Proximal structured sparsity regularization for online reconstruction in high-resolution accelerated Magnetic Resonance imaging

Loubna El Gueddari

NNT : 2019SACLS573

Thèse de doctorat

Proximal structured sparsity regularization for online reconstruction in high-resolution accelerated Magnetic Resonance Imaging.

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°575 Physique et ingénierie: électrons, photons et sciences du vivant (EOBE)
Spécialité de doctorat : Physique et imagerie médical

Thèse présentée et soutenue à Gif-sur-Yvette, le 13 Décembre 2019, par

LOUBNA EL GUEDDARI

Composition du Jury :

Dr. Rachid Deriche INRIA, (ATHENA)	Président du jury
Prof. Carola-Bibiane Schönlieb Université de Cambridge, (DAMPT)	Rapporteur
Prof. Angshul Majumdar IIIT Delhi, (ECE)	Rapporteur
Dr. Lotfi Chaari INP Toulouse, (TCI-IRIT)	Examinateur
Prof. Jean-Christophe Pesquet Central Supélec, Centre de Vision Numérique	Examinateur
Dr. Nicolas Boulant CEA NeuroSpin, Université Paris Saclay	Directeur de thèse
Dr. Philippe Ciuciu CEA NeuroSpin, Université Paris Saclay	Co-directeur de thèse

ipse se nihil scire id unum sciat
— Plato, 399 B.C.

Acknowledgements

The work presented in this thesis was only possible with the support and the help of many people. I would like to take the opportunity to express my gratitude.

First I would like to express the deepest appreciation to my supervisor Dr. **Philippe Ciuciu** for his continuous support during these three years. I really appreciated his patience, the trust and the freedom he gave me to conduct my work.

I would also like to express my gratitude to Dr. **Alexandre Vignaud** for his kind and regular advice. I am also really grateful to Dr. **Nicolas Boulant**, Pr. **Jean-Christophe Pesquet** and Dr. **Emilie Chouzenoux** for your advice during our insightful meetings. Thank you for your time, for sharing your knowledge with me.

I deeply thank my COSMIC colleagues especially Dr. **Jean-Luc Starck**, Dr. **Samuel Farrens**, **Zac Ramzi** and last but not least Dr. **Antoine Grigis** for sharing their knowledge and their vision on reproducible science. I am thankful to Dr. **Franck Mauconduit** for his help on the Gadgetron project. Thank you **Aurelien Massire**, **Chai GR**, **Mathilde Rippart** and **Carole Lazarus** for your help on the MANIAC project. In addition, I also wanted to thank the jury for the interest they have shown in my work, and the time they have spent reviewing and examining the manuscript.

In addition to the people directly involved in my work, I would like to thank my fellows at NeuroSpin, who contributed making the time spent here memorable. To **Caroline Le Ster**, I could never thank you enough for spending those periods of doubts with me, you were always here and ready to give me some precious advice, I hope that our friendship will last for years. To **Arnaud Gloaguen** thank you for your help and for all the useful conversations we had on optimization and proximity operators. To **Bruno Pinho Menses** thank you for always laughing at my jokes and for the great conversations on *picanha*. To my former colleagues **Lisa Leroi**, **Raphaël Tomi-Tricot**, **Emilie Poirion** you were models. Not forgetting my formers interns **Hamza Cherkaoui**, **Hanaé Carrié**, hoping you will enjoy your research life in Singapore and **Benoit Sarthou**, I have learned more than I was expecting. I also wanted to thank the *aperitivo* team: **Nicolas Guigui**, **Arnaud Gloagen** and **Cathy Philippe**, Venice wouldn't have been the same without you. To **Léonie Borne**, **Slim Karkar** and **Arnaud Gloaguen** you were the perfect

running mates.

I also want to thank all my Parietal colleagues, it has been a great pleasure to work with you, you were a true inspiration to me. To **Olivier Grisel**, thank you for your kindness and for the great conversations we had, it has been a pleasure having you as professor for Deep Learning classes.

Finally I also want to thank my relatives and friends for supporting me during the PhD thesis and in my life in general. To **Maha**, **Kenza** and **Zineb** thank you for always being here for your kindness and support. To **Salma**, **Oussama** and **Narjiss** thank you for your presence during the defense, it meant a lot to me. To my parents, my brother and my sister this thesis is the result of the many sacrifices you have made and the guidance with which you have provided me for all these years. A very warm thank you to my father-in-law **Michel** for always supporting me and being the perfect father in law. And of course thank you **Pierre**, I could never find words to express how grateful I am to have you in my life.

Contents

List of Figures

List of Tables

Symbols Used

Symbol	Constant name	Value
c	Speed of light in a vacuum	$299\,792\,458\,\mathrm{m}\cdot\mathrm{s}^{-1}$ (exact)
$\hbar$	Planck constant (reduced)	$1.055\times10^{-34}\,\mathrm{J}\cdot\mathrm{s}$
k_B	Boltzmann constant	$1.381\times10^{-23}\,\mathrm{J}\cdot\mathrm{K}^{-1}$
γ_{H}	^{1}H gyromagnetic ratio	$267.5\times10^{6}\,\mathrm{rad}\cdot\mathrm{s}^{-1}\cdot\mathrm{T}^{-1}$
$\frac{1}{2\pi}\gamma_{\mathrm{H}}$	^{1}H gyromagnetic ratio (reduced)	$42.57\times10^{6}\,\mathrm{Hz}\cdot\mathrm{T}^{-1}$

Symbol	Description	Unit
B_0	static main field	$\mathrm{kg}\cdot\mathrm{s}^{-2}\cdot\mathrm{A}^{-1}$ (T)
B_1^+	coil transmission field	T
B_1^-	coil reception field	T
BW	radiofrequency bandwidth	s^{-1} (Hz)
E	electric field	$\mathrm{V}\cdot\mathrm{m}^{-1}$
$\mathcal{E}$	energy	$\mathrm{kg}\cdot\mathrm{m}^2\cdot\mathrm{s}^{-2}$ (J)
f	temporal frequency	s^{-1} (Hz)
f_0	Larmor temporal frequency	s^{-1} (Hz)
G	spatial static field gradients	$\mathrm{T}\cdot\mathrm{m}^{-1}$
k	spatial frequency	m^{-1}
M	macroscopic magnetisation	$\mathrm{A}\cdot\mathrm{m}^{-1}$
M_o	macroscopic equilibrium magnetisation	$\mathrm{A}\cdot\mathrm{m}^{-1}$
S	spatial static field gradients slew rate	$\mathrm{T}\cdot\mathrm{m}^{-1}\cdot\mathrm{s}^{-1}$
T	temperature	K
T_1	spin-lattice relaxation time	s
T_2	spin-spin relaxation time	s
T_2^*	effective spin-spin relaxation time	s
γ	gyromagnetic ratio	$\mathrm{rad}\cdot\mathrm{s}^{-1}\cdot\mathrm{T}^{-1}$
$\Delta\mathrm{B}_0$	static field inhomogeneity	T
Δf_0	offset from carrier frequency	s^{-1} (Hz)

$\Delta\omega$	offset from carrier angular frequency	$\mathrm{rad} \cdot \mathrm{s}^{-1}$
ε_r	relative permittivity	—
ω_0	Larmor angular frequency	$\mathrm{rad} \cdot \mathrm{s}^{-1}$
σ	electrical conductivity	$\mathrm{kg}^{-1} \cdot \mathrm{m}^{-2} \cdot \mathrm{s}^3 \cdot \mathrm{A}^2$ ($\mathrm{S} \cdot \mathrm{m}^{-1}$)
φ	radiofrequency phase	degree (°), rad
χ	magnetic susceptibility	—

General mathematical notations

The mathematical notations, in this manuscripts, follows general typographical conventions.

Symbol	Description
$\mathbb{N}$	The set of Natural number.
$\mathbb{R}$	The set of Real number.
$\mathbb{C}$	The set of Complex number.
$\imath$	The imaginary unit is expressed, such that $\imath^2 = -1$.
$\Re(z)$	The real parts of a complex number z.
$\Im(z)$	The imaginary parts of a complex number z.
$\boldsymbol{v}$	Lower-case bold letters denote vectors.
$\boldsymbol{M}$	Upper-case bold letters denote matrices.
$(\cdot)^*$	Complex conjugation is conveyed by an exponent asterisk.
$(\cdot)^\top$	Vector or matrix transpose is represented by an exponent $\top$ following the object, *e.g.* $\boldsymbol{v}^\top$.
$(\cdot)^{\mathrm{H}}$	Hermitian transpose, is denoted by an exponent following the object, *e.g.* $\boldsymbol{M}^{\mathrm{H}} = (\boldsymbol{M}^*)^\top$.
$\times$	Cross product.
$\langle \cdot, \cdot \rangle$	Scalar product defined as follows: $\forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{C}^{n \times n} \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}$
$\|\boldsymbol{M}\|_2$	The Frobenius norm of a matrix defined as $\|\boldsymbol{M}\|_2 = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \|m_{i,j}\|^2}$.
$\|\|\|\boldsymbol{M}\|\|\|$	The spectral norm of a matrix defined as $\|\|\|\boldsymbol{M}\|\|\| = \max_{\|\boldsymbol{x}\|_2 \neq 0} \frac{\|\boldsymbol{M}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$.
$\Gamma_0(\mathbb{C}^p)$	The set of convex, proper, lower semi-continuous functions on $\mathbb{C}^p \to \mathbb{R} \cup +\infty$.

Abbreviations and Acronyms

2D	two-dimensional
3D	three-dimensional
AFI	actual flip angle imaging
CAIPIRINHA	Controlled Aliasing In Parallel Imaging Results In Higher Acceleration
CEA	Commissariat à l'énergie atomique et aux énergies alternatives
CLEAR	Calibration-free locally Low-rank EncourAging Reconstruction
CNR	contrast-to-noise ratio
COSMIC	Compressed Sensing for Magnetic Resonance Imaging and Cosmology
CP	circular polarisation
CS	Compressed Sensing
DRF	Fundamental Research Division
DWI	diffusion-weighted imaging
EPI	echo-planar imaging
ESPIRiT	Eigenvalue Self-consistent Parallel Imaging Reconstruction using iTerative algorithm
FA	flip angle
FB	Forward-Backward
FID	free induction decay
FISTA	Fast Iterative Soft-Thresholding Algorithm
fMRI	functional MRI
FOV	Field Of View

GPGPU	general-purpose computing on graphics processing units
GRAPPA	generalised autocalibrating partially parallel acquisition
GRASP	golden-angle radial sparse parallel MRI
GRE	gradient-recalled echo
GROG	GRAPPA Gridding Operator
ISTA	Iterative Soft-Thresholding Algorithm
LASSO	Least Absolute Shrinkage and Selection Operator
LORAKS	Low-rank matrix modeling of local k-space neighborhoods
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
NRMSE	Normalised Root-Mean-Square Error
OP	out of phase (complete phase opposition)
OSCAR	Octagonal Shrinkage and Clustering Algorithm for Regression
PDw	proton-density-weighted
PF	partial Fourier
p-MRI	parallel-MRI
POGM	Proximal Optimized Gradient Method
PRUNO	Parallel Reconstruction Using Null Operations
pTx	parallel transmission
PySAP	Python Sparse Data Analysis Package
RF	radiofrequency
ROI	region of interest
SAG	sagittal plane
SAKE	Simultaneous Autocalibrating and K-space Estimation
SAR	specific absorption rate
SENSE	SENsitivity Encoding
SNR	Signal-to-Noise Ratio
SOS	square-root of the Sum of Square
Sparse-BLIP	Sparse BLind Iterative Parallel imaging reconstruction
STD	sample standard deviation

sTx	single-channel transmission
T_1w	T_1-weighted
T_2w	T_2-weighted
T_2^*w	T_2^*-weighted
TA	acquisition time
TE	echo time
TGV	total generalized variation
TR	repetition time
TV	total variation
UHF	ultra-high field
VDS	Variable Density Sampling
VIBE	volumetric interpolated breath-hold examination
WT	wavelet transform

General Introduction

Context & Motivations

MAGNETIC Resonance Imaging (MRI) is the key imaging modality for exploring the most fascinating organ in the human body, namely *the brain*. With a huge diversity of contrasts, MRI can leverage multiple information to ease medical diagnosis. Besides diagnosis, longitudinal studies take advantage of MR non-invasiveness to understand the brain or predict the outcome of diseases using appropriate biomarkers. In order to improve accuracy in the diagnosis and help uncover structural abnormalities at an early stage, high resolution MRI could be beneficial at the expense of scan time. Long scans (> 15 min) are more sensitive to subject's motion which impedes the resolution and increases the number of discarded exams. Although dedicated acquisition strategies combined with motion correction algorithms may reduce artifacts, the scan time remains too long to fit within the 20 min dedicated to the whole clinical MRI exam, during which, multiple acquisitions are performed to collect complementary information on tissues. Consequently, specific acceleration and encoding strategies (*e.g.* parallel imaging techniques, navigator echoes, propeller scheme, Partial Fourier, etc.) have emerged over the last two decades to limit scan duration and gain in robustness to motion. Yet clinical imaging remains limited to millimetric resolution.

During the last decade, the Compressed Sensing theory has made a major breakthrough to accelerate high-resolution scans. This theory combines efficient acquisition schemes and reconstruction algorithm. Recently, the MR community has seen the emergence of new efficient sampling patterns, first using Cartesian straight lines drawn pseudo-randomly, then with simple non-Cartesian sampling such as spiral or radial spokes that allow to reach higher acceleration factors compared to Cartesian acquisition. Even more recently, a new mathematically driven and hardware-adaptive sampling framework, called SPARKLING, has been proposed. SPARKLING stands for Spreading Projection Algorithm for Rapid K-space samplING, it is the outcome of 6 years of collaborative research between mathematicians (P. Weiss and J. Kahn, CNRS/ITAV Toulouse), physicists (A. Vignaud and F. Mauconduit, CEA/NeuroSpin), a researcher in signal processing for neuroimaging (P. Ciuciu, CEA/NeuroSpin) and former PhD students (C. Lazarus and N. Chauffert,

CEA/NeuroSpin). The collaboration led to an efficient optimization algorithm which takes into account MR hardware constraints specific to the magnetic field gradients and a target sampling distribution. The outcome of this algorithm is a set of sampling trajectories whose temporal derivatives, namely the gradient profiles, can then be injected into a MR pulse sequence. SPARKLING has been proved to drastically accelerate scan time at 7 Tesla by a factor of 20 in 2D and 60 for 3D acquisitions with a sub-millimetric isotropic resolution. Despite the tremendous acceleration provided by the latter trajectory, SPARKLING acquisition are unsuitable for clinical routines. The very reason is that several issues related to **image reconstruction** are still pending in order to deliver reliable MR images shortly after the end of acquisition.

This PhD thesis has been actually dedicated to bring the SPARKLING acquisition schemes to the clinical routine. The primary goal of this PhD thesis was to develop an online reconstruction framework in order to provide MR images by the end of acquisition. To do so, we need first to accelerate reconstruction algorithms but also to interleave image acquisition and reconstruction in order to feed the reconstruction part with new data. The main advantage of the proposed approach is to provide some feedback to the physician on the MR image quality during data acquisition. In this thesis, we will illustrate all the developments on T_2^* imaging at 7 Tesla but the corresponding algorithmic contributions will be soon available on the 3 Tesla scanner at NeuroSpin as SPARKLING has been deployed on this MR system as well.

Contributions

The fist goal of this thesis was to improve the reconstruction accuracy in the high-resolution highly accelerated setup – *i.e.* using multi-channel phased array acquisition and non-Cartesian encoding schemes. The latter means that the data are no longer collected over a grid or lattice of samples but instead that the measurements lie on the continuous Fourier domain, also called k-space in MRI.

While various reliable reconstruction strategies have been successfully developed for Cartesian acquisitions using different model-based approaches (*e.g.* parallel imaging, regularization for inverse problems in imaging, etc.) and different kinds of prior knowledge (*e.g.* sparsity, low-rank structure), non-Cartesian reconstruction is often addressed using a gridding step, which impedes the final image quality. Based on an inverse problem formulation, we take into account non-Cartesian multi-channel data with and propose a dedicated reconstruction algorithm that leverages *structured sparsity* information across the multiple channels. We also get rid of the complicated and painful calibration step of the sensitivity maps associated with the multiple channels, which makes our CS reconstruction algorithm a single-step approach. This approach remains in the convex optimization setting which guarantees convergence to a stable solution. Our results demonstrate an enhanced image quality achieved in a more time-efficient manner.

While the scan time has been drastically reduced using non-Cartesian trajectories, the application of compressed sensing to daily-routine exams remains limited owing to the

long iterative reconstruction process. Therefore the second objective of this thesis was to propose a method to speed up this reconstruction. Taking into account the segmented acquisition properties of k-space data, a novel online framework with interleaved data acquisition and image reconstruction has been proposed. Mathematically, we used a mini-batch formulation to adapt existing optimization algorithms to this online reconstruction setting. The online reconstruction allows one to give partial feedbacks during the scan time making Compressed Sensing and high-resolution MRI available in clinical routines.

A third significant contribution of this thesis was software development as all pieces of work have been implemented in an open source package called `PySAP` (Python Sparse data Analaysis Package), which is developed at CEA between the NeuroSpin and CosmoStat teams in the context of the DRF Impulsion COSMIC project led by my main supervisor, Philippe Ciuciu. On top of that, we interfaced the algorithmic developments done in `PySAP` with the client/slave Gadgetron interface system in order to push the online reconstruction to a remote workstation. Asynchronous communication between the MR system and the computer dedicated to image reconstruction was thus implemented.

Thesis Outline

This thesis is organized as follows:

Chapter 1: *Background in Magnetic Resonance Imaging*

In the first chapter, the principles of MR physics that will be needed in the rest of the manuscript will be summarized. This allows to make the document self-contained. In particular, we discuss the principles of signal detection, sampling and image formation using various encoding schemes as they are at the heart of accelerated acquisition strategies. As the NMR signal is collected on the k-space – *i.e.* Fourier conjugate space – we briefly describe tools for image reconstruction when the samples do not fall into a Cartesian grid. Furthermore, we present specific reception coils that first allow to increase the signal-to-noise ratio (SNR) and second empower scan acceleration with dedicated acquisition and reconstruction methods, *i.e. parallel-imaging*. However, the accelerations achieved by parallel imaging are limited and the goal is to go beyond this acceleration. Compressed Sensing theory is the right framework to drastically shorten scan time and will be naturally discussed in the next chapter.

Chapter 2: *Compressed Sensing in MRI*

In this chapter, we briefly introduce Compressed Sensing theory and its application to MRI. Therefore, its three main ingredients are discussed in the following order: first, the way to collect k-space data as *incoherent* as possible using variable density sampling, first along Cartesian lines and then using more sophisticated non-Cartesian trajectories. Second, CS theory relies on the sparsity of the sought object (*i.e.* MR image) in an appropriate domain. Although MR images are not sparse in the image domain, they are at least compressible in a transform domain– *i.e.* they can be efficiently represented or approximated by a few non-zeros coefficients in this domain. We review different classes of

sparsifying transforms, whether they are orthogonal/complete, overcomplete embedding redundant information, fixed or data driven. Third, we review the different classes of optimization problems associated with different assumptions on sparsity (analysis vs synthesis prior) and present the efficient optimization algorithms. Last but not least, as this thesis is dedicated to image reconstruction, we will present and choose a suitable image quality metric to assess the performances of the algorithms proposed along this manuscript.

Chapter 3: *Multi-channel MRI reconstruction*

Achieving high-resolution MR imaging requires high-input SNR, which is often obtained using specific phased array coil to collect the data. Phased array coils are composed of small elementary coils that are combined together to boost the SNR. However this combination at the reconstruction stage is non-trivial, therefore this chapter will cover the way image reconstruction is performed in the setting of multi-channel array coil acquisition. In particular, we present a first contribution to this field, namely a simple yet efficient two-step approach for coil calibration and image reconstruction in the SENSitivity Encoding (SENSE) framework. Also, we will pay attention to the compatibility of the different algorithms to on-line reconstruction constraints, which are at the heart of this PhD thesis.

Chapter 4: *Calibrationless multi-channel reconstruction*

Due to the lack of compatibility of two-step SENSE-based approaches with online constraints and the limited applicability of alternative methods (*e.g.* LORAKS) to non-Cartesian sampling schemes, this chapter will explore a different track, called *calibrationless* reconstruction methods. To avoid the calibration step, these approaches aim to reconstruct an image per channel. As multiple images of the same object, *i.e.* organ, are reconstructed, they share information therefore and in order to handle the redundant information shared by the coil images, structured sparsity prior knowledge is imposed. In this field, one key contribution of this thesis is the application of new convex regularization schemes based on mixed norms (*e.g.* OSCAR, k-support norm) that promote structured sparsity in a more adapted way than conventional group-LASSO (which requires the explicit knowledge of the group-structure). The set of MR images is thus given by the minimization of a global nonsmooth objective function and primal-dual algorithm is used to perform the optimization. The validation – on magnitude ad phase images – is conducted on prospectively highly accelerated T_2^* 2D MR imaging from ex-vivo human and baboon brain with in-plane resolution of 400μm.

Chapter 5: *Online MR image reconstruction*

Finally the *online reconstruction* framework is introduced in this final contributed chapter. A general problem proposes a data splitting into mini-batches scheme, and develops the optimization. Then, we adopt a pedagogical approach to first address the simpler case of single-channel data acquisition. In this case, a single image is reconstructed, either from Cartesian and non-Cartesian data. In particular, we explore the optimal way for tuning the batch size in order to deliver a reliable solution by the end of acquisition. Last, we extend the online approach to multi-channel acquisition by considering the calibration-less method we described in Chapter 4. For given computational power, we show how the cost of the proximity operator associated with structured sparsity may prevent us from fulfilling the online reconstruction constraints. Therefore we introduced some *online trick* in the reconstruction process to remain compatible with these online constraints. Finally, we validate the proposed method on the same data set as in Chapter 4.

Part I

Context

Chapter 1

Background in Magnetic Resonance Imaging

Chapter Outline

Magnetic resonance imaging (MRI) is a powerful technique to non-invasively probe the human body with a vast variety of contrasts. As the underlying phenomenon of magnetic resonance may be thought of rather complex, the goal of the first part of this chapter is to give a brief overview of the basic physical principles underlying MRI[1]. Importantly, information contained in the pixels of a given MR image is not directly probed. Instead, the measurements are collected in the k-space, which in its simplest and ideal (*i.e.* artifact-free) form corresponds to the spatial frequency domain or Fourier space. In Section 1.2, we will thus review the spatial encoding principle of MR images in k-space. In a second step we recall the principle behind the recording of the magnetic signal. As these two processes of spatial encoding and data recording are at the heart of the complex nature of MR images, a particular attention will be paid to the demodulation. Next, as MRI is hampered by the long scan time, accelerating techniques allowing to shorten MRI

[1]Further details about MR physics can be found in [Haacke 1999, Bernstein 2004, McRobbie 2006].

exams have been developed for the last two decades. It has actually been a major field of research with outcomes nowadays available in clinical routine. Based on the use of multi-channel receiver coils, the parallel imaging method and the acceleration it provides will be described in the last section.

1.1 Nuclear Magnetic Resonance phenomenon

Magnetic resonance imaging principles are based on Nuclear Magnetic Resonance (NMR) phenomenon, that was first described in 1938 by Isidor Isaac Rabi [Rabi 1938], work for which he won the Nobel Pize in 1944. The applications to molecular structure exploration quickly followed with the independent works of Felix Bloch [Bloch 1946] and Edward Mills Purcell [Purcell 1946], for which they were jointly awarded the Physics Nobel Prize in 1952. The principles described by these scientists are briefly explained in Sections 1.1.1 and 1.1.2.

It is only three decades later that the works of Paul Lauterbur [Lauterbur 1973] and Sir Peter Mansfield [Mansfield 1977a] allowed to encode the Nuclear Magnetic Resonance (NMR) signal spatially, thus enabling NMR imaging, or MRI. For this, they shared the Nobel Prize in Medicine and Physiology in 2003. Spatial encoding is the subject of Section 1.2.

The NMR phenomenon can be observed in any atom that has a non-zero *spin*, a quantum-mechanical property conveying an intrinsic form of angular momentum. Atoms composed of an odd number of nucleons all have a non-zero spin. This is the case for the naturally abundant isotopes of hydrogen, fluorine, sodium and phosphorus (^{1}H, ^{19}F, ^{23}Na and ^{31}P, respectively) as well as for rarer isotopes of carbon (^{13}C), sodium (^{15}N) or oxygen (^{17}0).

As human body tissues are mostly made of water and fat compartments, they are extremely rich in hydrogen atoms ^{1}H, whose nucleus consists of a unique proton with a $\frac{1}{2}$-spin. For this reason, most clinical applications of MRI focus on proton imaging. Non-proton – also called "exotic" or "X"- – nuclei spectroscopy and imaging is a wide and rich area of research, and their imaging is a dynamic and emerging field with a promising future for clinical research [Coste 2017], especially at ultra-high magnetic field (7 Tesla). Indeed, their lower abundance in the human body associated with a low signal-to-noise ratio (SNR) may be compensated by a strong static magnetic field. However, their analysis remains beyond the scope of this thesis, which focuses solely on ^{1}H imaging.

1.1.1 Magnetization

The fundamental concept behind nuclear magnetic resonance imaging is the interaction of a proton spin with an external magnetic field $\mathbf{B}_0$. Nuclei with a non-zero spin can be characterized by their magnetic moment, and behave like microscopic magnets. When the nuclei is exposed to a static magnetic field $\mathbf{B}_0$, its moment aligns with the static field, either pointing in the same direction (called *parallel*) or in the opposite direction (called

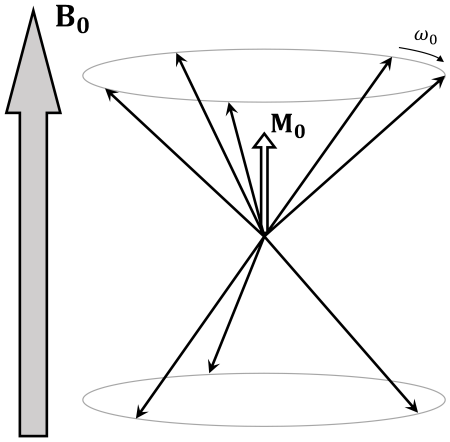


Figure 1.1-1: Classical representation of macroscopic net magnetization $\mathbf{M_0}$ (hollow vector) in an isochromat consisting of several adjacent spins immersed in an external static magnetic field ($\mathbf{B_0}$, grey vector). It is induced by the slight excess of magnetic moments (black vectors) parallel to $\mathbf{B_0}$, compared to anti-parallel ones. ω_0 is the Larmor angular frequency at which each spin precesses around $\mathbf{B_0}$. From [Tomi-Tricot 2018, ch. 1].

anti-parallel), and start spinning like a gyroscope. This gyroscopic movement is called *precession*. The angular frequency of this gyroscopic spinning, also referred as angular precession frequency or Larmor frequency, is described by the Larmor equation given by:

$$\omega_0 = \gamma \mathbf{B}_0 \tag{1.1}$$

with γ being a nuclei-specific constant called gyromagnetic ratio. In the case of proton ^1H, the constant is equal to $\gamma_{\mathrm{H}} = 267.5 \times 10^6\,\mathrm{rad/s/T}$ which corresponds to $\frac{1}{2\pi}\gamma_{\mathrm{H}} = 42.57\,\mathrm{MHz/T}$. At the thermal equilibrium, the spins are distributed over the two existing parallel and anti-parallel states corresponding respectively to low and high-level energy, the proportion of which being given by the Boltzmann statistics. The latter depends on the Boltzmann constant, the temperature, the gyromagnetic ratio and the field strength. Due to thermal energy the difference between spin populations at low- and high-energy level is extremely small. This difference, called the *spin excess*, is about 1 over 100 000 at 1.5 T. Even though the spin excess is a million time smaller than the number of protons, the NMR experiments consider a mesoscopic scale called *isochromat*, which consists of million of billions of protons. In average this isochromat creates a mesoscopic magnetic moment called net magnetization $\mathbf{M}$ also denoted at the equilibrium in a static field $\mathbf{B_0}$ by $\mathbf{M_0}$. This is particularly convenient as this net magnetization operates under the laws of classical physics.

At thermal equilibrium, the distribution of low and high-energy spins within the isochromat follows a Maxwell-Boltzmann law. If we denote by N the number of spins considered, T the temperature, $\hbar$ the reduced Planck constant and k_B the Boltzmann constant (see 'Symbols Used' section on page xiii), and if we consider that the energy difference is negligible compared to k_BT, then the macroscopic magnetization at equilibrium pictured

in Figure 1.1-1 can be written as:

$$\mathbf{M_0} \approx N \frac{\gamma^2 \hbar^2}{4 k_B T} \mathbf{B_0} \tag{1.2}$$

During NMR experiments the observable variable is the net magnetization $\mathbf{M}_0$. Interestingly, Eq. (1.2) shows that $\mathbf{M}_0$ depends on the strength of $\mathbf{B_0}$. This constitutes the major reason for the development of ultra-high field MR systems as the stronger $\mathbf{B_0}$, the larger the magnetization and so the potential NMR signal.

1.1.2 Excitation and Relaxation

Even in a macroscopic body, spin excess does not guarantee a detectable signal, the magnetization vector must be tipped away from the static field direction $\mathbf{B}_0$ in order to set it into precession. This step is performed by applying a Radio Frequency (RF) magnetic field, $\mathbf{B_1^+}$ generated by specific RF coils, for a short period of time also referred as pulse, the latter step is called "Excitation". In order to interact with the spins the generated $\mathbf{B_1^+}$ magnetic field must oscillate at the Larmor frequency. On top of this frequency, the generated field is orthogonal to $\mathbf{B_0}$ which creates a transverse component of the net magnetization as depicted in Figure 1.1-2. The magnetization vector can then be decomposed into two components, namely the transverse and longitudinal magnetizations, denoted respectively by $\mathbf{M_{xy}}$ and $\mathbf{M}_z$ and defined as follows:

$$\mathbf{M} = \mathbf{M}_{xy} + \mathbf{M}_z \quad \text{with:} \quad \mathbf{M_{xy}} = \begin{bmatrix} M_x \\ M_y \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{M_z} = \begin{bmatrix} 0 \\ 0 \\ M_z \end{bmatrix} \tag{1.3}$$

Once the magnetization has been tilted, the $\mathbf{B_1^+}$ field is stopped and the system starts returning to its initial equilibrium state. This phenomena is called "Relaxation", the rate for which the system returns to its equilibrium state defines the relaxation time. On one hand, the longitudinal rate is characterized by the time needed for the $\mathbf{M}_z$ component to retrieve 63% of its initial value, this time is called T_1. Since the T_1 is related to the spin interactions with their surrounding, it is therefore referred to as the *spin-lattice* relaxation. On the other hand, the transverse decay is characterized by the time T_2 needed for $\mathbf{M}_{xy}$ to decay to 37% of its initial value. The phenomena emanates from the loss of phase coherence between spins and is therefore referred to as the *spin-spin* relaxation (see Figure 1.1-3). In practice, the $\mathbf{B_0}$ field is not uniform and the transverse decay is faster than expected, hence the accessible time is not directly the T_2 but rather T_2^* that is linked through the following relation:

$$\frac{1}{\mathrm{T}_2^*} = \frac{1}{\mathrm{T}_2} + \frac{1}{\mathrm{T}_{2i}} \tag{1.4}$$

where $1/\mathrm{T}_{2i} = \gamma \Delta \boldsymbol{B}_i$ is the relaxation rate contribution attributable to field inhomogeneities ($\Delta \boldsymbol{B}_i$) across a voxel. Both T_1 and T_2 are intrinsic tissue properties, however only T_1 depends on the $\mathbf{B}_0$ field strength.

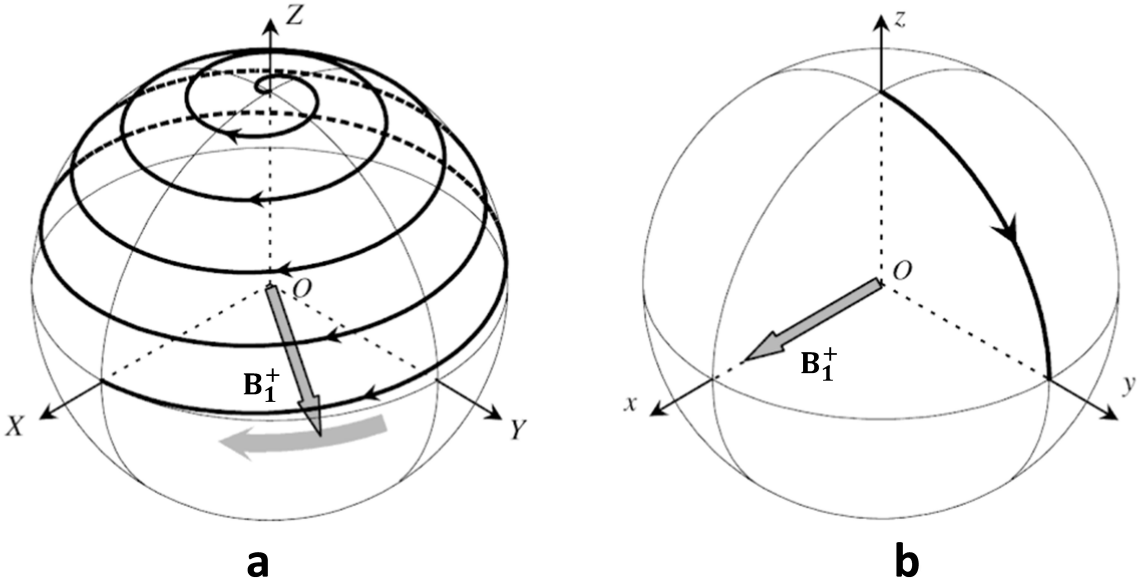


Figure 1.1-2: Effect on macroscopic magnetization of a $\mathbf{B_1^+}$ field rotating at the Larmor frequency $\omega_\mathbf{0}$ viewed (a) from laboratory frame $(\mathbf{XYZ})$ where $\mathbf{B_0}$ is aligned with the $\mathbf{Z}$ axis. (b) It is generally convenient to represent it in a rotating frame of reference $(\mathbf{xyz})$ of frequency $\omega_\mathbf{rf}$, where $\mathbf{B_1^+}$ lies along with the $\mathbf{x}$ axis. The $\mathbf{z}$ axis is aligned with $\mathbf{Z}$. From [Tomi-Tricot 2018].

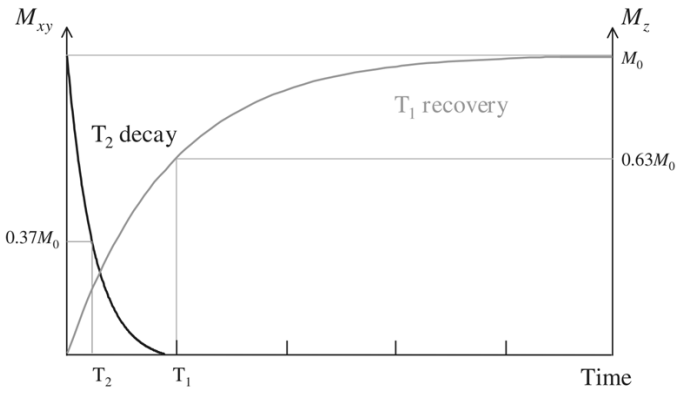


Figure 1.1-3: Longitudinal (grey) and transverse (black) relaxation of magnetization for a given $\mathbf{T_1}$ and $\mathbf{T_2}$. From [McRobbie 2006].

While magnetization relaxes to its equilibrium state, it is possible to measure the evolution of its transverse component, using either the same coil that was used for transmission (Tx) of $\mathbf{B_1^+}$, or a dedicated receive coil (Rx). Indeed, the rotation of $\mathbf{M_{xy}}$ in the transverse plane induces a measurable voltage in the coil [Hoult 1997]. The recorded magnetic field – also called $\mathbf{B_1^-}$ – generates a signal, called free induction decay (FID) [Hahn 1950]; as shown in Figure 1.1-4, it oscillates at ω_0, with an exponential attenuation corresponding to T_2^*. Note that the amplitude of FID depends on the initial longitudinal magnetization $\mathbf{M}_0$ that is tipped into the transverse plane: it is therefore related to Eq. (1.2).

The behavior of magnetization exposed to a magnetic field $\mathbf{B}$ was formalized by Felix

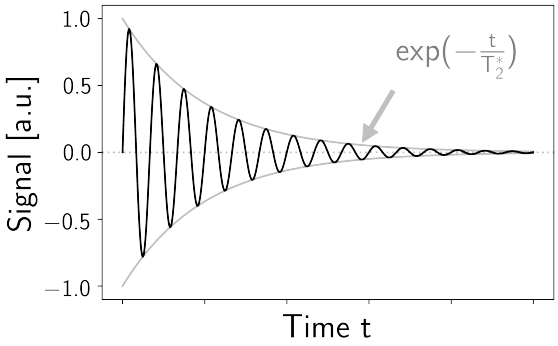


Figure 1.1-4: Free induction decay signal representation, showing the signal that oscillates at the carrier frequency (128 MHz at 3 T), weighted by a decreasing exponential caused by $\mathbf{T_2^*}$ decay. No values are shown on the time axis, as in this simple representation, for the sake of readability, the frequency is orders of magnitude lower than in reality.

Bloch [Bloch 1946] and written in the laboratory frame (xyz) as:

$$\left(\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t}\right)_{\mathrm{lab}} = \gamma\mathbf{M}\times\mathbf{B} - \frac{1}{\mathrm{T}_2}\mathbf{M_{xy}} - \frac{1}{\mathrm{T}_1}(\mathbf{M_z} - M_0) \tag{1.5}$$

As pictured in Figure 1.1-2, it is convenient to describe the evolution of magnetization in a frame (xyz) rotating according to the rotation vector $\mathbf{\Omega}$:

$$\mathbf{\Omega} = \begin{bmatrix} 0 & 0 & \omega \end{bmatrix}^{\mathrm{T}} \tag{1.6}$$

Equation (1.5) then becomes:

$$\left(\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t}\right)_{\mathrm{rot}} = \gamma\mathbf{M}\times\left(\mathbf{B} - \frac{\mathbf{\Omega}}{\gamma}\right) - \begin{bmatrix} M_x/\mathrm{T}_2 \\ M_y/\mathrm{T}_2 \\ (M_z - M_0)/\mathrm{T}_1 \end{bmatrix} \tag{1.7}$$

For instance, Figure 1.1-2b is obtained, ignoring relaxation, with $\mathbf{B} = \mathbf{B_0} + \mathbf{B_1^+}$ and $\omega = \omega_0 = \omega_{\mathrm{rf}}$: in ($xyz$), $\mathbf{B_0}$ disappears from Eq. (1.7), and $\mathbf{B_1^+}$ is fixed.

While the Bloch equations describe the spins dynamics for a large voxel in a homogeneous sample, they do not describe how signals generated from different voxels are discriminated. Hence, the following section will explain how these signals can be distinguished using spatial encoding.

1.2 Image formation

Spatial encoding in MRI is not based on a direct measurement of the pixels values. The reason for that is rather simple, the maximum achievable spatial resolution is governed by the Rayleigh criterion. In MRI, the wavelength is in the order of dozens of centimeters making the direct encoding of the body infeasible in practice. It is only in 1973 that the works of Paul Lauterbur [Lauterbur 1973] and Sir Peter Mansfield [Mansfield 1977a] allowed to encode the NMR signal spatially in a clever way. The goal of this section is to provide tools to understand the concepts underlying spatial encoding.

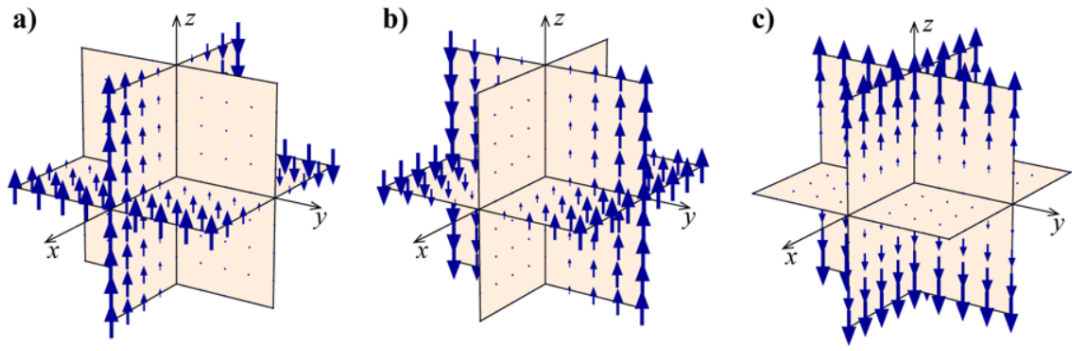


Figure 1.2-5: Diagrams of the B_z component of the 3D magnetic field variation of a) G_x-, b) G_y-, and c) G_z-gradients. The field is aligned with the direction of the arrows. The scale of the arrows indicates the magnitude of the magnetic field strength at that point. From [Poole 2007].

1.2.1 Spatial Encoding

The signals in MR imaging arise from multiple voxels recorded all at once. These voxels contain different materials with different spin densities and relaxation times. In order to discriminate spatial location of the different voxels Lauterbur and Mansfield [Mansfield 1977b] have proposed to change spatially the magnetic field, making the Larmor frequency a function of space. The local changes of the magnetic field is due to the use of three gradients (G_x, ,G_y ,G_z, one for each spatial dimension as shown in Figure 1.2-5). For example, when G_x is applied, the magnetic field $B(x)$ will vary with respect to the position according to $B(x) = \mathrm{B}_0 + G_x x$. This variation causes the precession angular frequency to vary linearly in space as follows:

$$\omega(x) = \gamma(\mathrm{B}_0 + G_x x) \tag{1.8}$$

Since the precession frequency linearly depends on the position, therefore positive positions (with respect to the iso-center located at the origin) will get higher frequency than negative ones.

2D Imaging

In 2D MRI, the slice to be imaged is first selectively excited with a *slice-selecting* gradient G_{Slice} (usually in the z direction with G_z). The data corresponding to this slice is then encoded in terms of spatial frequencies along the two other dimensions. Two techniques are used together to fill the 2D k-space, line by line:

- Applying a frequency-encoding (or *readout*) gradient G_{Read} changes the precession frequency of isochromats while signal is acquired, thus allowing to discriminate their position in that direction. We collect as such a line in the acquisition space (also called k-space as defined in Section 1.2.3).
- For now on, we have only acquired one line at the center of k-space. To fill the whole k-space we need to encode spatial frequencies along the remaining dimension, therefore to cover multiple lines. We thus apply a third gradient, G_{Phase}, called

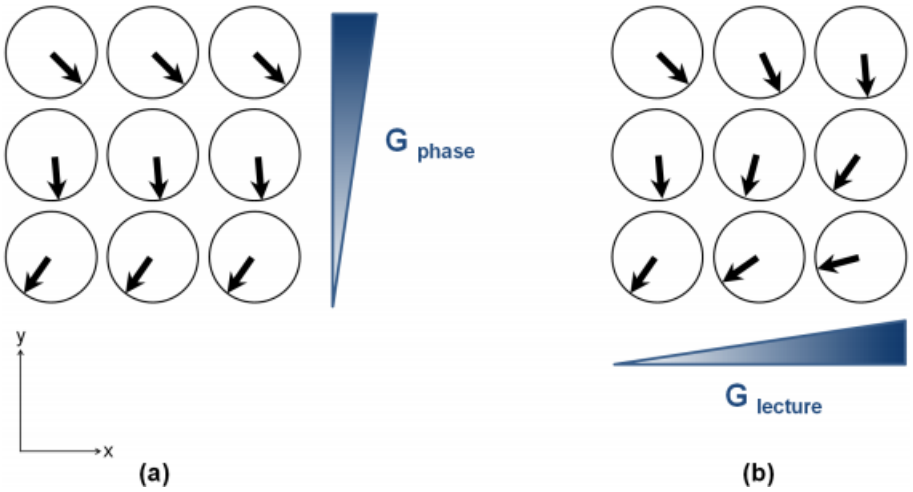


Figure 1.2-6: Encoding the k-space line by line. From [Le Ster 2017].

phase-encoding gradient, orthogonal to G_{Read}. It is applied between excitation and readout and induces a variation of phase along its direction, which will hold during the signal acquisition (readout) as described above.

To acquire a full 2D k-space, it is necessary to repeat the $G_{Phase} \rightarrow G_{Read}$ pattern as many times as there are lines in k-space, changing the intensity of G_{Phase} every time, so as to explore all needed spatial frequencies in this direction. The relation between a gradient $\mathbf{G}$ – over any combination of axes – applied for a certain duration t and the corresponding encoded spatial frequency $\mathbf{k}$ (in m^{-1}) is:

$$\mathbf{k}(t) = \gamma \int_0^t \mathbf{G}(\tau)\, d\tau \tag{1.9}$$

At the end of the 2D imaging process, we have encoded k-space data for one slice, and we are able to reconstruct the corresponding image (using a 2D transform). We will study the reconstruction process in detail in Section 1.2.3. In 2D imaging, one can repeat the process described above several times, selecting different slices, to acquire a three-dimensional Field Of View (FOV).

However, resolution is limited in the third dimension (slice thickness), because of three main factors. First, selecting a thinner slice requires higher intensity gradients, which eventually run into hardware limitations. Secondly, it requires increasingly sharp excitation profiles to avoid overlap between adjacent slices (at a lower slice resolution, we usually leave a gap between slices to prevent this overlap). Last but not least, when a slice is excited, the acquired signal comes from all isochromats embodied in it: a lower signal thus comes from a thinner slice resulting in a loss of image quality. Therefore to overcome those issues 3D imaging have to be considered.

3D Imaging

All the mentioned factors make high isotropic resolution hard – even impossible – to achieve with 2D acquisitions. Fortunately, it is possible to excite the whole FOV at once

and perform 3D imaging by acquiring a three-dimensional k-space. To do, so we apply the same steps as described earlier and we add a second phase-encoding gradient, G_{Part}, along which the slice – also called partition – direction is encoded by inducing phase variations through the *partition-encoding* gradient. The latter is generally played simultaneously with G_{Phase}. Hence , to acquire a full 3D k-space, we have to repeat the $G_{Part} \rightarrow G_{Phase} \rightarrow G_{Read}$ pattern $N_{\text{Part}} \times N_{\text{Phase}}$ times, with N_{Part} the number of partitions to be encoded and N_{Phase} the number of lines. As shown later on, the imaged volume will be recovered by applying an inverse 3D Fourier transform.

One drawback of 3D imaging is its sensitivity to patient motion, as the whole 3D k-space must be filled – which takes time – before reconstruction. Any spatial information corrupted by movement occurring during acquisition affects the quality of all slices within the volume. In comparison, in 2D imaging, each slice is reconstructed separately: motion during the acquisition of one slice has no effect on the rest of the FOV.

The next section, which explains how the NMR signal is detected and will clarify the origin of Fourier encoding.

1.2.2 Reception: Detection, Conversion & Demodulation

This step is at heart of MRI as it justifies the complex nature of MR images. In NMR experiments, we seek to measure the time-varying magnetization of the sample. For this purpose we rely on Faraday's law of electromagnetic induction, where a variation of the flux magnetic field creates an electromotive force which is converted to current in a reception coil[2].

The signal acquired by the coil is first amplified with a low-noise amplifier (LNA). If we assume that the scanned sample is made up of a single material, *i.e.* the resonance frequency is unique, and we apply a gradient varying such that the magnetic field varies linearly with the position, then the signal acquired by the coil can be expressed as follows:

$$s(t) \;\propto\; -\frac{\mathrm{d}}{\mathrm{d}t}\int_{sample} \mathbf{M}(\boldsymbol{r},t)\cdot\mathbf{B_1^-}(\boldsymbol{r}) \tag{1.10}$$

with $\mathbf{B_1^-}$ the magnetic field per unit current that would be produced by the coil at a point $\boldsymbol{r}$, the latter field is an example of the reciprocity principle, and $\mathbf{M}$ the magnetization produced by the spin precession and defined as follows [Haacke 1999]:

$$\mathbf{M}(\boldsymbol{r},t) = \begin{bmatrix} e^{-t/\mathrm{T_2}}\; M_\perp(\boldsymbol{r})\;\cos(\gamma \mathrm{B}_0 t + \varphi_0(\boldsymbol{r})) \\ e^{-t/\mathrm{T_2}}\; M_\perp(\boldsymbol{r})\;\sin(\gamma \mathrm{B}_0 t + \varphi_0(\boldsymbol{r})) \\ e^{-t/\mathrm{T_1}} M_z(\boldsymbol{r},0) + (1-e^{-t/\mathrm{T_1}})M_0 \end{bmatrix}, \tag{1.11}$$

with $\mathbf{M_{xy}}(\boldsymbol{r},0) = M_\perp(\boldsymbol{r})e^{-\imath\varphi_0(\boldsymbol{r})}$.

We recall that depending on the tissue γ can be either positive or negative and that the phase φ_0 and the magnitude $M_\perp$ are determined by the initial RF pulse conditions. Hence after interchanging the integral and time derivative operators and neglecting the

[2]While it is more common to have dedicated coils for transmission (Tx) (excitation) and reception (Rx), a single coil could ensure both roles, namely excitation and reception (Rx/Tx).

derivative of the e^{-t/T_1} and e^{-t/T_2} factors compared to that of the $e^{-\omega_0 t}$ factor, the signal gets proportional to:

$$s(t) \propto \omega_0 \int_{V_s} e^{-t/T_2(\boldsymbol{r})} \left|\mathbf{M_{xy}}(\boldsymbol{r})\right| \left|\mathbf{B_{xy}}(\boldsymbol{r})\right| \sin\left(\Phi_M(\boldsymbol{r},t) - \Phi_B(\boldsymbol{r})\right) d\boldsymbol{r} \tag{1.12}$$

where $\omega_0 = \gamma \mathrm{B}_0$ is the Larmor frequency in radians per second; $\mathbf{B_{xy}}(\boldsymbol{r}) = B_\perp(\boldsymbol{r}) e^{-\imath \Phi_B(\boldsymbol{r})}$ is the transverse component of the receive coil $\mathbf{B_1^-}$ field; ; $B_\perp(\boldsymbol{r})$ is the magnitude and $\Phi_B(\boldsymbol{r})$ the phase of $\mathbf{B_{xy}}$ in the laboratory frame $\Phi_M(\boldsymbol{r},t)$ is the phase of $\mathbf{M_{xy}}$, *i.e.* its angle with the x-axis.

The spectrum of the signal in Eq. (1.12) induced in the reception coil includes two frequency bands around ω_0. For proton imaging (*i.e.* ^{1}H), these frequencies are located around 128 and 298 MHz at 3 T and 7 T, respectively. As working with such high frequency level can be problematic, the first step in the reception chain consists in removing this fast oscillation – the *demodulation* step – by multiplying the signal with a cosinusoid and a sinusoid whose frequency is around the transmit one. Figure 1.2-7 illustrates the reception chain. After the low-noise amplification, the signal is split into two channels. In the first one, the signal is multiplied by an signal in-phase (cosinusoid) and in the second one by a signal in quadrature (sinusoid). The in-phase component leads to the real part of the MR signal while the in-quadrature component is considered as its imaginary part. Before being digitally converted the signal is low-pass filtered to get rid of any aliasing artifacts due to digital conversion. Then each channel is converted separately and recombined as real and imaginary components of k-space samples.

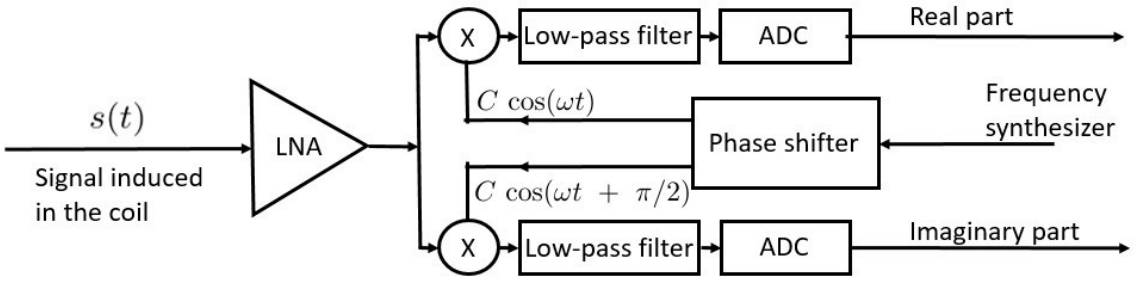


Figure 1.2-7: Block diagram of the reception chain. The signal from the coil is amplified by a low-noise amplifier before it gets split into two channels. Each channel will be multiplied by an in-phase and in quadrature signal generated by a frequency synthesizer, leading to a complex multiplication of the signal. The low-pass filter corresponds to an anti-aliasing step which prevents aliasing before the Analog-to-Digital conversion with a frequency being at least twice the maximal sampling frequency provided by the ADC. Adapted from [Décorps 2012].

1.2.3 Reconstruction

Fourier Transform

After the demodulation step the sampled signal $s^{\mathrm{dem}}(t)$ is recombined into a complex expression that can be read as follows:

$$s^{\text{dem}}(t) \propto \int_{sample} \rho(\boldsymbol{r}) e^{-\imath \boldsymbol{r}\cdot\gamma \int_0^t \boldsymbol{G}(\tau) d\tau} \, d\boldsymbol{r} \tag{1.13}$$

with ρ being the spin density proportional to ω_0, the transverse magnetization $\mathbf{M_{xy}}$ and the transverse receptive field $B_{1\perp}^{-}$. The integral $\gamma \int_0^t \boldsymbol{G}(\tau)\, d\tau$ is often denoted by:

$$\boldsymbol{k}(t) = \gamma \int_0^t \boldsymbol{G}(\tau)\, d\tau \tag{1.14}$$

Eq. (1.14) is nothing but the Fourier adjoint variable of the spatial location $\boldsymbol{r}$. By combining Eqs. (1.13) and (1.14), we get

$$s^{\text{dem}}(\boldsymbol{k}(t)) \propto \int_{sample} \rho(\boldsymbol{r}) e^{-\imath \boldsymbol{r}\cdot\boldsymbol{k}(t)} \, d\boldsymbol{r} = \mathcal{F}[\rho(\boldsymbol{r})] \tag{1.15}$$

which tells us that the *k-space* measurements[3] actually correspond to the spatial Fourier transform of the spin density $\rho(\boldsymbol{r})$. The acquisition space is thus called the *k-space*. Hence the gradients G_x, G_y and G_z are driving the frequency samples to be collected in the k-space. The simplest way to acquire an image in MRI is to design gradient such that the k-space samples are falling into a Cartesian grid (as depicted in Section 1.2.1). The image is then reconstructed by applying an inverse Fourier transform that can be computed using the Fast Fourier Transform (FFT) algorithm originally proposed by [Cooley 1965].

Non-Cartesian Fourier Transform

We have previously seen that the gradients are driving the k-space trajectory. Although it is more common to fill a Cartesian grid and then to apply an inverse Fourier Transform to recover the image, in some cases the k-space trajectory leaves the grid. In such scenario, the sampling is called non-Cartesian and there are several possibilities to deal with non-Cartesian acquisitions. First, the data can be interpolated and resampled on a Cartesian grid before applying a conventional inverse FFT. This technique is referred to as "gridding" [Pauly 2012], and it is a much faster than computing the Discrete Fourier Transform (DFT). Alternatively, one can deal with continuous measurements out of the grid using the Nonuniform or Nonequispaced FFT (NFFT) [Fessler 2003, Keiner 2009]. Hereafter, we present first two categories of gridding methods and then provide details about the NFFT.

Grid-driven. The idea behind grid-driven approach is to estimate the value at each grid point based on the neighboring data (*i.e.* k-space samples). Figure 1.2-8 illustrates the principle of grid-driven methods. Although they are easy to implement, they are suboptimal as they do not exploit the whole input data to perform interpolation, especially in heavily sampled k-space regions, loosing part of the available SNR in those area. To overcome this issue, the target Cartesian grid can be oversampled leading to a finer grid, and then the data is interpolated with a simple kernel such as a bilinear interpolator. Using the finer grid, grid-driven interpolation can yield high-fidelity reconstructions especially when the sampling density is varying or when the non-Cartesian samples do not fall far from the finer Cartesian grid.

[3] In an ideal artifact-free acquisition scenario.

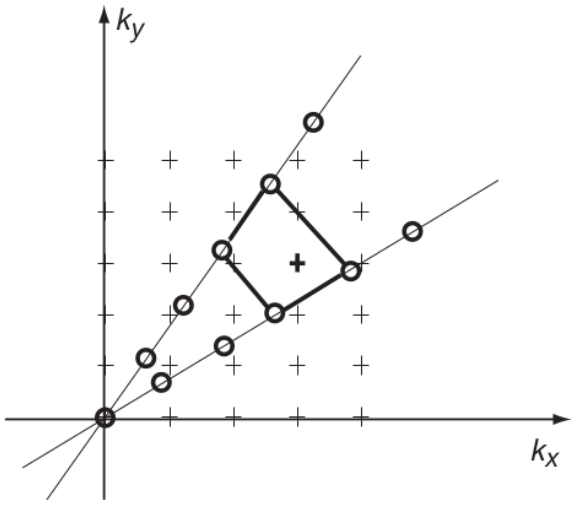
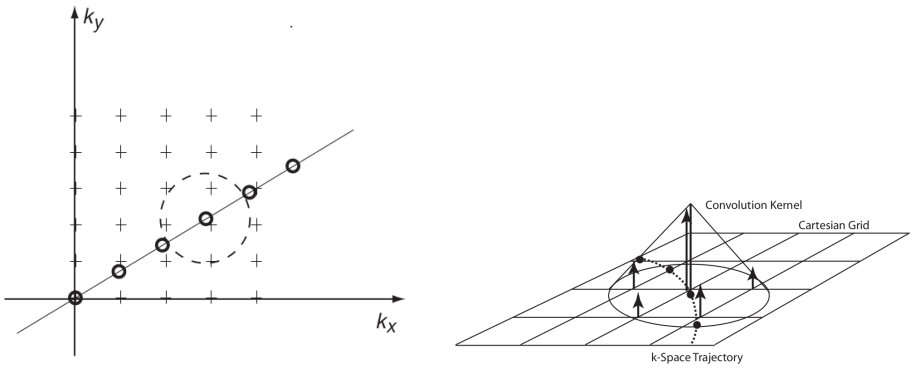


Figure 1.2-8: Grid-driven interpolation, the value at any Cartesian grid point (+) is computed as the interpolation (*e.g.* bilinear) of the four surrounding non-Cartesian points (o). The interpolation kernel is a user-defined parameter of grid-driven methods, tat can be either linear, cubic, or based on the nearest neighbor algorithm.



(a) (b)

Figure 1.2-9: (a) Data samples (o) lie on diameters in k-space, each data point is considered convolved with a small kernel and the value of that convolution is added to the adjacent k-space grid point (+). (b) Each sampled point (dashed line) is convolved with a gridding kernel and that convolution is evaluated at the adjacent grid points.

In practice, this approach is seldom if ever used, as the MRI community converges on the use of data-driven interpolations, which are discussed next.

Data-driven. The idea behind data-driven approach is to add the contribution of each data point to the neighboring grid points. Figure 1.2-9 illustrates the data-driven gridding operation. Consequently, each input sample is considered convolved with a gridding kernel [Jackson 1991, O'sullivan 1985]. The latter is chosen wide enough to cover the neighboring grid points. In this way, each data sample is distributed over adjacent grid locations. The striking difference between this approach and the grid-driven one lies on the use of the whole data set in data-driven methods.

However to mitigate the fact that some k-space region (*e.g.* the center) might be over

represented, one needs to introduce a density compensation step to correct for potential bias arising from high concentration of samples in those regions. Akin to grid-data methods, the reconstruction accuracy is a trade-off between the interpolator complexity and k-space oversampling. A fair compromise consists in choosing a Kaiser-Bessel kernel with a limited (*e.g.* from 1.2 to 1.5-times) oversampled grid, see [Beatty 2005] for details.

The non-uniform Fast Fourier Transform. Others existing data-driven gridding method have been proposed in the literature [Dutt 1993, Fessler 2003, Greengard 2004, Keiner 2009], hereafter we will only remind the nonequispaced Fast Fourier Transform (NFFT) introduced by [Keiner 2009].

Let us consider a d-dimensional nonequispaced Discrete Fourier Transform defined by a set of arbitrary spatial node χ and a frequency bandwidth vector $\boldsymbol{N}$. Each nodes x_j in the nonequispaced sampling set $\chi := \left\{x_j \in \mathbb{T}^d : j = 0, \ldots, M-1\right\}$ is drawn from the d-dimensional taurus $\mathbb{T}^d \cong \left[-\frac{1}{2}, \frac{1}{2}\right)^d$ with the number of nodes equal to $|\chi| = M$. For each dimension $t = 0, \ldots, d-1$, the bandwidth $N_t \in 2\mathbb{N}$ is defined as a fixed even number. Let us define the multi-index set $\boldsymbol{I}_N$ which is a representation of all possible frequencies in a transform:

$$\boldsymbol{I}_N := \mathbb{Z}^d \cap \prod_{t=0}^{d-1} \left[-\frac{N_t}{2}, \frac{N_t}{2}\right) = \left\{\boldsymbol{k} = (k_t)_{t=0,\ldots,d-1} \in \mathbb{Z}^d : -\frac{N_t}{2} \leq k_t < \frac{N_t}{2}, t = 0, \ldots, d-1\right\}$$

Given Fourier coefficients $\hat{f}_k \in \mathbb{C}, \boldsymbol{k} \in \boldsymbol{I}_N$ as input, the nonequispaced Discrete Fourier Transform (NDFT) is defined as the evaluation of the corresponding trigonometric polynomial $f \in T_N$ at the set of M arbitrary nodes χ, *i.e.*, the calculation of the sums:

$$f_j = \sum_{\boldsymbol{k} \in \boldsymbol{I}_N} \widehat{f}_{\boldsymbol{k}} \mathrm{e}^{-2\imath\pi \boldsymbol{k}\cdot\boldsymbol{x}_j}, \quad (\forall j = 0, \ldots, M-1) \tag{1.16}$$

The NDFT can also be represented as a matrix-vector product defined as follows:

$$\boldsymbol{f} = \boldsymbol{A}\widehat{\boldsymbol{f}} \tag{1.17}$$

with the columnwise vectors $\boldsymbol{f} := (f_j)_{j=0,\ldots,M-1}, \widehat{\boldsymbol{f}} := \left(\widehat{f}_{\boldsymbol{k}}\right)_{\boldsymbol{k}\in\boldsymbol{I}_N}$, and the nonequispaced Fourier matrix $\boldsymbol{A} := \left(\mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{k}\boldsymbol{x}_j}\right)_{j=0,\ldots,M-1}^{\boldsymbol{k}\in\boldsymbol{I}_N}$, typically $\boldsymbol{A}$ is not square. The adjoint operator (or matrix) is then defined by:

$$\widehat{\boldsymbol{h}} = \boldsymbol{A}^H \boldsymbol{f} \tag{1.18}$$

which is equivalent to the sums:

$$\forall \boldsymbol{k} \in \boldsymbol{I}_N, \quad \widehat{h}_{\boldsymbol{k}} = \sum_{j=0}^{M-1} f_j e^{2\imath\pi \boldsymbol{k}\cdot\boldsymbol{x}_j}. \tag{1.19}$$

The NFFT C-library is a fast approximation algorithm that computes the sums in Eqs. (1.16) and (1.19). This library uses only $\mathcal{O}\left(|\boldsymbol{I}_N| \log |\boldsymbol{I}_N| + |\log \varepsilon|^d M\right)$ instead of $\mathcal{O}\left(M\,|\boldsymbol{I}_N|\right)$, floating point operations, with ε being the desired computation accuracy of the approximation. The key idea of the library is to use standard FFTs (more precisely

the FFTW[4] [Frigo 1998]) in combination with an approximation scheme that is based on a windowing function φ. The latter must be mutually well localized in spatial and frequency domains. Two parameters control the accuracy of the NFFT: the oversampling factor σ and a truncation parameter m.

Nyquist criterion in k-space sampling

In the field of digital signal processing, the sampling theorem is fundamental as it establishes a sufficient condition on the sampling rate that allows a discrete sequence of samples to capture all information contained in a band-limited signal. The Nyquist-Shannon theorem states that the spectrum of a discretely sampled signal is replicated in the Fourier conjugate domain. It provides a prescription for the nominal sampling rate required to avoid aliasing. It may be stated as follows:

Theorem 1.2.1. *A band-limited continuous-time signal can be sampled and perfectly reconstructed from its discrete samples if the waveform is sampled at least twice as fast as its highest frequency component.*

Spatial encoding in MRI might not deviate from this rule, hence to prevent wrap-around artifact or *aliasing* the theorem should be applied to the k-space domain. Let us consider a 2D Cartesian acquisition with an image field of view of $\mathrm{FOV}_x \times \mathrm{FOV}_y$ and a matrix size of $N_x \times N_y$ the image resolution is defined by $\Delta_x = \frac{\mathrm{FOV}_x}{N_x}$ and $\Delta_y = \frac{\mathrm{FOV}_y}{N_y}$. To comply with the Nyquist-Shannon criterion and get the desired resolution, the k-space sampling should respect the following relations:

$$\Delta k_x \leq \frac{1}{\mathrm{FOV}_x} \quad \text{and} \quad \Delta k_y \leq \frac{1}{\mathrm{FOV}_y} \tag{1.20a}$$

$$K_x^{\max} = \frac{1}{2\Delta_x} \quad \text{and} \quad K_y^{\max} = \frac{1}{2\Delta_y} \tag{1.20b}$$

where Δk_x (resp. Δk_y) is the frequency interval in the x (resp. y) k-space direction, and $K_x^{\max}$ (resp. $K_y^{\max}$) is the maximum frequency along the x (resp. y) k-space direction.

Due to the short lifespan of the NMR signal (few tens of ms) and the requirements in (1.20) related to the sampling theorem, the acquisition time gets often too long in MRI. Consequently, reducing this time has become a major issue especially for high resolution imaging, as the latter means larger $K_x^{\max}$ and $K_y^{\max}$. Therefore, over the last two decades, many contributions have tried to dramatically reduce the scan time. In the next section, we will review one of the most widely used techniques to shorten the acquisition time, which is called *parallel imaging* or *parallel-MRI*[5]. The latter is based on *phased array coils*. Consequently, we will first explain the physics principles underlying the way these coils work and second we will highlight their pros and cons. Last, we will summarize the acceleration and reconstruction methods in the parallel imaging field.

[4] which is accelerated for computing FFT for integers that do not read as 2^n with $n \in \mathbb{N}_\star$

[5] Parallel-MRI might be confusing since it can refer to both, the use of multi-channel receiver coil to acquire the signal and the acceleration.

1.3 Multi-channel receiver coil and Parallel Imaging

In Section 1.2.2, we only considered a coil with a quadrature design to record the NMR signal. Since the magnitude of the electric voltage induced in the receiver coil is pretty small, namely in the order of *millivolts*, the most important properties of an RF receiver coil are the maximal achievable signal-to noise ratio (SNR) and a large coverage of the RF response over the imaged volume. Hereafter, we discuss both aspects.

1.3.1 Signal-to-Noise Ratio increase

Many factors determine the SNR available in a nuclear magnetic resonance experiment. In the early days of MRI, two somewhat parallel paths towards more efficient MRI experiments were drawn: the first consisted of improving the gradient technology and pulse sequence design to increase spatial resolution and decrease imaging time, and the second relied on the development of the hardware RF coil technology. Regardless of the field strength, the key requirement of any receiver coil is achieving the maximum SNR in order to obtain the best possible image quality. This section provides tools to understand how phased array coils maximize the input-SNR.

Noise origin

As it is not straightforward to exhibit a simple relation between the sample scanned and the noise associated with the RF coil, the discussion hereafter will suppose perfectly homogenized fields. The expression for the noise in the coil is based upon thermodynamic principle and is given by [Haacke 1999]:

$$noise \propto \sqrt{4kT_{coil}\Delta f \mathcal{R}_{eff}} \tag{1.21}$$

where T_{coil} refers to the temperature of the coil, Δf the bandwidth used for the experiment, and $\mathcal{R}_{eff}$ is the effective resistance, which includes contributions from the coil $\mathcal{R}_{coil}$, the electronics $\mathcal{R}_{electronics}$ and the sample being imaged $\mathcal{R}_{sample}$ [Haacke 1999], such that:

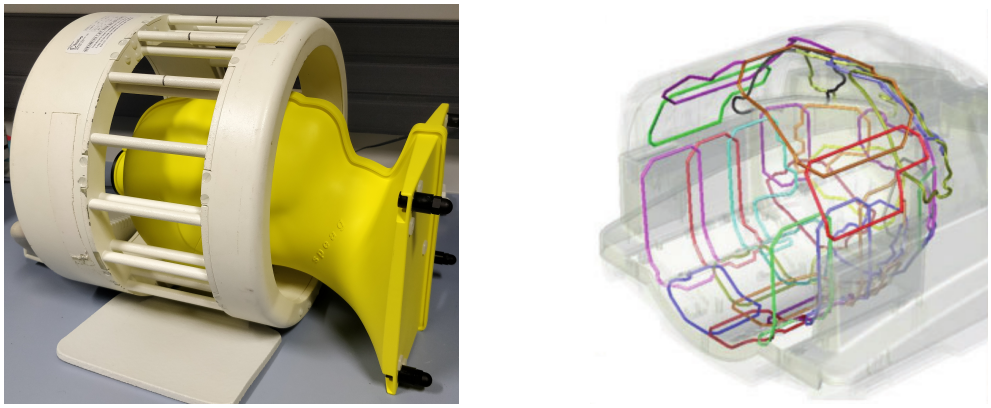
$$\mathcal{R}_{eff} = \mathcal{R}_{coil} + \mathcal{R}_{electronics} + \mathcal{R}_{sample} \tag{1.22}$$

Although $\mathcal{R}_{coil} \propto V_{coil}$, with V_{coil} being the coil volume, the latter could be neglected owing to the recent improvement in electronics. Hence, it is seen that $\mathcal{R}_{eff} \approx \mathcal{R}_{sample}$, which is proportional to the volume of the region of the body ($V_{sensitivity}$) from which the signal arises. From the signal expression described in Eq. (1.12) and the noise definition given in Eq. (1.21), the SNR associated with a particular coil reads as follows[6]:

$$SNR \propto \frac{\omega^{7/4}\boldsymbol{B}_{\perp}}{\sqrt{T_{coil}\Delta f V_{sensitivity}}} \tag{1.23}$$

From Eq. (1.23) one can deduce that optimal coils are the ones that only cover the imaging region of interest.

[6] This relation holds only for small objects compared to the wavelength, in the other cases the expression of the signal is still debated.



(a) (b)

Figure 1.3-10: (a): Single channel birdcage coil used on the 7 T scanner at Neurospin. (b) 32-channel array coil in transparency, the different channels are color-coded (from [Paolini 2015]).

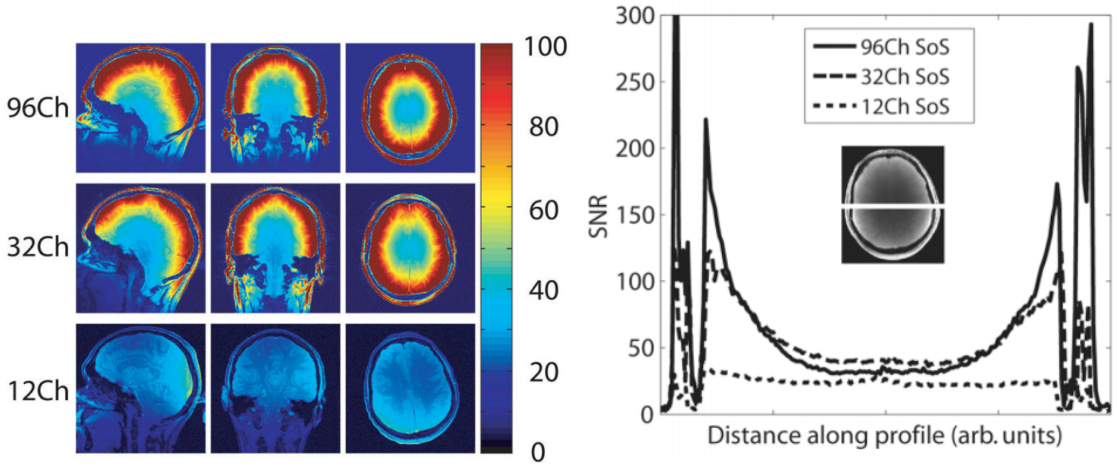
Phased array coil

In MRI, a phased array coil generally refers to a set of receiver coils whose signals are combined to obtain a uniform image over a region larger than any individual coil could cover while taking advantage of the high SNR available from the smaller individual coils, as shown in Eq. (1.21). Figure 1.3-10 shows the single channel birdcage coil available at Neurospin and an example of multi-channel array head coil, where the small elements provide a higher input-SNR compared to the single channel coil.

In order to obtain optimal SNR from phased array coil, it is necessary to make sure that the noise from coil to coil is largely uncorrelated. On top of dedicated electronic circuit for each coil, minimal electromagnetic interaction should exists between the coils to make sure that the noise samples remain uncorrelated. The SNR is not the only parameter to take into account for the design of phased array coils, the *penetration depth* is as important as the SNR. The penetration depth refers to the depth at which the coils sensitivity drops to 37% of that at the coil center. As a rule of thumb, the penetration depth of a circular, sample noise-dominated loop-coil is approximately equal to its diameter [Haase 2000]. Figure 1.3-11 illustrates the trade-off between penetration depth and sensitivity. While a large number of channels, *e.g.* 96, with small diameter provides a significant SNR gain on the cortical surface, the central SNR shows 20% SNR loss compared to the 32-channel coil for root-sum-of-square combination (cf Section 1.3.1 for more details on coil combination).

Channel combination: retrieving magnitude and phase information

When multiple receivers are used, a complex-valued image is reconstructed from each receiver coil using the inverse FFT. However the individual coil images are not really useful as each image is sensitive to a particular brain region. Instead, one prefers to combine all of them in order to cover the whole FOV. In the next part, we address the problem of coil combination for both the phase and magnitude images.



(a) (b)

Figure 1.3-11: Comparison of 12, 32, and 96-channel phased array coils at 3 T. (a) SNR map for acquisitions with the different receiver coils. Large number of channels with small diameter are more sensitive to the region next to the coils. (b) SNR profile along the central line in the brain for receiver coils with different number of channels, while the SNR gain is less significant in the middle of the brain. From [Wiggins 2009].

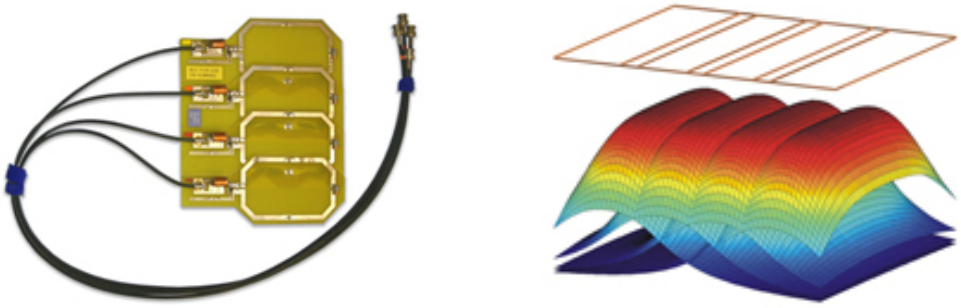
The quest of optimal combination of coil specific images $(I_k)_{1\leq k\leq L}$ has been a long-standing issue. It can be optimally tackled by injecting the knowledge of the receiver field $(\mathrm{B}_{1\,\ell}^{-})_{1\leq \ell\leq L}$ associated with the ℓ^{th}-coil in the following relation:

$$I(x,y) = \frac{\sum_{k,\ell=1}^{L} \mathrm{B}_{1\,\ell}^{-*}(x,y)\ \boldsymbol{\Sigma}_{k,\ell}^{-1}\ I_k(x,y)}{\sum_{k,\ell=1}^{L} \mathrm{B}_{1\,\ell}^{-*}(x,y)\ \boldsymbol{\Sigma}_{k,\ell}^{-1}\ \mathrm{B}_{1\,\ell}^{-}(x,y)} \tag{1.24}$$

where $\boldsymbol{\Sigma}$ is the noise correlation matrix associated with the coils, and each entry $\boldsymbol{\Sigma}_{k,\ell}$ is the correlation between the k^{th} and the ℓ^{th} coil given by:

$$\boldsymbol{\Sigma}_{k,\ell} = \mathbb{E}\left[N_\ell^* N_k\right] - \mathbb{E}\left[N_\ell^*\right]\mathbb{E}\left[N_k\right] \tag{1.25}$$

with $\mathbb{E}\left[\cdot\right]$ being the expectation operator, taken over the noise measurement N_ℓ during a short acquisition without gradient or RF pulse and for which the configuration of the readout bandwidth must be the same as the one used for the "real" scan. The receiver field $\mathrm{B}_{1\,\ell}^{-}$ is also known as the sensitivity profile of the coil. Figure 1.3-12 illustrates the sensitivity profile of a 4-channel reception coil. Since the coil sensitivity profiles depend on the loading of the receiver coils (cf Eq. (1.22)), they need to be estimated for each scan (Section 3.2 will discuss methods that estimate those maps). In many situations, to avoid the estimation of the coil sensitivity profile, and potential errors due to this estimation, a method that combines the data without detailed knowledge of the receiver fields and that preserves,at the same time, a high SNR for the phased array, is really appealing.



(a) (b)

Figure 1.3-12: (a) Four coil phased array showing overlapping coils, (b) their associated sensitivity maps. From [Gruber 2018].

In [Roemer 1990], the authors have proposed an approximation for the coil combination in Eq. (1.24) by proposing to end up with a single image $I(x, y)$ computed as the sum-of squares (SOS) of coil-specific images $I_k(x, y)$, $\forall k = 1, \ldots, L$:

$$I(x,y) = \sqrt{\sum_{\ell=1}^{L} I_\ell^*(x,y) I_\ell(x,y)}\,. \tag{1.26}$$

Due to the inconvenience of estimating the receiver fields $\mathrm{B}_{1\,\ell}^-$, the sum-of-squares approximation is more commonly used today. the authors of [Roemer 1990] have also shown that the SNR loss resulting from this SOS approximation usually ranges around a few percent only. Moreover, it should be noted that the optimal coil combination in (Eq. (1.24)) provides a complex-valued MR image, whereas the SOS method only yields a magnitude image. This makes its application problematic with methods that requires phase information, in applications where phase brings relevant knowledge such as flow/velocity, temperature mapping or shimming. Combining the phase information is therefore an critical aspect for the mentioned techniques.

Based on the slowly varying phase assumption, [Parker 2014] have proposed to create a virtual coil reference to which the measurements of each receiver coil are aligned and then combined to obtain an optimal phase distribution estimate. The method is called *virtual coil reconstruction* and defines a virtual coil as follows:

$$I_{virtual} = \sum_{\ell=1}^{L} w_\ell I_\ell = \sum_{\ell=1}^{L} |w_\ell|\, e^{\imath\varphi_{ref\,\ell}} |I_\ell| e^{\imath(\vartheta_\ell + \theta + \eta_\ell)} \tag{1.27}$$

where the weights $w_\ell = |p_\ell|\, e^{-i\varphi_{ref\,\ell}} / \sum_k |p_k|$, I_ℓ is the reconstructed image of the ℓ^{th} coil, $\varphi_{ref\,\ell}$ is the phase of the weights used to avoid SNR loss, ϑ_ℓ is the coil phase, θ is the true tissue phase and η_ℓ is a random component due to the thermal noise in the individual RF coil measurement. While θ and ϑ_ℓ are supposed to be slowly varying, *i.e.* smooth in space, the phase noise η_ℓ is expected to be uncorrelated from one voxel to another. Because the coil phase $\varphi_{ref\,\ell}$ varies slowly in space and in order to reduce measurement errors the

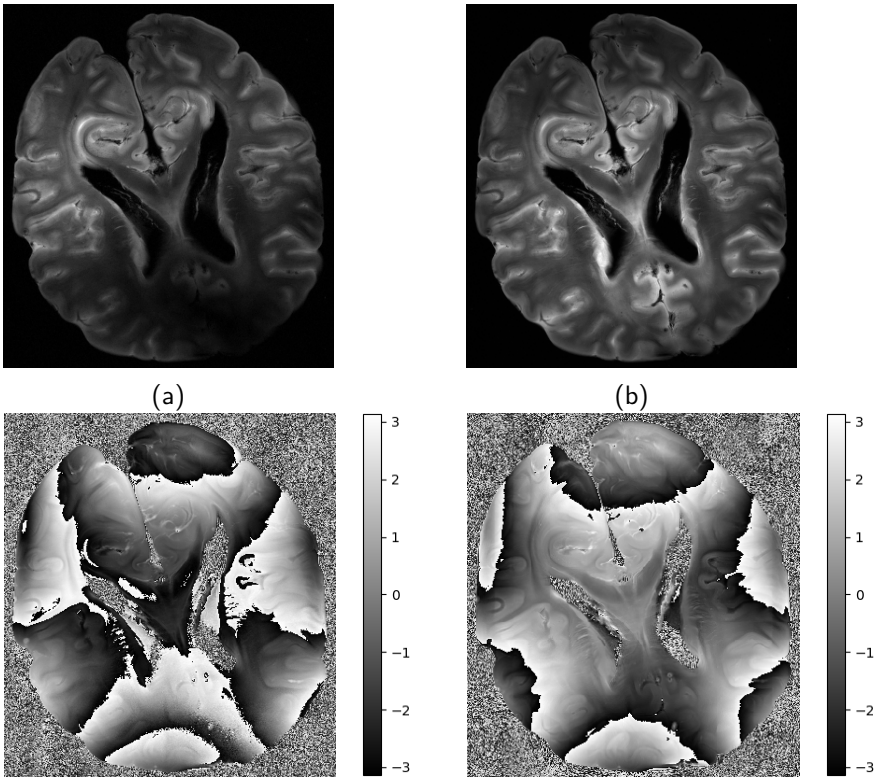


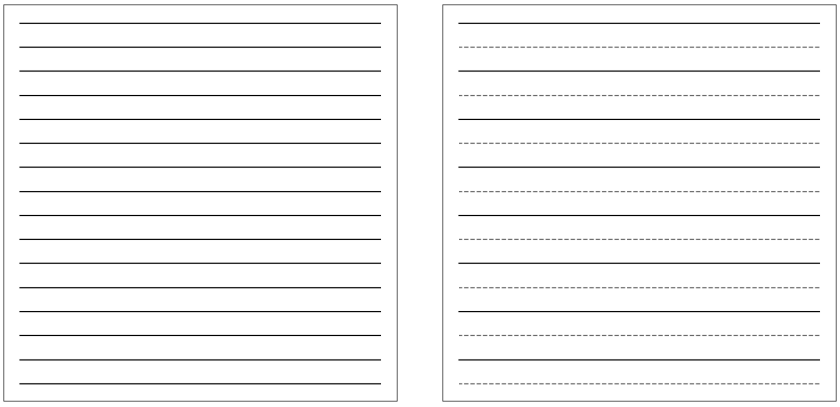
Figure 1.3-13: Top: Coil combination methods on an ex-vivo human brain acquired at 7 T with a 32-channel head coil: (a) magnitude image associated with the complex summation, (b) approximate combination proposed by [Roemer 1990] using the square root of the SOS. Bottom: (c) phase image associated with the complex summation, (d) phase obtained with the virtual coil method as proposed by [Parker 2014].

averaged phase can be considered instead:

$$\varphi_{ref_\ell} = -angle\left(\frac{\sum_{x,y} p_\ell(x,y)}{\sum_{x,y} |p_\ell(x,y)|}\right) \tag{1.28}$$

where x, y are 2D positions over the entire image. This can be easily extended to 3D imaging of course. Using the virtual coil image, coil-specific phase images are subtracted to the virtual phase and low-pass filtered to remove the noise phase. Hence, the virtual coil is used as phase reference.

Figure 1.3-13 shows the coil-combination of an ex-vivo human brain acquired at 7 T with a 32-channel head coil, using either a simple complex sum of the coil-specific images or the SOS [Roemer 1990] to combine the magnitude images and the virtual coil [Parker 2014] to retrieve phase information. The magnitude images depicted in Figures 1.3-13a and 1.3-13b demonstrate the good performances of the SOS method while the simple complex-valued sum yields a strong variation of the SNR across pixels, hence poorer performances. Also, the virtual coil method shows the benefits for phase combination as the resulting phase image gets smoother. It should be also noted that the phase is wrapped therefore variations form $-\pi$ to π are actually smooth.



(a) (b)

Figure 1.3-14: (a) Fully sampled Cartesian k-space. (b) Acquired k-space for parallel imaging with an acceleration factor $R = 2$. Dashed lines represent the missing samples while solid lines correspond to the acquired k-space values. Horizontal axis represents the frequency encoding direction and vertical axis is the phase-encoding one.

1.3.2 Parallel MRI: first MR accelerations

In the early 2000's, it has been noticed that phased array coils might not only be dedicated to improve the SNR, but also to accelerate the acquisition [Pruessmann 1999], yielding to the known *parallel imaging* techniques. The basis of these methods is that the scan time is proportional to the number of phase encoding lines in a Cartesian acquisition. Increasing the distance between phase-encoding lines in k-space by a factor of R, while keeping the maximal spatial resolution fixed, reduces the scan time by the same factor. However increasing the distance between phase-encoding lines also decreases the FOV (as shown in Eq. (1.20)), leading to aliasing artifacts. In parallel imaging, dedicated reconstruction techniques have been proposed to prevent aliasing artifact to occur. These methods are based on the spatial dependence of the B_1^- field on the receiver coils (also called sensitivity maps). Generalised autocalibrating partially parallel acquisition (GRAPPA) [Griswold 2002, Blaimer 2004] and SENsitivity Encoding (SENSE) [Pruessmann 1999] methods are the most popular parallel imaging reconstruction techniques that will be described hereafter[7]. In short, SENSE methods perform image reconstruction in the image space whereas GRAPPA methods operate in the native k-space domain.

Sensitivity Encoding methods for fast MRI

With parallel imaging the scan time is reduced by a factor of R if we collect one line every R lines in the phase-encoded direction. The FOV of the reconstructed image gets therefore reduced by the same factor. Since Nyquist criterion is no longer satisfied, some pixels will be aliased. For each spatial location of reduced FOV images, the pixel value

[7]Other complementary acceleration methods exist such as Controlled Aliasing In Parallel Imaging Results In Higher Acceleration (CAIPIRINHA) [Breuer 2006], however those are only specific to 3D acquisitions.

is actually the superposition of the original object and R shifted replicates, the distance between two replicates being equal to FOV_y/R if we assume k_y to be the phase encoding direction.

The key idea to separate the signals lies in the fact that in each single-coil image $\boldsymbol{I}_\ell$, the signal superposition occurs with different weights according to local coil sensitivities $\boldsymbol{S}_\ell$, with $\ell \in \{1, \ldots, L\}$. The forward model in the image domain relating the unfolded image $\boldsymbol{I}$ to the folded ones $(\boldsymbol{I}_\ell)_{\ell=1}^{L}$ thus reads:

$$\forall = 1, \ldots, L, \quad \boldsymbol{I}_\ell = \boldsymbol{S}_\ell \boldsymbol{I} \; (+\boldsymbol{N}_\ell) \tag{1.29}$$

where $\boldsymbol{N}_\ell$ stands for some additional noise related to receiver coil ℓ. As matrices $\boldsymbol{S}_\ell$ are rectangular of dimensions $N_x N_y/R \times N_x N_y$, the SENSE solution $\widehat{\boldsymbol{I}}_{sense}$, introduced by [Pruessmann 1999], is thus given by the least squares solution:

$$\widehat{\boldsymbol{I}}_{sense} = \left(\sum_{\ell=1}^{L} \boldsymbol{S}_\ell^{\mathrm{H}} \boldsymbol{S}_\ell \right)^{-1} \boldsymbol{S}_\ell^{\mathrm{H}} \boldsymbol{I}_\ell \, . \tag{1.30}$$

Of course, it is worth mentioning that this method relies on the extraction and prior knowledge of sensitivity profiles $(\boldsymbol{S}_\ell)_{\ell=1}^{L}$. This topic will be discussed in Section 3.2.

Generalized Auto-calibrating Partially Parallel Acquisition

In order to avoid sensitivity profile estimation, methods like GRAPPA [Griswold 2002] formulate the parallel imaging reconstruction as an interpolation problem in k-space. This method relies on the acquisition of auto-calibration central lines (ACL) that are used to calibrate an interpolation kernel in k-space. It is worth mentioning that the acquisition of those extra-lines slows down the overall scan time. To shortly describe the GRAPPA method, let us define a set of block operators.

The operator $\boldsymbol{R}_r$ represents the selection a block of k-space samples in the neighborhood of the position r over all coils. GRAPPA was originally developed for Cartesian under-sampling, so the k-space neighborhood was defined over the grid. Also, the operator $\boldsymbol{P}_r$ represents the local sampling pattern that selects the collected k-space samples from a given block. Let $\boldsymbol{y}$ be the multi-coil k-space samples concatenated into a columnwise vector in which unseen data are zero filled, and let $\widehat{\boldsymbol{y}}$ be the reconstructed k-space. The product $\boldsymbol{P}_r \boldsymbol{R}_r \boldsymbol{y}$ thus defines a vector containing only the the measurements located in the k-space neighborhood around position r. The recovery of the missing values in the ℓ^{th}-coil at position r is given by:

$$\widehat{\boldsymbol{y}}_\ell(r) = \left(\boldsymbol{P}_r \boldsymbol{R}_r \boldsymbol{y} \right)^\top \boldsymbol{g}_{r,\ell} \tag{1.31}$$

where $\boldsymbol{g}_{r,\ell}$ is the GRAPPA kernel, *i.e.* a set of reconstruction weights. The full grid is reconstructed by the evaluation of Eq. (1.31) for all coils at each missing k-space position, while the GRAPPA kernel $\boldsymbol{g}_{r,\ell}$ are obtained by solving the same relation at different position where $\widehat{\boldsymbol{y}}_i$ is known. Typically this is done from the ACL region in the center of k-space.

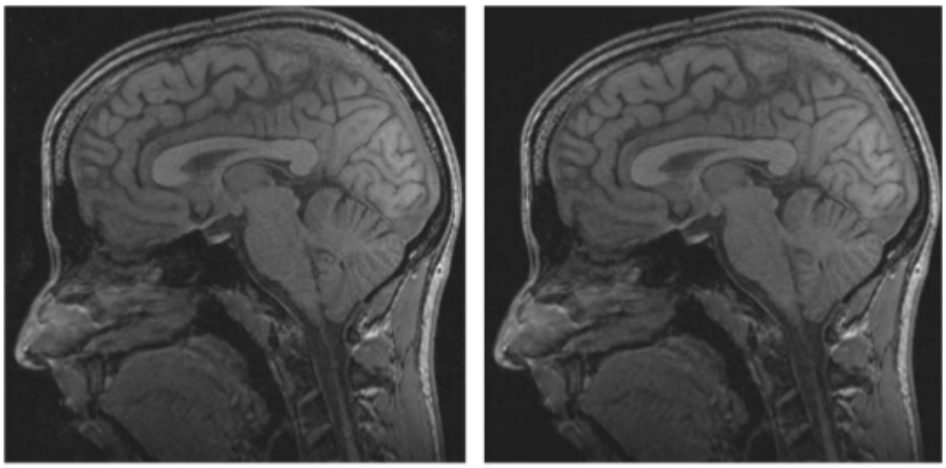


Figure 1.3-15: Comparison of the image quality of SENSE and GRAPPA reconstructions with an acceleration factor $R = 3$. **Left**: SENSE reconstruction with accurate coil sensitivity extraction. **Right**: GRAPPA reconstruction. Visually, no difference can be seen between both methods. From [Blaimer 2004].

Comparison SENSE vs GRAPPA

In their study, [Blaimer 2004] have proposed an extensive comparison of different reconstruction methods including SENSE and GRAPPA for several body organs, acquisition sequences and acceleration factors.

Although both methods provide nearly identical image quality (as illustrated in Figure 1.3-15), SENSE method gives a better SNR provided that accurate coil sensitivity maps are estimated. Regarding GRAPPA, the method is more robust to imperfect kernel estimation, which makes this reconstruction technique particularly beneficial in organ regions where accurate estimation of coil sensitivity maps is difficult to reach. This typically happens in regions associated with low spin density (*e.g.* lung imaging or deep brain structures such as basal ganglia) or for specific fast imaging sequences such as GRE Echo-Planar-Imaging (EPI) [Mansfield 1977b] which suffers from strong distortion. Yet the SENSE method is more robust to high acceleration factors as illustrated in Figure 1.3-16, where no aliasing artifact is reported for SENSE reconstruction. However, local noise enhancement can be seen in the same reconstruction.

Since its first application in clinics, MRI has always been subject to intensive research for shortening long scan times. Based on the development of phased array coils and dedicated image reconstruction methods, the k-space has been under-sampled to speed up the acquisition with parallel MRI methods. In 2D imaging the typical acceleration factors used in practice are lower than 4, whereas in 3D imaging acceleration can be achieved along the phase and partition encoding directions, leading to higher acceleration factors (*e.g.* 3×3). Although this scan time reduction may turn out large enough for some applications, high-resolution imaging (*e.g.* $\simeq 500 \times 500\mu m$) requires further acceleration. In the next chapter, we investigate a promising alternative based on the Compressed Sensing (CS) theory [Donoho 2006, Candès 2006a], with the ultimate goal of speeding

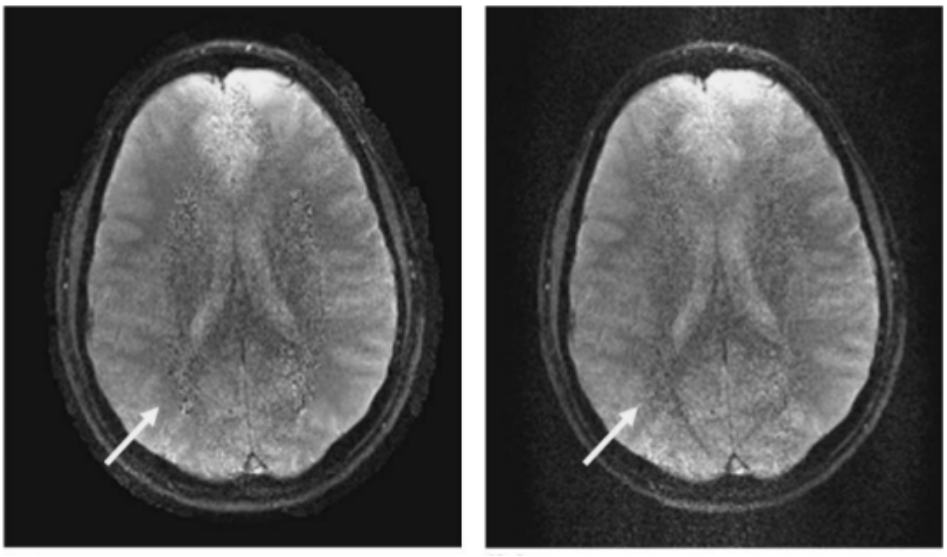


Figure 1.3-16: Comparison of artifacts in SENSE and GRAPPA reconstructions at high acceleration factors R (chosen close to the number of coils). In this example, $R = 4$ leads to a four-fold acceleration of the scan time for an equal number of coils ($L = 4$). **Left**: The SENSE image shows a local noise enhancement due to non-ideal conditioning for the reconstruction. **Right**: The noise enhancement in GRAPPA is distributed more evenly over the FOV. Additionally, aliasing artifacts can be seen due to inaccurate calculation of missing k-space lines. From [Blaimer 2004].

up MR acquisition even further at the cost of more complex image reconstruction but with limited impact on image quality.

* * *
* *
*

Chapter 2

Compressed Sensing in MRI

Chapter Outline

Minor contribution of this chapter (Section 2.2.3) was accepted for publication as:

C. Lazarus, P. Weiss, N. Chauffert, F. Mauconduit, L. El Gueddari, C. Destrieux, I. Zemmoura, A. Vignaud and P. Ciuciu. *SPARKLING: variable-density k-space filling curves for accelerated T_2^*-weighted MRI.* Magnetic resonance in medicine, vol. 81, pages 3643–3661, 2019.

Minor contributions of this chapter (Sections 2.2.3, 2.3.1 and 2.3.1) were presented at international conferences as:

C. Lazarus, P. Weiss, L. El Gueddari, F. Mauconduit, , A. Vignaud and P. Ciuciu. *Distribution-controlled and optimally spread non-Cartesian sampling curves for accelerated in vivo brain imaging at 7 Tesla.* In Proceedings of the 26th Annual Meeting of ISMRM, page 2666, Paris, France, 2018.

H. Cherkaoui, L. El Gueddari, C. Lazarus, A. Grigis, F. Poupon, A. Vignaud, S. Farrens, J.-L. Starck and P. Ciuciu. *Analysis vs synthesis-based regularization for combined compressed sensing and parallel MRI reconstruction at 7 Tesla.* In 2018 26th European Signal Processing Conference, pages 36–40. EUSIPCO, 2018.

H. Carrié, L. El Gueddari, H. Cherkaoui, E. Dohmatob, L. Leroi and P. Ciuciu. *Multi-Contrast Dictionary Learning for 2D Compressed Sensing MRI Reconstruction.* In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington D.C., USA, 2018. IEEE ISBI.

This section introduces some theoretical and practical concepts underlying Compressed sensing (CS) for MRI.

2.1 Compressed Sensing theory

In this thesis, we consider the case of discrete images, *i.e.* defined on a grid with finite number of pixels denoted by $\boldsymbol{x} \in \mathbb{C}^n$ with n being the total number of pixels. Let $\boldsymbol{y} \in \mathbb{C}^m$ be the observed signal and $\boldsymbol{E} \in \mathbb{C}^{m \times n}$ the encoding matrix which encodes the physical principles underlying MRI acquisition, either Fourier transform in the easiest case treated in this thesis or more complex models including off-resonance effects, chemical shift, etc, see for instance [Doneva 2020] for an extensive review on computational models in MRI. The relation between $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{E}$ – also called the forward model – reads as follows:

$$\boldsymbol{y} = \boldsymbol{E}\boldsymbol{x} \tag{2.1}$$

If $\boldsymbol{E}$ is invertible (in particular for $m = n$) then the reconstruction is performed through the inversion of matrix $\boldsymbol{E}$. However, the Compressed Sensing (CS) theory supposes that $m \ll n$. This assumption will hold in the rest of this thesis.

2.1.1 Sparsity and incoherence

In the CS context (*i.e.* $m \ll n$) the number of solutions of Eq. (2.1) is infinite and the problem is considered *ill-posed*. To solve Eq. (2.1), prior information must be injected to guarantee the uniqueness of the solution. CS theory works with sparsity assumptions on $\boldsymbol{x}$. We recall here what this means. We define a s-sparse vector on $\mathbb{C}^n$ as a vector with only $s \ll n$ non-zero entries. If the image $\boldsymbol{x}$ is not naturally sparse then one can find an orthonormal transform $\boldsymbol{\Psi} : \mathbb{C}^n \to \mathbb{C}^p$ which transforms $\boldsymbol{x} \mapsto \boldsymbol{z} = \boldsymbol{\Psi}\boldsymbol{x}$ such that $\boldsymbol{z}$ is s-sparse (or at least compressible), and $\boldsymbol{\Psi}^{\mathrm{H}}$ being its adjoint *i.e.* transpose conjugate which in the case of orthonormal transforms corresponds to its inverse. The choice of the sparsifying transform will be discussed in Section 2.3.1. Then the following optimization tells us how to reconstruct the image:

$$\widehat{\boldsymbol{z}} = \underset{\boldsymbol{z} \in \mathbb{C}^p}{\arg\min} \|\boldsymbol{z}\|_0 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}. \tag{2.2}$$

where ℓ_0 stands for the pseudo-norm corresponding to the number of non-zeros coefficients and $\boldsymbol{A} = \boldsymbol{E}\boldsymbol{\Psi}^{\mathrm{H}}$. Once a solution $\widehat{\boldsymbol{z}}$ of Eq. (2.2) is found the image is reconstructed as follows: $\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}^{\mathrm{H}}\widehat{\boldsymbol{z}}$. Due to the combinatorial properties of the ℓ_0 pseudo-norm and its non-convexity, problem Eq. (2.2) is NP-complete meaning that, any solution of Eq. (2.2) can be tested quickly in polynomial time. However the existence of an algorithm that solves problem (2.2) in a efficient manner (in polynomial time too) has not been demonstrated yet[1]. Due to the mentioned impediments of the ℓ_0-pseudo norm, problem (2.2) is often relaxed and its tightest convex envelop [Boyd 2004] is considered instead – *i.e.* the tightest convex hull of the ℓ_0-pseudo-norm is the ℓ_1-norm $\|\boldsymbol{x}\|_1 = \sum_{j=1}^{n} |x_i|$ – leading to the following optimization problem:

$$\widehat{\boldsymbol{z}} = \underset{\boldsymbol{z}\in\mathbb{C}^p}{\arg\min} \|\boldsymbol{z}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}. \tag{2.3}$$

Based on the restricted isometry property (RIP), which characterizes matrices nearly orthonormal, the first results of the Compressed Sensing theory [Candès 2006b, Donoho 2006, Candès 2008] guarantee exact recovery of s-sparse signals. The RIP conditions state that for all s-sparse $\boldsymbol{z}$ vector, it exists $\delta_s \geq 0$ such that:

$$(1 - \delta_s)\|\boldsymbol{z}\|_2 \leq \|\boldsymbol{A}\boldsymbol{z}\|_2 \leq (1 + \delta_s)\|\boldsymbol{z}\|_2 \,. \tag{2.4}$$

Assuming the noiseless case, [Candès 2008] proved the following results:

Theorem 2.1.1. *if $\delta_{2s} \leq \sqrt{2} - 1$, then the solution $\widehat{\boldsymbol{z}}$ of Eq. (2.3) is unique.*

However, due to the restricted assumptions of the RIP condition only a small subset of encoding matrices – such as random matrices with identically independently distributed (i.i.d.) entries – fulfill the RIP conditions and the RIP constant δ_s cannot be calculated explicitly.

Over the same period of time, [Rauhut 2010, Candès 2011] have extended the theory to random linear projections from orthogonal bases. They considered a sensing matrix $\boldsymbol{A} = (\boldsymbol{a}_i^*)_{1\leq i\leq n} \in \mathbb{C}^{m\times n}$ to be constructed by randomly drawing rows of an orthogonal matrix $\boldsymbol{A}_0 \in \mathbb{C}^{n\times n}$ given by:

$$\boldsymbol{A}_0 = \begin{pmatrix} \boldsymbol{a}_1^* \\ \vdots \\ \boldsymbol{a}_n^* \end{pmatrix} \tag{2.5}$$

Then, matrix $\boldsymbol{A}$ can be constructed by randomly drawing rows of $\boldsymbol{A}_0$ as follows:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_{J_i}^* \\ \vdots \\ \boldsymbol{a}_{J_m}^* \end{pmatrix} \tag{2.6}$$

[1]The"P=NP" problem is one of the seven Millennium Prize Problems, for which the Clay Mathematics Institute offered a $1 million prize for the first correct solution.

where $\{J_1, \cdots, J_m\}$ are drawn from an independent and identically distributed manner over the set $\{1, \ldots, n\}$: this means they are copies of a uniform random variable X such that $\mathbb{P}(X = i) = \pi_i = 1/n$, $\forall 1 \leq i \leq n$. The coherence κ of matrix $\boldsymbol{A}_0$ can be defined by:

$$\kappa(\boldsymbol{A}_0) = n \cdot \max_{1 \leq i \leq n} \|a_i\|_\infty^2 \tag{2.7}$$

Then the following theorem holds:

Theorem 2.1.2. *Let $\boldsymbol{z}$ be an s-sparse vector. If the number of measurements m satisfies:*

$$m \geqslant C \cdot s \cdot \kappa(\boldsymbol{A}_0) \cdot \log\left(\frac{n}{\eta}\right) \tag{2.8}$$

where C is a universal constant, then $\boldsymbol{z}$ is the unique minimizer of Eq. (2.3) with a probability exceeding $1 - \eta$.

The latter theorem is particularly interesting since it bridges the gap between two fundamental principles underlying CS theory, namely *sparsity* and *incoherence*. First, from Theorem 2.1.2 it can be noticed that the number of measurements required is proportional to the sparsity level s. Second, the other important value is the coherence $\kappa(\boldsymbol{A}_0)$ which varies from 1 to n for orthogonal matrices. The coherence measures how a vector represented in the sparse space spreads in the acquisition space. In the case of discrete Fourier transform $\boldsymbol{F}$ the coherence is minimal since $\kappa(\boldsymbol{F}) = 1$. However, in MRI (as for most applications), the sensing matrix $\boldsymbol{A}_0 = \boldsymbol{F}\boldsymbol{\Psi}^{\mathrm{H}}$ is coherent ($\kappa(\boldsymbol{A}_0) = \mathcal{O}(n)$), with $\boldsymbol{\Psi}$ is the sparse representation such as wavelet transform (synthesis operator) and $\boldsymbol{\Psi}^{\mathrm{H}}$ its adjoint (conjugate-transpose as defined Page xv) operator that performs image reconstruction from a set of wavelet coefficients. This is often called the "coherence barrier" [Adcock 2017].

2.1.2 Breaking the coherence barrier

The above mentioned result does not assume any structure – apart from sparsity – on the signal of interest. Recovering arbitrary sparse vectors is a very demanding property that precludes the use of CS in many practical settings. If some structure of the sought signal is assumed, then it can be recovered with fewer measurements than those previously uncovered, however dedicated sensing matrices such as variable density sampling should be considered [Chauffert 2014, Adcock 2017]. In that case, the coherence of the matrix $\boldsymbol{A}_0$ is defined as follows:

$$\kappa(\boldsymbol{A}_0, \pi) := \max_{1 \leq i \leq n} \frac{\|\boldsymbol{a}_i\|_\infty^2}{\pi_i} \tag{2.9}$$

where $\pi = (\pi_1, \cdots, \pi_n)$ denotes the probability distribution on $\{1, \cdots, n\}$. It has been shown that the probability density π minimizing $\kappa(\boldsymbol{A}_0, \pi)$ (in Eq. (2.9)) and thus that minimizes the number of measurements m in Eq. (2.8) reads as follows:

$$\pi_i = \frac{\|\boldsymbol{a}_i\|_\infty^2}{\sum_{k=1}^n \|\boldsymbol{a}_k\|_\infty^2} \tag{2.10}$$

The density π defined above, samples more frequently the measurements (i.e. k-space positions in MRI) associated with larger $\|\boldsymbol{a}_i\|_\infty^2$ values, which happens in the low frequency regime. Consequently, the higher frequencies are sparsely sampled compared to the lower ones. Thus, variable density sampling (VDS) is achieved by drawing m raws $(J_1, \ldots, J_m)$ from $\boldsymbol{A}_0$ according to π to build up $\boldsymbol{A}$. The next theorem thus gives us tighter bounds for exact recovery [Chauffert 2014]:

Theorem 2.1.3. *Let $\boldsymbol{z}$ an s-sparse vector. If the number of measurement m satisfies:*

$$m \geqslant C \cdot s \cdot \left(\sum_{k=1}^{n} \|\boldsymbol{a}_k\|_\infty^2 \right) \cdot \log\left(\frac{n}{\eta}\right) \tag{2.11}$$

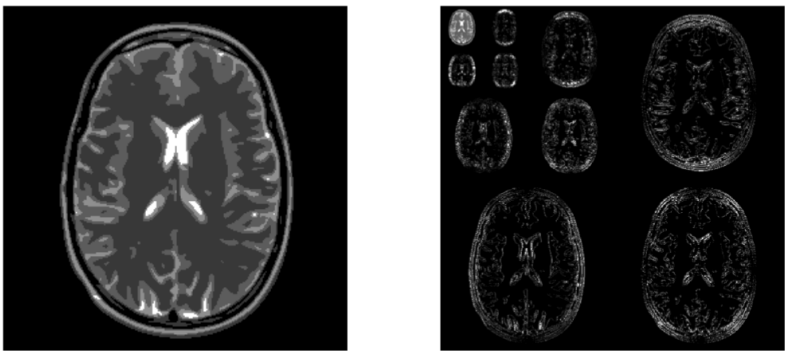
where C is a universal constant, then $\boldsymbol{z}$ is the unique minimizer of Eq. (2.3) *with a probability exceeding $1 - \eta$.*

This theorem is actually more appealing for MRI as one can show that $\sum_{k=1}^{n} \|\boldsymbol{a}_k\|_\infty^2 = \mathcal{O}(\log(n))$. Hence the required number of measurements to get an exact reconstruction in MRI gets closer to $\mathcal{O}(s \log(n)^2)$. Although Theorem 2.1.3 allows to partially break the coherence barrier, if we look closer at the numerical values for an image size of 356×256 we get $\log(n)^2 \simeq 123$, which makes the above result less attractive. If we assume on top that $C \geq 1$ and $s \simeq 0.2 \cdot n$, then m gets larger than n. However, in practice we may observe at least on simulations that a number of measurements $m \approx 2 \cdot s$ provides excellent reconstruction results [Adcock 2017]. The reason is related to the underlying structure of the signal, more precisely its structured sparsity.

Figure 2.1-1 illustrates the structured sparsity pattern of brain images decomposed within the Symmlet basis with $J = 3$ decomposition scales, each scale being composed of four subbands. Normally, the scale is inversely proportional to the spatial frequency. So, $j = 1$ refers to the high-frequency details and corresponds to the subbands depicted in the bottom row and top right corner of Figure 2.1-1, while $j = 3$ is associated with low-resolution information and contains both the low-frequency approximation as well as the three detail sub-bands located both in the top left corner of Figure 2.1-1.

Unfortunately, for many applications – including MRI – the sampling matrices cannot be constructed at random and the collected measurements are dependent one another. Therefore, concrete applications of CS are often based on sampling strategies that may strongly deviate from theory. Despite having no solid theoretical foundation, these heuristic strategies work very well in practice. Recently [Boyer 2017], derived CS theoretical results for structured acquisitions and structured signals (i.e. structured sparsity), with a recovery probability that explicitly depends on their support. The main contribution can be formulated informally as follows:

Let $x \in \mathbb{C}^n$ denote a vector with support $S \subset \{1, \ldots, n\}$, if we consider m-blocks of measurement drawing with a distribution $\pi \in \mathbb{R}^{m_p}$, with m_p being the number of available blocks. If $m \geq \Gamma(S, \pi) \ln\left(\frac{n}{\varepsilon}\right)$, with $\Gamma(S, \pi)$ being the maximum value between the inter-block coherence and the local coherence in the ideal case, i.e. when no block structure is



(a) Brain phantom **(b)** Wavelet representation

Figure 2.1-1: Wavelet decomposition of a brain phantom on 3 scales ($J = 3$). The low-frequency or higher scale of decomposition are less sparse than the lower or high-frequency scales. This illustrates the structured sparsity pattern of brain images. From [Lazarus 2018a].

assumed. Then vector $\boldsymbol{z}$ *can then be recovered by the* ℓ_1 *minimization with probability greater than* $1 - \varepsilon$.

Although the latest works by [Boyer 2017, Adcock 2018] derived theorems that guarantee exact reconstruction using block sampling sensing matrices in the ℓ_1-minimization and two major properties, *i.e. extra* and *intra* block coherence, few major drawbacks can be identified:

i) the evaluation of the extra and intra coherence can be hard;
ii) only blocks of parallel lines are considered;
iii) the design of the optimal sampling strategy is still an open issue, especially in non-Cartesian acquisitions;
iv) only the ℓ_1 minimization with orthonormal basis have been proposed, whereas we could expect an improvement of the reconstruction assuming more sophisticated norms or transforms.

Overall the existing theoretical results only provide some guidelines on the design of sampling schemes (*e.g.* variable density with a locally uniform coverage) and regarding reconstruction (nonlinear reconstructions promoting sparsity). In the following sections we focus on the application of CS to MRI by first considering the acquisition of k-space following the aforementioned considerations, before handling the reconstruction.

2.2 MR acquisition consideration

Compressed sensing (CS) theory has been applied to MRI by Lustig [Lustig 2007] shortly after its invention by Candès group [Candès 2006a] and Donoho [Donoho 2006]. Since the design of the sensing matrix is at the heart of CS theory, it is crucial to understand the constraints that govern it. Therefore, the first part of this section will be dedicated to the hardware constraints involved on magnetic field gradients for data acquisition. Next, we will review traditional methods to under-sample k-space according to CS recipes, before

introducing the SPARKLING algorithm to generate physically plausible and maximally efficient k-space trajectories [Lazarus 2017, Lazarus 2019a].

2.2.1 Gradients constraints

The k-space acquisition is governed by different gradients as it has been described in Section 1.2.1. Due to the hardware and physiological constraints – such as peripheral nerve stimulation – the coils used for the generation of those magnetic fields are bounded by their maximum current intensity in Eq. (2.12a) and their first-order derivative in Eq. (2.12b), also called *slew-rate* which are defined $\forall t \in [0, \mathrm{T_{obs}}]$ as follows:

$$\|G(t)\| \leq G_{max} \tag{2.12a}$$

$$\left\| \frac{\mathrm{d}G(t)}{\mathrm{d}t} \right\| \leq S_{max} \tag{2.12b}$$

where $\mathrm{T_{obs}}$ is the readout time. The norm used is either the ℓ_2 norm or the ℓ_∞ depending on the electronic configuration of the coil, *i.e.* if for each coil the current is generated either by the same generator or not. Since the trajectory is driven by the gradients following the relation defined in Eq. (1.14), the k-space constraints are derived $\forall t \in [0, \mathrm{T_{obs}}]$ as follows:

$$\left\| \frac{\mathrm{d}k(t)}{\mathrm{d}t} \right\| \leq \gamma\, G_{max} \tag{2.13a}$$

$$\left\| \frac{\mathrm{d}^2k(t)}{\mathrm{d}t^2} \right\| \leq \gamma\, S_{max} \tag{2.13b}$$

Eq. (2.13a) defines the maximum sampling speed which is given by $\gamma\, G_{max}$ while Eq. (2.13b) defines the maximum acceleration bounded by $\gamma\, S_{max}$. Since these constraints are gradient-specific, it might explain why most of the widely used trajectories are based on simple geometrical patterns such as lines (Cartesian or not) or spirals. Also, the more complex the trajectories, the more likely the actual curves played by the scanner may deviate from the originally prescribed ones (see [Vannesjo 2013, Vannesjo 2017]). In addition, the rapid decay of the MR signal ($\backsim$50 ms) usually prevents the measurement of all the needed data at once. For these reasons, k-space trajectories are generally composed of multiple segments, called *shots*. The latter sequentially fill the considered k-space, either on the grid for Cartesian acquisitions (spin-warp) or out of the grid (radial, spiral) for non-Cartesian sampling. Therefore the global acquisition time (TA) in segmented acquisitions is equal to the product of the number of shots (n_s) and the value of TR, the time of repetition: $\mathrm{TA} = n_s \cdot \mathrm{TR}$. At each and every TR, a radio-frequency (RF) pulse is used to tip the global magnetization for a given slice (2D imaging) or volume (3D imaging) and then a new shot is collected during $\mathrm{T_{obs}}$ [Bernstein 2004, Chap.11 2.2].

It is worth mentioning that single shot trajectories exist too – such as echo planar imaging (EPI) – however, those trajectories may induce distortions. Therefore their usage is limited to situations where fast scan is mandatory such as in functional or diffusion-weighted MRI. Consequently single shot EPI is never used for high-resolution anatomical imaging.

2.2.2 State-of-the art on accelerated trajectories

As explained before, owing to the physical constraints over the magnetic field gradients, it is easier to use simple geometrical patterns such as straight lines than more sophisticated curves. For that reason, first acceleration methods reduce the number of shots n_s. Hence for 2D imaging, 1D acceleration was achieved by subsampling the collected shots along the phase encoding direction, for instance using a 1D variable density centered around the central line ($k_y = 0$). In 3D imaging, shots were drawn on a 2D plane – defined by the phase and partition encoding directions – following a variable density according to the Poisson disk-sampling principle[2], while the readout (frequency encoding) was performed along the third orthogonal direction. However, the two major drawbacks of those strategy are:

i) the sampling efficiency – *i.e.* the k-space portion covered per unit of time – is not maximal;
ii) the optimization of the sampling lies on a lower-dimensional subspace, *i.e.* the 2D cross-section for 3D imaging. Hence, variable density is not implemented along the third dimension and consequently redundant information is collected.

Therefore to maximize sampling efficiency, non-Cartesian trajectories need to be considered. 2D variable density sampling can either be achieved thanks to radial or spiral patterns, with the maximum sampling efficiency achieved for spiral imaging.

The next paragraph summarizes the most popular approach to perform k-space undersampling, originally introduced in [Lustig 2007] in the MRI field, namely Poisson-disk sampling.

3D Poisson disk sampling

Poisson disk under-sampling basically consists first in drawing *independent* random points from a 2D sampling density typically with a radially decaying profile in order to perform variable density and promote low frequencies. Second, the k-space measurements are collected over orthogonal lines (readout) whose cross-section is actually given by the drawn points in 2D [Lustig 2007]. Hence, this strategy can only be implemented on 3D acquisitions as illustrated in Figure 2.2-2. Nonetheless this paradigm has been intensively used in retrospective 2D CS validations.Its advantage is threefold:

i) Poisson-disk actually achieves locally uniform coverage as the k-space samples stay far apart from each other, the distance between them being tuned by the disk radius;
ii) it actually fully samples the central region of k-space [Levine 2017, Vasanawala 2011];
iii) it avoids to play with complex sampling trajectories and thus to manipulate non-uniform or non-equispaced Fourier transforms.

However, Poisson-disk sampling is barely tested in 3D acquisition setups but rather in 2D simulations, *i.e.* on retrospective under-sampling scenarios. Instead, its performances are most often demonstrated in 2D imaging regardless of the MR acquisition specification.

[2]It is actually particular case of SPARKLING trajectories where we no longer activate the projection constraints over the magnetic field gradients.

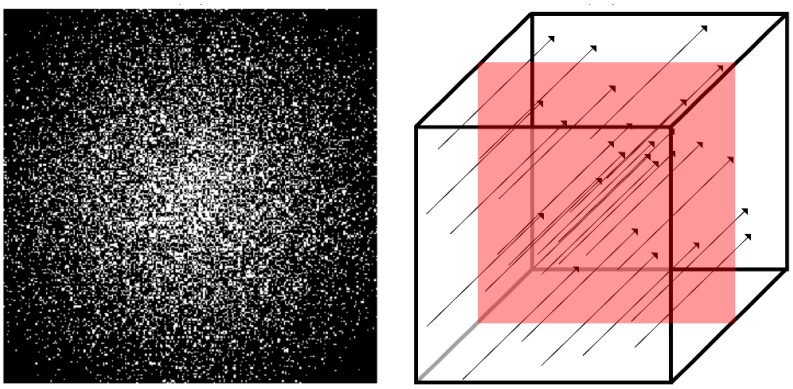


Figure 2.2-2: 3D acquisition with 2d Poisson disk sampling. From [Chauffert 2015]. The readout is performed along the horizontal lines wih black arrows, namely orthogonally to the 2d cross-section along which isolated points have been drawn.

2D variable density lines

2D acquisitions can also fulfill the requirements of CS theory, by adopting similar idea to those used in parallel imaging: reducing the number of shots – *i.e.* TR– therefore considering fewer lines in the phase encoding direction. However, in contrast to parallel MRI the lines are not drawn deterministically but rather randomly with a variable density distribution along the phase encoding direction, *i.e.* in that case only 1D VDS is implemented. An example of the reconstruction using variable density lines is presented Figure 2.2-3.

However, both 1D VDS on Cartesian lines and 3D Poisson disk sampling are hampered by the same drawback: they do not exploit the full dimensions of the ambient space. So their performances are limited in terms of acceleration factor. To go one step further, we summarize in the next part the non-Cartesian sampling strategies that fully implement VDS strategies in all dimensions of the native space, 2D or 3D VDS in respectively 2D and 3D imaging. The major difficulty lies in the ability to be compliant with the gradient constraints.

Spiral and Radial trajectories

Before the advent of Compressed Sensing theory and besides *parallel-imaging*, few trajectories were used to under-sample the k-space and thus accelerate scans. Among all k-space trajectories, radial patterns are the most widely and successfully used in MRI and bridge the gap with tomographic imaging. Radial sampling has a particular appeal for compressed sensing since the k-space center is naturally oversampled compared to higher frequencies. Moreover its repeated sampling of the k-space center results in a signal averaging in the image space making this trajectory robust to motion artifacts. For dynamic imaging purposes, radial trajectories have been stacked together to form a 3D stack of stars, which is known to be efficient for dynamic application (e.g. cardiac MRI) as it has been proved in [Chandarana 2014] for free-breathing abdominopelvic exams.

Under-sampled spiral trajectories were also used to accelerate the scan as they present

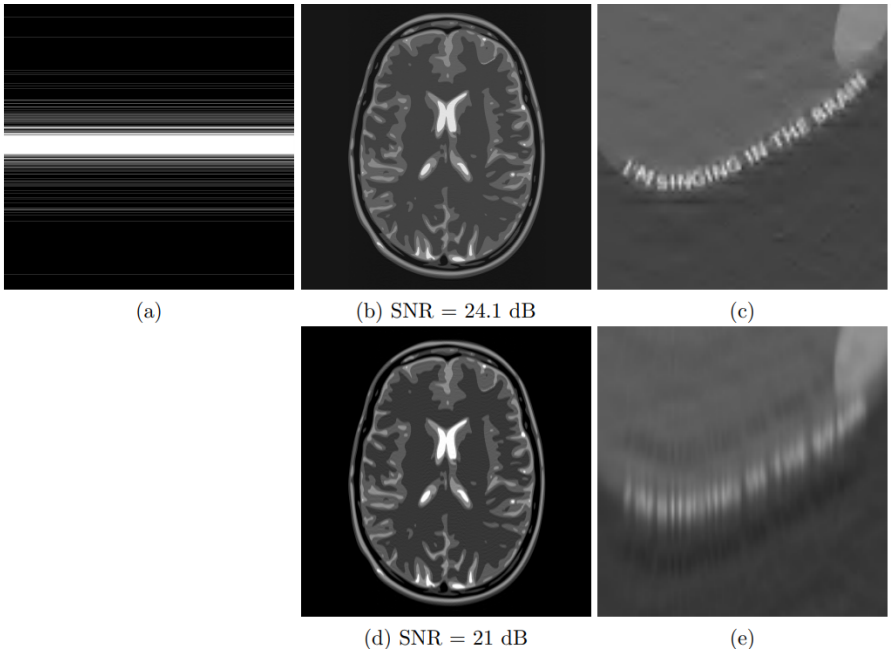
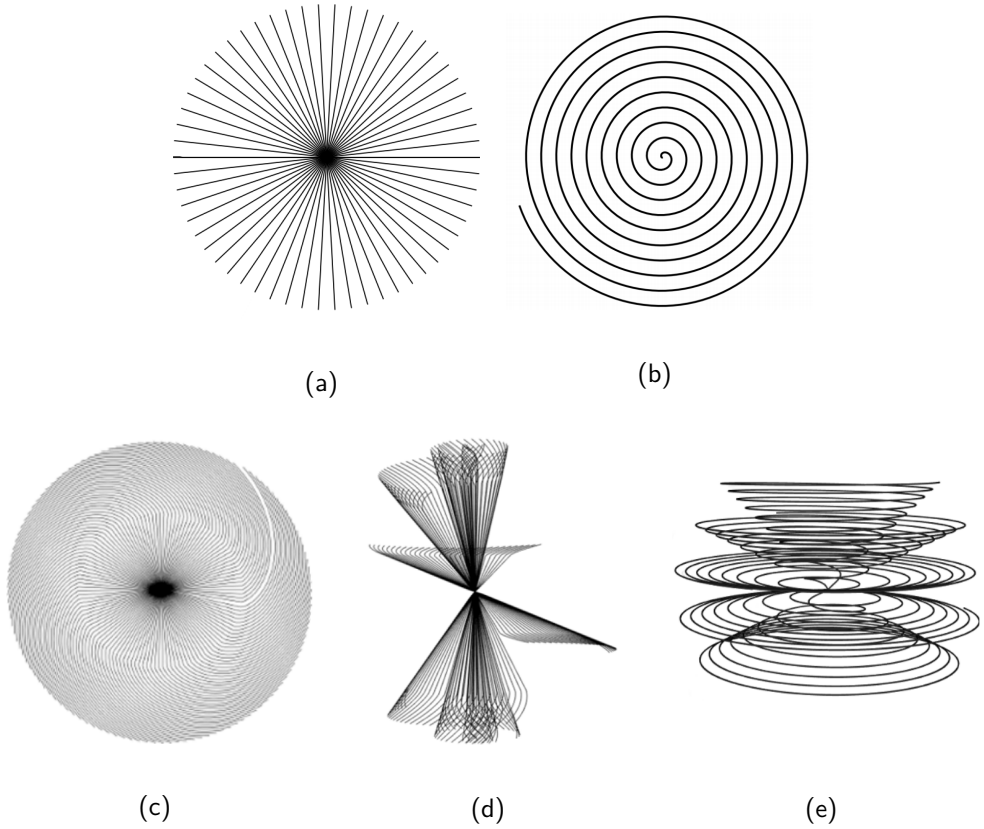


Figure 2.2-3: An example of reconstructions of a 2048×2048 MR image from block of measurements. (a) Sampling pattern with horizontal lines – *i.e.* variable density in the phase encoding direction – with an under-sampling factor $m/n \approx 8$. (b) Corresponding reconstruction using the ℓ_1-minimization. (c) A zoom on a part of the reconstructed image. (d) Image obtained by using the pseudo-inverse transform. (e) A zoom o a part of the image. From [Boyer 2017].

non-aliased artifacts and robustness to motion and flow. While spiral trajectories are known to take full advantage of the gradient hardware capabilities, two regimes are commonly defined: a slew-rate-limited and an amplitude-limited one [Delattre 2010]. Near the center of k-space, the trajectory is only limited by the maximal gradient slew rate $S_{\max}$, i.e. the curvature of the trajectory. Then, when moving to the outer part of k-space, the trajectory gets limited by the maximal gradient amplitude $G_{\max}$. Although classical spirals are Archimedean and keep constant distance between consecutive revolutions, variable density spirals are feasible [Lee 2003] to comply with CS theory [Lee 2003]. Radial [Lauterbur 1973] and spiral [Ahn 1986] trajectories have been originally proposed separately to speed up the acquisition, later on they have been elegantly combined yielding the TWisting Radial Lines (TWIRL) [Jackson 1992] sampling scheme. Further, for 3D imaging, an extension called the Twisted Projection Imaging (TPI) [Boada 1997] has been pushed forward to combine the advantages of high under-sampling of radial imaging in the center of k-space while spreading the points in the high-frequencies as spirals do. More recently, the Fermat Loop ORthogonaly Encoded Trajectory (FLORET) [Pipe 2011] strategy has been developed to provide more robustness to motion by covering more densely the center of k-space; see an illustration of these patterns in Figure 2.2-4.

Other optimization based method based on random perturbations have been proposed [El-Metwally 2008, Lustig 2005] to sample the k-space more efficiently.



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

Figure 2.2-4: Examples of some non-Cartesian trajectories used within CS theory to accelerate the scan time(a) Originally proposed by [Lauterbur 1973], Radial trajectory inherently follows a variable density. (b) Spiral trajectories are one of the most efficient curve as it can be implemented so as to take fully advantage of the encoding gradients (*i.e.* saturation of $S_{\max}$ in the k-space center and $G_{\max}$ in the periphery). (c) TWIRL: the first combination of radial and spiral trajectories for 2D acquisitions [Jackson 1991]. (d) TWIST: the 3D extension of the TWIRL method [Boada 1997]. (e) FLORET: recent advances on the design of 3D non-Cartesian sampling patterns based on the combination of radial and spiral.

2.2.3 Design of the SPARKLING trajectories

Recently a new method to design sampling patterns, called SPARKLING, has been proposed [Lazarus 2019a]. SPARKLING stands for Spreading Projection Algorithm for Rapid K-space samplING. Its objective is to yield multi-shot trajectories to efficiently undersample the k-space while being compliant with the gradient constraints, implementing 2D variable density in 2D imaging, performing locally uniform k-space coverage and eventually manipulating contrast affine constraints by specifying for instance the echo time TE, *i.e.* the k-space center crossing time.

The SPARKLING method relies on optimization to automatically generate k-space trajectories $\boldsymbol{k}(t)$[3] under the aforementioned hardware constraints on the gradients as defined in Eq. (2.12) – which can be expressed in terms of trajectory constraints as defined in Eq. (2.13) – by minimizing a tailored distance between the sample distribution and any prescribed density π. The objective becomes fitting the density π over the trajector-

[3]In 2D we get $\boldsymbol{k}(t) = [k_x(t), k_y(t)]$.

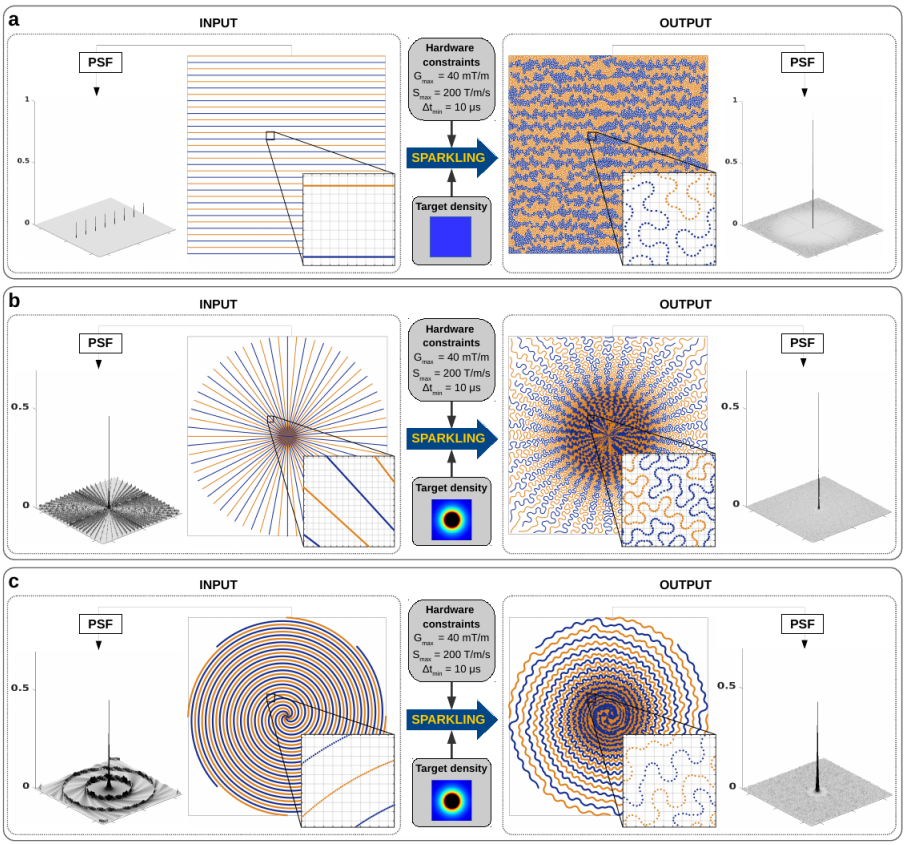


Figure 2.2-5: Generated SPARKLING trajectories for different initializations and target densities. Top tow: The SPARKLING method applied to Cartesian lines with uniform density. Mid row: The SPARKLING method applied to radial spokes with variable radial density. Bottom row: The SPARKLING method applied to the centered-out Archimedean spiral initialization with variable radial density. For each initialization the Point Spread Function (PSF) gives a rough idea on the artifact structure for the under-sampled factor. From [Lazarus 2018a].

ies. The projection algorithm behind the SPARKLING trajectory has originally been developed by [Boyer 2016, Chauffert 2017] and further extended in [Lazarus 2017] to handle echo time constraints. As the objective in SPARKLING is to minimize the distance between the target density π and the distribution over the trajectories $\boldsymbol{k}(t)$ under the convex and affine constraints involved on the gradients, the locations of the samples and the trajectories themselves are found simultaneously. Mathematically speaking, the following constrained optimization problem is solved:

$$\min_{k \in \mathcal{Q}_p} \operatorname{dist}(\pi, v(\boldsymbol{k})) = \min_{k \in \mathcal{Q}_p} \frac{1}{2} \| h \star (v(\boldsymbol{k}) - \pi) \|_2^2 \tag{2.14}$$

where h is an interpolation kernel, $v(k)$ is the probability of measure supported by the trajectory k and $\mathcal{Q}_p$ is the set of admissible curves respecting to the constraints defined in Eq. (2.13). An ℓ^2 distance has been chosen for the sake of simplicity. The discretization of Eq. (2.14) leads to the following formulation:

$$\min_{k\in\mathcal{Q}_p} \underbrace{\frac{1}{p^2}\sum_{1\leq i,j\leq p}\|\boldsymbol{k}[i]-\boldsymbol{k}[j]\|_2^2}_{F_r(k)} - \underbrace{\frac{1}{p}\sum_{i=1}^{p}\int_\Omega \|x-\boldsymbol{k}[i]\|_2^2\pi(x)dx}_{F_a(k)} \tag{2.15}$$

The latter formulation can be interpreted as the minimization of a potential energy, composed of a repulsive term F_r (to avoid clusters) and an attractive term F_a (to concentrate the samples according to the target density π). Due to the non-convexity of Eq. (2.15), this optimization is sensitive to the initialization as there as potentially many local minimizers. Figure 2.2-5 presents the outputs of SPARKLING algorithm for different initializations, in particular the Point Spread Function (PSF) gives an idea on the performances of the different sampling patterns. While for the Cartesian sampling presented in Figure 2.2-5 the spikes in the PSF predict the aliasing artifacts associated with under-sampling, the SPARKLING output is not affected, *i.e.* its PSF presents a pure central peak without any small replica.

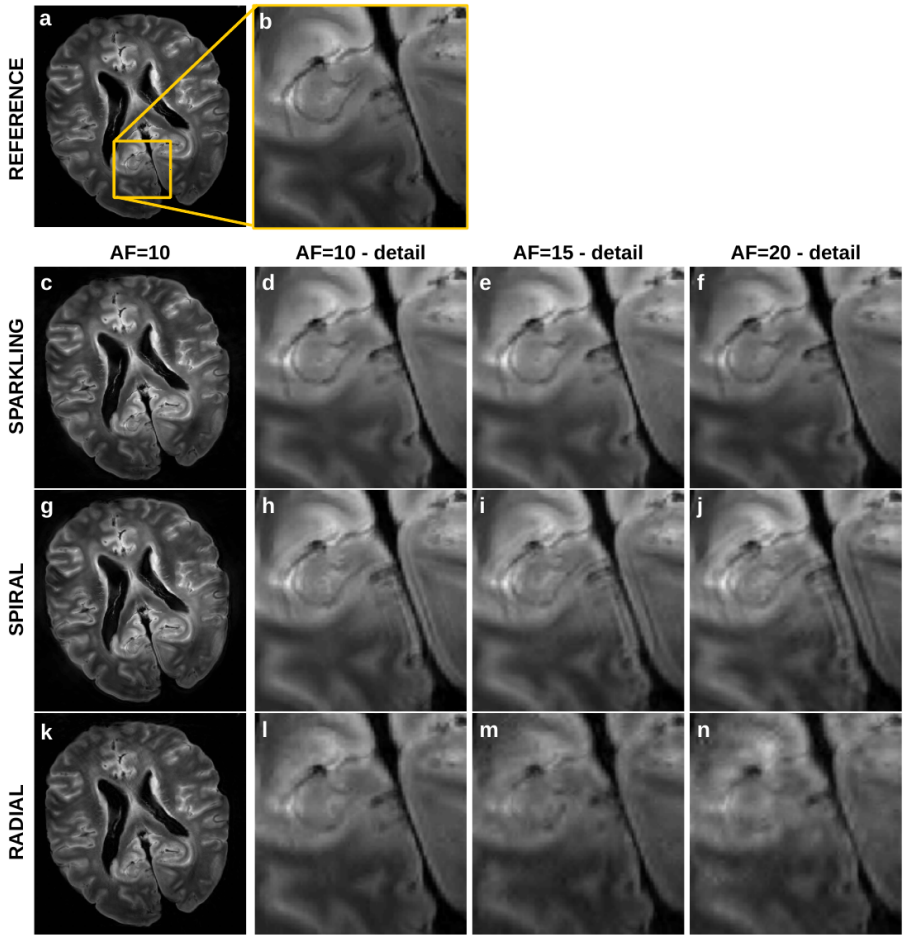


Figure 2.2-6: Comparison of different acquisition strategies on an T_2^* ex-vivo human brain acquired at 7 T for various acceleration factor $AF = 10, 15$ and 20. From [Lazarus 2018a].

The SPARKLING trajectories have also been implemented at 7 T for prospective acquisition within a gradient recalled echo (GRE) Fast Low Angle SHot (FLASH) sequence.

The method has been compared to variable density centered-in/out radial – used as initialization – and spiral [Lee 2003] imaging. The results of this comparison are presented in Figure 2.2-6. They indicate that SPARKLING trajectories maintain high image quality even for large acceleration factors, *i.e.* high similarity to the fully sampled Cartesian reference[4]. Due to the good performances of SPARKLING trajectories, the experimental setup used in the rest of this thesis will be the radially initialized SPARKLING trajectories.

2.3 Reconstruction

Due to the under-determination of the MR image reconstruction inverse problem from highly under-sampled data, there exists an infinite number of solutions. CS theory has provided the right framework to compute a unique solution to the inverse problem by imposing sparsity in the sought image. Originally, in the noise-free context, the image solution was computed by minimizing the ℓ_1 norm subject to the data consistency problem (see Eq. (2.3)). As we the data is noisy and due to its robustness to perturbation, the regularized version of reconstruction is often preferred, it reads as follows:

$$\widehat{\boldsymbol{z}} = \underset{\boldsymbol{z}\in\mathbb{C}^{N_\Psi}}{\operatorname{argmin}} \underbrace{\frac{1}{2}\|\boldsymbol{y} - \mathcal{F}_\Omega \boldsymbol{\Psi}_s \boldsymbol{z}\|_2^2}_{F(\boldsymbol{z})} + \underbrace{\lambda\|\boldsymbol{z}\|_1}_{R(\boldsymbol{z})} \tag{2.16}$$

where $\boldsymbol{y}$ is a vectorized version of the k-space measurements, $\boldsymbol{z}$ is the sparse representation in the dictionary $\boldsymbol{\Psi}_s$ of the image to be estimated, $\mathcal{F}_\Omega$ is the under-sampled Fourier operator over the binary mask or support Ω that picks up the collected measurements: it can be a single FFT for Cartesian under-sampling or the NFFT operator for non-Cartesian acquisitions. $\boldsymbol{\Psi}_s : \mathbb{C}^{N_\Psi} \rightarrow \mathbb{C}^N$ is the synthesis version of the sparsifying transform. Therefore, the problem formulated in Eq. (2.16) is also called the *synthesis* formulation as the estimated image is given by $\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}_s \widehat{\boldsymbol{z}}$. However, one can also introduce an *analysis* formulation that computes the solution directly in the image space as follows:

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{C}^{N}}{\operatorname{argmin}} \underbrace{\frac{1}{2}\|\boldsymbol{y} - \mathcal{F}_\Omega \boldsymbol{x}\|_2^2}_{F(\boldsymbol{x})} + \underbrace{\lambda\|\boldsymbol{\Psi}_a \boldsymbol{x}\|_1}_{R(\boldsymbol{x})} \tag{2.17}$$

where $\boldsymbol{x}$ corresponds to the image and $\boldsymbol{\Psi}_a : \mathbb{C}^N \rightarrow \mathbb{C}^{N_\Psi}$ the analysis version of the sparsifying transform. The difference between the synthesis and the analysis formulations (Eq. (2.16) *i.e.* vs Eq. (2.17)) will be discussed in the following section. Moreover it should be mentioned that their adjoint operator is defined as the conjugate-transpose $\boldsymbol{\Psi}_s^{\mathrm{H}}$ for the synthesis transform and $\boldsymbol{\Psi}_a^{\mathrm{H}}$ for the analysis one.

Whatever the retained formulation, the reconstruction problem relies on two main ingredients that will be discussed in depth in this part:

i) the sparsifying transform $\boldsymbol{\Psi}$, which should be chosen as incoherent as possible with respect to the acquisition basis;

[4] An extensive comparison including several image contrasts can be found in [Lazarus 2018a].

ii) the optimizer – solver – that affects the convergence speed, hence the computation time.

In what follows, these two points will be discussed in detail while the last subsection will be dedicated to the definition of quantitative image quality metrics. Indeed, as this work focuses on image reconstruction, it gets critical to choose an appropriate quantitative metric to assess the performances of the reconstruction algorithms proposed along this work.

2.3.1 Sparse decomposition

One of the key tools in CS theory is the incoherence between the sparse representation and the acquisition basis. However, the natural MR images are not sparse in the canonical basis and need therefore to be sparsified through a decomposition that can either have an analytical fixed formulation – such as Wavelet Transform – or learned from the data itself such as in the Transform or Dictionary Learning (TL vs DL) [Ravishankar 2010, Ravishankar 2015, Wen 2019] framework. In what follows, we will study the impact of the *synthesis* or *analysis* formulation on fixed dictionaries, as well as the choice of complete vs over-complete transforms. In a second step we explore the performances of the DL on the reconstructed images.

Analytical fixed sparsifying dictionaries

A basic known result states that if $\boldsymbol{\Psi}$ is an orthonormal transform – *i.e.* $\boldsymbol{\Psi}^{\mathrm{H}}\boldsymbol{\Psi} = \boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{H}} = \boldsymbol{I}_N$ and $\boldsymbol{\Psi}_a\boldsymbol{\Psi}_s = \boldsymbol{I}_{N_\psi}$ with necessary $N_\psi = N$ – then both analysis and synthesis formulations Eqs. (2.16) and (2.17) are equivalent. The same property holds more generally for invertible transforms $\boldsymbol{\Psi}$. However, in other situations, the two formulations provide different solutions [Elad 2007]. Noticeably, for over-complete transforms – *i.e.* a transform that maps the vectorized image $\boldsymbol{x} \in \mathbb{C}^N$ to a coefficient vector $\boldsymbol{z} \in \mathbb{C}^{N_{\boldsymbol{\Psi}}}$ with $N < N_{\boldsymbol{\Psi}}$ – the analysis formulation is preferred as it has been shown in [Majumdar 2012b, Cherkaoui 2018, Selesnick 2009]. The choice of an appropriate fixed decomposition has been intensively studied in the literature [Selesnick 2009, Majumdar 2012b, Cherkaoui 2018]. Importantly, it has been proved that over-complete dictionaries such as curvelets [Starck 2002], shearlets [Guo 2007] or undecimated Bi-orthogonal WT [Starck 2007] combined with the analysis formulation outperform the standard orthogonal wavelet transform. However owing to the redundant information manipulated in that case – $N < N_{\boldsymbol{\Psi}}$ – optimization is performed in the image space to save memory usage and limit the computational load, especially in the high-resolutions context.

Adaptive data driven representations

While analytical over-complete dictionaries already provide good results, one could expect an improvement on the image quality using learned dictionaries. DL methods [Ravishankar 2010, Ravishankar 2015] extract and learn atoms which represent the image

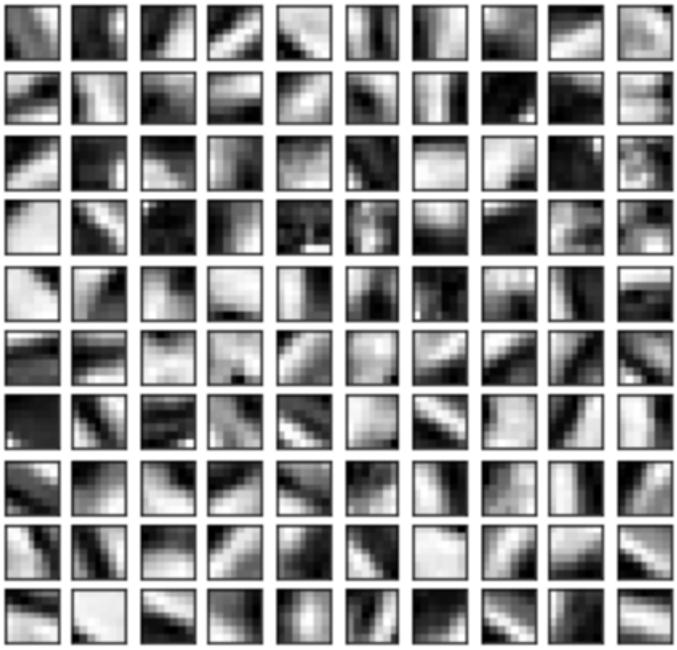


Figure 2.3-7: Set of atoms learned from the CamCan database [Taylor 2017b] where only magnitude information has been learned. From our work published in [Carrié 2018].

to be reconstructed with few non-zero coefficients. Two approaches have been developed so far: either the dictionary is learned off-line and then injected as a sparsifying transform or it can be directly inferred in a blind bilinear reconstruction problem where both the reconstruction and the dictionary estimation steps are interleaved [Ravishankar 2010, Ravishankar 2015].

The two step procedure can be described as follows: first, the dictionary learning step can be formulated as in Eq. (2.18a) then the reconstruction step Eq. (2.18b) is defined using the previously learned dictionary.

$$(\hat{\boldsymbol{D}}, \hat{\boldsymbol{z}}) = \underset{\substack{\boldsymbol{D}\in\mathbb{C}^{N\times n_c}, \boldsymbol{z}\in\mathbb{C}^{n_c} \\ \forall k\leq n_c,\ \|\boldsymbol{D}_k\|_2=1}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}\|_2^2 + \lambda\|\boldsymbol{z}\|_1 \tag{2.18a}$$

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{C}^N}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{y} - \mathcal{F}_\Omega\boldsymbol{x}\|_2^2 + \lambda\|\hat{\boldsymbol{D}}\boldsymbol{x}\|_1 \tag{2.18b}$$

where n_c is the number of components used to represent the vectorized image $\boldsymbol{x} \in \mathbb{C}^n$ [Ravishankar 2010, Mairal 2010], the learned dictionary $\hat{\boldsymbol{D}}$ is composed of atoms that are illustrated in Figure 2.3-7. However these approaches either require a large database of images for learning atoms or are computationally slow. Another drawback concerns the reconstruction time [Carrié 2018], since at each iteration, the image needs to be decomposed in the dictionary, hence for non-Cartesian acquisitions like SPARKLING – *i.e.* using NFFT operator in Eq. (2.18b) – and single channel acquisition setting, the reconstruction takes approximately 30min against 3min with the wavelet decomposition for $N = 512 \times 512$.

2.3.2 Proximal optimization method

In general CS-based reconstruction – either based on the analysis Eq. (2.17) or synthesis Eq. (2.16) model – is composed of a sum of data-consistency term – denoted by $F(\cdot)$ – and regularization – denoted by $R(\cdot)$ – which promotes sparsity through nonsmooth penalty. In this thesis we mainly considered convex methods as it guarantees convergence to the global minimizer of a convex but nonsmooth objective function $(F + R)(\cdot)$ [Combettes 2011]. Hereafter we define the notation and the class of algorithms we have used to solve the reconstruction problem.

Definitions

We denote by $\mathbb{C}^N$ the usual N-dimensional complex Hilbert space. Standard definitions and notations from convex analysis will be assumed [Rockafellar 1970][5]. The domain of a function $f : \mathbb{C}^N \to]-\infty, +\infty]$ is denoted by $\mathrm{dom} f$ and defined such as $\mathrm{dom} f = \left\{\boldsymbol{x} \in \mathbb{C}^N | f(x) < +\infty\right\}$. A convex function is defined as follows:

Definition 2.3.1. *A function $f : \mathbb{C}^N \to \mathbb{R}$ is* convex *if $domf$ is a convex set and if $\forall \boldsymbol{x}, \boldsymbol{y}, \in domf$ and $\theta \in \mathbb{R} \mid 0 \leq \theta \leq 1$, we have*

$$f(\theta\ \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})\,.$$

Let $\Gamma_0(\mathbb{C}^N)$ be the class of lower semi-continuous convex functions from $\mathbb{C}^N \to]-\infty, +\infty]$. The convex conjugate of the function f denoted by $f^\star \in \Gamma_0(\mathbb{C}^N)$, is defined as follows:

$$\begin{aligned} f^\star : \mathbb{C}^N &\to]-\infty, +\infty] \\ \boldsymbol{u} &\mapsto \sup_{\boldsymbol{x} \in \mathbb{C}^N} \{\langle \boldsymbol{x}, \boldsymbol{u} \rangle - f(\boldsymbol{x})\} \end{aligned} \tag{2.19}$$

The Moreau *subdifferential* is defined as follows:

Definition 2.3.2. *The Moreau subdifferential of a function $f : \mathbb{C}^N \to]-\infty, +\infty[$ at $\boldsymbol{x} \in \mathbb{C}^N$ is defined as :*

$$\begin{aligned} \partial f : \mathbb{C}^N &\to 2^{\mathbb{C}^N} \\ \boldsymbol{x} &\mapsto \left\{\boldsymbol{u} \in \mathbb{C}^N \mid (\forall \boldsymbol{y} \in \mathbb{C}^N)\ f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{u} \rangle\right\} \end{aligned}$$

Any vector $\boldsymbol{u}$ in $\partial f(\boldsymbol{x})$ is called subgradient of f at $\boldsymbol{x}$.

We define the proximity operator as follows:

Definition 2.3.3. *Let $f \in \Gamma_0(\mathbb{C})^N$. For every $\boldsymbol{x} \in \mathbb{C}^N$, the minimization problem:*

$$\min_{\boldsymbol{y} \in \mathbb{C}^N} \left\{ f(\boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2 \right\} \tag{2.20}$$

admits a unique solution called proximity operator of f denoted by $\mathrm{prox}_f(x) : \mathbb{C}^N \to \mathbb{C}^N$.

[5] On top of the notations defined page xv.

Moreover the proximity operator of f can be characterized by the following inclusion:

$$\left(\forall(\boldsymbol{x},\boldsymbol{p}) \in \mathbb{C}^N \times \mathbb{C}^N\right) \quad \boldsymbol{p} = \text{prox}_f(\boldsymbol{x}) \quad \Leftrightarrow \quad \boldsymbol{x} - \boldsymbol{p} \in \partial f(\boldsymbol{p}) \tag{2.21}$$

We also define a β-Lipschitz gradient of a function f as:

$$\forall(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{C}^{N\times N} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2 \leq \beta \|\boldsymbol{x} - \boldsymbol{y}\|_2 \tag{2.22}$$

Since the data-fidelity term ($F(\cdot)$ as defined in Eqs. (2.16) and (2.17)) is smooth and has a β-Lipschitz gradient and the regularization term ($R(\cdot)$ as defined in Eqs. (2.16) and (2.17)) is proper (*i.e.* $\text{dom}R \neq \emptyset$) closed convex function, the reconstruction problem can be efficiently solved using proximal splitting methods that are described in the next part.

Proximal majorize minimize method: Forward-Backward Splitting method

A classical approach to solve Eq. (2.16) is to use the Forward-Backward Splitting algorithm (FBS) [Levitin 1966, Combettes 2011]. It can be seen from the Taylor expansion of $F(\boldsymbol{y})$ and simple linear algebra that the iterations sequence to solve Eq. (2.16) takes the form:

$$\boldsymbol{z}_{k+1} := \underbrace{\text{prox}_{\gamma_k R}}_{\text{backward step}} \underbrace{(\boldsymbol{z}_k - \gamma_k \nabla F(\boldsymbol{z}_k))}_{\text{forward step}}, \quad \text{with} \quad \gamma_k \in \left]0, 2/\beta\right[\tag{2.23}$$

with $\boldsymbol{z}_0 \in \mathbb{C}^N$. In the particular case where the ℓ_1-norm is used for R the algorithm correspond to the Iterative Soft-Thresholding Algorithm (ISTA) [Daubechies 2004]. Moreover, it is well established that FBS iterates converge as long as $\gamma_k \in \left]0, 2/\beta\right[$. However the convergence speed is quite slow (with a rate in the order of $\mathcal{O}(1/k)$). In the literature, two complementary paths were developed to speed up convergence, namely adaptive restart [O'donoghue 2015] and the use of a momentum strategies [Nesterov 1983], both being discussed hereafter. In order to speed the convergence, many variants of Eq. (2.23) have been proposed in the literature, most of them rely on the setting of a momentum term [Nesterov 1983] that can be summarized as follows:

Algorithm 1: Accelerated Forward Backward

Set $\boldsymbol{z}_0 \in \mathbb{C}^{N_\Psi}$, $\boldsymbol{w}_0 = \boldsymbol{z}_0$, $\theta_0 = 1$, $\gamma_k \in \left]0, \beta\right]$ and $q \in [0,1]$;
for $k = 0, \ldots$ **do**
  θ_{k+1} solves $\theta_{k+1}^2 = (1-\theta_{k+1})\,\theta_k^2 + q\theta_{k+1}$;
  $\alpha_{k+1} = \theta_k\,(1-\theta_k) \,/\, (\theta_k^2 + \theta_{k+1})$;
  $\boldsymbol{w}_{k+1} = \boldsymbol{z}_{k+1} + \alpha_{k+1}\,(\boldsymbol{z}_{k+1} - \boldsymbol{z}_k)$;
  $\boldsymbol{z}_{k+1} := \text{prox}_{\gamma_k R}\,(\boldsymbol{w}_{k+1} - \gamma_k \nabla F\,(\boldsymbol{w}_{k+1}))$;
end

One of the most popular techniques associated with Nesterov acceleration scheme is the Fast Iterative Soft-Thresholding Algorithm (FISTA) [Beck 2009] which has a convergence

rate of $\mathcal{O}\left(1/k^2\right)$. The convergence proof of the iterates of FISTA haven't been demonstrated yet, however [Chambolle 2015] proposed a variant of FISTA with a convergence proof of the iterates. On drawback of the Nesterov acceleration schemes is the presence of ripples in the trace of the objective function which slows down the convergence as illustrated in Figure 2.3-8. In [O'donoghue 2015] the authors exhibit two regimes depending on the value of the momentum and noticed that the ripples arise when the momentum term was too high. Therefore, by simply restarting the momentum term – *i.e.* setiting it to zero – they could accelerate the different algorithms. The methods proposed by [O'donoghue 2015] is a very general and could potentially be applied to different optimization algorithms that present the same behavior.

Until now, in Eq. (2.23) we have only considered the synthesis formulation of the reconstruction problem as defined in Eq. (2.16). The analysis formulation can be handled similarly when Ψ_a is an orthogonal sparsifying transform as the proximity operator could be calculated explicitly [Combettes 2011, Table 10.1-x], for more general decompositions this is no longer the case. Hence, the proximity operator has to be computed iteratively which slows down the reconstruction again. In these specific cases, to avoid the iterative calculation of proximity operators, primal-dual methods have been developed.

Proximal optimization: Primal-Dual algorithm

We have previously seen that the use of over-complete dictionaries could noticeably improve image reconstruction accuracy. However, solving it with one of the FBS variants is time consuming. Hopefully, the analysis problem Eq. (2.17) can be solved by minimizing both the primal and dual formulation as suggested by [Komodakis 2015][6]. The dual problem can be formulated as follows:

$$\widehat{\boldsymbol{z}} \in \underset{\boldsymbol{z}\in\mathbb{C}^{N_{\boldsymbol{\Psi}}}}{\operatorname{argmin}}\ F^{\star}(-\boldsymbol{\Psi}^{\mathrm{H}}\boldsymbol{z}) + R^{\star}(\boldsymbol{z}) \tag{2.24}$$

where $F^{\star}$ and $R^{\star}$ are the convex conjugate of F and G, respectively. It should be noted that the primal and dual problems can be reformulated so as to be solved together, hence the Lagrangian saddle point is found [Bauschke 2011, Chap. 19]:

$$(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{z}}) \in \underset{\boldsymbol{x}\in\mathbb{C}^{N}}{\operatorname{argmin}} \left(\underset{\boldsymbol{z}\in\mathbb{C}^{N_{\boldsymbol{\Psi}}}}{\operatorname{argmax}}\ F(\boldsymbol{x}) - R^{\star}(\boldsymbol{z}) + \langle \boldsymbol{\Psi}\boldsymbol{x}, \boldsymbol{z} \rangle \right) \tag{2.25}$$

Moreover if we suppose that :

i) Eq. (2.17) admits at least one solution

ii) $\operatorname{ri}(\operatorname{dom} R) \cap \boldsymbol{\Psi}\ \operatorname{dom} F \neq \emptyset$, with ri the restricted interior;

then the dual problem admits a solution and at convergence, the dual gap is null meaning that:

[6] In this case we only considered the sum of smooth and non-smooth functions however the primal-dual approaches are more general and could be applied to any problem defined as follow: $F(\boldsymbol{x}) + G(\boldsymbol{x}) + R(\boldsymbol{\Psi}\boldsymbol{x})$, where F, R and $\boldsymbol{\Psi}$ are defined as before and G is is proper closed convex function.

$$F(\widehat{\boldsymbol{x}}) + R(\boldsymbol{\Psi}\widehat{\boldsymbol{x}}) = -F^{\star}(-\boldsymbol{\Psi}^{\mathrm{H}}\widehat{\boldsymbol{z}}) - R^{\star}(\widehat{\boldsymbol{z}}) \tag{2.26}$$

The first approach developed to solve the primal-dual problem was the Alternating Direction Method of Multiplier (ADMM) [Komodakis 2015, Boyd 2004, Gabay 1975]. Although it has been widely used in various fields, the setting of its parameters may be tricky. It should also be mentioned that a variant of ADMM in the dual space has been proposed in [Eckstein 1992].

Another primal-dual algorithm has been developed simultaneously by [Condat 2013] and [Vũ 2013]. The sequence of its iterates for solving Eq. (2.17) can be summarized as follows:

Algorithm 2: Primal-Dual Condat-Vũ Algorithm

1 Set $\boldsymbol{z}_0 \in \mathbb{C}^N$, $\boldsymbol{z}_0 \in \mathbb{C}^{N_{\boldsymbol{\Psi}}}$, $(\sigma, \tau) \in \left]0, +\infty\right[^2$, and $\forall k \in \mathbb{N},\ \lambda_k \in \left]0, +\infty\right[$;
2 **for** $k = 0, \ldots$ **do**
3 $\quad \boldsymbol{u}_k := \boldsymbol{x}_k - \tau(\nabla F(\boldsymbol{x}_k) + \boldsymbol{\Psi}^{\mathrm{H}}\boldsymbol{z}_k + e_{F,k})$;
4 $\quad \boldsymbol{w}_k := \mathrm{prox}_{\sigma R^{\star}}\left(\boldsymbol{z}_k + \sigma\boldsymbol{\Psi}(2\boldsymbol{u}_k - \boldsymbol{x}_k)\right) + e_{R,k}$;
5 $\quad \boldsymbol{x}_{k+1} := \lambda_k \boldsymbol{u}_k + (1 - \lambda_k)\,\boldsymbol{x}_k$;
6 $\quad \boldsymbol{z}_{k+1} := \lambda_k \boldsymbol{w}_k + (1 - \lambda_k)\,\boldsymbol{z}_k$;
7 **end**

A symmetric version of the Condat-Vũ algorithm [7] has also been proposed and consists in switching Lines 2. and 3. in Algorithm 2. Moreover, the convergence proof of the iterates has been established by the following theorem:

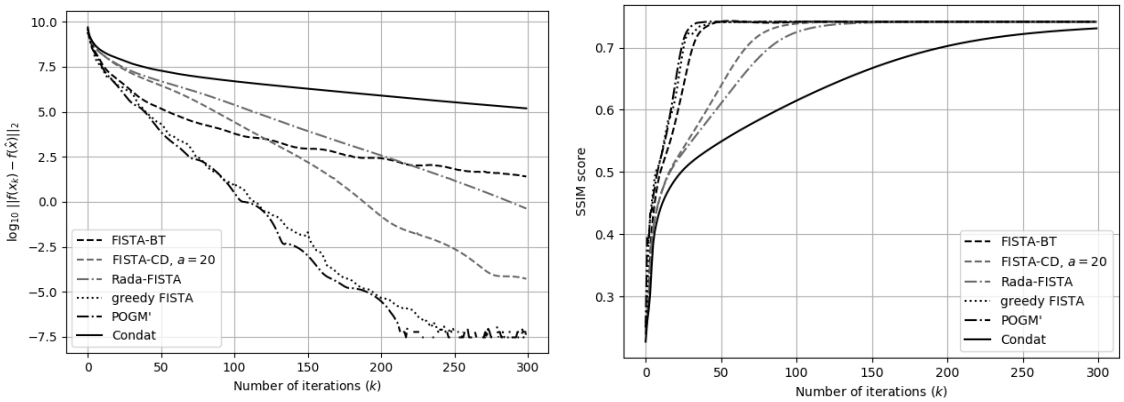
Theorem 2.3.4. *[Condat 2013, Theorem 3.1] Let $\gamma > 0$, $\tau > 0$ and the sequences $(\lambda_k)_{\forall k \in \mathbb{N}}$, $(e_{F,k})_{\forall k \in \mathbb{N}}$ and $(e_{R,k})_{\forall k \in \mathbb{N}}$ be the parameters of Algorithm 2. Let $\beta > 0$ be the Lipschitz constant of the gradient of F, and suppose that the following holds:*

i) $\frac{1}{\tau} - \sigma\|\boldsymbol{\Psi}\|_2 \geq \frac{\beta}{2}$
ii) $\forall k \in \mathbb{N}, \lambda_k \in \left]0, \delta\right[$, *where we set* $\delta := 2 - \frac{\beta}{2}\left(\frac{1}{\tau} - \sigma\|\boldsymbol{\Psi}\|^2\right)^{-1} \in [1, 2]$
iii) $\sum_{k \in \mathbb{N}} \lambda_k\,(\delta - \lambda_k) = +\infty$
iv) $\sum_{k \in \mathbb{N}} \lambda_k\,\|e_{F,k}\| < +\infty \quad$ *and* $\quad \sum_{k \in \mathbb{N}} \lambda_k\,\|e_{R,k}\| < +\infty$

Then there exists a pair $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{z}}) \in \mathbb{C}^N \times \mathbb{C}^{N_{\boldsymbol{\Psi}}}$ solution to Eq. (2.25), such that the Algorithm 2 (and its symmetric version), the sequences $(\boldsymbol{x}_k)_{\forall k \in \mathbb{N}}$ and $(\boldsymbol{z}_k)_{\forall k \in \mathbb{N}_{\boldsymbol{\Psi}}}$ converge weakly to respectively $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{z}}$.

Recently a benchmark has been proposed by [Ramzi 2019] to assess the performances of the different algorithms, where $\boldsymbol{\Psi}$ was an orthogonal wavelet transform. The benchmark included FISTA [Beck 2009] denoted by FISTA-BT, the convergent version of FISTA proposed by [Chambolle 2015] denoted by FISTA-CD, the Restart and Adaptive α-FISTA (Rada-FISTA) and the greedy-FISTA proposed in [Liang 2019]. Besides all these versions

[7] When F, and the error terms $(e_{F,k})_{\forall k \in \mathbb{N}}$, $(e_{R,k})_{\forall k \in \mathbb{N}}$ is null the Condat-Vũ sequence corresponds to the Chambolle-Pock algorithm [Chambolle 2011, Boyer 2012].



(a) (b)

Figure 2.3-8: Comparison of the convergence speed of different proximal algorithms to solve Eq. (2.17). The Nesterov ripples can be seen from the FISTA-CD proposed by [Chambolle 2015]. From [Ramzi 2019].

of FISTA, the benchmark also comprised other acceleration schemes such as the adaptive restart strategy of Proximal-Optimized Gradient Method (POGM') [Taylor 2017a, Kim 2017a, Kim 2018a]. Finally, the comparison included the Condat-Vũ algorithm too. Figure 2.3-8 **(a)** presents the evolution of the objective function for the different algorithms while Figure 2.3-8 **(a)** displays the Structural Similarity index (SSIM) evolution over iterations. The results of [Ramzi 2019] benchmark show that the greedy-FISTA and POGM' were comparable, however the latter was twice more memory demanding which could potentially be critical for high-resolution 3D reconstruction in the parallel imaging setup.

2.3.3 Metric for Image comparison

As the quantification of the MR image quality is important for the evaluation of imaging techniques both at the acquisition and reconstruction levels, this part shortly reviews some standard metrics to assess image quality when a ground truth or reference image is available.

Mean Square Error (MSE) and Structural Similarity Index (SSIM) are the most frequently used indices for assessing image quality between a reference and a reconstructed image. The implicit assumption is that they correlate well with the physicians assessment. In [Zanforlin 2014], the authors have compared the SSIM score with the mean observation error (MOS). Interestingly they drew a comparison between the similarity score and the MOS and proposed a table that summarizes the correlation between SSIM and the visual perception, see Table 2.1,

However the results stated by [Zanforlin 2014] were considered for natural images and not for medical ones. Recently, [Mason 2019] has assessed the correlation between various quality metrics and physician' views. To do so, the experimental protocol was defined as follows: nine brain scans selected from hospital images and corrupted versions

Table 2.1: Correspondence between SSIM score ans Mean Opinion Scores (MOS) between an image and its reference.

SSIM	MOS	Quality	Impairments
≥ 0.99	5	Excellent	Imperceptible
$[0.95, 0.99($	4	Good	Perceptible but not annoying
$[0.88, 0.95($	3	Fair	Slightly annoying
$[0.5, 0.88($	2	Poor	Annoying
≤ 0.5	1	Bad	Very annoying

with either Gaussian blurring, wavelet compression, Rician noise [Gudbjartsson 1995] or motion artifacts were presented to two neuroradiologists who carefully read and rank the images. The physicians were asked to score overall image quality on a 1-5 Likert scale that was determined by consensus on a training set before scoring the images included in the results. Correlations between each image quality metrics and the physician' scores were measured using the Pearson linear correlation coefficients(PLCC) and the Spearman rank order correlation coefficients (SROCC). According to [Mason 2019] the best quality metric should be linear with respect to the Likert index of the physician. Fortunately, for the SSIM score above 0.9 this was the case. Hence in the rest of the thesis, we will stick to the SSIM score on the regime where it fits the physician perception *i.e.* SSIM > 0.9.

2.3.4 Conclusion

As seen in the previous chapters, MRI is a powerful non-invasive technique that suffers from long scan time. Although methods such as parallel-imaging permit to shorten this time, they acceleration power remains too limited for high-resolution MRI. The very reason for that is twofold: one one hand, under-sampling is performed in a deterministic manner with a target uniform sampling density. On the other hand, most often the reconstruction itself is purely linear. To go one step further, the application of CS theory to MRI has revolutionized the way data can be collected and images are reconstructed. CS provides higher acceleration factors compared to parallel imaging or partial Fourier techniques but at the cost of more computationally demanding reconstruction. However, to become even more efficient regarding the way SNR is preserved, CS has also been combined with multi-channel acquisition over phased array coils. Then a natural question has emerged on the optimal manner to combine CS and multi-channel acquisition for making image reconstruction more efficient and accurate. These issues will be discussed in the next chapters.

* * *
* *
*

Part II

Methodological Developments

Chapter 3

Compressed Sensing reconstruction from multi-channel acquisition

Chapter Outline

The method developed in this chapter (Section 3.2.1) was used as reconstruction technique in the following works :

C. Lazarus, P. Weiss, N. Chauffert, F. Mauconduit, L. El Gueddari, C. Destrieux, I. Zemmoura, A. Vignaud and P. Ciuciu. *SPARKLING: variable-density k-space filling curves for accelerated T_2^*-weighted MRI.* Magnetic resonance in medicine, vol. 81, pages 3643–3661, 2019.

C. Lazarus, P. Weiss, L. El Gueddari, F. Mauconduit, C. Destrieux, I. Zemmoura, A. Vignaud and P. Ciuciu. *3D SPARKLING trajectories for high-resolution* T_2^**-weighted Magnetic Resonance imaging.* NMR biomedicine, 2019. *Under review.*

Parts of this chapter (Section 3.2.1) were presented at International conferences as:

L. El Gueddari, C. Lazarus, H. Carrié, A. Vignaud and P. Ciuciu. *Self-Calibrating Nonlinear Reconstruction Algorithms for Variable Density Sampling and Parallel Reception MRI.* In 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), pages 415–419, July 2018.

L. El Gueddari, C. Lazarus, H. Carrié, A. Vignaud and P. Ciuciu. *Self-calibrating nonlinear MR image reconstruction algorithms for variable density sampling and parallel imaging.* In Proceedings of the 27th Annual Meeting of ISMRM, page 3547, Paris, France, 2018.

This chapter introduces and discusses the MR image reconstruction methods based on multi-channel array coil acquisition. A particular attention will be paid to their compatibility with non-Cartesian sampling schemes.

3.1 Introduction

Compressed Sensing provides a way to drastically reduce the scan time, however the maximum under-sampling factor depends on the available input-SNR and the image resolution[1]. An empirical study on the maximum under-sampling factor for T_2^* MR images [Lazarus 2018c] illustrates several important results, for a targeted image quality:

- at a given resolution, the under-sampling factor can be increased as the input-SNR improves until a stationary regime is reached for high input-SNR.
- high resolution images offers the possibility to reach high under-sampling factors as long as the input-SNR remains sufficient, while the under-sampling ratio is limited for low-resolution images.
- at each resolution, an optimal point, which combines the minimum requisite input SNR with the maximum under-sampling factor allowing to maintain the targeted image quality (*e.g.* SSIM score of 0.9), can be determined.

Figure 3.1-1 presents those relations in a slightly different way: the 3D map provides the evolution of the maximum under-sampling factor achievable for a range of image resolution and input-SNR which preserves a fair image quality, corresponding to an SSIM score above 0.9 according to the mean opinion score (Table 2.1). The highlighted red line corresponds to the available input-SNR on the Magnetom 7 T scanner (Siemens-Healthineers, Erlangen, Germany) using the available single-channel coil.

To achieve high acceleration factors in the high resolution setting, the input SNR has to be boosted. This can be reached using a multi-channel coil array, as demonstrated in [Roemer 1990] and presented Section 1.3.1. However such reception coil will increase the reconstruction complexity and raise questions of Compressed Sensing theory in such a configuration. This chapter will review CS reconstruction methods for multi-channel acquisitions. We will focus on the non-Cartesian case since the latter permits higher ac-

[1] or image dimension as we assume the FOV fixed.

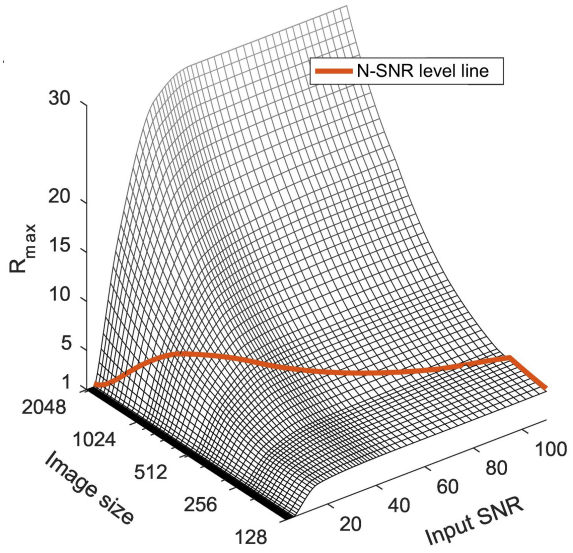


Figure 3.1-1: A view of the maximum under-sampling factor allowing SSIM scores above 0.9 as a function of image size (or resolution as the FOV is fixed) and input SNR. Orange line: Available input-SNR on the 7 T scanner (Siemens-Healthineers, Erlangen, Germany). [Lazarus 2018c].

celeration factors. Those reconstruction algorithms can be split in two different categories, whether they rely on the explicit knowledge of coil sensitivity profiles or not.

Notation

Let us define $\mathcal{F}$ the 2D Fourier transform operator, possibly non-uniform, and Ω the support of the k-space under-sampling scheme, such that $\mathcal{F}_\Omega$ corresponds to the under-sampled Fourier operator over Ω. We also introduce L the total number of channel used to acquire the NMR signal and we denote $\boldsymbol{y}_{\Omega,\ell} \in \mathbb{C}^m$, with $\forall \ell \in \{1, \dots, L\}$, as the under-sampled k-space measurements collected by the ℓ^{th}-channel. With n the base resolution and $N = n \times n$ is the total number of pixels, the vectorized image is represented by $\boldsymbol{x} \in \mathbb{C}^N$. As the k-space is massively under-sampled, we get $m \ll N$. The sensitivity matrix of the ℓ^{th}-coil is denoted by $\boldsymbol{S}_\ell$, $\forall \ell \in \{1, \dots, L\}$, and $\boldsymbol{S}_\ell$ being diagonal.

3.2 Parallel MR image reconstruction based on sensitivity maps

In the Compressed Sensing (CS) context, sparse multi-coil image reconstruction has first been applied using a variant of the SENSE formulation (Section 1.3.2) called sparse-SENSE and originally proposed by [Liang 2009]. The reconstruction problem reads as follows:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathbb{C}^N}{\arg\min} \frac{1}{2} \sum_{\ell=1}^{L} \|\mathcal{F}_\Omega \boldsymbol{S}_\ell \boldsymbol{x} - \boldsymbol{y}_{\Omega,\ell}\|_2^2 + \lambda \|\boldsymbol{\Psi} \boldsymbol{x}\|_1, \quad \lambda > 0. \tag{3.1}$$

with $\boldsymbol{\Psi}$ being a sparsifying transform. Note however that a prior work in the parallel imaging context – with deterministic under-sampling along the phase encoding direction – has been proposed by [Chaâri 2008] and further extended in [Chaâri 2011].

In the SENSE formulation, the calibration of the spatial sensitivity profiles associated with the multiple channels gets critical as they heavily affect the final reconstructed image. Originally those sensitivity profiles were estimated using a dedicated pre-scan (low-resolution acquisition). However, this method suffers from two limitations: first it increases the acquisition time and second it is sensitive to the patient's motion between the calibration and accelerated scans. To overcome those issues, many methods have been proposed to estimate the sensitivity profiles. One sensible approach consists of selecting a calibration region in the k-space domain where the signal has been sampled (at least) at the Nyquist rate. As these maps are spatially smooth, one can actually extract the center of k-space (low frequency domain) and then apply the inverse Fourier transform to recover an estimate. In what follows, we describe the different strategies in detail. Then, we propose a simple yet efficient extraction method well suited for non-Cartesian sampling patterns.

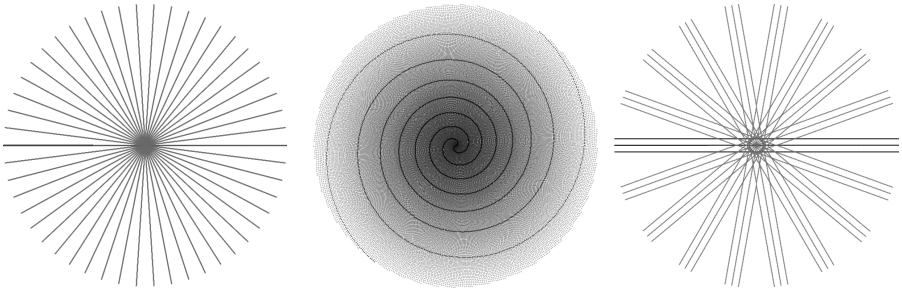
3.2.1 Coil sensitivity extraction

Direct extraction

First let us consider methods that extract the sensitivity information beforehand and use it as a prior in the optimization of Eq. (3.1). As originally proposed by [McKenzie 2002], extra k-space central lines can be added to the accelerated scan and used to extract low-resolution coil images which are then used as sensitivity profiles. While those lines can easily be added to Cartesian acquisitions, it is less obvious for non-Cartesian sampling. However because most of non-Cartesian schemes (such as radial [Spuentrup 2004, Feng 2016], variable density spiral [Lee 2003], PROPELLER [Pipe 1999], or even SPARKLING [Lazarus 2019a] as presented in Figure 3.2-2) are heavily sampled in the k-space center region, they could be considered as inherently calibrated as it has been mentioned by [Yeh 2005]. Using a gridding operation, low-resolution coil-specific images are then reconstructed using a zero-filled inverse Fast Fourier transform and used as sensitivity profiles for non-Cartesian sampling.

In the same context, we have proposed a similar approach to extract the sensitivity information. First, akin to [Yeh 2005], we selected only a portion θ of the central surface of k-space and set the other values to zero. Second, instead of using a gridding step, the NFFT [Keiner 2009] adjoint operator $\mathcal{F}^{\mathrm{H}}$ was applied to the data $\boldsymbol{y}_{\tilde{\Omega},\ell} = \boldsymbol{y}_{\Omega,\ell}^{\mathrm{LF}}$ in order to get the $n \times n$ low-resolution images: $\boldsymbol{x}_\ell^{\mathrm{LR}} = \mathcal{F}^{\mathrm{H}}_{[\Omega_{|\theta\%},\mathbf{0}]} \boldsymbol{y}_{\tilde{\Omega},\ell}$ where LR stands for low resolution. Third, the square root of the Sum of Squares (sSOS) was computed: $\widehat{\boldsymbol{x}}^{\,\mathrm{LR}} = \sqrt{\sum_{\ell=1}^{L} \|\boldsymbol{x}_\ell^{\mathrm{LR}}\|^2}$. Finally, the sensitivity maps were given by the pixel-wise ratio between the coil-specific images and the sSOS: $\left[\boldsymbol{s}_\ell^{\,\mathrm{LR}}\right]_{ij} = \mathrm{diag}\left[\boldsymbol{S}_\ell\right]_{ij} = \left[\boldsymbol{x}_\ell^{\,\mathrm{LR}}\right]_{ij} / \left[\widehat{\boldsymbol{x}}^{\,\mathrm{LR}}\right]_{ij}, \forall \ell \in \{1,\ldots,L\}$and $\forall i,j \in \{1,\ldots,N\}$. Because of this sSOS operation, our method is less dependent on the threshold θ compared to [Yeh 2005], who directly exploited the $\boldsymbol{x}_\ell^{\mathrm{LR}}$ images as sensitivity maps. In order to attenuate the noisy background, one could mask

the estimated sensitivity maps by a binary mask computed by thresholding $\widehat{\boldsymbol{x}}_{\mathrm{LR}}$, where the actual threshold value is given by a 2-cluster k-means algorithm. The binary mask is eventually defined as the largest connected component.



(a) (b) (c)

Figure 3.2-2: Examples of non-Cartesian trajectories: (a) radial sampling, (b) spiral sampling, (c) PROPELLER sampling. Oversampling the center of k-space brings a gain in robostness to the patient's motion.

Indirect extraction

While the sensitivity maps estimation in the previous works relied on a direct k-space extraction, [Uecker 2014] proposed an indirect approach, based on eigenvector decomposition called Eigenvalue Self-consistent Parallel Imaging Reconstruction using iTerative algorithm (ESPIRiT). Because the latter is considered as the state-of-the art in parallel imaging reconstruction we will explain it hereafter. In the original paper, the proposed method rests upon a calibration matrix $\boldsymbol{A}$ to extract the sensitivity information (such as SENSE) but it also enforces self-consistency with respect to the calibrated data (such as GRAPPA). Hence the authors claim to bridge the gap between SENSE and GRAPPA.

Construction of a calibration matrix. First let us introduce how the calibration matrix $\boldsymbol{A}$ is built up. Considering a calibration region AC[2] with non-missing k-space values $\boldsymbol{y}_{\Omega}^{\mathrm{AC}}$, the method proposed by [Uecker 2014] constructs a block-Hankel matrix $\boldsymbol{A}$ (also called calibration matrix). Block-Hankel matrices are known to possess well-defined subspaces [Heinig 1992] which in this case can be used to extract sensitivity information. The construction of matrix $\boldsymbol{A}$ is based on data concatenation along a 2D sliding window that moves over the AC region and across all channels as illustrated in Figure 3.2-3.

Extraction of the coil sensitivity profiles. With highly correlated rows, matrix $\boldsymbol{A}$ is supposed to be low-rank, hence only a small set of vectors span the k-space. To analyze the calibration matrix, the authors perform its singular value decomposition, which decomposes $\boldsymbol{A}$ into $\boldsymbol{U}$, Σ and $\boldsymbol{V}^H$ such that $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^H$, where the columns of $\boldsymbol{V}$ are the basis of the rows of $\boldsymbol{A}$ and therefore a basis for the overlapping blocks in the calibration data. The authors also show that $\boldsymbol{V}$ can be split into $\boldsymbol{V}_{\|}$ and $\boldsymbol{V}_{\perp}$ that respectively span for the row-space and the null-space of $\boldsymbol{A}$. Similarly to SENSE, the

[2] AC stands for Auto Calibrated.

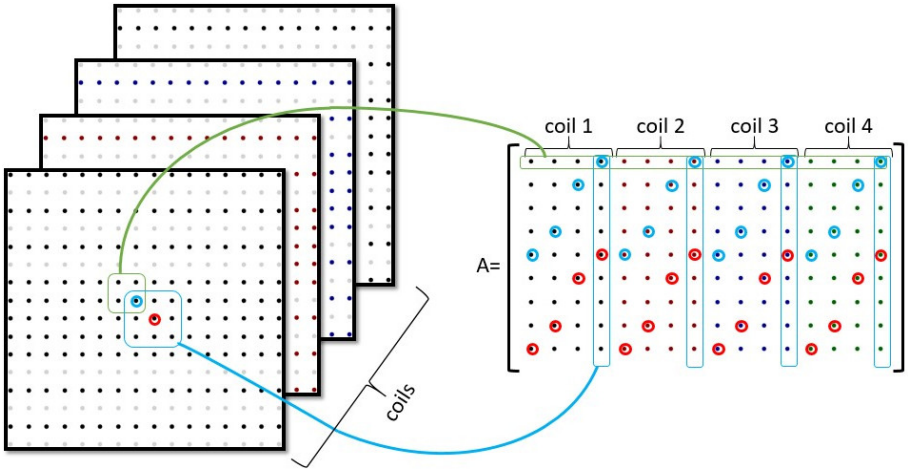


Figure 3.2-3: Left: Set of k-spaces obtained within the multi-coil setting. Right: the calibration matrix $\boldsymbol{A}$ is obtained by sliding a window (small green frame) on the calibration data $\boldsymbol{y}_{\Omega}^{\mathsf{AC}}$ (blue frame).

authors propose to explicitly identify the sub-space spanned by the coil sensitivities from the calibration matrix. Given the operator $\boldsymbol{R}_r$, which extracts and flattens the patches around the r^{th} k-space position, and $\hat{\boldsymbol{Y}}_{\Omega} = [\boldsymbol{y}_{\Omega,1}, \ldots, \boldsymbol{y}_{\Omega,L}]$ the unknown k-space data, one can re-write the null space condition thanks to the decomposition of subspace $\boldsymbol{V}$: $\boldsymbol{V}_{\|}\boldsymbol{V}_{\|}^{\mathrm{H}} + \boldsymbol{V}_{\perp}\ \boldsymbol{V}_{\perp}^{\mathrm{H}} = \mathbf{Id}$. The null space condition reads as follows:

$$\sum_r \boldsymbol{R}_r^{\mathrm{H}} \boldsymbol{V}_{\perp} \boldsymbol{V}_{\perp}^{\mathrm{H}} \boldsymbol{R}_r \hat{\boldsymbol{Y}} = 0$$

$$\underbrace{\left(\sum_r \boldsymbol{R}_r^{\mathrm{H}} \boldsymbol{R}_r\right)^{-1} \sum_r \boldsymbol{R}_r^{\mathrm{H}} \boldsymbol{V}_{\|} \boldsymbol{V}_{\|}^{\mathrm{H}} \boldsymbol{R}_r^{\mathrm{H}}}_{\mathcal{W}}\ \hat{\boldsymbol{Y}} = \hat{\boldsymbol{Y}}$$

Assume that image $\boldsymbol{x}$ is not null and satisfies the forward model, *i.e.* $\mathcal{F}\boldsymbol{S}\boldsymbol{x} = \hat{\boldsymbol{Y}}$, it follows that the sensitivity matrices are represented by the eigenvectors of $\mathcal{F}^{\mathrm{H}}\mathcal{W}\mathcal{F}$ associated with the eigenvalue 1. Those eigenvectors could be used to compute the sensitivity profiles, however as reported by the authors this leads to a high computation cost. Consequently, they have proposed to simplify the eigenvalue decomposition of the operator $\mathcal{W}$ by directly applying it in the image domain. While the ESPIRiT method represents an elegant way of extracting the sensitivity matrices, this method remains suited for Cartesian sampling due to the calibration matrix $\boldsymbol{A}$. For non-Cartesian sampling, the construction of this matrix requires a gridding step which might degrade the extraction and thus the final image quality.

While fast estimation of coil sensitivity profiles could be thought of in the context of online reconstruction, the Lipschitz constant of the forward model depends on the estimated sensitivity profiles. Hence, it cannot be computed beforehand. This constitutes a major bottleneck for online MR image reconstruction.

3.2.2 Blind bi-linear reconstruction

Although for "calibrated"-Cartesian sampling schemes the extraction of sensitivity matrices can be efficiently performed using one of the above mentioned methods, their adaptation to non-Cartesian, random or uncalibrated Cartesian sampling patterns is far less straightforward. In these cases, sensitivity matrices cannot be extracted prior to reconstruction. However to extend the SENSE formulation, a new category of methods, which aims at recovering both the coil sensitivity profiles and the image solution, has emerged. Such methods are often referred to as *blind bi-linear* inverse approaches.

Hereafter we define the general regularized blind bi-linear reconstruction problem as follows:

$$(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{S}}_1, \dots, \widehat{\boldsymbol{S}}_L) = \underset{\substack{\boldsymbol{x}\in\mathbb{C}^N \\ \forall \ell\in\{1,\dots,L\},\ \boldsymbol{S}_\ell\in\mathbb{C}^N}}{\operatorname{argmin}} \left\{ \sum_{\ell=1}^{L} \frac{1}{2} \|\mathcal{F}_\Omega \boldsymbol{S}_\ell \boldsymbol{x} - \boldsymbol{y}_\ell\|_2^2 + \mathcal{R}_{\text{sens}}(\boldsymbol{S}_\ell) + \mathcal{R}_{\text{im}}(\boldsymbol{x}) \right\} \quad (3.2)$$

with $\mathcal{R}_{\text{sens}}$ being the penalty term on the coil sensitivity profiles $(\boldsymbol{S}_\ell)_\ell$, which promotes their spatial smoothness, while $\mathcal{R}_{\text{im}}$ encodes the regularization term on the image $\boldsymbol{x}$, which can enforce sparsity in the CS framework.

The regularized blind bi-linear reconstruction was first introduced by [Uecker 2008]. They originally proposed a Tikhonov regularization [Tikhonov 1943, Tikhonov 1963] as a prior on $\boldsymbol{x}$ and the Sobolev regularization [Sobolev 1950] for $(\boldsymbol{S}_\ell)_{\ell\in\{1,\dots,L\}}$. The Sobolev penalty term is a special case of Tikhonov regularization since it penalizes the weighted ℓ_2-norm of the Fourier coefficients instead of the ℓ_2-norm in the original domain (*e.g.* image space). The regularization terms proposed by [Uecker 2008] were defined as follows:

$$\forall \boldsymbol{x} \in \mathbb{C}^N, \quad \mathcal{R}_{\text{im}}(\boldsymbol{x}) = \frac{\alpha}{2} \|\boldsymbol{x}\|_2^2 \quad (3.3a)$$

$$\forall \ell \in \{1,\dots,L\},\ \boldsymbol{S}_\ell \in \mathbb{C}^{N\times N}, \quad \mathcal{R}_{\text{sens}}(\boldsymbol{S}_\ell) = \frac{\beta}{2} \| \left(1 + \|\boldsymbol{k}\|_2^2\right)^{\gamma/2} \mathcal{F}\boldsymbol{S}_\ell\|_2^2 \quad (3.3b)$$

with $\|\boldsymbol{k}\|_2^2$ being the distance to the origin of k-space, $\mathcal{F}$ the fully sampled Fourier transform and α, β and γ (the polynomial degree) being positive hyper-parameters that need to be set.

In the original paper [Uecker 2008], the authors have proposed to decrease the value of α and β parameters over iterations, leading at the end to the minimization to a non-regularized, ill-conditioned problem. To overcome this issue, [Knoll 2012] have recommended to fix it during the minimization. Based on a previous work [Knoll 2011], they have also proposed to replace Tikhonnov regularization $\mathcal{R}_{\text{im}}$ in the image domain by total variation (TV) [Rudin 1992], which promotes piece-wise constant images (cf Eq. (3.4a)) or its generalization, namely the total generalized variation (TGV), which assumes non piece-wise constant images (cf Eq. (3.4b)). This eventually leads to the following penalty terms:

$$\text{TV:} \quad \mathcal{R}_{\text{im}}(x) = \alpha \|\nabla \boldsymbol{x}\|_1 \quad (3.4a)$$

$$\text{TGV:} \quad \mathcal{R}_{\text{im}}(x) = \alpha \min_{\boldsymbol{v}} \left(\|\nabla \boldsymbol{x} - \boldsymbol{v}\|_1 + \|\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^{\text{H}}\|_1 \right) \quad (3.4b)$$

with $\nabla \boldsymbol{x}$ being the spatial gradient of $\boldsymbol{x}$ *i.e.* $(\nabla \boldsymbol{x})_{i,j} = (x_{i+1,j} - x_{i,j},\ x_{i,j+1} - x_{i,j})$, hence $\nabla \boldsymbol{v}$ is linked to the second order derivative of $\boldsymbol{x}$. While the overall criterion in Eq. (3.2) is not jointly convex in $\big(\boldsymbol{x}, (\boldsymbol{S}_\ell)_\ell\big)$, the authors of both works [Uecker 2008, Knoll 2012] initialize $\boldsymbol{x}$ and $(\boldsymbol{S}_\ell)_\ell$ to zero-valued vector and matrices and claim that the method was not sensitive to the initialization.

Besides the Sobolev regularization on the coil sensitivity profiles, [She 2014] has proposed the Sparse BLind Iterative Parallel imaging reconstruction (Sparse-BLIP) method based on TV regularization of the sensitivity profiles as defined in Eq. (3.5b). On top of this new regularization, following the seminal paper in sparse MRI reconstruction [Lustig 2007], they have also regularized the image $\boldsymbol{x}$ by combining its ℓ_1-norm in the wavelet domain (*i.e.* $\boldsymbol{\Psi x}$) with the TV penalty in the image domain as shown in Eq. (3.5a):

$$\forall \boldsymbol{x} \in \mathbb{C}^N, \quad \mathcal{R}_{\text{im}}(\boldsymbol{x}) = \lambda \|\boldsymbol{\Psi x}\|_1 + \alpha \|\nabla \boldsymbol{x}\|_1 \tag{3.5a}$$

$$\forall \ell \in \{1, \dots, L\},\ \boldsymbol{S}_\ell \in \mathbb{C}^{N \times N}, \quad \mathcal{R}_{\text{sens}}(\boldsymbol{S}_\ell) = \beta \|\nabla \boldsymbol{S}_\ell\|_1 \tag{3.5b}$$

Contrary to [Knoll 2012, Uecker 2008], [She 2014] set up the image $\boldsymbol{x}$ with zeros and the sensitivity profiles $(\boldsymbol{S}_\ell)_\ell$ with low resolution coil images. They also studied the impact of hyper-parameters setting and showed the benefits of ℓ_1-regularization.

One of the main limitation of blind-bi-linear methods concerns their computational cost. For example, in the benchmark presented in [She 2014], the authors considered 2D Cartesian data collected over a eight-channel coil and reconstructed with a target resolution of 256×256, for a FOV of $20.4 \times 20.4\text{cm}^2$. In this scenario, the computation time for blind bi-linear methods (*i.e.* J-SENSE, Sparse-BLIP and IRGN-TV) was between 4.5 and 10 times slower than Sparse-SENSE (see Table 3.1). This constitutes the major drawback of blind bi-linear methods. In highly accelerated non-Cartesian sampling schemes, we can anticipate even more demanding computational load and memory usage as:

i) the target base resolution is often larger than the one proposed by [Shin 2014] for the same FOV;
ii) the FFT is replaced by the NUFFT wich has a higher numerical complexity (Section 1.2.3);
iii) a larger number of receivers is used in multi-channel coil acquisition to boost the SNR.

Table 3.1: Reported computation time for 2D-Cartesian reconstruction (acquired with an 8-channel coil) with a target image resolution equal to 256×256 and an under-sampling factor of 2.78, for different methods including: Sparse-BLIP [She 2014], J-SENSE [Ying 2007], Sparse-SENSE [Liang 2009] and IRGN-TV [Knoll 2012]. Originally presented in [Shin 2014].

Sparse-BLIP	J-SENSE	Sparse-SENSE	IRGN-TV
113 s	172 s	25 s	213 s

While most recent blind methods rely on the regularization of sensitivity profiles, their dependency on the initial guess, the hard setting of hyper-parameters and the slow conver-

gence of algorithms make them less appealing for online image reconstruction purposes.

3.3 Coil-by-coil MR image reconstruction

While multi-channel MR image reconstruction methods based on coil sensitivity profiles are well suited for non-Cartesian acquisition, the extraction of sensitivity profiles is a critical step with a severe impact on the final reconstructed image. In some cases, such as the Cartesian scan with calibration lines, the estimation problem is easier. However, it gets harder when no-calibration data is available or for non-Cartesian acquisitions. In the last case, blind reconstruction methods have been proposed to estimate both the final image and the coil sensitivity profiles, using priors such as sparsity for the combined image and smoothness for the sensitivity profiles. However, due to the non-convexity of the underlying objective function, blind bi-linear methods might have unexpected artifacts depending for instance on the initial estimate of the coil sensitivity profiles. In what follows, we introduce coil-by-coil MR image reconstruction methods. The latter can be split in two different categories depending on the considered optimization domain either the optimization lies in the k-space or in the image or sparse transform domain. Since those methods reconstruct one image per coil, for comparison purposes we will use the square root of the sum-of squares to combine all coil-specific images.

3.3.1 K-space driven methods promoting data low-rankness

K-space methods were originally motivated by the good performances of GRAPPA [Griswold 2002] reconstruction (presented in Section 1.3.2). While early methods are based on the robust extraction of an interpolation kernel, k-space driven reconstructions have met a shift in their principle. In fact, instead of finding an explicit interpolation kernel to generate the missing samples, state-of-the-art methods enforce the low-rankness of k-space data. This was made possible due to the following assumptions:

i) All the sensitivity maps are band-limited hence they can be represented by a small number of Fourier coefficients. Because of the forward model (element wise multiplication in the image domain) this corresponds to a convolution in the k-space with a small kernel.

ii) K-space sampling is locally uniform and the parallel MRI reconstruction problem is in fact over-determined due to the redundancy between channels.

In particular, [Zhang 2011] were the first to exploit Assumption i) and proposed a reconstruction method called Parallel Reconstruction Using Null-Operations (PRUNO) that exploits a k-space low-rank constraint. We explain it in detail hereafter.

Let us consider a Cartesian acquisition scenario. If we denote by $\boldsymbol{y}_a$ the k-space measurements concatenated over the channels and $\boldsymbol{y}_m$ the missing ones over all channels, then $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_a \\ \boldsymbol{y}_m \end{bmatrix}$ corresponds to the concatenation of the measured and missing samples. The missing k-space samples are supposed to be estimated using an interpolation kernel

$\boldsymbol{C}$ that is calibrated from the calibration region in k-space (usually around its center). This convolution can be written as follows:

$$\boldsymbol{y}_m = \boldsymbol{C}\boldsymbol{y}_a \tag{3.6}$$

$$\begin{bmatrix}\mathbf{Id} & -\boldsymbol{C}\end{bmatrix}\begin{bmatrix}\boldsymbol{y}_m \\ \boldsymbol{y}_a\end{bmatrix} = \mathbf{0} \tag{3.7}$$

The optimization algorithm alternates between a conjugate gradient step to solve Eq. (3.6) and a regularization step that enforces the low-rankness of $\begin{bmatrix}\mathbf{Id} & \boldsymbol{C}\end{bmatrix}$.

While the previously mentioned method requires a calibration region to extract the kernel $\boldsymbol{C}$, recent reconstruction algorithms such as SAKE [Shin 2014] or p-LORAKS [Haldar 2016][3], have relaxed this assumption and proposed to use a low-rank constraint over a constructed block-Hankel matrix. For the sake of simplicity, we will rely on the SAKE formalism. The construction of this matrix is based on the vectorization of a sliding window that spans the whole k-space. This matrix is actually constructed like the ESPIRiT calibration matrix (presented in Figure 3.2-3) however the sliding window is not restricted to the calibration data (blue frame in Figure 3.2-3) but covers the entire k-space. If we denote by $\boldsymbol{H}$ the linear operator that generates the block-Hankel matrix $\boldsymbol{A}$ from the k-space $\boldsymbol{Y}$, and $\boldsymbol{H}^{\mathrm{H}}$ its adjoint, then the reconstruction problem solved by SAKE (or p-LORAKS) is defined as:

$$\begin{aligned}&\underset{\boldsymbol{W}\in\mathbb{C}^{m\times L}}{\arg\min} \quad rank(\boldsymbol{A}) \\ &\text{subject to}\begin{cases}\boldsymbol{W} &= \boldsymbol{H}^{\mathrm{H}}\boldsymbol{A}, \\ \Omega\boldsymbol{W} &= \boldsymbol{Y}\end{cases}\end{aligned} \tag{3.8}$$

with $\boldsymbol{W}$ the estimated k-space and Ω the sampling mask, hence, $\Omega\boldsymbol{W} = \boldsymbol{Y}$ is a consistency term with respect to the collected k-space measurements. To solve Eq. (3.8), the authors rely on a projected gradient descent algorithm. Recently, [Lobos 2019] has studied the optimal window shape (green frame in Figure 3.2-3) and proved that circular shapes maximize the reconstruction quality. Although these methods can be easily applied to any Cartesian sampling (phase-encoded variable density sampling or 3D Cartesian Poisson disk sampling etc...), their application to non-Cartesian readouts is less trivial, mainly because of the $\boldsymbol{H}$ operator, which is more complex to calculate in that case. This limitation could be relaxed using a gridding procedure such as GROG [Seiberlich 2007], however the latter may dramatically impair the reconstruction.

[3]The two methods are identical, yet they have been derived independently and are based on different theoretical assumptions. While SAKE [Shin 2014] is based on multi-channel k-space relationships, p-LORAKS [Haldar 2016] was originally motivated by the k-space support modeling and phase assumptions of the images.

3.3.2 Domain based methods

While k-space based methods are efficient for calibrationless image reconstruction from Cartesian data, they are prone to artifacts for non-Cartesian readouts as they require a gridding step. Hence, for non-Cartesian acquisitions an inverse problem formulation is beneficial for maximizing image quality. However as seen in Section 3.2.1, those reconstruction methods require the explicit knowledge of coil sensitivity profiles. In some cases, the extraction of the latter can be either computationally demanding, unstable or inaccurate (depending on the initialization), leading to unexpected artifacts on the magnitude or phase images. Some alternatives have proposed to reconstruct channel-by-channel images using an inverse problem formulation, which basically consists of minimizing a variational form that regularize the reconstruction of multiple images through a joint sparsity penalization over the channels. Consequently, we focus here on methods that formulate the problem as follows:

$$\widehat{\boldsymbol{X}} = \underset{\boldsymbol{X}\in\mathbb{C}^{N\times L}}{\arg\min} \Big\{ \frac{1}{2} \sum_{\ell=1}^{L} \|\mathcal{F}_\Omega \boldsymbol{x}_\ell - \boldsymbol{y}_{\Omega,\ell}\|_2^2 + \mathcal{R}(\boldsymbol{Z}) \Big\}. \tag{3.9}$$

with $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L]$, $\boldsymbol{Z} = [\boldsymbol{\Psi}\boldsymbol{x}_1, \ldots \boldsymbol{\Psi}\boldsymbol{x}_L]$ and $\mathcal{R}$ the regularization term that promotes the joint-sparsity. Dictionary $\boldsymbol{\Psi}$ can be either the identity matrix, a fixed sparsifying decomposition (*e.g.* wavelet transform) or any data-driven dictionary. In their work, [Trzasko 2011] have proposed to use a locally low-rank formulation for image reconstruction called Calibration-free locally Low-rank EncourAging Reconstruction (CLEAR). They have introduced the use of a patch-based nuclear norm in the image domain to solve the multi-channel reconstruction problem. The nuclear norm $\|\boldsymbol{A}\|_*$ of a matrix $\boldsymbol{A}$ corresponds to the ℓ_1-norm of the singular value of $\boldsymbol{A}$. The CLEAR method actually defines $\boldsymbol{\Psi}$ as the image patch selection operator $\boldsymbol{P}_{s,p}$ requiring two other hyper-parameters to be set: the stride s and the patch size p, hence the regularization reads as follows: $\mathcal{R}(\boldsymbol{\Psi}\boldsymbol{X}) = \lambda\|\boldsymbol{P}_{s,p}\boldsymbol{X}\|_*$. The operator $\boldsymbol{P}_{s,p}$ is illustrated in Figure 3.3-4. In the original work, [Trzasko 2011] have proposed to use a stride equal to the patch size (*i.e.* non-overlapping patches), however to avoid the well-know patching artifact this work can be easily extended to overlapping patches where $s < p$.

This method has two major drawbacks: the large number of hyper-parameters to be tuned (λ, s and p) and the computational load that depends on the stride setting. While small strides prevent patching artifacts they increase the computational cost.

The group-sparsity regularization has also been used for MR reconstruction by [Majumdar 2012a]. In the latter work, the authors have used a constraint optimization method leading to solving the following problem:

$$\hat{\boldsymbol{X}} = \underset{\boldsymbol{X}\in\mathbb{C}^{N\times L}}{\operatorname{argmin}} \|\boldsymbol{\Psi}\boldsymbol{X}\|_{2,1}, \text{ such that } \|\mathcal{F}_\Omega\boldsymbol{X} - \boldsymbol{Y}\|_2 \leq \varepsilon$$

Using similar ideas, we have relaxed the constraint to adopt a penalized formulation and derive more sophisticated norms that will be extensively described in Chapter 4 for

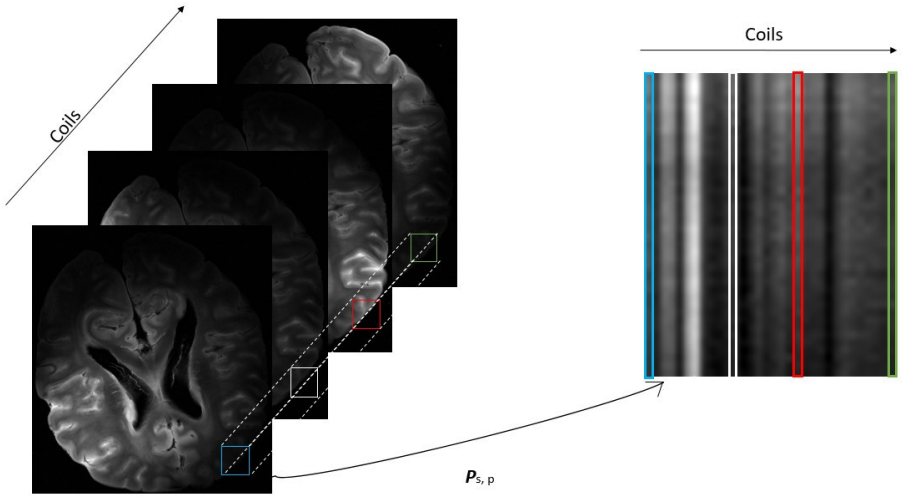


Figure 3.3-4: Illustration of the $\boldsymbol{P}_{s,p}\boldsymbol{X}$ operation performed by the CLEAR method, implemented on our ex vivo human brain data set.

calibrationless domain based multi-channel MR image reconstruction. In what follows, we describe how we implemented the above described methods and in particular our contribution presented in Section 3.2.1. The objective is to perform a benchmark of these competing methods in the non-Cartesian acquisition setting.

3.4 Experiments & Results

3.4.1 Experimental setup

Acquisition

For validation purposes, we acquired 5 different brain (ex- and in-vivo) anatomical MR images on a Magnetom 7 T scanner (Siemens Healthineers, Erlangen, Germany) equipped with $L = 32$-channels receiver coil (Nova Medical Inc., Wilmington, MA, USA). A modified 2D T2*-weighted GRE sequence was implemented to perform prospective CS based on the multi-shot Sparkling trajectories [Lazarus 2019a]. The acquisition parameters were set as follows: TR = 550 ms, TE = 30 ms and FA = 25° with in-plane resolution of $400 \times 400\,\mu\text{m}^2$ corresponding to an image matrix size of $N = 512 \times 512$. Sparkling shots were generated all together using the algorithm proposed in [Boyer 2016] to draw samples according to a variable density with a polynomial decay of 2 (*i.e.*, $h(\boldsymbol{k}) = 1/\|\boldsymbol{k}\|^2$ with $\boldsymbol{k} = (k_x, k_y)$). A Cartesian reference scan composed of 512 shots of 512 samples each has also been acquired using the same coil. For comparison purposes, we used the square root of the sum-of-squares. The acceleration factor in time AF, corresponding to the time ratio between the accelerated acquisition and the Cartesian reference, the slice thickness and the object under investigation are reported in Table 3.2. While ex-vivo acquisitions are relevant for quantitative comparisons, quantitative comparisons for in-vivo acquisitions should be considered carefully due to possible motion artifacts or mis-registration between the Cartesian reference scan and the accelerated one. Hence, the

latter will be only compared visually for illustration purposes. The reader is referred to Appendix A for in-vivo comparison.

Table 3.2: Ex-vivo baboon and in-vivo human brain data collected using various slice thickness and acceleration factors A.

	Object	Slice thickness	AF
1.	Baboon brain (ex-vivo)	3 mm	15 (34 shots)
2.		2 mm	15 (34 shots)
3.		1 mm	15 (34 shots)
4.	Human brain (in-vivo)	3 mm	8 (64 shots)
5.			15 (34 shots)

Reconstruction

In this benchmark we considered:

1. Sparse-SENSE formulation with the proposed extraction [El Gueddari 2018b];
2. ℓ_1-ESPIRiT [Uecker 2014] implemented within BART toolbox [Uecker 2015];
3. p-LORAKS [Haldar 2016] implementation provided by the authors of [Kim 2018b] using a gridding step before the reconstruction;
4. CLEAR method [Trzasko 2011] with various patch shape with overlaps.

We omitted blind-bi-linear reconstruction since the reported computation time (see Table 3.1) was considered excessively long for relatively simple cases, *i.e.* Cartesian sampling, 8-channel coil, and 256×256 image size. The CALM method was also omitted since Chapter 4 proposes an extensive comparison with this method and its variants. All the regularization parameters were tuned using a grid-search procedure to maximize the Structural Similarity Index (SSIM) score [Wang 2004]. For channel-by-channel MR reconstruction the coil combination was performed using the SOS. We considered the Symmlet basis as sparsifying transform when needed. All experiments were run on a machine with 128 GB of RAM and an 8-core (2.40 GHz) Intel Xeon E5-2630 v3 Processor.

3.4.2 Results

To study the impact of different reconstruction methods on prospectively accelerated non-Cartesian acquisitions, we proceeded in two steps. First, we considered the SENSE-like reconstruction to evaluate the impact of sensitivity profiles on the reconstructed image. As no ground-truth exists for these sensitivity maps, in a second step we validated the results on the reconstructed images where we compared the SSIM scores for the different techniques (*i.e.*, Sparse-SENSE with the developed sensitivity profile extraction, ℓ_1-ESPIRiT, p-LORAKS and CLEAR methods).

Sensitivity matrix impact on the reconstructed images

First, we studied the impact of the sensitivity profiles over the reconstructed image using a SENSE-like prior, with the extraction method proposed [El Gueddari 2018b]. Using our method, we investigated the impact of the k-space portion used to extract sensitivity information. The results are presented in Figure 3.4-5, where we reported the SSIM score of the reconstructed image for different values of $\theta \in \{1\%, 5\%, 25\%, 50\%, 100\%\}$. As regards the setting of parameter λ, we performed a grid-search to find its optimal value. This figure suggests that for an optimal λ value –$\simeq 5 \times 10^{-6}$– small k-space portions – *i.e.* $\theta < 25\%$ – The SSIM score is maximized. However, images reconstructed with sensitivity profiles extracted from small k-space portion are more sensitive to the tuning of λ. In fact when parameter λ is underestimated (*e.g.* $\lambda < 10^{-6}$), medium to large k-space coverage – $\theta > 25\%$ – provide better reconstruction. In contrast, a slight overestimation of λ impacts less image quality after reconstruction.

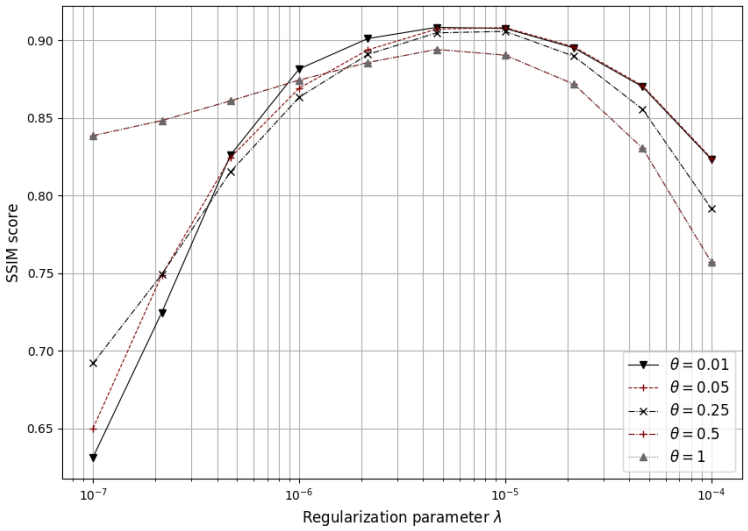


Figure 3.4-5: Impact of sensitivity profiles estimation on the final reconstructed image quality. The last two settings ($\theta \in \{0.5, 1\}$) yield superimposed traces indicating that there is no gain in collecting high-frequency information for the recovery of sensitivity profiles.

Visually, examples of the extracted sensitivity profiles, estimated with ℓ_1-ESPIRiT and our approach are presented in Figure 3.4-6 for the 2 mm slice-thickness baboon brain acquisition (cf Table 3.2 **(2.)**). The striking difference between the two approaches lies in the patterns of $\boldsymbol{S}_\ell$ images (see Figure 3.4-6): because of the SVD decomposition involved in ℓ_1-ESPIRiT, the corresponding sensitivity profiles are less structured and smoother than the maps yielded by our approach. The latter appear sharper and clearly show the sensitivity area of each channel. Moreover, our approach is much more efficient in terms of computing time since the average cost is about 1 min using a Matlab R2017a-based implementation as compared to 10 min for ℓ_1-ESPIRiT on the same hardware configuration.

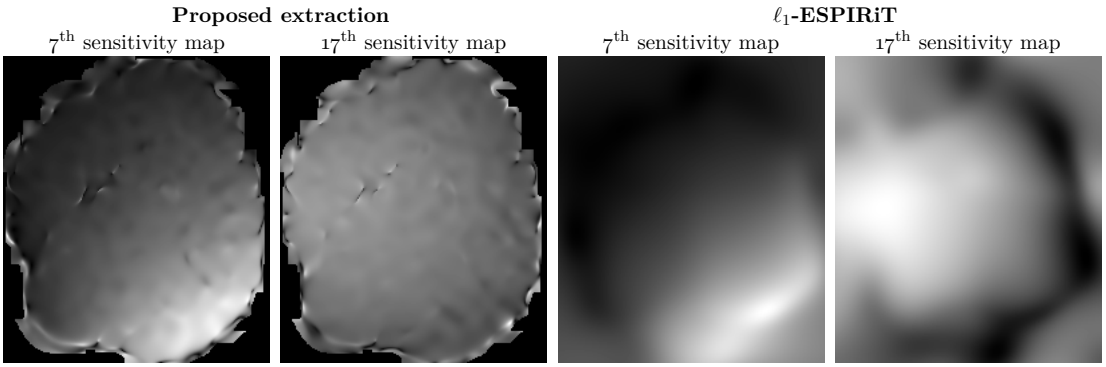


Figure 3.4-6: 2 out of 32 sensitivity maps extracted using our method (left) and ℓ_1-ESPIRiT (right).

Reconstructed images

Table 3.3 presents the quantitative comparison with respect to the ground truth based on the SSIM score. According to the SSIM score, the Sparse-SENSE reconstruction provides the best results for the three data sets under study. A visual illustration is provided in Figure 3.4-7 on the 2mm ex-vivo baboon brain which confirms these quantitative results.

Table 3.3: SSIM values within the (0, 1) range computed over three data sets. The larger the SSIM score the better the image quality. **Bold font** indicates the best score.

Dataset	ℓ_1-ESPIRiT	Sparse-SENSE	p-LORAKS	CLEAR
1.	0.873	**0.880**	0.854	0.872
2.	0.897	**0.908**	0.860	0.894
3.	0.848	**0.890**	0.793	0.834

The reconstruction proposed by the Sparse-SENSE is very similar to the Cartesian reference scan, while the ℓ_1-ESPIRiT method presents a smooth background with vertical lines, particularly visible on the zoom depicted in Figure 3.4-7 (h). Regarding the coil-by-coil p-LORAKS reconstruction method the results seems affected by high frequency noise, although no blurring effect is visible. The reconstruction method is hampered by the gridding step required to build the calibration matrix $\boldsymbol{A}$. This can be seen when comparing the output result of p-LORAKS with the SOS of the gridded k-space displayed in Figure 3.4-8. Both figures present the same noise pattern and p-LORAKS does not improve the reconstruction. During this experiment two major impediments were reported regarding p-LORAKS: the first one reflects its poor performance on non-Cartesian acquisitions while the second one concerns the construction of matrix $\boldsymbol{A}$ for high-resolution data collected over a large number of coils (*e.g.* $L = 32$) which requires a huge memory load. Hence given our experimental set-up and the code provided by the authors [Kim 2017b], we could only consider a sliding window of size 3 and 5 (Figure 3.2-3 green frame), since we were running out of memory on a 128 GB RAM workstation for larger window sizes. Last, CLEAR results presents very similar results compared to ℓ_1-ESPIRIT in term of

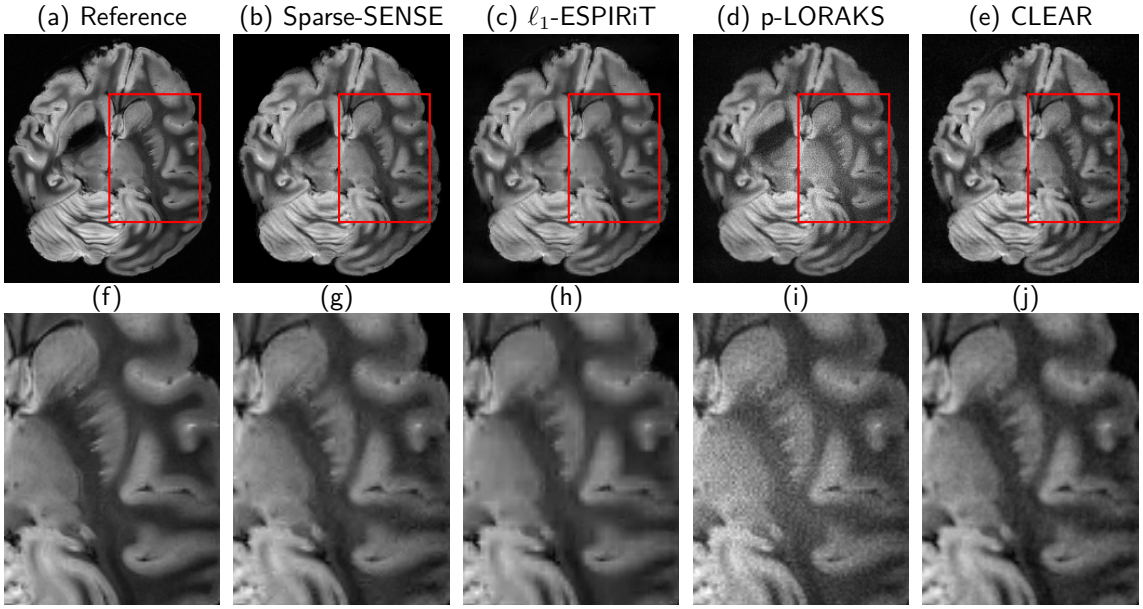


Figure 3.4-7: Comparison of different parallel MRI reconstruction methods on a prospectively 15-fold accelerated Sparkling acquisition on the ev-vivo baboon brain (2mm slice thickness). (a) Cartesian reference, (b) Sparse-SENSE reconstruction with the proposed sensitivity maps extraction method, (c) ℓ_1-ESPIRiT reconstruction (d) p-LORAKS solution after the gridding step and (e) CLEAR reconstruction and (f-j) their respective zoom over the red frames.

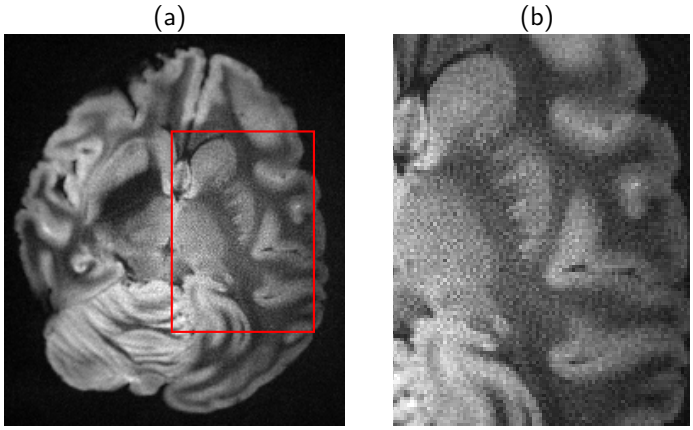


Figure 3.4-8: (a) Square root of the Sum-Of-Squares (sSOS) of the gridded k-space reconstruction, (corresponding to the result before p-LORAKS reconstruction), (b) zoom in the red frame.

SSIM score, however the reconstruction is more affected by the noise. Figure 3.4-9 presents the hyper-parameter setting for CLEAR. The optimal parameter values were obtained with high overlapping factor of $= 4$ – corresponding to a stride s of 8 – and a patch size p of 32×32. Also, large values of p provide a significant gain in robustness with respect to the setting of parameter λ irrespective of the overlapping factor: a high SSIM score is maintained overs almost five orders of magnitudes ($\lambda \in [10^{-6}, 10^{-2}]$). Importantly, while large overlapping factors prevent from patching artifact, they are more computationally demanding although they can be parallelized.

Moreover when we visually investigated the results of CLEAR method, we found some

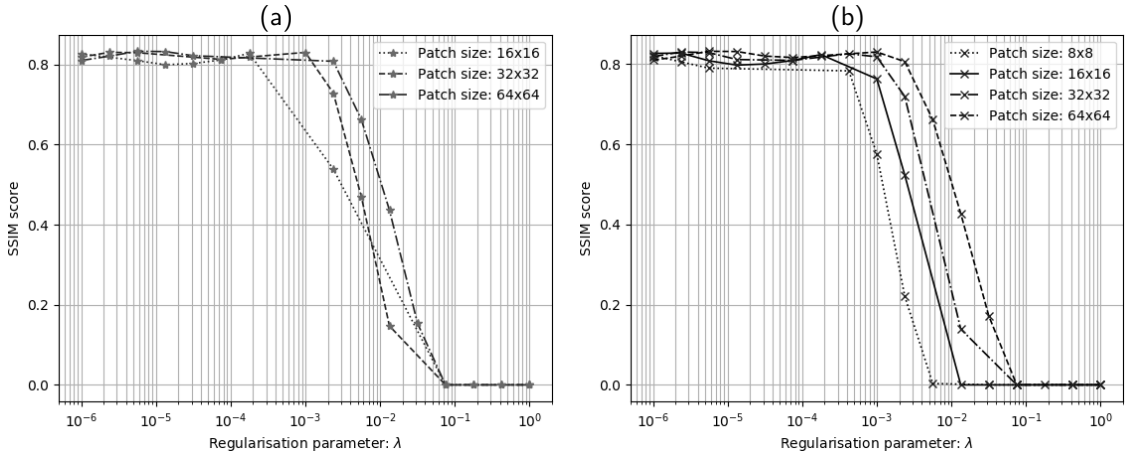


Figure 3.4-9: SSIM score of the sSOS results for the CLEAR method (a) for of $= 2$ (corresponding to a stride s of 8×8 and a patch size p of 16×16) and (b) for of $= 4$ (corresponding to a stride s of 4×4 and a patch size p of 16×16).

inconsistency between the reported SSIM score and the visual perception. For instance, Figure 3.4-10(a) provides the highest SSIM score while Figure 3.4-10(b) gives the best visual perception.

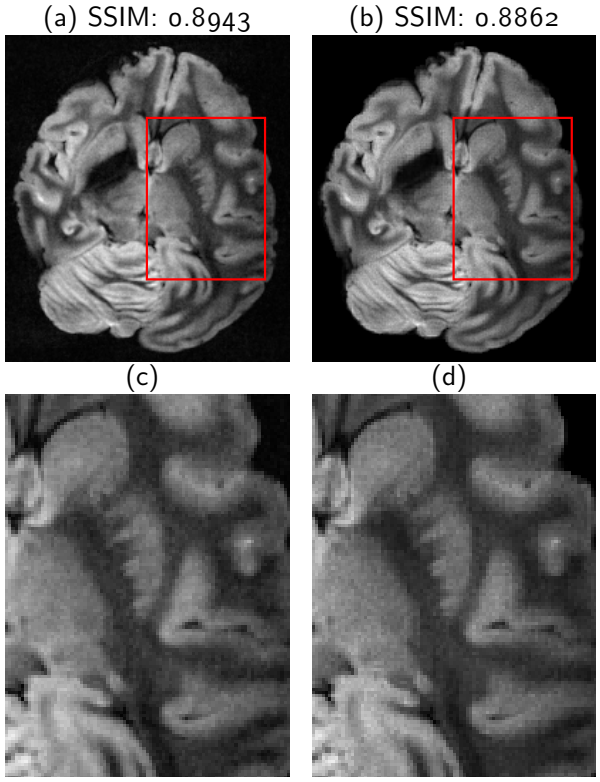


Figure 3.4-10: (a-b): CLEAR results for two settings of hyper-parameters. (a): provides the reconstruction with highest SSIM score (0.8943) and (b) yields the best visual perception but with a lower SSIM score (0.8862). (c-d): their respective zooms.

3.5 Conclusion

For high-resolution MR images, the input-SNR is a crucial issue as it allows to increase both the spatial resolution and the acceleration factor. As getting higher input-SNR often means collecting the signal with a multi-channel receiver coil, the input data set dimension increases and the reconstruction problem becomes harder. Multi-channel reconstruction algorithms can be split in two categories: either they deliver a single image from all the coils or one image per coil.

The first category relies on the knowledge of the coil sensitivity profiles that need to be estimated for each scan since they are scan-dependent. The estimation of coil sensitivity profiles is often based on a heavily sampled k-space region (also termed a calibration region) where much data have been acquired. Direct calibration methods use the coil-specific images as sensitivity maps (see *e.g.* [El Gueddari 2018b]) whereas indirect approaches estimate the sensitivity maps using eigenvalue decomposition (cf ESPIRiT [Uecker 2014] and its ℓ_1 extension), still the direct approaches are more efficient in terms of computational time and memory footprint. As an alternative, the blind bi-linear method also relies on SENSE reconstruction but jointly reconstructs the coil sensitivity profiles and a combined image. However, due to the non-convexity of the overall objective function and the underlying numerical complexity, the computation time often remains the major bottleneck.

The second category performs coil-by-coil reconstruction and implements regularization in the k-space or image space. While k-space regularization provides good performances for Cartesian sampling, its application to non-Cartesian acquisitions often requires a gridding procedure which degrades the final image quality as shown in this chapter. Alternative image space coil-by-coil reconstruction (such as CLEAR or CALM), relies on joint sparsity promotion over the multiple channels. Image based coil-coil reconstruction presents three main advantages:

i) *Efficiency* for non-Cartesian k-space acquisition as image reconstruction is solved using an inverse problem formulation.
ii) *Independence* with respect to the coil sensitivity profiles, no need to be extracted/estimated prior to reconstruction.
iii) *Reliability* as this formulation relies on the minimization of a convex criterion: any reached local minimum is global.

In terms of contribution, in this chapter we presented a new method to extract sensitivity maps and used a Sparse-SENSE formulation leading to a new self-calibrated method for non-Cartesian variable density sampling schemes. Compared to the state-of-the-art, our approach is much more efficient and the sensitivity profiles are easier to interpret. We demonstrated on several anatomical T_2^* data collected at 7 T that our global self-calibrated method is both more accurate and efficient than the competing alternatives such as ℓ_1-ESPIRiT, p-LORAKS and CLEAR. The proposed method has also been extended to various other contrast such as T_1-weighting [Lazarus 2018a, Chapter 3.6] and also to 3D imaging with a direct application to susceptibility weighted imaging (SWI) [Laz-

arus 2018a, Chapter 4.4].

3.6 Outlook

In the context of this thesis, *i.e.* non-Cartesian high-resolution imaging, SENSE-like methods and coil-by-coil image reconstructions are the only potential competitors. However, owing to the presence of coil sensitivity profiles in SENSE-like formulation, this kind of reconstruction methods is hardly online compliant. A potential alternative could be the coil-by-coil image reconstruction promoting joint sparsity across the different channels, such as CLEAR. In this thesis we have not addressed the choice of regularization parameters which may be a tricky problem in various scientific fields (such as data-science, statistical inference, and image restoration or reconstruction) depending on the overall similarity/global cost addressed. Instead, we chose to maximize the SSIM score, however sometimes we found some inconsistencies between visual perception of reconstruction images and the corresponding SSIM scores. Since no artifact related to the sampling scheme was present on the coil-specific images, one can wonder whether sensitivity information really matters for non-Cartesian reconstruction. In the rest of this thesis we review, propose and evaluate new regularization schemes based on structured sparsity for coil-by-coil high-resolution non-Cartesian MR image reconstruction.

* * *
* *
*

Chapter 4

Calibrationless multi-channel reconstruction: from group-sparsity to structured sparsity

Chapter Outline

The methods and principles contained in this chapter were also presented at international conferences as:

L. El Gueddari, P. Ciuciu, E. Chouzenoux, A. Vignaud and J.-C. Pesquet. *OSCAR-based*

reconstruction for compressed sensing and parallel MR imaging. In Proceedings of the 27th Annual Meeting of ISMRM, page 4766, Montreal, QC, Canada, 2019.

L. El Gueddari, P. Ciuciu, E. Chouzenoux, A. Vignaud and J.-C. Pesquet. *Calibrationless OSCAR-based image reconstruction in compressed sensing parallel MRI.* In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 2019. IEEE ISBI.

The purposes of this chapter is to assess the impact of regularization on the image quality in a calibration-less reconstruction framework. We emphasize the importance of promoting structured sparsity across the multiple receivers to improve the overall image quality.

4.1 Introduction

Multi-channel image reconstruction becomes challenging when it has to comply with timing constraints such as the ones involved for online reconstruction: basically, the SENSE formulation is no longer the most convenient as the sensitivity maps cannot be extracted from incomplete data, which does not cover a wide portion of the center of k-space. Hence, the self-calibrating methods presented in Chapter 3 are no longer suitable to extract the sensitivity maps associated with the multiple receivers. Also, a second bottleneck exposed in Section 3.2.1 is due to the dependence of the Lipschitz constant to those maps. In contrast, calibration-less methods do not require this prior knowledge to perform image reconstruction in the parallel imaging setting. This makes them more appealing for online reconstruction purposes. Calibration-less methods may operate either in the k-space domain or in the image or any transform (*e.g.* wavelet) domain. In the context of non-Cartesian acquisition, the existing k-space calibration-less methods in the literature do not perform as well as the self-calibrating ones mainly owing to the gridding step which is usually applied to perform regularization in the k-space. For this reason, we will investigate domain-based calibration-less methods and look at their advantages for imposing some structured sparsity across channels. Then we will evaluate the impact of those regularization terms on the final image quality and propose an extended comparison to other domain-based reconstruction methods, mainly, CALM [Majumdar 2012a] as it has been discussed in Chapter 3.

4.2 Calibration-less reconstruction: general problem statement

4.2.1 General problem statement

Let us first define the general calibration-less multi-channel coil image reconstruction problem. We recall n being the resolution of the sought image per channel and $N = n^d$ the image size where d is the dimension[1], L the number of channels used to acquire the NMR

[1] In 2D imaging $d = 2$, whereas in 3D imaging $d = 3$.

signal and M the number of k-space measurements per channel, with $M \ll N$. For the sake of compactness, we denote the complete data set $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L] \in \mathbb{C}^{M \times L}$ which stacks all vectors $\boldsymbol{y}_\ell \in \mathbb{C}^M$, the latter gathering the k-space samples collected in the ℓ^{th} channel.

From the noisy under-sampled data $(\boldsymbol{y}_\ell)_{1 \leq \ell \leq L}$, the goal is to retrive L MR images stacked in $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L] \in \mathbb{C}^{N \times L}$ such that each $\boldsymbol{x}_\ell \in \mathbb{C}^N$ is associated with the $\ell^{\textbf{th}}$ channel of the phased array coil.

$$\widehat{\boldsymbol{X}} = \underset{\boldsymbol{X} \in \mathbb{C}^{N \times L}}{\arg\min} \Big\{ \sum_{\ell=1}^{L} \frac{1}{2\sigma_\ell^2} \|\mathcal{F}_{\boldsymbol{M}} \boldsymbol{x}_\ell - \boldsymbol{y}_\ell\|_2^2 + \mathcal{R}(\boldsymbol{\Psi} \boldsymbol{X}) \Big\}. \tag{4.1}$$

Hereabove, $\mathcal{R} \in \Gamma_0(\mathbb{C}^{N_\Psi \times L})$ is a regularization function composed with a linear operator $\boldsymbol{\Psi} \in \mathbb{C}^{N_\Psi \times N}$, with the aim to enforce sparsity of the solution within a given multiscale decomposition (*e.g.*, wavelet transform). In the following we suppose that $\mathcal{R}$ have a closed form proximity operator. This formulation enables the use of over-complete dictionaries [Elad 2007]. We will assume that $\boldsymbol{\Psi}$ decomposes the stack of L images $\boldsymbol{X} \in \mathbb{C}^{N \times L}$ into a stack of coefficients $\boldsymbol{\Psi X} \in \mathbb{C}^{N_\Psi \times L}$ with C scales. Each scale $c \in \{1, \ldots, C\}$ is composed of S_c sub-bands. Each sub-band $s \in \{1, \ldots, S_c\}$ has $K_{s(c)}$ coefficients, so that finally $N_\Psi = \sum_{c=1}^{C} \sum_{s=1}^{S_c} K_{s(c)}$. We also introduce $\boldsymbol{Z}$ as the sparsifying decomposition $\boldsymbol{Z} = \boldsymbol{\Psi X} = [\boldsymbol{z}_1 \ldots \boldsymbol{z}_\ell] \in \mathbb{C}^{N_\Psi \times L}$ of the muti-channel image $\boldsymbol{X} = [\boldsymbol{x}_1 \ldots \boldsymbol{x}_\ell] \in \mathbb{C}^{N \times L}$. The resolution of Problem (4.1) delivers L channel images $(\widehat{\boldsymbol{x}}_\ell)_{1 \leq \ell \leq L}$, stacked in $\widehat{\boldsymbol{X}}$. To retrieve the magnitude and phase information many methods exist as discussed in Section 1.3, in this work we will use the square-root of the sum-of-squares (sSOS) [Roemer 1990], $\widehat{\boldsymbol{x}}_{sSOS} = \sqrt{\sum_{\ell=1}^{L} \|\widehat{\boldsymbol{x}}_\ell\|_2^2}$, to form a single *magnitude* image and the virtual coil method [Parker 2014] for phase image.

4.2.2 Optimization algorithm

In order to solve Problem (4.1) we adapted a primal-dual algorithm proposed in [Condat 2013, Vũ 2013] and recalled Section 2.3, leading to the Algorithm 3

Algorithm 3: Condat-Vú algorithm

Set $\tau > 0$, $\kappa > 0$, $\boldsymbol{X}_0 \in \mathbb{C}^{N \times L}$, $\boldsymbol{Z}_0 \in \mathbb{C}^{N_\Psi \times L}$;
for $k = 0, \ldots, T$ **do**
  $\boldsymbol{X}_{k+1} := \boldsymbol{X}_k - \tau \left(\nabla f(\boldsymbol{X}_k) + \boldsymbol{\Psi}^* \boldsymbol{Z}_k \right)$;
  $\boldsymbol{W}_{k+1} := \boldsymbol{Z}_k + \kappa \, \boldsymbol{\Psi} \left(2\boldsymbol{X}_{k+1} - \boldsymbol{X}_k \right)$;
  $\boldsymbol{Z}_{k+1} := \boldsymbol{W}_{k+1} - \kappa \, \text{prox}_{\mathcal{R}/\kappa} \left(\frac{\boldsymbol{W}_{k+1}}{\kappa} \right)$;
end

According to [Condat 2013, Theorem 3.1], the sequence $(\boldsymbol{X}_k)_{k \in \mathbb{N}}$ generated by Algorithm 3 weakly converges to a solution of Eq. (4.1) as soon as $\frac{1}{\tau} - \kappa |||\boldsymbol{\Psi}|||^2 \geq \frac{\beta}{2}$, with $\beta = \sum_{l=1}^{L} (\sigma_\ell^2)^{-1} |||\mathcal{F}_{\boldsymbol{M}}|||^2$ the Lipschitz constant of the gradient of the data consistency term. In practice, the hyper-parameters of this algorithm are set as follows: $\tau := \frac{1}{\beta}$,

$\kappa := \frac{\beta}{2|||\mathbf{\Psi}|||}$. Note that when $\mathbf{\Psi}$ defines a basis, we get $|||\mathbf{\Psi}||| = 1$. The main advantage of Algorithm 3 is that it does not involve the computation of $\text{prox}_{\mathcal{R}\circ\mathbf{\Psi}}$. The latter does not usually have closed form, in particular when $\mathbf{\Psi}$ is overcomplete (*e.g.* undecimated wavelet transform), and would require the use of an inner iterative solver [Combettes 2011].

4.3 From group-sparsity

4.3.1 Least Absolute Shrinkage and Selection Operator

The ℓ_1-norm has been successfully used for promoting sparsity of the solution in many different fields such as generalized linear models, Bayesian statistics and dictionary learning. The success of such norm in the regression/regularization domain is mainly due to its good properties as variable selection operator and regularization term as it has been shown [Tibshirani 1996] via the Least Absolute Shrinkage and Selection Operator (LASSO). In the context of Compressed Sensing the usage of such norm has often been justified as being the tightest convex envelope of the cardinality function (*i.e.* the ℓ_0 pseudo-norm) [Lustig 2007, Donoho 2006, Candès 2008]. When the sampling pattern presents large incoherence with respect to the sparsifying dictionary (as it is the case of Sparkling trajectories), the ℓ_1-norm can be perfectly applied to multi-channel image reconstruction. Hence, in the calibration-less setting the regularization term involvoing the Wavelet coefficients $\boldsymbol{Z}$ can be summarized as follows:

$$\forall \boldsymbol{Z} \in \mathbb{C}^{N\Psi\times L},\ \lambda \in \mathbb{R}_+,\ \mathcal{R}_{LASSO}(\boldsymbol{Z}) = \lambda\|\boldsymbol{Z}\|_1 = \lambda \sum_{n=1}^{N_\Psi}\sum_{\ell=1}^{L}|z_{n,\ell}| \tag{4.2}$$

with λ being a positive hyper-parameter that needs to be set. We recall that the proximity operator of the LASSO corresponds to a soft-thresholding step [Bertsekas 2015]:

$$\left(\text{prox}_{\mathcal{R}_{\text{LASSO}}}(\boldsymbol{Z})\right)_{n,\ell} = \begin{cases} z_{n,\ell}\left(1 - \frac{\lambda}{|z_{n,\ell}|}\right) & \text{, if } |z_{n,\ell}| \geq \lambda \\ 0 & \text{, otherwise} \end{cases} \tag{4.3}$$

Although the good performances of the LASSO have been demonstrated in many domain, such as dictionnary learning [Mairal 2010], this regularization also has some well known drawbacks such as its poor behavior in the case of highly correlated variables [Zhao 2006, Zou 2005, Argyriou 2012]. In this setup, the regularization tends to select one of those variables and discard the others, whereas one can expect to treat all the correlated variables the same way. To overcome this issue many regularization terms have been proposed, such as Elastic-Net [Zou 2005], group-LASSO [Yuan 2006], or based on clustering methods such as octagonal shrinkage and clustering algorithm for regression (OSCAR) [Bondell 2008]. In the context of multi-channel acquisition, the k-space seen by the different channels is actually highly correlated due to the coil configuration. Therefore in the following we will describe alternatives to LASSO regularization and their application to the multi-channel reconstruction.

4.3.2 Elastic-Net

The first alternative that has been proposed is the Elastic-net regularization [Zou 2005]. [Zou 2005] presented Elastic-net as the right counterpart to the LASSO penalty when variables are highly correlated. Elastic-net regularization performs a linear interpolation between ℓ_1 (LASSO) and ℓ_2 (Ridge regression [Hoerl 1970] or Tikhonov regularization [Tikhonov 1943]) norms and is defined as follows:

$$\forall \boldsymbol{Z} \in \mathbb{R}^{N_\Psi \times L}, (\lambda_1, \lambda_2) \in \mathbb{R}_+^2,\ \mathcal{R}_{\text{Elastic-net}}(\boldsymbol{Z}) = \lambda_1 \|\boldsymbol{Z}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{Z}\|_2^2 \tag{4.4}$$

The proximity operator of such norm is easy to compute, although the authors [Zou 2005] noticed that the naive version of Elastic net causes a double amount of shrinkage which led to poor predictions. Hence they proposed to rescale the thresholded coefficients by a factor of $(1 + \lambda_2/2)$ and thus the proximity operator is reduced to:

$$\left(\text{prox}_{\mathcal{R}_{\text{Elastic-net}}}(\boldsymbol{Z})\right)_{n,l} = \begin{cases} z_{n,\ell}\left(1 - \frac{\lambda_1}{(1+\lambda_2)|z_{n,\ell}|}\right) \times \left(1 + \frac{\lambda_2}{2}\right) & \text{, if } |z_{n,\ell}| \geq \frac{\lambda_1}{1+\lambda_2} \\ 0 & \text{, otherwise} \end{cases} \tag{4.5}$$

Although the LASSO and Elastic-net regularization terms do not explicitly exploit redundant information provided by the different coils, in the following we will explore various extensions that handle the structural information and redundancy across channels. To begin with, we study the group-LASSO penalty.

4.3.3 Group-LASSO

Definition and application to image reconstruction

In order to overcome the above mentioned drawbacks associated with LASSO regularization and select a group of variables, the group-LASSO [Yuan 2006] has been proposed. Given a set of groups $\mathcal{G}$ the group-lasso $\Omega^{\mathcal{G}}_{\text{group}}$ is basically defined as a $\ell_{2,1}$ mixed-norm:

$$\forall \boldsymbol{w} \in \mathbb{C}^p,\ \Omega^{\mathcal{G}}_{\text{group}}(\boldsymbol{w}_g) = \sum_{g\ \in \mathcal{G}} \|\boldsymbol{w}_g\|_2 \tag{4.6}$$

with $\mathcal{G}$ being the non-overlapping groups that need to be specified a priori. In the context of MR image reconstruction, a group gathers the same wavelet coefficients over all channels [Majumdar 2012a], the regularization can thus be defined as:

$$\forall\ \boldsymbol{Z} \in \mathbb{C}^{N_\Psi \times L},\ \lambda \in \mathbb{R}_+\ \mathcal{R}_{\text{gLASSO}}(\boldsymbol{Z}) = \lambda \|\boldsymbol{Z}\|_{2,1} = \lambda \sum_{n=1}^{N_\Psi} \sqrt{\sum_{\ell=1}^{L} |Z_{n,\ell}|^2} \tag{4.7}$$

The proximity operator of the $\ell_{2,1}$ mixed-norm is explicit and equal to:

$$\left(\text{prox}_{\mathcal{R}_{\text{gLASSO}}}(\boldsymbol{Z})\right)_{n,\ell} = \begin{cases} z_{n,\ell}\left(1 - \frac{\lambda}{\sqrt{\sum_{\ell=1}^{L}|z_{n,\ell}|^2}}\right) & \text{, if } \sqrt{\sum_{\ell=1}^{L}|z_{n,\ell}|^2} \geq \lambda \\ 0 & \text{, otherwise} \end{cases} \tag{4.8}$$

Mixed norm in other fields. The group penalty has arisen a strong interest in the inverse problem community as it has been used in many applications such as spatio-temporal MEG/EEG source reconstruction [Strohmeier 2016, Bekhti 2018], where the authors proposed an iterative scheme to approach the $\ell_{2,0.5}$-pseudo norm.

Overlapping Group-LASSO. An overlapping version of the group-LASSO also has been introduced by [Jacob 2009]. The regularization term is formulated as follow:

$$\boldsymbol{w} \in \mathbb{C}^p, \quad \Omega^{\mathcal{G}}_{overlapp}(\boldsymbol{w}) = \left\{ \sum_{g\in\mathcal{G}} \|\boldsymbol{w}_g\|_2 : \text{supp}(\boldsymbol{w}_g) \subseteq g, \boldsymbol{w} = \sum_{g\in\mathcal{G}} \boldsymbol{w}_g \right\} \tag{4.9}$$

with $\mathcal{G}$ the set of all groups. The major bottleneck of the use of such norm is related to the definition of the groups as the latter need to be defined a priori.

Expected improvement for multi-channel image reconstruction

The theory behind multi-channel sparse recovery problems has been extensively studied by the CS community as multiple-measurement vector (MMV) [Chen 2006, Eldar 2009a, Eldar 2009b, Duarte 2011, Cotter 2005] and uses the multi-channel basis pursuit:

$$\forall \boldsymbol{Y} \in \mathbb{C}^{p\times L}, \ \min_{\boldsymbol{Z}\in\mathbb{C}^{N\times L}} \|\boldsymbol{Z}\|_{2,1} \text{ s.t. } \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} \tag{4.10}$$

Although in the worst case scenario [Chen 2006] the use of the group-sparsity norm may not be beneficial in terms of recovery guarantees, [Eldar 2009b] has proposed an average case study and shown the potential benefit of adding the group structure with a probability of recovery failure decaying exponentially with the number of channels.

In the MRI context, [Chun 2015] has developed similar results. The authors have shown that for the case of disjoint sensitivity matrices the recovery guarantees of the multi-channel basis pursuit can be divided by a factor of L, with L the total number of channels. However those results were given for the case of disjoint sensitivity matrices, *i.e.* when the sensitivity profiles do not overlap and thus bring complementary information, which is not true in the real life Section 1.3. Therefore the benefits of such norms must be mitigated for real case applications.

4.4 To structured sparsity

In order to exploit the redundant information provided by each coil, the group-LASSO regularization requires a prior definition of the group-structure. While a straight-forward prior for calibrationless reconstruction consist of defining a group by the same coefficients

for all the channels, a more appropriate approach would be to consider different group-structure that leverages the group definition to maximize the final image quality. The following section will be dedicated to more sophisticated regularization terms that infer the structure of the groups. In particular we focus on two competitors, the first one presented in Section 4.4.1 is based on a Clustering algorithm, and finds the structure promoting a pairwise ℓ_∞-norm, while the second one in Section 4.4.2 based on the ℓ_2-norm defines the sets of all possible overlapping groups with a cardinality at most equal to k.

4.4.1 Octagonal Shrinkage and Clustering Algorithm for Regression

In the context of highly correlated variables, the first method one could think of consist of clustering those variables before the selection/shrinkage operation. This idea has been exploited in a new regularization called Octagonal Shrinkage and Clustering Algorithm for Regression (OSCAR) [Bondell 2008]. In [Argyriou 2012] the authors proved a link between the OSCAR norm and the Ordered Weighted ℓ_1 (OWL) norm [Zeng 2014b, Bogdan 2015]. Appendix B recalls the derivation of this link and the computation of its proximity operator.

In what follows, we propose four choices for function $\mathcal{R}$ in Problem (4.1) relying on OSCAR norm, with the aim to perform efficient calibration-less MR image reconstruction. The main difference between these formulations lies in the way the OSCAR norm is applied to the sparsifying decomposition $\boldsymbol{Z}$ of the muti-channel image $\boldsymbol{X}$.

General definition

Let $\boldsymbol{z} \in \mathbb{C}^p$ with $p \geq 1$. We introduce the magnitude sorting operator $S_p : \mathbb{C}^p \to \mathbb{C}^p$ such that vector $S_p(\boldsymbol{z}) = (S_p(\boldsymbol{z})_j)_{1 \leq j \leq p}$ contains the p entries of $\boldsymbol{z}$ sorted in decreasing order in magnitude, *i.e.* such that

$$|S_p(\boldsymbol{z})_1| \geq |S_p(\boldsymbol{z})_2| \geq \cdots \geq |S_p(\boldsymbol{z})_p|. \tag{4.11}$$

Then, OSCAR norm is defined as follows:

$$\Omega_{\lambda,\gamma}(\boldsymbol{z}) = \lambda \|\boldsymbol{z}\|_1 + \gamma \sum_{1 \leq j < k \leq p} \max(|z_j|, |z_k|). \tag{4.12}$$

with λ, γ two positive hyper-parameters. The ℓ_1-norm term in $\Omega_{\lambda,\gamma}$ promotes the sparsity of $\boldsymbol{z}$ while the second term, corresponding to a pairwise ℓ_∞-norm, encourages the equality of each pair of entries in $\boldsymbol{z}$. As pointed out in [Zeng 2014a, Sec. II. A.], OSCAR norm has a closed relation with the OWL norm defined below:

$$\Theta_{\boldsymbol{w}}(\boldsymbol{z}) = \sum_{j=1}^{p} w_j |S_p(\boldsymbol{z})_j|. \tag{4.13}$$

with $\boldsymbol{w} \in \mathbb{R}_+^p$ a vector of hyper-parameters such that $w_1 \geq \cdots \geq w_p \geq 0$. More precisely, OWL and OSCAR become equivalent if one sets the OWL weights as follows: $w_j = \lambda + \gamma(p - j)$ for $j = 1, \ldots, p$.

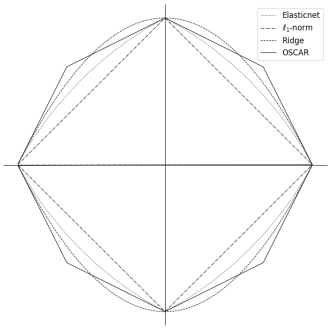


Figure 4.4-1: ℓ_1, Elastic net, Ridge and OSCAR ball in $\mathbb{R}^2$.

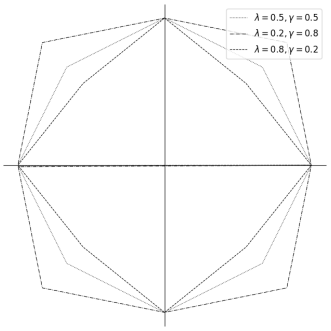


Figure 4.4-2: OSCAR ball in $\mathbb{R}^2$ for various pairs of (λ, γ).

Figure 4.4-1 compares the unit ball of the ℓ_1, Elastic-net and OSCAR norms. When γ is small, the OSCAR norm corresponds to the ℓ_1 norm. In contrast when $\lambda \to 0$ the OSCAR norm tends to the pair-wise ℓ_∞ norm, Figure 4.4-2 depicts the unit-balls for different pairs of (λ, γ) hyper-parameters. Note: in the 2D case the pair-wise ℓ_∞-norm correspond to the global ℓ_∞ norm.

Proximity operator

The proximity operator of $\boldsymbol{z} \in \mathbb{C}^p$ can be efficiently computed thanks to the following algorithm as shown in [Zeng 2014a, Sec.III A]:

Algorithm 4: Proximity operator of the OWL norm.

1 **Input:** $\boldsymbol{z} \in \mathbb{C}^p/\{0\}$, $\boldsymbol{w} \in \mathbb{R}^p$;
2 $\boldsymbol{n} = |\boldsymbol{z}|/\boldsymbol{z}$;
3 Let $\boldsymbol{P} \in \mathbb{R}^{p \times p}$ s.t. $S_p(\boldsymbol{n}) = \boldsymbol{P}\boldsymbol{n}$;
4 **Return:** $\text{prox}_{\Theta_{\boldsymbol{w}}}(\boldsymbol{z}) = \boldsymbol{n} \odot \boldsymbol{P}^\top \text{PAV}(S_p(\boldsymbol{n}) - \boldsymbol{w})$;

Hereabove, PAV refers to the Pool Adjacent Violator Algorithm [Mair 2009]. For $\boldsymbol{z}$ is equal to zero, then the proximity operator of the OWL norm at $\boldsymbol{z}$ is also equal to zero. The proximity operator of OSCAR can thus be easily deduced by setting the appropriate value for $\boldsymbol{w}$ mentioned above, so that OSCAR and OWL match together.

In the following we apply different versions of OSCAR regularization to the multi-channel wavelet coefficients $\boldsymbol{Z} \in \mathbb{C}^{N_\Psi \times L}$.

Global-OSCAR regularization

The most straightforward way to implement OSCAR-based regularization consists of flattening all wavelet coefficients and thus discarding the multiscale and multi-channel structure (*e.g.* in scales, sub-bands, coefficients and channels) of $\boldsymbol{Z}$. For that reason, we call this version global OSCAR (g-OSCAR) regularization. The wavelet coefficients are stacked together, leading to a single but large vector with entries $(z_j)_{1 \leq j \leq LN_\Psi}$, where we remind that

$N_\Psi = S\sum_{c=1}^{C} K_c$ and $N_\Psi = N$ when $\boldsymbol{\Psi}$ is orthogonal. The g-OSCAR regularizer then reads:

$$\begin{aligned}\mathcal{R}_{\text{g-OSCAR}}(\boldsymbol{Z}) &= \Omega_{\lambda,\gamma}(\boldsymbol{Z}) \\ &= \sum_{j=1}^{LN_\Psi} \lambda|z_j| + \gamma \sum_{1\le j<k\le LN_\Psi} \max\{|z_j|,|z_k|\} \\ &= \sum_{j=1}^{LN_\Psi} \left(\lambda + \gamma(LN_\Psi - j)\right) |S_{LN_\Psi}(\boldsymbol{z})_j| \,. \end{aligned} \tag{4.14}$$

with S_{LN_Ψ}, the magnitude decreasing sorting operator of dimension LN_Ψ.

Scale-wise OSCAR regularization

We now propose a scalewise formulation, where OSCAR norm is applied to each specific scale c of the wavelet decomposition, hence to each vector $\boldsymbol{z}_{c,:}$ separately where $\boldsymbol{z}_{c,:}$ gathers the LSK_c wavelet coefficients across all channels in a specific scale $c \in \{1,\dots,C\}$. This leads to the so-called s-OSCAR regularizer:

$$\begin{aligned}\mathcal{R}_{\text{s-OSCAR}}(\boldsymbol{Z}) &= \sum_{c=1}^{C} \Omega_{\lambda,\gamma}(\boldsymbol{z}_{c,:}) \\ &= \sum_{c=1}^{C} \sum_{j=1}^{LSK_c} \left(\lambda + \gamma(LSK_c - j)\right) |S_{LSK_c}(\boldsymbol{z}_{c,:})_j| \,, \end{aligned} \tag{4.15}$$

with S_{LSK_c}, the LSK_c-dimensional magnitude decreasing sorting operator. In that way, the wavelet coefficients can be clustered together regardless the subband they belong to, their position and their channel dependence. Thus, C sorting operations are required, each of them involving LSK_c parameters. As the s-OSCAR regularization is separable by scales, the computation of its proximity operator can be performed efficiently using parallelization over scales.

Band-wise OSCAR regularization

The present formulation applies OSCAR regularization to each specific subband of the wavelet decomposition, hence to each vector $\boldsymbol{z}_{cs,:}$ separately where $\boldsymbol{z}_{cs,:}$ gathers the K_cL wavelet coefficients across all channels in a given subband s of scale c. The band-wise (b-OSCAR) regularization thus reads:

$$\begin{aligned}\mathcal{R}_{\text{b-OSCAR}}(\boldsymbol{Z}) &= \sum_{c=1}^{C}\sum_{s=1}^{S} \Omega_{\lambda,\gamma}(\boldsymbol{z}_{cs,:}) \\ &= \sum_{c=1}^{C}\sum_{s=1}^{S}\sum_{j=1}^{K_cL} \left(\lambda+\gamma(K_cL - j)\right) |S_{LK_c}(\boldsymbol{z}_{cs,:})_j| \,, \end{aligned} \tag{4.16}$$

with S_{LK_c}, the LK_c-dimensional magnitude decreasing sorting operator. Here again, the separability of the regularizer can be exploited for an efficient implementation of the proximity operator.

Coefficient-wise OSCAR regularization

Finally we propose to apply OSCAR norm to each wavelet coefficient separately and thus to each vector $\boldsymbol{z}_{csk,:}$, where $\boldsymbol{z}_{csk,:}$ gathers the L wavelet coefficients across all channels for coefficient k in subband s of scale c. The coefficient-wise (c-OSCAR) regularization thus reads:

$$\begin{aligned}\mathcal{R}_{\text{c-OSCAR}}(\boldsymbol{Z}) &= \sum_{c=1}^{C}\sum_{s=1}^{S}\sum_{k=1}^{K_c}\Omega_{\lambda,\gamma}(\boldsymbol{z}_{csk,:}) \\ &= \sum_{c=1}^{C}\sum_{s=1}^{S}\sum_{k=1}^{K_c}\sum_{\ell=1}^{L}\left(\lambda+\gamma(L-\ell)\right)\left|S_L(\boldsymbol{z}_{csk,:})_\ell\right|,\end{aligned} \tag{4.17}$$

with S_L, the L-dimensional magnitude decreasing sorting operator. This formulation is the closest to the usual application of the group-LASSO penalty [Chun 2015, Majumdar 2012a] as it operates separately on each pixel in the transform domain. However, instead of implicitly assuming constant noise level over all receivers by taking the ℓ_2-norm, the c-OSCAR regularization allows one to weight these receivers differently thanks to the sorting step and to adapt regularization to stay robust to a space varying noise.

4.4.2 k-support norm

The ℓ_1 norm has been intensively and successfully used for solving sparse estimation problems. Its usage has often been justified as being the tightest convex envelope of the ℓ_0 pseudo-norm. However this holds only in the case of vectors with upper bounded entries by 1 as it is easy to show that $\forall \boldsymbol{z} \in \mathbb{C}^p, \|\boldsymbol{z}\|_1 \leq \|\boldsymbol{z}\|_\infty\|\boldsymbol{z}\|_0$. [Argyriou 2012] claims that exploiting this relation between the ℓ_1-norm and the ℓ_∞ norm might be sub-optimal (the relation with the ℓ_∞ norm has been studied in Section 4.4.1 with OSCAR). Instead of exploiting the ℓ_∞-norm they define a new surrogate based on the ℓ_2-norm such as: $\|\boldsymbol{z}\|_1 \leq \|\boldsymbol{z}\|_2\sqrt{\|\boldsymbol{z}\|_0}$ and studied the Gauge function[2] associated with the following convex hull:

$$C_k := \text{conv}(S_k) = \text{conv}(\boldsymbol{z}, \|\boldsymbol{z}\|_0 \leq k, \|\boldsymbol{z}\|_2 \leq 1). \tag{4.18}$$

General definition

Although the formulation 4.18 could be difficult to solve [Argyriou 2012] proposed a variational formulation defined as follows:

$$\forall \boldsymbol{z} \in \mathbb{C}^p, \|\boldsymbol{z}\|_k^{sp} := \min\{\sum_{I\in\mathcal{G}_k}\|\boldsymbol{v}_I\|_2 : \text{supp}(\boldsymbol{v}_I) \subseteq I, \sum_{I\in\mathcal{G}_k}\boldsymbol{v}_I = \boldsymbol{z}\} \tag{4.19}$$

where $\mathcal{G}_k$ denotes the set of all subsets of $\{1,\cdots,p\}$ of cardinality at most k[3] . To compute the k-support norm the authors proposed an algorithm with a complexity of $\mathcal{O}(p\log p)$. We recall the proposition hereafter:

[2] The Gauge function γ_{C_k} associated with a convex set C_k is defined as $\gamma_{C_k}(x) = \inf\{\lambda \in \mathbb{R}_+ : x \in \lambda C_k\}$

[3] The k-support norm is therefore defined as the norm whose unit-ball is the convex-hull of the set of vectors of cardinality at most equal to k and with an ℓ_2-norm not greater than 1.

Proposition 4.4.1. *For $\boldsymbol{z} \in \mathbb{C}^p$, and $k \in \mathbb{N}, k \leq p$,*

$$\|\boldsymbol{z}\|_k^{sp} = \left(\sum_{j=1}^{k-q-1} (S_p(|\boldsymbol{z}|)_j)^2 + \frac{1}{q+1} \left(\sum_{j=k-q}^{p} S_p(|\boldsymbol{z}|)_j \right)^2 \right)^{\frac{1}{2}}, \tag{4.20}$$

with: S_p the magnitude sorting operator, $S_p : \mathbb{C}^p \to \mathbb{C}^p$ such that vector $S_p(\boldsymbol{z}) = (S_p(\boldsymbol{z})_j)_{1 \leq j \leq p}$ contains the p entries of $\boldsymbol{z}$ sorted in decreasing order in magnitude and q being the unique integer in $\{0, \cdots, k-1\}$, and $z_0 = \infty$, that satisfies:

$$S_p(|\boldsymbol{z}|)_{k-q-1} \geq \frac{1}{q+1} \sum_{j=k-q}^{p} S_p(|\boldsymbol{z}|)_j \geq S_p(|\boldsymbol{z}|)_{k-q} \tag{4.21}$$

when $k = 1$ then $q = 0$ and the remaining term is the second one wich correspond to the ℓ_1-norm, in a similar way when $k = p$ then q should be equal to 0 to verify the following inequality: $S_p(|\boldsymbol{z}|)_{p-1} \geq S_p(|\boldsymbol{z}|)_p \geq S_p(|\boldsymbol{z}|)_p$ leading to the definition of the ℓ_2 norm.

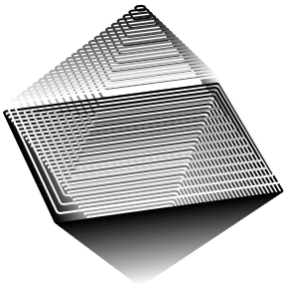
[McDonald 2016] generalized the k-support norm using the box norm, and derived the same computation for the norm. Using the variational formulation 4.4.1, one can notice that the k-support norm for $k = 1$ is equivalent to the ℓ_1-norm (unit ball with $p = 3$ is presented Figure 4.4-3a), and for $k = p$ it is equivalent to the ℓ_2-norm (unit ball with $p = 3$ is presented Figure 4.4-3c). When $k \notin \{1, p\}$, its unit ball differs from a linear combination between the ℓ_1 and ℓ_2 norm, *i.e.* elastic-net regularization. Taking for instance $p = 3$, Figure 4.4-3(a)-(c) present the unit balls for $k = 1, 2, 3$ Figure 4.4-3(d)-(e) depicts elastic-net and OSCAR norms. A comparison with Elastic-net unit ball shows that the 2-support norm has more rounded corners.

Relation to elastic-net. [Argyriou 2012] shows that the k-support norm provides tighter relaxation bound compared to elastic-net by exactly a factor of $\sqrt{2}$. For prediction purposes, [Gkirtzou 2013] demonstrated good performances of the k-support norm compared to the elastic-net regularization in functional MRI data analysis (supervised learning task).

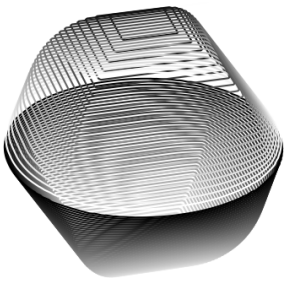
Relation to the overlapping group-LASSO. Interestingly the variational definition of the k-support norm in Eq. (4.19) shows similarities to the overlapping group-LASSO in Eq. (4.9) when the sets of groups are chosen to be $\binom{p}{k}$. Since the potential number of groups is exponential this makes the overlapping group-LASSO untractable in the case of all the possible subsets of cardinality at least k. On the other hand, the k-support norm can be efficiently computed with a complexity of $\mathcal{O}(p \log(p))$.

Dual norm. Although the variational formulation in Eq. (4.19) seems complex, the dual of the k-support norm is surprisingly easy to compute and corresponds to the ℓ_2-norm of the k largest entries:

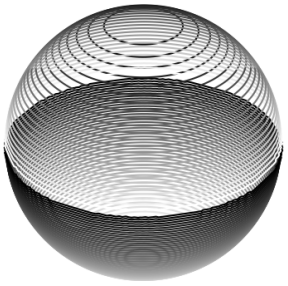
$$\|\boldsymbol{u}\|_k^{sp^\star} := \max\{\langle \boldsymbol{z}, \boldsymbol{u} \rangle : \|\boldsymbol{z}\|_k^{sp} \leq 1\} = \max\{(\sum_{i \in I} |\boldsymbol{u}|_i^2)^{\frac{1}{2}} : I \in \mathcal{G}_k\} = \left(\sum_{i=1}^{k} S_p(|\boldsymbol{u}|)_i^2 \right)^{\frac{1}{2}} \tag{4.22}$$



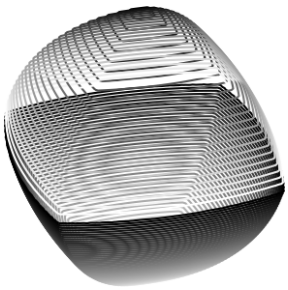
(a) 1-support norm (ℓ_1-norm)



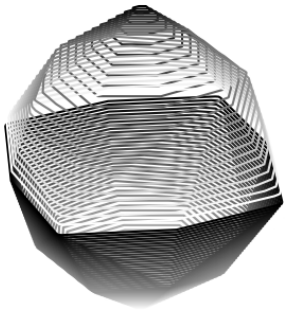
(b) 2-support norm



(c) 3-support norm (ℓ_2-norm)



(d) Elastic-net



(e) OSCAR

Figure 4.4-3: **Top row:** Unit balls of the k-support norm in $\mathbb{R}^3$ for different k values. (a) presentq the unit ball of the 1-support norm in $\mathbb{R}^3$, which corresponds to the LASSO regularization, (b) represents the unit ball of the 2-support norm, (c) shows the 3-support unit ball which correspond to the ridge regularization. **Bottom row:** Elastic-net regularization and unit ball of OSCAR norm in $\mathbb{R}^3$.

The dual of the the k-support norm could be seen as an interpolation between the ℓ_∞ and ℓ_2 since for $k = p$ we get the dual of the ℓ_2 norm and for $k = 1$ the dual takes the largest entries of the $\boldsymbol{u}$ vector which corresponds to the ℓ_∞ norm (*i.e.* the dual of the ℓ_1 norm).

Proximity Operator

[Argyriou 2012, McDonald 2016] proposed to use the k-support norm as a sparse regularization for a learning task[4], in particular they consider the square of the k-support norm, which led to the following regularizer:

$$\forall \boldsymbol{z} \in \mathbb{C}^p,\ \mathcal{R}_{\frac{\lambda}{2}\|\cdot\|_k^{sp2}}(\boldsymbol{z}) = \frac{\lambda}{2}\min\{\sum_{I\in\mathcal{G}_k} \|\boldsymbol{v}_I\|_2^2 : \mathrm{supp}(\boldsymbol{v}_I) \subseteq I, \sum_{I\in\mathcal{G}_k} \boldsymbol{v}_I = \boldsymbol{z}\} \tag{4.23}$$

where $\mathcal{G}_k$ denotes the set of all subsets of $\{1, \cdots, p\}$ of cardinality at most k. According to [McDonald 2016, Corollary 16] the proximity operator of the k-support norm is defined

$$\left(\mathrm{prox}_{\frac{\lambda}{2}\|\cdot\|_k^{sp2}}(\boldsymbol{z})\right)_i = \frac{\theta_i z_i}{\theta_i + \lambda} \tag{4.24}$$

$$\theta_i = \begin{cases} 1, & \text{if } \alpha|z_i| > \lambda + 1 \\ \alpha|z_i| - \lambda, & \text{if } \lambda + 1 \geq \alpha|z_i| \geq \lambda \\ 0, & \text{if } \lambda > \alpha|z_i| \end{cases}$$

[4]The k-support norm encourages $\boldsymbol{z}$ to be a sum of limited number of vectors with small support.

where α is chosen such that $\sum_{i=1}^{p} \theta_i(\alpha) = k$. Even though the computation of this proximity operator is complicated it can be done efficiently using Algorithm 5.

Algorithm 5: Proximity operator of $\frac{\lambda}{2}\|\cdot\|_k^{sp2}$.

Input: $\boldsymbol{z} \in \mathbb{C}^p/\{0\}$, $\lambda \in \mathbb{R}$;

Output: $\boldsymbol{x} = \text{prox}_{\frac{\lambda}{2}\|\cdot\|_k^{sp2}}(\boldsymbol{z})$;

Compute $\{\alpha^{2i}, \alpha^{2i+1}\}_{i=1}^{p} = \left\{\frac{\lambda}{|z_i|}, \frac{\lambda+1}{|z_i|}\right\}_{i=1}^{p}$;

Sort α^i such that $\alpha^i \leq \alpha^{i+1}$;

Find i such that:

$\sum_{j=1}^{p} \theta_j(\alpha^i) \leq k$ and $\sum_{j=1}^{p} \theta_j(\alpha^{i+1}) \geq k$;

Interpolate α^* between α^i and α^{i+1} such that $\sum_{j=1}^{p} \theta_j = k$;

$$\theta_i(\alpha^*) = \begin{cases} 1, & \text{if } \alpha^*|z_i| > \lambda + 1 \\ \alpha^*|z_i| - \lambda, & \text{if } \lambda + 1 \geq \alpha^*|z_i| \geq \lambda \\ 0, & \text{if } \lambda > \alpha^*|z_i| \end{cases}$$

$$x_i = \left(\text{prox}_{\frac{\lambda}{2}\|\cdot\|_k^{sp2}}(\boldsymbol{z})\right)_i = \frac{\theta_i z_i}{\theta_i + \lambda};$$

Application to MR image reconstruction

We implemented the straightforward version of the k-support norm regularization, that consists of flattening all wavelet coefficients and thus discarding the multi-scale and multi-channel structure (*e.g.* in scales, sub-bands, coefficients and channels) of $\boldsymbol{Z}$. Following the wording used for OSCAR-norm regularization, this implementation corresponds to the global version of k-support norm and it models the signal as the g-OSCAR regularization (cf Section 4.4.1). The wavelet coefficients are stacked together, leading to a single but large vector with entries $(z_j)_{1 \leq j \leq LN_\Psi}$, where we remind that $N_\Psi = S\sum_{c=1}^{C} K_c$ and $N_\Psi = N$ when $\boldsymbol{\Psi}$ is orthogonal. The k-support regularization then reads:

$$\mathcal{R}_{\frac{\lambda}{2}\|\cdot\|_k^{sp2}}(\boldsymbol{Z}) = \frac{\lambda}{2}\min\{\sum_{I \in \mathcal{G}_k} \|\boldsymbol{v}_I\|_2^2 : \text{supp}(\boldsymbol{v}_I) \subseteq I, \sum_{I \in \mathcal{G}_k} \boldsymbol{v}_I = \boldsymbol{Z}\} \tag{4.25}$$

4.5 Experimental setting & Results

To assess the performances of the different regularization penalties in terms on MR image quality, we used prospectively accelerated non-Cartesian real acquisitions and proceed in four steps. First, we compared the global regularization regardless of the wavelet decomposition structure, in particular we compared the LASSO Section 4.3.1, Elastic-net Section 4.3.2, the global version of OSCAR (g-OSCAR) and the k-support norm regularization. A particular attention was paid to the performances of the ℓ_∞ vs ℓ_2 norm assumptions behind respectively OSCAR and k-support norm. Second, we made some comparisons between the coefficient formulation of OSCAR (c-OSCAR) and group-LASSO (i.e. CALM method [Majumdar 2012a]), since both rely on the same assumptions

in terms of structured sparsity. Third, we compared the four OSCAR-based formulations, described in Section 4.4.1. Finally, an overall comparison was conducted between self-calibrating and calibration-less state-of-the-art methods, such as ℓ_1-ESPIRiT [Uecker 2014] and AC-LORAKS [Haldar 2016, Kim 2017b].
Hereafter, we describe the experimental setup used for validation purposes.

4.5.1 Experimental setup

All numerical experiments were conducted on 2D k-space data even though the proposed framework could be extended to 3D imaging (i.e. 3D k-space data) directly. Hence, we only report results on slices.

Prospective non-Cartesian acquisition

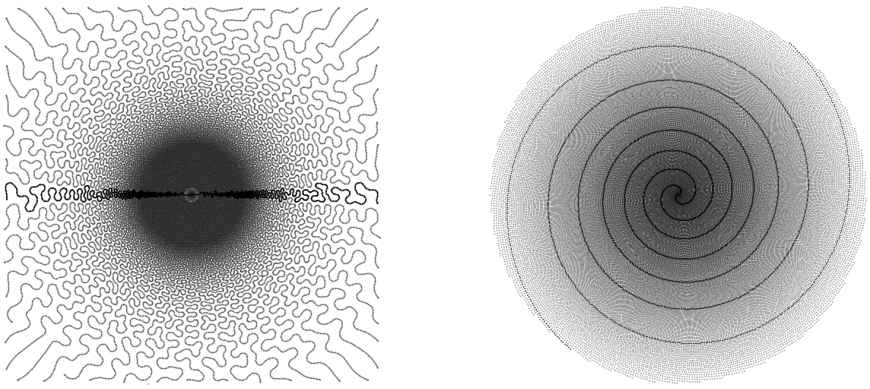
We consider the reconstruction of an ex vivo human brain with an in plane resolution of $0.39 \times 0.39\text{mm}^2$ and a slice thickness of 3mm for different acquisition scenarios prospectively accelerated on the same 7T MR system and with a $L = 32$-channel coil (Nova Medical). Ex vivo imaging offers the best imaging conditions to remain insensitive to motion when validation is performed on truly accelerated acquisitions. In the literature, for the sake of simplicity, most of reconstruction algorithms are usually validated on retrospectively under-sampled k-space in vivo data. However in this context their robustness to actual acquisition set ups is overlooked. Spiral trajectories [Lee 2003] were generated for different under-sampling factors as detailed in Tab. 4.1. More advanced sampling schemes were also designed using Sparkling (Spreading Projection Algorithm for Rapid K-space samplING) [Lazarus 2019a, Boyer 2016]. Sparkling method generates physically plausible trajectories with improved sampling efficiency and robustness to gradient imperfections and lower sensitivity to off-resonance artifacts as demonstrated in [Lazarus 2019a]. The resulting sampling schemes are known to reach higher image quality for a given scan time, compared to state-of-the art trajectories (*e.g.* spiral or radial in 2D imaging). We explain hereafter while in Tab. 4.1 we specifically achieve in the non-Cartesian setting an higher acceleration in time as compared to the under-sampling factor expressed in the number of measurements over the number of image pixels. We recall the difference between acceleration factor (AF) in time and under-sampling factor (UF) in the number of measurements: $\text{AF} = n/n_c$ where n_c is the number of shots. For instance, we used $n = 512$ and varied n_c between 64 down to 26 corresponding to acceleration in time from 8 to 20, as reported in Tab. 4.1. In the same time, an oversampling (with a factor of 6) along each shot was performed ($n_s = 3,072$ samples per shot) as usually done in non-radial non-Cartesian trajectories, hence the under-sampling factor defined as $\text{UF} = N/M$ with $M = n_c \times n_s$ varied in a more limited range from 1.3 to 3.3.

The other acquisition parameters were set as follows: $\text{FOV} = 200 \times 200$ mm^2, TR = 550 ms (for 11 slices), TE = 30 ms, BW = 100 kHz, $T_{\text{obs}} = 30.72$ms (long readout, i.e. each shot is traversed in T_{obs}) and FA=25°. In Tab. 4.1), we summarized the different data sets we considered by varying the under-sampling or acceleration factor for both Sparkling

Table 4.1: Acquisition parameters used for prospective CS ex vivo T_2^* acquisition on the same human brain.

Idx	Sampling scheme	Acceleration Factor	Undersampling Factor
1.	Sparkling, cf Fig. 4.5-4(a)	8	1.3
2.		10	1.7
3.		12	2
4.		15	2.5
5.		20	3.3
6.	Spiral, cf Fig. 4.5-4(b)	8	1.3
7.		10	1.7
8.		12	2
9.		15	2.5
10.		20	3.3

and Spiral readouts (trajectories are displayed Figure 4.5-4). A fully sampled Cartesian dataset was acquired and the corresponding image was reconstructed using inverse FFT for each receiver coil and the square root of the sum of squares was eventually computed over all channels to form a single image Figure 4.5-5 **(a)** which will serve as ground truth using the same sequence parameters (matrix size: $N = 512 \times 512$ or $n = 512$).



(a) Sparkling trajectory (b) Spiral trajectory

Figure 4.5-4: Non-Cartesian in-out variable density sampling schemes used in prospective CS at 7 Tesla. (a): 20-fold accelerated-in-time Sparkling trajectories (26 shots). (b): 20-fold accelerated-in-time spiral trajectories with the same number of shots and measurements as Sparkling. For details about their generation, see [Lazarus 2019a] for Sparkling and [Lee 2003] for spiral imaging. The trace of a single in-out shot is highlighted.

Reconstruction parameters

Algorithm 3 was run until the maximum number of iterations ($T = 150$) or a tolerance threshold on the image solution was reached, which appears sufficient to converge. Moreover, we used for $\mathbf{\Psi}$ a Daubechies 4 orthogonal wavelet transform (OWT) with $C = 4$ decomposition scales (i.e., $N_\Psi = N$). Note that MR image quality could probably be improved using redundant multiscale transforms such as undecimated bi-orthogonal wavelet transforms or curvelets as shown in [Cherkaoui 2018, Ma 2017] but those transforms may increase the time and memory requirements of the overall algorithm. The hyper-parameters were set using a grid-search procedure such that the SSIM [Wang 2004] score was maximized. In the next section we also report the peak Signal to Noise Ratio (p-SNR) and the normalized root mean square error (NRMSE) scores.

4.5.2 Results

Before presenting in more details the results corresponding to the different regularization schemes, we first computed the zero-order solution (iFT). Since we used non-Cartesian trajectories, the forward model relies on the non-Uniform or Non-equispaced Fast Fourier Transform (NUFFT or NFFT) [Fessler 2003, Keiner 2009], which is not invertible, a conjugate gradient descent was performed on the data-fidelity term Eq. (4.1). In what follows, we call the corresponding estimator the *least squares solution*. Although not perfect, those results presented in Table 4.2 will be taken as a baseline for comparison purposes.

Table 4.2: Performances of the least squares (LS) solution (zero-order reconstruction) on the dataset Table 4.1. The results in red box are displayed in Figure 4.5-5 .

	Idx	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
iFT	SSIM	0.908	0.906	0.902	0.894	0.884	0.927	0.925	0.922	0.916	0.911
	p-SNR	30.57	29.19	28.89	28.39	28.25	31.55	31.64	31.47	29.86	29.79
	NRMSE	0.1463	0.1715	0.1774	0.1880	0.1911	0.1307	0.1297	0.1391	0.1588	0.1599

The similarity scores (SSIM and p-SNR) decreases as the under-sampling factor increases for both Sparkling and Spiral reconstructions, which is not surprising as the problem gets harder with fewer measurement. We illustrate the results provided by the least square solution on the 20-fold accelerated scans for both Sparkling and spiral acquisition on Figure 4.5-5. Despite the high image quality metrics for Spirals compared to Sparkling (for same acceleration factor), Spiral MR reconstructions are contaminated by severe off-resonance artifacts (cortical ribbon highlighted with red arrow in Figure 4.5-5 **(c)**) whereas Sparkling ones are not. These results illustrate the much better point spread function (PSF) associated with Sparkling sampling schemes as shown in [Lazarus 2019a] and highlighted Figure 4.5-5.

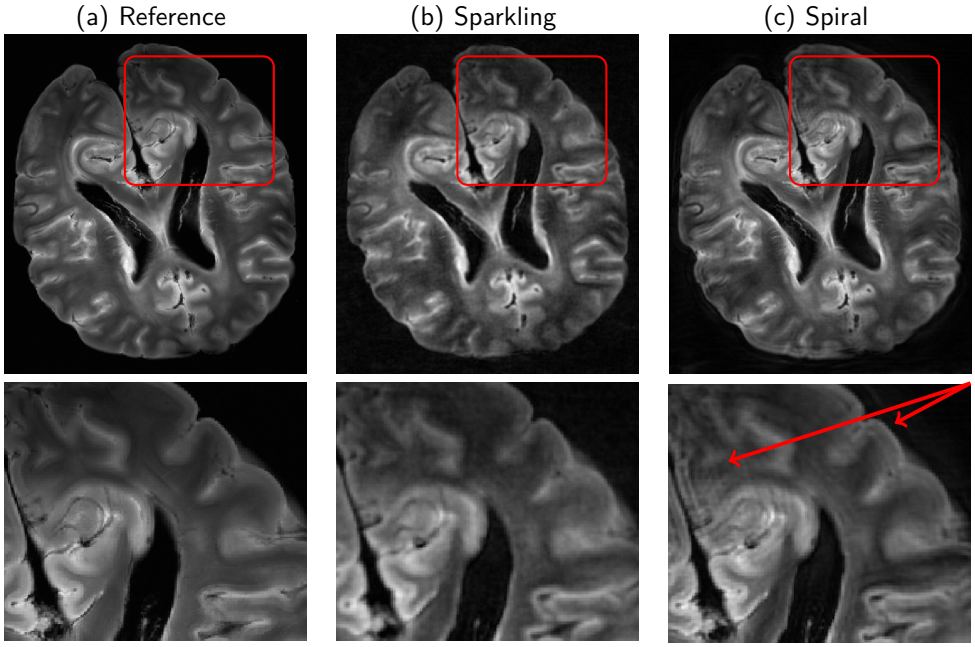


Figure 4.5-5: **Top:** (a) Cartesian reference image. Zero-order Fourier reconstructions based on the minimization of the least square data consistency term for 20-fold accelerated-in-time (26 shots) (b) Sparkling (26 shots) and (c) spiral trajectories. **Bottom:** Respective zoom over the red frame. Red arrows show the presence of gradient-related artifacts.

Comparison of global regularization schemes

Table 4.3 summarizes the performances of different global regularization terms (regardless of the structure of the wavelet decomposition), it includes the LASSO (Eq. (4.2)), Elastic net (Eq. (4.4)), global version of OSCAR (g-OSCAR) (Eq. (4.14)) and the k-support norm (Eq. (4.25)). The k-support norm achieves the highest SSIM score for the entire dataset, as well as the best results in terms of p-SNR (with up to 2dB above its competitors) and NRMSE. Those results are confirmed visually as shown in Figure 4.5-6 for both spiral and Sparkling images. They also support the claim of [Argyriou 2012] in Section 4.4.2 *i.e.* that the relation between the ℓ_1-norm and the ℓ_∞ is actually sub-optimal when it comes to finding the tightest convex envelope of the ℓ_0-norm.

Despite the optimization of hyper-parameters for all the regularization schemes, the Elastic-net and the LASSO present the lowest similarity scores. In particular, when comparing the LASSO and Elastic-net solutions, while both show the same SSIM score the p-SNR value is actually lower for Elastic-net. This confirms that these two metrics capture complementary information on the image quality. Visually, the Elastic-net reconstruction appears to be smoother. The potential reason is the high value of the ridge parameter, however Figure 4.5-7 presents the evolution of the SSIM score as a function of the two hyper-parameters for Elastic-net regularization, where the selected one is depicted with a black cross on Figure 4.5-7. The latest figure illustrate the compromise between the ℓ_1 and ℓ_2 regularization performed by the Elastic-net regularization.

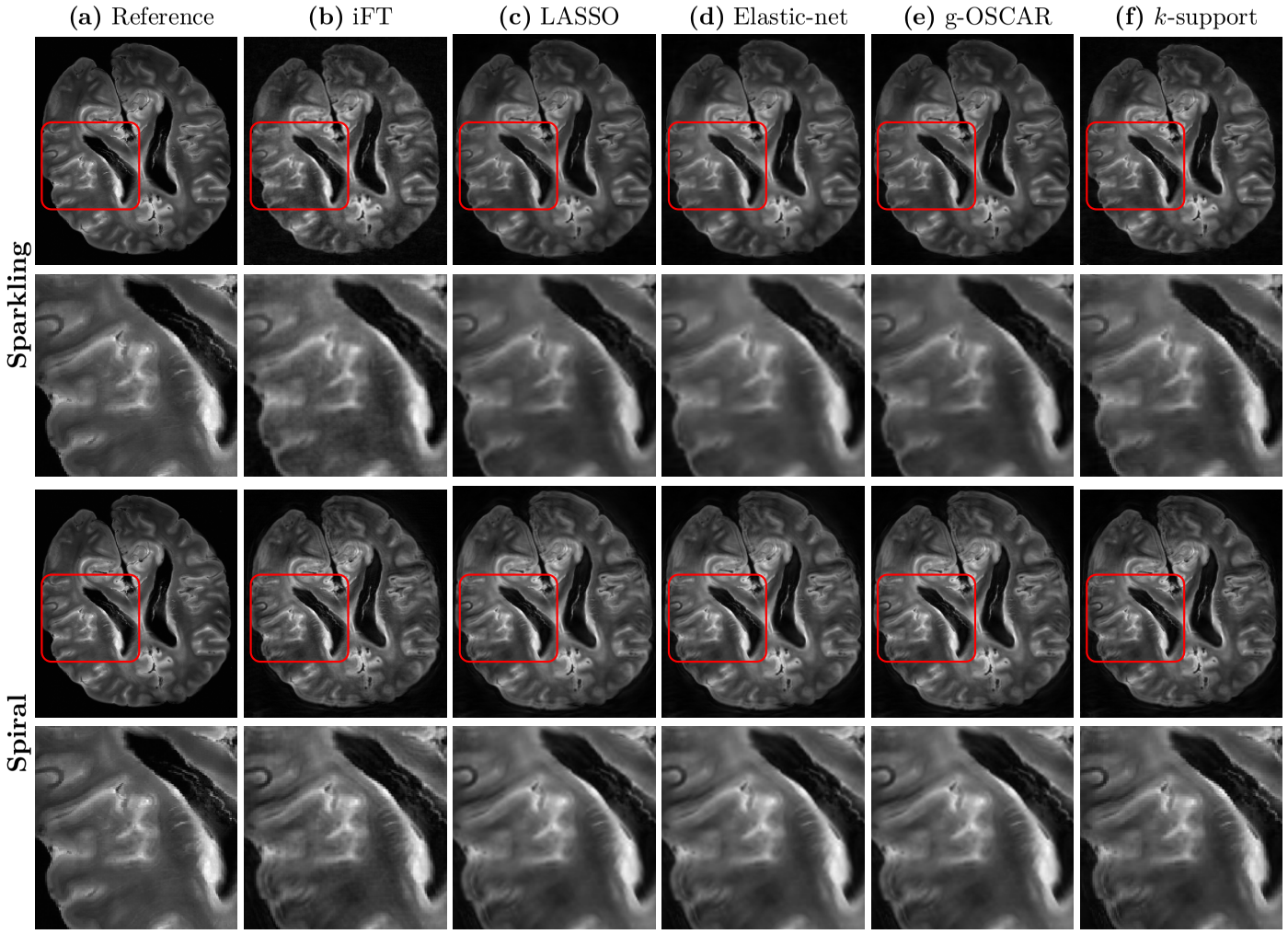


Figure 4.5-6: **First-third rows:** Reconstructed MR images (square-root of the sum-of-squares) from 20-fold accelerated Sparkling (first two rows) and spiral acquisitions (last two rows) using different regularization schemes. **(a)** Cartesian reference. **(b)** zero-order inverse Fourier Transform. Reconstruction based on **(c)** LASSO regularization $\mathcal{R}_{\text{LASSO}}$. **(d)** Elastic-net regularization $\mathcal{R}_{\text{Elastic-net}}$, **(e)** global version of OSCAR $\mathcal{R}_{\text{g-OSCAR}}$ and **(f)** square of k-support norm $\mathcal{R}_{\frac{\lambda}{2}\|\cdot\|_k^{sp2}}$. **Second-fourth rows:** Respective zooms in the red frames .

Table 4.3: Reconstruction of 20-fold accelerated scans acquired with sparkling and spiral trajectories. Comparison of global regularization functions (LASSO, Elastic-net, g-OSCAR and k-support norm) with hyper-parameters chosen to as to maximize the SSIM score. Best values appear in **bold font**. The images associated with the scores highlighted in red boxes are displayed in Figure 4.5-6.

Idx	LASSO			Elastic-Net			g-OSCAR			k-support norm		
	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE
1.	**0.923**	30.34	0.1501	**0.923**	30.25	0.1517	**0.923**	30.52	0.1471	**0.923**	**31.23**	**0.1356**
2.	0.920	29.27	0.1699	0.920	29.20	0.1712	0.920	29.21	0.1711	**0.922**	**31.03**	**0.1386**
3.	0.916	28.81	0.1790	0.916	28.75	0.1803	0.916	28.81	0.1792	**0.919**	**30.89**	**0.1410**
4.	0.899	29.09	0.1686	0.900	28.75	0.1840	0.912	29.28	0.1700	**0.913**	**30.46**	**0.1486**
5.	0.899	29.09	0.1734	0.899	28.99	0.1757	0.899	29.12	0.1728	**0.900**	**30.29**	**0.1510**
6.	0.932	30.52	0.1470	0.932	30.08	0.1548	0.932	30.12	0.1540	**0.933**	**31.58**	**0.1302**
7.	0.927	29.66	0.1624	0.928	29.44	0.1667	0.928	29.76	0.1606	**0.928**	**30.82**	**0.1420**
8.	0.922	**31.46**	**0.1320**	0.923	30.94	0.1402	0.922	29.35	0.1683	**0.924**	31.16	0.1366
9.	0.919	29.06	0.1751	0.920	28.81	0.1790	0.920	29.18	0.1715	**0.922**	**30.83**	**0.1420**
10.	0.916	29.69	0.1619	0.916	29.60	0.1635	0.916	29.72	0.1614	**0.917**	**30.85**	**0.1416**

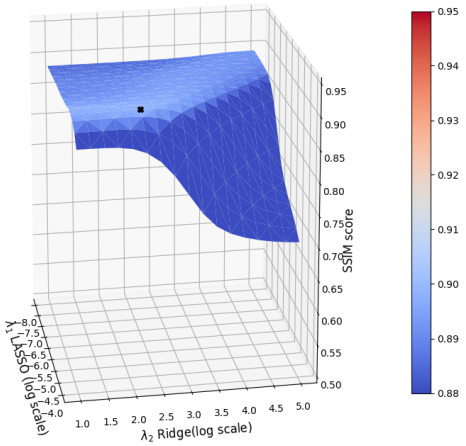


Figure 4.5-7: Evolution of the SSIM score computed between the reconstructed image from 20-fold accelerated Sparkling trajectory using Elastic-net regularization and the reference image. The black cross displays the hyper-parameters that maximize the SSIM score.

Comparison of coefficient-based regularization schemes

Additionally, we compare the group-LASSO and the c-OSCAR regularization as both impose a group-structure on the same wavelet coefficient across channels. Table 4.4 summarizes the results of the magnitude image obtained using the square root of the sum of square. Once again the hyper-parameters were optimized such that the SSIM index was set to its maximal value, however we also reported the p-SNR and NRMSE scores. Although for all tested sampling patterns the best results were obtained using the c-OSCAR regularization, the difference is not significant enough to notice a notable impact on the final images as presented in Figure 4.5-8.

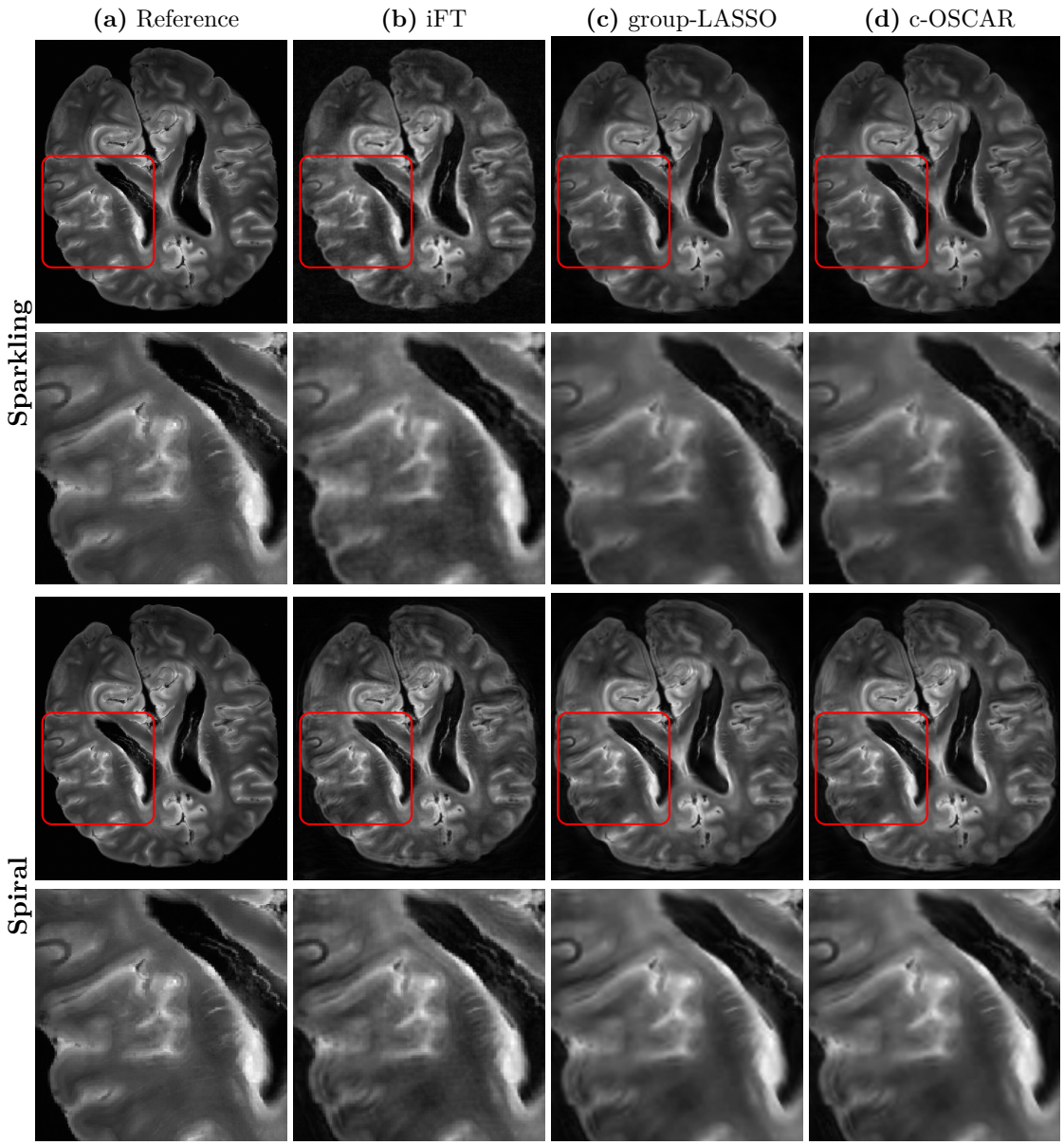


Figure 4.5-8: Reconstructed MR images (square-root of the sum-of-squares) from 20-fold accelerated Sparkling (first two rows) and spiral acquisitions (last two rows) using different regularization. **(a)** Cartesian reference. **(b)** inverse Fourier Transform. Reconstruction based on **(c)** group-LASSO regularization $\mathcal{R}_{\text{group-LASSO}}$. **(d)** coefficient version of OSCAR $\mathcal{R}_{\text{c-OSCAR}}$. **Second and Forth rows:** Sparkling and Spiral respective zooms in the red frame .

Comparison of OSCAR-norm regularization schemes

As regards the different versions of OSCAR-based regularization, Table 4.5 summarizes their respective performances.

For the five Sparkling data sets, the best scores were achieved with the b- and c-OSCAR versions, where each subband and each wavelet coefficient are treated separately, respectively. This is not really surprising as the other versions mix up different orientation or scale details together. The fact that the coefficient-wise version performs slightly better

Table 4.4: Quantitative comparison of coefficient based regularization schemes (c-OSCAR and group-LASSO) with hyper-parameters set to maximize the SSIM score. Best values appear in **bold font**.

Idx	group-LASSO			c-OSCAR		
	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE
1.	**0.924**	30.51	0.1473	**0.924**	**30.96**	**0.1398**
2.	0.921	29.54	0.1647	**0.922**	**29.59**	**0.1636**
3.	0.917	29.05	0.1741	**0.918**	**29.33**	**0.1688**
4.	0.912	28.87	0.1778	**0.913**	**29.66**	**0.1624**
5.	0.897	28.59	0.1836	**0.901**	**29.77**	**0.1604**
6.	0.932	**30.95**	**0.1400**	**0.933**	30.66	0.1447
7.	**0.928**	29.87	0.1585	**0.928**	**29.92**	**0.1577**
8.	**0.922**	29.55	0.1645	**0.922**	**31.50**	**0.1319**
9.	0.920	29.21	0.1711	**0.921**	**29.72**	0.1613
10.	0.915	29.52	0.1650	**0.916**	**29.66**	**0.1623**

Table 4.5: Quantitative comparison of OSCAR-norm regularization schemes with hyper-parameters set to maximize the SSIM score. Best values appear in **bold font**.

Idx	g-OSCAR			s-OSCAR			b-OSCAR			c-OSCAR		
	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE
1.	0.923	30.52	0.1471	0.925	31.66	0.1290	**0.926**	**31.68**	**0.1287**	0.924	30.96	0.1398
2.	0.920	29.21	0.1711	0.921	29.62	0.1632	**0.922**	**30.28**	**0.1512**	**0.922**	29.59	0.1636
3.	0.916	28.81	0.1792	**0.918**	28.40	0.1878	**0.918**	**29.78**	**0.1602**	**0.918**	29.33	0.1688
4.	0.912	29.28	0.1700	0.912	29.05	0.1742	**0.913**	29.52	0.1650	**0.913**	**29.66**	**0.1624**
5.	0.899	29.12	0.1728	0.896	28.35	0.1889	0.899	29.52	0.1650	**0.901**	**29.77**	**0.1604**
6.	0.932	30.12	0.1540	0.931	**30.70**	0.1440	**0.933**	30.36	0.1498	**0.933**	30.66	**0.1447**
7.	0.928	29.76	0.1606	0.927	**30.50**	0.1474	**0.928**	29.92	**0.1576**	**0.928**	29.92	0.1577
8.	0.922	29.35	0.1683	**0.923**	29.74	0.1609	0.921	31.22	0.1358	0.922	**31.50**	**0.1319**
9.	0.920	29.18	0.1715	0.920	29.89	0.1581	0.919	**30.22**	**0.1522**	**0.921**	29.72	0.1613
10.	**0.916**	29.72	0.1614	0.914	29.40	0.1673	0.915	**29.84**	**0.1592**	**0.916**	29.66	0.1623

than the subband-wise one suggests that localized regularization in space preserves much better details as the acceleration factor increases. Similar findings were replicated on an ex-vivo baboon brain (results not shown). For the five spiral data sets, the best SSIM scores were yielded in the vast majority of cases by the b- and c-OSCAR regularization schemes illustrating the consistency between the two readout scenarios.

Comparison with other reconstruction methods

Tab. 4.6 summarizes the quantitative performances (SSIM, p-SNR and NRMSE) of competing reconstruction approaches, computed over the magnitude images. These competitors embed i) zero-order inverse NFFT (no regularization, i.e. LS solution), ii) CALM solution which implements a group-LASSO regularization and self-calibrating techniques, namely ℓ_1-ESPIRiT as well as k-space based calibration-less method AC-LORAKS. As in sparkling and spiral imaging, the k-space center is heavily sampled, ℓ_1-ESPIRiT performs well for extracting the sensitivity maps (i.e. low resolution information) which are then used for solving the ℓ_1-norm regularized CS-SENSE reconstruction problem. The bottleneck in LORAKS method lies in the gridding step required to project the non-Cartesian sampling pattern onto a binary Cartesian mask. This step degrades the image quality in the end and demonstrates that AC-LORAKS may not be really compliant with non-Cartesian sampling.

Table 4.6: Quantitative comparison of state-of-the-art methods for non-Cartesian PI-CS reconstruction using either a calibration-less or self-calibrating approach. The sparkling and spiral solutions on the red frame are shown in Figure 4.5-9 and Figure 4.5-10, respectively.

Idx	No regularization			CALM			ℓ_1-ESPIRiT			AC-LORAKS		
	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE	SSIM	p-SNR	NRMSE
1.	0.908	30.57	0.1463	0.924	30.51	0.1473	0.911	27.82	0.1946	0.894	26.09	0.2375
2.	0.906	29.19	0.1715	0.921	29.54	0.1647	0.906	26.58	0.2246	0.897	26.23	0.2340
3.	0.902	28.89	0.1774	0.917	29.05	0.1741	0.904	27.17	0.2099	0.893	26.25	0.2333
4.	0.894	28.39	0.1880	0.912	28.87	0.1778	0.900	26.29	0.2323	0.884	25.94	0.2418
5.	0.884	28.25	0.1911	0.897	28.59	0.1836	0.885	26.48	0.2272	0.753	25.52	0.2536
6.	0.927	31.55	0.1307	0.932	30.95	**0.1400**	0.927	26.37	0.2300	0.921	27.55	0.2008
7.	0.925	31.64	**0.1293**	**0.928**	29.87	0.1585	0.925	26.07	0.2382	0.921	27.54	0.2010
8.	0.922	31.47	0.1391	**0.922**	29.55	0.1645	0.922	26.27	0.2328	0.919	27.23	0.2084
9.	0.916	29.86	0.1588	**0.920**	29.21	0.1711	0.916	26.33	0.2311	0.911	26.67	0.2221
10.	0.911	29.79	0.1599	**0.915**	29.52	0.1650	0.910	25.86	0.2439	0.902	26.23	0.2338

In terms of SSIM and p-SNR scores, CALM regularization outperforms all other techniques. However, CALM performances remain below those of OSCAR, whether it is the subband- or coefficient-wise version. This confirms that extending the group-LASSO penalization using a pairwise ℓ_∞-norm between channels instead of the global the ℓ_2-norm brings more flexibility to account for varying SNR across channels. Last but not least, in Figs. 4.5-9-4.5-10 we show that calibration-less reconstruction techniques such as OSCAR and CALM better preserve phase information in comparison with ℓ_1-ESPIRiT, AC-LORAKS. In this comparison we also provide the phase information related to each reconstructions. As there is neither sensitivity map extraction nor gridding step in domain-based calibration-less approaches, any phase-related error (*e.g.* phase shift or wrapping) in those steps does not propagate in the image reconstruction itself. Such artifacts are particularly visible in ℓ_1-ESPIRiT for spiral imaging (see red arrows in Fig. 4.5-10). Phase images based on AC-LORAKS are also corrupted by artifacts that may be due to the

gridding operation, especially for Sparkling imaging (see red arrows in Fig. 4.5-9).

4.6 Conclusion

Non-Cartesian trajectories have been used for many different applications and have recently met a renew of interest to accelerate scan time. However, in the multi-receiver coil context most of reconstruction methods are more suited for Cartesian acquisitions than for non-Cartesian ones. This is especially the case of k-space based calibration-less techniques.

Indeed, we formulated CS-PI image reconstruction as an inverse ill-posed problem and made use of NFFT for dealing with non-Cartesian trajectories and structured sparse induced norm for regularization purposes. We instantiated different versions of structured sparse norm regularization to exploit the redundant information provided by each coil. Then, we relied on state-of-the-art convex nonsmooth optimization tools, namely the Condat-Vù algorithm with sound convergence properties to compute the global minimizer of the derived cost function. The main advantages of this algorithm are first that it can efficiently deal with analysis-based regularization which is known to provide better results than synthesis-based priors. Second, this algorithm is highly flexible, as it can be used the same way for a large range of penalty terms and the parameters that control its convergence speed are easier to tune than those involved in ADMM (aka, split Bregman) methods [Ghadimi 2014], and the poor convergence rate [Deng 2016].

The experimental validation was made on prospectively collected ex vivo human brain data at 7 Tesla using different acquisition setups, *i.e.* various sampling patterns and acceleration factors. Image quality reaches a maximum for two regularization schemes, namely c-OSCAR and k-support norm regularization schemes. A comparison with self-calibrating (ℓ_1-ESPIRiT) and calibration-less (K-space based AC-LORAKS and wavelet based CALM) methods was conducted. Quantitative structural similarity scores show that OSCAR-based approaches improved the overall image quality with Sparkling and Spiral trajectories compared to their competitors. However, the visual impression show that the k-support norm outperformed the c-OSCAR regularization. On top of the improvement in magnitude images, phase information was better preserved for sparsity-based calibration-less reconstruction methods. Interestingly, the use of a gridding step in AC-LORAKS reconstruction constitutes the major bottleneck of k-space based calibration-less reconstruction methods when applied to non-Cartesian acquisitions.

In this work, both parameters have been set by maximizing an image quality score – here the SSIM – between a ground truth image reconstructed from a non-accelerated Cartesian acquisition and the image solution of Eq. (4.1).

Figure 4.5-9: **Top:** Reconstructed MR images (magnitude) from 20-fold accelerated Sparkling acquisitions using different methods. **(a)** Cartesian reference. **(b)** Zero-order reconstruction obtained by solving the least square criterion (data consistency term) involving the NFFT operator. **(c)** Calibration-less regularized reconstruction based on the subband-wise OSCAR formulation. **(d)** Calibration-less regularized reconstruction based on k-support penalization. **(e)** Self-calibrating ℓ_1-ESPIRiT reconstruction. **(f)** Auto-calibrated (AC) LORAKS reconstruction. **Middle:** Respective zooms in the red frames. Reconstructed MR images (phase). **Bottom:** Wrapped phase images extracted using the virtual coil approach.

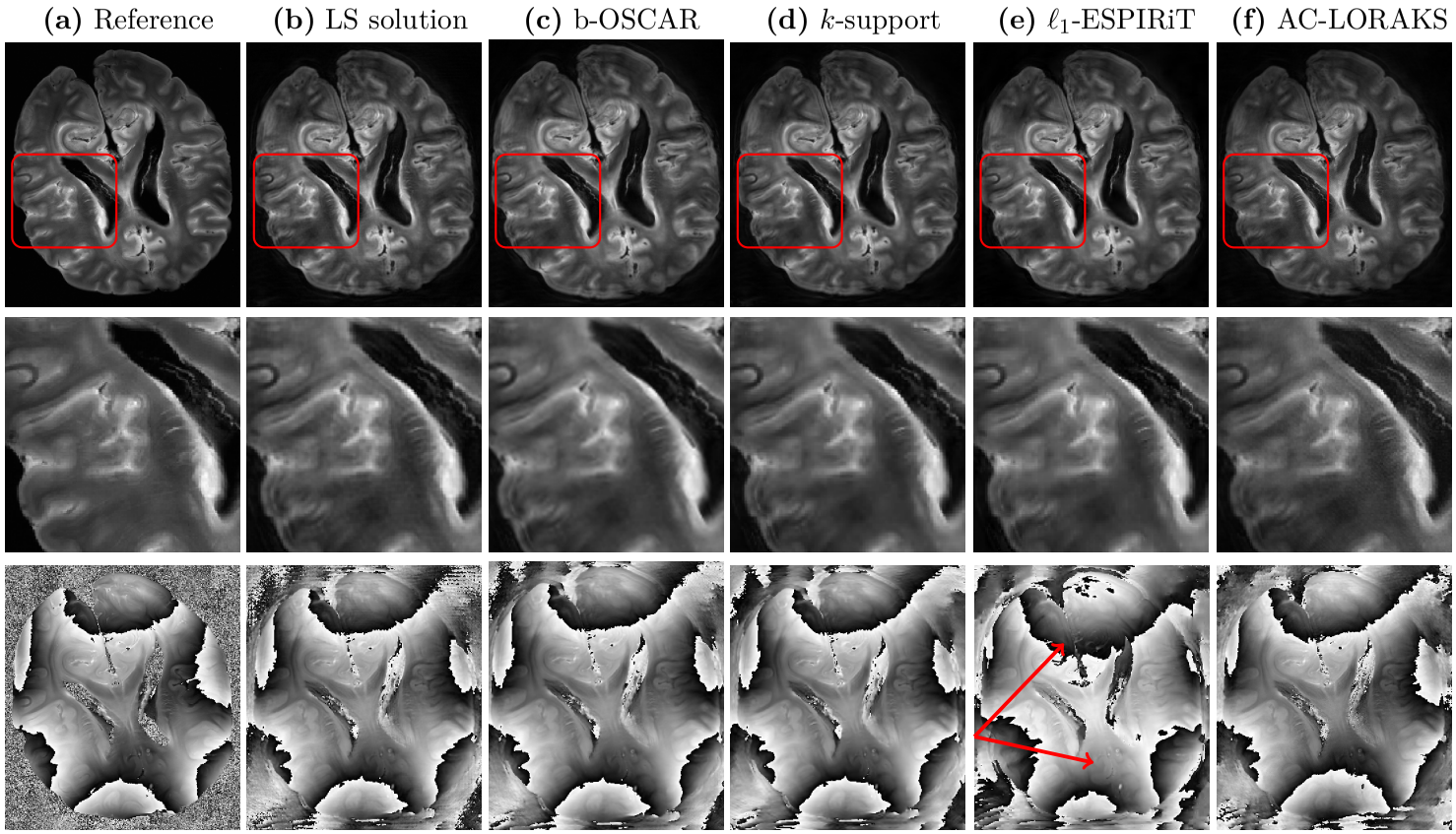


Figure 4.5-10: **Top:** Reconstructed MR images (magntiude and phase) from 20-fold accelerated spiral acquisitions using different methods. **(a)** Cartesian reference. **(b)** Zero-order reconstruction obtained by solving the least square criterion (data consistency term) involving the NFFT operator. **(c)** Calibration-less regularized reconstruction based on the subband-wise OSCAR formulation. **(d)** Calibration-less regularized reconstruction based on k-support penalization. **(e)** Self-calibrating ℓ_1-ESPIRiT reconstruction. **(f)** Auto-calibrated (AC) LORAKS reconstruction. **Middle:** Respective zooms in the red frames. **Bottom:** Wrapped phase images extracted using the virtual coil approach [Parker 2014] .

4.7 Discussion

In the future, we will explore alternative automatic hyper-parameters settings either based on (generalized) cross-validation [Hastie 2005] across slices for instance as a larger data set is necessary to implement this strategy, or on statistical inference (*e.g.* SURE estimator [Stein 1981]), which requires the knowledge of noise statistics. The overall SSIM scores may also be improved using redundant wavelet transforms either considering undecimated multiscale decompositions or more sophisticated transforms such as curvelet or shearlet transforms [Cherkaoui 2018, Ma 2017]. Although the aforementioned approaches might improve MR image quality for each reconstruction, it could also increase the computation time and the memory loading that can be critical especially for 3D high-resolution multi-channel reconstruction.

This work did not address the case of accelerated Cartesian acquisition and reconstruction. In this setup, the proposed approach can be useful as far as low acceleration factors are considered. However when the acceleration is pushed further and the sampling is kept over the Cartesian grid, the proposed calibration-less methodology – including CALM and CLEAR methods – may perform worse than state-of-the-art self-calibrating methods (ℓ_1-ESPIRiT) or domain based calibration-less (P-LORAKS). Nevertheless one could think of a self-calibrating approach with a more sophisticated regularization (*e.g.* OSCAR norm or group-LASSO) as done in [Chun 2015].

In order to speed-up reconstruction, many different method have been proposed by the Deep-Learning community, in particular unrolled optimization [Kamilov 2016, Yang 2016, Adler 2018] methods seems the most promising one. With this perspective, [Wen 2016] generalizes the structured sparsity norm to use it on deep neural network. However deep-learning methods requires a large database of raw data (*i.e.* high resolution multi-channel complex k-spaces unavailable until recently [Zbontar 2018]), whereas traditional methods such as the one proposed are not. Importantly the proposed method has theoretical guarantees of convergence and is made of fully interpretive steps which make it more reliable and robust to small perturbations than recent deep learning approaches (such as [Zhu 2018, Mardani 2018]) proposed for medical image reconstruction as it has been shown by [Antun 2019].

In terms of applications, the proposed method can be easily extended to 3D imaging and used as such in isotropic high-resolution susceptibility weighted imaging (SWI). In this context, the scan time is lengthy and in the clinical realm SWI acquisitions will benefit from highly accelerated non-Cartesian encoding schemes to reach 800 μm in a scan time of 1-2 min. Noticeably, in SWI imaging, post-processing is applied to phase information in order to reveal potential alteration of the microvascular brain network. The fact that our calibration-less regularized reconstruction better preserves phase information is thus an asset for its utilization in SWI application for diagnostic purposes.

* * *
* *
*

Chapter 5

Online MR image reconstruction

Chapter Outline

Parts of this chapter were presented at international conference in 2019:

L. El Gueddari, P. Ciuciu, E. Chouzenoux, A. Vignaud and J.-C. Pesquet. *Online compressed sensing MR image reconstruction for high resolution T_2^* imaging.* In Proceedings of the 27th Annual Meeting of ISMRM, page 4679, Montreal, QC, Canada, 2019.

L. El Gueddari, E. Chouzenoux, A. Vignaud, J.-C. Pesquet and P. Ciuciu. *Online MR image reconstruction for compressed sensing acquisition in $T2^*$ imaging.* In Wavelets: Applications in Signal and Image Processing XVIII. International Society for Optics and Photonics, 2019.

This chapter propose a new way for accelerating MR image reconstruction in the context of CS-accelerated acquisitions.

5.1 Introduction

Instead of performing offline image reconstruction by minimizing a sparsity promoting regularized objective function as most of existing works [Lustig 2007, Guerquin-Kern 2011, Chaâri 2011] do, we introduce an online image reconstruction approach so that acquisition and reconstruction processes become interleaved. This method is well-suited for high-resolution MR acquisition. Indeed, the short lifespan of the MR signal makes the spatial encoding in large k-spaces (*i.e.* high-resolution setting) necessarily segmented. The ultimate goal is actually to reduce the cumulative time of acquisition and reconstruction. To this end, we adopt a mini-batch formalism, that consist of stacking together collected k-space samples to form a so-called mini-batch. Once a mini-batch is available, the reconstruction algorithm is run from an incomplete k-space data set for a few iterations. Hereafter, the current image solution is then used as a warm restart for the next mini-batch processing. This new framework allows us to reach a decent image quality by the end of acquisition and can be implemented throughout the Gadgetron [Hansen 2013] project. It thus enables us to display the resulting images directly to the MR system console, making online CS implementations available beyond any vendor solution.

The rest of the Chapter is organized as follows. We first recall the link between previous work related to online image reconstruction in the context of dynamic MRI. Then we derive the problem for anatomical imaging in a simplified setup, when then data is collected over a single channel receiver coil. The proposed method is evaluated on a retrospective Cartesian acquisition as well as on prospective non-Cartesian k-space data, we scanned on a Magnetom 7 T scanner (Siemens Healthineers , Erlangen, Germany) an ex-vivo baboon brain using a 2D high-resolution (400μm and 3mm slice thickness) T_2^*-weighted imaging with single channel birdcage coil. Although of pedagogical interest, the practical application of this preliminary study remains limited for high-resolution imaging as we need multiple receivers coil to boost the input SNR in this context. For that reason, the online formalism is then extended to deal with multiple-channel or multi-receiver coil. To this end, we actually rely on the *calibrationless* image reconstruction approach presented in Chapter 4. The proposed online reconstruction method is evaluated on prospective non-Cartesian k-space data collected with acceleration factor up to 15 in time, where an ex-vivo human brain was scanned using Sparkling trajectories at 7T for 2D high-resolution (400μm in plane and 1.5mm slice thickness) T_2^*-weighted imaging using a multi-receiver phased array.

Related works in dynamic MRI

Nowadays, real time MR image reconstruction is mainly implemented for dynamic imaging and more precisely for cardiac MRI. Indeed, the idea of real-time MR image reconstruc-

tion is not new and has raised a huge interest in the dynamic MRI community. Especially [Majumdar 2012c] has proposed a real-time dynamic reconstruction based on the reconstruction of the difference between two consecutive frames. Yet the authors mentioned a reduction of the accuracy compared to the offline methods. Over the last years, dynamic MR image reconstruction has been solved by applying a sparse prior to the images and a low-rank one on the temporal variations from one scan to the next [Otazo 2015]. Using the same ideas [Dietz 2017] has proposed an online approach combining CS and Principal Component Analysis, in particular they have demonstrated the benefit of this method for tracking lung cancer.

However this work does not rely on the same assumption as for dynamic MR reconstruction. In the case of dynamic MRI, there is a trade-off between spatial and temporal resolution, whereas in high-resolution anatomical imaging the only one that matters is spatial resolution. Moreover due to the resolution and in order to improve the signal to noise ratio acquisition is necessary segmented which makes the acquisition even lengthier. The purpose of this chapter is to take advantage of this time to start the reconstruction and to deliver a decent image as fast as possible while maintaining the same image quality compared to the offline reconstruction.

5.2 Extension to anatomical imaging: general online problem statement

To take advantage of the sequential aspect of segmented acquisition, the reconstruction process starts during acquisitions' dead times associated with the time intervals separating consecutive shots (*i.e.* TR when the magnetization grows back). The acquired shots are stacked to form a mini-batch and the reconstruction starts from this incomplete data set. Hence, the two processes get interleaved and partial feedback may be delivered to the MR technician along the exam with the aim to provide a nearly optimal image by the end of acquisition[1] . Two acquisition setups will be studied hereafter. First, we will consider the case of a single-channel receiver coil. Though simpler, this case is rarely met in the high resolution context since multi-channel receiver coils are traditionally used to boost the input SNR. The second setup corresponds to online reconstruction from data collected over a multi-channel receiver coil. While state-of-the art methods often rely on the coil sensitivity information that requires an estimation step for each scan as described in Chapter 3, we will rely on a calibration-less method proposed in Chapter 4 that tackles the image reconstruction problem with the great advantage of not requiring the prior knowledge of sensitivity matrices.

[1] In the sense that the recovered image at this stage is close to the final one obtained from the whole data set after complete convergence of the reconstruction algorithm.

5.2.1 Acquisition considerations

Let us first recall some acquisition considerations presented in Chapter 1. In segmented acquisition, multiple radio-frequency (RF) pulses are used to tip the global magnetization of a given volume. More precisely, at each repetition time (TR), a new RF pulse is delivered and a "new shot is collected" [Bernstein 2004, Chap.11 2.2].

In the following, S represents the number of shots used to fill the k-space hence, the total scan time is equal to $S\times$TR. For each $i \in \{1, \dots, S\}$, we refer to the k-space support of the i^{th}-shot as Γ_i and $\boldsymbol{y}_{\Gamma_i} \in \mathbb{C}^C$ gathers C samples measured over this shot. Next, let $k \in \{1, \dots, S\}$, we define Ω_k as the concatenation of the k first collected shots. Hence $\Omega_k = \cup_{i=1}^{k} \Gamma_i$, and the k-space measurements associated with Ω_k read $\boldsymbol{y}_{\Omega_k} \in \mathbb{C}^{kC}$. Using the aforementioned notation, we recall the definition of the acceleration and undersampling factors (AF and UF, respectively) with respect to the Cartesian reference as respectively the ratios n/S and $N/(SC)$. Note that in the CS context, we get both $S < n$ and $SC < N$. Although AF and UF usually evolve similarly (*i.e.* AF = UF), the Sparkling sampling scheme [Lazarus 2019a] introduced in Section 2.2.3 breaks down this relation as AF > UF.

5.2.2 Variational formulation of image reconstruction

In standard offline approaches, image reconstruction from the k-space data is performed by minimizing an objective function that sums a data consistency term f_{Ω_S} (depending on the trajectory support Ω_S) and a regularization term g, which usually promotes sparsity in a given multiscale decomposition $\boldsymbol{\Psi} \in \mathbb{C}^{N_{\boldsymbol{\Psi}} \times N}$ such as a wavelet transform [Pustelnik 1999]. The general offline image reconstruction problem aims at finding:

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x} \in \mathbb{C}^N}{\arg\min} f_{\Omega_S}(\boldsymbol{x}) + g(\boldsymbol{\Psi}\boldsymbol{x}), \tag{5.1}$$

where the following standard assumptions are made: (i) $g \in \Gamma_0(\mathbb{C}^{N_{\boldsymbol{\Psi}}})$ with a closed form proximity operator, and (ii) f_{Ω_S} is convex, differentiable on $\mathbb{C}^N$ and its gradient ∇f_{Ω_S} is β_S-Lipschitz *i.e.*:

$$(\forall (\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{C}^N) \quad \|\nabla f_{\Omega_S}(\boldsymbol{x}) - \nabla f_{\Omega_S}(\boldsymbol{x}')\| \leq \beta_S \|\boldsymbol{x} - \boldsymbol{x}'\|. \tag{5.2}$$

Problem (5.1) can be efficiently solved using iterative solvers as presented in Section 2.3.2 Note that for over-complete dictionaries [Elad 2007] the use of primal-dual approaches is more efficient in terms of computation time since these methods do not require the use of an inner iterative solver [Combettes 2011] to compute the proximity operator of $g \circ \boldsymbol{\Psi}$.

For online reconstruction, the data consistency term is progressively filled during the scan and reconstruction starts with incomplete data. Concretely, incomplete versions of

Problem (5.1) are solved in order to compute:

$$\widehat{\boldsymbol{x}}_k \in \underset{\boldsymbol{x}\in\mathbb{C}^N}{\arg\min} f_{\Omega_k}(\boldsymbol{x}) + g(\boldsymbol{\Psi}\boldsymbol{x}), \tag{5.3}$$

with k and $\widehat{\boldsymbol{x}}_k$ being respectively the number of available shots so far and $\widehat{\boldsymbol{x}}_k$ the resulting solution. For the processing of the next shot (*i.e.* $k+1$), $\widehat{\boldsymbol{x}}_k$ will be used as initialization. In an ideal case, the shots are processed one by one (*i.e.*, k is incremented by one at each step), however, such strategy is only feasible in practice if TR is longer than the time T_{it} needed to solve SubProblem (5.3), which is rarely met in the non-Cartesian multi-channel acquisition. Therefore to adapt online reconstruction to this specific case, we adopt a mini-batch formulation presented in Algorithm 6 where b_s consecutive spokes are stacked together to form a mini-batch. Then, a warm restart strategy is implemented once an complete mini-batch is available, *i.e.* SubProblem (5.3) is not solved for any integer k but only for multiples of b_s. In this setting, to comply with the online timing requirements, the maximum number of *allowed* iterations n_b per subproblem in Algorithm 6 is given by:

$$n_b \times T_{\text{it}} \approx b_s \times \text{TR}. \tag{5.4}$$

In Algorithm 6, the result of n_b iterations of the chosen proximal algorithm for solving SubProblem (5.3), initialized with $\boldsymbol{z}_0$, is denoted by $\mathcal{A}_{k,n_b}(\boldsymbol{z}_0)$. The variable $\boldsymbol{z}_k$ usually includes an approximation $\boldsymbol{x}_k$ to $\widehat{\boldsymbol{x}}_k$, but it may also embody additional information *e.g.* a dual variable delivered by the optimization algorithm. Note that when the last mini-batch is acquired ($k = S$), n_b must be relaxed and set large enough to ensure convergence. The basic online formulation is recovered for a batch size $b_s = 1$.

Algorithm 6: Online mini-batch reconstruction algorithm for solving Problem 5.3.

```
initialize k = b_s, z_0;
while k ≤ S do
    z_k = A_{k,n_b}(z_{k−b_s});
    k ← k + b_s;
end
```

It is worth noting that, since SubProblem (5.3) is convex, any good optimization algorithm is guaranteed to converge to a global solution whatever its initialization. However, the time to converge to the solution may be very sensitive to the initial guess. Therefore the warm restart procedure is beneficial to reduce the overall reconstruction time. To solve Subproblem (5.3), we propose to make use of the primal-dual Condat-Vù [Condat 2013, Vũ 2013, Combettes 2016] approach summarized in Algorithm 7. This choice was originally motivated by the usage in Eq. (5.3) of an *analysis-based* regularization term, which allows us to consider overcomplete dictionaries for $\boldsymbol{\Psi}$ such as undecimated wavelet transforms [Cherkaoui 2018]. If a synthesis-based prior had been chosen instead, one would have probably considered a fast proximal gradient method such as the POGM [Taylor 2017a] or greedy FISTA algorithms [Liang 2019] as improved convergence rates can be reached by

those algorithms. Note that β_k denotes here the Lipschitz constant of ∇f_{Ω_k}. According to [Condat 2013, Theorem 3.1], Algorithm 7 converges to a solution of SubProblem (5.3).

Algorithm 7: Condat-Vù algorithm $\mathcal{A}_{k,n_b}(\boldsymbol{x}_{k,0}, \boldsymbol{v}_{k,0})$ for solving SubProblem (5.3).

initialize $k \leq S$, $(\boldsymbol{x}_{k,0}, \boldsymbol{v}_{k,0})$;
$\kappa_k = \frac{\beta_k}{2|||\boldsymbol{\Psi}|||^2}$;
$\tau_k = \frac{1}{\beta_k}$;
for $t = 1, 2, \ldots, n_b$ **do**
  $\boldsymbol{x}_{k,t} = \boldsymbol{x}_{k,t-1} - \tau_k \left(\nabla f_{\Omega_k}(\boldsymbol{x}_{k,t-1}) + \boldsymbol{\Psi}^* \boldsymbol{v}_{k,t-1}\right)$;
  $\boldsymbol{w}_{k,t} = \boldsymbol{v}_{k,t-1} + \kappa_k \boldsymbol{\Psi} \left(2\boldsymbol{x}_{k,t} - \boldsymbol{x}_{k,t-1}\right)$;
  $\boldsymbol{v}_{k,t} = \boldsymbol{w}_{k,t} - \kappa_k \, \mathrm{prox}_{g/\kappa_k} \left(\frac{\boldsymbol{w}_{k,t}}{\kappa_k}\right)$;
end
return $(\boldsymbol{x}_{k,n_b}, \boldsymbol{v}_{k,n_b})$;

In Sections 5.3.3 and 5.4.3 we will discuss the choice of the mini-batch size b_s and the setting of hyper-parameters for respectively single and multi-channel receiver coils. Hereafter, we first derive the reconstruction problem for the single channel coil acquisition before analyzing the more challenging one, namely the multi-channel acquisition setup.

5.3 Single-channel receiver coil

5.3.1 Problem statement

In the single-channel context, Algorithm 7 is applied to the following mini-batch formulation of SubProblems (5.3):

$$(\forall k \in \{1, \ldots, S\}), \quad \hat{\boldsymbol{x}}_k = \underset{\boldsymbol{x} \in \mathbb{C}^N}{\arg\min} \frac{S}{2k} \|\mathcal{F}_{\Omega_k} \boldsymbol{x} - \boldsymbol{y}_{\Omega_k}\|_{\mathrm{F}}^2 + \lambda \|\boldsymbol{\Psi} \boldsymbol{x}\|_1, \tag{5.5}$$

where λ is the positive hyper-parameter that controls the sparsity level and can be efficiently set in an offline manner. In the case of Cartesian acquisition, the following relation describes the operator $\mathcal{F}_{\Omega_k} = \Omega_k \mathcal{F}$ where Ω_k is the under-sampling binary mask and $\mathcal{F}$ the fast Fourier transform (FFT). In the case of non-Cartesian acquisition, $\mathcal{F}_{\Omega_k}$ refers to non-equispaced or non-uniform FFT [Keiner 2009, Fessler 2003]. While the Lipschitz constant is $\beta_k = S/k$ for Cartesian acquisition, its value has to be estimated for each value of k when dealing with the non-Cartesian case. This can be performed offline – *i.e.* prior to the acquisition – using the power iterative method [Press 2007]. Importantly, all these computations can be done offline and prior to running the acquisition on the scanner.

Note that when $\boldsymbol{\Psi}$ defines an orthonormal basis, we get $|||\boldsymbol{\Psi}||| = 1$. Nevertheless, SubProblem (5.5) makes use of an analysis-based prior which enables the use of over-complete dictionaries and guarantees enhanced image quality at reconstruction at the cost of longer reconstruction times [Florescu 2014, Cherkaoui 2018].

5.3.2 Parameters setting

Acquisition parameters

An ex-vivo baboon brain was scanned on a 7 Tesla MR system (Siemens Healthineers, Erlangen, Germany) using a birdcage 1Tx/1Rx coil. All animal studies were conducted in accordance with the European convention for animal care and the NIHs Guide for the Care and Use of Laboratory Animals. The acquisition parameters were set as follows: TR=550 ms (for collecting 11 slices), TE=30 ms and flip angle FA=25° with an in-plane resolution of 400 μm for a field of view (FOV) of 20.4 cm which leads to a matrix size of $N = 512 \times 512$, to maintain a high SNR, we considered a slice thickness of 3 mm and 20 excitation have been averaged. A fully sampled Cartesian reference scan composed of 11 slices, each slice consisting of 512 lines with 512 samples each, was collected in an acquisition time of 281.6 s[2]. In what follows, we will restrict our numerical experiments to the central slice, *i.e.* to 2D imaging. The image was reconstructed using an FFT and used as reference for computing image quality scores (*e.g.*. SSIM metric). Fig. 5.3-1 shows (a) the Cartesian reference, (b) the retrospective Cartesian variable density sampling mask and (c) the prospective 15-fold accelerated non-Cartesian Sparkling [Lazarus 2019a] pattern. While the prospective reconstruction is more challenging due to potential B_0 inhomogeneities and real artifacts, the retrospective reconstruction is not affected, hence the retrospective validation is addressed first and the prospective one, which is more challenging, is presented afterwards.

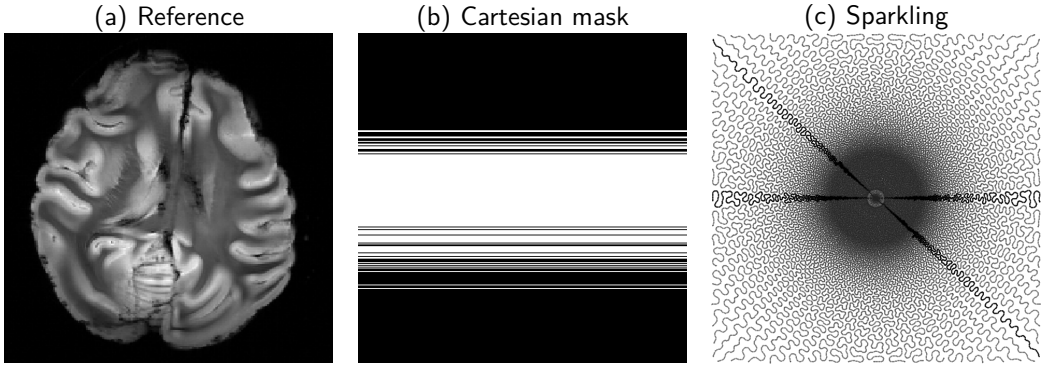


Figure 5.3-1: Single-channel coil acquisition: (a) ex vivo baboon brain Cartesian reference image, (b) retrospective Cartesian under-sampling scheme (AF = UF = 2.9), (c) prospective non-Cartesian Sparkling sampling scheme (implemented with a golden-angle ordering, two consecutive shots are highlighted in black) (AF = 15, UF = 2.5), the first collected being the horizontal one and the rotation being counterclockwise.

The Cartesian mask (Fig. 5.3-1(b)) was composed of $S = 176$ lines of $C = 512$ samples each, leading to UF = AF = 2.9 and an acquisition time of 97 s for the 11 slices. The segmented acquisition was ordered in time by considering that central lines of k-space were collected first and then that others were acquired in a pseudo-random order. The

[2]Strictly speaking, in 2D imaging we cannot divide the global acquisition time of multiple slices by the number of slices to get that of a single slice because the usage of a large TR enable to concatenate multiple slices and shortening the TR would change the imaging contrast.

Sparkling sampling pattern (Fig. 5.3-1(c)) was generated with $S = 34$ shots from a radial initialization, each shot being composed of $C = 3072$ samples leading to UF $= 2.5$ and AF $= 15$. The total scan time for the 11 slices using Sparkling was reduced to 18.7 s. The acquisition was segmented in time according to the golden-angle ordering scheme, as it was proven [Winkelmann 2006] to increase the maximize the k-space coverage during the scan. Hence, between two consecutive shots a rotation of 137° was applied as highlighted in Figure 5.3-1(c).

Reconstruction parameters

The decimated Symmlet 8 Wavelet Transform with 4 decomposition scales was used as sparsifying transform. Regarding hyper-parameter λ in Eq. (5.5) it has been set retrospectively so as to maximize the structural similarity score (SSIM) [Wang 2004] to the reference image once all data are available (*i.e.* offline scenario). Note that in a more clinically plausible setting, the value of λ needs either to be estimated using the noise statistical properties [Stein 1981] based on a pre-calibration scan or set using cross-validation across slices in 2D imaging [Hastie 2005]. For non-Cartesian acquisition, the GPU [Lin 2018] implementation of the NUFFT [Fessler 2003] was used. In Section 5.3.3, we compare the performances of several batch sizes with a given number of iterations defined by Eq. (5.4). Once all shots are considered in Algorithm 6, *i.e.* $k = S$, then the number of iterations is set to $n_b = 200$ to guarantee that convergence was reached.

Computing parameters setting

All experiments were run on a machine with 128 GB of RAM and an 8-core (2.40 GHz) Intel Xeon E5-2630 v3 Processor. All the codes have been developed in Python using the PySAP package[3]. The values of computing times have been obtained using 5 epochs of 10 Condat-Vú iterations each, the mean and standard deviation being summarized in Table 5.1. Using multiple epochs permits to account for potential variability in computing times due to concomitant processes running on the machine. Although the NUFFT is usually slower than the FFT, here we observed the converse as we used the GPU implementation of the NUFFT and the CPU implementation of the FFT. A more fair comparison might be achieved using the GPU version of FFT but this was not mandatory for the present online study as the FFT was not the computational bottleneck.

Table 5.1: Computing time for one iteration estimated on 5 epochs of 10 iterations each using `timeit`..

Cartesian single-channel acquisition	Non-Cartesian single-channel acquisition
93.9 ms ± 14.4 ms	78.2 ms ± 8.9 ms

[3] https://github.com/CEA-COSMIC/pysap

5.3.3 Results

Retrospective Cartesian sampling

First we implemented the online reconstruction pipeline on single-channel acquisition using the retrospectively under-sampled Cartesian k-space mask shown in Fig. 5.3-1(b). Mini-batches of increasing size ($b_s \in \{1,\ 4,\ 16,\ 22,\ 44,\ 88\}$) were tested against the offline reconstruction scenario ($b_s = 176$). In Fig. 5.3-2(a)-(b), we show the evolution over time of the global cost function and the SSIM score. The time origin corresponds to the beginning of the scan. All settings eventually converge to the same value both in terms of cost function and SSIM score. This confirms that the final image is the same. Moreover Fig. 5.3-3(top-row) depicts partial reconstructions for the tested mini-batch sizes by the end of acquisition (*i.e.* before taking into account the last mini-batch). While large mini-batches show aliasing artifacts (see Fig. 5.3-3 for $b_s = 44$, and $b_s = 88$), the small batches deliver pretty accurate images, which tends to demonstrate the benefits of using small batch sizes for online reconstruction purposes. The reason for which we observed aliasing artifacts in large batch sizes is the varying amount of available k-space data in the top row of Fig. 5.3-3. Indeed, by the end of the acquisition (*i.e.* before the process of the last mini-batch *i.e.* at $T = \mathrm{TR} - \varepsilon$) only a single shot is missing for $b_s = 1$ while for larger b_s (such as for $b_s = 88$) a large number of shots are missing in the data-fidelity term (when $b_s = 88$ half of the spoke are not processed).

(a) Evolution of the cost function (5.5) over time (b) Evolution of the SSIM score over time

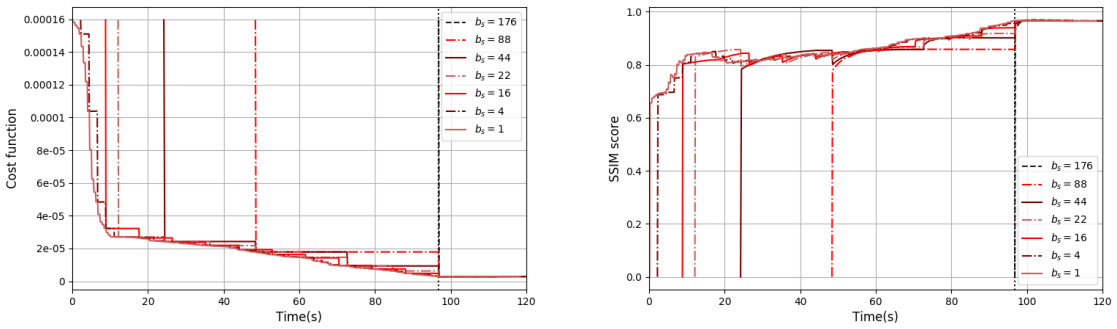


Figure 5.3-2: Single-channel retrospective Cartesian reconstruction: (a) evolution over time of cost function in Eq. (5.5) and (b) of the SSIM score for different batch sizes b_s. The dark dashed line marks the end of acquisition.

Prospective Non-Cartesian acquisition

The non-Cartesian case was also tested using 15-fold accelerated Sparkling trajectories as shown in Fig. 5.3-1(c). As this sampling pattern comprises $S = 34$ shots, the only possible tunings of the mini-batch size b_s are its factors $1, 2, 17$. In that context, the GPU implementation of the NUFFT was really helpful to comply with online reconstruction constraints and maintain a short time per iteration. The GPU-NUFFT actually enable to iterate over a sufficiently large number of iterations in each subproblem and to explore

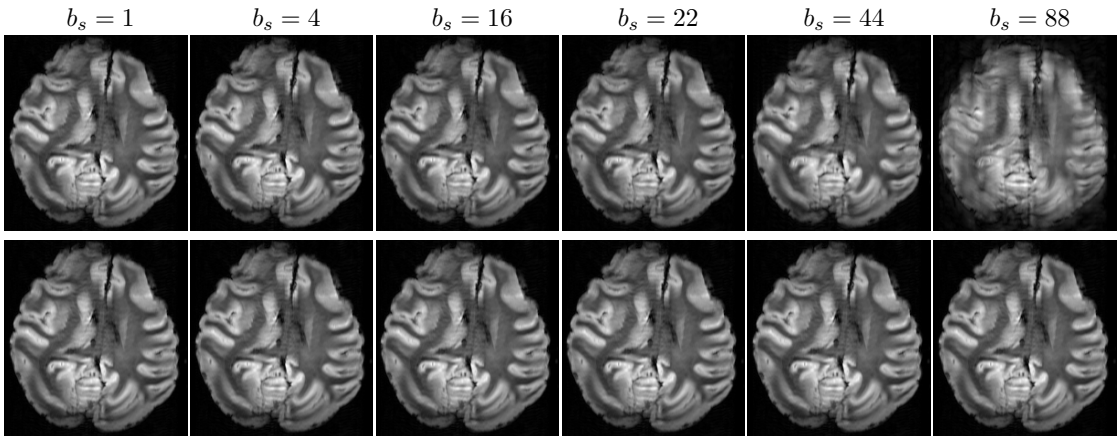


Figure 5.3-3: Online reconstruction of 15-fold ($S = 34$ shots) retrospectively accelerated Cartesian scan of ex vivo baboon brain in a single-channel coil acquisition setup. Images reconstructed for increasing values of mini-batch size b_s at the end of acquisition (**top**) and at convergence (**bottom**).

small batch sizes too. We respectively used $n_b = 8,\ 16,\ 24$ for increasing batch sizes $b_s = 1,\ 2,\ 17$. Fig. 5.3-4 demonstrates even more clearly than in Cartesian sampling that the evolution over time of both the global cost function and the SSIM score benefit from small batch sizes to reach almost convergence by the end of acquisition (depicted in dotted black line). The fastest converging online scenario corresponds to $b_s = 1$ (black dashed trace in Fig. 5.3-4) as $n_b = 8$ are sufficient to significantly decrease the cost function. This is confirmed by the partial solutions obtained by the end of acquisition (see Fig. 5.3-5). The reconstructed images for $b_s = 1$ or $b_s = 2$ are already very close to the Cartesian reference as compared to the one obtained for the larger batch size $b_s = 17$. Once again, the aliasing artifacts shown in Fig. 5.3-5(top-row) for $b_s = 17$ are due to a larger part of missing data (half in this case).

These results confirm the feasibility of online MR image reconstruction from non-Cartesian k-space data in a single-channel coil acquisition scenario.

(a) Evolution of the cost function (5.5) over time (b) Evolution of the SSIM score over time

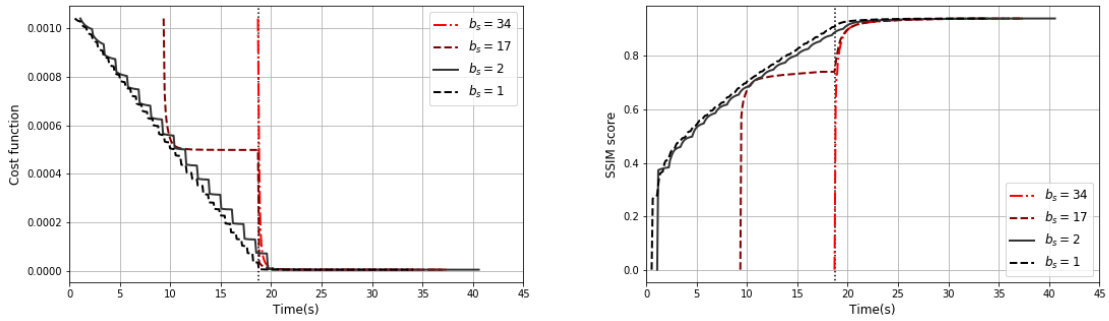


Figure 5.3-4: Online reconstruction of 15-fold ($S = 34$ shots) prospectively accelerated Spark-ling scan of ex vivo baboon brain in a single-channel coil acquisition setup. Evolution over time of (a) the cost function in Eq. (5.5), and of the SSIM score in (b) for batch sizes b_s corresponding to the primary factors of S. The dark dashed line marks the end of acquisition.

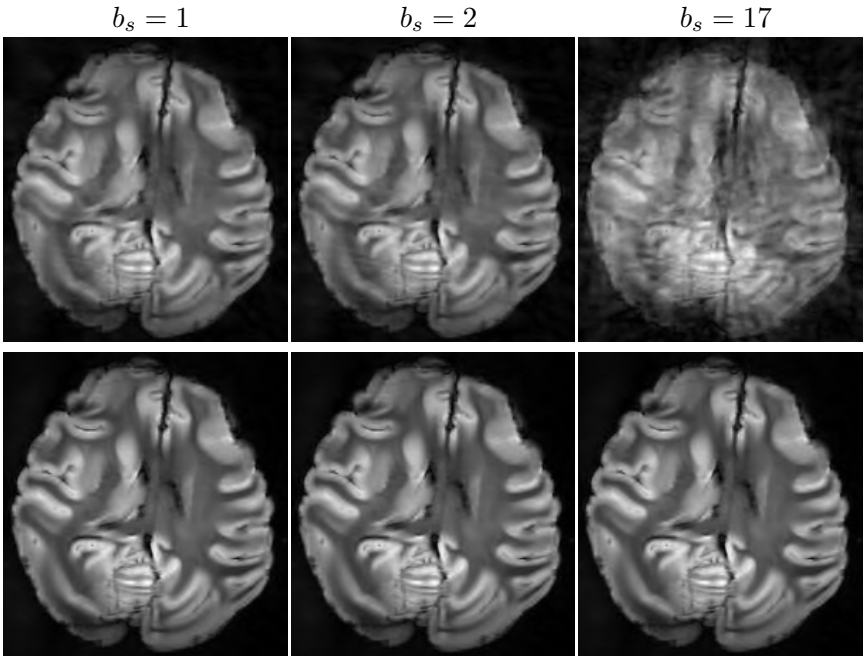


Figure 5.3-5: Online reconstruction of 15-fold ($S = 34$ shots) prospectively accelerated Spark-ling scan of ex vivo baboon brain in a single-channel coil acquisition setup. Images recon-structed for increasing values of mini-batch size b_s by the end of acquisition (**top**) and at convergence (**bottom**).

5.4 Multi-channel receiver coil

5.4.1 Problem statement

As shown Section 1.3.1 the use of the multi-channel array coil allows an increase of the input SNR. In what follows, we denote by L the number of channels in this kind of coils. In this context, the k-space measurements associated with Ω_k collected by the ℓ^{th} channel are defined as $\boldsymbol{y}_{\Omega_k,\ell} \in \mathbb{C}^{kC}$. All collected k-space data are then stacked in

$\boldsymbol{Y}_{\Omega_k} = [\boldsymbol{y}_{\Omega_k,1}, \ldots, \boldsymbol{y}_{\Omega_k,L}] \in \mathbb{C}^{kC \times L}$.

As described in Chapter 3, the multi-channel reconstruction methods can be split in two classes. The first one models the coil sensitivity maps $(\boldsymbol{S}_\ell)_{1 \leq \ell \leq L}$ where $\boldsymbol{S}_\ell \in \mathbb{C}^{N \times N}$ represents the ℓ^{th} channel sensitivity profile. In these techniques, a single image is recovered from the combination of all channels [Chaâri 2011, Florescu 2014]. The reconstruction problem is often formulated as follows:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathbb{C}^N}{\arg\min} \frac{1}{2} \sum_{\ell=1}^{L} \|\mathcal{F}_{\Omega_S} \boldsymbol{S}_\ell \boldsymbol{x} - \boldsymbol{y}_{\Omega_S,\ell}\|_2^2 + \lambda \|\boldsymbol{\Psi} \boldsymbol{x}\|_1, \quad \lambda > 0. \tag{5.6}$$

However, the coil sensitivities depend on the scanned subject, therefore they have to be calibrated for each subject. The calibration step can be performed beforehand [Uecker 2014, El Gueddari 2018b] and then some estimates $(\widehat{\boldsymbol{S}}_\ell)_{1 \leq \ell \leq L}$ can be injected in Eq. (5.6). Alternatively, the reconstruction can be viewed as a blind bilinear inverse problem [She 2014] where the optimization alternates between sensitivity profile estimation and image recovery steps. While the second approach is more computationally demanding, the first one does not really fit the constraints of online reconstruction. Indeed, as any shot is able to entirely cover the center of k-space, low frequency information associated with spatially smooth sensitivity profiles cannot be extracted in a straight manner. Moreover, a second weakness of this formalism lies in the fact that the gradient Lipschitz constant β_k depends on the coil sensitivities and cannot be computed in advance if we assume that no calibration scan has been run beforehand.

The second set of approaches falls in the class of *calibrationless* methods, which by definition do not require any prior knowledge on the sensitivity profiles. As a consequence, these methods try to reconstruct an image per channel with some regularizing constraints across channels to impose consistency. Either the regularization is performed in the k-space domain where a low-rank penalty is applied to a Hankel matrix [Haldar 2014, Shin 2014, Lee 2016], or it is imposed in a sparse transformed domain for instance in order to promote group sparsity [Majumdar 2012a, Trzasko 2011, El Gueddari 2019b]. While k-space-based methods demonstrate good image recovery for Cartesian sampling [Haldar 2014] , their application to non-Cartesian trajectories implies a gridding step such as GROG [Seiberlich 2007] that strongly degrades the final image quality. For that reason, domain-based *calibrationless* reconstruction seems better suited to online processing of multi-channel non-Cartesian k-space data.

For the aforementioned reasons, we will use a domain-based *calibrationless* formulation that reads as follows:

$$\widehat{\boldsymbol{X}} = \underset{\boldsymbol{X} \in \mathbb{C}^{N \times L}}{\arg\min} \frac{1}{2} \|\mathcal{F}_{\Omega_S} \boldsymbol{X} - \boldsymbol{Y}_{\Omega_S}\|_2^2 + \mathcal{R}(\boldsymbol{\Psi} \boldsymbol{X}). \tag{5.7}$$

In Eq. (5.7), $\mathcal{R}$ refers to a structured sparsity promoting term. As shown in Section 4.4, this formulation is able to recover images with a quality competitive with state-of-the art methods [She 2014], especially for non-Cartesian sampling schemes. Subsequently, the Octagonal Shrinkage and Clustering Algorithm for Regression (OSCAR) [El Gueddari 2019b] based calibrationless reconstruction method is implemented, where the wavelet transform

$\boldsymbol{\Psi}$ decomposes the stack of images into a stack of coefficients $\boldsymbol{C} = \boldsymbol{\Psi}\boldsymbol{X} = \Big[\boldsymbol{c}_1, \dots, \boldsymbol{c}_L\Big] = \Big[\boldsymbol{\Psi}\boldsymbol{x}_1, \dots, \boldsymbol{\Psi}\boldsymbol{x}_L\Big] \in \mathbb{C}^{N_\Psi \times L}$. Each vector of wavelet coefficients $\boldsymbol{c}_\ell$ with $\ell \in \{1, \dots, L\}$ is made up of B bands. The $b \in \{1, \dots, B\}$ band contains P_b coefficients $(c_{\ell,b,j})_{1 \leq j \leq P_b}$. The OSCAR penalty is expressed as follows:

$$\mathcal{R}_{\text{b-OSCAR}}(\boldsymbol{C}) = \sum_{b=1}^{B} \sum_{j=1}^{P_b} \Big(\sum_{\ell=1}^{L} \lambda |c_{\ell,b,j}| + \gamma \sum_{\ell' < \ell} \max\{|c_{\ell,b,j}|, |c_{\ell',b,j}|\Big), \quad \lambda > 0,\ \gamma > 0. \tag{5.8}$$

Here, we can introduce the straightforward extension of our online formulation to the matrix case. The online *calibrationless* reconstruction Subproblem (5.3) thus reads:

$$(\forall k \in \{1, \dots, S\}), \quad \hat{\boldsymbol{X}}_k = \underset{\boldsymbol{X} \in \mathbb{C}^{N \times L}}{\arg\min} \frac{S}{2k} \|\mathcal{F}_{\Omega_k} \boldsymbol{X} - \boldsymbol{Y}_{\Omega_k}\|_2^2 + \mathcal{R}_{\text{b-OSCAR}}(\boldsymbol{\Psi}\boldsymbol{X}). \tag{5.9}$$

On the one hand, the gradient Lipschitz constant β_k only depends on the sampling scheme Ω_k, hence it can be computed offline and loaded at the beginning of each scan. On the other hand, the proximity operator of $\mathcal{R}_{\text{b-OSCAR}}$ is explicit and can be computed efficiently [Zeng 2014a, Eq. 24].

The performance of this new multi-channel image reconstruction formulation and its ability to comply with online processing constraints will be discussed hereafter.

5.4.2 Parameters setting

Acquisition parameters

An ex vivo human brain was scanned using a 1Tx/32Rx (Nova Medical Inc., Wilmington, MA, USA), *i.e.* $L = 32$ with $4000\mu\text{m}^2$ in-plane resolution and a slice thickness of 1.5 mm. The donor gave his written consent before death to the donation program of University of Tours, France. The reference slice is displayed in Fig. 5.4-6(a). Prospectively accelerated Sparkling acquisitions were considered with AF = 8 and AF = 15 corresponding to $S = 64$ and $S = 34$ shots, respectively and to acquisition times of 35.2 s and 18.7 s for the 11 slices as compared to 281.6 s for the Cartesian reference. Both are shown in Fig. 5.4-6(b)-(c). The 15-fold accelerated sampling pattern was the same as the one used in single-channel acquisition. We also investigated less accelerated acquisition (AF=8) using radially-initialized Sparkling trajectories to see to what extent this may impact online reconstruction. A larger number of shots actually offers more degrees of freedom to define the mini-batch size b_s in online reconstruction.

Reconstruction parameters.

The decimated Symmlet 8 Wavelet Transform with 4 decomposition scales was used as sparsifying transform. As mentioned previously, the hyper-parameters (λ, γ) in Eq. (5.9) have been set retrospectively so as to maximize the SSIM score [Wang 2004] to the reference image once all data are available (*i.e.* offline scenario). Note that in a more clinically

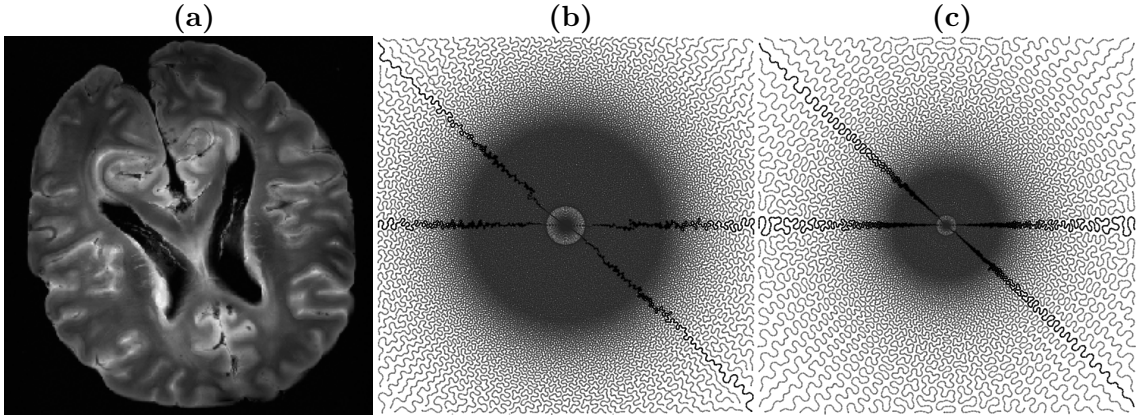


Figure 5.4-6: Multi-channel coil acquisition: **(a)** ex vivo human brain Cartesian reference image, **(b)** prospective Sparkling sampling pattern with $S = 64$ spokes (*i.e.* AF $= 8$) and $C = 3072$ points (*i.e.* UF $= 1.3$), **(c)** prospective Sparkling sampling pattern with $S = 34$ spokes (*i.e.* AF $= 15$), each one being composed of $C = 3072$ points (*i.e.* UF $= 2.5$).

plausible setting, at least hyper-parameter λ needs either to be estimated using the noise statistical properties [Stein 1981] or set using cross-validation across slices in 2D imaging [Hastie 2005]. For non-Cartesian acquisition the GPU [Lin 2018] implementation of the NUFFT [Fessler 2003] was used. In Section 5.4.3, we compare the performances of several batch sizes with a given number of iterations defined by Eq. (5.4). Once all shots are considered in Algorithm 6, *i.e.* $k = S$, then the number of iterations is set to $n_b = 200$ to guarantee that convergence was reached.

Computing parameters setting

All experiments were run on a machine with 128 GB of RAM and an 8-core (2.40 GHz) Intel Xeon E5-2630 v3 Processor. All the codes have been developed in Python using the PySAP package[4]. The values of computing times have been obtained using 5 epochs of 10 Condat-Vú iterations each, the mean and standard deviation being summarized in Table 5.2. Using multiple epochs permits to account for potential variability in computing times due to concomitant processes running on the workstation.

Table 5.2: Computing time for one iteration estimated on 5 epochs of 10 iterations each using timeit for Sparkling with, first row: $S = 34$ (presented Fig. 5.4-6(b)) and second row $S = 64$ (presented Fig. 5.4-6(c)).

	Gradient step	Proximity op. step	Linear Operator		Total time per iteration
			direct	adjoint	
$S = 34$	750 ms ± 32.2 ms	847 ms ± 17.9 ms	998 ms ± 15.8 ms	667 ms ± 16.2 ms	4.29 s ± 111 ms
$S = 64$	1.11 s ± 18.5 ms				4.92 s ± 102 ms

[4]For more details on the PySAP package please refer to Section C.2.

5.4.3 Results

As multi-channel online reconstruction is more demanding from a computational viewpoint, the way constraint (5.4) can be satisfied requires a larger number of shots. Because Algorithm 6 assumes uniform batch sizes over iterations, we decided to explore a drastically less accelerated (*i.e.* 8-fold) Sparkling acquisition in which $S = 64$ are collected. This allows us to get more flexibility in the setting of the batch size b_s as compared to the 15-fold accelerated Sparkling sampling scheme where $S = 34$ can be split only in $b_s = 2$ or $b_s = 17$.

Fig. 5.4-7 shows that there is only a little advantage for small batch sizes (*i.e.* $b_s = 8$ and $b_s = 16$) in the context of multi-channel receiver coils. This results from the fact that the overall computing time per iteration and constraint (5.4) allow to perform only $n_b = 1$ ($n_b = 2$, respectively) iteration(s) when $s_b = 8$ ($s_b = 16$, respectively). The zoom in the curve of SSIM score in Fig. 5.4-7(b) confirms this tiny advantage for small batch sizes. Stronger evidence in favor of small values of b_s is shown in Fig. 5.4-8(top) where the ventricles (especially the right one) appear darker in $b_s = 8$ as compared to $b_s = 32$. Images obtained at convergence (Fig. 5.4-8(bottom)), *i.e.* once $n_b = 200$ iterations have been run from $k = S$ mini-batches, confirm the convergence to the Cartesian reference.

Overall, in this multi-channel acquisition scenario the main bottleneck remains the computation cost per iteration. With a diminished cost, we could perform a larger number of iterations in particular for small batch sizes (for instance $n_b = 4$ for $b_s = 8$) and get better MR images by the end of acquisition.

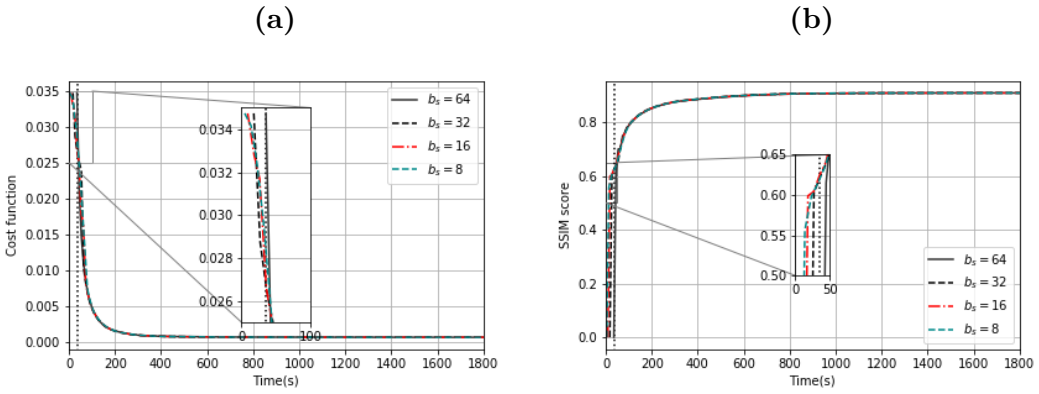


Figure 5.4-7: Online reconstruction of 8-fold ($S = 64$ shots) prospectively accelerated Sparkling scan of ex vivo human brain in a multi-channel coil acquisition setup. Evolution over time of **(a)** the cost function in Eq. (5.9) and of **(b)** the SSIM score for increasing values of batch size b_s. The dark dashed line marks the end of acquisition.

Last, we investigated a Sparkling acquisition scenario corresponding to the same sampling pattern (Fig. 5.4-6(c)) as the one studied in the single-channel setup ($S = 34$, 15-fold acceleration). In this setup, the online reconstruction approach is not really viable as shown first in Fig. 5.4-9: a single mini-batch configuration ($b_s = 17$) was explored given the timing constraint and it does not converge by the end of acquisition both in terms of cost

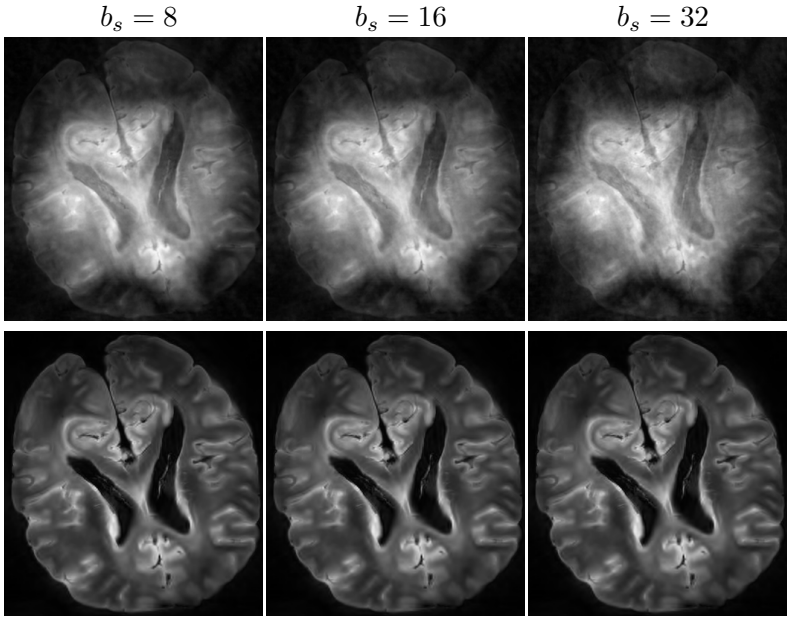


Figure 5.4-8: Online reconstruction of 8-fold ($S = 64$ shots) prospectively accelerated Sparkling scan of ex vivo human brain in a multi-channel coil acquisition setup. Images reconstructed for increasing values of mini-batch size b_s by the end of acquisition (**top**) and at convergence (**bottom**).

function and SSIM score as only $n_b = 2$ iterations were carried out. This observation is then confirmed on the reconstructed MR image reported in Fig. 5.4-10(a), which is severely corrupted by aliasing artifacts. Nevertheless, the final solution at convergence (Fig. 5.4-10(b)) exactly matches the Cartesian reference.

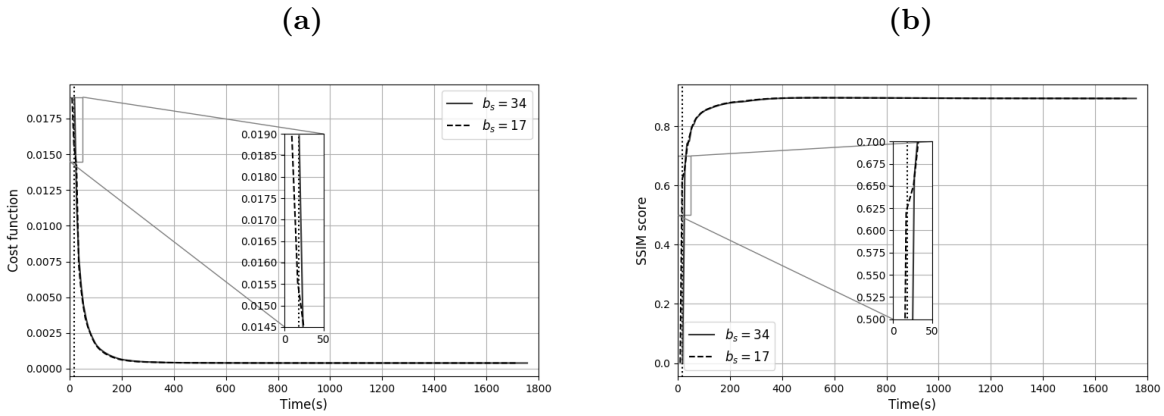


Figure 5.4-9: Online reconstruction of 15-fold ($S = 34$ shots) prospectively accelerated Sparkling scan of ex vivo human brain in a multi-channel coil acquisition setup. Evolution over time of **(a)** the cost function in Eq. (5.9) and of **(b)** the SSIM score for batch sizes $b_s = 17$ and $b_s = 34$. The dark dashed line marks the end of acquisition.

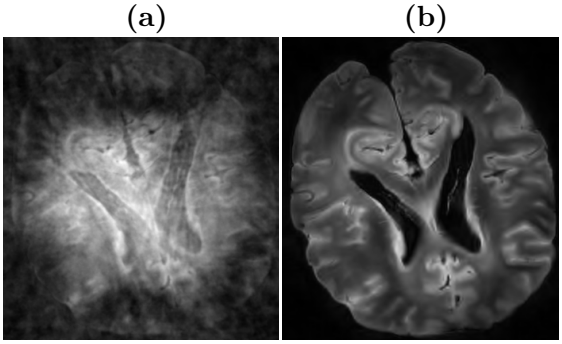


Figure 5.4-10: Online reconstruction of 15-fold ($S = 34$ shots) prospectively accelerated Sparkling scan of ex vivo human brain in a multi-channel coil acquisition setup using $b_s = 17$. **(a)** Partial solution obtained by the end of acquisition and **(b)** at convergence.

5.4.4 Online multi-channel reconstruction: improvements

Although the single-channel online reconstruction provides a reliable image by the end of the acquisition, the multi-channel results are partly disappointing for the given computational power. Instead of picking up a more powerful machine to reach good results under the online timing constraints, we investigated another reconstruction scheme. In fact when analyzing the contribution of the different terms to the global cost function for offline reconstruction (see Figure 5.4-11), we realized that the weight of the data consistency term was predominant in the global cost function over the first iterations.

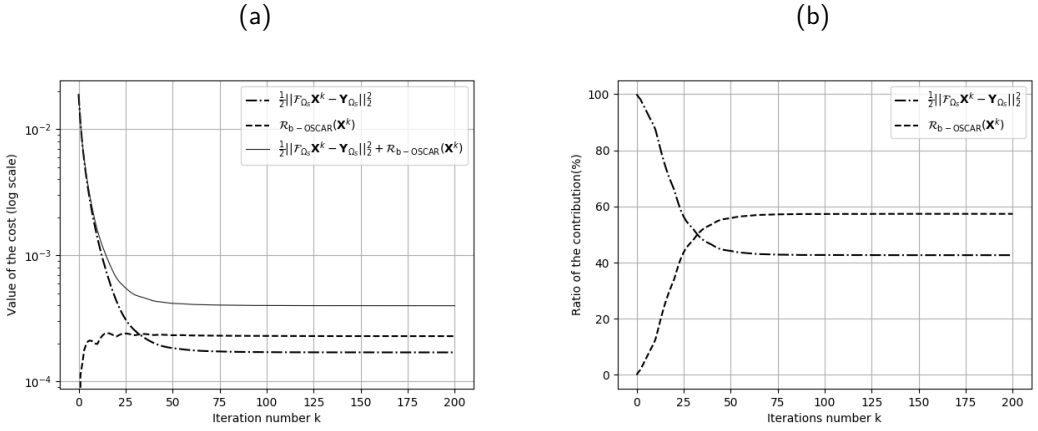


Figure 5.4-11: Split of the contributions to the global cost function **(a)** for offline reconstruction using the 15-fold accelerated Sparkling acquisition, and **(b)** evolution of the contributions ratio over iterations.

While the main bottleneck is the computation of the proximity operator and since the largest contribution corresponds to the data consistency term during the first iterations, we logically propose to solve the least squares calibrationless problem during the scan time, and then to switch to the complete regularized formulation. The reduction of the computing time per iteration has two main advantages: first, it enables to reach larger numbers of iterations and second, it enables the batch size reduction as shown in Section 5.3.3 where

small batches provide better image quality by the end of scan as the number of k-space samples not already taken into account in the reconstruction gets smaller. Consequently, during the exact acquisition time we considered solving the following problem:

$$(\forall k \in \{1, \dots, S - b_s\}), \quad \hat{\boldsymbol{X}}_k = \underset{\boldsymbol{X} \in \mathbb{C}^{N \times L}}{\arg\min} \frac{S}{2k} \|\mathcal{F}_{\Omega_k} \boldsymbol{X} - \boldsymbol{Y}_{\Omega_k}\|_2^2. \tag{5.10}$$

In this formulation, the computation time is dedicated to the evaluation of the gradient step as presented Table 5.2. This permits the batch size deflation down to 2 spokes for the acquisition schemes under consideration.

Figures 5.4-12 and 5.4-13, show the improvement of this new approach on the reconstruction of the 8-fold accelerated Sparkling trajectory (depicted in Figure 5.4-6 **(a)**). Compared to Figure 5.4-8, the partial solutions show an improvement in terms of image quality especially for smaller batch sizes ($b_s = 2,\ 4$). The same results hold of course for 15-fold accelerated scans, *i.e.* with Sparkling $S = 34$ (illustrated in Figure 5.4-6 **(b)**).

(a) **(b)**

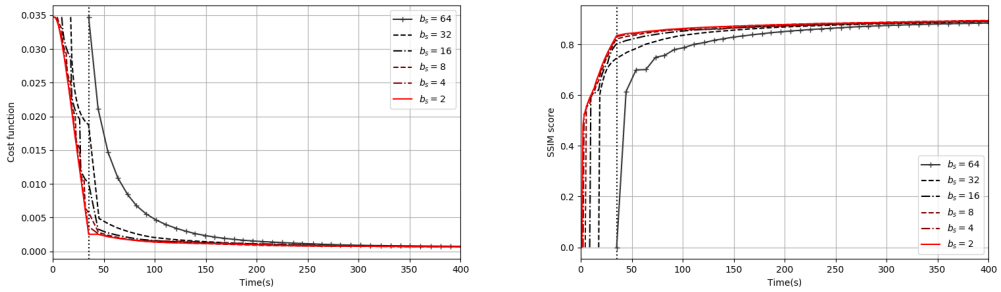


Figure 5.4-12: Fast online reconstruction of 8-fold ($S = 64$ shots) prospectively accelerated Sparkling scan of ex vivo human brain in a multi-channel coil acquisition setup. Evolution over time of **(a)** the cost function in Eq. (5.9) and of **(b)** the SSIM score for increasing values of batch size b_s. The dark dashed line marks the end of acquisition.

(a) **(b)**

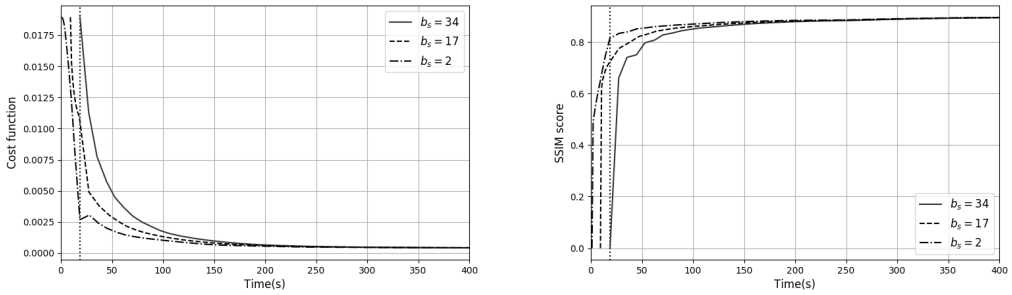


Figure 5.4-13: Fast online reconstruction of 15-fold ($S = 34$ shots) prospectively accelerated Sparkling scan of ex vivo human brain in a multi-channel coil acquisition setup. Evolution over time of **(a)** the cost function in Eq. (5.9) and of **(b)** the SSIM score for batch sizes $b_s = 2$ $b_s = 17$ and $b_s = 34$. The dark dashed line marks the end of acquisition.

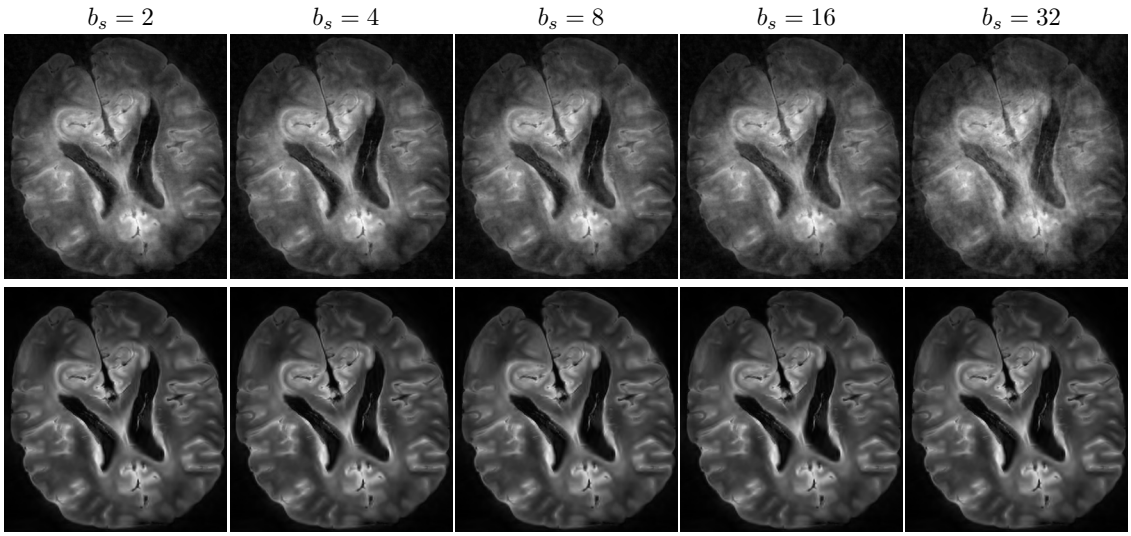


Figure 5.4-14: Fast-online reconstruction of 8-fold ($S = 64$ shots) prospectively accelerated Sparkling scan of ex vivo human brain in a multi-channel coil acquisition setup. Images reconstructed for increasing values of mini-batch size b_s by the end of acquisition (**top**) and at convergence (**bottom**). There is a clear advantage for short batch sizes..

5.5 Conclusion

Acceleration of MRI acquisition has been a major area of research for the last decade, especially in the high-resolution context where motion becomes critical. In this context, k-space data is collected in a segmented manner through multiple consecutive shots separated by times of repetition. Among the different acceleration strategies proposed in the literature (*e.g.* partial Fourier, parallel imaging, etc.), Compressed Sensing has been the most appealing as it enables larger acceleration factors without degrading the image quality at the reconstruction step. However, the price to pay lies in a long iterative process for image reconstruction.

Our online reconstruction approach relies on a mini-batch formulation which consists of aggregating the acquired shots in multiple subsets over time on mini-batches processing. We take advantage of the period of time between the mini-batches acquisitions to launch the reconstruction from incomplete data. The use of a warm-restart mechanism to set up new (primal and dual) variables in a given mini-batch reconstruction from the solution obtained at the previous iteration allowed time savings and faster convergence. Still in this context, CS reconstruction was formulated as the minimization of a regularized criterion that combines a data consistency term and a sparsity promoting penalization. Although the ℓ_1-norm term is the most straightforward in CS reconstruction, it only properly addresses the single-channel reconstruction scenario as it performances on multi-channel acquisition are limited (Section 4.5). A second contribution of this chapter was thus to extend the online framework to parallel imaging reconstruction in which k-space data are collected over multiple channels. For this purpose we made use of the *calibrationless* method [El Gueddari 2019b] developed in Chapter 4 and promoted group sparsity across

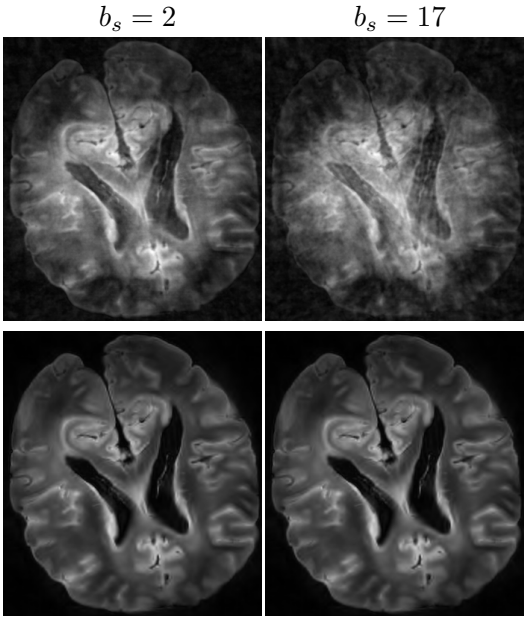


Figure 5.4-15: Fast-online reconstruction of 15-fold ($S = 34$ shots) prospectively accelerated Sparkling scan of ex vivo human brain in a multi-channel coil acquisition setup. Images reconstructed for increasing values (only 2 possible values here as 34 is only divisible by 2 and 17) of mini-batch size b_s by the end of acquisition (**top**) and at convergence (**bottom**).

channels using OSCAR-norm regularization. This *calibrationless* formulation presents the advantage of being well-suited to non-Cartesian sampling and does not require prior knowledge on the sensitivity profiles associated with the multiple channels. Importantly, the proposed online approach relies on convex optimization and its convergence to the global solution is guaranteed once all the spokes have been acquired.

The proposed framework was validated on single and multi-channel acquisitions at 7 T. We considered ex vivo imaging to avoid any motion-related concern. In the single-channel acquisition setup, a retrospective Cartesian sampling was implemented using variable density over phase encoding lines with an acceleration factor of 2.9. When the shots are split in small batches, the online reconstruction showed a significant gain in terms of speedup as the SSIM score of the image solution computed by the end of acquisition was very close to that of the final image delivered by offline reconstruction. This finding also holds for non-Cartesian acquisitions performed with a single-channel coil.

In the multi–channel acquisition setup, online reconstruction gets more challenging as both the computing time per iteration and memory usage increase. Hence, it gets harder to comply with the timing constraints using small batch sizes without significantly decreasing the number of iterations per mini-batch processing. We still observed some advantages for MR online reconstruction in 8-fold accelerated Sparkling acquisitions using a 32-receiver coil. In contrast, in highly accelerated acquisitions (*i.e.* 15-fold Sparkling) we showed that the current approach was not able to recover artifact-free images by the end of acquisition as only large batch sizes can be managed in this acquisition scenario. This

intermediate conclusion calls for exploring new directions to overcome this difficulty. We already investigated an imperfect solution that is based on solving the data-fidelity term as the non-Cartesian Fourier transform involved in the gradient step is poorly conditioned and hence requires several iterations to converge to a descent image quality. Then the regularization term is added to the reconstruction as the problem is ill-posed thus needs regularization.

5.6 Outlook

A first attempt for improving numerical efficiency of online reconstruction would consist of improving the initialization $\boldsymbol{x}_0$ using a density compensated regridded solution [Fessler 2007]. A second direction that could be relevant to explore is related to coil compression. With the current trend of enlarging the number of channels in reception coils in order to continuously improve the input SNR (*e.g.* 64 channel Nova coil now available on 3 Tesla systems), coil compression could drastically limit memory usage and computing time [Buehrer 2007]. However, the question of optimal combination of coil compression with calibrationless reconstruction is still an open issue and there is no contribution in the literature that addresses this point. A third direction would be to consider stochastic implementation [Combettes 2016], where the global gradient is approximated by the sum of a subset of partial gradients.

Overall, the proposed framework enables to reduce the reconstruction times in CS MRI applications. Importantly, for practical implementation, it is compatible with the Gadgetron system [Hansen 2013] that provides the right interface to make those reconstruction techniques available to the scanner. A description of the Gadgetron system is provided Section C.1. This may be really helpful to guide the physician for selecting further pulse MR sequences in order to sharpen his/her medical diagnosis.

Future work will be devoted to the adaptation of this framework to other imaging contrasts (*e.g.*, T_2-weighted imaging) and pulse sequences (*e.g.*, turbo spin echo) for which larger times of repetition (*e.g.* 5 s) give us the opportunity to successfully apply the proposed approach in the multi-channel acquisition setting. A pending challenge remains how we can extend this approach to 3D imaging in a time-efficient way. A nice feature we have recently implemented in PySAP to deal with 3D imaging applies to the situation when the k-space data are collected along parallel and uniformly-spaced planes, as done in z-variable density SPARKLING [Lazarus 2019b]. In this case, we can restrict the usage of the NFFT to 2D only and make use of the FFT along the third dimension. This typically allow us to decrease the memory usage by a factor of 50 and the computation time by a factor of 5.

* * *
* *
*

General Conclusions and Perspectives

At the end of these three years, we hope we gave a boost to the clinical application of Compressed Sensing for high-resolution MRI. The aim of this thesis was not merely to propose an efficient reconstruction method for highly accelerated non-Cartesian acquisition particularly suited for high-resolution imaging, but also to speed up the reconstruction in order to bring those trajectories into clinical practice.

Context

Improving the spatial resolution in MRI while limiting the acquisition time has been a crucial challenge for the MRI community over the last two decades. Therefore, a major research field has emerged with multiple contributions dedicated to scan time accelerations. In the early 2000s parallel imaging has been proposed as a key step and first results were obtained using dedicated phased array receiver coils and deterministic undersampling, while the reconstruction relying on the spatial discrimination of the coil sensitivity profiles. However the maximum accelerations factor was fairly low. Over the last 10 years, the development of compressed sensing theory and more specifically, the gain in acquisition times and the recovery guarantees on the reconstruction made a breakthrough in the MRI field. Although the practical implementation slightly departs from theory, the resulting acceleration factors exceeded the ambition, up to 20 in 2D imaging and 70 in 3D imaging for the SPARKLING trajectories proposed by [Lazarus 2019a]. Despite this acceleration, the high-resolution imaging context remains unsuited for clinical purposes, the reason for that is rather simple: the long acquisition time was transferred to an expensive reconstruction and thus the gain in acquisition provided by CS theory was not converted into a global acceleration in the whole acquisition/reconstruction pipeline. The purpose of this PhD thesis was thus to reduce this reconstruction time and therefore to bring the high-resolution MRI to the clinical practice. For this purpose, we have developed a new approach that interleaves acquisition and reconstruction called *online reconstruction*.

Contributions & Limitations

High-resolution imaging requires a certain level of input signal to noise ratio (SNR), hence the acquisition is often performed with multi-channel receiver coils, to ensure the signal quality. Before the beginning of this work none of the reconstruction algorithm in the laboratory was considered for the multi-channel case, and the SPARKLING acquisitions were hampered by the lack of SNR. Therefore the first step was to implement a simple yet efficient method to estimate the coil sensitivity profiles involved in the SENSE-like multi-channel reconstruction problem. The proposed method has been used for many contrasts and multiple sampling strategies, additionally it has been extended to the 3D imaging case and has been used as the reference reconstruction technique involved in the 2D and 3D SPARKLING papers [Lazarus 2019a, Lazarus 2019b]. However the sensitivity estimation constitutes the major limitation for the application to online reconstruction. Hence, after a deep review of the literature on image reconstruction for non-Cartesian sampling and multiple trials, calibrationless reconstruction appeared as the most appealing technique. Calibrationless methods have been preferred since they do not require any prior knowledge on the coil sensitivity profiles and do not involve any gridding step – which are known to deteriorate the final image quality. Consequently, the quest was to design a new regularization scheme that achieves at least the same results as the self-calibrating methods. The outcomes were beyond our expectations since calibrationless reconstruction, with a dedicated structured sparsity norms, outperforms the self-calibration techniques, both with respect to the image magnitude and more importantly regarding the recovery of phase information. The proposed calibrationless reconstruction method constitutes the second contribution of this thesis. Finally, the ultimate goal was to extend this calibrationless method to the online framework. However due to the computational load, the first online results were disappointing, consequently we investigated a faster solution for which only the data-fidelity term was considered and optimized over the scan time. Using this new strategy, we improved the image quality of the estimate obtained by the end of acquisition. Online reconstruction methods definitely constitutes the last contribution of this thesis.

Besides those scientific contributions, the goal was also to propose a new package for image reconstruction. Hence with the help of CosmoStat laboratory, we have developed a new python package dedicated to the reconstruction named *pySAP*. The pySAP package was also adapted to be included within the Gadgetron project and thus enable the reconstruction to be performed in the scanner room and the results sent back to the scanner console.

Although the results of our calibrationless method seem promising, the set-up of hyper-parameters remains an open question, constituting one of the main limitations of these methods. On top of these considerations, the major limitation remains the heavy computation load associated with the non-Cartesian Fourier operator. Although we used GPU accelerations for Non-Cartesian Fourier operations in 2D imaging we did not fully implement the whole algorithm in GPU. This might be helpful in the future, especially for 3D imaging combined with multiple receiver phased arrays, for which the increasing number

of channels directly affects the computation cost.

Perspectives

Many more developments could be thought of using the calibrationless approaches proposed in this PhD thesis. For instance, structured sparsity norms could be adapted to deal with multi-contrast or dynamic MRI. Moreover, susceptibility weighting imaging (SWI) could also benefit from the good performances of calibrationless reconstruction, especially owing to the well preservation of phase information involved in the post-processing of SWI images. This work is currently addressed in the MANIAC project where patients and healthy volunteers are scanned on a 3 T MR system using SPARKLING acquisitions and the proposed calibrationless reconstruction method.

Regarding the computational aspects, the non-Cartesian Fourier transform could benefit from a faster implementation on GPUs. Also to avoid the time loss during the communication between the GPU and CPU, our calibrationless method could be made scalable by implementing it in a pure GPU framework such as the CuPy module.

Last, the reconstruction methods proposed in this thesis are inherently slow due to their iterative nature, therefore to speed up reconstruction, a promising avenue lies in the usage of Deep Learning (DL) approaches instead. In particular unrolled model-based optimization methods [Kamilov 2016, Yang 2016, Adler 2018] seem the most promising. The reason for this is rather simple, they are less mysterious than data-driven counterparts.However DL methods may require a large database of raw data (*i.e.* high resolution multi-channel complex-valued k-space measurements unavailable until recently [Zbontar 2018]). Although the application of these methods are promising, it is worth mentioning that they are sensitive to small perturbations or adversarial attacks as it has been proved in [Antun 2019].

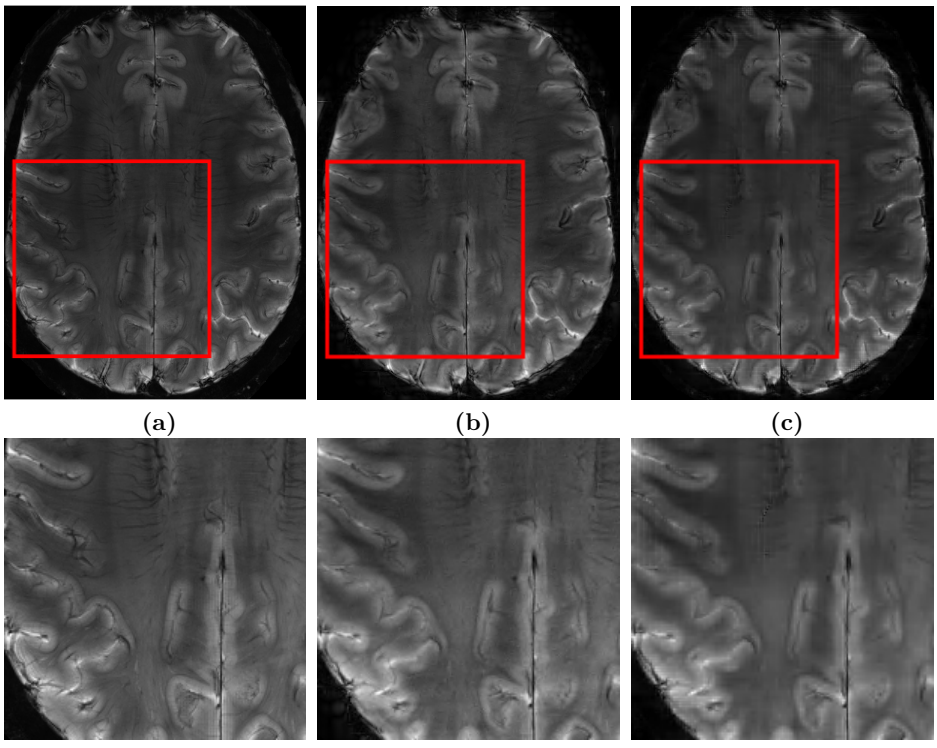
All in all, we have proposed a new framework to adapt MR image reconstruction algorithms to the context of high-resolution highly accelerated non-Cartesian and parallel imaging, and deploy them in a clinical environment.

* * *
* *
*

Appendices

Appendix A

Additional images Chapter 3 In vivo acquisitions



(a) (b) (c)

(d) (e) (f)

Figure A.1: **(a)** Cartesian reference, **(b)** Sparse-SENSE reconstruction with the proposed sensitivity extraction method and **(c)** ℓ_1-ESPIRiT reconstructions from 8-fold accelerated prospective CS based on Sparkling trajectories. **(d)-(f)** respective zooms in the red square.

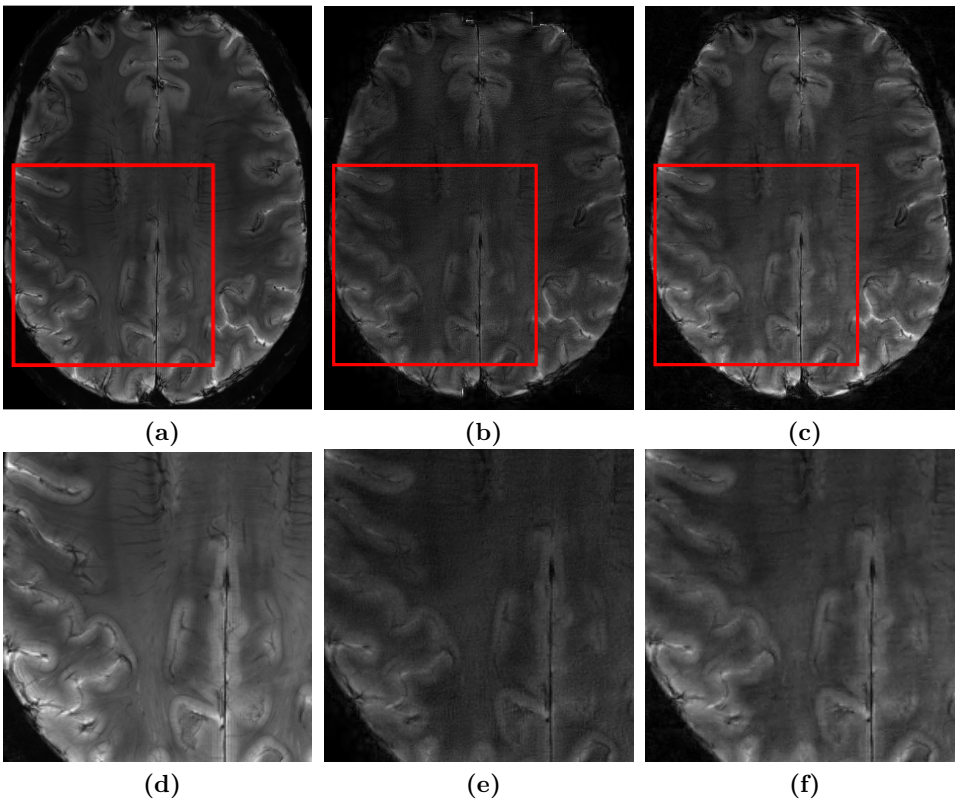


Figure A.2: **(a)** Cartesian reference, **(b)** Sparse-SENSE reconstruction with the proposed sensitivity extraction method and **(c)** ℓ_1-ESPIRiT reconstructions from 15-fold accelerated prospective CS based on Sparkling trajectories. **(d)**-**(f)** respective zooms in the red square.

Appendix B

Proximity Operator details of the Ordered Weighted ℓ_1 norm

Let $\boldsymbol{z} \in \mathbb{C}^p$ with $p \geq 1$. We introduce the magnitude sorting operator $S_p : \mathbb{C}^p \to \mathbb{C}^p$ such that vector $S_p(\boldsymbol{z}) = (S_p(\boldsymbol{z})_j)_{1 \leq j \leq p}$ contains the p entries of $\boldsymbol{z}$ sorted in decreasing order in magnitude, *i.e.* such that

$$|S_p(\boldsymbol{z})_1| \geq |S_p(\boldsymbol{z})_2| \geq \cdots \geq |S_p(\boldsymbol{z})_p|. \tag{B.1}$$

For the sake of compactness, we renamed the sorting operator S_p as the permutation matrix $\boldsymbol{P}_\downarrow$, by definition, matrix $\boldsymbol{P}_\downarrow$ is orthogonal. We recall the definition of the OSCAR norm:

$$\Omega_{\lambda,\gamma}(\boldsymbol{z}) = \lambda \|\boldsymbol{z}\|_1 + \gamma \sum_{1 \leq j < k \leq p} \max(|z_j|, |z_k|). \tag{B.2}$$

with λ, γ two positive hyper-parameters. As pointed out in [Zeng 2014a, Sec. II. A.], OSCAR norm has a closed relation with the OWL norm defined below:

$$\Theta_{\boldsymbol{w}}(\boldsymbol{z}) = \sum_{j=1}^{p} w_j |(\boldsymbol{P}_\downarrow \boldsymbol{z})_j|. \tag{B.3}$$

with $\boldsymbol{w} \in \mathbb{R}_+^p$ a vector of hyper-parameters such that $w_1 \geq \cdots \geq w_p \geq 0$. More precisely, OWL and OSCAR become equivalent if one sets the OWL weights as follows: $w_j = \lambda + \gamma(p - j)$ for $j = 1, \ldots, p$.

Proposition B.0.1. *The OSCAR norm defined in Eq.* (B.2) *is equivalent to the OWL norm in Eq.* (B.3) *with weights* $w_j = \lambda + \gamma(p - j)$ *for* $j = 1, \ldots, p$.

Proof. Set $\boldsymbol{z} \in \mathbb{C}^p$. First, let us remark that, for any permutation matrix $\boldsymbol{P} \in \mathbb{R}^{p \times p}$, $\Omega_{\lambda,\gamma}(\boldsymbol{z}) = \Omega_{\lambda,\gamma}(\boldsymbol{P}\boldsymbol{z})$. Let us consider now the particular permutation matrix $\boldsymbol{P}_\downarrow$ such that

$\boldsymbol{P}_\downarrow \boldsymbol{z} = \boldsymbol{z}_\downarrow$. Then, $\Omega_{\lambda,\gamma}(\boldsymbol{P}_\downarrow \boldsymbol{z}) = \Omega_{\lambda,\gamma}(\boldsymbol{z}_\downarrow)$ and we get:

$$\begin{aligned}\Omega_{\lambda,\gamma}(z_\downarrow) &= \lambda \sum_{j=1}^{p} |(\boldsymbol{z}_\downarrow)_j| + \gamma \sum_{1\leq j<k\leq p} \max(|(\boldsymbol{z}_\downarrow)_j|, |(\boldsymbol{z}_\downarrow)_k|) \\ &= \lambda \sum_{j=1}^{p} |(\boldsymbol{z}_\downarrow)_j| + \gamma \sum_{j=1}^{p} (p-j)|(\boldsymbol{z}_\downarrow)_j| \\ &= \sum_{j=1}^{p} (\lambda + \gamma(p-j))|(\boldsymbol{z}_\downarrow)_j| = \Theta_{\boldsymbol{w}}(\boldsymbol{z})\end{aligned} \tag{B.4}$$

with the above mentioned definition of $(w_j)_{1\leq j\leq p}$. □

Proximity operator

By definition, the proximity operator of the OSCAR norm at $\boldsymbol{z}$ reads:

$$\begin{aligned}\operatorname{prox}_{\Omega_{\lambda,\gamma}}(\boldsymbol{z}) = \underset{\boldsymbol{v}\in\mathbb{C}^p}{\arg\min} \Big\{ &\frac{1}{2}\|\boldsymbol{z}-\boldsymbol{v}\|^2 + \lambda\|\boldsymbol{v}\|_1 \\ &+\gamma \sum_{1\leq j<k\leq p} \max\left(|v_j|, |v_k|\right) \Big\}.\end{aligned} \tag{B.5}$$

Since the $\boldsymbol{P}_\downarrow$ operator is orthogonal and using the property of composition with semi-orthogonal linear transform as defined in [Combettes 2011, Table 10.1], and Proposition B.0.1 it allows to deduce:

Proposition B.0.2. *For any permutation matrix $\boldsymbol{P}$ the following equality holds:*

$$\operatorname{prox}_{\Omega_{\lambda,\gamma}}(\boldsymbol{z}) = \boldsymbol{P}^\top \operatorname{prox}_{\Omega_{\lambda,\gamma}}(\boldsymbol{P}\boldsymbol{z}) \tag{B.6}$$

Proof. By definition, matrix $\boldsymbol{P}$ is orthogonal, *i.e.* $\boldsymbol{P}^\top\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}^\top = \boldsymbol{I}_p$, the p-dimensional identity matrix. Moreover $\Omega_{\lambda,\gamma}(\boldsymbol{z}) = \Omega_{\lambda,\gamma}(\boldsymbol{P}\boldsymbol{z})$. Using the property of composition with semi-orthogonal linear transform as defined in [Combettes 2011, Table 10.1] we get $\operatorname{prox}_{\Omega_{\lambda,\gamma}}(\boldsymbol{z}) = \boldsymbol{z} + \boldsymbol{P}^\top(\operatorname{prox}_{\Omega_{\lambda,\gamma}}(\boldsymbol{P}\boldsymbol{z}) - \boldsymbol{P}\boldsymbol{z})$ and finally the result as $\boldsymbol{P}$ is orthogonal.

□

$$\operatorname{prox}_{\Omega_{\lambda,\gamma}}(\boldsymbol{z}) = \boldsymbol{P}_\downarrow^\top \operatorname{prox}_{\Theta_{\boldsymbol{w}}}(\boldsymbol{P}_\downarrow \boldsymbol{z}) \tag{B.7}$$

with $\boldsymbol{w}$ defined as Proposition B.0.1. The proximity operator of OWL norm can be solved efficiently using the following proposition.

Proposition B.0.3. *Let $\boldsymbol{z}$ in $\mathbb{C}^p$ and the OWL norm defined in Eq.* (B.3)*. Then, the following equality holds:*

$$\operatorname{prox}_{\Theta_{\boldsymbol{w}}}(\boldsymbol{z}_\downarrow) = \operatorname{sign}(\boldsymbol{z}_\downarrow) \odot \operatorname{PAV}(|\boldsymbol{z}_\downarrow| - \boldsymbol{w}) \tag{B.8}$$

*with $\odot$ being the Hadamard product (*i.e. *element-wise),* $\operatorname{sign}(v) = \frac{v}{|v|}, \forall v \in \mathbb{C}^*$ *and* $\operatorname{sign}(0) = 0$.

Proof. Let $\boldsymbol{w} \in \mathbb{R}_+^p$ being a vector of coefficients sorted in non-increasing order,

$$
\begin{aligned}
\operatorname{prox}_{\Theta_{\boldsymbol{w}}}(\boldsymbol{z}_{\downarrow}) &= \underset{\boldsymbol{v}\in\mathbb{C}^p}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{v}-\boldsymbol{z}_{\downarrow}\|_2^2 + \Theta_{\boldsymbol{w}}(\boldsymbol{v}) \right\} \\
&= \underset{\boldsymbol{v}\in\mathbb{C}^p}{\arg\min} \left\{ \frac{1}{2}\|\operatorname{sign}(\boldsymbol{z}_{\downarrow}) \odot (\operatorname{sign}(\boldsymbol{z}_{\downarrow}) \odot \boldsymbol{v} - |\boldsymbol{z}_{\downarrow}|)\|_2^2 + \Theta_{\boldsymbol{w}}(\operatorname{sign}(\boldsymbol{z}_{\downarrow})\boldsymbol{v}) \right\} \\
&= \operatorname{sign}(\boldsymbol{z}_{\downarrow}) \odot \underset{\boldsymbol{v}\in\mathbb{R}^p}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{v}-|\boldsymbol{z}_{\downarrow}|\|_2^2 + \Theta_{\boldsymbol{w}}(\boldsymbol{v}) \right\} \\
&= \operatorname{sign}(\boldsymbol{z}_{\downarrow}) \odot \underset{\substack{\boldsymbol{v}\in[0,+\infty)^p \\ \boldsymbol{v}_1\geq\ldots\boldsymbol{v}_p\geq 0}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{v}-|\boldsymbol{z}_{\downarrow}|\|_2^2 + \boldsymbol{w}^\top \boldsymbol{v} \right\} \\
&= \operatorname{sign}(\boldsymbol{z}_{\downarrow}) \odot \underset{\substack{\boldsymbol{v}\in[0,+\infty)^p \\ \boldsymbol{v}_1\geq\ldots\boldsymbol{v}_p\geq 0}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{v}-|\boldsymbol{z}_{\downarrow}|\|_2^2 \right. \\
&\left. + \boldsymbol{w}^\top(\boldsymbol{v}-|\boldsymbol{z}_{\downarrow}|) + \frac{1}{2}\|\boldsymbol{w}\|_2^2 \right\} \\
&= \operatorname{sign}(\boldsymbol{z}_{\downarrow}) \odot \underbrace{\underset{\substack{\boldsymbol{v}\in[0,+\infty)^p \\ \boldsymbol{v}_1\geq\ldots\boldsymbol{v}_p\geq 0}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{v}+\boldsymbol{w}-|\boldsymbol{z}_{\downarrow}|\|_2^2 \right\}}_{\text{Pool Adjacent Violator}} .
\end{aligned}
$$

□

Combining Propositions B.0.2 and B.0.3, it allows to deduce the proximity operator of the OSCAR norm.

* * *
* *
*

Appendix C

Softwares

C.1 Gadgetron

The Gadgetron is an open source framework for medical image reconstruction developed at the National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA and at the Department of Computer Science and Department of Clinical Medicine, Aarhus University, Denmark. It is made freely available to the medical image reconstruction community. It has originally been proposed for reconstruction purposes and offers many advantages listed hereafter:

- *Facilitate cross platform Development*: The platform is operating-system independent such that the users can then deploy it on their platform of choice, which may be dictated by the medical imaging device or their local computing environment.
- *Flexibility*: The project has been designed to be compatible over several vendors, and is not only restricted to one modality. It can be used either for CT, PET or MRI modality.
- *Modular*: Different modules are common to all the modalities and can then be shared across vendors.
- *Facilitate deployment*: Once the communication between the Gadgetron host and the scanner is set the deployment of the new algorithm can be easily performed and since the framework is used directly in the scanner loop, it allows online reconstruction but also online post-processing to be completely transparent to the user view point.
- *Various programming languages*: One of the key strengths of such method is its independence with respect to the development language hence C++, Matlab or even Python code can be run during the same pipeline.
- *Free*: The platform is freely available to all researchers.

Software architecture

In order to optimize the development of new reconstruction algorithms the idea of the Gadgetron project is to pool common functions. Each of these functions or modules (called Gadget) has an input fed with data and an output. Multiple Gadgets are wired together (connected) to assemble a reconstruction process (a Gadget stream). The Gadget

stream is configured dynamically at runtime using a plugin architecture. More specifically, the Gadgets of a particular reconstruction stream can reside in multiple independently compiled binaries (shared libraries) and are loaded at runtime based on a Gadget stream configuration in an extensible markup language (XML) file. Data are passed from a given client into the Gadgetron streaming architecture through a Transmission Control Protocol/Internet Protocol (TCP/IP) connection through which the resulting images are eventually returned to the client application.

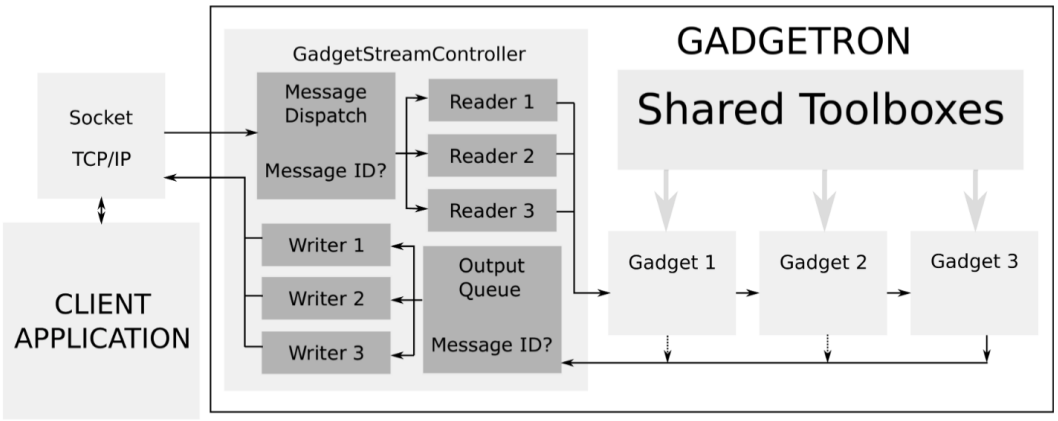


Figure C.1-1: Gadgetron communicates with the client through a TCP/IP protocol. The client application sends data to the Gadgetron and associated with each data package is a Message ID. Based on the Message ID, control of the socket is handed over to a specific Reader, which is capable of deserializing the incoming data package. The data package is converted to message blocks that are added to the first Gadget's queue. Data are then passed down the Gadget stream where each Gadget can modify and transform the data. Any Gadget can return images (or partially processed data) to the Gadgetron framework. Based on the MessageID of this return data package, the control of the socket and the data are handed to a particular Writer, which is responsible for writing the return message to the client. From [Hansen 2013].

In general the Gadgetron reconstruction process (Gadget stream) consists of three main components: "Readers", "Writers", and "Gadgets". These components are assembled and controlled by a GadgetStreamController, which is also responsible for maintaining the connection with the client. The role of the Reader is to receive and deserialize data from the client, and the Writer serves the (opposite) role of serializing and returning data to the client. Once the data has been received, it is passed down the chain of Gadgets. Each Gadget can modify or transform the data before passing it on to the subsequent Gadget.

Adding Python modules

To make the transition from prototyping to deployment easier, the Gadgetron is designed so that the functionality of individual Gadgets can be implemented using Python as a scripting language. The Python scripting functionality can be accessed by the user through a PythonGadget, and the framework provides certain hooks so that the user can implement callback functions for in a Python module. The user specifies which Python module

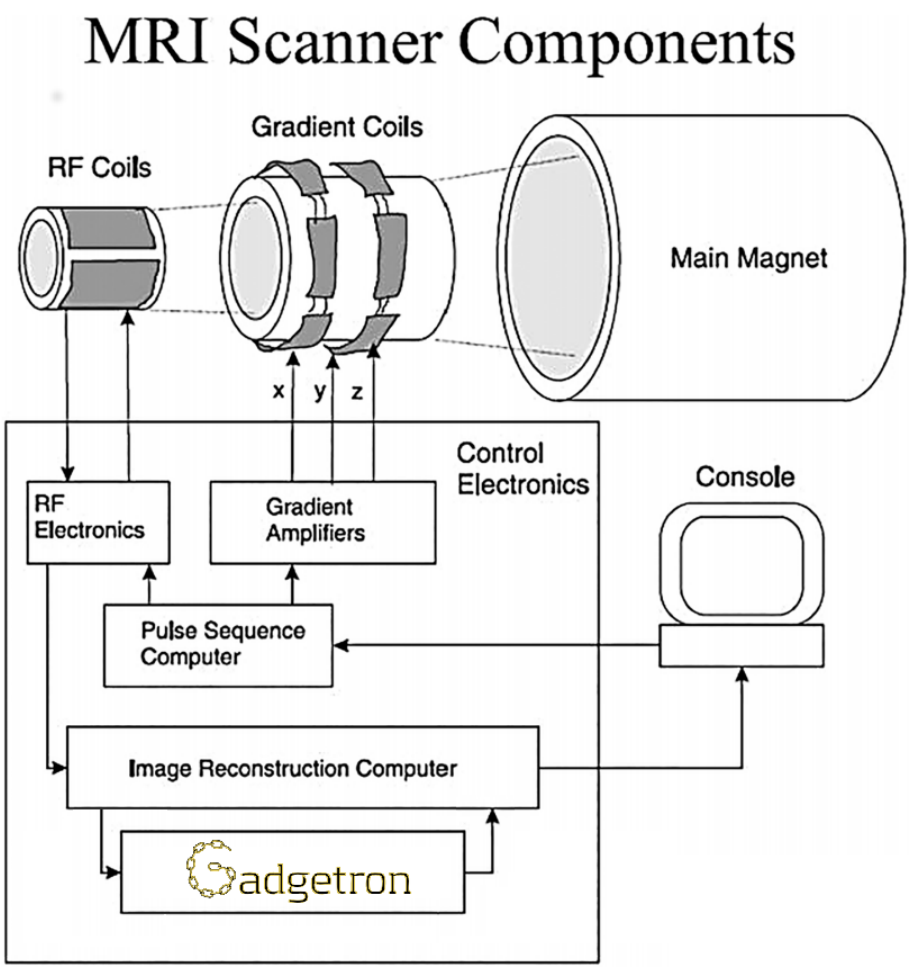


Figure C.1-2: Schematic view of the Gadgetron installation at the 7T scanner.

the Gadget should load and which functions in the Python module should be called in response to which events. It is possible to have an arbitrary number of PythonGadgets in the reconstruction pipeline. Moreover, the PythonGadgets can be mixed with standard Gadgets implemented purely in C/C++. This enables the user to reuse existing, efficient implementations of reconstruction steps while maintaining the capability of prototyping in Python.

Communication with the 7T scanner

In order to send the reconstruction proposed during this thesis directly on the scanner console, we rely on the Gadgetron framework as the latter allows rapid prototyping. But the first step consists in making the communication between the scanner and the Gadgetron host. The schematic Figure C.1-2 view presents the network architecture set-up on the 7T scanner.

C.2 PySAP: Python Sparse data Analysis Package

REPRODUCIBILITY is among the main objectives of the modern scientific approach and relies nowadays on open source software development. The latter guarantees the user trust in new approaches, facilitates their spread, helps benchmark different competing methods to solve the same problem and eventually provides educational tools through legible source code. In this context, the Compressed Sensing for Magnetic Resonance Imaging and Cosmology (COSMIC) project[1] was funded by the Fundamental Research Division (DRF) at the French Alternative Energies and Atomic Energy Commission (CEA) in 2016. This project is a collaboration between two computer scientist CEA teams, namely CosmoStat[2] (cosmologist) and NeuroSpin (MR specialist), working together to develop and implement new statistical signal processing methods to recover artifact-free high-resolution images from under-sampled, potentially blurred and noisy data. To this end, they have shared efforts to provide a reliable open-source python package called PySAP[3], which stands for Python Sparse data Analysis Package.

General Architecture

The PySAP project is organized around a core architecture and plugins for end-user applications. The front-end provides a framework for combining modules, managing IO files, displaying images and handling exceptions. Two modules compose the core of PySAP, namely Sparse2D and ModOpt. Figure C.2-3 depicts the schematic view of the functionalities in PySAP.

On one hand, Sparse2D proposes many sparsifying multi-scale image transforms such as decimated and undecimated wavelets but also more uncommon ones such as Curvelets, Shearlets or Starlets transforms [Starck 2007]. On the other hand, ModOpt provides a Modular Optimization module that gathers algorithms for nonsmooth convex optimization such as the ones exposed in Section 2.3.2 (e.g.proximal gradient methods and primal-dual algorithms), with corresponding proximity operators (cf the ones proposed in Chapter 4). Currently, the existing plugins combine the features provided by Sparse2D and ModOpt for a given application, hence, pysap-MRI and pysap-astro are respectively dedicated to MR image reconstruction and astronomical image processing.

Over the last three years, my contributions were essentially on the MR-related plugins namely PySAP-MRI and PySAP-Gadgetron.

PySAP-MRI

PySAP-MRI [4] is an MR related Python plugin, which aims at reconstructing complex-valued MR images from under-sampled k-space data. This module provides reconstruction algorithms for Cartesian and non-Cartesian data by embedding the NFFT [Keiner 2009]

[1] http://cosmic.cosmostat.org/
[2] http://www.cosmostat.org/
[3] http://github.com/cea-cosmic/pysap/
[4] https://github.com/CEA-COSMIC/pysap-mri

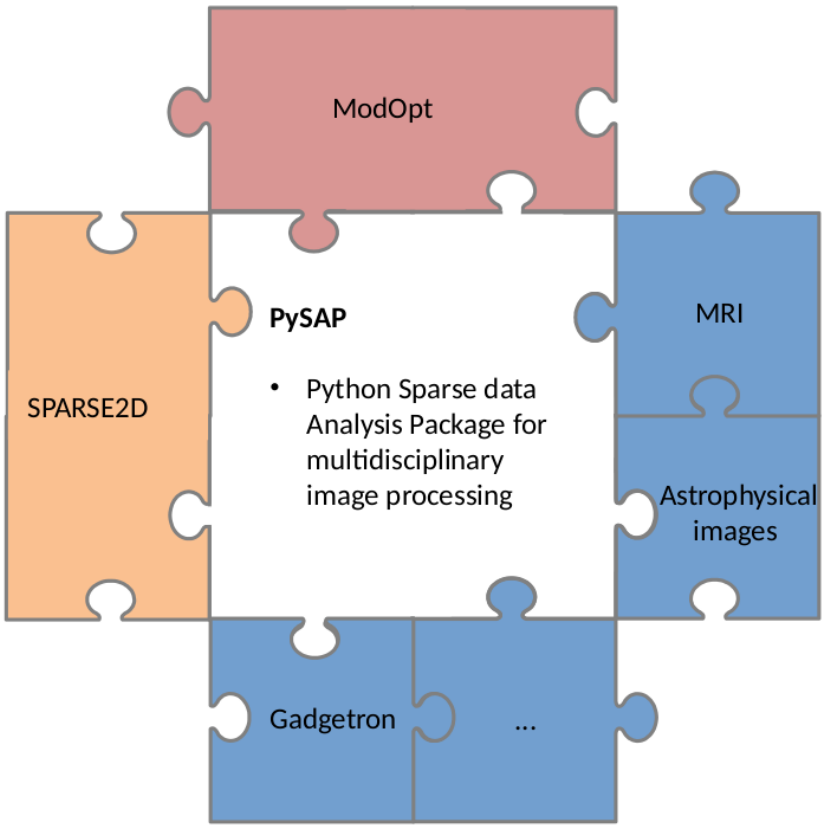


Figure C.2-3: Schematic view of the PySAP architecture.

operator, but also multi-channel data collected over phased array coils in parallel imaging. In this context, both two-step self-calibrating and one-step calibrationless reconstruction methods have been developed in this plugin. For the sake of conciceness, we remind here that the former perform sensitivity matrix extraction prior to the reconstruction of a single full FOV image while the latter reconstructs as many images as receiver channels. Among all available features, a key aspect is to guarantee the reliability of the proposed classes and methods. Hence unit-tested have been written for each and every proposed operator as well as for the overall reconstruction methods. This makes continuous integration with Travis and CircleCI smoother. For instance, the non-uniform Fourier operator and its adjoint have been tested so as to check that the adjoint relation Eq. (C.1) is fulfilled. Also, in the simpler case when the collected k-space samples lie on the Cartesian grid, we guarantee that the output of NFFT and FFT operators provide the same result.

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}, \quad \langle \boldsymbol{A}\boldsymbol{x}\,, \boldsymbol{y} \rangle = \langle \boldsymbol{x}\,, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} \rangle \tag{C.1}$$

Concerning the optimization part, we have tested the gradient descent of all optimizers for the smooth part in the cost functions. We have also checked that the value of the results should be equal to the inverse Fourier transform for Cartesian sampling regardless of the Fourier operator used. Last, The presence of unit-test permits to increase the code reliability and provide use-case to end-users.

PySAP-gadgetron

The PySAP-gadgetron plugin [5] proposes an interface between the MR scanner and the Python reconstruction code. It has been implemented through the Gadgetron project (see Appendix C.1 for details on the Gadgetron).

* * *
* *
*

[5] https://github.com/CEA-COSMIC/pysap-gadgetron

Appendix D

Résumé en français

Abstract in French

Sujet : Reconstruction d'images haute résolution acquises par échantillonnage compressif et régularisé à l'aide d'opérateurs de parcimonie structurée.

Nous résumons ici les différents aspects abordés au travers de cette thèse. Après avoir décrit les enjeux et motivations qui nous ont poussé au développement des méthodes abordées dans ce travail, nous résumerons chacune des contributions.

Motivations et contextes

L'imagerie par résonance magnétique (IRM) est une technique qui a révolutionné, de par son caractère non-ionisant, l'imagerie des tissues mous et en particulier celle du cerveau. Malgré une bonne résolution spatiale, la résolution temporelle en IRM reste problématique lorsque l'on considère l'acquisition d'images très haute résolution ($<400\mu$m). Afin de pallier à ce problème, la théorie de l'échantillonnage compressif propose une solution élégante pour raccourcir le temps d'acquisition. En effet, elle allie l'incohérence du schéma de sous-échantillonnage à la structure de parcimonie du signal pour garantir des preuves de reconstruction exacte de ce dernier. Bien que les premiers schémas de sous-échantillonnage aient prouvé leur efficacité, ces derniers reposent principalement sur des structures simples, comme un sous-échantillonnage cartésien, radial ou encore spiral. Récemment l'équipe dans laquelle j'effectue ma thèse a développé une méthode d'optimisation du schéma de sous-échantillonnage prometteuse appelée SPARKLING. Cette méthode d'optimisation permet de déplacer les points en entrée de manière à générer des trajectoires physiquement plausibles – *i.e.* qui répondent aux contraintes des gradients et donc implémentables sur le scanner – tout en respectant une densité de distribution cible. La méthode a originellement été proposée par [Boyer 2016] et adaptée par la suite par [Lazarus 2019a] afin de prendre en compte les contraintes liées au contraste souhaité – *e.g.* T_1w, T_2^*... – c'est-à-dire les contraintes liées au temps d'écho. Bien que les temps d'acquisition furent raccourcis,

l'application de cette théorie dans le monde clinique reste problématique. La raison en est toute simple : les temps de reconstructions restent rédhibitoires. L'objectif de ce travail est donc de proposer des méthodes efficaces de reconstruction. Pour cela nous nous appuierons sur la reconstruction en-ligne, c'est-à-dire, lorsque celle-ci débute durant l'acquisition à partir de données incomplètes.

Introduction à l'IRM

Un scanner IRM est organisé autour d'un puissant aimant caractérisé par la puissance de son champ magnétique. Lorsqu'un volume est placé dans un champ magnétique statique noté B_0, les spins de protons d'hydrogène composant le volume s'alignent ou s'anti-alignent le long du champ, laissant néanmoins une légère propension pour l'état de plus faible énergie, créant ainsi à l'échelle macroscopique un moment magnétique, appelé $\boldsymbol{M}_0$. Afin d'obtenir un signal variant dans le temps, l'aimantation $\boldsymbol{M}_0$ est basculée grâce à un champ radio-fréquence (RF), noté B_1^+ transverse au champ statiques B_0, c'est le phénomène d'excitation. Une fois l'aimantation basculée, l'impulsion RF le champs B_1^+ est arrêtée et l'aimantation revient à son état d'équilibre en tournant autour de B_0 (phénomènes de relaxation et de précession). La fréquence de rotation – aussi appelée fréquence de Larmor – dépend de la composition du tissu mais également de l'intensité du champ magnétique généré. Durant la relaxation un signal appelé FID est mesuré. Celui-ci peut-être caractérisé par deux constantes de temps :

- T_1, correspondant à la constante de temps décrivant le retour à l'état d'équilibre de la composante longitudinale ;
- T_2, correspondant à la constante de temps décrivant la disparition de la composante transverse.

Ces deux constantes sont caractéristiques du tissu sondé. Afin de discriminer spatialement le signal acquis, il est important d'encoder spatialement celui-ci. Pour cela, des gradients d'encodage, permettent de faire changer localement la fréquence et phase encodant ainsi le domaine conjugué de l'espace de Fourier aussi appelé espace k. Afin de reconstruire l'image à partir du signal magnétique électrique, la transformée de Fourier y est appliquée. Ainsi pour acquérir une image, il est d'usage de répéter l'excitation pour différentes valeurs de fréquence et de phase. De ce fait l'IRM est une modalité lente car les constantes de temps impliquées sont à minima de l'ordre de la dizaine de millisecondes. Pour accélérer les acquisitions, l'imagerie parallèle a été créée. En effet en sous-échantillonnant l'espace k de manière à acquérir une ligne sur deux, le temps d'acquisition peut se voir en effet réduit. Cependant, lors de la reconstruction le théorème de Shannon-Nyquist n'étant plus vérifié, du repliement est observé sur l'image. Pour s'affranchir de ce repliement, l'idée est de se baser sur l'information spatiale apportée par les canaux de réception de l'antenne en résolvant un système linéaire. Bien que les premières méthodes d'accélération développés fassent maintenant parti des séquences utilisées en clinique, les facteurs d'accélération correspondant restent faibles. Pour réduire d'avantage ces temps, l'échantillonnage compressif permet d'accélérer l'acquisition tout en garantissant des preuves de reconstruction

exactes faisant de cette nouvelle théorie un candidat potentiel pour massivement accélérer les examens IRM.

L'échantillonnage compressif : quand l'incohérence rencontre la parcimonie

Contrairement à la théorie de Shannon-Nyquist qui modélise un signal à bande limitée, l'échantillonnage compressif peut s'exprimer comme suit :

> « Il est possible de sous-échantillonner des signaux compressibles, c'est-à-dire parcimonieux dans une représentation donnée et de les reconstruire exactement. »

Cette théorie ne s'applique que dans certaines conditions résumées ci-dessous :

i) il existe une transformée dans laquelle on peut représenter le signal ou l'image de manière parcimonieuse (ou compressible) ;
ii) le schéma de sous-échantillonnage est incohérent avec la cette transformée ;
iii) la reconstruction doit promouvoir la parcimonie du signal.

En IRM, le signal acquis n'est pas nécessairement parcimonieux dans son espace d'origine, par conséquent le premier point consiste à trouver une représentation parcimonieuse du signal. La transformée en ondelette sera utilisée tout au long de cette thèse pour remplir cette fonction. Le deuxième point concerne les schémas de sous échantillonnage. Afin que l'acquisition soit physiquement implémentable sur le scanner, il est nécessaire de prendre en compte les contraintes liées à celui-ci. Ainsi l'objectif est de concevoir des trajectoires qui allient à la fois les contraintes des gradients ainsi qu'une bonne incohérence. De précédents sujets de thèse ont proposé un algorithme – appelé SPARKLING – permettant de générer des trajectoires efficaces pour l'accélération en IRM. A partir d'un schéma d'échantillonnage en entrée, l'algorithme SPARKLING génère une trajectoire physiquement plausible, à densité d'échantillonnage variable tout en évitant les amas ou les trous de points dans l'espace k. La méthode proposée a été appliquée à des acquisitions en pondération T_2^* avec succès, atteignant des facteurs d'accélération jusqu'à 20 en 2D et 60 en 3D.

Afin d'accélérer l'acquisition en IRM nous nous reposerons sur cette théorie. Nous utiliserons en particulier la méthode SPARKLING afin de générer des trajectoires implémentables sur le scanner tout en respectant les préceptes de l'échantillonnage compressif. Cependant, pour de l'imagerie haute résolution il est nécessaire d'avoir un bon rapport signal à bruit en entrée. De ce fait l'acquisition doit se faire à l'aide d'un réseaux d'antenne. La partie suivante explique comment combiner échantillonnage compressif avec utilisation d'un réseau d'antennes.

Reconstruction de données multi-canaux : approches classiques

Pour reconstruire les images acquises à l'aide d'un réseau d'antennes, les algorithmes de reconstruction d'images peuvent être classer en deux catégories, la première estime le résultat dans le domaine image, contrairement à la deuxième catégorie qui estime la

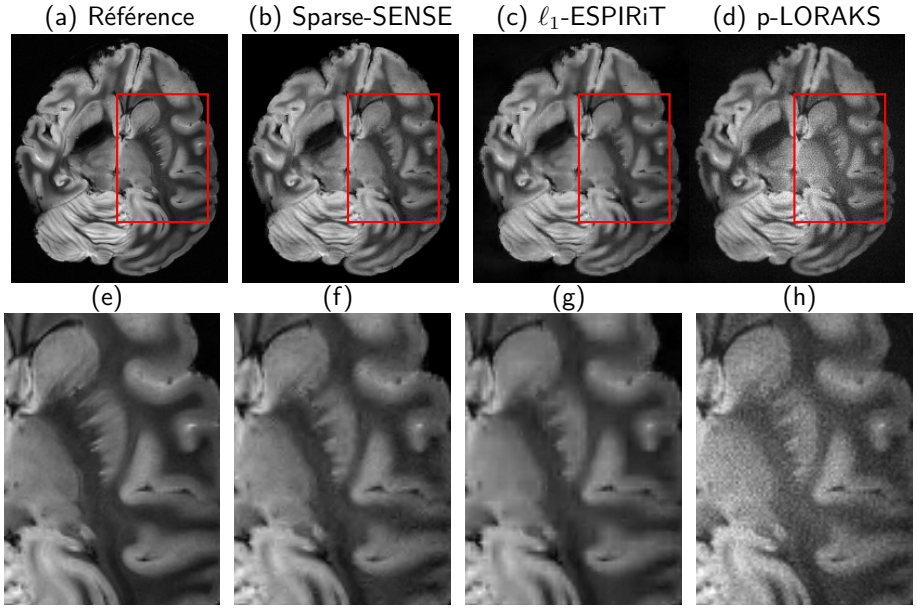


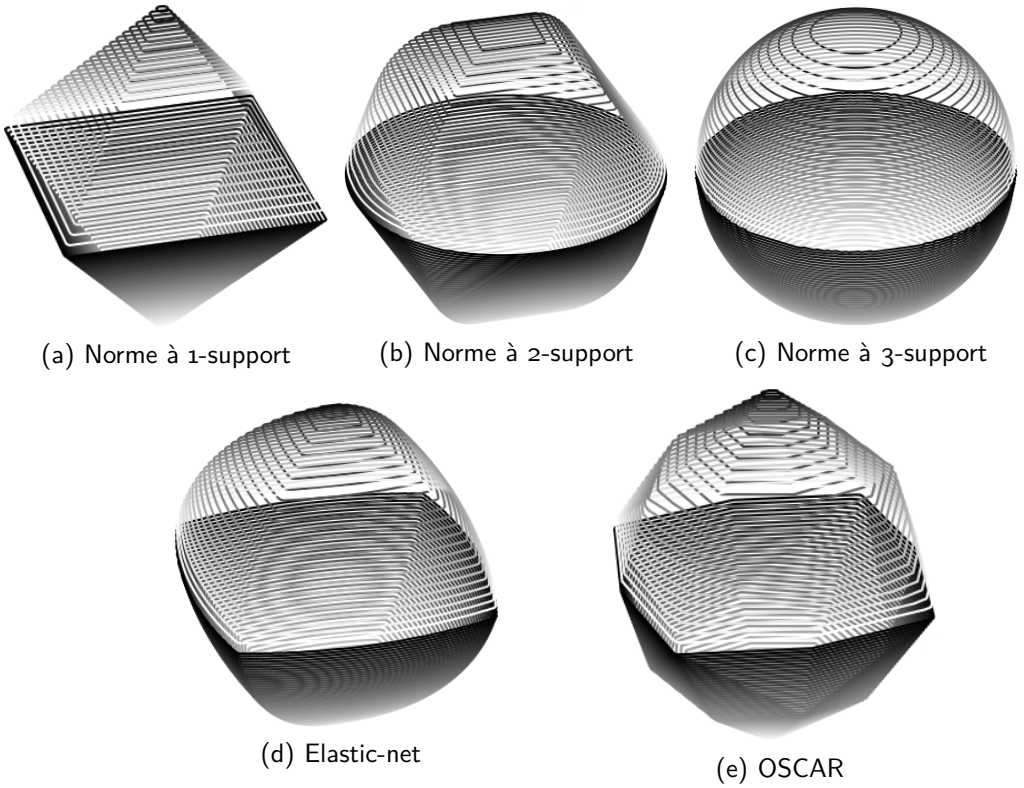
Figure D-1 : Comparaison des différentes méthodes de reconstructions d'images acquises à l'aide d'une trajectoire SPARKLING – dont le facteur d'accélération est de 15 en temps – et d'antenne multicanaux. (a) La référence Cartésienne, (b) reconstruction Sparse-SENSE avec notre méthode d'extraction des matrices de sensibilité, (c) la reconstruction ℓ_1-ESPIRiT, (d) La solution p-LORAKS après la projection sur la grille cartésienne. (e-h) leurs zooms respectifs.

solution dans l'espace k. Alors que les estimations dans l'espace k – *e.g.* p-LORAKS [Haldar 2016] – requièrent une étape de projection sur la grille détériorant ainsi la qualité finale de l'image, les estimations dans l'espace image – *e.g.* ℓ_1-ESPIRiT – reposent sur l'extraction de matrices de sensibilité, *i.e.* la connaissance à priori du profil de sensibilité de chacun des canaux. C'est dans ce but que nous proposons une méthode rapide d'estimation de ces matrices basée sur l'extraction du centre de cet espace, puis normalisées à l'aide de la somme des carrés. Cette méthode a été testée à 7 T sur un cerveau de babouin (à une résolution de $400\mu\text{m}^2$ sur le plan) et dont les résultats sont présentés Figure D-1. Dans cette figure on peut apercevoir le gain en terme de qualité images de l'approche proposée – correspondant au (b) – durant cette thèse.

Étant donnée la dépendance aux matrices de sensibilité, l'application de tels algorithmes à la reconstruction en-ligne nécessite l'estimation conjointe de l'image et des matrices de sensibilité rendant la reconstruction lente voire inefficace. Un moyen simple de contourner ce problème consiste à faire abstraction de cette information. Ce dernier point est discuté dans la prochaine section.

Reconstruction de données multi-canaux : approches sans calibration

Afin de s'affranchir de l'étape d'estimation des matrices de sensibilité, il est possible d'inférer une parcimonie de groupe à l'aide d'une régularisation promouvant une parcimonie structurée. L'objectif de ces nouvelles normes étant de sélectionner tout un ensemble de



(a) Norme à 1-support (b) Norme à 2-support (c) Norme à 3-support

(d) Elastic-net (e) OSCAR

Figure D-2 : **Première ligne :** Boule unitaire de la norme à k-support en 3D pour les différentes valeurs de k. (a) présente la boule unitaire pour k=1 – correspondant à la régularization ℓ_1, (b) présente la boule unitaire pour k=2, (c) présente la boule unitaire pour k=2 – correspondant à la régularisation ℓ_2. **Seconde ligne :** Régularisation Elastic-net – combinaison linéaire entre la norm ℓ_1 et ℓ_2 – et la boule unitaire d'OSCAR en 3D.

coefficients partageant le même degré de parcimonie. La parcimonie structurée permet ainsi de palier à l'inconvénient de la norme ℓ_1 lors de la présence de variables fortement corrélées, comme c'est le cas en IRM parallèle. En particulier, l'impact des normes OSCAR [Bondell 2008] a été étudié ainsi que de la norme à k-support [Argyriou 2012] – les boules unitaires sont représentées Figure D-2, l'effet de regroupement se manifeste par les sommets plats de leurs boules unitaires. Les différentes normes de parcimonie structurée sont alors adaptées, implémentées et testées sur des données haute résolution (400μm isotrope et 1.5 mm d'épaisseur de coupe) acquises à l'aide d'une trajectoire SPARKLING dont le facteur d'accélération est fixé à 20. Les résultats de ces reconstructions sont visibles sur le Figure D-3. En se basant sur les résultats illustrés Figure D-3, il est aisé de constater qu'avec une régularisation à k-support, la qualité de l'image – magnitude et phase – est préservée. Ainsi grâce à la combinaison de la trajectoire et de la reconstruction proposée, nous pouvons en déduire que l'information des matrices de sensibilité est facultative.

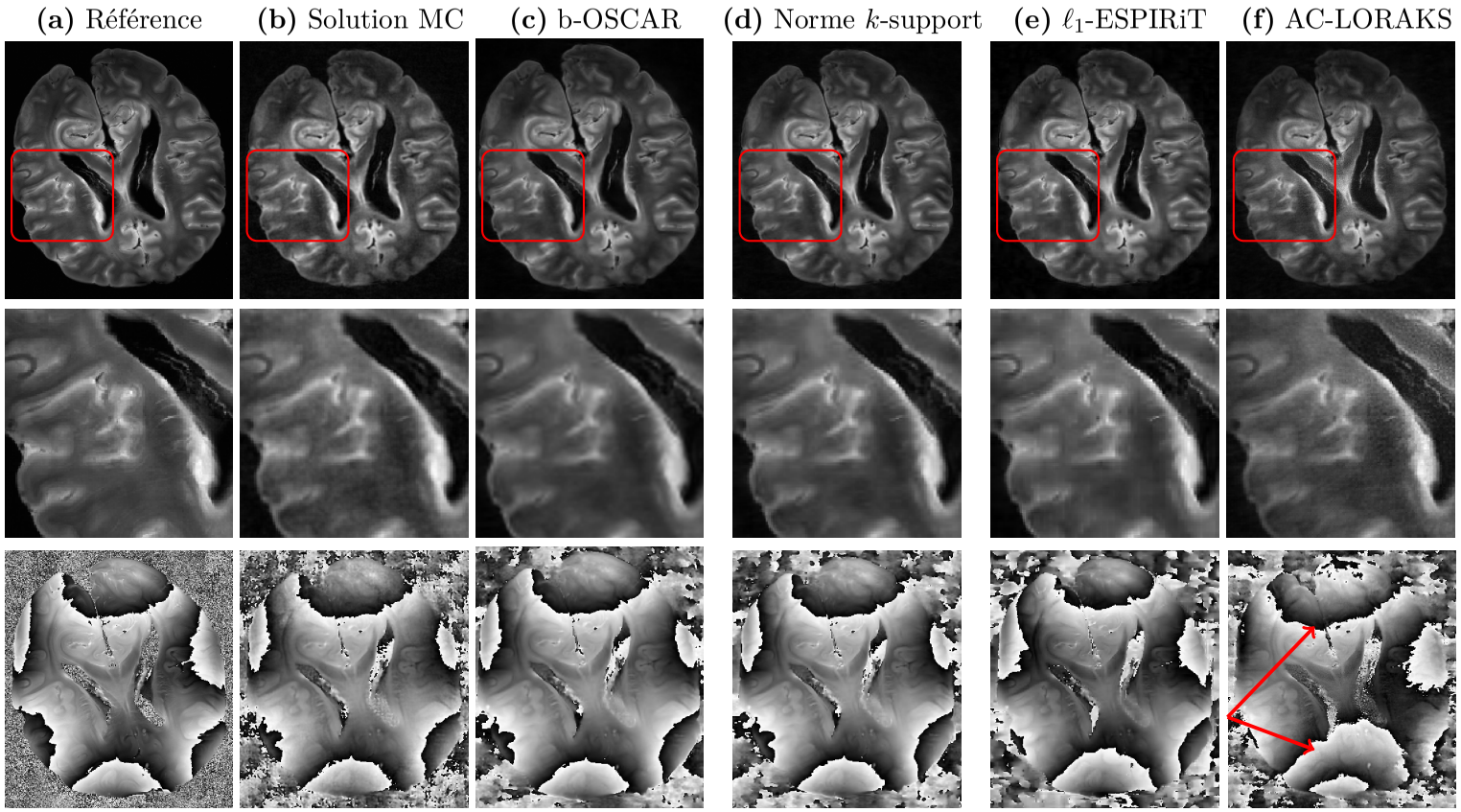


Figure D-3 : **Première ligne :** Magnitude des images acquises à l'aide des trajectoires SPARKLING avec un facteur d'accélération égal à 20. **(a)** Référence Cartésienne. **(b)** La transformée de Fourier inverse obtenue grâce à la méthode des moindres carrés. **(c)** Reconstruction obtenue par la méthode sans calibration basée sur la norme OSCAR. **(d)** Reconstruction obtenue par la méthode sans calibration basée sur la norme à k-support. **(e)** Reconstruction obtenue par la méthode auto-calibrée ℓ_1-ESPIRiT. **(f)** Reconstruction obtenue par la méthode auto-calibrée (AC) LORAKS basée sur la régularisation de l'espace k. **Deuxième ligne :** Zooms respectifs de la partie encadré en rouge. **Dernière ligne :** Images de phase extraites pour chacune des méthodes grâce à l'algorithme du canal virtuelle [Parker 2014].

Reconstruction en ligne

Dans la section précédente, nous avons proposé un algorithme de reconstruction d'images acquises par échantillonnage compressif à l'aide d'un réseau d'antennes. Il est maintenant temps de s'intéresser à la reconstruction en-ligne. Afin de se conformer aux constantes de temps impliquées dans l'acquisition, les différents segments sont concaténés par lots. Une fois un lot complet acquis, la reconstruction peut commencer. Dès lors qu'un nouveau lot est acquis, le résultat de la reconstruction précédente sert d'initialisation au traitement des nouvelles données entrantes. Un exemple des résultats obtenus à la fin de l'acquisition – c'est-à-dire avant le traitement du dernier lot – est présenté Figure. D-4. Lorsque tous les paquets ont été acquis et traités, la solution converge vers l'unique solution du problème qui correspond aux résultats hors-ligne. Ce dernier point constitue le principal avantage de cette méthode. De plus, en analysant les résultats de la reconstruction en ligne, nous pouvons clairement apercevoir l'impact de la taille des paquets. En effet avec des paquets plus petits l'image reconstruite à la fin de l'acquisition est de meilleure qualité.

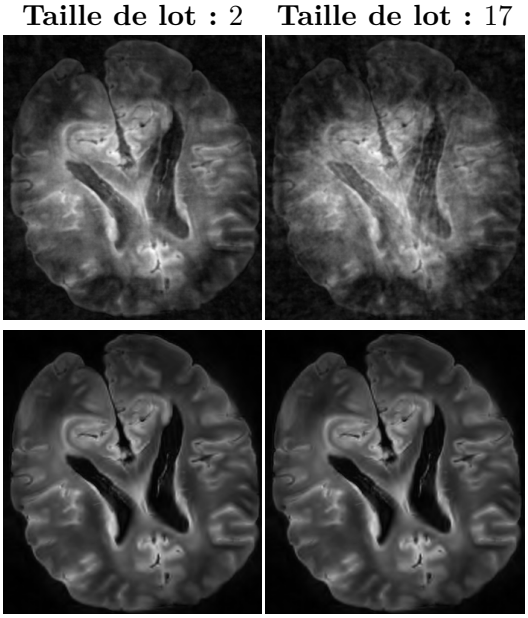


Figure D-4 : Reconstruction rapide en ligne de données SPARKLING accélérées 15 fois (S=34 segments) d'un cerveau humain ex vivo acquise par un réseau d'antennes. **Première ligne :** Les images reconstruites à la fin de l'acquisition, pour de tailles de paquets différentes (2 et 17 segments par paquets). **Seconde ligne :** Les images obtenues à convergence.

Conclusion

Au terme de ces trois années, nous espérons avoir amélioré l'application clinique de l'acquisition comprimé pour l'IRM haute résolution. L'objectif de cette thèse n'était pas seulement de proposer une méthode de reconstruction efficace pour une acquisition non carté-

sienne hautement accélérée, particulièrement adaptée à l'imagerie haute résolution, mais aussi d'accélérer la reconstruction afin de mettre ces trajectoires en pratique clinique.

L'imagerie à haute résolution nécessite un certain niveau de rapport signal à bruit en entrée (RSB), c'est pourquoi l'acquisition est souvent effectuée à l'aide d'un réseau d'antennes. Avant le début de ces travaux, aucun algorithme de reconstruction n'était disponible au sein du laboratoire pour ce cas là. De ce fait les acquisitions SPARKLING étaient entravées par le manque de RSB. Par conséquent, la première étape a consisté à mettre en œuvre une méthode simple mais efficace pour estimer les profils de sensibilité des antennes impliquées dans le problème de reconstruction multicanal de type SENSE [Pruessmann 1999]. La méthode proposée a été utilisée pour de nombreux contrastes et de multiples stratégies d'échantillonnage. Elle a ensuite été étendue au cas de l'imagerie 3D et a été utilisée comme technique de reconstruction de référence pour les articles scientifiques de SPARKLING 2D et 3D [Lazarus 2019a, Lazarus 2019b].

Cependant, l'estimation de la sensibilité constitue la principale limite pour la reconstruction en ligne. Ainsi, après un examen approfondi de la littérature sur la reconstruction d'image pour l'échantillonnage non cartésien et de multiples essais, la reconstruction sans calibration est apparue comme la technique la plus intéressante. Par conséquent, l'objectif était de concevoir de nouvelles régularisations ayant à minima les mêmes résultats que ceux obtenus avec les méthodes d'autocalibration. Les résultats ont été au-delà de nos espérances puisque la reconstruction sans calibration, se basant sur des normes de parcimonie structurée, surpasse les techniques d'autocalibration, aussi bien en termes d'images d'amplitude que de récupération de l'information de phase. Ces méthodes de reconstruction constituent la deuxième contribution de cette thèse.

Enfin, le but principal était d'étendre cette méthode sans calibration au framework en ligne, et grâce à cette nouvelle stratégie, nous avons amélioré la qualité de l'image estimée obtenue à la fin de l'acquisition. Les méthodes de reconstruction en ligne constituent finalement la dernière contribution de cette thèse.

Outre ces contributions scientifiques, l'objectif était également de proposer un nouveau package pour la reconstruction d'image. Ainsi, avec l'aide du laboratoire CosmoStat, nous avons développé un package python dédié à la reconstruction nommé *pySAP*. Le package pySAP a également été adapté pour être inclus dans le projet Gadgetron et permettre ainsi d'effectuer la reconstruction dans la salle du scanner et de renvoyer les résultats directement à la console du scanner, pour traitement supplémentaire, visualisation et archivage.

Dans l'ensemble, nous avons proposé un nouveau cadre algorithmique permettant l'application clinique de l'échantillonnage compressif et l'avons appliqué avec succès à la reconstruction de l'image IRM.

* * *
* *
*

Publications

Articles in Peer-Reviewed Journals

Loubna El Gueddari, Emilie Chouzenoux, Alexandre Vignaud, Jean-Christophe Pesquet and Philippe Ciuciu. *Calibrationless parallel imaging Compressed Sensing reconstruction based on OSCAR regularization.* Submitted.

Carole Lazarus, Pierre Weiss, Nicolas Chauffert, Franck Mauconduit, **Loubna El Gueddari**, Christophe Destrieux, Ilyess Zemmoura, Alexandre Vignaud and Philippe Ciuciu. *SPARKLING: variable density k-space filling curves for accelerated* T_2^**-weighted MRI.* Journal of Magnetic Resonance in Medicine, vol. 81, no. 6, pp. 3643-3661, 2019.

Carole Lazarus, Pierre Weiss, **Loubna El Gueddari**, Franck Mauconduit, Alexandre Vignaud and Philippe Ciuciu. *3D SPARKLING trajectories for high-resolution* T_2^**-weighted Magnetic Resonance imaging.* NMR in Biomedicine. In revision.

Antoine Grigis, Samuel Farrens, **Loubna El Gueddari**, Jean-Luc Starck, Philippe Ciuciu. *PySAP: Python Sparse Data Analysis Package for Multidisciplinary Image Processing.* To be submitted

International Conferences Paper Presented with Reading Committee and Proceedings

Loubna El Gueddari, Emilie Chouzenoux, Alexandre Vignaud, Jean-Christophe Pesquet, Philippe Ciuciu. *Online MR image reconstruction for compressed sensing acquisition in T2* imaging.* Proceedings of the Wavelets: Applications in Signal and Image Processing XVIII, San Diego, CA, USA, 2019. Oral.

Loubna El Gueddari, Philippe Ciuciu, Emilie Chouzenoux, Alexandre Vignaud, Jean-Christophe Pesquet. *Calibrationless OSCAR-based image reconstruction in compressed sensing parallel MRI.* Proceedings of the IEEE International Symposium of Biomedical Imaging, Venice, Italy, 2019. Oral.

Loubna El Gueddari, Carole Lazarus, Hanaé Carrié, Alexandre Vignaud, Philippe Ciuciu. *Self-calibrating nonlinear reconstruction algorithms for variable density sampling and parallel reception MRI.* Proceedings of the 10th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Shefield, UK, 2018. Oral.

Hamza Cherkaoui, **Loubna El Gueddari**, Carole Lazarus, Antoine Grigis, Fabrice Poupon, Alexandre Vignaud, Samuel Farrens, Jean-Luc Starck, Philippe Ciuciu. *Analysis Vs Synthesis-based Regularization for Combined Compressed Sensing and Parallel MRI Reconstruction at 7 Tesla.* Proceedings of the 26th Europeean Signal Processing Conference, Rome, Italy, 2018

Abstracts Presented at International Conferences with Reading Committee and Proceedings

Loubna El Gueddari, Emilie Chouzenoux, Jean-Christophe Pesquet, Alexandre Vignaud and Philippe Ciuciu. *OSCAR-based reconstruction for compressed sensing and parallel MR imaging.* Proceedings of the 27th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Montreal, Canada, 2019. ePoster.

Loubna El Gueddari, Emilie Chouzenoux, Jean-Christophe Pesquet, Alexandre Vignaud and Philippe Ciuciu. *Online compressed sensing MR image reconstruction for high resolution T2* imaging.* Proceedings of the 27th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Montreal, Canada, 2019. ePoster.

Carole Lazarus, Pierre Weiss, Nicolas Chauffert, Franck Mauconduit, **Loubna El Gueddari**, Christophe Destrieux, Ilyess Zemmoura, Alexandre Vignaud and Philippe Ciuciu. *SPARKLING: variable-density k-space filling curves for accelerated MRI.* Proceedings of the 27th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Montreal, Canada, 2019. ePoster.

Loubna El Gueddari, Carole Lazarus, Hanaé Carrié, Alexandre Vignaud and Philippe Ciuciu. *Self-calibrating nonlinear MR image reconstruction algorithms for variable*

density sampling and parallel imaging. Proceedings of the 28th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Paris, France, 2018. ePoster.

Carole Lazarus, Pierre Weiss, **Loubna El Gueddari**, Franck Mauconduit, Alexandre Vignaud and Philippe Ciuciu. *Distribution-controlled and optimally spread non-Cartesian sampling curves for accelerated in vivo brain imaging at* 7 T. Proceedings of the 28th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Paris, France, 2018. ePoster.

Loubna El Gueddari, Carole Lazarus, Hamza Cherkaoui, Elvis Dohmatob, Alexandre Vignaud, Philippe Ciuciu. *Self-calibrating non-linear reconstruction algorithm for variable density sampling and parallel reception MRI.* Proceedings of the IEEE International Symposium of Biomedical Imaging, Washington D.C., USA, 2018. Poster

Hanaé Carrié, **Loubna El Gueddari**, Hamza Cherkaoui, Elvis Dohmatob, Lisa Leroi, Philippe Ciuciu. *Multi-Contrast Dictionary Learning for 2D Compressed Sensing MRI Reconstruction.* Proceedings of the IEEE International Symposium of Biomedical Imaging, Washington D.C., USA, 2018.

Martin Jacob, **Loubna El Gueddari**, Gabriele Navarro, Cyrille Marie-Claire, Pascale Bayle-Guillemaud, Philippe Ciuciu, Zineb Saghi. *Statistical Machine Learning and Compressed Sensing Approaches for Analytical Electron Tomography - Application to Phase Change Materials.* Proceedings of Microscopy and Microanalysis, Portland, Oregon, 2019.

Bibliography

[Adcock 2017] B. Adcock, A. Hansen, C. Poon and B. Roman. *Breaking the coherence barrier: A new theory for compressed sensing.* Forum of Mathematics, Sigma, vol. 5, page e4, 2017. (Cited on pages 36 and 37.)

[Adcock 2018] B. Adcock, C. Boyer and S. Brugiapaglia. *On oracle-type local recovery guarantees in compressed sensing.* arXiv preprint arXiv:1806.03789, 2018. (Cited on page 38.)

[Adler 2018] J. Adler and O. Öktem. *Learned primal-dual reconstruction.* IEEE transactions on medical imaging, vol. 37, no. 6, pages 1322–1332, 2018. (Cited on pages 102 and 127.)

[Ahn 1986] C. Ahn, J. Kim and Z. Cho. *High-speed spiral-scan echo planar NMR imaging-I.* IEEE transactions on medical imaging, vol. 5, no. 1, pages 2–7, 1986. (Cited on page 42.)

[Antun 2019] V. Antun, F. Renna, C. Poon, B. Adcock and A. C. Hansen. *On instabilities of deep learning in image reconstruction-Does AI come at a cost?* arXiv preprint arXiv:1902.05300, 2019. (Cited on pages 102 and 127.)

[Argyriou 2012] A. Argyriou, R. Foygel and N. Srebro. *Sparse Prediction with the k-Support Norm.* In Advances in Neural Information Processing Systems, pages 1457–1465, 2012. (Cited on pages 80, 83, 86, 87, 88, 93, and 147.)

[Bauschke 2011] H. H. Bauschke, P. L. Combettes *et al.* Convex analysis and monotone operator theory in hilbert spaces, volume 408. Springer, 2011. (Cited on page 51.)

[Beatty 2005] P. J. Beatty, D. G. Nishimura and J. M. Pauly. *Rapid gridding reconstruction with a minimal oversampling ratio.* IEEE transactions on medical imaging, vol. 24, no. 6, pages 799–808, 2005. (Cited on page 21.)

[Beck 2009] A. Beck and M. Teboulle. *A fast iterative shrinkage-thresholding algorithm for linear inverse problems.* SIAM journal on imaging sciences, vol. 2, no. 1, pages 183–202, 2009. (Cited on pages 50 and 52.)

[Bekhti 2018] Y. Bekhti, F. Lucka, J. Salmon and A. Gramfort. *A hierarchical Bayesian perspective on majorization-minimization for non-convex sparse regression: application to M/EEG source imaging.* Inverse Problems, vol. 34, no. 8, page 085010, 2018. (Cited on page 82.)

[Bernstein 2004] M. A. Bernstein, K. F. King and X. J. Zhou. Handbook of mri pulse sequences. Elsevier, 2004. (Cited on pages 9, 39, and 106.)

[Bertsekas 2015] D. P. Bertsekas. Convex optimization algorithms. Athena Scientific Belmont, 2015. (Cited on page 80.)

[Blaimer 2004] M. Blaimer, F. Breuer, M. Mueller, R. M. Heidemann, M. A. Griswold and P. M. Jakob. *SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method.* Topics in Magnetic Resonance Imaging, vol. 15, no. 4, pages 223–236, 2004. (Cited on pages 28, 30, and 31.)

[Bloch 1946] F. Bloch. *Nuclear Induction.* Phys. Rev., vol. 70, pages 460–474, Oct 1946. (Cited on pages 10 and 14.)

[Boada 1997] F. E. Boada, J. S. Gillen, G. X. Shen, S. Y. Chang and K. R. Thulborn. *Fast three dimensional sodium imaging.* Magnetic resonance in medicine, vol. 37, no. 5, pages 706–715, 1997. (Cited on pages 42 and 43.)

[Bogdan 2015] M. Bogdan, E. Van Den Berg, C. Sabatti, W. Su and E. J. Candès. *SLOPE—adaptive variable selection via convex optimization.* The annals of applied statistics, vol. 9, no. 3, page 1103, 2015. (Cited on page 83.)

[Bondell 2008] H. D. Bondell and B. J. Reich. *Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR.* Biometrics, vol. 64, no. 1, pages 115–123, 2008. (Cited on pages 80, 83, and 147.)

[Boyd 2004] S. Boyd and L. Vandenberghe. Convex optimization. Cambridge university press, 2004. (Cited on pages 35 and 52.)

[Boyer 2012] C. Boyer, P. Ciuciu, P. Weiss and S. Mériaux. *HYR^{2}PICS: Hybrid regularized reconstruction for combined parallel imaging and compressive sensing in MRI.* In 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), pages 66–69. IEEE, 2012. (Cited on page 52.)

[Boyer 2016] C. Boyer, N. Chauffert, P. Ciuciu, J. Kahn and P. Weiss. *On the generation of sampling schemes for magnetic resonance imaging.* SIAM Journal on Imaging Sciences, vol. 9, no. 4, pages 2039–2072, 2016. (Cited on pages 44, 68, 90, and 143.)

[Boyer 2017] C. Boyer, J. Bigot and P. Weiss. *Compressed sensing with structured sparsity and structured acquisition.* Applied and Computational Harmonic Analysis, 2017. (Cited on pages 37, 38, and 42.)

[Breuer 2006] F. A. Breuer, M. Blaimer, M. F. Mueller, N. Seiberlich, R. M. Heidemann, M. A. Griswold and P. M. Jakob. *Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA).* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 55, no. 3, pages 549–556, 2006. (Cited on page 28.)

[Buehrer 2007] M. Buehrer, K. P. Pruessmann, P. Boesiger and S. Kozerke. *Array compression for MRI with large coil arrays.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 57, no. 6, pages 1131–1139, 2007. (Cited on page 123.)

[Candès 2006a] E. J. Candès, J. Romberg and T. Tao. *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.* IEEE Transactions on information theory, vol. 52, no. 2, pages 489–509, 2006. (Cited on pages 30 and 38.)

[Candès 2006b] E. J. Candès, J. K. Romberg and T. Tao. *Stable signal recovery from incomplete and inaccurate measurements.* Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, vol. 59, no. 8, pages 1207–1223, 2006. (Cited on page 35.)

[Candès 2008] E. J. Candès and M. B. Wakin. *An introduction to compressive sampling – a sensing/sampling paradigm that goes against the common knowledge in data acquisition.* IEEE signal processing magazine, vol. 25, no. 2, pages 21–30, 2008. (Cited on pages 35 and 80.)

[Candès 2011] E. J. Candès and Y. Plan. *A probabilistic and RIPless theory of compressed sensing.* IEEE transactions on information theory, vol. 57, no. 11, pages 7235–7254, 2011. (Cited on page 35.)

[Carrié 2018] H. Carrié, L. El Gueddari, H. Cherkaoui, E. Dohmatob, L. Leroi and P. Ciuciu. *Multi-Contrast Dictionary Learning for 2D Compressed Sensing MRI Reconstruction.* In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington D.C., USA, 2018. IEEE ISBI. (Cited on pages viii, 34, 48, and 153.)

[Chaâri 2008] L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia and P. Ciuciu. *Autocalibrated regularized parallel MRI reconstruction in the wavelet domain.* In 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pages 756–759. IEEE, 2008. (Cited on page 60.)

[Chaâri 2011] L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia and P. Ciuciu. *A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging.* Medical image analysis, vol. 15, no. 2, pages 185–201, 2011. (Cited on pages 60, 104, and 114.)

[Chambolle 2011] A. Chambolle and T. Pock. *A first-order primal-dual algorithm for convex problems with applications to imaging.* Journal of mathematical imaging and vision, vol. 40, no. 1, pages 120–145, 2011. (Cited on page 52.)

[Chambolle 2015] A. Chambolle and C. Dossal. *On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm".* Journal of Optimization theory and Applications, vol. 166, no. 3, pages 968–982, 2015. (Cited on pages 51, 52, and 53.)

[Chandarana 2014] H. Chandarana, K. T. Block, M. J. Winfeld, S. V. Lala, D. Mazori, E. Giuffrida, J. S. Babb and S. S. Milla. *Free-breathing contrast-enhanced T1-weighted gradient-echo imaging with radial k-space sampling for paediatric abdominopelvic MRI.* European radiology, vol. 24, no. 2, pages 320–326, 2014. (Cited on page 41.)

[Chauffert 2014] N. Chauffert, P. Ciuciu, J. Kahn and P. Weiss. *Variable density sampling with continuous trajectories. Application to MRI.* SIAM Journal on Imaging Sciences, vol. 7, no. 4, pages 1962–1992, November 2014. (Cited on pages 36 and 37.)

[Chauffert 2015] N. Chauffert. *Echantillonnage compressé le long de trajectoires physiquement plausibles en IRM.* PhD thesis, University of Paris Sud - Paris XI, 2015. (Cited on page 41.)

[Chauffert 2017] N. Chauffert, P. Ciuciu, J. Kahn and P. Weiss. *A projection method on measures sets.* Constructive Approximation, vol. 45, no. 1, pages 83–111, 2017. (Cited on page 44.)

[Chen 2006] J. Chen and X. Huo. *Theoretical results on sparse representations of multiple-measurement vectors.* IEEE Transactions on Signal processing, vol. 54, no. 12, pages 4634–4643, 2006. (Cited on page 82.)

[Cherkaoui 2018] H. Cherkaoui, L. El Gueddari, C. Lazarus, A. Grigis, F. Poupon, A. Vignaud, S. Farrens, J.-L. Starck and P. Ciuciu. *Analysis vs synthesis-based regularization for combined compressed sensing and parallel MRI reconstruction at 7 Tesla.* In 2018 26th European Signal Processing Conference, pages 36–40. EUSIPCO, 2018. (Cited on pages 34, 47, 92, 102, 107, 108, and 152.)

[Chun 2015] I. Y. Chun, B. Adcock and T. M. Talavage. *Efficient compressed sensing SENSE pMRI reconstruction with joint sparsity promotion.* IEEE transactions on medical imaging, vol. 35, no. 1, pages 354–368, 2015. (Cited on pages 82, 86, and 102.)

[Combettes 2011] P. L. Combettes and J.-C. Pesquet. *Proximal splitting methods in signal processing.* In Fixed-point algorithms for inverse problems in science and engineering, pages 185–212. Springer, 2011. (Cited on pages 49, 50, 51, 80, 106, and 134.)

[Combettes 2016] P. L. Combettes and J.-C. Pesquet. *Stochastic approximations and perturbations in forward-backward splitting for monotone operators.* Pure and Applied Functional Analysis, vol. 1, no. 1, pages 13–37, 2016. (Cited on pages 107 and 123.)

[Condat 2013] L. Condat. *A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms.* Journal of Optimization Theory and Applications, vol. 158, no. 2, pages 460–479, 2013. (Cited on pages 52, 79, 107, and 108.)

[Cooley 1965] J. W. Cooley and J. W. Tukey. *An algorithm for the machine calculation of complex Fourier series.* Mathematics of computation, vol. 19, no. 90, pages 297–301, 1965. (Cited on page 19.)

[Coste 2017] A. Coste. *Methodological Developments for Sodium, Phosphorus and Lithium MRI at high magnetic field: Applications to clinical research at 3 and 7 Tesla.* PhD thesis, University of Paris Saclay, 2017. (Cited on page 10.)

[Cotter 2005] S. F. Cotter, B. D. Rao, K. Engan and K. Kreutz-Delgado. *Sparse solutions to linear inverse problems with multiple measurement vectors.* IEEE Transactions on Signal Processing, vol. 53, no. 7, pages 2477–2488, 2005. (Cited on page 82.)

[Daubechies 2004] I. Daubechies, M. Defrise and C. De Mol. *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.* Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, vol. 57, no. 11, pages 1413–1457, 2004. (Cited on page 50.)

[Décorps 2012] M. Décorps. Imagerie de résonance magnétique. EDP Sciences, 2012. (Cited on page 18.)

[Delattre 2010] B. M. Delattre, R. M. Heidemann, L. A. Crowe, J.-P. Vallée and J.-N. Hyacinthe. *Spiral demystified.* Magnetic resonance imaging, vol. 28, no. 6, pages 862–881, 2010. (Cited on page 42.)

[Deng 2016] W. Deng and W. Yin. *On the global and linear convergence of the generalized alternating direction method of multipliers.* Journal of Scientific Computing, vol. 66, no. 3, pages 889–916, 2016. (Cited on page 99.)

[Dietz 2017] B. Dietz, E. Yip, J. Yun, B. G. Fallone and K. Wachowicz. *Real-time dynamic MR image reconstruction using compressed sensing and principal component analysis (CS-PCA): Demonstration in lung tumor tracking.* Medical physics, vol. 44, no. 8, pages 3978–3989, 2017. (Cited on page 105.)

[Doneva 2020] M. Doneva. *An overview of mathematical models for MRI reconstruction.* IEEE Signal Processing Magazine, page to appear, 2020. (Cited on page 34.)

[Donoho 2006] D. Donoho. *Compressed sensing.* IEEE Transactions on Information Theory, vol. 52, no. 4, pages 1289–1306, 2006. (Cited on pages 30, 35, 38, and 80.)

[Duarte 2011] M. F. Duarte and Y. C. Eldar. *Structured compressed sensing: From theory to applications.* IEEE Transactions on signal processing, vol. 59, no. 9, pages 4053–4085, 2011. (Cited on page 82.)

[Dutt 1993] A. Dutt and V. Rokhlin. *Fast Fourier transforms for nonequispaced data.* SIAM Journal on Scientific computing, vol. 14, no. 6, pages 1368–1393, 1993. (Cited on page 21.)

[Eckstein 1992] J. Eckstein and D. P. Bertsekas. *On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators.* Mathematical Programming, vol. 55, no. 1-3, pages 293–318, 1992. (Cited on page 52.)

[El Gueddari 2018a] L. El Gueddari, C. Lazarus, H. Carrié, A. Vignaud and P. Ciuciu. *Self-calibrating nonlinear MR image reconstruction algorithms for variable density sampling and parallel imaging.* In Proceedings of the 27th Annual Meeting of ISMRM, page 3547, Paris, France, 2018. (Cited on pages 58 and 153.)

[El Gueddari 2018b] L. El Gueddari, C. Lazarus, H. Carrié, A. Vignaud and P. Ciuciu. *Self-Calibrating Nonlinear Reconstruction Algorithms for Variable Density Sampling and Parallel Reception MRI.* In 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), pages 415–419, July 2018. (Cited on pages 58, 69, 70, 74, 114, and 152.)

[El Gueddari 2019a] L. El Gueddari, E. Chouzenoux, A. Vignaud, J.-C. Pesquet and P. Ciuciu. *Online MR image reconstruction for compressed sensing acquisition in T2* imaging.* In Wavelets: Applications in Signal and Image Processing XVIII. International Society for Optics and Photonics, 2019. (Cited on pages 103 and 152.)

[El Gueddari 2019b] L. El Gueddari, P. Ciuciu, E. Chouzenoux, A. Vignaud and J.-C. Pesquet. *Calibrationless OSCAR-based image reconstruction in compressed sensing parallel MRI.* In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 2019. IEEE ISBI. (Cited on pages 78, 114, 121, and 152.)

[El Gueddari 2019c] L. El Gueddari, P. Ciuciu, E. Chouzenoux, A. Vignaud and J.-C. Pesquet. *Online compressed sensing MR image reconstruction for high resolution T_2^* imaging.* In Proceedings of the 27th Annual Meeting of ISMRM, page 4679, Montreal, QC, Canada, 2019. (Cited on pages 103 and 152.)

[El Gueddari 2019d] L. El Gueddari, P. Ciuciu, E. Chouzenoux, A. Vignaud and J.-C. Pesquet. *OSCAR-based reconstruction for compressed sensing and parallel MR imaging.* In Proceedings of the 27th Annual Meeting of ISMRM, page 4766, Montreal, QC, Canada, 2019. (Cited on pages 77 and 152.)

[El Gueddari 2019e] L. El Gueddari, P. Ciuciu, E. Chouzenoux, A. Vignaud and J.-C. Pesquet. *Self-calibrating non-linear reconstruction algorithm for variable density sampling and parallel reception MRI.* In 2018 IEEE 15th International Symposium

on Biomedical Imaging (ISBI 2018), Washington D.C, USA, 2019. IEEE ISBI. (Cited on page 153.)

[El-Metwally 2008] S. M. El-Metwally, K. Z. Abd-Ehnoniem, A.-B. M. Youssef and Y. M. Kadah. *Rapid cardiac MRI using random radial trajectories.* In 2008 National Radio Science Conference, pages 1–7. IEEE, 2008. (Cited on page 42.)

[Elad 2007] M. Elad, P. Milanfar and R. Rubinstein. *Analysis versus synthesis in signal priors.* Inverse problems, vol. 23, no. 3, page 947, 2007. (Cited on pages 47, 79, and 106.)

[Eldar 2009a] Y. C. Eldar and M. Mishali. *Robust recovery of signals from a structured union of subspaces.* IEEE Transactions on Information Theory, vol. 55, no. 11, pages 5302–5316, 2009. (Cited on page 82.)

[Eldar 2009b] Y. C. Eldar and H. Rauhut. *Average case analysis of multichannel sparse recovery using convex relaxation.* IEEE Transactions on Information Theory, vol. 56, no. 1, pages 505–519, 2009. (Cited on page 82.)

[Feng 2016] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson and R. Otazo. *XD-GRASP: golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing.* Magnetic resonance in medicine, vol. 75, no. 2, pages 775–788, 2016. (Cited on page 60.)

[Fessler 2003] J. A. Fessler and B. P. Sutton. *Nonuniform fast Fourier transforms using min-max interpolation.* IEEE transactions on signal processing, vol. 51, no. 2, pages 560–574, 2003. (Cited on pages 19, 21, 92, 108, 110, and 116.)

[Fessler 2007] J. A. Fessler and D. C. Noll. *Iterative reconstruction methods for non-Cartesian MRI.* In Proc. ISMRM Workshop on Non-Cartesian MRI, 2007. (Cited on page 123.)

[Florescu 2014] A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu and S. Ciochina. *A majorize-minimize memory gradient method for complex-valued inverse problems.* Signal Processing, vol. 103, pages 285–295, Jan 2014. (Cited on pages 108 and 114.)

[Frigo 1998] M. Frigo and S. G. Johnson. *FFTW: An adaptive software architecture for the FFT.* In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181), volume 3, pages 1381–1384. IEEE, 1998. (Cited on page 22.)

[Gabay 1975] D. Gabay and B. Mercier. A dual algorithm for the solution of non linear variational problems via finite element approximation. Institut de recherche d'informatique et d'automatique, 1975. (Cited on page 52.)

[Ghadimi 2014] E. Ghadimi, A. Teixeira, I. Shames and M. Johansson. *Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic*

problems. IEEE Transactions on Automatic Control, vol. 60, no. 3, pages 644–658, 2014. (Cited on page 99.)

[Gkirtzou 2013] K. Gkirtzou, J. Honorio, D. Samaras, R. Goldstein and M. B. Blaschko. *fMRI analysis of cocaine addiction using k-support sparsity*. In 2013 IEEE 10th International Symposium on Biomedical Imaging, pages 1078–1081. IEEE, 2013. (Cited on page 87.)

[Greengard 2004] L. Greengard and J.-Y. Lee. *Accelerating the nonuniform fast Fourier transform*. SIAM review, vol. 46, no. 3, pages 443–454, 2004. (Cited on page 21.)

[Griswold 2002] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer and A. Haase. *Generalized autocalibrating partially parallel acquisitions (GRAPPA)*. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 47, no. 6, pages 1202–1210, 2002. (Cited on pages 28, 29, and 65.)

[Gruber 2018] B. Gruber, M. Froeling, T. Leiner and D. W. Klomp. *RF coils: A practical guide for nonphysicists*. Journal of Magnetic Resonance Imaging, vol. 48, no. 3, pages 590–604, 2018. (Cited on page 26.)

[Gudbjartsson 1995] H. Gudbjartsson and S. Patz. *The Rician distribution of noisy MRI data*. Magnetic resonance in medicine, vol. 34, no. 6, pages 910–914, 1995. (Cited on page 54.)

[Guerquin-Kern 2011] M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann and M. Unser. *A fast wavelet-based reconstruction method for magnetic resonance imaging*. IEEE transactions on medical imaging, vol. 30, no. 9, pages 1649–1660, 2011. (Cited on page 104.)

[Guo 2007] K. Guo and D. Labate. *Optimally sparse multidimensional representation using shearlets*. SIAM journal on mathematical analysis, vol. 39, no. 1, pages 298–318, 2007. (Cited on page 47.)

[Haacke 1999] E. M. Haacke, R. W. Brown, M. R. Thompson and R. Venkatesan. Magnetic Resonance Imaging - Physical Principles And Sequence Design. Wiley-Liss, 1999. (Cited on pages 9, 17, and 23.)

[Haase 2000] A. Haase, F. Odoj, M. Von Kienlin, J. Warnking, F. Fidler, A. Weisser, M. Nittka, E. Rommel, T. Lanz, B. Kalusche*et al.* *NMR probeheads for in vivo applications*. Concepts in Magnetic resonance, vol. 12, no. 6, pages 361–388, 2000. (Cited on page 24.)

[Hahn 1950] E. L. Hahn. *Nuclear induction due to free Larmor precession*. Physical Review, vol. 77, no. 2, page 297, 1950. (Cited on page 13.)

[Haldar 2014] J. Haldar. *Low-Rank Modeling of Local k-Space Neighborhoods (LORAKS) for Constrained MRI.* IEEE transactions on Medical Imaging, vol. 33, no. 3, pages 668–681, 2014. (Cited on page 114.)

[Haldar 2016] J. P. Haldar and J. Zhuo. *P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data.* Magnetic resonance in medicine, vol. 75, no. 4, pages 1499–1514, 2016. (Cited on pages 66, 69, 90, and 146.)

[Hansen 2013] M. S. Hansen and T. S. Sørensen. *Gadgetron: an open source framework for medical image reconstruction.* Magnetic resonance in medicine, vol. 69, no. 6, pages 1768–1776, 2013. (Cited on pages 104, 123, and 138.)

[Hastie 2005] T. Hastie, R. Tibshirani, J. Friedman and J. Franklin. *The elements of statistical learning: data mining, inference and prediction.* The Mathematical Intelligencer, vol. 27, no. 2, pages 83–85, 2005. (Cited on pages 102, 110, and 116.)

[Heinig 1992] G. Heinig and P. Jankowski. *Kernel structure of block Hankel and Toeplitz matrices and partial realization.* Linear algebra and its applications, vol. 175, pages 1–30, 1992. (Cited on page 61.)

[Hoerl 1970] A. E. Hoerl and R. W. Kennard. *Ridge regression: Biased estimation for nonorthogonal problems.* Technometrics, vol. 12, no. 1, pages 55–67, 1970. (Cited on page 81.)

[Hoult 1997] D. I. Hoult and B. Bhakar. *NMR signal reception: Virtual photons and coherent spontaneous emission.* Concepts in Magnetic Resonance: An Educational Journal, vol. 9, no. 5, pages 277–297, 1997. (Cited on page 13.)

[Jackson 1991] J. I. Jackson, C. H. Meyer, D. G. Nishimura and A. Macovski. *Selection of a convolution function for Fourier inversion using gridding (computerised tomography application).* IEEE transactions on medical imaging, vol. 10, no. 3, pages 473–478, 1991. (Cited on pages 20 and 43.)

[Jackson 1992] J. I. Jackson, D. G. Nishimura and A. Macovski. *Twisting radial lines with application to robust magnetic resonance imaging of irregular flow.* Magnetic resonance in medicine, vol. 25, no. 1, pages 128–139, 1992. (Cited on page 42.)

[Jacob 2009] L. Jacob, G. Obozinski and J.-P. Vert. *Group lasso with overlap and graph lasso.* In Proceedings of the 26th annual international conference on machine learning, pages 433–440. ACM, 2009. (Cited on page 82.)

[Jacob 2019] M. Jacob, L. El Gueddari, G. Navarro, M.-C. Cyrille, P. Bayle-Guillemaud, P. Ciuciu and Z. Saghi. *Statistical Machine Learning and Compressed Sensing Approaches for Analytical Electron Tomography - Application to Phase Change Materials.* Microscopy and Microanalysis, vol. 25, no. S2, page 156–157, 2019. (Cited on page 153.)

[Kamilov 2016] U. S. Kamilov and H. Mansour. *Learning optimal nonlinearities for iterative thresholding algorithms.* IEEE Signal Processing Letters, vol. 23, no. 5, pages 747–751, 2016. (Cited on pages 102 and 127.)

[Keiner 2009] J. Keiner, S. Kunis and D. Potts. *Using NFFT 3—a software library for various nonequispaced fast Fourier transforms.* ACM Transactions on Mathematical Software (TOMS), vol. 36, no. 4, page 19, 2009. (Cited on pages 19, 21, 60, 92, 108, and 140.)

[Kim 2017a] D. Kim and J. A. Fessler. *On the convergence analysis of the optimized gradient method.* Journal of optimization theory and applications, vol. 172, no. 1, pages 187–205, 2017. (Cited on page 53.)

[Kim 2017b] T. H. Kim, K. Setsompop and J. P. Haldar. *LORAKS makes better SENSE: Phase-constrained partial Fourier SENSE reconstruction without phase calibration.* Magnetic resonance in medicine, vol. 77, no. 3, pages 1021–1035, 2017. (Cited on pages 71 and 90.)

[Kim 2018a] D. Kim and J. A. Fessler. *Adaptive restart of the optimized gradient method for convex optimization.* Journal of Optimization Theory and Applications, vol. 178, no. 1, pages 240–263, 2018. (Cited on page 53.)

[Kim 2018b] T. Kim and J. Haldar. *LORAKS software version 2.0: faster implementation and enhanced capabilities.* Los Angeles, CA: University of Southern California, 2018. (Cited on page 69.)

[Knoll 2011] F. Knoll, K. Bredies, T. Pock and R. Stollberger. *Second order total generalized variation (TGV) for MRI.* Magnetic resonance in medicine, vol. 65, no. 2, pages 480–491, 2011. (Cited on page 63.)

[Knoll 2012] F. Knoll, C. Clason, K. Bredies, M. Uecker and R. Stollberger. *Parallel imaging with nonlinear reconstruction using variational penalties.* Magnetic resonance in medicine, vol. 67, no. 1, pages 34–41, 2012. (Cited on pages 63 and 64.)

[Komodakis 2015] N. Komodakis and J.-C. Pesquet. *Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems.* IEEE Signal Processing Magazine, vol. 32, no. 6, pages 31–54, 2015. (Cited on pages 51 and 52.)

[Lauterbur 1973] P. C. Lauterbur*et al.* *Image formation by induced local interactions: examples employing nuclear magnetic resonance.* Nature, vol. 242, no. 5394, pages 190–191, 1973. (Cited on pages 10, 14, 42, and 43.)

[Lazarus 2017] C. Lazarus, P. Weiss, N. Chauffert, F. Mauconduit, M. Bottlaender, A. Vignaud and P. Ciuciu. *SPARKLING: Novel non-Cartesian sampling schemes*

for accelerated 2D anatomical imaging at 7T using compressed sensing. In Proceedings of the 25th Annual Meeting of ISMRM, page 1529, Honolulu, HI, USA, 2017. (Cited on pages 39, 44, and 152.)

[Lazarus 2018a] C. Lazarus. *Compressed Sensing in MRI: optimization-based design of k-space filling curves for accelerated MRI.* PhD thesis, University Paris-Saclay, doctoral school: Electrical, Optical, Bio-Physics and Engineering, September 2018. (Cited on pages 38, 44, 45, 46, and 74.)

[Lazarus 2018b] C. Lazarus, P. Weiss, L. El Gueddari, F. Mauconduit, , A. Vignaud and P. Ciuciu. *Distribution-controlled and optimally spread non-Cartesian sampling curves for accelerated in vivo brain imaging at 7 Tesla.* In Proceedings of the 26th Annual Meeting of ISMRM, page 2666, Paris, France, 2018. (Cited on pages 33 and 153.)

[Lazarus 2018c] C. Lazarus, P. Weiss, A. Vignaud and P. Ciuciu. *An empirical study of the maximum degree of undersampling in compressed sensing for T2*-weighted MRI.* Magnetic resonance imaging, vol. 53, pages 112–122, 2018. (Cited on pages 58 and 59.)

[Lazarus 2019a] C. Lazarus, P. Weiss, N. Chauffert, F. Mauconduit, L. El Gueddari, C. Destrieux, I. Zemmoura, A. Vignaud and P. Ciuciu. *SPARKLING: variable-density k-space filling curves for accelerated T_2^*-weighted MRI.* Magnetic resonance in medicine, vol. 81, pages 3643–3661, 2019. (Cited on pages 33, 39, 43, 57, 60, 68, 90, 91, 92, 106, 109, 125, 126, 143, 150, and 151.)

[Lazarus 2019b] C. Lazarus, P. Weiss, L. El Gueddari, F. Mauconduit, C. Destrieux, I. Zemmoura, A. Vignaud and P. Ciuciu. *3D SPARKLING trajectories for high-resolution T_2^*-weighted Magnetic Resonance imaging.* NMR biomedicine, 2019. (Cited on pages 57, 123, 126, 150, and 151.)

[Le Ster 2017] C. Le Ster. *Exploration de la moelle osseuse en Imagerie par Résonance Magnetique: quantification de biomarqueurs.* PhD thesis, University of Bretagne-Loire, doctoral school: Life-Agronomy-Health, April 2017. (Cited on pages vii and 16.)

[Lee 2003] J. H. Lee, B. A. Hargreaves, B. S. Hu and D. G. Nishimura. *Fast 3D imaging using variable-density spiral trajectories with applications to limb perfusion.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 50, no. 6, pages 1276–1285, 2003. (Cited on pages 42, 46, 60, 90, and 91.)

[Lee 2016] D. Lee, K. H. Jin, E. Y. Kim, S.-H. Park and J. C. Ye. *Acceleration of MR parameter mapping using annihilating filter-based low rank hankel matrix (ALOHA).* Magnetic resonance in medicine, vol. 76, no. 6, pages 1848–1864, 2016. (Cited on page 114.)

[Levine 2017] E. Levine, B. Daniel, S. Vasanawala, B. Hargreaves and M. Saranathan. *3D Cartesian MRI with compressed sensing and variable view sharing using complementary poisson-disc sampling.* Magnetic resonance in medicine, vol. 77, no. 5, pages 1774–1785, 2017. (Cited on page 40.)

[Levitin 1966] E. S. Levitin and B. T. Polyak. *Constrained minimization methods.* USSR Computational mathematics and mathematical physics, vol. 6, no. 5, pages 1–50, 1966. (Cited on page 50.)

[Liang 2009] D. Liang, B. Liu, J. Wang and L. Ying. *Accelerating SENSE using compressed sensing.* Magnetic resonance in medicine, vol. 62, no. 6, pages 1574–1584, 2009. (Cited on pages 59 and 64.)

[Liang 2019] J. Liang and C.-B. Schönlieb. *Improving "Fast Iterative Shrinkage-Thresholding Algorithm": Faster, Smarter and Greedier.* Convergence, vol. 50, page 12, 2019. (Cited on pages 52 and 107.)

[Lin 2018] J.-M. Lin. *Python Non-Uniform Fast Fourier Transform (PyNUFFT): An Accelerated Non-Cartesian MRI Package on a Heterogeneous Platform (CPU/GPU).* Journal of Imaging, vol. 4, no. 3, page 51, 2018. (Cited on pages 110 and 116.)

[Lobos 2019] R. Lobos and J. P. Haldar. *Improving the Performance of Accelerated Image Reconstruction in K-Space: The Importance of Kernel Shape.* In Proceedings of the 27th Annual Meeting of ISMRM, page 2407, Montreal, QC, Canada, 2019. (Cited on page 66.)

[Lustig 2005] M. Lustig, J. H. Lee, D. L. Donoho and J. M. Pauly. *Faster imaging with randomly perturbed, under-sampled spirals and l1 reconstruction.* In Proceedings of the 13th annual meeting of ISMRM, Miami Beach, page 685. Citeseer, 2005. (Cited on page 42.)

[Lustig 2007] M. Lustig, D. Donoho and J. Pauly. *Sparse MRI: The application of compressed sensing for rapid MR imaging.* Magnetic Resonance in Medicine, vol. 58, no. 6, pages 1182–1195, 2007. (Cited on pages 38, 40, 64, 80, and 104.)

[Ma 2017] J. Ma. *Generalized sampling reconstruction from Fourier measurements using compactly supported shearlets.* Applied and Computational Harmonic Analysis, vol. 42, no. 2, pages 294–318, 2017. (Cited on pages 92 and 102.)

[Mair 2009] P. Mair, K. Hornik and J. de Leeuw. *Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods.* Journal of statistical software, vol. 32, no. 5, pages 1–24, 2009. (Cited on page 84.)

[Mairal 2010] J. Mairal, F. Bach, J. Ponce and G. Sapiro. *Online learning for matrix factorization and sparse coding.* Journal of Machine Learning Research, vol. 11, no. Jan, pages 19–60, 2010. (Cited on pages 48 and 80.)

[Majumdar 2012a] A. Majumdar and R. Ward. *Calibration-less multi-coil MR image reconstruction.* Magnetic Resonance Imaging, vol. 30, no. 7, pages 1032–1045, 2012. (Cited on pages 67, 78, 81, 86, 89, and 114.)

[Majumdar 2012b] A. Majumdar and R. K. Ward. *On the choice of Compressed Sensing priors and sparsifying transforms for MR image reconstruction: An experimental study.* Signal Processing: Image Communication, vol. 27, no. 9, pages 1035–1048, 2012. (Cited on page 47.)

[Majumdar 2012c] A. Majumdar, R. Ward and T. Aboulnasr. *Compressed sensing based real-time dynamic MRI reconstruction.* IEEE Transactions on Medical Imaging, vol. 31, no. 12, pages 2253–2266, 2012. (Cited on page 105.)

[Mansfield 1977a] P. Mansfield and A. A. Maudsley. *Medical imaging by NMR.* The British Journal of Radiology, vol. 50, no. 591, pages 188–194, March 1977. (Cited on pages 10 and 14.)

[Mansfield 1977b] P. Mansfield. *Multi-planar image formation using NMR spin echoes.* Journal of Physics C: Solid State Physics, vol. 10, no. 3, page L55, 1977. (Cited on pages 15 and 30.)

[Mardani 2018] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing and J. M. Pauly. *Deep generative adversarial neural networks for compressive sensing MRI.* IEEE transactions on medical imaging, vol. 38, no. 1, pages 167–179, 2018. (Cited on page 102.)

[Mason 2019] A. Mason, J. Rioux, A. Costa, M. Schmidt, V. Keought, T. Huynh and S. Beyea. *Are Root Mean Squared Error (RMSE) and Structural Similarity Index (SSIM) the most appropriate metrics for assessment of MR image quality?* In Proceedings of the 27th Annual Meeting of ISMRM, page 2465, Montreal, Canada, 2019. (Cited on pages 53 and 54.)

[McDonald 2016] A. M. McDonald, M. Pontil and D. Stamos. *New perspectives on k-support and cluster norms.* The Journal of Machine Learning Research, vol. 17, no. 1, pages 5376–5413, 2016. (Cited on pages 87 and 88.)

[McKenzie 2002] C. A. McKenzie, E. N. Yeh, M. A. Ohliger, M. D. Price and D. K. Sodickson. *Self-calibrating parallel imaging with automatic coil sensitivity extraction.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 47, no. 3, pages 529–538, 2002. (Cited on page 60.)

[McRobbie 2006] D. W. McRobbie, E. A. Moore, M. J. Graves and M. R. Prince. MRI from Picture to Proton. Cambridge university press, 2006. (Cited on pages 9 and 13.)

[Nesterov 1983] Y. E. Nesterov. *A method for solving the convex programming problem with convergence rate* $\mathcal{O}(1/k^2)$. In Soviet Math. Dokl., volume 269, pages 543–547, 1983. (Cited on page 50.)

[O'sullivan 1985] J. O'sullivan. *A fast sinc function gridding algorithm for Fourier inversion in computer tomography.* IEEE transactions on medical imaging, vol. 4, no. 4, pages 200–207, 1985. (Cited on page 20.)

[Otazo 2015] R. Otazo, E. Candes and D. K. Sodickson. *Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components.* Magnetic Resonance in Medicine, vol. 73, no. 3, pages 1125–1136, 2015. (Cited on page 105.)

[O'donoghue 2015] B. O'donoghue and E. Candes. *Adaptive restart for accelerated gradient schemes.* Foundations of computational mathematics, vol. 15, no. 3, pages 715–732, 2015. (Cited on pages 50 and 51.)

[Paolini 2015] M. Paolini, D. Keeser, M. Ingrisch, N. Werner, N. Kindermann, M. Reiser and J. Blautzik. *Resting-state networks in healthy adult subjects: a comparison between a 32-element and an 8-element phased array head coil at 3 Tesla.* Acta Radiologica, vol. 56, no. 5, pages 605–613, 2015. (Cited on page 24.)

[Parker 2014] D. L. Parker, A. Payne, N. Todd and J. R. Hadley. *Phase reconstruction from multiple coil data using a virtual reference coil.* Magnetic Resonance in Medicine, vol. 72, no. 2, pages 563–569, 2014. (Cited on pages 26, 27, 79, 101, and 148.)

[Pauly 2012] J. M. Pauly. *Gridding and the NUFFT for non-cartesian image reconstruction.* ISMRM Educational Course on Image Reconstruction, vol. 45, 2012. (Cited on page 19.)

[Pipe 1999] J. G. Pipe. *Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging.* Magnetic Resonance in Medicine, vol. 42, no. 5, pages 963–969, 1999. (Cited on page 60.)

[Pipe 2011] J. G. Pipe, N. R. Zwart, E. A. Aboussouan, R. K. Robison, A. Devaraj and K. O. Johnson. *A new design and rationale for 3D orthogonally oversampled k-space trajectories.* Magnetic resonance in medicine, vol. 66, no. 5, pages 1303–1311, 2011. (Cited on page 42.)

[Poole 2007] M. Poole. *Improved equipment and techniques for dynamic shimming in high field MRI.* PhD thesis, University of Nottingham, 2007. (Cited on page 15.)

[Press 2007] W. Press, S. Teukolsky, W. Vetterling and B. Flannery. Numerical recipes 3rd edition: The art of scientific computing. Cambridge University Press, New York, NY, USA, 3 édition, 2007. (Cited on page 108.)

[Pruessmann 1999] K. Pruessmann, P. Weiger, M. Scheidegger and P. Boesiger. *SENSE: sensitivity encoding for fast MRI.* Magnetic Resonance in Medicine, vol. 42, no. 5, pages 952–962, 1999. (Cited on pages 28, 29, and 150.)

[Purcell 1946] E. M. Purcell, H. C. Torrey and R. V. Pound. *Resonance absorption by nuclear magnetic moments in a solid.* Physical review, vol. 69, no. 1-2, page 37, 1946. (Cited on page 10.)

[Pustelnik 1999] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng and J.-C. Pesquet. *Wavelet-Based Image Deconvolution and Reconstruction.* Wiley Encyclopedia of Electrical and Electronics Engineering, pages 1–34, 1999. (Cited on page 106.)

[Rabi 1938] I. I. Rabi, J. R. Zacharias, S. Millman and P. Kusch. *A new method of measuring nuclear magnetic moment.* Physical Review, vol. 53, no. 4, page 318, 1938. (Cited on page 10.)

[Ramzi 2019] Z. Ramzi, P. Ciuciu and J.-L. Starck. *Benchmarking proximal methods acceleration enhancements for CS-acquired MR image analysis reconstruction.* In SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations Workshop, Toulouse, France, July 2019. (Cited on pages 52 and 53.)

[Rauhut 2010] H. Rauhut. *Compressive sensing and structured random matrices.* Theoretical foundations and numerical methods for sparse recovery, vol. 9, pages 1–92, 2010. (Cited on page 35.)

[Ravishankar 2010] S. Ravishankar and Y. Bresler. *MR image reconstruction from highly undersampled k-space data by dictionary learning.* IEEE transactions on medical imaging, vol. 30, no. 5, pages 1028–1041, 2010. (Cited on pages 47 and 48.)

[Ravishankar 2015] S. Ravishankar and Y. Bresler. *Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging.* SIAM Journal on Imaging Sciences, vol. 8, no. 4, pages 2519–2557, 2015. (Cited on pages 47 and 48.)

[Rockafellar 1970] R. T. Rockafellar. Convex analysis, volume 28. Princeton university press, 1970. (Cited on page 49.)

[Roemer 1990] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza and O. M. Mueller. *The NMR phased array.* Magnetic Resonance in Medicine, vol. 16, no. 2, pages 192–225, 1990. (Cited on pages 26, 27, 58, and 79.)

[Rudin 1992] L. I. Rudin, S. Osher and E. Fatemi. *Nonlinear total variation based noise removal algorithms.* Physica D: nonlinear phenomena, vol. 60, no. 1-4, pages 259–268, 1992. (Cited on page 63.)

[Seiberlich 2007] N. Seiberlich, F. A. Breuer, M. Blaimer, K. Barkauskas, P. Jakob and M. Griswold. *Non-Cartesian data reconstruction using GRAPPA operator gridding*

(GROG). Magnetic Resonance in Medicine, vol. 58, no. 6, pages 1257–1265, 2007. (Cited on pages 66 and 114.)

[Selesnick 2009] I. W. Selesnick and M. A. Figueiredo. *Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors.* In Wavelets XIII, volume 7446, page 74460D. International Society for Optics and Photonics, 2009. (Cited on page 47.)

[She 2014] H. She, R. Chen, D. Liang, E. DiBella and L. Ying. *Sparse BLIP: BLind Iterative Parallel imaging reconstruction using compressed sensing.* Magnetic Resonance in Medicine, vol. 71, no. 2, pages 645–660, 2014. (Cited on pages 64 and 114.)

[Shin 2014] P. J. Shin, P. E. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron and M. Lustig. *Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion.* Magnetic resonance in medicine, vol. 72, no. 4, pages 959–970, 2014. (Cited on pages 64, 66, and 114.)

[Sobolev 1950] S. L. Sobolev. Applications of functional analysis in mathematical physics. Leningrad University, first édition, 1950. (Cited on page 63.)

[Spuentrup 2004] E. Spuentrup, M. Katoh, A. Buecker, W. J. Manning, T. Schaeffter, T.-H. Nguyen, H. P. Kühl, M. Stuber, R. M. Botnar and R. W. Günther. *Free-breathing 3D steady-state free precession coronary MR angiography with radial k-space sampling: comparison with cartesian k-space sampling and cartesian gradient-echo coronary MR angiography—pilot study.* Radiology, vol. 231, no. 2, pages 581–586, 2004. (Cited on page 60.)

[Starck 2002] J.-L. Starck, E. J. Candès and D. L. Donoho. *The curvelet transform for image denoising.* IEEE Transactions on image processing, vol. 11, no. 6, pages 670–684, 2002. (Cited on page 47.)

[Starck 2007] J.-L. Starck, J. Fadili and F. Murtagh. *The undecimated wavelet decomposition and its reconstruction.* IEEE Transactions on Image Processing, vol. 16, no. 2, pages 297–309, 2007. (Cited on pages 47 and 140.)

[Stein 1981] C. M. Stein. *Estimation of the mean of a multivariate normal distribution.* The annals of Statistics, pages 1135–1151, 1981. (Cited on pages 102, 110, and 116.)

[Strohmeier 2016] D. Strohmeier, Y. Bekhti, J. Haueisen and A. Gramfort. *The iterative reweighted mixed-norm estimate for spatio-temporal MEG/EEG source reconstruction.* IEEE transactions on medical imaging, vol. 35, no. 10, pages 2218–2228, 2016. (Cited on page 82.)

[Taylor 2017a] A. B. Taylor, J. M. Hendrickx and F. Glineur. *Exact worst-case performance of first-order methods for composite convex optimization.* SIAM Journal on Optimization, vol. 27, no. 3, pages 1283–1313, 2017. (Cited on pages 53 and 107.)

[Taylor 2017b] J. R. Taylor, N. Williams, R. Cusack, T. Auer, M. A. Shafto, M. Dixon, L. K. Tyler, R. N. Henson *et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample.* Neuroimage, vol. 144, pages 262–269, 2017. (Cited on pages viii and 48.)

[Tibshirani 1996] R. Tibshirani. *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pages 267–288, 1996. (Cited on page 80.)

[Tikhonov 1943] A. N. Tikhonov. *On the stability of inverse problems.* In Dokl. Akad. Nauk SSSR, volume 39, pages 195–198, 1943. (Cited on pages 63 and 81.)

[Tikhonov 1963] A. N. Tikhonov. *On the solution of ill-posed problems and the method of regularization.* In Doklady Akademii Nauk, volume 151, pages 501–504. Russian Academy of Sciences, 1963. (Cited on page 63.)

[Tomi-Tricot 2018] R. Tomi-Tricot. *Clinical Applications of the K_T-points Method to Homogenise Spin Excitation in* 3 T *MRI.* PhD thesis, University of Paris Saclay, 2018. (Cited on pages 11 and 13.)

[Trzasko 2011] J. D. Trzasko and A. Manduca. *Calibrationless parallel MRI using CLEAR.* In 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pages 75–79, Nov 2011. (Cited on pages 67, 69, and 114.)

[Uecker 2008] M. Uecker, T. Hohage, K. T. Block and J. Frahm. *Image reconstruction by regularized nonlinear inversion—joint estimation of coil sensitivities and image content.* Magnetic Resonance in Medicine, vol. 60, no. 3, pages 674–682, 2008. (Cited on pages 63 and 64.)

[Uecker 2014] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala and M. Lustig. *ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA.* Magnetic resonance in medicine, vol. 71, no. 3, pages 990–1001, 2014. (Cited on pages 61, 69, 74, 90, and 114.)

[Uecker 2015] M. Uecker, F. Ong, J. I. Tamir, D. Bahri, P. Virtue, J. Y. Cheng, T. Zhang and M. Lustig. *Berkeley advanced reconstruction toolbox.* In Proc. Intl. Soc. Mag. Reson. Med, volume 23, page 2486, 2015. (Cited on page 69.)

[Vannesjo 2013] S. J. Vannesjo, M. Haeberlin, L. Kasper, M. Pavan, B. J. Wilm, C. Barmet and K. P. Pruessmann. *Gradient system characterization by impulse response measurements with a dynamic field camera.* Magnetic resonance in medicine, vol. 69, no. 2, pages 583–593, 2013. (Cited on page 39.)

[Vannesjo 2017] S. J. Vannesjo, Y. Duerst, L. Vionnet, B. E. Dietrich, M. Pavan, S. Gross, C. Barmet and K. P. Pruessmann. *Gradient and shim pre-emphasis by inversion*

of a linear time-invariant system model. Magnetic resonance in medicine, vol. 78, no. 4, pages 1607–1622, 2017. (Cited on page 39.)

[Vasanawala 2011] S. Vasanawala, M. Murphy, M. T. Alley, P. Lai, K. Keutzer, J. M. Pauly and M. Lustig. *Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients.* In 2011 ieee international symposium on biomedical imaging: From nano to macro, pages 1039–1043. IEEE, 2011. (Cited on page 40.)

[Vũ 2013] B. Vũ. *A splitting algorithm for dual monotone inclusions involving cocoercive operators.* Advances in Computational Mathematics, vol. 38, no. 3, pages 667–681, Apr 2013. (Cited on pages 52, 79, and 107.)

[Wang 2004] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al. Image quality assessment: from error visibility to structural similarity.* IEEE transactions on image processing, vol. 13, no. 4, pages 600–612, 2004. (Cited on pages 69, 92, 110, and 115.)

[Wen 2016] W. Wen, C. Wu, Y. Wang, Y. Chen and H. Li. *Learning structured sparsity in deep neural networks.* In Advances in neural information processing systems, pages 2074–2082, 2016. (Cited on page 102.)

[Wen 2019] B. Wen, S. Ravishankar, L. Pfister and Y. Bresler. *Transform Learning for Magnetic Resonance Image Reconstruction: From Model-based Learning to Building Neural Networks.* arXiv preprint arXiv:1903.11431, 2019. (Cited on page 47.)

[Wiggins 2009] G. C. Wiggins, J. R. Polimeni, A. Potthast, M. Schmitt, V. Alagappan and L. L. Wald. *96-Channel receive-only head coil for 3 Tesla: design optimization and evaluation.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 62, no. 3, pages 754–762, 2009. (Cited on page 25.)

[Winkelmann 2006] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers and O. Doessel. *An optimal radial profile order based on the Golden Ratio for time-resolved MRI.* IEEE transactions on medical imaging, vol. 26, no. 1, pages 68–76, 2006. (Cited on page 110.)

[Yang 2016] Y. Yang, J. Sun, H. Li and Z. Xu. *Deep ADMM-Net for compressive sensing MRI.* In Advances in neural information processing systems, pages 10–18, 2016. (Cited on pages 102 and 127.)

[Yeh 2005] E. N. Yeh, M. Stuber, C. A. McKenzie, R. M. Botnar, T. Leiner, M. A. Ohliger, A. K. Grant, J. D. Willig-Onwuachi and D. K. Sodickson. *Inherently self-calibrating non-Cartesian parallel imaging.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 54, no. 1, pages 1–8, 2005. (Cited on page 60.)

[Ying 2007] L. Ying and J. Sheng. *Joint image reconstruction and sensitivity estimation in SENSE (JSENSE).* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 57, no. 6, pages 1196–1202, 2007. (Cited on page 64.)

[Yuan 2006] M. Yuan and Y. Lin. *Model selection and estimation in regression with grouped variables.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 68, no. 1, pages 49–67, 2006. (Cited on pages 80 and 81.)

[Zanforlin 2014] M. Zanforlin, D. Munaretto, A. Zanella and M. Zorzi. *SSIM-based video admission control and resource allocation algorithms.* In 2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), pages 656–661. IEEE, 2014. (Cited on page 53.)

[Zbontar 2018] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana*et al. fastmri: An open dataset and benchmarks for accelerated mri.* arXiv preprint arXiv:1811.08839, 2018. (Cited on pages 102 and 127.)

[Zeng 2014a] X. Zeng and M. Figueiredo. *The ordered weighted l1 norm: Atomic formulation, dual norm, and projections.* arXiv preprint ArXiv:1409.4271, 2014. (Cited on pages 83, 84, 115, and 133.)

[Zeng 2014b] X. Zeng and M. A. Figueiredo. *Decreasing Weighted Sorted ℓ_1 Regularization.* IEEE Signal Processing Letters, vol. 21, no. 10, pages 1240–1244, 2014. (Cited on page 83.)

[Zhang 2011] J. Zhang, C. Liu and M. E. Moseley. *Parallel reconstruction using null operations.* Magnetic resonance in medicine, vol. 66, no. 5, pages 1241–1253, 2011. (Cited on page 65.)

[Zhao 2006] P. Zhao and B. Yu. *On model selection consistency of Lasso.* Journal of Machine learning research, vol. 7, no. Nov, pages 2541–2563, 2006. (Cited on page 80.)

[Zhu 2018] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen and M. S. Rosen. *Image reconstruction by domain-transform manifold learning.* Nature, vol. 555, no. 7697, page 487, 2018. (Cited on page 102.)

[Zou 2005] H. Zou and T. Hastie. *Regularization and variable selection via the elastic net.* Journal of the royal statistical society: series B (statistical methodology), vol. 67, no. 2, pages 301–320, 2005. (Cited on pages 80 and 81.)

université PARIS-SACLAY
ÉCOLE DOCTORALE
Physique et ingénierie : électrons, photons, sciences du vivant (EOBE)

Titre : Algorithmes de structures parcimonieuses pour la reconstruction en-ligne d'image haute résolution en IRM.

Mots clés : Echantillonnage compressif, reconstruction en-ligne, parcimonie structurée, IRM, acquisition multi-canal

Résumé :
L'imagerie par résonance magnétique (IRM) est la technique d'imagerie médicale de référence pour sonder in vivo et non invasivement les tissus mous du corps humain, en particulier le cerveau. L'amélioration de la résolution de l'IRM en un temps d'acquisition standard (*e.g.*400μm isotrope en 15 minutes) permettrait aux médecins d'améliorer considérablement leur diagnostic et le suivi des patients. Cependant, le temps d'acquisition en IRM reste long. Pour réduire ce temps, la récente théorie de l'échantillonnage comprimée (EC) a révolutionné la façon d'acquérir des données dans plusieurs domaines dont l'IRM en surmontant le théorème de Shannon-Nyquist. Avec l'EC, les données peuvent alors être massivement sous-échantillonnées tout en assurant des conditions optimales de reconstruction des images.
Dans ce contexte, les thèses de doctorat précédemment soutenue au sein du laboratoire ont été consacrées à la conception et à la mise en œuvre de scénarios d'acquisition physiquement plausibles pour accélérer l'acquisitions. Un nouvel algorithme d'optimisation pour la conception de trajectoire non cartésienne avancée appelée SPARKLING pour *Spreading Projection Algorithm for Rapid K-space sampIING* en est né. Les trajectoires SPARKLING générées ont conduit à des facteurs d'accélération allant jusqu'à 20 en 2D et 70 pour les acquisitions 3D sur des images à haute résolution pondérées en T_2^* acquises à 7 Tesla. Ces accélérations n'étaient accessibles que grâce au rapport signal/bruit d'entrée élevé fourni par l'utilisation de bobines de réception multi-canaux (IRMp). Cependant, ces résultats ont été obtenus au détriment d'une reconstruction longue et complexe. Dans cette thèse, l'objectif est de proposer une nouvelle approche de reconstruction en-ligne d'images acquises par IRMp non Cartésienne. Pour atteindre cet objectif, nous nous appuyons sur une approche en ligne où reconstruction et acquisition s'entremèlent. Par conséquent, la reconstruction débute avant la fin de l'acquisition et un résultat partiel est délivré au cours de l'examen. L'ensemble du pipeline est compatible avec une implémentation réelle à travers l'interface Gadgetron pour produire les images reconstruites à la console du scanner.
Ainsi, après avoir exposé la théorie de l'échantillonage comprimé, nous présentons l'état de l'art de la méthode dédiée à la reconstruction en imagerie multi-canaux. En particulier, nous nous concentrerons d'abord sur les méthodes d'autocalibration qui présentent l'avantage d'être adaptées à l'échantillonnage non cartésien et nous proposons une méthode simple mais efficace pour estimer le profil de sensibilité des différents cannaux. Cependant, en raison de leur dépendance au profil de sensibilité, ces méthodes ne sont pas adapatable à la reconstruction en ligne. Par conséquent, la deuxième partie se concentre sur la suppression des ces profils et celà grâce à l'utilisation de normes mixtes promouvant une parcimonie structurée. Ensuite, nous adaptons différentes régularizations basées sur la parcimonie structurée pour reconstruire ces images fortement corrélées. Enfin, la méthode retenue sera appliquée à l'imagerie en ligne.

Title : Proximal structured sparsity regularization for online reconstruction in high-resolution accelerated MRI

Keywords : Compressed Sensing ; online reconstruction ; structured sparsity ; MRI ; multi-channel acquisition

Abstract : Magnetic resonance imaging (MRI) is the reference medical imaging technique for probing in vivo and non-invasively soft tissues in the human body, notably the brain. MR image resolution improvement in a standard scanning time (e.g., 400μm isotropic in 15 min) would allow medical doctors to significantly improve both their diagnosis and patients' follow-up. However the scanning time in MRI remains long, especially in the high resolution context. To reduce this time, the recent Compressed Sensing (CS) theory has revolutionized the way of acquiring data in several fields including MRI by overcoming the Shannon-Nyquist theorem. Using CS, data can then be massively under-sampled while ensuring conditions for optimal image recovery.
In this context, previous Ph.D. thesis in the laboratory were dedicated to the design and the implementation of physically plausible acquisition scenarios to accelerate the scan. Those projects delivered a new optimization algorithm for the design of advanced non-Cartesian trajectory called Spreading Projection Algorithm for Rapid K-space sampling (SPARKLING). The generated trajectories led to acceleration factor up to 20 in 2D and 60 for 3D-acquisitions on highly resolved T_2^* weighted images acquired at 7 Tesla. Those accelerations were only accessible thanks to the high input Signal-to-Noise Ratio delivered by the usage of multi-channel reception coils. However, such high under-sampling factors were coming at a price of long and complex reconstruction. In this thesis the objective is to propose an online approach for non-Cartesian multi-channel MR image reconstruction. To achieve this goal we rely on a progressive online approach where the reconstruction starts from incomplete data. Hence acquisition and reconstruction are interleaved, and a partial feedback is given during the scan. After exposing the CS theory, we present state-of the art methods dedicated to multi-channel coil reconstruction. In particular we first focus on self-calibrating methods that present the advantage to be adapted to non-Cartesian sampling and we propose a simple yet efficient method to estimate the coil sensitivity profiles based on the data itself, i.e. without any pre-calibration scan. However, owing to its dependence to user-defined parameters, this two-step approach (extraction of sensitivity maps and then image reconstruction) is not compatible with the timing constraints associated with online reconstruction. Then we study the case of calibration-free reconstruction methods and splits them into two categories, the k-space based and the domain-based. While the k-space calibration-free methods are sub-optimal for non-Cartesian reconstruction, due to the gridding procedure, we retain the domain-based reconstruction and prove their interests and performances for online purposes. Hence in the second part, we first prove the advantage of mixed norm to improve the recovery guarantee in the parallel MRI setting. Then we investigate the impact of structured sparse induced norm on the reconstruction multi-channel purposes, where we adapt different penalty to handle those highly correlated images. Finally, the retained method is applied to online purposes. The entire pipeline, is compatible with an implementation through the Gadgetron pipeline to deliver the reconstruction at the scanner console.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France