Parallel Magnetic Resonance Imaging reconstruction problems using wavelet representations

Lotfi Chaari

To cite this version:

Lotfi Chaari. Parallel Magnetic Resonance Imaging reconstruction problems using wavelet representations. Signal and Image processing. Université Paris-Est, 2010. English. NNT : . tel-00575136v1

HAL Id: tel-00575136
https://theses.hal.science/tel-00575136v1

Submitted on 9 Mar 2011 (v1), last revised 20 Apr 2011 (v2)

UNIVERSITÉ
PARIS-EST

THÈSE

soutenue le 05/11/2010 pour obtenir

le grade de **Docteur en Sciences** de l'Université de Paris-Est Marne-la-Vallée

Spécialité: Traitement du Signal et des images

par

Lotfi Chaari

Problèmes de reconstruction en Imagerie par Résonance Magnétique parallèle à l'aide de représentations en ondelettes

Composition de la commission d'examen :

Président : Jean-Yves Tourneret
Rapporteurs : Dimitri Van de Ville
Jean-François Giovannelli
Examinateurs : Florence Forbes
Amel Benazza-Benyahia
Directeur de thèse : Jean-Christophe Pesquet
Co-directeur de thèse : Philippe Ciuciu

Remerciments

Je tiens en premier lieu à remercier mon admirable directeur de thèse, M. Jean-Christophe PESQUET (Laboratoire d'Informatique de l'Institut Gaspard Monge, Univ. Paris-Est Marne-la-Vallée), auprès duquel j'ai énormément appris. Je le remercie pour toute l'attention et pour le soutien qu'il m'a portés durant ces trois années. Il m'a communiqué sa rigueur et sa volonté d'aller toujours plus loin, et j'ai pris un réel plaisir à effectuer cette thèse sous sa direction.

Je tiens également à exprimer ma reconnaissance à M. Philippe CIUCIU (CEA-NeuroSpin) et Mme. Amel BENAZZA-BENYAHIA (SUP'COM Tunis) pour avoir co-dirigé cette thèse. Je les remercie d'avoir toujours répondu présent lorsque j'avais besoin de leurs conseils. Je les remercie pour leurs qualités humaines et leurs encouragements.

Je voudrais aussi remercier M. Jean-Yves TOURNERET (INP-ENSEEIHT Toulouse) d'avoir bien voulu présider le jury de thèse, M. Jean-François GIOVANNELLI (Univ. de Bordeaux) et M. Dimitri VAN DE VILLE (Univ. de Genève) pour avoir accepté de rapporter sur cette thèse, ainsi que Mme Florence FORBES (INRIA Rhône Alpes) d'avoir bien voulu participer à ce jury.

Cette thèse a été effectuée dans l'Équipe Signal et Communications (elle-même faisant partie du Laboratoire d'Informatique de l'Institut Gaspard Monge) dans laquelle j'ai été très bien accueilli et au sein de laquelle j'ai pu excercer mes travaux de recherche dans de très bonnes conditions. Je remercie tous les membres de cette équipe pour leur aide, pour les discussions que nous avons eues, pour leurs précieux conseils. Je tiens surtout à saluer mes collègues de cette équipe avec qui j'ai eu énormément de plaisir pendant mes trois années de thèse: Caroline, Mounir, Nelly, Elena, Florian, Pascal, Mireille, Anna.

Durant ma thèse, j'ai aussi en l'occasion de collaborer avec des chercheurs de NeuroSpin comme M. Sébastien MÉRIAUX, M. Thomas VINCENT et Mlle. Solveig BADILLO, que je remercie également pour les discussions et les reflexions menées ensemble.

Mes remerciment à tous mes amis qu'il soient à Paris ou en Tunisie !

Je dédie cette thèse à mes parents Mohsen et Sania, ainsi qu'à mes frères Mohamed, Fahmi et Sami qui m'ont toujours soutenu de très prés durant ce long chemin. Je la dédie également à ma cher fiancée Salma pour son soutien et sa patience malgré les distances qui nous séparent.

Abstract

To reduce scanning time or improve spatio-temporal resolution in some MRI applications, parallel MRI acquisition techniques with multiple coils have emerged since the early 90's as powerful methods. In these techniques, MRI images have to be reconstructed from acquired undersampled "k-space" data. To this end, several reconstruction techniques have been proposed such as the widely-used SENSitivity Encoding (SENSE) method. However, the reconstructed images generally present artifacts due to the noise corrupting the observed data and coil sensitivity profile estimation errors. In this work, we present novel SENSE-based reconstruction methods which proceed with regularization in the complex wavelet domain so as to promote the sparsity of the solution. These methods achieve accurate image reconstruction under degraded experimental conditions, in which neither the SENSE method nor standard regularized methods (e.g. Tikhonov) give convincing results. The proposed approaches relies on fast parallel optimization algorithms dealing with convex but non-differentiable criteria involving suitable sparsity promoting priors. Moreover, in contrast with most of the available reconstruction methods which proceed by a slice by slice reconstruction, one of the proposed methods allows 4D (3D + time) reconstruction exploiting spatial and temporal correlations. The hyperparameter estimation problem inherent to the regularization process has also been addressed from a Bayesian viewpoint by using MCMC techniques. Experiments on real anatomical and functional data show that the proposed methods allow us to reduce reconstruction artifacts and improve the statistical sensitivity/specificity in functional MRI.

Résumé

Pour réduire le temps d'acquisition ou bien améliorer la résolution spatio-temporelle dans certaines application en IRM, de puissantes techniques parallèles utilisant plusieurs antennes réceptrices sont apparues depuis les années 90. Dans ce contexte, les images d'IRM doivent être reconstruites à partir des données sous-échantillonnées acquises dans le "k-space" . Plusieurs approches de reconstruction ont donc été proposées dont la méthode SENSitivity Encoding (SENSE). Cependant, les images reconstruites sont souvent entâchées par des artéfacts dus au bruit affectant les données observées, ou bien à des erreurs d'estimation des profils de sensibilité des antennes. Dans ce travail, nous présentons de nouvelles méthodes de reconstruction basées sur l'algorithme SENSE, qui introduisent une régularisation dans le domaine transformé en ondelettes afin de promouvoir la parcimonie de la solution. Sous des conditions expérimentales dégradées, ces méthodes donnent une bonne qualité de reconstruction contrairement à la méthode SENSE et aux autres techniques de régularisation classique (e.g. Tikhonov). Les méthodes proposées reposent sur des algorithmes parallèles d'optimisation permettant de traiter des critères convexes, mais non nécessairement différentiables contenant des a priori parcimonieux. Contrairement à la plupart des méthodes de reconstruction qui opèrent coupe par coupe, l'une des méthodes proposées permet une reconstruction 4D (3D + temps) en exploitant les corrélations spatiales et temporelles. Le problème d'estimation d'hyperparamètres sous-jacent au processus de régularisation a aussi été traité dans un cadre bayésien en utilisant des techniques MCMC. Une validation sur des données réelles anatomiques et fonctionnelles montre que les méthodes proposées réduisent les artéfacts de reconstruction et améliorent la sensibilité/spécificité statistique en IRM fonctionnelle.

Contents

1 Introduction **25**
1.1 Motivations . . . 25
1.2 Organization of the manuscript . . . 27
1.3 Publications . . . 29
1.4 Patent . . . 30
1.5 Software . . . 30

2 Magnetic Resonance Imaging background **31**
2.1 Introduction . . . 31
2.2 Magnetic Resonance Imaging . . . 31
2.2.1 Nuclear Magnetic resonance . . . 31
2.2.2 NMR signal measurement . . . 32
2.3 Image contrast and acquisition parameters . . . 37
2.3.1 TR effect . . . 37
2.3.2 TE effect . . . 38
2.3.3 Acquiring a T_1 or T_2-weighted MRI image . . . 38
2.4 Functional Magnetic Resonance Imaging . . . 39
2.4.1 Blood Oxygen Level Dependent effect . . . 40
2.4.2 Data acquisition in fMRI . . . 40
2.4.3 Artifacts in fMRI . . . 41
2.5 Parallel Magnetic Resonance Imaging . . . 44
2.5.1 Parallel MRI basics . . . 45
2.5.2 Parallel MRI reconstruction . . . 48
2.6 Conclusion . . . 55

3 Regularization and convex analysis for inverse problems **57**
3.1 Introduction . . . 57
3.2 Regularization for inverse problems . . . 58
3.2.1 Inverse problems . . . 58
3.2.2 Regularization . . . 59
3.2.3 The frame concept . . . 60
3.2.4 Bayesian approach using frame representations . . . 61
3.3 Convex optimization . . . 63
3.3.1 Some convex optimization algorithms . . . 64
3.3.2 For those who see life in pink . . . 69
3.4 Numerical illustrations . . . 72
3.4.1 Comparison of the AA and SA performance . . . 72
3.4.2 Convergence speed comparison . . . 74
3.4.3 Convergence speed gain when calculating $\|HF^*\|^2$ . . . 75
3.5 Conclusion . . . 80

4 Regularized SENSE reconstruction 81
4.1 Introduction . . . 81
4.2 Regularization in pMRI . . . 82
  4.2.1 Tikhonov regularization . . . 82
  4.2.2 Total variation regularization . . . 84
4.3 Regularization in the WT domain . . . 87
  4.3.1 Motivation . . . 87
  4.3.2 Definitions and notations . . . 88
  4.3.3 Wavelet-based regularized reconstruction . . . 89
  4.3.4 Regularization using bivariate wavelet prior . . . 100
  4.3.5 Constrained wavelet-based regularization . . . 106
4.4 Combined wavelet-TV regularization . . . 114
4.5 Spatio-temporal regularization . . . 120
4.6 Conclusion . . . 123

5 Hyper-parameter estimation 125
5.1 Introduction . . . 125
5.2 Problem Formulation . . . 126
5.3 Hierarchical Bayesian Model . . . 127
5.4 Sampling strategies . . . 129
  5.4.1 Hybrid Gibbs Sampler . . . 129
  5.4.2 Hybrid MH sampler using algebraic properties of frame representations 134
5.5 Toward a more general Bayesian Model . . . 137
5.6 Numerical illustrations . . . 141
  5.6.1 Validation experiments . . . 141
  5.6.2 Convergence results . . . 144
  5.6.3 Application to image denoising . . . 145
  5.6.4 Hyper-parameter estimation in parallel MRI . . . 158
5.7 Conclusion . . . 158

6 Experimental validation in fMRI 159
6.1 Introduction . . . 159
6.2 Data processing in fMRI . . . 159
  6.2.1 Exploratory vs. hypothesis-driven methods . . . 160
  6.2.2 The General Linear Model . . . 162
  6.2.3 Subject-level analysis . . . 163
  6.2.4 Group-level analysis . . . 166
6.3 Validation of the proposed methods . . . 168
  6.3.1 Experimental data . . . 168
  6.3.2 Subject-level analysis . . . 169
  6.3.3 Group-level analysis . . . 178
6.4 Discussion . . . 184
6.5 Conclusion . . . 184

7 Conclusion **187**

Acronyms

$\tilde{A}$ Fourier transform of A
A Norm of the vector $\mathbf{A}$
$\mathbb{R}$ Set of reals
$\mathbb{C}$ Set of complex numbers
$\mathbb{N}$ Set of non negative integers
T Wavelet transform operator
$\mathrm{Re}(x)$ Real part of x
$\mathrm{Im}(x)$ Imaginary part of x
H^* Adjoint of the linear operator H
T^{-1} Inverse of the linear operator T
$(\cdot)^{\mathsf{H}}$ Transposed complex conjugate
$(\cdot)^{\sharp}$ Pseudo-inverse
$\boldsymbol{r}$ Vector
$\boldsymbol{x}^{\mathsf{T}}$ The transpose of the vector $\boldsymbol{x}$
$\overline{\rho}$ Image to be estimated
$\widehat{\rho}$ Estimated image

AA Analysis Approach

A-V Audio-Video

BOLD Blood Oxygen Level Dependent

CS Compressed Sensing

CSF Cerebral Spinal Fluid

CWR-SENSE Constrained Walevel Regularized-SENSE

DR Douglas Rachford

EPI Echo Planar Imaging

EVI Echo Volume Imaging

EM Expectation-Maximization

fMRI functional Magnetic Resonance Imaging

FISTA Fast Iterative Soft Thresholding Algorithm

FOV Field of View

FB Forward Backward

FID Free Induction Decay

GG Generalized Gaussian

GGL Generalized Gauss-Laplace

GRE GRadient-Echo

GLM General Linear Model

GRAPPA Generalized Autocalibrating Partially Parallel Acquisitions

HRF Haemodynamic Response Function

ICA Independent Component Analysis

Lc-Rc Left click-Right click

MAP Maximum A Posteriori

MMSE Minimum Mean Square Error

ML Maximum Likelihood

MRI Magnetic Resonance Imaging

NMR Nuclear Magnetic Resonance

OLS Ordinary Least Squares

pMRI parallel Magnetic Resonance Imaging

PPXA Parallel ProXimal Algorithm

prox Proximity operator

PET Positron Emission Tomography

PCA Principal Component analysis

pdf probability density function

PWLS Penalized Weighted Least Squares

RF Radio Frequency

SVD Singular Value Decomposition

SNR Signal to Noise Ratio

SSIM Structural SIMilarity

SENSE SENSitivity Encoding

SMASH Simultaneous Acquisition of Spatial Harmonics

SA Synthesis Approach

TE Time of Echo

TR Time of Repetition

TIW Translation Invariant Wavelet

TV Total Variation

U2OB Union of 2 Orthonormal Bases

UWR-SENSE Unconstrained Walevel Regularized-SENSE

WT Wavelet Transform

List of Figures

1.1 Reconstruction artifacts using the SENSE method for $R = 4$. 26

2.1 (a): magnetic moment at the initial equilibrium state; (b): magnetic moment in the presence of a stationary magnetic field of magnitude B_0. 32
2.2 Precession movement of the magnetic moment. 33
2.3 T_1 relaxation curve. 33
2.4 T_2 relaxation curve. 34
2.5 Slice by slice acquisition in MRI. 35
2.6 GRE sequence diagram (ADC denotes the Analog-to-Digital Converter). . . 37
2.7 Example of experimental paradigms: (a) block paradigm; (b) event-related paradigm. 41
2.8 Used stimuli with a block paradigm (top) and induced BOLD signal (bottom). 41
2.9 Example of metal artifacts; yellow array: hyper-signal areas; red array: hypo-signal areas; green arrays: image deformation [Gerardin, 2006]. 42
2.10 Example of motion artifacts; right: blur artifacts; left: phantom artifacts [Gerardin, 2006]. 42
2.11 Example of magnetic susceptibility artifacts [Gerardin, 2006]. 43
2.12 Example of ghosting artifacts [Gerardin, 2006]. 43
2.13 Example of ghosting artifacts. 44
2.14 Parallel acquisition in MRI using an 8 coils array. 46
2.15 Sampling schemes for conventional and parallel acquisitions on a Cartesian grid. Here $R = 2$ indicates a subsampling along the phase encoding direction (y-axis) by a factor of two. 47
2.16 Spiral encoding trajectory; (1): square spiral, (2): circular spiral, (3): circular spiral for segmented acquisition. 47
2.17 Parallel acquisition in MRI using an 8 coils array: (a) coil sensitivity profiles; (b) acquired images without subsampling the k-space; (c) acquired reduced FOV images; (d) reconstructed full FOV image. 48
2.18 GE anatomical data: reconstructed adjacent slices using SENSE for $R = 2$. 53
2.19 GE anatomical data: reconstructed adjacent slices using SENSE for $R = 4$. 54
2.20 Six EPI reconstructed slices using basic SENSE for $R = 4$. 55

3.1 Curves of the convex function defined in Eq. (3.20). 65
3.2 Curves of the proximity operator associated with the functions plotted in Fig. 3.1. 65
3.3 Summary diagram of iterative convex optimization algorithms. 66
3.4 Original cameraman image (a), degraded one (b), restored images using AA (c) and SA (d) with the contourlet frame. 74
3.5 Original barbara image (a), degraded one (b), restored images using AA (c) and SA (d) with the contourlet frame. 75

3.6 Original boat image (a), degraded one (b), restored images using AA (c) and SA (d) with the contourlet frame. 76
3.7 Original straw image (a), degraded one (b), restored images using AA (c) and SA (d) with the contourlet frame. 77
3.8 Evaluation of the optimality criterion w.r.t iteration number with the AA. . 78
3.9 Evaluation of the optimality criterion w.r.t iteration number with the SA. . 79
3.10 Convergence curve of Algorithm 5. 79
3.11 Evaluation of the optimality criterion w.r.t iteration number for the FB algorithm when calculating (dot line) and approximating $\|HF^*\|^2$ (solid line). 79

4.1 Reconstructed anatomical slices using Tikhonov regularization for $R = 2$. . 84
4.2 Reconstructed anatomical slices using Tikhonov regularization for $R = 4$. . 85
4.3 Two EPI reconstructed slices using Tikhonov regularization for $R = 4$. . . . 86
4.4 Reconstructed anatomical (left) and functional EPI (right) images using TV regularization in the image domain with moderate (a)-(b) and strong (c)-(d) regularization levels ($R = 4$). 87
4.5 Example of normalized empirical histograms of wavelet coefficients and associated pdfs using GG (red) and GGL (blue) distributions, the hyperparameters being estimated using the proposed MCMC approach in Chapter 5. 90
4.6 Convergence speed of the optimization algorithm w.r.t. the choice of the relaxation parameter λ for $j_{\max} = 3$. 96
4.7 Convergence speed comparison for the TWIST, FISTA and FB algorithms. 96
4.8 Reconstructed anatomical slices using UWR-SENSE for $R = 2$. 97
4.9 Rreconstructed anatomical slices using UWR-SENSE for $R = 4$. 98
4.10 Six EPI reconstructed slices using our UWR-SENSE algorithm for $R = 4$. . 99
4.11 Influence of the number of resolution levels on the anatomical reconstructed images using our CWR-SENSE method for $R = 4$; from left to right: reconstructed images with $j_{\max} = 2$ - $j_{\max} = 1$, $j_{\max} = 3$ - $j_{\max} = 2$ and $j_{\max} = 4$ - $j_{\max} = 3$. 100
4.12 Influence of the number of resolution levels on the EPI reconstructed images using our CWR-SENSE method for $R = 4$; from left to right: reconstructed images with $j_{\max} = 2$ - $j_{\max} = 1$, $j_{\max} = 3$ - $j_{\max} = 2$ and $j_{\max} = 4$ - $j_{\max} = 3$. 100
4.13 Joint 2D empirical histogram $\boldsymbol{\zeta}_{1,2}$ (left) and pdfs of the independent (middle) and proposed bivariate (right) models. 101
4.14 Top: 3D plot of the modulus of a complex-valued number for which the real and imaginary parts belong to a uniform 2D grid; bottom: 3D plot of the modulus of the proximity operator associated to Example 4.3.4 with $\alpha = 2$, $\beta = 1$, $\gamma = 0.5$ and $p = 1$, and where the real and imaginary parts belong to the same 2D grid as the top figure. 105
4.15 Reconstructed anatomical slices using UWR-SENSE with the bivariate prior for $R = 4$. 106
4.16 Reconstructed EPI slices using UWR-SENSE with the bivariate prior for $R = 4$. 107

4.17 CWR-SENSE reconstruction steps. 111
4.18 Reconstructed anatomical slices using CWR-SENSE for $R = 2$ 111
4.19 Reconstructed anatomical slices using CWR-SENSE for $R = 4$ 112
4.20 Reconstructed EPI slices using CWR-SENSE for $R = 4$. 113
4.21 Influence of the wavelet basis on the anatomical reconstructed images using our CWR-SENSE method for $R = 4$. 114
4.22 Reconstructed anatomical slices using Algorithm 8 with the *Symmlet* 8 orthonormal basis for $R = 4$. 118
4.23 Reconstructed anatomical slices using Algorithm 8 with the U2OB for $R = 4$.119

5.1 Examples of empirical approximation (left) and detail (right) histograms and pdfs of frame coefficients corresponding to a synthetic image. 143
5.2 NMSE between the reference and current MMSE estimators w.r.t iteration number corresponding to $v_{1,1}$ in $\mathcal{B}_1$. 145
5.3 NMSE between the reference and current MMSE estimators w.r.t iteration number corresponding to $v_{2,2}$ in $\mathcal{B}_2$. 145
5.4 Ground truth values (dashed line) and posterior distributions (solid line) of the sampled hyper-parameters γ and β, for the subbands $h_{1,2}$ and $h_{2,2}$ in $\mathcal{B}_1$ and $\mathcal{B}_2$, rsepectively. 146
5.5 Original 256×256 *Boat* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d). 149
5.6 Original 128×128 *Sebal* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d). 150
5.7 Original 128×128 *Tree* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d). 151
5.8 Original 256×256 *Peppers* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d). . . 152
5.9 Original 128×128 *Kodim* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d). 153
5.10 Original 256×256 *House* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d). 154
5.11 Original 128×128 *Tire* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d). 155
5.12 Original 128×128 *cameraman* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d). . . 156
5.13 Original 128×128 *Straw* image (a), noisy image (b) and denoised images using the variational approach (c) and the proposed MMSE estimator (d). . 157

6.1 Required pre-processing steps before data analysis in fMRI. 162
6.2 Illustration of the student distribution thresholding: (a) fixed threshold probability; (b) the threshold corresponds to the observed t-score (t_n) allowing to obtain the maximum probability (p-value). If $t_n > t_\alpha$, or equivalently $p < \alpha$, $\mathcal{H}_0$ is rejected, otherwise (if $t_n \leq t_\alpha$) it is retained. 165
6.3 Reconstructing reduced FOV images from the raw data. 168

6.4 **Axial** reconstructed slices using mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$ and $R = 4$ with 2×2 mm^2 in-plane spatial resolution. Red circles and ellipsoids indicate reconstruction artifacts. . . . 169

6.5 **Coronal** reconstructed slices using mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$ and $R = 4$ with 2×2 mm^2 in-plane spatial resolution. Red circles and ellipsoids indicate reconstruction artifacts. . . . 170

6.6 **Sagittal** reconstructed slices using mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$ and $R = 4$ with 2×2 mm^2 in-plane spatial resolution. Red circles indicate reconstruction artifacts. . . . 170

6.7 (a): design matrix and the `Lc-Rc` contrast involving two conditions (grouping auditory and visual modalities); (b): design matrix and the `A-V` contrast involving four conditions (sentence, computation, left click, right clisk). The design matrix is made up of twenty regressors corresponding to the canonical HRF and its time derivative for each of the ten experimental conditions, in addition to a mean signal regressor. . . . 171

6.8 Subject-level student-t maps superimposed to anatomical MRI for the `A-V` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE. Neurological convention: left is left. The blue cross indicates the activation peak. . . . 174

6.9 Subject-level student-t maps superimposed to anatomical MRI for the `Lc-Rc` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE. Neurological convention: left is left. The blue cross indicates the activation peak. . . . 175

6.10 3D plots of the detected activated area for the `Lc-Rc` contrast superimposed to a 3D mesh of the brain where $R = 2$. Activation foci appear in the right motor cortex whatever the reconstruction method, but with larger spatial extent using our regularized reconstructions. . . . 176

6.11 3D plots of the detected activated area for the Lc-Rc contrast superimposed to a 3D mesh of the brain where $R = 4$. Only our reconstruction algorithms recover activation foci in the right motor cortex. . . . 177

6.12 Group-level student-t maps for the `A-V` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$. Neurological convention: left is left. Red arrows indicate the global maxima. . . . 179

6.13 Group-level student-t maps for the `A-V` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 4$. Neurological convention: left is left. Red arrows indicate the global maxima. . . . 180

6.14 Group-level student-t maps for the `Lc-Rc` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$. Neurological convention: left is left. Red arrows indicate the global maxima. . . . 182

6.15 Group-level student-t maps for the `Lc-Rc` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 4$. Neurological convention: left is left. Red arrows indicate the global maxima. 183

List of Tables

2.1 Classification of the main pMRI reconstruction methods depending of the reconstruction space. 50

3.1 SNR evaluation for different images and frames in image deblurring. 78
3.2 Encountered difficulties and retained solutions when using the FB, FISTA and PPXA algorithms with the AA or SA. 78

4.1 SNR (in dB) evaluation for reconstructed images using different methods for $R = 4$. 112
4.2 SNR (in dB) evaluation for reconstructed images using Algorithm 8 with the OB and U2OB wavelet representations for $R = 4$. 119

5.1 Example 1: NMSEs for the estimated hyper-parameters using the MMSE and MAP estimators. 142
5.2 Example 2: NMSEs for the estimated hyper-parameters using the MMSE and MAP estimators with Sampler 2. 143
5.3 Example 3: NMSEs for the estimated hyper-parameters using the MMSE and MAP estimates with Sampler 2. 144
5.4 SNR and SSIM values for the noisy and denoised images. 147
5.5 Computational time (in minutes) for the used methods. 148
5.6 SNR and SSIM values for the noisy and denoised *Straw* images. 149

6.1 Significant statistical results for the `A-V` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 2$. 172
6.2 Significant statistical results for the `A-V` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 4$. 173
6.3 Significant statistical results for the `Lc-Rc` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 2$. 173
6.4 Significant statistical results for the `Lc-Rc` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 4$. 173
6.5 Significant statistical results at the group-level for the `A-V` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 2$. 178

6.6 Significant statistical results at the group-level for the `A-V` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 4$. 181
6.7 Significant statistical results at the group-level for the `Lc-Rc` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 2$. 181
6.8 Significant statistical results at the group-level for the `Lc-Rc` contrast (corrected for multiple comparisons). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 4$. 182

Chapter 1

Introduction

Contents

1.1 Motivations 25
1.2 Organization of the manuscript 27
1.3 Publications 29
1.4 Patent 30
1.5 Software 30

1.1 Motivations

Magnetic Resonance Imaging (MRI) is a modality which has already proven to be very useful since many decades, and continues to be investigated in various clinical and research activities such as neuroimaging. High resolution images provided in non-invasive manner are of high interest to clinicians and researchers whatever their research area (cognitive neuroscience, clinical research,...). Although the significance of MRI is rising, one may want to minimize time of patient's exposition to MRI environment and to have more spatial and/or temporal details in the acquired data. These requirements may be met by improving spatio-temporal resolution and reducing the global imaging time. The above mentioned improvements were intensively studied in numerous applications such as neuroimaging [Rabrait et al., 2008], cardiac [Weiger et al., 2000] and abdominal [Zhu et al., 2004] imaging. They can be achieved without significantly degrading the image quality by using parallel MRI (pMRI) systems, developed in the 1990's. These systems rely on two main technical features. First, MRI data are simultaneously collected in the frequency domain by multiple receiver coils with complementary sensitivity profiles located around the underlying object. Second, the acquisition time is reduced since data are sampled at a frequency rate R times lower than the Nyquist sampling rate along at least one spatial direction, i.e. usually the phase encoding direction or typically the Y axis. However, if one wants to increase the spatial resolution while keeping the acquisition time almost fixed, the acquired data can be sampled at a higher rate along one or more encoding directions. A reconstruction step is then necessary to build up a full Field of View (FOV) image. It is performed by unfolding the undersampled coil-specific data and exploiting the complementarity of the sensitivity profiles of the used coils. It is also a challenging task because of three main artifact sources which decrease the Signal to Noise Ratio (SNR): aliasing artifacts related to the undersampling rate, acquisition noise and also errors in the estimation of coil sensitivity profiles.

Many reconstruction methods have been proposed since the early 1990's such as SMASH

[Sodickson and Manning, 1999], SPACE-RIP [Kyriakos et al., 2000], AUTO-SMASH [Jakob et al., 1998], GRAPPA [Griswold et al., 2002], SENSE [Pruessmann et al., 1999a]... Some of these methods have been designed to operate directly in the Fourier domain like SMASH and GRAPPA, while some others operate in the spatial domaine like SENSE. However, when the experimental conditions become quite severe (i.e. high noise level, low sampling rate or poor estimation of the coil sensitivity profiles), the reconstruction quality of all of these methods is strongly degraded. Early works tried to optimize the coil geometry [Weiger et al., 2001], improve the coil sensitivity profile estimation [Blaimer et al., 2004] or optimize the encoding trajectory in order to alleviate this problem [Yan et al., 1999].

In the same context, we have focused in this PhD on the first proposed pMRI reconstruction method operating in the spatial domain, namely SENSE. We have been interested in improving its reconstruction performance even under severe experimental conditions by resorting to non smooth regularization techniques. We have also chosen to design our regularization method in the wavelet transform domain regarding to the encountered artifact features. These artifacts generally appear in localized area with either very high or very low intensity levels (see Fig. 1.1).

Figure 1.1: Reconstruction artifacts using the SENSE method for $R = 4$.

Indeed, using wavelet transforms allows us to have a sparse representation of the image to be reconstructed. The wavelet domain enables also extracting spatial and frequency informations at the same time, which helps in detecting the reconstruction artifacts and smoothing them.

As any regularization method, a set of regularization parameters has to be estimated so that the method can be autocalibrated. However, when using redundant wavelet representations, this problem becomes difficult to handle due to the non-bijectivity of the wavelet operator. This problem has been addressed from a Bayesian viewpoint in this PhD thesis. Our regularized approach proceeds by a slice by slice reconstruction like the previously cited reconstruction methods since a 3D volume is generally acquired slice by slice. Iterating over slices is then necessary to reconstruct an entire 3D volume. The different slices are therefore assumed independent and no correlation between them is taken into account although it does exist in practice. A first extension of our method has been developed in order to exploit correlation between adjacent slices using 3D wavelet transforms. On the other hand, in dynamic MRI applications like functional MRI (fMRI), the brain is imaged several times in order to track the BOLD contrast, which is an indirect consequence of neuronal activity. Another source of correlation would consequently be helpful: temporal correlation between successive acquired volumes. This additional source of information has been exploited in this PhD in order to improve the reconstruction quality in pMRI SENSE imaging. The resulting 4D regularization approach involves also the 3D regularized one by turning off the temporal regularization.

1.2 Organization of the manuscript

- **MRI background**

  The second chapter of this manuscript summarizes the main state of the art in MRI, pMRI and fMRI. It helps the reader to better understand the rest of the manuscript by giving the background of MRI from the physical to the technical principle. Parallel MRI is also introduced to better emphasize the motivations and the challenges of this PhD. Finally, since we are focusing here on brain imaging, fMRI is also briefly introduced as it will be the main application in the validation task.

- **Convex optimization tools for the regularization of inverse problems**

  When the considered inverse problem is ill-posed, one generally resorts to regularization techniques. These techniques may inhere easy optimization tasks as well as complicated ones. Although the regularization literature is quite abundant, we will focus here on regularization techniques which lead to optimization problems involving convex but not necessarily differentiable optimality criteria. Chapter 3 of this manuscript aims at presenting the theoretical framework of convex analysis and optimization tools which will be used for the proposed regularization approaches in Chapter 4. The main hurdles which may be encountered are outlined, and some solutions are also provided in this chapter.

- **Regularized parallel MRI reconstruction**

  The SENSE method on which we will focus in this manuscript does not perform well when the experimental conditions become quite severe (c.f. Chapter 2). For this reason, regularizing the considered inverse problem would be a suitable solution to achieve better reconstruction quality. Chapter 4 of this manuscript describes in details the proposed regularized reconstruction methods. The first proposed method proceeds by a regularization in the wavelet domain using appropriate priors for the

wavelet coefficients of the image to be reconstructed. Besides the wavelet regularization, the second proposed algorithm uses an additional constraint on the artifact regions to better smooth them.
The two first proposed approaches proceed by a slice by slice reconstruction. However, adjacent slices are actually spatially correlated. Moreover, the whole brain is imaged several times in dynamic MRI applications like fMRI, and dependencies between different acquisitions do exist. A third regularization approach is therefore proposed taking into account 3D spatial correlations and temporal dependencies between the acquired volumes.

- **Estimating the regularization hyper-parameters**

  As any regularization method, the proposed approaches require a hyper-parameter estimation step in order to be fully automatic. This task may be easy to perform when a reference image is available so that the hyper-parameters can be estimated based on this image. However, since the proposed regularization methods operate in the wavelet domain, the hyper-parameter estimation task becomes much more complicated when redundant representations are used. Indeed, since no bijection exists between the original and transformed spaces, even if the reference image is perfectly known, the frame coefficients are not. Chapter 5 of this manuscript describes a Bayesian framework and the associated sampling strategies to handle this problem and estimate the hyper-parameters from a noisy observation when overcomplete representations are used even if no reference image is available.

- **Experimental validation**

  Besides the visual inspection of reconstructed images, a more complete validation of the proposed approaches is presented in Chapter 6. An fMRI study is conducted in this chapter on real data at the subject and group levels. These data have been acquired in the neuroimaging center NeuroSpin-CEA (Commissariat à l'Energie Atomique). NeuroSpin is one of the partners of the OPTIMED project which has been supported by the french Agence Nationale de la Recherche (ANR). This PhD has been involved in the ANR-OPTIMED project and took place at the Laboratoire d'Informatique Gaspard Monge (LIGM), the coordinator of this project. In addition, this PhD has been supported by the Région Île de France.
  This validation shows that the proposed approaches allow us to improve the statistical sensitivity/specificity of the activation detection in fMRI.

1.3 Publications

The main publications are available at the following URL:
http://lotfi-chaari.net/en/publications.html

- Journal articles:
    - Lotfi Chaari, Jean-Christophe Pesquet, Jean-Yves Tourneret, Philippe Ciuciu and Amel Benazza-Benyahia, “A Hierarchical Bayesian Model for Frame Representation”, IEEE transactions on Signal Processing, vol. 18, no. 11, pp. 5560-5571, November 2010.
    - Lotfi Chaari, Jean-Christophe Pesquet, Amel Benazza-Benyahia and Philippe Ciuciu, “A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging”, Medical Image Analysis, In Press, DOI 10.1016/j.media.2010.08.001, 2010.
- Conference articles:
    - Lotfi Chaari, Sébastien Mériaux, Jean-Christophe Pesquet, Philippe Ciuciu, “Impact of the parallel imaging reconstruction algorithm on brain activity detection in fMRI”, International Symposium on Applied Sciences in Biomedical and Communication Technologies, Rome, Italy, November 7-10, 2010.
    - Lotfi Chaari, Sébastien Mériaux, Jean-Christophe Pesquet, Philippe Ciuciu, “Impact of the parallel imaging reconstruction algorithm on the statistical sensitivity in fMRI”, 16th Annual Meeting of the Organization for Human Brain Mapping, Barcelona, Spain, June 6-10, 2010.
    - Lotfi Chaari, Jean-Christophe Pesquet, Jean-Yves Tourneret, Philippe Ciuciu and Amel Benazza-Benyahia, “A Hierarchical Bayesian Model for Frame Representation”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Dallas, USA, pp. 4086-4089, March 14 - 19, 2010
    - Lotfi Chaari, Amel Benazza-Benyahia , Jean-Christophe Pesquet and Philippe Ciuciu, “Wavelet based parallel MRI regularization using bivariate sparsity promoting priors.”, IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, pp. 1725-1728, November 7-11, 2009.
    - Lotfi Chaari, Nelly Pustelnik, Caroline Chaux and Jean-Christophe Pesquet, “Solving inverse problems with overcomplete transforms and convex optimization techniques”, Society of Photo-optical Instrumentation Engineers (SPIE) Conference, San Diego, Ca, USA, August 2009, 14 p.
    - Lotfi Chaari, Philippe Ciuciu, Amel Benazza-Benyahia and Jean-Christophe Pesquet, “Performance of three parallel MRI reconstruction mthods in the presence of coil sensitivity map errors”, International Society for Magnetic Resonance in Medicine (ISMRM) Meeting, Honolulu, USA, 1 p., April 18-24 2009.
    - Lotfi Chaari, Jean-Christophe Pesquet, Amel Benazza-Benyahia and Philippe Ciuciu, “Minimization of a sparsity promoting criterion for the recovery of

complex-valued signals", SPARS Workshop, Saint-Malo, France, 4 p., April 6-9 2009.

- Lotfi Chaari, Jean-Christophe Pesquet, Amel Benazza-Benyahia and Philippe Ciuciu, "Autocalibrated regularized parallel MRI reconstruction in the wavelet domain", IEEE International Symposium on Biomedical Imaging (ISBI), Paris, France, pp. 756-759, May 14-17 2008.

1.4 Patent

A patent was recently submitted to the United States patent office under the reference **61/379,105**. It involves the proposed 3D and 4D regularized reconstruction algorithms.
Title: "Spatio-temporal regularized reconstruction for parallel MRI acquisition systems"
Date: September 2, 2010
Inventors: Lotfi Chaari, Jean-Christophe Pesquet, Sébastien Mériaux and Philippe Ciuciu.

1.5 Software

This PhD work has led to the design of a free software called **pMRILab**. *pMRILab 1.0.0 is* a free software for parallel MRI reconstruction. It can be freely used, modified and distributed under a CeCILL v2 licence. It allows the standard SENSE and the 2D regularized reconstructions from the reduced FOV data. It is available online at the following URL:
`http://lotfi-chaari.net/pMRILab/pMRILab.tar.gz`.
Note that this version of the software include only the 2D reconstruction apprach. The 3D and 4D regularized approaches are not yet available online for confidentiality reasons. The current version is implemented in Matlab and can be used with any platform provided that the Wavelab software is installed.
For more details, the interested reader can refer to the *Plume* web page availabe at the following URL:
`http://www.projet-plume.org/relier/pmrilab`.
The second version of pMRILab (pMRILab 2.0.0) has been implemented in the C language and is installed in NeuroSpin for validation. However, this version is still not available online.
In addition to its previous functionalities, pMRILab 2.0.0 implements also the spatio-temporal regularized version of the proposed approach. Its current implementation in C runs faster on multi-core machines due to its implementation which supports the OpenMP library for multithread processing.

Chapter 2

Magnetic Resonance Imaging background

Contents

2.1 Introduction . . . 31
2.2 Magnetic Resonance Imaging . . . 31
2.3 Image contrast and acquisition parameters . . . 37
2.4 Functional Magnetic Resonance Imaging . . . 39
2.5 Parallel Magnetic Resonance Imaging . . . 44
2.6 Conclusion . . . 55

2.1 Introduction

This chapter briefly describes the technical and physical principles of Magnetic Resonance Imaging (MRI), functional Magnetic Resonance Imaging (fMRI) and parallel Magnetic Resonance Imaging (pMRI). More precisely, Section 2.2 describes how to get an MRI image from the Nuclear Magnetic Resonance (NMR) signal and physical phenomena behind it. Image contrast and acquisition parameters are detailed in Section 2.3. FMRI is then briefly introduced in Section 2.4 and finally, parallel MRI, the main topic of this thesis, is introduced in Section 2.5. However, readers familiar with this topic can directly go through Chapter 3.

2.2 Magnetic Resonance Imaging

2.2.1 Nuclear Magnetic resonance

The Nuclear Magnetic Resonance (NMR) phenomenon was described for the first time by Edward Purcell in 1946. It relies on the fact that each proton of a tissue sample has a quantic property called *spin*, which refers to its angular momentum. Each spin has a microscopic magnetic moment $\overrightarrow{\mu}$. If no external magnetic field is applied, the global aimantation is cancelled out: $\overrightarrow{M} = \sum \overrightarrow{\mu} = \overrightarrow{0}$ (see Fig. 2.1(a)). However, if an external magnetic field $\overrightarrow{B}_0$ is applied, protons will align along its orientation and a resulting aimantation $\overrightarrow{M} \neq \overrightarrow{0}$ will appear like shown in Fig. 2.1(b). Actually, spins are not perfectly aligned along $\overrightarrow{B}_0$, but have a circular movement around it at the angular *Larmor* frequency $f_0 = \frac{\gamma}{2\pi} B_0$, where γ is the gyromagnetic ratio related to the specific atom and B_0 is the norm of the magnetic field $\overrightarrow{B}_0$. For instance, the hydrogen atom, which

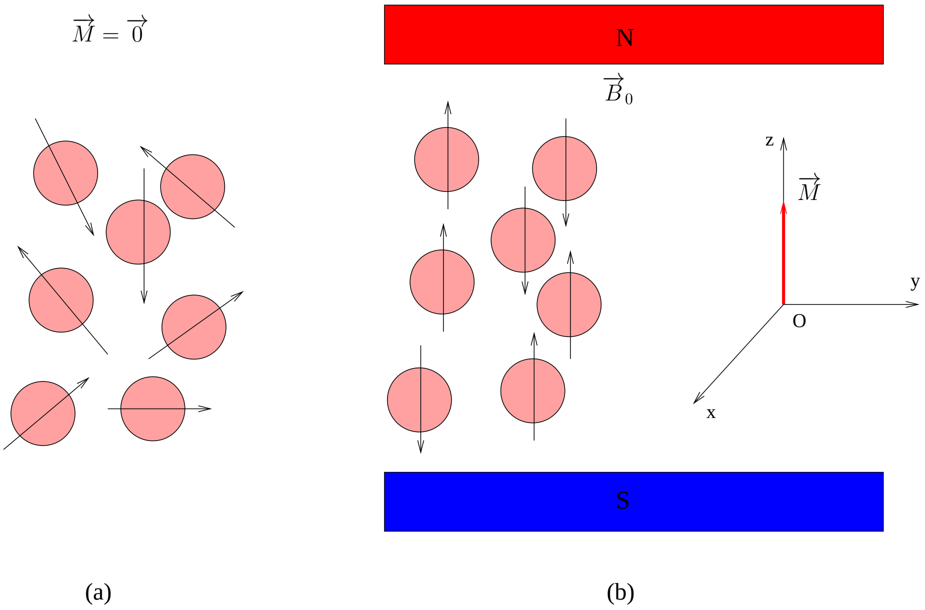


Figure 2.1: (a): magnetic moment at the initial equilibrium state; (b): magnetic moment in the presence of a stationary magnetic field of magnitude B_0.

is abundantly present in the human body and especially in the brain, has a gyromagnetic ratio $\gamma = 42.58$ Mhz/Tesla. In general, $\overrightarrow{M}$ may be split into two components as follows:

$$\overrightarrow{M} = \overrightarrow{M}_{xy} + \overrightarrow{M}_z, \tag{2.1}$$

where $\overrightarrow{M}_{xy}$ and $\overrightarrow{M}_z$ are the transversal and longitudinal components, respectively. The transversal component $\overrightarrow{M}_{xy}$ is contained into the plane (xOy), while the longitudinal component is colinear to the (Oz) axis.

At the new equilibrium state, $\overrightarrow{M}$ is aligned along (Oz) without transversal component (see Fig. 2.1(b)). Therefore, we have $\overrightarrow{M}_{xy} = \overrightarrow{0}$ and $\overrightarrow{M}_z$ which grows with both the concentration of protons in the tissue and the intensity of $\overrightarrow{B}_0$.
However, it is not possible to directly measure $\overrightarrow{M}$ since it is infinitesimal compared to $\overrightarrow{B}_0$. Fortunately, indirect measurement is possible through switching its orientation in the plane xOy using an additional magnetic field $\overrightarrow{B}_1$ (see Fig. 2.2(a)) applied due to a Radio Frequency (RF) pulse. This magnetic field should have an angular frequency $f_1 = f_0$ in order to make the spins resonate and change their energy level.

2.2.2 NMR signal measurement

Subject to excitations, spins are not stable from an energetic viewpoint. When returning to their equilibrium state, the spins will lead to a *precession* movement of the magnetic moment $\overrightarrow{M}$ as shown in Fig. 2.2(b), which allows the measurement of the NMR signal.

As illustrated in Fig. 2.2(b), the magnetic moment $\overrightarrow{M}$ follows a spiral trajectory when returning to its equilibrium position ($\overrightarrow{M}_{xy} = \overrightarrow{0}$). During this time, the two components of the magnetic moment ($\overrightarrow{M}_{xy}$ and $\overrightarrow{M}_z$) behave differently to reach their equilibrium

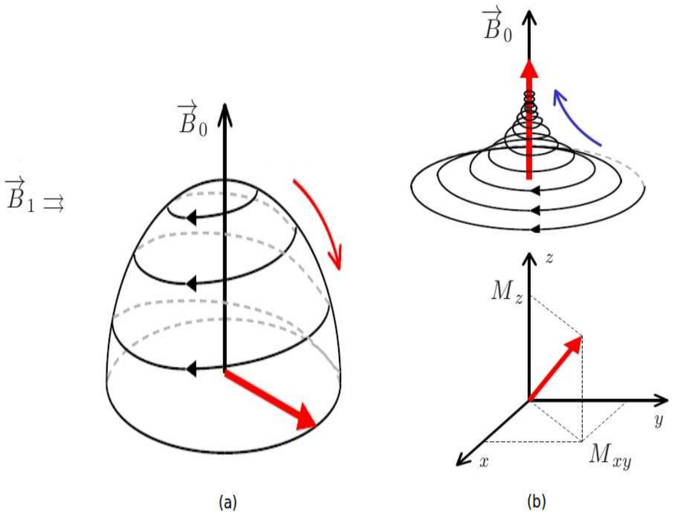


Figure 2.2: Precession movement of the magnetic moment.

state. The longitudinal module M_z returns back to its initial state in an exponential way parameterized by a time constant T_1 called the *spin-lattice* relaxation time:

$$M_z(t) = M_0(1 - e^{-t/T_1}). \tag{2.2}$$

This movement curve is illustrated in Fig. 2.3. In practice, the constant T_1, which generally takes its values between 100 ms and 1000 ms, corresponds to the time when M_z returns to 63% of its original value.

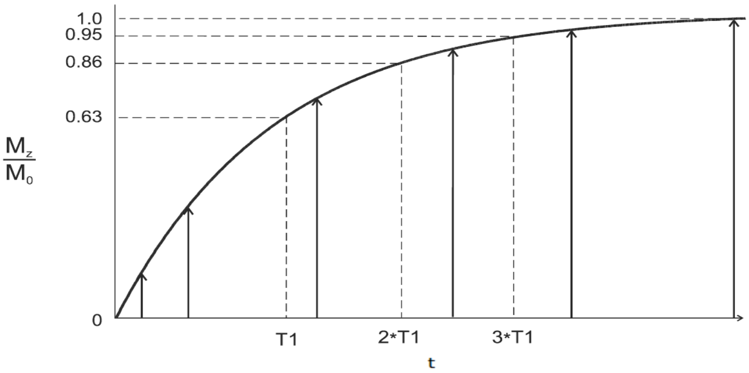


Figure 2.3: T_1 relaxation curve.

At the same time, and as illustrated in Fig. 2.4, the transversal module M_{xy} decreases exponentially as a function of a time constant T_2 called *spin-spin* relaxation time:

$$M_{xy}(t) = M_0 e^{-t/T_2}. \tag{2.3}$$

This constant T_2 corresponds to the time when M_{xy} is at 37% of its value just after the spin excitation. The T_2 relaxation time is typically around 40-100 ms, and it is always

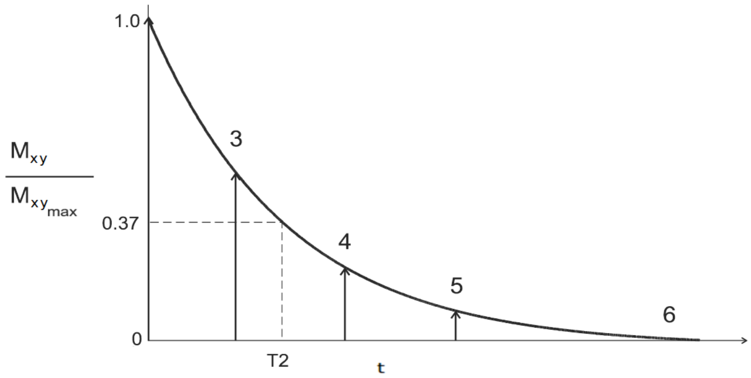


Figure 2.4: T_2 relaxation curve.

shorter than T_1.
However, in practice, inhomogeneities of the main magnetic field $\overrightarrow{B}_0$ and magnetic susceptibility differences make the spins precess at different frequencies. If we denote by $T_{2_{B_0}}$ and $T_{2_{MS}}$ the dephasing times caused by the main magnetic field inhomogeneities and magnetic susceptibility differences, the total decay transversal relaxation time T_2^* is often twice shorter than T_2 and is given by $1/T_2^* = 1/T_2 + 1/T_{2_{B_0}} + 1/T_{2_{MS}}$.

Each tissue is characterized by a given couple of time constants T_1 and T_2, which allow one to distinguish between the different tissues within the same MRI image. More precisely, in brain imaging for instance, these time constants allow us to distinguish between the white matter, gray matter and Cerebral Spinal Fluid (CSF) [Brown and Semelka, 2003]. In fact, the pixel intensity in an MRI image is proportional to the number of protons contained within the voxel weighted by the T_1 and T_2 relaxation times for the tissues within the voxel. These time constants define what is called a *contrast* and we talk about T_1 and T_2-weighted MRI images depending on the importance of the T_1 or T_2 constants in the images. This fact outlines the importance of the relaxation movement. The impact of the precession movement is also of great interest since it is the direct source of the measured NMR signal. In fact, the spiral rotation of the transversal component $\overrightarrow{M}_{xy}$ induces a magnetic field in the plane xOy. This *Free Induction Decay* (FID) signal is registered using a receiver coil put in the xOy plane. The receiver coil then transforms the FID into a measurable electrical signal.
However, at this stage, it is impossible to spatially localize the NMR signal since the dimensions of the brain (or the imaged organ) are small compared to the used wavelength at standard magnetic field strength, i.e. 1.5 or 3 Tesla. This assumption is no longer valid at ultra high magnetic fields such as 7 or 11 Tesla, where the used wavelength become smaller than the brain size.
In order to be able to encode an MRI image, it is necessary to proceed by a spatial coding of the measured NMR signal. This was proposed by Lauterbur in 1973 [Lauterbur, 1973] who demonstrated that it is possible to reconstruct images from the NMR signal using a superposition of linear magnetic field gradients. Doing so, the MRI technique consists of three steps: slice selection, phase encoding and frequency encoding.

2.2.2.1 Slice selection

Although 3D imaging techniques have been developed more recently (3D- EPI [Afacan et al., 2009] or EVI [Rabrait et al., 2008]), the MRI technique generally proceeds by a slice by slice (2D) acquisition as illustrated in Fig. 2.5. For this reason, the first step in the acquisition process is the slice selection. Suppose that we are interested in acquiring a slice located at some z position in the Cartesian coordinate system (x, y, z) having the same orientation as in Fig. 2.5. In order to excite the spins belonging to this slice, a magnetic field gradient $\overrightarrow{G}_z$ orthogonal to the plane xOy is applied. For the excited spins, the related Larmor frequency writes as follows:

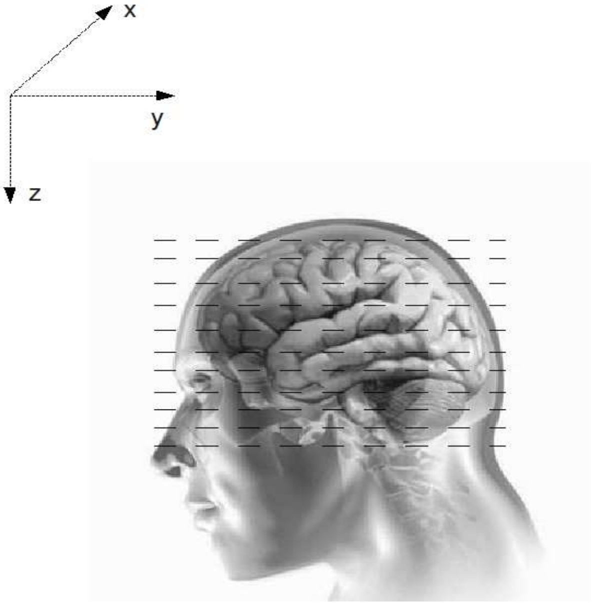


Figure 2.5: Slice by slice acquisition in MRI.

$$f_z(t) = \frac{\gamma}{2\pi}(B_0 + zG_z(t)), \tag{2.4}$$

where t is the acquisition time and G_z is the norm of $\overrightarrow{G}_z$. Hence, the Larmor frequency of these spins will allow us to distinguish them from spins belonging to the rest of the brain. The next step aims at spatially encoding the NRM signal measured from a given slice. For doing so, two other encoding steps are applied: frequency and phase encodings.

2.2.2.2 Frequency encoding

This step is called frequency encoding because the spatial position along the x-axis is encoded using the precessing frequency of the spins. This direction (x-axis) is also called the read-out because the frequency encoding gradient is turned on during the signal read-out (acquisition of the NMR signal). Subject to this second magnetic field gradient $\overrightarrow{G}_x$ applied in the x direction, the spins will precess at a Larmor frequency which depends on

x as follows:

$$f_x = \frac{\gamma}{2\pi}(B_0 + xG_x), \tag{2.5}$$

where G_x is the norm of $\overrightarrow{G}_x$. The longer the gradient $\overrightarrow{G}_x$ applied, the higher the spatial frequency in the measured 1D signal. Using the expression of the Larmor frequency of the excited spins in Eq. (2.5), the measured signal from the excited Field of View (FOV) writes:

$$\begin{aligned}\widetilde{d}(k_x) &\propto \int_{\text{FOV}} \overline{\rho}(x)e^{-\imath 2\pi f_x t}dx \\ &\propto \int_{\text{FOV}} \overline{\rho}(x)e^{-\imath x k_x}dx \end{aligned} \tag{2.6}$$

where t is the acquisition time, $k_x = \gamma G_x t$ and $\overline{\rho}$ is the spin density in the imaged volume[1].

2.2.2.3 Phase encoding

After slice selection and frequency encoding, it is necessary to encode the received NMR signal along the third spatial direction, i.e. the y-axis. In this respect, the phase of the NMR signal is used, performing therefore what is called the phase encoding step. To this end, a magnetic field gradient $n\overrightarrow{G}_y$ is applied in the y direction during a short time T_y (w.r.t the acquisition time) before the readout, where the integer n changes for each acquisition. As a result, during this short time, the spins precess with a spatially dependent phase:

$$\varphi(y) = \frac{\gamma}{2\pi} y n G_y T_y, \tag{2.7}$$

where G_y is the norm of $\overrightarrow{G}_y$. Within a given slice, spins located at a spatial position (y, x) will therefore precess at a unique frequency. However, in contrast with frequency encoding where the gradient $\overrightarrow{G}_x$ is applied once, in order to cover all the imaged space, the repetition of the phase encoding step by changing either the gradient's module nG_y or duration T_y is essential.

Regarding the new precession frequency of the excited spins after a frequency and phase encoding steps, the 2D measured signal will therefore be expressed as:

$$\begin{aligned}\widetilde{d}(k_y, k_x) &\propto \int\int_{\text{FOV}} \overline{\rho}(y, x)e^{-\imath 2\pi(f_x t + \varphi(y))}dxdy \\ &\propto \int\int_{\text{FOV}} \overline{\rho}(y, x)e^{-\imath(yk_y + xk_x)}dxdy, \end{aligned} \tag{2.8}$$

with $k_y = \gamma n G_y T_y$.

After slice selection, frequency and phase encoding, it is possible to fully encode the measured NMR signal and provide an MRI image for a given slice, and therefore for all the imaged volume by iterating over slices. The sequencing of the encoding steps is illustrated in Fig. 2.6 where the diagram of a GRadien-Echo (GRE) sequence is given with

---

[1] The overbar is used to distinguish the "true" data from a generic variable.

the following definitions of the sequence parameters:

- TE: Echo Time: the time between the RF excitation and the signal acquisition time,
- TR: Time of Repetition: the time between successive excitation pulses for a given slice.

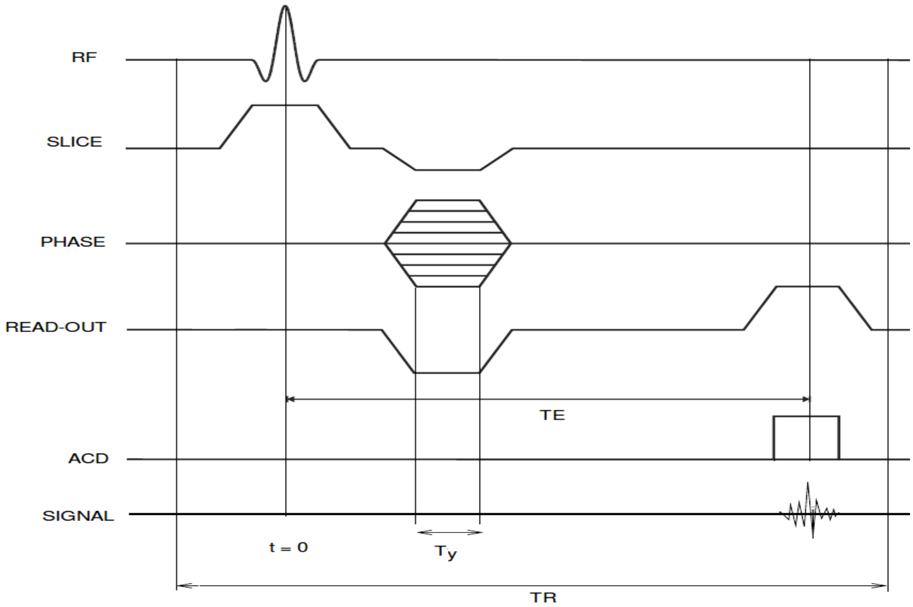


Figure 2.6: GRE sequence diagram (ADC denotes the Analog-to-Digital Converter).

For a given slice, the measured signal writes as in Eq. (2.8). However, in practice, only a finite number of N_x samples is acquired at equidistant time intervals Δt during the frquency encoding step, which means that a discrete sampling of the k-space is performed at equidistant frequency intervals $\Delta k_x = \gamma G_x \Delta t$. Similarly, only N_y samples are acquired along the phase encoding direction at equidistant frequency intervals $\Delta k_y = \gamma G_y T_y$ by changing the integer n.

2.3 Image contrast and acquisition parameters

An MRI contrast consists of transcoding the acquired NMR signal into gray levels. It reflects the relaxation times and spin density differences between the image tissues. The main two factors in an MRI contrast are T_1 and T_2. Since these factors always contribute in different ways to define a contrast, changing the sequence parameters (i.e. TE and TR) allows indirectly defining a contrast through managing the respective contributions of these main factors.

2.3.1 TR effect

The module of the longitudinal aimantation M_z returns nearly to its original maximum value M_0 function to TR, and then it is re-switched to the transversal plane (xOy) due to the RF pulse so that a new cycle is iterated. Depending on TR, one of the following cases may be encountered:

- TR is long w.r.t. T_1 of the imaged tissues: the longitudinal aimantation returns completely to its equilibruim position at the end of each cycle (after each TR),
- TR is short w.r.t. T_1 of the imaged tissues: the longitudinal relaxation curve is interrupted and M_z will not completely return to its initial value.

TR influences therefore the longitidunal relaxation, and so the T_1 contrast (also called T_1 weighting) of a given MRI sequence.
In fact, let us consider two different tissues belonging to the same imaged FOV, each of them has its own relaxation time T_1. We will suppose that the first (F) and second (S) tissues have a fast and slow relaxation times T_1^F and T_1^S, respectively. If TR is long comparing to T_1^F and T_1^S, spins of the two tissues will recover their initial longitudinal aimantations. In this case, it will be impossible to distinguish the two tissues although they have different longitidunal relaxation times. On the other hand, if TR is short, the aimantation of the tissue F will return faster than S to its initial situation. The measured signal from the tissue F will therefore be stronger than the one from S, which defines the T_1 contrast.

2.3.2 TE effect

At the begining of each cycle, the transversal aimantation will appear due to the RF pulse. As mentioned above, TE determines the time when the NRM signal is measured, or equivalently the time during which the transversal aimantation returns to its initial value. Let us consider now the same example of Section 2.3.1, but with tissues having fast and slow T_2 relaxation times. If TE is long comparing to T_2^F and T_2^S, the two tissues will recover their initial transversal aimantations, and it will be impossible to distinguish them although they have different transversal relaxation times. However, if TE is short, the aimantation of the tissue F will return faster than S to its initial situation. The measured signal from the tissue S will therefore be higher than the one from F, which defines the T_2 contrast.

2.3.3 Acquiring a T_1 or T_2-weighted MRI image

Based on the roles of TR and TE, acquiring a T_1 or T_2-weighted MRI image will be performed by fixing the values of TR and TE while accounting for the image tissue features (T_1 and T_2). To acquire a T_1-weighted image, the sequence parameters have to be fixed as follows:

- short TR in order to make appearing the T_1 contrast,
- short TE in order to cancel the T_2 contrast.

However, to acquire a T_2-weighted image, the sequence parameters have to be fixed as follows:

- long TR in order to reduce the T_1 contrast,
- long TE in order to enhance the T_2 contrast.

2.4 Functional Magnetic Resonance Imaging

In order to understand the structure and function of the human brain, neuro-scientists have been probing brain's structure and function using indirect methods for a very long time. Starting from post-mortem analysis, scientists succeeded to establish functional cerebral maps of the human brain due to electrical stimulations applied directly to the brain during a neuro-surgery by Wilder Penfield [Penfield and Rasmussen, 1952]. Since the early 90's, different imaging modalities have revolutionized this active research field making it possible to collect real-time information about what is happening into the human brain without opening the cerebral box.

Generally speaking, we talk about *anatomical* and *functional* imaging. The first modality is designed to highlight the cerebral structures and outline any disorder or damage (tumor, deformation, hemorrhage ...).

The second modality is rather designed to measure brain activity during some specific tasks (also called *stimuli*) or to probe intrinsic or ongoing activity (*resting state*). Besides fundamental research aiming at understanding the organization of cerebral structures, functional imaging is also used for epileptic diagnosis, pre-surgery investigations in order to preserve some specific brain areas, studying the effect of new drugs designed to cure some brain diseases or neurological disorders like Alzheimer, schizophrenia,... However, for simple brain functions, these two imaging modalities may be coupled in order to draw functional cerebral maps by linking each brain area to its functional role. For more complicated brain functions, connections and interactions between the involved brain areas may also be studied.

In this context, functional Magnetic Resonance Imaging (fMRI) is a recent neuroimaging technique for measuring brain activity. Generally speaking, it is used to detect brain areas which are involved in a specific task (e.g. simple auditory visual or motor) or more complex cognitive process (e.g. language, computation,...). It can also be used to study emotions, attention, memory, the intrinsic activity of the brain,... This technique is being widely developed due to its strength especially when coupled with anatomical MRI [Ogawa et al., 1990; Bandettini et al., 1993; Chen and Ogawa, 1999]. The main advantage of fMRI is its non-ionisant property which allows one to non-invasively establish functional maps of activated areas. Moreover, compared to other imaging modalities like Positron Emission Tomography (PET), fMRI has a better spatial and temporal resolution. However, it should be noted that at standard magnetic field intensities such as 1.5 or 3 Tesla, fMRI has a lower chemical/molecular resolution. This drawback is being alleviated due to the use of ultra high magnetic fields such as 7 Tesla.

2.4.1 Blood Oxygen Level Dependent effect

Signal acquisition in fMRI is based on oxygenation variation and blood flow. In fact, it has been observed by Roy and Sherrington [Roy and Sherrington, 1890] since 1890 that there is a blood flow increase in the cortex of subjects exposed to stimulation. A link between neuronal activity and blood flow has therefore been established: cerebral activity changes the oxygen concentration in the blood within the cortex. It has also been highlighted that hemoglobin, the molecule that carries oxygen in the blood, is mainly responsible for this concentration variation. In fact, oxyhemoglobin (hemoglobin loaded of oxygen) is a diamagnetic molecule and does not affect the local magnetic field, whereas the deoxyhemoglobin (hemoglobin unloaded of oxygen) is paramagnetic and disturbs the relaxation phenomena of the spins, which is the source of the BOLD (*Blood Oxygen Level Dependent*) signal [Ogawa et al., 1990]. A change of the BOLD signal in a given brain region indicates then that this region is involved in the stimulation delivered to the subject. The BOLD signal hence allows indirect measurement of neuronal activity [Logothetis et al., 2001].

2.4.2 Data acquisition in fMRI

During an fMRI *run*, the whole brain volume is imaged many times. The acquisition of each volume is called a *scan*. The requested time to acquire a scan is called *Time-of-Repetition*. Generally speaking, a scan is performed in about two seconds, while a run may last up to ten minutes. During an fMRI *session*, several runs may be acquired in addition do eventual anatomical or diffusion [Filler et al., 1991] MRI acquisitions. However, an fMRI *session* should last less than 90 minutes for adults and 45 minutes for children. During this acquisition time, and unless a resting state fMRI study is conducted, the subject is exposed to an experimental paradigm. Two kinds of paradigms may be used: event-related or block paradigms. The first one consists of presenting isolated and random stimuli with jittering to the subject. Each of these isolated stimuli lasts less than two or three seconds. However, in a block paradigm, repeated tasks every two seconds are presented to the subject during 30 to 40 seconds. Fig. 2.7 illustrates an example of an event-related and block paradigms where several conditions (i.e. A, B, C,...) are differently delivered. In a given experimental paradigm, one or more tasks can actually be involved. These tasks (stimuli) may be visual (e.g. face recognition), motor (e.g. pressing a button), auditory (e.g. listening to sentences) or cognitive (e.g. mental calculation, reading,...)... Subject to the presented stimuli, the time course of the BOLD signal represents the *haemodynamic response function* (HRF), which describes the brain response to the related stimuli. Fig. 2.4.2 illustrates an example of stimuli and the related BOLD signal acquired in a voxel which belongs to the activated area.

Ideally, if the concerned voxel belongs to the activated brain area involved in processing of the the stimulus, the measured BOLD signal should be close to the expected response. However, because of low Signal-to-Noise Ratio (SNR) of MRI images, the BOLD signal is degraded, leading to a decrease of the Contrast to Noise Ratio (CNR). Based on this degraded signal, activated voxels are detected using a statistical analysis. This procedure will be presented in more detail in Chapter 6.

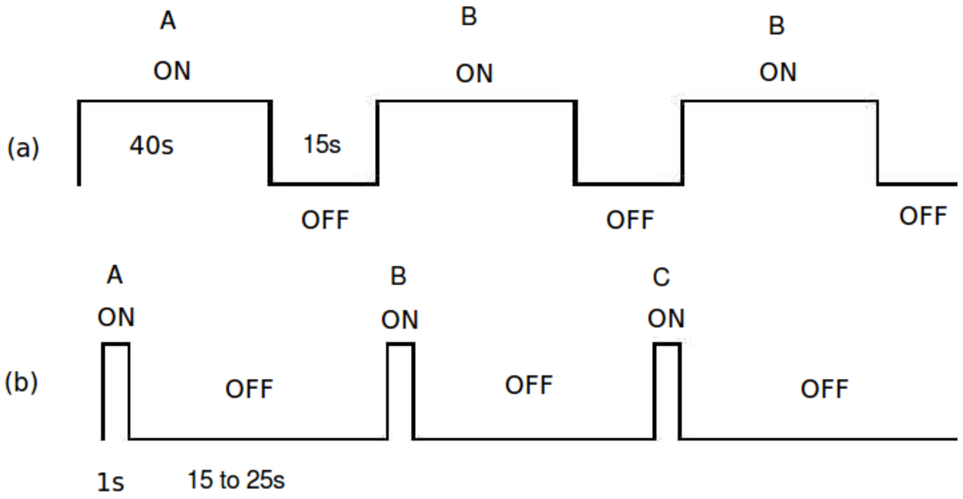


Figure 2.7: Example of experimental paradigms: (a) block paradigm; (b) event-related paradigm.

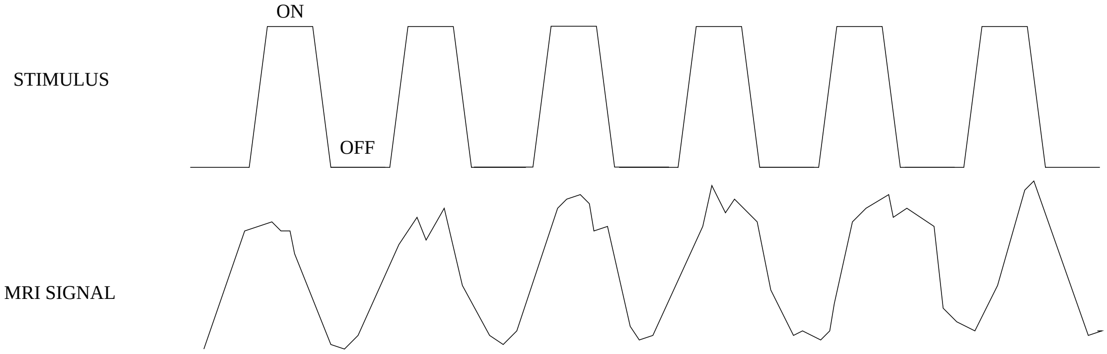


Figure 2.8: Used stimuli with a block paradigm (top) and induced BOLD signal (bottom).

To achieve a good activation detection performace and sentitivity, a trade off has to be made between the spatio-temporal resolution and SNR of the fMRI data. However, acquiring fMRI images of high temporal resolution requires fast imaging sequences like the Echo Planar Imaging (EPI) one, which has been proposed by Mansfield in 1997 [Mansfield, 1997]. This sequence on which rely most of the fMRI studies in the literature allows to reduce the imaging time due to the application of an echo-train after the RF pulse, which enables the acquisition of all the phase encoding steps at once. However, acquired images generally suffer from several forms of artifacts.

2.4.3 Artifacts in fMRI

Generally speaking, two kinds of artifacts may be encountered in EPI fMRI images: artifacts encountered in conventional MRI, and those specific to the EPI sequence. A non exhaustive describtion of these artifacts is given herebelow.

2.4.3.1 Artifacts in conventional MRI

- Metal artifacts:
  These artifacts are caused by the presence of ferromagnetic objects within or next to the imaged organ. They can consist of an increase (hyper-signal) or loss (hypo-signal) of the signal in some areas, or also a deformation of the image (see Fig. 2.9).

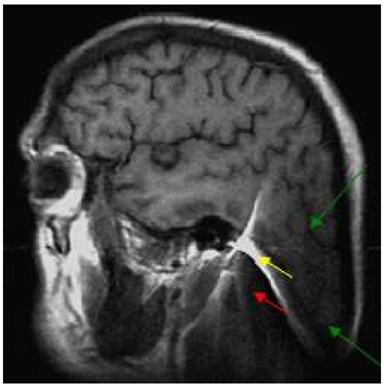


Figure 2.9: Example of metal artifacts; yellow array: hyper-signal areas; red array: hypo-signal areas; green arrays: image deformation [Gerardin, 2006].

- Motion artifacts:
  Motion artifacts occur because of the subject movement between two TRs or during the phase encoding. This kind of artifacts often occurs in dynamic imaging such as abdomen or cardiac imaging. As illustrated in Fig. 2.10 (right), a blur may occur in the reconstructed image if the movement is random. However, phantom images may degrade the image if the movement is periodic like illustrated in Fig. 2.10 (left).

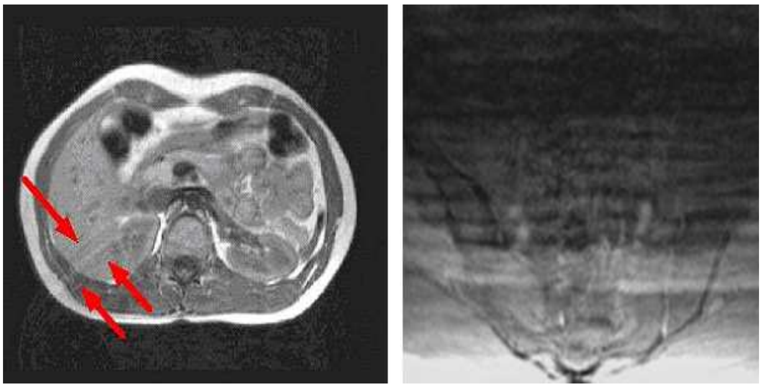


Figure 2.10: Example of motion artifacts; right: blur artifacts; left: phantom artifacts [Gerardin, 2006].

- Magnetic susceptibility artifacts:
  These artifacts are observed when two structures of different magnetic susceptibilities intersect such as the brain/air interface. Next to such regions, we observe an

inhomogeneity of the main magnetic field $\overrightarrow{B}_0$. These inhomogeneities are resposible for two kinds of magnetic susceptibility artifacts: first, they result in a frequency shift of the resoning frequency of the spins, which leads to a shift in the reconstructed images. This kind of artifacts is also called *geometrical distortion* artifacts and it consists of a compression or dilation of the image along the phase encoding direction (see Fig. 2.11 (right)). Second, these $\overrightarrow{B}_0$ inhomogeneities result in an intra-voxel dephasing, which leads to signal loss in the considered areas (see Fig. 2.11 (left)).

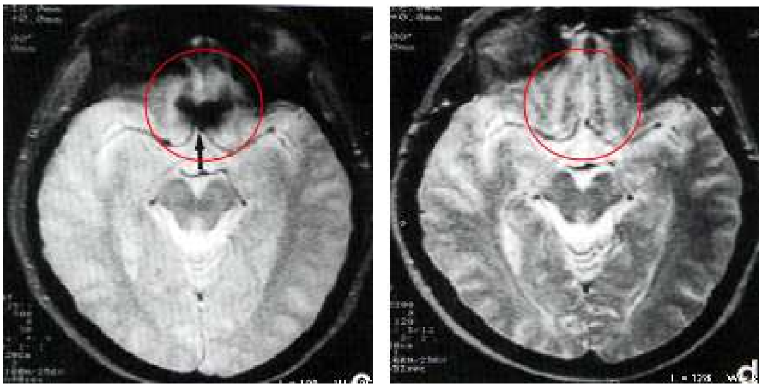


Figure 2.11: Example of magnetic susceptibility artifacts [Gerardin, 2006].

It should be mentioned here that this kind of artifacts is increased in EPI because of the long TE (which increases the dephasing time).

2.4.3.2 Specific EPI artifacts: ghosting artifacts

Ghosting artifacts in EPI are caused by the fact that the Fourier plane is acquired alternatively in two directions (left to right and right to left) as illustrated in Fig. 2.12. This encoding scheme requires a temporal correction of even or odd line before reconstruction because of the gradient dephasing introduced when inverting the encoding pathway. As illustrated in Fig. 2.13, ghosting artifacts appear as additional shifted parasite images.

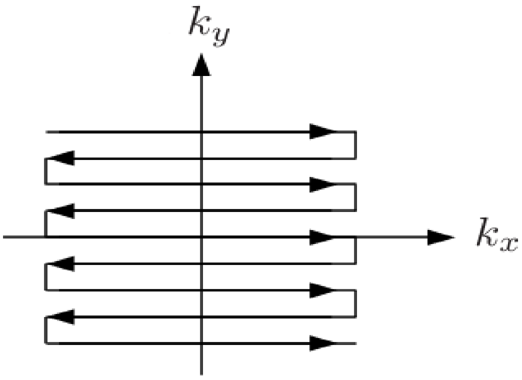


Figure 2.12: Example of ghosting artifacts [Gerardin, 2006].

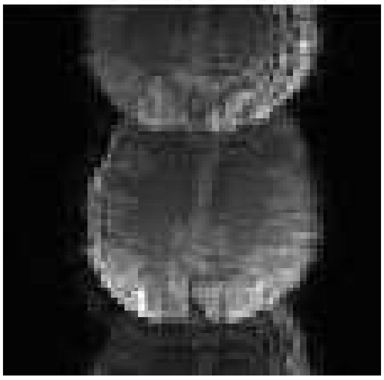


Figure 2.13: Example of ghosting artifacts.

2.5 Parallel Magnetic Resonance Imaging

As detailed in Section 2.2, the frequency encoding step has to be repeated several times to encode a whole line of the k-space. Moreover, in order to encode all the lines of the k-space, the phase encoding step has also to be iterated. However, for fMRI experiments, reducing the global imaging time without significantly degrading the image quality is of great interest for the final goal, i.e. studying the brain functions. Indeed, since an fMRI study requires the acquisition of the brain volume several times in order to track brain activity, reducing the acquisition time (which may lead to reducing TR) allows faster repetition of the brain imaging, which leads to better temporal resolution. It allows hence getting more accurate knowledge about the brain response dynamics. On the other hand, increasing the spatial resolution is also benifical for fMRI since it allows more precise spatial localization of activations. Improving this resolution requires the acquisition of more k-space points. If reducing the TR is not the goal, shortening the acquisition time can be exploited to increase the spatial resolution by acquiring additional k-space points while maintaining the TR almost fixed. In multishot acquisitions, reducing the acquisition time may also be exploited to acquire more images, and hence to increase the acquisition SNR. It leads also to shorter read-out duration, and hence allows the reduction of reconstruction artifacts such as magnetic susceptibility ones even for anatomical MRI.

Since switching on/off the magnetic field gradients is technically time-consuming, a solution to reduce the global imaging time lies in the use of parallel imaging systems. In such systems, multiple receiver coils with complementary sensitivity profiles located around the underlying object are used to simultaneously collect MRI data in the k-space. In order to speed up the acquisition or reduce reconstruction artifacts, the data are sampled at a frequency rate R times lower than the Nyquist sampling rate along at least one spatial direction, i.e. usually the phase encoding one. Parallel MRI presents therefore the following advantages for both anatomical and functional MRI:

- reducing geometrical distortion artifacts by increasing the phase-encoding bandwidth [Weiger et al., 2002; Lin et al., 2005];
- reducing magnetic susceptibility artifacts by reducing TE;

- improving the patient comfort during an MRI exam by shortening the global acquisition time and the acoustic noise by reducing the gradient switching [De Zwart et al., 2002] (since not all the k-space is acquired);
- improving the spatio-temporal resolution of fMRI images;
- reducing motion artifacts by shortening the echo-train in EPI (functional).

However, as any imaging technique, resorting to pMRI is not completely costless. Using pMRI systems generally leads to lower acquisition Signal to Noise Ratio (SNR) than with conventional MRI. This SNR decrease is less important in functional EPI images since an important gain is achieved due to the reduction of the previously described artifacts. Using pMRI systems requires also a specific and more complicated reconstruction step to build up a full FOV image by unfolding the undersampled coil-specific data sampled under the Nyquist rate. This pMRI reconstruction is a challenging task because of the low SNR due to four main artifact sources:

- aliasing artifacts related to the undersampling rate;
- acquisition noise;
- errors in the estimation of coil sensitivity maps;
- sub-optimal coil geometry to provide sufficient independent complementary information from all the array.

Using parallel imaging to reduce global imaging time at fixed spatial resolution, the TR is approximately R times shorter than with conventional acquisitions. Hence, in imaging sequences involving a single volume acquisition, the global imaging time is directly impacted: acquiring a high resolution T_1-weighted image takes about 9 min using convential settings and only 5 min (respectively, 3 min) using $R = 2$ (resp. $R = 4$) in pMRI. This allows to reduce the exposition time to the magnetic fields for the subject when only anatomical data have to be acquired. In dynamic imaging like fMRI, the TR reduction is usually exploited in another direction, namely the improvement of spatial or temporal resolution instead of the decrease of the global scanning time. One underlying reason is that the EPI sequence on which fMRI data acquisition relies has low spatial resolution in conventional whole brain imaging (typically 64×64 pixels by slice). In that case, pMRI enables high resolution imaging up to 128×128 for the same TR, provided that a high enough R-factor is used.

2.5.1 Parallel MRI basics

As illustrated in Fig. 2.14, an array of L coils with complementary spatial sensitivity profiles is used to measure the spin density $\overline{\rho}$ in the object under investigation. The signal $\widetilde{d}_\ell$ received by each coil ℓ $(1 \leq \ell \leq L)$ is the Fourier transform of the desired 2D field $\overline{\rho}$ on the specified FOV weighted by the coil sensitivity profile s_ℓ, evaluated at some locations

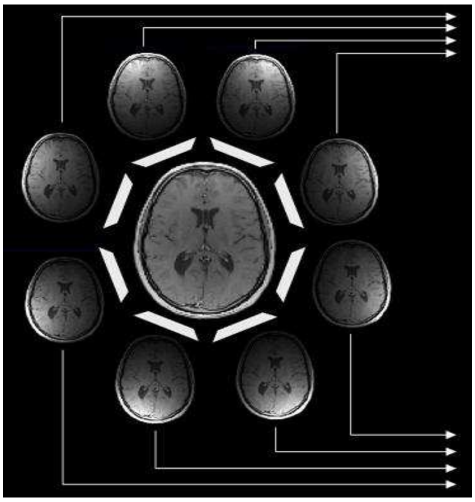


Figure 2.14: Parallel acquisition in MRI using an 8 coils array.

$\boldsymbol{k} = (k_y, k_x)^\intercal$ in the k-space:

$$\widetilde{d}_\ell(\boldsymbol{k}) = \int_{\mathrm{FOV}} \overline{\rho}(\boldsymbol{r}) s_\ell(\boldsymbol{r}) e^{-\imath \boldsymbol{k}^\intercal \boldsymbol{r}} d\boldsymbol{r} + \widetilde{n}_\ell(\boldsymbol{k}), \tag{2.9}$$

where $\widetilde{n}_\ell(\boldsymbol{k})$ is a coil-dependent additive zero-mean Gaussian noise [Pruessmann et al., 1999a; Sijbers et al., 2007], which is independent and identically distributed (iid) in the k-space, and $\boldsymbol{r} = (y, x)^\intercal$ is the spatial position in the image domain. The size of the acquired data $\widetilde{d}_\ell$ in the k-space clearly depends on the sampling scheme. For the sake of simplicity, a Cartesian coordinate system is generally adopted in the neuroimaging context. In this case, Fig. 2.15 clearly illustrates the differences between conventional and parallel acquisitions in terms of k-space sampling. In parallel MRI, the sampling period along the phase encoding direction is R times larger than the one used for conventional acquisition, $R \leq L$ being the reduction factor.

Note however that other sampling schemes have been used in the pMRI literature like the spiral one. Spiral trajectories are in fact less sensitive to motion and blood flow artifacts [Noll et al., 1995; Yan et al., 1999], which explains the fact that they are usually used for cardiac imaging for instance. However, other artifacts such as the geometrical distorsion ones are more complicated to reduce when spiral encoding trajectories are adopted. Spiral sequences are also more sensitive to magnetic field inhomogeneities. Fig. 2.16 illustrates three kinds of spiral trajectories. The advantage of square spiral trajectories is that no interpolation is requested for a Cartesian grid. However, circular spiral trajectories require lower magnetif field gradient power. In this manuscript, we will focus on the Cartesian encoding scheme as it is often used in brain imaging, the application that we are interested in.

Using an eight coils array whose sensitivity profiles are illustrated in Fig. 2.17(a),

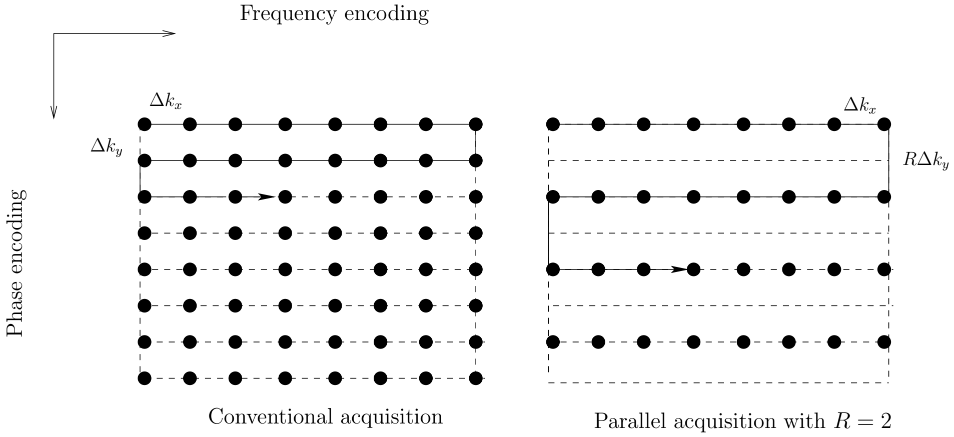


Figure 2.15: Sampling schemes for conventional and parallel acquisitions on a Cartesian grid. Here $R = 2$ indicates a subsampling along the phase encoding direction (y-axis) by a factor of two.

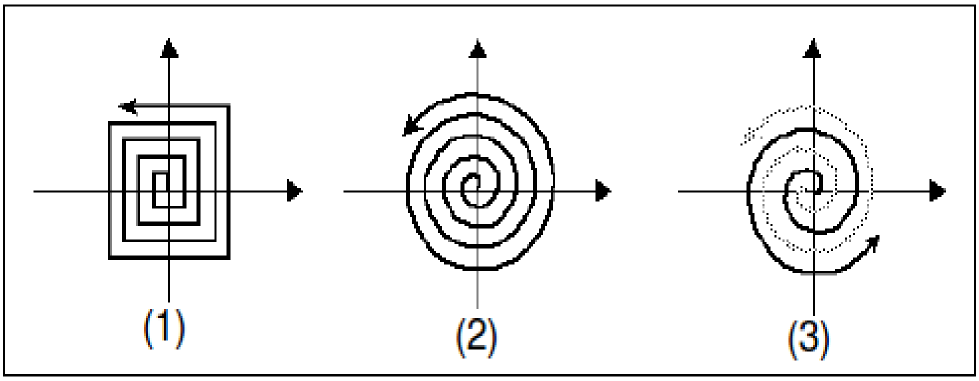


Figure 2.16: Spiral encoding trajectory; (1): square spiral, (2): circular spiral, (3): circular spiral for segmented acquisition.

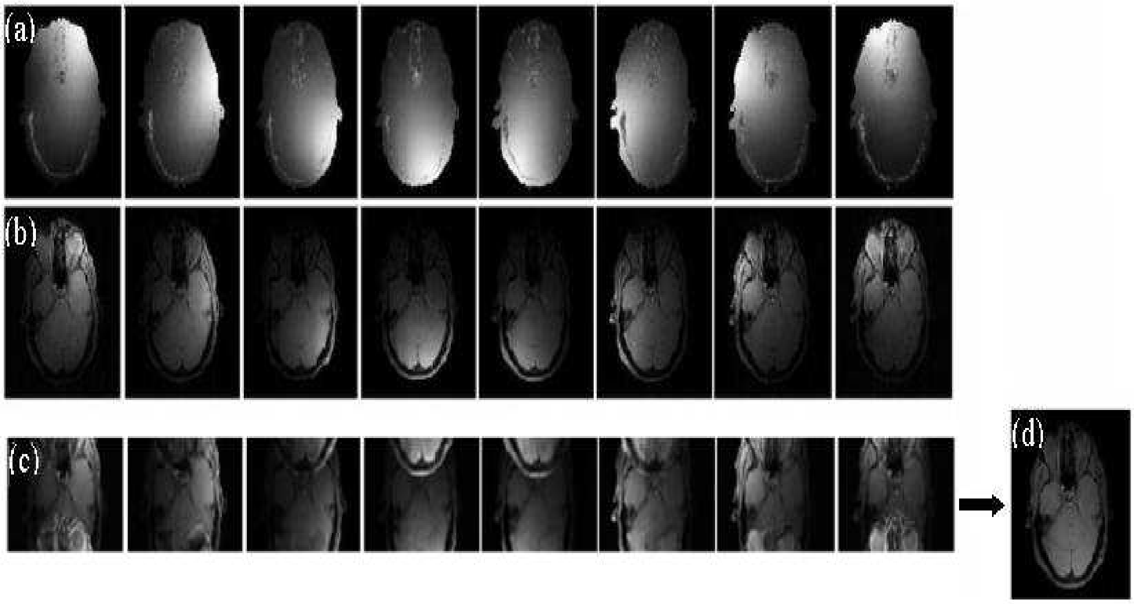


Figure 2.17: Parallel acquisition in MRI using an 8 coils array: (a) coil sensitivity profiles; (b) acquired images without subsampling the k-space; (c) acquired reduced FOV images; (d) reconstructed full FOV image.

conventional sampling of the k-space would lead to the coil images shown in Fig. 2.17(b). However, in practice, and because of subsampling the k-space under the Nyquist rate, aliasing appears in the spatial domain after an inverse Fourier transform as illustrated in Fig. 2.17(c). The reconstruction process then consist of recovering the full FOV image such as in Fig. 2.17(d) based on:

- reduced FOV images;
- coil sensitivity profiles;
- acquisition noise properties.

This step may be performed either in the image domain or directly in the k-space.

2.5.2 Parallel MRI reconstruction

2.5.2.1 Brief state of art

The Simultaneous Acquisition of Spatial Harmonics (SMASH) [Sodickson and Manning, 1999], introduced by Sodickson and Manning in 1997, was the first reconstruction method to operate in the k-space. It uses a linear combination of pre-estimated coil sensitivity maps to generate the missing phase encoding steps. Some other k-space based reconstruction techniques have also been proposed like GRAPPA (Generalized Autocalibrating Partially Parallel Acquisitions) [Griswold et al., 2002] or generalizations of the SMASH method (AUTO-SMASH) [Jakob et al., 1998; Heidemann et al., 2001]. The general idea of these reconstruction methods operating in the k-space is that the missing k-space lines (not acquired) are reconstructed using linear combinations of acquired lines. However, the specificity of SMASH is that it needs a separate coil sensitivity map estimation in contrast with these most recent generalizations, which proceed by acquiring a few additional k-space lines to derive coil sensitivity maps without a reference scan. This explains why these extensions are referred to as *autocalibrated*.

Indeed, GRAPPA and its alternatives may be preferable to non-autocalibrated methods when accurate coil sensitivity maps may be difficult to extract. This occurs when reference scans are difficult to acquire either because of limited global imaging time or spatial resolution changes. For instance, in lung and abdomen imaging, the numerous inhomogeneous regions with a low spin density make the estimation of the sensitivity information inaccurate. Reference scans may also appear inappropriate in other circumstances, typically when they are not able to account for artifacts that will appear later during the parallel imaging sequence. This is especially critical in dynamic imaging in the case of motion artifacts. However, all the reported methods may suffer from phase cancellation problems (since the missing k-space lines are recovered using linear combination of acquired lines), low SNR during the acquisition process and limited reconstruction quality.

In [Pruessmann et al., 1999a], an alternative reconstruction method called SENSitivity Encoding (SENSE) has been introduced. In its simplified form, which relies on Cartesian k-space sampling, SENSE is a two-step procedure involving first a reconstruction of reduced FOV images and second a spatial unfolding technique, which amounts to a weighted least squares estimation. This technique requires a precise estimation of coil sensitivity maps using a reference scan (usually a 2D GRE). Note that for non-Cartesian sampling schemes, like the spiral one, SENSE reconstruction is more complicated and requires iterative reconstruction [Pruessmann et al., 2001].

Extensions of the SENSE method such as in vivo SENSE [Sodickson, 2000], JSENSE [Ying and Sheng, 2007] or mSENSE [Wang et al., 2001] have also been proposed to avoid separately estimating the coil sensitivity maps based on a reference scan. Table 2.1 gives a summarized overview of pMRI reconstruction methods.

To the best of our knowledge, in actual clinical daily routines, only GRAPPA and SENSE-like algorithms are available on scanners: Siemens provides GRAPPA and mSENSE techniques which roughly correspond to ASSET and ARC methods on General Electric Healthcare machines. In practice, SENSE and GRAPPA perform similarly and provide comparable reconstruction SNR. The main difference is that GRAPPA is autocalibrated while SENSE requires separable coil sensitivity estimation. This difference is no longer significant since autocalibrated methods for SENSE like mSENSE are available. However, it is known that SENSE performs better than GRAPPA when reconstructing image details (high frequency content) provided that a good estimation of the sensitivity maps is available. On the other hand, GRAPPA allows better reconstruction when no precise sensitivity maps are available (artifacts, low spin densities,...) or in the presence of subject movement. For a general overview of reconstruction methods in pMRI the reader may refer to [Hoge et al., 2005; Schoenberg et al., 2007].

In the rest of this manuscript we will limit ourselves to the SENSE method since the considered application (i.e. brain imaging) allows deriving precise sensitivity maps and

Table 2.1: Classification of the main pMRI reconstruction methods depending of the reconstruction space.

Method	folded data domain	unfolded data domain
SMASH [Sodickson and Manning, 1999]	Fourier	Fourier
Cartesian SENSE [Pruessmann et al., 1999a]	Spatial	Spatial
Generalized SENSE [Pruessmann et al., 2001]	Fourier	Spatial
SPACE-RIP [Kyriakos et al., 2000]	Fourier	Spatial
AUTO-SMASH [Jakob et al., 1998]	Fourier	Fourier
GRAPPA [Griswold et al., 2002]	Fourier	Fourier

requires better reconstruction of image details. Although 2D extensions of SENSE have been investigated in some recent works as in [Van Der Zwaag et al., 2006; Rabrait, 2007; Rabrait et al., 2008], we will also focus here on the 1D Cartesian version of SENSE.

2.5.2.2 The pMRI acquisition model in the image domain

After discretization of the measured signal in Eq. (2.9) with samplig steps Δk_y and Δk_x, a 2D inverse Fourier transform allows us to recover the measured signal by a coil ℓ (of sensitivity profile s_ℓ) in the spatial domain, which is defined as

$$d_\ell(y,x) = \frac{\Delta k_y \Delta k_x}{4\pi^2} \sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} \widetilde{d}_\ell(p\Delta k_y, q\Delta k_x) e^{i(py\Delta k_y + qx\Delta k_x)} \tag{2.10}$$

where $y \in [0, \dfrac{2\pi}{\Delta k_y}]$ and $x \in [0, \dfrac{2\pi}{\Delta k_x}]$.
Using the Poisson summation formula, we have

$$\sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} \widetilde{d}_\ell(p\Delta k_y, q\Delta k_x) e^{i(py\Delta k_y + qx\Delta k_x)} =$$

$$\frac{\Delta k_y \Delta k_x}{4\pi^2} \sum_{l=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \overline{\rho}(y + \frac{l2\pi}{\Delta k_y}, x + \frac{m2\pi}{\Delta k_x}) s_\ell(y + \frac{l2\pi}{\Delta k_y}, x + \frac{m2\pi}{\Delta k_x}), \tag{2.11}$$

which leads to the following expression of the measured signal by a coil ℓ in the spatial domain:

$$d_\ell(y,x) = \sum_{l=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \overline{\rho}(y + \frac{l2\pi}{\Delta k_y}, x + \frac{m2\pi}{\Delta k_x}) s_\ell(y + \frac{l2\pi}{\Delta k_y}, x + \frac{m2\pi}{\Delta k_x}), \tag{2.12}$$

By accounting for the reduction factor R along the phase encoding direction, the dimensions of the bounded support imaged FOV will be $\dfrac{\text{FOV}_y}{R} = \dfrac{2\pi}{\Delta k_y}$ and $\text{FOV}_x = \dfrac{2\pi}{\Delta k_x}$ along the phase and frequency encoding directions, respectively. In this case, the undersampled measured signal in the image domain reads:

$$d_\ell(y,x) = \sum_{l=0}^{R-1} \overline{\rho}(y + \frac{l\text{FOV}_y}{R}, x) s_\ell(y + \frac{l\text{FOV}_y}{R}, x). \tag{2.13}$$

In the image domain, this signal has to be sampled using some sampling steps Δy and Δx. Let $Y = \dfrac{\text{FOV}_y}{\Delta y}$ and $X = \dfrac{\text{FOV}_x}{\Delta x}$ be the sizes of the full FOV image along the phase and frequency encoding directions, respectively assumed to be integer values. We have for any $(i,j) \in \{0,\dots,\frac{Y}{R}-1\} \times \{0,\dots,X-1\}$,

$$d_\ell(i\Delta y, j\Delta x) = \sum_{l=0}^{R-1} \overline{\rho}((i + l\frac{Y}{R})\Delta y, j\Delta x) s_\ell((i + l\frac{Y}{R})\Delta y, j\Delta x). \tag{2.14}$$

By normalizing $\Delta y = \Delta x = 1$, and by a slight abuse of notations substituting i for y and j for x, we retrieve the following discrete form of the measured signal:

$$\forall (y,x) \in \{0,\dots,\frac{Y}{R}-1\} \times \{0,\dots,X-1\},\ d_\ell(y,x) = \sum_{l=0}^{R-1} \overline{\rho}(y + \frac{lY}{R}, x) s_\ell(y + \frac{lY}{R}, x). \tag{2.15}$$

Accounting for the whole set of L coils, the observation model at a spatial position $\boldsymbol{r}$ can finally be expressed in a matrix form as follows:

$$\boldsymbol{d}(\boldsymbol{r}) = \boldsymbol{S}(\boldsymbol{r})\overline{\boldsymbol{\rho}}(\boldsymbol{r}) + \boldsymbol{n}(\boldsymbol{r}), \tag{2.16}$$

where

$$\boldsymbol{S}(\boldsymbol{r}) \triangleq \begin{pmatrix} s_1(y,x) & \dots & s_1(y + (R-1)\frac{Y}{R}, x) \\ \vdots & \vdots & \vdots \\ s_L(y,x) & \dots & s_L(y + (R-1)\frac{Y}{R}, x) \end{pmatrix},$$

$$\overline{\boldsymbol{\rho}}(\boldsymbol{r}) \triangleq \begin{pmatrix} \overline{\rho}(y,x) \\ \overline{\rho}(y + \frac{Y}{R}, x) \\ \vdots \\ \overline{\rho}(y + (R-1)\frac{Y}{R}, x) \end{pmatrix}, \quad \boldsymbol{d}(\boldsymbol{r}) \triangleq \begin{pmatrix} d_1(y,x) \\ d_2(y,x) \\ \vdots \\ d_L(y,x) \end{pmatrix} \tag{2.17}$$

$$\text{and} \quad \boldsymbol{n}(\boldsymbol{r}) \triangleq \begin{pmatrix} n_1(y,x) \\ n_2(y,x) \\ \vdots \\ n_L(y,x) \end{pmatrix}.$$

In Eq. (2.16), each coil-dependent noise n_ℓ is a spatially iid circular zero-mean complex-valued Gaussian process defined over the FOV of size $Y \times X$. However, the sequence $\boldsymbol{n}$ is identically distributed and spatially independent, but it is not independent at a given location $\boldsymbol{r}$ in the sense that its components $(n_\ell(\boldsymbol{r}))_\ell$ are correlated, which also reads $\boldsymbol{n}(\boldsymbol{r}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is the between-coil $L \times L$ covariance matrix [Pruessmann et al., 1999a; Sijbers et al., 2007]. In practice, $\boldsymbol{\Psi}$ is estimated by acquiring L images $(\underline{\mathrm{d}}_\ell)_{1 \leq \ell \leq L}$ from all coils without radio frequency pulse, and its generic entry $\boldsymbol{\Psi}(\ell_1, \ell_2)$ corresponding to the covariance between the two coils ℓ_1 and ℓ_2 is given by:

$$\forall (\ell_1, \ell_2) \in \{1, \dots, L\}^2, \; \widehat{\boldsymbol{\Psi}}(\ell_1, \ell_2) = \frac{1}{Y \times X} \sum_{(y,x)} \underline{\mathrm{d}}_{\ell_1}(y, x) \underline{\mathrm{d}}^*_{\ell_2}(y, x), \tag{2.18}$$

where $(\cdot)^*$ stands for the complex conjugate. The reconstruction procedure then consists of recovering $\overline{\boldsymbol{\rho}}(\boldsymbol{r})$ based on the observation $\boldsymbol{d}(\boldsymbol{r})$, the knowledge about the sensitivity matrix $\boldsymbol{S}(\boldsymbol{r})$ and noise assumptions.

2.5.2.3 The SENSE method

In its simplest form, the SENSE reconstruction amounts to solving a one-dimensional inversion problem due to the separability of the Fourier transform. Note however that this inverse problem admits a two-dimensional extension in 3D imaging sequences like Echo Volume Imaging (EVI) [Rabrait et al., 2008; Afacan et al., 2009] where undersampling occurs in two k-space directions. The standard 1D-SENSE reconstruction method [Pruessmann et al., 1999a] is simply the maximum likelihood estimate, which amounts to minimizing a Weighted Least Squares (WLS) criterion from a deterministic viewpoint given the Gaussian noise assumptions. The objective is to find a vector $\widehat{\boldsymbol{\rho}}_{\mathrm{WLS}}(\boldsymbol{r})$ at each spatial location $\mathbf{r}$ such that:

$$\begin{aligned}\widehat{\boldsymbol{\rho}}_{\mathrm{WLS}}(\boldsymbol{r}) &= \arg\min_{\boldsymbol{\rho}(\boldsymbol{r}) \in \mathbb{C}^L} \mathcal{J}_{\mathrm{WLS}}(\boldsymbol{\rho}(\boldsymbol{r})) = \arg\min_{\boldsymbol{\rho}(\boldsymbol{r}) \in \mathbb{C}^L} \| \, \boldsymbol{d}(\boldsymbol{r}) - \boldsymbol{S}(\boldsymbol{r}) \boldsymbol{\rho}(\boldsymbol{r}) \, \|^2_{\boldsymbol{\Psi}^{-1}} \\ &= \left(\boldsymbol{S}^{\mathsf{H}}(\boldsymbol{r}) \boldsymbol{\Psi}^{-1} \boldsymbol{S}(\boldsymbol{r}) \right)^{\sharp} \boldsymbol{S}^{\mathsf{H}}(\boldsymbol{r}) \boldsymbol{\Psi}^{-1} \boldsymbol{d}(\boldsymbol{r})\end{aligned} \tag{2.19}$$

where $(\cdot)^{\mathsf{H}}$ (resp. $(\cdot)^{\sharp}$) stands for the transposed complex conjugate (resp. pseudo-inverse) and, $\| \cdot \|_{\boldsymbol{\Psi}^{-1}} = \sqrt{(\cdot)^{\mathsf{H}} \boldsymbol{\Psi}^{-1} (\cdot)}$ defines a norm on $\mathbb{C}^L$.

In practice, the performance of the 1D-SENSE method is limited because of the presence of:

- distortions in the measurements $\boldsymbol{d}(\boldsymbol{r})$;
- the putative ill-conditioning of $\boldsymbol{S}(\boldsymbol{r})$ specifically at locations $\boldsymbol{r}$ close to the image center where most of the useful information is contained;
- the presence of errors in the estimation of $\boldsymbol{S}(\boldsymbol{r})$ mainly at brain/air interfaces.

These limitations make the considered inverse problem *ill-posed* [Hadamard, 1902] and reconstructed images using the SENSE method generally suffer from severe artifacts. To illustrate these undesirable effects, experiments have been conducted on real data sets comprising $256 \times 256 \times 14$ GRE anatomical and $64 \times 64 \times 30$ fMRI EPI images with respectively $0.93 \times 0.93 \times 8$ (mm^3) and $3.75 \times 3.75 \times 3$ (mm^3) spatial resolution. Note also that these images have been acquired using acceleration factors $R = 2$ and $R = 4$ on a Signa 1.5 Tesla GE Healthcare scanner with an eight-channel head coil. Interestingly, the scanning time of anatomical data lasted 5 min in non-parallel imaging, while acquisition duration was decreased to 3 min 10 s and 2 min 20 s in parallel imaging with $R = 2$ and $R = 4$, respectively.

Figs. 2.18 and 2.19 show aliasing artifacts in the SENSE reconstructed anatomical images (9 different slices) for two values of the reduction factor: $R = 2$ and $R = 4$.

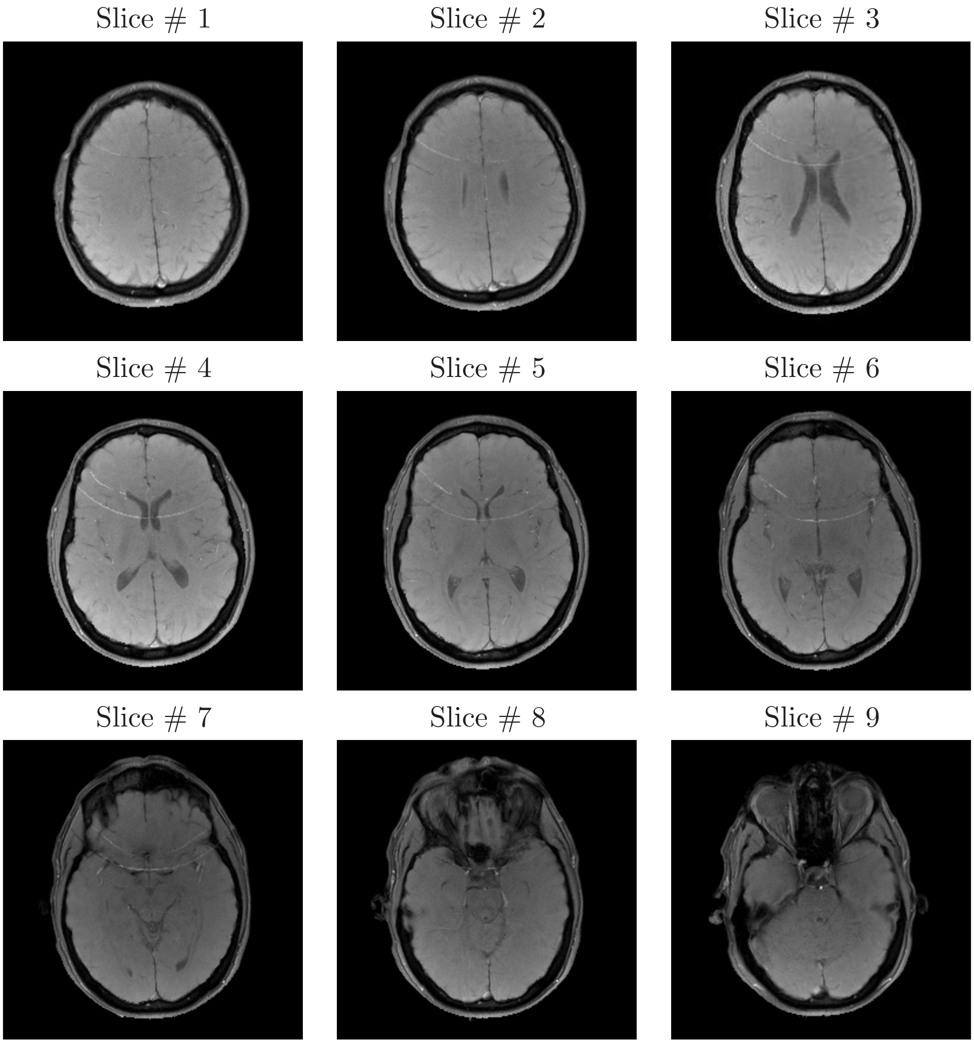


Figure 2.18: GE anatomical data: reconstructed adjacent slices using SENSE for $R = 2$.

The illustrated images show that aliasing artifacts increase with the reduction factor and are generally well spatially localized with either very high or very low intensity levels. The same effects are also observed on functional reconstructed images illustrated in Fig. 2.20.

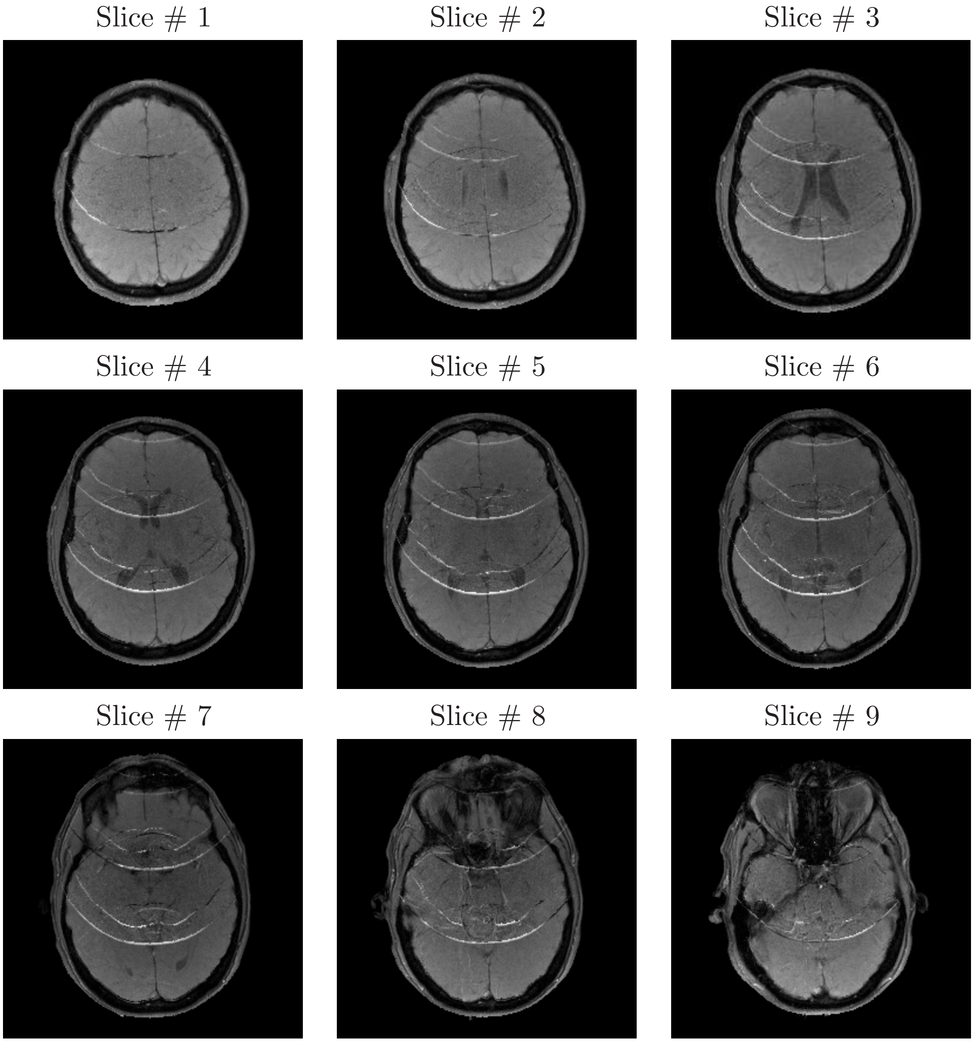


Figure 2.19: GE anatomical data: reconstructed adjacent slices using SENSE for $R = 4$.

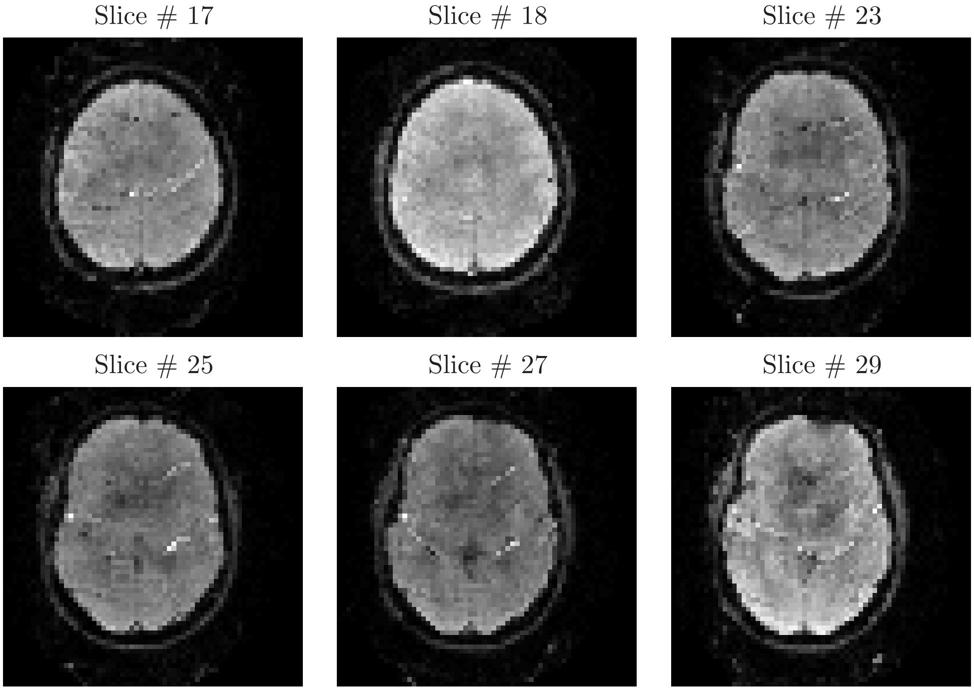


Figure 2.20: Six EPI reconstructed slices using basic SENSE for $R = 4$.

2.6 Conclusion

This chapter presented a brief recall of the MRI principles and basics, and how MRI images are created from acquired signal. It also detailed the main basics of pMRI, as a promizing technique used for fMRI. For pMRI recostruction methods, it mainly focused on the SENSE algorithm which will be investigated in this manuscript, and which do not perform very well under severe experimental conditions. To achieve better SENSE reconstruction results, and since the reonstruction problem is ill-posed, regularization techniques are generally used in such a situation in order to enhance the target solution. The regularization issue is addressed in a general way in Chapter 3 and applied to the pMRI reconstruction problem in Chapter 4.

Chapter 3

Regularization and convex analysis for inverse problems

Contents

3.1 Introduction **57**

3.2 Regularization for inverse problems **58**

3.3 Convex optimization **63**

3.4 Numerical illustrations **72**

3.5 Conclusion **80**

3.1 Introduction

Because of sensor imperfection and acquisition mode, the measured data are often noisy and degraded by a linear operator. These operator and noise properties may actually depend on the application under consideration. However, in many real world applications the involved inverse problem is ill-posed. For instance, in pMRI where the noise is Gaussian, the linear operator is simply the sensitivity matrix whose ill-posedness depends on the acceleration factor regarding the number of channels in the coil.

The general goal of this chapter is to provide the reader with a brief introduction to regularization techniques for ill-posed inverse problems, as well as to describe the inherent optimization challenges and algorithmic issues while proposing solutions to some of the encountered practical difficulties. Although some inverse problems may involve a non-linear operator like in diffraction tomography [Pan, 1998], we will limit ourselves in this chapter to inverse problems where the operator is linear, which is the case in pMRI. We will also restrict our presentation in this chapter to real-valued signals. However, all the definitions and algorithms presented in this chapter can be extended to the complex-valued case like in Chapter 4 where the pMRI reconstruction problem is addressed.

Section 3.2 gives an introduction to ill-posed inverse problems and regularization issues to achieve stable and efficient solutions. We then go through the inherent optimization problems in Section 3.3 and solve some of the practical hurdles encountered when using iterative optimization algorithms. Some numerical illustrations and conclusions are finally given in Sections 3.4 and 3.5, respectively.

3.2 Regularization for inverse problems

3.2.1 Inverse problems

In many real world problems where one is interested in measuring a physical signal, this signal is not directly available, but only measured through some physical laws linking the sought signal to the measurements. The inherent *inverse problem* consists of inverting these physical laws and recovering the target signal. This problem can be formulated as follows. Let $\overline{\boldsymbol{y}} \in \mathbb{R}^N$ be the real-valued signal/image to be recovered from its degraded observation $\boldsymbol{z} \in \mathbb{R}^M$ through a linear operator $H : \mathbb{R}^N \longrightarrow \mathbb{R}^M$. This linear operator simply describes the physical laws linking $\overline{\boldsymbol{y}}$ to the measurement $\boldsymbol{z}$. Moreover, an observation noise $\boldsymbol{n} \in \mathbb{R}^M$ generally affects the measurements. When this noise is additive, the resulting observation model reads:

$$\boldsymbol{z} = H\overline{\boldsymbol{y}} + \boldsymbol{n}. \tag{3.1}$$

Other non-additive noise models may be encountered in applications like Poisson noise in PET [Padfield and Manjeshwar, 2006], Gamma distributed noise in Synthetic Aperture Radar (SAR) [Weibin and Mingyi, 2009] ...
However, the model in Eq. (3.1) is an exact match of the one in pMRI reconstruction since the noise is zero-mean Gaussian and additive [Pruessmann et al., 1999a; Sijbers et al., 2007]. For this reason, this model will be investigated in the rest of this manuscript.
We recall that an inverse problem is called ill-posed [Hadamard, 1902] if one of the following conditions is not met:

i) existence: a solution $\boldsymbol{y}$ exists for each observation $\boldsymbol{z}$;

ii) unicity: this solution is unique;

iii) stability: a small perturbation of the observation induces a small perturbation of the solution.

The inverse problem in Eq. (3.1) can be solved using conventional estimators by minimizing some distance D between the solution and the observation. In this case, the estimation procedure relies on the optimization of the following criterion:

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} D(H\boldsymbol{y}, \boldsymbol{z}). \tag{3.2}$$

When the considered distance is quadratic, which means that the observation noise is Gaussian, the Weighted Least Squares (WLS) estimator is often used in signal/image recovery literature. However, a drawback of the WLS estimation is its lack of stability with respect to imprecise measurements. Indeed, let $\delta\boldsymbol{z}$ be the observation error and $\delta\widehat{\boldsymbol{y}}$ the inherent error affecting the solution $\widehat{\boldsymbol{y}}$. When the weighting matrix in the WLS estimation reduces to the identity, the perturbation study shows that [Golub and Van Loan, 1996, p. 81]:

$$\frac{\| \delta\widehat{\boldsymbol{y}} \|}{\| \widehat{\boldsymbol{y}} \|} \leq c\frac{\| \delta\boldsymbol{z} \|}{\| \boldsymbol{z} \|}, \tag{3.3}$$

where $c \geq 1$ is the *condition number* of the problem, defined as the ratio of largest and smallest singular values of H. As shown in Eq. (3.3), this condition number is directly related to the stability of the solution: the larger it is, the higher the influence of the measurement error $\delta \boldsymbol{z}$ on the solution $\widehat{\boldsymbol{y}}$. The problem is called *ill-conditioned* when $c \gg 1$, which is generally the case in many real world image restoration and reconstruction problems like the pMRI reconstruction one where c can be equal to 9 in the center of the k-space, and can even reach higher values especially in the brain/air interfaces. To summarize, when $M \geq N$, the problem is overdetermined and well-posed, unless H is ill-conditioned. However, when $M < N$, the problem is underdetermined with an infinity of solutions which may match the observation model, and hence it is ill-posed.
To alleviate this problem and get a satisfactory solution, regularization techniques [Tikhonov, 1963] have been widely investigated in the inverse problems literature. For more details about inverse problems and regularization techniques, the interested reader can refer to [Demoment and Idier, 2008a; Demoment and Idier, 2008b].

3.2.2 Regularization

In practice, regularization simply consists of introducing some prior knowledge about the target solution, which is generally user-dependent.
Phylosophically speaking, regularization techniques can be split into two different groups: those which proceed by reducing the search space where a solution is looked for, and those which minimize some penalized criterion, eventually under additional constraints. However, some regularization methods can enter either in the first or the second class. Equivalence between approaches belonging to different classes can even be demonstrated in some cases.

Reducing the search space

This regularization approach consists of controlling the solution dimensions: the solution is decomposed into sub-spaces of reduced dimensions and then reconstructed by eliminating sub-spaces dominated by noise. This may be performed using a Singular Value Decomposition (SVD) [Demoment and Idier, 2001], Fourier [Ciuciu et al., 2001], or wavelet transforms [Mallat, 1998], eliminating high frequency components, and then reconstructing a denoised solution.

Penalized regularization

A stabilization of the target solution is here performed by adding a prior information about it. This additional information may be modelled by a penalization term added to the criterion measuring the closeness of the estimate to the data (likelihood). For instance, a penalized regularization of the inverse problem in Eq. (3.1) is simply achieved by minimizing the following criterion,

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} [D(H\boldsymbol{y}, \boldsymbol{z}) + \alpha g(\boldsymbol{y})], \tag{3.4}$$

which is obtained from Eq. (3.2) by adding the penalty term $g(\cdot)$. The parameter $\alpha > 0$ is called the *regularization parameter*, which balances the solution between the data fidelity term $D(H\cdot, \boldsymbol{z})$ and the penalization $g(\cdot)$. When α tends to zero, the regularized solution tends to the one obtained using the WLS estimator. When $\alpha \longrightarrow \infty$, the solution tends to the one obtained by minimizing only the penalization term, called also the *regularization function*. A probabilistic interpretation of this regularization approach is given in Section 3.2.4. Note that in the rest of this manuscript, we will focus only on the penalized regularization approach.

The regularization penalty may be designed in order to emphasize some features of the signal under investigation like sparsity or local smoothness properties. Prior information about these features can also be supplied whether in the original space or in a transformed one.

In this context, and since the 90's, Wavelet Transforms (WT) [Mallat, 1998] have been widely used in regularization literature [Guerrero-Colon et al., 2008; Pesquet et al., 2009; Vonesch and Unser, 2009] since they allow one to detect local image features and details through the sparse representation they provide. WTs allow one to decompose signals into insightful scale-space elements which are easier to interpret and process. Nevertheless, it has been observed especially in the image processing literature that using overcomplete wavelet representations is generally more advantageous than wavelet bases. In fact, some useful properties cannot be obtained using classical bases transforms. For instance, shift invariance and directionality are of great interest in image processing and allow to separate the information included in the image into different sub-spaces. Each sub-space can then be processed differently depending on the target application. In this context, various redundant wavelet representations have been introduced. The undecimated wavelet transform [Coifman and Donoho, 1995] has been proposed having the fruitful shift-invariance properties, but at the expense of high redundancy. More recently, geometrical transforms have been proposed such as curvelets [Candès et al., 2006], contourlet [Do and Vetterli, 2005], dual-trees [Selesnick et al., 2005; Chaux et al., 2006b] or grouplets [Mallat, 2009], which are often able to achieve the shift-invariance property and perform accurate directional analyses. The next section will give a brief overview of the *frame* concept in general, which is not reduced to wavelet transforms and can involve other linear representations.

3.2.3 The frame concept

Let $\mathcal{H}$ and $\mathcal{G}$ denote two separable Hilbert spaces with scalar product $\langle .,. \rangle$ and norm $\|.\|$. A family of vectors $(\boldsymbol{e}_k)_{k\in\mathbb{I}}$ in $\mathcal{H}$ with $\mathbb{I} \subset \mathbb{Z}$ is called a frame when there exists two constants $\underline{\mu}$ and $\overline{\mu}$ in $]0, +\infty[$ such that

$$(\forall \boldsymbol{y} \in \mathcal{H}), \qquad \underline{\mu}\|\boldsymbol{y}\|^2 \leq \sum_{k\in\mathbb{I}} |\,\langle \boldsymbol{y}|\boldsymbol{e}_k\rangle\,|^2 \leq \overline{\mu}\|\boldsymbol{y}\|^2. \tag{3.5}$$

If $\underline{\mu} = \overline{\mu}$, $(\boldsymbol{e}_k)_{k\in\mathbb{I}}$ is called a *tight* frame. The bounded linear frame analysis operator F and its adjoint synthesis frame operator F^* are defined as

$$\begin{aligned} F\colon \mathcal{H} &\to \mathcal{G} \\ y &\mapsto (\langle \boldsymbol{y} | \boldsymbol{e}_k \rangle)_{k\in\mathbb{I}}, \end{aligned} \tag{3.6}$$

$$\begin{aligned} F^*\colon \mathcal{G} &\to \mathcal{H} \\ (\xi_k)_{k\in\mathbb{I}} &\mapsto \sum_{k\in\mathbb{I}} \xi_k \boldsymbol{e}_k. \end{aligned} \tag{3.7}$$

When F is invertible and $F^{-1} = F^*$, $(\boldsymbol{e}_k)_{k\in\mathbb{I}}$ is an orthonormal basis. A simple example of a redundant frame is the union of $\mu \in \mathbb{N}^*$ orthonormal bases. In this case, the frame is tight with $\mu = \underline{\mu} = \overline{\mu}$ and thus, we have $F^* \circ F = \mu \mathrm{Id}$ where Id is the identity operator. Note that in the rest of this manuscript where the Hilbert spaces $\mathcal{H}$ and $\mathcal{G}$ are finite-dimentional (for example $\mathcal{H} = \mathbb{R}^N$ or $\mathcal{H} = \mathbb{R}^K$), the upper bound $\overline{\mu}$ always exists.

3.2.4 Bayesian approach using frame representations

To deal with the regularization task using wavelet frame representations for example, two strategies can be adopted: regularization with respect to the image (Analysis Approach - AA) or to the wavelet coefficients (Synthesis Approach - SA). Comparisons of these two competing approaches have been investigated in some recent works [Elad et al., 2007b; Carlavan et al., 2009]. The study in [Elad et al., 2007b], where a general inverse problem involving Gaussian noise has been considered, shows that the two approaches are no longer equivalent when overcomplete transforms are used. However, the two approaches are not completely disconnected: depending on the considered problem, it can be easier to express the prior information in the image or transformed space. Duality properties allow one to establish relationships between the two inherent optimization problems (see [Combettes et al., 2010] where a denoising example involving Gaussian noise is investigated).

Under a Bayesian framework, the design of the regularization function is performed through the choice of a prior distribution that models the signal under investigation in a given space. Recovery of the unknown signal is then performed based on its posterior distribution.

To go through this Bayesian formulation in more detail, we will first assume the following.

- the signal to recover $\overline{\boldsymbol{y}}$ is a realization of a random variable $\boldsymbol{Y}$ with prior probability distribution $p(\boldsymbol{y})$,
- the observed signal $\boldsymbol{z}$ is a realization of a random variable $\boldsymbol{Z}$ with conditional likelihood $p(\boldsymbol{z}|\boldsymbol{y})$.

Using a frame representation, and combining the likelihood and the appropriate prior distribution, the posterior probability distribution may be easily derived even using AA or SA.

3.2.4.1 The Analysis Approach

If one adopts AA, the prior distribution will be designed based on a direct transformation of the signal to recover. In this case, the posterior probability distribution of the unknown signal is given by:

$$p(\boldsymbol{y}|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{y})\ p(\boldsymbol{y}), \tag{3.8}$$

where the likelihood and prior probability distributions are given by

$$p(\boldsymbol{z}|\boldsymbol{y}) \propto \exp\left(-D(H\boldsymbol{y}, \boldsymbol{z})\right) \tag{3.9}$$

and

$$p(\boldsymbol{y}) \propto \exp\left(-g(F\boldsymbol{y})\right), \tag{3.10}$$

F being the bounded linear frame analysis operator. In practice, F may be chosen as a WT, Fourier transform, DCT [Ahmed and Natarajen, 1974],...
The prior defines a proper probability density if for example,

$$\forall \boldsymbol{x} \in \mathbb{R}^N, \qquad g(\boldsymbol{x}) \geq \rho\|\boldsymbol{x}\|^p \tag{3.11}$$

where $\rho > 0$ and $p > 0$. We have indeed

$$\int_{\mathbb{R}^N} e^{-g(F\boldsymbol{y})} d\boldsymbol{y} \leq \int_{\mathbb{R}^N} e^{-\rho\|F\boldsymbol{y}\|^p} d\boldsymbol{y} \leq \int_{\mathbb{R}^N} e^{-\rho\underline{\mu}^{p/2}\|\boldsymbol{y}\|^p} d\boldsymbol{y} < \infty. \tag{3.12}$$

The posterior probability in Eq. (3.8) can then be reexpressed as:

$$p(\boldsymbol{y}|\boldsymbol{z}) \propto \exp\left(-D(H\boldsymbol{y}, \boldsymbol{z}) - g(F\boldsymbol{y})\right). \tag{3.13}$$

3.2.4.2 The Synthesis Approach

When dealing with SA, the prior distribution will be designed based on the representation of the signal in a given dictionary such that $\overline{\boldsymbol{y}} = F^*\overline{\boldsymbol{x}}$. In this case, an estimate $\widehat{\boldsymbol{x}}$ of $\overline{\boldsymbol{x}}$ will first be found before getting $\widehat{\boldsymbol{y}} = F^*\widehat{\boldsymbol{x}}$. The posterior probability of the unknown signal is given by:

$$p(\boldsymbol{x}|\boldsymbol{z}) \propto p(\boldsymbol{z}|F^*\boldsymbol{x})\ p(\boldsymbol{x}). \tag{3.14}$$

Assuming the same form for the likelihood and the following shape of the prior

$$p(\boldsymbol{x}) \propto \exp\left(-g(\boldsymbol{x})\right), \tag{3.15}$$

the posterior probability distribution in this case is given by:

$$p(\boldsymbol{x}|\boldsymbol{z}) \propto \exp\left(-D(HF^*\boldsymbol{x}, \boldsymbol{z}) - g(\boldsymbol{x})\right). \tag{3.16}$$

3.2.4.3 Variational formulation of the AA/SA estimation procedure

Based on the posterior probability distribution in each case, the Minimum Mean Square Error (MMSE) (i.e. the Posterior Mean) or the Maximum A Posteriori (MAP) estimators may be used to recover an estimate $\widehat{\boldsymbol{y}}$ of $\overline{\boldsymbol{y}}$. The MMSE can be used only when easy mathematical integration of the posterior distribution can be performed. However, its approximation may be obtained by resorting to Monte Carlo simulation methods. This approach will not be detailed in this chapter, but more details can be found in [Robert and Castella, 2004] and references therein, or by referring to Chapter 5. Our attention will be focused on the MAP estimator since it is somehow more familiar than the PM one. If one wants to use the MAP estimator by maximizing the posterior distribution established with AA and SA, the estimation procedure can be written in the following variational form:

- AA:

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}\in\mathbb{R}^N} [D(H\boldsymbol{y},\boldsymbol{z}) + g(F\boldsymbol{y})], \tag{3.17}$$

- SA:

$$\widehat{\boldsymbol{y}} = F^* \arg\min_{\boldsymbol{x}\in\mathbb{R}^K} [D(HF^*\boldsymbol{x},\boldsymbol{z}) + g(\boldsymbol{x})]. \tag{3.18}$$

Philosophically, AA and SA are quite different. However, when the used frame is an orthonormal basis, the two approaches are strictly equivalent. For more general frames, the equivalence is not always ensured unless for special cases of frames or regularization functions. In fact, Elad et al. [Elad et al., 2007b] has proved the equivalence between AA and SA when a redundant frame is used for the case of Gaussian noise and quadratic regularization functions. However, for other noise classes or sparser penalty terms like an ℓ_1 norm for instance (Laplace prior), the two problems become quite different and their respective resolution may lead to different results. Moreover, it has been reported in the same work that the AA may be more robust to estimation errors in each solution component apart. However, good results in image deconvolution have been obtained using even AA in [Weiss et al., 2008] or SA in [Chaux et al., 2007]. Based on this variational formulation, the regularization process reduces to the optimization problem in Eqs. (3.17) and (3.18) for AA and SA, respectively. The next section addresses the optimization issue inherent to the regularization process.

3.3 Convex optimization

For the rest of this chapter, we will denote by $\Gamma_0(\mathcal{H})$ (resp. $\Gamma_0(\mathcal{G})$) the class of lower semicontinuous convex functions from $\mathcal{H}$ (resp. $\mathcal{G}$) to $]-\infty,+\infty]$ which are not identically equal to $+\infty$. We will also recall the concept of *proximity* operator [Moreau, 1962; Moreau, 1965] as follows.

Definition 3.3.1 *[Moreau, 1965]*
For a function $\phi \in \Gamma_0(\mathcal{H})$, the proximity operator is such that:

$$(\forall \xi \in \mathcal{H}), \qquad \operatorname{prox}_{\phi}(\xi) = \arg\min_{\zeta\in\mathcal{H}} \phi(\zeta) + \frac{1}{2}\|\zeta - \xi\|^2. \tag{3.19}$$

In recent convex optimization literature, many useful properties of the proximity operator are provided in recent works like [Combettes and Wajs, 2005; Chaux et al., 2007; Combettes and Pesquet, 2010].

Note also that for a function $\phi : \mathbb{R} \to \mathbb{R}$, the convex conjugate [Rockafellar, 1970] of

$\xi^2/2 + \phi(\xi)$ reads simply as $\xi^2/2 - \operatorname{prox}_\phi(\xi)$. This convex conjugate function has been used in some recent works for half-quadratic optimization [Ciuciu and Idier, 2002].
For illustration purposes, we give here an example of a convex function and the associated proximity operator.

Example 3.3.1

Let ϕ be the function defined by:

$$\begin{aligned} \phi \,:\, \mathbb{R} &\to \mathbb{R} \\ \xi &\mapsto \frac{|\xi|^\beta}{\alpha} \end{aligned} \tag{3.20}$$

where $\beta \in [1, \infty[$ and $\alpha \in \mathbb{R}_+^*$ are the shape and scale parameters, respectively. The associated proximity operator writes:

- if $\beta = 1$, the considered function is simply the anti-logarithm of a Laplacian distribution and we have $\operatorname{prox}_\phi(\xi) = \operatorname{soft}_{1/\alpha}(\xi)$ where $\operatorname{soft}_{1/\alpha}$ is simply the soft thresholding operator defined by:

$$\forall \xi \in \mathbb{R},\ \operatorname{soft}_\tau(\xi) = \operatorname{sign}(\xi) \max\{|\xi| - 1/\alpha, 0\} \tag{3.21}$$

- if $\beta \neq 1$, this function corresponds to the anti-logarithm of a Generalized Gaussian (GG) distribution with $\operatorname{prox}_\phi(\xi) = \operatorname{sign}(\xi)\eta$, where η is the unique solution in $[0, +\infty[$ to $\eta + \beta\eta^{\beta-1}/\alpha = |\xi|$.

Fig. 3.1 illustrates the curve of the function ϕ defined in Eq. (3.20) with different values of the shape and scale parameters. Fig. 3.2 displays the associated proximity operator plots for the same values of the shape and scale parameters. It can be easily noticed through these figures that in fact, for $\beta = 1$ the proximity operator behaves as a soft thresholder. This is actually a nice property in image/signal restoration problems which is not achieved for a wide variety of penalty terms like quadratic or half-quadratic ones [Geman and Yang, 1995; Ciuciu et al., 2001; Ciuciu and Idier, 2004].

The next section describes some optimization schemes from the convex optimization literature which may be used for the optimization of problems in Eqs. (3.17)-(3.18).

3.3.1 Some convex optimization algorithms

In the rest of this section, we will focus on the case where the optimality criteria in Eqs. (3.17)-(3.18) are convex, which is generally inherent to the choice of the regularization function g, provided that the data fidelity term is also convex. Generally speaking, convex penalizations induce easier optimization problems than non convex ones since every local minimizer is global. However, even within the class of convex penaltizations, smooth penalties [Phillips, 1962; Twomey, 1963; Geman and Reynolds, 1992] (such as the Tikhonov one) have been the most frequently used in regularization literature for a long time, since no efficient algorithms allowing the minimization of non-smooth criteria have

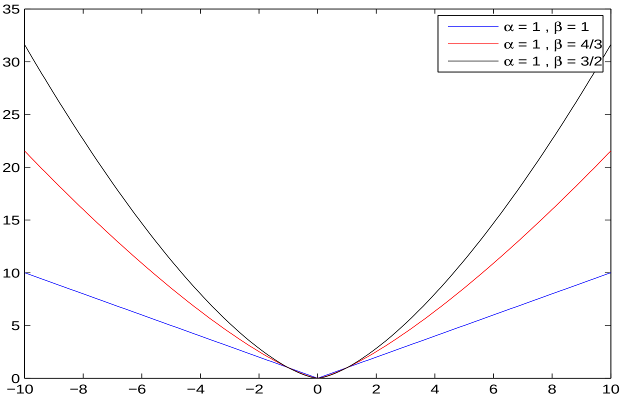


Figure 3.1: Curves of the convex function defined in Eq. (3.20).

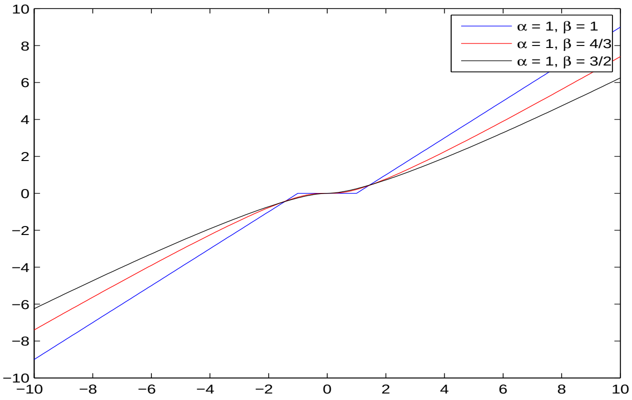


Figure 3.2: Curves of the proximity operator associated with the functions plotted in Fig. 3.1.

been available. Due to some advances in the convex optimization literature, non-smooth penalties (such as half-quadratic or ℓ_1 functions) started to be investigated [Geman and Reynolds, 1992; Geman and Yang, 1995]. This emergence in the image/signal literature has been catalyzed by the emergence of optimization algorithms having as an idea to optimize an approximation of the non-smooth criterion instead of the criterion itself [Nashed and Scherzer, 1997; Ciuciu et al., 2001; Lustig et al., 2007; Carlavan et al., 2009]. Using more general non-smooth penalizations has been made possible due to recent advances in convex optimization due to algorithms dealing directly with non-smooth criteria. Among the recent efficient optimization algorithms dealing with convex but non-smooth criteria, we can cite the Forward-Backward (FB) algorithm [Daubechies et al., 2004; Chaux et al., 2007], Fast Iterative Soft Thresholding Algorithm (FISTA) [Beck and Teboulle, 2009] or Parallel ProXimal Algorithm (PPXA) [Combettes and Pesquet, 2008].
Except PPXA which is designed to minimize the sum of $J \in \mathbb{N}^*$ convex functions, FB and FISTA can only deal with criteria involving two convex functions. However, the common

property of these algorithms is that they are iterative schemes and their convergence is guaranteed under appropriate choice of their parameters. Their structure may be summarized in a unique diagram involving two main steps: parameters setting and update. This diagram is illustrated in Fig. 3.3.

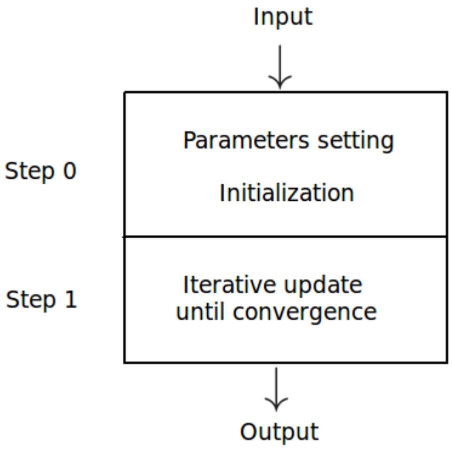


Figure 3.3: Summary diagram of iterative convex optimization algorithms.

We consider here that a good iterative optimization algorithm should satisty the following conditions:

1) **Convergence**: an algorithm which is not guaranteed to converge cannot be relied on, especially for applications which are supposed to be user-independent and where the final results have to be trusted.

2) **Stability**: one has to pay attention to numerical errors which usually occur when evaluating some functions, variables ..., which may be nasty especially in iterative schemes where the error may be increased in each iteration and lead to poor results.

3) **Efficiency**: a trade-off between convergence speed, computational time and requested computer resources has to be made.

4) **Simplicity**: the complexity in terms of implementation and parameter settings should be taken into account.

In what follows, our attention will be paid to convex optimization algorithms which are widely used in literature and referenced as the fastest available ones that satisfy the four previously cited conditions. We will namely give a quick overview of the FB, FISTA and PPXA algorithms.

3.3.1.1 Forward-Backward algorithm

Let us first define the function $f \in \Gamma_0(\mathbb{R}^N)$ from D (the data fidelity term) as follows:

$$\forall \boldsymbol{y} \in \mathbb{R}^N, \quad f(\boldsymbol{y}) = D(H\boldsymbol{y}, \boldsymbol{z}). \tag{3.22}$$

In the following, the functions $D(\cdot,\boldsymbol{z})$ and $g(\cdot)$ are assumed to be in $\Gamma_0(\mathbb{R}^M)$ and $\Gamma_0(\mathbb{R}^K)$, respectively. When one of these functions is differentiable with a Lipschitz gradient, the FB algorithm can be used. The FB iterations are given by Algorithm 1 when $D(\cdot,\boldsymbol{z})$ is assumed to be differentiable with α-Lipschitz gradient. The convergence [Combettes and Wajs, 2005; Chaux et al., 2007] of $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ (resp. $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$) to a solution of Problem (3.17) (resp. Problem (3.18)) is ensured when $\inf_{n\in\mathbb{N}}\lambda_n > 0$, $\sum_{n=0}^{+\infty}\|\boldsymbol{a}_n\| < +\infty$ and $\sum_{n=0}^{+\infty}\|\boldsymbol{b}_n\| < +\infty$. The sequences $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$ correspond to some error tolerances in the evaluation of the gradient and the proximity operator, respectively.

Algorithm 1 The FB algorithm

AA

```
• Select $\boldsymbol{y}_0 \in \mathbb{R}^N$
for $n \in \mathbb{N}^*$ do
    • Set $\gamma_n \in ]0, 2/(\alpha\,\|H\|^2)[$ and $\lambda_n \in ]0,1]$
    • $\boldsymbol{y}_{n+\frac{1}{2}} = \boldsymbol{y}_n - \gamma_n\Big(\nabla f(\boldsymbol{y}_n) + \boldsymbol{b}_n\Big)$
    • $\boldsymbol{p}_n = \mathrm{prox}_{\gamma_n g\circ F}(\boldsymbol{y}_{n+\frac{1}{2}})$
    • $\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n(\boldsymbol{p}_n + \boldsymbol{a}_n - \boldsymbol{y}_n)$
end for
• After convergence, take $\widehat{\boldsymbol{y}} = \boldsymbol{y}_{n+1}$.
```

SA

```
• Select $\boldsymbol{x}_0 \in \mathbb{R}^K$
for $n \in \mathbb{N}^*$ do
    • Set $\gamma_n \in ]0, 2/(\alpha\,\|HF^*\|^2)[$ and $\lambda_n \in ]0,1]$
    • $\boldsymbol{x}_{n+\frac{1}{2}} = \boldsymbol{x}_n - \gamma_n\Big(F\nabla f(F^*\boldsymbol{x}_n) + \boldsymbol{b}_n\Big)$
    • $\boldsymbol{p}_n = \mathrm{prox}_{\gamma_n g}(\boldsymbol{x}_{n+\frac{1}{2}})$
    • $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n(\boldsymbol{p}_n + \boldsymbol{a}_n - \boldsymbol{x}_n)$
end for
• After convergence, take $\widehat{\boldsymbol{y}} = F^*\boldsymbol{x}_{n+1}$.
```

3.3.1.2 Fast Iterative Soft Thresholding Algorithm

The FISTA algorithm is also adapted to the case when one of the two functions is differentiable with Lipschitz gradient. We will therefore keep using the same roles for the functions $f(\cdot)$ and $g(\cdot)$ as for the FB algorithm. The FISTA iterations are summarized by Algorithm 2 when $D(\cdot,\boldsymbol{z})$ is assumed to be differentiable with α-Lipschitz gradient. The

Algorithm 2 The FISTA algorithm

AA

```
• Select $\boldsymbol{y}_0 \in \mathbb{R}^N$, $t_1 = 1$ and buid the sequence $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ as follows
for $n \in \mathbb{N}^*$ do
    • $\boldsymbol{y}_{n+\frac{1}{2}} = \boldsymbol{y}_n - \frac{1}{\alpha\|H\|^2}\nabla f(\boldsymbol{y}_n)$
    • $\boldsymbol{p}_n = \mathrm{prox}_{g\circ F}\boldsymbol{y}_{n+\frac{1}{2}}$
    • $t_{n+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$
    • $\boldsymbol{y}_{n+1} = \boldsymbol{p}_n + \frac{t_k-1}{t_{k+1}}(\boldsymbol{p}_n - \boldsymbol{p}_{n-1})$
end for
• After convergence, take $\widehat{\boldsymbol{y}} = \boldsymbol{y}_{n+1}$.
```

SA

```
• Select $\boldsymbol{x}_0 \in \mathbb{R}^K$, $t_1 = 1$ and buid the sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ as follows
for $n \in \mathbb{N}^*$ do
    • $\boldsymbol{x}_{n+\frac{1}{2}} = \boldsymbol{x}_n - \frac{1}{\alpha\|HF^*\|^2}F\nabla f(F^*\boldsymbol{x}_n)$
    • $\boldsymbol{p}_n = \mathrm{prox}_g(\boldsymbol{x}_{n+\frac{1}{2}})$
    • $t_{n+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$
    • $\boldsymbol{x}_{n+1} = \boldsymbol{p}_n + \frac{t_k-1}{t_{k+1}}(\boldsymbol{p}_n - \boldsymbol{p}_{n-1})$
end for
• After convergence, take $\widehat{\boldsymbol{y}} = F^*\boldsymbol{x}_{n+1}$.
```

convergence to a solution to Problem (3.17) (resp. Problem (3.18)) of the sequence $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ (resp. $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$) generated by Algorithm 2 is no longer guaranteed. However, the convergence of the objective values $(D(H\boldsymbol{y}_n, \boldsymbol{z}) + g(F\boldsymbol{y}_n))_{n\in\mathbb{N}}$ (resp. $(D(HF^*\boldsymbol{x}_n, \boldsymbol{z}) + g(\boldsymbol{x}_n))_{n\in\mathbb{N}}$) is ensured [Beck and Teboulle, 2009].

3.3.1.3 Parallel ProXimal Algorithm

We will use here the same function $f(\cdot)$ as defined in Eq. (3.22). With respect to the FB context, we will relax the assumption that one of the two functions ($f(\cdot)$ or $g(\cdot)$) is differentiable with Lipschitz gradient. The involved functions, for which we have to calculate the proximity operator, have just to be convex. In fact, the PPXA allows the minimization of criteria involving $J \in \mathbb{N}^*$ convex functions, provided that the computation of the proximity operator of each of them is possible.
The PPXA iterations minimize the sum of J convex functions ($\sum_{j=1}^{J} f_j$) as explained in Algorithm 3. However, a solution to Problem (3.17) (resp. (3.18)) can be obtained by setting $J = 2$ and choosing $f_1 = f$ and $f_2 = g \circ F$ (resp. $f_1 = f \circ F^*$ and $f_2 = g$). The sequences $(\boldsymbol{a}_{j,n})_{n\in\mathbb{N}}$ in $\mathbb{R}^N$ or $\mathbb{R}^K$ for $j \in \{1,\dots,J\}$ introduced in Algorithm 3 correspond to possible errors (numerical errors for instance) in the computation of the proximity operators, which show that convergence is assured in spite of these errors.

Algorithm 3 The PPXA algorithm

AA

```
• Let $\gamma \in ]0,+\infty[$
• Set $(\omega_j)_{1\le j\le J} \in ]0,1]^J$ such that $\sum_{j=1}^{J} \omega_j = 1$
• Set $(\boldsymbol{u}_{j,0})_{1\le j\le J} \in (\mathbb{R}^N)^J$ and $\boldsymbol{y}_0 = \sum_{j=1}^{J} \omega_j \boldsymbol{u}_{j,0}$
for $n \in \mathbb{N}$ do
    • for $j \in \{1,\dots,J\}$, calculate
    $\boldsymbol{p}_{j,n} = \mathrm{prox}_{\gamma/\omega_j f_j} \boldsymbol{u}_{j,n} + \boldsymbol{a}_{j,n}$
    • $\boldsymbol{p}_n = \sum_{j=1}^{J} \omega_j \boldsymbol{p}_{j,n}$
    • $\lambda_n \in ]0,2[$
    • for $j \in \{1,\dots,J\}$, calculate
    $\boldsymbol{u}_{j,n+1} = \boldsymbol{u}_{j,n} + \lambda_n\,(2\,\boldsymbol{p}_n - \boldsymbol{y}_n - \boldsymbol{p}_{j,n})$

    • $\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n(\boldsymbol{p}_n - \boldsymbol{y}_n)$
end for
• After convergence, take $\widehat{\boldsymbol{y}} = \boldsymbol{y}_{n+1}$.
```

SA

```
• Let $\gamma \in ]0,+\infty[$
• Set $(\omega_j)_{1\le j\le J} \in ]0,1]^J$ such that $\sum_{j=1}^{J} \omega_j = 1$
• Set $(\boldsymbol{u}_{j,0})_{1\le j\le J} \in (\mathbb{R}^K)^J$ and $\boldsymbol{x}_0 = \sum_{j=1}^{J} \omega_j \boldsymbol{u}_{j,0}$
for $n \in \mathbb{N}$ do
    • for $j \in \{1,\dots,J\}$, calculate
    $\boldsymbol{p}_{j,n} = \mathrm{prox}_{\gamma/\omega_j f_j} \boldsymbol{u}_{j,n} + \boldsymbol{a}_{j,n}$
    • $\boldsymbol{p}_n = \sum_{j=1}^{J} \omega_j \boldsymbol{p}_{j,n}$
    • $\lambda_n \in ]0,2[$
    • for $j \in \{1,\dots,J\}$, calculate
    $\boldsymbol{u}_{j,n+1} = \boldsymbol{u}_{j,n} + \lambda_n\,(2\,\boldsymbol{p}_n - \boldsymbol{x}_n - \boldsymbol{p}_{j,n})$

    • $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n(\boldsymbol{p}_n - \boldsymbol{x}_n)$
end for
• After convergence, take $\widehat{\boldsymbol{y}} = F^*\boldsymbol{x}_{n+1}$.
```

The sequence $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ (resp. $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$) generated by Algorithm 3 converges to a solution to Problem (3.17) (resp. Problem (3.18)) under the following assumption [Combettes and Pesquet, 2008]:

Assumption 3.3.2

(i) $\lim_{\|\boldsymbol{y}\|\to+\infty} f_1(\boldsymbol{y}) + \ldots + f_J(\boldsymbol{y}) = +\infty$ (resp. $\lim_{\|\boldsymbol{x}\|\to+\infty} f_1(\boldsymbol{x}) + \ldots + f_J(\boldsymbol{x}) = +\infty$).

(ii) $\mathcal{H}$ is finite-dimensional and $\cap_{j=1}^{J} \operatorname{rint} \operatorname{dom} f_j \neq \varnothing$.[1]

(iii) $\sum_{n \in \mathbb{N}} \lambda_n (2 - \lambda_n) = +\infty$.

(iv) $(\forall j \in \{1, \ldots, J\}) \quad \sum_{n \in \mathbb{N}} \lambda_n \|\boldsymbol{a}_{j,n}\| < +\infty$.

3.3.2 For those who see life in pink

Actually, using such iterative optimization algorithms is not completely costless. Indeed, by considering the proximal algorithms described in the previous section to solve the minimization problems (3.17) and (3.18), some hurdles can actually be encountered. These difficulties may be encountered whether in Step 1 or Step 2 of the diagram in Fig. 3.3. For example, the convergence rate of the FB algorithm [Chaux et al., 2007] when used to solve Problem (3.18) depends on the upper frame constant $\overline{\mu}$. For other optimization schemes such as PPXA [Combettes and Pesquet, 2008], the difficulty may stem from the computation of the proximity operator associated with $\varphi \circ T$ where $\varphi \in \Gamma_0(\mathcal{H})$ and $T : \mathcal{G} \to \mathcal{H}$ is a bounded linear operator.
In this section we will focus on these two main difficulties:

i) calculating the proximity operator associated with the composition of a function and a linear operator;

ii) calculating the frame constant.

3.3.2.1 Proximity operator and frame representation

As reported in the previous section, using FB, FISTA or PPXA may require the calculation of the proximity operator associated with the composition of a convex function and a bounded linear operator. To perform this calculation, three methods are described: explicit form computation, splitting approach and iterative approach.

Explicit form

The described method here is completely based on the proposition 3.3.3 introduced in [Combettes and Pesquet, 2007] and is restricted to the case where the linear operator T satisfies a specific property.

Proposition 3.3.3 *Let $\mathcal{G}$ be a real Hilbert space, let $\varphi \in \Gamma_0(\mathcal{G})$, and let $T\colon \mathcal{H} \to \mathcal{G}$ be a bounded linear operator. Suppose that the composition of T and T^* satisfies $T \circ T^* = \chi \mathrm{Id}$, for some $\chi \in]0, +\infty[$. Then $\varphi \circ T \in \Gamma_0(\mathcal{H})$ and*

$$\operatorname{prox}_{\varphi \circ T} = \mathrm{Id} + \frac{T^*}{\chi} \circ (\operatorname{prox}_{\chi \varphi} - \mathrm{Id}) \circ T. \tag{3.23}$$

For instance, when a denoising problem ($H = \mathrm{Id}$) is considered with SA and the PPXA algorithm is used for the optimization task, we mainly need to calculate the proximity

---

[1] The relative interior of a set S of $\mathcal{H}$ is designated by $\operatorname{rint} S$ and the domain of a function $f : \mathcal{H} \to]-\infty; +\infty]$ is $\operatorname{dom} f = \{\xi \in \mathcal{H} | f(\xi) < +\infty\}$.

operator $\operatorname{prox}_{D(.,z)\circ F^*}$. This calculation may be performed using Proposition 3.3.3 in the restrictive framework of a tight frame ($F^* \circ F = \overline{\mu}\mathrm{Id}$) with $T = F^*$ and $\chi = \overline{\mu}$.

Splitting approach [Pustelnik et al., 2010]
When the function $\varphi \in \Gamma_0(\mathbb{R}^N)$ is separable, which means that

$$(\forall \xi = (\xi^{(n)})_{1\leq n\leq N}), \quad \varphi(\xi) = \sum_{n=1}^{N} \varphi_n(\xi^{(n)}), \tag{3.24}$$

the splitting approach consists of decomposing the operator T in operators $(T_i)_{1\leq i\leq I}$ such that $T_i \circ T_i^* = \chi_i \mathrm{Id}$. We subsequently assume that $(\mathbb{I}_i)_{1\leq i\leq I}$ is a partition of $\{1,\dots,N\}$ in nonempty sets. For every $i \in \{1,\dots,I\}$, let N_i be the cardinality of $\mathbb{I}_i$ and let

$$\Upsilon_i \; : \mathbb{R}^{N_i} \to \,]0,+\infty[\tag{3.25}$$

$$(\xi^{(n)})_{n\in\mathbb{I}_i} \mapsto \sum_{n\in\mathbb{I}_i} \varphi_n(\xi^{(n)}). \tag{3.26}$$

Assume that $\varphi \circ T = \sum_{i=1}^{I} \Upsilon_i \circ T_i$ where T_i is the linear operator from $\mathcal{G}$ to $\mathbb{R}^{N_i}$ associated with the matrix

$$\begin{bmatrix} t_{n_1}^\top \\ \vdots \\ t_{n_{N_i}}^\top \end{bmatrix}$$

with $\mathbb{I}_i = \{n_1,\dots,n_{N_i}\}$. The following assumption plays a prominent role in the splitting approach.

Assumption 3.3.4 For all $i \in \{1,\dots,I\}$, $(t_n)_{n\in\mathbb{I}_i}$ is an orthogonal family having the same norm $\chi_i^{1/2}$ where $\chi_i > 0$.

Consider AA when the frame representation is associated with a union of $I = \mu$ orthogonal bases and the regularization function g is separable. The minimization problem (3.17) can be rewritten as

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} D(H\boldsymbol{y},\boldsymbol{z}) + \sum_{i=1}^{\mu} g_i(F_i\boldsymbol{y}), \tag{3.27}$$

where for every $i \in \{1,\dots,I\}$, $F_i^* \circ F_i = \mathrm{Id}$ and $g_i \in \Gamma_0(\mathbb{R}^{K/I})$ (assuming that K/I and μ are integers). By considering this form of Problem (3.17), we are able to compute the proximity operators associated with each of the $\operatorname{prox}_{g_i\circ F_i}$ by (3.23). Algorithm 3 can thus be applied to find the optimal solution to Problem (3.17).

Iterative approach
Unfortunately, if the used frame is not tight, the two previous methods to calculate the proximity operator $g \circ F$ cannot be used. However, although deriving a closed form for this proximity operator in this case is not obvious, it can be iteratively calculated using the duality principle [Zălinescu, 2002]. By using the definition of the proximity operator

given in (3.19), the primal problem consists of finding

$$\min_{\boldsymbol{p}\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{p}\|^2 + g(F\boldsymbol{p}) \tag{3.28}$$

and the associated dual problem is then

$$\min_{\boldsymbol{v}\in\mathbb{R}^K} \frac{1}{2}\|\boldsymbol{y}-F^*\boldsymbol{v}\|^2 + g^*(\boldsymbol{v}) \tag{3.29}$$

where $g^* \in \Gamma_0(\mathbb{R}^K)$ is the conjugate of g.[2] This formulation is a particular case of the one proposed in [Combettes et al., 2010] and can be solved using the Splitting Dual-Primal algorithm [Combettes et al., 2010]. In Algorithm 4, under the assumption that $\sum_{n\in\mathbb{N}} \|\boldsymbol{a}_n\| \leq +\infty$ and $\sum_{n\in\mathbb{N}} \|\boldsymbol{b}_n\| \leq +\infty$, the sequence $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ converges to a solution $\widehat{\boldsymbol{v}}$ of the dual Problem (3.29) and $\operatorname{prox}_{g\circ F}(\boldsymbol{y}) = \boldsymbol{y} - F^*\widehat{\boldsymbol{v}}$.

Algorithm 4 Splitting Dual-Primal algorithm to compute $\operatorname{prox}_{g\circ F}(\boldsymbol{y})$

```
Select $\epsilon \in ]0, \min\{1, 1/\overline{\mu}\}[$
Set $\boldsymbol{v}_0 \in \mathbb{R}^K$
for $n \in \mathbb{N}$ do
    $\boldsymbol{y}_n = \boldsymbol{y} - F^*\boldsymbol{v}_n + \boldsymbol{b}_n$
    $\gamma_n \in [\epsilon, 2/\overline{\mu} - \epsilon]$
    $\lambda_n \in [\epsilon, 1]$
    $\boldsymbol{v}_{n+1} = \boldsymbol{v}_n + \lambda_n\big(\gamma_n F\boldsymbol{y}_n - \gamma_n \operatorname{prox}_{\frac{1}{\gamma_n}g}(\frac{\boldsymbol{v}_n}{\gamma_n} + F\boldsymbol{y}_n) + \boldsymbol{a}_n\big)$
end for
After convergence, take $\widehat{\boldsymbol{v}} = \boldsymbol{v}_{n+1}$.
```

3.3.2.2 Frame constant

Another issue related to proximal algorithms when used to solve the SA optimization problem consists of computing the norm $\|HF^*\|^2$ (see Section 3.3.1). The problem is clearly present in Algorithms 1 and 2 (SA) where the step-size γ_n is inversely proportional to this norm. If the norms of the linear operator $\|H\|$ and the frame operator $\|F\|$ can be calculated, the norm $\|HF^*\|^2$ can be approximated by its upper bound $\|H\|^2\|F\|^2$. However, despite its simplicity, this solution remains sub-optimal and the inherent step-size γ_n will not lead to the fastest convergence rate. Since the norm $\|HF^*\|^2$ is usually not easy to calculate, we propose herebelow an algorithm which converges to this value. Consider the objective function (3.18) for which $h = D(., \boldsymbol{z})$ is α-Lipschitz differentiable. By definition of Lipschitz differentiability, we have

$$(\forall(\boldsymbol{y}_1,\boldsymbol{y}_2)\in(\mathbb{R}^M)^2), \qquad \|\nabla h(\boldsymbol{y}_1) - \nabla h(\boldsymbol{y}_2)\| \leq \alpha\|\boldsymbol{y}_1 - \boldsymbol{y}_2\| \tag{3.30}$$

When using the FB or the FISTA algorithms to solve the SA optimization Problem (3.18), the gradient of the function $f_S = f\circ F^* = h\circ H\circ F^* = D(H\circ F^*., \boldsymbol{z})$ has to be calculated

---

[2] We recall that the conjugate of a function $h : \mathbb{R}^K \to \mathbb{R}$ is the function $h^* : \mathbb{R}^K \to \mathbb{R}$ defined by $h^*(\xi) = \sup_{\boldsymbol{x}\in\mathbb{R}^K}\{\langle\xi|\boldsymbol{x}\rangle - h(\boldsymbol{x})\}$ (see [Rockafellar, 1970] for details).

and is defined as

$$(\forall \boldsymbol{x} \in (\mathbb{R}^K)), \qquad \nabla f_{\mathrm{S}}(\boldsymbol{x}) = FH^*(\nabla h(HF^*\boldsymbol{x})), \tag{3.31}$$

which yields to

$$(\forall (\boldsymbol{x}_1, \boldsymbol{x}_2) \in (\mathbb{R}^K)^2) \qquad \|\nabla f_{\mathrm{S}}(\boldsymbol{x}_1) - \nabla f_{\mathrm{S}}(\boldsymbol{x}_2)\| \le \alpha \|HF^*\|^2 \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|. \tag{3.32}$$

Consequently f_{S} is $\alpha\|HF^*\|^2$-Lipschitz differentiable. Let $B = HF^*$ and perform the eigenvalue decomposition $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ of the matrix associated with the positive semi-definite operator B^*B, where $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \dots, \lambda_K\}$ and $\boldsymbol{U} = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_K] \in \mathbb{R}^{K\times K}$ is an orthogonal matrix. Moreover, $\|HF^*\|^2 = \lambda_{i_0}$ where $i_0 \in \mathrm{Arg}\ \max_{1\le \mathrm{i} \le \mathrm{K}} \lambda_{\mathrm{i}}$. Thus, let $\boldsymbol{x}_0 \in \mathbb{R}^K$ which does not belong to an eigenspace of B^*B, we can write,

$$\frac{\|B^n\boldsymbol{x}_0\|^2}{\|B^{n-1}\boldsymbol{x}_0\|^2} = \frac{\sum_{i=1}^K \lambda_i^n |\langle \boldsymbol{x}_0, \boldsymbol{u}_i\rangle|^2}{\sum_{i=1}^K \lambda_i^{n-1} |\langle \boldsymbol{x}_0, \boldsymbol{u}_i\rangle|^2}, \tag{3.33}$$

which yields to

$$\lim_{n\to+\infty} \frac{\|B^n\boldsymbol{x}_0\|^2}{\|B^{n-1}\boldsymbol{x}_0\|^2} = \lambda_{i_0} = \|HF^*\|^2. \tag{3.34}$$

The iterations of this algorithm are illustrated in Algorithm 5.

Algorithm 5 Frame constant computation

```
Select randomly $\boldsymbol{x}_0 \in \mathbb{R}^K$, set $\rho_0 = \epsilon + 1$, $n = 1$ and $\rho_n = 1$
while $\frac{|\rho_n - \rho_{n-1}|}{\rho_n} \le \epsilon$ do
    Set $\boldsymbol{x}_n = B^*B\boldsymbol{x}_{n-1}$ where $B = HF^*$
    Set $\rho_n = \frac{\|\boldsymbol{x}_n\|}{\|\boldsymbol{x}_{n-1}\|}$
end while
After convergence, take $\|HF^*\|^2 = \rho_n$
```

3.4 Numerical illustrations

In this section, a deconvolution problem is addressed with uniform blur H of size 5×5 and additive Gaussian noise of standard deviation $\sigma = 6$.

3.4.1 Comparison of the AA and SA performance

The AA and SA performance in image deblurring are compared here using in each case the appropriate tools described throughout this chapter to solve the inherent optimization problems. Four different redundant wavelet representations have been used to show how the results depend on the used frame. The used wavelet frames are:

1) translation invariant wavelet (TIW) [Coifman and Donoho, 1995];

2) GenLOT [De Queiroz et al., 1996; Gauthier et al., 2009] transform;

3) union of two orthonormal bases (U2OB);

4) contourlet transform [Do and Vetterli, 2005].

Since the blur operator H is poorly conditioned, the considered deblurring problem is necessarely ill-posed. A regularization strategy should then be used to enhance the target solution. Although many choices for the regularization function like Markovian edge-preserving priors [Ciuciu et al., 2001] can be made, a simple ℓ_1 regularization will be used as a penalty term, which means that a Laplace iid distribution will be retained as the prior to model the wavelet frame coefficients.
Based on the noise model and prior distribution, the optimality criteria in (3.17) and (3.18) may be reexpressed as:

- AA:

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}}[\frac{1}{2\sigma^2}\|H\boldsymbol{y}-\boldsymbol{z}\|^2+\kappa\|F\boldsymbol{y}\|_1]. \tag{3.35}$$

- SA:

$$\widehat{\boldsymbol{y}} = F^*\arg\min_{\boldsymbol{x}}[\frac{1}{2\sigma^2}\|HF^*\boldsymbol{x}-\boldsymbol{z}\|^2+\kappa\|\boldsymbol{x}\|_1]. \tag{3.36}$$

To minimize this criterion we used the FB algorithm (see Algorithm 1 for AA and SA) with $f = \dfrac{1}{2\sigma^2}\|H\cdot-\boldsymbol{z}\|^2$ and $g = \|\cdot\|_1$. The main difficulty when dealing with AA is to compute the proximity operator of $g \circ F$. This calculation was performed as explained in Section 3.3.2.1 using the iterative approach applicable for all the used frames. On the other hand, when SA is adopted, the main difficulty is to evaluate the Lipschitz constant of the gradient of $f \circ F^*$. In fact, since the gradient of f is $\dfrac{1}{\sigma^2}$-Lipschitz, $f_{\rm S}$ is then $\dfrac{1}{\sigma^2}\|HF^*\|^2$-Lipschitz differentiable. The difficulty is therefore to compute $\|HF^*\|^2$, which may be achieved using Algorithm 5 introduced in Section 3.3.2.2.
The comparison in this section is mainly based on the performance of AA and SA in terms of visual restoration quality, SNR and Structural SIMilarity (SSIM) [Wang et al., 2004]. Tests are conducted on four standard 256×256 images with different contents (smooth regions, contours, contract...): *cameraman*, *barbara*, *boat* and *straw*.
For illustration purpose, restoration results of AA and SA using the contourlet transform are visually compared. Figs. 3.4, 3.5, 3.6 and 3.7 show the reference images (a), the degraded (b) and restored ones using AA (c) and SA (d) for the camerama, barbara, boat and straw images, respectively. We notice that for the four test images, the restored images using AA and SA are very similar from a visual viewpoint.

As regards SNR and SSIM, quantitative results are given in Table 3.1. It is worth noticing through Table 3.1 that no clear preference can be attributed to one of the two competing approaches. The quantitative performance varies from an image to another and from a frame to another.

Figure 3.4: Original cameraman image (a), degraded one (b), restored images using AA (c) and SA (d) with the contourlet frame.

3.4.2 Convergence speed comparison

In this section, the convergence speed of the described FB, FISTA and PPXA algorithms is compared when used for the SA or AA. We address here the same deconvolution problem as in the previous section using the U2OB wavelet transform. Adopting the AA or SA, Table 3.2 summarizes the encountered difficulties and retained solutions when using the FB, FISTA and PPXA algorithms. In terms of convergence speed, Figs. 3.8-3.9 illustrate the evaluation of the optimality criterion w.r.t. the iteration number to show how the three algorithms compare. It can be easily noticed that when adopting the AA, FB has a faster convergence that FISTA and PPXA, wich perform similarly. However, when the SA is used, FISTA has the highest convergence speed, but FB remains faster than PPXA.

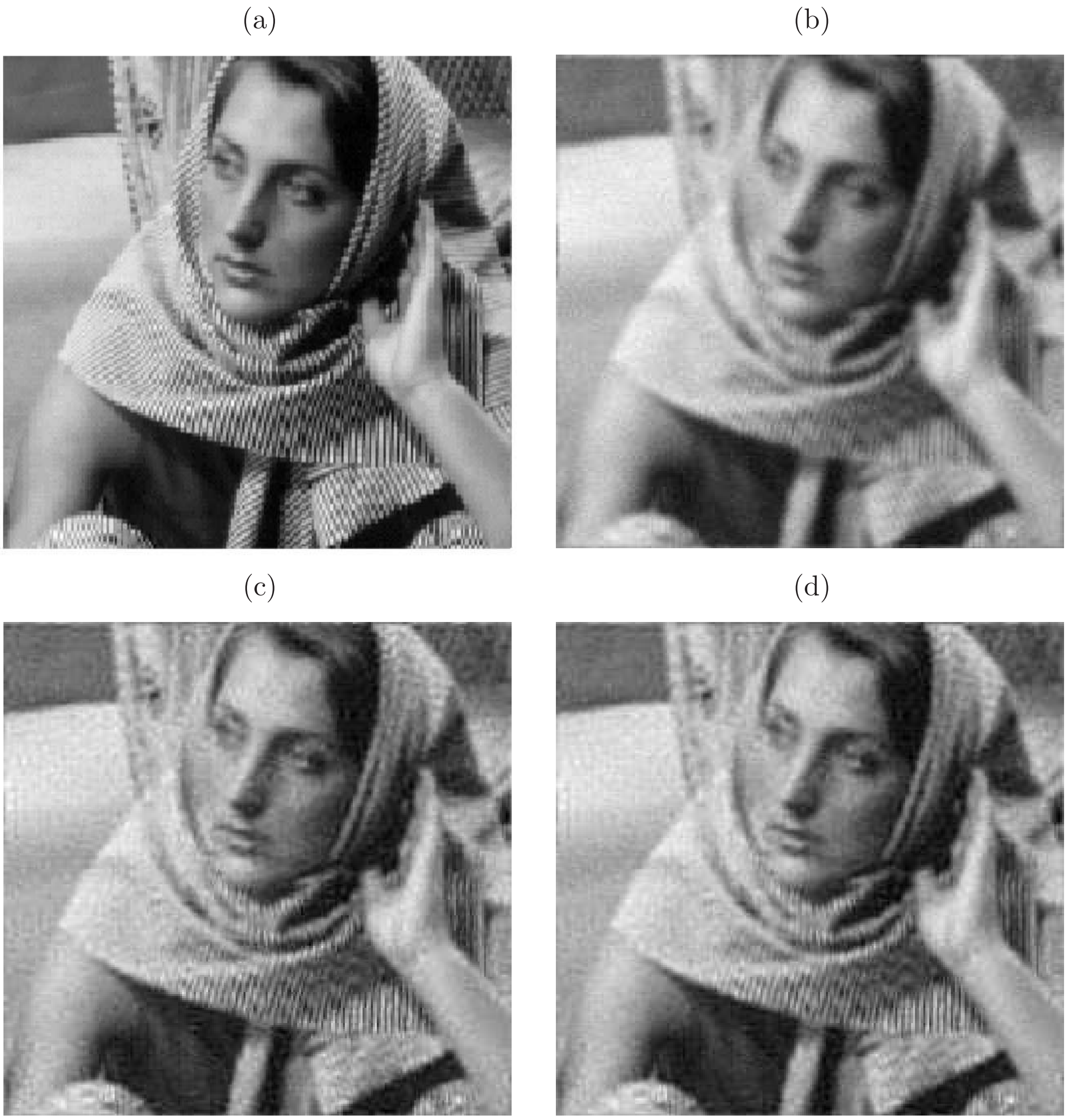


Figure 3.5: Original barbara image (a), degraded one (b), restored images using AA (c) and SA (d) with the contourlet frame.

It can therefore be concluded that for the considered optimization problems, the FB and FISTA algorithms behave similarly and tend to be faster than PPXA. Note however that PPXA is able to address additional regularization terms (e.g. hard constraints on the solution).

3.4.3 Convergence speed gain when calculating $\|HF^*\|^2$

In this section we illustrate the gain in convergence speed when using the FB algorithm for the SA and calculating the norm $\|HF^*\|^2$ using Algorithm 5 instead of approximating it by $\|H\|^2\|F\|^2$.
Fig. 3.10 illustrates first the convergence curve of Algorithm 5. It is clear through this figure

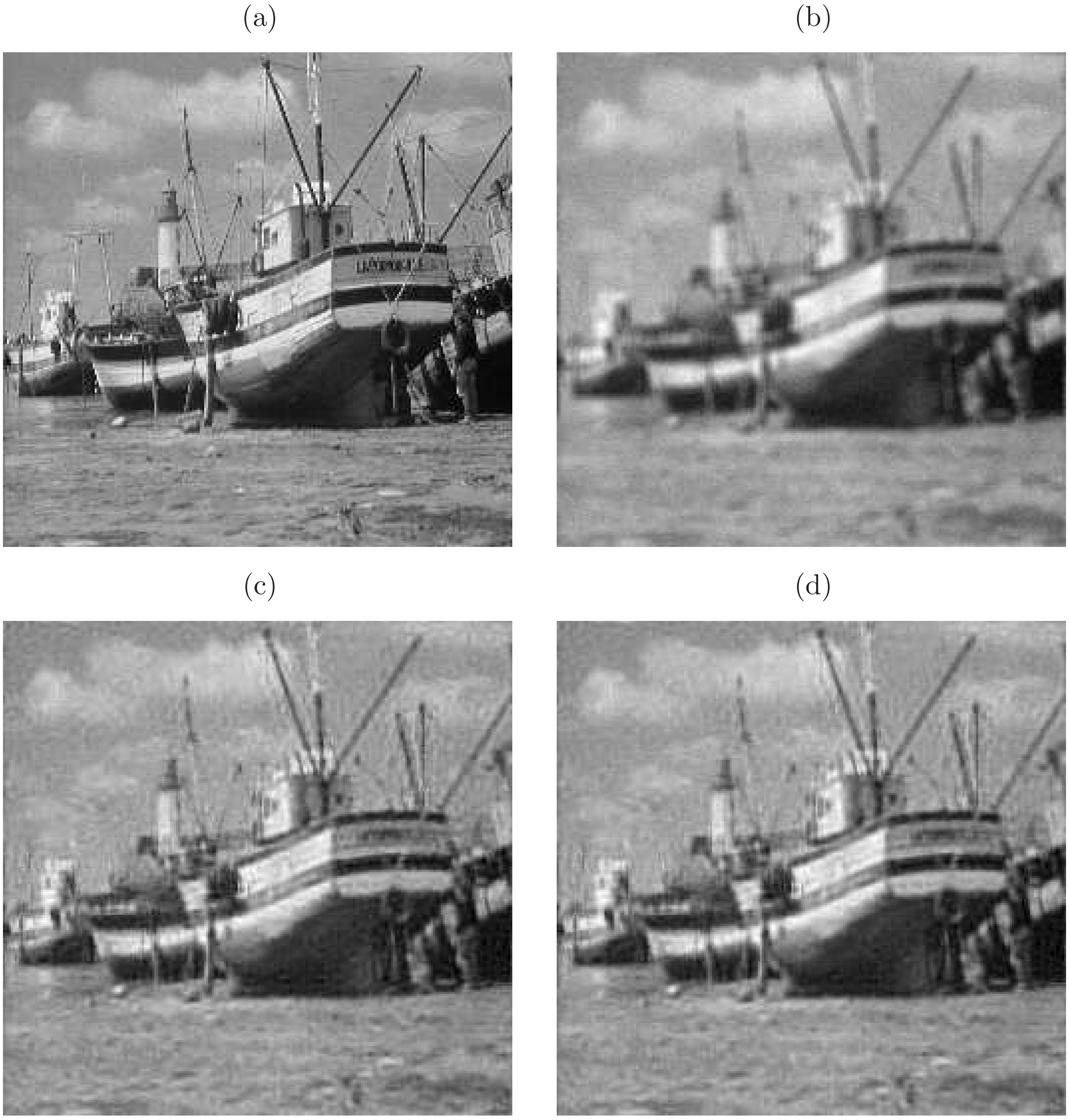


Figure 3.6: Original boat image (a), degraded one (b), restored images using AA (c) and SA (d) with the contourlet frame.

that convergence is reached after about 4 iterations. Fig. 3.11 illustrates the evaluation of the optimality criterion w.r.t the iteration number for these two cases, when the same deconvolution problem as in Section 3.4.1 is considered using the U2OB wavelet transform. It shows that the FB algorithm converges faster when the norm $\|HF^*\|^2$ is calculated using Algorithm 5 than when it is approximated. This confirms that using an optimal step-size for the FB algorithm, which is directly linked to the calculation of $\|HF^*\|^2$ when the SA is adopted, is essential to achieve the best performance of this algorithm in terms of convergence speed.

However, it has been noticed in practice through the same comparison using FISTA that the gain in convergence speed due to the calculation of the norm $\|HF^*\|^2$ is negligible for this algorithm.

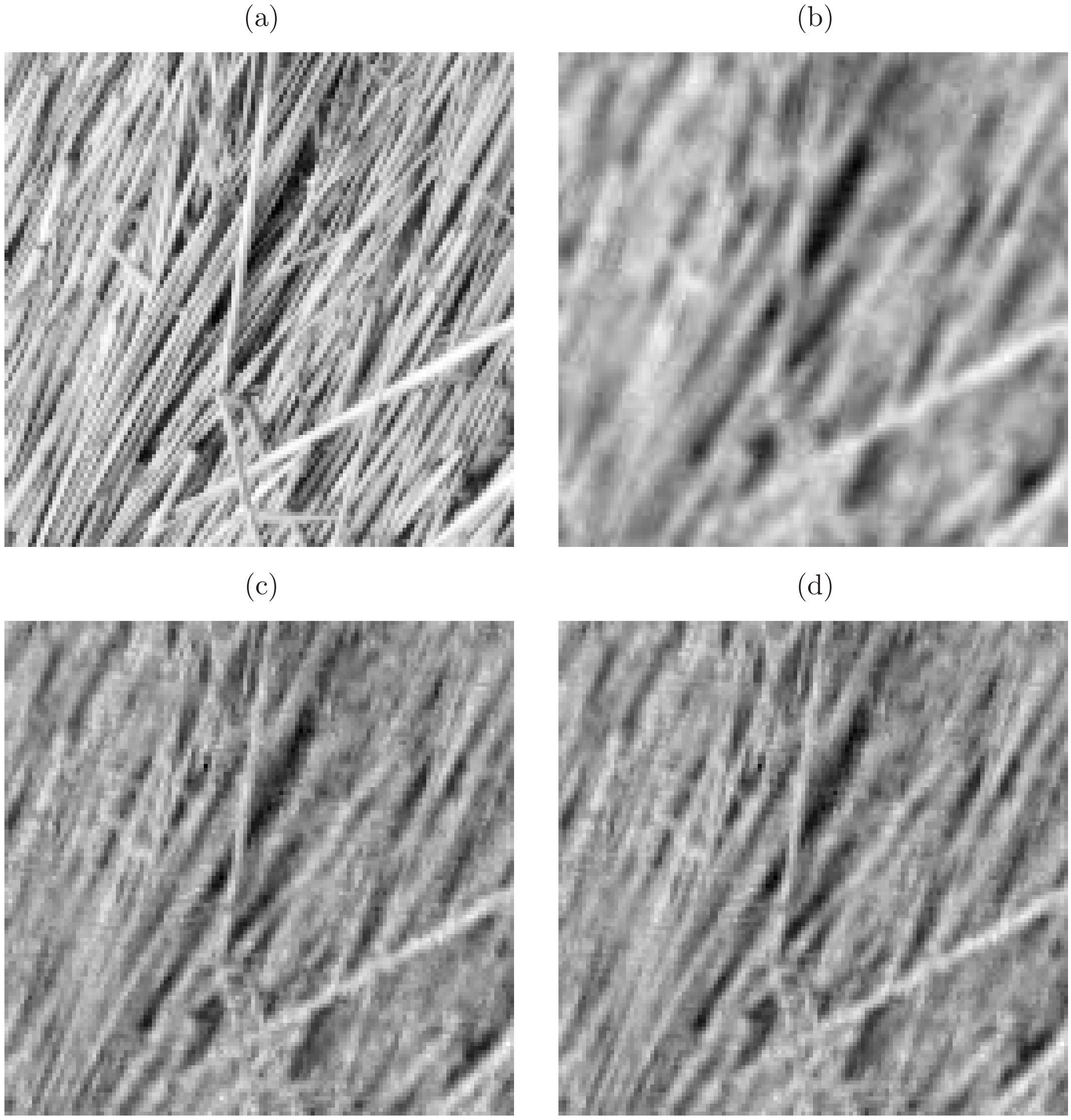


Figure 3.7: Original straw image (a), degraded one (b), restored images using AA (c) and SA (d) with the contourlet frame.

Table 3.1: SNR evaluation for different images and frames in image deblurring.

		SNR			SSIM		
		Degraded	AA	SA	Degraded	AA	SA
cameraman	TIW	17.05	19.93	**20.03**	0.502	0.764	0.764
	GenLot	17.05	20.00	19.90	0.502	0.757	0.745
	contourlet	17.05	19.81	19.72	0.502	0.685	0.687
	U2OB	17.05	19.41	20.02	0.502	0.736	**0.769**
barbara	TIW	18.01	20.59	20.50	0.574	0.718	0.719
	GenLot	18.01	**20.80**	20.76	0.574	0.729	0.732
	contourlet	18.01	20.22	20.25	0.574	0.706	0.708
	U2OB	18.01	20.13	20.33	0.574	0.726	**0.740**
boat	TIW	17.37	19.43	**19.92**	0.520	0.666	**0.694**
	GenLot	17.37	19.63	19.56	0.520	0.687	0.675
	contourlet	17.37	19.55	19.57	0.520	0.664	0.668
	U2OB	17.37	19.54	19.89	0.520	0.672	0.683
straw	TIW	15.22	16.78	17.00	0.353	0.557	0.589
	GenLot	15.22	16.87	17.02	0.353	0.570	0.587
	contourlet	15.22	**17.93**	17.01	0.353	**0.687**	0.595
	U2OB	15.22	17.13	16.98	0.353	0.616	0.609

Table 3.2: Encountered difficulties and retained solutions when using the FB, FISTA and PPXA algorithms with the AA or SA.

		$\operatorname{prox}_{g\circ F}$	$\operatorname{prox}_{f\circ F^*}$	$\|HF^*\|^2$
AA	FB	Iterative approach	——	——
	FISTA	Iterative approach	——	——
	PPXA	Splitting approach	——	——
SA	FB	——	——	Algorithm 5
	FISTA	——	——	Algorithm 5
	PPXA	——	Explicit form	——

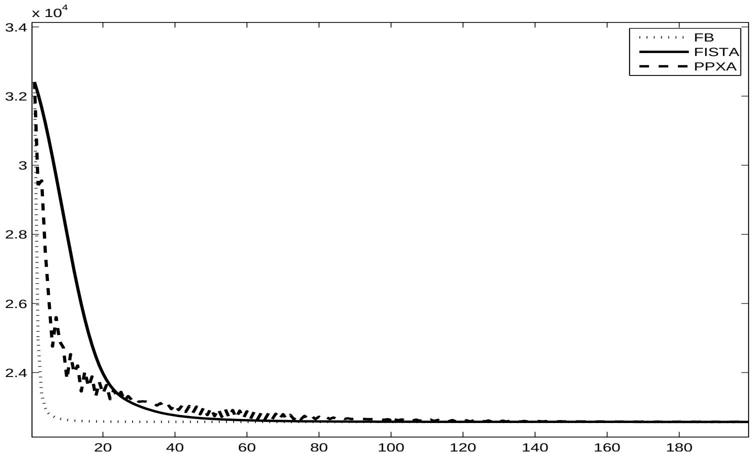


Figure 3.8: Evaluation of the optimality criterion w.r.t iteration number with the AA.

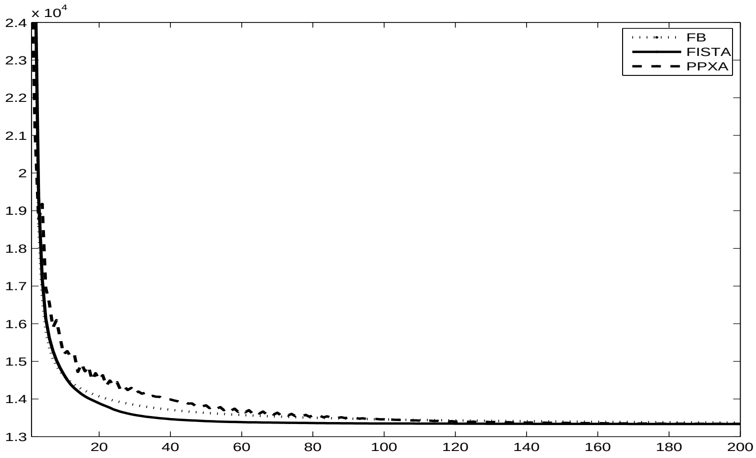


Figure 3.9: Evaluation of the optimality criterion w.r.t iteration number with the SA.

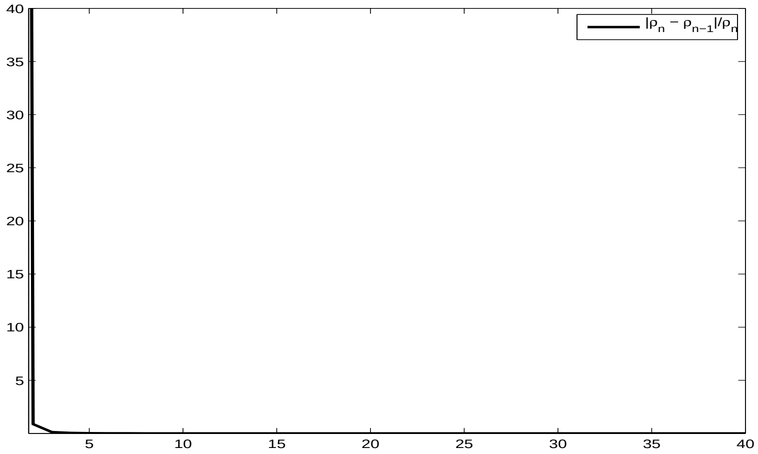


Figure 3.10: Convergence curve of Algorithm 5.

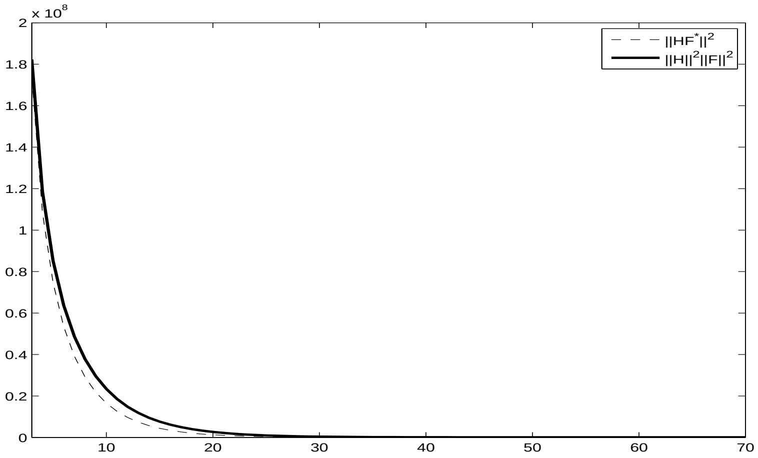


Figure 3.11: Evaluation of the optimality criterion w.r.t iteration number for the FB algorithm when calculating (dot line) and approximating $\|HF^*\|^2$ (solid line).

3.5 Conclusion

In this chapter, the general regularization framework has been introduced starting from the inverse problem up to the inherent optimization challenges. A particular interest has been paid to iterative proximal algorithms and related parameters which have to be carefully chosen: three proximal algorithms have been introduced and compared in terms of convergence speed. The two competing approaches (AA and SA) have been investigated by underlying in each case the difficulties which can be encountered when using redundant frames, and how best to solve them. An application to image deblurring showed that the performance of two methods depends on the used frame and even on the image itself. Choosing the frame representation may also depend on the image content. Consequently, it is quite difficult to conclude in a general context about the superiority of one approach over the other. Note however that similar comparisons have been conducted in [Carlavan et al., 2010] and do not lead to the same conclusions. In the next Chapter, the previously described tools are used when regularization is applied to the pMRI reconstruction problem.

Chapter 4

Regularized SENSE reconstruction

Contents

4.1 Introduction 81

4.2 Regularization in pMRI 82

4.3 Regularization in the WT domain 87

4.4 Combined wavelet-TV regularization 114

4.5 Spatio-temporal regularization 120

4.6 Conclusion 123

4.1 Introduction

SENSitivity Encoding (SENSE) is meant to achieve a perfect reconstruction in the case of noiseless data and perfect coil sensitivity maps knowledge. However, in practice, inaccuracies in the estimation of coil sensitivity maps (especially at the center of the FOV) and using high reduction factors make the reconstruction problem ill-posed. For these reasons, the SENSE performance is limited under such severe experimental conditions since, as reported in Chapter 2, this method reduces to a Weighted Least Squares (WLS) estimation. To overcome this limitation, some alternatives have been proposed like the optimization of the coil geometry [Weiger et al., 2001] and the improvement of coil sensitivity profiles estimation [Blaimer et al., 2004]. However, at low magnetic fields (such as 1.5 Tesla), the maximal reduction factor usually employed is $R = 2$ because severe aliasing artifacts affect the reconstructed images when larger R values are used.

To further improve the spatial or temporal resolution, or simply reduce the global imaging time while keeping the same acquistion parameters, it would be necessary to increase the reduction factor R. To compensate for the intrinsic degradation of the image quality, it is crucial to regularize the reconstruction process. For an introductory view of linear inverse problems with a special emphasis on pMRI reconstruction, the interested reader can refer to [Ribes and Schmitt, 2008].

This chapter aims at presenting the developed wavelet regularization approach and its different extensions. In Section 4.2, we first present different regularization techniques which have already beenreported in the pMRI reconstruction literature. Section 4.3 develops the proposed regularization strategies based on wavelet transforms. A combination of wavelet and Total Variation (TV) penalizations is also investigated in Section 4.4. For the specific fMRI concern using EPI sequences, an extension of this approach is presented in Section 4.5 taking into account temporal dependencies between acquired volumes during an fMRI session. Finally, a brief conclusion is drawn in Section 4.6.

4.2 Regularization in pMRI

Regularization techniques were first introduced in pMRI literature by Liang in 2002 using the standard Tikhonov regularization [Liang et al., 2002; Ying et al., 2004; Lin et al., 2004]. This kind of quadratic regularization can in fact improve the reconstructed image quality when the experimental conditions are not too degraded (high magnetic field intensity such as 3T, low reduction factor such as $R = 2$, low noise level ...). However, when it is not the case, artifacts are more severe and such regularization methods cannot be relied on as illustrated in Section 4.2.1. Since 2007, this underperformance of the Tikhonov regularization has motivated the use of edge-preserving penalizations in the regularization process such as TV [Liu et al., 2007; Block et al., 2007; Liu et al., 2009]. Regularization based on these kinds of penalties aims at preserving image discontinuities by adapting the penalization to the intensity variation level. As illustrated in Section 4.2.2 and reported in the later cited works, TV is well adapted to piecewise smooth images, which is not necessarily the case in most of the real MRI images. In the same context of edge-preserving regularization, wavelet-based penalizations have been introduced in the pMRI regularization literature since 2008 to alleviate this problem [Chaari et al., 2008; Liu et al., 2008; Guerquin-Kern et al., 2009] due to their hierarchical and sparse representations. This chapter will show that wavelet transforms allow us to achieve accurate reconstruction of full FOV images even under severe experimental conditions (low magnetic field intensity such as 1.5T, high reduction factor such as $R = 4$ and noise level ...).

These edge-preserving penalizations have also been used in conventional MRI [Boubertakh et al., 2000; Boubertakh et al., 2006] regularization and when Compressed Sensing (CS) and pMRI were combined in a number of recent works [Liang et al., 2009; Zou et al., 2008]. In these later works, a two-spets procedure is proposed. First, CS is used when reconstructing the reduced FOV images. A standard SENSE reconstruction is then performed to generate the full FOV image. The main reconstruction methods that rely on this procedure are SparseSENSE [Zou et al., 2008] and CS-SENSE [Liang et al., 2009]. It is worth noting that the first step of these methods (i.e. reconstructing reduced FOV images) relies on what is called Sparse MRI [Lustig et al., 2007], which is based on CS instead of the Fourier encoding in order to reduce the acquisition time.

However, in contrast to the Tikhonov regularization where a closed form is available for the regularized solution, edge-preserving penalties (TV or wavelets) may lead to non-differentiable optimality criteria. Hence, the optimal solution (or its approximation) is generally calculated iteratively, which means that the optimization algorithm will play an important role in the efficiency of the regularization process. For this reason, much attention has been paid in our regularization methods to design a rigorous optimization algorithm, in contrast to some of the previously cited works [Liu et al., 2008; Liang et al., 2009; Zou et al., 2008; Lustig et al., 2007] where only an approximation of the optimality criterion is minimized.

4.2.1 Tikhonov regularization

To improve the robustness of the solution to this ill-posed problem, Tikhonov regularization (also called quadratic regularization) has been widely used in the early pMRI regularization literature. As shown in [Ying et al., 2004; Liang et al., 2002], it consists of computing $\widehat{\boldsymbol{\rho}}_{\mathrm{PWLS}}(\boldsymbol{r})$ as the minimizer of a Penalized Weighted Least Squares (PWLS) criterion:

$$\begin{aligned}\widehat{\boldsymbol{\rho}}_{\mathrm{PWLS}}(\boldsymbol{r}) &= \arg\min_{\boldsymbol{\rho}(\boldsymbol{r})\in\mathbb{C}^L} \mathcal{J}_{\mathrm{PWLS}}(\boldsymbol{\rho}(\boldsymbol{r}))\\ &= \arg\min_{\boldsymbol{\rho}(\boldsymbol{r})\in\mathbb{C}^L} \big[\mathcal{J}_{\mathrm{WLS}}(\boldsymbol{\rho}(\boldsymbol{r})) + \kappa \,\| \boldsymbol{\rho}(\boldsymbol{r}) - \boldsymbol{\rho}_{\mathrm{r}}(\boldsymbol{r}) \|^2_{\boldsymbol{I}_R}\big],\end{aligned} \tag{4.1}$$

where $\boldsymbol{I}_R$ is the R-dimensional identity matrix and $\kappa > 0$ is the regularization parameter. Interestingly, $\widehat{\boldsymbol{\rho}}_{\mathrm{PWLS}}$ matches the Maximum A Posteriori (MAP) estimator in the Bayesian framework provided that a separable complex circular Gaussian prior $\mathcal{N}(\boldsymbol{\rho}_{\mathrm{r}}(\boldsymbol{r}), \frac{1}{\kappa}\boldsymbol{I}_R)$ is considered on $\boldsymbol{\rho}(\boldsymbol{r})$.

The regularization parameter κ ensures a balance between the closeness to the data and the penalty term, which controls the deviation from a given reference vector $\boldsymbol{\rho}_{\mathrm{r}}(\boldsymbol{r})$. The solution $\widehat{\boldsymbol{\rho}}_{\mathrm{PWLS}}(\boldsymbol{r})$ is given by:

$$\widehat{\boldsymbol{\rho}}_{\mathrm{PWLS}}(\boldsymbol{r}) = \boldsymbol{\rho}_{\mathrm{r}}(\boldsymbol{r}) + \big(\boldsymbol{S}^{\mathsf{H}}(\boldsymbol{r})\boldsymbol{\Psi}^{-1}\boldsymbol{S}(\boldsymbol{r}) + \kappa \boldsymbol{I}_R\big)^{-1}\boldsymbol{S}^{\mathsf{H}}(\boldsymbol{r})\boldsymbol{\Psi}^{-1}\big(\boldsymbol{d}(\boldsymbol{r}) - \boldsymbol{S}(\boldsymbol{r})\boldsymbol{\rho}_{\mathrm{r}}(\boldsymbol{r})\big).$$

Note that the accuracy of $\widehat{\boldsymbol{\rho}}_{\mathrm{PWLS}}(\boldsymbol{r})$ depends on the reference $\boldsymbol{\rho}_{\mathrm{r}}(\boldsymbol{r})$ and the choice of the regularization parameter κ. On the one hand, if κ tends to zero, the solution $\widehat{\boldsymbol{\rho}}_{\mathrm{PWLS}}(\boldsymbol{r})$ converges to $\widehat{\boldsymbol{\rho}}_{\mathrm{WLS}}(\boldsymbol{r})$. On the other hand, for large κ values the penalty term dominates in $\mathcal{J}_{\mathrm{PWLS}}$ and we get $\widehat{\boldsymbol{\rho}}_{\mathrm{PWLS}}(\boldsymbol{r}) \to \boldsymbol{\rho}_{\mathrm{r}}(\boldsymbol{r})$. To select a relevant value for κ, some works have proposed to resort to the discrepancy principle or the L-curve technique [Rabrait et al., 2008; Griesbaum et al., 2008; Engl, 1987; Lin et al., 2004; Ying et al., 2008; Block et al., 2007; Lin et al., 2005] in order to achieve the best trade-off between reconstruction noise removal and contrast reduction. Other contributions have promoted the Maximum Likelihood (ML) estimator using an Expectation-Maximization (EM) algorithm for its computation [Ciuciu et al., 2003; Ciuciu et al., 2004].
Quadratic regularization is actually known to produce smoothing effects [Phillips, 1962; Twomey, 1963]. This is well illustrated in Figs. 4.1, 4.2 and 4.3 where reconstructed images using Tikhonov regularization are shown. Note that illustrations here were made on the same datasets as in paragraph 2.5.2.3.

A comparison with the basic-SENSE reconstruction can be made by referring to Figs. 2.18, 2.19 for anatomical images, and 2.20 for functional ones. Note that in Tikhonov regularization, the reference image ρ_{r} was chosen as a mean image based on the basic-SENSE reconstruction, which contains the mean value of the signal of interest within the brain mask. The regularization parameter κ was manually fixed and different settings were tested in order to retain the best choice in terms of SNR/visual quality since a reference anatomical image acquired in non-parallel imaging was available in this experiment. As expected, some aliasing artifacts in the basic-SENSE reconstructed images are smoothed by the Tikhonov regularization at the expense of oversmoothing effects. To avoid introducing these oversmoothing effects and preserve tiny structures and image discontinuities, a solution lies in the use of edge-preserving penalty terms such as TV [Wang and Haomin, 2006; Keeling, 2003; Coulon et al., 2004; Liu et al., 2007] or half-quadratic ones [Geman

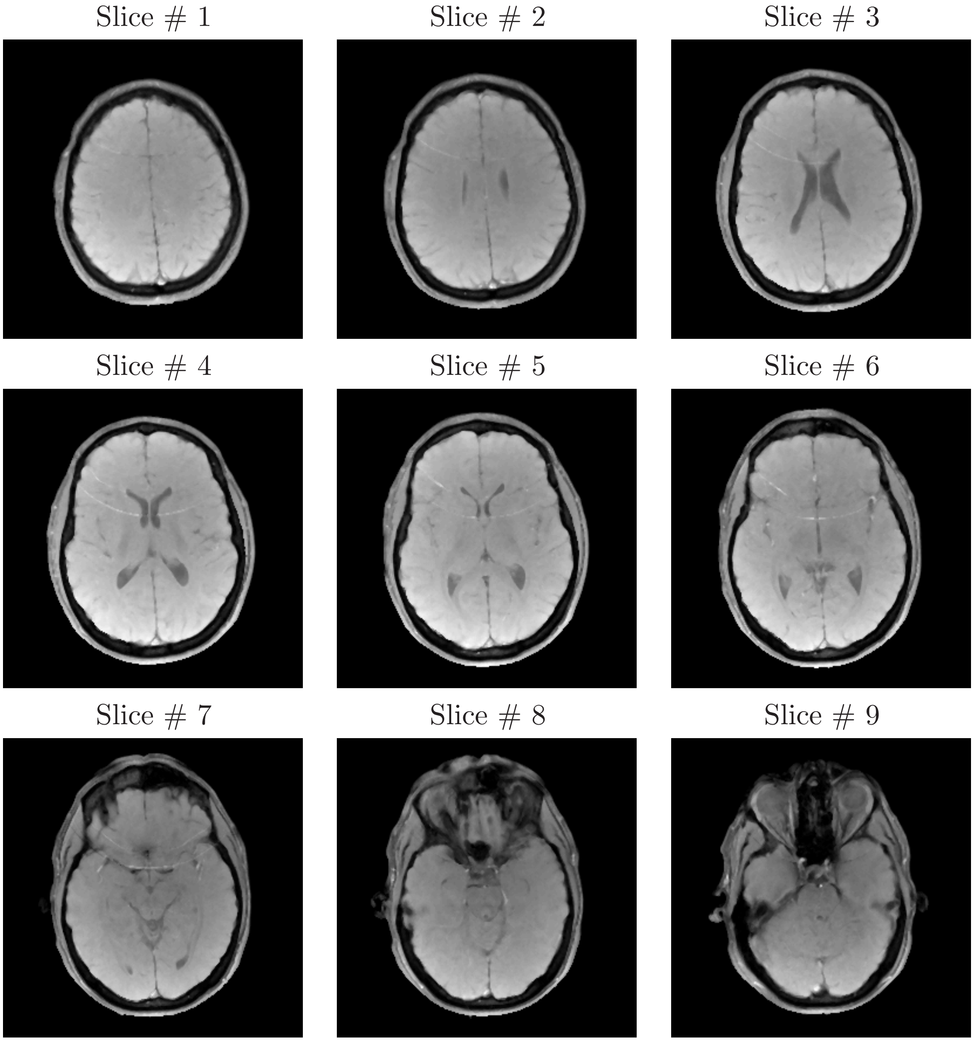


Figure 4.1: Reconstructed anatomical slices using Tikhonov regularization for $R = 2$.

and Reynolds, 1992; Geman and Yang, 1995; Ciuciu et al., 2001; Ciuciu and Idier, 2004].

4.2.2 Total variation regularization

Generally speaking, edge-preserving penalizations take place in the image domain and make the regularization more efficient by limiting blurring effects and preserving the image boundaries. One typical example of such penalization is the TV penalty. TV regularization has already been used in MRI literature, especially for removing artifacts occuring in fast acquisition schemes [Lustig et al., 2007] or parallel imaging [Raj et al., 2007; Liu et al., 2007; Zou et al., 2008; Liang et al., 2008; Block et al., 2007]. The ensuing global TV-

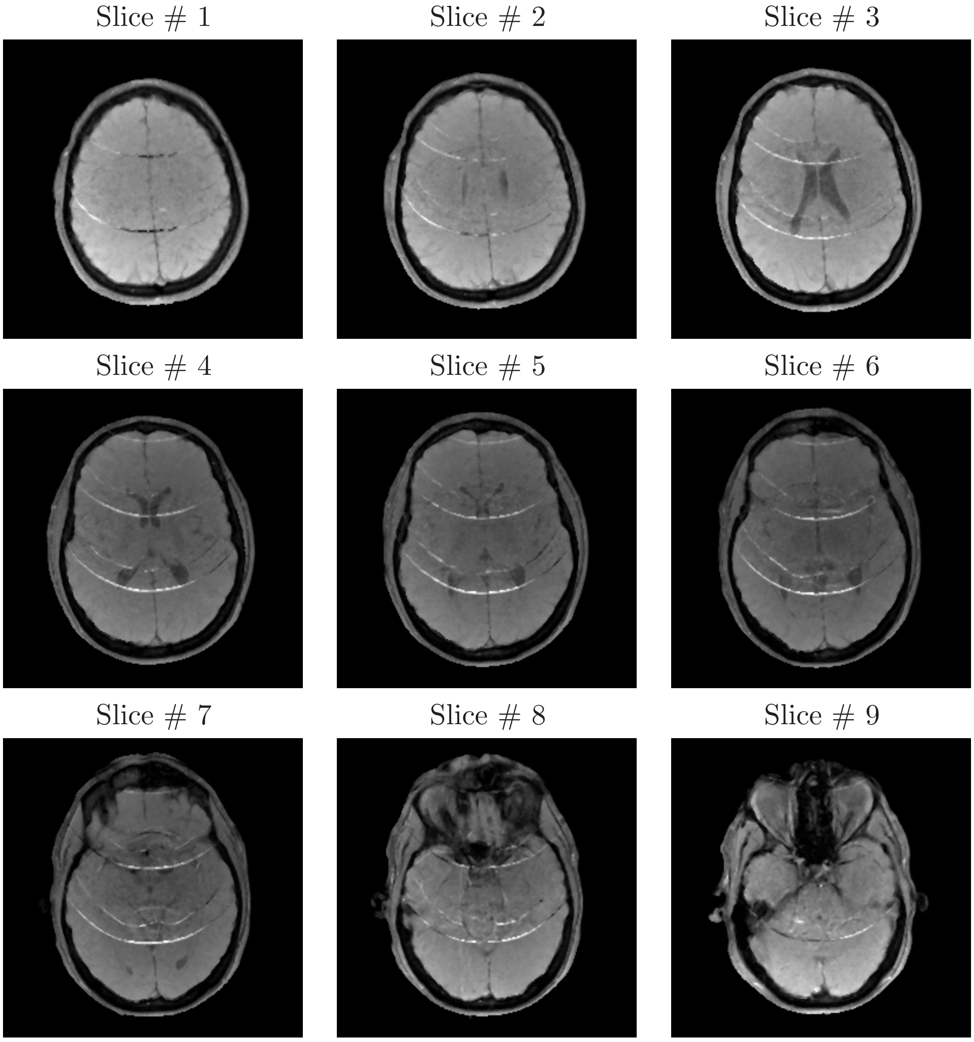


Figure 4.2: Reconstructed anatomical slices using Tikhonov regularization for $R = 4$.

regularized criterion reads as follows:

$$\begin{aligned} \widehat{\rho}_{\mathrm{TV}} &= \arg\min_{\rho \in \mathbb{C}^{Y \times X}} \mathcal{J}_{\mathrm{TV}}(\rho) \\ &= \arg\min_{\rho \in \mathbb{C}^{Y \times X}} \mathcal{J}_{\mathrm{L}}(\rho) + \kappa \|\rho\|_{\mathrm{TV}}, \end{aligned} \tag{4.2}$$

where $\| \cdot \|_{\mathrm{TV}}$ is the TV semi-norm [Liu et al., 2007; Liang et al., 2008], $\kappa > 0$ is the regularization parameter, and

$$\mathcal{J}_{\mathrm{L}}(\rho) = \sum_{\mathbf{r} \in \{1,\dots,Y/R\} \times \{1,\dots,X\}} \mathcal{J}_{\mathrm{WLS}}(\boldsymbol{\rho}(\mathbf{r})). \tag{4.3}$$

Note that unlike in Eq. (4.1), the regularization term here is no longer separable in $\boldsymbol{r}$.

It has been reported in some recent works from regularization literature [Liu et al.,

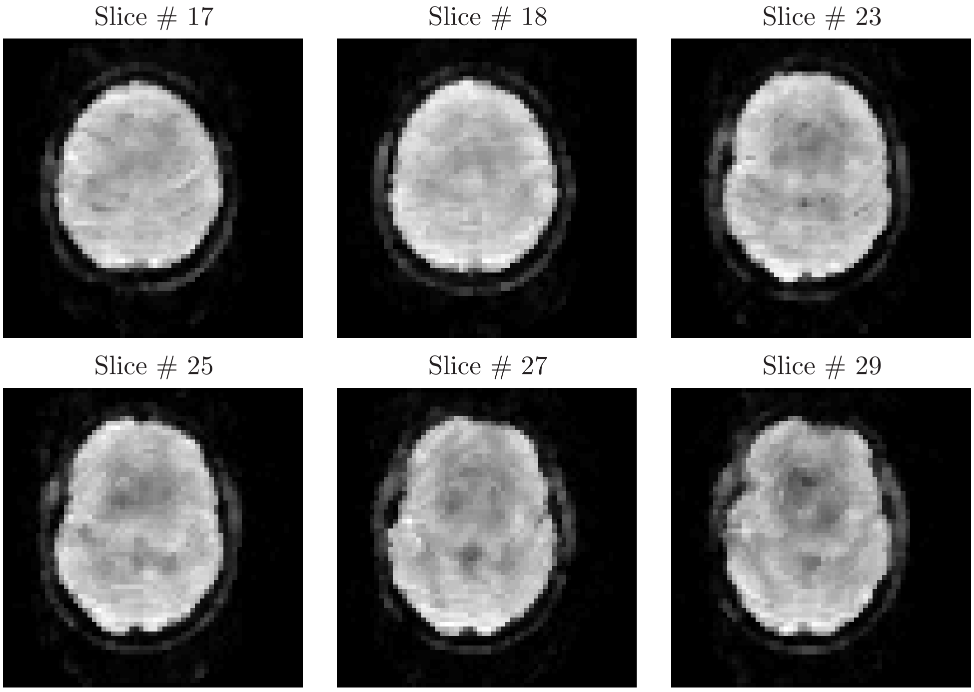


Figure 4.3: Two EPI reconstructed slices using Tikhonov regularization for $R = 4$.

2007] that TV performs well especially on piecewise smooth images. It has also been observed that TV regularization may sometimes lead to reconstruction noise increase [Liu et al., 2009]. Hence, under severe experimental conditions (at 1.5 Tesla and $R = 4$) as shown in Fig. 4.4, the most important artifacts appear as rings with sharp boundaries. For illustration and comparison purpose, TV regularization has been applied to anatomical and functional MRI data using moderate and large regularization levels (by tuning the value of the regularization parameter κ). TV regularized reconstruction results at $R = 4$ are presented in Fig. 4.4 on anatomical and functional images corresponding to the 9th anatomical and 27th functional slices, respectively. Figs. 4.4(a)-(b) show that some ring artifacts were smoothed but the strongest ones still exist. Moreover, the reconstructed images present some staircase effects. These effects are much more visible on functional EPI data at lower spatial resolution. To limit these undesirable effects, stronger TV penalization (by increasing the value of κ) can be used at the expense of the information content in the reconstructed images as shown in Figs. 4.4(c)-(d).

As shown by these experiments under severe acquisition conditions, image-based edge-preserving priors are not well-adapted to regularize the reconstruction of the full FOV image: severe aliasing artifacts are smoothed at the expense of the lost of useful information. On the contrary, if such information is preserved, aliasing artifacts will not be sufficiently smoothed. Therefore, in what follows, we make use of Wavelet Transforms (WT) to improve artifact localization both in space and scales and introduce some adequate priors in the wavelet space that promote the sparsity of such decomposition. We then derive efficient optimization algorithms that are able to cope with convex but non-differentiable criteria. As illustrated later, it will be shown that our penalization is better suited to the

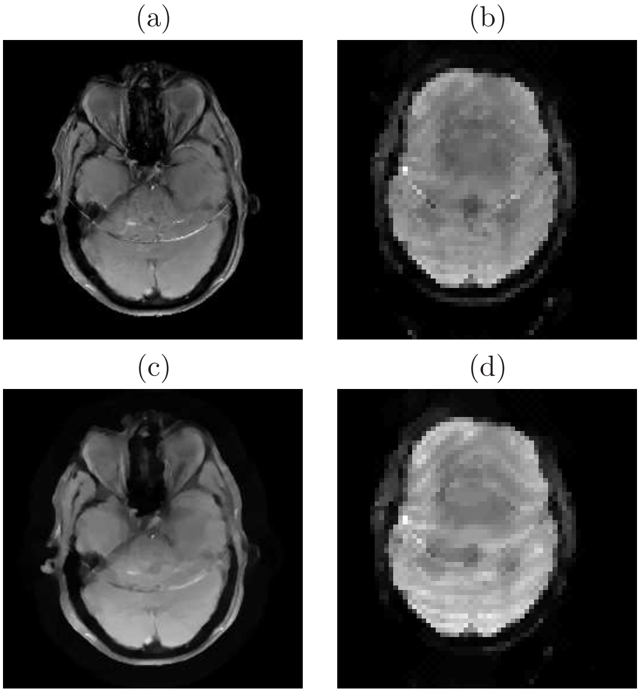


Figure 4.4: Reconstructed anatomical (left) and functional EPI (right) images using TV regularization in the image domain with moderate (a)-(b) and strong (c)-(d) regularization levels ($R = 4$).

regularization task than its image-based TV counterpart.

4.3 Regularization in the WT domain

4.3.1 Motivation

As mentioned earlier, the presence of sharp artifacts makes the basic SENSE reconstruction inefficient under severe experimental conditions. Therefore, it is highly desirable to resort to an image representation where these artifacts can be easily detected and hence attenuated. In this respect, the WT has been recognized as a powerful tool that enables a good space and frequency localization of useful informations [Mallat, 1998]. In previous literature, many wavelet decompositions and extensions have already been reported in the literature offering different features in order to provide sparse image representations. For instance, decompositions onto orthonormal dyadic wavelet bases [Daubechies, 1988] including the Haar transform [Haar, 1910] as a special simple case or decompositions onto biorthogonal dyadic wavelets [Cohen et al., 1992], M-band wavelet representations [Steffen et al., 1993] and wavelet packet representations [Coifman and Wickerhauser, 1992] have been extensively investigated in image denoising [Donoho and Johnstone, 1995; Leporini and Pesquet, 2001; Müller and Vidakovic, 1999; Heurta, 2005; Daubechies et al., 2004; Daubechies and Teschke, 2005] and deconvolution [Daubechies et al., 2004; Daubechies and Teschke, 2005; Vonesch and Unser, 2008; Chaux et al., 2006a].

In medical imaging, wavelet decompositions have also been widely used for image denoising [Weaver et al., 1991; Wang and Haomin, 2006; Pizurica et al., 2006], coil sensitivity map estimation and encoding schemes [Lin et al., 2003; Gelman and Wood, 1996; Wendt et al., 1998] in MRI, activation detection in fMRI [Ruttimann et al., 1998; Meyer, 2003; Van De Ville et al., 2004; Van De Ville et al., 2006], tissue characterization in ultrasound imaging [Mojsilovic et al., 1998] and tomographic reconstruction [Pustelnik et al., 2009].

An appealing property of the resulting decomposition is that the statistical distributions of approximation and detail wavelet coefficients can be easily modelled in a realistic way. Hence, the Bayesian framework can be adopted to capture relevant information in the data through the likelihood definition, deriving appropriate priors and selecting an efficient estimator to perform reconstruction in the wavelet transform domain.

4.3.2 Definitions and notations

In the following section, T stands for the WT operator and corresponds to a discrete decomposition onto a separable 2D M-band wavelet basis performed over $j_{\max}$ resolution levels. The full FOV image $\overline{\rho}$ of size $Y \times X$ can be seen as an element of the Euclidean space $\mathbb{C}^K$ with $K = Y \times X$ endowed with the standard inner product $\langle \cdot | \cdot \rangle$ and norm $\| \cdot \|$. As mentioned above, we are only interested in reconstructing one slice (2D image) for solving the 1D-SENSE problem. Hence, only 2D WT operators are investigated. However, an extension of the proposed approach to deal with 3D images using 3D WT operators will be presented in Section 4.5. In this context, the following notations are introduced.

Definition 4.3.1
Let $(e_k)_{1 \le k \le K}$ be the considered discrete wavelet basis of the space $\mathbb{C}^K$. The wavelet decomposition operator T is defined as the linear operator:

$$\begin{aligned} T\colon \mathbb{C}^K &\to \mathbb{C}^K \\ \rho &\mapsto (\langle \rho | e_k \rangle)_{1 \le k \le K}. \end{aligned} \tag{4.4}$$

The adjoint operator T^* used for reconstruction purpose is then defined as the bijective linear operator:

$$\begin{aligned} T^*\colon \mathbb{C}^K &\to \mathbb{C}^K \\ (\zeta_k)_{1 \le k \le K} &\mapsto \sum_{k=1}^{K} \zeta_k e_k. \end{aligned} \tag{4.5}$$

The resulting wavelet coefficient field of a target image function ρ is defined by $\zeta = \big(\boldsymbol{\zeta}_a, (\boldsymbol{\zeta}_{o,j})_{o \in \mathbb{O}, 1 \le j \le j_{\max}}\big)$ where $\boldsymbol{\zeta}_a = (\zeta_{a,k})_{1 \le k \le K_{j_{\max}}}$ and $\boldsymbol{\zeta}_{o,j} = (\zeta_{o,j,k})_{1 \le k \le K_j}$, $K_j = KM^{-2j}$ being the number of wavelet coefficients in a given subband at resolution j (by assuming that Y and X are multiple of $M^{j_{\max}}$). The coefficients have been reindexed in such a way that $\zeta_{a,k}$ denotes the k-th approximation coefficient at resolution level $j_{\max}$ and $\zeta_{o,j,k}$ denotes the k-th detail coefficient at resolution level j and orientation $o \in \mathbb{O} = \{0, \ldots, M-1\}^2 \setminus \{(0,0)\}$ where $k = 1, \ldots, K_j$. In the dyadic case ($M = 2$), there are three orientations corresponding to the horizontal, vertical or diagonal directions. Note that, when an orthonormal wavelet basis is considered, the adjoint operator T^* reduces to the inverse WT operator T^{-1} and the operator norm $\|T\|$ of T is equal to 1.

4.3.3 Wavelet-based regularized reconstruction

An estimate of the full FOV image $\overline{\rho}$ will be generated through the reconstruction wavelet operator T^*. Let $\overline{\zeta}$ be the unknown wavelet coefficients such that $\overline{\rho} = T^*\overline{\zeta}$. We aim at building an estimate $\widehat{\zeta}$ of the vector of coefficients $\overline{\zeta}$ from the observations $\mathrm{d} = (\boldsymbol{d}(\boldsymbol{r}))_{\mathbf{r}\in\{1,\ldots,Y/R\}\times\{1,\ldots,X\}}$.

To this end, we derive a Bayesian approach relying on suitable priors on the wavelet coefficients.

4.3.3.1 Likelihood

Given the observation model in Eq. (2.16) and the assumptions regarding the noise, the likelihood function factorizes over pixels lying in the $Y \times X$ FOV:

$$\begin{aligned} p(\mathrm{d} \mid T^*\zeta) &= \prod_{\mathbf{r}\in\{1,\ldots,Y/R\}\times\{1,\ldots,X\}} p(\boldsymbol{d}(\mathbf{r}) \mid \boldsymbol{\rho}(\mathbf{r})) \propto \prod_{\mathbf{r}\in\{1,\ldots,Y/R\}\times\{1,\ldots,X\}} \exp\left(-\mathcal{J}_{\mathrm{WLS}}(\boldsymbol{\rho}(\mathbf{r}))\right) \\ &\propto \exp\left(-\mathcal{J}_{\mathrm{L}}(T^*\zeta)\right) \end{aligned} \tag{4.6}$$

with $\mathcal{J}_{\mathrm{L}}$ and $\rho = T^*\zeta$ are defined from $\boldsymbol{\rho}$ as in Eq. (4.3).

4.3.3.2 Prior

Let f be the prior probability density function (pdf) of the image in the wavelet domain. We will assume here that the real and imaginary part of the wavelet coefficients are independent. We will also assume that the real (resp. imaginary) parts of the wavelet coefficients are independent and identically distributed (iid) in each subband. Their statistical characteristics may however vary between two distinct subbands. Furthermore, by looking at the empirical distributions of the real and imaginary parts of the considered wavelet coefficients, we have noticed that their empirical histograms are well-fitted by what we called a *Generalized Gauss-Laplace* (GGL) distribution. The histograms present a single mode and their shape vary between the Gaussian and Laplacian densities. The corresponding pdf reads:

$$\forall \xi \in \mathbb{R}, \quad f(\xi; \alpha, \beta) = \sqrt{\frac{\beta}{2\pi}} \frac{e^{-(\alpha|\xi| + \frac{\beta}{2}\xi^2 + \frac{\alpha^2}{2\beta})}}{\operatorname{erfc}(\frac{\alpha}{\sqrt{2\beta}})}, \tag{4.7}$$

where $\alpha \in \mathbb{R}_+$ and β in $\mathbb{R}_+^*$ are hyper-parameters to be estimated using the proposed MCMC approach in Chapter 5 for instance. Fig. 4.5 illustrates the empirical histograms of real and imaginary parts of the horizontal detail subband at the first resolution level using the dyadic ($M = 2$) wavelet decomposition with Daubechies filters of length 8. This figure shows also that the adopted GGL distribution better fits the empirical histogram than a Generalized Gaussian (GG) pdf [1]. This fact was confirmed by applying a Kolmogorov-Smirnov goodness-of-fit test.

Note that the anti-logarithm of the pdf in Eq. (4.7) is an exact match of the *elastic net* penalty used in statistics [Zou and Hastie, 2005; Zou and Zhang, 2009]. Interestingly, this kind of penalization is known to be an improvement of the ℓ_1 penalization for analyzing high-dimensional data [Zou and Zhang, 2009]: the ℓ_1 part allows automatic variable selection and promotes the sparsity of the solution, while the quadratic part improves the prediction and tends to preserve the solution regularity. When formulated from a Bayesian viewpoint, this elastic net penalization simply consists of assuming a prior on the signal under investigation which matches the pdf in Eq. (4.7).

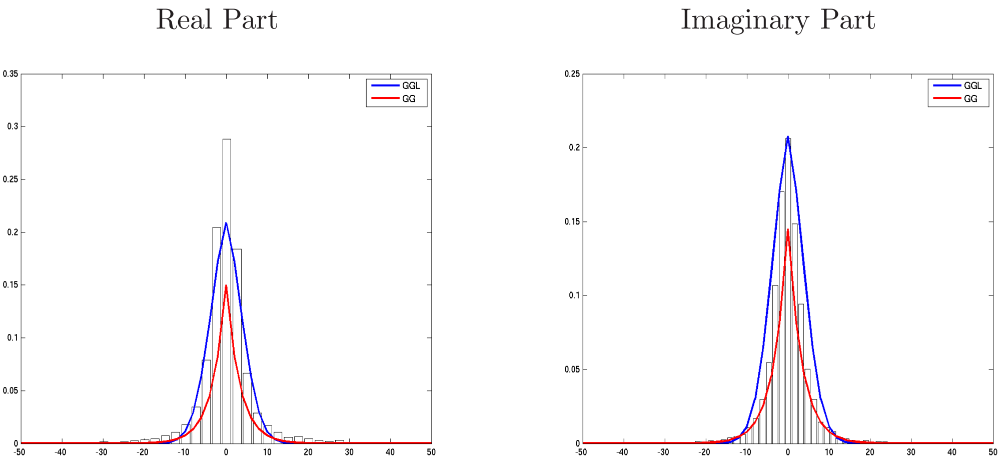

Real Part
Imaginary Part

Figure 4.5: Example of normalized empirical histograms of wavelet coefficients and associated pdfs using GG (red) and GGL (blue) distributions, the hyperparameters being estimated using the proposed MCMC approach in Chapter 5.

At the coarsest resolution level $j_{\max}$, as often used in the wavelet literature, the distributions of both the real and imaginary parts of the approximation coefficients is assumed to be Gaussian because of its low frequency content.

Due to its familiarity and simplicity, the MAP estimator will be used for the estimation purpose. As we will see later, using the MAP estimator is also easier to derive than many other estimators like the Minimum Meas Square Error (MMSE) since it is obtained through minimizing criteria which are more complicated to integrate.

---

[1] We recall that the pdf of a GG distribution is given by: $\forall x \in \mathbb{R}$, $f(x|\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\frac{|x|^\beta}{\alpha^\beta}\right)$ where $\alpha > 0$ and $\beta > 0$.

4.3.3.3 Bayesian inference

Based on the prior and the likelihood given hereabove, the MAP estimator is computed by maximizing the full posterior distribution or minimizing its negative log-likelihood:

$$\widehat{\zeta}^{\mathrm{MAP}} = \arg\max_{\zeta\in\mathbb{C}^K}\big(\,\ln p(\mathrm{d} \mid T^*\zeta) + \ln f(\zeta)\,\big),$$

or equivalently by minimizing the following criterion:

$$\begin{aligned}\widehat{\zeta}^{\mathrm{MAP}} &= \arg\min_{\zeta\in\mathbb{C}^K} \mathcal{J}_{\mathrm{WT}}(\zeta)\\ &= \arg\min_{\zeta\in\mathbb{C}^K}\big(\mathcal{J}_{\mathrm{L}}(T^*\zeta) + \mathcal{J}_{\mathrm{P}}(\zeta)\big) \qquad (4.8)\\ \text{with}\quad \mathcal{J}_{\mathrm{P}}(\zeta) &= \sum_{k=1}^{K_{j\max}} \Phi_a(\zeta_{a,k}) + \sum_{o\in\mathbb{O}}\sum_{j=1}^{j_{\max}}\sum_{k=1}^{K_j} \Phi_{o,j}(\zeta_{o,j,k}) \qquad (4.9)\end{aligned}$$

and

$$\Phi_a(\zeta_{a,k}) = \frac{(\mathrm{Re}(\zeta_{a,k}) - \mu^{\mathrm{Re}})^2}{(\sqrt{2}\sigma^{\mathrm{Re}})^2} + \frac{(\mathrm{Im}(\zeta_{a,k}) - \mu^{\mathrm{Im}})^2}{(\sqrt{2}\sigma^{\mathrm{Im}})^2}, \qquad (4.10)$$

$$\begin{aligned}\Phi_{o,j}(\zeta_{o,j,k}) = \alpha_{o,j}^{\mathrm{Re}}|\mathrm{Re}(\zeta_{o,j,k})| + \frac{\beta_{o,j}^{\mathrm{Re}}}{2}|\mathrm{Re}(\zeta_{o,j,k})|^2 + \alpha_{o,j}^{\mathrm{Im}}|\mathrm{Im}(\zeta_{o,j,k})|\\ + \frac{\beta_{o,j}^{\mathrm{Im}}}{2}|\mathrm{Im}(\zeta_{o,j,k})|^2. \qquad (4.11)\end{aligned}$$

Hereabove, $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ (or $\cdot^{\mathrm{Re}}$ and $\cdot^{\mathrm{Im}}$) stand for the real and imaginary parts, respectively. The prior parameters $\boldsymbol{\alpha}_{o,j} = (\alpha_{o,j}^{\mathrm{Re}}, \alpha_{o,j}^{\mathrm{Im}}) \in (\mathbb{R}_+^*)^2$, $\boldsymbol{\beta}_{o,j} = (\beta_{o,j}^{\mathrm{Re}}, \beta_{o,j}^{\mathrm{Im}}) \in (\mathbb{R}_+^*)^2$, $\boldsymbol{\mu} = (\mu^{\mathrm{Re}}, \mu^{\mathrm{Im}}) \in \mathbb{R}^2$ and $\boldsymbol{\sigma} = (\sigma^{\mathrm{Re}}, \sigma^{\mathrm{Im}}) \in \mathbb{R}_+^2$ are unknown and need to be estimated.

4.3.3.4 Optimization procedure

Although $\mathcal{J}_{\mathrm{WT}}$ is convex, its optimization cannot be performed by conventional descent algorithms like the pseudo-conjugate gradient because $\mathcal{J}_P$ is not differentiable even if $\mathcal{J}_L$ is differentiable with a Lipschitz-continuous gradient. This difficulty is frequently encountered in inverse problems involving sparsity promoting priors [Harikumar and Bresler, 1996; Tropp, 2006; Elad et al., 2007a; Zymnis et al., 2009]. Therefore, we propose to apply a generalized form of the iterative optimization procedure developed in [Daubechies et al., 2004; Chaux et al., 2007], which is based on the Forward-Backward (FB) algorithm.

The minimization of $\mathcal{J}_{\mathrm{WT}}$ is performed by resorting to the concept of proximity operators [Moreau, 1965] introduced in Chapter 3 (see Definition 3.3.1), which was found to be fruitful in a number of recent works in convex optimization [Chaux et al., 2007; Combettes and Wajs, 2005; Combettes and Pesquet, 2010]. In our context, and like in many other signal/image processing problems [Li and Adali, 2008; Calhoun et al., 2002], the observed data are complex-valued. For this reason, we generalize the definition of proximity operators to a class of convex functions defined for complex-valued variables.

For the function

$$\begin{aligned} \Phi\colon \mathbb{C}^K &\to]-\infty,+\infty] \\ x &\mapsto \phi^{\mathrm{Re}}(\mathrm{Re}(x)) + \phi^{\mathrm{Im}}(\mathrm{Im}(x)) \end{aligned} \tag{4.12}$$

where ϕ^{Re} and ϕ^{Im} are functions in $\Gamma_0(\mathbb{R}^K)$ and $\mathrm{Re}(x)$ (resp. $\mathrm{Im}(x)$) is the vector of the real parts (resp. imaginary parts) of the components of $x \in \mathbb{C}^K$, the proximity operator is defined as

$$\begin{aligned} \mathrm{prox}_{\Phi}\colon \mathbb{C}^K &\to \mathbb{C}^K \\ x &\mapsto \mathrm{prox}_{\phi^{\mathrm{Re}}}(\mathrm{Re}(x)) + \imath \mathrm{prox}_{\phi^{\mathrm{Im}}}(\mathrm{Im}(x)). \end{aligned} \tag{4.13}$$

An example of proximity operator for a function of a complex-valued variable is given below.

Example 4.3.1
Consider the following function:

$$\begin{aligned} \Phi \ : \ &\mathbb{C} \to \mathbb{R} \\ &\xi \mapsto \alpha^{\mathrm{Re}}|\mathrm{Re}(\xi-\mu)| + \frac{\beta^{\mathrm{Re}}}{2}\big(\mathrm{Re}(\xi-\mu)\big)^2 \\ &\quad + \alpha^{\mathrm{Im}}|\mathrm{Im}(\xi-\mu)| + \frac{\beta^{\mathrm{Im}}}{2}\big(\mathrm{Im}(\xi-\mu)\big)^2 \end{aligned} \tag{4.14}$$

with $(\alpha^{\mathrm{Re}}, \alpha^{\mathrm{Im}}) \in (\mathbb{R}_+)^2$, $(\beta^{\mathrm{Re}}, \beta^{\mathrm{Im}}) \in (\mathbb{R}_+^*)^2$ and $\mu \in \mathbb{C}$. It can be easily proved that the associated proximity operator reads:

$$\begin{aligned} \mathrm{prox}_{\Phi}\xi = &\frac{\mathrm{sign}(\mathrm{Re}(\xi-\mu))}{\beta^{\mathrm{Re}}+1}\max\{|\mathrm{Re}(\xi-\mu)| - \alpha^{\mathrm{Re}}, 0\} \\ &+ \imath \frac{\mathrm{sign}(\mathrm{Im}(\xi-\mu))}{\beta^{\mathrm{Im}}+1}\max\{|\mathrm{Im}(\xi-\mu)| - \alpha^{\mathrm{Im}}, 0\} + \mu. \end{aligned} \tag{4.15}$$

By extending the algorithm in [Chaux et al., 2007] to the complex case, a minimizer of $\mathcal{J}_{\mathrm{WT}}$ can be iteratively computed according to Algorithm 6 where the gradient of $\mathcal{J}_{\mathrm{L}}$ is first calculated, and then the frame coefficients are updated. Note that in this algorithm, the expressions of $\mathrm{prox}_{\gamma_n \Phi_a}$ and $\mathrm{prox}_{\gamma_n \Phi_{o,j}}$ at each iteration n are provided by Eq. (4.15) associated to Example 4.3.1. It can also be noticed that λ_n and γ_n correspond to relaxation and step-size parameters, respectively.

Again, we iterate Algorithm 6 over all slices to perform 3D volume reconstruction as required for anatomical data. The proposed reconstruction procedure in Algorithm 6 is also called the UWR-SENSE reconstruction method. Regarding functional data, we also iterate over volumes separately to get the reconstructed series of EPI full FOV volumes. Interestingly, Algorithm 6 allows a parallelization of the computation of the solution over resolution levels since the proximity operators can be computed separately.

Algorithm 6 UWR-SENSE: 2D-slice wavelet-based regularized reconstruction

```
Let $(\gamma_n)_{n>0}$ and $(\lambda_n)_{n>0}$ be sequences of positive reals.
1: Set $n = 0$ and $\varepsilon \geq 0$. Initialize $\zeta^{(n)}$ and set $\mathcal{J}^{(n)} = \mathcal{J}_{\mathrm{WT}}(\zeta^{(n)})$.
2: repeat
3:    Reconstruct the image by setting $\rho^{(n)} = T^*\zeta^{(n)}$.
4:    Compute the image $u^{(n)}$ such that:
      $\forall \mathbf{r} \in \{1, \dots, Y/R\} \times \{1, \dots, X\}$,
      $\boldsymbol{u}^{(n)}(\mathbf{r}) = 2\boldsymbol{S}^{\mathsf{H}}(\mathbf{r})\boldsymbol{\Psi}^{-1}\big(\boldsymbol{S}(\mathbf{r})\boldsymbol{\rho}^{(n)}(\mathbf{r}) - \boldsymbol{d}(\mathbf{r})\big)$,
      where the vector $\boldsymbol{u}^{(n)}(\mathbf{r})$ is defined from $u^{(n)}$ in the same way as $\overline{\boldsymbol{\rho}}(\mathbf{r})$ is defined from
      $\overline{\rho}$ (see Eq. (2.17)).
5:    Determine the wavelet coefficients $\upsilon^{(n)} = Tu^{(n)} = \big(\boldsymbol{\upsilon}_a, (\boldsymbol{\upsilon}_{o,j})_{o\in\mathbb{O}, 1\leq j\leq j_{\max}}\big)$ of $u^{(n)}$.
6:    Update the approximation coefficients of the reconstructed image $\rho^{(n+1)}$:
      $\forall k \in \{1, \dots, K_{j_{\max}}\}$, $\zeta_{a,k}^{(n+1)} = \zeta_{a,k}^{(n)} + \lambda_n\Big(\mathrm{prox}_{\gamma_n\Phi_a}(\zeta_{a,k}^{(n)} - \gamma_n\upsilon_{a,k}^{(n)}) - \zeta_{a,k}^{(n)}\Big)$.
7:    Update the detail coefficients of the reconstructed image $\rho^{(n+1)}$:
        $\forall o \in \mathbb{O}, \forall j \in \{1, \dots, j_{\max}\}, \forall k \in \{1, \dots, K_j\}$,
            $\zeta_{o,j,k}^{(n+1)} = \zeta_{o,j,k}^{(n)} + \lambda_n\Big(\mathrm{prox}_{\gamma_n\Phi_{o,j}}(\zeta_{o,j,k}^{(n)} - \gamma_n\upsilon_{o,j,k}^{(n)}) - \zeta_{o,j,k}^{(n)}\Big)$.
8:    Compute $\mathcal{J}^{(n+1)} = \mathcal{J}_{\mathrm{WT}}(\zeta^{(n+1)})$.
9:    $n \leftarrow n + 1$
10: until $|\mathcal{J}^{(n)} - \mathcal{J}^{(n-1)}| \leq \varepsilon\mathcal{J}^{(n-1)}$
11: return $\rho^{(n)} = T^*\zeta^{(n)}$
```

4.3.3.5 Convergence of Algorithm 6

For every $\mathbf{r} \in \{1, \dots, Y/R\} \times \{1, \dots, X\}$, let $\theta_{\mathbf{r}} \geq 0$ be the maximum eigenvalue of the Hermitian positive semi-definite matrix $\boldsymbol{S}^{\mathsf{H}}(\mathbf{r})\boldsymbol{\Psi}^{-1}\boldsymbol{S}(\mathbf{r})$ and let $\theta = \max_{\mathbf{r}\in\{1,\dots,Y/R\}\times\{1,\dots,X\}} \theta_{\mathbf{r}} > 0$. To guarantee the convergence of Algorithm 6, the step-size and relaxation parameters have to meet the following conditions:

Assumption 4.3.2

(i) $\inf_{n>0} \gamma_n > 0$ and $\sup_{n>0} \gamma_n < \frac{1}{\theta\|T\|^2}$,

(ii) $\inf_{n>0} \lambda_n > 0$ and $\sup_{n>0} \lambda_n \leq 1$.

More precisely, the following result can be shown:

Proposition 4.3.3 *Under Assumption 4.3.2, the sequence* $(\zeta^{(n)})_{n>0}$ *generated when iterating Steps 3 to 9 of Algorithm 6 converges linearly to the unique minimizer* $\widehat{\zeta}$ *of* $\mathcal{J}_{\mathrm{WT}}$.

Proof :

From Eqs. (4.9)-(4.11), it can be seen that $\mathcal{J}_{\mathrm{P}}$ is a convex function such that

$$\forall \zeta \in \mathbb{C}^K, \quad \mathcal{J}_P(\zeta) \geq \frac{\vartheta_1}{2}\|\zeta\|^2 - \vartheta_0$$

where

$$\begin{aligned}\vartheta_0 &= \frac{K_{j_{\max}}}{2}\Big(\big(\frac{\mu^{\mathrm{Re}}}{\sigma^{\mathrm{Re}}}\big)^2 + \big(\frac{\mu^{\mathrm{Im}}}{\sigma^{\mathrm{Im}}}\big)^2\Big)\\ \vartheta_1 &= \min\{(\sqrt{2}\sigma^{\mathrm{Re}})^{-2}, (\sqrt{2}\sigma^{\mathrm{Im}})^{-2}, (\beta^{\mathrm{Re}}_{o,j})_{o\in\mathbb{O},1\le j\le j_{\max}}, (\beta^{\mathrm{Im}}_{o,j})_{o\in\mathbb{O},1\le j\le j_{\max}}\}.\end{aligned}$$

This means that $\mathcal{J}_{\mathrm{P}}$ is a strongly convex function with modulus ϑ_1 [2]. Since $\mathcal{J}_{\mathrm{L}}$ is a finite convex function, $\mathcal{J}_{\mathrm{WT}}$ also is strongly convex. It is thus strictly convex and coercive (i.e. $\lim_{\|\zeta\|\to+\infty}\mathcal{J}_{\mathrm{WT}}(\zeta) = +\infty$) and, from standard result in convex analysis [Rockafellar, 1970; Ekeland and Témam, 1999], it can be deduced that $\mathcal{J}_{\mathrm{WT}}$ has a unique minimizer $\widehat{\zeta}$.

In addition, $\mathcal{J}_{\mathrm{L}}$ is a differentiable function and we have

$$\forall\zeta\in\mathbb{C}^K,\qquad \nabla\mathcal{J}_{LT}(\zeta) = \frac{\partial\mathcal{J}_{LT}(\zeta)}{\partial\mathrm{Re}(\zeta)} + \imath\frac{\partial\mathcal{J}_{LT}(\zeta)}{\partial\mathrm{Im}(\zeta)} = T\,\nabla\mathcal{J}_{\mathrm{L}}(T^*\zeta), \tag{4.16}$$

where $\mathcal{J}_{\mathrm{LT}}(\zeta) = \mathcal{J}_{\mathrm{L}}(T^*\zeta)$. Set $\rho = T^*\zeta$ and $u = \nabla\mathcal{J}_{\mathrm{L}}(\rho)$. It can be then deduced from Eq. (4.6) that

$$\forall\mathbf{r}\in\{1,\dots,Y/R\}\times\{1,\dots,X\},\qquad \boldsymbol{u}(\mathbf{r}) = 2\boldsymbol{S}^{\mathsf{H}}(\mathbf{r})\boldsymbol{\Psi}^{-1}\left(\boldsymbol{S}(\mathbf{r})\boldsymbol{\rho}(\mathbf{r}) - \boldsymbol{d}(\mathbf{r})\right).$$

where the vector $\boldsymbol{u}(\mathbf{r})$ is defined from u in the same way as $\overline{\boldsymbol{\rho}}(\mathbf{r})$ is defined from $\overline{\rho}$ in Eq. (2.17). Furthermore, for every $\zeta'\in\mathbb{C}^K$,

$$\|\nabla\mathcal{J}_{\mathrm{LT}}(\zeta) - \nabla\mathcal{J}_{\mathrm{LT}}(\zeta')\| \le \|T\|\|u-u'\| \tag{4.17}$$

where $u' = \nabla\mathcal{J}_{\mathrm{L}}(\rho')$ and $\rho' = T^*\zeta'$. We have then

$$\begin{aligned}\|u-u'\|^2 &= \sum_{\mathbf{r}\in\{1,\dots,Y/R\}\times\{1,\dots,X\}} \|\boldsymbol{u}(\mathbf{r}) - \boldsymbol{u}'(\mathbf{r})\|^2\\ &= 4\sum_{\mathbf{r}\in\{1,\dots,Y/R\}\times\{1,\dots,X\}} \|\boldsymbol{S}^{\mathsf{H}}(\mathbf{r})\boldsymbol{\Psi}^{-1}\boldsymbol{S}(\mathbf{r})\left(\boldsymbol{\rho}(\mathbf{r}) - \boldsymbol{\rho}'(\mathbf{r})\right)\|^2\\ &\le 4\sum_{\mathbf{r}\in\{1,\dots,Y/R\}\times\{1,\dots,X\}} \theta_{\mathbf{r}}^2\|\boldsymbol{\rho}(\mathbf{r}) - \boldsymbol{\rho}'(\mathbf{r})\|^2\\ &\le 4\theta^2\|\rho-\rho'\|^2\\ &\le 4\theta^2\|T\|^2\|\zeta-\zeta'\|^2.\end{aligned} \tag{4.18}$$

Altogether, Eq. (4.17) and Eq. (4.18) yield

$$\|\nabla\mathcal{J}_{\mathrm{LT}}(\zeta) - \nabla\mathcal{J}_{\mathrm{LT}}(\zeta')\| \le 2\theta\|T\|^2\|\zeta-\zeta'\| \tag{4.19}$$

which shows that $\mathcal{J}_{\mathrm{LT}}$ has a Lipschitz continuous gradient with constant $2\theta\|T\|^2$.

Based on these observations and the fact that,

$$\begin{aligned}&\forall\zeta = \big((\zeta_{a,k})_{1\le k\le K_{j_{\max}}}, (\zeta_{o,j,k})_{1\le j\le j_{\max},1\le k\le K_j}\big),\\ &\qquad\mathrm{prox}_{\gamma_n\mathcal{J}_{\mathrm{P}}}\zeta = \big((\mathrm{prox}_{\gamma_n\Phi_a}\zeta_{a,k})_{1\le k\le K_{j_{\max}}}, (\mathrm{prox}_{\gamma_n\Phi_{o,j}}\zeta_{o,j,k})_{1\le j\le j_{\max},1\le k\le K_j}\big),\end{aligned} \tag{4.20}$$

---

[2] A function $f\colon\chi\to]-\infty,+\infty]$, where χ is a Hilbert space, is said *strongly convex* on χ with modulus $\vartheta_1>0$ if there exists some $g\in\Gamma_0(\chi)$ such that $f = g + \frac{\vartheta_1\|\cdot\|^2}{2}$.

the sequence $(\zeta^{(n)})_{n>0}$ built by Algorithm 6 can be rewritten under the more classical Forward-Backward iterative form [Chaux et al., 2007]:

$$\zeta^{(n+1)} = \zeta^{(n)} + \lambda_n \Big(\mathrm{prox}_{\gamma_n \mathcal{J}_\mathrm{P}} \big(\zeta^{(n)} - \gamma_n \nabla \mathcal{J}_\mathrm{LT}(\zeta^{(n)}) \big) - \zeta^{(n)} \Big) \tag{4.21}$$

and, due to the Lipschitz differentiability of $\mathcal{J}_\mathrm{LT}$, the convergence of the algorithm is guaranteed under Assumption 4.3.2 (see [Chaux et al., 2007; Combettes and Wajs, 2005]). Furthermore, since $\mathcal{J}_\mathrm{P}$ is strongly convex with modulus ϑ_1, we have (see [Chaux et al., 2009] and references therein)

$$\forall n > 0, \qquad \|\zeta^{(n)} - \widehat{\zeta}\| \leq \Big(1 - \frac{\underline{\lambda}\underline{\gamma}\vartheta_1}{1 + \underline{\gamma}\vartheta_1} \Big)^{n-1} \|\zeta^{(1)} - \widehat{\zeta}\| \tag{4.22}$$

where $\underline{\gamma} = \inf_{n>0} \gamma_n$ and $\underline{\lambda} = \inf_{n>0} \lambda_n$. This proves that $(\zeta^{(n)})_{n>0}$ converges linearly to $\widehat{\zeta}$. ■

The results we obtained using Algorithm 6 are discussed in Section 4.3.3.6.

4.3.3.6 Reconstruction results

Results on anatomical data

In this experiment, dyadic ($M = 2$) *Symmlet* orthonormal wavelet bases [Daubechies, 1992] associated with filters of length 8 were used over $j_\mathrm{max} = 3$ resolution levels. Regarding the wavelet coefficients, the prior described in Subsection 4.3.3.3 has been employed. The related hyper-parameters [3] were estimated using the Bayesian approach described in Chapter 5. Full FOV image reconstruction was then performed using our UWR-SENSE method. For the sake of simplicity, constant values of relaxation and step-size parameters (λ_n and γ_n, respectively) have been adopted along the algorithm iterations and will be denoted by λ and γ: first, we experimentally observed that $\lambda \equiv 1$ is the best value of the relaxation parameter among those tested in terms of convergence rate (see Fig. 4.6). Second, a step-size parameter γ close to the allowed maximum value in Assumption 4.3.2 provided the fastest convergence rate. After computing the constant θ related to the considered sensivity map, γ was thus chosen equal to $0.99/\theta = 12.83$.

The algorithm was stopped when $\mathcal{J}_\mathrm{WT}$ no longer significantly varied, by choosing $\varepsilon = 10^{-4}$ in Algorithm 6. For different values of λ, Fig. 4.6 illustrates the evolution of the optimized criterion $\mathcal{J}_\mathrm{WT}$ w.r.t. the iteration number for a 2D-slice reconstruction.

In Fig. 4.6, it is emphasized that after about 20 iterations the minimizer $\widehat{\zeta}^\mathrm{MAP}$ was reached. In terms of computation time, the approach was implemented using the C language and took 6 seconds when running on an Intel Core 2 (2.26 GHz) architecture. Note that accelerated algorithms such as TWIST or FISTA have been proposed in the recent convex optimization literature [Bioucas-Dias and Figueiredo, 2007; Beck and Teboulle, 2009] for minimizing the same optimality criterion. In Fig. 4.7, we compare these algorithms and their convergence speed on the pMRI application. With respect to the number of iterations required to achieve convergence, we did not observe any improvement in terms of convergence speed using FISTA instead of the proposed FB algorithm. Also, our FB implementation as well as FISTA perform faster than TWIST.

---

[3] A couple of hyper-parameters is fitted for real/imaginary parts of each subband, i.e. each approximation/detail coefficients at each resolution level and orientation.

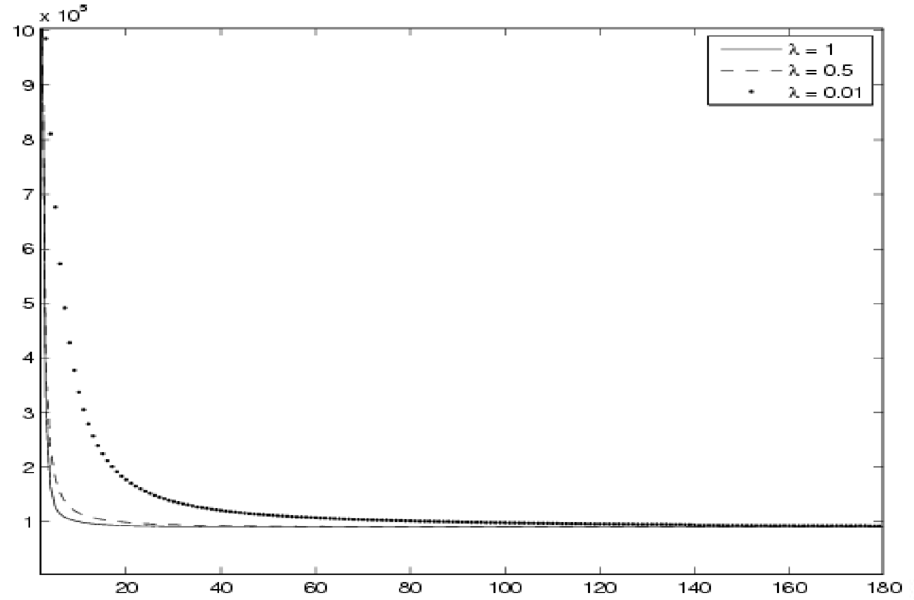


Figure 4.6: Convergence speed of the optimization algorithm w.r.t. the choice of the relaxation parameter λ for $j_{\max} = 3$.

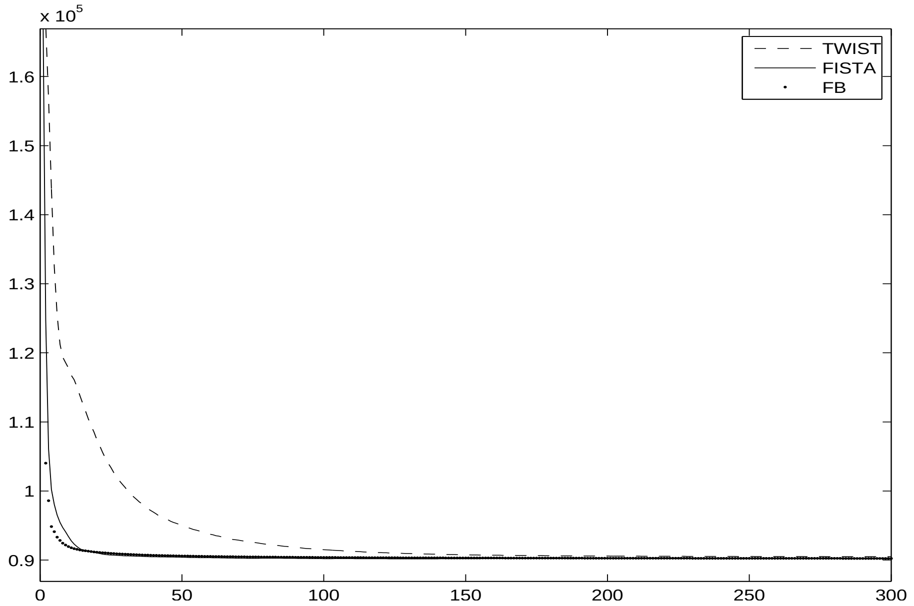


Figure 4.7: Convergence speed comparison for the TWIST, FISTA and FB algorithms.

Since our approach is designed to reconstruct 2D slices, we should mention here that Algorithm 6 behaves similarly for all slices.

Figs. 4.8 and 4.9 shows reconstructed full FOV anatomical images using the proposed approach (UWR-SENSE) with $R = 2$ and $R = 4$.

The smoothing effects observed in Figs. 4.1 and 4.2 with Tikhonov regularization no longer exist in the WT regularized images in Figs. 4.8 and 4.9, where a quite accurate reconstruction is performed within the brain mask without introducing staircase effects

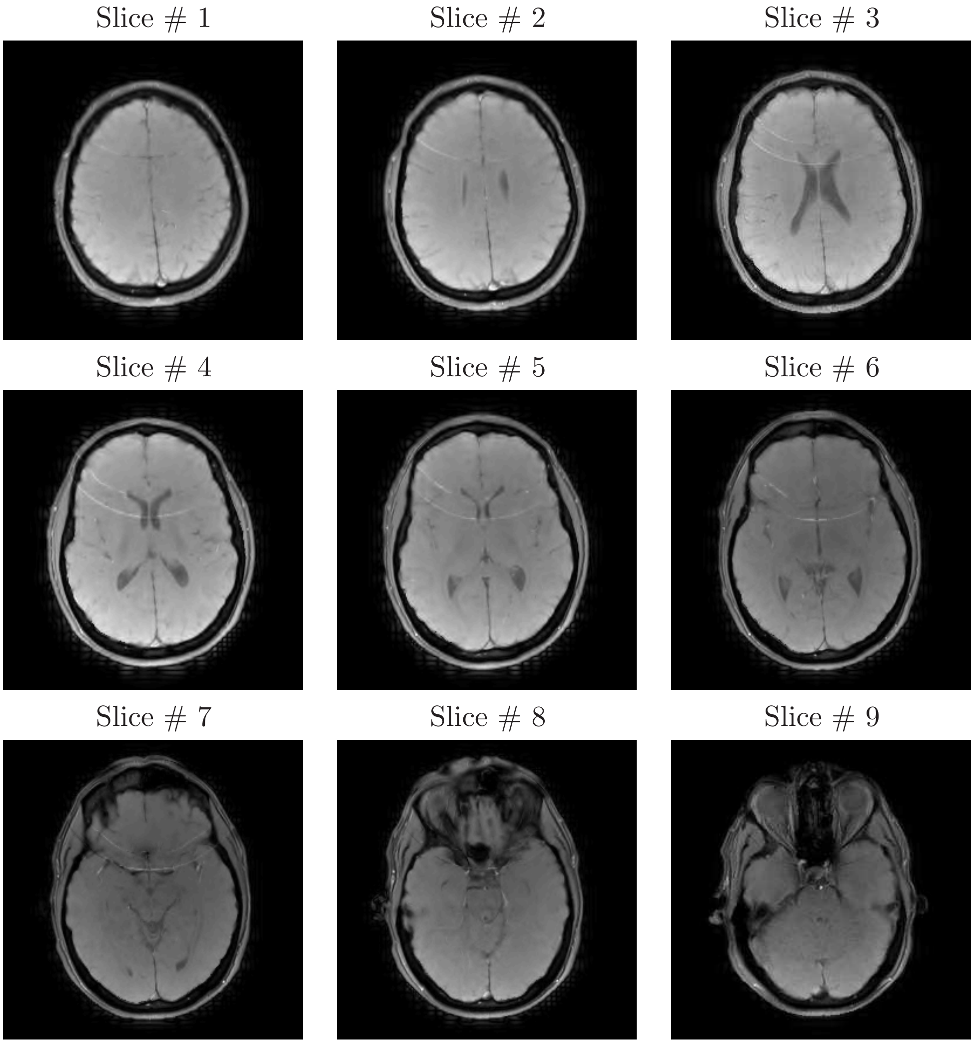


Figure 4.8: Reconstructed anatomical slices using UWR-SENSE for $R = 2$.

as in the case of TV regularization (see Fig. 4.4). Nevertheless, some strong aliasing artifacts still exist in the reconstructed images using the UWR-SENSE algorithms. A more sophisticated reconstruction approach will be developed in Section 4.3.5 to handle these remaining artifacts.

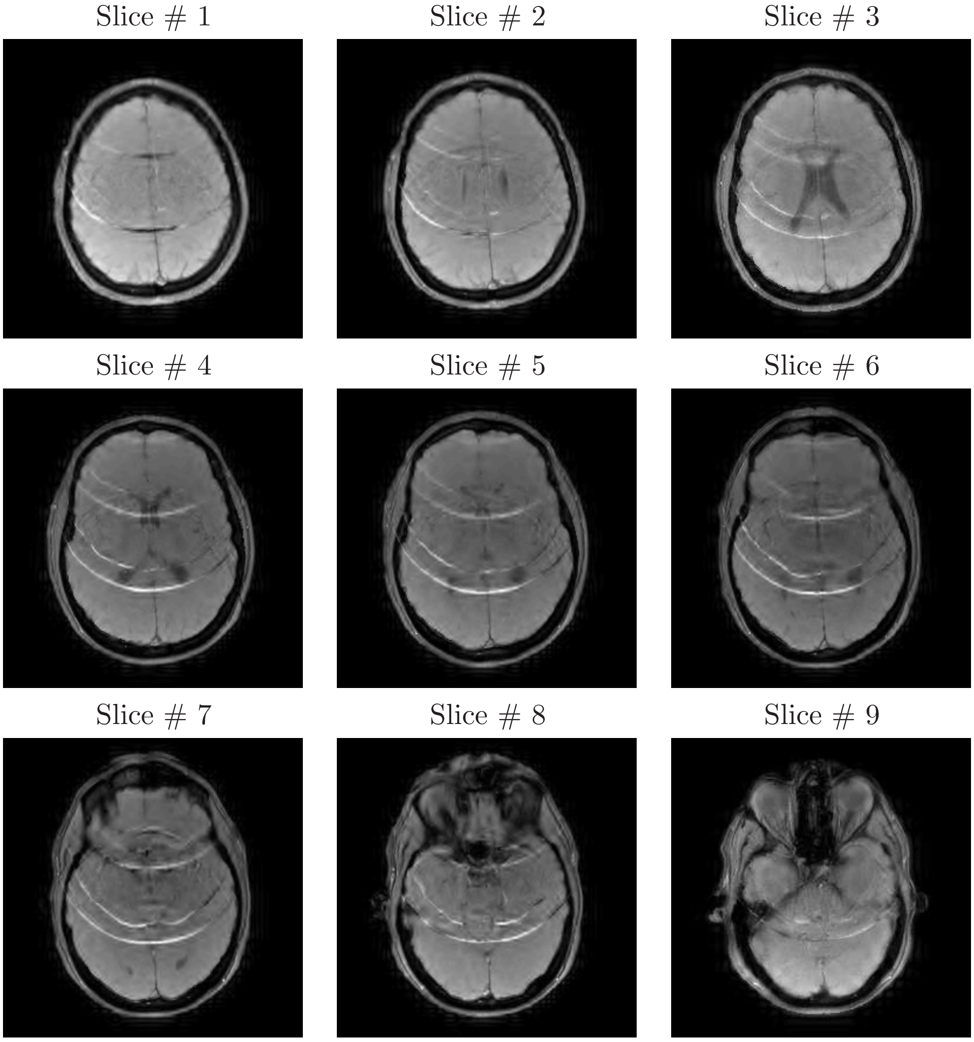


Figure 4.9: Rreconstructed anatomical slices using UWR-SENSE for $R = 4$.

Results on functional EPI data

This experiment on functional 64×64 EPI data was conducted using the same wavelet basis and priors. Algorithm parameters (i.e. relaxation λ and step-size γ parameters) were adjusted according to the same rules as those for anatomical data: $\lambda = 1$ and $\gamma = 0.99/\theta = 20.63$ were chosen after deriving the θ constant. Again, these EPI fMRI data were acquired during a short resting state session of 1 min 12 s. Fig. 4.10 illustrates two reconstructed full FOV slices using our UWR-SENSE algorithm.

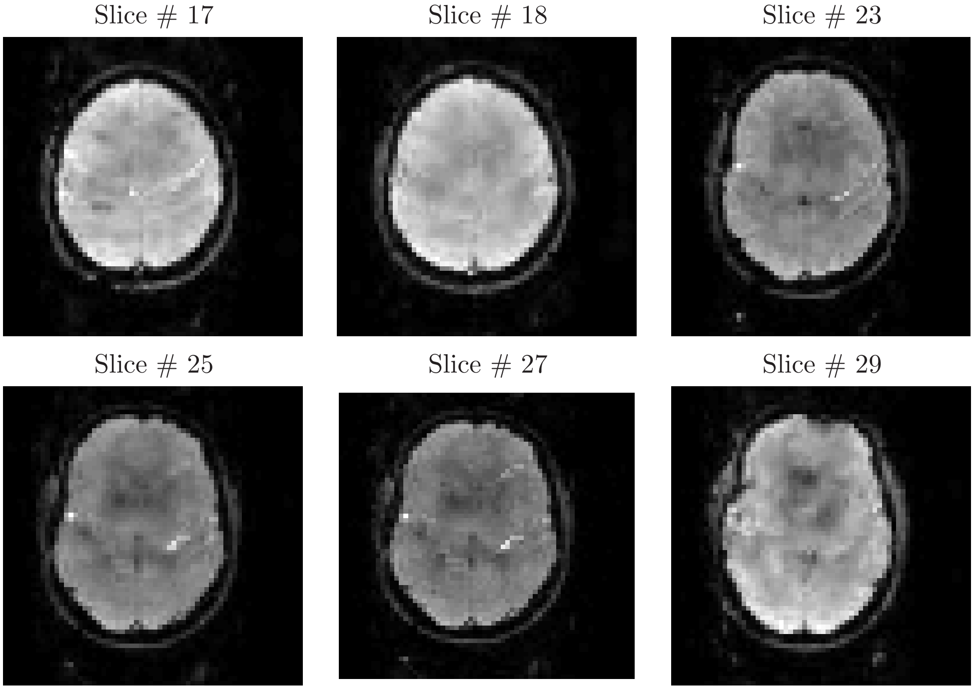


Figure 4.10: Six EPI reconstructed slices using our UWR-SENSE algorithm for $R = 4$.

It can be shown that many defective pixels (i.e. which have locally very high or very low intensity level) were corrected when using the proposed WT regularization in UWR-SENSE without introducing additional artifacts, in contrast to Tikhonov regularization.

Choice of the maximum resolution level

In this paragraph we focus on the effect of the choice of the maximum resolution level $j_{\max}$ in terms of reconstruction quality. The impact on reconstructed full FOV images can be emphasized through the difference between reconstructed images using 1, 2, 3 and 4 resolution levels. Fig. 4.11 illustrates the difference between anatomical reconstructed images using one and two (left), two and three (middle), or three and four (right) resolution levels.

The difference between anatomical reconstructed images at different resolution levels is significant since it can reach a value of 40 within the intensity range $[0, 255]$ (around 15%). Moreover, the intensity difference seems correlated to the presence of distortions since it

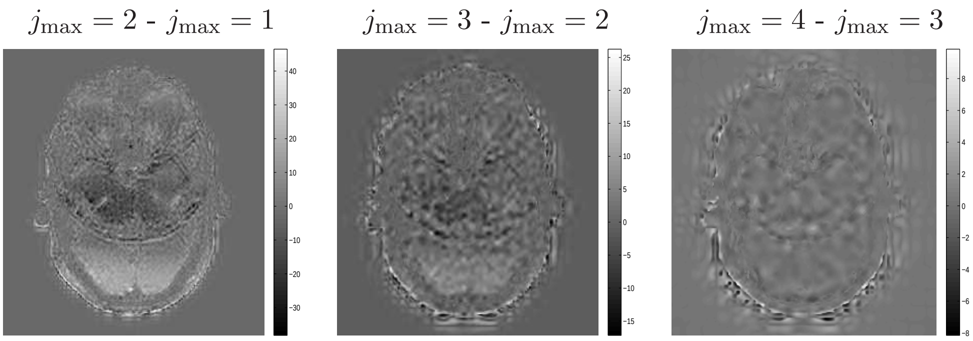

$j_{\max} = 2$ - $j_{\max} = 1$ $j_{\max} = 3$ - $j_{\max} = 2$ $j_{\max} = 4$ - $j_{\max} = 3$

Figure 4.11: Influence of the number of resolution levels on the anatomical reconstructed images using our CWR-SENSE method for $R = 4$; from left to right: reconstructed images with $j_{\max} = 2$ - $j_{\max} = 1$, $j_{\max} = 3$ - $j_{\max} = 2$ and $j_{\max} = 4$ - $j_{\max} = 3$.

appears particularly important in distorted areas. Hence, the higher the maximum resolution level, the better regularized the artifacts are. However, only slight improvements are obtained beyond three resolution levels. Note also that by increasing the number of resolution levels, boundary effects become more visible but they do not affect the brain region. Clearly, $j_{\max} = 3$ appears as a fair compromise to achieve an acceptable reconstruction quality. Similar tests were conducted on EPI images sized 64×64 (see Fig. 4.12), and led to the same conclusions.

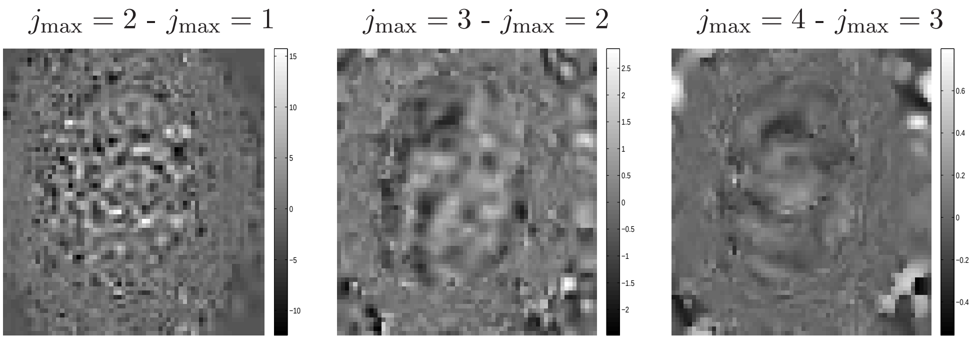

$j_{\max} = 2$ - $j_{\max} = 1$ $j_{\max} = 3$ - $j_{\max} = 2$ $j_{\max} = 4$ - $j_{\max} = 3$

Figure 4.12: Influence of the number of resolution levels on the EPI reconstructed images using our CWR-SENSE method for $R = 4$; from left to right: reconstructed images with $j_{\max} = 2$ - $j_{\max} = 1$, $j_{\max} = 3$ - $j_{\max} = 2$ and $j_{\max} = 4$ - $j_{\max} = 3$.

4.3.4 Regularization using bivariate wavelet prior

In the proposed approach, as well as in any wavelet regularization problem, choosing the prior model for the wavelet coefficients at each resolution level and orientation is a key issue. As our signal is complex-valued as well as its wavelet coefficients, the simplest way to model it is, as we considered in Section 4.3.3.2, to assume that the real and imaginary parts are mutually independent. Despite its simplicity, this model does not account for the statistical dependence between the real and imaginary parts. In fact, kernel statistical tests of independence [Gretton et al., 2007] applied on our experimental data sets indicate that the hypothesis "the real and imaginary parts are independent" is rejected with an error level less than 2%. In this section, we will use a more accurate modelling by considering the *joint* distribution of the two parts through an appropriate bivariate prior probability density function (pdf). From the examination of the joint empirical histograms (see Fig. 4.13 below), it seems more appropriate to consider the following bivariate pdf form:

$$\forall \xi \in \mathbb{C}, \quad f_{\alpha,\beta,\gamma,p}(\xi) = C_f e^{-(\alpha|\mathrm{Re}(\xi)|+\beta|\mathrm{Im}(\xi)|+\gamma|\xi|^p)} \tag{4.23}$$

where $C_f \in \mathbb{R}_+^*$ is a normalization constant, and where the hyperparameters α, β, γ belong to $\mathbb{R}_+^*$ and $p \in [1, \infty[$. It can be noticed that the proposed class of pdfs includes those considered in [Sendur and Selesnick, 2002] for denoising purpose as particular cases when $\alpha = \beta$ and $p = 1$.

For illustration, Fig. 4.13 shows the empirical histogram of the coefficients $\boldsymbol{\zeta}_{1,2}$, the independent pdf used in [Chaari et al., 2008] and the adopted bivariate pdf in Eq. (4.23).

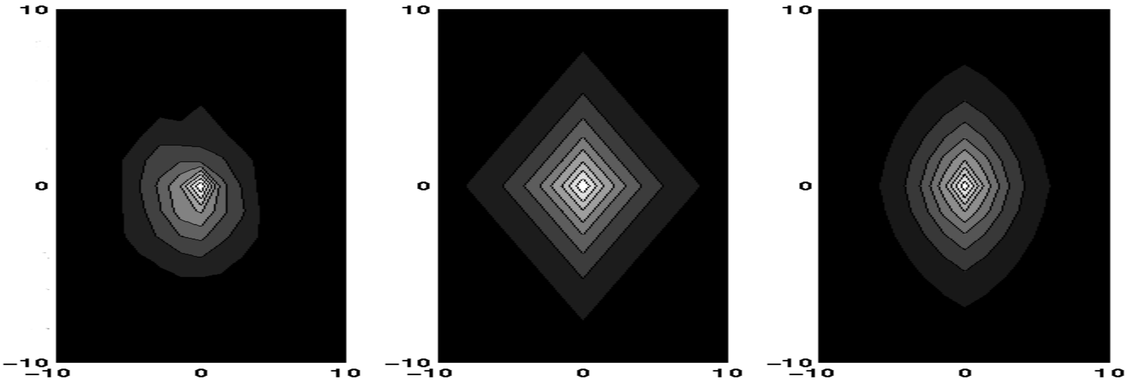


Figure 4.13: Joint 2D empirical histogram $\boldsymbol{\zeta}_{1,2}$ (left) and pdfs of the independent (middle) and proposed bivariate (right) models.

It is clear that, due to the elliptical shape, the bivariate model fits the empirical histogram better than the independent one. For instance, according to the independent model, some pairs of values of $(\mathrm{Re}(\zeta_{1,2,k}), \mathrm{Im}(\zeta_{1,2,k}))$ have an over/under-estimated occurrence frequency w.r.t. the empirical joint histogram and the bivariate model.
Based on this bivariate prior, the new MAP estimator $\widehat{\zeta}^{\mathrm{MAP}}$ can be rewritten as follows:

$$\begin{aligned} \widehat{\zeta}^{\mathrm{MAP}} &= \arg\min_{\zeta \in \mathbb{C}^K} \mathcal{J}_{\mathrm{WTbiv}}(\zeta) \\ &= \arg\min_{\zeta \in \mathbb{C}^K} \big(\mathcal{J}_{\mathrm{L}}(T^*\zeta) + \mathcal{J}_{\mathrm{Pbiv}}(\zeta)\big), \end{aligned} \tag{4.24}$$

$$\text{with} \quad \mathcal{J}_{\mathrm{Pbiv}}(\zeta) = \sum_{k=1}^{K_{j_{\max}}} \Phi_a(\zeta_{a,k}) + \sum_{o \in \mathbb{O}} \sum_{j=1}^{j_{\max}} \sum_{k=1}^{K_j} \Phi_{o,j}(\zeta_{o,j,k}) \tag{4.25}$$

and

$$\begin{aligned} \Phi_a(\zeta_{a,k}) &= \alpha_a|\mathrm{Re}(\xi)| + \beta_a|\mathrm{Im}(\xi)| + \gamma_a|\xi|^{p_a} \\ \Phi_{o,j}(\zeta_{o,j,k}) &= \alpha_{o,j}|\mathrm{Re}(\xi)| + \beta_{o,j}|\mathrm{Im}(\xi)| + \gamma_{o,j}|\xi|^{p_{o,j}}. \end{aligned} \tag{4.26}$$

The proximity operators of Φ_a and $\Phi_{o,j}$ are given by Example 4.3.4.

Example 4.3.4

Consider the following function:

$$\begin{aligned}\Phi\colon \mathbb{C} &\to \mathbb{R} \\ \xi &\mapsto \alpha|\mathrm{Re}(\xi)| + \beta|\mathrm{Im}(\xi)| + \gamma|\xi|^p\end{aligned} \tag{4.27}$$

where $(\alpha, \beta, \gamma) \in (\mathbb{R}_+^*)^3$ and $p \geq 1$.
The associated proximity operator is given by:

$$\mathrm{prox}_\Phi(\xi) = \mathrm{prox}_\varphi\Big(\mathrm{soft}_\alpha(\mathrm{Re}(\xi)) + \imath\, \mathrm{soft}_\beta(\mathrm{Im}(\xi))\Big) \tag{4.28}$$

where $\mathrm{soft}_\tau(\cdot)$ with $\tau \geq 0$ is the real-valued soft-thresholding operator defined in Eq. (3.21). and prox_φ is the proximity operator of $\varphi = \gamma|\cdot|^p$ which is given by:

- if $p = 1$,
$$\forall \xi \in \mathbb{C}, \quad \mathrm{prox}_\varphi(\xi) = \begin{cases} \Big(1 - \dfrac{\gamma}{|\xi|}\Big)\xi & \text{if } |\xi| > \gamma \\ 0 & \text{otherwise} \end{cases} \tag{4.29}$$
- if $p > 1$,
$$\forall \xi \in \mathbb{C}, \quad \mathrm{prox}_\varphi(\xi) = \begin{cases} \Big(1 - \dfrac{\nu(\xi)}{|\xi|}\Big)\xi & \text{if } \xi \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.30}$$
  where $\nu(\xi)$ is the unique non-negative real solution of the equation
$$\nu(\xi) + \big(\nu(\xi)/(\gamma p)\big)^{1/(p-1)} = |\xi|. \tag{4.31}$$

Proof :

Let $\chi = [-\alpha, \alpha] \times [-\beta, \beta]$ be a closed convex subset of $\mathbb{R}^2$. Let also $\varphi \in \Gamma_0(\mathbb{R})$ be the even and nonconstant function defined by:

$$\begin{aligned}\varphi : \mathbb{R} &\to \mathbb{R} \\ a &\mapsto \gamma|a|^p.\end{aligned} \tag{4.32}$$

We recall that the support function of the closed convex subset $\chi \subset \mathbb{R}^2$ is defined by [Nesterov, 2004]:

$$\begin{aligned}\sigma_\chi : \mathbb{R}^2 &\to]-\infty, +\infty] \\ u = (u_1, u_2)^\top &\mapsto \sup_{v \in \chi} \langle v | u \rangle = \alpha|u_1| + \beta|u_2|.\end{aligned} \tag{4.33}$$

If we consider the function Φ defined by:

$$\begin{aligned}\Phi : \mathbb{R}^2 &\to \mathbb{R} \\ u = (u_1, u_2)^\top &\mapsto \alpha|u_1| + \beta|u_2| + \gamma\|u\|^p\end{aligned} \tag{4.34}$$

where $\|\cdot\|$ is the euclidean norm, we have therefore $\forall u \in \mathbb{R}^2,\ \Phi(u) = \sigma_\chi(u) + \varphi(\|u\|)$. Using

[Briceño-Arias and Combettes, 2009, Prop. 2.2], it tuns out that:

$$\forall u \in \mathbb{R}^2,\ \operatorname{prox}_\Phi(u) = \begin{cases} \dfrac{\operatorname{prox}_\varphi d_\chi(u)}{d_\chi(u)}(u - P_\chi(u)) & \text{if } u \notin \chi \\ 0 & \text{if } u \in \chi, \end{cases} \tag{4.35}$$

where $d_\chi(u)$ is the distance from u to χ, and $P_\chi(u)$ is the projection of u onto χ [Briceño-Arias and Combettes, 2009].
If we define $\chi_1 = [-\alpha, \alpha]$ and $\chi_2 = [-\beta, \beta]$, we can write $\forall u = (u_1, u_2)^\intercal \in \mathbb{R}^2,\ P_\chi(u) = (P_{\chi_1}(u_1), P_{\chi_2}(u_2))$ where

$$P_{\chi_1}(u_1) = \begin{cases} \alpha & \text{if } u_1 > \alpha \\ u_1 & \text{if } |u_1| \leq \alpha \\ -\alpha & \text{if } u_1 < -\alpha \end{cases} \tag{4.36}$$

and a similar expression for $P_{\chi_2}(u_2)$. It turns out that

$$u_1 - P_{\chi_1}(u_1) = \begin{cases} u_1 - \alpha & \text{if } u_1 > \alpha \\ 0 & \text{if } |u_1| \leq \alpha \\ u_1 + \alpha & \text{if } u_1 < -\alpha, \end{cases} \tag{4.37}$$

or equivalently $u_1 - P_{\chi_1}(u_1) = \operatorname{soft}_\alpha(u_1)$, and we have similarly $u_2 - P_{\chi_2}(u_2) = \operatorname{soft}_\beta(u_2)$. Consequently, we can write

$$\begin{aligned} \forall u = (u_1, u_2)^\intercal \in \mathbb{R}^2,\ d_\chi(u) &= \sqrt{d^2_{\chi_1}(u_1) + d^2_{\chi_2}(u_2)} \\ &= \sqrt{(u_1 - P_{\chi_1}(u_1))^2 + (u_2 - P_{\chi_2}(u_2))^2} \\ &= \sqrt{\operatorname{soft}^2_\alpha(u_1) + \operatorname{soft}^2_\beta(u_2)}. \end{aligned} \tag{4.38}$$

Eq. (4.35) can therefore be rewritten as:

$$\forall u \in \mathbb{R}^2,\ \operatorname{prox}_\Phi(u) = \begin{cases} \dfrac{\operatorname{prox}_\varphi\left(\sqrt{\operatorname{soft}^2_\alpha(u_1) + \operatorname{soft}^2_\beta(u_2)}\right)}{\sqrt{\operatorname{soft}^2_\alpha(u_1) + \operatorname{soft}^2_\beta(u_2)}}(\operatorname{soft}_\alpha(u_1), \operatorname{soft}_\beta(u_2))^\intercal & \text{if } u \notin \chi \\ 0 & \text{if } u \in \chi. \end{cases} \tag{4.39}$$

Based on the proximity operator form of the function φ given in Eqs. (4.29)-(4.30), we can write:

- if $u \notin \chi$,
  - if $p = 1$,
    - if $\sqrt{\operatorname{soft}^2_\alpha(u_1) + \operatorname{soft}^2_\beta(u_2)} > \gamma$,

      $$\operatorname{prox}_\Phi(u) = \left(1 - \frac{\gamma}{\left(\sqrt{\operatorname{soft}^2_\alpha(u_1) + \operatorname{soft}^2_\beta(u_2)}\right)}\right)(\operatorname{soft}_\alpha(u_1), \operatorname{soft}_\beta(u_2))^\intercal$$

    - if $\sqrt{\operatorname{soft}^2_\alpha(u_1) + \operatorname{soft}^2_\beta(u_2)} \leq \gamma$,

      $$\operatorname{prox}_\Phi(u) = 0$$

  - if $p > 1$,

since $\sqrt{\mathrm{soft}_\alpha^2(u_1)+\mathrm{soft}_\beta^2(u_2)} \neq 0$,

$$\mathrm{prox}_\Phi(u) = \left(1 - \frac{\nu\left(\sqrt{\mathrm{soft}_\alpha^2(u_1)+\mathrm{soft}_\beta^2(u_2)}\right)}{\left(\sqrt{\mathrm{soft}_\alpha^2(u_1)+\mathrm{soft}_\beta^2(u_2)}\right)}\right)(\mathrm{soft}_\alpha(u_1), \mathrm{soft}_\beta(u_2))^\intercal$$

where $\nu\left(\sqrt{\mathrm{soft}_\alpha^2(u_1)+\mathrm{soft}_\beta^2(u_2)}\right)$ is the unique non-negative real solution of Eq. (4.31) with $\xi = \sqrt{\mathrm{soft}_\alpha^2(u_1)+\mathrm{soft}_\beta^2(u_2)}$.

- <u>if $u \in \chi$</u>,

$$\mathrm{prox}_\Phi(u) = 0.$$

In summary, we have

$$\forall u \in \mathbb{R}^2,\ \mathrm{prox}_\Phi(u) = \mathrm{prox}_\varphi\Big(\mathrm{soft}_\alpha(u_1), \mathrm{soft}_\beta(u_2)\Big). \tag{4.40}$$

Note that this result can also be generalized in $\mathbb{R}^N$ with $N \geq 2$.
For the complex-valued cas, the proximity operator in Eq. (4.28) can be easily retrieved by taking $u_1 = \mathrm{Re}(\xi)$ and $u_2 = \mathrm{Im}(\xi)$ in Eq. (4.40) for every $\xi \in \mathbb{C}$.

■

As a consequence, for every $\xi \in \mathbb{C}$ such that $(\mathrm{Re}(\xi), \mathrm{Im}(\xi)) \in [-\alpha, \alpha] \times [-\beta, \beta]$, $\mathrm{prox}_\Phi(\xi) = 0$, which means that prox_Φ is a bivariate *proximal thresholder* [Combettes and J.-C. Pesquet, 2007]. Fig. 4.14 illustrates 3D plots of the moduli of a complex valued number and the associated proximity operator related to the function Φ in Example 4.3.4. The proximity operator plot shows that it behaves as a soft thresholder since close to the origin, a subset of values are set to zero whereas the moduli of the others are attenuated. Note that in Fig. 4.14 the opposite of the moduli are shown (and not the moduli themselves) in order to better illustrate the bivariate soft thresholder behaviour of the bivariate proximity operator. This shows that the proposed prior distribution should be helpful in promoting the sparsity of the wavelet representation of complex-valued data.
Finally, to calculate the MAP estimator $\widehat{\zeta}^{\mathrm{MAP}}$, the optimality criterion $\mathcal{J}_{\mathrm{WTbiv}}$ is optimized using Algorithm 6. Reconstructed anatomical images using Algorithm 6 with the bivariate prior are illustrated in Fig. 4.15.

A comparison with results obtained using the independent prior can be made from Fig. 4.9. Quantitative comparisons between the two prior models were also made based on the Signal-to-Noise Ratio (SNR) computed as follows:

$$\mathrm{SNR} = 20\log_{10}\left(\|\rho_{\mathrm{True}}\|/\|\rho_{\mathrm{True}} - \widehat{\rho}\|\right), \tag{4.41}$$

where ρ_{True} stands for the reference image and $\widehat{\rho}$ is the reconstructed image obtained with any algorithm. Hence, the denominator gives us the amount of residual obtained by any reconstructor. Over the whole data volume (9 slices) for $R = 4$, we noticed no important gain in term of SNR average (12.63 dB and 12.73 dB for the independent and bivariate models, respectively).
Note here that SNR values for reconstructed anatomical images using the UWR-SENSE

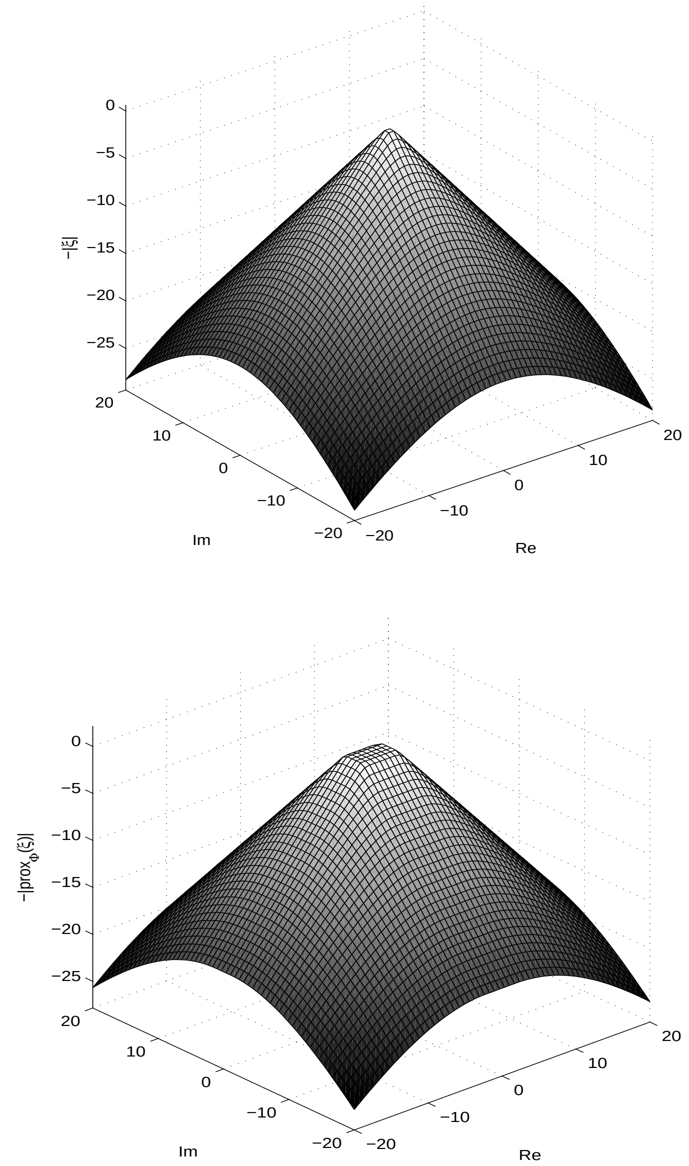
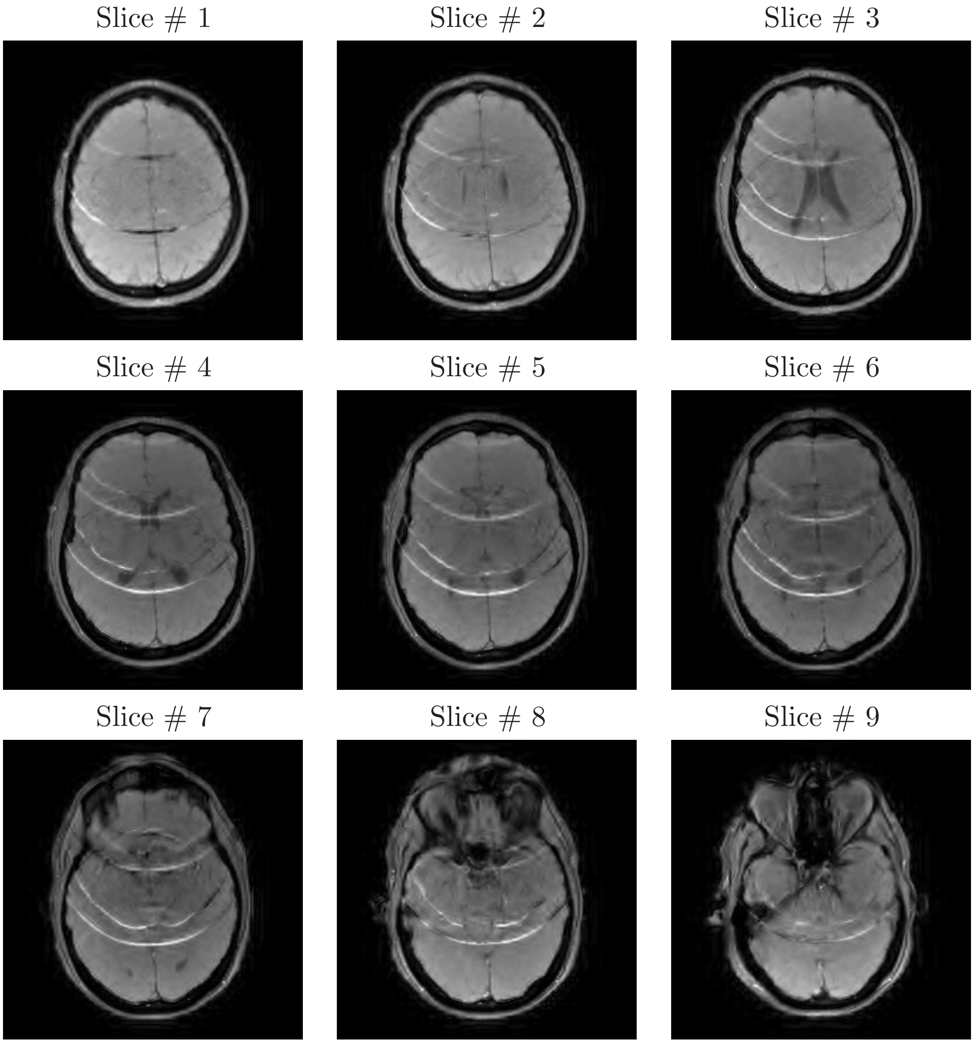


Figure 4.14: Top: 3D plot of the modulus of a complex-valued number for which the real and imaginary parts belong to a uniform 2D grid; bottom: 3D plot of the modulus of the proximity operator associated to Example 4.3.4 with $\alpha = 2$, $\beta = 1$, $\gamma = 0.5$ and $p = 1$, and where the real and imaginary parts belong to the same 2D grid as the top figure.

algorithm with the independent prior are also provided for $R = 4$ in Table 4.1 of Section 4.3.5. However, SNR values for EPI images cannot be calculated since no reference image is available.

Reconstructed EPI images using Algorithm 6 with the bivariate prior are illustrated in Fig. 4.16. Results also seem very similar to the ones obtained using the independent prior.

Based on these comparisons, it appears that no significant improvement of the reconstruction quality is achieved by using the bivariate prior despite its accounting for



Slice # 1
Slice # 2
Slice # 3
Slice # 4
Slice # 5
Slice # 6
Slice # 7
Slice # 8
Slice # 9

Figure 4.15: Reconstructed anatomical slices using UWR-SENSE with the bivariate prior for $R = 4$.

dependencies between real and imaginary parts of the wavelet coefficients. Moreover, using the bivariate prior makes the hyper-parameter estimation step much more complicated in practice than using the independent one since standard estimators such as the ML one are not easy to be deriven.

In conclusion, our experiments show that using the bivariate prior increases the computational complexity of the proposed algorithm without significant gain in reconstruction quality. For these practical reasons, we will keep on using the independent prior in the rest of this manuscript.

4.3.5 Constrained wavelet-based regularization

Since only artifacts of moderate size were removed using Algorithm 6, we will therefore present an extension which accounts for additional constraints leading to a better reconstruction quality and cancellation of more severe artifacts. We propose to extend our

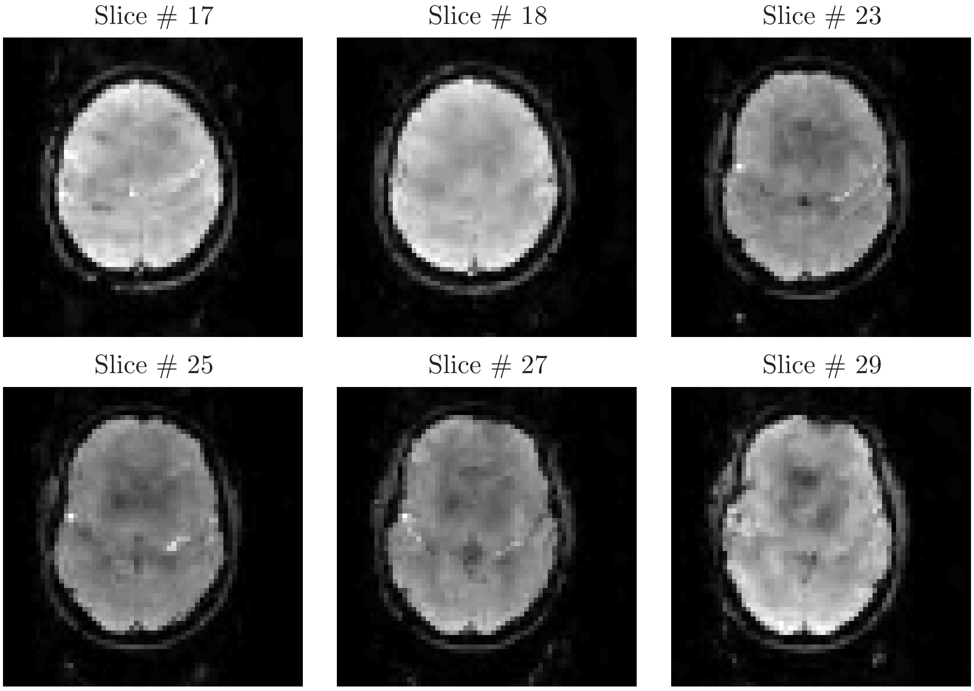


Figure 4.16: Reconstructed EPI slices using UWR-SENSE with the bivariate prior for $R = 4$.

approach by incorporating an additional constraint in the method described hereabove in order to better regularize artifact regions. The ensuing algorithm is called CWR-SENSE method hereafter.

4.3.5.1 New optimality criterion

Here, we set local lower and upper bounds on the image intensity values in artifact areas, regardless of their shape and/or location. These bounds define the nonempty closed convex set:

$$C = \{\rho \in \mathbb{C}^K \mid \forall \boldsymbol{r} \in \{1, \dots, Y/R\} \times \{1, \dots, X\}, \rho(\boldsymbol{r}) \in C_{\boldsymbol{r}}\} \tag{4.42}$$

where the constraint introduced on the range values at position $\boldsymbol{r} \in \{1, \dots, Y/R\} \times \{1, \dots, X\}$ is modelled by:

$$C_{\boldsymbol{r}} = \{\xi \in \mathbb{C} \mid \mathrm{Re}(\xi) \in \mathbb{I}^{\mathrm{Re}}_{\boldsymbol{r}}, \mathrm{Im}(\xi) \in \mathbb{I}^{\mathrm{Im}}_{\boldsymbol{r}}\}, \tag{4.43}$$

with $\mathbb{I}^{\mathrm{Re}}_{\boldsymbol{r}} = [\mathrm{I}^{\mathrm{Re}}_{\mathrm{min},\boldsymbol{r}}, \mathrm{I}^{\mathrm{Re}}_{\mathrm{max},\boldsymbol{r}}]$ and $\mathbb{I}^{\mathrm{Im}}_{\boldsymbol{r}} = [\mathrm{I}^{\mathrm{Im}}_{\mathrm{min},\boldsymbol{r}}, \mathrm{I}^{\mathrm{Im}}_{\mathrm{max},\boldsymbol{r}}]$. When taking into account additional constraints defined in Eq. (4.43), the optimized criterion in Eq. (4.24) becomes:

$$\mathcal{J}_{\mathrm{CWT}}(\zeta) = \mathcal{J}_{\mathrm{WT}}(\zeta) + i_{C^*}(\zeta), \tag{4.44}$$

where

$$C^* = \{\zeta \in \mathbb{C}^K \mid T^*\zeta \in C\}$$

and i_{C^*} is the indicator function of the closed convex set C^* defined by:

$$\forall \zeta \in \mathbb{C}^K, \qquad i_{C^*}(\zeta) = \begin{cases} 0 & \text{if } \zeta \in C^* \\ +\infty & \text{otherwise.} \end{cases}$$

Hence, the constrained MAP estimator satisfies:

$$\widehat{\zeta}^{\text{MAP}} = \arg\min_{\zeta \in C^*} \mathcal{J}_{\text{WT}}(\zeta) = \arg\min_{\zeta \in \mathbb{C}^K} \mathcal{J}_{\text{CWT}}(\zeta) \tag{4.45}$$

An open question consists of choosing the reference image for deriving the convex set C^*. As illustrated in Section 4.3.5.3, the pixelwise constraints are computed on the reconstructed SENSE image for simplicity and numerical reasons. This result is straightforward to obtain and does not require additional numerical steps like hyper-parameter estimation.

4.3.5.2 Computation of the constrained MAP estimator

Theoretically, to solve the minimization problem in Eq. (4.45), the FB iteration has to be updated according to:

$$\zeta^{(n+1)} = \zeta^{(n)} + \lambda_n \Big(\text{prox}_{\gamma_n \mathcal{J}_P + i_{C^*}} \big(\zeta^{(n)} - \gamma_n \nabla \mathcal{J}_{\text{L}}(\zeta^{(n)}) \big) - \zeta^{(n)} \Big). \tag{4.46}$$

The main difficulty here lies in the fact that the proximity operator of $\gamma_n \mathcal{J}_{\text{P}} + i_{C^*}$ does not admit a closed form. However, from its definition, we get:

$$\forall \zeta \in \mathbb{C}^K, \qquad \text{prox}_{\gamma_n \mathcal{J}_{\text{P}} + i_{C^*}}(\zeta) = \arg\min_{\zeta' \in \mathbb{C}^K} \gamma_n \mathcal{J}_{\text{P}}(\zeta') + \mathcal{J}'_{\zeta}(\zeta') \tag{4.47}$$

where

$$\mathcal{J}'_{\zeta}(\cdot) = \frac{1}{2} \| \cdot - \zeta \|^2 + i_{C^*}(\cdot). \tag{4.48}$$

Although $\text{prox}_{\gamma_n \mathcal{J}_{\text{P}} + i_{C^*}}$ cannot be expressed simply, the proximity operator of $\gamma_n \mathcal{J}_{\text{P}}$ is given by Eq. (4.20) and the proximity operator of $\mathcal{J}'_{\zeta}$ is easily determined. In fact, it is quite straightforward to show that:

$$\forall \zeta' \in \mathbb{C}^K, \qquad \text{prox}_{\mathcal{J}'_{\zeta}}(\zeta') = P_{C^*}\left(\frac{\zeta' + \zeta}{2} \right) \tag{4.49}$$

where P_{C^*} is the projection onto the convex set C^*. In turn, provided that the considered wavelet basis is orthonormal, the projection onto C^* of $\zeta' \in \mathbb{C}^K$ is obtained by performing the wavelet decomposition of the projection of $\rho' = T^* \zeta'$ onto C. The latter projection reads:

$$P_C(\rho') = \big(P_{C_{\boldsymbol{r}}}(\rho'(\boldsymbol{r})) \big)_{\boldsymbol{r} \in \{1,\ldots,Y/R\} \times \{1,\ldots,X\}} \tag{4.50}$$

where, for every $\boldsymbol{r} \in \{1, \ldots, Y/R\} \times \{1, \ldots, X\}$,

$$\forall \xi \in \mathbb{C}, \qquad \mathrm{Re}\big(P_{C_{\boldsymbol{r}}}(\xi)\big) = \begin{cases} \mathrm{I}^{\mathrm{Re}}_{\min,\boldsymbol{r}} & \text{if } \mathrm{Re}(\xi) < \mathrm{I}^{\mathrm{Re}}_{\min,\boldsymbol{r}} \\ \mathrm{I}^{\mathrm{Re}}_{\max,\boldsymbol{r}} & \text{if } \mathrm{Re}(\xi) > \mathrm{I}^{\mathrm{Re}}_{\max,\boldsymbol{r}} \\ \xi & \text{otherwise,} \end{cases} \tag{4.51}$$

a similar expression being used to calculate $\mathrm{Im}\big(P_{C_{\boldsymbol{r}}}(\xi)\big)$. Knowing $\mathrm{prox}_{\gamma_n \mathcal{J}_{\mathrm{P}}}$ and $\mathrm{prox}_{\mathcal{J'}_\zeta}$, $\mathrm{prox}_{\gamma_n \mathcal{J}_{\mathrm{P}} + i_{C^*}} \zeta$ can be iteratively computed by solving the optimization problem in Eq. (4.47) using the Douglas-Rachford (DR) algorithm [Chaux et al., 2009; Combettes and Pesquet, 2007]. More precisely, we apply the following proposition:

Proposition 4.3.5
Set $\eta^{(0)} \in \mathbb{C}^K$ *and construct for all* $m \in \mathbb{N}$*:*

$$\begin{cases} \eta^{(m+\frac{1}{2})} = \mathrm{prox}_{\mathcal{J'}_\zeta} \eta^{(m)} \\ \eta^{(m+1)} = \eta^{(m)} + \tau\big(\mathrm{prox}_{\gamma_n \mathcal{J}_{\mathrm{P}}}(2\eta^{(m+\frac{1}{2})} - \eta^{(m)}) - \eta^{(m+\frac{1}{2})}\big), \end{cases} \tag{4.52}$$

where $\tau \in]0, 2[$. *Then,* $(\eta^{(m+\frac{1}{2})})_{m \in \mathbb{N}}$ *converges to* $\mathrm{prox}_{\gamma_n \mathcal{J}_{\mathrm{P}} + i_{C^*}} \zeta$.

Inserting this extra iterative step in the forward-backward algorithm and using the expressions of $\mathrm{prox}_{\gamma_n \mathcal{J}_{\mathrm{P}}}$ and $\mathrm{prox}_{\mathcal{J'}_\zeta}$ in Eqs. (4.49)-(4.51) leads to Algorithm 7 hereafter called the CWR-SENSE method. At iteration n, M_n is the number of times the DR step is run. According to the convergence analysis conducted in [Chaux et al., 2009, Prop. 4.2], if M_n is large enough (e.g. $M_n = 10$) and Assumption 4.3.2 holds, iterating Steps 3 to 25 of the CWR-SENSE method guarantees the convergence to the unique solution of $\mathcal{J}_{\mathrm{CWT}}$. Note however that [Chaux et al., 2009, Prop. 4.2] does not provide a practical guideline for chosing M_n. The practical rule we chose is explained in Section 4.3.3.6. The improvements in terms of accuracy drawn from the use of the CWR-SENSE algorithm are illustrated in the next section, but at the expense of the computation time.

4.3.5.3 Reconstruction results

Our CWR-SENSE constrained algorithm presented in Section 4.3.5 was applied to the anatomical data with the *Symmlet 8* wavelet basis. The parameters λ and γ (since λ_n and γ_n have been fixed through iterations) have been set to the same values as for the UWR-SENSE, while we chose $\tau = 2$ for the underlying Douglas-Rachford iterations as it was practically observed that this value gives the best convergence rate. In practice, the value of M_n was defined as the minimum integer value such that

$$\frac{|\eta^{(n,M_n-1)} - \eta^{(n,M_n-2)}|}{|\eta^{(n,M_n-2)}|} < 10^{-4},$$

which results in about $M_n = 4$ iterations of the Douglas-Rachford algorithm.

A morphological gradient [Serra, 1982] was used to detect artifact regions on which we apply the additional convex constraints. The upper and lowed bounds that define the

Algorithm 7 CWR-SENSE: Constrained 2D-slice wavelet-based regularized reconstruction.

Let $(\gamma_n)_{n>0}$ and $(\lambda_n)_{n>0}$ be sequences of positive reals, let $(M_n)_{n>0}$ be a sequence of positive integers and set $\tau \in]0,2]$.

```
1:  Set $n = 0$ and $\varepsilon \geq 0$. Initialize $\zeta^{(n)}$ and set $\mathcal{J}^{(n)} = \mathcal{J}_{\mathrm{WT}}(\zeta^{(n)})$.
2:  repeat
3:      Same as for Algorithm 6
4:      Same as for Algorithm 6
5:      Same as for Algorithm 6
6:      Initialize the Douglas-Rachford algorithm by setting
        $\eta^{(n,0)} = \zeta^{(n)} - \gamma_n \upsilon^{(n)}$.
7:      Douglas-Rachford iterations:
8:      for $m = 0$ to $M_n - 1$ do
9:          Compute $\eta^{(n,m+\frac{1}{2})} = P_{C^*}\Big(\dfrac{\eta^{(n,m)} + \zeta^{(n)}}{2}\Big)$;
10:         Update the approximation components of $\eta^{(n,m)}$:
            $\forall k \in \{1,\ldots,K_{j_{\max}}\}$,
            $\eta_{a,k}^{(n,m+1)} = \eta_{a,k}^{(n,m)} + \tau\big(\mathrm{prox}_{\gamma_n \Phi_a}(2\eta_{a,k}^{(n,m+\frac{1}{2})} - \eta_{a,k}^{(n,m)}) - \eta_{a,k}^{(n,m+\frac{1}{2})}\big)$,
11:         Update the detail components of $\eta^{(n,m)}$:
              $\forall o \in \mathbb{O}, \forall j \in \{1,\ldots,j_{\max}\}, \forall k \in \{1,\ldots,K_j\}$,
              $\eta_{o,j,k}^{(n,m+1)} = \eta_{o,j,k}^{(n,m)} + \tau\big(\mathrm{prox}_{\gamma_n \Phi_{o,j}}(2\eta_{o,j,k}^{(n,m+\frac{1}{2})} - \eta_{o,j,k}^{(n,m)}) - \eta_{o,j,k}^{(n,m+\frac{1}{2})}\big)$;
12:         If $\eta^{(n,m+1)} = \eta^{(n,m)}$, goto 14.
13:     end for
14:     Update the wavelet coefficients of the reconstructed image:
        $\zeta^{(n+1)} = \zeta^{(n)} + \lambda_n(\eta^{(n,m+\frac{1}{2})} - \zeta^{(n)})$.
15:     Compute $\mathcal{J}^{(n+1)} = \mathcal{J}_{\mathrm{CWT}}(\zeta^{(n+1)})$.
16:     $n \leftarrow n + 1$
17: until $|\mathcal{J}^{(n)} - \mathcal{J}^{(n-1)}| \leq \varepsilon \mathcal{J}^{(n-1)}$
18: return $\rho^{(n)} = T^* \zeta^{(n)}$
```

convex sets $C_{\boldsymbol{r}}$ in Eq. (4.43) were computed using a morphological opening and closing operations applied to the basic-SENSE reconstructed image in order to discard very low and high intensities. The CWR-SENSE reconstruction steps are illustrated in Fig. 4.17.

Clearly, the accuracy of the artifact region detection (see the mask image on the top of Fig. 4.17) impacts the performance of our CWR-SENSE method since it defines the areas on which the convex constraints are applied as lower and upper thresholds on the image intensities. As illustrated in Figs. 4.18 and 4.19, inspection of the CWR-SENSE reconstruction results shows that the surviving artifacts in Fig. 4.2 have now been removed due to the anisotropic smoothing using the additional convex constraint.

From a quantitative point of view, significant improvements were achieved by our UWR/CWR-SENSE algorithms in comparison with basic-SENSE and Tikhonov reconstructions. Table 4.1 reports the SNR values in dB corresponding to the basic-SENSE, Tikhonov regularization and the proposed UWR/CWR-SENSE techniques for the illus-

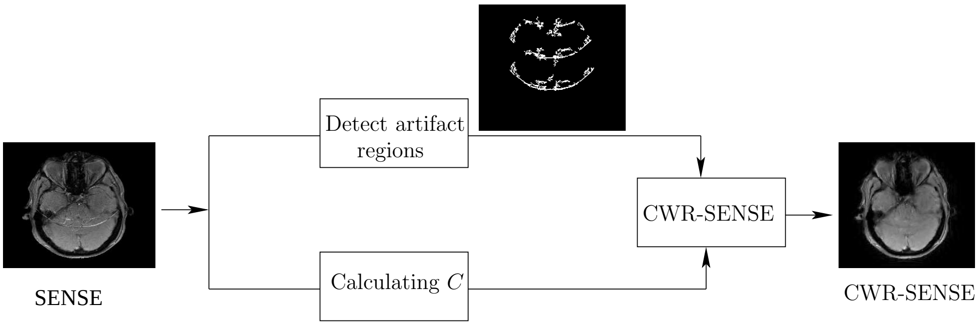


Figure 4.17: CWR-SENSE reconstruction steps.

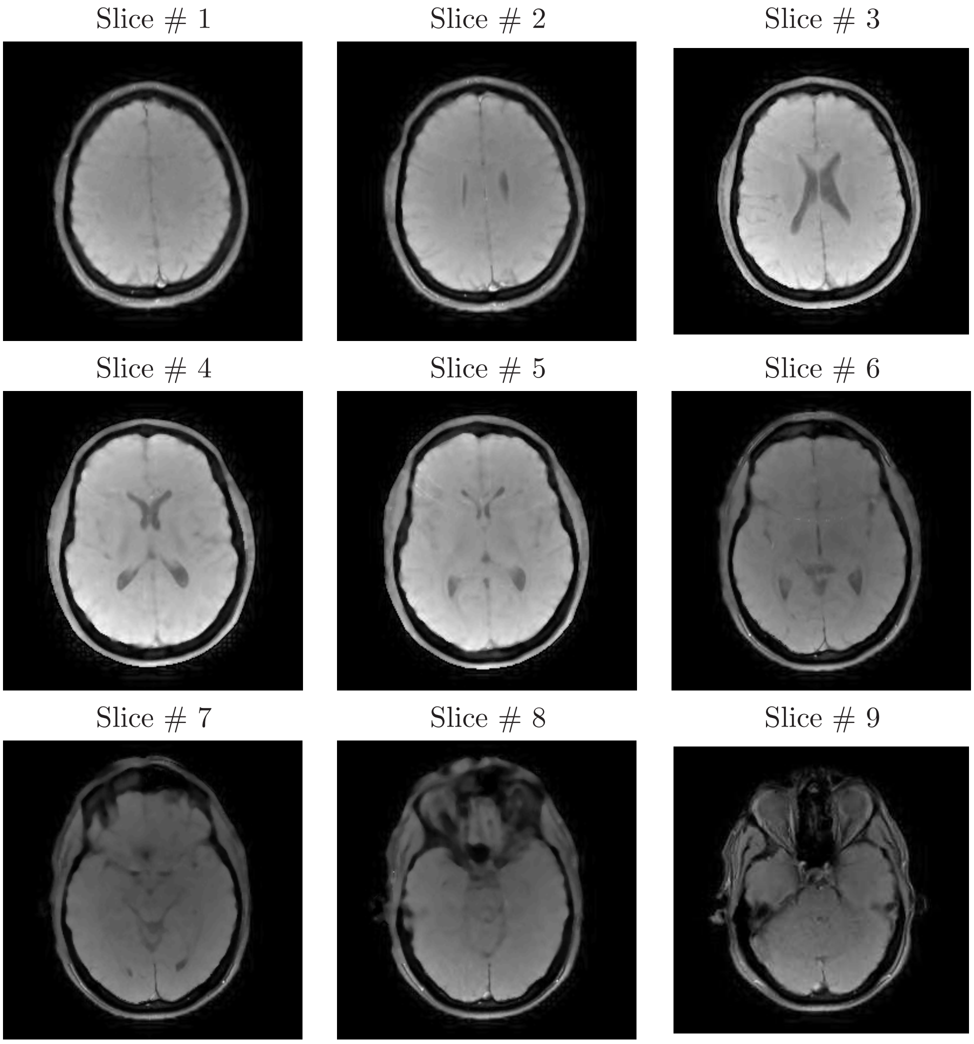
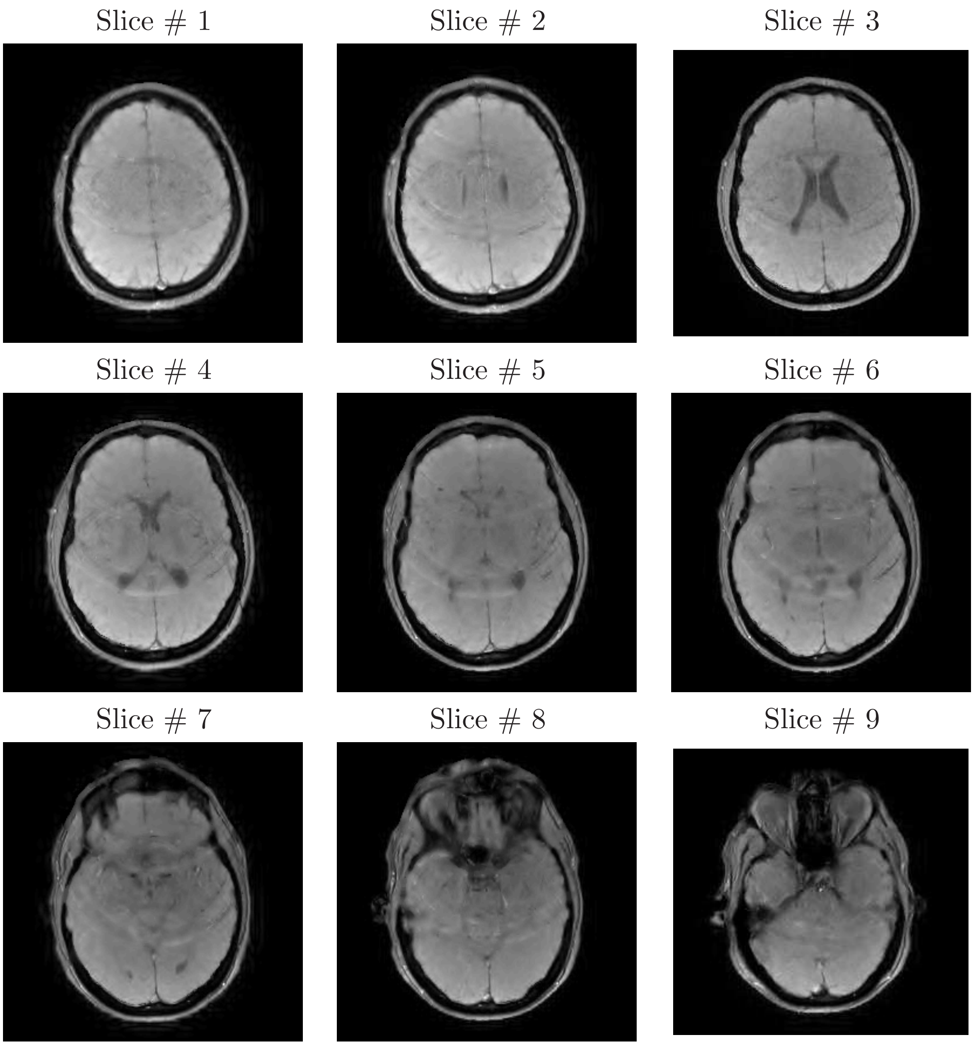


Figure 4.18: Reconstructed anatomical slices using CWR-SENSE for $R = 2$.

trated slices of the anatomical brain volume shown in Figs. 2.19, 4.2, 4.9 and 4.19. In average, the gain drawn from the proposed constrained regularization strategy amounts to 1.08 dB and a better visual quality.



Slice # 1 Slice # 2 Slice # 3

Slice # 4 Slice # 5 Slice # 6

Slice # 7 Slice # 8 Slice # 9

Figure 4.19: Reconstructed anatomical slices using CWR-SENSE for $R = 4$.

Table 4.1: SNR (in dB) evaluation for reconstructed images using different methods for $R = 4$.

	SENSE	Tikhonov	UWR-SENSE	CWR-SENSE
Slice #1	14.36	14.63	14.77	14.95
Slice #2	11.55	11.62	12.01	12.53
Slice #3	12.95	13.44	14.02	14.22
Slice #4	9.24	9.48	10.01	10.30
Slice #5	11.50	11.79	12.06	12.25
Slice #6	9.68	9.87	10.12	10.32
Slice #7	11.00	11.56	11.85	12.00
Slice #8	12.16	12.49	13.00	13.36
Slice #9	13.78	14.77	15.83	16.04
Volume average	11.80	12.18	12.63	12.88

Results on functional EPI data

This experiment on functional EPI data was conducted using the same wavelet basis and priors. The algorithm parameters were adjusted according to the same rules as applied for anatomical data: $\lambda = 1$ and $\gamma = 0.99/\theta = 20.63$ were chosen after deriving the θ constant. Note that the same approach as that applied to anatomical data was adopted to detect artifact regions and compute the upper and lower bounds defining the convex sets $C_{\boldsymbol{r}}$. Fig. 4.20 illustrates six reconstructed full FOV EPI slices using our CWR-SENSE algorithm.

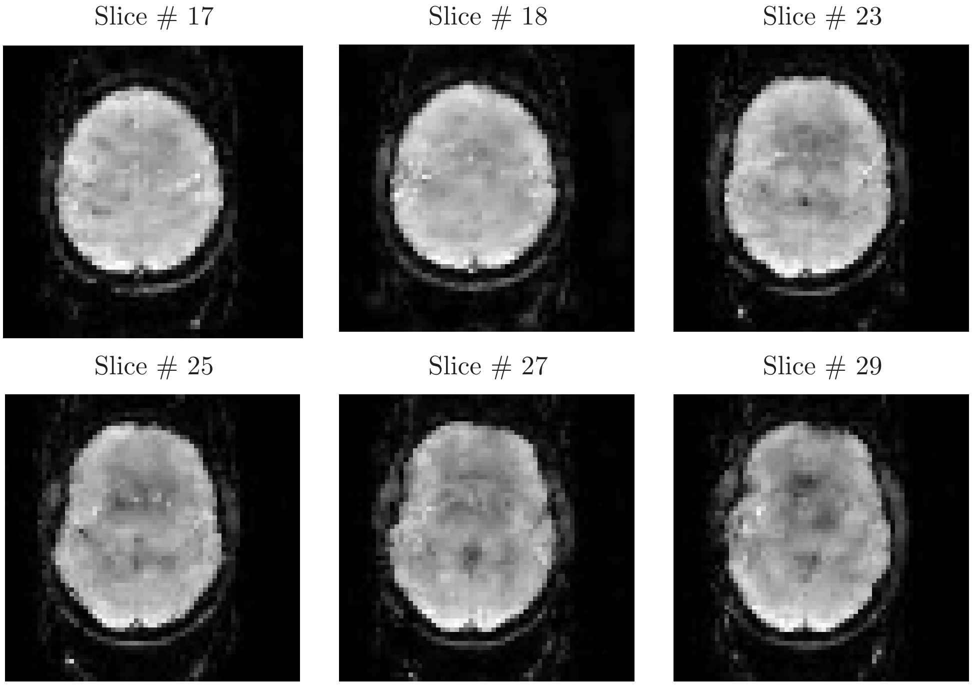


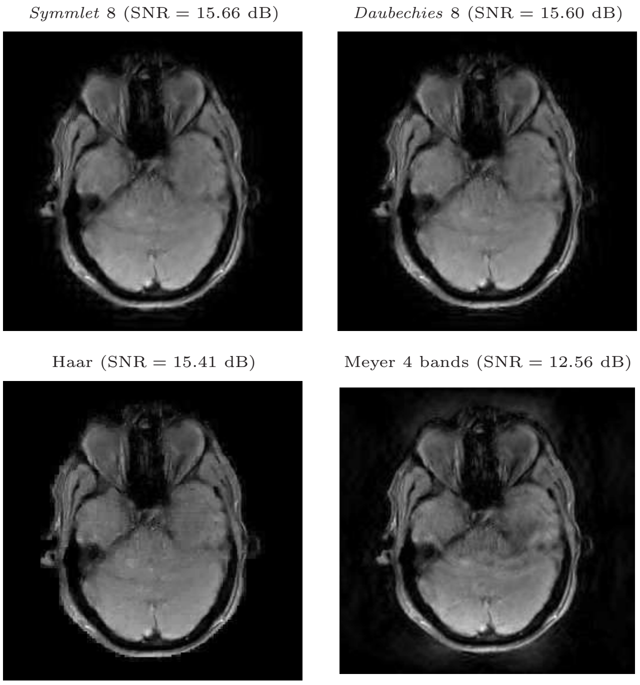
Figure 4.20: Reconstructed EPI slices using CWR-SENSE for $R = 4$.

We noticed that reconstructed EPI images obtained using CWR-SENSE show that the initial very high intensities were smoothed more than those retrieved using basic-SENSE and Tikhonov regularization. Residual defective pixels belonging to distorted areas in SENSE reconstruction were also removed due to the convex constraints introduced in these areas.

Choice of the wavelet basis

Here, we study how the choice of the wavelet basis may influence the reconstruction performance. For comparison purposes, we present the results obtained with four different wavelet bases: dyadic *Symmlet* 8, dyadic *Daubechies* 8, dyadic *Haar* and *Meyer* with $M = 4$ bands [Chaux et al., 2006b]. In Fig. 4.21, reconstructed anatomical images using the different wavelet bases with $j_{\max} = 3$ are displayed. Some boundary effects appear in the reconstructed images, but with a low intensity level. Moreover, they do not affect the brain mask when using the *Symmlet*, *Daubechies* and *Haar* wavelet bases. These additional artifacts become much more important with the *Meyer* 4-band wavelet basis because of its large spatial support. Hence, they drastically decrease the SNR of the reconstructed full FOV image. Note also that the *Haar* wavelet basis introduces some blocking effects caused by its discontinuities that do not occur using alternative bases. Among the latter, *Symmlet* 8 gives slightly more accurate regularized results than *Daubechies* 8 but none of them generates significant additional artifacts.

We would like here to mention that using the *Haar* wavelets with quadratic penalization is somehow equivalent to Tikhonov regularization where no reference image is needed (i.e. by setting $\rho_{\mathrm{r}} = 0_{\mathbb{C}^{X\times Y}}$ in Eq. (4.1)) with a weighting matrix which allows us to penalize the inter-pixel differences. This may be performed by substituting the identity matrix $\boldsymbol{I}_R$ in Eq. (4.1) to the R dimensional square matrix which allows us to calculate the second order finite differences in the image.


Symmlet 8 (SNR = 15.66 dB)

Daubechies 8 (SNR = 15.60 dB)

Haar (SNR = 15.41 dB)

Meyer 4 bands (SNR = 12.56 dB)

Figure 4.21: Influence of the wavelet basis on the anatomical reconstructed images using our CWR-SENSE method for $R = 4$.

4.4 Combined wavelet-TV regularization

It has been shown hereabove that using wavelet ragularization allows preserving the image details while smoothing reconstruction artifacts. However, it has also been noticed that wavelet regularization may introduce some irregularity in homogeneous area of the image. On the other hand, it is well known that TV regularization is well adapted to regularize smooth regions, but may oversmooth image details (see Section 4.2.2).

In the pMRI regularization literature, these two different edge-preserving penalizations have always been used separately. However, combining them in a joint regularization framework would be fruitful in order to take advantage of both of them since they may alleviate the drawbacks of each other. A joint Wavelet Transform-Total Variation (WT-TV) regularization can reduce to the optimization of the following penalized criterion:

$$\begin{aligned}\widehat{\rho}_{\mathrm{WT-TV}} &= T^* \arg\min_{\zeta} \mathcal{J}_{\mathrm{WT-TV}}(\zeta) \\ &= T^* \arg\min_{\zeta} \mathcal{J}_{\mathrm{L}}(T^*\zeta) + \kappa_1 \mathcal{J}_{\mathrm{P}}(\zeta) + \kappa_2 \|T^*\zeta\|_{\mathrm{TV}},\end{aligned} \tag{4.53}$$

where $\mathcal{J}_{\mathrm{L}}$ and $\mathcal{J}_{\mathrm{P}}$ are defined in Eqs. (4.3) and (4.9), respectively, $\kappa_1 > 0$ and $\kappa_2 > 0$ are regularization parameters, and T^* is the wavelet adjoint operator. The discrete total variation norm of an image $\rho \in \mathbb{C}^{Y\times X}$ is given by:

$$\|\rho\|_{\mathrm{TV}} = \sum_{y=1}^{Y}\sum_{x=1}^{X} \sqrt{|(\nabla_1\rho)(y,x)|^2 + |(\nabla_1\rho^{\intercal})^{\intercal}(y,x)|^2}, \tag{4.54}$$

where for every $\rho \in \mathbb{C}^{Y\times X}$, ∇_1 is the horizontally smoothed gradient operator defined by

$$\nabla_1(\rho) = \frac{1}{2}(\rho(y+1,x+1) - \rho(y,x+1) + \rho(y+1,x) - \rho(y,x))_{1\le y\le Y, 1\le x\le X}. \tag{4.55}$$

Note here that for the sake of presentation, we assumed that the image ρ is periodic. Although the optimality criterion $\mathcal{J}_{\mathrm{WT-TV}}(\zeta)$ is an exact match of the kind of criteria that can be minimized using the Parallel ProXimal algorithm (PPXA), the difficulty here stems from the calculation of the proximity operator of $\|\cdot\|_{\mathrm{TV}} \circ T^*$. To circumvent this difficulty, the TV penalization in Eq. (4.53) can be split into four terms as in [Combettes and Pesquet, 2008]. The TV penalty term therefore reads:

$$\forall \rho \in \mathbb{C}^{Y\times X},\ \|\rho\|_{\mathrm{TV}} = \sum_{i=0}^{3} tv_i(\rho), \tag{4.56}$$

where for every $(q,r) \in \{0,1\}^2$,

$$tv_{2q+r}(\rho) = \sum_{y=1}^{Y/2}\sum_{x=1}^{X/2} \sqrt{|(\nabla_1\rho)(2y+q,2x+r)|^2 + |(\nabla_1\rho^{\intercal})^{\intercal}(2y+q,2x+r)|^2}. \tag{4.57}$$

For every q and r in $\{0,1\}$, let $\downarrow_{q,r}$ be the decimation operator defined by

$$\begin{aligned}\downarrow_{q,r}\colon \mathbb{C}^{2Y\times 2X} &\to \mathbb{C}^{Y\times X} \\ \nu = (\nu_{y,x})_{1\le y\le 2Y, 1\le x\le 2X} &\mapsto (\nu_{2y+q,2x+r})_{1\le y\le 2Y, 1\le x\le 2X}.\end{aligned} \tag{4.58}$$

and U_{q+2r} be the following operator:

$$\begin{aligned}U_{q+2r} : \mathbb{C}^{Y\times X} &\to \mathbb{C}^{Y\times X} \\ \rho &\mapsto \downarrow_{q,r} \begin{bmatrix} \nabla_0\rho & \nabla_1\rho \\ (\nabla_1\rho^{\intercal})^{\intercal} & \nabla_2\rho \end{bmatrix},\end{aligned} \tag{4.59}$$

where for every $\rho \in \mathbb{C}^{Y\times X}$, the operators ∇_0 and ∇_2 are defined by

$$\nabla_0(\rho) = \frac{1}{2}(\rho(y+1,x+1)+\rho(y,x+1)+\rho(y+1,x)+\rho(y,x))_{1\leq y\leq Y,1\leq x\leq X} \tag{4.60}$$

and

$$\nabla_2(\rho) = \frac{1}{2}(\rho(y+1,x+1)-\rho(y,x+1)-\rho(y+1,x)+\rho(y,x))_{1\leq y\leq Y,1\leq x\leq X}, \tag{4.61}$$

respectively. Let also h be the function defined on $\mathbb{C}^{Y\times X}$ by:

$$h(\rho) = \sum_{y=1}^{Y/2}\sum_{x=1}^{X/2}\sqrt{|(\nabla_1\rho)(y,x+X/2)|^2+|(\nabla_1\rho^{\intercal})^{\intercal}(y+Y/2,x)|^2}. \tag{4.62}$$

It turns out from Eqs. (4.57) and (4.59) that

$$\forall i\in\{0,1,2,3\},\ tv_i = h\circ U_i. \tag{4.63}$$

Consequently, the optimization problem in Eq. (4.53) can be rewritten as:

$$\begin{aligned}\widehat{\rho}_{\mathrm{WT-TV}} &= T^*\arg\min_{\zeta}[\mathcal{J}_{\mathrm{L}}(T^*\zeta)+\kappa_1\mathcal{J}_{\mathrm{P}}(\zeta)+\kappa_2\sum_{i=0}^{3}tv_i(T^*\zeta)]\\ &= T^*\arg\min_{\zeta}[\mathcal{J}_{\mathrm{L}}(T^*\zeta)+\kappa_1\mathcal{J}_{\mathrm{P}}(\zeta)+\kappa_2\sum_{i=0}^{3}h(U_i(T^*\zeta))].\end{aligned} \tag{4.64}$$

When the used wavelet transform is a tight frame ($T^*T=\mu\mathrm{Id}$), The proximity operator of $tv_i = h\circ U_i\circ T^*$ for every $i\in\{0,1,2,3\}$ can be calculated according to Proposition 4.4.1 [Combettes and Pesquet, 2008, Prop. 4.1].

Proposition 4.4.1 *[Combettes and Pesquet, 2008]*
Set

$$\begin{aligned}\Pi : \mathbb{C}^{Y\times X} &\to \mathbb{C}^{Y\times X}\\ \nu = (\nu_{y,x})_{1\leq y\leq 2Y,1\leq x\leq 2X} &\mapsto (\pi_{y,x})_{1\leq y\leq Y,1\leq x\leq X},\end{aligned}$$

where $\forall(y,x)\in\{1,\dots,Y/2\}\times\{1,\dots,X/2\}$, $\begin{cases}\pi_{y,x}=\nu_{y,x}\\ \pi_{y+Y/2,x+X/2}=\nu_{y+Y/2,x+X/2}\\ \pi_{y,x+X/2}=\Omega_{y,x}(\nu)\nu_{y,x+X/2}\\ \pi_{y+Y/2,x}=\Omega_{y,x}(\nu)\nu_{y+Y/2,x}\end{cases}$

with

$$\Omega_{y,x}(\nu) = \begin{cases}1-\dfrac{\kappa_2\mu}{\sqrt{|\nu_{y,x+X/2}|^2+|\nu_{y+Y/2,x}|^2}} & \text{if } \sqrt{|\nu_{y,x+X/2}|^2+|\nu_{y+Y/2,x}|^2}\geq\mu\kappa_2\\ 0 & \text{otherwise.}\end{cases}$$

Then, for every $i \in \{0, 1, 2, 3\}$,

$$\operatorname{prox}_{tv_i} = \mathrm{Id} + T \circ (U_i^* \circ \Pi \circ U_i - \mathrm{Id}) \circ T^* \tag{4.65}$$

The regularized WT-TV reconstruction approach is summarized in Algorithm 8 where the PPXA algorithm [Combettes and Pesquet, 2008] is used to minimize the optimality criterion in Eq. (4.64) which is made up of $J = 6$ convex functions.

Algorithm 8 WT-TV-SENSE: 2D-slice WT-TV-based regularized reconstruction.

```
Set $\gamma \in ]0, +\infty[$, $n = 0$, $(\omega_i)_{1 \le i \le 6} \in [0,1]^6$ such that $\sum_{i=0}^{6} \omega_i = 1$, $(\zeta_i^{(n)})_{1 \le i \le 6} \in (\mathbb{C}^K)^6$ where $\zeta_i^{(n)} = \big((\boldsymbol{\zeta}_a^{i,(n)}), ((\boldsymbol{\zeta}_{o,j}^{i,(n)}))_{o \in \mathbb{O}, 1 \le j \le j_{\max}}\big)$ for every $i \in \{1, \dots, 6\}$. Set also $\varepsilon \ge 0$, initialize $\zeta^{(n)} = \sum_{i=1}^{6} \omega_i \zeta_i^{(n)}$ and $\mathcal{J}^{(n)} = 0$.
repeat
    Compute the image $u^{(n)}$ such that:
    $\forall \mathbf{r} \in \{1, \dots, Y/R\} \times \{1, \dots, X\}$,
    $\boldsymbol{u}^{(n)}(\mathbf{r}) = (\boldsymbol{I}_R + \frac{2\gamma}{\omega_1}\boldsymbol{S}^{\mathsf{H}}(\mathbf{r})\boldsymbol{\Psi}^{-1}\boldsymbol{S}(\mathbf{r}))^{-1}(\boldsymbol{\rho}_1^{(n)}(\mathbf{r}) + \frac{2\gamma}{\omega_1}\boldsymbol{S}^{\mathsf{H}}(\mathbf{r})\boldsymbol{\Psi}^{-1}\boldsymbol{d}(\mathbf{r}))$ where $\rho_1^{(n)} = T^*\zeta_1^{(n)}$.
    Compute the wavelet coefficients $p_1^{(n)} = Tu^{(n)}$ of $u^{(n)}$.
    Calculate $p_2^{(n)} = \big(\operatorname{prox}_{\gamma\kappa_1\Phi_a/\omega_2}(\boldsymbol{\zeta}_a^{2,(n)}), (\operatorname{prox}_{\gamma\Phi_{o,j}/\omega_2}(\boldsymbol{\zeta}_{o,j}^{2,(n)}))_{o \in \mathbb{O}, 1 \le j \le j_{\max}}\big)$.
    For every $i \in \{0, 1, 2, 3\}$, calculate $p_{i+3}^{(n)} = \operatorname{prox}_{\gamma\kappa_2 tv_i/\omega_i}(\zeta_{i+3}^{(n)})$.
    Set $P^{(n)} = \sum_{i=1}^{6} \omega_i p_i^{(n)}$.
    Set $\lambda_n \in [0, 2]$.
    for $i = 1$ to $6$ do
        $\zeta_i^{(n)} = \zeta_i^{(n)} + \lambda_n(2P^{(n)} - \zeta^{(n)} - p_i^{(n)})$ .
    end for
    $\zeta^{(n+1)} = \zeta^{(n)} + \lambda_n(P^{(n)} - \zeta^{(n)})$.
    Compute $\mathcal{J}^{(n+1)} = \mathcal{J}_{\mathrm{WT-TV}}(\zeta^{(n+1)})$.
    $n \leftarrow n + 1$.
until $|\mathcal{J}^{(n)} - \mathcal{J}^{(n-1)}| \le \varepsilon \mathcal{J}^{(n-1)}$
return $\rho^{(n)} = T^*\zeta^{(n)}$
```

Experiments were conducted on the same dataset of anatomical and functional images used previously. For $R = 4$, Fig. 4.22 shows the anatomical reconstructed slices using Algorithm 8 when a dyadic ($M = 2$) *Symmlet* orthonormal wavelet basis associated with filters of length 8 is used over $j_{\max} = 3$ resolution levels. Fig. 4.23 illustrates reconstructed images corresponding to the same slices when a Union of 2 Orthonormal Bases (U2OB) WT is used with *Symmlet* 8 and *Symmlet* 4 filters. Note here that the hyper-parameters of the wavelet prior and the TV regularization parameter were estimated using the proposed approach in Chapter 5. These figures show that the reconstructed images present better regularity than the ones reconstructed using the UWR-SENSE algorithm. However, from a visual viewpoint, using redundant WT do not necessarily lead to better reconstruction quality. From a quantitative viewpoint, SNR values are provided in Table 4.2. Comparisons with SNR values in Table 4.1 confirm the usefulness of combining WT and TV in a joint regularization framework. This table shows that a slight improvement of 0.02 dB is obtained compared to the UWR-SENSE algorithm when using the *Symmlet* WT. However, more isignificant improvement (0.24 dB) is reached when using the U2OB WT. It turns out then that even if the reconstruction performances of WT-TV-SENSE using the OB and U2OB are visually similar, SNR values indicate that using redundant WTs is fruitful.

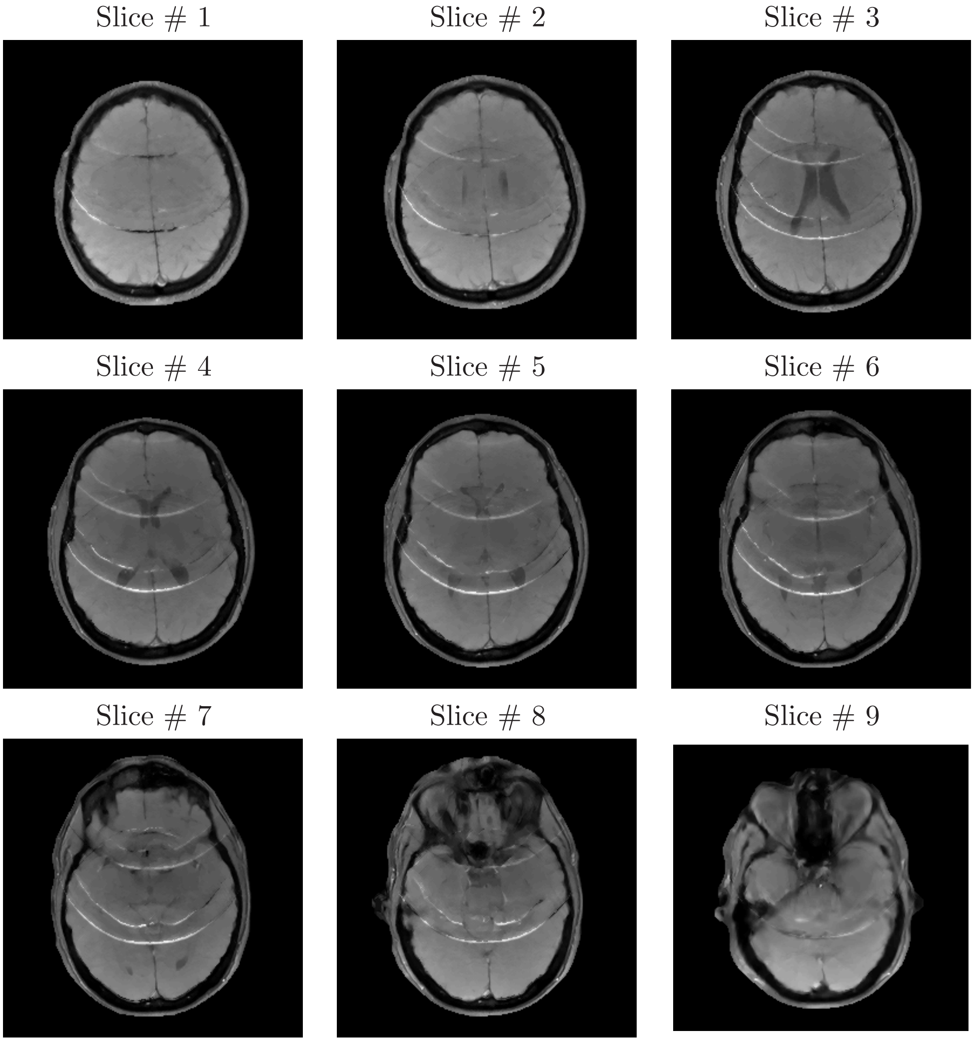


Figure 4.22: Reconstructed anatomical slices using Algorithm 8 with the *Symmlet* 8 orthonormal basis for $R = 4$.

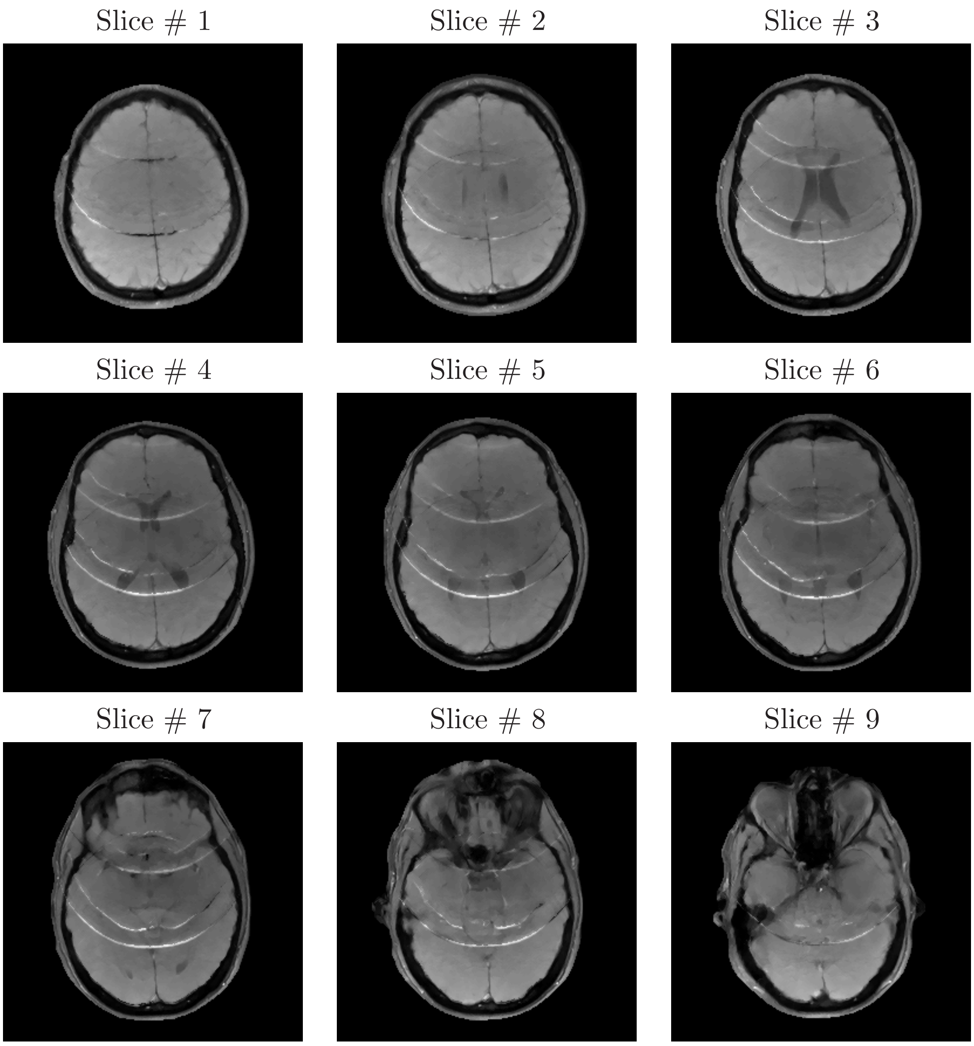


Figure 4.23: Reconstructed anatomical slices using Algorithm 8 with the U2OB for $R = 4$.

Table 4.2: SNR (in dB) evaluation for reconstructed images using Algorithm 8 with the OB and U2OB wavelet representations for $R = 4$.

	OB	U2OB
Slice #1	14.85	15.16
Slice #2	12.18	12.35
Slice #3	13.85	14.32
Slice #4	9.78	9.85
Slice #5	12.17	12.62
Slice #6	10.20	10.17
Slice #7	11.75	12.04
Slice #8	13.15	13.44
Slice #9	15.94	15.88
Volume average	12.65	12.87

4.5 Spatio-temporal regularization

All the proposed approaches up to now proceed by a slice by slice reconstruction. Iterating over slices is therefore necessary in order to reconstruct the whole volume. However, in an fMRI study, the whole brain volume has to be acquired several times. Iterating over all the acquired 3D volumes is then possible in order to reconstruct a 4D data volume corresponding to an fMRI session. Consequently, the 3D volumes are supposed independent although they belong to the same fMRI session. However, in practice, the 3D temporal volumes are somehow dependent since they belong to the same fMRI session involving the same experimental paradigm. The BOLD time-series and the acquisition noise are in fact correlated in time in an fMRI session. For this reason, taking into account temporal dependencies between 3D volumes may be fruitful for the reconstruction process. This ambition motivated the extension of the proposed approach in order to exploit the dependencies between acquired volumes. These dependencies are exploited by applying an additional regularization term in the temporal dimension of the 4D dataset. This additional regularization should help to increase the SNR through the acquired volumes, and therefore enhance the reliability of the statistical analysis in fMRI. These temporal dependencies have also been used in the dynamic MRI literature in order to improve the reconstruction quality in conventional MRI [Sümbül et al., 2009]. However, since in dynamic MRI the imaged object geometry generally changes during the acquisition, joining the reconstruction process to the temporal regularization is very difficult.
To deal with a 4D reconstruction of the N_r acquired volumes (N_r is usually even), we will first rewrite the observation model in Eq. (2.16) as follows:

$$\boldsymbol{d}_t(\boldsymbol{r}) = \boldsymbol{S}(\boldsymbol{r})\boldsymbol{\rho}_t(\boldsymbol{r}) + \boldsymbol{n}_t(\boldsymbol{r}), \tag{4.66}$$

where $t \in \{1, \ldots, N_r\}$ is the acquisition time and $\boldsymbol{r} = (y, x, z)$ is the 3D spatial position, $z \in \{1, \ldots, Z\}$ being the position along the third direction (slice selection one).
Using a dyadic 3D wavelet operator T, the coefficients will be reindexed as in Section 4.3.2 so that $\zeta^t = \big(\boldsymbol{\zeta}^t_a, (\boldsymbol{\zeta}^t_{o,j})_{o\in\mathbb{O},1\le j\le j_{\max}}\big)$ with $o \in \mathbb{O} = \{0,1\}^3 \setminus \{(0,0,0)\}$. Using 3D dyadic wavelet transforms represents also an extension of the previously proposed approaches since 3D wavelets allow us to smooth reconstruction artifacts along the slice selection direction, which have always been inaccessible using any slice by slice operating approach.
Accounting for the additional temporal ℓ_p regularization term, reconstructing the 4D volume is performed through the minimization of the following optimality criterion:

$$\begin{aligned}\hat{\zeta} &= \arg\min_{\zeta} \mathcal{J}_{\mathrm{ST}}(\zeta)\\ &= \arg\min_{\zeta} \sum_{t=1}^{N_r} \sum_{\mathbf{r}\in\{1,\ldots,Y/R\}\times\{1,\ldots,X\}\times\{1,\ldots,Z\}} \|\boldsymbol{d}_t(\boldsymbol{r}) - \boldsymbol{S}(\boldsymbol{r})(T^*\zeta^t)(\boldsymbol{r})\|^2_{\boldsymbol{\Psi}^{-1}} + \sum_{t=1}^{N_r} \mathcal{J}_{\mathrm{P}}(\zeta^t) \\ &+ \kappa \sum_{t=2}^{N_r} \|T^*\zeta^t - T^*\zeta^{t-1}\|_p^p,\end{aligned} \tag{4.67}$$

where $\zeta = (\zeta^1, \zeta^2, \dots, \zeta^{N_r})^\intercal$, $\kappa > 0$ is a regularization parameter and $\mathcal{J}_\mathrm{P}$ is defined similarly to Eq. (4.9). The adjoint wavelet operator T^* is then applied to each component ζ^t of ζ to obtain the reconstructed 3D volume ρ_t at the acquisition time t by taking into account the time dependencies with the other acquired volumes.

Adopting this formulation, the minimization step will play a prominent role in the reconstruction process. Since the optimality criterion is made up of more than two terms, to the best of our knowledge, the sole algorithm allowing the minimization of such optimality criteria without using internal iterative loops is PPXA. In fact, PPXA can be used to minimize the criterion in Eq. (4.67) once we are able to calculate the proximity operator of each of three involved terms. This task is quite simple for the two first terms of Eq. (4.67) since they are separable w.r.t. the time variable t and the spatial position $\boldsymbol{r}$. However, since this is not the case for the time penalization term (third term in Eq. (4.67)), calculating its proximity operator is not obvious. For this reason, we propose to rewrite the optimality criteria in $\mathcal{J}_\mathrm{ST}$ by decomposing the time penalization into two terms which are somehow separable w.r.t. the time variable t (a given acquisition time t is either involved in the first or in the second term), and for which the proximity operators are easy to calculate:

$$\begin{aligned}\mathcal{J}_\mathrm{ST}(\zeta) = &\sum_{t=1}^{N_r} \sum_{\mathbf{r}\in\{1,\dots,Y/R\}\times\{1,\dots,X\}\times\{1,\dots,Z\}} \|\boldsymbol{d}_t(\boldsymbol{r}) - \boldsymbol{S}(\boldsymbol{r})(T^*\zeta^t)(\boldsymbol{r})\|^2_{\boldsymbol{\Psi}^{-1}} + \sum_{t=1}^{N_r} \mathcal{J}_\mathrm{P}(\zeta^t) \\ &+ \mathcal{J}_\mathrm{T}^1(\zeta) + \mathcal{J}_\mathrm{T}^2(\zeta),\end{aligned} \tag{4.68}$$

where

$$\mathcal{J}_\mathrm{T}^1(\zeta) = \kappa \sum_{t=1}^{N_r/2} \|T^*\zeta^{2t} - T^*\zeta^{2t-1}\|_p^p \tag{4.69}$$

and

$$\mathcal{J}_\mathrm{T}^2(\zeta) = \kappa \sum_{t=1}^{N_r/2-1} \|T^*\zeta^{2t+1} - T^*\zeta^{2t}\|_p^p. \tag{4.70}$$

Since $\mathcal{J}_\mathrm{T}^1$ (resp. $\mathcal{J}_\mathrm{T}^2$) is separable w.r.t the time variable t, its proximity operator can easily be calculated based on the proximity operator of each of the involved terms in the sum of Eq. (4.69) (resp. Eq. (4.70)).

Let us consider the following function

$$\begin{aligned}\Phi : \mathbb{C}^K \times \mathbb{C}^K &\longrightarrow \mathbb{R} \\ (\zeta^t, \zeta^{t-1}) &\mapsto \kappa\|T^*\zeta^t - T^*\zeta^{t-1}\|_p^p = \phi \circ H((\zeta^t, \zeta^{t-1})),\end{aligned} \tag{4.71}$$

where $\phi(\cdot) = \kappa\|T^* \cdot\|_p^p$ and H is a linear operator defined by

$$\begin{aligned}H : \mathbb{C}^K \times \mathbb{C}^K &\longrightarrow \mathbb{C}^K \\ (a, b) &\mapsto a - b.\end{aligned} \tag{4.72}$$

Its associated adjoint operator H^* is therefore given by

$$\begin{aligned} H^* : \mathbb{C}^K &\longrightarrow \mathbb{C}^K \times \mathbb{C}^K \\ a &\mapsto (a, -a). \end{aligned} \tag{4.73}$$

The proximity operator of Φ can easily be calculated using Assumption 3.3.3 in the explicit form described in Section 3.3.2.1 of Chapter 3 since we have $HH^* = 2\mathrm{Id}$:

$$\mathrm{prox}_\Phi = \mathrm{prox}_{\phi \circ H} = \mathrm{Id} + \frac{1}{2} H^* (\mathrm{prox}_{2\phi} - \mathrm{Id}) \circ H. \tag{4.74}$$

The resulting algorithm for the minimization of the optimality criterion in Eq. (4.68) is given in Algorithm 9. Since it is impossible to apply the 4D-UWR-SENSE algorithm to anatomical data, and since the resting state EPI dataset acquired at 1.5 Tesla does not contain enough repetitions ($N_r = 14$), the validation of this apatio-temporal regularized reconstruction approach is discussed in Chapter 6. This validation includes an fMRI study on 4D EPI data acquired at different reduction factors at a 3 Tesla magnetic field.

Algorithm 9 4D-UWR-SENSE: 4D spatio-temporal regularized reconstruction.

Set $\gamma \in]0,+\infty[$, $n=0$, $(\omega_i)_{1\le i\le 4}\in[0,1]^4$ such that $\sum_{i=0}^{4}\omega_i = 1$, $(\zeta_i^{(n)})_{1\le i\le 4}\in(\mathbb{C}^{K\times N_r})^4$ where $\zeta_i^{(n)} = (\zeta_i^{1,(n)},\zeta_i^{2,(n)},\dots,\zeta_i^{N_r,(n)})$ and $\zeta_i^{t,(n)} = \big((\boldsymbol{\zeta}_{i,a}^{t,(n)}),((\boldsymbol{\zeta}_{i,o,j}^{t,(n)}))_{o\in\mathbb{O},1\le j\le j_{\max}}\big)$ for every $i\in\{1,\dots,4\}$ and $t\in\{1,\dots,N_r\}$. Set also $\varepsilon\ge 0$ and initialize $\zeta^{(n)} = \sum_{i=1}^{4}\omega_i\zeta_i^{(n)}$ and $\mathcal{J}^{(n)} = 0$.

```
repeat
  Set $p_4^{1,(n)} = \zeta_4^{1,(n)}$.
  for $t = 1$ to $N_r$ do
    Calculate the image $u_t^{(n)}$ such that: $\forall \mathbf{r}\in\{1,\dots,Y/R\}\times\{1,\dots,X\}\times\{1,\dots,Z\}$,
    $\boldsymbol{u}_t^{(n)}(\mathbf{r}) = (\boldsymbol{I}_R + \frac{2\gamma}{\omega_1}\boldsymbol{S}^{\mathsf{H}}(\mathbf{r})\boldsymbol{\Psi}^{-1}\boldsymbol{S}(\mathbf{r}))^{-1}(\boldsymbol{\rho}_{t,1}^{(n)}(\mathbf{r}) + \frac{2\gamma}{\omega_1}\boldsymbol{S}^{\mathsf{H}}(\mathbf{r})\boldsymbol{\Psi}^{-1}\boldsymbol{d}_t(\mathbf{r}))$ where
    $\rho_{t,1}^{(n)} = T^*\zeta_1^{t,(n)}$.
    Compute the wavelet coefficients $p_1^{t,(n)} = Tu_t^{(n)}$.
    Compute $p_2^{t,(n)} = \big(\mathrm{prox}_{\gamma\Phi_a/\omega_2}(\boldsymbol{\zeta}_{2,a}^{t,(n)}),(\mathrm{prox}_{\gamma\Phi_{o,j}/\omega_2}(\boldsymbol{\zeta}_{2,o,j}^{t,(n)}))_{o\in\mathbb{O},1\le j\le j_{\max}}\big)$.
    if $t$ is even then
      calculate $(p_3^{t,(n)},p_3^{t-1,(n)}) = \mathrm{prox}_{\gamma\kappa\Phi/\omega_3}((\zeta^{t,(n)},\zeta^{t-1,(n)}))$
    else if $t$ is odd and $t>1$ then
      calculate $(p_4^{t,(n)},p_4^{t-1,(n)}) = \mathrm{prox}_{\gamma\kappa\Phi/\omega_4}((\zeta^{t,(n)},\zeta^{t-1,(n)}))$.
    end if
    if $t>1$ then
      Set $P^{t-1,(n)} = \sum_{i=1}^{4}\omega_i p_i^{t-1,(n)}$.
    end if
  end for
  Set $p_4^{N_r,(n)} = \zeta_4^{N_r,(n)}$.
  Set $P^{N_r,(n)} = \sum_{i=1}^{4}\omega_i p_i^{N_r,(n)}$.
  Set $p_1^{(n)} = (p_1^{1,(n)},p_1^{2,(n)},\dots,p_1^{N_r,(n)})$, $p_2^{(n)} = (p_2^{1,(n)},p_2^{2,(n)},\dots,p_2^{N_r,(n)})$,
  $p_3^{(n)} = (p_3^{1,(n)},p_3^{2,(n)},\dots,p_3^{N_r,(n)})$, $p_4^{(n)} = (p_4^{1,(n)},p_4^{2,(n)},\dots,p_4^{N_r,(n)})$ and
  $P^{(n)} = (P^{1,(n)},P^{2,(n)},\dots,P^{N_r,(n)})$.
  Set $\lambda_n\in[0,2]$.
  for $i = 1$ to $4$ do
    $\zeta_i^{(n)} = \zeta_i^{(n)} + \lambda_n(2P^{(n)} - \zeta^{(n)} - p_i^{(n)})$ .
  end for
  $\zeta^{(n+1)} = \zeta^{(n)} + \lambda_n(P^{(n)} - \zeta^{(n)})$.
  Compute $\mathcal{J}^{(n+1)} = \mathcal{J}_{\mathrm{ST}}(\zeta^{(n+1)})$.
  $n \leftarrow n+1$.
until $|\mathcal{J}^{(n)} - \mathcal{J}^{(n-1)}| \le \varepsilon\mathcal{J}^{(n-1)}$.
Set $\hat{\zeta} = \zeta^{(n)}$.
return $\hat{\rho}_t = T^*\hat{\zeta}^t$ for every $t\in\{1,\dots,N_r\}$.
```

4.6 Conclusion

In this chapter, we have presented the proposed wavelet-based regularized reconstruction approach, as well as its different extensions. The first extension involved convex constraints on artifact regions, which allowed us to achieve precise reconstruction even under severe experimental conditions. The proposed approach was also combined with Total Variation regularization to take advantage of the two edge-preserving penalties. The last extension consisted of using 3D wavelet transforms and temporal regularization in order to reconstruct 4D fMRI data while exploiting temporal dependencies between acquired 3D volumes. It is worth noting that the proposed reconstruction strategy involves the estimation of several hyper-parameters. It is a delicate task especially when redundant WT are employed. The next chapter deals with the hyper-parameters estimation problem in such a regularization framework using frame representations.

Chapter 5

Hyper-parameter estimation

Contents

5.1 Introduction 125
5.2 Problem Formulation 126
5.3 Hierarchical Bayesian Model 127
5.4 Sampling strategies 129
5.5 Toward a more general Bayesian Model 137
5.6 Numerical illustrations 141
5.7 Conclusion 158

5.1 Introduction

Results reported in Chapter 4 emphasize the high interest of using wavelet representations in pMRI regularization. More generally, redundant linear decomposition families called *frames* are very useful in many signal and image processing applications like reconstruction [Kawahara, 1993; Chaari et al., 2010], restoration [Miller, 1999; Chaux et al., 2005], compression [Martin and Bell, 2001; Meuleneire, 2008]... The main advantage of frames lies in their flexibility to capture local features of the signal. Hence, they may result in sparse representations as shown in the literature on curvelets [Candès et al., 2002], contourlets [Do and Vetterli, 2005], bandelets [Le Pennec and Mallat, 2005] or dual-trees [Chaux et al., 2006b] in image processing. However, as it has been shown in Chapter 4, a major difficulty when using frame representations in a statistical framework is to estimate the parameters of the frame coefficient probability distribution. Actually, since frame synthesis operators are generally not injective, even if the signal is perfectly known, the determination of its frame coefficients is an underdetermined problem. This chapter studies a hierarchical Bayesian approach to estimate the frame coefficients and their hyper-parameters. Although this approach is conceptually able to deal with any desirable distribution for the frame coefficients, we focus in this chapter on generalized Gaussian (GG) priors. The developed approach can easily be adapted to handle other priors such as the Generalized Gauss-Laplace (GGL) one used in Eq. (4.23) for pMRI regularization. Note however that we do not restrict our attention to log-concave GG prior probability density functions (pdf), which may be limited for providing accurate models of sparse signals [Seeger et al., 2007]. In addition, the proposed method can be applied to noisy data when imprecise measurements of the signal are only available. One of the contributions of this work is to address the case of uniform noise distributions. Such distributions are useful in many applications. For example, they can be used to model quantization noise arising in data compression [Watson et al., 1997] and they are often employed when dealing with bounded error measurements [Zhang et al., 1992; Rangan and Goyal, 2001; Breuel, 1996; Alghoniemy and Tewfik, 2004].

Our work takes advantage of the current developments in Markov Chain Monte Carlo (MCMC) algorithms [Robert and Castella, 2004; Cappé, 2002; Andrieu et al., 2001] that have already been investigated for instance in image separation [Ichir and Mohammad-Djafari, 2003], image restoration [Jalobeanu et al., 2002; Orieux et al., 2010; Giovannelli, 2008] and brain activity detection in functionnal MRI [Makni et al., 2005; Khalidov et al., 2007]. These algorithms have also been investigated for signal/image processing problems with sparsity constraints. These constraints may be imposed in the original space like in [Dobigeon et al., 2009], where a sparse image reconstruction problem is assessed in the image domain. They may also be imposed on some redundant representation of the signal like in [Blumensath and Davies, 2007], where a time-series sparse coding problem is considered.

Hybrid MCMC algorithms [Zeger and Karim, 1991; Tierney, 1994] are designed combining Metropolis-Hastings (MH) [Hastings, 1970] and Gibbs [Geman and Geman, 1984] moves to sample according to the posterior distribution of interest. MCMC algorithms and WT have been jointly investigated in some works dealing with signal denoising under a Bayesian framework [Leporini and Pesquet, 2001; Müller and Vidakovic, 1999; Heurta, 2005; Ichir and Mohammad-Djafari, 2003]. However, in contrast with the present work where overcomplete frame representations are considered, these works are limited to wavelet bases for which the hyper-parameter estimation problem is much easier to handle. Other interesting works concerning the use of MCMC methods for generating sparse representations [Févotte and Godsill, 2006; Wolfe and Godsill, 2003] assume Gaussian noise models, which may facilitate the derivation of the proposed sampler, especially when a mixture of Gaussian prior is chosen. Alternative Bayesian approaches have also been proposed in [Kowalski and Torrésani, 2008; Molla and Torrésani, 2005] for some specific forms of frame representations.

This chapter is organized as follows. Section 5.2 presents the general problem formulation. The hierarchical Bayesian model proposed for frame representation is introduced in Section 5.3. Two algorithms for sampling the posterior distribution are proposed in Section 5.4. An extension of the proposed approach to account for more general priors is also presented in Section 5.5. To illustrate the effectiveness of these algorithms, experiments on both synthetic and real world data are presented in Section 5.6. In this section, applications to image recovery problems are also considered. Finally, some conclusions are drawn in Section 5.7.

5.2 Problem Formulation

An observed signal $\boldsymbol{y} \in \mathbb{R}^L$ can be written according to its frame representation (FR) involving coefficients $\boldsymbol{x} \in \mathbb{R}^K$ as follows

$$\boldsymbol{y} = F^* \boldsymbol{x} + \boldsymbol{n} \tag{5.1}$$

where $\boldsymbol{n}$ is the error between the observed signal $\boldsymbol{y}$ and its FR $F^*\boldsymbol{x}$. We recall here that F and F^* are the linear frame and adjoint frame operators, respectively (see Section 3.2.3). This error is modeled by imposing that $\boldsymbol{x}$ belongs to the closed convex set

$$C_\delta = \{\boldsymbol{x} \in \mathbb{R}^K \mid N(\boldsymbol{y} - F^*\boldsymbol{x}) \leq \delta\} \tag{5.2}$$

where $\delta \in [0, \infty[$ is some error bound and $N(.)$ can be any norm on $\mathbb{R}^L$.
In signal/image recovery problems, $\boldsymbol{n}$ is nothing but an additive noise that corrupts the measured data. By adopting a probabilistic approach, $\boldsymbol{y}$ and $\boldsymbol{x}$ are assumed to be realizations of random vectors $\boldsymbol{Y}$ and $\boldsymbol{X}$. In this context, our goal is to characterize the probability distribution of $\boldsymbol{X}|\boldsymbol{Y}$, by considering some parametric probabilistic model and by estimating the associated hyper-parameters.
A useful example where this characterization may be of great interest is frame-based signal/image denoising in a Bayesian framework. Actually, denoising in the wavelet domain using wavelet frame decompositions has already been investigated since the seminal work in [Donoho, 1995] as this kind of representation provides sparse description of regular signals. The related hyper-parameters have then to be estimated. When F is bijective and $\delta = 0$, this estimation can be performed by inverting the transform so as to deduce $\boldsymbol{x}$ from $\boldsymbol{y}$ and by resorting to standard estimation techniques on $\boldsymbol{x}$. However, for redundant frames, F^* is not bijective, which makes the hyper-parameter estimation problem more difficult. This chapter presents hierarchical Bayesian algorithms to address this issue.

5.3 Hierarchical Bayesian Model

In a Bayesian framework, we first need to define prior distributions for the frame coefficients. For instance, this prior may be chosen so as to promote the sparsity of the representation. In the following, $f(\boldsymbol{x}|\boldsymbol{\theta})$ denotes the pdf of the frame coefficients that depends on an unknown hyper-parameter vector $\boldsymbol{\theta}$ and $f(\boldsymbol{\theta})$ is the a priori pdf for the hyper-parameter vector $\boldsymbol{\theta}$. In compliance with the observation model in Eq. (5.1) and the constraint in Eq. (5.2), $\boldsymbol{n}$ is assumed to be uniformly distributed on the ball

$$B_{\boldsymbol{0},\delta} = \{\boldsymbol{a} \in \mathbb{R}^L \mid N(\boldsymbol{a}) \leq \delta\}. \tag{5.3}$$

From Eq. (5.1), it can be deduced that $f(\boldsymbol{y}|\boldsymbol{x})$ is the uniform pdf on the closed convex ball $B_{F^*\boldsymbol{x},\delta}$ defined as

$$B_{F^*\boldsymbol{x},\delta} = \{\boldsymbol{y} \in \mathbb{R}^L \mid N(\boldsymbol{y} - F^*\boldsymbol{x}) \leq \delta\}. \tag{5.4}$$

Denoting by $\boldsymbol{\Theta}$ the random vector associated with the hyper-parameter vector $\boldsymbol{\theta}$ and using the hierarchical structure between $\boldsymbol{Y}, \boldsymbol{X}$ and $\boldsymbol{\Theta}$, the conditional pdf of $(\boldsymbol{X}, \boldsymbol{\Theta})$ given $\boldsymbol{Y}$ can be written as

$$f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{x}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) \tag{5.5}$$

where $\propto$ means *proportional to.*

In this work, we assume that frame coefficients are a priori independent with marginal GG distributions. This assumption has been successfully used in many studies [Mallat, 1989; Joshi et al., 1997; Simoncelli and Adelson, 1996; Moulin and Liu, 1998; Do and Vetterli, 2002] and leads to the following frame coefficient prior

$$f(x_k|\alpha_k, \beta_k) = \frac{\beta_k}{2\alpha_k \Gamma(1/\beta_k)} \exp\left(-\frac{|x_k|^{\beta_k}}{\alpha_k^{\beta_k}}\right) \tag{5.6}$$

where $\alpha_k > 0, \beta_k > 0$ (with $k \in \{1, \ldots, K\}$) are the scale and shape parameters associated with x_k, which is the kth component of the frame coefficient vector $\boldsymbol{x}$ and $\Gamma(.)$ is the Gamma function. Note that small values of the shape parameters are appropriate for modeling sparse signals. When $\forall k \in \{1, \ldots, K\}$, $\beta_k = 1$, a Laplace prior is obtained, which was shown to play a central role in sparse signal recovery [Seeger and Nickisch, 2008] and compressed sensing [Babacan et al., 2009].

By introducing $\forall k \in \{1, \ldots, K\}$, $\gamma_k = \alpha_k^{\beta_k}$, the frame prior can be rewritten as[1]

$$f(x_k|\gamma_k, \beta_k) = \frac{\beta_k}{2\gamma_k^{1/\beta_k} \Gamma(1/\beta_k)} \exp\left(-\frac{|x_k|^{\beta_k}}{\gamma_k}\right). \tag{5.7}$$

The distribution of a frame coefficient generally differs from one coefficient to another. However, some frame coefficients can have very similar distributions (that can be defined by the same hyper-parameters β_k and γ_k). As a consequence, we propose to split the frame coefficients into G different groups. The gth group will be parameterized by a unique hyper-parameter vector denoted as $\boldsymbol{\theta}_g = (\beta_g, \gamma_g)$ (after the reparameterization mentioned above). In this case, the frame prior can be expressed as

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{g=1}^{G} \left[\left(\frac{\beta_g}{2\gamma_g^{1/\beta_g} \Gamma(1/\beta_g)} \right)^{n_g} \exp\left(-\frac{1}{\gamma_g} \sum_{k \in S_g} |x_k|^{\beta_g} \right) \right] \tag{5.8}$$

where the summation covers the index set S_g of the elements of the gth group containing n_g elements and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_G)$. Note that in our simulations, each group g will correspond to a given wavelet subband. A coarser classification may be made when using multiscale frame representations by considering that all the frame coefficients at a given resolution level belong to a same group.

The hierarchical Bayesian model for the frame decomposition is completed by the following improper hyperprior

$$\begin{aligned} f(\boldsymbol{\theta}) &= \prod_{g=1}^{G} f(\boldsymbol{\theta}_g) = \prod_{g=1}^{G} [f(\gamma_g) f(\beta_g)] \\ &\propto \prod_{g=1}^{G} \left[\frac{1}{\gamma_g} 1_{\mathbb{R}^+}(\gamma_g) 1_{[0,3]}(\beta_g) \right] \end{aligned} \tag{5.9}$$

where $1_A(\xi)$ is the function defined on $A \subset \mathbb{R}$ by $1_A(\xi) = 1$ if $\xi \in A$ and $1_A(\xi) = 0$ otherwise. The motivations for using this kind of prior are summarized below:

---

[1] The interest of this new parameterization will be clarified in Section 5.4.

- the interval $[0, 3]$ covers all possible values of β_g encountered in practical applications. Moreover, there is no additional information about the parameter β_g.
- The prior for the parameter γ_g is a Jeffrey's distribution that reflects the absence of knowledge about this parameter. This kind of prior is often used for scale parameters [Jeffreys, 1946].

The resulting posterior distribution is therefore given by

$$f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto 1_{C_\delta}(\boldsymbol{x}) \prod_{g=1}^{G} \left[\left(\frac{\beta_g}{2\gamma_g^{1/\beta_g}\Gamma(1/\beta_g)} \right)^{n_g} \exp\left(-\frac{1}{\gamma_g} \sum_{k \in S_g} |x_k|^{\beta_g} \right) \left(\frac{1}{\gamma_g} 1_{\mathbb{R}^+}(\gamma_g) 1_{[0,3]}(\beta_g) \right) \right]. \tag{5.10}$$

The Bayesian estimators (e.g., the maximum a posteriori (MAP) or minimum mean square error (MMSE) estimators) associated with the posterior distribution in Eq. (5.10) have no simple closed-form expression. The next section studies different sampling strategies that allow one to generate samples distributed according to the posterior distribution in Eq. (5.10). The generated samples will be used to estimate the unknown model parameter and hyper-parameter vectors $\boldsymbol{x}$ and $\boldsymbol{\theta}$.

5.4 Sampling strategies

This section proposes different MCMC methods to generate samples distributed according to the posterior $f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ defined in Eq. (5.10).

5.4.1 Hybrid Gibbs Sampler

A very standard strategy to sample according to Eq. (5.10) is provided by the Gibbs sampler (GS). GS iteratively generates samples distributed according to conditional distributions associated with the target distribution. More precisely, the basic GS iteratively generates samples distributed according to $f(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ and $f(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y})$.

5.4.1.1 Sampling the frame coefficients

Straightforward calculations yield the following conditional distribution

$$f(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \propto 1_{C_\delta}(\boldsymbol{x}) \prod_{g=1}^{G} \exp\left(-\frac{1}{\gamma_g} \sum_{k \in S_g} |x_k|^{\beta_g} \right) \tag{5.11}$$

where C_δ is defined in Eq. (5.2). This conditional distribution is a product of GG distributions truncated on C_δ. Actually, sampling according to this truncated distribution is not always easy to perform since the adjoint frame operator F^* is usually of large dimension. However, two alternative sampling strategies are detailed in what follows.

Naive sampling This sampling method proceeds by sampling according to independent GG distributions

$$\prod_{g=1}^{G} \exp\Big(-\frac{1}{\gamma_g}\sum_{k\in S_g}|x_k|^{\beta_g}\Big) \tag{5.12}$$

and then accepting the proposed candidate $\boldsymbol{x}$ only if $N(\boldsymbol{y}-F^*\boldsymbol{x})\le\delta$. This method can be used for any frame decomposition and any norm. However, it can be quite inefficient because of a very low acceptance ratio, especially when δ takes small values.

Gibbs sampler This sampling method is designed to sample more efficiently from the conditional distribution in Eq. (5.11) when the considered frame is the union of M orthonormal bases and $N(\cdot)$ is the Euclidean norm. In this case, the analysis frame operator and the corresponding adjoint can be written as $F=\begin{bmatrix}F_1\\ \vdots\\ F_M\end{bmatrix}$ and $F^*=[F_1^*\dots F_M^*]$, respectively, where $\forall m\in\{1,\dots,M\}$, F_m is the decomposition operator onto the mth orthonormal basis such as $F_m^*F_m=F_mF_m^*=\mathrm{Id}$. In what follows, we will decompose every $\boldsymbol{x}\in\mathbb{R}^K$ with $K=ML$ as $\boldsymbol{x}=[\boldsymbol{x}_1^\top,\dots,\boldsymbol{x}_M^\top]^\top$ where $\boldsymbol{x}_m\in\mathbb{R}^L$, for every $m\in\{1,\dots,M\}$. The GS for the generation of frame coefficients draws vectors according to the conditional distribution $f(\boldsymbol{x}_n|\boldsymbol{x}_{-n},\boldsymbol{y},\boldsymbol{\theta})$ under the constraint $N(\boldsymbol{y}-F^*\boldsymbol{x})\le\delta$, where $\boldsymbol{x}_{-n}$ is the reduced size vector of dimension $\mathbb{R}^{K-L}$ built from $\boldsymbol{x}$ by removing the nth vector $\boldsymbol{x}_n$. If $N(\cdot)$ is the Euclidean norm, we have for every $n\in\{1,\dots,M\}$,

$$\begin{aligned}
&N(\boldsymbol{y}-\sum_{m=1}^{M}F_m^*\boldsymbol{x}_m)\le\delta\\
\Leftrightarrow&\parallel F_n^*(F_n\boldsymbol{y}-\sum_{m=1}^{M}F_nF_m^*\boldsymbol{x}_m)\parallel\le\delta\\
\Leftrightarrow&\parallel F_n\boldsymbol{y}-\sum_{m\ne n}F_nF_m^*\boldsymbol{x}_m-\boldsymbol{x}_n\parallel\le\delta\quad(\text{since }\ \forall\boldsymbol{z}\in\mathbb{R}^L,\ \ \parallel F_n^*\boldsymbol{z}\parallel=\parallel\boldsymbol{z}\parallel)\\
\Leftrightarrow&N(\boldsymbol{x}_n-\boldsymbol{c}_n)\le\delta,
\end{aligned}\tag{5.13}$$

where

$$\boldsymbol{c}_n=F_n\boldsymbol{y}-\sum_{m\ne n}F_nF_m^*\boldsymbol{x}_m. \tag{5.14}$$

To sample each $\boldsymbol{x}_n$, we propose to use an MH step whose proposal distribution is supported on the ball $B_{\boldsymbol{c}_n,\delta}$ defined by

$$B_{\boldsymbol{c}_n,\delta}=\{\boldsymbol{a}\in\mathbb{R}^L\mid N(\boldsymbol{a}-\boldsymbol{c}_n)\le\delta\}. \tag{5.15}$$

Random generation from a pdf q_δ defined on $B_{\boldsymbol{0},\delta}$ is described in Section 5.4.1.3. Having a closed form expression of this pdf is important to be able to calculate the acceptance ratio of the MH move. To take into account the value of $\boldsymbol{x}_n^{(i-1)}$ obtained at the previous iteration $(i-1)$, it may however be preferable to choose a proposal distribution supported on a restricted ball of radius $\eta \in]0, \delta[$ containing $\boldsymbol{x}_n^{(i-1)}$. This strategy similar to the random walk MH algorithm [Robert and Castella, 2004, p. 287] results in a better exploration of regions associated with large values of the conditional distribution $f(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$. More precisely, we propose to choose a proposal distribution defined on $B_{\hat{\boldsymbol{x}}_n^{(i-1)}, \eta}$, where

$$\hat{\boldsymbol{x}}_n^{(i-1)} = P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n) + \boldsymbol{c}_n \tag{5.16}$$

and P is the projection onto the ball $B_{\boldsymbol{0}, \delta-\eta}$ defined as

$$\forall \boldsymbol{a} \in \mathbb{R}^L, \qquad P(\boldsymbol{a}) = \begin{cases} \boldsymbol{a} & \text{if } N(\boldsymbol{a}) \leq \delta - \eta \\ \dfrac{\delta - \eta}{N(\boldsymbol{a})} \boldsymbol{a} & \text{otherwise.} \end{cases} \tag{5.17}$$

This choice of the center of the ball guarantees that $B_{\hat{\boldsymbol{x}}_n^{(i-1)}, \eta} \subset B_{\boldsymbol{c}_n, \delta}$.

Proof :

Let $\boldsymbol{x} \in B_{\hat{\boldsymbol{x}}_n^{(i-1)}, \eta}$. By using Eq. (5.16), we have therefore :

$$\|\hat{\boldsymbol{x}}_n^{(i-1)} - \boldsymbol{x}\| \leq \eta \Longleftrightarrow \|P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n) + \boldsymbol{c}_n - \boldsymbol{x}\| \leq \eta. \tag{5.18}$$

Consequently, we can write:

$$\begin{aligned} \|\boldsymbol{c}_n - \boldsymbol{x}\| &= \|\hat{\boldsymbol{x}}_n^{(i-1)} - P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n) - \boldsymbol{x}\| \\ &= \|\hat{\boldsymbol{x}}_n^{(i-1)} - \boldsymbol{c}_n + \boldsymbol{c}_n - P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n) - \boldsymbol{x}\| \\ &\leq \|\hat{\boldsymbol{x}}_n^{(i-1)} - \boldsymbol{c}_n\| + \|\boldsymbol{c}_n - P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n) - \boldsymbol{x}\| \\ &\leq \|P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n)\| + \|\boldsymbol{c}_n - P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n) - \boldsymbol{x}\| \\ &\leq \delta - \eta + \eta, \end{aligned} \tag{5.19}$$

which means that $\boldsymbol{x} \in B_{\boldsymbol{c}_n, \delta}$. ■

Moreover, any point of $B_{\boldsymbol{c}_n, \delta}$ can be reached after consecutive draws in $B_{\hat{\boldsymbol{x}}_n^{(i-1)}, \eta}$. Note that the radius η has to be adjusted to ensure a good exploration of $B_{\boldsymbol{c}_n, \delta}$. In practice, it may also be interesting to fix a small enough value of η (compared with δ) so as to improve the acceptance ratio.

Remark:

Alternatively, a GS can be used to draw successively the L elements $(x_{n,l})_{1 \leq l \leq L}$ of $\boldsymbol{x}_n$ under the following constraint for every $l \in \{1, \ldots, L\}$

$$\| \boldsymbol{x}_n - \boldsymbol{c}_n \| \leq \delta \Leftrightarrow -\sqrt{\delta^2 - \sum_{k \neq l} (x_{n,k} - c_{n,k})^2} \leq x_{n,l} - c_{n,l} \leq \sqrt{\delta^2 - \sum_{k \neq l} (x_{n,k} - c_{n,k})^2},$$

where $c_{n,k}$ is the kth element of the vector $\boldsymbol{c}_n$. However, this method is very time-consuming since it proceeds sequentially for each component of the high dimensional vector $\boldsymbol{x}$.

5.4.1.2 Sampling the hyper-parameter vector

Instead of sampling $\boldsymbol{\theta}$ according to $f(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})$, we propose to iteratively sample according to $f(\gamma_g|\beta_g,\boldsymbol{x},\boldsymbol{y})$ and $f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y})$. Straightforward calculations allow us to obtain the following results

$$f(\gamma_g|\beta_g,\boldsymbol{x},\boldsymbol{y}) \propto \gamma_g^{-\frac{n_g}{\beta_g}-1} \exp\Big(-\frac{1}{\gamma_g}\sum_{k\in S_g}|x_k|^{\beta_g}\Big)\mathbf{1}_{\mathbb{R}^+}(\gamma_g), \tag{5.20}$$

$$f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y}) \propto \frac{\beta_g^{n_g}\mathbf{1}_{[0,3]}(\beta_g)}{\gamma_g^{n_g/\beta_g}\Big[\Gamma\Big(1/\beta_g\Big)\Big]^{n_g}} \exp\Big(\sum_{k\in S_g}\frac{-|x_k|^{\beta_g}}{\gamma_g}\Big). \tag{5.21}$$

Consequently, due to the new parameterization introduced in Eq. (5.7), $f(\gamma_g|\beta_g,\boldsymbol{x},\boldsymbol{y})$ is the pdf of the inverse gamma distribution $\mathcal{IG}\left(\frac{n_g}{\beta_g},\sum_{k\in S_g}|x_k|^{\beta_g}\right)$ that is easy to sample. Conversely, it is more difficult to sample according to the truncated pdf $f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y})$. This is achieved by using an MH move whose proposal $q(\beta_g \mid \beta_g^{(i-1)})$ is a Gaussian distribution truncated on the interval $[0,3]$ with standard deviation $\sigma_{\beta_g}=0.05$ [Dobigeon and Tourneret, 2007]. Note that the mode of this distribution is the value of the parameter $\beta_g^{(i-1)}$ at the previous iteration $(i-1)$. The resulting method is the hybrid GS summarized in Algorithm 10.

Although this algorithm is intuitive and simple to implement, it must be pointed out that it was derived under the restrictive assumption that the considered frame is the union of M orthonormal bases. When these assumptions do not hold, another algorithm proposed in Section 5.4.2 allows us to sample frame coefficients and the related hyper-parameters by exploiting algebraic properties of frames.

5.4.1.3 Sampling on the unit ℓ_p ball

This section explains how to sample vectors in the unit ℓ_p ball ($p\in]0,+\infty]$) of $\mathbb{R}^L$. First, it is interesting to note that sampling on the unit ball can be easily performed in the particular case $p=+\infty$, by sampling independently along each space coordinate according to a distribution on the interval $[-1,1]$. Thus, this section focuses on the more difficult problem associated with a finite value of p. In the following, $\|\cdot\|_p$ denotes the ℓ_p norm. We recall the following theorem:

Theorem 5.4.1 [Gupta and Song, 1997]
Let $\boldsymbol{A}=[A_1,\ldots,A_{L'}]^\top$ be the random vector of iid components which have the following $\mathrm{GG}(p^{1/p},p)$ pdf

$$\forall a\in\mathbb{R},\ f(a)=\frac{p^{1-1/p}}{2\Gamma(1/p)}\exp\Big(-\frac{|a|^p}{p}\Big). \tag{5.22}$$

Algorithm 10 Sampler 1: hybrid GS to simulate according to $f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ (superscript $\cdot^{(i)}$ indicates values computed at iteration number i).

```
Initialize with some $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_g^{(0)})_{1 \leq g \leq G} = (\gamma_g^{(0)}, \beta_g^{(0)})_{1 \leq g \leq G}$ and $\boldsymbol{x}^{(0)} \in C_\delta$, and set $i = 1$.
repeat
    Sampling $\boldsymbol{x}$:
    for $n = 1$ to $M$ do
        Compute $\boldsymbol{c}_n^{(i)} = F_n\Big(\boldsymbol{y} - \sum_{m<n} F_m^* \boldsymbol{x}_m^{(i)} - \sum_{m>n} F_m^* \boldsymbol{x}_m^{(i-1)}\Big)$
        and $\hat{\boldsymbol{x}}_n^{(i-1)} = P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n^{(i)}) + \boldsymbol{c}_n^{(i)}$.
        Simulate $\boldsymbol{x}_n^{(i)}$ as follows:
          • Generate $\widetilde{\boldsymbol{x}}_n^{(i)} \sim q_\eta(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n^{(i-1)})$ where $q_\eta$ is defined on $B_{\boldsymbol{0},\eta}$ (see Section 5.4.1.3).
          • Compute the ratio
              $$r(\widetilde{\boldsymbol{x}}_n^{(i)}, \boldsymbol{x}_n^{(i-1)}) = \frac{f(\widetilde{\boldsymbol{x}}_n^{(i)}|\boldsymbol{\theta}^{(i-1)},(\boldsymbol{x}_m^{(i)})_{m<n},(\boldsymbol{x}_m^{(i-1)})_{m>n},\boldsymbol{y})\ q_\eta\Big(\boldsymbol{x}_n^{(i-1)}-P(\widetilde{\boldsymbol{x}}_n^{(i)}-\boldsymbol{c}_n^{(i)})-\boldsymbol{c}_n^{(i)}\Big)}{f(\boldsymbol{x}_n^{(i-1)}|\boldsymbol{\theta}^{(i-1)},(\boldsymbol{x}_m^{(i)})_{m<n},(\boldsymbol{x}_m^{(i-1)})_{m>n},\boldsymbol{y})\ q_\eta\Big(\widetilde{\boldsymbol{x}}_n^{(i)}-\hat{\boldsymbol{x}}_n^{(i-1)}\Big)}$$
            and accept the proposed candidate with the probability $\min\{1, r(\widetilde{\boldsymbol{x}}_n^{(i)}, \boldsymbol{x}_n^{(i-1)})\}$.
    end for
    Sampling $\boldsymbol{\theta}$:
    for $g = 1$ to $G$ do
        Generate $\gamma_g^{(i)} \backsim \mathcal{IG}\left(\frac{n_g}{\beta_g^{(i-1)}}, \sum_{k \in S_g} |x_k^{(i)}|^{\beta_g^{(i-1)}}\right)$.
        Simulate $\beta_g^{(i)}$ as follows:
          • Generate $\widetilde{\beta}_g^{(i)} \backsim q(\beta_g \mid \beta_g^{(i-1)})$.
          • Compute the ratio
              $$r(\widetilde{\beta}_g^{(i)}, \beta_g^{(i-1)}) = \frac{f(\widetilde{\beta}_g^{(i)}|\gamma_g^{(i)}, \boldsymbol{x}^{(i)}, \boldsymbol{y}) q(\beta_g^{(i-1)} \mid \widetilde{\beta}_g^{(i)})}{f(\beta_g^{(i-1)}|\gamma_g^{(i)}, \boldsymbol{x}^{(i)}, \boldsymbol{y}) q(\widetilde{\beta}_g^{(i)} \mid \beta_g^{(i-1)})}$$
            and accept the proposed candidate with the probability $\min\{1, r(\widetilde{\beta}_g^{(i)}, \beta_g^{(i-1)})\}$.
    end for
    Set $i \leftarrow i + 1$.
until Convergence
```

Let $\boldsymbol{U} = [U_1, \ldots, U_{L'}]^\intercal = \boldsymbol{A}/\|\boldsymbol{A}\|_p$. Then, the random vector $\boldsymbol{U}$ is uniformly distributed on the surface of the ℓ_p unit sphere of $\mathbb{R}^{L'}$ and the joint pdf of $U_1, \ldots, U_{L'-1}$ is

$$f(u_1, \ldots, u_{L'-1}) = \frac{p^{L'-1}\Gamma(L'/p)}{2^{L'-1}(\Gamma(1/p))^{L'}} \left(1 - \sum_{k=1}^{L'-1} |u_k|^p\right)^{(1-p)/p} 1_{D_{p,L'}}(u_1, ..., u_{L'-1}) \quad (5.23)$$

where $D_{p,L'} = \{(u_1, ..., u_{L'-1}) \in \mathbb{R}^{L'-1} \mid \sum_{k=1}^{L'-1} |u_k|^p < 1\}$.

The uniform distribution on the unit ℓ_p sphere of $\mathbb{R}^{L'}$ will be denoted by $\mathcal{U}(L', p)$. The construction of a random vector distributed within the ℓ_p ball of $\mathbb{R}^L$ with $L < L'$ can be derived from Theorem 5.4.1 as expressed below:

Theorem 5.4.2 [Gupta and Song, 1997]
Let $\boldsymbol{U} = [U_1, \ldots, U_{L'}]^\intercal \sim \mathcal{U}(L', p)$. For every $L \in \{1, \ldots, L'-1\}$, the pdf of $\boldsymbol{V} = [U_1, \ldots, U_L]^\intercal$ is given by

$$q_1(u_1, \cdots, u_L) = \frac{p^L\Gamma(L'/p)\left(1 - \sum_{k=1}^{L} |u_k|^p\right)^{(L'-L)/p-1}}{2^L(\Gamma(1/p))^L\Gamma((L'-L)/p)} 1_{D_{p,L+1}}(u_1, ..., u_L). \quad (5.24)$$

In particular, if $p \in \mathbb{N}^*$ and $L' = L + p$, we obtain the uniform distribution on the unit ℓ_p ball of $\mathbb{R}^L$.
Sampling from a distribution q_η on the ℓ_p ball of radius $\eta > 0$ is straightforwardly deduced by scaling $\boldsymbol{V}$.

5.4.2 Hybrid MH sampler using algebraic properties of frame representations

As a direct generation of samples according to $f(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ is generally impossible, we propose here an alternative that replaces the Gibbs move by an MH move. This MH move aims at sampling globally a candidate $\boldsymbol{x}$ according to a proposal distribution. This candidate is accepted or rejected with the standard MH acceptance ratio. The efficiency of the MH move strongly depends on the choice of the proposal distribution for $\boldsymbol{x}$. We denote as $\boldsymbol{x}^{(i)}$ the ith accepted sample of the algorithm and $q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)})$ the proposal that is used to generate a candidate at iteration i. The main difficulty for choosing $q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)})$ stems from the fact that it must guarantee that $\boldsymbol{x} \in C_\delta$ (as mentioned in Section 5.2) while yielding a tractable expression of $q(\boldsymbol{x}^{(i-1)} \mid \boldsymbol{x})/q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)})$.

For this reason, we propose to exploit the algebraic properties of frame representations. More precisely, any frame coefficient vector can be decomposed as $\boldsymbol{x} = \boldsymbol{x}_H + \boldsymbol{x}_{H^\perp}$, where $\boldsymbol{x}_H$ and $\boldsymbol{x}_{H^\perp}$ are realizations of random vectors taking their values in $H = \mathrm{Ran}(F)$ and $H^\perp = [\mathrm{Ran}(F)]^\perp = \mathrm{Null}(F^*)$, respectively[2]. The proposal distribution used here allows us to generate samples $\boldsymbol{x}_H \in H$ and $\boldsymbol{x}_{H^\perp} \in H^\perp$. More precisely, the following separable

---

[2] We recall that the range of F is $\mathrm{Ran}(F) = \{\boldsymbol{x} \in \mathbb{R}^K | \exists \boldsymbol{y} \in \mathbb{R}^L, F\boldsymbol{y} = \boldsymbol{x}\}$ and the null space of F^* is $\mathrm{Null}(F^*) = \{\boldsymbol{x} \in \mathbb{R}^K | F^*\boldsymbol{x} = \boldsymbol{0}\}$.

form of the proposal pdf will be considered

$$q(\boldsymbol{x} \mid \boldsymbol{x}^{(i)}) = q\left(\boldsymbol{x}_H \mid \boldsymbol{x}_H^{(i-1)}\right) q\left(\boldsymbol{x}_{H^\perp} \mid \boldsymbol{x}_{H^\perp}^{(i-1)}\right) \tag{5.25}$$

where $\boldsymbol{x}_H^{(i-1)} \in H$, $\boldsymbol{x}_{H^\perp}^{(i-1)} \in H^\perp$ and $\boldsymbol{x}^{(i-1)} = \boldsymbol{x}_H^{(i-1)} + \boldsymbol{x}_{H^\perp}^{(i-1)}$. In other words, independent sampling of $\boldsymbol{x}_H$ and $\boldsymbol{x}_{H^\perp}$ will be performed.

If we consider the decomposition $\boldsymbol{x} = \boldsymbol{x}_H + \boldsymbol{x}_{H^\perp}$, sampling $\boldsymbol{x}$ in C_δ is equivalent to sampling $\boldsymbol{\lambda} \in \overline{C}_\delta$, where $\overline{C}_\delta = \{\boldsymbol{\lambda} \in \mathbb{R}^L | N(\boldsymbol{y} - F^*F\boldsymbol{\lambda}) \leq \delta\}$. Indeed, we can write $\boldsymbol{x}_H = F\boldsymbol{\lambda}$ where $\boldsymbol{\lambda} \in \mathbb{R}^L$ and, since $\boldsymbol{x}_{H^\perp} \in \text{Null}(F^*)$, $F^*\boldsymbol{x} = F^*F\boldsymbol{\lambda}$. Sampling $\boldsymbol{\lambda}$ in $\overline{C}_\delta$ can be easily achieved, e.g., by generating $\boldsymbol{u}$ from a distribution on the ball $B_{\boldsymbol{y},\delta}$ and by taking $\boldsymbol{\lambda} = (F^*F)^{-1}\boldsymbol{u}$.
To make the sampling of $\boldsymbol{x}_H$ at iteration i more efficient, taking into account the sampled value at the previous iteration $\boldsymbol{x}_H^{(i-1)} = F\boldsymbol{\lambda}^{(i-1)} = F(F^*F)^{-1}\boldsymbol{u}^{(i-1)}$ may be interesting. Similarly to random walk generation techniques, we proceed by generating $\boldsymbol{u}$ in $B_{\hat{\boldsymbol{u}}^{(i-1)},\eta}$ where $\eta \in]0,\delta[$ and $\hat{\boldsymbol{u}}^{(i-1)} = P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) + \boldsymbol{y}$. This allows us to draw a vector $\boldsymbol{u}$ such that $\boldsymbol{x}_H = F(F^*F)^{-1}\boldsymbol{u} \in C_\delta$ and $N(\boldsymbol{u} - \boldsymbol{u}^{(i-1)}) \leq 2\eta$.
Proof :

i) Proof that $\boldsymbol{x}_H = F(F^*F)^{-1}\boldsymbol{u} \in C_\delta$:
We can first write that $N(\boldsymbol{y} - F^*\boldsymbol{x}_H) = N(\boldsymbol{y} - F^*F(F^*F)^{-1}\boldsymbol{u}) = N(\boldsymbol{y} - \boldsymbol{u})$. On the other hand, we can write

$$\begin{aligned} N(\boldsymbol{y} - \boldsymbol{u}) =& N(\boldsymbol{y} + P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) - P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) - \boldsymbol{u}) \\ &\leq N(\boldsymbol{y} + P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) - \boldsymbol{u}) + N(P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y})) \\ &\leq N(\hat{\boldsymbol{u}}^{(i-1)} - \boldsymbol{u}) + N(P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y})). \end{aligned} \tag{5.26}$$

And since $\boldsymbol{u} \in B_{\hat{\boldsymbol{u}}^{(i-1)},\eta}$, which means that $N(\hat{\boldsymbol{u}}^{(i-1)} - \boldsymbol{u}) \leq \eta$, we have therefore

$$\begin{aligned} N(\boldsymbol{y} - \boldsymbol{u}) &\leq \eta + \delta - \eta \\ &\leq \delta. \end{aligned} \tag{5.27}$$

It turns out that $N(\boldsymbol{y} - \boldsymbol{x}_H) \leq \delta$, which means that $\boldsymbol{x}_H \in C_\delta$.

ii) Proof that $N(\boldsymbol{u} - \boldsymbol{u}^{(i-1)}) \leq 2\eta$:
Let $\boldsymbol{u} \in B_{\hat{\boldsymbol{u}}^{(i-1)},\eta}$, which means that $N(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)}) \leq \eta$. We can therefore write:

$$\begin{aligned} N(\boldsymbol{u} - \boldsymbol{u}^{(i-1)}) &\leq N(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)} + \hat{\boldsymbol{u}}^{(i-1)} - \boldsymbol{u}^{(i-1)}) \\ &\leq N(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)}) + N(P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) + \boldsymbol{y} - \boldsymbol{u}^{(i-1)}). \end{aligned} \tag{5.28}$$

If we denote by $\boldsymbol{a} = \boldsymbol{u}^{(i-1)} - \boldsymbol{y}$, we have

$$N(\boldsymbol{a} - P(\boldsymbol{a})) = \begin{cases} N(\boldsymbol{a} - \boldsymbol{a}) = 0 & \text{if } P(\boldsymbol{a}) = \boldsymbol{a} \\ N(\boldsymbol{a} - \dfrac{\delta - \eta}{N(\boldsymbol{a})}\boldsymbol{a}) = N(\dfrac{N(\boldsymbol{a}) - \delta + \eta}{N(\boldsymbol{a})}\boldsymbol{a}) = N(\boldsymbol{a}) - \delta + \eta & \text{otherwise.} \end{cases} \tag{5.29}$$

And since at the previous iteration $\boldsymbol{x}_H^{(i-1)} = F(F^*F)^{-1}\boldsymbol{u}^{(i-1)} \in C_\delta$, we have

$$N(\boldsymbol{a}) = N(\boldsymbol{y} - \boldsymbol{u}^{(i-1)}) = N(\boldsymbol{y} - F^*F(F^*F)^{-1}\boldsymbol{u}^{(i-1)}) = N(\boldsymbol{y} - F^*\boldsymbol{x}_H^{(i-1)}) \leq \delta. \tag{5.30}$$

It turns out from Eqs. (5.29) and (5.30) that $N(\boldsymbol{a} - P(\boldsymbol{a})) \leq \eta$. Eq. (5.28) can then be rewritten as:

$$N(\boldsymbol{u} - \boldsymbol{u}^{(i-1)}) \leq \eta + N(\boldsymbol{a} - P(\boldsymbol{a})) \leq 2\eta. \tag{5.31}$$

■

The generation of $\boldsymbol{u}$ can then be performed as explained in Section 5.4.1.3 provided that $N(\cdot)$ is an ℓ_p norm with $p \in [1, +\infty]$.

Once we have simulated $\boldsymbol{x}_H = F\boldsymbol{\lambda} \in H \cap C_\delta$ (which ensures that $\boldsymbol{x}$ is in C_δ), $\boldsymbol{x}_{H^\perp}$ has to be sampled as an element of $H^\perp$. Since $\boldsymbol{y} = F^*\boldsymbol{x} + \boldsymbol{n} = F^*\boldsymbol{x}_H + \boldsymbol{n}$, there is no information in $\boldsymbol{y}$ about $\boldsymbol{x}_{H^\perp}$. As a consequence, we propose to sample $\boldsymbol{x}_H$ by drawing $\boldsymbol{z}$ according to the Gaussian distribution $\mathcal{N}(\boldsymbol{x}^{(i-1)}, \sigma_{\boldsymbol{x}}^2\boldsymbol{I})$ and by projecting $\boldsymbol{z}$ onto $H^\perp$, i.e.,

$$\boldsymbol{x}_{H^\perp} = \Pi_{H^\perp}\boldsymbol{z} \tag{5.32}$$

where $\Pi_{H^\perp} = \mathrm{I} - F(F^*F)^{-1}F^*$ is the orthogonal projection operator onto $H^\perp$.[3]

Let us now derive the expression of the proposal pdf. It can be noticed that, if $K > L$, there exists a linear operator $F_\perp$ from $\mathbb{R}^{K-L}$ to $\mathbb{R}^K$ which is semi-orthogonal (i.e., $F_\perp^*F_\perp = \mathrm{I}$) and orthogonal to F (i.e., $F_\perp^*F = 0$), such that

$$\boldsymbol{x} = \underbrace{F\boldsymbol{\lambda}}_{\boldsymbol{x}_H} + \underbrace{F_\perp\boldsymbol{\lambda}_\perp}_{\boldsymbol{x}_{H^\perp}} \tag{5.33}$$

and $\boldsymbol{\lambda}_\perp = F_\perp^*\boldsymbol{x} \in \mathbb{R}^{K-L}$. Standard rules on bijective linear transforms of random vectors lead to

$$q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)}) = |\det\left([F \quad F_\perp]\right)|^{-1} q(\boldsymbol{\lambda} \mid \boldsymbol{x}^{(i-1)}) q(\boldsymbol{\lambda}_\perp \mid \boldsymbol{x}^{(i-1)}) \tag{5.34}$$

where, due to the bijective linear mapping between $\boldsymbol{\lambda}$ and $\boldsymbol{u} = F^*F\boldsymbol{\lambda}$

$$q(\boldsymbol{\lambda} \mid \boldsymbol{x}^{(i-1)}) = \det(FF^*)\; q_\eta(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)}) \tag{5.35}$$

and $q(\boldsymbol{\lambda}_\perp \mid \boldsymbol{x}^{(i-1)})$ is the pdf of the Gaussian distribution $\mathcal{N}(\boldsymbol{\lambda}_\perp^{(i-1)}, \sigma_{\boldsymbol{x}}^2\boldsymbol{I})$ with mean $\boldsymbol{\lambda}_\perp^{(i-1)} = F_\perp^*\boldsymbol{x}^{(i-1)}$. Recall that q_η denotes a pdf defined on the ball $B_{\boldsymbol{0},\eta}$ as expressed in Section 5.4.1.3. Due to the symmetry of the Gaussian distribution, it can be deduced that

$$\frac{q(\boldsymbol{x}^{(i-1)} \mid \boldsymbol{x})}{q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)})} = \frac{q_\eta(\boldsymbol{u}^{(i-1)} - P(\boldsymbol{u} - \boldsymbol{y}) - \boldsymbol{y})}{q_\eta(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)})}. \tag{5.36}$$

This expression remains valid in the degenerate case when $K = L$ (yielding $\boldsymbol{x}_{H^\perp} = \boldsymbol{0}$). Finally, it is important to note that, if q_η is the uniform distribution on the ball $B_{\boldsymbol{0},\eta}$, the above ratio reduces to 1, which simplifies the computation of the MH acceptance

---

[3] Note here that using a tight frame makes the computation of both $\boldsymbol{x}_H$ and $\boldsymbol{x}_{H^\perp}$ much easier due to the relation $F^*F = \mu\mathrm{I}$.

ratio. The final algorithm is summarized in Algorithm 11. Note that the sampling of the hyper-parameter vector is performed as for the hybrid GS in Section 5.4.1.2.

Algorithm 11 Sampler 2: hybrid MH sampler using algebraic properties of frame representations to sample according to $f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$.

```
Initialize with some $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_g^{(0)})_{1 \le g \le G} = (\gamma_g^{(0)}, \beta_g^{(0)})_{1 \le g \le G}$ and $\boldsymbol{u}^{(0)} \in B_{\boldsymbol{y},\delta}$. Set
    $\boldsymbol{x}^{(0)} = F(F^*F)^{-1}\boldsymbol{u}^{(0)}$ and $i = 1$.
repeat
    Sampling $\boldsymbol{x}$:
        • Compute $\hat{\boldsymbol{u}}^{(i-1)} = P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) + \boldsymbol{y}$.
        • Generate $\widetilde{\boldsymbol{u}}^{(i)} \backsim q_\eta(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)})$ where $q_\eta$ is defined on $B_{\boldsymbol{0},\eta}$ (see Section 5.4.1.3).
        • Compute $\widetilde{\boldsymbol{x}}_H^{(i)} = F(F^*F)^{-1}\widetilde{\boldsymbol{u}}^{(i)}$.
        • Generate $z^{(i)} \backsim \mathcal{N}(\boldsymbol{x}^{(i-1)}, \sigma_{\boldsymbol{x}}^2 \boldsymbol{I})$.
        • Compute $\widetilde{\boldsymbol{x}}_{H^\perp}^{(i)} = \Pi_{H^\perp} z^{(i)}$ and $\widetilde{\boldsymbol{x}}^{(i)} = \widetilde{\boldsymbol{x}}_H^{(i)} + \widetilde{\boldsymbol{x}}_{H^\perp}^{(i)}$.
        • Compute the ratio $r(\widetilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(i-1)}) = \frac{f(\widetilde{\boldsymbol{x}}^{(i)}|\boldsymbol{\theta}^{(i-1)}, \boldsymbol{y})\; q_\eta\big(\boldsymbol{u}^{(i-1)} - P(\widetilde{\boldsymbol{u}}^{(i)} - \boldsymbol{y}) - \boldsymbol{y}\big)}{f(\boldsymbol{x}^{(i-1)}|\boldsymbol{\theta}^{(i-1)}, \boldsymbol{y})\; q_\eta\big(\widetilde{\boldsymbol{u}}^{(i)} - \hat{\boldsymbol{u}}^{(i-1)}\big)}$
          and accept the proposed candidates $\widetilde{\boldsymbol{u}}^{(i)}$ and $\widetilde{\boldsymbol{x}}^{(i)}$ with probability
          $\min\{1, r(\widetilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(i-1)})\}$.
    Sampling $\boldsymbol{\theta}$:
    for $g = 1$ to $G$ do
        Generate $\gamma_g^{(i)} \backsim \mathcal{IG}\left(\frac{n_g}{\beta_g^{(i-1)}}, \sum_{k \in S_g} |x_k^{(i)}|^{\beta_g^{(i-1)}}\right)$.
        Simulate $\beta_g^{(i)}$ as follows:
            • Generate $\widetilde{\beta}_g^{(i)} \backsim q(\beta_g \mid \beta_g^{(i-1)})$.
            • Compute the ratio $r(\widetilde{\beta}_g^{(i)}, \beta_g^{(i-1)}) = \frac{f(\widetilde{\beta}_g^{(i)}|\gamma_g^{(i)}, \boldsymbol{x}^{(i)}, \boldsymbol{y}) q(\beta_g^{(i-1)}|\widetilde{\beta}_g^{(i)})}{f(\beta_g^{(i-1)}|\gamma_g^{(i)}, \boldsymbol{x}^{(i)}, \boldsymbol{y}) q(\widetilde{\beta}_g^{(i)}|\beta_g^{(i-1)})}$
              and accept the proposed candidate with the probability
              $\min\{1, r(\widetilde{\beta}_g^{(i)}, \beta_g^{(i-1)})\}$.
    end for
    Set $i \leftarrow i + 1$.
until Convergence
```

Experimental estimation results and applications to some image recovery problems of the proposed stochastic sampling techniques are provided in Section 5.6.

5.5 Toward a more general Bayesian Model

In this section, we extend the previously developed hierarchical Bayesian model in Section 5.3 to include an additional term in the prior depending on the Total Variation (TV)

of the image under investigation [Rudin et al., 1992; Chambolle and Lions, 1997]. This kind of prior has been used in numerous works in image restoration since it is known to be edge-preserving and helps in recovering homogeneous image areas. In [Louchet and Moisan, 2008; Louchet and Moisan, 2010], a TV Bayesian denoising algorithm based on the Least Square Error estimate (LSE) was proposed when a TV prior is used to regularize the considered denoising problem. However, like redundant frame representations, using TV priors leads also to a hyper-parameter estimation problem. This difficulty has been outlined in [Pustelnik et al., 2010] where a hybrid variational regularization approach is adopted combining an ℓ_1 sparsity term and a TV penalization. We extend here our proposed Bayesian approach to deal with denoising problems when a hybrid prior is considered to introduce simultaneously prior informations about the frame coefficients and the TV of the processed image. As it will be outlined hereafter, the proposed approach using this more general prior allows us to estimate the related hyper-parameters.
Assuming an exponential shape, the new prior can be expressed as:

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left(-\kappa \|F^*\boldsymbol{x}\|_{\mathrm{TV}}\right) \prod_{g=1}^{G} \left[\left(\frac{1}{\gamma_g^{1/\beta_g}} \right)^{n_g} \exp\left(-\frac{1}{\gamma_g} \sum_{k \in S_g} |x_k|^{\beta_g} \right) \right] \tag{5.37}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_G, \kappa)$ is the new hyper-parameter vector with $\kappa > 0$, $\|.\|_{\mathrm{TV}}$ is the TV semi-norm [Rudin et al., 1992; Chambolle and Lions, 1997], and $Z(\boldsymbol{\theta})$ is a normalization constant. The new hierarchical Bayesian model for the frame decomposition is completed by the following improper hyperprior

$$\begin{aligned} f(\boldsymbol{\theta}) &= Z(\boldsymbol{\theta}) \varphi(\kappa) \prod_{g=1}^{G} \varphi(\boldsymbol{\theta}_g) = Z(\boldsymbol{\theta}) \varphi(\kappa) \prod_{g=1}^{G} [\varphi(\gamma_g) \varphi(\beta_g)] \\ &\propto Z(\boldsymbol{\theta}) 1_{[0,\kappa_{\max}]}(\kappa) \prod_{g=1}^{G} \left[\frac{1}{\gamma_g} 1_{\mathbb{R}^+}(\gamma_g) 1_{[0,3]}(\beta_g) \right], \end{aligned} \tag{5.38}$$

where $\kappa_{\max}$ is a positive real to be fixed (fixed to $\kappa_{\max} = 10$ in practice).
It should be noted here that the hyperprior $f(\boldsymbol{\theta})$ in Eq. (5.38) has a stable asymptotic behaviour when $\gamma_g \to +\infty$. Indeed, we have

$$Z(\boldsymbol{\theta}) = \int \exp\left(-\kappa \|F^*\boldsymbol{x}\|_{\mathrm{TV}}\right) \prod_{g=1}^{G} \left[\left(\frac{1}{\gamma_g^{1/\beta_g}} \right)^{n_g} \exp\left(-\frac{1}{\gamma_g} \sum_{k \in S_g} |x_k|^{\beta_g} \right) \right] d\boldsymbol{x}. \tag{5.39}$$

Since $\exp\left(-\kappa \|F^*\boldsymbol{x}\|_{\mathrm{TV}}\right) \leq 1$, it turns out that

$$Z(\boldsymbol{\theta}) \leq \int \prod_{g=1}^{G} \left[\left(\frac{1}{\gamma_g^{1/\beta_g}} \right)^{n_g} \exp\left(-\frac{1}{\gamma_g} \sum_{k \in S_g} |x_k|^{\beta_g} \right) \right] d\boldsymbol{x}. \tag{5.40}$$

By performing the change of variable $\forall k \in S_g,\ y_k = \frac{x_k}{\gamma_g^{1/\beta_g}}$ with $g \in \{1, \dots, G\}$, we can write

$$Z(\boldsymbol{\theta}) \leq \int \prod_{g=1}^{G} \left[\exp\left(-\sum_{k \in S_g} |y_k|^{\beta_g} \right) \right] d\boldsymbol{y} < +\infty. \tag{5.41}$$

It is therefore ensured that the right term of the inequality in Eq. (5.41) is bounded, and thus $Z(\boldsymbol{\theta})$ is uniformly bounded w.r.t. γ_g.
The resulting new posterior distribution is therefore given by

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) =& 1_{C_\delta}(\boldsymbol{x}) \prod_{g=1}^{G} \left[\left(\frac{1}{\gamma_g^{1/\beta_g}} \right)^{n_g} \exp\left(-\frac{1}{\gamma_g} \sum_{k \in S_g} |x_k|^{\beta_g} \right) \left(\frac{1}{\gamma_g} 1_{\mathbb{R}+}(\gamma_g) 1_{[0,3]}(\beta_g) \right) \right] \\ & \times \exp\left(-\kappa \|F^* \boldsymbol{x}\|_{\mathrm{TV}} \right) 1_{[0,\kappa_{\max}]}(\kappa). \end{aligned} \tag{5.42}$$

The Bayesian estimators associated with the posterior distribution in Eq. (5.42) still have no simple closed-form expression. For this reason, we will apply the same sampling strategy as in Section 5.4.2 and sample the frame coefficients as in Algorithm 11. However, for the hyper-parameters vector, straightforward calculations show that the posterior distribution for the hyper-parameters γ_g, β_g and κ will be expressed as:

$$f(\gamma_g|\beta_g, \kappa, \boldsymbol{x}, \boldsymbol{y}) \propto \gamma_g^{-\frac{n_g}{\beta_g}-1} \exp\Big(-\frac{1}{\gamma_g} \sum_{k \in S_g} |x_k|^{\beta_g} \Big) 1_{\mathbb{R}+}(\gamma_g), \tag{5.43}$$

$$f(\beta_g|\gamma_g, \kappa, \boldsymbol{x}, \boldsymbol{y}) \propto \exp\Big(\sum_{k \in S_g} -\frac{1}{\gamma_g} |x_k|^{\beta_g} \Big) 1_{[0,3]}(\beta_g), \tag{5.44}$$

and

$$f(\kappa|\gamma_1, \dots, \gamma_G, \beta_1, \dots, \beta_G, \boldsymbol{x}, \boldsymbol{y}) \propto \exp\left(-\kappa \|F^* \boldsymbol{x}\|_{\mathrm{TV}} \right) 1_{[0,\kappa_{\max}]}(\kappa), \tag{5.45}$$

respectively.

Consequently, as in Section 5.4.1.2, $f(\gamma_g|\beta_g, \kappa, \boldsymbol{x}, \boldsymbol{y})$ is the pdf of the inverse gamma distribution $\mathcal{IG}\Big(\frac{n_g}{\beta_g}, \sum_{k \in S_g} |x_k|^{\beta_g}\Big)$. Sampling γ_g will therefore be performed exactly as in Algorithm 10. Conversely, it is more difficult to sample according to $f(\beta_g|\gamma_g, \kappa, \boldsymbol{x}, \boldsymbol{y})$ and $f(\kappa|\gamma_1, \dots, \gamma_G, \beta_1, \dots, \beta_G, \boldsymbol{x}, \boldsymbol{y})$. This task is achieved by using two MH moves whose proposal distributions $q(\beta_g \mid \beta_g^{(i-1)})$ and $q(\kappa \mid \kappa^{(i-1)})$ are Gaussian distributions truncated on the intervals $[0, 3]$ and $[0, \kappa_{\max}]$ with standard deviations $\sigma_{\beta_g} = 0.05$ and $\sigma_\kappa = 0.01$, respectively. Note that these standard deviation values have been fixed based on our practical observations. The resulting method to sample according to the posterior distribution in Eq. (5.42) is summarized in Algorithm 12.

Algorithm 12 Hybrid MH sampler using algebraic properties of frame representations to simulate according to $f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ using the general Bayesian model.

```
Initialize with some $\boldsymbol{\theta}^{(0)} = \left((\boldsymbol{\theta}_g^{(0)})_{1\le g\le G}, \kappa^{(0)}\right) = \left((\gamma_g^{(0)}, \beta_g^{(0)})_{1\le g\le G}, \kappa^{(0)}\right)$ and $\boldsymbol{u}^{(0)} \in B_{\boldsymbol{y},\delta}$. Set $\boldsymbol{x}^{(0)} = F(F^*F)^{-1}\boldsymbol{u}^{(0)}$ and $i = 1$.
repeat
    Sampling $\boldsymbol{x}$:
    • Compute $\hat{\boldsymbol{u}}^{(i-1)} = P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) + \boldsymbol{y}$.
    • Generate $\widetilde{\boldsymbol{u}}^{(i)} \backsim q_\eta(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)})$ where $q_\eta$ is defined on $B_{\boldsymbol{0},\eta}$ (see Section 5.4.1.3).
    • Compute $\widetilde{\boldsymbol{x}}_H^{(i)} = F(F^*F)^{-1}\widetilde{\boldsymbol{u}}^{(i)}$.
    • Generate $z^{(i)} \backsim \mathcal{N}(\boldsymbol{x}^{(i-1)}, \sigma_{\boldsymbol{x}}^2 \boldsymbol{I})$.
    • Compute $\widetilde{\boldsymbol{x}}_{H^\perp}^{(i)} = \Pi_{H^\perp} z^{(i)}$ and $\widetilde{\boldsymbol{x}}^{(i)} = \widetilde{\boldsymbol{x}}_H^{(i)} + \widetilde{\boldsymbol{x}}_{H^\perp}^{(i)}$.
    • Compute the ratio $r(\widetilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(i-1)}) = \dfrac{f(\widetilde{\boldsymbol{x}}^{(i)}|\boldsymbol{\theta}^{(i-1)}, \boldsymbol{y})\; q_\eta\big(\boldsymbol{u}^{(i-1)} - P(\widetilde{\boldsymbol{u}}^{(i)} - \boldsymbol{y}) - \boldsymbol{y}\big)}{f(\boldsymbol{x}^{(i-1)}|\boldsymbol{\theta}^{(i-1)}, \boldsymbol{y})\; q_\eta\big(\widetilde{\boldsymbol{u}}^{(i)} - \hat{\boldsymbol{u}}^{(i-1)}\big)}$
      and accept the proposed candidates $\widetilde{\boldsymbol{u}}^{(i)}$ and $\widetilde{\boldsymbol{x}}^{(i)}$ with probability $\min\{1, r(\widetilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(i-1)})\}$.
    Sampling $\boldsymbol{\theta}$:
    for $g = 1$ to $G$ do
        Generate $\gamma_g^{(i)} \backsim \mathcal{IG}\left(\frac{n_g}{\beta_g^{(i-1)}}, \sum_{k\in S_g} |x_k^{(i)}|^{\beta_g^{(i-1)}}\right)$.
        Simulate $\beta_g^{(i)}$ as follows:
        • Generate $\widetilde{\beta}_g^{(i)} \backsim q(\beta_g \mid \beta_g^{(i-1)})$.
        • Compute the ratio $r(\widetilde{\beta}_g^{(i)}, \beta_g^{(i-1)}) = \frac{f(\widetilde{\beta}_g^{(i)}|\gamma_g^{(i)},\kappa^{(i)},\boldsymbol{x}^{(i)},\boldsymbol{y})q(\beta_g^{(i-1)}|\widetilde{\beta}_g^{(i)})}{f(\beta_g^{(i-1)}|\gamma_g^{(i)},\kappa^{(i)},\boldsymbol{x}^{(i)},\boldsymbol{y})q(\widetilde{\beta}_g^{(i)}|\beta_g^{(i-1)})}$ and accept the proposed candidate with the probability $\min\{1, r(\widetilde{\beta}_g^{(i)}, \beta_g^{(i-1)})\}$.
    end for
    Simulate $\kappa^{(i)}$ as follows:
    • Generate $\widetilde{\kappa}^{(i)} \backsim q(\kappa \mid \kappa^{(i-1)})$.
    • Compute the ratio
      $$r(\widetilde{\kappa}^{(i)}, \kappa^{(i-1)}) = \frac{f(\widetilde{\kappa}^{(i)}|\gamma_1^{(i)}, \dots, \gamma_G^{(i)}, \beta_1^{(i)}, \dots, \beta_G^{(i)}, \boldsymbol{x}^{(i)}, \boldsymbol{y})q(\kappa^{(i-1)} \mid \widetilde{\kappa}^{(i)})}{f(\kappa^{(i-1)}|\gamma_1^{(i)}, \dots, \gamma_G^{(i)}, \beta_1^{(i)}, \dots, \beta_G^{(i)}, \boldsymbol{x}^{(i)}, \boldsymbol{y})q(\widetilde{\kappa}^{(i)} \mid \kappa^{(i-1)})}$$
      and accept the proposed candidate with the probability $\min\{1, r(\widetilde{\kappa}^{(i)}, \kappa^{(i-1)})\}$.
    Set $i \leftarrow i + 1$.
until Convergence
```

This approach is applied to image denoising in Section 5.6.3.1. However, in contrast to the first model where it is possible to generate frame coefficients according to their prior distribution (see Section 5.6.1), generating frame coefficients according to the prior distribution in Eq. (5.37) is not possible because of the TV term. For this reason, no validation experiments on synthetic data will be provided here for the hyper-parameters estimation. However, to evaluate the performance of the proposed approach in estimating the hyper-parameters when a hybrid prior is used, we can refer to Section 4.4 of Chapter 4 where reconstruction examples in pMRI are provided. The good reconstruction performance achieved in that experiments where the hyper-parameters have been estimated using Algorithm 12 shows that our approach provides an accurate estimation of the hyper-parameters.

5.6 Numerical illustrations

5.6.1 Validation experiments

5.6.1.1 Example 1

To show the effectiveness of our algorithm, a first set of experiments is carried out on synthetic images. As a frame representation, we use the union of two 2D separable wavelet bases $\mathcal{B}_1$ and $\mathcal{B}_2$ using Daubechies and shifted Daubechies filters of length 8 and 4, respectively. The ℓ_2 norm is used for $N(\cdot)$ in Eq. (5.1) with $\delta = 10^{-4}$. To generate a synthetic image (of size 128×128), we synthesize wavelet frame coefficients $\boldsymbol{x}$ from known prior distributions. Let $\boldsymbol{x}_1 = (a_1, (h_{1,j}, v_{1,j}, d_{1,j})_{1\leq j\leq 2})$ and $\boldsymbol{x}_2 = (a_2, (h_{2,j}, v_{2,j}, d_{2,j})_{1\leq j\leq 2})$ be the sequences of wavelet basis coefficients generated in $\mathcal{B}_1$ and $\mathcal{B}_2$, where a, h, v, d stand for approximation, horizontal, vertical and diagonal coefficients and the index j is the resolution level. Wavelet frame coefficients are generated from a GG distribution in accordance with the chosen priors. The coefficients in each subband are modeled with the same values of the hyper-parameters α_g and β_g, which means that each subband forms a group of index g. The number of groups (i.e., the number of subbands) G is therefore equal to 14. A uniform prior distribution over $[0, 3]$ is chosen for parameter β_g whereas a Jeffrey's prior is assigned to each parameter γ_g. For each group, the hyper-parameters β_g and γ_g are first generated from a uniform prior distribution over $[0, 3]$ and a beta distribution, respectively. Drawing the hyper-parameters from different distributions than the priors allows us to evaluate the robustness of our approach to modeling errors. A set of frame coefficients is then randomly generated to synthesize the observed data. The hyper-parameters are then supposed unknown, sampled using the proposed algorithm, and estimated based on the generated samples by:

(i): computing the mean according to the MMSE principle;

(ii): computing the MAP estimate.

Having reference values, the normalized mean square errors (NMSEs) related to the estimation of each hyper-parameter belonging to a given group (here a given subband) are computed from 30 Monte Carlo runs. The NMSEs computed for the estimators associated with the two samplers of Sections 5.4.1 and 5.4.2 are reported in Table 5.1. Table 5.1 shows that the proposed algorithms (using Sampler 1 of Section 5.4.1 and Sampler 2 of Section 5.4.2) provide accurate estimates of the hyper-parameters using the MMSE or the MAP estimator (with a slightly better performance for the MMSE estimator). The two samplers perform similarly for this experiment. However, one advantage of Sampler 2 is that it can be applied to different kinds of redundant frames, unlike Sampler 1. Indeed, as reported in Section 5.4.1, the conditional distribution in Eq. (5.11) is generally difficult to sample when the frame representation is not a union of orthonormal bases.

Table 5.1: Example 1: NMSEs for the estimated hyper-parameters using the MMSE and MAP estimators.

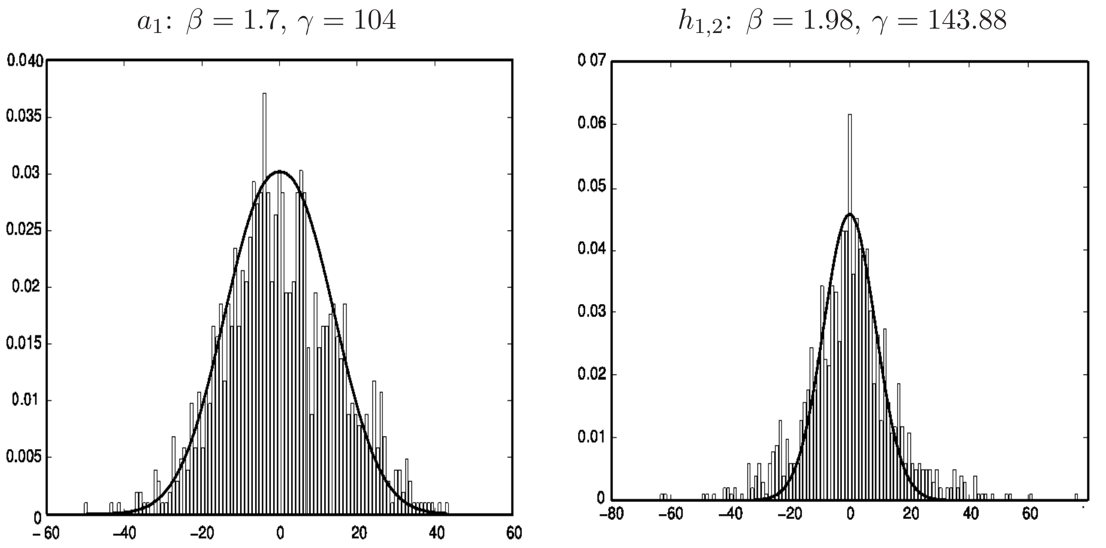
	MMSE				MAP			
	Sampler 1		Sampler 2		Sampler 1		Sampler 2	
	β	α	β	α	β	α	β	α
$h_{1,1}$	0.019	0.016	0.012	0.030	0.025	0.021	0.013	0.039
$v_{1,1}$	0.022	0.021	0.022	0.026	0.029	0.032	0.034	0.051
$d_{1,1}$	0.007	0.030	0.011	0.044	0.013	0.037	0.025	0.051
$h_{1,2}$	0.042	0.044	0.021	0.026	0.055	0.051	0.033	0.037
$v_{1,2}$	0.011	0.018	0.020	0.019	0.021	0.027	0.031	0.022
$d_{1,2}$	0.009	0.012	0.023	0.041	0.017	0.020	0.024	0.038
a_1	0.040	0.043	0.039	0.023	0.046	0.050	0.052	0.034
$h_{2,1}$	0.036	0.043	0.015	0.025	0.045	0.051	0.019	0.038
$v_{2,1}$	0.041	0.057	0.025	0.031	0.049	0.056	0.034	0.042
$d_{2,1}$	0.008	0.017	0.029	0.023	0.021	0.026	0.037	0.035
$h_{2,2}$	0.019	0.021	0.016	0.034	0.025	0.029	0.024	0.041
$v_{2,2}$	0.011	0.009	0.013	0.022	0.020	0.015	0.019	0.030
$d_{2,2}$	0.018	0.019	0.011	0.040	0.023	0.027	0.019	0.041
a_2	0.025	0.031	0.010	0.028	0.033	0.038	0.017	0.032

To further illustrate the good performance of the proposed estimator, Fig. 5.1 shows two examples of empirical histograms of wavelet frame coefficients (corresponding to $\mathcal{B}_1$) that are in good agreement with the corresponding pdfs obtained after replacing the hyper-parameters by their estimates.

5.6.1.2 Example 2

In this experiment, another frame representation is considered, namely a tight frame version of the TIW transform [Coifman and Donoho, 1995] with Daubechies filters of length 8. The ℓ_2 norm is also used for $N(\cdot)$ in Eq. (5.1) with $\delta = 10^{-4}$. We use the same process to generate frame coefficients as for Example 1. The coefficients in each subband (i.e., each group) are modeled with the same values of the hyper-parameters γ_g and β_g, the number of groups being equal to 7. The same priors for the hyper-parameters γ_g and β_g as for Example 1 are used.

After generating the hyper-parameters and frame coefficients, the hyper-parameters are


a_1: $\beta = 1.7$, $\gamma = 104$

$h_{1,2}$: $\beta = 1.98$, $\gamma = 143.88$

Figure 5.1: Examples of empirical approximation (left) and detail (right) histograms and pdfs of frame coefficients corresponding to a synthetic image.

then sampled using the proposed algorithm, and estimated using the MMSE estimator. Table 5.2 shows NMSEs based on reference values of each hyper-parameter, where the frame coefficient vector is denoted by $\boldsymbol{x} = (a, (h_j, v_j, d_j)_{1 \leq j \leq 2})$. Note that Sampler 1 is difficult to be implemented in this case since the used frame is not the union of orthonormal bases. Consequently, only NMSE values for Sampler 2 have been reported in Table 5.2.

Table 5.2: Example 2: NMSEs for the estimated hyper-parameters using the MMSE and MAP estimators with Sampler 2.

	MMSE		MAP	
	β	α	β	α
h_1	0.050	0.027	0.056	0.035
v_1	0.024	0.007	0.029	0.011
d_1	0.050	0.014	0.051	0.021
h_2	0.037	0.028	0.044	0.033
v_2	0.051	0.044	0.057	0.050
d_2	0.040	0.012	0.043	0.021
a	0.040	0.050	0.046	0.055

5.6.1.3 Example 3

A third frame is considered in this experiment to show the versatility of our approach with respect to the choice of the frame representation: the contourlet transform [Do and Vetterli, 2005] with ladder filters over two resolution levels. The ℓ_∞ norm is used for $N(\cdot)$

in Eq. (5.1) with $\delta = 10^{-4}$. We use the same procedure to generate frame coefficients as for Examples 1 and 2. The coefficients in each of the eight groups are modeled with the same values of the hyper-parameters γ_g and β_g and the same hyperparameter priors. After generating the hyper-parameters and frame coefficients, the hyper-parameters are then supposed unknown and estimated using the MMSE estimator based on samples drawn with Sampler 2. Table 5.3 shows NMSEs based on reference values of each hyper-parameter.

Table 5.3: Example 3: NMSEs for the estimated hyper-parameters using the MMSE and MAP estimates with Sampler 2.

	MMSE		MAP	
	β	α	β	α
SB_1	0.007	0.027	0.0120	0.071
SB_2	0.002	0.032	0.011	0.056
SB_3	0.004	0.011	0.009	0.023
SB_4	0.001	0.018	0.008	0.022
SB_5	0.001	0.006	0.006	0.012
SB_6	0.010	0.040	0.028	0.048
SB_7	0.009	0.020	0.018	0.07
SB_8	0.002	0.021	0.009	0.033

5.6.2 Convergence results

To be able to automatically stop the simulated chain and ensure that the last simulated samples are appropriately distributed according to the posterior distribution of interest, a convergence monitoring technique based on the Potential Scale Reduction Factor (PSRF) is used by simulating several chains in parallel (see [Gelman and Rubin, 1992] for more details). This convergence monitoring technique indicates that sample convergence arises as soon as PSRF < 1.2. Using the union of two orthonormal bases as a frame representation, Figs. 5.2 and 5.3 illustrate the variations w.r.t. the iteration number of the NMSE between the MMSE estimator and a reference estimator (computed by using a large number of burn-in and computation iterations, so as to guarantee that convergence has been achieved). The NMSE plots show that convergence is reached after about $150,000$ iterations (burn-in period of $100,000$ iterations), which corresponds to about 4 hours of computational time using Matlab 7.7 on an Intel Core 4-3 GHz architecture. When comparing the two proposed samplers in terms of convergence speed, it turns out from our simulations that Sampler 1 shows faster convergence than Sampler 2. Indeed, Sampler 1 needs about $110,000$ iterations to converge, which reduces the global computational time to about 3 hours.

The posterior distributions of the hyper-parameters β and γ related to the subbands $h_{1,2}$ and $h_{2,2}$ in $\mathcal{B}_1$ and $\mathcal{B}_2$ are shown in Fig. 5.4, as well as the known original values. It is clear that the modes of the posterior distributions are around the ground truth value, which confirms the good estimation performance of the proposed approach.

Note that when the resolution level increases, the number of subbands also increases, which

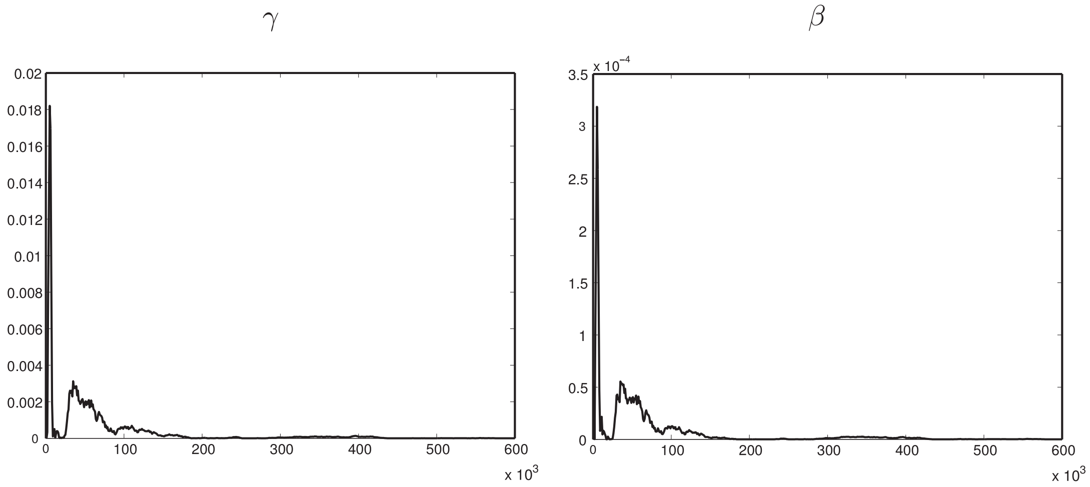


Figure 5.2: NMSE between the reference and current MMSE estimators w.r.t iteration number corresponding to $v_{1,1}$ in $\mathcal{B}_1$.

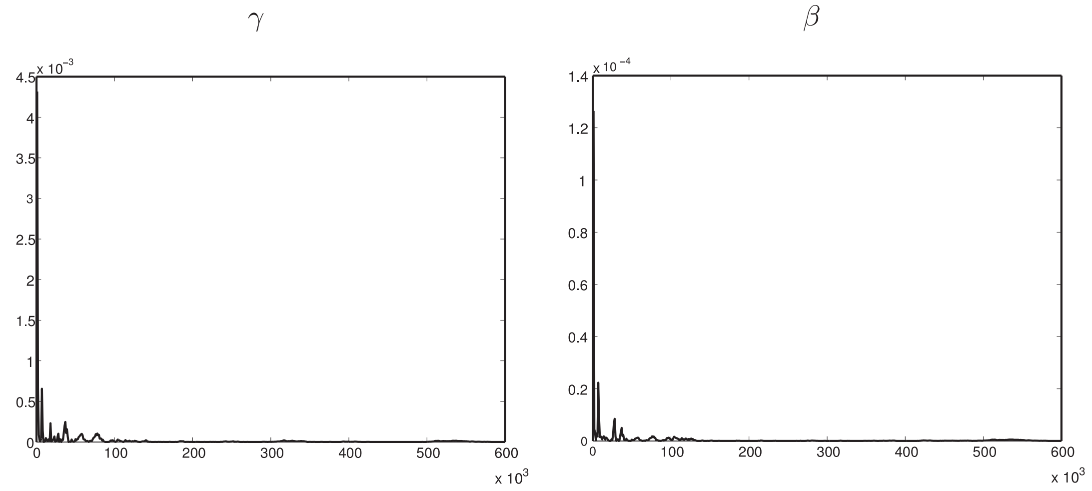


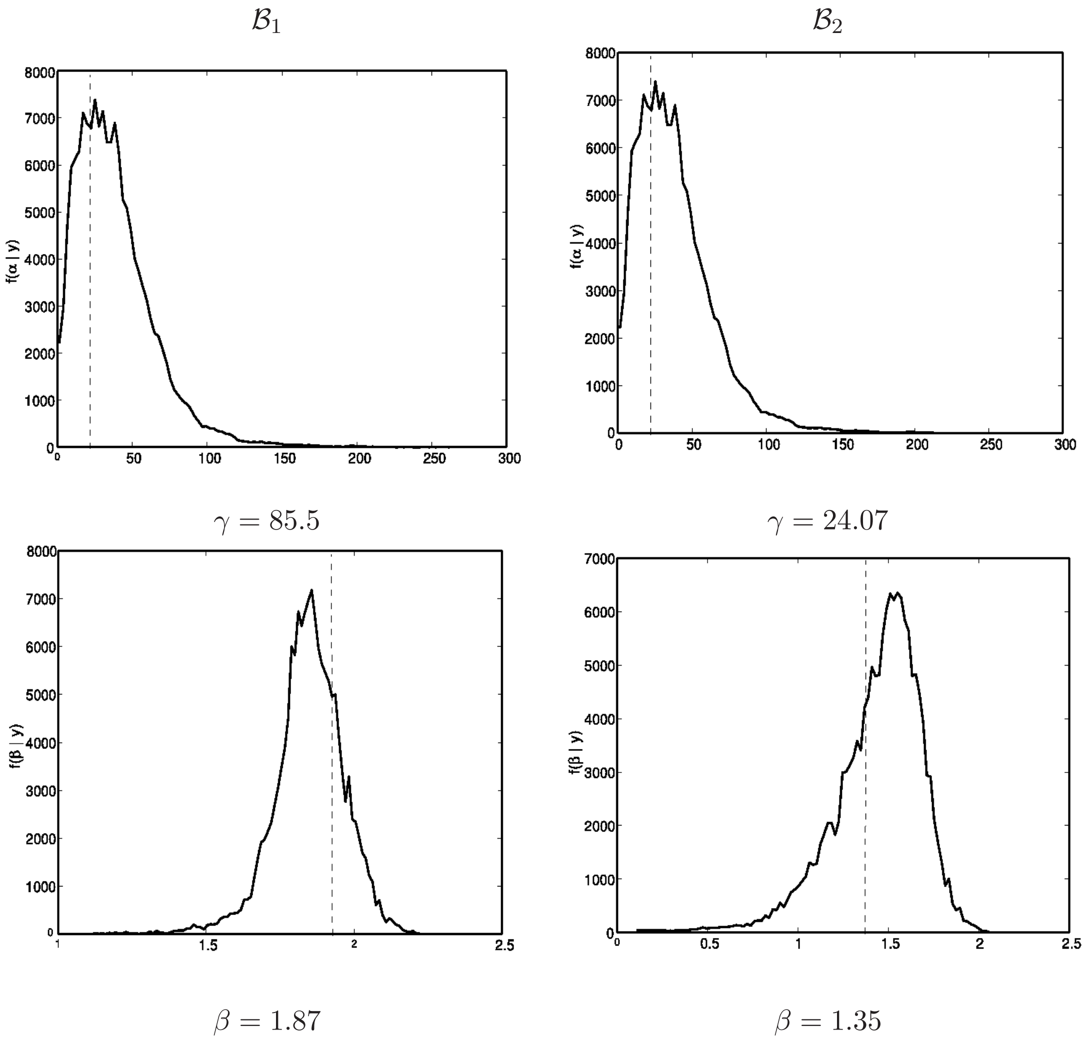
Figure 5.3: NMSE between the reference and current MMSE estimators w.r.t iteration number corresponding to $v_{2,2}$ in $\mathcal{B}_2$.

leads to a higher number of hyper-parameters to be estimated and a potential increase of the required computational time to reach convergence. For example, when using the union of two dyadic orthonormal wavelet bases with two resolution levels, the number of hyper-parameters to estimate is $G = 28$.

5.6.3 Application to image denoising

5.6.3.1 Example 1

In this experiment, we are interested in recovering an image (the *Boat* image of size 256×256 coded at 8 bpp) from its noisy observation affected by a noise $\boldsymbol{n}$ uniformly



$\gamma = 85.5$ $\gamma = 24.07$

$\beta = 1.87$ $\beta = 1.35$

Figure 5.4: Ground truth values (dashed line) and posterior distributions (solid line) of the sampled hyper-parameters γ and β, for the subbands $h_{1,2}$ and $h_{2,2}$ in $\mathcal{B}_1$ and $\mathcal{B}_2$, rsepectively.

distributed over the ball $[-\delta, \delta]^{256\times256}$ with $\delta = 30$. We recall that the observation model for this image denoising problem is given by Eq. (5.1). The noisy image in Fig. 5.5 (b) is simulated using the available reference image $\boldsymbol{y}_{\text{ref}}$ in Fig. 5.5 (a) and the noise properties described above.

The union of two 2D separable wavelet bases $\mathcal{B}_1$ and $\mathcal{B}_2$ using Daubechies and shifted Daubechies filters of length 8 and 4 (as for validation experiments in Section 5.6.1) is used as a tight frame representation. Denoising is performed using the MMSE estimator denoted as $\hat{\boldsymbol{x}}$ computed from sampled wavelet frame coefficients. The adjoint frame operator is then applied to recover the denoised image from its denoised estimated wavelet frame coefficients ($\hat{\boldsymbol{y}} = F^*\hat{\boldsymbol{x}}$). The obtained denoised image is depicted in Fig. 5.5 (d). For comparison purpose, the denoised image using a variational approach [Chaux et al., 2007; Combettes and Pesquet, 2008] based on a MAP criterion using the estimated values of the hyper-parameters with our approach is illustrated in Fig. 5.5 (c). This comparison

shows that, for denoising purposes, the proposed method gives better visual quality than the other reported methods. Signal to noise ratio (SNR $= 20\log_{10}\left(\|\boldsymbol{y}_{\text{ref}}\|/\|\boldsymbol{y}_{\text{ref}} - \hat{\boldsymbol{y}}\|\right)$) and structural similarity (SSIM) [Wang et al., 2004] values are also given in Table 5.4 to quantitatively evaluate denoising performance. Additional comparisons with respect to Wiener filtering and the algorithm developed in [Févotte and Godsill, 2006] (denoted here by SLR) are given in this table. Note that SLR can be applied only when the employed frame is the union of orthonormal bases, while our approach remains valid for any frame representation. Note also that SLR and Wiener filtering are basically designed to deal with Gaussian noise. This comparison shows that assuming the right noise model is essential to achieve good denoising performance. On the other hand, comparisons with the variational approach, which accounts for the right uniform noise model and uses the same frame representation and coefficient groups, show that the improvement achieved by our algorithm is not only due to the model choice.
The SNR and SSIM values are given for seven additional test images (*Sebal*, *Tree*, *Peppers*, *Kodim*, *House*, *Tire* and *Cameraman*) with different textures and contents to better illustrate the good performance of the proposed approach and its robustness to model mismatch. The corresponding original, noisy and denoised images are displayed in Figs. 5.6 to 5.12.

Table 5.4: SNR and SSIM values for the noisy and denoised images.

		Noisy	Wiener	Variational	SLR	MCMC	
						Algorithm 10	Algorithm 12
Boat	SNR	16.67	18.02	18.41	18.40	19.20	19.16
	SSIM	0.521	0.553	0.570	0.563	0.614	0.601
Sebal	SNR	13.85	14.40	15.04	14.98	15.69	15.91
	SSIM	0.642	0.695	0.704	0.697	0.701	0.748
Tree	SNR	17.19	19.27	19.29	19.38	19.82	19.62
	SSIM	0.662	0.768	0.765	0.776	0.785	0.802
Peppers	SNR	21.23	21.64	22.40	22.19	22.67	22.97
	SSIM	0.754	0.781	0.807	0.790	0.811	0.826
Kodim	SNR	19.89	23.25	23.69	23.80	24.32	24.19
	SSIM	0.444	0.742	0.749	0.795	0.875	0.843
House	SNR	18.54	19.57	21.09	20.49	22.41	22.56
	SSIM	0.488	0.532	0.651	0.639	0.703	0.741
Tire	SNR	16.16	17.86	19.23	19.26	19.86	19.57
	SSIM	0.733	0.816	0.811	0.815	0.854	0.860
Cameraman	SNR	17.79	20.47	20.76	20.53	21.03	20.98
	SSIM	0.524	0.703	0.765	0.718	0.703	0.767

It is worth noticing that the visual quality and quantitative results show that the denoised image based on the MMSE estimator of the wavelet frame coefficients is better than the one obtained with the Wiener filtering or the variational approach. For the latter approach, it must be emphasized that the choice of the hyper-parameters always constitute

a delicate problem, for which our algorithm brings a numerical solution. It should also be noted that compared with the variational approach, our algorithm recovers sharper and better denoised edges. However, our approach seems to be less performant in smooth regions, even if it does not introduce blurring effects like the variational approach.
In terms of computational time, in contrast with Wiener filtering and the variational approach which are very fast, SLR and our approach are more time-consuming. Table 5.5 gives the iteration numbers and computational times for the used methods on an Intel Core 4-3 GHz architecture using a Matlab implementation. However, a high gain in computational time can be expected through code optimization and parallel implementation using multiple CPU cores. In fact, since the frame coefficients are split into G groups with a couple of hyper-parameters for each of them, a high number of loops is required, which is detrimental to the computational time in a Matlab implementation.

Table 5.5: Computational time (in minutes) for the used methods.

	Wiener	Variational	SLR	MCMC
Iterations	1	100	100,000	150,000
Computational time	0.002	3	60	130

5.6.3.2 Example 2

In this experiment, we are interested in recovering an image (the *Straw* image of size 128×128 coded at 8 bpp) from its noisy observation affected by a noise $\boldsymbol{n}$ uniformly distributed over the centered ℓ_p ball of radius δ when $p \in \{1, 2, 3\}$. Experiments are conducted using two different frame representations: the TIW transform with a Symmlet filter of length 8 and the contourlet transform with ladder filters, both over 3 resolution levels. The ℓ_p norm ($p \in \{1, 2, 3\}$) was used for $N(\cdot)$ in Eq. (5.1). Figs. 5.13 (a) and 5.13 (b) show the original and noisy images using a uniform noise over the ℓ_2 ball of radius 3000. When using the TIW transform, Figs. 5.13 (c) and 5.13 (d) illustrate the results generated by the denoising strategies based on the variational approach and the MMSE estimator using frame coefficients sampled with our algorithm.

Table 5.6 shows the SNR and SSIM values for noisy and denoised images using the proposed MMSE estimator for different values of p and δ.

This second set of image denoising experiments shows that the proposed approach performs well when using different kinds of frame representations and various noise properties, which emphasizes its robustness to modelling errors.

Table 5.6: SNR and SSIM values for the noisy and denoised *Straw* images.

				TIW		contourlet	
		Noisy	Wiener	Variational	MCMC	Variational	MCMC
$\delta = 3.10^5$	SNR	15.56	16.42	16.67	18.11	17.76	18.79
$p = 1$	SSIM	0.719	0.705	0.730	0.755	0.678	0.803
$\delta = 3.10^3$	SNR	16.46	17.03	17.84	19.02	18.61	19.21
$p = 2$	SSIM	0.749	0.720	0.758	0.796	0.719	0.808
$\delta = 7.10^2$	SNR	16.14	17.05	17.65	19.29	18.28	19.44
$p = 3$	SSIM	0.734	0.720	0.671	0.771	0.698	0.788

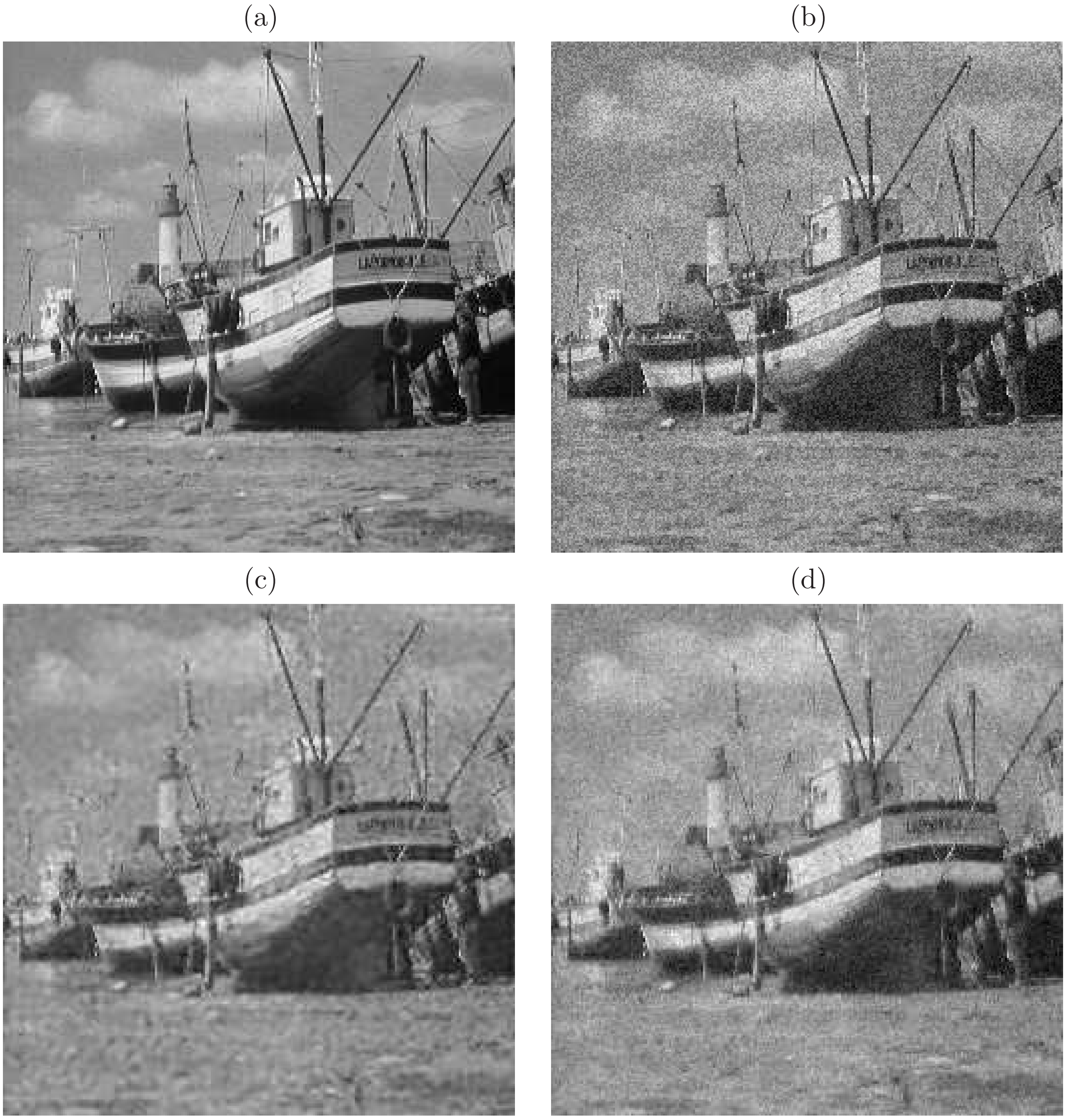

(a)
(b)
(c)
(d)

Figure 5.5: Original 256 × 256 *Boat* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).

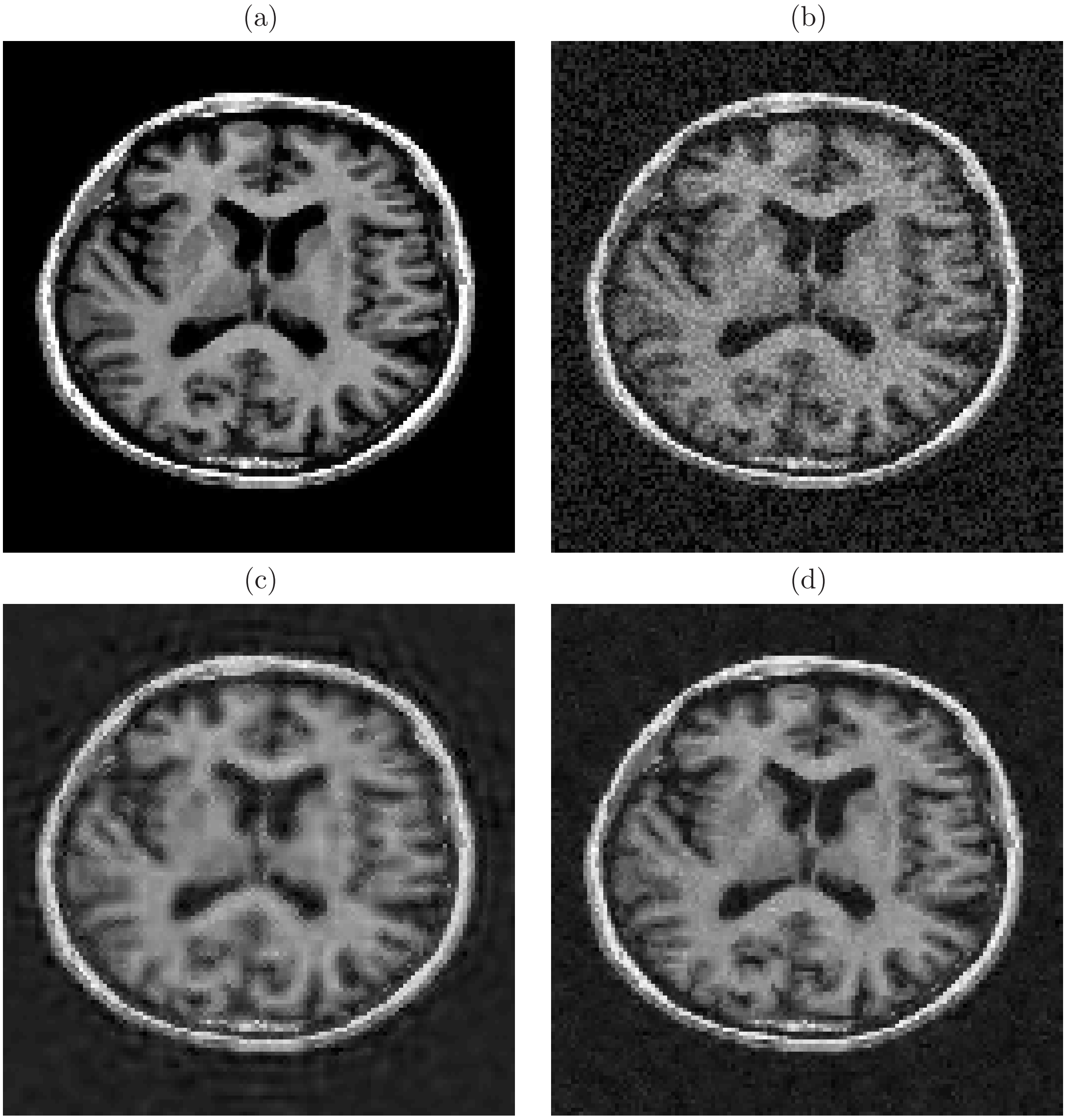
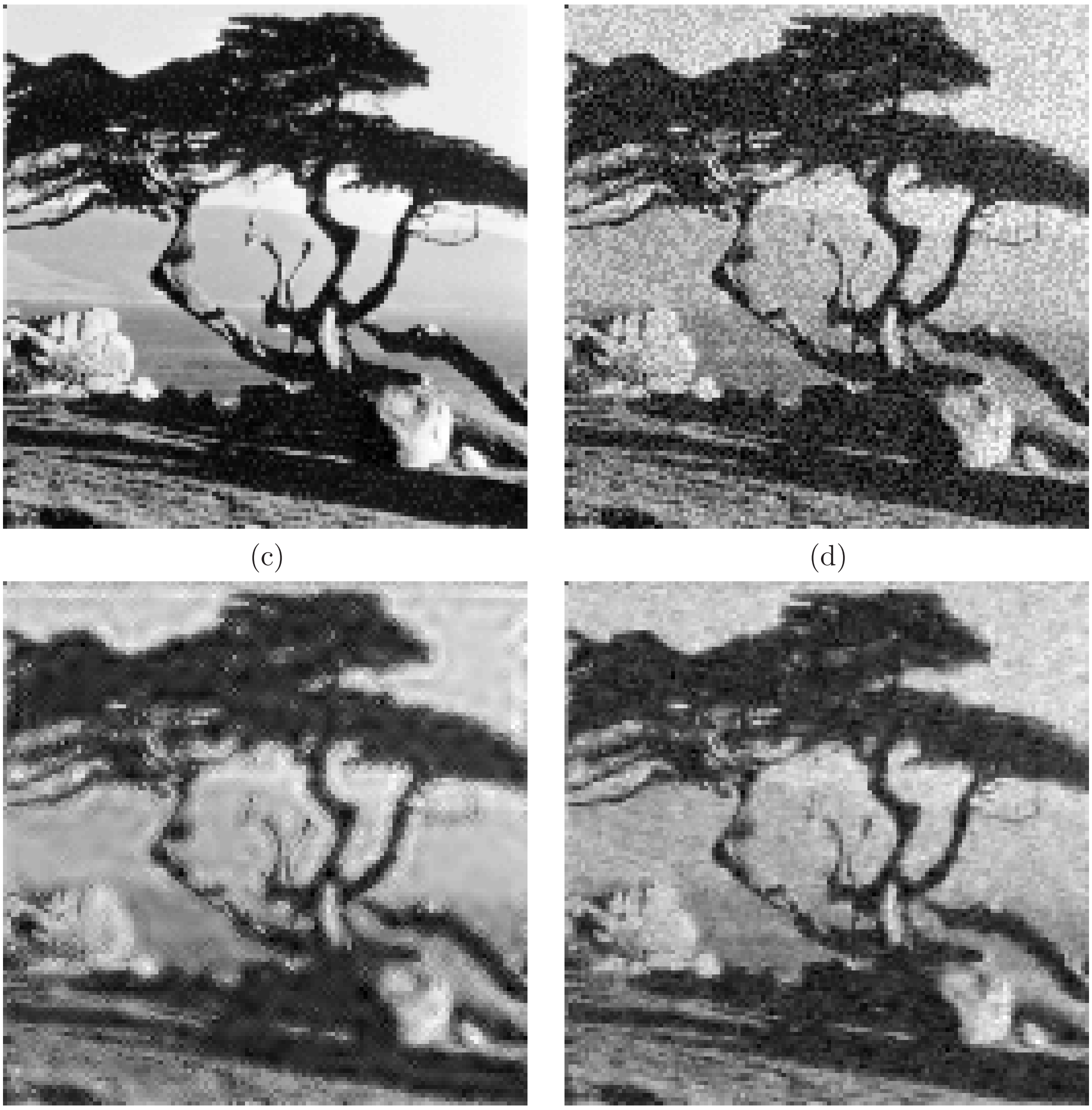


Figure 5.6: Original 128 × 128 *Sebal* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).



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

Figure 5.7: Original 128 × 128 *Tree* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).

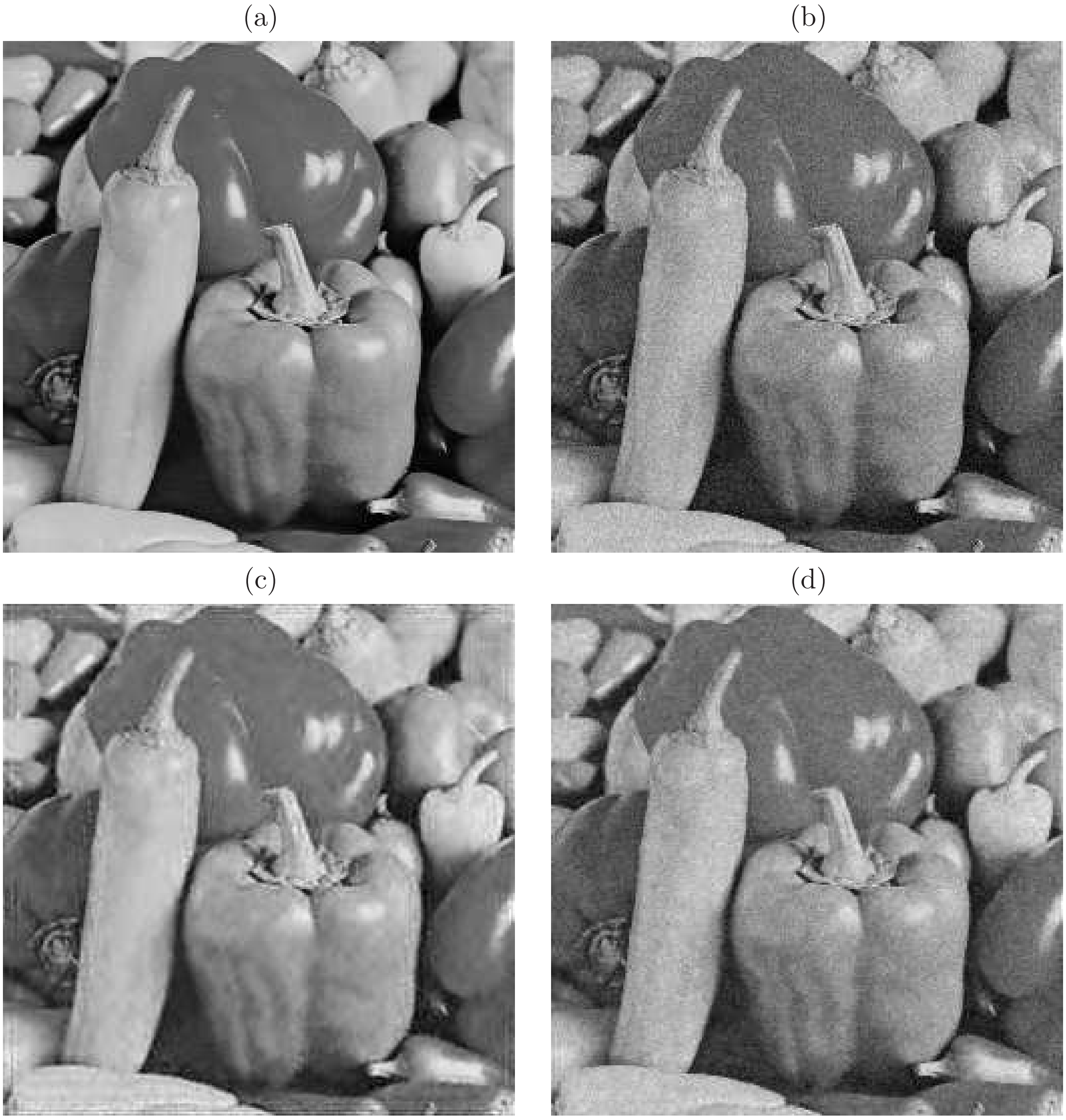


Figure 5.8: Original 256×256 *Peppers* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).

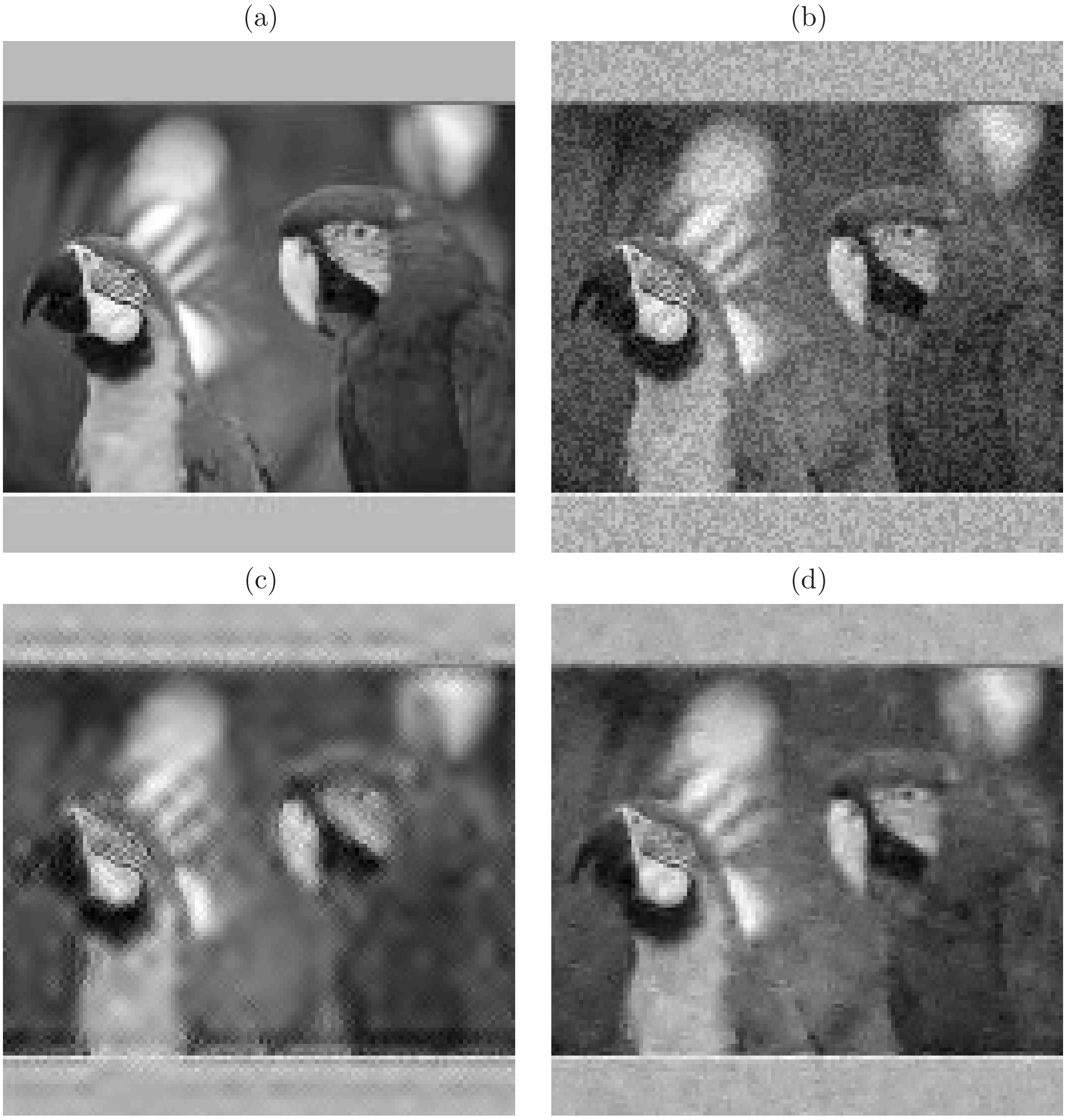
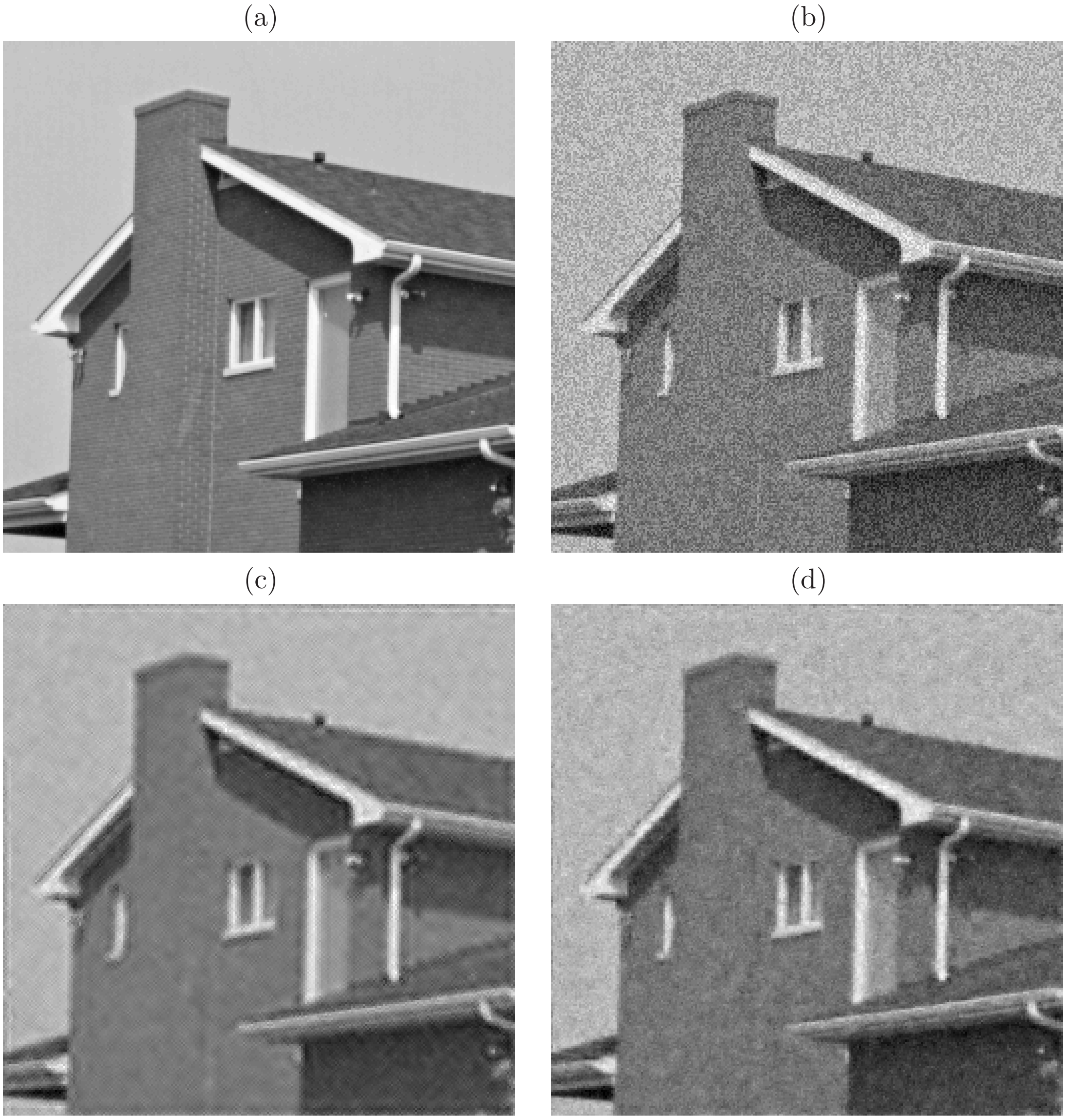


Figure 5.9: Original 128 × 128 *Kodim* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).



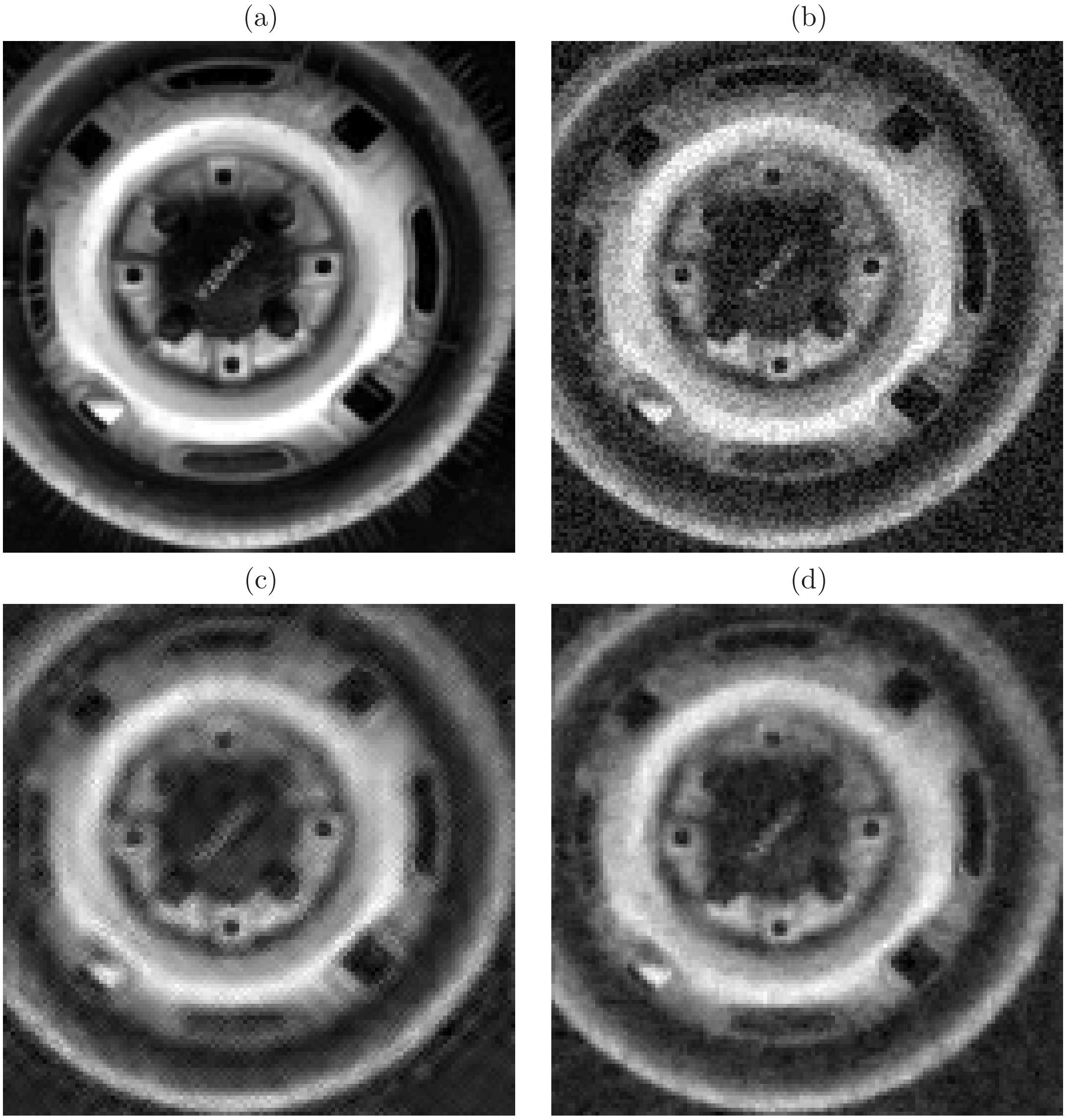
(a)

(b)

(c)

(d)

Figure 5.10: Original 256 × 256 *House* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).



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

Figure 5.11: Original 128 × 128 *Tire* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).

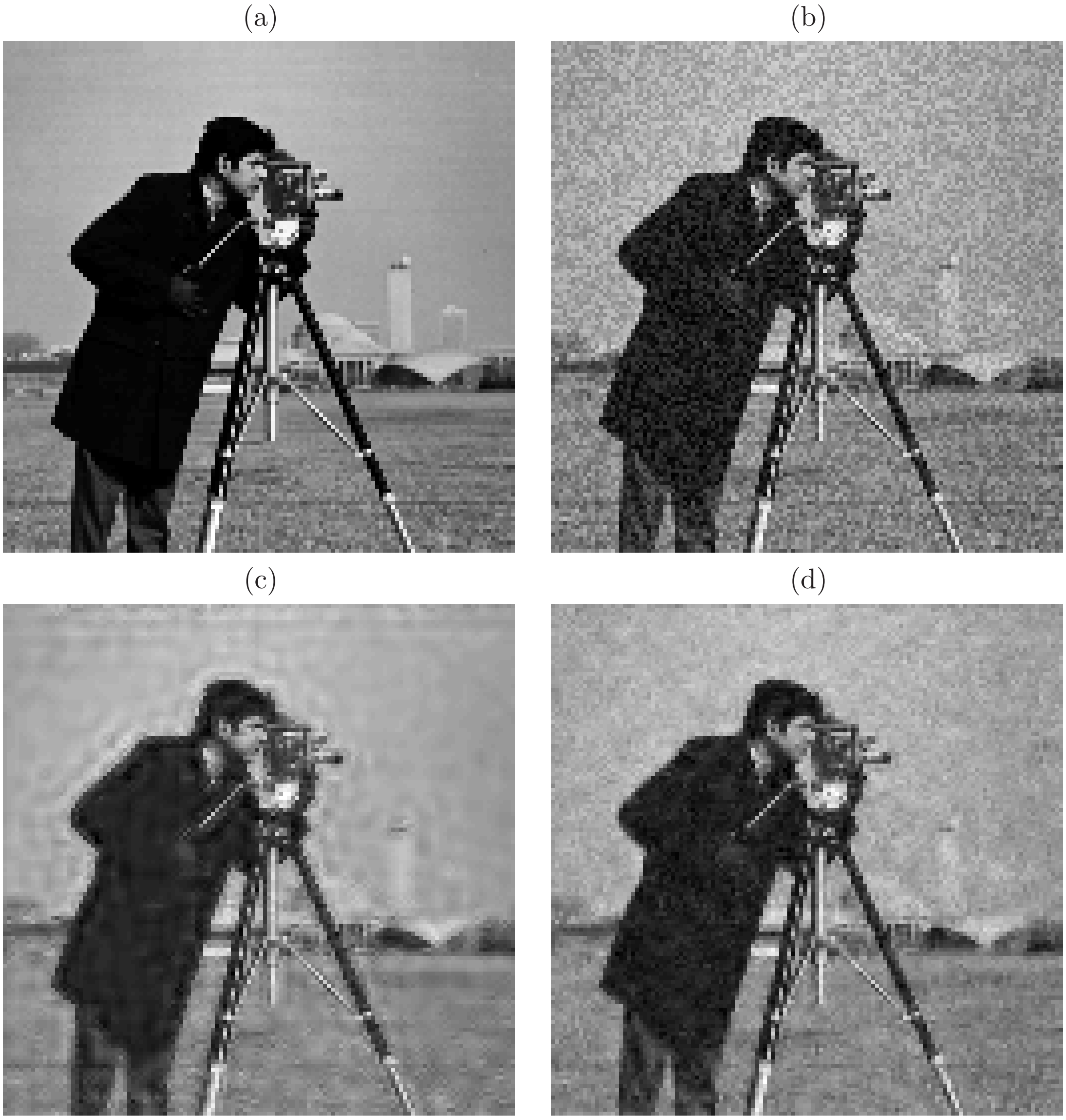


Figure 5.12: Original 128 × 128 *cameraman* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).

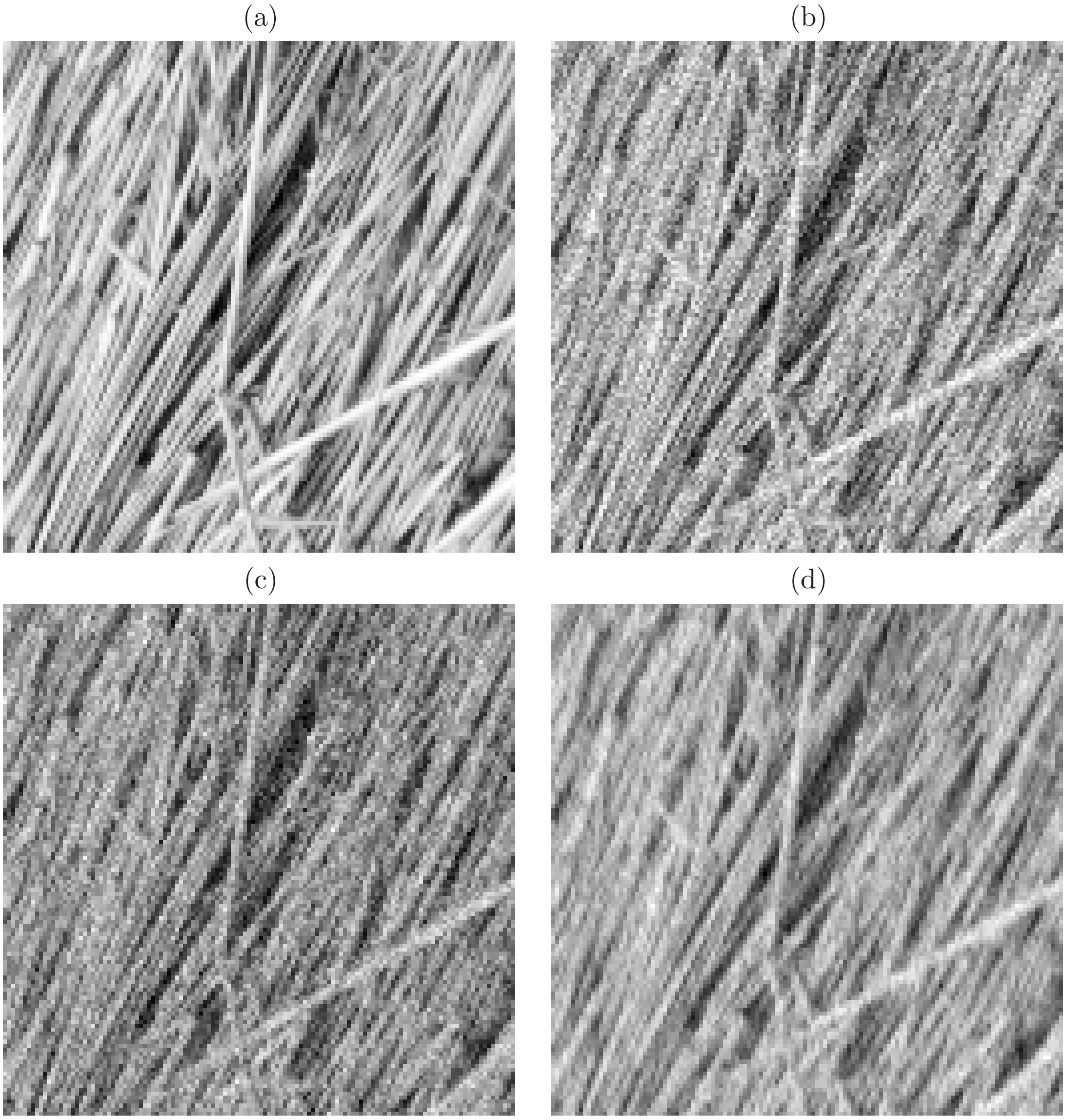


Figure 5.13: Original 128 × 128 *Straw* image (a), noisy image (b) and denoised images using the variational approach (c) and the proposed MMSE estimator (d).

5.6.4 Hyper-parameter estimation in parallel MRI

This section briefly explains how the proposed algorithm can be used to estimate the hyper-parameters in the parallel MRI regularization problem addressed in Chapter 4. We should first recall that the processed data is complex-valued. In this context, a prior pdf has been used for each of the real and imaginary parts of the wavelet coefficients of the image to be reconstructed. Estimating the hyper-parameters of these pdfs can be performed using the Bayesian approach proposed in this chapter. For doing so, the GG prior pdf in Eq. (5.6) has to be replaced by the prior in Eq. (4.7), and the hierarchical Bayesian model established in Eq. (5.10) has to be updated accordingly by adopting Jefrey's hyper-priors on the hyper-parameters α and β to be estimated (see Eq. (4.7)). The hyper-parameter estimation can therefore be performed by using the ℓ_∞ norm in Algorithm 12 with $\delta = 10^{-4}$ for example since a reference image is available for the estimation task.

5.7 Conclusion

This chapter proposed a hierarchical Bayesian algorithm for frame coefficient from a noisy observation of a signal or image of interest. The signal perturbation was modeled by introducing a bound on a distance between the signal and its observation. A hierarchical model based on this maximum distance property was then defined. This model assumed flexible GG priors for the frame coefficients, and has been generalized to involve a Total Variation term. Vague priors were assigned to the hyper-parameters associated with the frame coefficient priors. Different sampling strategies were proposed to generate samples distributed according to the joint distribution of the parameters and hyper-parameters of the resulting Bayesian models. The generated samples were finally used for estimation purposes. Our validation experiments indicated that the proposed algorithms provide an accurate estimation of the frame coefficients and hyper-parameters. The good quality of the estimates was confirmed on statistical processing problems in image denoising. Clearly, the proposed Bayesian approach outperforms the other methods because it allows us to use the right noise model and an appropriate frame coefficient prior. The numerous experiments which were conducted also showed that the proposed algorithm is robust to model mismatch. This developed Bayesian approach can be extended to handle complex-values signals/images assuming complex-valued priors like the complex GG [Novey et al., 2010] or the one in Eq. (4.23).

Chapter 6

Experimental validation in fMRI

Contents

6.1 Introduction 159
6.2 Data processing in fMRI 159
6.3 Validation of the proposed methods 168
6.4 Discussion 184
6.5 Conclusion 184

6.1 Introduction

The goal of this chapter is to experimentally validate the proposed UWR-SENSE and 4D-UWR-SENSE reconstruction approaches (see Chapter 4) in functional MRI (fMRI). We first recall in Section 6.2 the main fMRI data analysis methods at the subject and group-level. After describing the real data involved in this validation and the way they have been reconstructed from the raw data, this chapter details in Section 6.3 the experimental validation of the proposed reconstruction approaches at the subject and group-level. This validation allows us to study the impact of the pMRI reconstruction algorithm on activation detection in fMRI. Some conclusions are finally drawn based on the validation results.

6.2 Data processing in fMRI

Depending on the functional study context, data analysis in fMRI may be performed at different levels. In fact, during an fMRI session, data is generally acquired while the subject is submitted to an experimental paradigm. Analysing these data allows to detect and to localize haemodynamic variations or the Blood Oxygen Level Dependent (BOLD) signal in the brain inherent to the experimental paradigm for the involved subject: a *subject-level* analysis is therefore performed. These haemodynamic variations indirectly reflect neuronal activation, which can be used either in cognitive neuroscience or in clinical research. The first application aims at understanding brain function on healthy subjects, while the second is usually focused on specific neurodegenerative diseases or neurological disorders. In cognitive neurosciences where a given cognitive process is studied, one has to go further in the functional analysis and detect activations inherent to the same experimental paradigm applied to different subjects belonging to one or more given populations: we talk about a *group analysis*. In such an analysis, data are acquired for each subject during one or more fMRI sessions. Increasing the number of sessions allows to improve the statistical

sensitivity.
In practice, the statistical analysis of the collected data faces multiple difficulties: huge data size (up to hundreds of gigabytes), low SNR, acquisition artifacts and distortions (see Section 2.4.3), ... In the fMRI literature, two families of methods can be found:

- univariate methods: the processed voxels are assumed independently in space and processed separately. Each voxel is represented by its own signal which is used to decide whether the corresponding voxel is activated or not. These methods generally rely on some generative model [Friston et al., 1995; Smith, 2004].
- multivariate methods: all the data are processed globally and spatial dependencies between the voxels are exploited. The early multivariate methods are model-free [McKoewn et al., 1998; Calhoun et al., 2001b; Thirion and Faugeras, 2003; Kherif et al., 2004], but more recent ones are model-based. [Benali et al., 1997; Makni et al., 2005; Penny et al., 2005; Flandin and Penny, 2007; Van De Ville et al., 2007].

Although the multivariate methods seem more appropriate to analyse fMRI data where the voxels are indeed correlated, the univariate methods provide a good trade-off in terms of efficiency and computational simplicity.
In the literature, another distinction can be made between the available methods: exploratory and hypothesis-driven methods. Since we will rely on the popular hypothesis-driven methods for the validation experiments, these methods will be described in details in Section 6.2.1.2. However, we will also give a brief overview of exploratory methods in Section 6.2.1.1. For more details about these methods, the interested reader can refer to [Varoquaux et al., 2010].

6.2.1 Exploratory vs. hypothesis-driven methods

6.2.1.1 Exploratory methods

These methods generally aim at detecting eventual structures or significant forms either in the spatial or temporal domain in order to characterize the registered signal and study functional connectivities. They do not rely on any statistical model or prior knowledge [McKoewn et al., 1998; Petersson et al., 1999; Calhoun et al., 2001b; Beckmann and Smith, 2004]. These methods allow to extract the main patterns in the images and detect regions involving activations or unexpected parasitic effects (respiratory, cardiac,...). However, in contrast to recent exploratory methods allowing to derive statistical inference [Calhoun et al., 2001a; Varoquaux et al., 2010], early methods do not derive any decision/conclusion about the activation status of voxels. Nevertheless, when combined with other methods which allow to take such a decision (as we will see for the hypothesis-driven methods in the next sub-section), exploratory methods may help in the model selection task before the data analysis, or even for a validation purpose to confirm whether the selected model fits well to the data dynamics or not. Exploratory methods have been widely investigated first in resting-state data processing in order to study the intrinsic brain activation. More sophisticated versions relying on Principal Component Analysis (PCA) [Thirion and Faugeras, 2003; Kherif et al., 2004] and Independent Component Analysis

(ICA) [McKoewn et al., 1998; Calhoun et al., 2001b; Calhoun et al., 2001a; Beckmann and Smith, 2004] were afterwords in the fMRI literature. Although exploratory methods used to be applied for subject-level analysis, more recent methods such as Canonical ICA (CanICA) or NEtwork Detection using ICA (NEDICA) [Varoquaux et al., 209; Perlbarg et al., 2008] were proposed to deal with the statistical aspects at the group-level analysis. It should be noted here that, by nature, exploratory methods are multivariate since all the voxels are jointly processed.

In this manuscript, we will mainly focus on hypothesis-driven methods which will be introduced in the following section.

6.2.1.2 Hypothesis-driven methods

Hypothesis-driven methods rely on some assumptions about a given form of the response to the experimental stimulation [Worsley and Friston, 1995]. Based on these assumptions which are introduced as a prior hypothesis on the signal dynamics, a parameterized General Lineal Model (GLM) [Friston et al., 1994] is established and the inherent parameters are then estimated. Finally, a statistical test is necessary to estimate the response and conclude about the activation state by rejecting or not the null hypothesis of no activation.

These methods are usually univariate, i.e. the parameter estimation and statistical test are performed independently for each voxel. The main strength of these methods is that the statistical test allows to provide a precise answer to the fundamental question with a type I error (also called false positive) rate α (usually $\alpha = 10^{-3}$): “Is the time course of the measured signal in this voxel correlated with the expected BOLD response induced by the experimental paradigm?” The answer would be either “Yes”, which means that the concerned voxel is activated, or “Not”, which means that no activation is detected in that voxel. However, this kind of conclusion does not make sense when one is conducting a group analysis due to the brain anatomical and functional variabilities between subjects. In this context, hypothesis-driven methods can even answer a second important question: “Are there any clusters of connected activated voxels?” Indeed, detecting activated clusters is more significant than detecting activated voxels in a group analysis.

Our attention will be focused here on the GLM [Friston et al., 1994] which is the most used hypothesis-driven method, and which has been popularized in the fMRI literature due to the Statistical Parametric Mapping (SPM) software [1].

Note however that a series of pre-processing steps must be applied to the fMRI images before analysing them. These steps are summarized in Fig. 6.1 and include:

- distortion correction: the acquisition artifacts, and especially geometrical distortion have to be corrected;
- motion correction: the subject may move during the acquisition;
- slice timing correction: slices are actually acquired at slightly different acquisition times in sequential or interleaved order;

---

[1] http://www.fil.ion.ucl.ac.uk

- coregistration: BOLD fMRI images have to be registered with an anatomical image so that a joint visualisation is possible;
- spatial normalisation: when a group analysis is conducted, data of the different subjects have to be put in a common space using a template after applying a non-rigid transform;
- spatial smoothing: a convolution using a Gaussian filter is performed in order to improve the SNR and reduce the anatomical/functional variability between subjects.

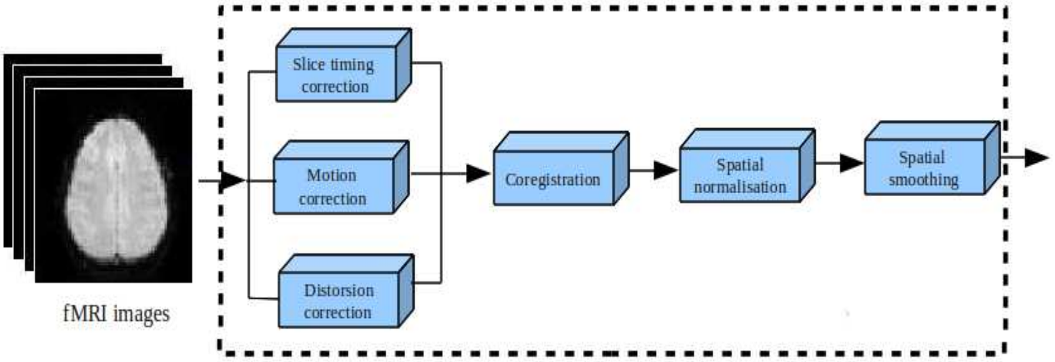


Figure 6.1: Required pre-processing steps before data analysis in fMRI.

For further details about these pre-processing steps, the interested reader can refer to [Flandin, 2004; Makni, 2006]. In the following section, the GLM is described in more details.

6.2.2 The General Linear Model

The GLM is based on three main properties:

- linearity: a group of events (different or not) leads to cumulative effects (if many stimuli are grouped, their effects sum up);
- time-invariance: a delay of the stimulation leads to the same delay of the inherent response;
- causality: an effect depends only on previous causes.

Let $\boldsymbol{Y}$ be an $N_v \times N_r$ matrix denoting an fMRI session dataset, where N_v is the number of voxels and N_r is the number of repetitions (length of the time series). We will also denote by $\boldsymbol{y}_n$ the measured signal at a voxel n [2]. We will assume that during the acquisition, the experimental paradigm is made up of N_c different conditions. In its current version, the GLM assumes that the measured signal at a given voxel is a linear weighted combination of

---

[2] For the sake of presentation clarity, the spatial 3D voxel position (y, x, z) is denoted here by n.

N_c components $(R_c)_{1\leq n\leq N_c}$ describing the stimuli presented in the experimental paradigm, called *regressors*, and a residual noise component $\epsilon_n \sim \mathcal{N}(0, \sigma_n^2)$. The signal model is therefore given by:

$$y_n(k) = \sum_{c=1}^{N_c} \beta_c^n R_c(k) + \epsilon_n(k), \tag{6.1}$$

where $k \in \{k_1, k_2, \dots, k_{N_r}\}$ is the acquisition time[3] and $n \in \{1, \dots, N_v\}$. For every $c \in \{1, 2, \dots, N_c\}$, β_c^n is the amplitude of the response to the c$^{\text{th}}$ stimulus and R_c is the corresponding regressor. When accounting jointly for the N_r repetitions, Eq. (6.1) may be rewritten as

$$\begin{pmatrix} y_n(k_1) \\ \vdots \\ y_n(k_{N_r}) \end{pmatrix} = \begin{pmatrix} R_1(k_1) & \dots & R_{N_c}(k_1) \\ \vdots & \vdots & \vdots \\ R_1(k_{N_r}) & \dots & R_{N_c}(k_{N_r}) \end{pmatrix} \begin{pmatrix} \beta_1^n \\ \vdots \\ \beta_{N_r}^n \end{pmatrix} + \begin{pmatrix} \epsilon_n(k_1) \\ \vdots \\ \epsilon_n(k_{N_r}) \end{pmatrix}, \tag{6.2}$$

or equivalently in a matrix form at the voxel level:

$$\boldsymbol{y}_n = \boldsymbol{R}\boldsymbol{\beta}_n + \boldsymbol{\epsilon}_n. \tag{6.3}$$

All the prior information about the design and the model will be plugged into the well matrix $\boldsymbol{R}$, also called the *design matrix* which is assumed to be known. Every column of this matrix corresponds to a regressor. The regressor associated with an experimental condition is simply the convolution of a stimuli vector x^c with the HRF h (see Chapter 2). For every acquisition time k, the regressor can be written as follows:

$$R_c(k) = h(k) * x^c(k), \tag{6.4}$$

In addition to the stimuli regressors, other regressors describing the first and second temporal derivatives of the first regressors may be added to the design matrix in order to model the haemodynamic variability. Based on the observations and prior hypothesis on the HRF, the main task will now consist of estimating the parameter vector $\boldsymbol{\beta}_n$ in Eq. (6.3) which is the only unknown in the observation model. Actually, the use of a canonical HRF assumes that the impulse response during an activation remains constant throughout the brain, which is a quite idealistic scenario that recent works like [Goutte et al., 2000; Ciuciu et al., 2003; Makni et al., 2008] challenge. In fact, such a response may considerably vary from one region to another, from one subject to another... A solution may lie in estimating it at each voxel [Marrelec et al., 2004], or better within functionally homogeneous regions of interest [Makni et al., 2008; Vincent et al., 2010].

6.2.3 Subject-level analysis

This section describes the activation detection procedure at the subject level using the GLM. The different steps are presented below.

---

[3] We do not use the same notation as in the first part of the manuscript since k denotes the acquisition time instead of t in order to avoid confusions with the Student t distribution used in Section 6.2.3.2.

6.2.3.1 Estimation of the model parameters

Estimating the parameter vector $\boldsymbol{\beta}_n$ at a voxel n may be performed using a least squares estimation. If the noise is assumed to be white in time, the Ordinary Least Squares (OLS) estimation of $\boldsymbol{\beta}_n$ reads:

$$\widehat{\boldsymbol{\beta}}_n = (\boldsymbol{R}^{\mathsf{T}}\boldsymbol{R})^{-1}\boldsymbol{R}^{\mathsf{T}}\boldsymbol{y}_n, \tag{6.5}$$

and the variance of this linear unbiased estimator is simply $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_n) = \sigma_n^2(\boldsymbol{R}^{\mathsf{T}}\boldsymbol{R})^{-1}$. However, it has been shown that time series in an fMRI session are temporally correlated [Bullmore et al., 1996]. We therefore have to consider a non-diagonal covariance matrix $\boldsymbol{\Sigma}_n = \sigma_n^2\boldsymbol{V}_n$ for the auto-correlated noise entries $\boldsymbol{\epsilon}_n \sim \mathcal{N}(0, \boldsymbol{\Sigma}_n)$. The Weighted Least Squares (WLS) estimator can then be written as:

$$\widehat{\boldsymbol{\beta}}_n = (\boldsymbol{R}^{\mathsf{T}}\boldsymbol{V}_n^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^{\mathsf{T}}\boldsymbol{V}_n^{-1}\boldsymbol{y}_n \tag{6.6}$$

with $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_n) = \sigma_n^2(\boldsymbol{R}^{\mathsf{T}}\boldsymbol{V}_n^{-1}\boldsymbol{R})^{-1}$. The obtained estimator is unbiased, but requests the knowledge of the matrix $\boldsymbol{V}_n$. Estimating this matrix is actually quite difficult since the problem is under-determined. In [Friston et al., 2000; Woolrich et al., 2001; Penny et al., 2003], an efficient technique taking into account the temporal correlations between the time series has been proposed. It consists of a pre-*whitening* approach. In other words, the covariance matrix is estimated by fixing a strong constraint on its structure (e.g. autoregressive model). However, a poor approximation of the auto-correlation coefficients may lead to inaccurate parameter estimates $\boldsymbol{\beta}_n$. After estimating the parameter vector $\widehat{\boldsymbol{\beta}}_n$ for each voxel n, the statistical maps have now to be calculated based on the N_c effect maps.

6.2.3.2 Calculating the statistical maps

In order to compare two given effects (e.g. a test condition vs. a control condition) and to detect voxels being activated due to one of the two effects, a *contrast* has to be defined. A contrast γ is defined by a linear combination of the design matrix regressors. The coefficients of this linear combination can be organized in a vector $\gamma = (0, \ldots, 0, -1, 1, 0, \ldots, 0)^{\mathsf{T}}$ in which all the terms are set to 0, except the terms involving the effects to be compared. If we consider that the vector $\widehat{\boldsymbol{\beta}}_n$ is the response to the effects in the design matrix $\boldsymbol{R}$, the defined contrast allows one to check whether the response to a given effect is more or less significant than another one. If no significant effect is observed for a given contrast γ, we have $\gamma^{\mathsf{T}}\boldsymbol{\beta}_n = 0$. We define therefore the null hypothesis $\mathcal{H}_0$ for each voxel n as follows:

$$\mathcal{H}_0 \ : \ \gamma^{\mathsf{T}}\boldsymbol{\beta}_n = 0. \tag{6.7}$$

After calculating the linear combination $\gamma^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_n$, one assesses the significance of the estimated response to the contrast by comparing it to its estimated standard deviation $\sqrt{\widehat{\mathrm{Var}}(\gamma^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_n)}$

through calculating the subsequent statistic

$$T_n = \frac{\gamma^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_n}{\sqrt{\widehat{\mathrm{Var}}(\gamma^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_n)}}, \tag{6.8}$$

which is distributed according to a *Student* distribution t_d with $d = N_r - N_c$ degrees of freedom [Worsley and Friston, 1995]. For each voxel n, we test if its statistical value t_n (t_n being a realisation of the random variable T_n) belongs to the t_d distribution in order to decide about the validity of the null hypothesis $\mathcal{H}_0$.

The test value t_n of each voxel n allows one to draw a statistical t map for each contrast and each subject describing the significance of the effects related to a given contrast γ.

6.2.3.3 Statistical test and activation detection

After calculating the statistical maps, the final step will now consist of detecting the activated voxels. The standard method relies on thresholding the t maps at a given level of significance while controlling the error rate. To this end, we consider the probability α controlling the false positive rate (also called type I error) which is defined as follows:

$$\alpha = P(T_n > t_\alpha | \mathcal{H}_0), \tag{6.9}$$

where t_α is the statistical value associated to the risk α (see Fig. 6.2) over which the probability that t_n comes from the distribution t_d is low. As illustrated in Fig. 6.2, a *p-value* is associated with each voxel, indicating the statistical threshold for the t-score allowing to reject the null hypothesis. Hence, computing the significance at the voxel level amounts to comparing $p = P(T_n > t_\alpha|\mathcal{H}_0)$ with α. If $t_n > t_\alpha$, or equivalently $p < \alpha$, $\mathcal{H}_0$ is

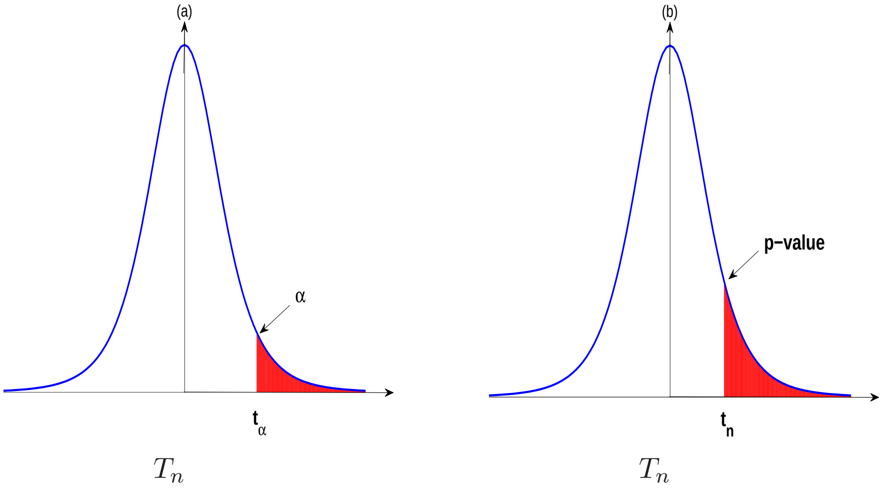


Figure 6.2: Illustration of the student distribution thresholding: (a) fixed threshold probability; (b) the threshold corresponds to the observed t-score (t_n) allowing to obtain the maximum probability (p-value). If $t_n > t_\alpha$, or equivalently $p < \alpha$, $\mathcal{H}_0$ is rejected, otherwise (if $t_n \leq t_\alpha$) it is retained.

rejected in this voxel: the voxel is therefore detected as an activated one. Otherwise, the null hypothesis $\mathcal{H}_0$ cannot be rejected.

After applying the statistical test to all voxels, the result of a functional analyis are simply represented as images where only voxels of a p-value lower than a threshold α (generally set to 10^{-3}) are with nonzero intensities. However, if one fixes the threshold $\alpha = 10^{-3}$ for example, the number of false positives will be equal to α multiplied by the number of voxels involved in the functional analysis, which is around $N_v = 5 \times 10^4$ for each volume. In such a situation, the number of false positives would be $\alpha \times N_v = 50$. To better control the false positive rate in the search volume, a correction for multiple comparisons has to be performed, which amounts to a Family Wise Error Rate (FWER) control problem. The Bonferroni correction [Holmes, 1994] allows to fix a global threshold taking into account the high number of voxels by simply dividing α by the number of applied statistical tests, which consists of reapplying the thresholding step with new p-values calculated from the old ones multiplied by the number of tests. The drawback of this method lies in its lack of sensitivity since it leads to rejecting activated voxels (too conservative). A more sophisticated approach is based on the Random Field theory (RFT) [Ashburner et al., 2004] and accounts for the spatial correlation of voxels. In this method, the statistical map is assumed to be a Gaussian random field. It can therefore be applied only when the data are smoothed using a Gaussian filter. Some other methods have been more recently proposed such as the False Discovery Rate (FDR) correction [Genovese, 2000; Nichols and Holmes, 2003] or Wavelet-based Statistical Parametric Mapping (WSPM) [Van De Ville et al., 2007]. Note also that nonparametric alternatives have been specifically designed for group-level analysis [Mériaux et al., 2006; Roche et al., 2007].

6.2.4 Group-level analysis

The functional analysis technique presented in Section 6.2.3 allows one to derive activation maps for a given subject from data acquired when this subject is submitted to an experimental paradigm. In order to determine if the detected activation for a subject can be extended to a whole population, a second or group-level is required. This level of analysis allows to study an entire cognitive process, and not only one realisation for a given subject. Clearly speaking, a group analysis aims at detecting a mean effect within a group of subjects by answering the following question: “what regions have been significantly activated for the group of subject in response to a given contrast?” Group activation maps are generally provided as a result. Even at this analysis level, the goal may be to study the cognitive process just for the considered group of subjects, or to extend the conclusions to the whole population. As we will see in the two following sections, differently formulating the null hypothesis allows to achieve the two previously cited goals.
We should note that a series of pre-processing steps must be applied to the data before activation detection. For instance, data has to be normalized to a common reference space because of anatomo-functional variability between subjects. For more details about pre-processing steps, the reader can refer to [Makni, 2006; Operto, 2009]. In what follows, we briefly recall the principle of three main group analysis methods.

6.2.4.1 Fixed effect model

This model assumes that the subjects involved in the functional analysis represent the exhaustive cohort. In this context, a subject is assumed to be a fixed effect and the inter-subject variance is not taken into account. In practice, the problem is addressed as if all the subjects correspond to the same one submitted many times to the same experiment. The null hypothesis can thus be formulated as: "no activation has been detected for any subject in response to the considered contrast". Grouping the data corresponding to the S subjects allows to retrieve the GLM with other parameters to estimate. Therefore, Eq. (6.3) can be rewritten as:

$$\begin{pmatrix} \boldsymbol{y}_n^1 \\ \boldsymbol{y}_n^2 \\ \vdots \\ \boldsymbol{y}_n^S \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_1 & 0 & \dots & 0 \\ 0 & \boldsymbol{R}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \boldsymbol{R}_S \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_n^1 \\ \boldsymbol{\beta}_n^2 \\ \vdots \\ \boldsymbol{\beta}_n^S \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_n^1 \\ \boldsymbol{\epsilon}_n^2 \\ \vdots \\ \boldsymbol{\epsilon}_n^S \end{pmatrix}, \tag{6.10}$$

where $\boldsymbol{y}_n^s$, $\boldsymbol{\beta}_n^s$ and $\boldsymbol{\epsilon}_n^s$ are respectively the observation, parameter and noise vectors defined in Eq. (6.3) for the s^{th} subject with $s \in \{1, \dots, S\}$. As regards $\boldsymbol{R}_s$, it denotes the design matrix corresponding the s^{th} subject.

If an effect is sufficiently present for at least one subject, the group statistics become significant. However, these statistics are only associated with the considered current subjects and cannot be generalized to any other group.

6.2.4.2 Random effect model

The random effect model [Laird and Ware, 1982] assumes that the considered group of subjects is simply a sample of a population of interest consisting of an infinity of subjects. It assumes that the observed effects for different subjects are due to the inter-subject variability and appear randomly. Each subject is assumed to be a random effect while taking into account the inter-subject variance. In contrast to the fixed effect model, the null hypothesis $\mathcal{H}_0$ can be formulated in this case as follows: "The mean effect is equal to zero". Consequently, the detected activated voxels suggest that a mean effect has been observed at the group level. The main advantage of the random effect model with respect to the fixed effect one is that it allows to generalize the results obtained from the group of studied subjects to the population of interest.

Generally speaking, an fMRI group analysis aims at drawing functional conclusions at the population level. However, some hurdles may be encountered in such studies when using the random effect model [Petersson et al., 1999]:

i) random effect analysis is less sensitive than fixed effect analysis and generally request more subjects to be established in a reliable way (typically 15 subjects);

ii) random effect models assume gaussianity of the data, which is generally not easy to ensure because of the low number of samples;

iii) the group of subjects used in the random effect analysis rarely represents a perfect sampling of the same population.

6.3 Validation of the proposed methods

This section is dedicated to the experimental validation of the proposed methods in Chapter 4. The validation will be based on an fMRI study involving subject and group-level analyses. The next section describes the used real data that we reconstructed using our own pipeline at the neuroimaging center NeuroSpin.

6.3.1 Experimental data

The used fMRI data for these validation experiments were recorded at 3 Tesla on a Siemens Trio magnet using a Gradient-Echo EPI (GE-EPI) sequence ($TE = 30$ ms, $TR = 2400$ ms, slice thickness = 3 mm, transversal orientation, FOV = 192 mm^2) during a cognitive *localizer* [Pinel et al., 2007; Vincent, 2010] experiment designed to map auditory, visual and motor brain functions as well as higher cognitive tasks such as number processing and language comprehension. It consisted of a single session of $N_r = 128$ scans. The paradigm was a fast event-related design comprising sixty auditory, visual and motor stimuli, defined in ten experimental conditions (auditory and visual sentences, auditory and visual calculations, left/right auditory and visual clicks, horizontal and vertical checkerboards). An $L = 32$ channels coil was used to enable parallel imaging. Sixteen healthy subjects gave informed consent to be scanned. For each subject, fMRI data were collected at different in-plane spatial resolutions (3×3, 2×2 and 1.5×1.5 mm^2). However, in this manuscript we only report results on 2×2 mm^2, since the processing of the other dataset is still in progress. fMRI data were also collected at different reduction factors ($R = 2$ or $R = 4$) but only for the two highest spatial resolutions.

Based on the raw data files delivered by the scanner, reduced FOV images have been reconstructed as detailed in the diagram of Fig. 6.3.

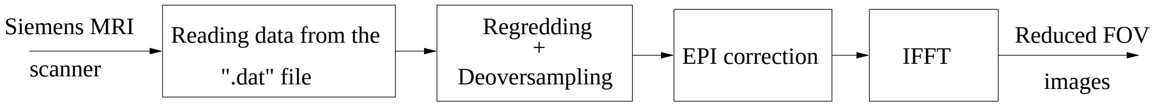


Figure 6.3: Reconstructing reduced FOV images from the raw data.

This first reconstruction requires two special steps:

i) *k-space regredding and deoversampling* to take into account non-uniform k space sampling, which occurs in fast MRI sequences like GE-EPI;

ii) *EPI correction* to remove the EPI Nyquist ghost artifacts due to the odd-even echo inconsistencies.

Once the reduced FOV images are available, the proposed reconstruction approaches have to be used in order to reconstruct the full FOV images which will be used for the fMRI study. The proposed reconstruction algorithms also require the estimation of the coil sensitivity maps. These sensitivity maps were obtained using the method in [Pruessmann et al., 1999b] by dividing the coil-specific images by the module of the Sum Of Squares (SOS) images, which are calculated based on an acquisition of the k-space center (24 lines) before the N_r scans. For comparison purpose, our reconstruction pipeline and the Siemens reconstruction have been used to generate EPI volumes in order to measure their impact on brain activation detection. Figs. 6.4-6.6 illustrate reconstructed slices using the mSENSE algorithm (SENSE reconstruction implemented in the Siemens scanner), as well as our pipeline including the UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithms. In order to have a general view of the brain, these figures illustrate the reconstructed axial, coronal and sagittal slices, respectively. These figures easily show that our pipeline allows to achieve more accurate reconstruction than the mSENSE algorithm in terms of reconstruction artifacts. In fact, the mSENSE reconstructed images present artifacts located at the center and image borders, which may be nasty in activation detection in these area such as temporal lobes. Note that these conclusions are reproducible across subjects although the artifacts may appear on different slices (see red circled). Note also that, in contrast to the Siemens reconstruction, our pipeline do not involve any signal homogeneity filter that may introduce some bias in the voxel intensities.

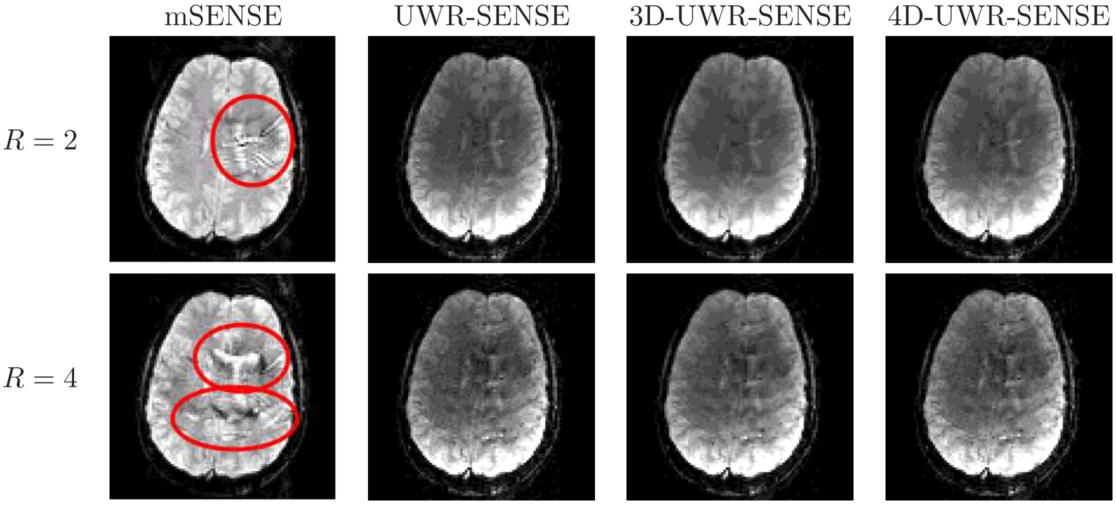


Figure 6.4: **Axial** reconstructed slices using mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$ and $R = 4$ with 2×2 mm^2 in-plane spatial resolution. Red circles and ellipsoids indicate reconstruction artifacts.

Results and comparisons in terms of statistical sensitivity are given in the two following sections.

6.3.2 Subject-level analysis

We quantitatively compare here the performance of the mSENSE and our parallel MRI reconstruction algorithms. Comparisons will concern the UWR-SENSE and 4D-UWR-SENSE algorithms. Moreover, to better investigate the temporal regularization effect, another comparison with what we will denote here by 3D-UWR-SENSE algorithm will be provided. The 3D-UWR-SENSE algorithm is simply obtained from the 4D-UWR-SENSE algorithm by turning off the temporal regularization, i.e. by setting $\kappa = 0$ in Eq. (4.67). fMRI data analysis has been conducted using the GLM framework proposed in SPM5 in which the design matrix (see Fig. 6.7) relying on ten experimental conditions has been built

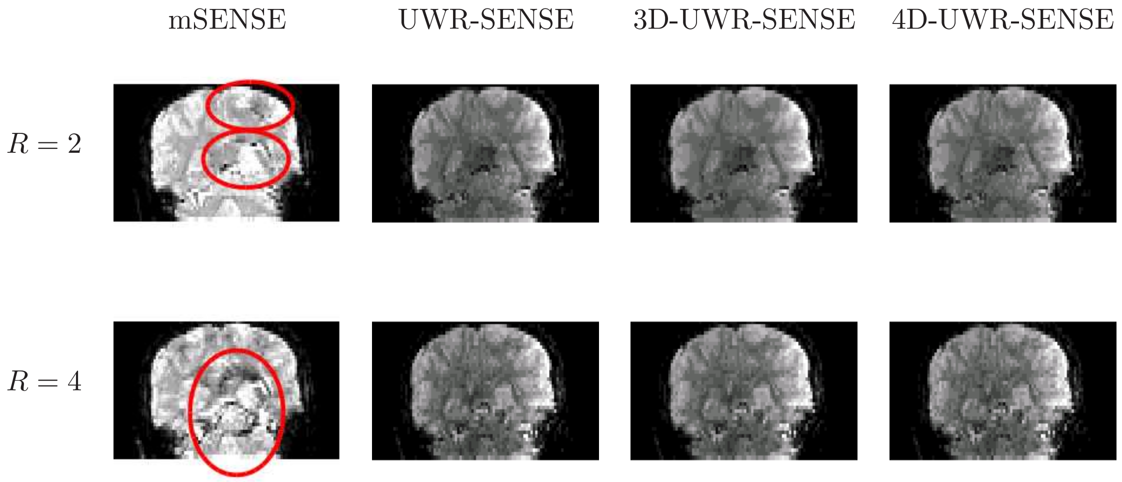


Figure 6.5: **Coronal** reconstructed slices using mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$ and $R = 4$ with 2×2 mm^2 in-plane spatial resolution. Red circles and ellipsoids indicate reconstruction artifacts.

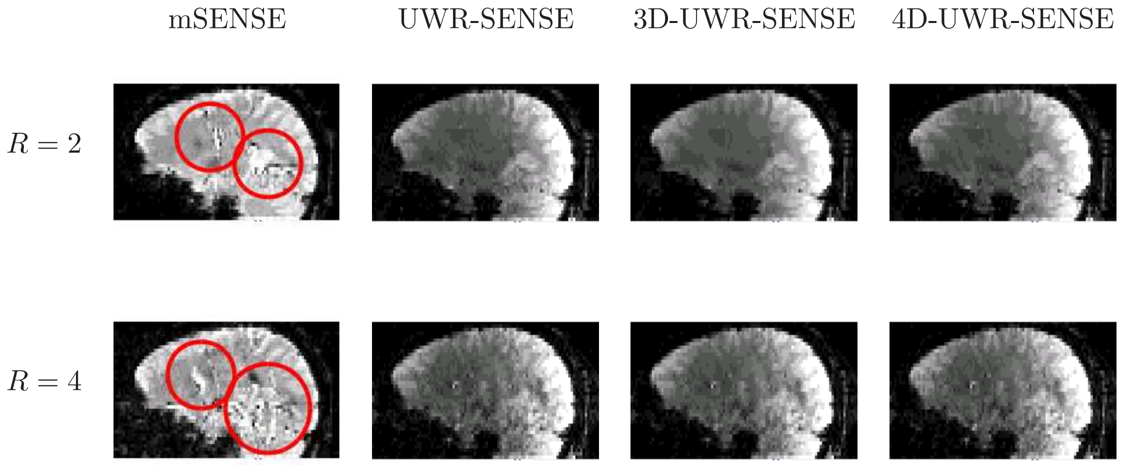


Figure 6.6: **Sagittal** reconstructed slices using mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$ and $R = 4$ with 2×2 mm^2 in-plane spatial resolution. Red circles indicate reconstruction artifacts.

up. This design matrix is made up of twenty regressors corresponding to the canonical HRF and its time derivative related to the ten used experimental conditions, in addition to the mean signal regressor ($N_c = 21$). Here, we only report results involving the Left click vs. Right click (`Lc-Rc`) and Auditory vs. Visual (`A-V`) contrasts since the expected activations, which lie in different parts of the brain, can be putatively corrupted by reconstruction artifacts. The `A-V` contrast defines a compound comparison which involves four stimuli. These stimuli are presented either in the visual or auditory modality. In this sense, this comparison aims only at localizing sensory brain areas, i.e. the primary auditory cortices. On the other hand, the `Lc-Rc` contrast defines a compound comparison which involves two stimuli which are also presented either in the visual or auditory

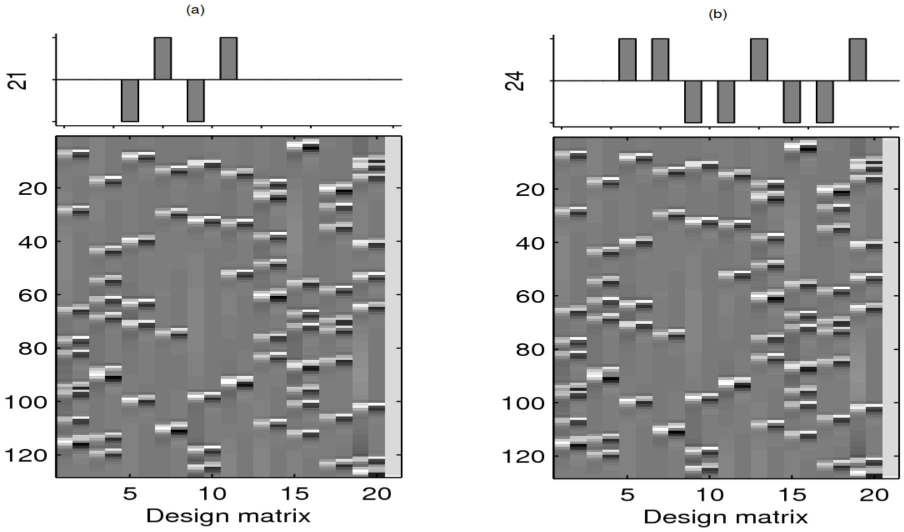


Figure 6.7: (a): design matrix and the `Lc-Rc` contrast involving two conditions (grouping auditory and visual modalities); (b): design matrix and the `A-V` contrast involving four conditions (sentence, computation, left click, right clisk). The design matrix is made up of twenty regressors corresponding to the canonical HRF and its time derivative for each of the ten experimental conditions, in addition to a mean signal regressor.

modality. This comparison aims therefore at detecting lateralization effect in the motor cortex. These two contrasts are illustrated in Fig. 6.7 (a) and (b), respectively. Results here are reported in terms of Student-t statistical maps thresholded at a $p = 0.05$ p-value corrected for multiple comparisons [Ashburner et al., 2004] as well as statistical tables that provide cluster and voxel-level p-values, maximal T-scores and corresponding locations of these peaks. For the `A-V` contrast, it is shown in Fig. 6.8 that the proposed algorithms allow to retrieve expected bilateral activations in the temporal lobes elicited by speech perception and comprehension involved in the `A-V` contrast. Conversely, the mSENSE method recovered smaller clusters: activation cluster in the left hemisphere is somehow lost due to strong reconstruction artifacts. This result holds both for $R = 2$ and $R = 4$. From a quantitative point of view, the proposed algorithms recover larger clusters whose excursion or local maxima is close to the one obtained using mSENSE. Concerning the largest clusters, our algorithms give very often the highest T-score of the local maxima. Tables 6.1-6.2 give the p-values at the voxel and cluster levels, as well as the size, T-score and position of the most significant cluster maxima. Note that clusters in these tables are listed in a decreasing significance order. These tables also indicate that the 4D-UWR-SENSE often gives larger detected clusters than 3D-UWR-SENSE and UWR-SENSE, which also always outperform the mSENSE algorithm in terms of detected cluster size. Interestingly, when looking at the activation peak by referring to Tables 6.1-6.2 or the blue cross in Fig. 6.8, it can be noticed that the mSENSE reconstruction always leads to an activation peak which is situated in the right hemisphere both for $R = 2$ and $R = 4$. However, for $R = 2$, the 4D-UWR-SENSE reconstruction allows to detect an activation peak in the left hemisphere. For $R = 4$, all our proposed methods retrieve also an activation peak in the left hemisphere. This is actually more coherent with the experimental circumstances since only one headphone has been used in the right ear because of the small coil dimensions which did not allow to use two headphones. For this reason, the right ear has been better protected from acoustic noise. It is easier to detect evoked activity in the right temporal lobe related to the auditory stimuli. However, since all the subjects are right handed, speech perception usually elicits stronger activation in the left temporal lobe. Here, using the mSENSE algorithm, this activation is lost because of low SNR. Our methodology enables the recovery of activation in the left superior temporal sulcus. These comparisons show therefore that our methods are also more robust to acoustic noise. Moreover, results for $R = 2$ show that the 4D-UWR-SENSE algorithm is also more robust than UWR-SENSE and 3D-UWR-SENSE to acoustic noise.

Table 6.1: Significant statistical results for the `A-V` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 2$.

	cluster-level		voxel-level		
	p-value	Size	p-value	T-score	Position
mSENSE	$< 10^{-3}$	695	$< 10^{-3}$	10.59	-66 -38 12
	$< 10^{-3}$	550	$< 10^{-3}$	11.76	64 -8 3
	0.009	28	$< 10^{-3}$	5.12	48 -26 12
UWR-SENSE	$< 10^{-3}$	847	$< 10^{-3}$	10.82	-52 -12 3
	$< 10^{-3}$	804	$< 10^{-3}$	11.37	60 -20 9
	0.006	32	0.001	4.53	-48 28 9
3D-UWR-SENSE	$<10^{-3}$	886	$< 10^{-3}$	11.60	58 -20 9
	$< 10^{-3}$	796	$< 10^{-3}$	10.69	-54 -12 6
4D-UWR-SENSE	$<10^{-3}$	**889**	$<10^{-3}$	11.55	50 -14 3
	$<10^{-3}$	881	$<10^{-3}$	**12.11**	-60 -20 9
	0.002	38	0.001	4.56	44 30 9

For the `Lc-Rc` contrast, Fig. 6.9 shows that the proposed algorithms allow the detection of expected activated areas missed by the mSENSE algorithm especially for $R = 4$. Moreover, the detected activation peak using the mSENSE reconstruction for $R = 4$ moves far away from the right motor cortex (see blue cross in Fig. 6.9), which is not the case of our algorithms that always allow to detect significant activations in the expected brain area. Quantitative results in Tables 6.3-6.4 show that larger clusters with higher local T-scores maxima are detected using our pipeline with the UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithms, both at $R = 2$ and $R = 4$.

In order to further investigate the effect of the reconstruction algorithm in pMRI on the statistical analysis, Figs. 6.10-6.11 illustrate 3D plots of the detected activated area for the `Lc-Rc` contrast for $R = 2$ and $R = 4$, respectively. It is clear that the proposed approaches outperform the mSENSE algorithm in terms of statistical sensitivity and size of the detected clusters. We can also notice through Fig. 6.11 where $R = 4$ that the proposed UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithms allow to detect activated area which are missed using the mSENSE reconstruction.

Table 6.2: Significant statistical results for the `A-V` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 4$.

	cluster-level		voxel-level		
	p-value	Size	p-value	T-score	Position
mSENSE	$< 10^{-3}$	264	$< 10^{-3}$	9.21	-66 -40 12
	$< 10^{-3}$	197	$< 10^{-3}$	10.02	66 -8 6
	0.009	28	$< 10^{-3}$	5.12	48 -26 12
UWR-SENSE	$< 10^{-3}$	441	$< 10^{-3}$	8.97	62 -8 12
	$< 10^{-3}$	139	$< 10^{-3}$	11.26	-60 -42 9
	0.001	28	0.022	5.44	64 -20 6
3D-UWR-SENSE	$< 10^{-3}$	**472**	$< 10^{-3}$	9.26	62 -8 9
	$< 10^{-3}$	184	$< 10^{-3}$	**11.34**	-62 -38 12
	0.005	23	$< 10^{-3}$	7.10	60 -34 6
4D-UWR-SENSE	$< 10^{-3}$	447	$< 10^{-3}$	10.11	-60 -42 12
	$< 10^{-3}$	174	$< 10^{-3}$	9.04	62 -6 6
	$< 10^{-3}$	54	$< 10^{-3}$	7.65	60 -32 3

Table 6.3: Significant statistical results for the `Lc-Rc` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 2$.

	cluster-level		voxel-level		
	p-value	Size	p-value	T-score	Position
mSENSE	$< 10^{-3}$	172	0.001	6.10	36 -22 66
UWR-SENSE	$< 10^{-3}$	169	0.007	5.68	46 -26 66
3D-UWR-SENSE	$<10^{-3}$	198	0.042	5	40 -32 69
4D-UWR-SENSE	$<10^{-3}$	**208**	<0.001	**6.12**	46 -26 66
	<0.066	21	<0.005	5.28	36 -8 66

Table 6.4: Significant statistical results for the `Lc-Rc` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 4$.

	cluster-level		voxel-level		
	p-value	Size	p-value	T-score	Position
mSENSE	$< 10^{-3}$	81	0.001	6.08	34 -24 69
UWR-SENSE	$< 10^{-3}$	**122**	0.001	6.04	32 -22 72
3D-UWR-SENSE	$< 10^{-3}$	105	$< 10^{-3}$	6.33	32 -24 72
	$< 10^{-3}$	20	0.086	4.21	38 -20 48
4D-UWR-SENSE	$< 10^{-3}$	114	$< 10^{-3}$	**6.73**	48 -18 63
	$< 10^{-3}$	33	0.008	5.16	42 -30 66

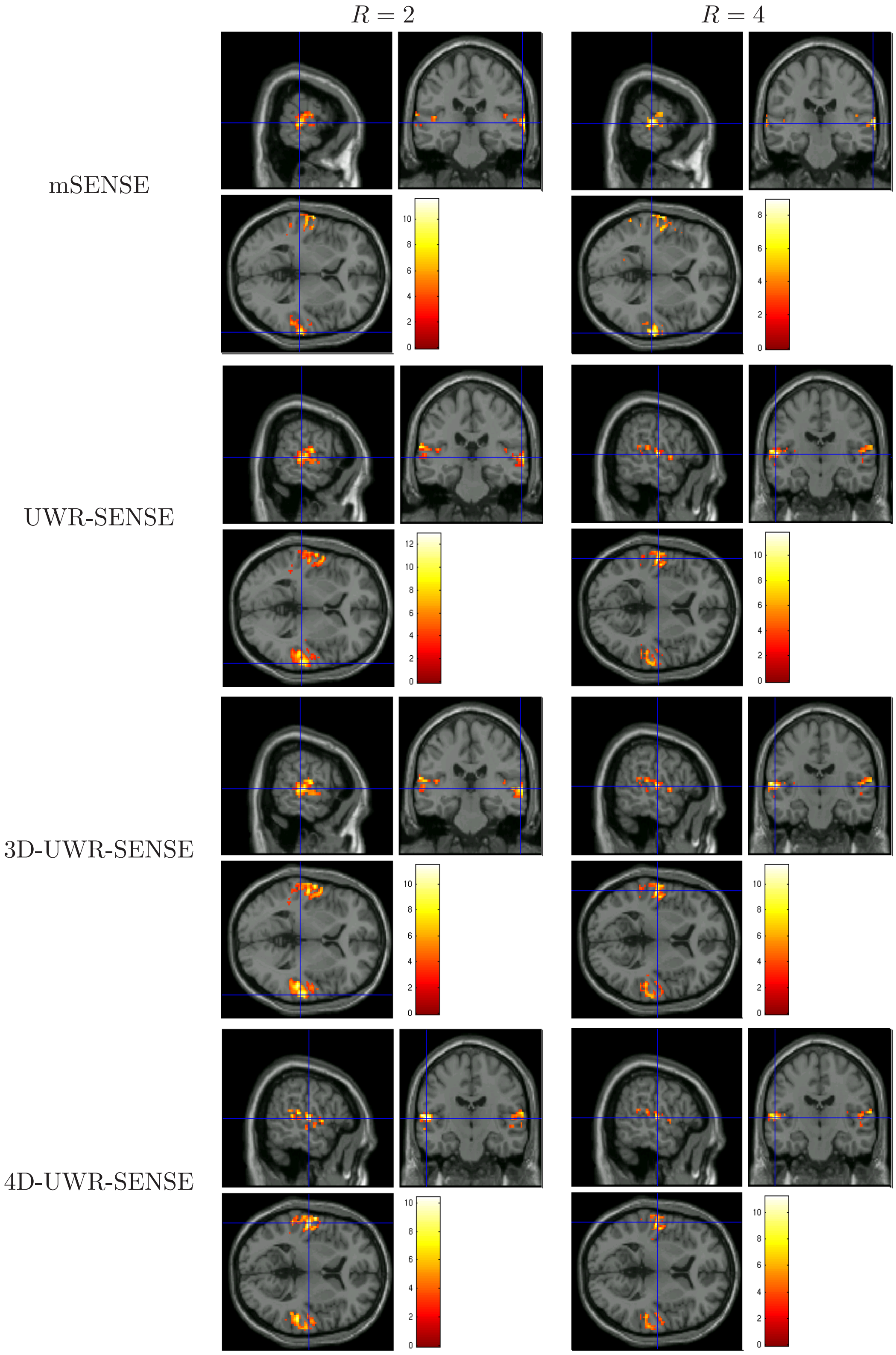


Figure 6.8: Subject-level student-t maps superimposed to anatomical MRI for the `A-V` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE. Neurological convention: left is left. The blue cross indicates the activation peak.

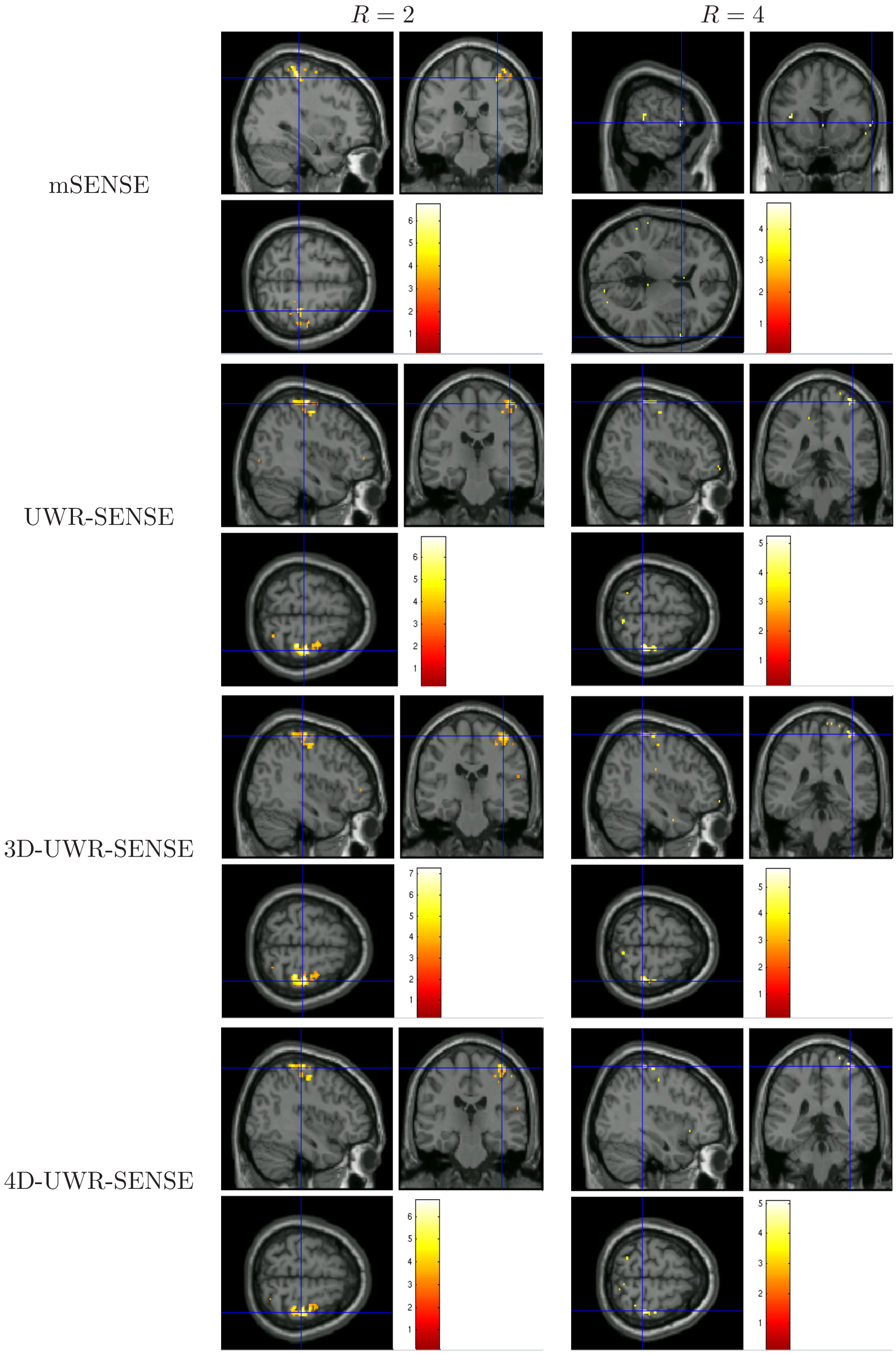


Figure 6.9: Subject-level student-t maps superimposed to anatomical MRI for the `Lc-Rc` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE. Neurological convention: left is left. The blue cross indicates the activation peak.

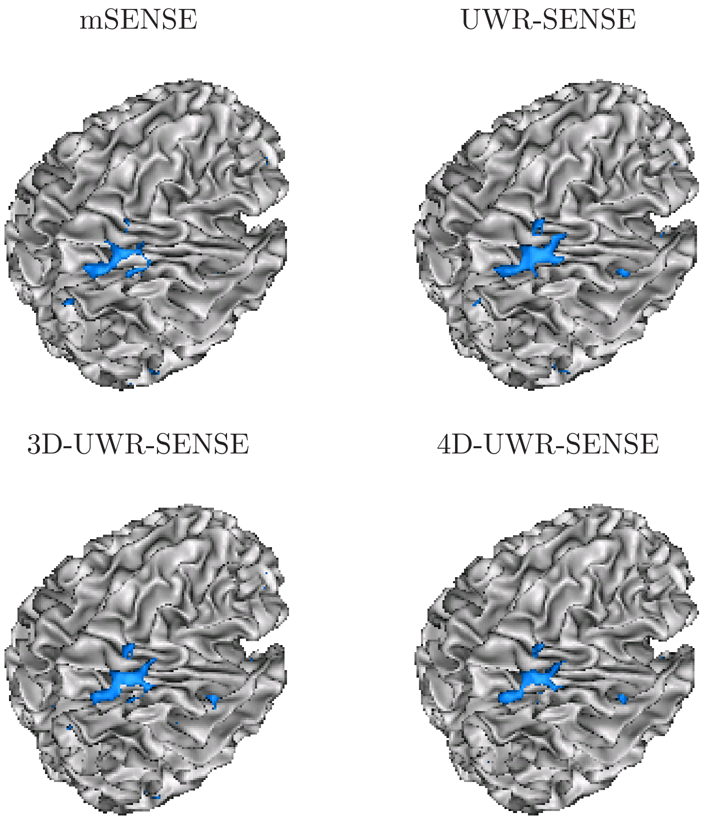


Figure 6.10: 3D plots of the detected activated area for the `Lc-Rc` contrast superimposed to a 3D mesh of the brain where $R = 2$. Activation foci appear in the right motor cortex whatever the reconstruction method, but with larger spatial extent using our regularized reconstructions.

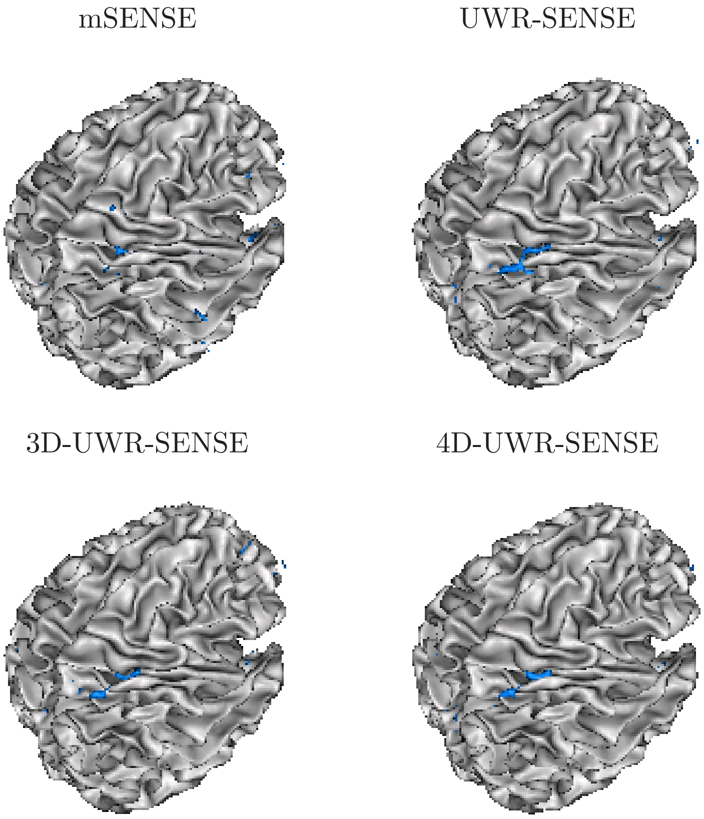


Figure 6.11: 3D plots of the detected activated area for the Lc-Rc contrast superimposed to a 3D mesh of the brain where $R = 4$. Only our reconstruction algorithms recover activation foci in the right motor cortex.

6.3.3 Group-level analysis

In this fMRI study, random effect analysis has been used for group-level validation involving fifteen healthy subjects of the acquired database. As for the subject-level analysis, our reconstruction pipeline and the Siemens reconstruction have been used to generate EPI volumes in order to compare these methods in terms of activation detection at the group level. For the `A-V` contrast, Maximum Intensity Projection (MIP) student-t maps for $R = 2$ and $R = 4$ are provided in Figs. 6.12-6.13. These figures clearly show that, both for for $R = 2$ and $R = 4$, our pipeline allows to better detect expected significant bitaleral activations than the Siemens pipeline. From a quantitative viewpoint, results for the most significant clusters are provided for the `A-V` contrast in Tables 6.5-6.6 for $R = 2$ and $R = 4$, respectively. Voxel and cluster-level results show that our pipeline outperforms the mSENSE reconstruction in terms of clusters size and T-score maxima. It should be noted here that, as reported in Section 6.3.2, stronger activation is expected in the left temporal lobe due to the use of only one headphone at the right ear. When looking at the position of the largest detected clusters for $R = 2$ (see Table 6.5), we can notice that indeed, our methods allow the detection of higher activations in the left hemisphere in contrast to the mSENSE reconstruction. However, for $R = 4$ (see Table 6.6), all the methods detect higher activations in the left hemisphere.

Table 6.5: Significant statistical results at the group-level for the `A-V` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 2$.

	cluster-level		voxel-level		
	p-value	Size	p-value	T-score	Position
mSENSE	$< 10^{-3}$	539	0.001	10.73	58 -10 3
	0.001	390	0.031	8.11	-44 -22 6
UWR-SENSE	$< 10^{-3}$	964	$< 10^{-3}$	12.80	-52 -10 6
	$< 10^{-3}$	906	$< 10^{-3}$	12.79	62 -6 0
3D-UWR-SENSE	$< 10^{-3}$	**1075**	$< 10^{-3}$	**14.15**	-42 -28 12
	$< 10^{-3}$	992	0.001	13.05	60 -12 0
4D-UWR-SENSE	$< 10^{-3}$	913	0.001	11.44	-42 -24 12
	$< 10^{-3}$	762	0.003	10.37	58 -10 3

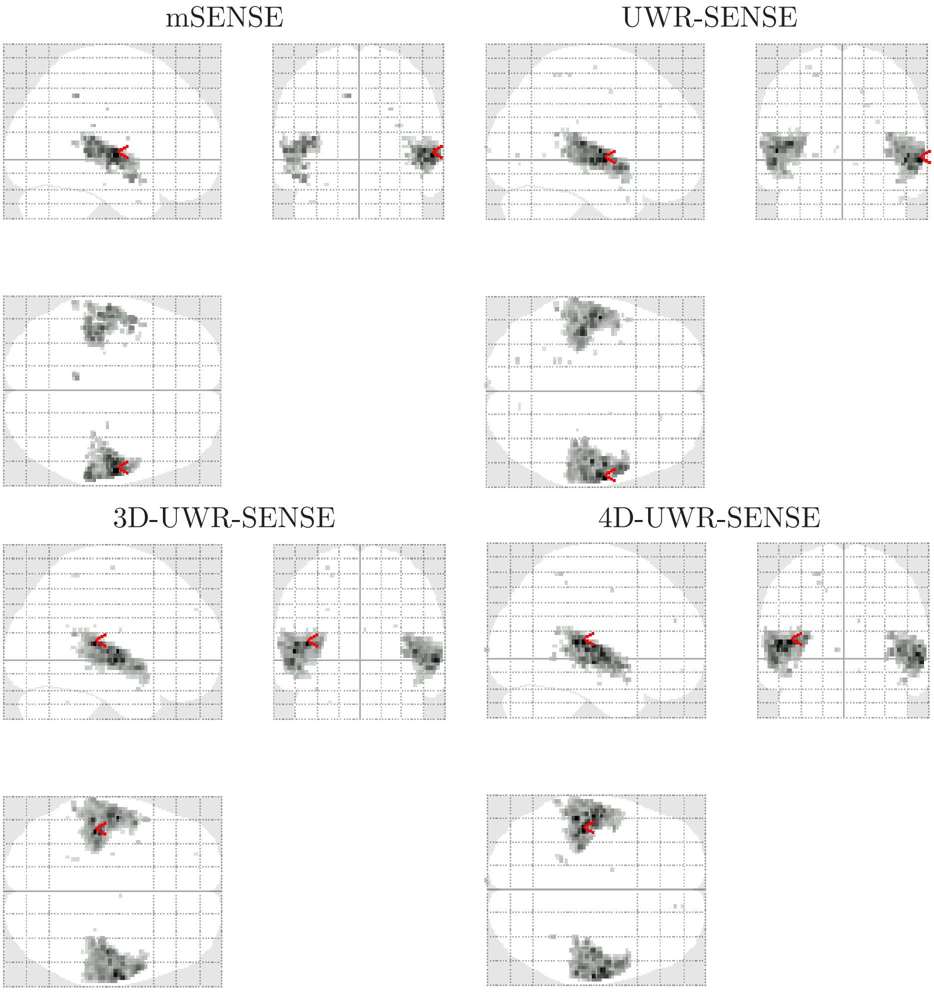
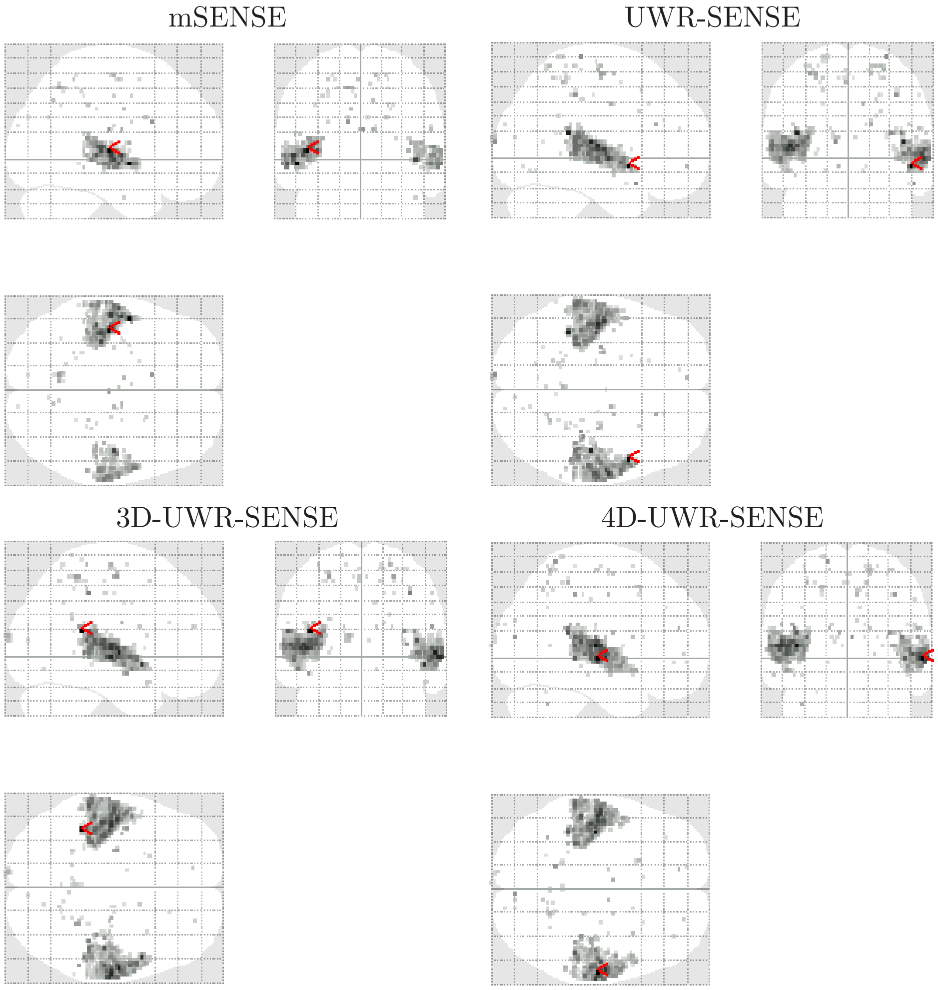


Figure 6.12: Group-level student-t maps for the `A-V` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$. Neurological convention: left is left. Red arrows indicate the global maxima.



mSENSE

UWR-SENSE

3D-UWR-SENSE

4D-UWR-SENSE

Figure 6.13: Group-level student-t maps for the `A-V` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 4$. Neurological convention: left is left. Red arrows indicate the global maxima.

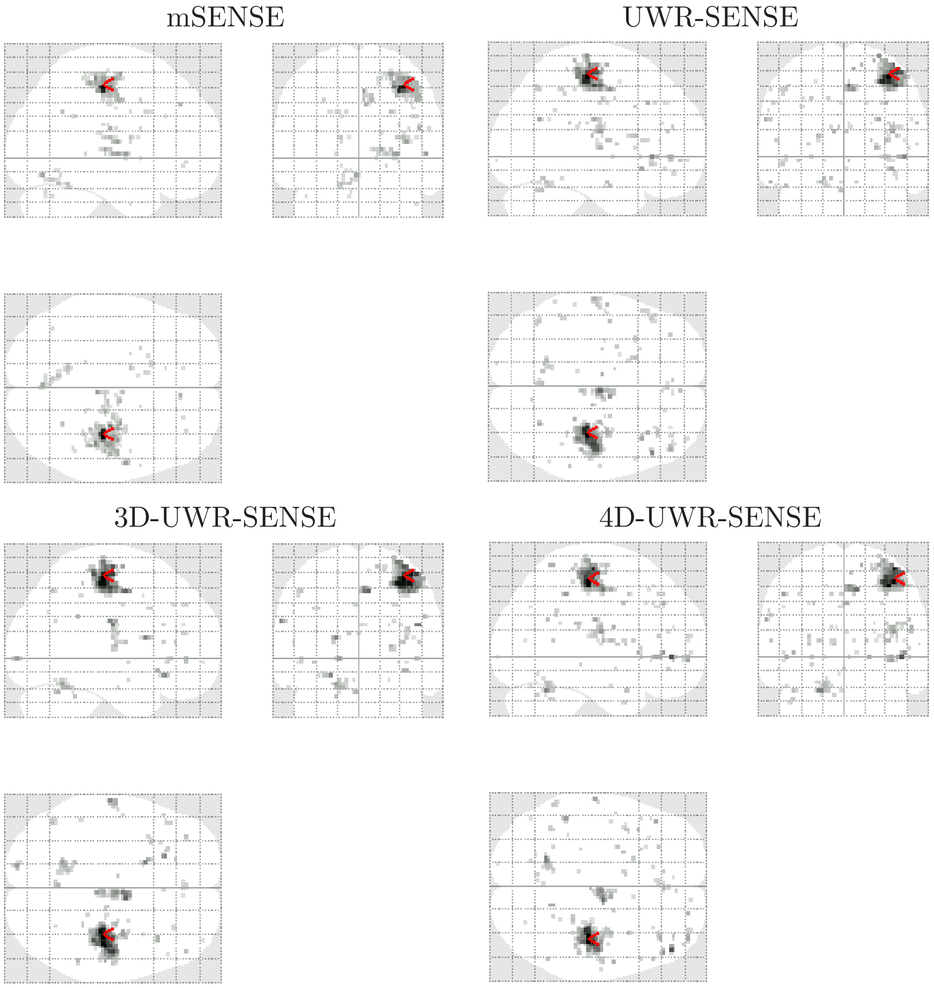
Table 6.6: Significant statistical results at the group-level for the `A-V` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 4$.

	cluster-level		voxel-level		
	p-value	Size	p-value	T-score	Position
mSENSE	$< 10^{-3}$	424	0.002	9.86	-42 -18 6
	$< 10^{-3}$	222	0.048	7.76	44 -14 0
UWR-SENSE	$< 10^{-3}$	639	0.008	10.64	-40 -40 18
	$< 10^{-3}$	555	0.002	9.52	50 8 -6
3D-UWR-SENSE	$< 10^{-3}$	**784**	0.001	**11.06**	-40 -40 18
	$< 10^{-3}$	658	0.012	9.93	58 -10 3
4D-UWR-SENSE	$< 10^{-3}$	653	0.010	10.11	-40 -18 9
	$< 10^{-3}$	447	0.004	9.35	60 -16 0
	$< 10^{-3}$	34	0.193	7.24	46 -34 0

For the `Lc-Rc` contrast, student-t maps for $R = 2$ and $R = 4$ are provided in Figs. 6.14-6.15. These figures clearly show that, both for for $R = 2$ and $R = 4$, our pipeline allows to detect much more spatially extended activation area in the motor cortex. Quantitatively speaking, Tables 6.7-6.8 show that the detected clusters using our pipeline reconstruction are of larger size and higher T-score maxima than the ones detected based on the mSENSE reconstruction. These results are valid both for $R = 2$ and $R = 4$. It is also worth noticing that detected activated area using our pipeline reconstruction are more excentric that with the mSENSE reconstruction, which has actually been expected regarding to the motor cortex location.

Table 6.7: Significant statistical results at the group-level for the `Lc-Rc` contrast (corrected for multiple comparisons at $p = 0.05$). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 2$.

	cluster-level		voxel-level		
	p-value	Size	p-value	T-score	Position
mSENSE	$< 10^{-3}$	229	$< 10^{-3}$	12.21	36 -22 51
	0.001	22	0.665	6.25	30 -10 3
UWR-SENSE	$< 10^{-3}$	350	0.005	9.83	36 -22 57
	$< 10^{-3}$	35	0.286	7.02	4 -12 51
3D-UWR-SENSE	$< 10^{-3}$	**418**	0.011	9.28	36 -22 57
	$< 10^{-3}$	55	0.437	**13.05**	8 -4 48
4D-UWR-SENSE	$< 10^{-3}$	304	0.001	9.77	40 -22 54
	$< 10^{-3}$	38	0.439	6.74	10 -10 48



mSENSE

UWR-SENSE

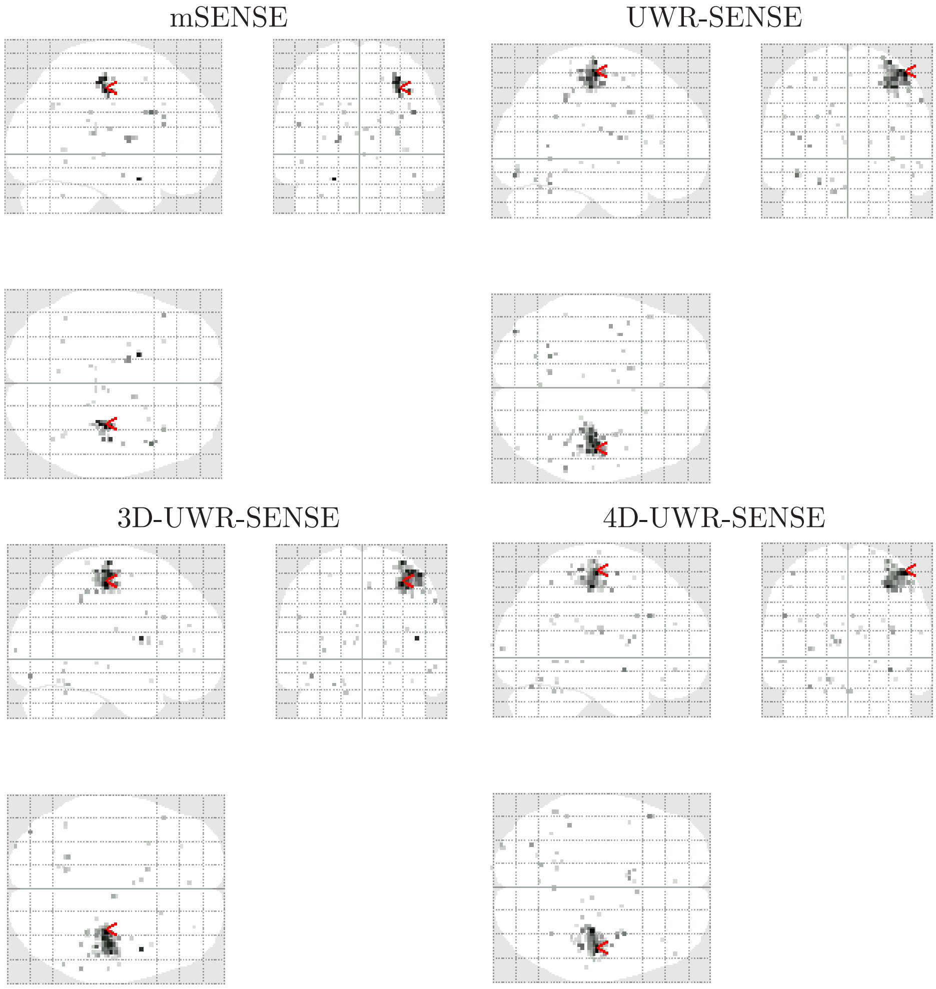
3D-UWR-SENSE

4D-UWR-SENSE

Figure 6.14: Group-level student-t maps for the `Lc-Rc` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 2$. Neurological convention: left is left. Red arrows indicate the global maxima.

Table 6.8: Significant statistical results at the group-level for the `Lc-Rc` contrast (corrected for multiple comparisons). Images were reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithm for $R = 4$.

	cluster-level		voxel-level		
	p-value	Size	p-value	T-score	Position
mSENSE	$< 10^{-3}$	38	0.990	5.97	32 -20 45
UWR-SENSE	$< 10^{-3}$	163	0.128	7.51	46 -18 60
3D-UWR-SENSE	$< 10^{-3}$	**174**	0.182	7.27	32 -22 54
4D-UWR-SENSE	$< 10^{-3}$	162	0.111	**7.61**	46 -18 60



mSENSE

UWR-SENSE

3D-UWR-SENSE

4D-UWR-SENSE

Figure 6.15: Group-level student-t maps for the `Lc-Rc` contrast where data have been reconstructed using the mSENSE, UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE for $R = 4$. Neurological convention: left is left. Red arrows indicate the global maxima.

6.4 Discussion

The illustrated results in this chapter clearly show that the proposed UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE algorithms outperform the mSENSE reconstruction algorithm available in the Siemens scanner pipeline. Better results using our algorithms have been obtained both quantitatively and qualitatively. First, our reconstruction allows us to detect larger clusters with higher T-score maxima, both at the subject and group-level. These results were confirmed for two different reduction factors $R = 2$ and $R = 4$ and showed that for $R = 4$, the gain in statistical sensitivity provided by our reconstruction is much more important compared with the mSENSE reconstruction. This observation was also confirmed for two different contrasts. Second, it was shown that our reconstruction algorithms allow always to retrieve expected significant activations in the right brain region in contrast to the mSENSE reconstruction. In accordance with the experimental circumstances, comparisons also showed that our algorithms are more robust to acoustic noise. Moreover, among our algorithms, it turns out from presented expreimental validations that the 4D-UWR-SENSE algorithm is more robust to acoustic noise than the UWR-SENSE and 3D-UWR-SENSE ones. It should also be noted that some comparisons showed that the 3D-UWR-SENSE algorithm may outperform the 4D-UWR-SENSE one in terms of detected cluster size or T-score maxima. This observation suggests that the temporal regularization using a given setup of hyper-parameters may be more adapted to some brain regions than others, to some stimulations than others (inducing different BOLD signal contrasts, and thus statistical properties)... It is also worth noticing that the same spatial and temporal regularization hyper-parameters have been used to reconstruct data relative to all the fifteen subjects, which may actually explain the underperformance of the 4D-UWR-SENSE algorithm in some cases compared with the 3D-UWR-SENSE one.

6.5 Conclusion

In this chapter, we examined the impact of the parallel imaging reconstruction algorithm on the statistical performance for brain activity detection in BOLD fMRI data acquired using a GE-EPI sequence. At the subject and group-levels (at 3 Tesla), we showed that our proposed algorithms (UWR-SENSE, 3D-UWR-SENSE and 4D-UWR-SENSE) outperform the mSENSE method both qualitatively and quantitatively from a statistical viewpoint. We showed how the choice of the parallel imaging reconstruction algorithm impacts the statistical sensitivity in fMRI data analysis and enables whole brain neuroscience studies at high spatial resolution, even when comparing subtle contrasts. Future comparisons would involve higher resolved images such as 1.5×1.5 mm^2 in-plane ones, which can only be reached using parallel imaging with a reduction factor $R = 4$ if one wants to keep reasonable TR values (typically TR = 2.4 s). Ongoing work will also concern the combination with the Joint Detection Estimation (JDE) approach [Makni et al., 2008; Vincent et al., 2010] to enable new fMRI data analysis pipeline. Another extension would also be to combine our wavelet regularized reconstruction with the WSPM approach [Van De Ville et al., 2007] where the statistical analysis is directly performed in the wavelet transform domain.

Chapter 7

Conclusion

MRI is an imaging modality which became widely used in clinical daily routines due to its non-invasivity and good spatio-temporal resolution. In addition, parallel imaging is a more recent technique for medical imaging. It presents numerous advantages when used in MRI like reducing the global imaging time or improving the spatio-temporal resolution. Unfortunately, standard reconstruction methods such as the SENSE algorithm do not perform well when experimental conditions become quite severe.

This PhD thesis contributed to the development of new reconstruction methods for parallel MRI which allow to achieve good reconstruction quality even under degraded experimental setups. The developed approaches take part of the pMRI regularized reconstruction literature like Tikhonov and Total Variation regularization for instance. They are mainly based on the SENSE algorithm which goes back to 1999. The ill-posed inverse problem of pMRI reconstruction is regularized by providing suitable information about the images to be reconstructed in the appropriate space. The general regularization issue was first properly addressed in Chapter 3 while outlining the main practical difficulties which may be encountered and proposing appropriate solutions. In Chapter 4, the pMRI regularization problem was then handled from a Bayesian viewpoint. The injected information consisted of prior knowledge about the solutions in the wavelet transform domain. The inherent optimization problem for each developed regularization approach was carefully addressed while taking the complex-valued nature of the data into account. Efficient iterative optimization algorithms were then selected to satisfy the particularities of each optimization problem. The retained algorithms were extended to the complex-valued case in order to be used under proven convergence guaranties. Experimental validations of the developed approaches in fMRI were also provided in Chapter 6, and results show that they allow to improve the statistical sensitivity/specificity of activation detection in fMRI.

Another issue which was addressed in this PhD (Chapter 5 of this PhD) is the hyper-parameter estimation in variational regularization when using overcomplete dictionaries. This problem was also handled from a Bayesian viewpoint by proposing a new approach to estimate the hyper-parameters based on a noisy observation of the image/signal whose frame coefficients are characterized by a suitable prior distributions.

In summary, the main contributions of this PhD are the followings:

1) The proposed UWR-SENSE and CWR-SENSE algrithms are among the most recent reconstruction methods for SENSE imaging in pMRI. The regularized reconstruction they provide is designed in a Bayesian framework and uses sparsity-promoting priors for the wavelet coefficients of the image to be reconstructed. They also rely on one of the most recent and fast iterative convex optimization algorithms. This proximal optimization algorithm has been carefully chosen and adapted depending on the specificities of the criteria to be optimized.

2) Wavelet and Total Variation penalizations have been combined in a joint regularization framework for pMRI reconstruction in order to take advantage of both of them. Results show that this combination leads to slightly better reconstruction quality.

3) A reconstruction method using spatio-temporal regularization has also been proposed for applications where series of images are acquired and have to be reconstructed such as in fMRI. This method accounts for temporal dependencies between the acquired volumes and proceeds by a joint temporal and 3D wavelet regularization.

4) A Bayesian algorithm has been proposed to estimate the hyper-parameters involved in the regularization process from a noisy observation of the image to be reconstructed. This algorithm was also extended to estimate the hyper-parameters in the case of joint Wavelet-Total Variation prior.

5) The proposed reconstruction methods were validated in fMRI. This validation allowed us to study the impact of our reconstruction algorithms on statistical sensitivity/specificity of activation detection in fMRI. Results also showed that, compared with the mSENSE reconstruction algorithm provided with the Siemens scanner, our reconstruction helps achieving more accurate activation detection.

As a future work, this PhD opens several perspectives both from a methodological and application viewpoints among which we mention:

1) **Using non-convex penalizations**:

   Designing the regularization method while using frame representations deeply depends on the chosen prior for the frame coefficients. Promoting the sparsity of the considered representation is therefore of great interest in the reconstruction procedure. In this work, we used sparsity-promoting priors which guarantee the convexity of the final optimality criterion. This allowed us to take advantage of the availability of fast and efficient convex optimization algorithms. However, sparser but not necessarily convex priors can potentially be used to achieve better reconstruction quality. In this case, the optimization framework has to be adapted by resorting to non-convex optimization algorithms.

2) **Tissue-dependent regularization**:

   In neuroimaging, it is well known that the brain is made up of three main components: white matter, gray matter and cerebral spinal fluid. Moreover, many efficient segmentation algorithms are now available and allow to precisely extract these three components. Based on such a segmentation, the CWR-SENSE algorithm may be extended to account for the differences between these three tissue types inside the brain, which may allow one to adapt the regularization process to the image content.

This differenciation would also help to outline the different statistical properties of the three tissues either in the original of the transform space. Moreover, extending the proposed methods by combining the segmentation and regularization steps in a joint framework would also lead to a fully automatic method exploiting the brain tissue varieties.

3) **2D-SENSE imaging**:

In this work we focused on the SENSE imaging technique where a sub-sampling of the k-space is performed along only one direction, i.e. the phase encoding direction. In 3D imaging, sub-sampling the k-space may also be performed along the partition direction i.e. the slice encoding one: we talk about 2D-SENSE imaging. In this context, extending the proposed regularization approaches to deal with this more complicated problem would drastically improve the reconstruction performance for such imaging sequences. In fact, sub-sampling the k-space along two directions enables the use of sparser acquisition schemes corresponding for instance to 3 acceleration factors, and hence allows us to reduce the global imaging time and/or to improve the spatio-temporal resolution of the acquired images. However, the counterpart is necessarily a loss in the reconstruction quality. Since the coil sensitivity profile estimation will also be more complicated, the problem will become severely ill-posed and a regularization procedure will be necessary. In this context, extending the developed regularization approaches will be fruitful especially when using 3D wavelet transforms, which would allow to smooth reconstruction artifacts along the three encoding directions.

4) **Fully Bayesian reconstruction**:

In the current regularization framework, the hyper-parameters are estimated separately, and then used in the regularization process. Moreover, the hyper-parameters of the real and imaginary parts are also estimated separately since the designed algorithm for the hyper-parameter estimation can only deal with real-valued signals. A first extension of this algorithm to deal with more sophisticated complex-valued models would therefore be fruitful. This would be of primary interest either in pMRI or even in other applications where data are complex-valued.
On the other hand, the adopted observation model in the hyper-parameter estimation algorithm reads:

$$\boldsymbol{y} = F^*\boldsymbol{x} + \boldsymbol{n}. \tag{7.1}$$

Except the frame adjoint operator, this model does not account for any other linear operator through which the reasurements may be degraded. In fact, the observation model in pMRI involves such a linear operator, which matches the sensitivity matrix. In this context, a more general and realistic model would read as follows:

$$\boldsymbol{y} = HF^*\boldsymbol{x} + \boldsymbol{n}, \tag{7.2}$$

where H is an arbitrary linear operator (which reduces to the sensitivity matrix in pMRI). Extending the proposed approach to account for this more general model enables also direct estimation of the full FOV image in pMRI. In fact, instead of estimating the hyper-parameters and then plugging them into the regularization process, this extension would lead to a fully Bayesian regularization of the pMRI inverse problem. This would be possible since the proposed approach allows also to estimate the frame coefficients.

5) **Other MRI applications**:

We mainly focused in this thesis on brain imaging. However, parallel imaging is widely used even in other MRI applications like cardiac and abdominal imaging. Applying our reconstruction methods to such applications would face a major problem: the used encoding trajectories in these application is not necessarily Cartesian. In this case, more attention has to be paid to the observation model.
On the other hand, the proposed spatio-temporal regularization would also be of great interest in cardiac imaging for instance where the heart dynamics are studied, and where several 3D images are acquired. However, if one wants to perform a temporal regularization using our spatio-temporal reconstruction approach, it is necessary to take the non-rigidity of the imaged object into account. In fact, in neuroimaging, the brain is not supposed to move and change its geometry between successive acquisitions. However, in cardiac imaging, this condition is no longer satisfied and the intrinsic deformation of the heart has to be taken into account.

Bibliography

Afacan, O., Hoge, S., Brooks, D. H., and Morocz, I. A. (2009). Increasing temporal resolution in hybrid 3D EPI fMRI studies using unfold. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 654–657, Boston, USA.

Ahmed, N. and Natarajen, R. (1974). Discrete Cosine Transform. *IEEE Transactions on Computers*, C(32).

Alghoniemy, M. and Tewfik, A. H. (2004). A sparse solution to the bounded subset selection problem : a network flow model approach. In *IEEE International Conference on Acoustics, Speech, and Signal (ICASSP)*, pages 89–92, Montreal, Canada.

Andrieu, C., Djuric, P. M., and Doucet, A. (2001). Model selection by MCMC computation. *Signal Processing*, 81:19–37.

Ashburner, J., Friston, K. J., and Penny, W. (2004). Introduction to random field theory. In Ashburner, J., Friston, K., and Penny, W., editors, *Human Brain Function*. Academic Press, 2nd edition.

Babacan, S. D., Molina, R., and Katsaggelos, A. K. (2009). Fast Bayesian compressive sensing using Laplace priors. In *IEEE International Conference on Acoustics, Speech, and Signal (ICASSP)*, pages 2873–2876, Taipei, Taiwan.

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., and Hyde, J. S. (1993). Processing strategies for time-course data sets in functional MRI of human brain. *Magnetic Resonance in Medicine*, 30(2):161–173.

Beck, A. and Teboulle, M. (2009). A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202.

Beckmann, C. F. and Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Medical Imaging*, 23(2):137–152.

Benali, H., Buvat, I., Anton, J. L., Pélégrini, M., Di Paola, M., Bittoun, J., Burnod, Y., and Di Paola, R. (1997). Space-time statistical model for functional MRI image sequences. In Duncan, J. and Gindi, G., editors, *Information Processing in Medical Imaging*, pages 285–298. Norwell, USA.

Bioucas-Dias, J. M. and Figueiredo, M. A. T. (2007). A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004.

Blaimer, M., Breuer, F., Mueller, M., Heidemann, R. M., Griswold, M. A., and Jakob, P. M. (2004). SMASH, SENSE, PILS, GRAPPA: How to choose the optimal method. *Magnetic Resonance in Medicine*, 15(4):223–236.

Block, K. T., Uecker, M., and Frahm, J. (2007). Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine*, 56(7):1086–1098.

Blumensath, T. and Davies, M. E. (2007). Monte Carlo methods for adaptive sparse approximations of time-series. *IEEE Transactions on Signal Processing*, 55(9):4474–4486.

Boubertakh, R., Giovannelli, J.-F., De Cesare, A., and Herment, A. (2006). Non-quadratic convex regularized reconstruction of MR images from spiral acquisitions. *Signal Processing*, 86(9):2479–2494.

Boubertakh, R., Herment, A., Giovannelli, J.-F., and De Cesare, A. (2000). MR image reconstruction from sparse data and spiral trajectories. In *17th Annual meeting of the European Society for Magnetic Resonance in Medicine and Biology*, page 85, Paris, France.

Breuel, T. M. (1996). Finding lines under bounded error. *Pattern recognition*, 29(1):167–178.

Briceño-Arias, L. M. and Combettes, P. L. (2009). Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery. *Numerical Mathematics: Theory, Methods, and Applications*, 2(4):485–508.

Brown, M. A. and Semelka, R. C. (2003). *MRI: Basic Principles and Applications*. Wiley-Liss, NJ, USA, 3rd edition edition.

Bullmore, E. T., Brammer, M., Williams, S. C., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., and Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, 35(2):261–277.

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001a). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(2):140–151.

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001b). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Human Brain Mapping*, 13(1):43–53.

Calhoun, V. D., Pekar, J. J., McGinty, V. B., Adali, T., Watson, T. D., and Pearlson, G. D. (2002). Independent component analysis of fMRI data in the complex domain. *Magnetic Resonance in Medicine*, 48(1):180–192.

Candès, E., Demanet, L., Donoho, D., and Ying, L. (2006). Fast discrete curvelet transforms. *SIAM Multiscale Model. Simul.*, 5(3):861–899.

Candès, E. J., , and Donoho, D. L. (2002). Recovering edges in ill-posed inverse problems: optimality of curvelet frames. *Ann. Stat.*, 30(3):784–842.

Cappé, O. (2002). A Bayesian approach for simultaneous segmentation and classification of count data. *IEEE Transactions on Signal Processing*, 50(2):400–410.

Carlavan, M., Weiss, P., and Blanc-Féraud, L. (2010). Régularité et parcimonie pour les problèmes inverses en imagerie : algorithmes et comparaisons. *Traitement du Signal*. To appear.

Carlavan, M., Weiss, P., Blanc-Féraud, L., and Zerubia, J. (2009). Algorithme rapide pour la restauration d'image régularisée sur les coefficients d'ondelettes. In *GRETSI Symposium on Signal and Image Processing*, Dijon, France.

Chaari, L., Pesquet, J.-C., Benazza-Benyahia, A., and Ciuciu, P. (2008). Autocalibrated parallel MRI reconstruction in the wavelet domain. In *IEEE International Sympsium on Biomedical Imaging (ISBI)*, pages 756–759, Paris, France.

Chaari, L., Pesquet, J.-C., Benazza-Benyahia, A., and Ciuciu, P. (2010). A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging. *Medical Image Analysis*. In Press.

Chambolle, A. and Lions, P. L. (1997). Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188.

Chaux, C., Combettes, P., Pesquet, J.-C., and Wajs, V. (2007). A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4):1495–1518.

Chaux, C., Combettes, P. L., Pesquet, J.-C., and Wajs, V. (2005). A forward-backward algorithm for image restoration with sparse representations. In *Signal Processing with Adaptative Sparse Structured Representations*, pages 49–52, Rennes, France.

Chaux, C., Combettes, P. L., Pesquet, J.-C., and Wajs, V. (2006a). Iterative image deconvolution using overcomplete representations. In *European Signal Processing Conference*, Florence, Italy. 5 pp.

Chaux, C., Duval, L., and Pesquet, J.-C. (2006b). Image analysis using a dual-tree m-band wavelet transform. *IEEE Transactions on Image Processing*, 15(8):2397–2412.

Chaux, C., Pesquet, J.-C., and Pustelnik, N. (2009). Nested iterative algorithms for convex constrained image recovery problems. *SIAM Journal on Imaging Sciences*, 82(2):730–762.

Chen, W. and Ogawa, S. (1999). Principles of BOLD functional MRI. In C., M. and Bandettini, P., editors, *Functional MRI*, pages 103–113. Springer.

Ciuciu, P. and Idier, J. (2002). A half-quadratic block-coordinate descent method for spectral estimation. *Signal Processing*, 82(7):941–959.

Ciuciu, P. and Idier, J. (2004). Regularized Doppler radar imaging for target identification in atmospheric clutter. In *IEEE International Conference Acoustics, Speech, and Signal (ICASSP)*, pages 265–268, Montreal, Canada.

Ciuciu, P., Idier, J., and Giovannelli, J.-F. (2001). Regularized estimation of mixed specta using a circular Gibbs-Markov model. *IEEE Transactions on Signal Processing*, 49(10):2202–2213.

Ciuciu, P., Idier, J., Roche, A., and Pallier, C. (2004). Outlier detection for robust region-based estimation of the hemodynamic response function in event-related fMRI. In *IEEE International Symposium on Biomedical Imaging*, pages 392–395, Arlington, VA USA.

Ciuciu, P., Poline, J., Marrelec, G., Idier, J., Pallier, C., and Benali, H. (2003). Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Transactions on Medical Imaging*, 22(10):1235–1251.

Cohen, A., Daubechies, I., and Feauveau, J.-C. (1992). Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5):485–560.

Coifman, R. and Donoho, D. (1995). Translation-invariant de-noising. In *Wavelets and Statistics*, Springer Lecture Notes in Statistics 103, pages 125–150. Springer, New York, NY, USA.

Coifman, R. R. and Wickerhauser, W. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718.

Combettes, P. L., Dung, D., and Vu, B. C. (2010). Dualization of signal recovery problems. *Set-Valued and Variational Analysis*. To appear.

Combettes, P. L. and J.-C. Pesquet (2007). Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Optim.*, 18:1351–1376.

Combettes, P. L. and Pesquet, J.-C. (2007). A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(1):564–574.

Combettes, P. L. and Pesquet, J.-C. (2008). A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6):27.

Combettes, P. L. and Pesquet, J.-C. (2010). Proximal splitting methods in signal processing. In Bauschke, H. H., Burachik, R., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, chapter 1. New York.

Combettes, P. L. and Wajs, V. R. (2005). Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4:1168–1200.

Coulon, O., Alexanderb, D. C., and Arridgeb, S. (2004). Diffusion tensor magnetic resonance image regularization. *Medical Image Analysis*, 8(1):47–67.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, Philadelphia.

Daubechies, I., Defrise, M., and DeMol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457.

Daubechies, I. and Teschke, G. (2005). Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising. *Applied and Computational Harmonic Analysis*, 19(1):1–16.

De Queiroz, R. L., Nguyen, T. Q., and Rao, K. R. (1996). The GenLOT: generalized linear-phase lapped orthogonal transform. *IEEE Transactions on Signal Processing*, 40:497–507.

De Zwart, J. A., Van Geldern, P., Kellman, P., and Duny, J. H. (2002). Reduction of gradient acoustic noise in MRI using SENSE-EPI. *Neuroimage*, 16(4):1151–1155.

Demoment, G. and Idier, J. (2001). Régularisation d'un problème mal-posé : approches génériques. In Idier, J., editor, *Approche bayésienne pour les problèmes inverses*, chapter 2, pages 41–58. Hermès - Lavoisier, Paris.

Demoment, G. and Idier, J. (2008a). Inverse Problems, Ill-posed Problems. In Idier, J., editor, *Bayesian Approach to Inverse Problems*, chapter 1, pages 25–40. Wiley-ISTE.

Demoment, G. and Idier, J. (2008b). Main approaches to the regularization of ill-posed problems. In Idier, J., editor, *Bayesian Approach to Inverse Problems*, chapter 2, pages 41–58. Wiley-ISTE.

Do, M. N. and Vetterli, M. (2002). Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 11(2):146–158.

Do, M. N. and Vetterli, M. (2005). The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12):2091–2106.

Dobigeon, N., Hero, A. O., and Tourneret, J.-Y. (2009). Hierarchical Bayesian sparse image reconstruction with application to MRFM. *IEEE Transactions on Image Processing*, 19(9):2059–2070.

Dobigeon, N. and Tourneret, J.-Y. (2007). Truncated multivariate Gaussian distribution on a simplex. Technical report, University of Toulouse.

Donoho, D. L. (1995). Denoising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

Ekeland, I. and Témam, R. (1999). *Convex analysis and variational problems*. Society for Industrial and Applied Mathematics, Philadelphia, USA.

Elad, M., Matalon, B., and Zibulevsky, M. (2007a). Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization. *Applied and Computational Harmonic Analysis*, 23(3):346–367.

Elad, M., Milanfar, P., and Ron, R. (2007b). Analysis versus synthesis in signal priors. *Inverse Problems*, 23:947–968.

Engl, H. W. (1987). Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Journal of Optimization Theory and Applications*, 52(2):209–215.

Févotte, C. and Godsill, S. J. (2006). Sparse linear regression in unions of bases via Bayesian variable selection. *IEEE Signal Processing Letters*, 13(7):441–444.

Filler, A. G., Winn, H. R., Howe, F. A., Griffiths, J. R., Bell, B. A., and Deacon, T. W. (1991). Axonal transport of superparamagnetic metal oxide particles: Potential for magnetic resonance assessments of axoplasmic flow in clinical neurosciece. In *International Society for Magnetic Resonance in Medicine*, San Francisco, USA. 1p.

Flandin, G. (2004). *Utilisation d'informations géométriques pour l'analyse statistique des données d'IRM fonctionnelle*. PhD thesis, Université de Nice-Sophia Antipolis.

Flandin, G. and Penny, W. D. (2007). Bayesian fMRI data analysis with sparse spatial basis function priors. *Neuroimage*, 34(3):1108–1125.

Friston, K. J., Holmes, A. P., Poline, J.-B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., and Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage*, 2(1):45–53.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. (1994). Statistical parametric map in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210.

Friston, K. J., Josephs, O., Zarahn, E., Holmes, A. P., Rouquette, S., and Poline, J.-B. (2000). To smooth or not to smooth? bias and efficiency in fMRI time series analysis. *Neuroimage*, 12(2):196–208.

Gauthier, J., Duval, L., and Pesquet, J.-C. (2009). Optimization of synthesis oversampled complex filter banks. *IEEE Transactions on Signal Processing*, 57(10):3827–3843.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Gelman, N. and Wood, L. M. (1996). Wavelet encoding for 3D gradient-echo MR imaging. *Magnetic Resonance in Medicine*, 36(4):613–619.

Geman, D. and Reynolds, G. (1992). Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):367–383.

Geman, D. and Yang, C. (1995). Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4(7):932–946.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Genovese, C. R. (2000). A Bayesian time-course model for functional magnetic resonance imaging data. *J. Amer. Statist. Assoc.*, 95:691–719.

Gerardin, E. (2006). Artéfacts en imagerie par résonance magnétique. `www.chu-rouen.fr/radiologie/artefacts.ppt`.

Giovannelli, J.-F. (2008). Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function. *IEEE Transactions on Image Processing*, 17(1):16–23.

Golub, C. and Van Loan, C. F. (1996). *Matrix Computations*. Johns Hopkins University Press.

Goutte, C., Nielsen, F., and Hansen, L. K. (2000). Modeling the haemodynamic response in fMRI using smooth FIR filters. *IEEE Transactions on Medical Imaging*, 19(12):1188–1201.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2007). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems*, pages 585–592, Red Hook, USA.

Griesbaum, A., Kaltenbacher, B., and Vexler, B. (2008). Efficient computation of the Tikhonov regularization parameter by goal-oriented adaptive discretization. *Inverse Problems*, 24(2). 20 pp.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., and Haase, A. (2002). Generalized autocalibrating partially parallel acquisitions GRAPPA. *Magnetic Resonance in Medicine*, 47(6):1202–1210.

Guerquin-Kern, M., Van De Ville, D., Vonesch, C., Baritaux, J.-C., Pruessmann, K. P., and Unser, M. (2009). Wavelet regularized reconstruction for rapid MRI. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 193–196, Boston, USA.

Guerrero-Colon, J., Mancera, L., and Portilla, J. (2008). Image restoration using space-variant gaussian scale mixtures in overcomplete pyramids. *IEEE Transactions on Image Processing*, 17(1):27–41.

Gupta, A. K. and Song, D. (1997). Lp-norm uniform distribution. *J. Statist. Plann. Inference*, 125(2):595–601.

Haar, A. (1910). Zur Theorie der Orthogonalen Funktionen Systeme. *Mathematische Annalen*, 69:331–371.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52.

Harikumar, G. and Bresler, Y. (1996). A new algorithm for computing sparse solutions to linear inverse problems. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1331–1334, Atlanta, USA.

Hastings, W. K. (1970). Monte Carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109.

Heidemann, R. M., Griswold, M. A., Haase, A., and Peter, M. J. (2001). VD AUTO-SMASH imaging. *Magnetic Resonance in Medicine*, 45(6):1066–1074.

Heurta, G. (2005). Multivariate Bayes wavelet shrinkage and applications. *J. of Applied Statistics*, 32(5):529–542.

Hoge, W. S., Brooks, D. H., Madore, B., and Kyriakos, W. E. (2005). A tour of accelerated parallel MR imaging from a linear systems perspective. *Concepts in Magnetic Resonance*, 27A(1):17–37.

Holmes, A. P. (1994). *Statistical issues in functional brain mapping.* PhD thesis, University of Glasgow.

Ichir, M. and Mohammad-Djafari, A. (2003). Wavelet domain blind image separation. In *SPIE Technical Conference on Wavelet Applications in Signal and Image Processing X*, San Diego, USA. 10p.

Jakob, P. M., Griswold, M. A., Edelman, R. R., and Sodickson, D. K. (1998). AUTO-SMASH: a self-calibrating technique for SMASH imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 7(1):42–54.

Jalobeanu, A., Blanc-Féraud, L., and Zerubia, J. (2002). Hyperparameter estimation for satellite image restoration using a MCMC Maximum Likelihood method. *Pattern Recognition*, 35(2):341–352.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. of the Royal Society of London. Series A*, 186(1007):453–461.

Joshi, R. L., Jafarkhani, H., Kasner, J. H., Fischer, T. R., Farvardin, N., Marcellin, M. W., and Bamberger, R. H. (1997). Comparison of different methods of classification in subband coding of images. *IEEE Transactions on Image Processing*, 6(11):1473–1486.

Kawahara, H. (1993). Signal reconstruction from modified auditory wavelet transform. *IEEE Transactions on Signal Processing*, 41(12):3549–3554.

Keeling, S. L. (2003). Total variation based convex filters for medical imaging. *Applied Mathematics and Computation*, 139(1):101–119.

Khalidov, I., Van De Ville, D., Fadili, J., and Unser, M. (2007). Activelets and sparsity: a new way to detect brain activation from fMRI data. In *SPIE, the International Society for Optical Engineering*, San Diego, USA. 8p.

Kherif, F., Poline, J.-B., Mériaux, S., Benali, H., Flandin, G., and Brett, M. (2004). Group analysis in functional neuroimaging. *Neuroimage*, 20(4):2197–2208.

Kowalski, M. and Torrésani, B. (2008). Random models for sparse signals expansion on unions of bases with application to audio signals. *IEEE Transactions on Signal Processing*, 56(8):3468–3481.

Kyriakos, W. E., Panych, L. P., Kacher, D. F., Westin, C. F., Bao, S. M., Mulkern, R. V., and Jolesz, F. A. (2000). Sensitivity profiles from an array of coils for encoding and reconstruction in parallel (SPACE RIP). *Magnetic Resonance in Medicine*, 44(2):301–308.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for Longitidunal data. *Biometrics*, 38(4):963–974.

Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191.

Le Pennec, E. and Mallat, S. (2005). Sparse geometric image representations with bandelets. *IEEE Transactions on Image Processing*, 14(4):423–438.

Leporini, D. and Pesquet, J.-C. (2001). Bayesian wavelet denoising: Besov priors and non-Gaussian noises. *Signal Processing*, 81(1):55–67.

Li, H. and Adali, T. (2008). Complex-valued adaptove signal processing using nonlinear functions. *Journal on Advances in Signal Processing, Special issue on Emerging Machine Learning Techniques in Signal Processing*, 2008. 9p.

Liang, D., Liu, B., Wang, J.-J., and Ying, L. (2009). Accelerating SENSE using compressed sensing. *Magnetic Resonance in Medicine*, 62(6):1574–1584.

Liang, D., Liu, B., and Ying, L. (2008). Accelerating sensitivity encoding using compressed sensing. In *IEEE Engineering in Medicine and Biology*, pages 1667–1670, Vancouver, Canada.

Liang, Z. P., Bammer, R., Ji, J., Pelc, N. J., and Glover, G. H. (2002). Making better SENSE: wavelet denoising, Tikhonov regularization, and total least squares. In *International Society for Magnetic Resonance in Medicine*, page 2388, Hawaï, USA.

Lin, F.-H., Chen, Y.-J., Belliveau, J. W., and Wald, L. L. (2003). A wavelet-based approximation of surface coil sensitivity profiles for correction of image intensity inhomogeneity and parallel imaging reconstruction. *Human Brain Mapping*, 19(2):96–111.

Lin, F.-H., Huang, T.-Y., Chen, N. K., Wang, F.-N., Stufflebeam, S. M., Belliveau, J. W., Wald, L. L., and Kwong, K. K. (2005). Functional MRI using regularized parallel imaging acquisition. *Magnetic Resonance in Medicine*, 54(2):343–353.

Lin, F.-H., Kwong, K. K., Belliveau, J. W., and Wald, L. L. (2004). Parallel imaging reconstruction using automatic regularization. *Magnetic Resonance in Medicine*, 51(3):559–567.

Liu, B., Abdelsalam, E., Sheng, J., and Ying, L. (2008). Improved spiral SENSE reconstruction using a multiscale wavelet model. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1505–1508, Paris, France.

Liu, B., King, K., Steckner, M., Xie, J., Sheng, J., and Ying, L. (2009). Regularized sensitivity encoding (SENSE) reconstruction using Bregman iterations. *Magnetic Resonance in Medicine*, 61(1):145–152.

Liu, B., Ying, L., Steckner, M., Jun, X., and Sheng, J. (2007). Regularized SENSE reconstruction using iteratively refined total variation method. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 121–124.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and oeltermann oeltermann oeltermann oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150–157.

Louchet, C. and Moisan, L. (2008). Total variation denoising using posterior expectation. In *European Signal Processing Conference*, Lausanne, Switzerland. 5p.

Louchet, C. and Moisan, L. (2010). Total variation denoising using posterior expectation. *hal-00258849*. `http://hal.archives-ouvertes.fr/hal-00258849/fr/`.

Lustig, M., Donoho, D., and Pauly, J. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195.

Makni, S. (2006). *Détection-estimation conjointe de l'activité cérébrale en imagerie par résonance magnétique fonctionnelle*. PhD thesis, Université de Paris Sud.

Makni, S., Ciuciu, P., Idier, J., and Poline, J.-B. (2005). Joint detection-estimation of brain activity in functional MRI: a multichannel deconvolution solution. *IEEE Transactions on Signal Processing*, 53(9):3488–3502.

Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., and Ciuciu, P. (2008). A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3):941–969.

Mallat, S. (1998). *A wavelet tour of signal processing*. Academic Press, San Diego, CA, USA, 2nd edition.

Mallat, S. (2009). Geometrical grouplets. *Appl. Comput. Harm. Anal.*, 26(2):161–180.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Mansfield, P. (1997). Multi-planar image formation using NMR spin echos. *Journal of Physics C*, 10(3):55–58.

Marrelec, G., Ciuciu, P., Pélégrini-Issac, M., and Benali, H. (2004). Estimation of the hemodynamic response function in event-related functional MRI: direct acyclic graphs for general Bayesian inference framework. *IEEE Transactions on Medical Imaging*, 23(8):959–967.

Martin, M. B. and Bell, A. E. (2001). New image compression techniques using multiwavelets and multiwavelet packets. *IEEE Transactions on Image Processing*, 10(4):500–510.

McKoewn, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., and Sejnowsky, T. J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6(3):160–188.

Mériaux, S., Roche, A., Dehaene-Lambertz, G., Thirion, B., and Poline, J.-B. (2006). Combined permutation test and mixed-effect model for group average analysis in fMRI. *Human Brain Mapping*, 27(5):402–410.

Meuleneire, M. D. (2008). Algebraic quantization of transform coefficients for embedded audio coding. In *IEEE International Conference on Acoustics, Speech, and Signal (ICASSP)*, pages 4789–4792, Las Vegas, USA.

Meyer, F. G. (2003). Wavelet-Based Estimation of a Semiparametric Generalized Linear Model of fMRI Time-Series. *IEEE Transactions on Medical Imaging*, 22(3):315–322.

Miller, E. L. (1999). Efficient computational methods for wavelet domain signal restoration problems. *IEEE Transactions on Signal Processing*, 47(2):1184–1188.

Mojsilovic, A., Popovic, M., Markovic, S., and Krstic, M. (1998). Characterization of visually similar diffuse diseases from B-scan liver images using nonseparable wavelet transform. *IEEE Transactions on Medical Imaging*, 17(4):541–549.

Molla, S. and Torrésani, B. (2005). A hybrid scheme for encoding audio signal using hidden Markov models of waveforms. *Applied and Computational Harmonic Analysis*, 18(2):137–166.

Moreau, J. J. (1962). Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci.*, 255:2897–2899.

Moreau, J.-J. (1965). Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299.

Moulin, P. and Liu, J. (1998). Analysis of multiresolution image denoising schemes using generalized-Gaussian priors. In *IEEE International Symposium on Time-Frequency Time-Scale Analysis*, pages 633–636, Pittsburgh, USA.

Müller, P. and Vidakovic, B. (1999). MCMC methods in wavelet shrinkage: non-equally spaced regression, density and spectral density estimation. In Müller, P. and Vidakovic, B., editors, *Bayesian Inference in Wavelet-Based Models*, pages 187–202.

Nashed, M. Z. and Scherzer, O. (1997). Stable approximation of nondifferentiable optimization problems with variational inequalities. *J. Amer. Math. Soc.*, 204:155–170.

Nesterov, Y. (2004). *Introductory Lectures on Convex Optimization: A Basic Course.* Kluwer Academic Publishers, New York.

Nichols, T. E. and Holmes, A. P. (2003). Controlling the Familywise Error Rate in Functional Neuroimaging. *Human Brain Mapping*, 15(5):1–25.

Noll, C., Cohen, J. D., Meyer, C. H., and Schneider, W. (1995). Spiral k-space MR imaging of cortical activation. *Magnetic Resonance in Medicine*, 5(1):49–56.

Novey, M., Adali, T., and Roy, A. (2010). A complex Generalized Gaussian distribution - Characterization, generation, and estimation. *IEEE Transactions on Signal Processing*, 58(3):1427–1433.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87:9868–9872.

Operto, G. (2009). *Analyse structurelle surfacique de données fonctionnelles cérébrales.* PhD thesis, Université Paul Cézanne.

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

Padfield, D. R. and Manjeshwar, R. (2006). Adaptive conductance filtering for spatially varying noise in PET images. *Progress in biomedical optics and imaging*, 7(30):61446N.

Pan, X. (1998). A unified reconstruction theory for diffraction tomography, with consideration of noise control. *J. Opt. Soc. Amer.*, 15(9):2312–2326.

Penfield, W. and Rasmussen, T. (1952). The cerebral cortex of man. *Fleming Journal of Mental Science*, 98:188–189.

Penny, W. D., Kiebel, S., and Friston, K. J. (2003). Variational Bayesian inference for fMRI time series. *Neuroimage*, 19(3):727–741.

Penny, W. D., Trujillo-Barreto, N., and Friston, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *Neuroimage*, 23(2):350–362.

Perlbarg, V., Marrelec, G., Doyon, J., Pélégrini-Issac, M., Lehéricy, S., and Benali, H. (2008). NEDICA: Detection of group functional networks in fMRI using spatial independent component analysis. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1247–1250, Paris, France.

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

Petersson, K. M., Nichols, T. E., Poline, J. B., and Holmes, A. P. (1999). Statistical limitations in functional neuroimaging. I. Non-inferential methods and statistical models. *Philosophical transactions-Royal Society of London. Biological sciences*, 354(1387):1239–1260.

Phillips, D. L. (1962). A technique for the numerical solution of certain integral equation of the first kind. *J. Ass. Comput. Mach.*, 9:84–97.

Pinel, P., Thirion, B., Mériaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J.-B., and Dehaene, S. (2007). Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci.*, 8(1):91.

Pizurica, A., Wink, A. M., Vansteenkiste, E., Philips, W., and Roerdink, J. (2006). A review of wavelet denoising in MRI and ultrasound brain imaging. *Current Medical Imaging Reviews*, 2(2):247–260.

Pruessmann, K. P., Weiger, M., Börnert, P., and Boesiger, P. (2001). Advances in sensitivity encoding with arbitrary k-space trajectories. *Magnetic Resonance in Medicine*, 46(4):638–651.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. (1999a). SENSE : sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. (1999b). SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962.

Pustelnik, N., Chaux, C., and Pesquet, J.-C. (2010). Parallel proximal algorithm for image restoration using hybrid regularization. *arXiv:0911.1536v1*. Submitted.

Pustelnik, N., Chaux, C., Pesquet, J.-C., Sureau, F. C., Dusch, E., and Comtat, C. (2009). Adapted convex optimization algorithm for wavelet-based dynamic PET reconstruction. In *Fully 3D*, Bejing, China. 4p.

Rabrait, C. (2007). *Imagerie par Résonance Magnétique à haute résolution temporelle: développement d'une méthode d'acquisition parallèle tridimensionnelle pour l'imagerie fonctionnelle cérébrale.* PhD thesis, Université de Paris Sud.

Rabrait, C., Ciuciu, P., Ribès, A., Poupon, C., Leroux, P., Lebon, V., Dehaene-Lambertz, G., Bihan, D. L., and Lethimonnier, F. (2008). High temporal resolution functional MRI using parallel echo volume imaging. *Magnetic Resonance Imaging*, 27(4):744–753.

Raj, A., Singh, G., Zabih, R., Kressler, B., Wang, Y., Schuff, N., and Weiner, M. (2007). Bayesian parallel imaging with edge-preserving priors. *Magnetic Resonance in Medicine*, 57(1):8–21.

Rangan, S. and Goyal, V. K. (2001). Recursive consistent estimation with bounded noise. *IEEE Transactions on Information Theory*, 47(1):457–464.

Ribes, A. and Schmitt, F. (2008). Linear inverse problems in imaging. *IEEE Signal Processing Magazine*, 25(2):84–99.

Robert, C. and Castella, G. (2004). *Monte Carlo statistical methods*. Springer, New York.

Roche, A., Mériaux, S., Keller, M., and Thirion, B. (2007). Mixed-effect statistics for group analysis in fMRI: a nonparametric maximum likelihood approach. *Neuroimage*, 38(3):501–510.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press.

Roy, C. S. and Sherrington, C. S. (1890). On the regulation of the blood supply of the brain. *Journal of Physiology*, 11:85–108.

Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithm. *Physica*, 60(1-4):259–268.

Ruttimann, U. E., Unser, M., Rawlings, R. R., Rio, D., Ramsey, N. F., Mattay, V. S., Hommer, D. W., Frank, J. A., and Weinberger, D. R. (1998). Statistical analysis of functional MRI data in the wavelet domain. *IEEE Transactions on Medical Imaging*, 17(2):142–154.

Schoenberg, S. O., Dietrich, O., and Reiser, M. F. (2007). *Parallel imaging in clinical MR applications*. Springer, New York.

Seeger, M., Gerwinn, S., and Bethge, M. (2007). Bayesian inference for sparse generalized linear models. *Machine Learning*, 4701:298–309.

Seeger, M. W. and Nickisch, H. (2008). Compressed sensing and Bayesian experimental design. In *International Conference on Machine Learning*, pages 912–919, Helsinki, Finland.

Selesnick, I. W., Baraniuk, R. G., and Kingsbury, N. G. (2005). The dual-tree complex wavelet transform. 22(6):123–151.

Sendur, L. and Selesnick, I. W. (2002). Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transactions on Signal Processing*, 50(11):2744–2756.

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.

Sijbers, J., Poot, D., Dekker, A. J. D., and Pintjens, W. (2007). Automatic estimation of the noise variance from the histogram of a magnetic resonance image. *Physics in Medicine and Biology*, 52(5):1335–1348.

Simoncelli, E. P. and Adelson, E. H. (1996). Noise removal via Bayesian wavelet coring. In *IEEE International Conference on Image Processing (ICIP)*, pages 379–382, Lausanne, Switzerland.

Smith, S. M. (2004). Overview of fMRI analysis. *British Journal of Radiology*, 77(2):167–175.

Sodickson, D. K. (2000). Tailored SMASH image reconstructions for robust in vivo parallel mr imaging. *Magnetic Resonance in Medicine*, 44(2):243–251.

Sodickson, D. K. and Manning, W. J. (1999). Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603.

Steffen, P., Heller, P. N., Gopinath, R. A., and Burrus, C. S. (1993). Theory of regular m-band wavelet bases. *IEEE Transactions on Signal Processing*, 41(12):3497–3511.

Sümbül, U., Santos, J. M., and Pauly, J. M. (2009). Improved time series reconstruction for dynamic magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 28(7):1093–1104.

Thirion, B. and Faugeras, O. (2003). Dynamical components analysis for fMRI data through kernel PCA. *Neuroimage*, 20(1):34–49.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Stat.*, 22(4):1701–1762.

Tikhonov, A. (1963). Tikhonov regularization of incorrectly posed problems. *Soviet Mathematics Doklady*, 4:1624–1627.

Tropp, J. A. (2006). Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051.

Twomey, S. (1963). On the numerical solution of fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature. *J. ACM*, 10:97–101.

Van De Ville, D., Blu, T., and Unser, M. (2004). Integrated wavelet processing and spatial statistical testing of fMRI data. *Neuroimage*, 23(4):1472–1485.

Van De Ville, D., Blu, T., and Unser, M. (2006). Surfing the brain - An overview of wavelet-based techniques for fMRI data analysis. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):65–78.

Van De Ville, D., Seghier, M., Lazeyras, F., Blu, T., and Unser, M. (2007). WSPM: Wavelet-based statistical parametric mapping. *Neuroimage*, 37(4):1205–1217.

Van Der Zwaag, W., Francis, S., and Bowtell, R. (2006). Improved echo volumar imaging (EVI) for functional MRI. *Magnetic Resonance in Medicine*, 56(6):1320–1327.

Varoquaux, G., Sadaghiani, S., Pinel, P., Kleinschmidt, A., Poline, J.-B., and Thirion, B. (2010). A group model for stable multi-subject ICA on fMRI datasets. *Neuroimage*, PMID:20153834.

Varoquaux, G., Sadaghiani, S., Poline, J.-B., and Thirion, B. (209). Canica: Model-based extraction of reproducible group-level ICA patterns from fMRI time series. *hal-00435262*.

Vincent, T. (2010). *Modèles hémodynamiques spatiaux adaptatifs pour l'imagerie cérébrale fonctionnelle*. PhD thesis, Univ. Paris-Sud.

Vincent, T., Risser, L., and Ciuciu, P. (2010). Spatially adaptive mixture modeling for analysis of within-subject fMRI time series. *IEEE Transactions on Medical Imaging*, 29(4):1059–1074.

Vonesch, C. and Unser, M. (2008). A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolutions. *IEEE Transactions on Image Processing*, 17(4):539–549.

Vonesch, C. and Unser, M. (2009). A fast multilevel algorithm for wavelet-regularized image restoration. *IEEE Transactions on Image Processing*, 18(3):509–523.

Wang, J., Nittka, T., Jellus, V., Kuhn, B., and Kiefer, B. (2001). Parallel acquisition techniques with modified SENSE reconstruction (mSENSE). In *First Wurzburg Workshop on Parallel Imaging Basics and Clinical Applications*, page 92, Wurzburg.

Wang, Y. and Haomin, Z. (2006). Total variation wavelet-based medical image denoising. *International Journal of Biomedical Imaging*. 6 pp.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Watson, A. B., Yang, G. Y., Solomon, J. A., and Villasenor, J. (1997). Visibility of wavelet quantization noise. *IEEE Transactions on Image Processing*, 6(8):1164–1175.

Weaver, J. B., Yansun, X., Healy, D. M., and Cromwell, L. D. (1991). Filtering noise from images with wavelet transforms. *Magnetic Resonance in Medicine*, 21(2):288–295.

Weibin, L. and Mingyi, H. (2009). A novel adaptive noise filtering method for SAR images. In *Proceedings of SPIE, the International Society for Optical Engineering*, San Diego, USA.

Weiger, M., Pruessmann, K. P., and Boesiger, P. (2000). Cardiac real-time imaging using SENSE. *Magnetic Resonance in Medicine*, 43(2):177–184.

Weiger, M., Pruessmann, K. P., Lessler, C., Roschmann, P., and Boeisiger, P. (2001). Specific coil design for SENSE: a six-element cardiac array. *Magnetic Resonance in Medicine*, 45(3):495–504.

Weiger, M., Pruessmann, K. P., Osterbauer, R., Bornert, P., Boesiger, P., and Jezzard, P. (2002). Sensitivity-encoded single-shot spiral imaging for reduced susceptibility artifacts in BOLD fMRI. *Magnetic Resonance in Medicine*, 48(5):860–866.

Weiss, P., Blanc-Féraud, L., and Aubert, G. (2008). Efficient schemes for total variation minimization under constraints in image processing. *SIAM journal on Scientific Computing*, 31:2047–2080.

Wendt, M., Busch, M., Lenz, G., Duerk, J. L., Lewin, J. S., Seibel, R., and Grnemeyer, D. (1998). Dynamic tracking in interventional MRI using wavelet-encoded gradient-echo sequences. *IEEE Transactions on Medical Imaging*, 17(5):803–809.

Wolfe, P. J. and Godsill, S. J. (2003). Bayesian estimation of time-frequency coefficients for audio signal enhancement. *Advances in Neural Information Processing Systems 15*, pages 1197–1204.

Woolrich, M., Ripley, B., Brady, M., and Smith, S. (2001). Temporal autocorrelation in univariate linear modelling of fMRI data. *Neuroimage*, 14(6):1370–1386.

Worsley, K. J. and Friston, K. J. (1995). Analysis of fMRI time-series revisited - again. *Neuroimage*, 2(3):173–181.

Yan, Y., Wen, H., Mattay, V. S., Balaban, R., Frank, J. A., and Duyn, J. H. (1999). Compression of 3D BOLD functional MRI with spiral acquisition at 1.5 T and 4.0 T. *Neuroimage*, 9(4):446–451.

Ying, L., Liu, B., Steckner, M., Wu, G., Wu, M., and Li, S.-J. (2008). A statistical approach to SENSE reconstruction with arbitrary trajectories. *Magnetic Resonance in Medicine*, 60(2):414–421.

Ying, L. and Sheng, J. (2007). Joint image reconstruction and sensitivity estimation in SENSE (JSENSE). *Magnetic Resonance in Medicine*, 57(6):1196–1202.

Ying, L., Xu, D., and Liang, Z.-P. (2004). On Tikhonov regularization for image reconstruction in parallel MRI. In *IEEE Engineering in Medicine and Biology Society*, pages 1056–1059, San Francisco, USA.

Zeger, S. and Karim, R. (1991). Generalized linear models with random effects a Gibbs sampling approach. *J. Amer. Statist. Assoc.*, 86(413):79–86.

Zhang, Y., Hero, A. O., and Rogers, W. L. (1992). A bounded error estimation approach to image reconstruction. In *Nuclear Science Symposium and Medical Imaging Conference*, pages 966–968, Orlando, USA.

Zhu, Y., Hardy, C. J., Sodickson, D. K., Giaquinto, R. O., Dumoulin, C. L., Kenwood, G., Niendorf, T., Lejay, H., McKenzie, C. A., Ohliger, M. A., and Rofsky, N. M. (2004). Highly parallel volumetric imaging with a 32-element RF coil array. *Magnetic Resonance in Medicine*, 52(4):869–877.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. of the Royal Statistical Society*, 67(4):301–320.

Zou, H. and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, 37(4):1733–1751.

Zou, Y. M., Ying, L., and Liu, B. (2008). SparseSENSE: application of compressed sensing in parallel MRI. In *IEEE International Conference on Technology and Applications in Biomedicine*, pages 127–130, Shenzhen, China.

Zălinescu, C. (2002). *Convex analysis in general vector spaces.* World Scientific, NJ, USA.

Zymnis, A., Boyd, S., and Candes, E. (2009). Compressed sensing with quantized measurements. *IEEE Signal Processing Letters*, 17(2):149–152.