemical studies of interstellar matter. Observations in the far infra-red are now possible thanks to new types of sensors (Ge:Ga Si:As semiconductors and bolometer arrays). The properties of these new sensors encouraged the astrophysicists of the Institut d'Astrophysique Spatiale (IAS) to work with researchers at the Laboratoire des Signaux et Systèmes (L2S) in order to develop suitable processing methods. The spectral imaging work presented here was carried out in the framework of this cooperative effort. The aim is to reconstruct an over-resolved object having two spatial dimensions (α, β)[1] and one spectral dimension λ. Data provided by the Infrared Spectrograph (IRS) [1] on board the american Spitzer Space Telescope launched in 2003 are used to illustrate our work. Several sets of two-dimensional data are delivered by the Spitzer Science Center (SSC), each set being the result of an acquisition for a given satellite pointing direction. The data were acquired using a slit spectrograph, the operation of which is described in detail in part II. This instrument is located in the focal plane of the telescope. When the telescope is pointed towards a region of the sky, the spectrograph slit selects a direction of space α. The photon flux is then dispersed perpendicularly to the slit direction with a diffraction grating. The measurement is made using a two-dimensional sensor. A signal containing one spatial dimension α and the spectral dimension λ is thus obtained. The second spatial dimension β is obtained by scanning the sky (modifying telescope pointing). This scanning has two notable characteristics:

- it is irregular, because the telescope control is not perfect;
- it is, however, measured with sub-pixel accuracy (eighth of a pixel).

In addition, for a given pointing direction, : the telescope optics, the slit width and the sensor integration limit the spatial resolution while the grating, the slit and the sensor integration limit the spectral resolution. The specificity of systems of this type is that the width of impulse response depends on the wavelength. A phenomenon of spectral also aliasing appears for the shortest wavelengths. Finally, the scanning results in irregular sampling along the spatial direction β. The problem to be solved is thus one of inverting the spectral aliasing (i.e. the over-resolution) using a finite number of discrete data provided by a complex system. The solution proposed here is based on precise modelling of the instrument and, in particular, the integral equations containing the continuous variables (α, β and λ) of the optics and sensing system. The model input is naturally a function of these continuous variables $\phi(\alpha, \beta, \lambda)$ and the output is a finite set $\boldsymbol{y}$ of discrete data items.

The approach used for solving the inverse problem, i.e. reconstructing an object having three continuous variables from the discrete data

- comes within the framework of regularization by penalization;
- uses a semi-parametric format where the object is decomposed into a family of functions.

There is a multitude of families of functions available (possibly forming a basis of the chosen functional space). The most noteworthy are Shannon, Fourier, wavelet and pixel-indicator families or those of spline, Gaussian Kaiser-Bessel, etc. Work on 3D tomographic reconstruction has used a family of Kaiser-Bessel functions having spherical symmetry in order to calculate the projections more efficiently [2], [3], [4], [5]. In a different domain, the signal processing community has been working on the reconstruction of over-resolved images

---

T. Rodet, F. Orieux, J.-F. Giovannelli are with the Laboratoire des Signaux et Systèmes (CNRS – Supélec – UPS), Supélec, Plateau de Moulon, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette Cedex, France. E-mail: `orieux,rodet,giova@lss.supelec.fr`.

A. Abergel is with the Institut d'Astrophysique Spatiale, batiment 121, Université Paris Sud 11, 91405 Orsay Cedex. E-mail: `abergel@ias.u-psud.fr`.

[1] In this paper, the spatial dimensions are angles in radian

from a series of low resolution images [6]. A generic direct model can be described [6], starting with a continuous scene, to which are applied k shift or deformation operators including at least one translation. This step gives k deformed, high-resolution images. A convolution operator modelling the optics and sensor cells is then applied to each of the images. After subsampling, the k low-resolution images that constitute the data are obtained. Recent work on the direct model has mainly concerned modelling the shift by introducing a rotation of the image [7], [8] and a magnifying factor [9]. Other works have modelled the shift during sensor integration by modifying the convolution operator [10]. To the best of our knowledge, in most works, the initial discretization step is performed on pixel indicators [11], [7], [10], [8], [6], [12]. On this point, a noteworthy contribution has been made by P. Vandewalle *et al.* who discretize the scene on a truncated discrete Fourier basis [13]. However, their decomposition tends to make the images periodic leading to create artefacts on the image side. Thus we have decided not to use this approach. Recently, the problem of X-ray imaging spectroscopy has been solved in the Fourier space [14], but each spectral component has been estimated independantly.

The two major contributions of our paper are (1) the modelling of the measurement system as a whole with continuous variables and (2) the continuous variable decomposition of the three dimensional object over a family of Gaussian functions. Modelling with continuous variables enables a faithful description to be made of the physical phenomena involved in the acquisition and avoids to carry out any prior data interpolation. In our case, computing the model output requires six integrals (two for the response of the optics, two for the grating response, and two for the sensor integration) and the choice of a Gaussian family allows five of these six integrals to be explicitly stated. Our paper is organised as follows:

Part II describes the continuous model of the instrument comprising: the diffraction at the aperture, the truncation by the slit, the response of the grating, the distortion of the light flux, the sensor integration and the scanning of the sky. In part III, the object with continuous variables is decomposed over a family of Gaussian functions. The aperture and grating responses are approximated by Gaussian functions. This part concludes with the obtention of a precise, efficient model of the measuring system. The inverse problem is solved in a regularized framework in part IV. Finally, part V gives an evaluation of the resolving power of the method and a comparison with a standard data co-addition method using both simulated and real data.

II. CONTINUOUS DIRECT MODEL

The aim of the instrument model is to reproduce the data, $\boldsymbol{y}$, acquired by the spectral imager from a flux $\phi(\alpha, \beta, \lambda)$ of incoherent light. Fig. 1 illustrates the instrument model for one acquisition (the telescope remains stationary). To simplify, we present the scanning procedure in section II-D. First, we have the response of the primary mirror (aperture), which corresponds to a convolution. Second, there is a truncation due to a rectugular slit. Third, a grating disperses the light. Finnally, the sensor integration provides the discrete data $\boldsymbol{y}$. Distortion of the luminous flux is modelled in the sensor integration.

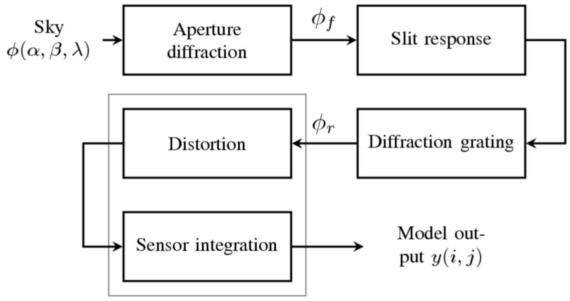


Fig. 1. Block diagram of the direct model for one acquisition: from a continuously defined sky ϕ to a discrete output $\boldsymbol{y}$ describing the data. The flux ϕ_f is a convolution of the flux ϕ and the PSF of the primary miror. ϕ_f is truncated by a rectangular slit and is dispersed by the gratting. Finally, the sensor provide a discrete output $\boldsymbol{y}$.

A. Aperture diffraction

Under some hypotheses, the propagation of a light wave which passes through an aperture is determined by FRESNEL diffraction [15] and the result in the focal plane is a convolution of the input flux ϕ with the Point Spread Function (PSF) h_a illustrated in Fig. 2 for a circular aperture. This PSF, which is a low pass filter, has a width proportional to the wavelength of the incident flux. For a circular aperture, it can be written:

$$h_a(\alpha, \beta, \lambda) = A \left[2 \frac{J_1(\pi D \sqrt{\alpha^2 + \beta^2} / \lambda)}{\pi D \sqrt{\alpha^2 + \beta^2} / \lambda} \right]^2 \tag{1}$$

where J_1 is the first order Bessel function of the first kind, A is an amplitude factor and D is the diameter of the mirror.

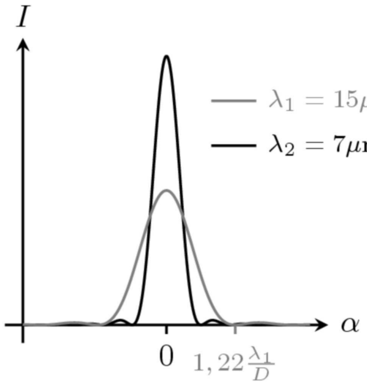


Fig. 2. Profile of an *Airy* disk (PSF for a circular aperture) for two wavelengths.

The flux in the focal plane, ϕ_f, is written in integral form:

$$\phi_f(\alpha', \beta', \lambda) = \iint_{\alpha, \beta} \phi(\alpha, \beta, \lambda)\, h_a(\alpha - \alpha', \beta - \beta', \lambda)\, \mathrm{d}\alpha\, \mathrm{d}\beta \tag{2}$$

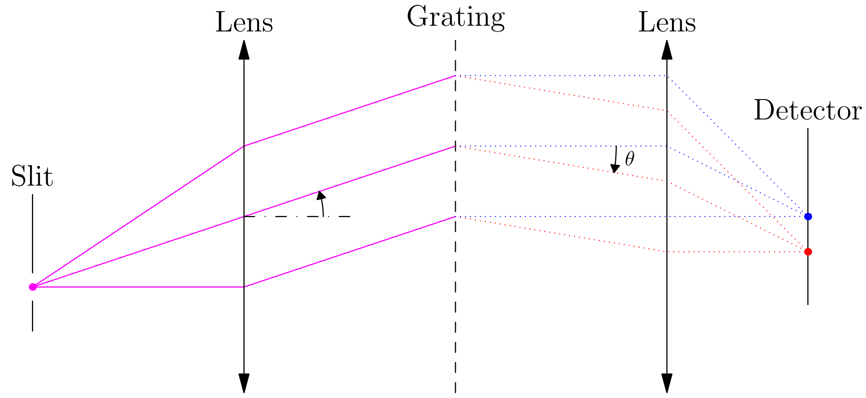


Fig. 3. Optical scheme of IRS instrument: the slit is on the focal plan of the telescope SPITZER. The gratting disperses the light and the detector collectes the dispersed flux

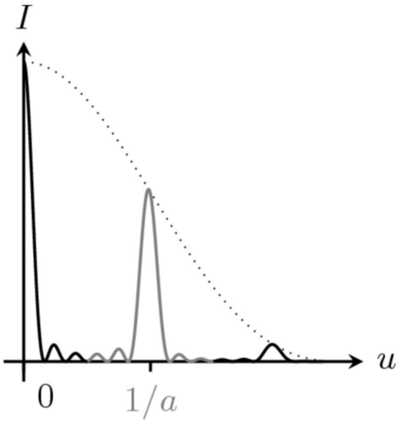


Fig. 4. Diffraction grating response. The grey curve corresponds to the first mode. In reality, the response width is smaller.

B. Slit and diffraction grating

a) Slit: Ideally, the slit and grating enable the dispersion of the wavelengths in spatial dimension β previously “suppressed” by the slit (see Fig. 3). In practice, the slit cannot be infinitely narrow because the flux would be zero. The slit thus has a width γ of about two pixels.

b) Diffraction grating: Ideally, the grating gives a diffracted wave with an output angle θ linearly dependent on the wavelength λ (see Fig. 3). In a more accurate model, the dependencies become more complex. Let us introduce a variable u in order to define an invariant response h_r of the system [16].

$$u = \frac{\sin\theta - \sin\beta'}{\lambda} \approx \frac{\sin\theta - \beta'}{\lambda} \tag{3}$$

where β' is the angle of incidence of the wave on the grating , and $|\beta'| \leq \gamma/2$ where γ is the angular slit width (5.6 arcseconds). The response of the grating centred on mode m ($m = 0, 1, \ldots$) can, with some approximations, be written as the square of a cardinal sine centred on m/a [16].

$$h_r(\theta, \beta', \lambda) = B \operatorname{sinc}^2\left(\pi L(u - m/a)\right) \tag{4}$$

where L is the width of the grating and a the grid step (distance between two grooves). This response centred on the first mode ($m = 1$) is plotted in Fig. 4.

As the flux is an incoherent light source, the expression for the signal at the output of the grating is written in the form of an integral over β' and λ:

$$\phi_r(\alpha', \theta) = \int_\lambda \int_{|\beta'| \leq \gamma/2} \phi_f(\alpha', \beta', \lambda) h_r(\beta', \lambda, \theta) \mathrm{d}\beta' \mathrm{d}\lambda \tag{5}$$

where l is the slit width.

C. Sensor integration

Once the flux has passed through the grating and the wavelengths have been dispersed according to θ, the light flux is focused on the sensor composed of square detectors. The sensor is simply modelled by integrating the flux ϕ_r on square areas of side d. The flux is integrated along the direction α, which is not modified by the diffraction grating, and the dimension θ, a combination of β and λ, to obtain the discrete values

$$y(i,j) = \int_{id}^{(i+1)d} \int_{jd + e_{ij}^1}^{(j+1)d + e_{ij}^2} \phi_r(\alpha', \theta) \mathrm{d}\alpha' \mathrm{d}\theta. \tag{6}$$

The integration limits are modified by the terms e_{ij}^n in order to take into account the data distortion as illustrated in Fig. 5.

D. Scanning procedure of the sky

In a direction parallel to the slit width a scanning procedure (illustrated Fig. 6) is applied. This scanning procedure is composed of Q acquisitions. Between the first and the q^{th} acquisitions, the instrument is moved by $\Delta_\alpha(q)$ (resp. $\Delta_\beta(q)$) in the direction α (resp. β). To taking into account the motion of the instrument, we substitue $\phi(\alpha, \beta, \lambda)$ for

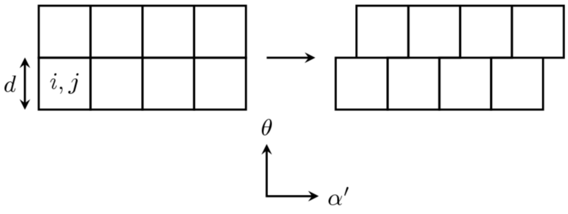


Fig. 5. Modelling the distortion: sensor integration limits are shifted according to the dimensions α'.

$\phi(\alpha - \Delta_\alpha(q), \beta - \Delta_\beta(q), \lambda)$ in the previous equations. In practice, we fix the α axis in the direction of the slit and the β axis perpendicular to the slit (see Fig. 6). In consequence, $\Delta_\alpha(q)$ is equal to zero.

E. Complete model

By combining expression (1), (2), (5),(4) and (6), we obtain a continuous direct model in the form

$$y(i,j,q) = \mathcal{A}\int_{id}^{(i+1)d}\int_{jd+e^1_{ij}}^{(j+1)d+e^2_{ij}}\int_\lambda\int_{|\beta'|\leq\gamma/2}\int_\alpha\int_\beta \phi(\alpha-\Delta_\alpha(q),\beta-\Delta_\beta(q),\lambda)h_a(\alpha-\alpha',\beta-\beta',\lambda)\mathrm{d}\alpha\mathrm{d}\beta\, h_r(\theta,\beta',\lambda)\mathrm{d}\beta'\mathrm{d}\lambda\mathrm{d}\alpha'\mathrm{d}\theta \quad (7)$$

where $\mathcal{A}$ is a scale factor.

The equation (7) can rewritten:

$$y(i,j,q) = \int_\alpha\int_\beta\int_\lambda \phi(\alpha,\beta,\lambda)h^{i,j,q}_{tot}(\alpha,\beta,\lambda)\mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\lambda \quad (8)$$

with

$$h^{i,j,q}_{tot}(\alpha,\beta,\lambda) = \mathcal{A}\int_{id}^{(i+1)d}\int_{jd+e^1_{ij}}^{(j+1)d+e^2_{ij}}\int_{-\gamma/2}^{\gamma/2} h_a(\alpha-\alpha'-\Delta_\alpha(q),\beta-\beta'-\Delta_\beta(q),\lambda)h_r(\theta,\beta',\lambda)\mathrm{d}\alpha'\mathrm{d}\theta\mathrm{d}\beta' \quad (9)$$

We have been developed a model relying the continuous sky $\phi(\alpha,\beta,\lambda)$ and discrete data $\boldsymbol{y}$. Our model is linear not-shift-invariant, because the aperture response and the grating response depend on the wavelength.

III. DECOMPOSITION OVER A FAMILY AND GAUSSIAN APPROXIMATION

In the previous part, we have seen that obtaining the output from the model requires the six integrals of equation (7) to be calculated. The estimation of $\hat{\phi}$ in $L^2(\mathbb{R})$ by inversion of this model is quite tricky, so we prefer to decompose the object over a family of functions. As we can see in the introduction, a lot of such decomposition functions can be used. The most traditional are Fourier bases, wavelets, cardinal sines, splines and pixel indicators. The choice does not have any great influence on the final result if the continuous object is decomposed over a sufficiently large number of functions. We therefore chose our decomposition functions in such a way as to reduce the computing time for the instrument model. First, we chose the α axis in the direction of the slit and the β axis perpendicular to the slit (see Fig. 6). Second, we have two spatial variables (α,β) and one spectral variable λ, so to simplify the calculus, we chose decomposition functions that are separable into (α,β) and λ. Thrid, the object is convolved by the response of the optics, which has circular symmetry. So we choose functions possessing the same circular symmetry in order to make this calculation explicit. Finally, the slit and the grating have an impact in the β direction only (5), which motivates us to choose functions that are separable into α and β. These considerations led us to choose Gaussian functions along the spatial directions. Finally the complexity of the λ dependence encouraged us to choose Dirac impulses for the spectral direction.

A. Decomposition over a family of Gaussian functions

The flux ϕ is a continous function decomposed over a family of separable functions:

$$\begin{aligned}\phi(\alpha,\beta,\lambda) &= \sum_k\sum_l\sum_p x(k,l,p)\\ &\quad \Pi(\alpha-kT_\alpha)\,\Phi(\beta-lT_\beta)\,\Gamma(\lambda-pT_\lambda)\\ &= \sum_k\sum_l\sum_p x(k,l,p)\Psi_{k,l,p}(\alpha,\beta,\lambda) \quad (10)\end{aligned}$$

where $x(k,l,p)$ are the decomposition coefficients, T_α, T_β and T_λ are the sampling steps, and with:

$$\Pi(\alpha)\Phi(\beta) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2}\frac{\alpha^2+\beta^2}{\sigma^2}\right) \quad (11)$$

$$\Gamma(\lambda) = \delta(\lambda) \quad (12)$$

With such decomposition, the inverse problem becomes one of estimating a finite number of coefficients $x(k,l,p)$ from discrete data $y(i,j,q)$. By combining equations (8) and (10), we obtain:

$$y(i,j,q) = \sum_k\sum_l\sum_p x(k,l,p)\int_\alpha\int_\beta\int_\lambda \Psi_{k,l,p}(\alpha,\beta,\lambda)\, h^{i,j,q}_{tot}(\alpha,\beta,\lambda)\mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\lambda \quad (13)$$

If the $y(i,j,q)$ and $x(k,l,p)$ are gathered in vectors $\boldsymbol{y}$ and $\boldsymbol{x}$[2] respectively, the equation (13) can be formalized as a vector matrix product.

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} \quad (14)$$

with each component of the matrix $\boldsymbol{H}$ is calculated using the integral part of the equation (13). The $n(k,l,p)$-th column of matrix $\boldsymbol{H}$ constitutes the output when the model is used with the n-th decomposition function ($\Psi_{k,l,p}$). The model output for $\Psi_{k,l,p}$ is calculated in the next two sections.

---

[2] In this paper, we use the following convention: bold, lower-case variables represent vectors and bold, upper-case variables represent matrices.

B. Impulse responses approximated by Gaussian functions

1) Approximation of the PSF : Equation (2) comes down to convolutions of a squared Bessel function and Gaussians. This integral is not explicit and, in order to carry out the calculations, the PSF is approximated by a Gaussian

$$\tilde{h}_a(\alpha,\beta,\lambda) = \frac{1}{2\pi\sigma_\lambda^2}\exp\left(-\frac{1}{2}\frac{\alpha^2+\beta^2}{\sigma_\lambda^2}\right)$$

with a standard deviation σ_λ depending on the wavelength. Indeed, the Bessel functions cross zero at the first time in $1.22\lambda/D$. σ_λ is determined numerically by minimizing the quadratic error between the Gaussian kernel and the squared Bessel function, which gives for our instrument $\sigma_\lambda \approx \lambda/2$. The relative quadratic error $err_{L2} = \|\text{Bessel} - \text{Gaussian}\|_2^2 \,/\, \|\text{Bessel}\|_2^2$ is equal to 0.15% for our instrument. If we caculate the relative absolute error $err_{L1} = \|\text{Bessel} - \text{Gaussian}\|_1 \,/\, \|\text{Bessel}\|_1$, we obtain 5 %. We can conclude that most of the energy of the squared bessel function is localized in the primary lobe. Another advantage of using the Gaussian approximation is that the convolution kernel is separable into α and β. Finally, the result of the convolution of two Gaussian functions is a standard one and is also a Gaussian:

$$\tilde{h}_a(\alpha,\beta,\lambda)\star\Pi(\alpha)\Phi(\beta) = \frac{1}{2\pi(\sigma_\lambda^2+\sigma^2)}\exp\left(-\frac{\alpha'^2+\beta'^2}{2(\sigma_\lambda^2+\sigma^2)}\right) \tag{15}$$

2) Approximation of the grating response: The presence of the slit means that integral (5) is bounded over β' and is not easily calculable. Since the preceding expressions use Gaussian functions, we approximate the squared cardinal sine by a Gaussian to make the calculations easier:

$$\operatorname{sinc}^2\left(\pi L\left(\frac{\sin\theta-\beta'}{\lambda}-\frac{m}{a}\right)\right) \approx \frac{1}{\sqrt{2\pi}\lambda\sigma_s}\exp\left(-\frac{1}{2}\frac{(\frac{\sin\theta-\beta'}{\lambda}-\frac{m}{a})^2}{\sigma_s^2}\right) \tag{16}$$

σ_s is determined numerically by minimizing the quadratic error between the Gaussian kernel and the squared cardinal sine, which gives for our instrument $\sigma_s \approx 25.5\ \text{m}^{-1}$. The relative errors made are larger than the bessel case ($err_{L2} = 0.43\%, err_{L1} = 10.7\%$), but this Gaussian approximation of the grating response allows the flux ϕ_r coming out of the grating to be known explicitly.

The error introduced here is larger than for the Gaussian approximation of the PSF described in the previous section. However, our goal is to have a good model of the spatial dimension of the array. Furthermore, with respect to the current method, the fact of taking the response of the grating into consideration, even as an approximation, is already a strong improvement.

$$\begin{aligned}
\phi_r(\alpha',\theta) &= \int_\lambda\int_{-\gamma/2}^{\gamma/2}\int_\alpha\int_\beta \Pi(\alpha-\alpha_k)\Phi(\beta-\beta_l-\Delta_\beta(q)) \\
&\quad \Gamma(\lambda-\lambda_p)\tilde{h}_a(\alpha-\alpha',\beta-\beta',\lambda)\tilde{h}_r(\theta,\beta',\lambda) \\
&\quad \mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\beta'\mathrm{d}\lambda \\
&= \mathcal{A}\exp\left(-\frac{(\alpha'-\alpha_k)^2}{2(\sigma_\lambda^2+\sigma^2)}\right)\exp\left(-\frac{(\sin\theta-\nu)^2}{2\Sigma^2}\right) \\
&\quad \times\left[\operatorname{erf}\left(\frac{\gamma/2-\mu}{\Sigma'\sqrt{2}}\right)-\operatorname{erf}\left(\frac{-\gamma/2-\mu}{\Sigma'\sqrt{2}}\right)\right]
\end{aligned} \tag{17}$$

with

$$\left\{\begin{array}{lcc}
\mathcal{A} & = & \dfrac{1}{4\pi}\sqrt{\dfrac{1}{(\sigma_\lambda^2+\sigma^2)(\sigma_\lambda^2+\sigma^2+\lambda_p^2\sigma_s^2)}} \\
\Sigma^2 & = & (\sigma_\lambda^2+\sigma^2)+\lambda_p^2\sigma_s^2 \\
\nu & = & \dfrac{m\lambda_p}{a}+\beta_l+\Delta_\beta(q) \\
\Sigma' & = & \dfrac{\sqrt{\sigma_\lambda^2+\sigma^2}\lambda_p\sigma_s}{\sqrt{\sigma_\lambda^2+\sigma^2+\lambda_p^2\sigma_s^2}} \\
\mu & = & \dfrac{\left(\sin\theta-\frac{m\lambda_p}{a}\right)(\sigma_\lambda^2+\sigma^2)}{(\sigma_\lambda^2+\sigma^2)+\lambda_p^2\sigma_s^2} \\
\operatorname{erf}(a) & = & \dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^a e^{-t^2}\mathrm{d}t
\end{array}\right.$$

In equation (17), it can be seen that ϕ_r is separable into α' and θ. Let us introduce the functions f and g such that:

$$\phi_r(\alpha',\theta) = \mathcal{A}f(\alpha')g(\theta) \tag{18}$$

C. Sensor integration

First, we calculate the sensor integration in the α' direction.

$$\begin{aligned}
&\int_{jd+e_{ij}^1}^{(j+1)d+e_{ij}^2} f(\alpha')\mathrm{d}\alpha' = \\
&\mathcal{K}\left[\operatorname{erf}\left(\frac{(j+1)d+e_{ij}^2-\alpha_k}{\sqrt{2(\sigma_\lambda^2+\sigma^2)}}\right)-\operatorname{erf}\left(\frac{jd+e_{ij}^1-\alpha_k}{\sqrt{2(\sigma_\lambda^2+\sigma^2)}}\right)\right]
\end{aligned} \tag{19}$$

with $\mathcal{K} = \sqrt{\pi(\sigma_\lambda^2+\sigma^2)/2}$

The integral of g is calculated numerically as the presence of erf functions in equation (17) does not allow analytical calculations.

We obtain the expression for the n-th column of matrix $\boldsymbol{H}$, which now contains only a single integral:

$$\begin{aligned}
&y(i,j,q) = \mathcal{A}\mathcal{K}\int_{id}^{(i+1)d} g(\theta)\mathrm{d}\theta \\
&\times\left[\operatorname{erf}\left(\frac{(j+1)d+e_{ij}^2-\alpha_k}{\sqrt{2(\sigma_\lambda^2+\sigma^2)}}\right)-\operatorname{erf}\left(\frac{jd+e_{ij}^1-\alpha_k}{\sqrt{2(\sigma_\lambda^2+\sigma^2)}}\right)\right]
\end{aligned} \tag{20}$$

Using expression (20), the elements of matrix $\boldsymbol{H}$ are precomputed relatively rapidly. Thanks to the sparsity of the matrix $\boldsymbol{H}$ to calculate the model output of Eq. (14).

IV. INVERSION

The previous sections build the relationship (14) between the object coefficients and the data: it describes a complex instrumental model but remains linear. The problem of input (sky) reconstruction is a typical inverse problem and the literarure on the suject is abundant.

The proposed inversion method resorts to linear processing. It is based on conventional approaches described in books such [17], [18] or more recently [19]. In this framework, the reader may also consider [20], [21] for inversion based on specific decomposition. These methods rely on a quadratic criterion

$$J(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2 + \mu_{\alpha\beta}||\boldsymbol{D}_{\alpha\beta}\boldsymbol{x}||^2 + \mu_\lambda||\boldsymbol{D}_\lambda\boldsymbol{x}||^2 \,. \quad (21)$$

It involves a least squares term and two penality terms concerning the differences between neighbouring coefficients: one for the two spatial dimensions and one for the spectral dimension. They are weighted by $\mu_{\alpha\beta}$ and μ_λ, respectively. The estimate $\hat{\boldsymbol{x}}$ is chosen as the minimizer of this criterion. It is thus explicit and linear with respect to the data:

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{H}^{\mathrm{t}}\boldsymbol{H} + \mu_{\alpha\beta}\boldsymbol{D}_{\alpha\beta}^{\mathrm{t}}\boldsymbol{D}_{\alpha\beta} + \mu_\lambda\boldsymbol{D}_\lambda^{\mathrm{t}}\boldsymbol{D}_\lambda\right)^{-1}\boldsymbol{H}^{\mathrm{t}}\boldsymbol{y} \quad (22)$$

and depends on the two regularization parameters $\mu_{\alpha\beta}$ and μ_λ.

Remark 1 — *This estimator can be interpreted in a Bayesian framework [22] based on Gaussian models for the errors and the object. As far as the errors are concerned, the model is a white noise. As far as the object is concerned, the model is correlated and the inverse of the correlation matrix is proportional to* $\mu_{\alpha\beta}\boldsymbol{D}_{\alpha\beta}^{\mathrm{t}}\boldsymbol{D}_{\alpha\beta} + \mu_\lambda\boldsymbol{D}_\lambda^{\mathrm{t}}\boldsymbol{D}_\lambda$*, i.e., it is a Gauss markov field. In this framework, the estimate maximizes the a posteriori law.*

Remark 2 — *Many works in the field of over-resolved reconstruction concern edge preserving priors [23], [11], [7], [24], [25], [12]. In our application here, smooth interstellar dust clouds are under study, so preservation of edges is not appropriate. For the sake of simplicity of implementation, we chose a Gaussian object prior.*

The minimizer $\hat{\boldsymbol{x}}$ given by relation (22) is explicit but, in practice, it cannot be calculated on standard computers, because the matrix to be inverted is too large. The solution $\hat{\boldsymbol{x}}$ is therefore computed by a numerical optimization algorithm. Practically, the optimization relies on a standard gradient descent algorithm [26], [27]. More precisely, the direction descent is a approximate conjugate gradient direction [28] and the optimal step of descent is used. Finally, we initialise the method with zero ($\boldsymbol{x} = 0$).

V. RESULTS

As we have presented in part III the α and β axis are fix (see Fig. 6, right). The real data is composed of 23 acquisitions having a spatial dimension α' and a spectral dimension θ of wavelength between 7.4 and 15.3 μm (each acquisition is an image composed of 38×128 detector cells, see Fig. 6, left). Between two acquisitions, the instrument is moved by half a slit width in the β direction. Fig. 6, right, shows the scanning procedure applied to the Horsehead nebula [29].

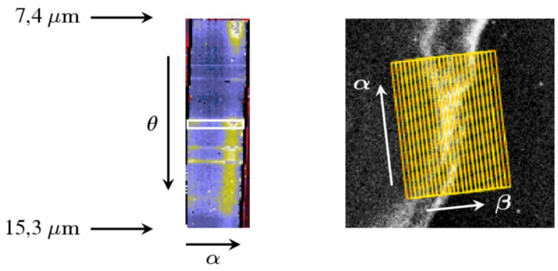


Fig. 6. Acquisition. The left-hand image represents the data acquired for one pointing position: the vertical axis shows the spectral dimension θ and the horizontal axis is the spatial dimension α'. The slit is represented schematically by a rectangle in the middle. The right-hand image illustrates the scanning strategy in the β direction.

Our results (Fig. 8(b), solid line on Fig. 9 and Fig. 10(b)) can be compared with those obtained with the conventional processing (Fig. 8(c), dotted line on Fig. 9 and Fig. 10(a)). For the conventional processing (described in Compiègne et al. 2007 [29]) an image of the slit is simply extracted for each wavelength from the data taken after each acquisition (e.g. left panel of Fig. 6) and projected and co-added on the output sky image, without any description of the instrument properties.

A. Simulated data

In our first experiment, we reconstruct data simulated using our direct model. We choose an object with the same spatial morphology and the same spectral content as the Horsehead nebula (see Fig. 8(a)). However, in order to tune the regularization coefficient, we perform a large number of reconstructions. Thus, we need to simulate a problem smaller than in our real case. The data are composed of 14 acquisitions, and the virtual detector contained 18×40 pixels. We choose to reconstruct a volume with 15870 gaussians distributed on a cartesian grid $23 \times 23 \times 30$. Finally, we add to the output of the model a white Gaussian noise with the same variance as the real data.

The results contain a set of 30 images (see Fig. 7). Fig. 8(b) and solid line on Fig.9 illustrate our result for one wavelength (8.27 μm) and one pixel, respectively. The image computed with our method (Fig. 8(b)) appears comparable to the true image (Fig. 8(a)), while the image computed with the conventional processing (Fig. 8(c)) is smoother. A comparison of solid line and dotted line in Fig. 9 clearly also shows that our method provides a spectrum comparable to the true spectrum, while the peaks obtained with the conventional processing are too broad.

Our sky estimation depends on the regularization coefficients $\mu_{\alpha\beta}$ and μ_λ. We tune this parameters by minimizing numerically the quadratic error between the estimated object and the real object were selected. In this experiment we obtain $\mu_{\alpha\beta} = 0.01,\ \mu_\lambda = 0.005$.

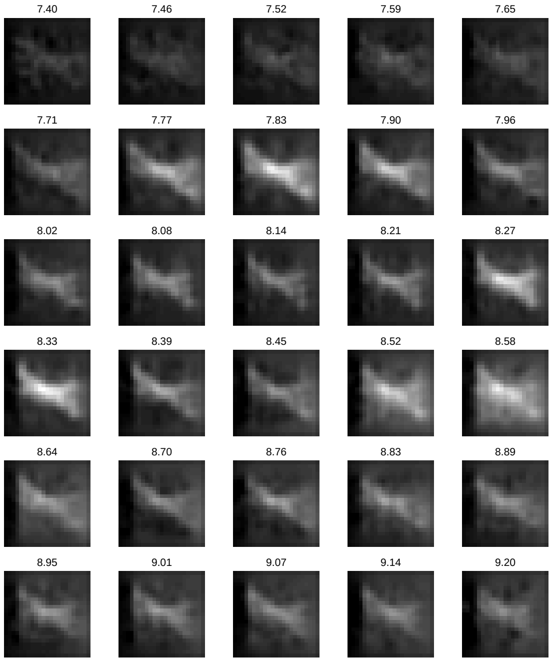
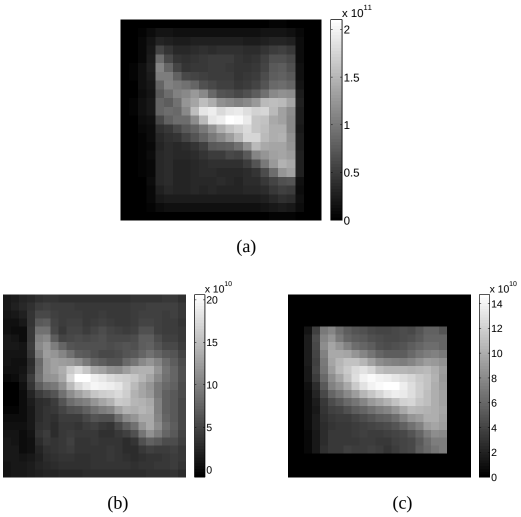


Fig. 7. Set of 30 images of our reconstruction from simulated data. Each image corresponds to one wavelength for 7.4 to 9.2 μm with a step of 0.062 μm.



(a)

(b)

(c)

Fig. 8. Image at $\lambda = 8.27\,\mu\text{m}$:, (a) simulated sky, (b) image estimated by our method, (c) image estimated by a conventional method.

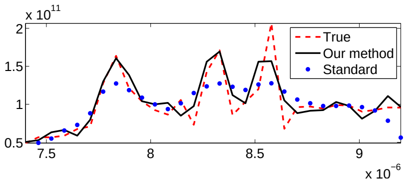


Fig. 9. Spectrum of a pixel. The curves abscissa is the wavelenght in meters: solid line: simulated sky, dashed line: our method, dotted line: conventional method.

B. *Real data*

Eeal data contain 23 acquisitions composed of 38×128 values. To obtain a over-resolved reconstruction, we describe our volume with 587264 gaussians destributes on a cartesian grid $74 \times 62 \times 128$. The spatial (α, β) sampling step is equal to a quater slit width, and the spectral dimension is uniformly sampled between the wavelength 7.4 and 15.3 μm. The reconstruction is computed after setting the regularization coefficients $\mu_{\alpha\beta}$ and μ_λ empirically. Too low a value for these coefficients produces an unstable method and a quasi explosive reconstruction. Too high a value produces images that are visibly too smooth. A compromise found by trial and error led us to $\mu_{\alpha\beta} = 0.3$ and $\mu_\lambda = 0.7$. The ratio between $\mu_{\alpha\beta}$ and μ_λ is also based on our simulation. However, we cannot compare the regularization coefficients between the simulated and the real case, since the size of the problem modifies the weigth of the norm in the Eq. (21). Pratically, we take large value for the regularization coefficients, and we gradually reduce the value up that we are seeing noise.

Our results (Fig. 10(b) and 11(b)) can be compared with those obtained with (Fig. 10(a) and 11(a), from [29]). A comparison of Fig. 10(a) and 10(b) clearly shows that our approach provides more resolved images that bring out more structures than the conventional approach. Note, in particular, the separation of the two filaments on the left part of the Fig. 10(b) obtained with our method, which remains invisible after conventional processing. For comparison, Fig. 10(c) shows the same object observed with the Infrared Array Camera (IRAC) of the Spitzer Space Telescope which has a better native resolution since it observes at a shorter wavelength (4.5 μm). Here the same structures are observed, providing a strong argument in favour of the reality of the results provided by our method.

A more precise analysis is done in section V-C. It provides a quantitative evaluation of the resolution.

Finally, the spectra reconstructed by our method (Fig. 11(b)) have a resolution slightly better than the one reconstructed by the conventional method (Fig. 11(a)). The peaks characterizing the observed matter (gas and dust) are well positioned, narrower and with a greater amplitude. However, ringing effects appear at the bases of the strongest peaks (Fig. 11(b)). They could be explained by an over-evaluation of the width σ_s of the response of the grating, or by the gaussian approximation.

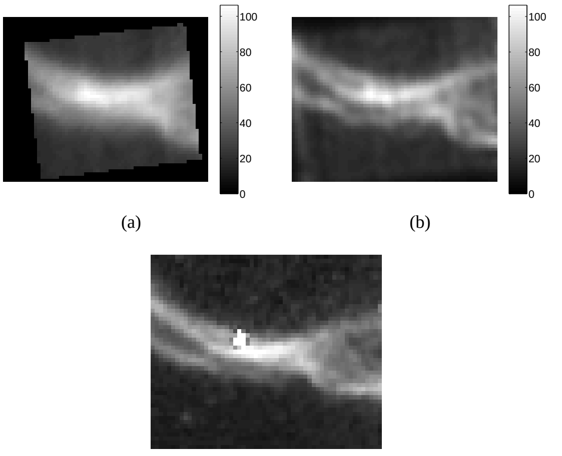


Fig. 10. Reconstruction of a sky ϕ representing the Horsehead nebula: (a) image estimated at 11.37 μm by the conventional method (b) image estimated at 11.37 μm by our method, (c) image obtained with the Infrared Array Camera IRAC on board the Spitzer Space Telescope having better resolution at 4.5 μm.

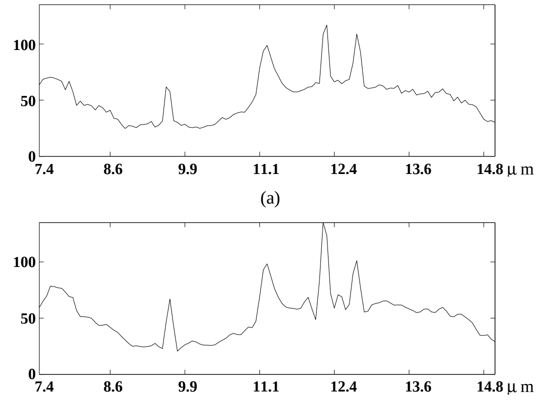


Fig. 11. Spectrum of the center point of the Fig. 10(b): (a) spectrum estimated with the conventional method (b) spectrum estimated with our method.

C. *Study of resolving power of our approach*

This part is devoted to numerical quantification of the gain in angular resolution provided with our method, using the Rayleigh criterion, which is frequently used by astrophysicists: for the smaller resolvable detail, the first minimum of the image of one point source coincides with the maximum of another. In practice, two point sources with the same intensity and a flat strectrum are considered to be separated if the minimal flux between the two peaks is lower than 0.9 times the flux at the peak positions. The resolution is studied in the β direction only as this is the direction in which the subslit scan is performed.

Two point sources are injected, at positions β_1 and β_2, respectively (see Fig. 13, top). The corresponding data are simulated, and the reconstruction $\hat{\phi}(\beta)$ is performed. As explained above, the two point sources are considered to be separated if $\hat{\phi}([\beta_1+\beta_2]/2) \leq 0.9 \times \hat{\phi}(\beta_1)$. The resolution is defined as the difference $\delta = \beta_2 - \beta_1$ at which the two point sources start to be separated.

Point sources are simulated for a set of differences δ between 2.4 and 5.4 arcseconds and simulations are performed in the configuration of the real data (signal to noise ratio, energy of the data). Moreover, we use the regularization parameters $\mu_{\alpha\beta}$ and μ_λ determined in section V-A. A number of reconstructions has been obtained. The ratio between the values of the reconstructed function at β_1 and $(\beta_1+\beta_2)/2$ is calculated as a function of the difference δ between the two peaks. Results are shown in Fig. 12.

The computed resolution is 3.4 arcseconds (see Fig. 12(a)) and 5 arcseconds (see Fig. 12(b)) for our method and the conventional method, respectively. Fig. 13 illustrates this gain in angular resolution. In the left column on Fig. 13 ($\delta = 3.4$) corresponds to the limit of resolution of our method. In this case, the peak is not separated with the conventional method (Fig. 13 (d)). In the middle column on Fig. 13, our algorithm clearly separates the peak (Fig. 13(h)) and not the conventional method (Fig. 13(e)). In the right column, we observe a stain with our method is smaller than the conventional method. Our method increases the resolution by a factor 1.5.

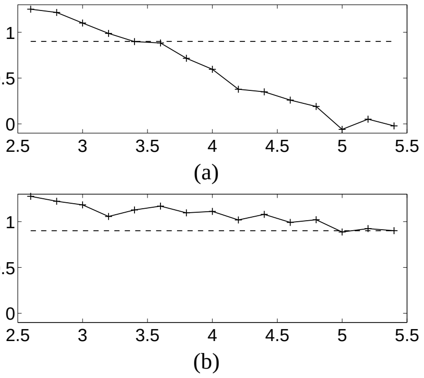
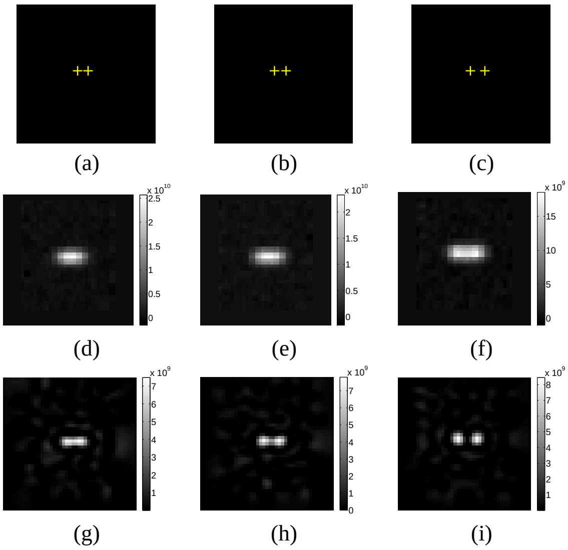


Fig. 12. Resolution of our method: the curve represents the ratio of the intensity at one peak to the intensity between the two peaks as a function of the distance between the peaks in arcseconds. The resolution is read at crossing of this curve and the dotted line (the ratio is 0.9). (a) Results obtained with our method. (b) Results obtained with the conventional method.

VI. CONCLUSIONS

We have developed an original method for reconstructing the over-resolved 3D sky from data provided by the IRS instrument. This method is based on:

1) a continuous variable model of the instrument based on a precise integral physical description,
2) a decomposition of the continuous variable object over a family of Gaussian functions, which results in a linear, semi-parametric relationship,
3) an inversion in the framework of deterministic regularization based on a quadratic criterion minimized by a gradient algorithm.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 13. Two peaks reconstruction for different δ : (a-c) Visualisation of the peaks position for $\delta = 3.4$, 4 and 5.4 arcsecond resp., (d-f) reconstruction with the conventional method for $\delta = 3.4$, 4 and 5.4 arcsecond resp., (g-i) reconstruction with our over-resolution method for $\delta = 3.4$, 4 and 5.4 arcsecond respectively.

The first results on real data show that we are able to evidence spatial structures not detectable using conventional methods. The spatial resolution is improved by a factor 1.5. This factor should increase using data with a motion between two acquisitions smaller than the half a slit width.

In the future, we plan to design highly efficient processing tools using our approach in particular for the systematic processing of the data which will be taken with the next generation of infrared to milimeter space observatory (HERSCHEL, PLANCK, ...).

REFERENCES

[1] J. R. Houck *et al.*, "The infrared spectrograph (IRS) on the Spitzer space telescope," *ApJS*, vol. 154, pp. 18–24, September 2004.

[2] R. M. Lewitt, "Multidimensional digital image representations using generalized Kaiser-Bessel window functions," *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 1834–1848, October 1990.

[3] ——, "Alternative to voxels for image representation in iterative reconstruction algorithms," *Physics in Medicine and Biology*, vol. 37, pp. 705–716, 1992.

[4] S. Matej and R. M. Lewitt, "Practical considerations for 3-D image reconstruction using spherically symmetric volume elements," *IEEE Trans. Medical Imaging*, vol. 15, pp. 68–78, January 1996.

[5] A. Andreyev, M. Defrise, and C. Vanhove, "Pinhole SPECT reconstruction using blobs and resolution recovery," *IEEE Trans. Nuclear Sciences*, vol. 53, pp. 2719–2728, October 2006.

[6] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Trans. Signal Processing Mag.*, pp. 21–36, May 2003.

[7] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1646–1658, December 1997.

[8] ——, "Superresolution restoration of an image sequence: Adaptive filtering approach," *IEEE Trans. Image Processing*, vol. 8, no. 3, pp. 387–395, March 1999.

[9] G. Rochefort, F. Champagnat, G. Le Besnerais, and J.-F. Giovannelli, "An improved observation model for super-resolution under affine motion," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3325–3337, November 2006.

[10] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1064–76, August 1997.

[11] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1621–1633, December 1997.

[12] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic methods for joint resgistration, restoration, and interpolation of multiple undersampled images," *IEEE Trans. Image Processing*, vol. 15, no. 1, pp. 201–13, January 2006.

[13] P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli, "Super-resolution from unregistrered and totally aliased signals using subspace methods," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3687–703, July 2007.

[14] M. Piana, A. M. Massone, G. J. Hurford, M. Prato, A. G. Emslie, E. P. Kontar, and R. A. Schwartz, "Electron flux imaging of solar flares through regularized analysis of hard X-ray source visibilities," *The Astrophysical Journal*, vol. 665, pp. 846–855, August 2007.

[15] J. W. Goodman, *Introduction à l'optique de Fourier et à l'holographie*. Paris, France: Masson, 1972.

[16] J.-P. Pérez, *Optique, fondements et applications*. Dunod, 2004.

[17] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.

[18] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[19] J. Idier, Ed., *Bayesian Approach to Inverse Problems*. London: ISTE Ltd and John Wiley & Sons Inc., 2008.

[20] M. Bertero, C. De Mol, and E. R. Pike, "Linear inverse problems with discrete data. I: General formulation and singular system analysis," *Inverse Problems*, vol. 1, pp. 301–330, 1985.

[21] ——, "Linear inverse problems with discrete data: II. stability and regularization," *Inverse Problems*, vol. 4, p. 3, 1988.

[22] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structure and problems," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-37, no. 12, pp. 2024–2036, December 1989.

[23] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 996–1011, June 1996.

[24] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient superresolution image reconstruction algorithm," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 573–83, April 2001.

[25] F. Humblot and A. Mohammad-Djafari, "Super-resolution using hidden Markov model and bayesian detection estimation framework," *EURASIP Journal on Applied Signal Processing, special issue on Super-Resolution Imaging Analysis, Algorithms, and Applications*, no. 10, pp. 1–16, 2006.

[26] D. P. Bertsekas, *Nonlinear programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[27] J. Nocedal and S. J. Wright, *Numerical Optimization*, ser. Series in Operations Research. New York: Springer Verlag, 2000.

[28] E. Polak, *Computational methods in optimization*. New York, NY: Academic Press, 1971.

[29] M. Compiègne, A. Abergel, L. Verstraete, W. T. Reach, E. Habart, J. D. Smith, F. Boulanger, and C. Joblin, "Aromatic emission from the ionised mane of the Horsehead nebula," *Astronomy & Astrophysics*, vol. 471, pp. 205–212, 2007.

Thomas Rodet was born in Lyon, France, in 1976. He received the Doctorat degree at Institut National Polytechnique de Grenoble, France, in 2002.

He is presently assistant professor in the Département de Physique at Université Paris-Sud 11 and researcher with the Laboratoire des Signaux et Systèmes (CNRS - Supélec - UPS). He is interested in tomography methods and Bayesian methods for inverse problems in astrophisical problems (inversion of data taken from space observatory: Spitzer, Herschel, SoHO, STEREO).

François Orieux was born in Angers, France, in 1981. He graduated from the École Supérieure d'Électronique de l'Ouest in 2006 and obtain the M. S. degree from the University of Rennes 1. He currently pursuing the Ph. D. degree at the Université Paris-Sud 11. His research interests are statistical image processing and Bayesian methods for inverse problems.

Jean-François Giovannelli was born in Béziers, France, in 1966. He graduated from the École Nationale Supérieure de l'Électronique et de ses Applications in 1990. He received the Doctorat degree in physics at Université Paris-Sud, Orsay, France, in 1995.

He is presently assistant professor in the Département de Physique at Université Paris-Sud and researcher with the Laboratoire des Signaux et Systèmes (CNRS - Supélec - UPS). He is interested in regularization and Bayesian methods for inverse problems in signal and image processing. Application fields essentially concern astronomical, medical and geophysical imaging.

Alain Abergel was born in Paris, France, in 1959. He received the Doctorat degree in astrophysics at the Paris 6 University, France, in 1987, and is presently Professor in Physics and Astronomy at the Paris-Sud University. He is interested in the interstellar medium of galaxies, and uses data taken from long wavelength space observatories (IRAS, COBE, ISO, Spitzer, ...). He is Co-investigator for the Spectral and Photometric Imaging Receiver (SPIRE) instrument on-board the European Space Agency's Herschel Space Observatory.

N. Barbey, F. Auchère, **T. Rodet** et J.-C. Vial, « A time-evolving 3D method dedicated to the reconstruction of solar plumes and results using extreme ultra-violet data », *Solar Physics*, vol. 248, n°2, pp. 409–423, avril 2008.

Solar Physics
DOI: 10.1007/●●●●●-●●●-●●●-●●●●-●

A Time-Evolving 3D Method Dedicated to the Reconstruction of Solar Plumes and Results Using Extreme Ultra-Violet Data

Nicolas Barbey[*†] · **Frédéric Auchère**[*] ·
Thomas Rodet[†] · **Jean -Claude Vial**[*]

Received: 9 May 2007 / Accepted: 24 January 2008

Abstract An important issue in the tomographic reconstruction of the solar poles is the relatively rapid evolution of the polar plumes. We demonstrate that it is possible to take into account this temporal evolution in the reconstruction. The difficulty of this problem comes from the fact that we want a 4D reconstruction (three spatial dimensions plus time) while we only have 3D data (2D images plus time). To overcome this difficulty, we introduce a model that describes polar plumes as stationary objects whose intensity varies homogeneously with time. This assumption can be physically justified if one accepts the stability of the magnetic structure. This model leads to a bilinear inverse problem. We describe how to extend linear inversion methods to these kinds of problems. Studies of simulations show the reliability of our method. Results for SOHO/EIT data show that we are able to estimate the temporal evolution of polar plumes in order to improve the reconstruction of the solar poles from only one point of view. We expect further improvements from STEREO/EUVI data when the two probes will be separated by about 60°.

1. Introduction

A method known as solar rotational tomography has been used to retrieve the 3D geometry of the solar corona (Frazin 2000; Frazin and Janzen 2002). This method assumes the stability of the structures during the time necessary to acquire the data. Since we generally have only one point of view at our disposal, about 15 days are required to have data for half a solar rotation at the poles. Here, we focus our study on solar polar plumes. They are bright, radial, coronal ray structures located at the solar poles in regions of open magnetic field. The study of plumes is of great interest since it may be the key to the understanding of the acceleration of the fast component of the solar wind (Teriaca *et al.*, 2003). However the three-dimensional shape of these structures is poorly known and different assumptions have

---

[*] Institut d'Astrophysique Spatiale, Université Paris-Sud, Orsay, France
email: nicolas.barbey@ias.u-psud.fr, frederic.auchere@ias.u-psud.fr, jean-claude.vial@ias.u-psud.fr

[†] Laboratoire des Signaux et Systèmes, Supéléc, Gif-sur-Yvette, France
email: nicolas.barbey@lss.supelec.fr, thomas.rodet@lss.supelec.fr

been made, *e.g.* Gabriel *et al.*, 2005; Llebaria, Saez, and Lamy, 2002. The plumes are known to evolve with a characteristic time of approximately 24 hours on spatial scales typical of *Extreme ultra-violet Imaging Telescope* (SOHO/EIT) data (2400 km) (DeForest, Lamy, and Llebaria, 2001). Consequently the stability assumption made in rotational tomography fails. Fortunately, the *Solar TErestrial RElations Observatory* (STEREO) mission consists of two identical spacecraft $\mathrm{STEREO}_\mathrm{A}$ and $\mathrm{STEREO}_\mathrm{B}$ which take pictures of the Sun from two different points of view. With the SOHO mission still operating, this results in three,simultaneous points of view. Three viewpoints help to improve the reconstruction of the plumes, but they are still not enough to use standard tomographic algorithms. The problem is underdetermined and consequently one has to add *a priori* information in order to overcome the lack of information. This leads to challenging and innovative signal analysis problems. There are different ways to deal with underdetermination depending on the kind of object to be reconstructed. Interestingly the field of medical imaging faces the same kind of issues. In cardiac reconstruction, authors make use of the motion periodicity in association with a high redundancy of the data (Grass *et al.*, 2003; Kachelriess, Ulzheimer, and Kalender, 2000). If one can model the motion as an affine transformation, and if one assumes that we know this transformation, one can obtain an analytic solution (Ritchie *et al.*, 1996; Roux *et al.*, 2004).

In solar tomography, the proposed innovative approaches involve the use of additional data such as magnetic-field measurements in the photosphere (Wiegelmann and Inhester, 2003) or data fusion (Frazin and Kamalabadi, 2005). Attempts have been made by Frazin *et al.* (2005) to treat temporal evolution using Kalman filtering.

Since polar plumes have apparently a local, rapid, and aperiodic temporal evolution, we developed as in the previously referenced work, a model based on the specifics of the object we intend to reconstruct (preliminary results can be found in Barbey *et al.*, (2007). Plumes have an intensity which evolves rapidly with time, but their position can be considered as constant. This hypothesis is confirmed by previous studies of the plumes such as DeForest, Lamy, and Llebaria (2001). The model is made up of an invariant morphological part ($\boldsymbol{x}$) multiplied by a gain term ($\boldsymbol{\theta}_t$) that varies with time. Only one gain term is associated with each plume in order to constrain the model. So we assume that the position of each plume in the scene is known. This model is justified if we consider polar plumes to be slowly evolving magnetic structures in which plasma flows.

Thanks to this model we can perform time-evolving three-dimensional tomography of the solar corona using only extreme ultra-violet images. Furthermore, there is no complex, underlying physical model. The only assumptions are the smoothness of the solution, the area-dependant evolution model, and the knowledge of the plume position. These assumptions allow us to consider a temporal variation of a few days, while assuming only temporal smoothness would limit variations to the order of one solar rotation (about 27 days). To our knowledge, the estimation of the temporal evolution has never been undertaken in tomographic reconstruction of the solar corona.

We first explain our reconstruction method in a Bayesian framework (Section 2). We then test the validity of our algorithm with simulated data (Section 3). An example of a reconstruction on real SOHO/EIT data is shown in Section 4. Results are discussed in Section 5. We conclude in Section 6.

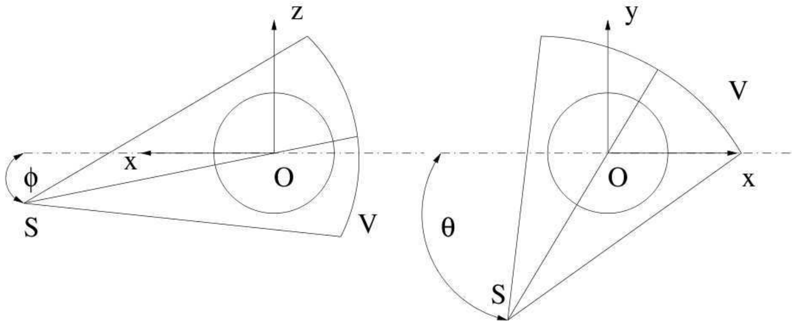


Figure 1. Scheme of the data acquisition geometry. $(\mathbf{O}; x, y, z)$ defines the Carrington heliocentric frame of reference. S is the spacecraft considered. ϕ is the latitude, and θ the longitude of this spacecraft. V is the virtual detector.

2. Method

Tomographic reconstruction can be seen as an inverse problem, the direct problem being the acquisition of data images knowing the emission volume density of the object (Section 2.1). If the object is evolving during the data acquisition, the inverse problem is highly underdetermined. So our first step is to redefine the direct problem thanks to a reparametrization, in order to be able to define more constraints (Section 2.2). Then, we place ourselves in the Bayesian inference framework in which data and unknowns are considered to be random variables. The solution of the inverse problem is chosen to be the maximum *a posteriori* (Section 2.3). This leads to a criterion that we minimize with an alternate optimization algorithm (Section 2.4).

2.1. Direct Problem

The geometrical acquisition is mathematically equivalent to a conical beam data acquisition with a virtual spherical detector (see Figure 1). In other words, the step between two pixels vertically and horizontally is constant in angle. The angle of the full field of view is around 45 minutes. In order to obtain an accurate reconstruction, we take into account the exact geometry, which means the exact position and orientation of the spacecraft relatively to Sun center. We approximate integration of the emission in a flux tube related to a pixel by an integration along the line of sight going through the middle of that pixel. We choose to discretize the object in the usual cubic voxels. $\boldsymbol{x}$ is a vector of size N containing the values of all voxels. In the same way, we define the vector of data $\boldsymbol{y}_t$ of size M at time t. Since the integration operator is linear, the projection can be described by a matrix $\boldsymbol{P}_t$. We choose $\boldsymbol{n}_t$ to be an additive noise:

$$\boldsymbol{y}_t = \boldsymbol{P}_t \boldsymbol{x}_t + \boldsymbol{n}_t, \forall t \in [1, ..., T] \tag{1}$$

$\boldsymbol{P}_t$ is the projection matrix at time t of size $M \times N$ which is defined by the position and the orientation of the spacecraft at this time. Its transpose is the backprojection matrix. Note that a uniform sampling in time is not required. In order to be able to handle large problems with numerous well-resolved data images and a large reconstruction cube, we chose not to store the whole projection matrix. Instead, we perform the projection operation ($\boldsymbol{Px}$) or its transpose each time it is needed at each iteration. Thus, we need a very efficient algorithm. We developed a code written in C which performs the projection operation. It makes use of the geometrical parameters given in the data headers in order to take into account the exact geometry (conicity, position, and orientation of the spacecraft). To keep this operation fast, we implemented the Siddon algorithm (Siddon, 1985). It allows a fast projection or backprojection in the case of cubic voxels (Cartesian grid). Since we focus on a small region at the poles, we consider that we do not need to use a spherical grid which would require a more time-consuming projection algorithm.

We take into account the fact that the field of view is conical. Despite the fact that the acquisition is very close to the parallel acquisition geometry, it is sufficient to introduce an error of several voxels of size 0.01 solar radius from one side to the other of a three solar radii reconstructed cube.

2.2. Modeling of the Temporal Evolution

With this model, the inverse problem is underdetermined since we have at most three images at one time and we want to reconstruct the object with its temporal evolution. In order to do so, we first redefine our unknowns to separate temporal evolution from spatial structure. We introduce a new set of variables $\boldsymbol{g}_t$ of size N describing the temporal evolution and require that $\boldsymbol{x}$ does not depend on time:

$$\boldsymbol{y}_t = \boldsymbol{P}_t(\boldsymbol{x} \circ \boldsymbol{g}_t) + \boldsymbol{n}_t \tag{2}$$

with $\circ$ being the term-by-term multiplication of vectors. This operator is clearly bilinear. However, this model would increase the number of variables excessively. So, we need to introduce some other kind of *a priori* into our model. We make the hypothesis that all of the voxels of one polar plume have the same temporal evolution:

$$\boldsymbol{g}_t = \boldsymbol{L}\boldsymbol{\theta}_t \tag{3}$$

The matrix $\boldsymbol{L}$ of size $N \times P$ (P being the number of areas) localizes areas where the temporal evolution is identical. Each column of $\boldsymbol{L}$ is the support function of one of the plumes. We would like to stress that in our hypothesis, those areas do not move relative to the object. In other words, $\boldsymbol{L}$ does not depend on time. Localizing these areas defines $\boldsymbol{L}$ and only leaves PT variables to estimate. We redefined our problem in a way that limits the number of parameters to estimate but still allows many solutions. Furthermore, the problem is linear in $\boldsymbol{x}$ knowing $\boldsymbol{\theta}$ and linear in $\boldsymbol{\theta}$ knowing $\boldsymbol{x}$. It will simplify the inversion of the problem as we shall see later. Note, however that the uniqueness of a solution $(\boldsymbol{x}, \boldsymbol{\theta})$ is not guaranteed with bilinearity despite its being guaranteed in the linear case. This example shows that A can be chosen arbitrarily without changing the closeness to the data: $\boldsymbol{x} \circ \boldsymbol{g} = (A\boldsymbol{x}) \circ (A^{-1}\boldsymbol{g})$, where A is a real constant. Introducing an *a priori* of closeness to $\mathbf{1}$ for $\boldsymbol{\theta}$ would allow us to deal with this indeterminacy in principle. But note that this indeterminacy is not critical since the physical quantity of interest is only the product $\boldsymbol{x} \circ \boldsymbol{g}$. Féron, Duchêne, and Mohammad-Djafari (2005) present a method which solves a bilinear inversion problem in the context of microwave tomography.

We do not deal with the estimation of the areas undergoing evolution, but we assume in this paper that the localization is known. This localization can be achieved using other sources of information, *e.g.* stereoscopic observations. We expect to be able to locate the areas using some other source of information.

We can regroup the equations of the direct problem. We have two ways to do so, each emphasizing the linearity throughout one set of variables.

$$\boldsymbol{y} = \boldsymbol{U_x}\boldsymbol{\theta} + \boldsymbol{n}$$
$$\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_T \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}_1\boldsymbol{X}\boldsymbol{L} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{P}_T\boldsymbol{X}\boldsymbol{L} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_T \end{pmatrix} + \begin{pmatrix} \boldsymbol{n}_1 \\ \vdots \\ \boldsymbol{n}_T \end{pmatrix} \tag{4}$$

with $\boldsymbol{X} = diag(\boldsymbol{x})$, the diagonal matrix defined by $\boldsymbol{x}$. $\boldsymbol{x}$ is of size N, $\boldsymbol{y}$ and $\boldsymbol{n}$ are of size MT, $\boldsymbol{\theta}$ is of size PT and $\boldsymbol{U_x}$ is of size $MT \times PT$.

Similarly,

$$\boldsymbol{y} = \boldsymbol{V_\theta}\boldsymbol{x} + \boldsymbol{n}$$
$$\text{with } \boldsymbol{V_\theta} = \begin{pmatrix} \boldsymbol{P}_1\text{diag}(\boldsymbol{L}\boldsymbol{\theta}_1) & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{P}_T\text{diag}(\boldsymbol{L}\boldsymbol{\theta}_T) \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_d \\ \vdots \\ \boldsymbol{I}_d \end{pmatrix} \tag{5}$$

with $\boldsymbol{I}_d$ the identity matrix of size $M \times M$. $\boldsymbol{V_\theta}$ is of size $MT \times N$.

2.3. Inverse Problem

In Bayes' formalism, solving an inverse problem consists in knowing the *a posteriori* (the conditional probability density function of the parameters, the data being given). To do so we need to know the likelihood (the conditional probability density function of the data knowing the parameters) and the *a priori* (the probability density function of the parameters). An appropriate model is a Gaussian, independent, identically distributed (with the same variance) noise $\boldsymbol{n}$. The likelihood function is deduced from the noise statistic:

$$f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}, \sigma_n, \mathcal{M}) = K_1 \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{U_x}\boldsymbol{\theta}\|^2}{2\sigma_{\boldsymbol{n}}^2}\right) \tag{6}$$

$\mathcal{M} = [\boldsymbol{P}, \boldsymbol{L}]$ describing our model (the projection algorithm and parameters and the choice of the plume position). We assume that the solution is smooth spatially and temporally, so we write the *a priori* as follows:

$$f(\boldsymbol{x}|\sigma_{\boldsymbol{x}}) = K_2 \exp\left(-\frac{\|\boldsymbol{D}_r\boldsymbol{x}\|^2}{2\sigma_{\boldsymbol{x}}^2}\right) \text{ and } f(\boldsymbol{\theta}|\sigma_{\boldsymbol{\theta}}) = K_3 \exp\left(-\frac{\|\boldsymbol{D}_t\boldsymbol{\theta}\|^2}{2\sigma_{\boldsymbol{\theta}}^2}\right) \tag{7}$$

$\boldsymbol{D}_r$ and $\boldsymbol{D}_t$ are discrete differential operators in space and time. Bayes' theorem gives us the *a posteriori* law if we assume that the model $\mathcal{M}$ is known as well as the hyperparameters $\mathcal{H} = [\sigma_{\boldsymbol{n}}, \sigma_{\boldsymbol{x}}, \sigma_{\boldsymbol{\theta}}]$:

$$f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}, \mathcal{H}, \mathcal{M}) = \frac{f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}, \sigma_{\boldsymbol{n}}, \mathcal{M}) f(\boldsymbol{x}|\sigma_{\boldsymbol{x}}) f(\boldsymbol{\theta}|\sigma_{\boldsymbol{\theta}})}{f(\boldsymbol{y}|\mathcal{H}, \mathcal{M})} \tag{8}$$

We need to choose an estimator. It allows us to define a unique solution instead of having a whole probability density function. We then choose to define our solution as the maximum *a posteriori.* which is given by:

$$(\boldsymbol{x}^{\mathrm{MAP}}, \boldsymbol{\theta}^{\mathrm{MAP}}) = \arg\max_{\boldsymbol{x},\boldsymbol{\theta}} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta},\sigma_{\boldsymbol{n}},\mathcal{M}) f(\boldsymbol{x}|\sigma_{\boldsymbol{x}}) f(\boldsymbol{\theta}|\sigma_{\boldsymbol{\theta}}) \tag{9}$$

since $f(\boldsymbol{y}|\mathcal{M})$ is a constant. Equation (9) can be rewritten as a minimization problem:

$$(\boldsymbol{x}^{\mathrm{MAP}}, \boldsymbol{\theta}^{\mathrm{MAP}}) = \arg\min_{\boldsymbol{x},\boldsymbol{\theta}} J(\boldsymbol{x},\boldsymbol{\theta}) \tag{10}$$

with:

$$J(\boldsymbol{x},\boldsymbol{\theta}) = -2\sigma_n \log f(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y},\mathcal{M},\mathcal{H}) = \|\boldsymbol{y} - \boldsymbol{U}_{\boldsymbol{x}}\boldsymbol{\theta}\|^2 + \lambda\|\boldsymbol{D}_r\boldsymbol{x}\|^2 + \mu\|\boldsymbol{D}_t\boldsymbol{\theta}\|^2 \tag{11}$$

$\lambda = \frac{\sigma_n^2}{\sigma_x^2}$ and $\mu = \frac{\sigma_n^2}{\sigma_\alpha^2}$ are user-defined hyperparameters.

The equivalence of Equations (9) and (10) has been proved by Demoment (1989).

Note that the solution does not have to be very smooth. It mostly depends on the level of noise since noise increases the underdetermination of the problem as it has been shown by the definition of λ and μ.

2.4. Criterion Minimization

The two sets of variables $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are very different in nature. However, thanks to the problem's bilinearity, one can easily estimate one set while the other is fixed. Consequently we perform an iterative minimization of the criterion, and we alternate minimization of $\boldsymbol{x}$ and $\boldsymbol{\theta}$. At each step n we perform:

$$\boldsymbol{\theta}^{n+1} = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{x}^n,\boldsymbol{\theta}) \text{ and } \boldsymbol{x}^{n+1} = \arg\min_{\boldsymbol{x}} J(\boldsymbol{x},\boldsymbol{\theta}^{n+1}) \tag{12}$$

The two subproblems are formally identical. However, $\boldsymbol{\theta}$ is much smaller than $\boldsymbol{x}$. This is of the utmost practical importance since one can directly find the solution on $\boldsymbol{\theta}$ by using the pseudo-inverse method. $\boldsymbol{x}$ is too big for this method, and we have to use an iterative scheme such as the conjugate-gradient to approximate the minimum. These standard methods are detailed in Appendices A and B.

2.5. Descent Direction Definition and Stop Threshold

We choose to use an approximation of the conjugate-gradient method that is known to converge much more rapidly than the simple gradient method (Nocedal and Wright, 2000; Polak and Ribière, 1969).

$$\begin{array}{rcl} \boldsymbol{d}^{p+1} & = & \boldsymbol{d}^p + b^p \nabla_{\boldsymbol{x}} J|_{\boldsymbol{x}=\boldsymbol{x}^p} \\ b^p & = & \frac{\left\langle \nabla_{\boldsymbol{x}} J|_{\boldsymbol{x}=\boldsymbol{x}^p}, \nabla_{\boldsymbol{x}} J|_{\boldsymbol{x}=\boldsymbol{x}^{p-1}} \right\rangle}{\|\nabla_{\boldsymbol{x}} J|_{\boldsymbol{x}=\boldsymbol{x}^{p-1}}\|^2} \end{array} \tag{13}$$

Since the minimum is only approximately found, we need to define a threshold which we consider to correspond to an appropriate closeness to the data in order to stop the iterations. Since the solution is the point at which the gradient is zero, we choose this threshold for updating x:

$$\text{mean}_{\boldsymbol{x}\in[\boldsymbol{x}^p,\boldsymbol{x}^{p-1},\boldsymbol{x}^{p-2}]} \|\nabla_{\boldsymbol{x}} J\|^2 < S_x \tag{14}$$

For the global minimization, the gradient is not computed, so we choose:

$$\text{mean}_{[n,n-1,n-2]} \|(\boldsymbol{x}_n, \boldsymbol{\theta}_n) - (\boldsymbol{x}_{n-1}, \boldsymbol{\theta}_{n-1})\ \|^2 < S_G \tag{15}$$

Note that this way to stop the iteration allows one to define how close one wants to be to the solution: if the difference between two steps is below this threshold, it is considered negligible. The algorithm can be summarized as shown in Figure 2.

```
initialize : $\boldsymbol{x} = 0$ and $\boldsymbol{\theta} = 1$
while Equation (15) is satisfied
    $\boldsymbol{x}$ minimization:
    while Equation (14) is satisfied
        * compute gradient at $\boldsymbol{x}_n$ with Equation (20)
        * compute descent direction with Equation (13)
        * compute optimum step with Equation (22)
        * update $\boldsymbol{x}$ with Equation (23)
    endwhile
    $\boldsymbol{\theta}$ minimization:
        * compute the matrix $\boldsymbol{U}_{\boldsymbol{x}^n}^T \boldsymbol{U}_{\boldsymbol{x}^n}$ and the vector $\boldsymbol{U}_{\boldsymbol{x}^n}^T \boldsymbol{y}$
        * inverse the matrix $\boldsymbol{U}_{\boldsymbol{x}^n}^T \boldsymbol{U}_{\boldsymbol{x}^n} + \mu \boldsymbol{D}_r^T \boldsymbol{D}_r$
        * compute Equation (19)
endwhile
```

Figure 2. Tomographic Reconstruction with Temporal Evolution Algorithm

3. Method Validation

In order to validate the principle of our method and test its limits, we simulate an object containing some plumes with temporal evolution and try to extract it from the data.

3.1. Simulation Generation Process

We generate an emission cube with randomly-placed, ellipsoidal plumes with a Gaussian shape along each axis:

$$E_p = A \exp\left(-\frac{1}{2}\left[\frac{\boldsymbol{r}.\boldsymbol{u}_\phi)}{a}\right]^2 - \frac{1}{2}\left[\frac{\boldsymbol{r}.\boldsymbol{u}_{\phi+\frac{\pi}{2}}}{b}\right]^2\right) \tag{16}$$

The plumes evolve randomly but smoothly by interpolating over a few randomly generated points. Once the object is generated, we compute a typical set of 60 images equally spaced along 180° using our projector algorithm. A Gaussian random noise is added to the projections with a signal to noise ratio (SNR) of five. The simulation parameters are summarized in Table 1.

Table 1. Simulation Definition: Plumes Parameters

Plume Number	Semimajor Axis a	Semiminor Axis b	ϕ	x_0	y_0	Intensity (A)
1	4.8	4.2	1.2	29	29	329
2	5.6	3.3	1.1	23	33	430
3	5.2	4.8	0.1	40	42	723

3.2. Results Analysis

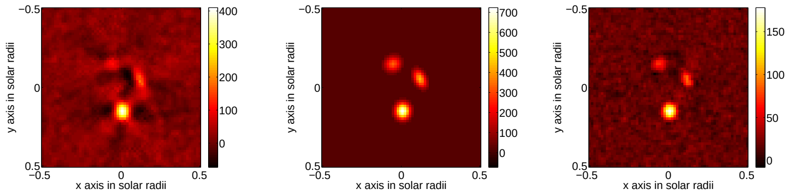
We now compare our results (Figure 3) with a filtered back-projection (FBP) algorithm. This method is explained by Natterer (1986) and Kak and Slaney (1987).

By comparing the simulation and the reconstruction in Figure 3, we can see the quality of the temporal evolution estimation. The shape of the intensity curves is well reproduced except for the first plume in the first ten time steps where the intensity is slightly underestimated. This corresponds to a period when plume 1 is hidden behind plume 2. Thus, our algorithm attributes part of the plume 1 intensity to plume 2. Let us note that this kind of ambiguity will not arise in the case of observations from multiple points of view such as STEREO/EUVI observations. The indeterminacy of the problem is due to its bilinearity discussed in Section 2.2. This allows the algorithm to attribute larger values to the $\boldsymbol{\theta}$ parameters and to compensate by decreasing the corresponding $\boldsymbol{x}$. This is not a drawback of the method since it allows discontinuities between plumes and interplumes. The only physical value of interest is the product $\boldsymbol{x} \circ \boldsymbol{g}$.

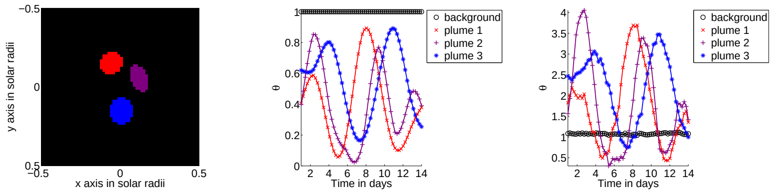
Figure 4 shows the relative intensity of the plumes at different times. One can compare with the reconstruction. One way to quantify the quality of the reconstruction is to compute the distance (quadratic norm of the difference) between the real object and the reconstructed one. Since the FBP reconstruction does not actually correspond to a reconstruction at one time, we evaluate the minimum of the distances at each time. We find it to be 3000. This is to be compared with a value of 700 with our algorithm, which is much better.

Table 2. Simulation Definition: Geometric Parameters

cube size (solar radii)	cube number of voxels	pixel size (radians)	projection number of pixels
$1 \times 1 \times 0.05$	$64 \times 64 \times 4$	$5 \times 10^{-5} \times 5 \times 10^{-5}$	128×8

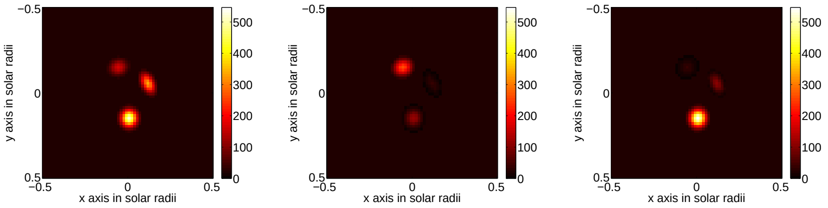


(a) $\boldsymbol{x}$ with a FBP algorithm (b) Simulated $\boldsymbol{x}$ (c) $\boldsymbol{x}$ with our algorithm

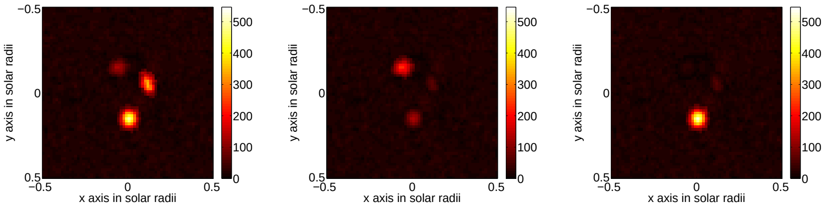


(d) Chosen temporal areas (e) Simulated $\boldsymbol{\theta}$ (f) $\boldsymbol{\theta}$ with our algorithm

Figure 3. Comparison of a standard FBP method (a), the real simulated object (b), and the object reconstructed with our method (c). The object is reconstructed using 60 projections regularly spaced over 180°. The areas of homogeneous temporal evolution (e) are the same in the simulation and the reconstruction. We associated one time per projection to define $\boldsymbol{\theta}$ in the simulation (e) and our reconstruction (f). The time scale is in days assuming a rotation speed of half a rotation in 14 days. $\boldsymbol{x}$ is the spatial distribution of the emission density volume. $\boldsymbol{\theta}$ is a gain representing the emission variation over time. Except for the FBP reconstruction, only the product $\boldsymbol{x} \circ \boldsymbol{\theta}$ has physical dimensions. The spatial scales are given in solar radii and centered on the solar axis of rotation. (a), (b) and (c) are slices of 3D cubes at the same $z = 0.1\,R_{\odot}$. Emission densities (arbitrary units) are scaled in the color bars in the right-end side of (a), (b), (c).



(a) Simulation at $10\Delta T$ (b) Simulation at $30\Delta T$ (c) Simulation at $50\Delta T$

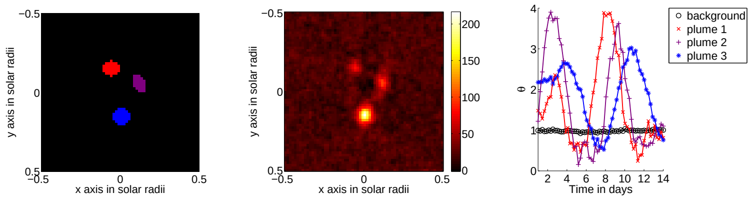


(d) Result at $10\Delta T$ (e) Result at $30\Delta T$ (f) Result at $50\Delta T$

Figure 4. Comparison of $\boldsymbol{x} \circ \boldsymbol{g}$ simulated and reconstructed at different times. ΔT is the time between two data images (5.6 hours). Distances are in solar radii. Values represent the volume emission density. All of this images are slices of 3D cubes at the same $z = 0.1 R_{\odot}$.

Table 3. Simulation Definition: Other Parameters

SNR	λ	μ	$S_{\boldsymbol{x}}$	S_G
5	2×10^{-2}	100	2×10^{-2}	1×10^{-2}



(a) New choice of areas (b) $\boldsymbol{x}$ with our algorithm (c) $\boldsymbol{\theta}$ with our algorithm

Figure 5. Reconstruction with smaller areas. To be compared with Figure 3. The new areas (a) do not correspond anymore to the ones used to generate the data. (b) is the emission map and (c) the temporal evolution estimated with our algorithm. (b) and (c) are slices of 3D cubes at the same $z = 0.1\, R_{\odot}$. Emission densities (arbitrary units) are scaled in the color bars in the right-end side of (b).

3.3. Choice of Evolution Areas

One can think that the choice of the evolution areas is critical to the good performance of our method. We show in this section that it is not necessarily the case by performing a reconstruction based on simulations with incorrect evolution areas. All parameters and data are exactly the same as in the previous reconstruction. The only difference is in the choice of the areas, *i.e.* the $\boldsymbol{L}$ matrix. These are now defined as shown in Figure 5*(a)*.

Although approximately 50 % of the voxels are not associated with their correct area, we can observe that the algorithm still performs well. The emission map of Figure 5*(b)* is still better than the emission reconstructed by a FBP method. Plus, the estimation of the temporal evolution in Figure 5*(c)* corresponds to the true evolution 3*(e)* even if less precisely than in Figure 3*(f)*.

4. Reconstruction of SOHO/EIT Data

4.1. Data Preprocessing

We now perform reconstruction using SOHO/EIT data. We have to be careful when applying our algorithm to real data. Some problems may arise due to phenomena not taken into account in our model; *e.g.* cosmic rays, or missing data.

Some of these problems can be handled with simple preprocessing. We consider pixels hit by cosmic rays as missing data. They are detected with a median filter. These pixels and missing blocks are labeled as missing data and the projector and the backprojector do not take them into account (*i.e.* the corresponding rows in the matrices are removed).

Table 4. EIT Data Reconstruction: Geometric Parameters

cube size (solar radii)	cube number of voxels	pixel size (radians)	projection number of pixels
$3 \times 3 \times 0.15$	$256 \times 256 \times 8$	$2.55 \times 10^{-5} \times 2.55 \times 10^{-5}$	512×38

Table 5. EIT Data Reconstruction : Other Parameters

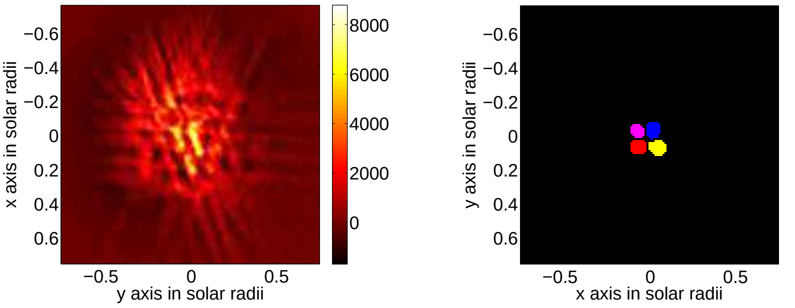
λ	μ	$S_{\boldsymbol{x}}$	S_G
2×10^{-2}	1×10^{4}	0.1	0.05

4.2. Results Analysis

In Figures 6 and 7, we present results from 17.1 nm EIT data between 1 and 14 November 1996. This period corresponds to the minimum of solar activity when one can expect to have less temporal evolution. 17.1 nm is the wavelength where the contrast of the plumes is the strongest. Some images are removed resulting in a sequence of 57 irregularly-spaced projections for a total coverage of 191°. We assume that we know the position of four evolving plumes as shown on Figure 6*(b)*. For each reconstructed image, we present subareas of the reconstructed cube of size 64×64 centered on the axis of rotation. We assume the rotation speed to be the rigid body Carrington rotation. All of the parameters given in Table 4 and 5 are shared by the different algorithms provided they are required by the method. The computation of this reconstruction on a Intel(R) Pentium(R) 4 CPU 3.00 GHz was 13.5 hours long.

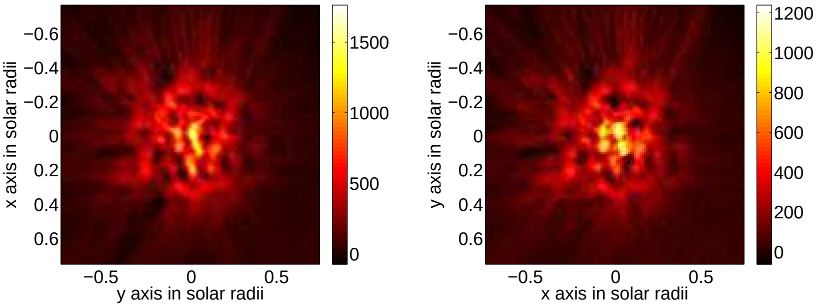
Presence of negative values is the indication of a poor behavior of the tomographic algorithm since it does not correspond to actual physical values. We can see in Figure 6 that our reconstruction has many fewer negative values in the $\boldsymbol{x}$ map than the FBP reconstruction. In the FBP reconstruction cube, 50% of the voxels have negative values; in the gradient-like reconstruction without temporal evolution 36% of the voxels are negative while in our reconstruction only 25 % are negative. This still seems like a lot but most of these voxels are in the outer part of the reconstructed cube. The average value of the negative voxels is much smaller also. It is -120 for the FBP, -52 for the gradient-like method without temporal evolution, and only -19 for our reconstruction with temporal evolution. However, we notice that the gain coefficients present a few slightly negative values.

In the reconstructions without temporal evolution, plumes three (upper right) and four (lower right) correspond to a unique elongated structure which we choose to divide. Note how our algorithm updated the $\boldsymbol{x}$ map reducing the emission values between these two plumes. It shows that what was seen as a unique structure was an artifact resulting from temporal evolution and it tends to validate the usefulness of our model. We note the disappearance of a plume located around (-0.2, -0.15) solar radii on the FBP reconstruction. It shows the utility of gradient-like methods to get rid of artifacts due to the non-uniform distribution of images. Another plume at (0.2, 0.2) solar radii has more intensity in the reconstruction without temporal evolution



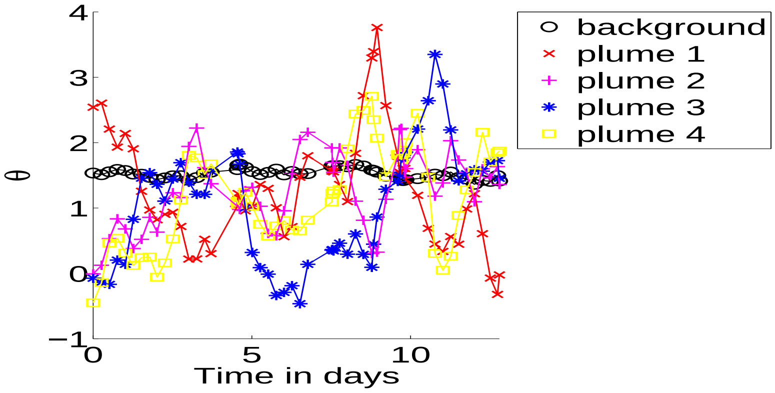
(a) $\boldsymbol{x}$ with a FBP algorithm

(b) Chosen temporal areas. Numbered clockwise starting from lower-left



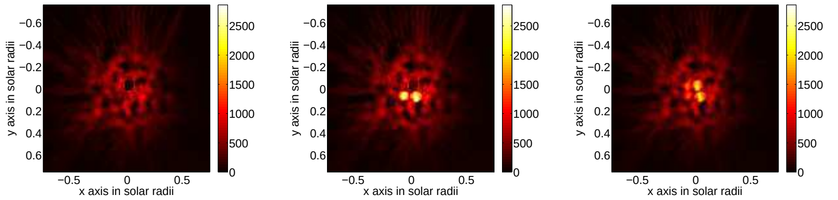
(c) $\boldsymbol{x}$ without temporal evolution

(d) $\boldsymbol{x}$ with our algorithm



(e) $\boldsymbol{\theta}_t$ with our algorithm.

Figure 6. A comparison of FBP (a), a gradient-like algorithm without temporal evolution (c), and our algorithm (d) with real EIT data. $\boldsymbol{x}$ is the spatial distribution of the volume emission density integrated over EIT 17.1 nm passband. The chosen areas are shown in (b). $\boldsymbol{\theta}$ is a gain representing the emission variation during time (e). The time scale is in days. In the case of our algorithm, only the product $\boldsymbol{x} \circ \boldsymbol{\theta}$ has physical meaning. The spatial scales are given in solar radii and centered on the solar axis of rotation. (a), (b), (c), and (d) are slices of 3D cubes at the same $z = 1.3R_{\odot}$. Emission densities (arbitrary units) are scaled in the color bars in the right-end side of (a), (c), (d).



(a) Reconstruction at t_{25} (6 days)

(b) Reconstruction at t_{35} (8.5 days)

(c) Reconstruction at t_{45} (9.8 days)

Figure 7. Reconstruction of $\boldsymbol{x} \circ \boldsymbol{g}$ at different times. Distances are in solar radii. Values represent the volume emission density integrated over the EIT 17.1 nm passband. All of these images are slices of 3D cubes at the same $z = 1.3\,R_{\odot}$.

than with our algorithm. It illustrates how temporal evolution can influence the spatial reconstruction.

5. Discussion

The major feature of our approach is the quality of our reconstruction, which is much improved with respect to FBP reconstruction, as demonstrated by the smaller number of negative values and the increased closeness to the data. Let us now discuss the various assumptions that have been made through the different steps of the method.

The strongest assumption we made, in order to estimate the temporal evolution of polar plumes, is the knowledge of the plume position. Here, we choose to define the plumes as being the brightest points in a reconstruction without temporal evolution. The choice is not based on any kind of automatic threshold. The areas are entirely hand-chosen by looking at a reconstruction. It is possible that these areas do not correspond to the actual physical plumes, they could correspond to areas presenting increased emission during half a rotation. Note that this is biased in favor of plumes closer to the axis of rotation since, along one slice of the reconstructed cartesian cube, their altitude is lower and thus, their intensity is higher. In order to have constant altitude maps one would have to carry out the computation on a spherical grid or to interpolate afterwards onto such a grid. For this reconstruction example we are aware that we did not locate all of the plumes but only tried to find a few. It would be interesting to try to locate the plumes using other data or with a method estimating their positions and shapes.

The method involves hyperparameters which we choose to set manually. There are methods to estimate hyperparameters automatically such as the L-curve method, the cross-validation method (Golub, Heath, and Wahba, 1979) or the full-bayesian method (Higdon *et al.*, 1997; Champagnat, Goussard, and Idier, 1996). We performed reconstructions using different hyperparameter values. We then looked at the reconstruction to see if the smoothness seemed exaggerated or if the noise were amplified in the results. This allowed us to reduce the computational cost and does not really put the validity of the method into question.

One possible issue with this algorithm is the non-convexity of our criterion. This can lead to the convergence to a local minimum that does not correspond to the desired solution defined as the global minimum of the criterion. One way to test this would be to change the initialization many times.

We chose the speed of rotation of the poles to be the Carrington rotation speed. But the speed of the polar structures has not been measured precisely to our knowledge and could affect drastically the reconstruction. This is an issue shared by all tomographic reconstructions of the Sun.

In the current approach, we need to choose on our own the position of the time-evolving areas which are assumed to be plumes. This is done by assuming that more intense areas of a reconstruction without temporal evolution correspond to plume positions. A more rigorous way would be to try to use other sources of information to try to localize the plumes. Another, self-consistent way, would be to develop a method that jointly estimates the position of the plumes in addition to the emission ($\boldsymbol{x}$) and the time evolution ($\boldsymbol{\theta}$). We could try to use the results of Yu and Fessler (2002) who propose an original approach in order to reconstruct a piece-wise homogeneous object while preserving edges. The minimization is alternated between an intensity map and boundary curves. The estimation of the boundary curves is made using level sets technics ((Yu and Fessler, 2002) and references therein). It would also be possible to use a Gaussian mixture model (Snoussi and Mohammad-Djafari, 2007).

6. Conclusion

We have described a method that takes into account the temporal evolution of polar plumes for tomographic reconstruction near the solar poles. A simple reconstruction based on simulations demonstrates the feasibility of the method and its efficiency in estimating the temporal evolution assuming that parameters such as plume position or rotation speed are known. Finally we show that it is possible to estimate the temporal evolution of the polar plumes with real data.

In this study we limited ourselves to reconstruction of images at 17.1 nm but one can perform reconstructions at 19.5 nm and 28.4 nm as well. It would allow us to estimate the temperatures of the electrons as in Frazin, Kamalabadi, and Weber (2005) or Barbey *et al.* (2006).

Acknowledgements Nicolas Barbey acknowledges the support of the Centre National d'Études Spatiales and the Collecte Localisation Satellites. The authors thank the referee for their useful suggestions for the article.

Appendix

A. Pseudo-Inverse Minimization

We want to minimize:

$$J = \|\boldsymbol{y} - \boldsymbol{U}_{\boldsymbol{x}^n}\boldsymbol{\theta}\|^2 + \lambda\|\boldsymbol{D}_r\boldsymbol{x}^n\|^2 + \mu\|\boldsymbol{D}_t\boldsymbol{\theta}\|^2 \tag{17}$$

The second term does not depend on $\boldsymbol{\theta}$. Due to the strict convexity of the criterion, the solution is a zero of the gradient. Since the criterion is quadratic, one can explicitly determine the solution:

$$\nabla_{\boldsymbol{\theta}} J|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{n+1}} = 2\boldsymbol{U}_{\boldsymbol{x}^n}^T\left(\boldsymbol{U}_{\boldsymbol{x}^n}\boldsymbol{\theta}^{n+1} - \boldsymbol{y}\right) + 2\mu\boldsymbol{D}_t^T\boldsymbol{D}_t\boldsymbol{\theta}^{n+1} = \boldsymbol{0} \tag{18}$$

from which we conclude:

$$\boldsymbol{\theta}^{n+1} = \left[\boldsymbol{U}_{\boldsymbol{x}^n}^T \boldsymbol{U}_{\boldsymbol{x}^n} + \mu \boldsymbol{D}_t^T \boldsymbol{D}_t\right]^{-1} \boldsymbol{U}_{\boldsymbol{x}^n}^T \boldsymbol{y} \tag{19}$$

B. Gradient-like Method

In this method we try to find an approximation of the minimum by decreasing the criterion iteratively. The problem is divided in two subproblems: searching for the direction and searching for the step of the descent. In gradient-like methods, the convergence is generally guaranteed ultimately to a local minimum. But since the criterion is convex, the minimum is global. To iterate, we start at an arbitrary point ($\boldsymbol{x}^0$) and go along a direction related to the gradient. The gradient at the p^{th} step is:

$$\nabla_{\boldsymbol{x}} J|_{\boldsymbol{x}=\boldsymbol{x}^p} = 2\boldsymbol{V}_{\boldsymbol{\theta}^{n+1}}^T \left(\boldsymbol{V}_{\boldsymbol{\theta}^{n+1}} \boldsymbol{x}^p - \boldsymbol{y}\right) + 2\lambda \boldsymbol{D}_r^T \boldsymbol{D}_r \boldsymbol{x}^p \tag{20}$$

Once the direction is chosen, searching for the optimum step is a linear minimization problem of one variable:

$$a_{\text{OPT}}^{p+1} = \arg\min_a J(\boldsymbol{x}^p + a\boldsymbol{d}^{p+1}) \tag{21}$$

which is solved by:

$$a_{\text{OPT}}^{p+1} = -\frac{1}{2} \frac{\boldsymbol{d}^{p+1}\, \nabla_{\boldsymbol{x}} J|_{\boldsymbol{x}=\boldsymbol{x}^p}}{\|\boldsymbol{V}_{\boldsymbol{\theta}^{n+1}} \boldsymbol{d}^{p+1}\|^2 + \lambda \|\boldsymbol{D}_r \boldsymbol{d}^{p+1}\|^2} \tag{22}$$

We can write the iteration:

$$\boldsymbol{x}^{p+1} = \boldsymbol{x}^p + a_{\text{OPT}}^{p+1} \boldsymbol{d}^{p+1} \tag{23}$$

References

Barbey, N., Auchére, F., Rodet, T., Bocchialini, K., Vial, J.C.: 2006, Rotational Tomography of the Solar Corona-Calculation of the Electron Density and Temperature. In: Lacoste, H. (ed.) *SOHO 17 - 10 Years of SOHO and Beyond. ESA Special Publications.* ESA Publications Division, Noordwijk, 66 – 68.

Barbey, N., Auchère, F., Rodet, T., Vial, .J.C.: 2007, Reconstruction Tomographique de Séquences d'Images 3D - Application aux Données SOHO/STEREO. In: *Actes de GRETSI 2007.*, 709 – 712.

Champagnat, F., Goussard, Y., Idier, J.: 1996, Unsupervised deconvolution of sparse spike trains using stochastic approximation. *IEEE Transactions on Signal Processing* **44**(12), 2988 – 2998.

DeForest, C.E., Lamy, P.L., Llebaria, A.: 2001, Solar Polar Plume Lifetime and Coronal Hole Expansion: Determination from Long-Term Observations. *Astrophys. J.* **560**, 490 – 498.

Demoment, G.: 1989, Image reconstruction and restoration: Overview of common estimation structure and problems. *IEEE Transactions on Acoustics, Speech and Signal Processing* **37**(12), 2024 – 2036.

Féron, O., Duchêne, B., Mohammad-Djafari, A.: 2005, Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data. *Inverse Problems* **21**(6), S95 – S115.

Frazin, R.A.: 2000, Tomography of the Solar Corona. I. A Robust, Regularized, Positive Estimation Method. *Astrophys. J.* **530**, 1026 – 1035.

Frazin, R.A., Butala, M.D., Kemball, A., Kamalabadi, F.: 2005, Time-dependent Reconstruction of Nonstationary Objects with Tomographic or Interferometric Measurements. *Astrophys. J.* **635**(2), L197 – L200.

Frazin, R.A., Janzen, P.: 2002, Tomography of the Solar Corona. II. Robust, Regularized, Positive Estimation of the Three-dimensional Electron Density Distribution from LASCO-C2 Polarized White-Light Images. *Astrophys. J.* **570**, 408 – 422.

Frazin, R.A., Kamalabadi, F.: 2005, Rotational Tomography For 3D Reconstruction Of The White-Light And Euv Corona In The Post-SOHO Era. *Solar Phys.* **228**, 219 – 237.

Frazin, R.A., Kamalabadi, F., Weber, M.A.: 2005, On the Combination of Differential Emission Measure Analysis and Rotational Tomography for Three-dimensional Solar EUV Imaging. *Astrophys. J.* **628**, 1070 – 1080.

Gabriel, A.H., Abbo, L., Bely-Dubau, F., Llebaria, A., Antonucci, E.: 2005, Solar Wind Outflow in Polar Plumes from 1.05 to 2.4 R_{solar}. *Astrophys. J.* **635**, L185 – L188.

Golub, G.H., Heath, M., Wahba, G.: 1979, Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**(2), 215 – 223.

Grass, M., Manzke, R., Nielsen, T., KoKen, P., Proksa, R., Natanzon, M., Shechter, G.: 2003, Helical cardiac cone beam reconstruction using retrospective ECG gating. *Phys. Med. Biol.* **48**, 3069 – 3083.

Higdon, D.M., Bowsher, J.E., Johnson, V.E., Turkington, T.G., Gilland, D.R., Jaszczak, R.J.: 1997, Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Transactions on Medical Imaging* **16**(5), 516 – 526.

Kachelriess, M., Ulzheimer, S., Kalender, W.: 2000, ECG-correlated imaging of the heart with subsecond multislice spiral CT. *IEEE Transactions on Medical Imaging* **19**(9), 888 – 901.

Kak, A.C., Slaney, M.: 1987, Principles of computerized tomographic imaging. *IEEE Press, New York.*

Llebaria, A., Saez, F., Lamy, P.: 2002, The fractal nature of the polar plumes. In: Wilson, A. (ed.) *ESA SP-508: From Solar Min to Max: Half a Solar Cycle with SOHO.*, 391 – 394.

Natterer, F.: 1986, The Mathematics of Computerized Tomography. John Wiley.

Nocedal, J., Wright, S.J.: 2000, Numerical optimization. *Series in Operations Research.* Springer Verlag, New York.

Polak, E., Ribière, G.: 1969, Note sur la convergence de méthodes de directions conjuguées. *Rev. Française d'Informatique Rech. Opérationnelle* **16**, 35 – 43.

Ritchie, C.J., Crawford, C.R., Godwin, J.D., King, K.F., Kim, Y.: 1996, Correction of computed tomography motion artifacts using pixel-specific backprojection. *IEEE Med. Imag.* **15**(3), 333 – 342.

Roux, S., Debat, L., Koenig, A., Grangeat, P.: 2004, Exact reconstruction in 2d dynamic CT: compensation of time-dependent affine deformations. *Phys. Med. Biol.* **49**(11), 2169 – 2182.

Siddon, R.L.: 1985, Fast calculation of the exact radiological path for a three-dimensional CT array. *Medical Physics* **12**, 252 – 255.

Snoussi, H., Mohammad-Djafari, A.: 2007, Estimation of structured Gaussian mixtures: the inverse EM algorithm. *IEEE Trans. on Signal Processing* **55**(7), 3185 – 3191.

Teriaca, L., Poletto, G., Romoli, M., Biesecker, D.A.: 2003, The Nascent Solar Wind: Origin and Acceleration. *Astrophys. J.* **588**, 566 – 577.

Wiegelmann, T., Inhester, B.: 2003, Magnetic modeling and tomography: First steps towards a consistent reconstruction of the solar corona. *Solar Physics* **214**, 287 – 312.

Yu, D., Fessler, J.: 2002, Edge-preserving tomographic reconstruction with nonlocal regularization. *IEEE Med. Imag.* **21**, 159 – 173.

T. Rodet, F. Noo et M. Defrise, « The cone–beam algorithm of Feldkamp, Davis, and Kress preserves oblique line integral », *Medical Physics*, vol. 31, n°7, pp. 1972–1975, juillet 2004.

The cone-beam algorithm of Feldkamp, Davis, and Kress preserves oblique line integrals

Thomas Rodet
Department of Nuclear Medicine, Vrije Universiteit Brussel, AZ-VUB, B-1090 Brussels, Belgium

Frédéric Noo
Department of Radiology, University of Utah, Salt Lake City, Utah 84112

Michel Defrise[a)]
Department of Nuclear Medicine, Vrije Universiteit, Brussel, AZ-VUB, B-1090 Brussels, Belgium

(Received 2 January 2004; revised 23 March 2004; accepted for publication 20 April 2004; published 17 June 2004)

The algorithm of Feldkamp, Davis, and Kress [J. Opt. Soc. Am. A **1**, 612–619 (1984)] is a widely used filtered-backprojection algorithm for three-dimensional image reconstruction from cone-beam (CB) projections measured with a circular orbit of the x-ray source. A well-known property of this approximate algorithm is that the integral of the reconstructed image along any *axial* line orthogonal to the plane of the orbit is exact when the cone-beam projections are not truncated. We generalize this result to oblique line integrals, thus providing an efficient method to compute synthetic radiographs from cone-beam projections. Our generalized result is obtained by showing that the FDK algorithm is invariant under transformations that map oblique lines onto axial lines. © *2004 American Association of Physicists in Medicine.* [DOI: 10.1118/1.1759828]

I. INTRODUCTION

The algorithm of Feldkamp, Davis, and Kress (FDK[1]) was introduced in 1984 to reconstruct cone-beam (CB) projections measured with a circular orbit of the x-ray source around the object. Despite considerable progress in the theory of CB reconstruction (see, e.g., Ref. 2), and despite its approximate character, the FDK algorithm is still widely used, either in its original form, or in various extensions developed for improved accuracy[3] or for helical cone-beam CT.[4–6] Applications include cardiac CT, small animal imaging, or radiotherapy. This lasting popularity stems from the simplicity and ease of implementation of the algorithm and from its good robustness to the CB artefacts caused by the incompleteness of the circular orbit.[7,8] The good accuracy of the FDK algorithm is often related to three well-known properties, which were derived in the original paper.[1]

(1) The reconstruction is exact in the plane of the circular orbit of the x-ray source.
(2) The reconstruction is exact for any object which is invariant for translations in the *axial* direction orthogonal to the plane of the orbit.
(3) The integral of the reconstructed image along any *axial* line orthogonal to the plane of the orbit is exact, i.e., the value of that integral coincides with the value of the corresponding line integral of the actual object.

In this paper, we generalize the latter property by showing that the FDK reconstruction also preserves the value of integrals along *oblique* lines that are not orthogonal to the circular orbit plane. This result provides additional insight into the accuracy of the FDK reconstructions and of synthetic projections [such as Digitally Reconstructed Radiographs (DRR) in radiotherapy] which can be calculated from these reconstructions. Although this property has never, to our knowledge, been mentioned, the very possibility of calculating oblique line integrals from circular CB data has previously been demonstrated by Chen, Patch, and Trofimov. Chen[9] reconstructs an oblique line integral by integrating the ramp filtered three-dimensional (3D) radon transform for all planes containing the line. The ramp filtered 3-D radon transform can be calculated for a given plane using any CB projection measured for a vertex located in the plane, by using Smith's intermediate function.[12] In Patch,[10] CB projections are synthetized for any vertex located within the disk defined by the circular orbit. The computation is achieved by solving John's equation for the three-dimensional x-ray transform, using the measured projections as boundary values. In Trofimov,[11] a theorem supporting the results in Patch, Chen, and this paper was given without the specification of an algorithm for the effective computation of line integrals. In this paper we show that the FDK algorithm allows the computation of oblique line integrals. After introducing the FDK algorithm in Sec. II, we demonstrate in Sec. III the invariance of this algorithm for some affine transformation. This serves in Sec. IV to prove the preservation of oblique line integrals when two conditions are satisfied.

(i) The cone-beam projections are not truncated;
(ii) the angle between the line and the axial direction does not exceed a maximum value that will be specified in Sec. IV.

II. THE FDK ALGORITHM

We consider the reconstruction of a smooth function f from CB projections measured for cone vertices on a centered circle in the plane $z=0$. To locate points in the field-of-view, vector notations $\vec{r}=x\vec{1}_x+y\vec{1}_y$ will be used for the transaxial components, while the axial coordinate z will be written separately. Thus, we write the density function as $f(\vec{r},z)$, and throughout the paper $\vec{r}$ will denote a vector orthogonal to the axis $\vec{1}_z$ of the scan.

We assume that the function f is smooth and vanishes outside the cylinder,

$$\Omega_f=\{(\vec{r},z)\in\mathbb{R}^3 | \quad \|\vec{r}\|\leqslant R_{\mathrm{FOV}},\ |z|\leqslant H_{\mathrm{FOV}}\}. \tag{1}$$

The position of the cone vertex is

$$\vec{a}(\lambda)=R\vec{1}_w(\lambda), \quad \lambda\in[0,2\pi), \tag{2}$$

where λ is the angle describing the rotation of the cone vertex, $R>R_{\mathrm{FOV}}$ is the radius of the orbit, and

$$\vec{1}_w(\lambda)=\cos\lambda\ \vec{1}_x+\sin\lambda\ \vec{1}_y \tag{3}$$

is a unit vector pointing from the origin of the coordinate system towards the cone vertex.

We consider a flat area detector which rotates with the cone vertex. At angular position λ, the detector is parallel to the z-axis and to the vector $\vec{a}'(\lambda)$ tangent to the circle. To simplify the derivations, we assume that the distance from $\vec{a}(\lambda)$ to the detector is R, i.e., the detector contains the rotation axis $\vec{1}_z$. Detector pixels are located with Cartesian coordinates u and v measured along two orthogonal unit vectors in the plane $\mathcal{D}(\lambda)$ containing the flat area detector:

$$\vec{1}_u(\lambda)=-\sin\lambda\ \vec{1}_x+\cos\lambda\ \vec{1}_y, \quad \vec{1}_v=\vec{1}_z. \tag{4}$$

Thus, u is the transaxial coordinate ("detector element") whereas v is the axial coordinate ("detector row"). The origin $(u,v)=(0,0)$ is at the orthogonal projection of $\vec{a}(\lambda)$ onto $\mathcal{D}(\lambda)$.

The CB projection for vertex position λ is the set of line integrals $g_{\mathrm{meas}}(u,v,\lambda)$ connecting $\vec{a}(\lambda)$ to points (u,v) in $\mathcal{D}(\lambda)$. Throughout the paper, we use modified data $g(\lambda,u,v)$ obtained by applying the usual FDK weight to the measured data:

$$\begin{aligned} g(\lambda,u,v)&=\frac{1}{\sqrt{R^2+u^2+v^2}}g_{\mathrm{meas}}(\lambda,u,v)\\ &=\int_0^\infty dl\, f(\vec{X}(\lambda,u,l),lv), \end{aligned} \tag{5}$$

with $(u,v)\in[-u_D,u_D]\times[-v_D,v_D]$ and

$$\vec{X}(\lambda,u,l)=(1-l)R\vec{1}_w(\lambda)+lu\vec{1}_u(\lambda). \tag{6}$$

In practice the limits of the integral over l in Eq. (5) are of course finite, and are determined by the radius of the FOV. We assume below that the projections are not truncated, i.e., the shadow of the cylindrical FOV is always contained within the detector. This requires

$$u_D\geqslant u_{\max}=\frac{RR_{\mathrm{FOV}}}{\sqrt{R^2-R_{\mathrm{FOV}}^2}}, \quad v_D\geqslant v_{\max}=\frac{RH_{\mathrm{FOV}}}{R-R_{\mathrm{FOV}}}. \tag{7}$$

The FDK algorithm[1] is an empirical 3-D extension of the standard 2-D fan-beam filtered-backprojection algorithm.[13] The reconstructed image is given by a CB backprojection:

$$f^{\mathrm{FDK}}(\vec{r},z)=\frac{1}{2}\int_0^{2\pi}d\lambda\,\frac{R^3}{(R-\vec{r}\cdot\vec{1}_w(\lambda))^2}g^F(\lambda,U,V), \quad \|\vec{r}\|<R, \tag{8}$$

where

$$\begin{aligned} U&=U(\vec{r},\lambda)=\frac{R\vec{r}\cdot\vec{1}_u(\lambda)}{R-\vec{r}\cdot\vec{1}_w(\lambda)},\\ V&=V(\vec{r},z,\lambda)=\frac{Rz}{R-\vec{r}\cdot\vec{1}_w(\lambda)}, \end{aligned} \tag{9}$$

are such that the ray defined by λ, U, V contains the point $(\vec{r},z)$. The filtered projections are given by

$$g^F(\lambda,u,v)=\int_{-u_{\max}}^{u_{\max}}du'\,h_R(u-u')g(\lambda,u',v). \tag{10}$$

This is a 1-D convolution of each detector row $v=$const with the kernel $h_R(u)$ of the ramp filter (the inverse 1-D Fourier transform of $|\nu|$).

The FDK reconstruction is defined only for $\|\vec{r}\|<R$ because the magnification weight in Eq. (8) is singular when $\vec{r}=\vec{a}(\lambda)$ for some λ. In addition, $f^{\mathrm{FDK}}(\vec{r},z)$ vanishes outside a volume Ω_{FDK} limited by two conical surfaces:

$$f^{\mathrm{FDK}}(\vec{r},z)=0 \ \text{ when } \ (\vec{r},z)\notin\Omega_{\mathrm{FDK}}:=\left\{(\vec{r},z) | \ \|\vec{r}\|<R \ \text{ and } \ |z|<v_{\max}\left(1+\frac{\|\vec{r}\|}{R}\right)\right\}. \tag{11}$$

This property is verified by noting that the filtering in Eq. (10) is done row by row, and that $|V(\vec{r},z,\lambda)|\geqslant R|z|/(R+\|\vec{r}\|)$.

Feldkamp *et al.*[1] showed that their algorithm preserves the value of axial line integrals when the CB projections are not truncated [i.e., when Eq. (7) is satisfied]. Specifically, they showed that for any $\vec{r}_0$ with $\|\vec{r}_0\|<R$,

$$\int_{(\vec{r}_0,z)\in\Omega_{\mathrm{FDK}}}dz\ \ f^{\mathrm{FDK}}(\vec{r}_0,z)=\int_{-H_{\mathrm{FOV}}}^{H_{\mathrm{FOV}}}dz\ \ f(\vec{r}_0,z). \tag{12}$$

On the left hand side of this equation, the FDK reconstruction must be integrated over the whole support Ω_{FDK}, i.e., over the range

$$-v_{\max}\left(1+\frac{\|\vec{r}_0\|}{R}\right)\leqslant z\leqslant v_{\max}\left(1+\frac{\|\vec{r}_0\|}{R}\right). \tag{13}$$

Thus, in order to use Eq. (12), the computation of the FDK reconstruction must not be restricted to the cylindrical support Ω_f of the object, but must include all "artifacts."

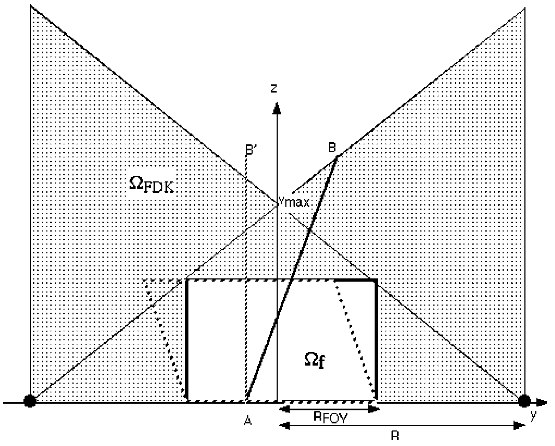


FIG. 1. A longitudinal section of the object support Ω_f, also showing the intersection with the circular orbit (small black disks). Only the $z\geqslant 0$ region is shown, and nontruncated projections are assumed. The region Ω_{FDK} in which the FDK reconstruction of f is nonzero is shown by the dotted pattern. The sheared object defined by the affine transform (14) is shown by the dashed parallelogram. The oblique line AB is transformed into the axial line AB' through the sheared object.

III. INVARIANCE OF THE FDK ALGORITHM FOR SOME AFFINE TRANSFORMS

To extend the property (12) to oblique lines, we first prove in this section that the FDK algorithm is invariant for affine transformations which map the object function f onto a "sheared" object (Fig. 1),

$$f_s(\vec{r},z):=f(\vec{r}+z\,\vec{s},z), \tag{14}$$

for some vector $\vec{s}\in\mathbb{R}^2$ orthogonal to $\vec{1}_z$. Restrictions on the vector $\vec{s}$ will be discussed later.

The weighted CB projections of f_s are obtained by inserting Eq. (14) into (5):

$$\begin{aligned} g_s(\lambda,u,v)&=\int_0^\infty dl\,f_s(\vec{X}(\lambda,u,l),lv)\\ &=\int_0^\infty dl\,f(\vec{X}(\lambda,u,l)+lv\vec{s},lv)\\ &=Sg(\lambda,S(u+v\vec{s}\cdot\vec{1}_u(\lambda)),Sv), \end{aligned}\tag{15}$$

where S is the following function of λ and v:

$$S=S(v,\lambda)=\frac{R}{R-v\,\vec{s}\cdot\vec{1}_w(\lambda)}. \tag{16}$$

To simplify notations, the arguments of S are omitted in Eq. (15) and in most equations below. The last line in Eq. (15) is obtained by noting that $\vec{X}(\lambda,u,l)+lv\vec{s}=\vec{X}(\lambda,S(u+v\vec{s}\cdot\vec{1}_u(\lambda)),S^{-1}l)$, by changing the integration variable to $l'=S^{-1}l$, and by comparing the result with Eq. (5).

We now apply the FDK algorithm to the sheared data. The first step is the ramp filtering of each detector row, according to Eq. (10):

$$\begin{aligned} g_s^F(\lambda,u,v)&=\int_{-\infty}^{\infty}du'h_R(u-u')g_s(\lambda,u',v)\\ &=\int_{-\infty}^{\infty}du'h_R(u-u')Sg(\lambda,S(u'+v\vec{s}\\ &\quad\cdot\vec{1}_u(\lambda)),Sv). \end{aligned}\tag{17}$$

Setting $\tilde{u}:=S(v,\lambda)(u'+v\vec{s}\cdot\vec{1}_u(\lambda))$, and using the scaling property $h_R(cu)=h_R(u)/c^2$, $c>0$, we get

$$\begin{aligned} g_s^F(\lambda,u,v)&=\int_{-\infty}^{\infty}d\tilde{u}h_R(u-S^{-1}\tilde{u}+v\vec{s}\cdot\vec{1}_u(\lambda))g(\lambda,\tilde{u},Sv)\\ &=S^2\int_{-\infty}^{\infty}d\tilde{u}h_R(S(u+v\vec{s}\cdot\vec{1}_u(\lambda))-\tilde{u})g(\lambda,\tilde{u},Sv)\\ &=S^2g^F(\lambda,S(u+v\vec{s}\cdot\vec{1}_u(\lambda)),Sv). \end{aligned}\tag{18}$$

The second step of the algorithm is the backprojection (8), yielding the reconstructed sheared image f_s^{FDK}:

$$\begin{aligned} f_s^{\text{FDK}}(\vec{r},z)&=\frac{1}{2}\int_0^{2\pi}d\lambda\,\frac{R^3}{(R-\vec{r}\cdot\vec{1}_w(\lambda))^2}\\ &\quad\times g_s^F(\lambda,U(\vec{r},\lambda),V(\vec{r},z,\lambda))\\ &=\frac{1}{2}\int_0^{2\pi}d\lambda\,\frac{R^3}{(R-\vec{r}\cdot\vec{1}_w(\lambda))^2}S^2\\ &\quad\times g^F(\lambda,S(U(\vec{r},\lambda)+V(\vec{r},z,\lambda)\vec{s}\\ &\quad\cdot\vec{1}_u(\lambda)),SV(\vec{r},z,\lambda)), \end{aligned}\tag{19}$$

with $S=S(V(\vec{r},z,\lambda),\lambda)$. Noting the three relations,

$$\begin{aligned} &\frac{S(V(\vec{r},z,\lambda),\lambda)}{R-\vec{r}\cdot\vec{1}_w(\lambda)}=\frac{1}{R-(\vec{r}+z\,\vec{s})\cdot\vec{1}_w(\lambda)},\\ &S(V(\vec{r},z,\lambda),\lambda)(U(\vec{r},\lambda)+V(\vec{r},z,\lambda)\vec{s}\cdot\vec{1}_u(\lambda))\\ &\quad=U(\vec{r}+z\vec{s},\lambda),\\ &S(V(\vec{r},z,\lambda),\lambda)V(\vec{r},z,\lambda)=V(\vec{r}+z\vec{s},z,\lambda), \end{aligned}\tag{20}$$

and comparing Eqs. (19) to (8), one obtains

$$f_s^{\text{FDK}}(\vec{r},z)=f^{\text{FDK}}(\vec{r}+z\vec{s},z). \tag{21}$$

This concludes the proof that the FDK algorithm is invariant for the affine mapping (14). This results holds for any $\vec{s},\vec{r},z$ such that $\|\vec{r}\|<R$ and $\|\vec{r}+z\vec{s}\|<R$ so that the FDK reconstructions on the left and right hand sides of Eq. (21) are defined. In view of the first line of Eq. (20), this condition also guarantees that $S(v,\lambda)$ is finite.

IV. PRESERVATION OF OBLIQUE LINE INTEGRALS BY THE FDK ALGORITHM

The property that the integral of the FDK reconstruction along an oblique line defined by $\vec{r}=\vec{r}_0+z\vec{s}$, for some $\vec{s}$ orthognal to $\vec{1}_z$, is exact is a direct consequence of the invariance property derived in the previous section. Indeed, we can write

$$\begin{aligned}\int_{-\infty}^{\infty} dz \;\; f^{\mathrm{FDK}}(\vec{r}_0+z\vec{s},z) &= \int_{-\infty}^{\infty} dz f_s^{\mathrm{FDK}}(\vec{r}_0,z)\\ &= \int_{-\infty}^{\infty} dz f_s(\vec{r}_0,z)\\ &= \int_{-\infty}^{\infty} dz f(\vec{r}_0+z\vec{s},z), \qquad (22)\end{aligned}$$

where the first equality uses Eq. (21). The second equality follows from the fact that f_s^{FDK} is obtained by applying the FDK algorithm to the CB projections of f_s, and from the preservation of axial line integrals by this algorithm. The last equality is obtained using the definition of the mapping (14).

Let us now discuss the conditions of validity of our main result. Equation (22) assumes that the relation (21) holds for all points of the oblique line for which $f^{\mathrm{FDK}}(\vec{r},z) \neq 0$. This condition requires that the intersections of the oblique line with the boundaries of the support Ω_{FDK} of $f^{\mathrm{FDK}}(\vec{r},z)$ are both on the lower or upper conical surface [see Eq. (11)], and not on the lateral surface of the cylinder of radius R, where f^{FDK} is not defined. Hence the intersections of the oblique line of parametric equation $(\vec{r}_0+z\vec{s},z)$ with the cylinder of radius R must both be located at $|z|>2v_{\max}$. One easily checks that this condition is satisfied when $\|\vec{r}_0\|<R$ and

$$2\|\vec{s}\|v_{\max} < \sqrt{R^2-\|\vec{r}_0\|^2+\frac{(\vec{r}_0\cdot\vec{s})^2}{\|\vec{s}\|^2}}-\frac{|\vec{r}_0\cdot\vec{s}|}{\|\vec{s}\|}. \qquad (23)$$

Equation (23) also implies that the support of the sheared object $f_s(\vec{r},z)$ is contained within the cylinder of radius R, thus guaranteeing that $f_s^{\mathrm{FDK}}(\vec{r},z)$ is well-defined. To verify this statement, note that

$$f_s(\vec{r},z)\neq 0 \rightarrow (\vec{r}+z\vec{s},z)\in\Omega_f \rightarrow |z|\leq H_{\mathrm{FOV}}$$

and

$$\|\vec{r}+z\vec{s}\|\leq R_{\mathrm{FOV}}. \qquad (24)$$

Therefore,

$$\begin{aligned}f_s(\vec{r},z)\neq 0 \rightarrow \|\vec{r}\|^2 &= \|\vec{r}+z\vec{s}\|^2+\|\vec{s}\|^2z^2-2z\vec{s}\cdot(\vec{r}+z\vec{s})\\ &\leq R_{\mathrm{FOV}}^2+\left(\frac{R-R_{\mathrm{FOV}}}{2H_{\mathrm{FOV}}}\right)^2 H_{\mathrm{FOV}}^2\\ &\quad +2\,\frac{R-R_{\mathrm{FOV}}}{2H_{\mathrm{FOV}}}H_{\mathrm{FOV}}R_{\mathrm{FOV}}\\ &= \frac{(R+R_{\mathrm{FOV}})^2}{4}<R^2, \qquad (25)\end{aligned}$$

where we have used the fact that condition (23), with $v_{\max}$ from Eq. (7), implies $\|\vec{s}\|<R/(2v_{\max})=(R-R_{\mathrm{FOV}})/(2H_{\mathrm{FOV}})$. We finally conclude that Eq. (23) is a sufficient condition of validity for Eq. (22).

V. CONCLUSION

We have shown that the filtered-backprojection algorithm of Feldkamp, Davis, and Kress for circular cone-beam reconstruction preserves the values of oblique line integrals, thereby generalizing the well-known property that axial line integrals are preserved. This property provides new insight into the accuracy of this widely used algorithm, and it also suggests an efficient way of calculating synthetic radiographs from CB data. The property only holds when the CB projections are not truncated and it requires an additional condition on the angle between the line and the axial direction. Its practical significance, therefore, mainly concerns the imaging of small animals and of isolated tissue samples, which can be put on blocks of material with very low attenuation to meet the hypothesis of nontruncation.

[a] Electronic mail: mdefrise@minf.vub.ac.be

[1] L. A. Feldkamp, L. C. Davis, and J. W. Kress, ‘‘Practical cone-beam algorithm,’’ J. Opt. Soc. Am. A **1**, 612–619 (1984).

[2] A. Katsevich, ‘‘Improved exact FBP algorithm for spiral CT,’’ Adv. Appl. Math. (submitted); also see ‘‘Analysis of an exact inversion algorithm for spiral cone-beam CT,’’ Phys. Med. Biol. **47**, 2583–2597 (2002).

[3] M. Grass, T. Köhler, and R. Proksa, ‘‘3D cone-beam CT reconstruction for circular trajectories,’’ Phys. Med. Biol. **45**, 329–348 (2000).

[4] H. Kudo and T. Saito, ‘‘Helical-scan computed tomography using cone-beam projections,’’ *Conference Record of the 1991 IEEE Medical Imaging Conference*, Santa Fe, NM, 1992, pp. 1958–1962.

[5] G. Wang, T. H. Lin, P. Cheng, and D. M. Shinozaki, ‘‘A general cone-beam reconstruction algorithm,’’ IEEE Trans. Med. Imaging **MI-12**, 486–496 (1993).

[6] X. Yan and R. M. Leahy, ‘‘Cone-beam tomography with circular, elliptical and spiral orbits,’’ Phys. Med. Biol. **37**, 493–506 (1992).

[7] H. Tuy, ‘‘An inversion formula for cone-beam reconstruction,’’ SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **43**, 546–552 (1983).

[8] P. Grangeat, ‘‘Mathematical framework of cone-beam 3D reconstruction via the first derivative of the Radon transform,’’ *Mathematical Methods in Tomography, Lecture Notes in Mathematics*, edited by G. T. Herman, A. K. Louis, and F. Natterer (Springer-Verlag, Berlin, 1991), Vol. 1497, pp. 66–97.

[9] J. Chen, ‘‘A theoretical framework of regional cone-beam tomography,’’ IEEE Trans. Med. Imaging **MI-11**, 342–351 (1992).

[10] S. K. Patch, ‘‘Consistency conditions upon 3D CT data and the wave equation,’’ Phys. Med. Biol. **47**, 2637–2650 (2002).

[11] O. E. Trofimov, ‘‘Virtual beam (x-ray) projections,’’ *Proceedings of the 2003 International Conference on Fully 3D Reconstruction in Radiology and Nuclear Medicine*, Saint-Malo, France, edited by Y. Bizais, 2003.

[12] B. D. Smith, ‘‘Image reconstruction from cone-beam projections—Necessary and sufficient conditions and reconstruction methods,’’ IEEE Trans. Med. Imaging **MI-4**, 14–25 (1985).

[13] F. Natterer, *The Mathematics of Computerized Tomography* (Wiley, New York, 1986).

H. Kudo, **T. Rodet**, F. Noo et M. Defrise, « Exact and approximate algorithms for helical cone-beam CT », *Physics in Medicine and Biology*, vol. 49, n°13, pp. 2913–2931, juillet 2004.

Exact and approximate algorithms for helical cone-beam CT

Hiroyuki Kudo[1], Thomas Rodet[2], Frédéric Noo[3], Michel Defrise[2]
[1] Dept. of Computer Science, Graduate School of Systems and Information Engineering,
University of Tsukuba, Japan
[2]Dept. of Nuclear Medicine, Vrije Universiteit Brussel, AZ-VUB, B-1090 Brussels, Belgium
[3] Dept. of Radiology, University of Utah, Salt-Lake City, Utah.

April 27, 2004

Abstract

This paper concerns image reconstruction for helical x-ray transmission tomography (CT) with multi-row detectors. We introduce two approximate cone-beam (CB) filtered-backprojection (FBP) algorithms of the Feldkamp type, obtained by extending to three dimensions (3D) two recently proposed exact FBP algorithms for 2D fan-beam reconstruction. The new algorithms are similar to the standard Feldkamp-type FBP for helical CT. In particular, they can reconstruct each transaxial slice from data acquired along an arbitrary segment of helix, thereby efficiently exploiting the available data. In contrast with the standard Feldkamp-type algorithm, however, the redundancy weight is applied *after* filtering, allowing a more efficient numerical implementation. To partially alleviate the CB artefacts, which increase with increasing values of the helical pitch, a frequency-mixing method is proposed. This method reconstructs the high frequency components of the image using the longest possible segment of helix, whereas the low frequencies are reconstructed using a minimal, short-scan, segment of helix to minimize CB artefacts. The performance of the algorithms is illustrated using simulated data.

1 Introduction

This paper concerns the image reconstruction problem for helical x-ray transmission tomography (CT) with multi-row detectors. An exact solution to this problem has been recently introduced by Katsevich (2002, 2002a, 2003) and leads to an efficient filtered-backprojection (FBP) algorithm for the reconstruction of a 3D function from axially truncated cone-beam (CB) projections acquired with a helical path of the x-ray source relative to the patient. In this algorithm, the CB projections are first differentiated with respect to the helical path length; a 1D Hilbert filter is then applied along one or several families of straight lines in the detector, and the filtered CB projections are finally backprojected as in the well-known algorithm of Feldkamp, Davis and Kress (1984).

One of the key points in the derivation of Katsevich's algorithm is the definition of families of filtering lines that guarantee exact reconstruction, allow the handling of axially truncated projections, and efficiently use the available detector area to optimize the signal-to-noise ratio (SNR). To date, families of filtering lines that yield exact reconstruction have been defined only for two specific cases, referred to here for convenience as the PI-reconstruction and the 3 PI-reconstruction. With the PI-reconstruction, each point in the image volume is reconstructed using the segment of helix bounded by the PI-line of the point[1]. This is the smallest helix segment compatible with exact reconstruction, and therefore, when the detector size is fixed, the PI-reconstruction allows to maximize the helical pitch. The 3 PI-reconstruction reconstructs each point P from a longer segment of helix bounded by a line containing P and two points of the helix with an axial separation comprised between one and two times the pitch. As with the PI reconstruction, however, the 3 PI-reconstruction makes efficient use of the available detector area only for a well defined value of the helical pitch.

Thus being restricted to these two specific cases, the exact algorithms known to date do not allow optimal data utilization with arbitrary values of the helical pitch, for a fixed size of the detector. This motivates continuing research on approximate algorithms that can more easily be adapted to general configurations. Two approaches have been proposed so far:

- Generalizations (Stierstorfer *et al* 2002, Tang 2003, Flohr *et al* 2003) of the advanced single slice rebinning algorithm (Larson *et al* 1998, Heuscher 1999, Noo *et al* 1999, Kachelriess *et al* 2000, Defrise *et al* 2001, 2003),
- Generalizations of the FDK algorithm (Feldkamp *et al* 1984) for helical data (Kudo and Saito 1992, Wang *et al* 1993, Yan and Leahy 1992).

The second approach will be discussed in this paper. The approximate algorithms of the FDK type are derived by extending to 3D standard algorithms for 2D fan-beam reconstruction. Specifically, an exact 2D fan-beam inversion formula is first written for each point $P = (x_0, y_0, z_0)$, assuming a virtual circular orbit contained in the transaxial plane $z = z_0$. The ray-sums required by this 2D fan-beam inversion formula are then replaced by measured ray-sums diverging from locations of the x-ray source along its circular or helical path. The 2D fan-beam backprojection is thereby converted into a 3D cone-beam backprojection, while the ramp filtering remains the same as in the 2D case (the nature of the lines in the detector along which the filter is applied will be discussed below).

This strategy has been applied to circular and to helical CT, starting either from a full-scan fan-beam reconstruction (360 degree) or from a short-scan reconstruction. In the latter case, data redundancy is handled using the weighting proposed by Parker (1982). Parker's weighting for data redundancy, however, can be generalized to arbitrary angular ranges, up to 360 degrees (Silver 2000), and possibly exceeding 360 degrees. Extending the resulting

---

[1] The PI-line of a point P is the unique line through P that connects two points of the helix separated axially by less than one pitch.

algorithms to 3D in the same way as for the standard FDK method leads to an approximate CB-CT algorithm in which each point $P = (x_0, y_0, z_0)$ is reconstructed from a segment of helix centered, axially, at $z = z_0$. The length of the helix segment is arbitrary, provided it is larger than the short-scan segment and does not exceed the maximum size allowed by the detector size.

Besides its approximate nature, the main drawback of the FDK type approach for helical CB-CT is that the cone-beam projections must be filtered more than once (Kudo *et al* 1999). This property is due to the fact that the projections must be multiplied by Parker's weight *prior* to filtering (see Parker 1982 and Wesarg *et al* 2002). But Parker's weight depends on the location of the x-ray source relative to the segment of helix used for reconstruction, and since this segment is centered on the transaxial slice z_0, the weight applied to a given detector row depends on the slice z_0 to which this row contributes. Therefore, each row must be weighted and then filtered, once for each image slice onto which it is backprojected. This added computational complexity is only avoided when the segment of helix used for reconstruction is an integer multiple of 2π since Parker's weight is constant in that case.

In this paper, we introduce two approximate algorithms for helical CB-CT which can be applied to arbitrary helix segments. These algorithms are obtained by extending to 3D, in the same way as above, two recently discovered 2D fan-beam reconstruction algorithms (Noo *et al* 2002, Kudo *et al* 2002) based on a relation between fan-beam projections and parallel-beam projections (Hamaker *et al* 1980). The main property of these algorithms is that the redundancy weight is applied *after* filtering. This property allows to filter each projection only once, hence considerably simplifying the extension to helical CT.

The paper is organized as follows. After introducing notations in section 2, the two fan-beam FBP algorithms are described in section 3, and are then extended in section 4 to helical CT. The link between the resulting approximate algorithms and exact CB-FBP methods is discussed in section 5, allowing a characterization of the approximation in terms of the 3D Radon transform. The link with the exact methods will also provide some motivation for selecting filtering lines in FDK type algorithms that are parallel to the tangent to the helix, as originally proposed by Yan and Leahy (1992). These results naturally lead to a new algorithm, described in section 6. This algorithm reconstructs the low- and the high-frequency components of the object using respectively a short and a longer helix segment; this technique efficiently exploits the available detector data while minimizing the cone-beam artefacts due to the approximate character of the FDK algorithms. Finally, in section 7, results with simulated data illustrate the accuracy of the proposed methods and their potential to reduce image noise by using redundant CB data in an efficient way.

2 The helical CT data and the 3D x-ray transform

Figure 1 illustrates the data acquisition geometry. We use $f(\vec{r})$ with $\vec{r} = (x, y, z)$ to describe the density function to be reconstructed and assume that this function is smooth and vanishes outside the cylinder

$$\Omega = \Big\{(x, y, z) \in \mathrm{I\!R}^3 \;\Big|\; x^2 + y^2 \le R_{FOV}^2\Big\}. \tag{1}$$

We will write equivalently $f(\vec{r}) = f(x, y, z)$. Physically, Ω represents the field-of-view of the scanner.

To reconstruct $f(\vec{r})$, we measure CB projections for cone vertices on the helical path $\vec{a}(\lambda)$ oriented along the z-axis:

$$\vec{a}(\lambda) = (R\cos\lambda, R\sin\lambda, h\lambda) \tag{2}$$

where $\lambda \in \mathrm{I\!R}$ is the angle describing the rotation of the cone vertex, $R > R_{FOV}$ is the radius of the helix and $P = 2\pi h$ is its pitch. In CT, the cone vertex represents the position of the x-ray source relative to the object or patient.

As illustrated in figure 1, we consider a flat area detector which moves with the cone vertex. At angular position λ, the detector is parallel to the z-axis and to the vector $\vec{a}'(\lambda) = d\vec{a}(\lambda)/d\lambda$ tangent to the helix. To simplify the derivations, we assume below that the distance from $\vec{a}(\lambda)$ to the detector is R, i.e. the detector contains the z-axis[2]. To locate detector pixels in space, we introduce cartesian coordinates u and v in the plane $\mathcal{D}(\lambda)$ containing the area detector. These coordinates are measured along two orthogonal unit vectors $\vec{1}_u$ and $\vec{1}_v$ in $\mathcal{D}(\lambda)$, with $\vec{1}_v$ parallel to the axis of the helix:

$$\vec{1}_u(\lambda) = (-\sin\lambda, \cos\lambda, 0) \qquad \vec{1}_v(\lambda) = (0, 0, 1) \tag{3}$$

The origin $(u, v) = (0, 0)$ is at the orthogonal projection of $\vec{a}(\lambda)$ onto $\mathcal{D}(\lambda)$, and the measured range is $(u, v) \in [-u_D, u_D] \times [-v_D, v_D]$. We assume that the detector is large enough to avoid truncation along the u axis, so that $u_D \ge u_m = R R_{FOV}/\sqrt{R^2 - R_{FOV}^2}$.

The CB projection for vertex position λ is the set of line integrals $g(u, v, \lambda)$ which connect $\vec{a}(\lambda)$ to points (u, v) in $\mathcal{D}(\lambda)$:

$$g(\lambda, u, v) = \sqrt{R^2 + u^2 + v^2} \int_{l_0(u)-\Delta l(u)}^{l_0(u)+\Delta l(u)} dl\; f\Big(X(\lambda, u, l), Y(\lambda, u, l), h\lambda + lv\Big) \tag{4}$$

with $(u, v) \in [-u_D, u_D] \times [-v_D, v_D]$ and

$$\begin{aligned} X(\lambda, u, l) &= R\cos\lambda + l(-R\cos\lambda - u\sin\lambda) \\ Y(\lambda, u, l) &= R\sin\lambda + l(-R\sin\lambda + u\cos\lambda) \end{aligned} \tag{5}$$

The limits of the integral in equation (4) are determined by the intersection of the ray with the boundary of the cylindrical field-of-view Ω:

$$l_0(u) = R^2/(R^2 + u^2) \quad , \quad \Delta l(u) = \sqrt{(u_m^2 - u^2)(R^2 - R_{FOV}^2)}/(R^2 + u^2) \tag{6}$$

---

[2] Scaling all equations by substituting $u \to uR/D$ and $v \to vR/D$ is sufficient to describe the case of a plane detector located at a distance D from the vertex.

This paper deals with the reconstruction of $f(x, y, z)$ from the cone-beam projections acquired on a helix segment $\lambda \in \Lambda(z) = [z/h - \Delta_\Lambda/2, z/h + \Delta_\Lambda/2]$ centered axially on the slice z and having a specified angular range Δ_Λ. As outlined in the introduction, approximate solutions to this problem will be defined by first considering in the next section the reconstruction of a transaxial slice $z = z_0$ from its 2D Radon transform.

3 Two-dimensional fan-beam reconstruction with redundant data

The 2D Radon transform of a slice $f(x, y)$[3] will be parametrized using the fan-beam sampling defined by restricting to 2D the 3D geometry of equation (4):

$$g(\lambda, u) = \sqrt{R^2 + u^2} \int_{l_0(u)-\Delta l(u)}^{l_0(u)+\Delta l(u)} dl \ f\Big(X(\lambda, u, l), Y(\lambda, u, l)\Big) \qquad |u| \leq u_m, \ \lambda \in \Lambda \tag{7}$$

These data correspond to a set of ray-sums diverging from locations of the x-ray source along a segment of circle $\vec{a}_{2D}(\lambda) = (R\cos\lambda, R\sin\lambda), \lambda \in \Lambda = [-\Delta_\Lambda/2, +\Delta_\Lambda/2]$. The length Δ_Λ of this segment must be such that any line crossing $\Omega_{2D} = \Big\{(x, y) \in \mathbb{R}^2 \ \Big| \ x^2 + y^2 \leq R_{FOV}^2\Big\}$ intersects the circle segment at least once (Natterer 1986). This 2D data sufficiency condition is satisfied when the segment is not smaller than the short-scan segment, i.e. when $\Delta_\Lambda \geq \pi + 2\arcsin(R_{FOV}/R)$. In this paper, however, we consider larger segments which may exceed 2π, even though this makes little sense in a 2D context. In this case, the data are redundant and the number $N(\lambda, u)$ of intersections between a ray-sum parameterized by (λ, u) and the circle segment is not constant. To account for this redundancy, one introduces a smooth normalization function $w(\lambda, u)$ such that for any pair $(\lambda \in \Lambda, u \in [-u_m, u_m])$,

$$\sum_{j=1}^{N(\lambda,u)} w(\lambda_j, u_j) = 1 \tag{8}$$

where $\vec{a}_{2D}(\lambda_j), j = 1, \cdots, N(\lambda, u)$ are the intersections of the ray (λ, u) with the circle segment $\vec{a}_{2D}(\Lambda)$, and $u_j, j = 1, \cdots, N(\lambda, u)$ are the corresponding detector coordinates, so that all rays $(\lambda_j, u_j), j = 1, \cdots, N(\lambda, u)$ coincide. The simplest normalization function is $w(\lambda, u) = 1/N(\lambda, u)$ but differentiable functions are prefered for better numerical stability, and are even required by some algorithms.

Starting from the standard FBP algorithm for the inversion of the Radon transform with parallel-beam parameterization, inversion formulae for fan-beam data have been derived by exploiting relations linking the fan-beam data $g(\lambda, u)$, equation (7), to parallel-beam data

$$p(s, \vec{n}) = \int_{\Omega_{2D}} d\vec{r} f(\vec{r}) \ \delta(s - \vec{n} \cdot \vec{r}) \tag{9}$$

---

[3] We consider in this section a fixed slice $z = 0$ and drop the third argument of f.

with the well-known relation between the parallel-beam and fan-beam parameters

$$\begin{aligned} \vec{n} &= (-R\sin\lambda + u\cos\lambda, R\cos\lambda + u\sin\lambda)/\sqrt{R^2+u^2} \\ s &= Ru/\sqrt{R^2+u^2} \end{aligned} \tag{10}$$

Three such 2D fan-beam inversion formulae are given below:

- **FBP-1:** Using the direct link $p(s,\vec{n}) = g(\lambda,u)$, one gets the *standard fan-beam FBP* (Parker 1982):

  $$f(x,y) = \int_\Lambda d\lambda \, \frac{R^2}{(R - x\cos\lambda - y\sin\lambda)^2} \, g_1^F(\lambda, U(x,y,\lambda)) \tag{11}$$

  where

  $$U(x,y,\lambda) = R(-x\sin\lambda + y\cos\lambda)/(R - x\cos\lambda - y\sin\lambda) \tag{12}$$

  is such that the ray defined by $(\lambda, U(x,y,\lambda))$ contains the point (x,y). The filtered projections are given by

  $$g_1^F(\lambda,u') = \int_{-u_m}^{u_m} du \; h_R(u'-u) w(\lambda,u) \frac{R}{\sqrt{R^2+u^2}} \, g(\lambda,u) \tag{13}$$

  where $h_R(u)$ is the kernel of the ramp filter, $h_R(u) = \int_{-\infty}^{\infty} d\nu \; |\nu| \; \exp(2\pi i\nu u)$.

- **FBP-2:** Using the relation of Hamaker (1980) between the Hilbert transforms of $p(s,\vec{n})$ and of $g(\lambda,u)$, one obtains (Kudo *et al* 2002):

  $$\begin{aligned} f(x,y) = \int_\Lambda d\lambda \, \frac{R^2}{(R - x\cos\lambda - y\sin\lambda)^2} \; \Big\{ \; & w(\lambda,U) \; g_{2R}^F(\lambda,U) \\ + \; & \frac{\partial w(\lambda,u)}{\partial u}\Big|_{u=U} \; g_{2H}^F(\lambda,U) \Big\} \end{aligned} \tag{14}$$

  where $U = U(x,y,\lambda)$ is as in equation (12), and the two filtered fan-beam projections are given by

  $$g_{2R}^F(\lambda,u') = \int_{-u_m}^{u_m} du \; h_R(u'-u) \frac{R}{\sqrt{R^2+u^2}} \, g(\lambda,u) \tag{15}$$

  $$g_{2H}^F(\lambda,u') = \frac{1}{2\pi} \int_{-u_m}^{u_m} du \; h_H(u'-u) \frac{R}{\sqrt{R^2+u^2}} \, g(\lambda,u) \tag{16}$$

  where $h_H(u) = \int_{-\infty}^{\infty} d\nu \; (-i) \; sign(\nu) \; \exp(2\pi i\nu u)$ is the kernel of the Hilbert transform.

- **FBP-3:** Finally, the derivative of Hamaker's relation with respect to λ leads to a third inversion formula (Noo *et al* 2002):

  $$f(x,y) = \int_\Lambda d\lambda \, \frac{1}{R - x\cos\lambda - y\sin\lambda} \, w(\lambda,U) \; g_3^F(\lambda,U) \tag{17}$$

where again $U = U(x, y, \lambda)$, and

$$g_3^F(\lambda, u') = \frac{1}{2\pi} \int_{-u_m}^{u_m} du \; h_H(u' - u) \frac{R}{\sqrt{R^2 + u^2}} \left(\frac{\partial}{\partial \lambda} + \frac{R^2 + u^2}{R} \frac{\partial}{\partial u} \right) g(\lambda, u) \qquad (18)$$

Note that this last formula involves the first power of the magnification weight in the backprojection.

Compared to standard fan-beam FBP, the last two algorithms have two important advantages. The first one is that the exact reconstruction of a point (x, y) requires the normalisation condition (8) to be satisfied only for all rays containing (a neighbourhood of) that point. Therefore, some subset of the FOV Ω_{2D} can be reconstructed from fan-beam data acquired along a segment of circle shorter than the short-scan segment $\pi + 2\arcsin(R_{FOV}/R)$ (Noo *et al* 2002). In contrast, the standard fan-beam reconstruction **FBP-1** requires measuring all rays through the object, even if only part of the FOV needs to be reconstructed. The length of the circle segment must then be at least a short-scan $\pi + 2\arcsin(R_{FOV}/R)$.

The second property of the algorithms **FBP-2** and **FBP-3** is that the redundancy weight $w(\lambda, u)$, or its derivative, is applied *after* filtering the data, in contrast with the standard fan-beam FBP of equation (13). As will be seen in the next section, this property simplifies the extension to helical cone-beam data.

4 Two new Feldkamp-type algorithms for helical CT

In this section we extend the 2D fan-beam FBP algorithms to reconstruct the helical cone-beam data defined by equation (4). This empirical extension follows the same approach as that leading to the classical Feldkamp algorithms (Feldkamp *et al* 1984).

Each transaxial slice z is reconstructed by backprojecting the filtered cone-beam projections acquired on the helix segment $\Lambda(z) = [z/h - \Delta_\Lambda/2, z/h + \Delta_\Lambda/2]$ according to equations (11), (14) and (17) for the three fan-beam algorithms respectively. Because we are dealing with cone-beam data $g(\lambda, u, v)$ instead of fan-beam data $g(\lambda, u)$, the axial detector variable v must be inserted into these equations. This is done by choosing v in such a way that $f(\vec{r})$ is reconstructed by summing filtered ray-sums containing the point $\vec{r}$.

The filtering of the fan-beam projections in equations (13), (15), (16) and (18) is extended to cone-beam projections with three modifications:

1. The weighting factors $R/\sqrt{R^2 + u^2}$ are replaced by $R/\sqrt{R^2 + u^2 + v^2}$,
2. The derivative $\partial/\partial\lambda + (R^2 + u^2)/R\,\partial/\partial u$ in equation (18) is a derivative with respect to the parameter λ for a fixed direction of the ray. It is replaced in 3D by the corresponding derivative at fixed *spatial* direction, which is $\partial/\partial\lambda + (R^2 + u^2)/R\,\partial/\partial u + (uv/R)\partial/\partial v$,

3. The direction in the detector along which the ramp and Hilbert filters must be applied is not defined unambiguously when extending empirically the algorithms to 3D. For the usual helical Feldkamp algorithm, the ramp filter has been applied either along the $\vec{1}_u$ axis orthogonal to the helix axis z (Kudo and Saito 1992, Wang *et al* 1993), or along an axis $\vec{1}_{ur}$ rotated by an angle $\eta = \arctan(h/R)$ with respect to the u axis (Yan and Leahy 1992):

$$\vec{1}_{ur} = \cos\eta\ \vec{1}_u + \sin\eta\ \vec{1}_v = (-\cos\eta\sin\lambda, \cos\eta\cos\lambda, \sin\eta) \tag{19}$$

This axis $\vec{1}_{ur}$ is parallel to the tangent $\vec{a}'(\lambda)$ to the helix at the vertex considered (we omit for simplicity the dependence of these unit vectors on λ). Several studies (Turbell and Danielsson 1999, Sourbelle and Kalender 2003) indicate that filtering along $\vec{1}_{ur}$ instead of $\vec{1}_u$ significantly reduces cone-beam artefacts, and this approach will be selected in this paper.

We are now ready to define the three Feldkamp-type algorithms:

- **FDK-1** The standard FDK algorithm is the 3D extension of **FBP-1**:

$$f(\vec{r}) \ = \ \int_{\Lambda(z)} d\lambda\ \frac{R^2}{(R - x\cos\lambda - y\sin\lambda)^2}\ g_1^F(\lambda, U, V) \tag{20}$$

where $U = U(x, y, \lambda)$ as in the previous section and

$$V = V(x, y, z, \lambda) = R(z - h\lambda)/(R - x\cos\lambda - y\sin\lambda) \tag{21}$$

is such that the ray defined by λ, U, V contains the point $\vec{r} = (x, y, z)$. The filtered projections are given by,

$$g_1^F(\lambda, u', v') \ = \ \int_{-u_m}^{u_m} du\ h_R(u' - u)\, g_w(\lambda, u, v' - (u' - u)\tan\eta) \tag{22}$$

where $g_w(\lambda, u, v)$ are the cone-beam data multiplied by the usual Feldkamp's weight and by the redundancy weight,

$$g_w(\lambda, u, v) = \frac{R}{\sqrt{R^2 + u^2 + v^2}}\ w(\lambda - z/h, u)\, g(\lambda, u, v) \tag{23}$$

where w is any solution to the normalization condition (8). Note that w was defined for an angular range $[-\Delta_\Lambda/2, +\Delta_\Lambda/2]$, and therefore its first argument in equation (23) is shifted by $-z/h$ to center the angular range on the center of the helix segment $\Lambda(z)$ used to reconstruct slice z. This dependence on z in equations (23) and (22) implies that each row in a cone-beam projection must be filtered more than once (unless $\Delta_\Lambda = 2\pi, 4\pi, \cdots$). This drawback is avoided with the two new algorithms presented below.

- **FDK-2** The first new algorithm is derived from **FBP-2**:

$$\begin{aligned} f(\vec{r}) &= \int_{\Lambda(z)} d\lambda \, \frac{R^2}{(R - x\cos\lambda - y\sin\lambda)^2} \times \qquad (24) \\ & \Big(w(\lambda - z/h, U) \; g_{2R}^F(\lambda, U, V) + \frac{\partial w(\lambda - z/h, u)}{\partial u}\Big|_{u=U} \; g_{2H}^F(\lambda, U, V) \Big) \end{aligned}$$

with $U = U(x, y, \lambda)$ and $V = V(x, y, z, \lambda)$ as before, and the filtered projections

$$\begin{aligned} g_{2R}^F(\lambda, u', v') &= \int_{-u_m}^{u_m} du \; h_R(u' - u) \; g_w(\lambda, u, v' - (u' - u)\tan\eta) \\ g_{2H}^F(\lambda, u', v') &= \frac{1}{2\pi} \int_{-u_m}^{u_m} du \; h_H(u' - u) \; g_w(\lambda, u, v' - (u' - u)\tan\eta) \qquad (25) \end{aligned}$$

where $g_w(\lambda, u, v)$ are the cone-beam data multiplied by the usual Feldkamp weight,

$$g_w(\lambda, u, v) = \frac{R}{\sqrt{R^2 + u^2 + v^2}} \; g(\lambda, u, v) \qquad (26)$$

- **FDK-3** The second new algorithm is derived from **FBP-3**:

$$f(\vec{r}) = \int_{\Lambda(z)} d\lambda \, \frac{1}{R - x\cos\lambda - y\sin\lambda} \; w(\lambda - z/h, U) \; g_3^F(\lambda, U, V) \qquad (27)$$

with the same U and V as above, and

$$g_3^F(\lambda, u', v') = \frac{1}{2\pi} \int_{-u_m}^{u_m} du \; h_H(u' - u) \; \tilde{g}_w(\lambda, u, v' - (u' - u)\tan\eta) \qquad (28)$$

where $\tilde{g}_w$ is obtained by differentiating and weighting the cone-beam projections:

$$\tilde{g}_w(\lambda, u, v) = \frac{R}{\sqrt{R^2 + u^2 + v^2}} \left(\frac{\partial}{\partial\lambda} + \frac{R^2 + u^2}{R}\frac{\partial}{\partial u} + \frac{uv}{R}\frac{\partial}{\partial v} \right) g(\lambda, u, v) \qquad (29)$$

The three algorithms above yield exact reconstruction when $f(\vec{r}) = f(x, y)$ is independent of z. To prove this assertion, it is sufficient to note that the fan-beam data $g_{2D}(\lambda, u)$ of any slice $z = const.$ of a z-homogeneous 3D object are related to the cone-beam data (4) by $g(\lambda, u, v) = \sqrt{R^2 + u^2 + v^2}\, g_{2D}(\lambda, u)/\sqrt{R^2 + u^2}$. Inserting this equation into the algorithms **FDK-1,2,3**, one obtains[4] the algorithms **FBP-1,2,3** respectively, applied to the data g_{2D}. The fact that these fan-beam algorithms are exact then proves the assertion.

---

[4] To guarantee this property, the filtering in equations (22),(25) and (28) is written with $u, v' - (u' - u)\tan\eta$ instead of $u' - (u - u')\cos\eta, v' - (u' - u)\sin\eta$. The two expressions give identical results with the Hilbert filter, but differ by a scaling factor $|\cos\eta|$ with the ramp filter.

5 Link with the exact algorithm of Katsevich

The three algorithms in the previous section can reconstruct each slice from data acquired along an arbitrary segment of helix. This versatility allows the detector data to be exploited in an efficient way without being restricted to specific configurations such as the PI- or the 3 PI-windows used by the Katsevich algorithms. In contrast with the latter algorithms, however, FDK-type methods are not exact and their accuracy decreases with increasing helical pitch or with increasing range Δ_Λ. A characterization of the approximation provides insight into the nature of the expected reconstruction artefacts. In this section, such a characterization is obtained for the **FDK-3** algorithm, using the link between this algorithm and the 3D Radon transform of $f(\vec{r})$.

We start from the general inversion formula of Katsevich (2002), which can be rewritten in terms of planar detector coordinates as (see Bontus *et al* 2003 and the appendix in Pack *et al* 2003):

$$f_M(\vec{r}) = \int_{\mathbb{R}} d\lambda \, \frac{1}{R - x\cos\lambda - y\sin\lambda} \, w_M(\lambda, \vec{r}) \; g^F(\vec{r}, \lambda, U, V) \tag{30}$$

with the same U and V as above, and

$$g^F(\vec{r}, \lambda, u', v') \;\; = \;\; \frac{1}{2\pi} \int_{-u_m}^{u_m} du \; h_H(u' - u) \; \tilde{g}_w(\lambda, u, v' - (u' - u)\tan\eta(\lambda, \vec{r})) \tag{31}$$

with $\tilde{g}_w$ given by equation (29). Comparison with equations (27) and (28) shows that this algorithm is similar to **FDK-3**, with two differences:

- The segment of helix $\Lambda(\vec{r})$ used to reconstruct a point $\vec{r}$ is allowed to depend on x, y, z and not only on the slice z. In equation (30) this segment is defined implicitly by the weighting function w_M, as $\Lambda(\vec{r}) = \{\lambda \in \mathbb{R} \mid w_M(\lambda, \vec{r}) \neq 0\}$.
- The direction along which the Hilbert filter is applied in the detector is no longer constant and is defined by the angle $\eta(\lambda, \vec{r})$[5].

As shown in (Katsevich 2003) the algorithm (30),(31) is equivalent to an inversion of the second derivative of the 3D Radon transform of f, in which each plane $\Pi(\vec{r}, \vec{\theta})$ containing $\vec{r}$ and normal to a unit vector $\vec{\theta} \in S^2$ is weighted by a factor $M(\vec{r}, \vec{\theta})$:

$$f_M(\vec{r}) = \frac{-1}{8\pi^2} \int_{S^2} d\vec{\theta} \; M(\vec{r}, \vec{\theta}) \; \frac{\partial^2 Rf(\vec{\theta}, \rho)}{\partial \rho^2}\Big|_{\rho = \vec{\theta}\cdot\vec{r}} \tag{32}$$

where S^2 is the unit sphere and $Rf(\vec{\theta}, \rho)$ is the 3D Radon transform of f. The link with equation (30) is given by

$$M(\vec{r}, \vec{\theta}) = \sum_{j=1}^{N(\vec{r}, \vec{\theta})} w_M(\lambda_j, \vec{r}) \; sign(\vec{a}'(\lambda_j) \cdot \vec{\theta}) \; sign(\vec{\xi}(\lambda_j, \vec{r}) \cdot \vec{\theta}) \tag{33}$$

---

[5] Via this dependence of η on $\vec{r}$, the filtered projection depends on the point $\vec{r}$, hence the algorithm entails in general repeated filtering of each cone-beam projection for each point $\vec{r}$, unless η depends on $\vec{r}$ only through $U(x, y, \lambda), V(x, y, z, \lambda)$.

where the sum is over all intersections $\vec{a}(\lambda_j), j = 1, \cdots, N(\vec{r}, \vec{\theta})$ of the plane $\Pi(\vec{r}, \vec{\theta})$ with the helix segment $\vec{a}[\Lambda(\vec{r})]$ defined by the support of $w_M(\lambda, \vec{r})$ with respect to λ. The vector $\vec{\xi}(\vec{r}, \lambda)$ in equation (33) is defined by

$$\vec{\xi}(\vec{r}, \lambda) = \cos\eta(\lambda, \vec{r})\vec{1}_u(\lambda) + \sin\eta(\lambda, \vec{r})\vec{1}_v(\lambda) \tag{34}$$

For a specific choice of $w_M(\lambda, \vec{r})$ and $\eta(\lambda, \vec{r})$, the corresponding weighting of the Radon plane can be calculated using equation (33). The algorithm provides an exact reconstruction at point $\vec{r}$ if $M(\vec{r}, \vec{\theta}) = 1$ for any $\vec{\theta} \in S^2$. When $M(\vec{r}, \vec{\theta}) \neq 1$, the algorithm is approximate and the reconstruction artefacts around a point $\vec{r}$ can be characterized geometrically by analysing the deviation $|M(\vec{r}, \vec{\theta}) - 1|$ as function of the orientation $\vec{\theta}$ of the Radon plane through $\vec{r}$. Specifically, a surface of discontinuity (or of rapid variation) of f at point $\vec{r}$, which is orthogonal to some direction $\vec{\theta}$, will be reconstructed accurately only if $M(\vec{r}, \vec{\theta})$ does not significantly differ from 1.

5.1 The Katsevich algorithm for PI-reconstruction

The exact algorithm of Katsevich corresponds to the following choice for w_M and $\vec{\xi}$:

- $\Lambda(\vec{r}) = [\lambda_b(\vec{r}), \lambda_t(\vec{r})]$ is the PI-segment of point $\vec{r}$, defined by the unique (Edholm *et al* 1997) pair of vertices such that $\lambda_b < \lambda_t < \lambda_b + 2\pi$ and the three points $\vec{a}(\lambda_b), \vec{a}(\lambda_t)$ and $\vec{r}$ are colinear. Thus,

  $$w_M(\lambda, \vec{r}) = \begin{cases} 1 & \lambda_b(\vec{r}) \leq \lambda \leq \lambda_t(\vec{r}) \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

  The planes $\Pi(\vec{r}, \vec{\theta})$ have either one or three intersections with $\Lambda(\vec{r})$, i.e. $N(\vec{r}, \vec{\theta}) = 1$ or 3.

- $\vec{\xi}(\vec{r}, \lambda)$ is a vector along the intersection of the detector plane with a plane containing the four points $\vec{r}, \vec{a}(\lambda), \vec{a}(\lambda + s)$ and $\vec{a}(\lambda + 2s)$. It can be shown that $\vec{\xi}$ is unique if s is restricted to the PI-interval $\lambda_b(\vec{r}) \leq \lambda + 2s \leq \lambda_t(\vec{r})$.

The algorithm defined by equations (30) and (31) with the above choice of w_M and $\vec{\xi}$ admits an efficient implementation because $w_M(\lambda, \vec{r})$ and the direction of filtering $\eta(\lambda, \vec{r})$ depend on $\vec{r}$ only through the coordinates U and V of the projection of $\vec{r}$ onto the detector. Specifically,

$$w_M(\lambda, \vec{r}) = \begin{cases} 1 & -V_{Tam}(-U) \leq \lambda \leq V_{Tam}(U) \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

with

$$V_{Tam}(U) = h(1 + \frac{U^2}{R^2})(\frac{\pi}{2} - \arctan\frac{U}{R}) \tag{37}$$

i.e. w_M is equal to 1 if the projection of $\vec{r}$ is inside Tam's window and zero otherwise, where the Tam window (Tam *et al* 1998) is the detector area bounded by the projection onto the detector of the helix turns directly below and above the source position. The Hilbert filter in equation (31) is applied along a family of straight lines in the detector. These lines depend on a parameter s and are defined by

$$\mathcal{L}_s : \quad v = hs + u\frac{h}{R} s \cot s \quad -\pi \leq s \leq \pi \tag{38}$$

This line is the intersection of the detector plane with the plane defined by $\vec{a}(\lambda)$, $\vec{a}(\lambda+s)$ and $\vec{a}(\lambda+2s)$. A point $\vec{r}$ is reconstructed using the data filtered along the line $\mathcal{L}_s$ that corresponds to the smallest value of $|s|$ such that $(U(\vec{r},\lambda), V(\vec{r},\lambda)) \in \mathcal{L}_s$, and $\tan\eta(\lambda,\vec{r}) = (hs\cot s)/R$.

5.2 The FDK-3 algorithm for an arbitrary helix segment

The Feldkamp-type algorithm **FDK-3** in section 4 can be seen by direct substitution to be the general algorithm of equations (30),(31) with $\eta(\lambda,\vec{r}) = \arctan(h/R)$ and

$$w_M(\lambda,\vec{r}) = \begin{cases} w(\lambda, U(x,y,\lambda)) & \text{if } \lambda \in \Lambda(z) \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

where $\Lambda(z) = [z/h - \Delta_\Lambda/2, z/h + \Delta_\Lambda/2]$ and $w(\lambda,u)$ is the weighting function introduced in section 3. Equation (33) then becomes

$$M(\vec{r},\vec{\theta}) = \sum_{j=1}^{N(\vec{r},\vec{\theta})} w(\lambda_j - z/h, U(x,y,\lambda_j)) \tag{40}$$

By construction of the fan-beam redundancy weight (see equation (8)), $M(\vec{r},\vec{\theta}) = 1$ for all planes parallel to the z axis $\vec{1}_z$, i.e. such that $\vec{\theta}\cdot\vec{1}_z = 0$. For other planes M may differ from 1, and the algorithm **FDK-3** is therefore approximate.

6 Frequency-Mixing approach

The FDK-type approximate algorithms (**FDK-1,2,3**) described in section 4 work well when the segment of helix $\Lambda(z)$ used to reconstruct each transaxial slice z is not too long and the pitch $2\pi h$ is moderate. In this case the rays used to reconstruct a point $\vec{r}$ can be treated to a good approximation as if they were all lying in the same plane. When the segment length Δ_Λ or the pitch $2\pi h$ become too large, this approximation breaks down, leading to unacceptable artefacts. This section proposes an empirical, but simple and effective, method which reduces these cone-beam artefacts and allows an improved compromise between image accuracy and image noise for a selected value of Δ_Λ. This approach was originally disclosed in (Kudo *et al* 2003), and independently proposed by Shechter *et al* (2003) for the rebinning-type algorithms.

The proposed method is based on the same idea as the pseudo-FBP algorithm developed by Tanaka *et al* (1992) for 3-D positron emission tomography. The key observation is that the accuracy of FDK-type reconstructions can be analysed in terms of the 3D Radon transform of the object. Specifically, as shown in section 5, the major artefacts caused by the algorithms **FDK-3** arise from the incorrect redundancy handling of Radon planes which are nearly orthogonal to the z-axis (see also figure 2 in section 7.1). Owing to the central section theorem for the 3D Radon transform, this means that the local modulation transfer function (MTF) of the reconstruction algorithm deviates from unity mainly within a cone of frequencies which is centered at the origin and is directed along the z-axis. This cone corresponds to low frequencies with respect to x and y. In contrast, the noise in tomographic reconstruction essentially affects the high spatial frequencies.

This observation suggests the following *frequency-mixing* method to reduce the cone-beam artefacts while at the same time optimizing image noise. First, the projection data $g(\lambda, u, v)$ are decomposed into high- and low-frequency components $g^H(\lambda, u, v)$ and $g^L(\lambda, u, v)$ with respect to the detector coordinate u. The high-frequency component $g^H(\lambda, u, v)$ is then reconstructed using a long segment of helix (i.e. a large value of Δ_Λ) which maximally utilizes the available detector data. On the contrary, the low-frequency component $g^L(\lambda, u, v)$ is reconstructed with a small value of Δ_Λ (e.g. a short-scan segment) so as to minimize the cone-beam artefacts. Finally, the two reconstructed components $f^H(\vec{r})$ and $f^L(\vec{r})$ are summed up to obtain $f(\vec{r})$. This approach can be applied to any of the **FDK-1,2,3** algorithms.

The cone-beam problem is not a convolution problem and therefore the decomposition into low- and high-frequency components is not strictly valid[6]. To minimize possible errors due by this limitation, and also to minimize discretization errors, the frequency splitting should not be sharp. Our current implementation uses the same Gaussian filter as Tanaka *et al* (1992) to decompose the projection data:

$$\begin{aligned} g^L(\lambda, u', v) &= \int_{-u_m}^{u_m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(u'-u)^2/(2\sigma^2)} g(\lambda, u, v) du \\ g^H(\lambda, u, v) &= g(\lambda, u, v) - g^L(\lambda, u, v) \end{aligned} \tag{41}$$

The filter parameter σ corresponds to the inverse of a smooth cut-off frequency, and is determined empirically.

7 Results

7.1 The Radon weighting in the FDK-3 algorithm

The Radon weighting function $M(\vec{r}, \vec{\theta})$ was calculated using equation (40). The radius of the helix was $R = 400$ mm and the pitch was $P = 46.28$ mm. Each image in figure 2 shows the Radon weight for all planes containing the point $\vec{r} = (x = 0, y = -90 \text{ mm }, z = 0)$, reconstructed from a helix segment centered on the slice $z = 0$ and of length 1.5π or 3π. The

---

[6] More precisely, $g^H(\lambda, u, v)$ and $g^L(\lambda, u, v)$ do not separately satisfy the data consistency conditions.

redundancy weight is defined as the following solution to equation (8):

$$w(\lambda, u) = \frac{c(\lambda)}{\sum_{j=1}^{N(\lambda,u)} c(\lambda_j)} \qquad \lambda \in \Lambda = [-\Delta_\Lambda/2, +\Delta_\Lambda/2] \tag{42}$$

where $N(\lambda, u)$ and λ_j are defined as in section 3, and $c(\lambda)$ is any smooth function which vanishes at the extremities of the helix segment. We use

$$c(\lambda) = \begin{cases} 1 & \text{if } |\lambda| \leq \Delta_\Lambda/2 - m \\ \sin^2\left(\frac{\pi}{2}\, \frac{\Delta_\Lambda/2 - |\lambda|}{m}\right) & \text{otherwise} \end{cases} \tag{43}$$

The width of the transition region was $m = 0.50$ radian. The same weighting function was used for all simulations.

In figure 2, the vertical axis in each image is $\vec{\theta} \cdot \vec{1}_z$, the cosine of the polar angle of the unit vector $\vec{\theta}$ perpendicular to a Radon plane. In all cases, $M(\vec{r}, \vec{\theta}) = 1$ on the central row, which corresponds to planes parallel to the axis of the helix: these planes are exactly normalized by the fan-beam weight w. The closer the plane is to an endpoint of the helix, the larger the deviation from $M = 1$, with the maximum deviations occuring for planes that are orthogonal to the axis of the helix (i.e. occuring in the first and last rows of each image).

7.2 The HEAD phantom

Helical data have been simulated for the FORBILD (www.imp.uni-erlangen.de/forbild) HEAD phantom (Schaller 1998). The HEAD phantom was scaled by a factor 0.736, and four co-centric thin disks (radius 14.7 mm, thickness 3 mm, spacing 6 mm) were added to better elicit the cone-beam artefacts. The simulated value of each detector pixel was the average of 9 analytically calculated line integrals of the phantom. These lines intersected the detector in 9 points uniformly distributed over the corresponding detector pixel, so as to roughly simulate the effect of the point spread function of the detector. The focus of the x-ray source was treated as a single point.

The parameters of the simulation and of the reconstruction are summarized in table 1. The ramp and Hilbert filters were apodized with a Hamming window cut-off at the Nyquist frequency. The derivatives were discretized with a kernel $-1, 1$. Figures 4 and 3 illustrate the good agreement between the reconstructions obtained with the two new algorithms and the expected degradation of image quality with increasing length of the helix segment. This degradation is most evident at the level of the disk structures, as expected from the result in section 6. Note that in this simulation, the pitch was fixed, and the cone-angle was modified so that the used helix segment varies between 1.5π and 3π. This is different from an actual CT system, in which the cone-angle is fixed and the maximum usable helix segment is determined by the pitch.

For comparison, figure 5 shows the same longitudinal section as in figure 4, reconstructed with the exact algorithm of Katsevich and with the usual helical FDK algorithm

Table 1: Simulation and reconstruction parameters.

phantom	HEAD	SPHERES
R (mm)	400	400
$P = 2\pi h$ (mm)	46.28	46.28
projections per turn	1200	1200
helix turns	3.5	3.5
image dimensions	400x400x200	128x128x128
voxel size $\Delta x = \Delta y = \Delta z$ (mm)	0.50	0.05
projection dimensions	400x200	400x200
proj. pixel size $\Delta u = \Delta v$ (mm)	0.50	0.50

Table 2: Reconstruction times (hours).

	1.5π	3π
FDK-1	3.7	9.2
FDK-2	2.4	4.9
FDK-3	2.2	4.6

FDK-1 applied to a helix segment of 3π. Note the similarity of the three algorithms **FDK-1,2,3**.

The reconstruction times are given in table 2 for an Alpha 21264 processor (667 MHz) with 512 Mbytes memory. The longer time required by **FDK-1** is due to the repeated filtering of each projection row. The Katsevich algorithm uses a minimum set of data and took 2.9 hours.

7.3 Resolution phantom

The reconstructed resolution was investigated by simulating helical data for 5 spheres of diameter 2 mm, with centers in the transaxial section $z = 0$. One sphere is located at the center, the other four are centered respectively at $x = 0, y = 90$ mm (North), $x = 90$ mm, $y = 0$ (East), $x = 0, y = -90$ mm (South), and $x = -90$ mm, $y = 0$ (West). The intersection of the helix with the slice $z = 0$ is located at $x = 0, y = -400$ mm and is therefore nearest to the "South" source. After reconstruction, each reconstructed sphere was compared with the exact image of the sphere that has been convolved with a 3D gaussian kernel. This kernel was characterized by its three half-axes (converted to FWHM values) and by the orientation of these axes, specified by three angles. These six parameters were determined to optimize the fit between the reconstruction and its model. The results are shown in tables 3 and 4 for the algorithms **FDK-2** and **FDK-3** and are summarized graphically in figure 6, where

Table 3: PSF 3D measured in different location for **FDK-2**. The last column gives the root mean square deviation between the reconstructed and modelled PSF.

Helix	location	σ_x (pixels)	σ_y (pixels)	σ_z (pixels)	θ (°)	ϕ (°)	ψ (°)	fit (in %)
1.5π	center	8.4	8.4	5.4	-3	3	9	2.63
1.5π	east	6.6	9.1	5.2	-3	37.0	3	3.01
1.5π	north	10.5	7.7	6.1	-3	0.0	0.0	2.29
1.5π	west	6.6	9.1	5.2	-3	-37.0	-3	2.72
1.5π	south	6.2	7.8	4.6	-3	0.0	0.0	2.75
3π	center	8.4	8.4	5.4	-3	3	-86	2.67
3π	east	7.5	8.4	5.4	-6	-25.8	0.0	2.57
3π	north	7.5	7.6	5.1	-3	3	-9	2.63
3π	west	7.5	8.4	5.4	-6	25.8	0.0	2.67
3π	south	9.0	7.6	5.5	-3	0.0	0.0	2.49

Table 4: PSF 3D measured in different location for **FDK-3**. The last column gives the root mean square deviation between the reconstructed and modelled PSF.

Helix	location	σ_x (pixels)	σ_y (pixels)	σ_z (pixels)	θ (°)	ϕ (°)	ψ (°)	fit (in %)
1.5π	center	9.2	9.1	6.2	-3	0.0	23	2.18
1.5π	east	7.6	10.6	5.6	-3	29.0	0.0	2.16
1.5π	north	12.3	8.4	6.5	0.0	0.0	0.0	1.70
1.5π	west	7.7	10.5	5.7	-3	-29.0	0.0	2.04
1.5π	south	8.1	8.5	5.0	-3	0.0	0.0	1.96
3π	center	9.2	9.1	6.3	-3	0.0	8.6	2.21
3π	east	8.3	10.2	5.8	-6	-11.5	0.0	1.86
3π	north	9.4	8.3	5.5	0.0	0.0	0.0	1.85
3π	west	8.3	10.2	5.7	-6	11.5	0.0	1.92
3π	south	10.8	8.2	5.9	-3	0.0	0.0	1.81

transaxial isocontours of the gaussian fit of the point response function are represented for the five positions and for two segment lengths $\Delta_\Lambda = 1.5\pi$ and 3π. The anisotropy and inhomogeneity of the resolution can be understood qualitatively by analyzing the relative position of the helix segment with respect to the various cylinders, taking into account the degradation of the spatial resolution at increasing distances from the source. Note that the fact that the algorithms **FDK-2** and **FDK-3** respectively apply a quadratic and a linear magnification weight for cone-beam backprojection does not seem to significantly influence the spatial variations or the anisotropy of the resolution.

7.4 Noise and the Frequency-Mixing algorithm

Increasing the helix segment used for reconstruction incorporates a larger fraction of the available detector data, and thereby improves the image SNR. This is illustrated for the **FDK-2** algorithm in the first row of figure 7, where pseudo-random Poisson noise has been added to the simulated data of the HEAD phantom, corresponding to an average incident flux of 200000 photons per detector pixel. Even though a quantitative analysis is beyond the scope of this paper, it is clear that image noise is significantly lower with the 3π acquisition. The noise behaviour of the algorithms **FDK-2** and **FDK-3** is similar (data not shown).

The Katsevich algorithm reconstructs each point using a PI-segment of helix, the length of which varies between π and 1.5π, resulting in an increased noise compared to the **FDK-2** reconstruction.

Figure 7 also illustrates the validity of the frequency-mixing approach (section 6) when applied to **FDK-2**. The parameter σ in equation (41) was fixed as $\sigma = \sqrt{2}/(0.5\,\pi\,\nu_N)$ where $\nu_N = 1/(2\Delta u)$ is the Nyquist frequency for the sampling in u. With the large pitch used in this simulation, the reconstruction from a 3π segment with the **FDK-2** algorithm produced artefacts around the high-contrast structures inside the phantom. These artefacts can be suppressed by using non-redundant short-scan data (**FDK-2** with a 1.5π segment, or the Katsevich algorithm), but the price to pay is a significant increase of the noise level. As illustrated, the Frequency-Mixing algorithm yields the best of the two methods by yielding a reconstruction with largely suppressed artefacts and with a low noise level. Worth noting is the fact that the computational time required by this algorithm was only 20% longer than for the algorithm with redundant data. This is thanks to the fact that the filtered high- and low-frequency components can be added after multiplication by the respective weights, and then backprojected simultaneously.

8 Discussions and conclusions

We have presented two approximate algorithms (**FDK-2** and **FDK-3**) for helical cone-beam data acquired along a helix segment of arbitrary length. These algorithms are similar to the standard Feldkamp method (**FDK-1**) and have similar performance in terms of accuracy and noise behaviour. In contrast with the Feldkamp method, however, the new algorithms apply the Parker-like redundancy weight *after* filtering; this property allows a more efficient numerical implementation (see table 2) because each cone-beam projection needs only be filtered once.

Just as for the standard method (**FDK-1**), the cone-beam artefacts observed with the two new algorithms increase when the length of the helix segment is increased. The second contribution of this paper aims at suppressing these artefacts without losing the benefit of an efficient data utilization. This is achieved by the frequency-mixing method proposed in section 6. This methods reconstructs separately the low and the high spatial frequency components of the image, using respectively a short helix segment (to minimize the cone-beam artefacts) and a longer helix segment (to minimize high frequency noise).

This approach can be applied to any approximate algorithm, although the evaluation in this paper is done with the **FDK-2** algorithm only. The results in section 7.4 demonstrate the potential of this method and its numerical efficiency.

Even though the Feldkamp-type algorithms allow to use an arbitrary segment of helix, data utilization is not yet optimal, because a fixed helix segment is used for each slice, and the number of detector rows must be sufficient to provide all required data for the *whole slice*. This constraint restricts the length of the usable helix segment. Taguchi *et al* (2003) recently proposed to overcome this limitation by extrapolating the data axially, thereby allowing the use of a longer segment of helix, such that each detected ray-sum is incorporated into the reconstruction. The two new algorithms presented here suggest an alternative approach: because the redundancy weight is applied after filtering, it becomes possible, numerically, to use a different weight for each pixel in the slice, so as to further improve data utilization. This approach will be investigated in future work.

References

[1] Bontus C, Köhler T, and Proksa R 2003 A quasiexact reconstruction algorithm for helical CT using a 3-Pi acquisition *Med. Phys.* **30** 2493-2502

[2] Danielsson P E, Edholm P, Eriksson J and Magnusson Seger M 1997 Towards exact reconstruction for helical cone-beam scanning of long objects. A new detector arrangement and a new completeness condition, *Proc. 1997 Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine (Pittbsurgh, PA)*, ed D W Townsend and P Kinahan, 141-4.

[3] Defrise M, Noo F and Kudo H 2001 Rebinning based algorithms for helical cone-beam CT *Phys. Med. Biol.* **46** 2911-2937

[4] Defrise M, Noo F and Kudo H 2003 Improved two-dimensional rebinning of helical cone-beam computerized tomography data using John's equation *Inverse Problems* **19** S41-S54

[5] Feldkamp L A, Davis L C and Kress J W 1984 Practical cone-beam algorithm *J. Opt. Soc. Am.* **A6** 612-619

[6] Flohr Th, Stierstorfer K, Bruder H, Simon J, Polacin A and Schaller S 2003 Image reconstruction and image quality evaluation for a 16-slice CT scanner *Med. Phys.* **30** 832-845

[7] Hamaker C, Smith K T, Solmon D C and Wagner S L 1980 The divergent beam x-ray transform *Rocky Mountain J. Math.* **10** 253-83

[8] Heuscher D 1999 Helical cone beam scans using oblique 2D surface reconstruction *Records of the 1999 Int. Meeting on 3D Image Reconstruction* 204-207

[9] Kachelriess M, Schaller S and Kalender W 2000 Advanced single-slice rebinning in cone-beam spiral CT *Med. Phys.* **27** 754-772

[10] Katsevich A 2002 Improved exact FBP algorithm for spiral CT, submitted to *Adv. Appl. Math.*

[11] Katsevich A 2002a Analysis of an exact inversion algorithm for spiral cone-beam CT *Phys. Med. Biol.* **47** 2583-2597.

[12] Katsevich A 2003 A general scheme for constructing inversion algorithms for cone-beam CT *Int. J. of Math. and Math. Sc.* **21** 1305-21

[13] Kudo H and Saito T 1992 Helical-scan computed tomography using cone-beam projections *Conf. Rec. 1991 IEEE Med. Imag. Conf. (Santa Fe, NM)* 1958-1962

[14] Kudo H, Park S, Noo F and Defrise M 1999 Performance of quasi-exact cone-beam filtered-backprojection algorithms for axially truncated helical data *IEEE Trans. Nucl. Sc.* **NS-46** 608-17

[15] Kudo H, Noo F, Defrise M and Clackdoyle R 2002 New super-short-scan algorithms for fan-beam and cone-beam reconstruction, *records of the 2002 IEEE Nuclear Science and Medical Imaging Symposium* (paper M5-3, CD-ROM, ISBN 0-7803-7637-4).

[16] Kudo H, Noo F, Defrise M and Rodet T 2003 New approximate filtered backprojection algorithm for helical cone-beam CT with redundant data *records of the 2003 IEEE Nuclear Science and Medical Imaging Symposium* (paper M14-330, CD-ROM, to appear).

[17] Larson G L, Ruth C C and Crawford C R 1998 Nutating slice CT image reconstruction. Patent Application WO 98/44847.

[18] Natterer F 1986 The Mathematics of Computerized Tomography, Wiley.

[19] Noo F, Defrise M and Clackdoyle R 1999 Single-slice rebinning for helical cone-beam CT *Phys. Med. Biol.* **44** 561-570

[20] Noo F, Defrise M, Clackdoyle R and Kudo H 2002 Image reconstruction from fan-beam projections on less than a short-scan *Phys. Med. Biol.* **47** 2525-46

[21] Pack J D, Noo F and Kudo H 2003 Investigation of saddle trajectories for cardiac CT imaging in cone-beam geometry, submitted to *Phys. Med. Biol.*

[22] Parker D L 1982 Optimal short scan convolution reconstruction for fan-beam CT *Med. Phys.* **9** 254-257

[23] Schaller S 1998 Practical image reconstruction for cone-beam tomography *PhD thesis* University of Erlangen

[24] Shechter G, Altman A, Koehler T and Proksa R 2003 High-resolution images of cone beam collimated CT scans *records of the 2003 IEEE Nuclear Science and Medical Imaging Symposium* (paper M11-323, CD-ROM, to appear).

[25] Silver M D 2000 A method for including redundant data in computed tomography, *Med. Phys.* **27** 773-774

[26] Sourbelle K and Kalender W 2003 Generalization of Feldkamp reconstruction for clinical spiral cone-beam CT *Proc. 2003 Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine (St Malo, France)*, ed Y Bizais.

[27] Stierstorfer K, Flohr T and Bruder H 2002 Segmented multi-plane reconstruction: a novel approximate reconstruction scheme for multi-slice spiral CT *Phys. Med. Biol.* **47** 2571-2581

[28] Taguchi K, Chiang B and Silver M 2003 New weighting scheme for cone-beam helical CT to reduce the image noise, *Proc. 2003 Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine (St Malo, France)*, ed Y Bizais.

[29] Tam K C, Samarasekera S and Sauer K 1998 Exact cone-beam CT with a spiral scan *Phys. Med. Biol.* **43** 1015-24

[30] Tanaka E, Mori S, Shimizu K, Yoshikawa E, Yamashita T and Murayama H 1992 Moving slice septa and pseudo 3-D reconstruction for multi-ring PET *Phys. Med. Biol.* **37** 661-672

[31] Tang X 2003 Matched view weighting in tilted-plane-based reconstruction algorithms to suppress helical artifacts and optimize noise characteristics *Med. Phys.* **30** 2912-2918

[32] Turbell H and Danielsson P E 1999 An improved PI-method for reconstruction from helical cone-beam projections *records of the 1999 IEEE Nuclear Science and Medical Imaging Symposium* (Seattle, WA), paper M2-7.

[33] Wang G, Lin T H, Cheng P and Shinozaki D M 1993 A general cone-beam reconstruction algorithm *IEEE Trans Med Imag* **MI-12** 486-496

[34] Wesarg S, Ebert M and Bortfeld T 2002 Parker weights revisited *Med. Phys.* **29** 372-378

[35] Yan X and Leahy R M 1992 Cone-beam tomography with circular, elliptical and spiral orbits *Phys. Med. Biol.* **37** 493-506

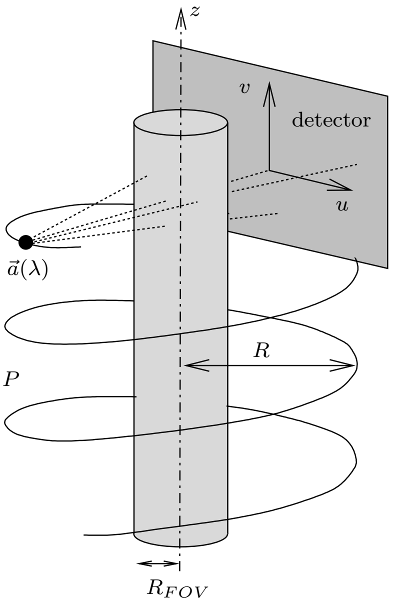
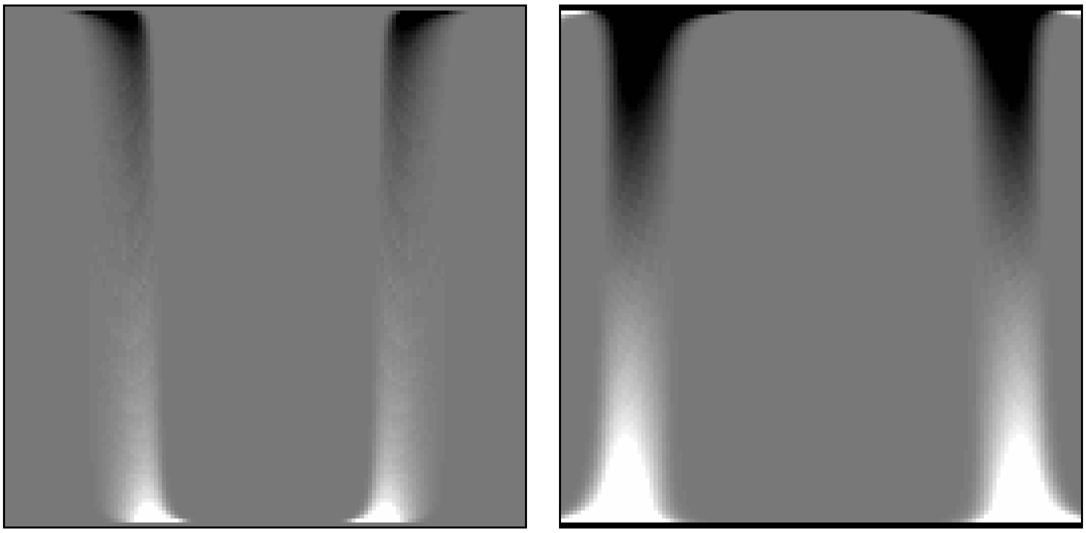


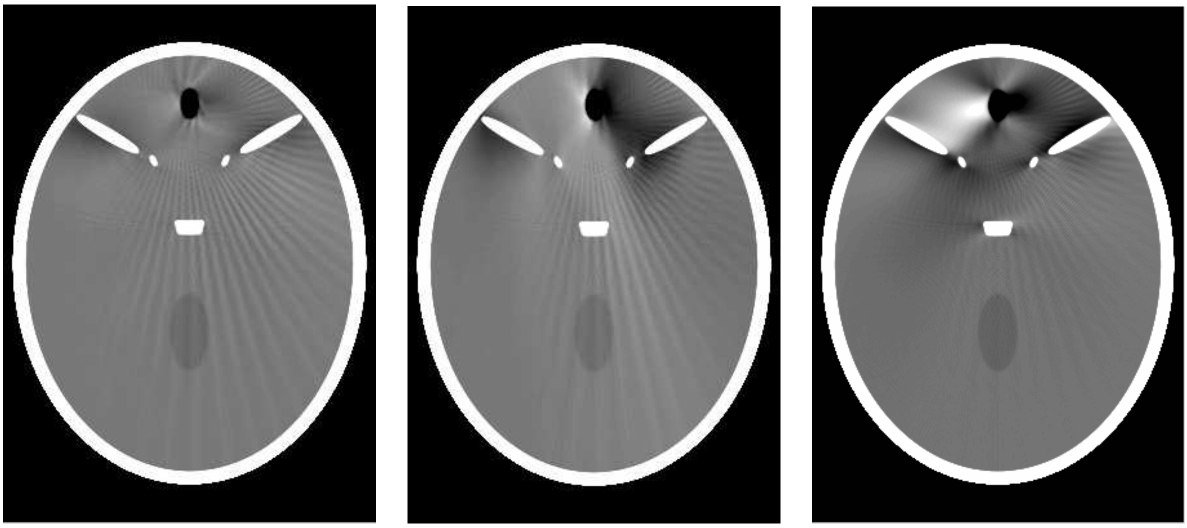
Figure 1: Geometry of the cone-beam acquisition.



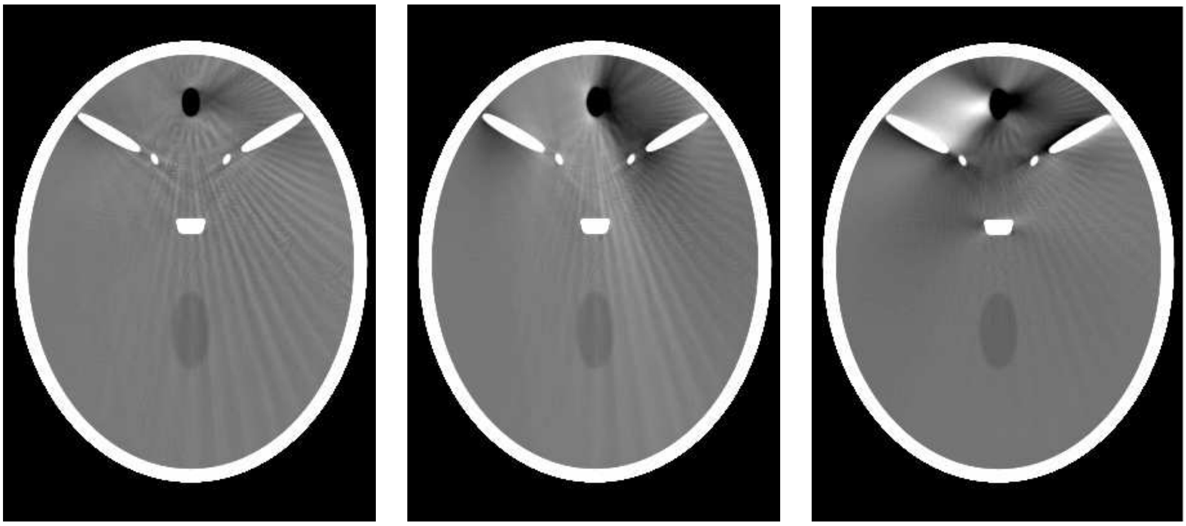
(a) helix segment 1.5π, min=0.77, max=1.24

(b) helix segment 3π, min=0.33, max=1.33

Figure 2: Total Radon weight $M(\vec{r},\vec{\theta})$ (eq. (40)) at $\vec{r} = (0, -90 \text{ mm}, 0)$. The horizontal and the vertical axes are respectively the azimutal angle and the cosine of the polar angle associated to $\vec{\theta}$. The grey scale range is $[0.9, 1.1]$. The minimum and the maximum of M are indicated in the sub-legend.

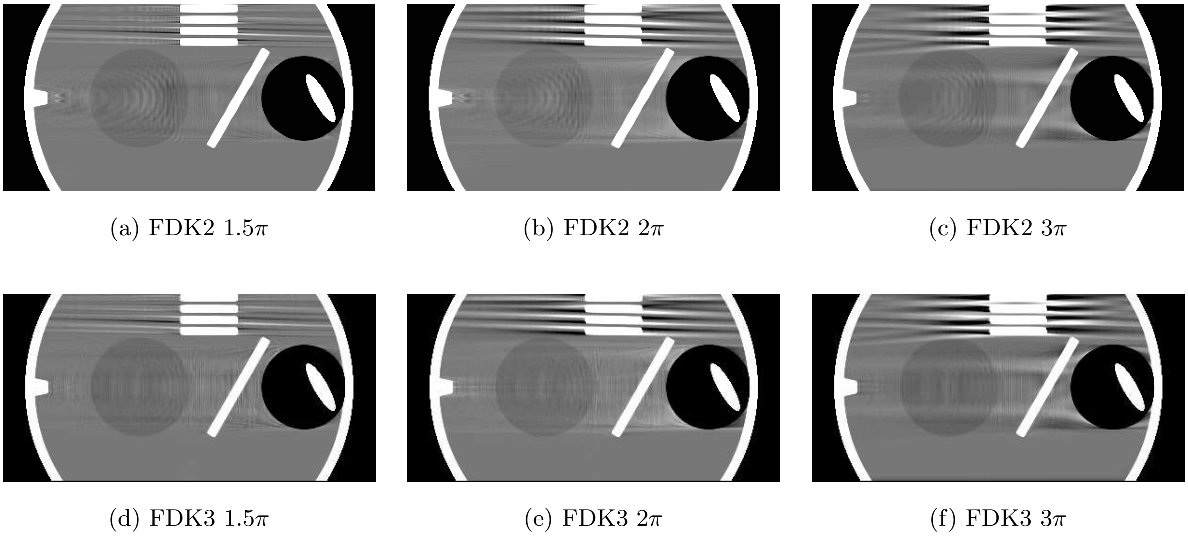


(a) FDK2 1.5π (b) FDK2 2π (c) FDK2 3π



(d) FDK3 1.5π (e) FDK3 2π (f) FDK3 3π

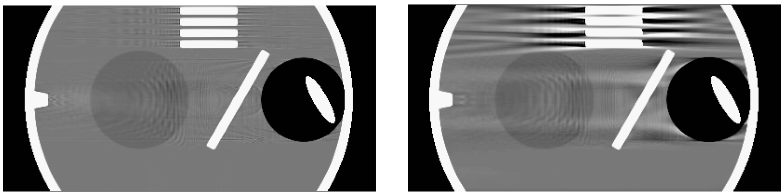
Figure 3: Transaxial slice $z = -21.5$ mm of the HEAD phantom reconstructed with the algorithms **FDK-2** and **FDK-3**. The grey scale range is $[1.01, 1.09]$.



(a) FDK2 1.5π (b) FDK2 2π (c) FDK2 3π

(d) FDK3 1.5π (e) FDK3 2π (f) FDK3 3π

Figure 4: Longitudinal slice $x = 0$ of the HEAD phantom reconstructed with the algorithms **FDK-2** and **FDK-3**. The grey scale range is $[1.01, 1.09]$.



(a) Katsevich (b) FDK1 3π

Figure 5: Longitudinal slice $x = 0$ of the HEAD phantom, reconstructed with the Katsevich algorithm (left) and with the standard **FDK-1** algorithm (right). The grey scale range is $[1.01, 1.09]$

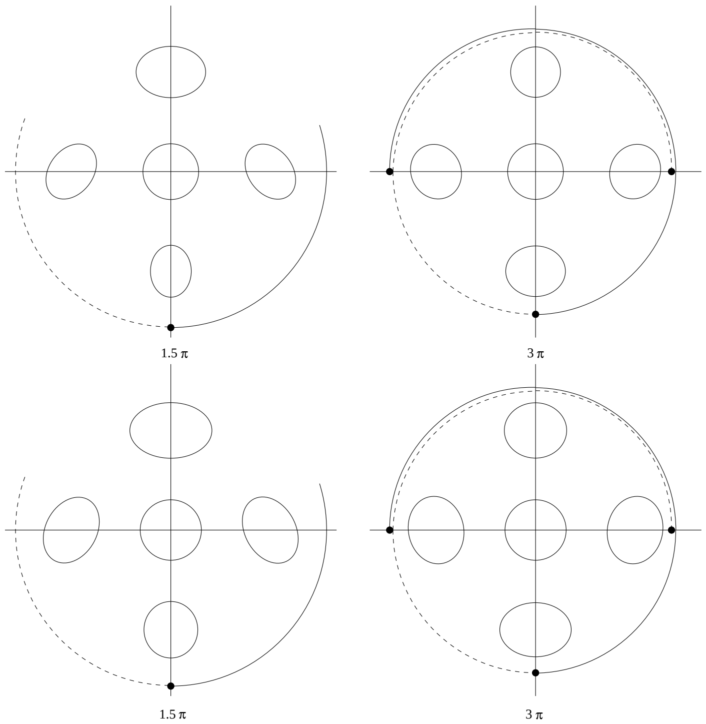
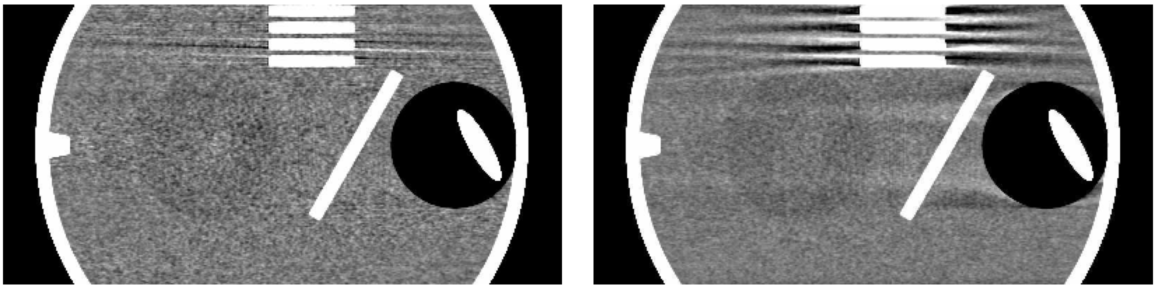
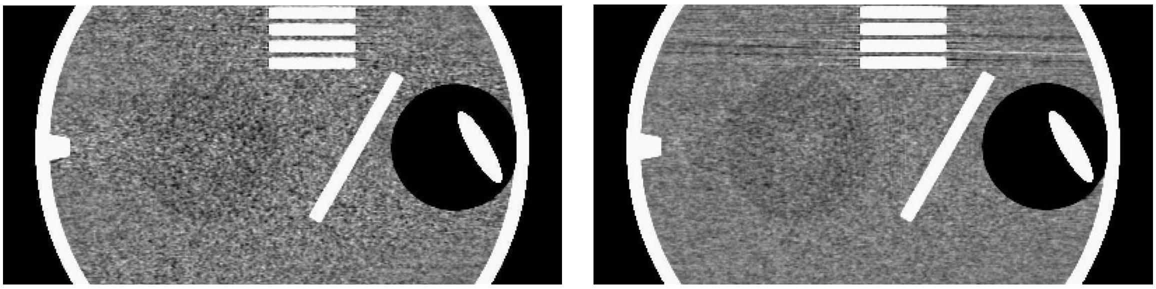


Figure 6: Schematic representation of the iso-contour lines of the response function of the algorithms **FDK-2** (top) and **FDK-3** (bottom). The segment of helix used to reconstruct slice $z = 0$ is represented, with the full and dotted lines corresponding respectively to the half segments located above and below the slice. In each case, the segment is centered at the South.



(a) FDK2 1.5π

(b) FDK2 3π



(c) Katsevich

(d) Frequency mixing 3π

Figure 7: Longitudinal slice $x = 0$ of the HEAD phantom, reconstructed from the noisy projection data. The grey scale range is $[1.0, 1.1]$.