Faculteit Wetenschappen
Departement Fysica

Improved MRI Relaxometry through Statistical Signal Processing

---

Verbeterde MRI Relaxometrie via Statistische Signaalverwerking

Proefschrift voorgelegd tot het behalen van de graad van

Doctor in de Wetenschappen: Fysica

aan de Universiteit Antwerpen, te verdedigen door

Gabriel RAMOS LLORDÉN

Promotoren:
Prof. Dr. Jan Sijbers
Prof. Dr. Ir. Arnold J. den Dekker

Antwerpen, 2018

Doctoral jury:

Prof. Dr. Floris Wuyts
Prof. Dr. Jan Sijbers
Prof. Dr. Ir. Arnold J. den Dekker
Prof. Dr. Marleen Verhoye

Prof. Dr. Santiago Aja Fernández
Prof. Dr. Aleksandra Pizurica
Dr. Ir. Dirk H. J. Poot

Contact information:

✉ **Gabriel Ramos Llordén**
imec - Vision Lab, Dept. of Physics
University of Antwerp (CDE)
Universiteitsplein 1, Building N1.15
B-2610 Wilrijk, Antwerpen, Belgium
☎ +32 3 265 24 45
✉ gabriel.Ramos-llorden@uantwerpen.be
🌍 http://visielab.uantwerpen.be/people/gabriel-ramos-llorden

“Life can only be understood backwards; but it must be lived forwards.”

SØREN KIERKEGAARD

Acknowledgments

Kierkegaard was right. If someone had asked me to predict how these four years and five months would look like, I would have preferred to square the circle. However, looking back now, it seems (almost) everything was part of a puzzle that looks complete.

I arrived at Antwerp on the twenty-ninth day of that August of 2013. The first week was already special. I was invited by my, at that moment, future supervisor, professor Jan Sijbers, to participate in a small workshop in medical imaging. It took place in a very beautiful building: Klooster van de Grauwzusters. I really enjoyed that scientific course. I met some of my colleagues of Vision Lab, and of course, I met Jan. I am truly honest if I say that such a week boosted my enthusiasm for the period of my life that was about to start. It is not a surprise then that I have chosen the same location to defend publically my PhD thesis. I like to be sentimental. The circle is closed.

The week after that 29th of August, I arrived at Vision Lab, where I was introduced to my colleagues, and I made my first contact to what "PhD research" signifies. I am quite happy to have performed my PhD studies at Vision Lab, and of course, I am in debt to my two supervisors, Jan, and also professor Arnold Jan den Dekker. By now, I consider Belgium my professional home. I really think I have become a true scientist here. I was given enough freedom to pursue my ambitions in a very proactive environment, where a huge amount of resources was available for our research. Of course, this does not fall from the sky magically but has to do with the people who supervised my research work.

I firmly believe that a PhD cannot be completed with full support, and this, Jan and Arjan, has come from you in all possible ways. I am not going to remark your scientific qualities since it is known by all people how outstanding they are. But yes, I would like to stress your empathy skills, which definitely made a difference when I had to hold on in the most difficult moments. Your infinite patience and comprehension have been essential ingredients for the success of this thesis. Perseverance, self-improvement, high standards pursuing, and most importantly, modesty and enthusiasm, are not just unmeaning words but elements of constant lessons during these four years, and that now have become part of me :)

I am very grateful to the jury members of my thesis: professors Floris Wuyts, Marleen Verhoye, Santiago Aja Fernández and Aleksandra Pizurica, as well as doctor Dirk H. J. Poot. This dissertation has substantially benefited from your constructive feedback and suggestions. Especial acknowledgments go to Marleen

for, since the very beginning of my PhD, being an important guidance in the applications of relaxometry in animal neuroimaging.

In these four years, there are a lot of to people at Vision Lab that I like to acknowledge. I am aware that in these three pages I am leaving out names, but you surely know that I've been very fortunate to have you as colleagues. I will not, however, omit people of the MRI group, those who are continuing his/her careers out of vision lab, and those who gather every week at the MRI meetings (last time at 9.00! :p). This list includes Jelle, Timo, Quinten, Ben, Gwendolyn, Quinten Beirinckx, Annelinde, Jan, Piet, and Diana. We have also shared very good moments around the world! There is always more to do at ISMRM conferences than presenting cutting-edge research. I still laugh a lot when I remember that canoeing trip with Ben and Quinten, in Canada. I knew from that trip that I would never win the world championship of canoeing. And of course, I will never forget our friends the "killers" mosquitoes.

One year later, we all again traveled to Singapore. Horrible (hot) weather but very great nightlife. Gwendolyn and Ben, probably you still remember that "karaoke" that we ended up with. If Toronto and Singapore were not enough, Hawaii was the third scientific destination. Hawaii always reminds me of Diana and Piet. Guys, we had very good times there! As a sign of gratefulness, I left my face on a window for the posterity! By the way, I did not tell you, but I have been contacted to promote travel trips to Pearl Harbor ;)

Really, Piet and Diana, I consider you my best friends at Vision Lab. You have made our office the coolest (sorry for the other offices). Piet, you are one of the best researchers I have collaborated with. Diana, you brought joyfulness and happiness to this office, and why not, to vision lab as well. Please both, do not change :)

My research would not have been the same without the help of fantastic collaborators. Thanks to people from BioImaging Lab: Marleen, Johan van Audekerke,... for helping me with the acquisition and also spending hours collecting data to test if my methods/algorithms have sense. Very thanks to the UZA as well, where we have acquired most of our human MR images. Thanks to prof. Paul Parizel and specially Floris Vanhevel, for staying on late-night shifts to fulfill my MRI acquisition wishlist. Collaborators around the world have been fundamental for my career as a scientist too. I want to give thanks to dr. Marcus Björk, from Uppsala Universiteit, for very fruitful discussions, advice and recommendations for NOVIFAST.

A milestone in my career was my three months living in Boston. My research stay at Harvard clearly shaped my perception of science, and convinced me that yes... I love this job. I would like to give thanks to prof. Carl-Fredrik Westin and prof. Yogesh Rathi for hosting me and introducing me in the world of diffusion MRI. I am really honored that such brilliant scientists found time in their busy schedules to help me during my research project abroad. I hope we can work together in the future! I also want to give thanks to prof. Raúl San José Estépar for managing all the bureaucratic procedures, offering me all the support, and plainly, making

me feel at home. Of course, I do not forget dr. Gonzalo Vegas Sánchez-Ferrero, a classic in all my theses' acknowledgments. Gonzalo, you have been included in all my three theses dissertations. That is a record, and you know it :) You always have advised me well, not only in career terms but also on a personal level. I am pretty sure you will still do so in the future. I do not forget the Spanish/Italian crew in Boston. Celtics, Bruins, Super Bowl, and of course, Cambridge nightlife. How many Samuel Adams did we enjoy?

Most of my conference travels, as well as my research stay at Harvard, were funded by competitive awards I received from the FWO. I am thankful to have been supported by this research agency.

There is always much more to life than work, and friends are one of the most important things. If I dedicated my MSc thesis to my friends from Spain, this time is the turn for my friends in Antwerp. I want to acknowledge all the people whom I have shared very great moments with, and I hope I can still do so. That PlanAntwerp group that saved Fridays where there were no plans at all! I would like to specially thanks those who have accompanied me from the very beginning of my PhD: Kike, Silvia, Juanjo, Mar, Anna, Eu, Victor, Patri, Irene,... It is not new that the first three have been probably my closest, and also those who have suffered the most from my frustrations and complaints :p. But also they have given their help at very crucial moments, especially the first two (I will visit you soon in Spain :p) It is a pity that some things have been turned around, and look difficult to overcome right now. Any case, I hope you enjoyed my presence, at least a bit. I have the fame of being hmm... let's say just a bit "insistent", but I think I am a cool guy, at least that is what people say ;) I hope our lives will allow us to meet each other again in any part of the world we finally end up.

I leave the best for the last part, and because the situation calls for it, I will turn to the language of Cervantes. Me gustaría dar las gracias a las personas más importantes de mi vida, que obviamente son mis padres y mi hermana. Incluso desde más de 1500 km, vuestro aliento, apoyo, y cariño se siente como si todavía estuviera viviendo en casa. Por supuesto, esta tesis no habría sido posible sin vosotros, como tantas buenas cosas que me han pasado y seguirán pasando en esta aventura llamada vida. Si quedaba alguna duda, esta tesis está dedicada a vosotros. Os quiero!

Really, thank you all, dank u wel, gracias,...

Gabriel Ramos Llordén,
Antwerpen,
February 2018

Contents

Summary

This PhD thesis tries to reduce the obstacles that preclude MRI relaxometry from being a fast, accurate, and precise quantitative MRI modality for clinical use, by improving the way MR relaxometry data are acquired, processed, and analyzed. The manuscript is divided into three parts. Part I covers the basics of quantitative MRI from a signal processing point of view, whereas part II presents a short introduction of MRI relaxometry as well as of statistical parameter estimation theory. The contributions of this dissertation are included in part III.

Quantitative MRI: a model-based imaging paradigm

Magnetic Resonance Imaging (MRI) is a versatile, non-invasive imaging modality that provides excellent soft tissue contrast without using ionizing radiation. Consequently, MRI is widely employed in clinics for medical diagnosis, staging, and follow-up of diseases. MRI is generally used in a qualitative way, with a radiologist interpreting the MR image, aiming to identify biomarkers that may be useful for monitoring the state of diseases. However, current radiology requires image biomarkers to be reliable, reproducible, and objective. Qualitative MRI often fails to meet these demands.

In MRI, the image intensity is related to quantitative parameters that are intrinsic descriptors of the biological state of tissues. The relation between the intensity and those parameters is given by a physical model. Such a physical model establishes a link between the "pictures" that the MR scanner takes, and the physical reality inside the tissues. This way, MR images are not merely considered as pictures, but as carriers of unique, tissue-specific information that can be measured, analyzed, and interpreted rigorously. This model-based imaging paradigm has received the name quantitative MRI (qMRI). Quantitative parameter maps can be estimated from MRI images, often providing more informative, controlled, and objective biomarkers than those obtained with qualitative MRI.

The reader can use the first two chapters of this thesis as an introduction to the basics of qMRI. Indeed, chapter 1 introduces the reader to the MRI field. Starting from the very basic physics that govern the spin dynamics, the reader is guided through the theory behind the MR signal generation and its localization

in space. Chapter 2 describes how MR images are reconstructed using signal processing algorithms.

MRI Relaxometry

A very popular modality in qMRI is relaxometry. MRI relaxometry deals with the estimation of the spin-lattice, T_1, and the spin-spin, T_2, relaxation times. Both relaxation times are fundamental parameters that describe the spin dynamics within a tissue during the relaxation process of the Nuclear Magnetic Resonance (NMR) phenomenon. During the last decades, spatial T_1 and T_2 maps have been analyzed to study and monitor the states of a multitude of diseases, to name a few: multiple sclerosis, Alzheimer's disease, dementia, schizophrenia or myocardial fibrosis. Those studies have shown that MRI relaxometry holds the promise of generating robust, objective image-based biomarkers for central nervous system pathologies, cardiovascular diseases and beyond.

However, though naturally preferable to those obtained from qualitative MRI, biomarkers that are obtained from MRI relaxometry are not yet sufficiently specific, sensitive, and robust to be routinely used in clinical practice. On top of that, high-resolution relaxation maps demand a clinically unfeasible long scanning time.

By adopting a quantitative approach, familiar terms such as accuracy and precision can be related to specificity and sensitivity, whereas scanning time is fundamentally linked to the data acquisition time. Terminology such as *accuracy, precision, and acquisition time* is essential for the scientific task of "measuring", a process which has two major components: data acquisition and data analysis. The accuracy and precision of quantitative MRI relaxometry are greatly influenced by errors that originate from inaccurate physical modeling, but also from inadequate signal processing or naïve parameter estimation techniques (data analysis). Besides, the acquisition of MR relaxometry data is time-consuming since a large number of data points must be sampled so as to reconstruct a set of high-quality MRI images needed to estimate T_1 and T_2 maps with a high accuracy and precision. The interested reader can use chapter 3 as a brief introduction to the basics of MRI relaxometry as well as the main challenges that emanate from this technique. Furthermore, a brief introduction to statistical parameter estimation theory, which is a must in modern qMRI, is presented in chapter 4.

Contributions

This PhD thesis collects three contributions that address current problems in MRI relaxometry, aiming at turning this modality into a robust, qMRI technique that generates biomarkers in a clinically feasible time. To that end, we adopt an approach based on a model-based signal processing perspective, with an important emphasis on statistical parameter estimation. Effort has been made into developing

model-based algorithms that are both theoretically grounded and efficient to be implemented in clinical MR scanner software.

Chapter 5 deals with the problem of accurate and precise T_1 map estimation in the presence of patient motion. Conventionally, motion is accounted for prior to T_1 map estimation, by spatially registering the dataset of MR images with general-purpose image registration techniques. This two-step approach precludes accurate and precise estimation of T_1 values since registration errors propagate to the estimation step, leading to biased T_1 estimates. However, there is no fundamental reason why motion correction must be performed with image registration techniques that were conceived for qualitative imaging, thereby leaving aside the abundant information that the relaxation model can provide. In chapter 5, it is demonstrated that substantially more accurate and precise T_1 mapping can be achieved if the image registration problem, which is an estimation problem, is embedded into a global unified approach, where also the T_1 map is jointly estimated. By **integrating the models of T_1 relaxation, motion, and noise into one statistical model of the dataset of MR images, the original motion-free T_1 map can be restored by using a joint Maximum Likelihood (ML) estimator**. The unified ML framework allows to accounting for the statistical noise model, the relaxation model, and the motion model simultaneously, exploiting, in addition to the temporal information, knowledge on data statistics.

Chapter 6 focuses on Variable Flip Angle (VFA) T_1 mapping, the most efficient T_1 mapping method to date, considering the trade-off between accuracy and scanning time. Apart from the short image acquisition time, another reason why VFA T_1 mapping is so popular is that the specific mathematical structure of the VFA relaxation signal model gives rise to linear and, hence, fast T_1 estimators. However, these estimators are not derived from theoretically-grounded statistical principles, as ML estimators are. Consequently, the statistical properties of the ML estimators are fairly superior though, ML algorithms, which are non-linear algorithms, are naturally much slower than linear ones. In this chapter, the apparent trade-off between speed and statistical optimality is reconciled with **a novel, fast VFA T_1 estimator, which is shown to provide the non-linear least squares (NLLS) estimates, (which equal the ML estimates in common clinical scanning conditions), but is yet computationally much faster than standard NLLS algorithms**, such as Levenberg-Marquardt, with a computational time of the same order of that of linear estimators.

Finally, in chapter 7, **a novel k-space reconstruction technique to accelerate the relatively long acquisition of MRI images**, and thereby, reduce the overall long protocol time of MRI relaxometry studies, is presented. Furthermore, a discussion is given about the possibility to include information from the relaxation signal model into the k-space reconstruction method so as to fully exploit all the information that is present in the collected images. This way, it is expected that even higher acceleration ratios can be achieved, aiming at placing MRI relaxometry in the front of the wave of qMRI modalities that can be applied in clinically feasible time.

Samenvatting

In dit proefschrift worden nieuwe methoden voorgesteld voor de acquisitie, verwerking en analyse van magnetische resonantie (MR) relaxometrie data, met als doel de obstakels weg te nemen die MR relaxometrie momenteel nog verhinderen een snelle, nauwkeurige en precieze kwantitatieve beeldvormingsmethode voor klinisch gebruik te zijn. Het manuscript is verdeeld in drie delen. Deel I behandelt de basis van kwantitatieve MRI vanuit het oogpunt van signaalverwerking. Deel II geeft een korte inleiding van MRI-relaxometrie en statistische parameterschattingstheorie. De bijdragen van dit proefschrift tot slot zijn opgenomen in deel III.

Kwantitatieve MRI: een op modellen gebaseerd beeldvormingsparadigma

Magnetische resonantie beeldvorming (MRI) is een veelzijdige, niet-invasieve beeldvormingsmodaliteit die zachte weefsels kan afbeelden met een uitstekend contrast, zonder dat daarbij ioniserende straling wordt gebruikt. Met name daarom wordt MRI veel toegepast in de kliniek voor medische diagnose, preventie en opvolging van ziekten. MRI wordt over het algemeen op een kwalitatieve manier gebruikt, waarbij een radioloog het MR-beeld interpreteert, gericht op het identificeren van biomarkers die nuttig kunnen zijn voor het opvolgen van de ziektetoestand. De huidige radiologie vereist echter dat biomarkers afgeleid uit beelden betrouwbaar, reproduceerbaar en objectief zijn. Kwalitatieve MRI voldoet vaak niet aan deze eisen.

In MRI is de beeldintensiteit gerelateerd aan kwantitatieve parameters die intrinsieke kenmerken zijn van de biologische toestand van weefsels. De relatie tussen de intensiteit en die parameters wordt gegeven door een wiskundig model. Zo'n model legt een verband tussen de "foto's" die de MR-scanner maakt en de fysieke realiteit in de weefsels. Op deze manier worden MR-beelden niet alleen als afbeeldingen beschouwd, maar als dragers van unieke, weefselspecifieke informatie die rigoureus kan worden gemeten, geanalyseerd en geïnterpreteerd. Dit op modellen gebaseerde beeldvormingsparadigma heeft de naam kwantitatieve MRI (qMRI) gekregen. Kwantitatieve parameters kunnen worden geschat op basis van MRI-beelden, die vaak meer informatieve, gecontroleerde en objectieve biomarkers bieden dan deze verkregen met kwalitatieve MRI.

De lezer kan de eerste twee hoofdstukken van dit proefschrift aanwenden als een in-

leiding tot de basisprincipes van qMRI. Hoofdstuk 1 introduceert MRI. Vertrekkend van de fundamentele fysische principes die de spindynamiek beschrijven, wordt de lezer door de theorie geloodst die ten grondslag ligt aan de MR-signaalvorming en de lokalisatie ervan in de ruimte. Hoofdstuk 2 beschrijft hoe MR-afbeeldingen worden gereconstrueerd met algoritmen voor signaalverwerking.

MRI Relaxometrie

Een veelgebruikte modaliteit in qMRI is relaxometrie. MRI-relaxometrie richt zich op de schatting van de spin-rooster, T_1, en de spin-spin, T_2, relaxatietijden. Beide relaxatietijden zijn fundamentele parameters die de spindynamiek in een weefsel beschrijven tijdens het relaxatieproces van het fenomeen van de kernmagnetische resonantie (NMR). Gedurende de laatste decennia zijn ruimtelijke T_1 en T_2 kaarten van de hersenen geanalyseerd om de toestanden van een groot aantal ziekten te bestuderen en te monitoren, zoals: multiple sclerose, de ziekte van Alzheimer, dementie, schizofrenie of myocardiale fibrose. Die studies hebben aangetoond dat MRI-relaxometrie de belofte in zich draagt van het genereren van robuuste, objectieve, op beelden gebaseerde biomarkers voor pathologieën van het centrale zenuwstelsel, hart- en vaatziekten en daarbuiten.

Echter, biomarkers die worden verkregen worden uit MRI-relaxometrie zijn nog niet voldoende specifiek, gevoelig en robuust om routinematig in de klinische praktijk te worden gebruikt. Daarbovenop vereisen hoge-resolutie relaxometrie-beelden een klinisch onhaalbaar lange scantijd.

Door een kwantitatieve benadering toe te passen, kunnen bekende termen zoals nauwkeurigheid en precisie worden gerelateerd aan specificiteit en sensitiviteit, terwijl de scantijd fundamenteel gekoppeld is aan de tijd voor het verzamelen van gegevens. Terminologie zoals *nauwkeurigheid, precisie en acquisitietijd* is essentieel voor de wetenschappelijke taak van “meten”, een proces dat twee belangrijke componenten heeft: data-acquisitie en data-analyse. De nauwkeurigheid en precisie van kwantitatieve MRI-relaxometrie worden sterk beïnvloed door fouten die voortkomen uit onnauwkeurige fysieke modellering, maar ook door ontoereikende signaalverwerking of naïve parameterschattingstechnieken (data-analyse). Daarnaast, de acquisitie van MR relaxometrie data is tijdrovend omdat een groot aantal datapunten moet worden bemonsterd om een set van hoogwaardige MRI-beelden te reconstrueren die nodig zijn om T_1 en T_2 parameters te schatten met een hoge nauwkeurigheid en precisie. De geïnteresseerde lezer kan het hoofdstuk 3 aanwenden als een korte inleiding tot de basis van MRI-relaxometrie en de belangrijkste uitdagingen die deze techniek met zich meebrengt. Verder wordt een korte inleiding tot statistische parameterschattingstheorie, een must voor moderne qMRI, gepresenteerd in hoofdstuk 4.

Bijdragen

Dit proefschrift bavat drie hoofdbijdragen die de huidige problemen in MRI-relaxometrie aanpakken, en gericht zijn op het transformeren van deze modaliteit tot een robuuste, qMRI-techniek die biomarkers genereert in een klinisch haalbare tijd. Hiertoe hanteren we een aanpak op basis van een modelgebaseerd signaalverwerkingsperspectief, met een belangrijke nadruk op statistische parameterschatting. Er is gewerkt aan de ontwikkeling van modelgebaseerde algoritmen die zowel theoretisch onderbouwd als efficiënt zijn om te worden geïmplementeerd in klinische MR-scannersoftware.

Hoofdstuk 5 behandelt het probleem van nauwkeurige en precieze schatting van T_1 parameters, zelfs bij beweging van de patiënt. Traditioneel wordt beweging, voorafgaand aan het schatten van de T_1 parameters, gecorrigeerd door de opgenomen serie van MR-beelden ruimtelijk te registreren met algemene beeldregistratietechnieken. Deze tweestapsbenadering sluit een nauwkeurige en precieze schatting van de T_1-waarden uit, omdat registratiefouten zich na de schattingsstap voortpropageren, wat leidt tot systematische fouten bij de T_1 parameterschattingen. Er is echter geen fundamentele reden waarom bewegingscorrectie moet worden uitgevoerd met beeldregistratietechnieken die oorspronkelijk werden ontwikkeld voor kwalitatieve beeldvorming, waarbij cruciale informatie die het relaxatiemodel kan bieden buiten beschouwing wordt gelaten. In hoofdstuk 5 wordt aangetoond dat aanzienlijk accuratere en preciezere T_1 waarden kunnen worden verkregen als het beeldregistratieprobleem, dat een schattingsprobleem is, wordt ingebed in een globale uniforme aanpak, waarbij ook de T_1 kaart gezamenlijk wordt geschat. Door **de modellen van T_1 relaxatie, beweging en ruis te integreren in een statistisch model van de dataset van MR-afbeeldingen, kan de originele, bewegingsvrije T_1-kaart worden hersteld met behulp van een gezamenlijk meest aannemelijke (ML) schatter**. Het uniforme ML-raamwerk maakt het mogelijk om tegelijkertijd rekening te houden met het statistisch ruismodel, het relaxatiemodel en het bewegingsmodel.

Hoofdstuk 6 richt zich op Variable Flip Angle (VFA) T_1 mapping, de meest efficiënte T_1 mapping-methode tot nu toe, gezien de wisselwerking tussen nauwkeurigheid en scantijd. Afgezien van de korte beeldopnametijd, is een andere reden waarom VFA T_1-toewijzing zo populair is, dat de specifieke mathematische structuur van het VFA-relaxatiesignaalmodel aanleiding geeft tot lineaire en dus snelle T_1-schatters. Deze schatters zijn echter niet afgeleid van theoretisch gefundeerde statistische principes, zoals de ML-schatters. De statistische eigenschappen van de ML-schatters zijn dan ook superieur, hoewel ML-algoritmen, die niet-lineair zijn, van nature veel langzamer zijn dan lineaire algoritmen. In dit hoofdstuk wordt de schijnbare wisselwerking tussen snelheid en statistische optimaliteit verzoend met **een nieuwe, snelle VFA T_1-schatter, waarvan wordt aangetoond dat deze de niet-lineaire kleinste-kwadraten schattingen (NLLS) oplevert (die gelijk zijn aan de ML schattingen in veel voorkomende klinische scan omstandigheden) maar nog steeds veel sneller is dan standaard NLLS-algoritmen**, zoals Levenberg-Marquardt, met een rekentijd van dezelfde orde als die van lineaire schatters.

Tenslotte wordt in hoofdstuk 7, **een nieuwe k-ruimte reconstructietechniek geïntroduceerd om de relatief lange acquisitie van MRI-beelden te verkorten**, en daardoor de lange opnametijd bij MRI-relaxometriestudies te verminderen. Verder wordt een discussie gewijd aan de mogelijkheid om informatie uit het relaxatiesignaalmodel op te nemen in de k-ruimte-reconstructiemethode om alle informatie die aanwezig is in de verzamelde beelden ten volle te benutten. Op deze manier kunnen naar verwachting zelfs nog hogere versnellingsratio's worden bereikt, waardoor MRI-relaxometrie een plaats zal kunnen gaan innemen in de voorlinie van de groep van qMRI-modaliteiten die kunnen worden toegepast in een klinisch haalbare tijd.

Part I

Quantitative MRI from a signal processing perspective

1

MRI: From spins' physics to image formation

Contents

1.1 Introduction

Magnetic Resonance Imaging (MRI) is a versatile non-invasive imaging modality that allows in vivo visualization of structure and function of the human body. MRI produces images with high contrast between different soft tissues without the need of ionizing radiation. By employing specific pulse sequences, MRI can be used to probe different physical phenomena that occur in the human body, leading to different MRI modalities with distinct clinical and biological applications. One of those modalities is the target application of this PhD, MRI relaxometry, the field that deals with the study or measurement of the relaxation variables in the Nuclear Magnetic Resonance (NMR) phenomenon. To provide a comprehensive approach for MRI relaxometry, an introduction of the underlying physical process of relaxation is needed. Moreover, in order to understand how an MR image is created, the signal detection and the underlying image formation theory are briefly reviewed. This chapter aims to provide the basic principles of MRI from a signal processing point of view. For more elaborate information on MRI, we refer the reader to the excellent references [Kuperman, 2000, Liang and Lauterbur, 2000, Bernstein et al., 2004, Tofts, 2004].

1.2 Historical overview

The very early achievements in the MRI discovery and development date back to the 1930s, when Isodor Isaac Rabi described the NMR phenomena for the first time [Rabi et al., 1938]. For this work, he was awarded the Nobel Prize in Physics a decade later. The next two important names in the journey of MRI are Felix Bloch and Edward Mills Purcell, who were jointly awarded the Nobel Prize in 1952. Bloch at Harvard University, and Purcell at Stanford University, independently demonstrated that the NMR phenomenon could be used to identify the specific atoms in any solid or liquid placed in a magnetic field [Bloch, 1946, Purcell et al., 1946]. Another milestone in the history of MRI may be attributed to Raymond Damadian, who discovered that the NMR signal of cancerous tissues is different from that of healthy tissues. This observation constituted the first application of NMR to medicine [Damadian, 1971], which in fact was patented just two years later.

The conceptual depart from one-dimensional (1D) NMR signals to two-dimensional (2D) images was accomplished by Paul Lauterbur (1973), who, relying on the work of Herman Carr (1954), developed the theory of spatial information encoding [Lauterbur, 1973, Carr and Purcell, 1954]. MRI as we know it nowadays was born. In parallel to Paul Lauterbur's seminal work, Peter Mansfield developed a method, currently known as echo planar imaging (EPI), so as to acquire 2D MR images in only a few seconds [Mansfield, 1977]. Both Lauterbur and Mansfield received the Nobel Prize in Medicine and Physiology in 2003, for their work on non-invasive imaging of the body. An important pioneer in the development of MRI was also Richard Ernst, who was awarded the Nobel Prize in Chemistry in 1991. Richard Ernst was the first to describe the use of the Fourier transform to reconstruct 2D

images, by using switched magnetic field gradients in the time domain for spatial encoding [Kumar et al., 1975]. The first MRI commercial systems were launched in the early 1980s. The interested reader may find a more extensive overview of the history of MRI in the reference [Mattson and Simon, 1996]. Of course, MRI would not have been possible without the work of scientists that may be well considered fundamental for the history of science, and not particularly confined to MRI. Perhaps the name of Nikola Tesla stands out among them. The discovery of the rotating magnetic field (1882) is due to the Serbian genius, and is, of course, critical for the functioning of MRI.

1.3 NMR signal generation

1.3.1 Spins' physics

Every physical object can be broken down into its molecules, then to atoms, and then to the nuclei and electrons. A fundamental property of nuclei is that those with an odd number of protons or neutrons possess an angular moment $\boldsymbol{J} = (J_x, J_y, J_z)^T$, which is often called spin, and whose modulus is given by

$$||\boldsymbol{J}||_2 = \hbar\sqrt{I(I+1)}, \tag{1.1}$$

where $\hbar$ is the reduced Planck's constant ($1.05 \cdot 10^{-34} J \cdot s$) and I is the nuclear spin quantum number. Like any spinning charged object, a nucleus whose $\boldsymbol{J}$ is nonzero creates a magnetic field around it. Such a magnetic field is represented by $\boldsymbol{\mu}$, the so-called magnetic dipole moment, which is related to $\boldsymbol{J}$ as [Lauterbur, 1973]

$$\boldsymbol{\mu} = \gamma \boldsymbol{J}, \tag{1.2}$$

with γ the gyromagnetic ratio of the nucleus. The direction of $\boldsymbol{\mu}$ is completely random in the absence of an external magnetic field. However, the direction $\boldsymbol{\mu}$ becomes deterministic when spins are exposed to a strong static external magnetic field $\boldsymbol{B_0} = (B_{0x}, B_{0y}, B_{0z})^T$ (Fig. 1.1). Indeed, if $\boldsymbol{\mu}$ is regarded as a classical magnetic moment, classical mechanics dictates that the magnetic moment $\boldsymbol{\mu}$ precesses about $\boldsymbol{B_0}$ with a precession angular frequency known as the Larmor frequency:

$$\omega_L = \gamma ||\boldsymbol{B_0}||_2, \tag{1.3}$$

and where the angle, θ, between $\boldsymbol{\mu}$ and $\boldsymbol{B_0}$ can only take $(2I+1)$ possible values, which are theoretically known [Lauterbur, 1973]. All nuclei with $I \neq 0$ exhibit this property, which is fundamental for the NMR phenomenon. Some of the nuclei that can be studied with NMR are: ^{1}H, ^{13}C, ^{19}F, ^{23}Na and ^{31}P. The most common nucleus that is considered in clinical MRI exams is the hydrogen nucleus, i.e., ^{1}H. For ^{1}H, the gyromagnetic ratio is $2.675 \cdot 10^8$ rad/s/T, and its nuclear spin number is 1/2. Therefore, $\boldsymbol{\mu}$ only precesses with two possible orientations: pointing up (parallel to $\boldsymbol{B_0}$) and pointing down (antiparallel to $\boldsymbol{B_0}$). In order to describe the collective behavior of a spin system or a spin ensemble, a macroscopic magnetization vector, $\boldsymbol{M} = (M_x, M_y, M_z)^T$, is used. This vector is often called the

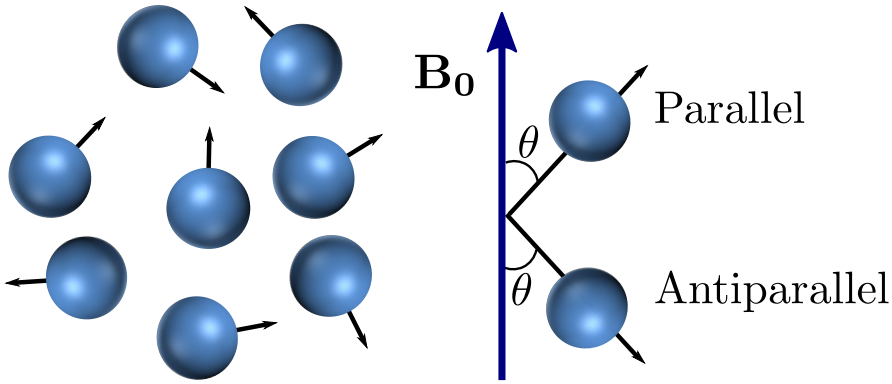


Figure 1.1: Nuclear magnetic dipole moments of ^{1}H (a) pointing in random directions, and (b) aligned with the only two possible directions.

net nuclear magnetization vector. The net nuclear magnetization vector of a spin ensemble is defined as the sum of the individual magnetic dipole moments of all spins which constitute that ensemble. In the absence of external magnetic field, $\boldsymbol{M} = \mathbf{0}$, since the direction of individual magnetic dipole moments is random. In the presence of $\boldsymbol{B_0}$, which we will assume that only points along the z-axis, that is, $\boldsymbol{B_0} = (0, 0, B_{0z})^T$, the transversal component of $\boldsymbol{M}$ is still zero since individual spins have a random phase in the x-y plane when they precess about the z-axis. However, M_z, is non-zero since it is the sum of the z-component of individual magnetic dipole moments whose individual directions are deterministic.

For a spin-$\frac{1}{2}$ system ($I = 1/2$), like ^{1}H, the orientation and strength of $\boldsymbol{M}$ can be determined by means of Boltzmann statistics. Indeed, it is well known that spins in different orientations have different energies of interaction with $\boldsymbol{B_0}$. According to quantum theory, spins-up possess negative energy whereas spins-down have positive energy. Since spins are more likely to take the lower energy state (with higher stability) than the higher-energy state, the population of pointing-up spins is slightly higher than that of pointing-down spins. Due to this difference in population, there is a net component of $\boldsymbol{M}$ in the positive direction of $\boldsymbol{B_0}$. On the contrary, there is no transverse component of $\boldsymbol{M}$ since, in the transverse plane, the average of the random phases of the magnetic dipole moments becomes zero. We say that such a spin system is magnetized and in thermal equilibrium.

1.3.2 Bloch equations

The NMR phenomenon can be studied in a tractable way when the macroscopic point view is adopted. Indeed, the evolution of a ensemble of spins in the presence of a magnetic field can be described by means of a differential equation formalism. The cornerstone in the macroscopic approach to NMR are the Bloch equations. In their more general form, Bloch equations describe the evolution of $\boldsymbol{M}$ along time

t under the presence of an arbitrary magnetic field $\boldsymbol{B} = (B_x, B_y, B_z)^T$ as [Bloch, 1946]

$$\frac{\partial \boldsymbol{M}}{\partial t} = \underbrace{\gamma \boldsymbol{M} \times \boldsymbol{B}}_{Excitation} - \underbrace{\begin{pmatrix} T_2^{-1} & 0 & 0 \\ 0 & T_2^{-1} & 0 \\ 0 & 0 & T_1^{-1} \end{pmatrix} \boldsymbol{M} + \boldsymbol{M_0} \begin{pmatrix} 0 \\ 0 \\ T_1^{-1} \end{pmatrix}}_{Relaxation}, \tag{1.4}$$

where $\times$ represents the cross-product operation, T_2 and T_1 are the transverse and longitudinal relaxation times, and $\boldsymbol{M_0} = (M_x(0), M_y(0), M_z(0))^T$ is the net nuclear magnetization vector at equilibrium. Bloch equations describe two fundamental processes of the NMR phenomenon, namely the *excitation process* due to an applied magnetic field, and the *relaxation process* of $\boldsymbol{M}$ towards the system equilibrium. Due to the fundamental importance the relaxation process has for this PhD thesis, we treat this concept separately in subsection 1.3.5. We now first focus on the *excitation* part of Eq. (1.4) when $\boldsymbol{B_0}$ is the only magnetic field applied.

1.3.3 Free precession

In the previous section, it was mentioned that, for spin-$\frac{1}{2}$ systems that are in thermal equilibrium, the vector $\boldsymbol{M}$ points out in the direction of $\boldsymbol{B_0}$. If that is so, $M_x = 0$ and $M_y = 0$. In a more general situation, $\boldsymbol{M}$ precesses about the z-axis with Larmor angular frequency ω_L [Lauterbur, 1973]. Mathematically, the precession of the vector $\boldsymbol{M}$ about the z-axis and with frequency ω_L can be described as

$$\boldsymbol{M} = \boldsymbol{R}_z(\omega_L t)\boldsymbol{M_0}, \tag{1.5}$$

where the rotation matrix $\boldsymbol{R}_z(\omega_L t) \in \mathbb{R}^{3\times 3}$ is defined as

$$\boldsymbol{R}_z(\omega_L t) = \begin{pmatrix} \cos\omega_L t & -\sin\omega_L t & 0 \\ \sin\omega_L t & \cos\omega_L t & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.6}$$

The precession of $\boldsymbol{M}$ about the z-axis can also be explained by means of the Bloch equations. Indeed, in equilibrium, the relaxation effects in Eq. (1.4) can be disregarded [Lauterbur, 1973]. Therefore, since the only magnetic field $\boldsymbol{B}$ that is present is $\boldsymbol{B_0}$, Eq. (1.4) transforms into

$$\frac{\partial \boldsymbol{M}}{\partial t} = \gamma \boldsymbol{M} \times \boldsymbol{B_0} = \begin{pmatrix} \gamma ||\boldsymbol{B_0}||_2 M_x \\ -\gamma ||\boldsymbol{B_0}||_2 M_y \\ 0 \end{pmatrix}, \tag{1.7}$$

whose solution is given by

$$\begin{pmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{pmatrix} = \begin{pmatrix} \cos\omega_L t & -\sin\omega_L t & 0 \\ \sin\omega_L t & \cos\omega_L t & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} M_x(0) \\ M_y(0) \\ M_z(0) \end{pmatrix}, \tag{1.8}$$

and which, according to Eq. (1.6), describes the precession of the vector $\boldsymbol{M}$ about the z-axis with frequency ω_L (Fig. 1.2).

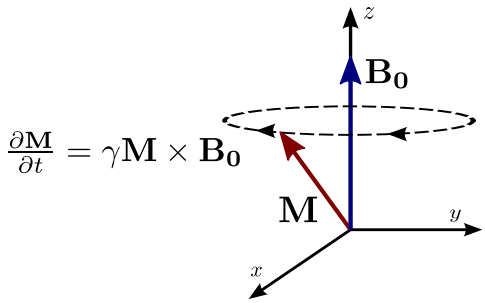


Figure 1.2: Precession of net nuclear magnetization vector $\boldsymbol{M}$ about $\boldsymbol{B_0}$ with angular frequency ω_L.

When the external magnetic field $\boldsymbol{B}$ is non-static, describing the temporal evolution of $\boldsymbol{M}$ becomes more complicated. The scope of the next subsection is to characterize the behavior of $\boldsymbol{M}$ when $\boldsymbol{B}$ is a dynamic magnetic field, namely, a Radio Frequency (RF) magnetic field.

1.3.4 RF excitation

1.3.4.1 Resonance condition

A necessary condition for the relaxation process to occur is that the spin system is perturbed from its equilibrium state. This is achieved by applying a second time-varying magnetic field, $\boldsymbol{B_1}(t)$, which is perpendicular to $\boldsymbol{B_0}$. If the oscillating frequency of $\boldsymbol{B_1}(t)$ equals the Larmor frequency, the individual spins that constitute the spin system present coherence in phase. This phenomenon is called the *resonance phenomenon*. The magnetic field $\boldsymbol{B_1}(t)$ is termed RF pulse since it oscillates in the RF range. The RF pulse, in contrast to $\boldsymbol{B_0}$, is normally turned on for just a few milliseconds, and typically takes the following form [Liang and Lauterbur, 2000],

$$\boldsymbol{B_1}(t) = B_1^e(t)\left(\cos(\omega_{RF}t)\mathbf{1}_x - \sin(\omega_{RF}t)\mathbf{1}_y\right), \tag{1.9}$$

where $B_1^e(t)$ is the envelope of the pulse, $\omega_{RF} = \omega_L$ (resonance condition), and $\mathbf{1}_x$ and $\mathbf{1}_y$ are the unitary vectors of the x and y-axis, respectively. The envelope function $B_1^e(t)$ is the heart of the RF pulse, since it uniquely determines its shape and duration, and in fact it determines how the spin-system is excited.

The resonance phenomenon can be studied by means of the Bloch equations as well. To do so, we can substitute $\boldsymbol{B} = \boldsymbol{B_0} + \boldsymbol{B_1}(t)$ in the excitation part of Eq. (1.4), and solve the equation for $\boldsymbol{M}$. While this is possible, a mathematical trick which greatly simplifies the analysis is the concept of the so-called rotating frame of reference.

1.3.4.2 RF-rotating frame of reference

A frame of reference that rotates along the z-axis with an angular frequency ω_{RF} is called a RF-rotating frame of reference. This new frame of reference, whose unit vectors along the orthogonal axes x', y', and z', are denoted as $\mathbf{1}_{x'}, \mathbf{1}_{y'}, \mathbf{1}_{z'}$, respectively, is related to the stationary (laboratory) frame by the following transformations:

$$\begin{cases} \mathbf{1}_{x'} = \cos(\omega_{RF}t)\mathbf{1}_x - \sin(\omega_{RF}t)\mathbf{1}_y \\ \mathbf{1}_{y'} = \sin(\omega_{RF}t)\mathbf{1}_x + \cos(\omega_{RF}t)\mathbf{1}_y \\ \mathbf{1}_{z'} = \mathbf{1}_z. \end{cases} \tag{1.10}$$

Since $\boldsymbol{B_1}(t)$ rotates with the same angular frequency as the RF-frame of reference, it appears to be a non-oscillatory field in this new coordinate system. Indeed, $\boldsymbol{B_1}(t)$ defined as in Eq. (1.9) points along the $\mathbf{1}_{x'}$ axis, and its component $B_{1x'}(t)$ is $B_1^e(t)$. The coordinates of the net nuclear magnetization vector $\boldsymbol{M}$ in the RF-rotation frame of reference, $\boldsymbol{M}_{Rot} = (M_{x'}, M_{y'}, M_{z'})^T$, can be succinctly derived by using complex notation. By defining the complex numbers $M_{xy} = M_x + iM_y$ and $M_{x'y'} = M_{x'} + iM_{y'}$, where i denotes the imaginary unit, it is possible to show that

$$M_{x'y'} = M_{xy}e^{i\omega_{RF}t}, \tag{1.11}$$

and $M_{z'} = M_z$. We can then solve the Bloch equations for $\boldsymbol{M}_{Rot}$ and go back to the laboratory frame with the relation given in Eq. (1.11). In the RF-rotating frame, (disregarding relaxation effects, which is a valid assumption if the RF pulse is sufficiently short [Lauterbur, 1973]), the Bloch equations take the form of

$$\frac{\partial \boldsymbol{M}_{Rot}}{\partial t} = \gamma \boldsymbol{M}_{Rot} \times (\boldsymbol{B}_{Rot} - \frac{\omega_{RF}\mathbf{1}_z}{\gamma}), \tag{1.12}$$

where $\boldsymbol{B}_{Rot}$ is the transformed $\boldsymbol{B}$ in the RF-rotating frame of reference, and which is given by

$$\boldsymbol{B}_{Rot} = \boldsymbol{B_0} + B_1^e(t)\mathbf{1}_{x'} = ||\boldsymbol{B_0}||_2\mathbf{1}_z + B_1^e(t)\mathbf{1}_{x'}. \tag{1.13}$$

Invoking the resonance condition, $\omega_{RF} = \omega_L = \gamma||\boldsymbol{B}_0||_2$, and substituting Eq. (1.13) in Eq. (1.12), we arrive at

$$\frac{\partial \boldsymbol{M}_{Rot}}{\partial t} = \gamma \boldsymbol{M}_{Rot} \times B_1^e(t)\mathbf{1}_{x'}. \tag{1.14}$$

Eq. (1.14) is the most general expression for the Bloch equations in the RF-rotating frame of reference. Observe that, under the resonance condition, the envelope of the RF pulse, $B_1^e(t)$, is the only external effect that controls how the spin-system is perturbed from equilibrium. A solution of Eq. (1.14) under initial conditions $M_{x'} = 0$, $M_{y'} = 0$ and $M_{z'} = M_z(0)$ is given by

$$\begin{cases} M_{x'}(t) = 0 \\ M_{y'}(t) = M_z(0)\sin\left(\int_0^t \gamma B_1^e(t')dt'\right) & 0 \le t \le \tau_p \\ M_{z'}(t) = M_z(0)\cos\left(\int_0^t \gamma B_1^e(t')dt'\right), \end{cases} \tag{1.15}$$

where τ_p is the duration of the RF pulse. Those relations describe a precession of $\boldsymbol{M}_{Rot}$ about the x' axis, with an angular frequency that varies with time and vanishes when $t > \tau_p$. In the laboratory frame, the net nuclear magnetization vector $\boldsymbol{M}$ precesses both about the $\boldsymbol{B_0}$ field but also about the $\boldsymbol{B_1}(t)$ field, a phenomenon which is called *forced precession*. To illustrate the forced precession phenomenon in the RF-rotating and laboratory frame of reference, let us consider the following simple case where $\boldsymbol{B_1}(t)$ is a square pulse whose time length is τ_p and amplitude B_1. Then, Eq. (1.15) becomes

$$\begin{cases} M_{x'}(t) = 0 \\ M_{y'}(t) = M_z(0)\sin\omega_1 t \quad 0 \leq t \leq \tau_p \\ M_{z'}(t) = M_z(0)\cos\omega_1 t, \end{cases} \tag{1.16}$$

with $\omega_1 = \gamma B_1$. It is evident that $\boldsymbol{M}_{Rot}$ precesses with a constant angular frequency ω_1 till $t = \tau_p$, when it stops. A graphical illustration of the precession of $\boldsymbol{M}_{Rot}$ about x' is presented in Fig. 1.3.(b). The forced precession of $\boldsymbol{M}$ as observed in the laboratory frame is depicted in Fig. 1.3.(a).

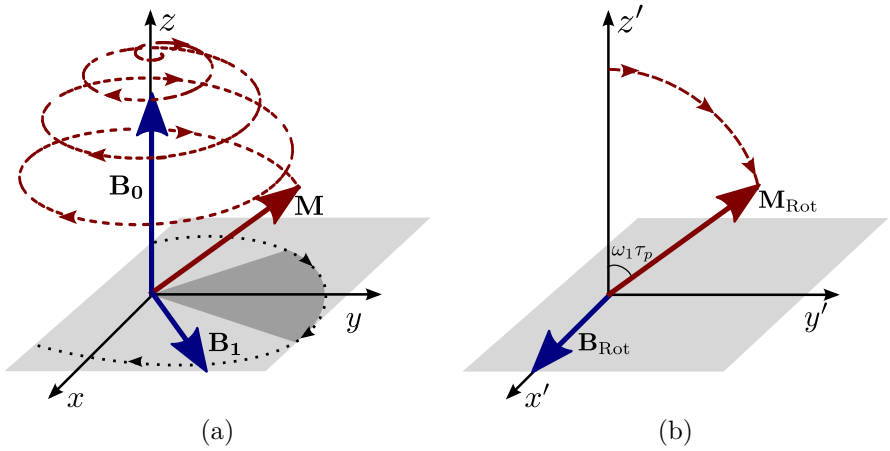


Figure 1.3: Motion of the net nuclear magnetization field in the presence of an RF field as observed in (a) the laboratory frame, and (b) the RF-rotating frame.

1.3.4.3 Flip angle

Due to the forced precession, the net nuclear magnetization is tipped away from the z-axis, creating a measurable transverse component. The angle between $\boldsymbol{M}_{Rot}$ and the z'-axis is known as the flip angle, and is denoted by α. According to Eq. (1.15), α is given by

$$\alpha = \int_0^{\tau_p} \gamma B_1^e(t')dt', \tag{1.17}$$

which, in the case of the rectangular pulse described above, $\alpha = \gamma B_1 \tau_p = \omega_1 \tau_p$.

The formalism of flip angle becomes useful since it allows to describe the effect of an RF pulse as a rotation of the vector $\boldsymbol{M}_{Rot}$ about the axis defined by $\boldsymbol{B_1}(t)$, and with an angle of α degrees with respect of the initial position. An RF pulse that causes a rotation of $\boldsymbol{M}_{Rot}$ with α degrees about the axis defined by $\boldsymbol{B_1}(t)$ is named an α-pulse. Popular choices are $\alpha = 90°$ and $\alpha = 180°$, with $\boldsymbol{B_1}(t)$ being aligned with either the x' or with the y'-axis. Those RF pulses are concisely indicated as $\alpha_{x'}$ and $\alpha_{y'}$, respectively. Like in Eq. (1.5), the net nuclear magnetization vector in the RF-rotating frame after the application of an $\alpha_{x'}$ or $\alpha_{y'}$ RF pulse, $\boldsymbol{M}_{Rot}^{t=0_+}$, can be calculated as

$$\begin{aligned} \boldsymbol{M}_{Rot}^{t=0_+} &= \boldsymbol{R_{x'}}(\alpha)\boldsymbol{M}_{Rot} \text{ or} \\ \boldsymbol{M}_{Rot}^{t=0_+} &= \boldsymbol{R_{y'}}(\alpha)\boldsymbol{M}_{Rot}, \end{aligned} \tag{1.18}$$

respectively, with

$$\boldsymbol{R_{x'}}(\alpha) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \text{ and } \boldsymbol{R_{y'}}(\alpha) = \begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix}. \tag{1.19}$$

1.3.5 Relaxation

After a magnetized spin system has been perturbed from its equilibrium state by an RF pulse, it will return to its equilibrium state according to the laws of thermodynamics. This process is called *relaxation*. Phenomenologically, the relaxation process can be described by the Bloch equations (second term in Eq. (1.4)), as a first-order dynamic process. Particularly, in the RF-rotating frame, we have

$$\begin{cases} \frac{\partial M_{z'}}{\partial t} = -\frac{M_{z'} - M_z(0)}{T_1} \\ \frac{\partial M_{x'y'}}{\partial t} = -\frac{M_{x'y'}}{T_2}. \end{cases} \tag{1.20}$$

The longitudinal relaxation time, T_1, characterizes the relaxation process of the longitudinal component $M_{z'}$, whereas the transverse relaxation time, T_2, describes the relaxation curve of the transverse component $M_{x'y'}$. We illustrate the recovery of the net nuclear magnetization vector towards the equilibrium state in Fig. 1.4. The justification of both relaxation process is based on the time-dependent microscopic magnetic fields that surround the nuclei and arise from random thermal motion. The specific causes of relaxation are rather diverse and we here just describe them in brief. The interested reader is referred to more specialized references [Cowan, 1997, Kimmich, 1997].

The **longitudinal or spin-lattice relaxation** stems from the exchange of energy with other degrees of freedom in the spin system in order to redistribute the population of the nuclear spin states. In the NMR jargon, these degrees of freedom are referred to as the *lattice* [Cowan, 1997]. This energy is often dissipated in terms of vibration and rotation motion. The spin-lattice relaxation process depends, among other things, on the gyro-magnetic ratio, γ, of the nucleus, and the mobility of the lattice. Phenomenologically, the spin-lattice relaxation process results in

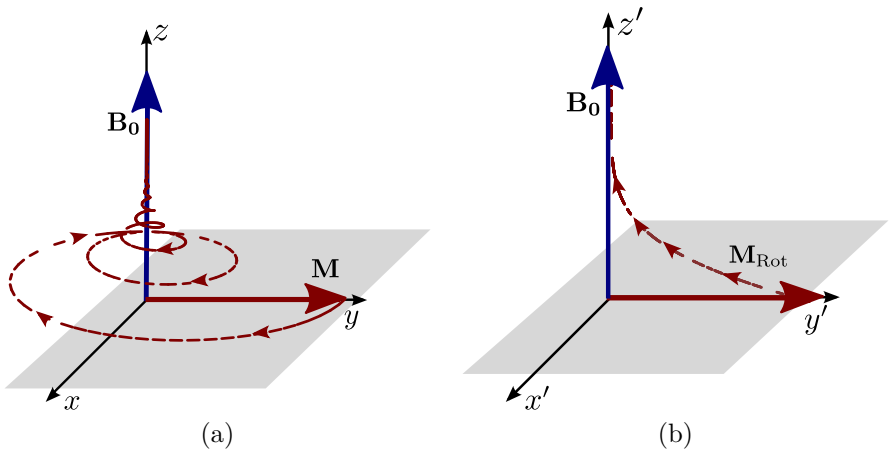


Figure 1.4: Relaxation of the net nuclear magnetization vector towards equilibrium as observed in (a) the laboratory frame, and (b) the RF-rotating frame.

a growth of M_z (Fig. 1.5), which is characterized by T_1, and whose evolution is fully determined by Eq. (1.20). An important consequence of the phenomenological

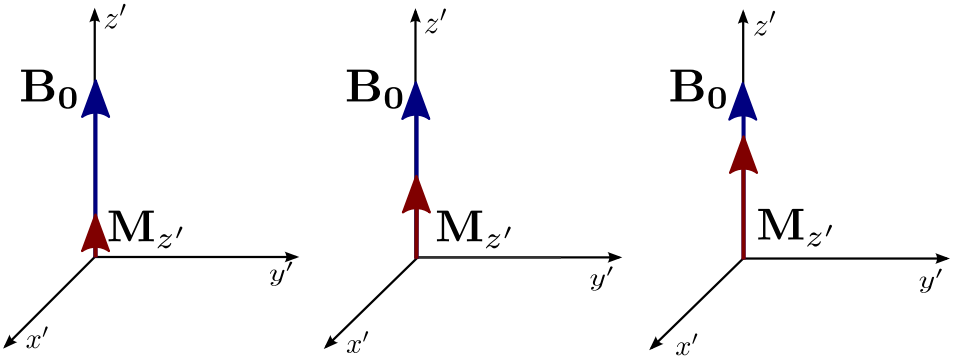


Figure 1.5: T_1 relaxation: Due to a loss of energy, the longitudinal component of the net nuclear magnetization vector M_z recovers towards equilibrium.

description of the spin-lattice relaxation as a first-order process is that the decay of M_z is exponential. Indeed, after an $\alpha_{x'} = 90°$ pulse, M_z can be written in function of time as [Lauterbur, 1973]

$$M_z(t) = M_{z'}(t) = M_z(0)(1 - e^{-\frac{t}{T_1}}) + M_{z'}^{t=0_+} e^{-\frac{t}{T_1}}, \tag{1.21}$$

where $M_{z'}^{t=0_+}$ is the longitudinal component of the net nuclear magnetization vector in the RF-rotating frame intermediately after the $\alpha_{x'} = 90°$ pulse.

The **transverse or spin-spin relaxation** refers to the recovery of the transverse component of the net nuclear magnetization vector $M_{x'y'}$. The transverse relaxation is caused by the interaction between spins of the same nuclear system, which causes

them to lose phase coherence. The loss in phase coherence of the spins results in a decrease of $M_{x'y'}$ (Fig. 1.6). Such a decrease is characterized by T_2, and described by Eq. (1.20). Another aspect that causes spin dephasing is the presence of an inhomogeneous magnetic field. The transversal relaxation due to the time independent field inhomogeneities and spin-spin interactions is called T_2^*, and is related to T_2 by the following inverse relation [Chavhan et al., 2009]:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}, \tag{1.22}$$

with $\frac{1}{T_2'} = \gamma \Delta \boldsymbol{B}_{\text{Inhomo.}}$, and where the term $\Delta \boldsymbol{B}_{\text{Inhomo.}} \geq 0$ contains contributions of the inhomogeneous magnetic field. Note that since $\gamma > 0$, T_2^* is always smaller or equal than T_2. The solution of Eq. (1.20) for $M_{x'y'}$ has also a exponential

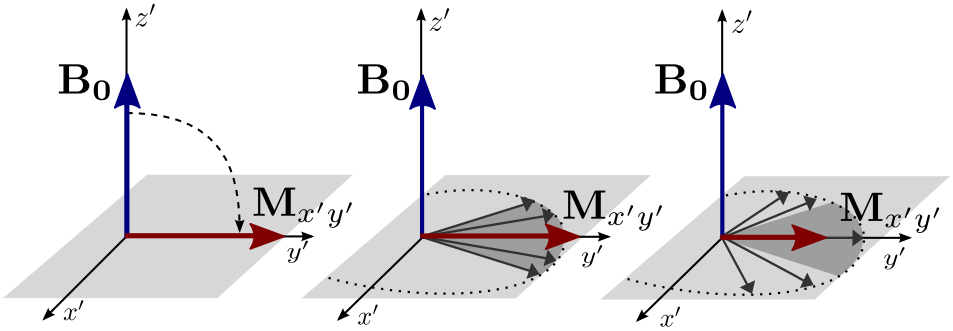


Figure 1.6: T_2 relaxation: The net nuclear magnetic dipoles associated with the spins dephase, causing a progressively decrease in $M_{x'y'}$. Figure adapted from [Van Steenkiste, 2016].

form,

$$M_{x'y'}(t) = M_{x'y'}^{t=0_+} e^{-\frac{t}{T_2^*}}, \tag{1.23}$$

with $T_2^* = T_2$ when field inhomogeneities are disregarded, and where $M_{x'y'}^{t=0_+}$ is the transverse component of the net nuclear magnetization vector in the RF-rotating frame immediately after the $\alpha_{x'} = 90°$ pulse. In contrast to the T_1 relaxation, the transverse relaxation does not involve any energy exchange, and is usually much faster (Fig. 1.7). In the laboratory frame, the net nuclear magnetization vector $\boldsymbol{M}$ can be written as:

$$\begin{cases} M_x(t) = \text{Re}\left(M_{x'y'}^{t=0_+} e^{-\frac{t}{T_2^*}} e^{-i\omega_{RF}(t+\tau_p)} \right) \\ M_y(t) = \text{Im}\left(M_{x'y'}^{t=0_+} e^{-\frac{t}{T_2^*}} e^{-i\omega_{RF}(t+\tau_p)} \right) \\ M_z(t) = M_z(0)(1 - e^{-\frac{t}{T_1}}) + M_{z'}^{t=0_+} e^{-\frac{t}{T_1}}, \end{cases} \tag{1.24}$$

where $\text{Re}\,(\cdot)$ and $\text{Im}\,(\cdot)$ denote the real and imaginary part operators, respectively. Observe that the longitudinal component of $\boldsymbol{M}$ is a slowly varying function compared to the transversal components $M_x(t)$ and $M_y(t)$. Since signal detection in a

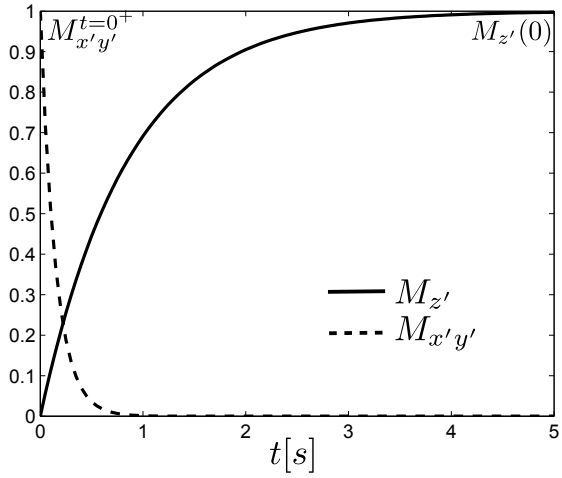


Figure 1.7: Longitudinal (solid line) and transversal (dashed line) relaxation after the $\alpha_{x'} = 90°$ pulse. Note that longitudinal relaxation is a much slower process than transversal relaxation.

MR scanner is based on the well known Faraday's law of induction, and Faraday's law dictates that the voltage (signal) induced in the coil depends linearly on the time change of the magnetic flux, the voltage depends solely on the transverse component given by Eq. (1.24). Such a signal is known as *free induction decay* (FID). FID signals are the most basic form of signals that can be detected from a spin system after pulse excitation. Though they are rarely used alone since they quickly disappear [Lauterbur, 1973], FID signals constitute the basic of other more complex form of MR signals, such as MR echoes. The next subsection is devoted to describing the concept of MR echo.

1.3.6 MR echo generation

FID signals originate from the spin-spin relaxation process, which as we have already described, is caused, among other things, by a loss in phase coherence of the nuclear system spins. In 1950, Erwin L. Hann presented a technique to refocus the spins that are progressively dephasing by means of the application of another 90° pulse after the initial 90° pulse, oriented along the same axis, either x' or y', thereby creating a measurable echo which is known in the NMR jargon as *spin echo* [Hahn, 1950]. The phenomenon of spin echo was further developed by Herman Carr and Edward Purcell four years later, who pointed out the advantages of using a $\alpha_{y'} = 180°$ (initial 90° pulse is assumed to be along the x' axis) instead of the second 90° pulse [Carr and Purcell, 1954]. The resulting pulse sequence is schematically shown in Fig. 1.8, and it receives the name of **spin echo** (SE) sequence. When the initial 90° pulse is applied, *fast* spins start to precess clockwise about the z-axis faster than *slow* spins, thereby progressively losing phase coherence.

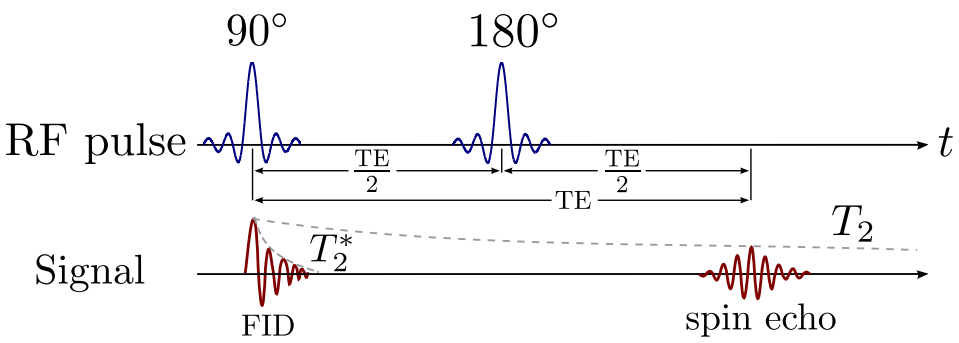


Figure 1.8: Spin echo sequence. The 180° pulse flips the nuclear spins to the other side of the traverse plane. Spins are rephased after TE, thereby producing a measurable echo.

After a time TE/2, where TE stands for *echo time*, the 180°-pulse rotates the net nuclear magnetization vector along the y'-axis, hence, flipping the nuclear spins to the other side of the transverse plane. Since spins continue to precess clockwise, slow spins now lead ahead whereas the fast ones trail behind. Progressively, fast spins catch slow spins, and hence complete refocusing is achieved at a time TE, when the spin echo is produced. Another advantage of the spin-echo sequence is that the spin dephasing due to the static magnetic field inhomogeneities is compensated with the 180° pulse. Consequently, the signal decay of the spin-echo only originates solely from the T_2 relaxation process.

Spin refocusing can also be achieved by applying gradient magnetic fields instead of a second RF pulse. A gradient magnetic field is a magnetic field whose components linearly vary along a specific direction. They are thoroughly described in subsection 1.4.1. Here, we just mention that the key concept of gradient magnetic fields for MR echo generation is that gradient magnetic fields can dephase and rephase a set of spins in a controlled fashion. The resulting echo signal is called a gradient re-called echo, and the corresponding pulse sequence is known as **gradient echo** (GE) sequence. In contrast to the spin echo signal, the gradient echo signal originates from T_2^* relaxation.

1.3.7 Signal detection

In MRI, the signal detection is based on Faraday's law of induction. The time-varying net nuclear magnetization vector induces a voltage $v(t)$ in the receiving coil, which is given by [Lauterbur, 1973]

$$v(t) = -\frac{\partial}{\partial t}\int_{\Omega} \langle \boldsymbol{B}^{\boldsymbol{R}}(\boldsymbol{r}), \boldsymbol{M}(\boldsymbol{r}, t)\rangle d\boldsymbol{r}, \tag{1.25}$$

where $\boldsymbol{r} = (x, y, z)^T$, Ω is the volume enclosed by that coil, $\boldsymbol{B}^{\boldsymbol{R}}(\boldsymbol{r})$ is the so-called coil sensitivity, and $\boldsymbol{M}(\boldsymbol{r}, t)$ is the net nuclear magnetization vector at point $\boldsymbol{r} \in \Omega \in \mathbb{R}^3$ and time t. As already mentioned, since the longitudinal component of

$\boldsymbol{M}(\boldsymbol{r},t)$ is a slow varying function compared to the transverse component, Eq. (1.25) can be simplified to [Lauterbur, 1973]

$$v(t) = -\int_{\Omega}\left(B_x^R(\boldsymbol{r})\frac{\partial M_x(\boldsymbol{r},t)}{\partial t} + B_y^R(\boldsymbol{r})\frac{\partial M_y(\boldsymbol{r},t)}{\partial t}\right)d\boldsymbol{r}. \tag{1.26}$$

By using the following complex notation, $B_{xy}^R(\boldsymbol{r}) = B_x^R(\boldsymbol{r}) - iB_y^R(\boldsymbol{r})$ and $M_{xy}(\boldsymbol{r},t) = M_x(\boldsymbol{r},t) + iM_y(\boldsymbol{r},t)$, Eq. (1.26) can be written as

$$v(t) = -\Re\left\{\int_{\Omega} B_{xy}^R(\boldsymbol{r})\frac{\partial M_{xy}(\boldsymbol{r},t)}{\partial t}d\boldsymbol{r}\right\}, \tag{1.27}$$

where $M_{xy}(\boldsymbol{r},t)$ can be related to the transverse component in the RF-rotating frame as $M_{xy}(\boldsymbol{r},t) = M_{x'y'}(\boldsymbol{r},t)e^{-i\omega_{RF}t}$. In all the cases, free precession is at much faster rate than relaxation, that is, the following assumption is valid:

$$\begin{aligned}\frac{\partial M_{xy}(\boldsymbol{r},t)}{\partial t} &= \frac{\partial M_{x'y'}(\boldsymbol{r},t)}{\partial t}e^{-i\omega_{RF}t} - i\omega_{RF}M_{x'y'}(\boldsymbol{r},t)e^{-i\omega_{RF}t} \\ &\approx -i\omega_{RF}M_{x'y'}(\boldsymbol{r},t)e^{-i\omega_{RF}t}.\end{aligned} \tag{1.28}$$

By doing so, Eq. (1.27) can be succinctly written as

$$v(t) = \Re\{i\omega_{RF}s(t)e^{-i\omega_{RF}t}\}, \tag{1.29}$$

where the complex signal $s(t)$ is given by

$$s(t) = \int_{\Omega} B_{xy}^R(\boldsymbol{r})M_{x'y'}(\boldsymbol{r},t)d\boldsymbol{r}. \tag{1.30}$$

Disregarding constants, the complex signal $s(t)$ receives, in communication theory, the name of *the complex envelope* of $v(t)$, whereas $e^{-i\omega_{RF}t}$ is the carrier signal [Haykin, 2001]. The complex envelope, $s(t)$, is the signal of interest in MRI, since it depends on the net nuclear magnetization vector. The complex envelope $s(t)$ can be written as $s(t) = s_I(t) + is_Q(t)$, where low-pass band signals $s_I(t)$ and $s_Q(t)$ are the (I)n phase and (Q)uadrature components, respectively. The technique that MR scanners include so as to extract $s_I(t)$ and $s_Q(t)$, and hence $s(t)$, from $v(t)$, is based on an I/Q demodulator [Haykin, 2001]. The schematic design of an I/Q demodulator is shown in Fig. 1.9.

1.4 MR image formation

Clearly, *all* spin systems in the object contribute to form the signal $s(t)$, resulting in the superposition integral of all the net nuclear magnetization vectors (Eq. (1.30)). Evidently, it is impossible to determine $M_{xy}(\boldsymbol{r},t)$ at every point $\boldsymbol{r}$ in the scanned object with information only from $s(t)$. In other words, more mechanisms are needed to form a graphical representation of the NMR phenomenon inside the scanned object. A graphical representation of the spatial distribution of the net nuclear magnetization vector is called an MR image. The essential concept that permits to move from NMR signals to MR images is the spatial localization or encoding concept.

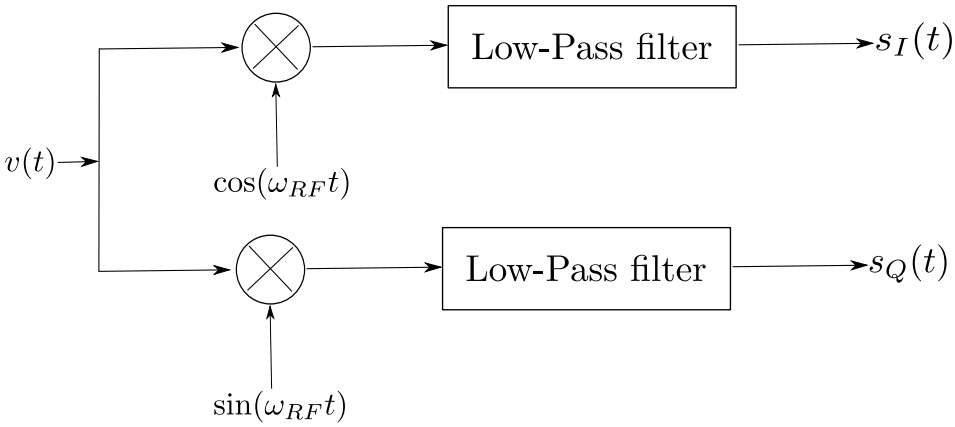


Figure 1.9: I/Q demodulator to extract the (I)n phase and (Q)uadrature components, $s_I(t)$ and $s_Q(t)$, of $v(t)$.

1.4.1 Spatial localization and k-space

Central to spatial localization is the use of magnetic gradient fields. A gradient field $\boldsymbol{B_G}(\boldsymbol{r})$ along the z-direction is a magnetic field whose z-component, $B_{G_z}(\boldsymbol{r})$, varies linearly with x, y and z, that is,

$$B_{G_z}(\boldsymbol{r}) = G_x x + G_y y + G_z z = \langle \boldsymbol{G}, \boldsymbol{r} \rangle, \tag{1.31}$$

where $\boldsymbol{G} = (G_x, G_y, G_z)^T$ is the gradient vector and $\langle \cdot, \cdot \rangle$ denotes the scalar product between vectors in an Euclidean space. A spin system located at point $\boldsymbol{r}$ that is placed under $\boldsymbol{B_G}(\boldsymbol{r})$ acquires a frequency which is given by

$$\omega(\boldsymbol{r}) = \omega_L + \gamma B_{G_z}(\boldsymbol{r}). \tag{1.32}$$

Consequently, $M_{xy}(\boldsymbol{r}, t) = M_{x'y'}(\boldsymbol{r}, t) e^{-i\omega_{RF}t} e^{-i\gamma B_{G_z}(\boldsymbol{r})t}$ and the MR signal transforms into

$$s(t) = \int_\Omega B^R_{xy}(\boldsymbol{r}) M_{x'y'}(\boldsymbol{r}, t) e^{-i\gamma \langle \boldsymbol{G}, \boldsymbol{r} \rangle t} d\boldsymbol{r}. \tag{1.33}$$

In the more general case where gradient fields depend on time, the MR signal turns out to be

$$s(t) = \int_\Omega B^R_{xy}(\boldsymbol{r}) M_{x'y'}(\boldsymbol{r}, t) e^{-i\gamma \int_0^t \langle \boldsymbol{G}(t'), \boldsymbol{r} \rangle dt'} d\boldsymbol{r}. \tag{1.34}$$

If we define the so-called k-space trajectories $\boldsymbol{k}(t) = (k_x(t), k_y(t), k_z(t))^T$ as

$$\boldsymbol{k}(t) = \frac{\gamma}{2\pi} \int_0^t \boldsymbol{G}(t') dt', \tag{1.35}$$

then Eq. (1.34) can be expressed as

$$s(t) = \int_\Omega B^R_{xy}(\boldsymbol{r}) M_{x'y'}(\boldsymbol{r}, t) e^{-2\pi i \langle \boldsymbol{k}(t), \boldsymbol{r} \rangle} d\boldsymbol{r}. \tag{1.36}$$

For a fixed time point t_1, the complex value $s(t_1)$ is nothing more than the three-dimensional spatial Fourier transform of the image $\rho(\boldsymbol{r}, t_1) = B^R_{xy}(\boldsymbol{r}) M_{x'y'}(\boldsymbol{r}, t_1)$, evaluated at vector frequency $\boldsymbol{k} = \boldsymbol{k}(t_1)$. By varying the acquisition time t_p in a time interval where $\rho(\boldsymbol{r}, t)$ remains relatively constant, different values of the Fourier transform of $\rho(\boldsymbol{r}, t)$ can be acquired. Different values of the Fourier transform can also be probed by changing the shape of $\boldsymbol{G}$ and hence of the k-space trajectories In the k-space terminology, the set of complex values $s(t_p)$ for every $p \in \mathbb{Z}$, where $\mathbb{Z}$ denotes the set of integer numbers, is denoted as k-space data, and $\boldsymbol{k}(t_p)$, $\forall p \in \mathbb{Z}$ the set of k-space points. Ideally, if all the values of the Fourier transform of $\rho(\boldsymbol{r}, t)$ are contained in the acquired k-space data set, the image $\rho(\boldsymbol{r}, t)$ can be analytically calculated by means of an inverse Fourier transform.

The observation that the MR signal can be mathematically expressed as a Fourier transform in the presence of time-varying gradient is probably the cornerstone of MR imaging since it allows to contemplate MR image formation as a *solvable* inverse Fourier problem. We italicize the word solvable because, in real life, attempting to solve the inverse Fourier problem to form an MR image possesses a huge number of challenges and mathematical considerations.

The field that is dedicated to solving the MR inverse problem from measured k-space data is called MR image reconstruction. Due to the fundamental importance of MR image reconstruction and, since this PhD thesis contains a novel contribution in that field, we fully dedicate the next chapter to guide the reader through the fundamentals of MR image reconstruction. Before ending this section, we briefly comment about common terminology that is often employed in the MR image formation literature: slice encoding and in-plane encoding.

1.4.2 Slice encoding

In subsection 1.4.1, we showed that MR images can be encoded in three dimensions with an appropriate k-space trajectory, thereby leading to a 3D Fourier transform. However, the classical approach in MR reconstruction is performing a tomographic acquisition, where a stack of 2D slices of the object is acquired sequentially. This type of acquisition is called a *multi-slice acquisition*. The rationale behind a multi-slice acquisition is that only the spin systems that are located in a plane given by $z = z_0$ are excited. More realistically, the spin systems located at $|z - z_0| \leq \Delta z$ are excited, where Δz is the slice thickness. To perform a selective excitation, two mechanisms are needed. First, a gradient field that varies along z is placed in the system, resulting in a spatially dependent Larmor frequency:

$$\omega_L = \gamma(||\boldsymbol{B_0}||_2 + G_z z). \tag{1.37}$$

Second, a slice-selective RF pulse is also required, whose resonance frequency is given by

$$\omega_{RF} = \gamma(||\boldsymbol{B_0}||_2 + G_z z_0), \tag{1.38}$$

where the shape of the envelope of the RF pulse determines Δz (Fig. 1.10).

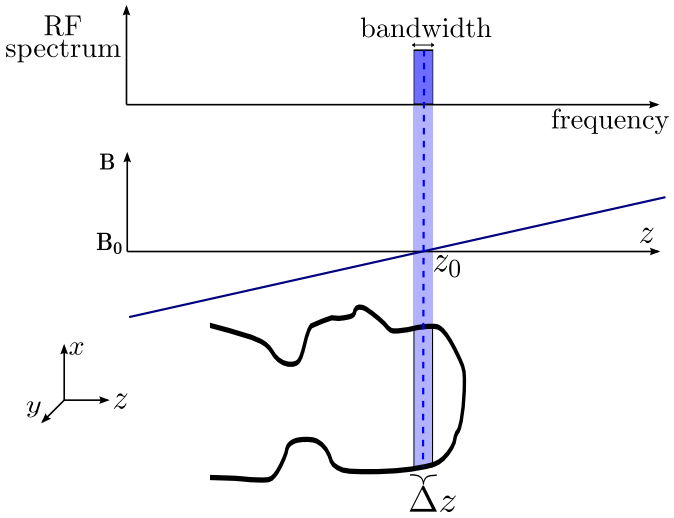


Figure 1.10: Illustration of the slice-encoding technique.

In this case, disregarding for simplicity practical considerations such as the effect of the slice profile, the MRI signal $s(t)$ can be seen interpreted as the 2D Fourier transform, also known as the in-plane Fourier transform, of the 2D function which results from the averaged slices $\boldsymbol{\rho}(x, y, z, t)$ within $|z - z_0| \leq \Delta z$.

1.4.3 Frequency and phase encoding

By doing slice encoding, only 2D k-space trajectories, namely, $k_x(t)$ and $k_y(t)$, are required to form 2D images. Though every 2D trajectory can be used, multi-slice images are often acquired by using phase and frequency encoding. To understand this specific terminology let us consider a 2D k-space trajectory constructed whose gradients G_x, G_y are sequences of square pulses. In this case, $k_x(t)$ and $k_y(t)$ are always linear in t. If, during the read-out step, that is, when the echo signal is received, only G_x is active and G_y has been applied after the initial RF pulse and then turned off before acquisition, we have that

$$k_x(t) = \frac{\gamma}{2\pi} G_x t, \tag{1.39}$$

but

$$k_y(t) = \frac{\gamma}{2\pi} G_y \tau_{G_y}, \tag{1.40}$$

where τ_{G_y} is the width of the pulses of G_y. Therefore, the complex exponential $e^{-2\pi i \langle \boldsymbol{k}(t), \boldsymbol{r} \rangle}$ in $s(t)$ adopts the form

$$e^{-2\pi i \langle \boldsymbol{k}(t), \boldsymbol{r} \rangle} = e^{-2\pi\gamma i G_x x t - 2\pi\gamma i G_y y \tau_{G_y}}. \tag{1.41}$$

The direction x is said to be frequency encoded since only the frequency of $e^{-2\pi i \langle \boldsymbol{k}(t), \boldsymbol{r} \rangle}$ depends linearly on x. Direction y, on the other hand, is said to

be phase encoded because the phase of $e^{-2\pi i\langle \boldsymbol{k}(t),\boldsymbol{r}\rangle}$ depends linearly on y. Different phase encoded lines can be acquired by just changing the value of height of the pulses in G_y in different repetitions.

1.5 MRI scanner

An MRI scanner (Fig. 1.11) is composed of three main hardware components: a large superconducting magnet, an RF system, and a magnetic field gradient system.

Figure 1.11: A 3T SIEMENS MAGNETOM Prisma scanner. Image courtesy of Siemens-Healthcare.

The **large superconducting magnet** is employed to generate the homogeneous, static, magnetic field $\boldsymbol{B_0}$. The strength of $\boldsymbol{B_0}$ is an important feature of every MR scanner. Typical strengths of $\boldsymbol{B_0}$ are 1.5T or 3T, however, some sites also operate a 7T scanner. The most important requisite for the larger magnet is its ability to produce a homogeneous field over the whole region of interest. In practice, the main magnet alone is not capable of generating such a completely homogeneous magnetic field. The way to overcome this problem is by using a secondary magnetic field generated by the so-called shim coils [Lauterbur, 1973].

The **RF system** consists of a transmitter coil so as to generate the $\boldsymbol{B_1}(t)$ field, and a receiver coil that converts the net nuclear magnetization vector into the complex MRI signal, $s(t)$. Sometimes, the same coil is used for both transmission and reception. Similar to $\boldsymbol{B_0}$, a desirable feature of a RF system is to provide a uniform $\boldsymbol{B_1}(t)$ field and high detection sensitivity [Lauterbur, 1973]. Several types of external coils can be used, namely, a body coil, surface coils, a bird-cage head coil or arrays of small coils [Lauterbur, 1973].

Finally, the **gradient system** is responsible for generating the time-varying magnetic field to provide both spatial encoding and GE-echo signals. Important specifications for a gradient system are the maximum gradient strength (mT/m) and the rise time, the time to achieve the maximum gradient strength. Both features are combined in a single specification name called slew rate, which is the maximum gradient strength divided by the rise time. Most superconducting clinical scanners operate with slew rates between 80 and 150 T/m/s, and most advanced ones can do so with slew rates around 200 T/m/s [GE, , SIE, b, SIE, a, PHI,]. Typical maximum gradient strengths are around 20-60 mT/m. Recent scanners reach values as high as 80 mT/m [GE, , SIE, b, SIE, a, PHI,].

2

MR image reconstruction

Contents

2.1 Introduction

In this chapter, the basics of MR image reconstruction are introduced. We start with fundamental concepts of sampling theory for Cartesian k-space sampling, such as aliasing and spatial resolution. As MRI is a slow imaging modality due to the strong sampling conditions, emphasis is given on the case of under-sampled k-space data, aiming at accelerated acquisition. To that end, we describe current parallel MRI techniques, and we further elaborate on modern algebraic reconstruction techniques that use prior knowledge and regularization. This chapter provides the theoretical background so as to understand one of the contributions of this PhD thesis (chapter 7).

2.2 The Continuous Inverse Fourier problem

In subsection 1.4.1, we described how the MR signal is localized in space with the concept of k-space encoding. It was shown that samples of the acquired signal $s(t)$ (Eq. (1.34)), that is, $s(t_p)$, $\forall p \in \mathbb{Z}$, can be seen as points of the spatial Fourier transform of $\rho(\boldsymbol{r}, t_p) = B^R_{xy}(\boldsymbol{r}) M_{x'y'}(\boldsymbol{r}, t_p)$. In MR image reconstruction, to end up with a proper Fourier Inverse problem, it is common to assume that all the k-space points correspond to the Fourier transform of a "snapshot" of $\rho(\boldsymbol{r}, t)$ at a fixed time t'. This way, $\rho(\boldsymbol{r}, t')$ can be treated as a 3D spatial image that is constant in time. Consequently, the samples $s(t_p), \forall p \in \mathbb{Z}$ can be seen as the amplitude of the frequencies of the Fourier spectrum of a *single* image $\rho(\boldsymbol{r}, t')$, which, for simplicity can be denoted as $\rho(\boldsymbol{r}) = B^R_{xy}(\boldsymbol{r}) M_{x'y'}(\boldsymbol{r})$. The time point t' is normally chosen to be the maximum point of the echo signal, that is, the echo time (TE).

The overarching assumption that we have made does apply under the following conditions. Let us suppose that $s(\boldsymbol{k}_p)$, with $\forall p \in \mathbb{Z}$, are acquired within a time interval around t' where the temporal change of $M_{xy}(\boldsymbol{r}, t)$ is negligible. This type of acquisition is called *single-shot acquisition*. If that is so, we can perfectly assume that all k-space points correspond to the *same* image $\rho(\boldsymbol{r})$.

On the other hand, this assumption is also valid for *multi-shot acquisition*. In this case, the complete NMR excitation process described in chapter 1 is repeated multiple times, and, in each "shot", different and small subsets of k-space points of $\boldsymbol{k}_p$, $\forall p \in \mathbb{Z}$ are acquired. The time between repetitions is called repetition time and is often denoted as TR. During the acquisition of the subset of k-space points, the temporal change in $M_{xy}(\boldsymbol{r}, t)$ is, again, assumed to be negligible.

We are then ready to define the MR image reconstruction problem but first, let us ease the notation. A k-space point that is mapped when a given k-space trajectory, $\boldsymbol{k}(t)$, is evaluated at $t = t_p$, with $p \in \mathbb{Z}$, that is, $\boldsymbol{k}(t_p)$, is denoted as $\boldsymbol{k}_p$. Furthermore, from now on, $s(\boldsymbol{k}_p)$ denotes the corresponding k-space data measurement, i.e., $s(t_p)$.

The MR image reconstruction problem is defined as [Lauterbur, 1973]

$$\text{Given } s(\boldsymbol{k}_p) = \int_{\Omega} \rho(\boldsymbol{r}) e^{-2\pi i \langle \boldsymbol{k}_p, \boldsymbol{r} \rangle} d\boldsymbol{r} \text{ with } \boldsymbol{k}_p, \forall p \in \mathbb{Z}$$
$$\text{determine } \rho(\boldsymbol{r}). \tag{2.1}$$

Observe that if $B^R_{xy}(\boldsymbol{r})$ is known, $M_{x'y'}(\boldsymbol{r})$ can be obtained exactly provided problem Eq. (2.1) is solved for $\rho(\boldsymbol{r})$. There are multiple techniques for estimating $B^R_{xy}(\boldsymbol{r})$ [Vemuri et al., 2005, Allison et al., 2012, Allison et al., 2013], so, in this dissertation, we will just assume that $B^R_{xy}(\boldsymbol{r})$ is given, and hence we only focus on reconstructing the image $\rho(\boldsymbol{r})$.

Problem 2.1 occurs in many scientific disciplines and, in fact, has been thoroughly studied long before MRI was conceived. It belongs to the class of discrete-to-continuous inverse problems, where the goal is to find a continuous function that matches a discrete finite or infinite set of samples [Lauterbur, 1973]. Such a problem is naturally ill-posed since it is undetermined, i.e., there are infinite possible choices for $\rho(\boldsymbol{r})$ to solve Eq. (2.1) [Fessler, 2010]. On top of that, in practice, the number of k-space points is always finite [Fessler, 2010]. Evidently, in order to end up with a tractable problem, some assumptions about the distribution of k-space points $\boldsymbol{k}_p$, and the target image $\rho(\boldsymbol{r})$ should be made. The branch of mathematics that deals with the recovery of a continuous function from a discrete set of samples is called sampling theory [Marks, 1993]. A brief incursion in sampling theory is required to understand the MR image reconstruction process. For simplicity, we will just focus on a 2D k-space case (the 3D case is simply a generalization of this particular case).

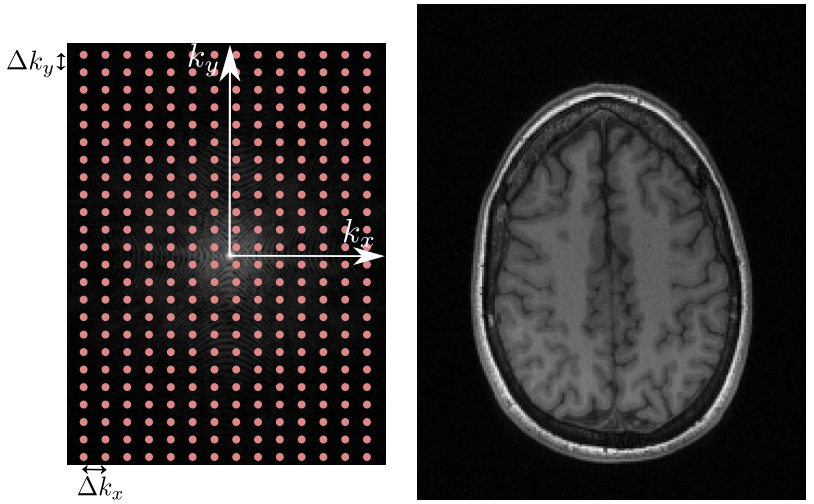
2.3 Cartesian k-space sampling

Most of the MR images are acquired with a Cartesian-based k-space scheme. The main reason why this type of sampling schemes is the most popular is based on two separate aspects. First, Cartesian k-space trajectories are relatively simple to implement and very robust in the presence of scanner hardware imperfections. Second, and perhaps, more importantly, there exist exact recovery methods for Cartesian k-space sampling [Marks, 2009, Marks, 1993]. This has motivated the conception of computational efficient reconstruction algorithm, which are nowadays broadly integrated into commercial MR scanners software.

An MR image is said to be sampled with a Cartesian k-space scheme if the k-space points $\boldsymbol{k}_p, \forall p \in \mathbb{Z}$, can be indexed with new indexes m and n as follows:

$$k_{x\,m} = m\Delta k_x, \quad k_{y_n} = n\Delta k_y, \tag{2.2}$$

with $m \in \mathbb{Z}$, $n \in \mathbb{Z}$, and where Δk_x and Δk_y are the sampling intervals along the x and y direction, respectively. In Fig. 2.1, the k-space data of an actual MR image sampled with a Cartesian scheme are graphically represented. An important concept in Cartesian k-space sampling is the so-called **aliasing effect**, and the Nyquist conditions. Those topics are covered below.



(a) (b)

Figure 2.1: Cartesian k-space scheme sampling (a) of the Fourier Transform of an actual MR image (b).

2.3.1 Nyquist conditions and aliasing

We define the periodic summation of the image $\rho(x, y)$, with periods $1/\Delta k_x$ and $1/\Delta k_y$, as the image $\bar{\rho}(x, y)$ given by [Pinsky, 2002]

$$\bar{\rho}(x, y) = \sum_{m\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} \rho(x - \frac{m}{\Delta k_x}, y - \frac{n}{\Delta k_y}). \tag{2.3}$$

Equipped with the Poisson summation formula [Marks, 2009], the following result can be shown to be true:

$$\bar{\rho}(x, y) = \Delta k_x \Delta k_y \sum_{m\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} s(m, n) e^{2\pi i m \Delta k_x x + 2\pi i n \Delta k_y y}, \tag{2.4}$$

where we have used the notation $s(m, n)$ for the discrete *k-space* image $s(m\Delta k_x, n\Delta k_y)$. The image $\bar{\rho}(x, y)$ is nothing more than the 2D Discrete-Time Fourier Transform (DTFT) of $s(m, n)$ [Marks, 2009]. Observe that whereas the *k-space image* is discrete or digital, the image $\bar{\rho}(x, y)$ is continuous. The word discrete-time is used deliberately to distinguish this type of Fourier transform from the popular Discrete Fourier Transform (DFT), where both domains are truly discrete [Marks, 2009].

Eq. (2.4) means that the value of $\bar{\rho}$ at every point x and y can be exactly recovered by applying a 2D DTFT to the k-space image. Without further assumptions, it is in principle not possible to extract $\rho(x, y)$ from $\bar{\rho}$. However, if $\rho(x, y)$ is an image of *finite support*, meaning that

$$\rho(x, y) = 0, \quad |x| > \frac{W_x}{2} \quad |y| > \frac{W_y}{2}, \tag{2.5}$$

then, there exist conditions to guarantee that $\rho(x, y)$ can be recovered from its periodic summation extension. Indeed, if we select the sampling intervals Δk_x and Δk_y that fulfill

$$\Delta k_x \leq \frac{1}{W_x}, \Delta k_y \leq \frac{1}{W_y} \tag{2.6}$$

then $\bar{\rho}(x, y) = \frac{1}{\Delta k_x \Delta k_y} \rho(x, y)$, and hence the 2D DTFT of the k-space image is simply a scaled version of $\rho(x, y)$. The conditions presented in Eq. (2.6) are known as the **Nyquist conditions**. When the Nyquist conditions do not hold, we say that the k-space data are under-sampled [Marks, 2009]. When sampling intervals are strictly smaller than $1/\Delta k_x$ and $1/\Delta k_y$, the term over-sampled is used instead [Marks, 2009]. The support of the MR image $\rho(x, y)$ is determined by the region of interest volume Ω (see Eq. (1.25)), which in turn is defined by the coil sensitivity profile. In the MRI jargon, the support of the MR image is called the Field of View (FOV). In practice, the user selects the desired FOV and then, Δk_x and Δk_y are chosen as the limit of the Nyquist conditions, that is, $\Delta k_x = 1/W_x$ and $\Delta k_y = 1/W_y$. We then say that k-space data has been **fully-sampled**. In Fig. 2.2, we graphically illustrate the concept of undersampling and the phenomenon of aliasing in the reconstructed image space.

2.3.2 Truncation of k-space and spatial resolution

The number of acquired k-space data points is always finite. Since most of the energy of the MR image is contained in its low-frequency components, the k-space is usually sampled in a Cartesian scheme up to some maximum value. That value is determined by the farthest k-space points that can be reached with a given trajectory, which in turn depends on the gradient strength [Fessler, 2010]. Being the maximum attainable k-space point in the x and y direction $k_{x\max}$ and $k_{y\max}$, respectively, the 2D DTFT of Eq. (2.4) transforms into a *truncated* version given by

$$S(x, y) = \Delta k_x \Delta k_y \sum_{m=-M/2}^{M/2-1} \sum_{n=-N/2}^{N/2-1} s(m, n) e^{2\pi i m \Delta k_x x + 2\pi i n \Delta k_y y}, \tag{2.7}$$

being M and N such that $(M/2-1) = \lfloor k_{x\max}/\Delta k_x \rfloor$ and $(N/2-1) = \lfloor k_{y\max}/\Delta k_y \rfloor$ respectively, where $\lfloor a \rfloor$ is the greatest integer that is less or equal to a. Eq. (2.7) can be seen as the truncated Fourier series of $\bar{\rho}(x, y)$ [Pinsky, 2002]. If the Nyquist conditions hold, $\bar{\rho}(x, y) = \rho(x, y)$, then [Pinsky, 2002]

$$S(x, y) = \rho(x, y) * D_{M,N}(x, y), \tag{2.8}$$

being $*$ the convolution operator and $D_{M,N}(x, y)$ the so-called Dirichlet periodic kernel:[1]

$$D_{M,N}(x, y) = \Delta k_x \Delta k_y \frac{\sin(M\pi\Delta k_x x)}{\sin(\pi\Delta k_x x)} \frac{\sin(N\pi\Delta k_y y)}{\sin(\pi\Delta k_y y)}. \tag{2.9}$$

[1] When the k-space is not symmetrically covered, like in Eq. (2.7), $D_{M,N}(x, y)$ should be multiplied by a complex exponential factor as well. However, this term is often left aside since its contribution is negligible [Lauterbur, 1973].

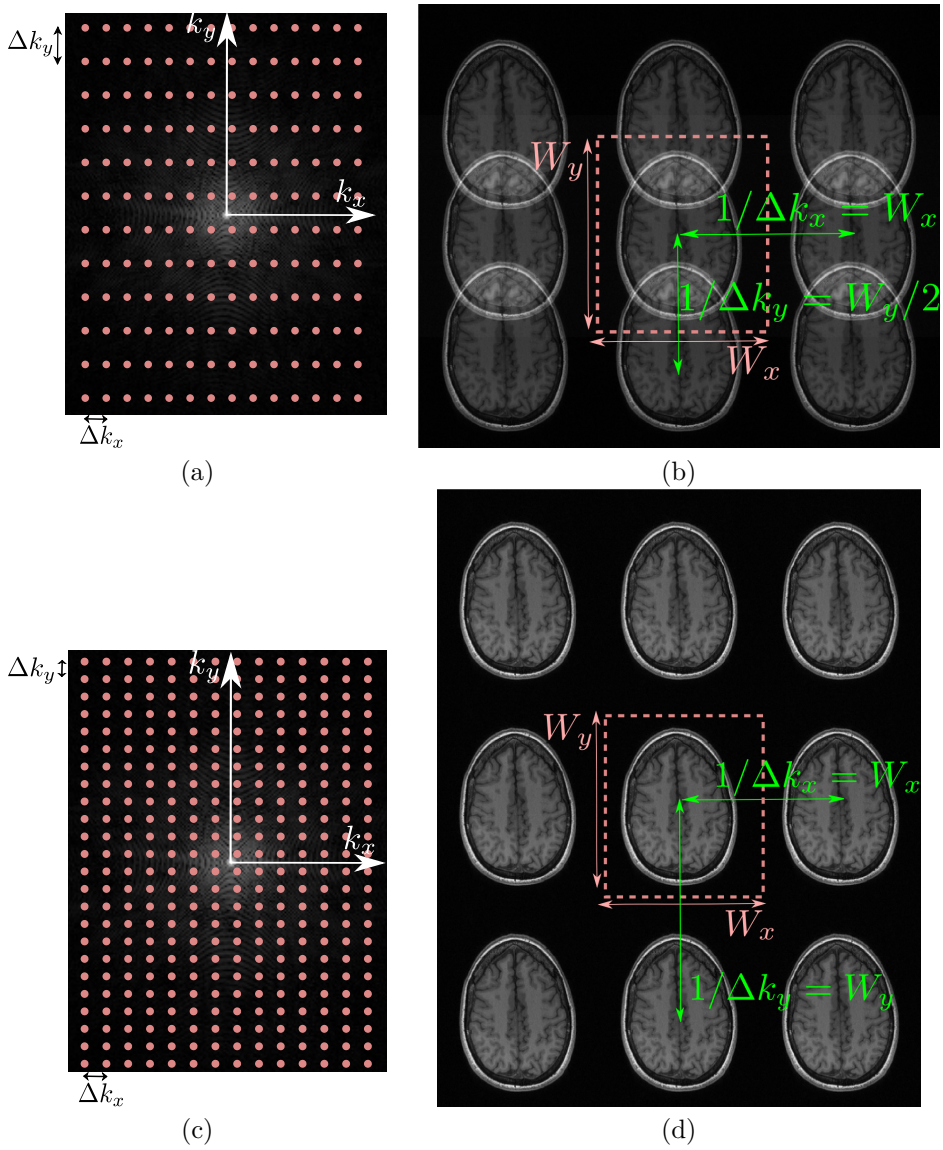


Figure 2.2: Cartesian k-space sampling scheme that violates the Nyquist criterion (a), and (b) reconstructed image after the 2DTFT , $\bar{\rho}(x, y)$. If the Nyquist conditions hold (c), no overlap exists in $\bar{\rho}(x, y)$ (d), and hence $\rho(x, y)$ is recoverable.

The function $D_{M,N}(x, y)$ can be interpreted as a Point Spread Function (PSF). The PSF $D_{M,N}(x, y)$ converges to the delta function when M and N tend to infinity. In the asymptotic case, $S(x, y) = \rho(x, y)$, as expected, since we have seen that for an infinite number of k-space points, exact recovery is possible if the Nyquist conditions are fulfilled. For finite M and N, the function $D_{M,N}(x, y)$ introduces distortion in $\rho(x, y)$. The image distortion comes in form of Gibbs ringing and,

more importantly, blurring. In general, the sharper the main lobule of the PSF function is, the less is then the blurring. Naturally, the main lobule becomes sharper when M and N increase. This intuitive explanation can be formalized by using classical concepts from imaging theory. In fact, a definition of **spatial resolution** can be given in terms of the width of the PSF. Two points can be "resolved" if they are separated more than the PSF width. The lower the width is, the higher is the spatial resolution. For PSFs that are not rectangular functions, the width is not uniquely defined. Two practical definitions are given. The Full Width at the Half Maximum (FWHM) and the so-called effective width. The effective width, W, is defined as the width of a rectangular function whose height is the maximum value of the actual PDF, and whose area is the area of the PSF. In particular, for the Dirichlet kernel PSF

$$W = \frac{1}{D_{M,N}(0,0)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} D_{M,N}(x,y)dxdy = \frac{1}{MN\Delta k_x \Delta k_y} \approx \frac{1}{k_{x\max} k_{y\max}}, \tag{2.10}$$

where it is assumed that $D_{M,N}(0,0)$ is the maximum point of $D_{M,N}(x,y)$. As a result, the higher $k_{x\max}$ and $k_{y\max}$ are, the higher the spatial resolution is. This could be achieved by increasing Δk_x and Δk_y. Since the Nyquist condition should be kept in mind, the only way to increase the spatial resolution is then to acquire more k-space data points, i.e., to increase M and N. Obviously the acquisition time then also increases. In Fig. 2.3, the effect that truncating the k-space has on the final spatial resolution is illustrated. The 2D DTFT of the truncated k-space data provides a continuous representation of $S(x,y)$, but MR images are always digital/discrete. In order to obtain a discrete image, Eq. (2.7) can be evaluated at points

$$\begin{aligned} x_m &= m\Delta x \quad \text{with } m = -M/2, ..., -1, 0, 1, ..., M/2-1, \\ y_n &= n\Delta y \quad \text{with } n = -N/2, ..., -1, 0, 1, ..., N/2-1, \end{aligned} \tag{2.11}$$

and hence it becomes

$$S(m,n) = \Delta k_x \Delta k_y \sum_{m'=-M/2}^{M/2-1} \sum_{n'=-N/2}^{N/2-1} s(m',n') e^{2\pi i m' \frac{m}{M} + 2\pi i n' \frac{n}{N}}, \tag{2.12}$$

which is the (scaled) 2D inverse DFT of $s(m,n)$, and is normally implemented with efficient Fast Fourier Transform (FFT) algorithms.

2.3.3 Limitations of Cartesian k-space sampling

While the advantages of Cartesian k-space sampling are evident (see subsection 2.3), it also presents serious limitations to efficient MR Imaging. Cartesian k-space sampling is often implemented with the frequency/phase encoding method (see subsection 1.4.3), which is also known as spin warping imaging [Edelstein et al., 1980]. Nowadays, most of the MRI acquisition protocols employ advanced upgraded versions of the original SE pulse sequence, where, instead of one single frequency line, several frequency lines are acquired per TR (Fig. 2.4). These sequences are often

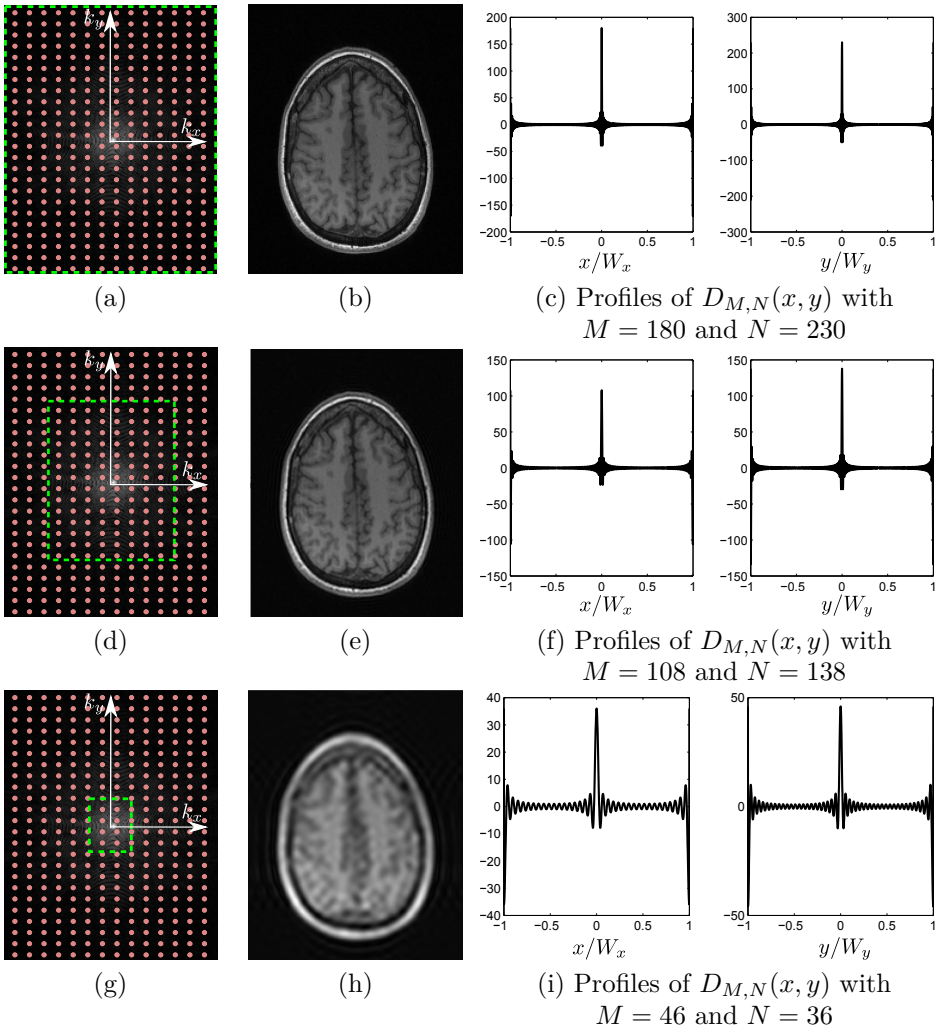


Figure 2.3: Loss of spatial resolution when the maximum attained k-space frequency is reduced, and hence M and N are reduced.

known by their commercial names Cartesian Fast Spin Echo (FSE) or Cartesian Turbo Spin Echo (TSE) [Bernstein et al., 2004]. If $N_{y/\mathrm{TR}}$ stands for the number of frequency lines that are acquired per TR with either an FSE or TSE sequence, and we assume that the acquisition time per frequency line is much shorter than TR, the total acquisition (T_{Acq}) time per 2D image can be shown to be

$$T_{\mathrm{Acq}} \approx \mathrm{TR}\frac{N}{N_{y/\mathrm{TR}}}. \tag{2.13}$$

Naturally, the user can tune the parameter $N_{y/\mathrm{TR}}$, however, $N_{y/\mathrm{TR}}$ cannot be made arbitrarily high without drastically suffering from T_2 decay. Likewise, we

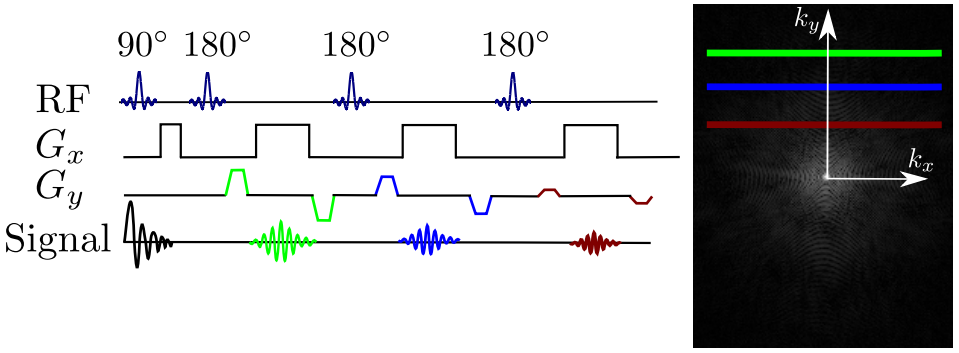


Figure 2.4: Schematic diagram of a TSE sequence with $N_{y/\mathrm{TR}} = 3$.

cannot decrease TR as much as we want. Indeed, in MRI relaxometry, we are interested in measuring the evolution of the net nuclear magnetization vector during the relaxation process. Therefore, consecutive RF pulses should not interfere with the relaxation process until the desired measurement has been made. In short, the time between pulses, i.e., TR, cannot be kept arbitrarily small. To illustrate the inadequacy of Cartesian TSE for fast MRI relaxometry, let us consider the following example. Let us suppose we employ the widespread Inversion Recovery (IR) pulse sequence with the goal of probing the T_1 relaxation time. For this particular sequence (see chapter 4 for more details), the TR value should be higher than four seconds, approximately. If $\mathrm{TR} = 4s$, the reconstruction process of a 2D slice MR image with $M = N = 256$ and $N_{y/\mathrm{TR}} = 10$ would take $T_{\mathrm{Acq}} = 4s \times 257/10 = 1.7$ min approximately! In an MRI relaxometry protocol, several images are acquired with different settings, therefore the total acquisition time for the image set becomes clearly infeasible.

If we adhere to fully-sampled k-space data, there is no other way than conceiving more efficient pulse sequences schemes to accelerate the acquisition. One of these pulse sequences is Cartesian Echo-Planar Imaging (EPI) [Mansfield, 1977]. With Cartesian EPI, *all* lines of the k-space are acquired simultaneously [Bernstein et al., 2004] during the readout. In its most basic version, Cartesian EPI corresponds to the type of single-shot acquisition [Bernstein et al., 2004], though segmented and interleaved EPI have also been proposed [Schmitt et al., 1998]. While much faster than Cartesian TSE, EPI presents numerous challenges for high image quality reconstruction since it often suffers from magnetic field inhomogeneities effects [Schmitt et al., 1998], leading to image distortions, often referred to EPI distortions. MRI reconstruction with EPI sequences deserves a special treatment on its own [Schmitt et al., 1998], and because in this PhD thesis EPI sequences have not been employed, we will not cover them here.

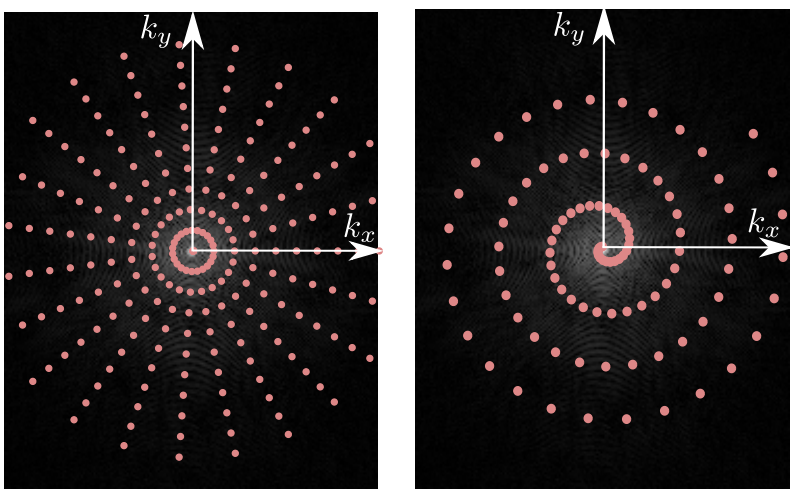
The main recognized disadvantage of Cartesian k-space schemes is intrinsic to the way the k-space is sampled. Cartesian k-space schemes are relatively inefficient at covering the whole k-space. This is because just a very small portion of the k-space is acquired during the readout step (a reduced number of lines). This explanation leads us naturally to conceive other different k-space trajectories resulting in

different k-sample sampling schemes.

2.4 Non-Cartesian k-space sampling

Historically, sampling along non-Cartesian trajectories has been largely ignored in favor of the simpler and well-understood spin-warping imaging or conventional Cartesian k-space sampling. This is true despite the seminal work of Paul Lauterbur [Lauterbur, 1973], where the concept of spatial encoding was illustrated with a radial sampling scheme. One of the most important properties of non-Cartesian trajectories is their potential for efficient use of MR gradient hardware and therefore rapid coverage of k-space [Wright et al., 2014]. Additionally, many non-Cartesian trajectories are much more tolerant to under-sampling [Peters et al., 2006], and are less affected by motion [Glover and Pauly, 1992, Liao et al., 1997, Feng et al., 2016]. Furthermore, they can be used with pulse sequences with ultra-short TE [Qian and Boada, 2008], since no preparatory step for phase encoding is needed. Perhaps, the two paradigmatic examples of non-Cartesian k-space sampling are radial and spiral sampling [Lauterbur, 1973]. Both radial and spiral k-space sampling can be applied in 3D MR image reconstruction. Here, for the sake of coherence with the rest of the chapter, we describe both k-space schemes for the 2D case.

With **radial sampling**, also known as projection sampling [Glover and Pauly, 1992, Altbach et al., 2002, Feng et al., 2016] the k-space is sampled by means of radial profiles or spokes, where these spokes always pass through the origin (Fig. 2.5.(a)). Similar to Cartesian k-space sampling, it is possible to derive conditions for exact



(a) (b)

Figure 2.5: Examples of non-Cartesian k-space sampling: (a) radial sampling (b) spiral sampling.

recovery. An important assumption is that image $\rho(x, y)$ has a finite circular support,

and that its Fourier transform is band-limited in the angular direction [Lauterbur, 1973]. If that is so, exact recovery conditions exist, although they require a slightly higher number of k-space data points than those for Cartesian k-space sampling [Lauterbur, 1973]. Additionally, the concept of the PSF to derive a spatial resolution criterion can be introduced as well. While conditions for exact recovery are more demanding than those for Cartesian k-space data, radial sampling has the benefit of performing better in the undersampling regime [Peters et al., 2006]. To understand this, observe that, by construction, in a radial scheme low-frequency components are acquired very often, whereas high- frequency components are acquired less frequently. Since most of the energy of the Fourier transform of an image is generally contained in the low-frequency region, k-space covering with radial schemes conveys more image information than with Cartesian k-space. Indeed, for Cartesian k-space sampling, undersampling leads to coherent artifacts in form of aliasing. For radial k-space sampling, reducing the number of spokes has a more clear effect in the high-frequency area rather than in the low-frequency region, where most of the energy is found. The resulting artifacts, which come in form of streaking artifacts, are shown to be incoherent, a property which may become useful for the concept of Compressed Sensing (CS) [Feng et al., 2014]. Another appealing property of radial sampling is its well-known robustness against motion artifacts during k-space acquisition. This explains the growing interest in radial sampling development for cardiac MR acquisitions. As a drawback, with radial scanning, it is important to account for anisotropic gradient delays [Peters et al., 2003].

Spiral sampling is known to be one of the most efficient schemes to sample the k-space [Ahn et al., 1986, Yudilevich and Stark, 1987b, Yudilevich and Stark, 1987a, Meyer et al., 1992, Irarrazabal and Nishimura, 1995]. With spiral sampling, the continuous k-space trajectory is parameterized by a spiral curve as is shown in Fig. 2.5.(b). Often, several spirals are interleaved, providing a more densely covered k-space. Guarantee conditions for exact recovery can also be given [Marks, 1993]. A spiral scan requires approximately 5-10 times fewer readouts than a Cartesian scan of similar resolution and FOV. So, for fully-sampled scans, the scan efficiency of spirals is the highest overall, followed by Cartesian and then by radial [Feng et al., 2016]. Like radial k-space sampling, spiral sampling is relatively robust to motion artifacts [Liao et al., 1997, Feng et al., 2016]. However, spiral schemes are less robust to under-sampling than radial schemes. For spiral scanning, it is typically desirable to perform a one-time calibration for gradient delays and eddy currents [Tan and Meyer, 2009, Addy et al., 2012, Campbell-Washburn et al., 2016].

The main disadvantage of non-Cartesian k-space sampling methods is the difficulty of reconstructing the resulting k-space data sets [Pauly, 2005]. Theorems for exact recovery in polar coordinates can be used to derive analytical formulas, but the numerical evaluation of those formulas is too slow [Yudilevich and Stark, 1987b, Yudilevich and Stark, 1987a,Marks, 1993]. Another option is the so-called regridding concept [Fessler and Sutton, 2003, Beatty, 2006]. Regridding reconstruction involves interpolating the k-space data into a Cartesian grid, and then reconstructing the final image with an FFT-based algorithm. Particular importance should be paid to the interpolating kernel since it affects the final quality of the reconstructed

image [Beatty et al., 2005,Beatty, 2006]. A more versatile option is reformulating the inverse Fourier Problem into a discrete-discrete inverse problem by a discretization the integral of Eq. (2.1). This type of reconstruction approach will be covered in subsection 2.7.

2.5 Phased-Array MR image reconstruction

Practically all of MR commercial scanners have the possibility to connect multiple RF detector coils, thereby allowing the simultaneous acquisition of the NMR signal. This technique is called phased-array technology (for similarities with phased-array radar), and it constitutes the standard way to reconstruct MR images. Phased-Array coil technology, introduced with the pioneering work of Roemer [Roemer et al., 1990], was conceived with the purpose of increasing the signal-to-noise ratio (SNR) of the reconstructed MR image, a terminology which will become clear in the next section when the concept of MR noise is introduced. Nevertheless, phased-array coil technology is most often used to reduce the scan time by means of undersampling k-space data. This technique receives the name of parallel MRI (pMRI), and this section is devoted to presenting its fundamental ideas and the most popular algorithms.

2.5.1 Multi-coil acquisition model

Let us consider an array of L coils, each one characterized by a transversal coil sensitivity profile $B^R_{xy_l}(\boldsymbol{r})$, with $l = 1, ..., L$. Like in subsection 2.2, if the coil sensitivities are known, and if we drop the temporal dependence of the magnetization vector, the MR image reconstruction problem in a phased-array or multi-coil system can be posed as

$$\begin{aligned} &\text{Given } s_l(\boldsymbol{k}_p) = \int_\Omega B^R_{xy_l}(\boldsymbol{r}) M_{x'y'}(\boldsymbol{r}) e^{-2\pi i \langle \boldsymbol{k}_p, \boldsymbol{r} \rangle} d\boldsymbol{r} \text{ with } \boldsymbol{k}_p, \forall p \in \mathbb{Z}, \text{ and } l = 1, ..., L \\ &\text{determine } M_{x'y'}(\boldsymbol{r}) \end{aligned} \tag{2.14}$$

In the origins of phased-array MR technology, Eq. (2.14) was "solved" independently. Indeed, each of the L fully-sampled k-space data sets, $s_l(\boldsymbol{k}_p)$, was inversely Fourier transformed (with either Cartesian or non-Cartesian k-space sampling), thereby providing the reconstructed image in the l-th coil, that is, $\rho_l(\boldsymbol{r}) = B^R_{xy_l}(\boldsymbol{r}) M_{x'y'}(\boldsymbol{r})$. Then, the next question was how to combine images $\rho_l(\boldsymbol{r})$, with $l = 1, ..., L$, aiming at getting $M_{x'y'}(\boldsymbol{r})$. The image resulting from this combination is called the composite MR image.

Though an advantage in terms of SNR, it seems like with this simple technique we are not fully exploiting the multiple parallel information given in every coil. In fact, extensions of the well-known Shannon sampling theorem have been given for a multichannel sampling problem like Eq. (2.14) [Papoulis, 1977, Cheung, 1993, Ying and Liang, 2010], indicating that exact reconstruction is possible even if the

individual sets $s_l(\boldsymbol{k}_p)$ are under-sampled. This justifies the technique of parallel MRI (pMRI), which we describe hereafter.

2.5.2 Parallel MRI (pMRI)

The first successful example of phased-array MR reconstruction with undersampled data was showcased by Daniel Sodickson [Sodickson and Manning, 1997]. In 1997, Sodickson and Manning proposed the simultaneous acquisition of spatial harmonics (SMASH) method so as to generate missing k-space lines by encoding with coil sensitives. The SMASH technique was followed by a plethora of acceleration methods that boosted pMRI to the state we know it today. Nowadays, the two most popular pMRI methods are by far SENSE and GRAPPA.

2.5.2.1 SENSitiviy Encoding (SENSE)

The SENSitiviy Encoding (SENSE) method was presented by Klaas Pruessmann [Pruessmann et al., 1999] to reconstruct MR images from under-sampled Cartesian k-space data. With SENSE, the extent of the k-space is kept the same, but the sampling interval in the phase direction Δ_{k_y} is augmented by a factor R, thereby violating the Nyquist criterion but successfully obtaining an acceleration of R times.

To understand how SENSE works, let us focus on the simple case of $R = 2$, that is, the k-space data set of every coil is sampled with

$$\Delta k_x = \frac{1}{W_x}, \quad \Delta k_y = \frac{R}{W_y} = \frac{2}{W_y}. \tag{2.15}$$

If that is so, when a 2D IDFT is applied to each of the L k-space data sets, aliased versions of $\rho_l(x, y)$, $\bar{\rho}_l(x, y)$, are reconstructed (see subsection 2.3.1). Confined to the image support, $|x| > \frac{W_x}{2}$, $|y| > \frac{W_y}{2}$, only $R = 2$ aliased replicates appear, thus

$$\bar{\rho}_l(x, y) = \frac{1}{\Delta k_x \Delta k_y}\big(B^R_{xy_l}(x, y_-)M_{x'y'}(x, y_-) + B^R_{xy_l}(x, y)M_{x'y'}(x, y)+$$
$$B^R_{xy_l}(x, y_+)M_{x'y'}(x, y_+)\big), \quad l = 1, ..., L, \tag{2.16}$$

with $y_- = y - W_y/2$ and $y_+ = y + W_y/2$. If the coil sensitivities are known, Eq. (2.16) represents, for each point (x, y) in the image support, a linear system of L equations, being $M_{x'y'}(x, y_-)$, $M_{x'y'}(x, y)$, and $M_{x'y'}(x, y_+)$ the unknowns variables.

Observe that the two aliased image versions, $M_{x'y'}(x, y_-)$ and $M_{x'y'}(x, y_+)$, never overlap each other. Hence, it suffices to solve Eq. (2.16) for points x and y, with either $y < 0$ or $y > 0$. In the first case, the upper half part is recovered by $M_{x'y'}(x, y_+)$ whereas in the second case, the lower half part is obtained with $M_{x'y'}(x, y_-)$. As we want digital MR images, the 2D IDTFT is replaced in favor

of an IDFT, and because the number of k-space points in the y direction is also reduced by a factor of R, the FOV of the discrete reconstructed image is reduced by R (see last part of subsection 2.3.2). In the SENSE terminology, the name of reduced FOV is sometimes used [Pruessmann et al., 1999, Blaimer et al., 2004], however, this FOV reduction is inherent to the DFT, and has nothing to do with the real FOV defined by W_x and W_y. A graphical representation of the SENSE method is shown in Fig. 2.6.

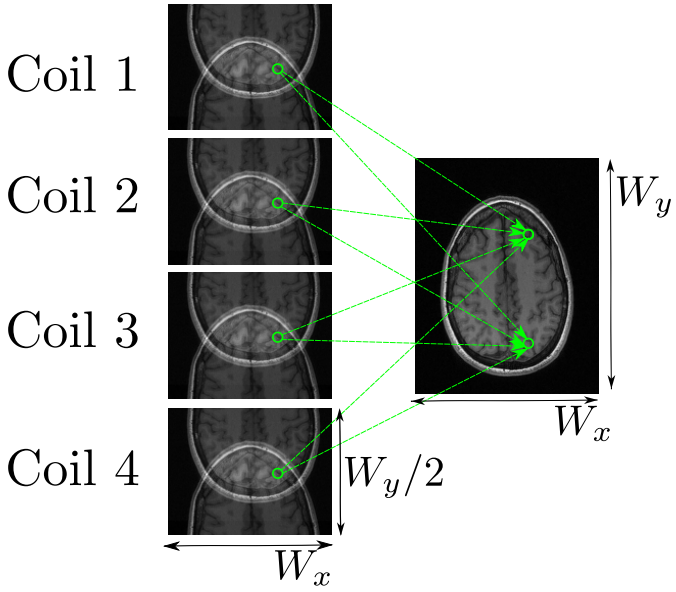


Figure 2.6: Schematic illustration of the SENSE method with $R = 2$ and $L = 4$ coils. Figure adapted from [Aja-Fernández and Vegas-Sánchez-Ferrero, 2016a].

Normally, current clinical MR scanners are equipped with $L = 32$ receiver coils, so the linear system of Eq. (2.16) is overdetermined for common acceleration factors of two or four [Blaimer et al., 2004]. In this situation, the least-square solution is calculated. Despite the required a priori knowledge of coil sensitivities, SENSE is, to date, the most widespread employed pMRI technique, and is offered by many companies in slightly modified implementations: Philips (SENSE), Siemens (mSENSE), General Electric (ASSET), Toshiba (SPEEDER) [Blaimer et al., 2004]. An important issue with SENSE is the required accurate estimation of the coil sensitivities profiles, which can be circumvented with GRAPPA since this technique works purely in the k-space domain.

2.5.2.2 GRAPPA

Three years after Pruessmann's contribution, Mark Griswold and colleagues [Griswold et al., 2002] presented the Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) method. GRAPPA, unlike SENSE, does not require knowledge of the coil sensitives and it works purely in the k-space domain. Though

initially conceived for Cartesian k-space data, extensions to non-Cartesian k-space data have also been proposed since then. GRAPPA reconstructs the missing k-space lines with information derived from the acquired k-space lines that are adjacent to the missing line. In particular, a missing k-space line in the l'-th coil is reconstructed by interpolating the k-space lines from all the L coils (Fig. 2.7). Mathematically, the missing k-space data point $s_{l'}(m', n')$ is approximated by

$$\hat{s}_{l'}(m', n') = \sum_{l=1}^{L} \sum_{m=-M/2}^{M/2-1} \sum_{n=\{-1,1\}} s_{l'}(m' - m, n' - n)\omega_{ll'}(m, n), \tag{2.17}$$

where $\omega_{ll'}(\cdot, \cdot)$ is the GRAPPA kernel which defines how the missing k-space data in the l'-th coil is interpolated from the acquired k-space data in the l-th coil. The GRAPPA kernel is learned from the so-called Auto Calibration Signal (ACS) lines, which conform the low-frequency spectrum of the k-space data of every coil [Griswold et al., 2002].

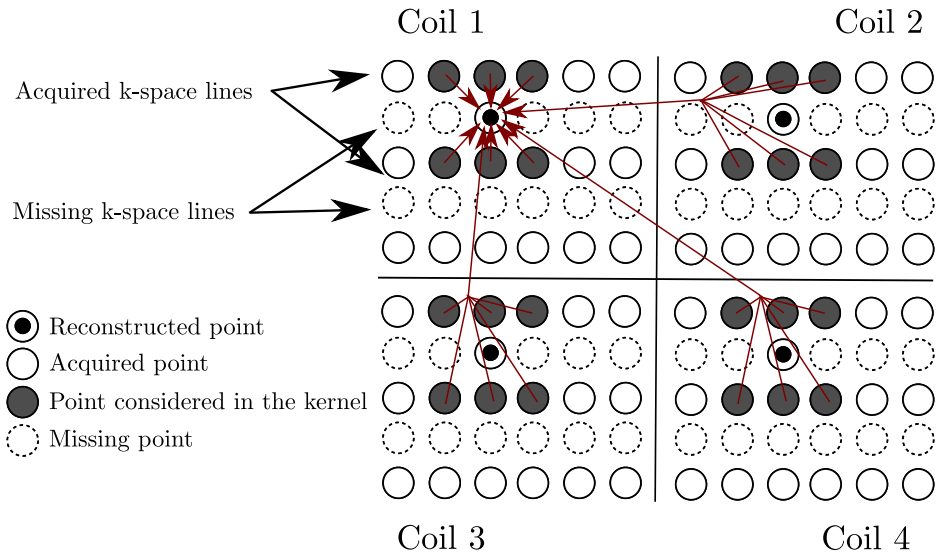


Figure 2.7: Schematic illustration of the GRAPPA method with $R = 2$ and $L = 4$ coils. Figure adapted from [Aja-Fernández and Vegas-Sánchez-Ferrero, 2016a].

Once the missing k-space lines have been filled up, an IDFT gives the reconstructed images $\hat{\rho}_l(x, y)$. The next pending question is how to form the composite image from $\hat{\rho}_l(x, y)$, aiming at estimating $M_{x'y'}(x, y)$. Several techniques may be applied. We briefly describe the most popular.

1. **Sum of Squares (SoS)**
   With the SoS technique [Roemer et al., 1990], the following composite image $S_{\text{comp}}(x, y)$ is created:

$$S_{\text{comp}}(x, y) = \sqrt{\sum_{l=1}^{L} |\hat{\rho}_l(x, y)|^2}. \tag{2.18}$$

A direct consequence of Eq. (2.18) is that $S_{\text{comp}}(x,y)$ is always a real and positive image. Note that, provided $\hat{\rho}_l(x,y)$ are good approximations of $\rho_l(x,y)$, with $l = 1, ..., L$, then

$$S_{\text{comp}}(x,y) = \sqrt{\sum_{l=1}^{L} |\hat{\rho}_l(x,y)|^2} \approx |M_{x'y'}(x,y)| \sqrt{\sum_{l=1}^{L} |B^R_{xy_l}(x,y)|^2}. \quad (2.19)$$

Hence, if the coil sensitivities are designed to fulfill

$$\sum_{l=1}^{L} |B^R_{xy_l}(x,y)|^2 = C \text{ for any } C > 0, \quad (2.20)$$

$S_{\text{comp}}(x,y)$ is simply a scaled version of the magnitude of the image $M_{x'y'}(x,y)$.

2. **Spatially Matched Filter (SMF)**
With the SMF technique [Roemer et al., 1990], the composite image $S_{\text{comp}}(x,y)$ is reconstructed as follows:

$$S_{\text{comp}}(x,y) = \sum_{l=1}^{L} m_l(x,y)\hat{\rho}_l(x,y), \quad (2.21)$$

where $m_l(x,y)$ is the l-th coefficient of the spatial filter at point (x,y). The optimal filter is derived from the array noise correlation matrix [Roemer et al., 1990, Walsh et al., 2000]. The SMF is known to be a computationally expensive technique. Several variations have been given where the coefficients of the filter are adaptively estimated [Walsh et al., 2000]. The name adaptive combine method is also given [Walsh et al., 2000].

2.6 Noise in MRI

In 1928, John Johnson demonstrated that thermal motion of electrons in a resistor R induces random fluctuation in the voltage across the resistor. The effective resistance is the sum of the coil resistance R_c and the resistance induced by the conductive losses in the scanned subject R_s [Hoult and Lauterbur, 1979]. Generally, the latter is the dominant source of noise. These voltage fluctuations are also known as thermal noise. Such value fluctuations are well-modeled as realizations of a Gaussian random variable with zero mean and variance

$$\sigma^2 = 4K_b T(R_c + R_s)\Delta f, \quad (2.22)$$

where K_b is the Boltzmann constant, T is the temperature, and Δf is the receiver bandwidth [Veraart, 2013, Aja-Fernández and Vegas-Sánchez-Ferrero, 2016a]. The existence of thermal noise has an important influence on the extraction of quantitative parameters in MRI. Indeed, the MR raw signal model of Eq. (1.25) is not valid anymore. Instead, a modified model can be considered by adding a zero mean white Gaussian noise $e(t)$,

$$\tilde{v}(t) = v(t) + e(t), \quad (2.23)$$

with $v(t)$ the voltage signal of Eq. (1.25), and $\tilde{v}(t)$, for each time point t, a realization of a Gaussian random variable with mean $v(t)$ and standard deviation σ (Eq. (2.22)). As a result, the complex MR signal, that is, $s(t)$ in Eq. (1.36), is also polluted by noise and therefore, we now have to deal with a noisy model for the k-space data sets acquired in each coil $s_l(\boldsymbol{k}_p)$, $p \in \mathbb{Z}$:

$$\tilde{s}_l(\boldsymbol{k}_p) = s_l(\boldsymbol{k}_p) + w_l(\boldsymbol{k}_p), \quad l = 1, ..., L, \tag{2.24}$$

with $w_l(\boldsymbol{k}_p)$ complex additive noise processes.

It is obvious then that the MR reconstructed images are always polluted by noise, that is, the intensities of MR images cannot be modeled as deterministic values anymore but as random variables. Furthermore, all the reconstruction techniques that were presented as "exact" are not so anymore, and can even amplify the noise effect when undersampled k-space data are in play. Therefore, a study of the effect of noise on the reconstruction of an MR image is fundamental for image quality assurance and particularly critical for qMRI estimation (see chapter 4).

Paramount to the characterization of a random variable and hence to quantitative MR estimation is the concept of the probability density function (PDF). In order to derive the PDF of $\tilde{s}_l(\boldsymbol{k}_p)$, we have to make assumptions of the statistics of $w_l(\boldsymbol{k}_p)$. More problematic is to take into account all the mathematical operations that are carried out by the reconstruction method of choice. A deeper, yet comprehensive analysis of the statistical distribution models can be found in the references [den Dekker and Sijbers, 2014, Aja-Fernández and Vegas-Sánchez-Ferrero, 2016a]. Here, we follow a more modest approach. Instead of focusing on a specific method, we derive the PDFs of the particular random variables that result from typical mathematical operations on Gaussian random variables. Then, for each case, we pinpoint to which actual scanning conditions those PDFs may apply.

2.6.1 Statistical assumptions of MRI noise

After I/Q demodulation, the resulting noise processes can accurately be described as a band-pass zero mean white Gaussian noise process, having a uniform power spectral density function symmetric about ω_{RF} [den Dekker and Sijbers, 2014]. An assumption that is of fundamental importance to continue is that temporal sampling of the signal $s_l(t)$, results in $w_l(\boldsymbol{k}_p)$, with $l = 1, ..., L$, being zero-mean white complex Gaussian processes which means their standard deviation does not depend on p [den Dekker and Sijbers, 2014]. Since, for each l, $w_l(\boldsymbol{k}_p)$ is a white Gaussian process, $w_l(\boldsymbol{k}_p)$ is, by definition, an uncorrelated process [Papoulis, 1977]. However, two different processes from different coils, let us say, from the l and l' coil, could be mutually correlated, that is, the random variables $w_l(\boldsymbol{k}_p)$ and $w_{l'}(\boldsymbol{k}_{p'})$ with $p \neq p'$ and with $l \neq l'$ can be correlated. The cause of this correlation is the existence of noise correlations in the phased-array system, [Brown et al., 2007, Hayes and Roemer, 1990]. For systems with a small number of coils, this effect is left aside, due to their minimal effect and practical considerations, as stated in [Constantinides et al., 1997]. However, for modern acquisition systems comprising

up to 32 or 64 coils, the receivers usually show a certain coupling [Aja-Fernández and Vegas-Sánchez-Ferrero, 2016a].

2.6.2 Common statistical data distributions

2.6.2.1 PDF of a Gaussian random variable

A random variable X is said to follow a Gaussian distribution with mean μ and standard deviation $\sigma > 0$ if its PDF $p_X(x;\mu,\sigma)$ has the following expression:

$$p_X(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{2.25}$$

With the assumptions made above, the real and imaginary part of each k-space data point $\tilde{s}_l(\boldsymbol{k}_p)$, for given p and l, follows a Gaussian distribution.

2.6.2.2 PDF of an N-dimensional Gaussian random variable

Let $X_1, X_2, ..., X_N$ be a collection of random variables. We say that the random vector $\boldsymbol{X} = (X_1, X_2, ..., X_N)^T$ follows a multi-variate Gaussian distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_N)^T$ and covariance matrix $\boldsymbol{C_X}$ $(N \times N)$ if its multidimensional PDF is given by

$$p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{C_X}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C_X}|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C_X}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \tag{2.26}$$

with $\boldsymbol{x} = (x_1, x_2, ..., x_N)^T$, and where $|\boldsymbol{C_X}|$ denotes the determinant of $\boldsymbol{C_X}$. The random variables $X_1, X_2, ..., X_N$ are also said to be jointly Gaussian distributed. The k-space data sets $\tilde{s}_l(\boldsymbol{k}_p)$ for $l = 1, ..., L$ and $\forall p \in \mathbb{Z}$ are jointly Gaussian distributed, with a covariance matrix depending on the coil correlations [Aja-Fernández and Vegas-Sánchez-Ferrero, 2016a].

2.6.2.3 Linear combinations of independent Gaussian random variables

Let $X_1, X_2, ..., X_N$ be a collection of independent Gaussian random variables each with mean μ_n, and standard deviation σ_n. Then, the random variable S defined as

$$S = \sum_{n=1}^{N} w_n X_n, \tag{2.27}$$

follows a Gaussian distribution with mean μ_S and σ_S given by

$$\mu_S = \sum_{n=1}^{N} w_n \mu_n, \quad \sigma_S = \sqrt{\sum_{n=1}^{N} w_n \sigma_n^2}. \tag{2.28}$$

This situation applies when **MR images $\bar{\rho}_l(x,y)$, $l = 1, ..., L$, are reconstructed independently with a 2D IDFT.** Observe that for each point (x, y), both the real and imaginary part of $\bar{\rho}_l(x,y)$ are linear combinations of independent Gaussian random variables, since the IDFT is a linear operator. Hence, $\bar{\rho}_l(x,y)$ is a complex Gaussian random variable whose mean vary with (x, y) but the standard deviation is constant.

2.6.2.4 Affine transformation of a Gaussian random vector

Let $\boldsymbol{X} = (X_1, X_2, ..., X_N)^T$ be a random vector that follows a multi-variate Gaussian distribution. If we form the following random vector

$$\boldsymbol{S} = \boldsymbol{c} + \boldsymbol{B}\boldsymbol{X}, \tag{2.29}$$

with $\boldsymbol{c}$ an $M \times 1$ vector and $\boldsymbol{B}$ an $M \times N$ matrix, then $\boldsymbol{S}$ $(M \times 1)$ follows a multivariate Gaussian distribution with

$$\boldsymbol{\mu_S} = \boldsymbol{c} + \boldsymbol{B}\boldsymbol{\mu}, \quad \boldsymbol{C_S} = \boldsymbol{B}\boldsymbol{C_X}\boldsymbol{B}^T. \tag{2.30}$$

This situation applies when the **MR image $M_{x'y'}(x,y)$ is reconstructed with SENSE**, since the linear-least-squares solution is an affine operation over the real and imaginary part of all the k-space data points, which are, by assumption, jointly Gaussian distributed. The covariance matrix $\boldsymbol{C_S}$ depends on the coil sensitivities as well as on the noise correlation matrix of the coils [Aja-Fernández and Vegas-Sánchez-Ferrero, 2016a]. For all points of the reconstructed image $M_{x'y'}(x,y)$, the intensity is a Gaussian random variable with spatially variant mean and standard deviation.

In addition, the multivariate Gaussian distribution also arises when the **the composite image, $S_{\mathbf{comp}}(x,y)$, is reconstructed with GRAPPA combined and SMF method**. Indeed, the convolution operator in GRAPPA can be formally considered as an affine transformation of the k-space data points. Therefore, individually, the intensity of each of the reconstructed coil images follow a Gaussian distribution, and all of them considered jointly together follow a multivariate Gaussian distribution. The SMF effectively combines all the information into one single image by means of an affine transformation. As a result, the intensity of the final composite image $M_{x'y'}(x,y)$ follows a spatially variant Gaussian distribution.

2.6.2.5 Magnitude of a complex independent Gaussian random variable

Let X_1 and X_2 be independent Gaussian random variables with mean μ_1 and μ_2, respectively, and identical standard deviation σ. Let S be the random variable obtained as

$$S = |X_1 + iX_2|, \tag{2.31}$$

where $|\cdot|$ the absolute value. The random variable S then follows a Rician distribution with signal parameter $\mu = |\mu_1 + i\mu_2|$, and scale parameter σ, where the PDF of S, $p_S(s;\mu,\sigma)$, has the formal expression

$$p_S(s;\mu,\sigma) = \frac{s}{\sigma^2} e^{\frac{-(s^2+\mu^2)}{2\sigma^2}} I(\frac{\mu s}{\sigma^2}) u(s), \tag{2.32}$$

with $I_0(\cdot)$ the zeroth order modified Bessel function of the first kind and where the unit step function $u(\cdot)$ is used to indicate that the expression for the PDF of S is valid for nonnegative values of s only. Note that for high signal-to-noise ratios $\frac{\mu}{\sigma} > 3$, the Rician PDF becomes quasi Gaussian [Sijbers et al., 1998c]. The Rician distribution constitutes the most ubiquitous probabilistic distribution in MRI since the magnitude operation is a very common operation in every MR image processing analysis where the phase information can be disregarded. The most trivial example where the Rician distribution arises is when the magnitude of each of the fully-sampled reconstructed image coils $\bar{\rho}_l(x,y)$ is taken, but also comes up naturally with undersampled k-space data. Indeed, when **the magnitude of the reconstructed image with SENSE is taken** and the real and imaginary part are independent, or when **the magnitude of the reconstructed image with GRAPPA+SMF is applied**, with the same assumption on the real and imaginary part, the intensity of the magnitude images will follow a spatially variant Rician distribution.

2.6.2.6 Sum of squares of independent Gaussian random variables

Let $X_1, X_2, ..., X_N$ be a collection of independent Gaussian random variables each one with mean μ_n and standard deviation σ_n. Then, the random variable S defined as

$$S = \sqrt{\sum_{n=1}^{N} X_n^2} \tag{2.33}$$

follows a non-central chi (χ) distribution with signal parameter

$$\mu = \sqrt{\sum_{n=1}^{N} \mu_n^2} \tag{2.34}$$

and scale parameter σ. The PDF of S, $p_S(s;\mu,\sigma)$, has the formal expression

$$p_S(s;\mu,\sigma) = \frac{\mu^{N/2-1} s}{\sigma^2} e^{\frac{-(s^2+\mu^2)}{2\sigma^2}} I_{N/2-1}(\frac{\mu s}{\sigma^2}) u(s), \tag{2.35}$$

with $I_{N/2-1}(\cdot)$ the $(N/2-1)$th order modified Bessel function of the first kind. Observe that the Rician distribution is a special case of the non-central chi (nc-χ) distribution for $N = 2$. The intensity of the **the composite image $S_{\mathbf{comp}}(x,y)$ in GRAPPA is constructed with the SoS method** follows a non-central chi (nc-χ) distribution.

2.6.3 Signal-to-noise ratio

Signal-to-noise ratio (SNR) is a measure commonly used in science and engineering. It compares the level of an expected/desired signal to the level of background noise. In its most basic formulation, it is defined as the ratio of the power of the signal (which is the information) and the power of the background noise (unwanted signal). Hence, a high SNR implies the level of background noise is quite small compared to the desired signal. Naturally, a high SNR is advantageous. Indeed, SNR can be used as a measure of quality control. It can be used as an optimization criterion for the design of the RF receptor coils, where the zero mean white Gaussian noise process $e(t)$ should be maintained low in comparison to the desired signal $v(t)$, see Eq. (2.23). While this is the most natural and rigorous way to define the SNR in MRI, it is more practical to provide quality measures that relate to the reconstructed MR image. An alternative SNR definition, which is widespread among the MRI community, is

$$\mathrm{SNR}(x,y) = \frac{I(x,y)}{\sigma_{\mathrm{Noise}}(x,y)}, \tag{2.36}$$

where $I(x,y)$ is a magnitude noise-free image, reconstructed with any method and evaluated at (x,y), and $\sigma_{\mathrm{Noise}}(x,y)$ the standard deviation of the "reconstructed" noise image at (x,y). In a very few cases, this SNR can be calculated analytically, for example in the case of fully-sampled coil by coil reconstruction. In practice, deriving the noise model of the final reconstructed image is often infeasible, and instead, this value is estimated with ad-hoc algorithms [Sijbers et al., 1998b, Sijbers et al., 2007, Aja-Fernández et al., 2008, Koay et al., 2009, Rajan et al., 2010, Maximov et al., 2012, Manjón et al., 2015, Aja-Fernández et al., 2015a, Poot and Klein, 2015, Veraart et al., 2016, Pieciak et al., 2017]. This constitutes the technique called *noise mapping*, which is out of the scope of this dissertation. We refer the reader to [Aja-Fernández and Vegas-Sánchez-Ferrero, 2016b] for a comprehensive review on noise mapping.

2.7 Algebraic reconstruction techniques

2.7.1 Linear Inverse Fourier problem

Let us focus on the single-coil reconstruction problem of Eq. (2.1), and assume that the continuous image $\rho(\boldsymbol{r})$ can be approximated with a *finite series expansion* as follows [Fessler, 2010]:

$$\rho(\boldsymbol{r}) = \sum_{n=1}^{N} x_n v(\boldsymbol{r} - \boldsymbol{r}_n), \tag{2.37}$$

where $v(\cdot)$ denotes the object basis function, $\boldsymbol{r}_n$ denotes the center of the nth translated basis function, and N is the number of parameters [Fessler, 2010]. Then,

by substituting Eq. (2.37) into Eq. (2.1) we arrive at

$$s(\boldsymbol{k}_p) = \sum_{n=1}^{N} a_{pn} x_n, \quad \forall p \in \mathbb{Z}, \tag{2.38}$$

with

$$a_{pn} = \int_{\Omega} v(\boldsymbol{r} - \boldsymbol{r}_n) e^{-2\pi i \langle \boldsymbol{k}_p, \boldsymbol{r} \rangle} d\boldsymbol{r}, \quad \forall p \in \mathbb{Z}, \quad n = 1, ..., N. \tag{2.39}$$

Commonly, the basis functions are highly localized (approximations of the Dirac delta in the limit), so it is very common to simplify Eq. (2.39) into

$$a_{pn} \approx e^{-2\pi i \langle \boldsymbol{k}_p, \boldsymbol{r}_n \rangle}, \quad \forall p \in \mathbb{Z}, \quad n = 1, ..., N. \tag{2.40}$$

If that is so, and if we further assume that the number of sampling points is finite, M then, by defining $\boldsymbol{y} = (s(\boldsymbol{k}_1), s(\boldsymbol{k}_2), ..., s(\boldsymbol{k}_M))^T$, $\boldsymbol{x} = (x_1, x_2, ..., x_N)^T$, and $\boldsymbol{A} = \{a_{m,n}\} \in \mathbb{C}^{M \times N}$ the so-called Fourier encoding matrix, where

$$a_{m,n} = e^{-2\pi i \langle \boldsymbol{k}_m, \boldsymbol{r}_n \rangle}, \tag{2.41}$$

the MR reconstruction problem (without noise) is often posed as an algebraic linear inverse problem

$$\begin{aligned} &\text{Given } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \\ &\text{determine } \boldsymbol{x}. \end{aligned} \tag{2.42}$$

Without further assumptions on the k-space data $\boldsymbol{y}$, this problem is not always solvable. If the k-space data are acquired in a Cartesian scheme, fulfilling the Nyquist conditions with M the total number of k-space points, and $\boldsymbol{r}_n$ samples of the FOV with $N = M$, then it can be shown that $\boldsymbol{A}$ is nothing more than a (scaled) DFT matrix (2D or 3D depending on whether $\boldsymbol{r} \in \mathbb{R}^2$ or $\boldsymbol{r} \in \mathbb{R}^3$). The inverse of $\boldsymbol{A}$ therefore exists, hence $\boldsymbol{x}$ can be calculated analytically, in fact with very efficient algorithms such as the IFFT.

When other non-Cartesian k-space schemes are used, this is no longer possible. In the interesting case of undersampling, $M << N$, and thus $\boldsymbol{A}$ is no longer invertible, since the problem is undetermined. Furthermore, the presence of noise makes that the equality $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ never holds. In this case, with undersampling k-space data and noise, the MR Image reconstruction problem can be cast as an optimization problem.

2.7.2 MR Image reconstruction as an optimization problem

Let us consider the more realistic MRI acquisition model [Fessler, 2010]

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \tag{2.43}$$

where $\boldsymbol{n} \in \mathbb{C}^N$ is white complex Gaussian noise. Such a model can easily mimic a multi-coil acquisition system by augmenting the matrix $\boldsymbol{A}$ with the coil sensitivities,

and also by considering that $\boldsymbol{y}$ now concatenates all k-space data that is acquired in the L coils. When necessary in this PhD thesis, we will differentiate between the encoding matrix $\boldsymbol{A}$ for one single-coil and for a multi-coil system (in chapter 7 for instance). In this chapter, however, we will use $\boldsymbol{A}$ indistinctly for both cases.

Naïvely, without further information, $\boldsymbol{x}$ is often reconstructed by solving the following linear least-squares (LLS) optimization problem:

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2. \tag{2.44}$$

Given the Gaussian noise statistics of $\boldsymbol{n}$, this formulation is equivalent to the Maximum Likelihood (ML) estimator [den Dekker and Sijbers, 2014], which we will further describe in chapter 4. The solution of Eq. (4.15) can sometimes be derived analytically as,

$$\boldsymbol{x} = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{y}, \tag{2.45}$$

where $\boldsymbol{A}^H$ is the Hermitian matrix of $\boldsymbol{A}$ and where we have assumed that $(\boldsymbol{A}^H\boldsymbol{A})$ is invertible. For the accelerated case (undersampled k-space data), the LS problem does not always provide a satisfactory solution since the problem is ill-conditioned [Fessler, 2010]. To circumvent this, the MR optimization problem often accommodates more complex cost functions than a simple l_2 norm as well as new constraints on the solution $\boldsymbol{x}$.

2.7.3 Phase-constraint formulation

A very powerful constraint on $\boldsymbol{x}$ is the so-called phase-constraint formulation [McGibney et al., 1993, Samsonov et al., 2004, Bydder and Robson, 2005, Samsonov et al., 2010, Blaimer et al., 2016]. Since $\boldsymbol{x}$ is a complex valued image, it can be written as

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{x}_+, \tag{2.46}$$

with $\boldsymbol{x}_+ \in \mathbb{R}_+^N$ the magnitude image, $\boldsymbol{\Psi} = \text{diag}(e^{i\psi_{\boldsymbol{x}}})$ being a diagonal matrix whose diagonal contains the entries of the vector $e^{i\psi_{\boldsymbol{x}}}$, and where $\psi_{\boldsymbol{x}} \in \mathbb{R}^N$ is the phase of $\boldsymbol{x}$. If the phase $\psi_{\boldsymbol{x}}$ is known, then the phase-constraint MR reconstruction problem is given by

$$\min_{\boldsymbol{x}_+\in\mathbb{R}_+^N} ||\boldsymbol{y} - \tilde{\boldsymbol{A}}\boldsymbol{x}_+||_2^2, \tag{2.47}$$

with $\tilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Psi}$. The required phase estimate, $\hat{\psi}_{\boldsymbol{x}}$, can be obtained from a low resolution recovered image in [Lustig et al., 2007] since normally the phase image varies slowly compared to magnitude images. Details are provided in subsection 7.3.3 of chapter 7.

2.7.4 Prior knowledge and regularization

In practically all situations, we have certain knowledge about the type of image we expect to recover. For example, we may expect that the target image is smooth

or that its Fourier spectrum obeys a particular type of pattern. There are infinite ways to model prior knowledge on $\boldsymbol{x}$, but practically all of them can be formally described by the so-called prior term, a mathematical function of the image $\boldsymbol{x}$, $\boldsymbol{\Phi}(\boldsymbol{x})$. The rationale behind the prior term is that the lower $\boldsymbol{\Phi}(\boldsymbol{x})$ is, the more $\boldsymbol{x}$ is in agreement with the prior knowledge we have.

If prior knowledge is incorporated, the reconstruction problem is often recast as a constrained optimization problem [Lustig et al., 2007, Figueiredo et al., 2007, Lustig et al., 2008] in the form of

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} \Phi(\boldsymbol{x}) \quad s.t. \quad ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 \leq \epsilon, \tag{2.48}$$

where the constraint on $\boldsymbol{x}$ is called the data fidelity condition. That is, we look for the solution $\boldsymbol{x}$ which adheres the most to our a priori knowledge, so effectively it minimizes $\Phi(\boldsymbol{x})$, but at the same time is a plausible solution to have generated the acquired k-space data $\boldsymbol{y}$ (data fidelity term). The value ϵ in the data fidelity condition is usually set below the expected noise level [Lustig et al., 2007].

Sometimes, prior knowledge is introduced in a different problem setting than that of Eq. (2.48). Instead of a constraint formulation, the following problem is also considered

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 + \lambda\Phi(\boldsymbol{x}), \tag{2.49}$$

where now, $\Phi(\boldsymbol{x})$ is often called a regularization term, and λ the corresponding regularization parameter. Such parameter controls the importance of the regularization compared to the data fidelity term. Both types of formulations, Eq. (2.48) and Eq. (2.49) could be shown to be equivalent by using tools from optimization theory. Such a link is established in subsection 7.3.1.

One of the main contributions of this PhD thesis is a novel prior knowledge for a specific type of MR images (see chapter 7). To put that work into context, we briefly describe below the most common types of prior knowledge or regularization terms in the MR image reconstruction field.

2.7.4.1 Quadratic regularization

One the oldest methods to incorporate regularization is through quadratic regularization, also known as Tikhonov regularization:

$$\Phi(\boldsymbol{x}) = ||\boldsymbol{R}(\boldsymbol{x} - \boldsymbol{x}_0)||_2^2, \tag{2.50}$$

where both the matrix $\boldsymbol{R}$ and the reference image $\boldsymbol{x}_0$ can be defined in multiple ways. If $\boldsymbol{R} = \boldsymbol{I}$, being $\boldsymbol{I}$ the identity matrix, and $\boldsymbol{x}_0 = \boldsymbol{0}$, we obtain the classical version of Tikhonov regularization. Observe that $\boldsymbol{x}$ is complex so the prior here is imposed on both the real and imaginary part. A reference image could be available from a previous MR scan or can be estimated iteratively. Another common option for $\boldsymbol{R}$ is the choice of a finite difference matrix [Fessler, 2010], which promotes a smooth image (if $\boldsymbol{x}_0 = \boldsymbol{0}$) where neighboring pixel values are similar. The resulting minimization problem is particularly well-behaving and simple. However, the choice of a finite difference matrix leads to smoothed image edges [Fessler, 2010].

2.7.4.2 Total Variation

Very popular in MR image reconstruction is the concept of Total Variation (TV). It was introduced in the image processing field in the nineties, by Rudin and Osher [Rudin et al., 1992]. Initially conceived for image denoising, TV has become a very powerful technique beyond simply image denoising, e.g., in image restoration, image inpainting, or image reconstruction. The TV of a continuous image $I(\boldsymbol{r})$ is defined as

$$I_{\mathrm{TV}} = \int ||\nabla I(\boldsymbol{r})||_1 d\boldsymbol{r}, \tag{2.51}$$

where $\nabla I(\boldsymbol{r})$ is the gradient of $I(\boldsymbol{r})$, and $||\cdot||_1$ denotes the l_1 norm. In practice, the integral is discretized, thereby giving a discrete TV in terms of absolute finite differences, which implements the gradient operator in a discrete setting. For example, in that discrete setting, the following prior is normally used:

$$\Phi(\boldsymbol{x}) = ||\boldsymbol{R}\boldsymbol{x}||_1, \tag{2.52}$$

where $\boldsymbol{R}$ is a finite difference matrix [Fessler, 2010]. Observe that the main fundamental difference of a discrete TV compared to a quadratic finite differences-based prior is just the use of the l_1 norm. In contrast to a quadratic regularizer, TV preserves edges much more accurate. It can be shown that minimizing the TV of an image leads to piece-wise constant images, with the appealing property that edges are sharply defined, and noise-type variations, are substantially removed. TV, however, can produce MR images with the undesirable appearance of cartoon-type images, also known, as staircase effect. To avoid such an effect, generalizations of TV have been proposed, for example, high order TV, where the gradient is replaced by differential operators of high order. Another interesting extension of TV is the so-called non-local TV, where the gradient in Eq. (2.51) is replaced by the non-local gradient. Non-local TV is described more in-depth in chapter 7 (subsection 7.3.5).

2.7.4.3 Sparsity

Many signals in nature are *compressible*. This means that they can be represented in a different domain quite accurately with much fewer coefficients than the dimension of the signal. Mathematically, if $\boldsymbol{P} \in \mathbb{C}^{D\times N}$ is the matrix that represents $\boldsymbol{x}$ in a given domain with coefficients $\boldsymbol{\theta} = \boldsymbol{P}\boldsymbol{x}$, then $\boldsymbol{x}$ is said to be compressible in the domain $\boldsymbol{P}$ if the vector $\boldsymbol{\theta}$ contains very few non-zero coefficients, being that number much smaller than D. A vector which contains a very few number of non-zero coefficients is informally called a *sparse* vector. Sparsity is intimately related to the l_0 pseudo-norm. The l_0 pseudo-norm is defined as the cardinal of the support of $\boldsymbol{x}$, that is, the number of non-zero elements. The lower l_0 is, the sparser the vector is.

That said, if $\boldsymbol{x}$ is compressible, $\boldsymbol{\theta}$ is sparse, and the prior term $\Phi(\boldsymbol{x})$ can be defined as

$$\Phi(\boldsymbol{x}) = ||\boldsymbol{\theta}||_0 = ||\boldsymbol{P}\boldsymbol{x}||_0. \tag{2.53}$$

Solving the resulting optimization problem is NP-hard (nondeterministic polynomial time - hard) in general. The NP concept plays a central role in computational complexity. The class of NP problems is formed by those computational decision problems whose solution can be given in a non-deterministic polynomial time. An optimization problem is NP-hard if it is at least as hard as the hardest problems in the NP class. In plain words, an NP-hard problem is, in practice, a computationally infeasible problem.

To alleviate the computational complexity, the l_0 pseudo-norm is replaced by the l_1 norm, giving a tractable optimization problem with very good results:

$$\Phi(\boldsymbol{x}) = ||\boldsymbol{\theta}||_1 = ||\boldsymbol{P}\boldsymbol{x}||_1. \tag{2.54}$$

Sparsity has been used massively in MR reconstruction with very satisfactory results in different image representations [Lustig et al., 2007] (e.g., Fourier [Lingala et al., 2011], Wavelet [Chen and Huang, 2012], [Lai et al., 2016], Curvelet [Candes et al., 2006], Shearlet [Aelterman et al., 2011], or redundant dictionaries [Ravishankar and Bresler, 2011, Caballero et al., 2014]). Generalizations of sparsity have also been given, for example, structural sparsity, where the sparse coefficients in a given representation domain adhere to certain structural patterns [Pizurica et al., 2011, Chen and Huang, 2014, Panic et al., 2017].

Undoubtedly the notion of sparsity is linked to the concept of Compressed-Sensing (CS). The theory behind CS is rather broad and by no means is covered here. A very extensive mathematical introduction can be found in [Foucart and Rauhut, 2010]. Briefly, the CS theory states that, if the encoding matrix $\boldsymbol{A}$ (see Eq. (2.42)) obeys the so-called Restricted Isometry Property (RIP) for all S-sparse vectors, (which are those with at most S non-zero entries), and if the unknown vector that solves Eq. (2.42), $\boldsymbol{x}^*$, is S-sparse, exact recovery of $\boldsymbol{x}^*$ is possible by solving an l_1 minimization problem [Candès and Wakin, 2008]. That is, the solution of the problem

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} ||\boldsymbol{x}||_1 \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{2.55}$$

is $\boldsymbol{x}^*$. For the noisy case model, (Eq. (2.43)), the l_2 norm error between the solution of the problem

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} ||\boldsymbol{x}||_1 \quad s.t. \quad ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 \leq \epsilon, \tag{2.56}$$

and the true vector $\boldsymbol{x}^*$, is below $C_1\epsilon$ where C_1 is a given constant [Candès and Wakin, 2008]. Similar results can be given when vectors are not sparse in the canonical basis but in the transformed domain defined by $\boldsymbol{P}$.

Besides sparsity, the RIP condition is critical for CS. The RIP condition is related to the so-called concept of incoherent sampling. The more incoherent is the sampling, the more likely the RIP condition holds. For further details about incoherent sampling, the reader is referred to [Candès and Wakin, 2008]. Before ending this subsection, it should be noted that though sparsity is an unavoidable condition for CS, sparsity-based priors could be and in fact have been used in different MR reconstruction problems where incoherence sampling is not required.

2.7.4.4 Low-rank modeling

Low-rank modeling is probably the most recent type of prior knowledge in MR image reconstruction. In many cases, the unknown image $\boldsymbol{x}$ can be mapped to a matrix whose rank is much smaller than the number of columns. Indeed, it has been recently demonstrated that the local k-space neighborhoods of a fully-sampled image $\boldsymbol{x}$ can be mapped to a matrix which is rank deficient [Haldar, 2014a, Haldar and Zhuo, 2015, Kim and Haldar, 2015, Kim et al., 2017b]. In addition, it has been shown that sparsity in a given domain results in rank-deficient Hankel structured matrices, which can be exploited in highly undersampled k-space data scenarios [Lee et al., 2016, Jin et al., 2016]. With all of these methods, the following prior is defined:

$$\Phi(\boldsymbol{x}) = \text{rank}(\boldsymbol{H}(\boldsymbol{x})), \tag{2.57}$$

where $\boldsymbol{H}(\boldsymbol{x})$ is the matrix that is expected to have low rank. A drawback of low-rank MR image reconstruction is that the resulting cost functions are non-convex and NP-hard to solve in general. To circumvent this, other priors that encourage rank-deficient matrices are used instead. Examples include the Schatten norm or the nuclear norm of $\boldsymbol{H}(\boldsymbol{x})$, and modifications of those [Haldar, 2014a]. Low-rank modeling has shown excellent results in MR reconstruction, and the assumptions the technique relies on are often much more relaxed than those of sparsity and TV. However, the main drawback of low-rank MR reconstruction is the complicated optimization problems that result when priors like Eq. (2.57) are employed, or simplified versions using the types of norm described above.

Part II

MRI Relaxometry

3

MRI Relaxometry: the basics

Contents

3.1 Introduction

This chapter is devoted to presenting the basics of MRI relaxometry, with special emphasis on T_1 mapping. First, the clinical applications of T_1 and T_2 relaxometry are described. Next, the most popular pulse sequences for T_1 mapping are covered. The relaxation signal models that result from those pulse sequences are also explained. At the end of the chapter, we give a short list of current bottlenecks/issues in MRI relaxometry, such as the most frequent sources of errors, the long acquisition time, and the specific absorption rate (SAR) limit.

3.2 Clinical applications of T_1 and T_2

In subsection 1.3.5, the physical basis of the spin-lattice relaxation time, T_1, and spin-spin relaxation time, T_2, were briefly described. As the fundamental causes of the relaxation process differ between types of matter, both the T_1 and T_2 values vary among tissues. Indeed, T_1 depends on the mobility of the lattice, which could be very different both across and within different tissues. For instance, the T_1 value in lipids is normally shorter than in other tissues since lipids molecules are less mobile [Van Steenkiste, 2016]. In general, as the T_1 value is related to the macromolecule concentration, water binding, and water content, biological tissues generally have distinct T_1 relaxation times. Furthermore, the same tissue in a different state (diseased or healthy) could lead to a different T_1 value.

3.2.1 Typical T_1 and T_2 values in the human body

In this section, we present measured T_1 and T_2 values (Tables 3.1 3.2 respectively) in different healthy organs and tissues of the human body. It should be noted that T_1 and T_2 depend as well on the strength of the $\boldsymbol{B_0}$ field since $\boldsymbol{B_0}$ is related to the Larmor frequency. Values obtained with a magnetic field strength of 3T are reported here. For a more complete list of measured T_1 and T_2 relaxation times, the reader is referred to [Bojorquez et al., 2017].

3.2.2 Human Brain pathologies characterization

So far, most of the clinical applications of MR relaxometry have focused on the characterization of pathologies in the human brain. Here we briefly describe the main findings when relaxometry is applied to the most common human brain diseases, and its relation to the reported T_1 and T_2 values. The description that we give below is by no means exhaustive. The interested reader is referred to [Tofts, 2004], and especially to the specific references that are listed for each type of disease.

Table 3.1: Reported T_1 at 3 T in healthy tissues and organs of the human body. The notation mean ± standard deviation (when provided) is used.

Tissue	T_1 [ms]
White matter	832 ± 10 [Wansapura et al., 1999]
Grey matter	1331 ± 13 [Wansapura et al., 1999]
Cerebrospinal fluid	4391 ± 545 [Shin et al., 2009]
Fat	450 ± 26 [Rakow-Penner et al., 2006]
Liver	677.5 ± 44.6 [Heye et al., 2012]
Myocardium	1052 ± 23 [Kim et al., 2017a]
Prostate	1530 ± 498 [Fennessy et al., 2012]
Bone Marrow	586 ± 73 [De Bazelaire et al., 2004]
Spleen	1328 ± 31 [De Bazelaire et al., 2004]
Blood	1932 ± 85 [Stanisz et al., 2005]
Muscle (knee)	1420 ± 38 [Han et al., 2003]
Cartilage (knee)	1240 ± 107 [Han et al., 2003]
Synovial fluid (knee)	3620 ± 320 [Han et al., 2003]

Table 3.2: Reported T_2 at 3 T in healthy tissues and organs of the human body. The notation mean ± standard deviation (when provided) is used.

Tissue	T_2 [ms]
White matter	79.6 ± 0.6 [Wansapura et al., 1999]
Grey matter	110 ± 2 [Wansapura et al., 1999]
Cerebrospinal fluid	2500 [Smith et al., 2008]
Fat	68 ± 4 [De Bazelaire et al., 2004]
Liver	71.7 ± 5.9 [Heye et al., 2012]
Myocardium	45.1 [von Knobelsdorff-Brenkenhoff et al., 2013]
Prostate	80 ± 34 [Bojorquez et al., 2016]
Bone Marrow	49 ± 4 [De Bazelaire et al., 2004]
Spleen	60± 19 [De Bazelaire et al., 2004]
Blood	275 ± 50 [Stanisz et al., 2005]
Muscle (knee)	31.7 ± 1.9 [Han et al., 2003]
Cartilage (knee)	36.9 ± 3.8 [Han et al., 2003]
Synovial fluid (knee)	767 ± 49 [Han et al., 2003]

3.2.2.1 Multiple Sclerosis

T_2-weighted qualitative MRI is the standard choice for the diagnosis of Multiple Sclerosis (MS) disease. However, the increase in T_2-weighted signal does not specifically depend on the pathology. In contrast, several studies have demonstrated that prolonged T_1 and T_2 relaxation times can be found in MS patients [Lacomis et al., 1986,Larsson et al., 1988,Miller et al., 1989,Barbosa et al., 1994] in comparison to those values in normal appearing white matter. Some of the variation of relaxation time within lesions is caused by the age of MS plaques [Ormerod et al., 1987]. The observed lesions are also quite heterogeneous. For example, those with axonal loss expand extracellular spaces, and hence the relaxation times associated with

those specific lesions are rather different than those related to lesions that are predominantly cellular with gliosis.

3.2.2.2 Intracranial Tumors

Several studies have demonstrated that T_1 values within tumors are significantly longer than normal white matter values [Englund et al., 1986, Just and Thelen, 1988, Kurki and Komu, 1995]. Although differentiation between tumors cannot be assessed by only probing T_1, important differences have been found: the shortest T_1 values are provided by gliomas and pituitary tumors, where the longest are reported in glioblastomas [Naruse et al., 1986].

3.2.2.3 Epilepsy

The standard choice for epilepsy studies is computed tomography (CT) and qualitative MRI. Nevertheless, MRI relaxometry has proved to be useful in the detection of subtle abnormalities [Tofts, 2004]. It has been reported that patients with epilepsy possess higher T_1 relaxation times than those in normal controls [Conlon et al., 1988]. Statistical significance has been found in the temporal lobes [Conlon et al., 1988]. Higher T_1 values were also found in the hemisphere containing the seizures focus. On the other hand, the T_2 relaxation time in the hippocampal region has proved to be an objective biomarker assessing the severity of signal abnormalities in patients with temporal lobe epilepsy [Jackson et al., 1993, Woermann et al., 1998, Namer et al., 1998, Okujava et al., 2002].

3.2.2.4 Stroke

With post-mortem MRI studies, it has been demonstrated that both T_1 and T_2 are consistently high within ischemic areas [DeWitt et al., 1987]. The detection of haemorrhagic change within infarcts has also benefited from probing the T_1 relaxation time. Finally, in [Lansberg et al., 2001], and related to the detection of strokes (within 24 h), it was shown that ROI-based analysis of T_2 spin-lattice relaxation times provides results more sensitive than those obtained by qualitative analysis.

3.2.2.5 Dementia

Studies have demonstrated that T_1 in the white and grey matter is prolonged in patients with Alzheimer's disease and especially in multi-infarct dementia, compared with normal controls [Besson et al., 1985, Erkinjuntti et al., 1987, Ebmeier et al., 1987]. It should be noted, however, that using T_1 quantification alone is not enough to distinguish between patient groups.

Besides, T_2 relaxometry has been performed in the hippocampus, showing an increase of the T_2 value compared to control subjects. Furthermore, the T_2 values

correlated with the severity of functional and cognitive impairment. Moreover, increased iron content in the substantia nigra has been reported in studies of patients with Parkinson's disease. Iron accumulation results in shortening of T_2 relaxation times. Several studies have shown shorter T_2 values in the basal ganglia [Ye et al., 1996]. Finally, it has been reported that T_2 relaxation times are shorter in the putamen and globus pallidus in Parkinson's-plus syndromes than in healthy subjects or subjects with classical Parkinson's disease [Drayer et al., 1986, Martin et al., 1998, Vymazal et al., 1999].

3.2.3 Other applications beyond the Human Brain

With the exception of the human brain, the heart is the most popular organ where MRI relaxometry has shown its potential. Recently, advancement has been made in T_1 relaxometry of the heart, allowing to characterize infiltrative and diffuse myocardial diseases [Li et al., 2012b, Xue et al., 2013, Kellman and Hansen, 2014, Jellis and Kwon, 2014]. T_1 relaxometry has also been applied to monitoring changes in contrast agent concentration [Caravan et al., 1999, Caravan et al., 2009]. T_1 relaxometry is also often used in perfusion studies [Peeters et al., 2004, Kershaw and Buckley, 2006]. Another application of T_1 relaxometry is the analysis of the human knee cartilage [Bron et al., 2013].

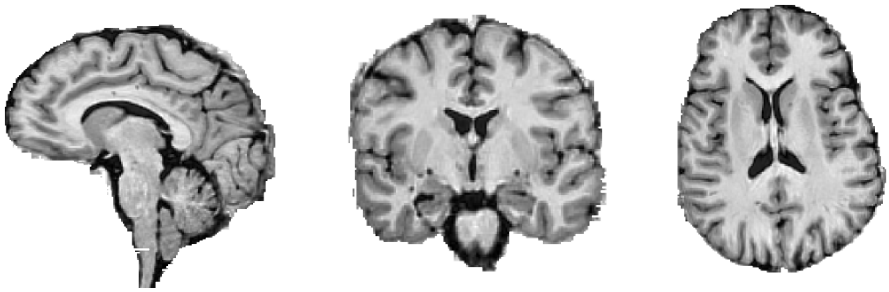
3.3 T_1 mapping: the ingredients

In this section, we will cover the basics of T_1 mapping, since most of the work in this PhD thesis has been focused on T_1 MRI relaxometry.

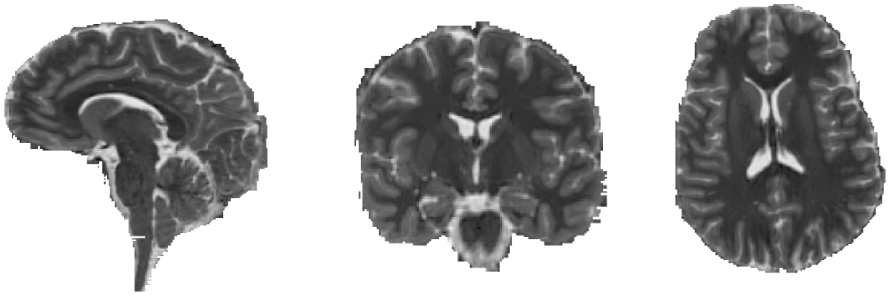
The goal of quantitative MR T_1 mapping is to probe the spin-lattice relaxation time T_1 at different spatial positions of the scanned object, thereby creating a spatial map of T_1. To that end, T_1 mapping involves three consecutive processes. First, the choice of a specific pulse sequence is needed. The applied pulse sequence disturbs the net nuclear magnetization vector from the equilibrium, and leads the spin ensemble system to the relaxation phase (see chapter 1). Different pulse sequences create different ways in which the longitudinal magnetization recovers, leading to different relaxation models. Those models always depend on T_1, but also on user-defined parameters, such as timing points or flip angles. The next step is to form the so-called T_1-weighted image. Indeed, since the T_1 should be probed at different spatial positions, spatial-encoding is necessary. Hence, the k-space data are acquired, and MR images are reconstructed with techniques described in chapter 2. Those images are acquired with different timing points or flip angles so as to "sample" the prescribed relaxation model, given by the pulse sequence at hand. This dataset of MR images is called the T_1-weighted data set (see Fig. 3.1). Finally, from this set of T_1-weighted images, T_1 is inferred by voxel-wise fitting the prescribed relaxation model to the intensity of the images.

In the succeeding section, we briefly review the most common pulse sequences that are used in T_1 mapping as well as the most popular T_1 relaxation signal models. It

is assumed that T_1-weighted images are reconstructed with the k-space techniques explained in the previous chapter, therefore no further discussion is given about this topic.



(a) T_1-weighted image (Sagittal-Coronal-Axial view)



(a) T_1 map (Sagittal-Coronal-Axial view)

Figure 3.1: T_1-weighted image of the normal human brain (a) and the corresponding T_1 map (b). Figure obtained with permission from [Van Steenkiste, 2016].

3.4 Typical pulse sequences for T_1 mapping

There are multiple ways to perform T_1 mapping [Stikov et al., 2015]. In this section, we cover the two most popular techniques, which in fact have been used in the work presented in chapters 5 and 6. Those are the Inversion Recovery (IR) and Variable Flip Angle (VFA) sequences, respectively.

3.4.1 Inversion Recovery (IR)

The gold standard pulse sequence to probe the T_1 value is the Inversion Recovery (IR) sequence [Stikov et al., 2015]. It dates back to the late 1940s, when initial NMR experiments were conducted [Drain, 1949, Hahn, 1949]. With this method, an inversion pulse, normally $\alpha_z = 180°$ pulse, flips the net nuclear magnetization vector from the z-axis to the $-z$-axis. Next, the system returns to the equilibrium obeying the Bloch relaxation equations (Eq. (1.20)). Since the MR receptor is only

sensitive to the transversal magnetization component, the longitudinal component is tipped in the transverse plane with an $\alpha_{x'} = 90°$ pulse. The time period between the 180° pulse and the 90° pulse is denoted as TI, which stands for Inversion Time. At t = TI, the longitudinal magnetization vector, $M_z(t)$, takes the expression of:

$$M_z(\mathrm{TI}) = M_{z'}(\mathrm{TI}) = M_z(0)(1 - e^{-\frac{\mathrm{TI}}{T_1}}) + M_{z'}^{t=0_+} e^{-\frac{\mathrm{TI}}{T_1}}. \tag{3.1}$$

With the assumption that $M_{z'}^{t=0_+} = -M_z(0)$, we obtain the popular formula in IR sequences:

$$M_z(\mathrm{TI}) = M_{z'}(\mathrm{TI}) = M_z(0)(1 - 2e^{-\frac{\mathrm{TI}}{T_1}}). \tag{3.2}$$

In Fig. 3.2, a schematic diagram of a typical IR sequence is shown, whereas the effect of the inversion recovery sequence on the longitudinal magnetization vector $M_z(t)$ is depicted in Fig. 3.3.

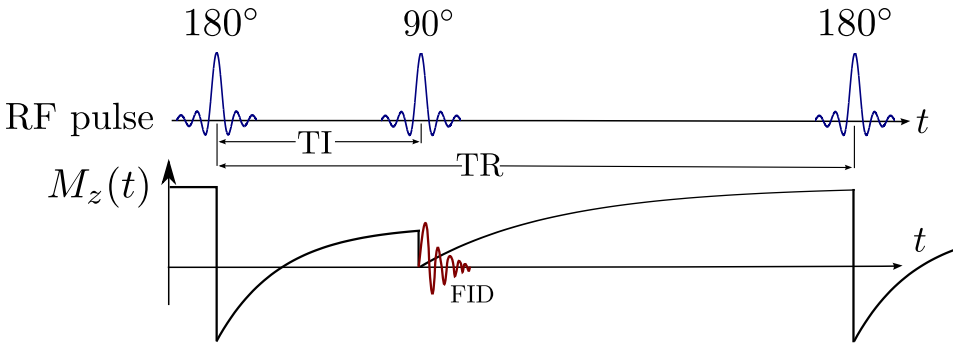


Figure 3.2: Inversion Recovery sequence. The longitudinal net nuclear magnetization vector is flipped by a 180° pulse. After an inversion time, TI, the longitudinal component is tipped into the transverse plane with a 90° pulse, thereby creating an FID signal.

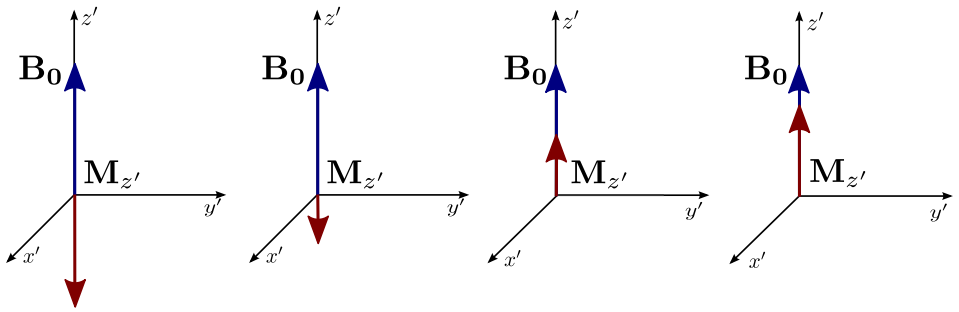


Figure 3.3: Effect of the inversion recovery sequence on the net nuclear longitudinal magnetization vector as seen in the RF-rotating frame.

The IR module, that is, the sequence of pulses shown in Fig. 3.3, is accompanied by an image module, or readout module, where the k-space data are acquired. Most popular readout modules are those based on a multi-shot type of acquisition, where

one or several k-space lines are acquired per TR, though single-shot IR sequences also exist [Ordidge et al., 1990, Clare and Jezzard, 2001]. These image readouts are most often based on elementary SE and GRE readout, which were described in subsection 1.3.6.

Probably, the most popular IR readout is the TSE sequence. With IR-TSE [Weigel et al., 2007], several numbers of phase encoding lines are acquired per TR (see Fig. 2.4). The number of phase encoding lines per TR, which in subsection 2.3.3 was denoted by $N_{y/\mathrm{TR}}$, is also called echo train length (ETL) or turbo factor (TF). The ETL cannot be arbitrarily increased since, apart from the T_2 decay already commented (see subsection 2.3.3), specific absorption rate (SAR) limits are undesirably reached [Weigel et al., 2007] (see subsection 3.6.2 for more details). For an extended coverage of IR readouts, the interested reader is referred to [Van Steenkiste, 2016] and especially [Stikov et al., 2015].

3.4.1.1 2D multi-slice and 3D IR sequences

IR T_1 mapping is often performed with 2D multi-slice sequences, and slightly less common, with 3D sequences as well.

In a 2D multi-slice IR sequence, while the spins of the first slice are reaching the equilibrium, additional slices are excited. Several 2D multi-slice schemes can be considered, which in turn determines the number of slices per TR that can be acquired. A more in-depth description of the type of multi-slice IR sequences can be found in [Van Steenkiste, 2016, Bernstein et al., 2004]. Here, we briefly summarize three of them: sequential, interleaved and distributed IR sequence.

In a **sequential IR acquisition**, the slices are arranged consecutively in an IR-module (see Fig. 3.4.(a)) in an odd-even type ordering. That is, immediately after the signal of the first slice have been measured in the corresponding readout module, the third slice is probed with a selective 180° pulse followed by a 90° pulse. After the readout module of the third slice, the fifth slice is probed and so on. When all odd slices have been probed, the process is repeated for even slices.

In a sequential IR acquisition, the number of slices that can be accommodated within one TR is severely restricted. In general, this type of acquisition is only efficient if TI is very short. With an **interleaved IR acquisition**, the time between the 180° pulse and the reception of the echo is used to play out IR modules for other slices, see Fig. 3.4.(b). The efficiency of an **interleaved IR acquisition** is substantially higher than that of a sequential IR acquisition.

In a **distributed IR acquisition**, the idle time that is yet present in an interleaved IR acquisition is used to play out the inversion module or the read-out sequences for slices that may not be covered during the same TR (see Fig. 3.4.(c)). While the most efficient approach, a drawback of this approach is that the TR, TI, and the number of slices are coupled together, and this limits the flexibility of choices for TR and TI [Listerud et al., 1996].

While less popular, there exist **3D IR sequences** [Szumowski et al., 2012,

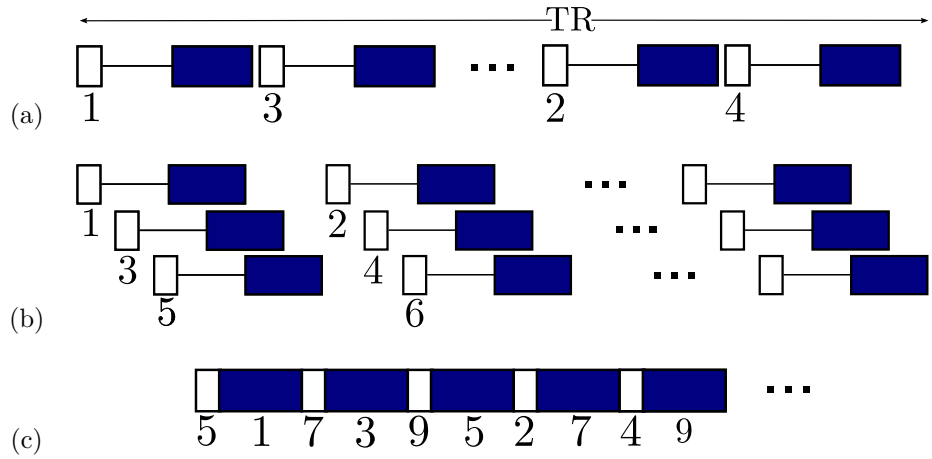


Figure 3.4: Acquisitions schemes for multi-slice IR sequences: (a) sequential IR acquisition, (b) interleaved IR acquisition, and (c) distributed IR acquisition. The white box represents the IR module, whereas the blue box represents the readout module. The solid line between boxes indicates the idle time, and the number denotes the slice index. Figure based on [Bernstein et al., 2004, Van Steenkiste, 2016].

Jablonowski et al., 2013, Hodel et al., 2014] which result in higher SNR, and avoid the partial volume effects caused by the slice gap in 2D multi-slice IR sequences. Nevertheless, the latter are yet faster than 3D IR sequences [McKenzie et al., 2006].

3.4.2 Variable Flip Angle (VFA)

VFA T_1 mapping consists of the acquisition of a range of steady-state spoiled gradient recalled (SPGR) echo MR images over a set of flip angles [Christensen et al., 1974, Homer and Beevers, 1985, Fram et al., 1987]. Since steady-state MR sequences can use much shorter repetition times (TR) [Nataraj et al., 2017] than classical inversion/saturation recovery sequences, high-resolution T_1 maps can be acquired in clinically feasible scanning time [Deoni et al., 2003].

Important in VFA T_1 mapping is the concept of spoiling. All of the T_1 mapping techniques assume that there is no residual transverse magnetization at the end of TR. This is a very reasonable assumption if TR $>>$ T_2, which is the case of an IR sequence. However, since VFA sequences use very short TR, it is necessary to get rid of the residual magnetization after TR. This technique is called spoiling [Stikov et al., 2015], and is a standard tool for SPGR sequences that are used for VFA.

In contrast to IR sequences, VFA T_1 mapping always employs GRE readout (Fig. 3.5). Moreover, VFA is often applied with 3D sequences [Jara et al., 1993, Han et al., 2015, Dietrich et al., 2015, Kemper et al., 2016].

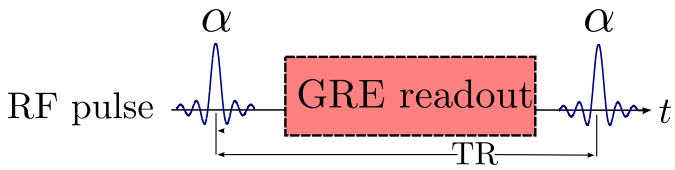


Figure 3.5: To achieve a steady state, a train of RF pulses with same flip angle α are sent with a very short TR. Figure based on [Stikov et al., 2015].

3.5 T_1 model fitting

Once the longitudinal net nuclear magnetization has been probed for different inversion times or flip angles, a T_1 model is voxel-wise fitted to the intensity of the images. The T_1 value which gives the "best" fit according to a specific criterion represents the estimate of the underlying T_1. Formally, inferring the T_1 value is an estimation problem, and in this regard, chapter 4 is fully dedicated to cover the basics of parameter estimation theory incorporating statistical knowledge. Apart from the fitting criterion or estimation procedures, a fundamental issue is the choice of an accurate relaxation signal model. The most popular relaxation models for IR and VFA sequences are reviewed below.

3.5.1 Relaxation signal models for T_1 mapping

T_1 relaxation signal models are dependent on the pulse sequence at hand. An assumption that we made is that the reconstructed MR images are, up to a constant factor, K, equal to the transversal component of the net nuclear magnetization vector, $M_{x'y'}$. Hence, the magnitude image f is proportional to the magnitude of $M_{x'y'}$ that is,

$$f \propto |M_{x'y'}|. \tag{3.3}$$

The transversal component $M_{x'y'}$ is related to the longitudinal component, which obviously depends on T_1, through a simple multiplication with a complex number, which models the projection of M_z on the x-y plane with a given flip angle. This complex-valued number also includes the effect of attenuation due to T_2 or T_2^* among other factors.

3.5.1.1 Inversion Recovery (IR)

With a GRE readout, if a perfect spoiling of M_{xy} after the 180° is assumed, the longitudinal component M_z at $t = \mathrm{TI}$ takes the expression of [Barral et al., 2010]

$$M_z = M_z(0)(1 + e^{-\frac{\mathrm{TR}}{T_1}} - 2e^{-\frac{\mathrm{TI}}{T_1}}). \tag{3.4}$$

Similarly, with an SE readout, M_z is given by [Barral et al., 2010]

$$M_z = M_z(0)(1 - e^{-\frac{\mathrm{TR}}{T_1}} + 2e^{-\frac{\mathrm{TR}-\mathrm{TE}/2}{T_1}} - 2e^{-\frac{\mathrm{TI}}{T_1}}). \tag{3.5}$$

Note that if TR >> T_1, both models converge into the popular model of Eq. (3.2). With a GRE readout, if the two initial flip angles, α_1 and α_2, are not exactly 180° and 90°, respectively, the following model is known to be more accurate

$$M_z = M_z(0)\frac{1-\cos(\alpha_1)e^{-\frac{\mathrm{TR}}{T_1}}-(1-\cos(\alpha_1))e^{-\frac{\mathrm{TI}}{T_1}}}{1-\cos(\alpha_1)\cos(\alpha_2)e^{-\frac{\mathrm{TR}}{T_1}}}. \tag{3.6}$$

Likewise, for an SE readout, when the third flip angle is not exactly 180°, the following model is preferable

$$\begin{aligned} M_z =& M_z(0)\frac{1-\cos(\alpha_1)\cos(\alpha_3)e^{-\frac{\mathrm{TR}}{T_1}}-\cos(\alpha_1)(1-\cos(\alpha_3))e^{-\frac{\mathrm{TR}-\mathrm{TE}/2}{T_1}}}{1-\cos(\alpha_1)\cos(\alpha_2)\cos(\alpha_3)e^{-\frac{\mathrm{TR}}{T_1}}} \\ &- M_z(0)\frac{(1-\cos(\alpha_1))e^{-\frac{\mathrm{TI}}{T_1}}}{1-\cos(\alpha_1)\cos(\alpha_2)\cos(\alpha_3)e^{-\frac{\mathrm{TR}}{T_1}}} \end{aligned} \tag{3.7}$$

In both cases, $M_{x'y'} \propto M_z \sin(\alpha_2)$. Since both the flip angles and the TR are constant for varying TI, the magnitude signal f can always be modeled as

$$f(\mathrm{TI}) = |a + be^{\frac{\mathrm{TI}}{T_1}}|, \tag{3.8}$$

where a and b are real values, and where $b = -2a$ when TR >> T_1 and perfect 180° is assumed. This signal model (Fig. 3.6) is used in chapter 5.

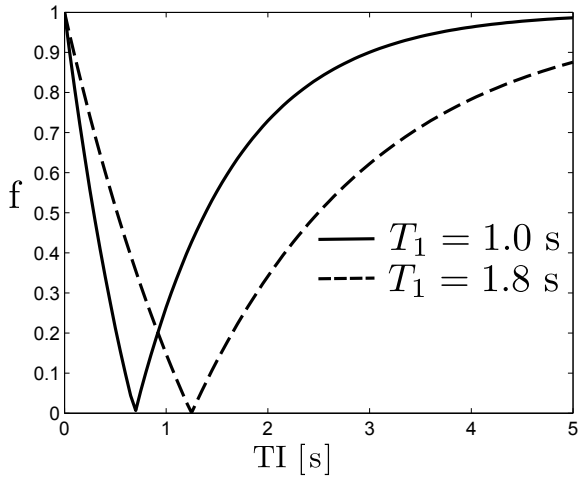


Figure 3.6: Two different curves of the IR signal model (Eq. (3.8)) for varying TI.

3.5.1.2 Variable Flip Angle (VFA)

In an SPGR echo sequence, the longitudinal component of the net nuclear magnetization vector, M_z, under a steady-state regime with a flip angle $\alpha \neq 180°$, can be

written as

$$M_z = \frac{M_z(0)(1 - e^{-\frac{\mathrm{TR}}{T_1}})}{1 - e^{-\frac{\mathrm{TR}}{T_1}} \cos(\alpha)} \tag{3.9}$$

and thus $M_{x'y'} \propto M_z \sin(\alpha)$. Hence, a popular model for the magnitude signal f is

$$f(\alpha) = \frac{K(1 - e^{-\frac{\mathrm{TR}}{T_1}}) \sin(\alpha)}{1 - e^{-\frac{\mathrm{TR}}{T_1}} \cos(\alpha)}, \tag{3.10}$$

where the unknown parameter $K > 0$ includes multiplicative factors such as $M_z(0)$, and the attenuation due to T_2^* relaxation for a fixed TE [Teixeira et al., 2017]. The signal model of Eq. (3.10) is plotted in Fig. 3.7 for two values of T_1 representing white and grey matter.

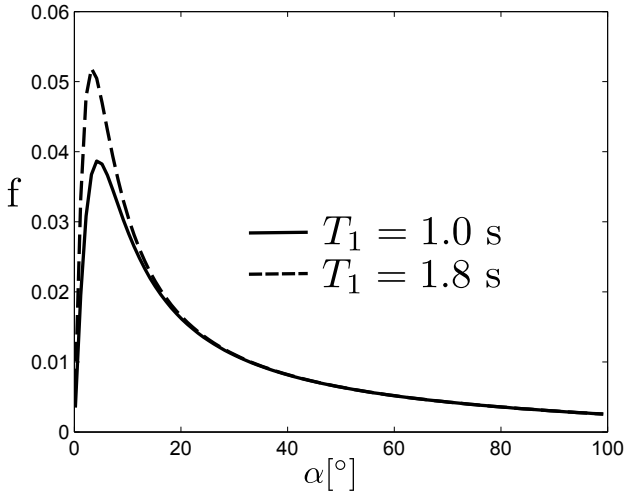


Figure 3.7: Two different curves of the SPGR/VFA signal model (Eq. (3.10)) for varying α.

3.6 Current challenges in MRI relaxometry

3.6.1 Sources of Errors

The clinical utility of MRI relaxometry depends fundamentally on the ability to provide relaxation data with high accuracy and reproducibility [Deoni, 2010]. Unfortunately, there are important sources of errors that prevent MRI relaxometry from fulfilling such requirements.

3.6.1.1 Motion

Subject motion is always a source of error in MRI. Subject motion can occur during the k-space acquisition process, having negative effects on image quality,

e.g., ghosting and blurring artifacts. Furthermore, in quantitative MRI, and in particular in MRI relaxometry, as the acquisition of the complete data set is relatively slow, the presence of motion is very likely, and hence images are prone to be misaligned.

The most straightforward way to correct for motion is to spatially register the T_1 or T_2-weighted images to a target image, using general-purpose image registration similarity measures. This is a challenging task since the intensity of the images varies spatially but also temporally. Furthermore, the inherent interpolation in the registration step invalidates statistical assumptions on the acquired images, which are needed to derive optimal statistical estimators, as described in chapter 4.

Recently, model-based approaches for inter-image motion correction have been introduced, where the signal model is included in the registration step [Xue et al., 2013, Hallack et al., 2014, Ramos-Llordén et al., 2015a, Ramos-Llordén et al., 2017]. One of the main contributions of this PhD thesis is a model-based unified approach for simultaneous motion and T_1 estimation by means of a Maximum Likelihood framework. In this approach, motion parameters and the motion-free T_1 map are optimally estimated by properly accounting for the data statistics. Chapter 5 is completely devoted to presenting this work.

3.6.1.2 Flip angle inhomogeneity

Accurate knowledge of flip angle is important for accurate T_1 and T_2 estimation. However, the nominal flip angles, those who are selected in the scanner software, may differ from those that are transmitted. The main causes are RF pulse profile errors and inhomogeneity effects in the tissues. Transmitted flip angles vary across the volume of the slice. For 3D imaging, the profile effect can be tolerated if the anatomy of interest lies in the center portion of the volume [Deoni, 2010], where the flip angle is approximately uniform. For 2D multi-slice imaging, the profile effect will create a variation through the image slice [Deoni, 2010]. Minimization of the flip angle related errors can be achieved by means of improved RF pulse design or calibration of the transmitted flip angle field [Deoni, 2010].

The transmitted flip angle can be also measured. The simplest technique is dubbed the *double-angle approach*, where two spin-echo images are acquired with flip angles of α and 2α, respectively, and very long TR . The transmitted flip angle is then calculated as the ratio between signal intensities. There are faster but more complex methods for flip angle correction. In most of them, an estimation of the amplitude of the $\boldsymbol{B_1}$ field is first obtained [Stollberger and Wach, 1996, Yarnykh, 2007, Sacolick et al., 2010, Chavez and Stanisz, 2012]. This technique is called $\boldsymbol{B_1}$ *mapping*. Then, the correct flip angle is obtained with a direct relation between the $\boldsymbol{B_1}$ pulse and the flip angle (see Eq. (1.17)).

3.6.1.3 Fitting procedures

There is no unique choice for the best criterion in MRI relaxometry model fitting. It is often common to find T_1 and T_2 mapping techniques that estimate the underlying T_1 or T_2 with closed-form formulas, resulting from the approximation of models. It is also common for relaxometry practitioners to use standard fitting algorithms as black-box procedures. Probably, the most popular type of fitting algorithms for MRI relaxometry are those based on least-squares fitting, such as Levenberg-Marquardt or Gaussian Newton methods. Those algorithms implicitly assume the least-squares approach as the criterion of goodness of fit, that is, they seek for the model that best fits the noisy relaxation signal in the l_2 sense.

Those solutions may provide adequate estimates of T_1 and T_2 at first sight, but when the quest for highly accurate and precise T_1/T_2 mapping is in play, substantially more accurate and precise estimators can be obtained if tools from the theory of statistical parameter estimation theory are learned. In fact, the naïve use of fitting procedures can also be considered an important source of errors in MRI relaxometry.

Indeed, still, a large set of algorithms disregard the knowledge of data statistics, in the sense that they are not derived from a statistical principle. Knowledge of MR noise is often only used to generate a realistic simulation framework in order to assess the accuracy and precision of given estimators in a Monte Carlo simulation setting, and nothing more. However, if the very nature of the intensity of MR images as random variables is considered right from the beginning, and the data distributions are accounted for, optimal estimators can be derived in a rigorous way, with theoretically proven qualities that have been there long before MRI relaxometry was born. Statistical parameter estimation theory provides scientists and engineers with "a guide of action" to estimate T_1 and T_2 values with rigorous estimators that are likely to have better performance than algorithms applied in a blind manner. Interestingly, this theory gives a justification for the use of least-squares approaches under specific conditions of the data distribution, where, for that specific case, they are shown to be statistically optimal. During this PhD thesis, we try to advocate an MRI relaxometry approach that uses sophisticated algorithms from statistical parameter estimation theory. The use of those algorithms has proved to improve the accuracy and precision of T_1 mapping substantially. While our inclinations for the application of statistically optimal estimators to relaxometry is evident, by no means that implies that these algorithms should be applied totally disregarding the relaxation signal models. A message we transmit along this dissertation is that if statistically optimal algorithms are tailored or adapted to the singularities of the relaxation model at hand, not only accuracy and precision but also computational efficiency can be gained. As a proof of this, we refer the reader to chapter 6.

3.6.2 Specific Absorption Rate (SAR)

The specific absorption rate (SAR) is a measure of the rate at which energy is absorbed when the human body is exposed to an RF pulse [Jin, 1998]. The SAR

value is proportional to $||\boldsymbol{B_0}||_2^2$, the square of α, and the fraction of duration of the RF pulse, D [Bottomley et al., 1985, Hennig, 1988, Ibrahim et al., 2001]:

$$\text{SAR} \propto ||\boldsymbol{B_0}||_2^2 \alpha^2 D. \tag{3.11}$$

With a TSE sequence, the echo train of RF pulses transfers a high RF energy, resulting in high values of SAR [Oshio and Feinberg, 1991]. To reduce the SAR, the number of slices that are acquired within one TR can be decreased. Another option is to use RF pulses with lower flip angles at the expense of a decrease in contrast and SNR [Weigel et al., 2007].

3.6.3 Acquisition time

The total acquisition time of a T_1-weighted data set is always proportional to the number of T_1-weighted images. The total scan time of a single slice of a T_1-weighted image with Cartesian TSE was already given in subsection 2.3.3. For a 2D multi-slice IR sequence with N phase-encoding lines, the total scan time T_{Scan} is given by

$$T_{\text{Scan}} = N \frac{\text{TR} N_{\text{ex}}}{\text{ETL}}, \tag{3.12}$$

where N_{ex} is the total number of slices divided by the number of slices acquired per TR [Van Steenkiste, 2016]. Though VFA sequences may use a much shorter TR than IR sequences, the latter, which are still the gold standard, requires a considerably long TR. Therefore, the acquisition time of a T_1-weighted data set is substantially long. The acquisition can be speeded up by acquiring less T_1-weighted image, however, this may come at the expense of a decrease in the accuracy and precision of the estimated T_1 map. Acquiring a T_2-weighted image often requires less acquisition time than acquiring a T_1 weighted image since much shorter TR values can be used. However, T_2 mapping is normally performed with a substantially higher number of images than T_1 mapping [Poon and Henkelman, 1992, Björk et al., 2016]. Therefore, T_2 relaxometry also demands a relatively long scan time.

4

Statistical parameter estimation theory

Contents

4.1 Introduction

The relaxation models that were described in the previous chapter are mathematical entities that relate the intensity of the voxels of the T_1-weighted images to the underlying relaxation parameters. Obviously, relaxation models cannot cover all aspects that occur in the relaxation process. As such, the task of estimating relaxation times is always prone to errors. Refining those models will eventually lead to more accurate relaxometry techniques. However, as was described in chapter 2, the intensities of the acquired MR images are always polluted with random noise, and should, therefore, be considered as random variables. In consequence, if we are

given another set of T_1-weighted images, acquired under the same conditions, the estimated relaxation parameters will vary, even though the estimator is based on exactly the same relaxation model. Random errors cannot be cured by improving the relaxation models, and are unavoidable in every scientific measurement. The theory that deals with the estimation of parameters from random variables is called statistical parameter estimation theory. This chapter is devoted to presenting to the reader the basics of this theory, with emphasis on statistical estimators, which have been extensively employed in this PhD thesis. We describe the desirable properties of statistical estimators after which we focus on two special estimators that have been employed in this work. Those estimators are the Maximum Likelihood (ML) estimator and the non-linear least squares (NLLS) estimator.

4.2 Statistical parameter estimation theory

4.2.1 Definition of statistical estimator

Suppose we have observed a realization of N random variables $\boldsymbol{X} = (X_1, X_2, ..., X_N)^T$ whose PDF is parametric in an unknown parameter vector $\boldsymbol{\theta} \in \mathbb{R}^P$. We would like to estimate that parameter vector. To do so, with the vector of realizations $\boldsymbol{x} = (x_1, x_2, ..., x_N)^T$, from now on called observations, we construct the vector $\hat{\boldsymbol{\theta}} = \boldsymbol{g}(x_1, x_2, ..., x_N)$, where $\boldsymbol{g}(\cdot)$ is any vector-valued function with P components and N variables. Vector $\hat{\boldsymbol{\theta}}$ is called an estimate of $\boldsymbol{\theta}$. If we repeat the experiment, we will obtain new observations, and we will then get a different estimate. Those estimates are realizations of the random variable $\hat{\boldsymbol{\theta}}(\boldsymbol{X}) = \boldsymbol{g}(X_1, X_2, ..., X_N)$, which is called a statistical estimator of $\boldsymbol{\theta}$ [Papoulis and Pillai, 2002].

4.2.2 Properties of statistical estimators

4.2.2.1 Accuracy and Precision

We say that an estimator is accurate if the associated estimates are *on average* close to the true value. The accuracy is formally described by the bias of the estimator, which is defined as

$$\text{bias}(\hat{\boldsymbol{\theta}}(\boldsymbol{X})) = \mathbb{E}\{\hat{\boldsymbol{\theta}}(\boldsymbol{X})\} - \boldsymbol{\theta}, \tag{4.1}$$

where $\mathbb{E}\{\hat{\boldsymbol{\theta}}(\boldsymbol{X})\}$ is the expectation of the random vector $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$. The lower the bias, the more accurate is the estimator. An estimator is unbiased if its bias is zero. An estimator is asymptomatically unbiased if the bias tends to zero when $N \to \infty$.

On the other hand, the precision is related to the *average spread* of the estimates, i.e., how much the estimates change when the experiment is repeated. Precision is a desirable property of an estimator. The precision of an estimator is formally described by its variance vector, that is by

$$\text{var}(\hat{\boldsymbol{\theta}}(\boldsymbol{X})) = ([\boldsymbol{C}_{\hat{\boldsymbol{\theta}}(\boldsymbol{X})}]_{11}, [\boldsymbol{C}_{\hat{\boldsymbol{\theta}}(\boldsymbol{X})}]_{22}, ..., [\boldsymbol{C}_{\hat{\boldsymbol{\theta}}(\boldsymbol{X})}]_{PP})^T, \tag{4.2}$$

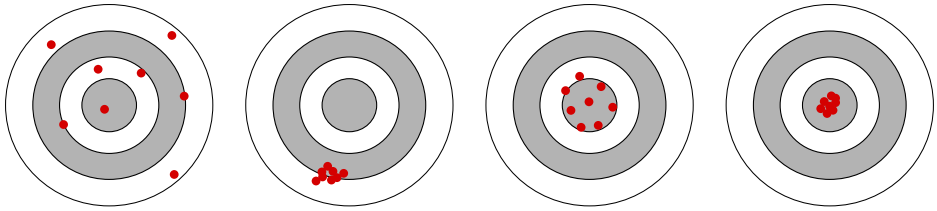


Figure 4.1: Difference between accuracy and precision when aiming at the center.

where $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}(\boldsymbol{X})}$ is the covariance matrix of the random vector $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$. The "smaller" $\mathrm{var}(\hat{\boldsymbol{\theta}}(\boldsymbol{X}))$, the more precise the estimator is. Hereafter, we just simply write variance of an estimator, though naturally we refer to the vector of variances of the estimator's components. Ideally, an estimator should have a high accuracy and precision. In Fig. 4.1, we illustrate the difference between accuracy and precision. Another popular performance measure of a statistical estimator is the mean squared error (MSE). The MSE combines both the bias and the variance in a single expression,

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}(\boldsymbol{X})) = \mathbb{E}\{||\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}||_2^2\} = \sum_{p=1}^{P}[\mathrm{var}(\hat{\boldsymbol{\theta}}(\boldsymbol{X}))]_p + \sum_{p=1}^{P}[\mathrm{bias}(\hat{\boldsymbol{\theta}}(\boldsymbol{X}))]_p^2. \quad (4.3)$$

Observe that an estimator can be more accurate than another one, while having a higher MSE. An estimator that has the lowest possible MSE is considered *optimal* in MSE sense. An overarching difficulty to obtain optimal estimators in MSE sense is that the criterion of Eq. (4.3) is dependent on the true but unknown parameter $\boldsymbol{\theta}$, and, in most of the cases, the optimal estimator in MSE sense depends as well on the unknown $\boldsymbol{\theta}$. Thus, though optimal, the estimator is unrealizable, and hence of no practical interest.

Instead, it is much more common to derive optimal estimators in terms of accuracy and precision by focusing on each property separately. A very generalized approach is to restrict the search for best estimators to those which are unbiased, and, from that set of estimators, select those or the one with the highest precision. In principle, it is also possible to seek for estimators with the highest possible precision, and next focus on increasing the accuracy. However, this approach leads to irrelevant results, since we can easily construct an estimator with an arbitrary low variance, in particular zero, but being the resultant estimator a constant, it would be severely biased for all but one possible value of $\boldsymbol{\theta}$. On the contrary, estimators with the highest possible accuracy, that is those which are unbiased, are necessarily not constant. From the class of unbiased estimators, we can perfectly seek for the one with the highest precision or, equivalently, the one with minimum variance for all $\boldsymbol{\theta}$. The reason is that, in contrast to the MSE optimality criterion, it is often possible to find *realizable* unbiased estimators that have the highest precision for all $\boldsymbol{\theta}$.

Such an estimator is called the uniformly minimum variance unbiased estimator (UMVUE). A rich part of statistical parameter estimation theory is devoted to the quest for that estimator. We will elaborate on this topic in subsection 4.2.2.3 and subsection 4.2.2.4, but we will first present another appealing property of every statistical estimator.

4.2.2.2 Consistency

An estimator is consistent if it converges in probability to $\boldsymbol{\theta}$ when the number of observations tends to infinity, that is, for every $\epsilon > 0$ [van den Bos, 2007],

$$\lim_{N\to\infty} \Pr(||\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}||_2 > \epsilon) = 0, \tag{4.4}$$

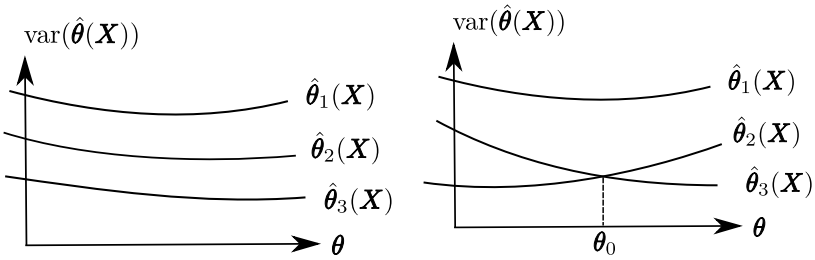
with $\Pr(\cdot)$ being the probability. An estimator is consistent if it is asymptotically unbiased and its variance tends to zero, though the converse is not true. Furthermore, an estimator may be consistent but biased for finite N.

4.2.2.3 Uniformly minimum variance unbiased estimator (UMVUE)

The UMVUE for the parameter $\boldsymbol{\theta}$ is formally defined as the estimator $\hat{\boldsymbol{\theta}}_{\text{UMVUE}}(\boldsymbol{X})$ that fulfills

$$\text{var}(\hat{\boldsymbol{\theta}}_{\text{UMVUE}}(\boldsymbol{X})) \leq \text{var}(\hat{\boldsymbol{\theta}}(\boldsymbol{X})) \quad \text{for all } \hat{\boldsymbol{\theta}}(\boldsymbol{X}) \text{ with } \text{bias}(\hat{\boldsymbol{\theta}}(\boldsymbol{X})) = 0. \tag{4.5}$$

The word uniformly implies that the condition must hold for all possible values that $\boldsymbol{\theta}$ takes. That strong condition may lead to situations where the UMVUE does not exist. Fig. 4.2 illustrates the importance of the uniform property for the UMVUE. In



(a) $\hat{\boldsymbol{\theta}}_{\text{UMVUE}}(\boldsymbol{X}) = \hat{\boldsymbol{\theta}}_3(\boldsymbol{X})$ (b) UMVUE does not exist.

Figure 4.2: Graphical illustration of the UMVUE.

Fig. 4.2.(a), the lowest variance for all values of $\boldsymbol{\theta}$ is achieved with estimator $\hat{\boldsymbol{\theta}}_3(\boldsymbol{X})$. Hence, it is the UMVUE. On the contrary, in Fig. 4.2.(b), $\hat{\boldsymbol{\theta}}_3(\boldsymbol{X})$ only possesses the lowest variance for $\boldsymbol{\theta} > \boldsymbol{\theta}_0$, where for $\boldsymbol{\theta} \leq \boldsymbol{\theta}_0$ the minimum variance is achieved with $\hat{\boldsymbol{\theta}}_2(\boldsymbol{X})$. Thus, we conclude that for the case (b), the UMVUE does not exist. If the UMVUE exists, the Rao-Blackwell-Lehmann-Scheffe theorem serves as a guide to obtain that estimator. The theorem, whose proof can be found in [Kay,

1993], dictates that, if $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ is an unbiased estimator of $\boldsymbol{\theta}$, and $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X})$ is a sufficient statistic for $\boldsymbol{\theta}$, then the estimator $\hat{\boldsymbol{\theta}}_{\text{Rao}}(\boldsymbol{X}) = \mathbb{E}\{\hat{\boldsymbol{\theta}}(\boldsymbol{X})|\boldsymbol{T}(\boldsymbol{X})\}$, where the expectation is taken over all possible realizations that the random variable $\boldsymbol{T}$ can take, does not depend on $\boldsymbol{\theta}$ (hence is a realizable estimator), and is still unbiased. Furthermore, it has a lower or equal variance than $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ for all values that $\boldsymbol{\theta}$ takes and, importantly, provided $\boldsymbol{T}$ is a complete statistic, it is the UMVUE. For a definition of a sufficient and complete statistic the reader is referred to [Papoulis and Pillai, 2002].

The Rao-Blackwell-Lehmann-Scheffe theorem gives us the necessary steps to construct the UMVUE, however, finding a complete sufficient statistic $\boldsymbol{T}$ may not be an easy task. In the search for the UMVUE, one can ask whether there exists a lower bound on the variance of all unbiased estimators. That bound may not necessarily be reached by any estimator, but if an estimator does, obviously that estimator is the $\hat{\boldsymbol{\theta}}_{\text{UMVUE}}(\boldsymbol{X})$. This line of thinking gives us another technique to seek for $\hat{\boldsymbol{\theta}}_{\text{UMVUE}}(\boldsymbol{X})$, but first we need to confirm the existence of such a lower bound.

4.2.2.4 Efficiency

Let us denote with $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$ the joint PDF of the random variables $(X_1, X_2, ..., X_N)^T$ evaluated at vector observation $\boldsymbol{x}$. Furthermore, let us assume that the following *regularity* condition holds:

$$\mathbb{E}\left\{\frac{\partial \log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{\partial[\boldsymbol{\theta}]_p}\right\} = 0, \text{ with } p = 1, ..., P. \tag{4.6}$$

If that is so, we have that for any unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$:

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}(\boldsymbol{X})} \succcurlyeq \boldsymbol{I}^{-1}(\boldsymbol{\theta}), \tag{4.7}$$

where $\boldsymbol{I}(\boldsymbol{\theta}) \in \mathbb{R}^{P\times P}$ is the so-called *Fisher information matrix* evaluated at $\boldsymbol{\theta}$, whose components are given by

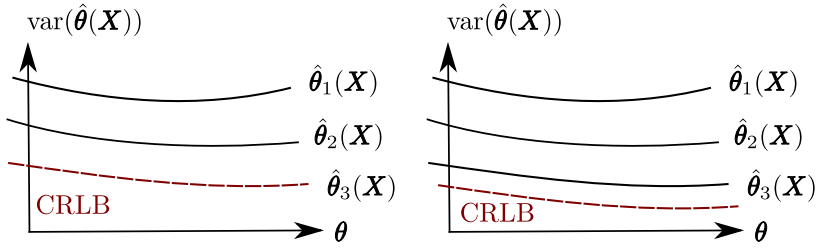
$$[\boldsymbol{I}(\boldsymbol{\theta})]_{ij} = \mathbb{E}\left\{\frac{\partial \log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{\partial[\boldsymbol{\theta}]_i}\frac{\partial \log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{\partial[\boldsymbol{\theta}]_j}\right\} \tag{4.8}$$

for $i = 1, ..., P$ and $j = 1, ..., P$, and where the symbol $\succcurlyeq$ means that the matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}(\boldsymbol{X})} - \boldsymbol{I}^{-1}(\boldsymbol{\theta})$ is semidefinite positive. The inequality of Eq. (4.7) is known as the *Cramér-Rao Lower Bound* (CRLB). A direct consequence of the CRLB is the following inequality on the variance:

$$[\text{var}(\hat{\boldsymbol{\theta}}(\boldsymbol{X}))]_p \geq [\boldsymbol{I}^{-1}(\boldsymbol{\theta})]_{pp}, \text{ with } p = 1, ..., P. \tag{4.9}$$

Therefore, the CRLB establishes a limit to the highest precision that can be obtained with an unbiased estimator. An estimator which reaches the CRLB, that is, that turns inequality Eq. (4.7) into an equality is said to be efficient. An

estimator that reaches the CRLB asymptotically is named asymptotically efficient. An efficient estimator is always the UMVUE, but the converse is not true. Indeed, an efficient estimator may not exist but still the UMVUE can be obtained, as it is depicted in Fig. 4.3.



(a) $\hat{\boldsymbol{\theta}}_3(\boldsymbol{X})$ is efficient and hence the UMVUE.

(b) $\hat{\boldsymbol{\theta}}_3(\boldsymbol{X})$ is not efficient but is the UMVUE.

Figure 4.3: Graphical illustration of the relation between efficiency and the UMVUE.

The CRLB is an extremely useful tool in estimation theory since it allows us to confirm that an estimator is the UMVUE, in a more practical way than the Rao-Blackwell-Lehmann-Scheffe theorem. Furthermore, if the bound is not reached, at least it provides a benchmark against which we can compare the performance of any unbiased estimator [Kay, 1993].

The CRLB can also be interpreted from a pure information theory point of view. Indeed, the Fisher information matrix is a way of measuring the amount of information that random variables $(X_1, X_2, ..., X_N)^T$ carry about parameter $\boldsymbol{\theta}$.

An intuitive, very common interpretation is based on the curvature of $\log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$ when considered to be a function of $\boldsymbol{\theta}$. Intuitively, the sharper $\log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$ is around its global maximum, the more information random variables $(X_1, X_2, ..., X_N)^T$ contains about $\boldsymbol{\theta}$. Indeed, when the curvature of $\log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$ is rather pronounced around that maximum, e.g, $\boldsymbol{\theta}^*$, this can be interpreted, from a probabilistic point of view, as if the neighborhood values were very unlikely to be the underlying parameter which would have produced the observed random vector $\boldsymbol{x}$, at least in comparison to $\boldsymbol{\theta}^*$. If the curvature is, instead, low, there are several values of $\boldsymbol{\theta}$ in the neighborhood that may have produced the same observation vector $\boldsymbol{x}$ with similar likelihood. The curvature of an N-dimensional function around its maximum is described by its Hessian matrix, and the Hessian matrix of the function $\log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$ (w.r.t. $\boldsymbol{\theta}$) is what is called the *observed* Fisher information matrix. The Fisher information matrix, $\boldsymbol{I}(\boldsymbol{\theta})$, is merely the expectation of the observed Fisher information matrix. Hence, the Fisher information matrix describes the *average* curvature of $\log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$. Other interpretations of the Fisher information matrix from an information theory point of view are related to the concepts of score vector and its covariance matrix, and dissimilarity measures between PDFs in information geometry [Duchi, 2016].

When function $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$ is interpreted as a function of $\boldsymbol{\theta}$ for fixed $\boldsymbol{x}$, it is called the *likelihood* function, which is denoted as $L(\boldsymbol{\theta};\boldsymbol{x})$. Its natural logarithm, $\log L(\boldsymbol{\theta};\boldsymbol{x})$,

is called the *log-likelihood* function. The likelihood function constitutes the main ingredient in the construction of the powerful and popular Maximum Likelihood (ML) estimator. Not surprisingly either, if an efficient estimator exist, it can only be the ML estimator [Papoulis and Pillai, 2002].

4.3 Maximum Likelihood estimators

ML estimators are based on the maximum likelihood principle. Unintentionally, we have already described this principle when we related the Fisher information matrix to the information that $(X_1, X_2, ..., X_N)^T$ contain about $\boldsymbol{\theta}$. According to the maximum likelihood principle, given a PDF $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$, it is reasonable to choose as estimate that value of $\boldsymbol{\theta}$ that most *likely* caused the observations $(x_1, x_2, ..., x_N)^T$ to occur. For fixed $\boldsymbol{x}$, such a value is the maximum point of the likelihood function, $L(\boldsymbol{\theta};\boldsymbol{x})$. Such a value is called an ML estimate for $\boldsymbol{\theta}$, and is defined as,

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta};\boldsymbol{x}), \tag{4.10}$$

or by using the negative log-likelihood function $\mathcal{L}(\boldsymbol{\theta};\boldsymbol{x}) = -\log L(\boldsymbol{\theta};\boldsymbol{x})$,

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg \min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta};\boldsymbol{x}). \tag{4.11}$$

The corresponding random vector $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\boldsymbol{X})$ is called the ML estimator of $\boldsymbol{\theta}$. In the literature one often encounters the following expression for $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$,

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg \min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \mathcal{L}_n(\boldsymbol{\theta}; x_n), \tag{4.12}$$

where $\mathcal{L}_n(\boldsymbol{\theta}; x_n)$ is the negative log-likelihood function of the random variable X_n. This particular expression arises when the elements of the vector of random variables $(X_1, X_2, ..., X_N)^T$ are independent. Then, the joint probability density function, $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$, is the product of the PDFs of the set of random variables $(X_1, X_2, ..., X_N)^T$, and the logarithm transforms such an expression into a finite sum. The reader should be aware, however, that this is a particular case of the more general expression of Eq. (4.11).

The maximum likelihood principle is the most popular approach to obtaining practical estimators, since it is a "turn-the-crank" procedure [Kay, 1993] and can be applied to a wide range of problems provided we know the formal expression for $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$. ML estimators have excellent asymptotic properties. In this section, we stress some of the particular features of the ML estimator, and we list its optimal asymptotic properties.

4.3.1 Distinctive properties of ML estimators

The ML estimator enjoys an invariance property [van den Bos, 2007]. Suppose that $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\boldsymbol{X})$ is the ML estimator of $\boldsymbol{\theta}$. Furthermore, suppose that $\boldsymbol{\phi}$ is an unknown

parameter that depends on $\boldsymbol{\theta}$ via the function relationship $\boldsymbol{\phi} = \boldsymbol{s}(\boldsymbol{\theta})$. Then, the ML estimator of $\boldsymbol{\phi}$ is the output of $s(\cdot)$ when the input is $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\boldsymbol{X})$.

Moreover, if an efficient estimator exists, it is necessarily the ML estimator. Indeed, the CRLB inequality becomes an equality when the elements of the statistical estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ fulfill [van den Bos, 2007]

$$[\hat{\boldsymbol{\theta}}]_p - [\boldsymbol{\theta}]_p = \sum_{j=1}^{P} [\boldsymbol{I}^{-1}(\boldsymbol{\theta})]_{pj} \frac{\partial \log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{\partial [\boldsymbol{\theta}]_j}, \text{ with } p = 1, ..., P, \tag{4.13}$$

for all possible values of $\boldsymbol{\theta}$. When evaluating at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\mathrm{ML}}$, the expression on the right becomes zero, and hence $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\mathrm{ML}}$. Therefore, the corresponding ML estimator $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\boldsymbol{X})$ is efficient.

4.3.2 Asymptotic optimality

Under some regularity conditions, the ML estimator converges in distribution to a multi-variate normal distribution with mean the unknown parameter $\boldsymbol{\theta}$, and with covariance matrix the inverse of the Fisher information matrix (evaluated at $\boldsymbol{\theta}$) [Papoulis, 1977]. As a consequence of convergence in distribution, the ML estimator is therefore asymptotically unbiased and efficient. Furthermore, the ML estimator is also a consistent estimator.

4.4 Non-linear least-squares estimators

Our line of thinking to design optimal statistical estimators has been to consider the class of unbiased estimators and determine the one with minimum variance. We have seen that for a large number of observations, the ML is the most appealing estimator because it is asymptotically unbiased and efficient, and consistent as well. In this last section, we present a class of estimators, called non-linear least-squares (NLLS) estimators which, though they do not share optimal statistical properties (except for particular cases that we will point out), are one of the most studied and employed estimators in the history of modern science. This class of estimators is based on the concept of least-squares, a method which dates back to 1795 when Gauss employed it in his studies about planetary motion [Kay, 1993]. NLLS estimators are rather simple to implement since they do not require knowledge about the joint PDF of the random variables $(X_1, X_2, ..., X_N)^T$, from which the parameter is estimated.

4.4.1 The Least-Squares approach

Let us suppose that the expectation of each of the random variables X_n depends on $\boldsymbol{\theta}$ as

$$\mathbb{E}\{X_n\} = \phi_n(\boldsymbol{\theta}), \tag{4.14}$$

where $\phi_n(\boldsymbol{\theta})$ for $n = 1, ..., N$ is a deterministic model. With the LS approach, we seek to minimize the squares of the differences between the realizations x_n and the model (Eq. (4.14)). Such a difference is often called the residual. Formally, an LS estimate $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ of $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = \arg \min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \left(x_n - \phi_n(\boldsymbol{\theta})\right)^2. \tag{4.15}$$

That is, the LS estimate $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ is that value of $\boldsymbol{\theta}$ that yields the *least* sum of *squares* of the differences or residuals. The corresponding random vector $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}(\boldsymbol{X})$ is the NLLS estimator if $\phi_n(\cdot)$ is a non-linear function of $\boldsymbol{\theta}$, and it is termed the Linear LS (LLS) estimator if $\phi_n(\cdot)$ is a linear function of $\boldsymbol{\theta}$. Since practically all of the models considered in this PhD thesis are non-linear, we will focus on the NLLS case. Sometimes, the squared differences in Eq. (4.15) are multiplied by deterministic values w_n in order to weight the contribution of every residual differently. If that is so, the term weighted NLLS is used. Unweighted (or equivalently, uniformly weighted) NLLS estimators are sometimes called ordinary NLLS estimators to make the distinction clear.

4.4.2 Properties and connection to ML estimators

NLLS estimators, both weighted and ordinary, share the invariance property of the ML estimators. Indeed, observe that the proof for the invariance property [van den Bos, 2007] is essentially valid for every extremum estimator (estimators that are obtained by minimizing a certain cost function). Nevertheless, we cannot make any claim about optimal properties of the NLLS estimator, unless statistical information of $(X_1, X_2, ..., X_N)^T$ is given. The interested reader can find some of those conditions in [van den Bos, 2007].

In certain cases, NLLS estimators are equal to ML estimators. If $p_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta})$ is the joint PDF of independent Gaussian random variables with mean $\phi_n(\boldsymbol{\theta})$ and standard deviation σ_n, the ML estimator of $\boldsymbol{\theta}$, with this set of Gaussian random variables, is identical to the WLLS estimator with $w_n = 1/\sigma_n^2$. If $\sigma_n = \sigma$ for every n, the ML estimator is equal to the ordinary NLLS estimator.

Part III

Contributions

5

A unified Maximum Likelihood framework for simultaneous motion and T_1 estimation in quantitative MR T_1 mapping

Contents

The work in this chapter has been published in:

- **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Jeurissen, B., Vanhevel, F., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2017). A unified maximum likelihood framework for simultaneous motion and T_1 estimation in quantitative MR T_1 mapping. *IEEE Trans. Med. Imag.*, 36(2):433–466.
- **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2015). Simultaneous motion correction and T_1 estimation in quantitative T_1 mapping: an ML restoration approach. *In Proc. IEEE ICIP*, pages 3160–3164.
- **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M. and Sijbers, J. (2015). Simultaneous group-wise registration and maximum likelihood T_1 estimation for T_1 mapping. *Proc. Intl. Soc. Mag. Reson. Med.*, 23:447.
- **Ramos-Llordén, G.**, den Dekker, A. J., and Sijbers, J. (2014). Joint Motion Correction and Estimation for T_1 mapping: proof of concept. *Proceedings of the workshop "Medical Imaging Summer School"*, 2014 https://doi.org/10.13140/rg.2.1.1859.0167 .
- **Ramos-Llordén, G.** and Sijbers, J. (2013). Misalignment correction for T1 maps using a maximum likelihood estimator approach. *Front. Neuroinform. Conf. Abstract: Imaging the brain at different scales: How to integrate multi-scale structural information?, Antwerp, Belgium.*

5.1 Introduction

Quantitative T_1 mapping is a Magnetic Resonance Imaging (MRI) technique in which the spin-lattice relaxation time T_1 of tissues is measured [Tofts, 2004]. Because T_1 depends on biophysical properties, it is used as biomarker in a broad range of diseases, such as multiple sclerosis [Larsson et al., 1989], epilepsy [Conlon et al., 1988] and Alzheimer's disease [Erkinjuntti et al., 1987], as well as in the measurement of perfusion [Detre et al., 1992] and blood flow [Cheng, 2007]. Hence, its accurate and precise estimation is of uttermost importance [Tofts, 2004, van den Bos, 2007]. In order to quantify T_1, a set of T_1-weighted images with different sequence settings needs to be acquired [Tofts, 2004, Trzasko et al., 2013, Van Steenkiste et al., 2017]. From this set, a spatial map of T_1 values can be calculated by fitting a known relaxation model at every voxel. Evidently, to obtain a meaningful T_1 map, spatial correspondence between the images in the acquired series is crucial [Huizinga et al., 2016]. However, due to patient motion and/or apparent spatial shifts introduced by the scanner (e.g., scanner drift [Foerster et al., 2005]), T_1-weighted images are often misaligned.

To deal with this problem, T_1-weighted images are commonly spatially registered prior to the estimation of the T_1 map [Deoni et al., 2003, Warntjes et al., 2007]. This is often done by choosing one T_1-weighted image as a target and subsequently registering the remaining T_1-weighted images to this target image by using a similarity measure such as Mutual Information (MI) [Studler et al., 2010, Bron et al., 2013].

Such an approach, however, suffers from inherent problems. First, the specific relation between the intensity value as a function of time of the (aligned) voxels is ignored. Second, the registration is not driven by a global optimization criterion that considers all T_1-weighted images simultaneously. Even more problematic is the fact that current motion correction is a *preprocessing* step prior to the estimation of the T_1 values. Such a two-step processing pipeline lacks a feedback mechanism between the image registration and the T_1 map estimation step. As a result, registration errors will propagate to the estimation step, leading to biased estimates [Ramos-Llordén et al., 2015a].

Recently, progress in registration of T_1-weighted images was made by the introduction of model-based approaches. Such techniques integrate the signal model connecting the series of images (such as a T_1 relaxation model) into the registration step. State-of-the-art model-based methods have shown to outperform the conventional two-step approach in terms of accuracy, for example, in myocardial T_1 mapping [Xue et al., 2013, Hallack et al., 2014]. Unfortunately, they all come with serious limitations for precise and accurate T_1 mapping, mainly because different criteria for registration and estimation are heuristically combined [Xue et al., 2013, Hallack et al., 2014]. Since they do not constitute a truly unified framework, the output of the algorithms cannot be related to the optimal value of a given global information-based criterion. As a consequence, it is doubtful whether all the information gathered in the series of T_1-weighted images, including the data statistics [Sijbers et al., 1998a, van den Bos, 2007], is optimally exploited.

In this chapter, we propose an integrated model-based image registration and T_1 estimation approach, where the motion parameters and T_1 map are jointly estimated using a unified global information criterion, more specifically, the maximum likelihood (ML) criterion [Sijbers et al., 1998a, Sijbers et al., 1999, den Dekker and Sijbers, 2014]. By combining models of T_1 relaxation, motion, and noise into one statistical model of the T_1-weighted images, we are able to restore the original motion-free T_1 map using a joint ML estimator. The unified ML framework allows accounting for the statistical noise model, the relaxation model, and the motion model simultaneously, exploiting, in addition to the temporal information, knowledge on data statistics. The large-scale ML optimization problem is solved by alternating between the estimation of motion and relaxation parameters in an efficient and robust manner, making use of block coordinate descent [Fessler and Kim, 2011] and Majorize-Minimize (MM) algorithms [Varadarajan and Haldar, 2015]. Exact convergence properties of the algorithm are presented, demonstrating that the proposed iterative procedure leads to the ML estimates in a computationally efficient way.

We thoroughly validate the proposed joint maximum likelihood estimator (MLE) with realistic Monte Carlo (MC) simulations and compare it with the conventional two-step approach as well as the newest state-of-the-art model-based approach of Hallack [Hallack et al., 2014]. We show that substantially more accurate T_1 maps as well as motion parameters can be obtained with our proposed joint MLE. Additionally, the T_1 maps estimated with the joint MLE are superior in terms of the root-mean-square error (RMSE). Apart from simulation experiments, we also quantitatively evaluate the performance of the joint MLE in a controlled experiment involving real T_1-weighted data. Further, we validate it with two in vivo human brain T_1-weighted data sets corrupted by patient motion, showing its applicability in real-life scenarios.

The remainder of this chapter is organized as follows. In section 5.2, the image model used to construct the joint MLE is presented. Section 5.3 is devoted to the joint MLE algorithm. Section 5.4 describes the experiments of which the results are presented in section 5.5, which is followed by a discussion in section 5.6. Finally, conclusions are drawn in section 5.7.

5.2 Theory

The derivation of the joint MLE requires a parametric statistical model of the images. This section is devoted to the derivation of such a model, which comprises a relaxation signal model, a motion model, and a statistical noise model.

5.2.1 Relaxation signal model

As explained in chapter 3, in the absence of noise, the evolution of the magnitude MRI signal in each voxel of a series of N T_1-weighted images can be described by a parametric model $\{f_n(\boldsymbol{\kappa}, T_1)\}_{n=1}^{N}$, where $\boldsymbol{\kappa}$ denotes a vector of nuisance parameters.

The exact expression for this T_1-relaxation model depends on the pulse sequence that is used. In this work, we use the Inversion Recovery (IR) sequence, being the gold standard for T_1-mapping [Tofts, 2004]. Note that signal models corresponding with other sequences, such as SPGR sequences or MOdified Look-Locker Inversion recovery (MOLLI) [Messroghli et al., 2004], can be accommodated within our framework as well. For the IR sequence, we employ the common three parameter magnitude relaxation model given in Eq. (3.8).

To model the noiseless T_1-weighted images, we use a vector notation for the spatially varying parameters T_1, a and b. Let $\boldsymbol{r} = (x, y, z)^T$ be a vector in the Cartesian coordinate system in which they are defined. Then, a 3D spatial T_1 map of M voxels can be defined as a column vector, $\boldsymbol{T_1} \in \mathbb{R}^{M\times 1}$, where $[\boldsymbol{T_1}]_m$ represents T_1 defined at the spatial point $\boldsymbol{r}_m$, indexed by voxel m. Similarly, we define $\boldsymbol{a} \in \mathbb{R}^{M\times 1}$ and $\boldsymbol{b} \in \mathbb{R}^{M\times 1}$ as the parameter maps of a and b. For ease of readability and to alleviate the notation, we introduce the parameter vector $\boldsymbol{\kappa} = (\boldsymbol{a}^T, \boldsymbol{b}^T)^T \in \mathbb{R}^{2M\times 1}$. The relaxation model for the noiseless n-th T_1-weighted image is then given by

$$\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) = |\boldsymbol{a} + \boldsymbol{b} \circ e^{-\frac{TI_n}{\boldsymbol{T_1}}}|, \tag{5.1}$$

with $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) \in \mathbb{R}^{M\times 1}$, where $\circ$ and $|\cdot|$ denote the point-wise or Hadamard product, and point-wise modulus operator, respectively.

5.2.2 Motion model

In what follows, we will restrict the motion model of the unified ML framework to inter-image motion, that is, motion between the 3D T_1-weighted images, as in [Hallack et al., 2014]. In section 5.6, we further elaborate on extensions of the unified ML framework in which intra-image motion is incorporated, in particular, motion between the slices of a multi-slice T_1-weighted image.

The effect of inter-image motion is modeled by assuming that $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$ is observed in a different Cartesian coordinate system $\boldsymbol{r}^n$ for each acquisition $n = 1, ..., N$. In this work, we illustrate the joint MLE with rigid motion. Hence, the spatial point $\boldsymbol{r}^n_m$, with $m = 1, ..., M$, is related to the reference-system point $\boldsymbol{r}_m$, through a rigid transformation matrix, $\boldsymbol{M}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{4\times 4}$ (in homogeneous coordinates), parameterized by

$$\boldsymbol{\theta}_n = (t_{x_n}, t_{y_n}, t_{z_n}, \alpha_n, \beta_n, \gamma_n)^T, \tag{5.2}$$

with $t_{x_n}, t_{y_n}, t_{z_n}$ the translation parameters and α_n, β_n, γ_n the Euler angles of the three elementary rotation matrices around axis x, y and z, respectively [Goldstein et al., 2014]. In our work, the reference system $\boldsymbol{r}$ is defined similarly as the intrinsic coordinate system which MATLAB uses to represent 3D images. That is, axis x points in the direction of increasing column index while y points in the direction of increasing row index. Finally, the axis z is aligned with the direction of increasing index of the third dimension. The origin of this coordinate system is the center of the 3D image. Furthermore, in multi-slice acquisitions the axis z is aligned with the slice-encoding direction.

The noiseless T_1-weighted image observed at $\boldsymbol{r}^n$ can be modeled as the output of a linear operator that applies rigid motion, $\mathcal{H}_{\boldsymbol{\theta}_n}\{\cdot\}$, and whose input is the unobserved $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$. Because $\mathcal{H}_{\boldsymbol{\theta}_n}\{\cdot\}$ is linear, the input-output relation can be concisely written in matrix form as:

$$\tilde{\boldsymbol{f}}_n(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}) = \boldsymbol{H}_{\boldsymbol{\theta}_n} \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}), \tag{5.3}$$

where $\tilde{\boldsymbol{f}}_n(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1})$ is the motion-corrupted noiseless T_1-weighted image acquired at TI_n and $\boldsymbol{H}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{M \times M}$ is the matrix representation of the linear motion operator $\mathcal{H}_{\boldsymbol{\theta}_n}\{\cdot\}$. To design $\boldsymbol{H}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{M \times M}$, we use the method proposed in [Larkin et al., 1997], where it was demonstrated that each of the rotation matrices of $\boldsymbol{M}_{\boldsymbol{\theta}_n}$ can be decomposed as the product of three shear matrices. Each of the shearings is implemented very efficiently with Fast Fourier Transforms (FFT). Translation is implemented using an FFT as well [1]. With the FFT approach, the motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ can be shown to be unitary, which means that its inverse is given by $\boldsymbol{H}_{\boldsymbol{\theta}_n}^H$, where the superscript H denotes the Hermitian conjugate. Hence, the motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ is reversible, i.e., when applied to an image, this image can be retrieved by applying $\boldsymbol{H}_{\boldsymbol{\theta}_n}^H$ to the output of this operation. The unitarity property of the motion operator will turn out to be useful in the derivation of the joint MLE algorithm. Details of the exact analytical expression of $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ and the proof of the unitarity property are provided in subsection A.1.

5.2.3 Statistical noise model

In practice, acquired T_1-weighted images are inherently disturbed by noise. A typical data distribution for magnitude T_1-weighted images is the Rice distribution [Gudbjartsson and Patz, 1995], whose expression was given in Eq. (2.32). In this work, we will illustrate the proposed joint MLE by deriving it for the case of independent Rician distributed voxels, with different noise standard deviation σ for each voxel m and for each acquisition n. As explained in subsection 2.6.2, this is an accurate noise model for magnitude images that are reconstructed with SENSE [Aja-Fernández and Tristán-Vega, 2013]. It is also a valid noise model for magnitude images that are reconstructed with GRAPPA jointly with the SMF method. If, instead of SMF, SoS is used in combination of GRAPPA, the data distribution can be well approximated at high SNR by a Gaussian distribution with a spatially variant variance [Aja-Fernández et al., 2015b]. The derivation of the joint MLE for Gaussian distributed data will be covered in subsection 5.3.5.

[1] The FFT approach implicitly assumes that images are of limited-support, both in the frequency and space domain. Since an image cannot be of limited-support in both domains simultaneously, aliasing may occur, which appears in the form of image distortions. In practice, however, for images that contain background and are not piece-wise constant, like those that are used in this work, aliasing becomes negligible.

5.3 Joint MLE

Let $\boldsymbol{s}_n \in \mathbb{R}^{M\times 1}$, with $n = 1, ..., N$, denote an actual, noisy T_1-weighted image acquired at inversion time TI_n. Assuming Rician distributed data, it follows from Eq. (2.32) and the motion-corrupted noiseless T_1-weighted model Eq. (5.3) that the PDF of the voxels $[\boldsymbol{s}_n]_m$, $m = 1, \ldots, M$, of this image is given by

$$p_{[\boldsymbol{s}_n]_m}([\boldsymbol{s}_n]_m|[\tilde{\boldsymbol{f}}_n]_m, [\boldsymbol{\sigma}_n]_m) = \frac{[\boldsymbol{s}_n]_m}{[\boldsymbol{\sigma}_n]_m^2} e^{\frac{-([\boldsymbol{s}_n]_m^2+[\tilde{\boldsymbol{f}}_n]_m^2)}{2[\boldsymbol{\sigma}_n]_m^2}} I_0\left(\frac{[\boldsymbol{s}_n]_m[\tilde{\boldsymbol{f}}_n]_m}{[\boldsymbol{\sigma}_n]_m^2}\right) u([\boldsymbol{s}_n]_m). \tag{5.4}$$

Furthermore, if all voxels are assumed to be independent, the joint PDF of the voxels constituting the image $\boldsymbol{s}_n$ is given by the product of the PDFs of the individual voxels, i.e., $p_{\boldsymbol{s}_n}(\boldsymbol{s}_n|\tilde{\boldsymbol{f}}_n, \boldsymbol{\sigma}_n) = \prod_{m=1}^{M} p_{[\boldsymbol{s}_n]_m}([\boldsymbol{s}_n]_m|[\tilde{\boldsymbol{f}}_n]_m, [\boldsymbol{\sigma}_n]_m)$. Similarly, the joint PDF of the supposedly independent voxels of a set of N T_1-weighted images $\{\boldsymbol{s}_n\}_{n=1}^{N}$ is given by

$$p_{\boldsymbol{s}}(\boldsymbol{s}|\tilde{\boldsymbol{f}}, \boldsymbol{\sigma}) = \prod_{n=1}^{N} p_{\boldsymbol{s}_n}(\boldsymbol{s}_n|\tilde{\boldsymbol{f}}_n, \boldsymbol{\sigma}_n) \tag{5.5}$$

with $\boldsymbol{s} = (\boldsymbol{s}_1^T, \ldots, \boldsymbol{s}_N^T)^T$, $\tilde{\boldsymbol{f}} = (\tilde{\boldsymbol{f}}_1^T, \ldots, \tilde{\boldsymbol{f}}_N^T)^T$ and $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1^T, \ldots, \boldsymbol{\sigma}_N^T)^T$. Note that this joint PDF depends on the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, ..., \boldsymbol{\theta}_N^T)^T$, $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$ via $\tilde{\boldsymbol{f}}$ and can hence be written as $p_{\boldsymbol{s}}(\boldsymbol{s}|\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T}_1, \boldsymbol{\sigma})$. To construct the MLE of these parameters, the *likelihood function* must be derived. The likelihood function is obtained from the joint PDF, Eq. (5.5), by replacing the independent variables $\boldsymbol{s}$ by the actual acquired voxel intensity values - that is, by numbers - and the supposedly fixed, exact parameters $\boldsymbol{\theta}, \boldsymbol{\kappa}$ and $\boldsymbol{T_1}$ by independent variables. The likelihood function is, therefore, a function of the parameters considered as independent variables and is parametric in the acquired voxel intensities, from now on called *observations* [van den Bos, 2007]. To express this, the likelihood function is written as $L(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s})$. Strictly speaking, the likelihood function also depends on $\boldsymbol{\sigma}$. However, in our work, we assume that $\boldsymbol{\sigma}$ can be estimated prior to the construction of the joint MLE using tailored noise estimation described in subsection 2.6.3. Hence, we omit the explicit σ-dependence in the notation.

To simplify the notation, let us define the parameter vector $\boldsymbol{\tau} = (\boldsymbol{\theta}^T, \boldsymbol{\kappa}^T, \boldsymbol{T_1}^T)^T$. The joint MLE $\hat{\boldsymbol{\tau}}_{\mathrm{ML}}$ of $\boldsymbol{\tau}$ from the observations $\boldsymbol{s}$ is that value of $\boldsymbol{\tau}$ that maximizes the likelihood function $L(\boldsymbol{\tau}|\boldsymbol{s})$, or equivalently, minimizes the so-called negative log-likelihood function $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\tau}|\boldsymbol{s}) \triangleq -\log L(\boldsymbol{\tau}|\boldsymbol{s})$ with respect to $\boldsymbol{\tau}$, i.e.,

$$\hat{\boldsymbol{\tau}}_{\mathrm{ML}} = \arg\min_{\boldsymbol{\tau}} \mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\tau}|\boldsymbol{s}). \tag{5.6}$$

It follows from Eq. (5.5) that $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\tau}|\boldsymbol{s})$ can be written as

$$\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\tau}|\boldsymbol{s}) = \sum_{n=1}^{N} \mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n), \tag{5.7}$$

with $\mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n) = -\log p_{\boldsymbol{s}_n}(\boldsymbol{s}_n|\tilde{\boldsymbol{f}}_n, \boldsymbol{\sigma}_n)$.

As the joint MLE can not be found analytically, one has to resort to numerical optimization algorithms. In order to solve this very-large-scale optimization problem, a cyclic block-coordinate descent (cBCD) method was used [Fessler and Kim, 2011]. cBCD methods work by iteratively minimizing the cost function $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s})$ with respect to a subset of the optimization variables, holding the remaining variables fixed, where in each iteration, the roles of the optimization and fixed variables are reversed [Hong et al., 2016]. The utility of the cBCD algorithm relies on a smart selection of the subset of optimization variables. In the case of our joint MLE, this subset is chosen to contain the motion parameters or the relaxation parameters. In this way, the very-large-scale optimization problem is separated into more easily solvable problems.

Indeed, alternating between the motion estimation problem and the relaxation estimation problem, the joint MLE is found in an efficient way. Moreover, the cBCD method assures that $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s})$ decreases at every iteration [Fessler and Kim, 2011]. Therefore, convergence to at least a local minimum is guaranteed [Fan et al., 1998]. In summary, the cBCD-based joint MLE is obtained by the following iterative recursive procedure:

$$\hat{\boldsymbol{\theta}}^{(t+1)} = \arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \hat{\boldsymbol{\kappa}}^{(t)}, \hat{\boldsymbol{T}_{\mathbf{1}}}^{(t)}|\boldsymbol{s}_n), \tag{P.1}$$

$$\{\hat{\boldsymbol{\kappa}}^{(t+1)}, \hat{\boldsymbol{T}_{\mathbf{1}}}^{(t+1)}\} = \arg\min_{\boldsymbol{\kappa}, \boldsymbol{T_1}} \sum_{n=1}^{N} \mathcal{L}_{\boldsymbol{s}_n}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n), \tag{P.2}$$

with $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}_{\text{ini}}$, $\hat{\boldsymbol{\kappa}}^{(0)} = \boldsymbol{\kappa}_{\text{ini}}$ and $\hat{\boldsymbol{T}_{\mathbf{1}}}^{(0)} = \boldsymbol{T_1}_{\text{ini}}$ the initial values of the parameters $\boldsymbol{\theta}$, $\boldsymbol{\kappa}$ and $\boldsymbol{T}_1$, respectively. This procedure is terminated when the number of iterations exceeds $t_{\max}$ or the relative decrease $\mathcal{E}^{(t)}$ of $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T}_1|\boldsymbol{s})$ between two consecutive iterations is below a fixed tolerance, $\mathcal{E}_{\min}$. A detailed description of the problems (P.1) and (P.2) is provided in subsections 5.3.1 and 5.3.2, respectively. Furthermore, a pseudo-code of the joint MLE algorithm is shown in subsection 5.3.3, whereas its implementation is described in subsection 5.3.4.

5.3.1 Problem 1 (P.1): estimation of the motion parameters

The motion estimation problem adopts a particularly simple structure when the relaxation parameters are fixed. If no dependence of $\{\boldsymbol{\theta}_n\}_{n=1}^{N}$ through index n is assumed, as is done here, the minimization can be decoupled into N optimization problems, which can be implemented very efficiently by parallel operations. The parameters $\{\boldsymbol{\theta}_n\}_{n=1}^{N}$ enter the linear motion operator in a non-trivial way, which renders the analytical calculation of the derivatives infeasible. Fortunately, the low dimensionality of the N minimization problems, involving only six variables each, allows us to use a derivative-free optimization method. In our approach, simulated annealing (SA) minimization is performed [Press et al., 2007], which is known for its ability to avoid being trapped in local minima and its robustness to functions

with complex structure [Eglese, 1990, Goffe et al., 1994, Press et al., 2007, Bertsimas and Nohadani, 2010]. Each time, motion estimates from the previous iteration t are used as initial guesses.

5.3.2 Problem 2 (P.2): estimation of the relaxation parameters

In contrast to P.1, the relaxation parameter estimation problem is a very-large-scale minimization problem. Derivative-free algorithms such as SA are therefore impractical to use. Quasi-Newton algorithms, on the other hand, may produce critical memory storage problems, due to the large dimensionality of the Hessian matrix approximation. Furthermore, line searches dramatically slow down the algorithm [Nocedal and Wright, 2006].

The optimization method of choice for solving this minimization problem is the use of a Majorize-Minimize (MM) framework [Hunter and Lange, 2004], which yields a voxel-wise independent algorithm, allowing a computationally efficient implementation. Here, we sketch the basics of the MM algorithm applied to the estimation problem at hand. The reader is referred to [Hunter and Lange, 2004] for further details.

Let $J(\boldsymbol{\kappa}, \boldsymbol{T_1}) = \sum_{n=1}^{N} \mathcal{L}_{\boldsymbol{s}_n}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{s}_n)$ be the cost function of P.2 that we seek to minimize w.r.t. $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$. MM algorithms are defined through the following recursive minimization problem:

$$\{\boldsymbol{\kappa}^{k+1}, \boldsymbol{T_1}^{k+1}\} = \arg\min_{\boldsymbol{\kappa}, \boldsymbol{T_1}} G(\boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k), \tag{5.8}$$

where $G(\boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ is a new user-designed cost function. It can be demonstrated that the sequence of iterates $\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k$ obtained from Eq. (5.8) converges to a local minimum of $J(\boldsymbol{\kappa}, \boldsymbol{T_1})$ if $G(\boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ is what is called in the optimization literature a *surrogate function* of $J(\boldsymbol{\kappa}, \boldsymbol{T_1})$. The properties that characterize a surrogate function are

1. $J(\boldsymbol{\kappa}, \boldsymbol{T_1}) \leq G(\boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) \quad \forall \boldsymbol{\kappa}, \boldsymbol{T_1}$
2. $J(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = G(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k | \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$.

Obviously, to really benefit from the MM algorithm, the surrogate function $G(\boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ should be easier to minimize than the original cost function $J(\boldsymbol{\kappa}, \boldsymbol{T_1})$. A key result is presented by Varadarajan and Haldar [Varadarajan and Haldar, 2015], showing that the following function is a valid surrogate function of $-\log p_{[\boldsymbol{s}_n]_m}([\boldsymbol{s}_n]_m | [\tilde{\boldsymbol{f}}_n]_m, [\boldsymbol{\sigma}_n]_m)$:

$$\frac{1}{2[\boldsymbol{\sigma}_n]_m^2} \left([\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1})]_m - [\breve{\boldsymbol{s}}_n^k]_m \right)^2 + C_n^m(k) \tag{5.9}$$

with $C_n^m(k)$ a constant independent of $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$,

$$\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1}) = \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}), \tag{5.10}$$

and

$$[\breve{\boldsymbol{s}}_n^k]_m = [\boldsymbol{s}_n]_m \frac{I_1\left(\frac{[\boldsymbol{s}_n]_m[\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)},\boldsymbol{\kappa}^k,\boldsymbol{T}_1^k)]_m}{[\boldsymbol{\sigma}_n]_m^2}\right)}{I_0\left(\frac{[\boldsymbol{s}_n]_m[\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)},\boldsymbol{\kappa}^k,\boldsymbol{T}_1^k)]_m}{[\boldsymbol{\sigma}_n]_m^2}\right)} \tag{5.11}$$

with $I_1(\cdot)$ the first order modified Bessel function of the first kind.

Note that $\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)},\boldsymbol{\kappa},\boldsymbol{T_1})$ describes the motion-corrupted synthetic T_1-weighted image, whereas $\breve{\boldsymbol{s}}_n^k$, from now on called the Bessel image, is the actual acquired image $\boldsymbol{s}_n$ corrected with a Bessel correction factor. A surrogate function for $\mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n,\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{s}_n)$ is now obtained by summing Eq. (5.9) over m, i.e.,

$$G_n(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k) = ||\boldsymbol{W}_n^{1/2}\big(\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa},\boldsymbol{T_1}) - \breve{\boldsymbol{s}}_n^k\big)||_2^2 + C_n(k) \tag{5.12}$$

with $\boldsymbol{W}_n = \text{diag}\{\frac{1}{2\boldsymbol{\sigma}_n^2}\}$. By summing $G_n(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$ over n, we would obtain a global surrogate function for $J(\boldsymbol{\kappa},\boldsymbol{T_1})$. At this point, the main benefit of applying the MM framework is that the relaxation estimation problem has now been transformed into a collection of weighted non-linear least squares (NLLS) problems, avoiding complicated minimization with Bessel functions. However, we still can further simplify the problem and convert it into a fully separable (voxel-wise independent) NLLS problem. To that end, we apply another surrogate function $G_n^*(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$ to each $G_n(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$. It is easy to demonstrate that if $G_n^*(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$ is a surrogate function for $G_n(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$, it is also a valid surrogate function for $\mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n,\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{s}_n)$. Therefore, we finally define $G(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$ as

$$G(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k) = \sum_{n=1}^{N} G_n^*(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k). \tag{5.13}$$

The choice for $G_n^*(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$ is a separable quadratic surrogate (SQS) function [Muckley et al., 2015], which, when applied to our problem at hand, takes the expression,

$$G_n^*(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k) = ||\boldsymbol{f}_n(\boldsymbol{\kappa},\boldsymbol{T_1}) - \boldsymbol{\rho}_n(\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)||_2^2 + C_n^*(k), \tag{5.14}$$

with

$$\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k) = \boldsymbol{f}_n(\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k) + \sigma^*\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n\big(\breve{\boldsymbol{s}}_n^k - \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)\big) \tag{5.15}$$

and $\sigma^* = 2\min_{n,m}[\boldsymbol{\sigma}_n]_m^2$. The complete derivation of $G_n^*(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$ can be found in section A.2 of appendix A. With $G_n^*(\boldsymbol{\kappa},\boldsymbol{T_1}|\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$, minimization of Eq. (5.13) is nothing more than fitting the relaxation model $\boldsymbol{f}_n(\boldsymbol{\kappa},\boldsymbol{T_1})$ to the "residual" images $\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k,\boldsymbol{T_1}^k)$ with $n = 1,...,N$ in a least squares sense. Therefore, it is a completely separable optimization problem and hence it can be implemented in parallel for every voxel m. This is the main distinct characteristic of the joint MLE that we present in this work, which makes it an efficient method to be used in practice. Once the model-fitting is performed, the new iterate serves to update the "residual" images. This process is repeated until $k > k_{\max}$ or $\mathcal{EJ}^{(k)} < \mathcal{EJ}_{\min}$ where $\mathcal{EJ}^{(k)}$ is the analogous of $\mathcal{E}^{(t)}$ for $J(\boldsymbol{\kappa},\boldsymbol{T_1})$. The final iterate yields the new $\hat{\boldsymbol{\kappa}}^{(t+1)}$ and $\hat{\boldsymbol{T}}_{\boldsymbol{1}}^{(t+1)}$, which are then used as input in the motion estimation problem P.1.

5.3.3 Initialisation

Although convergence to a local minimum is guaranteed, convergence to the global minimum, which corresponds with the MLE estimate, cannot be proved, since $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s})$ is non-convex. To increase the chances of finding the global minimum, providing good initial values is crucial. In our approach, initial values were obtained using the conventional approach (CA), which consists of image registration prior to voxel-wise relaxation model fitting. A sequential estimation to initialize unified motion model-based approaches was also used in [Fogtmann et al., 2014], with very good results, and we found it a robust method to initialize our joint MLE. In particular, firstly, the initial motion parameters $\boldsymbol{\theta}_{\text{ini}}$ were obtained by registering the set of T_1-weighted images based on maximization of MI between the images [Mattes et al., 2003, Pluim et al., 2003]. All images were pairwise registered to the reference system $\boldsymbol{r}$ with a pyramidal multi-resolution scheme of three levels. Bi-cubic interpolation was used. The number of iterations of the internal optimization algorithm was set to a very high value (> 900) to ensure convergence of the motion parameter estimation. Next, the relaxation parameters $\boldsymbol{\kappa}_{\text{ini}}$ and $\boldsymbol{T_1}_{\text{ini}}$ were voxel-wise estimated from the registered images using the MLE based on Rician distributed data [Sijbers et al., 1999]. To compute the MLE, a Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-newton algorithm [Nocedal and Wright, 2006] was used with exact analytical derivatives. The spatially variant standard deviation, which is required for the MLE, was estimated with the method of [Aja-Fernández et al., 2015b]. By using the CA as initialization, we have invariably found that the estimated T_1 maps, $\boldsymbol{T_1}_{\text{ML}}$, are superior in terms of accuracy and RMSE compared to those obtained with the CA. Furthermore, simulation results have shown that the joint MLE is stable and robust to occasional inaccuracies in the CA-based initial motion parameters.

Pseudo-code of the joint MLE algorithm is presented in Algorithm 1, and an illustrative flow-chart is shown in Fig. 5.1. In practice, the joint MLE requires as input, apart from the initial motion and relaxation parameters, an estimate of the spatially variant standard deviation $\boldsymbol{\sigma}$, $\hat{\boldsymbol{\sigma}}$. Such an estimate is obtained with the method of [Aja-Fernández et al., 2015b] directly applied to the acquired images $\{\boldsymbol{s}_n\}_{n=1}^{N}$.

5.3.4 Parameters selection, code implementation and computational cost

The proposed joint MLE was implemented in MATLAB and run on a computer with an Intel i7-4770K processor consisting of four cores at 3.5 GHz. The machine had 32 GB of RAM. The SA algorithm of P.1 was implemented using the MATLAB routine `simulannealbnd` with the default parameters. The NLLS fitting of P.2 was performed by the MATLAB routine `lsqnonlin` using the Levenberg-Marquardt (LM) [Nocedal and Wright, 2006] method, also with the default parameters. The tolerance criteria and the maximum number of iterations to halt the algorithm were chosen to be $\mathcal{EJ}_{\min} = 10^{-2}$ and $\mathcal{E}_{\min} = 10^{-3}$, and $t_{\max} = k_{\max} = 10$, respectively.

Algorithm 1 Pseudo-code of the joint MLE algorithm.

initialize $t = 0$, $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}$, $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}_{\text{ini}}$, $\hat{\boldsymbol{T}}_{\mathbf{1}}^{(0)} = \boldsymbol{T}_{\mathbf{1}\,\text{ini}}$ and $\hat{\boldsymbol{\kappa}}^{(0)} = \boldsymbol{\kappa}_{\text{ini}}$
while $\mathcal{E}^{(t)} \geq \mathcal{E}_{\min}$ and $t < t_{\max}$ **do**
    Solve (P.1) to get $\hat{\boldsymbol{\theta}}^{(t+1)}$
    Set $k = 0$ (P.2 begins)
    $\boldsymbol{\kappa}^k = \boldsymbol{\kappa}^{(t)}$ and $\boldsymbol{T}_{\mathbf{1}}^k = \boldsymbol{T}_{\mathbf{1}}^{(t)}$
    while $\mathcal{EJ}^{(k)} \geq \mathcal{EJ}_{\min}$ and $k < k_{\max}$ **do**
        Calculate $\hat{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}^k, \boldsymbol{T}_{\mathbf{1}}^k)$ with Eq. (5.10)
        Calculate $\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T}_{\mathbf{1}}^k)$ with Eq. (5.15)
        Voxel-wise NLLS fitting of $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T}_{\mathbf{1}})$ to $\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T}_{\mathbf{1}}^k)$ so as to get $\boldsymbol{\kappa}^{k+1}$ and $\boldsymbol{T}_{\mathbf{1}}^{k+1}$
        Calculate $\mathcal{EJ}^{(k)}$ and set $k = k + 1$
    end while
    Set $\hat{\boldsymbol{\kappa}}^{(t+1)} = \boldsymbol{\kappa}^k$ and $\hat{\boldsymbol{T}}_{\mathbf{1}}^{(t+1)} = \boldsymbol{T}_{\mathbf{1}}^k$ (P.2 ends)
    Calculate $\mathcal{E}^{(t)}$ and set $t = t + 1$
end while
$\boldsymbol{\theta}_{\text{ML}} = \hat{\boldsymbol{\theta}}^{(t)}$ and $\boldsymbol{T}_{\mathbf{1}\text{ML}} = \hat{\boldsymbol{T}}_{\mathbf{1}}^{(t)}$

To exploit the highly parallelizable structure of the joint MLE, MATLAB parallel computing tools were used to estimate $\boldsymbol{\theta}_n$ for each value of n separately. Similarly, the voxel-wise NLLS relaxation model fitting was performed in a parallel manner by dividing the spatial grid into eight non-overlapping 3D blocks. Finally, a mask was used to avoid calculation of the relaxation parameters in background areas, hence speeding up the implementation.

The computational time per iteration of the joint MLE algorithm is dominated by the voxel-wise T_1 fitting, which depends linearly on the number of voxels M, depending in turn on the Field-of-View (FOV) and the voxel size. With relatively little effort to optimize our code, and using the MATLAB parallel tools mentioned earlier, the voxel-wise T_1 fitting took approximately 8 min to process a series of $N = 8$ T_1-weighted images with $M \approx 10^5$ voxels. Overall, with the tolerance criterion described above, the average number of total iterations (external plus internal) was roughly 15, providing precise and accurate $\boldsymbol{\theta}_{\text{ML}}$ and $\boldsymbol{T}_{\mathbf{1}\text{ML}}$ in an average time of 2.2 hours. Note that migration of the MATLAB code to C++ would produce a much faster implementation, especially if multi-threading is used for the highly parallelizable relaxation estimation problem [Fogtmann et al., 2014].

5.3.5 Gaussian approximation for GRAPPA+SoS

When GRAPPA reconstructed data is combined with SoS, the statistical distribution of the composite magnitude image can be well approximated with a non-stationary nc-χ distribution, where both the variance and the (effective) degrees-of-freedom parameter, L_{eff}, are spatially non-stationary (i.e., vary from voxel to voxel) [Aja-

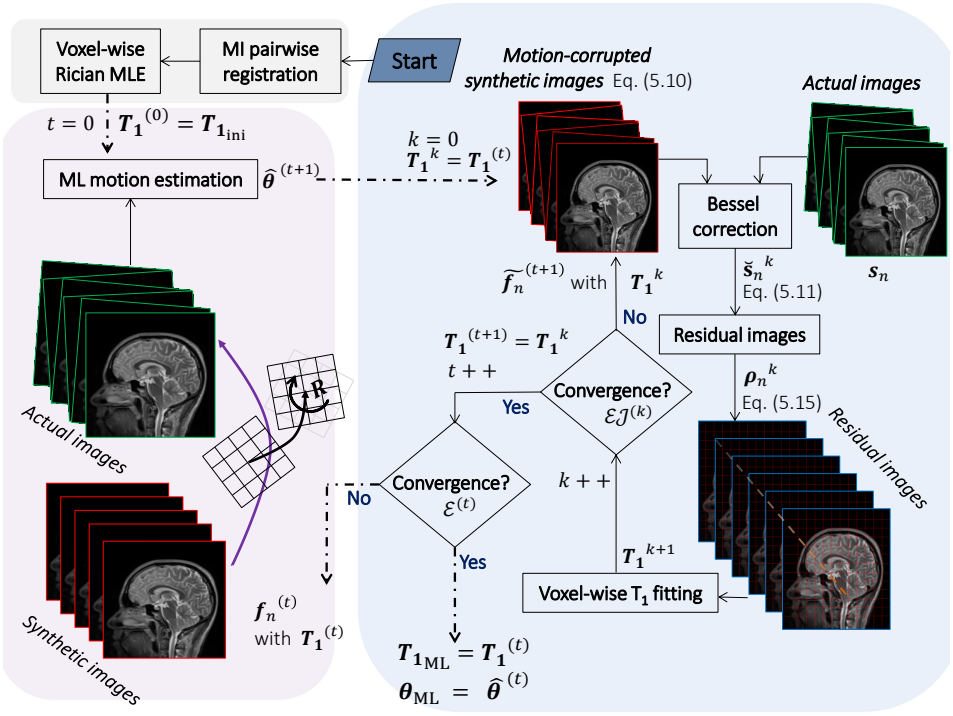


Figure 5.1: Flow-chart of the joint MLE algorithm.

Fernández and Tristán-Vega, 2013]. Since the MM framework was originally developed for the nc-χ distribution [Varadarajan and Haldar, 2015], the application of the proposed joint MLE is straightforward provided an estimate of L_{eff} for every voxel is available. Unfortunately, practical estimators of spatial maps of L_{eff} are, to the authors' knowledge, not yet available in the literature. Nevertheless, for high SNR, a Gaussian distribution with spatially variant σ has been proved to be an accurate model in replacement of the nc-χ model [Aja-Fernández et al., 2015b]. In this case, the joint MLE is even simpler than it was for the Rician case. Indeed, it can easily be shown that

$$\mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \hat{\boldsymbol{\kappa}}^{(t)}, \hat{\boldsymbol{T}}_{\boldsymbol{1}}^{(t)} | \boldsymbol{s}_n) = ||\boldsymbol{W}_n^{1/2}\big(\boldsymbol{H}_{\boldsymbol{\theta}_n} \boldsymbol{f}_n(\hat{\boldsymbol{\kappa}}^{(t)}, \hat{\boldsymbol{T}}_{\boldsymbol{1}}^{(t)}) - \boldsymbol{s}_n\big)||_2^2 \tag{5.16}$$

and

$$\mathcal{L}_{\boldsymbol{s}_n}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{s}_n) = ||\boldsymbol{W}_n^{1/2}\big(\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) - \boldsymbol{s}_n\big)||_2^2. \tag{5.17}$$

The same SA optimization algorithm as before can be used to minimize Eq. (5.16) for solving the motion estimation problem (P.1). To simplify the minimization of Eq. (5.17) for solving the relaxation parameter estimation problem (P.2), we can apply directly the SQS function on $\mathcal{L}_{\boldsymbol{s}_n}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{s}_n)$, avoiding the Bessel correction step. Indeed, the relaxation parameter estimation problem is again a NLLS fitting of $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$ to a different $\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ where just $\check{\boldsymbol{s}}_n^k$ in Eq. (5.15) has to be replaced by $\boldsymbol{s}_n$, the actual acquired images.

5.4 Experiments

The proposed joint motion and T_1 MLE was validated using both simulated and real data. Moreover, its performance was compared to that of the CA with MI-based registration [Pluim et al., 2003], and a recently proposed model-based approach of Hallack et al. [Hallack et al., 2014].

5.4.1 Simulated T_1-weighted data

A set of 3D IR-SE T_1-weighted images $\{\boldsymbol{s}_n\}_{n=1}^N$ affected by inter-image motion (as in Eq. (5.3)) and noise was simulated from ground truth T_1 and proton density maps. The ground truth T_1 map was created from the BrainWeb anatomical model, using reported T_1 values in human brain tissue at 3T [Gold et al., 2004, Wright et al., 2008]. For the three main brain tissues, white matter, grey matter and cerebrospinal fluid (CSF), the reference values were 838 ms, 1607 ms, and 4300 ms, respectively. The ground truth proton density map was created in a similar fashion. The size of both 3D maps was $111 \times 93 \times 71$ with an isotropic voxel size of 1.5 mm. From these maps, a set of IR-SE T_1-weighted images was simulated based on [Barral et al., 2010] with TR/TE = 10000 /14 ms, and $N = 8$ logarithmically equidistant inversion times $\{TI_n\}_{n=1}^N$ between 20 ms and 8000 ms. The three consecutive RF pulse angles were set to 180°, 90° and 180°. In the next step, we randomly generated ground truth motion parameters $\{\boldsymbol{\theta}_n\}_{n=1}^N$. Each of the six rigid motion parameters followed an independent Gaussian Random Walk (RW) [Fogtmann et al., 2014] along the temporal dimension n.

More precisely, the motion parameters were generated as

$$\boldsymbol{\theta}_n = \boldsymbol{c} + \boldsymbol{\theta}_{n-1} + \boldsymbol{w}_n, \tag{5.18}$$

where $\boldsymbol{c} \in \mathbb{R}^{6\times 1}$ denotes the motion drift and $\boldsymbol{w}_n \in \mathbb{R}^{6\times 1}$ a vector valued, zero mean, Gaussian random variable with covariance matrix $\boldsymbol{\Sigma} = \sigma_{\text{RW}}^2 \boldsymbol{I}$, with σ_{RW} the standard deviation of each of the elements of $\boldsymbol{w}_n$ and $\boldsymbol{I}$ the 6×6 identity matrix. The reference system $\boldsymbol{r}$ was chosen to be $\boldsymbol{r}_1$, hence $\boldsymbol{\theta}_1 = \boldsymbol{0}$. Finally, to account for noise, Rician distributed images $\{\boldsymbol{s}_n\}_{n=1}^N$ were simulated [Sijbers and den Dekker, 2004] with spatially variant noise maps. Synthetic spatially variant noise maps were generated based on a realistic noise pattern that was presented in [Aja-Fernández et al., 2015b]. This pattern was derived from a real parallel MRI acquisition [Aja-Fernández et al., 2014].

The proposed joint motion and T_1 MLE was compared to that of the CA with MI-based registration [Pluim et al., 2003] (the initialization technique for the joint MLE) and a recently proposed inter-image model-based approach of Hallack et al. [Hallack et al., 2014]. The MI-based rigid registration step of the CA was implemented using the first image of the series as a reference. Details of the implementation were already given in subsection 5.3.3. The remaining MI registration parameters were set to those provided in the MATLAB built-in code. Hallack's method was implemented by following the guidelines provided in [Hallack et al., 2014]. Just

like the joint MLE, it was initialized with the CA. The parameters $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$ were estimated with the LM algorithm. Hallack's algorithm was stopped when either the decrease of the cost function between iterations was below $\mathcal{E}_{\min}$, or the number of iterations exceeded $t_{\max}$. Two types of simulation experiments were performed:

5.4.1.1 Exp.1: Performance as a function of SNR

In a first set of experiments (Exp.1), the performance of the joint MLE as a function of the SNR of the T_1-weighted image data set was tested. To that end, motion parameters $\{\boldsymbol{\theta}_n\}_{n=1}^N$ were generated with $\sigma_{\text{RW}} = 0.4$ mm/degree and no drift. After fixing the motion parameters, T_1-weighted image data sets with SNR values between 20 and 100 were simulated, where the SNR is defined as the spatial mean of the ratio of the reference noise-free T_1-weighted image and the standard deviation noise map of this reference image. For each SNR, $N_{\text{MC}} = 15$ MC simulations were generated.

5.4.1.2 Exp.2: Performance as a function of the type of motion

In a second set of experiments (Exp.2), the performance of the joint MLE was evaluated for different types of motion (and fixed SNR = 40). For completeness, we also included a case without motion.

a) *Low Abrupt motion (LA-m).* The motion parameters were generated as in Exp. 1, i.e., without drift/tendency and with $\sigma_{\text{RW}} = 0.4$ mm/degree.

b) *High Abrupt motion (HA-m).* The motion parameters were generated without drift/tendency and with $\sigma_{\text{RW}} = 1.5$ mm/degree.

c) *Rotational motion (R-m).* The motion parameters were generated with drift parameter vector $\boldsymbol{c} = (0, 0, 0, 0.5, 0.5, 0.5)^T$ and $\sigma_{\text{RW}} = 0.4$ mm/degree. Note that it follows from Eq. (5.18) and Eq. (5.2) that in this scenario only the rotation parameters are affected by drift.

d) *Translational motion (T-m).* The motion parameters were generated with $\boldsymbol{c} = (0.5, 0.5, 0.5, 0, 0, 0)^T$ and $\sigma_{\text{RW}} = 0.4$ mm/degree.

e) *No motion (No-m).* No motion was applied, i.e., $\tilde{\boldsymbol{f}}_n(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}) = \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$.

For all types of motion, extreme values as well as mean values for each of the six rigid parameters along the temporal dimension n are shown in the second column of Table 5.1. For each type of motion, $N_{\text{MC}} = 15$ simulations, i.e., noisy T_1-weighted image data sets, were generated (with fixed motion parameters and SNR= 40).

To assess the ability of each method to estimate the T_1 map, the following performance measures were used:

(a) *Relative bias.* The bias quantifies the *accuracy* of the estimator. For each voxel, the relative sample bias was calculated as $(\bar{\hat{T}}_1 - T_1)/T_1$, where $\bar{\hat{T}}_1$ is the sample mean of the N_{MC} estimates $\hat{T}_1$ and T_1 is the true value. A measure of the overall accuracy of the T_1 map was obtained by calculating the spatial mean of the absolute value of the relative sample bias, using a brain mask to avoid the skull.

b) *Relative standard deviation.* The standard deviation quantifies the *precision* of the estimator. For each voxel the relative sample standard deviation was calculated as $\mathrm{std}(\hat{T}_1)/T_1$, and an overall precision measure was obtained by taking the spatial mean of these relative sample standard deviations, using the same brain mask.

c) *Relative root-mean-square error (relative RMSE).* The RMSE is a measure that incorporates both accuracy and precision. For each voxel, the relative sample RMSE was calculated as $\sqrt{\overline{(\hat{T}_1 - T_1)^2}}/T_1$. An overall RMSE measure was obtained by calculating the spatial mean of these relative sample RMSE values, again within the same brain mask.

To assess the ability of each method to estimate motion, the following performance measures were used:

d) *Relative motion error*, defined as

$$||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2 / ||\boldsymbol{\theta}||_2. \tag{5.19}$$

e) *Motion component relative bias*, defined as

$$\frac{1}{N}\sum_{n=1}^{N} |(\overline{[\hat{\boldsymbol{\theta}}_n]_j} - [\boldsymbol{\theta}_n]_j)/[\boldsymbol{\theta}_n]_j|, \tag{5.20}$$

with $[\boldsymbol{\theta}_n]_j$ the jth component of $\boldsymbol{\theta}_n$ and $\overline{[\hat{\boldsymbol{\theta}}_n]_j}$ the sample mean of the N_{MC} estimates $[\hat{\boldsymbol{\theta}}_n]_j$.

5.4.2 Ground-truth based real experiment

In order to assess the performance of the joint MLE with an actual T_1-weighted data set corrupted by motion, we performed a controlled experiment. The experiment comprised the acquisition of two data sets. Firstly, we acquired an IR T_1-weighted data set of a (static) watermelon. In a second step, we deliberately introduced motion between the acquisition of each of the acquired 2D multi-slice T_1-weighted images. In particular, we manually translated and rotated the watermelon after the complete acquisition for a fixed TI. From this data, estimated T_1 maps were obtained with the CA, Hallack's method and the joint MLE. We then quantitatively compared these T_1 maps to the estimated T_1 map of the first dataset, which was unaffected by motion and hence can be considered as a reasonable ground-truth.

Both IR T_1-weighted data sets were acquired with a 3T MRI scanner (MAGNETOM Prisma$^{\text{fit}}$, Siemens) using a 32-channel head coil. The IR T_1-weighted data sets comprised $N = 8$ T_1-weighted multi-slice images whose inversion times were logarithmically spaced between 300 and 6000 ms. For each inversion time, we acquired a 2D multi-slice image with a 2D interleaved multi slice IR TSE sequence [Bernstein et al., 2004] (see also subsection 3.4.1.1). The Echo Train Length (ETL) was 10 and TR/TE = 7920/8.8ms. Each multi-slice image was acquired within approximately 3 min. The total scan time was about 24 minutes. The acquisition plane was axial and the acquisition matrix was $256 \times 256 \times 40$ with an anisotropic voxel size of $1 \times 1 \times 4$ mm^3 and no slice gap. Magnitude data were reconstructed using the SENSE method (acceleration factor of 2) [Pruessmann et al., 1999]. The coordinate system of the image with the highest inversion time was chosen to be the reference system $\boldsymbol{r}$. An estimated SNR of the reference image, as defined in subsection 5.4.1.1, was found to be 26. The estimation of the SNR was performed with the Expectation-Maximization-based method of De Vore [DeVore et al., 2000] adapted for local estimation [Aja-Fernández et al., 2015b] to work with one single image (3×3 neighborhoods were used). The ground-truth T_1 map was estimated with a voxel-wise MLE based on Rician data [Sijbers et al., 1999], where the noise standard deviation map was estimated in a preprocessing step using the method described in [Aja-Fernández et al., 2015b].

5.4.3 In vivo T_1-weighted data

We validated the joint MLE with two in vivo human brain data sets suffering from involuntary patient motion.

5.4.3.1 In vivo axial human brain data

An IR T_1-weighted data set of a healthy 26-year old male volunteer was acquired with a 3T MRI scanner (MAGNETOM Skyra, Siemens) using a 20-channel head coil. For each inversion time, we acquired a 2D multi-slice image with an interleaved 2D multi-slice IR TSE sequence [Park et al., 1985, Bernstein et al., 2004]. The sequence parameters were: ETL = 4 and TR/TE = 8040/18 ms. Each multi-slice image was acquired within 2.5 min approximately. The IR T_1-weighted data set comprised $N = 7$ T_1-weighted multi-slice images whose inversion times were logarithmically spaced between 50 and 3200 ms. The total scan time was about 19 minutes. The acquisition plane was axial and the acquisition matrix was $128 \times 128 \times 25$ with an anisotropic voxel size of $1.9 \times 1.9 \times 6$ mm^3 and a slice gap of 10%. The SENSE method was employed to reconstruct the magnitude data with an acceleration factor of 3. Noise maps were obtained with the method of [Aja-Fernández et al., 2015b]. We estimated an SNR of 24.3 with the method of [DeVore et al., 2000]. In this case, the reference image was the one with lowest inversion time.

5.4.3.2 In vivo sagittal human brain data

An IR T_1-weighted data set of a healthy 28-year old male volunteer was acquired with a 3T MRI scanner (MAGNETOM Prisma^{fit}, Siemens) using a 32-channel head coil. As in the previous in vivo experiment, we acquired for each inversion time a 2D multi-slice image with an interleaved 2D multi-slice IR TSE sequence [Bernstein et al., 2004]. The sequence parameters were: ETL = 10 and TR/TE = 5000/4.8 ms. Each multi-slice image was acquired within 2 min approximately. The IR T_1-weighted data set comprised $N = 14$ T_1-weighted images whose inversion times were logarithmically spaced between 100 and 3000 ms, giving a total acquisition time of 28 min. The acquisition plane was sagittal and the acquisition matrix was $256 \times 256 \times 40$ with an anisotropic voxel size of $1 \times 1 \times 4$ mm^3 and no slice gap. Magnitude data were reconstructed with the GRAPPA method with SoS reconstruction (acceleration factor of 3) [Griswold et al., 2002]. The image with the lowest inversion time was chosen as a reference. We estimated an SNR of 55 with a locally adapted ML estimator (3×3 neighborhoods) assuming an nc-χ distribution. Due to the high SNR, we relied on results of [Aja-Fernández et al., 2015b] and used the version of the joint MLE algorithm adapted for spatially variant Gaussian noise (subsection 5.3.5). Noise maps were obtained with the method of [Aja-Fernández et al., 2015b]. The CA was implemented with a Gaussian MLE where the noise standard deviation was estimated with the method described in [Aja-Fernández et al., 2015b].

5.5 Results

5.5.1 Simulated T_1-weighted data

5.5.1.1 Exp.1: Performance as a function of SNR

Overall relative T_1 bias, standard deviation and RMSE results are shown in Fig.5.2.(a-c). For the whole range of SNR, the joint MLE allows a much more accurate estimation of the T_1 map than Hallack's method and especially than the CA. In terms of precision, CA obtains the best result, followed by the joint MLE and Hallack's method. However, in terms of the overall RMSE, the joint MLE performs best for all values of the SNR. Furthermore, the box-plot shown in Fig. 5.2.(d) demonstrates the superiority of the proposed joint MLE in terms of motion estimation. To complement the results, maps of the absolute value of the relative sample bias for the three methods are shown in Fig. 5.3.(b-d), along with the simulated ground truth in Fig. 5.3.(a). A close look at the bias maps corroborates the poor performance of the CA compared to the joint MLE. It is also clearly seen that the bias map of Hallack's method presents much higher values than that of the joint MLE, especially in white/grey matter surroundings.

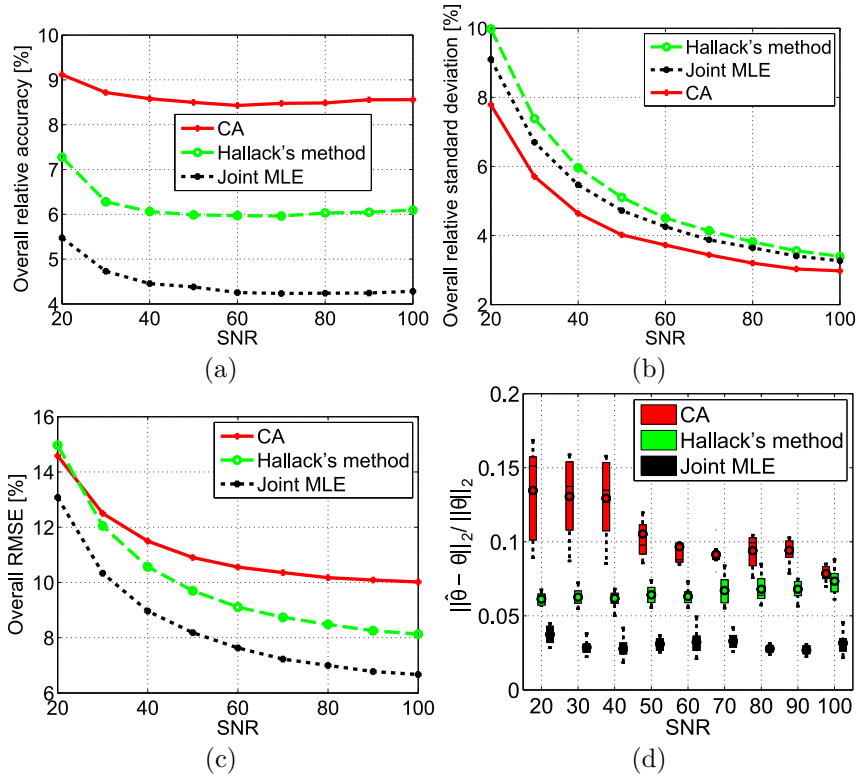


Figure 5.2: Results of Exp.1: (a) relative T_1 bias, (b) relative T_1 standard deviation, (c) relative T_1 RMSE and (d) box plots of the relative motion error, as a function of the SNR.

5.5.1.2 Exp.2: Performance as a function of the type of motion

Bar charts representing the overall T_1 accuracy, precision and RMSE for the four cases of motion and the no-motion scenario are shown in Fig. 5.4. In light of these results, it can be concluded that the joint MLE yields the most accurate T_1 maps in all the four considered motion scenarios, followed by Hallack's method. Furthermore, the performance of all methods seems to be fairly insensitive to the type of motion considered. Even though the highest precision was consistently obtained with the CA, its overall relative RMSE is much higher compared to Hallack's method and especially to the joint MLE, which again produces the best T_1 maps in RMSE sense. The case of no-motion is particularly interesting. In such a scenario, the CA performance drastically improves in terms of accuracy and RMSE, though its precision decreases. In the absence of motion, the error propagation of the CA approach is negligible, and hence, the overall relative accuracy of the CA approaches the one obtained with the joint MLE, which is yet the highest. The decrease in precision of the CA can be understood as follows. In the absence of motion, the interpolation (and hence smoothing) effects that are inherent to the registration step of the conventional two-step approach become marginal, hence

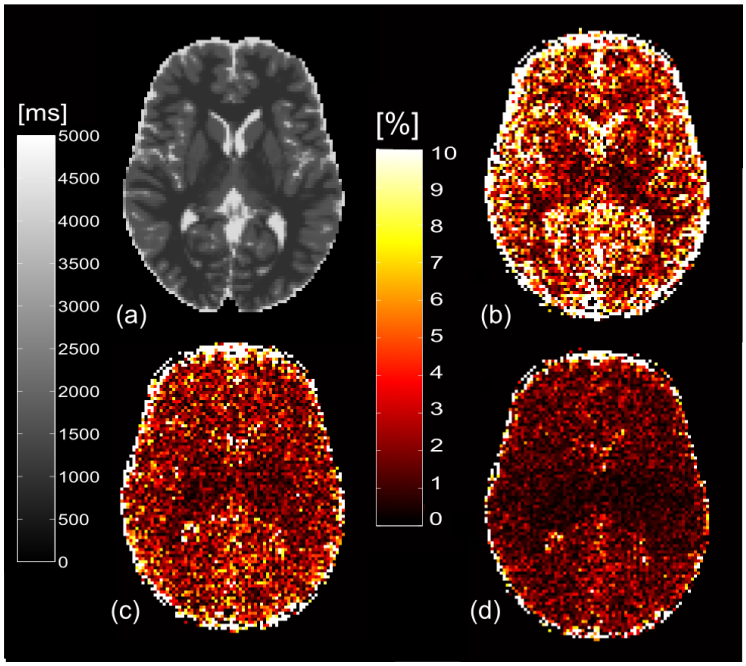


Figure 5.3: Results of Exp.1: (a) Mid-axial slice of the 3D ground truth T_1 map, and maps of the absolute value of the relative sample bias [in %] for (b) CA, (c) Hallack's method and (d) joint MLE, for SNR = 40.

not contributing to a reduction in the variability of the estimates.

The motion component relative bias for each of the six components are reported in Table 5.1. For the no-motion case, the motion component relative bias is not well-defined (division by zero). Instead, we report the motion component absolute bias.

The best results are highlighted in shaded gray. In 27 of 30 cases, the joint MLE achieved the highest accuracy in both the translation and rotation parameters. Sometimes, this improvement is even more than 5-fold compared to the CA. In general, Hallack's method provides more accurate T_1 estimates than CA, which is in agreement with previously reported results [Hallack et al., 2014, Huizinga et al., 2016]. Nonetheless, further substantial improvement can be obtained if the joint MLE is used. To illustrate the quality of the motion estimation, we have shown graphs, as a function of n, of the ground-truth and estimated motion parameters for one of the rotational motion (R-m) simulations in Fig. A.1.

It is also important to notice that the occasionally poor CA-based motion initialization does not prevent the joint MLE from producing the most accurate motion estimates. This highlights another feature of the joint MLE: it is fairly robust to scenarios where the CA-based motion initialization is relatively poor.

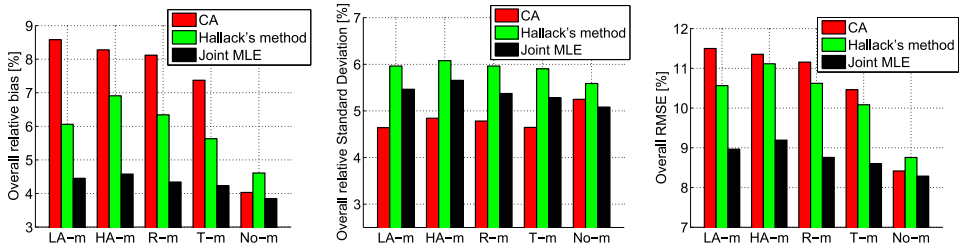


Figure 5.4: Results of Exp.2: (a) relative T_1 bias, (b) relative T_1 standard deviation and (c) relative T_1 RMSE.

5.5.2 Ground-truth based real experiment

A top-axial and a mid-axial slice of the estimated T_1 maps for the CA, Hallack's method and the joint MLE are shown in Fig. 5.5.(c-e) and Fig. 5.5.(m-o), whereas the ground truth T_1 map and the T_1 map estimated without motion correction are displayed in Fig. 5.5(a-b) for the top-axial slice and in Fig. 5.5.(k-l) for the mid-axial slice. From this experiment, it can be observed that more detailed T_1 maps can be obtained with the joint MLE in comparison to Hallack's method and especially to the CA. Aside from the presence of a large number of outliers in the T_1 maps obtained with the CA and Hallack's method, which are drastically reduced with the joint MLE, fine structural details seem better preserved with our proposed method. This observation is confirmed by inspecting the magnified regions. The heterogeneity of the T_1 values in those regions, as noticed from the ground-truth T_1 map, is better maintained with the joint MLE. See for instance the delineation of low T_1 value structures in Fig. 5.5.(f-j). Note as well that artifacts in the T_1 maps, as shown in Fig. 5.5.(p-t) (green arrow), are considerably mitigated with the joint MLE. Quantitative validation of the estimated T_1 maps was based on spatial maps of the absolute value of the relative errors [%]. Those maps are shown in Fig. 5.6. In accordance with our previous discussion, the spatial distributions of the relative errors further indicate the good performance of the joint MLE in comparison to competing methods. It is manifestly clear that the error maps of Hallack's method and the CA present much higher values than that of the joint MLE. To complement the quantitative analysis, we calculated an overall relative error, within a mask neglecting the background, in a similar fashion as done in Exp.1. Numerical results are in agreement with the observation made from the spatial maps. We found that the joint MLE produced the T_1 map with the lowest overall relative error. Indeed, the overall relative error for the without motion correction case, the CA, Hallack's method and the joint MLE was 87%, 20.2%, 19.7% and 15.1%, respectively. To further complement the quantitative analysis, graphs of the motion parameter estimates for the three methods are shown in Fig. A.2.

Table 5.1: Results of Exp.2: for four types of motion (column 1), the maximum and mean values of the motion parameters (column 2), and the motion component relative bias for each of the six components for CA, Hallack's and the joint MLE method (columns 3-5) are shown. For the no-motion case, the motion component absolute bias is reported instead of the motion component relative bias since the latter metric is not well-defined for parameters that are equal to zero.

Type	Motion (max / mean)	CA	Hallack	Joint MLE
LA-m	t_x (2.5 / 0.98 mm)	8.4 %	4 %	**1.1 %**
	t_y (1.4 / 0.88 mm)	1.4 %	2.1 %	**0.4 %**
	t_z (1 / 0.38 mm)	33.5 %	30.9 %	**5.3 %**
	α (0.2 / 0.003 degree)	63.5 %	20.6 %	**9.5 %**
	β (-0.2 / -0.06 degree)	90 %	43.8 %	**19.1 %**
	γ (0.57 / 0.26 degree)	75.4 %	57.5 %	**6.8 %**
HA-m	t_x (-9.4 / -4.4 mm)	2.3 %	3.9 %	**0.1 %**
	t_y (-2.3 / 0.08 mm)	30 %	12.9 %	**2.4 %**
	t_z (4.6 / 2 mm)	27 %	19.8 %	**13.5 %**
	α (-2.1 / -0.51 degree)	**4.9 %**	10.5 %	6.4 %
	β (3.7 / 1.9 degree)	1.4 %	2 %	**0.9 %**
	γ (1.3 / 0.5 degree)	**4.9 %**	18.2 %	19.6 %
R-m	t_x (-0.8 / -0.4 mm)	5.5 %	3.4 %	**1.2 %**
	t_y (1.1 / 0.8 mm)	2.3%	3.1 %	**1.1 %**
	t_z (0.3 / 0.02 mm)	466 %	166 %	**29.8 %**
	α (3.2 / 1.5 degree)	2.1 %	1.3 %	**0.7 %**
	β (3.1 / 1.5 degree)	1.6 %	**0.9 %**	1.1 %
	γ (1.9 / 1.1 degree)	2.9 %	5 %	**0.13 %**
T-m	t_x (2.9 / 1.4 mm)	1.3 %	1.9 %	**0.8 %**
	t_y (2.8 / 1.9 mm)	1.2 %	3 %	**1.1 %**
	t_z (2.1 / 1.2 mm)	2 %	2.6 %	**0.8 %**
	α (-0.3 / -0.2 degree)	26.7 %	64.7 %	**20.7 %**
	β (-0.8 / -0.3 degree)	121.15 %	17.2 %	**4.7 %**
	γ (-0.5 / -0.2 degree)	91.1 %	27.1 %	**21.7 %**
No-m	t_x (0 / 0 mm)	0.012	0.032	**0.009**
	t_y (0 / 0 mm)	0.112	0.077	**0.03**
	t_z (0 / 0 mm)	0.078	0.159	**0.086**
	α (0 / 0 degree)	0.0281	0.0133	**0.01**
	β (0 / 0 degree)	0.083	0.057	**0.019**
	γ (0 / 0 degree)	0.0201	0.0183	**0.011**

5.5.3 In vivo T_1-weighted data

5.5.3.1 In vivo axial human brain data

A mid-axial and a top-axial slice of the estimated T_1 map for the three methods are shown in Fig. 5.7.(b-d) and Fig. 5.7.(f-h), respectively. The estimated T_1 maps without motion correction are shown in Fig. 5.7.(a) and Fig. 5.7.(e). The presence of outliers in the T_1 map when motion correction is not applied is not completely

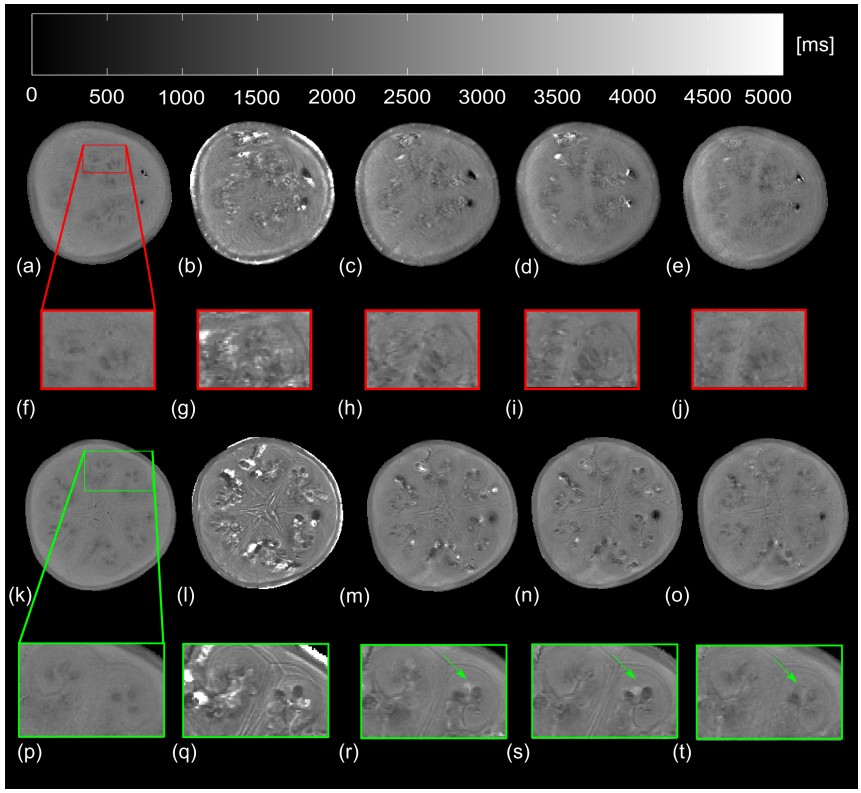


Figure 5.5: Two axial slices of a watermelon T_1 map. Top-axial slice: (a) ground-truth (no motion), (b) without motion correction, (c) CA, (d) Hallack's method, (e) joint MLE. Magnified regions are shown in (f-j). Mid-axial slice: (k) ground-truth (no motion), (l) without motion correction, (m) CA, (n) Hallack's method, (o) joint MLE. Magnified regions are shown in (p-t).

avoided with the CA. Indeed, it can be clearly observed that outliers are still present, in particular at the interfaces between ventricles and white matter (green arrow) for the mid-axial slice, and in the interfaces between white and grey matter for the top-axial slice Fig. 5.7(f). The T_1 map produced by Hallack's method seems free from outliers in the mid-axial but not in top-axial slice Fig. 5.7(g) (green arrow). The joint MLE provides T_1 maps which do not suffer from this issue. The motion parameter estimates obtained with the three methods are shown in Fig. A.3.

5.5.3.2 In vivo sagittal human brain data

Two mid-sagittal slice of the estimated T_1 map for the three methods are shown in Fig. 5.8(b-d) and Fig. 5.8(j-l), respectively. In Fig. 5.8(a) and Fig. 5.8(i) estimated T_1 maps for the no motion correction case are presented.

In this case, the estimated T_1 map with the CA is almost free from outliers, which are widespread when no motion correction is accomplished, see magnified regions in Fig. 5.8.(e-f) and Fig. 5.8.(m-n). However, the CA sacrifices the final resolution of the T_1 map. It is clear that in the T_1 map obtained with Hallack's method and especially in that obtained with the joint MLE, structure details are better defined and contours better delineated, particularly in the interfaces between white/grey matter and CSF. Visual differences between the T_1 maps provided by Hallack's method and the joint MLE can be discerned as well. For example, the CSF infiltrating the cortical folds (see Fig. 5.8.(g-h)) seems better defined with the joint MLE. In addition, yet a (reduced) number of outliers can be detected in the T_1 map estimated with the Hallack's method. See for example, Fig. 5.8.(o). Such outliers, as in the in vivo axial experiment, are not present provided the joint MLE is applied. The motion parameter estimates of this experiment are shown in Fig. A.4.

5.6 Discussion and Future work

We presented a unified model-based approach for simultaneous motion correction and T_1 mapping that jointly estimates the motion parameters and the T_1 map using a maximum likelihood estimator (MLE). The proposed joint MLE possesses optimal statistical properties, which are shared by neither the conventional two-step approach (image registration prior to T_1 estimation) nor other heuristic integrated model-based methods.

Using realistic MC simulation experiments, it was shown that the proposed joint MLE outperforms existing T_1 mapping methods in terms of both accuracy and RMSE, next to providing more accurate motion parameter estimates. Results of the controlled experiment based on ground-truth real data are in line with the findings of MC simulations. We have shown that detailed and meaningful T_1 map can be recovered with the joint MLE under the influence of manually induced, severe motion. Quantitative comparison against a ground-truth T_1 map demonstrates the superior quality in T_1 map restoration compared to CA and Hallack's method.

Furthermore, the optimal unified ML framework has been validated with in vivo human brain data experiments, suffering form involuntary motion. From these experiments, it has become evident that motion correction is indispensable in T_1 mapping, even when subject motion is relatively small. Interestingly, yet recognizing the limitation of visual assessment in quantitative MRI, some of the rigorously derived statistical conclusions from the MC simulations can be noticed in the two whole brain human data sets. For instance, the arguably poorer motion estimation performance of the CA compared to model-based registration approaches, already reported by [Hallack et al., 2014, Huizinga et al., 2016] and confirmed by our MC results (see Fig. 5.2.(d) and Table 5.1), may be the cause of the presence of outliers in the estimated T_1 map and the loss of fine details. The presence of a small number of outliers in the estimated T_1 map with Hallack's method can

be attributed to its non-optimal/heuristic design. While improvement in motion estimation compared to CA has been demonstrated, optimality in terms of T_1 estimation cannot be theoretically and empirically guaranteed. In contrast, our ML framework combines, in a single integrated approach, the benefits of model-based registration techniques with optimal T_1 map restoration based on statistical theory, where the noise statistics are properly accounted for.

On top of that, our careful algorithm design avoids heavy computational burden. Note that the voxel-wise T_1 fitting task, which contributes most to the computational cost of the proposed joint MLE, is often also included in the iterative loop of other model-based integrated methods. Consequently, the computation time per iteration is comparable.

The proposed ML framework can be extended in different ways without compromising its optimal statistical properties. Extension to other MRI sequences or modalities is straightforward by substituting a properly modified parametric signal model for Eq. (5.1). Such potential extensions include T_2 and T_2^* mapping [Sijbers et al., 1998c]. Moreover, multi-component T_2 mapping would benefit as well from the proposed ML framework [Björk et al., 2016]. Furthermore, it is worthwhile mentioning that when the joint MLE proposed in this work is applied to the particular case of T_1 mapping using the spoiled gradient recalled echo (SPGR) sequence, its computational efficiency can even be further improved by using the recently proposed fast non-linear least squares T_1 estimator NOVIFAST [Ramos-Llordén et al., 2016] for solving the voxel-wise NLLS problems in P.2.

Extensions towards the inclusion of different types of motion, e.g., non-rigid motion, are also possible but require further study, which is considered future work. In this work, we have assumed a motion model which accounts for inter-image motion, that is, motion between each of the 3D T_1-weighted images. Although this model left aside intra-motion effects, which is known to affect the k-space reconstruction process, we have not observed any derived ghosting artifacts in the in vivo reconstructed T_1-weighted images. It should be noted that, to deal with such kind of motion, navigators [Welch et al., 2002] or advanced k-space reconstruction methods with motion correction [Cordero-Grande et al., 2016] can be applied in conjunction with the joint MLE. However, though intra-image motion may be alleviated with such techniques, image registration, that is, inter-image motion correction, will remain necessary, which further emphasizes the relevance of the unified ML framework. Finally, it is relatively straightforward to extend the ML framework to cope with inter-slice motion, that is, motion occurring between the acquisition of 2D slices of a T_1-weighted dataset. An outlook to such an extension, which is especially relevant for image acquisition methods that acquire 3D volumes slice by slice sequentially, such as Echo Planar Imaging (EPI) sequences [Fogtmann et al., 2014], is given in subsection A.4.

5.7 Conclusions

In quantitative MR T_1 mapping, it is common practice to register the T_1-weighted images prior to T_1 map estimation. However, as demonstrated in this chapter, this conventional two-step approach lacks high accuracy motion estimation and leads to biased T_1 estimates. Hence, we have proposed a rigorous unified framework for simultaneous motion and T_1 estimation using a Maximum Likelihood (ML) estimator. It has been demonstrated that the proposed joint MLE outperforms the conventional approach as well as a recently proposed model-based method [Hallack et al., 2014] in terms of motion and T_1 estimation accuracy and RMSE. Our ML framework, which uses an efficient algorithm, has been validated in a controlled experiment with real T_1-weighted data and also with two in vivo human brain data sets. We believe that the unified ML framework possesses serious advantages over the conventional approach to replace it in clinical scenarios where precise and accurate T_1 estimates are the ultimate goal.

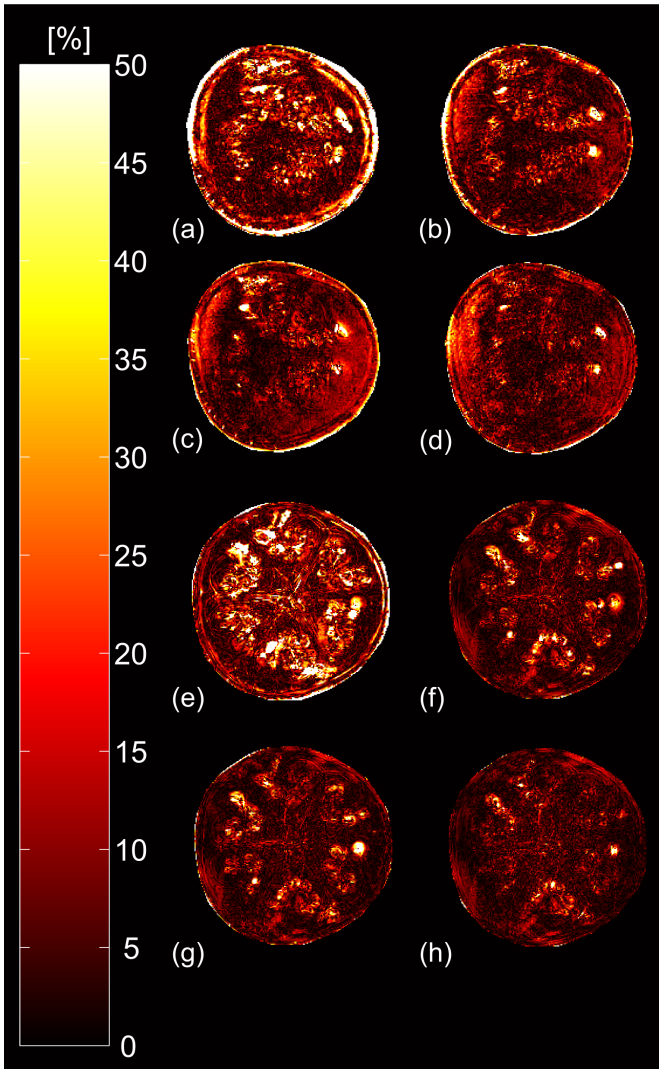


Figure 5.6: Maps of the absolute value of relative error with respect to the ground-truth T_1 map. Top-axial slice: (a) without motion correction, (b) CA, (c) Hallack's method, (d) joint MLE. Mid-axial slice: (e) without motion correction, (f) CA, (g) Hallack's method, (h) joint MLE.

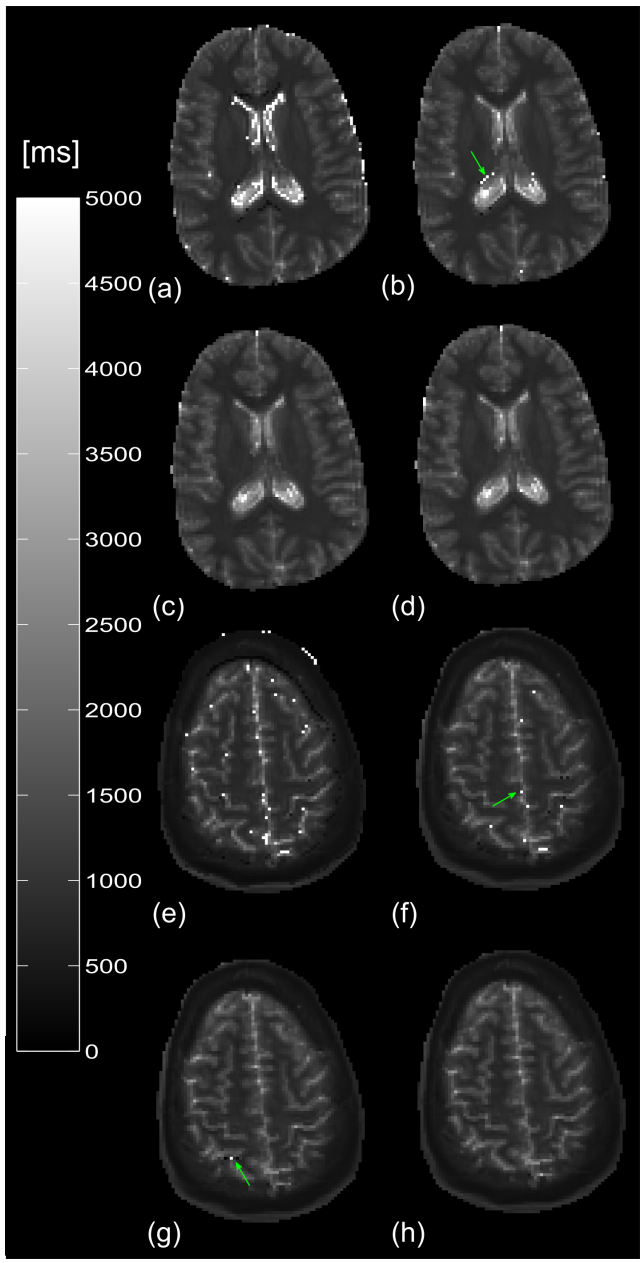


Figure 5.7: Two axial slices of a whole human brain T_1 map. Mid-axial slice: (a) without motion correction, and corrected with (b) CA, (c) Hallack's method, (d) joint MLE. Top-axial slice: (e) without motion correction, and corrected with (f) CA, (g) Hallack's method, (h) joint MLE.

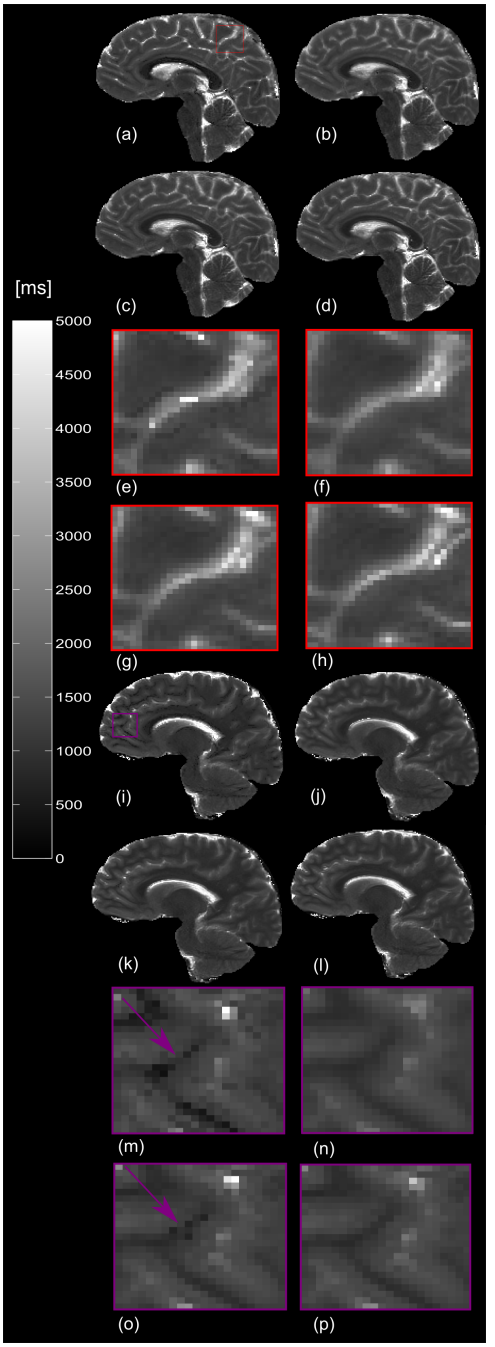


Figure 5.8: Two mid-sagittal slices of a whole human brain T_1 map. First mid-sagittal slice: (a) without motion correction, (b) CA, (c) Hallack's method, (d) joint MLE. Magnified regions are shown in (e), (f), (g) and (h), respectively. Second mid-sagittal slice: (i) without motion correction, (j) CA, (k) Hallack's method, (l) joint MLE. Magnified regions are shown in (m), (n), (o) and (p), respectively.

6

NOVIFAST: A fast algorithm for accurate and precise VFA MRI T_1 mapping

Contents

The work in this chapter is currently in review in:

- **Ramos-Llordén, G.**, Vegas-Sánchez-Ferrero, G., Björk, M., Vanhevel, F., Parizel, P. M., San José Estépar, R., den Dekker, A. J., and Sijbers, J. NOVIFAST: a fast algorithm for accurate and precise VFA MRI T_1 mapping. *IEEE Trans. Med. Imag.,* in review.

and it has been published in:

- **Ramos-Llordén, G.**, den Dekker, A. J., Björk, M., Verhoye, M., and Sijbers, J. (2016). NOVIFAST: A fast non-linear least squares method for accurate and precise estimation of T_1 from SPGR signals. *Proc. Intl. Soc. Mag. Reson. Med.*, 24:2820.

6.1 Introduction

Quantitative T_1 mapping is a Magnetic Resonance Imaging (MRI) technique that deals with the estimation of the spin-lattice relaxation time (T_1) in biological tissues [Tofts, 2004]. The spin-lattice relaxation time has proved to be an excellent biomarker in a broad range of diseases, such as multiple sclerosis [Larsson et al., 1989], epilepsy [Conlon et al., 1988] and Alzheimer's disease [Erkinjuntti et al., 1987]. Due to the large spectrum of potential applications, there is an increasing interest in turning quantitative T_1 mapping into a mature and robust MR modality that can be routinely used in clinical practice. In this effort, the longstanding goal is to provide accurate high-resolution spatial maps of T_1 in a short time frame [Deoni et al., 2003]. To that end, a plethora of T_1 mapping techniques have been proposed during the last decades (see [Stikov et al., 2015] for a recent review). Among them, the Variable Flip Angle (VFA) technique, also known as the variable nutation angle method, has gained increasing popularity [Deoni et al., 2003]. This is mainly due to its superior time efficiency compared to other T_1 mapping techniques, such as the traditional Inversion Recovery (IR) technique [Stikov et al., 2015].

VFA T_1 mapping consists of the acquisition of a range of steady-state spoiled gradient recalled echo (SPGR) MR images over a set of flip angles [Christensen et al., 1974, Homer and Beevers, 1985, Fram et al., 1987]. Since steady-state MR sequences can use much shorter repetition times (TR) [Nataraj et al., 2017] than classical inversion/saturation recovery sequences, high resolution T_1 maps can be acquired in real-time clinical acquisition [Deoni et al., 2003]. Importantly, the fact that the SPGR signal model can be easily linearized, an observation which dates back to 1977 [Gupta, 1977], has encouraged researchers to develop fast linear T_1 estimation algorithms [Deoni et al., 2003, Chang et al., 2008], rendering the estimation time of the T_1 map negligible in comparison to the acquisition time. With such computationally inexpensive algorithms, real-time high resolution T_1 mapping can be achieved. The simplicity and efficiency of the T_1 estimation step are among the main reasons why VFA T_1 mapping has drastically grown in popularity, with DESPOT1 (Driven Equilibrium Single Pulse Observation of T_1) being the most widespread algorithm [Deoni et al., 2003]. Unfortunately, the price to pay with such linear estimators is a loss of accuracy, since the linearization of the SPGR model becomes inexact in the presence of noise, thereby introducing a noise-induced bias. For that reason, some researchers still adhere to more accurate non-linear least squares (NLLS) estimators, which can be shown to have optimal statistical properties for clinically achievable signal-to-noise ratios [den Dekker and Sijbers, 2014]. However, NLLS estimators require the use of non-linear optimization algorithms, which are typically much slower than linear algorithms, can be difficult to implement, and are prone to convergence issues if not properly initialized.

Encouraged by such an apparent trade-off between speed and statistical optimality, we present a novel NLLS optimization algorithm for VFA T_1 mapping which combines the best of both worlds: high accuracy and precision in a low computation time. The method, which we dubbed NOVIFAST (being an acronym for NOn-linear VarIable Flip Angle data baSed T_1 estimator), shares the same favorable statistical properties as the de facto standard NLLS optimization algorithm, the

Levenberg-Marquard (LM) algorithm, but with a much lower computation time, being comparable to that of DESPOT1. To derive NOVIFAST, we follow a fundamentally different approach compared to conventional NLLS optimization algorithms. Instead of attempting to decrease the NLLS criterion in each iteration, we seek for the NLLS estimates by solving the root-finding problem that arises when the first-order conditions for optimality are imposed. In this derivation, we make use of the special structure of the SPGR signal model to come up with a surprisingly simple set of non-linear equations which resembles a purely linear system. Those non-linear equations are then solved iteratively in a very short number of iterations. We study NOVIFAST's convergence in a framework of fixed-point algorithms theory, thereby revealing its good convergence properties. Monte-Carlo based simulations are used to evaluate NOVIFAST's performance and compare it with that of NLLS optimization algorithms (i.e., the Gauss-Newton (GN) and LM algorithm), DESPOT1, and the Iterative Re-Weighted Linear Least Squares (IRWLLS) method proposed by Chang *et al.* [Chang et al., 2008]. Finally, NOVIFAST is validated with a set of in-vivo human brain SPGR MR images.

The structure of the remainder of the chapter is as follows. In section 6.2, the SPGR signal model is described as well as the most popular VFA T_1 mapping algorithms. In section 6.3, the complete derivation of NOVIFAST is given and its convergence properties are presented. Experiments are described in section 7.4 and the corresponding results are presented in section 6.5. Next, possible extensions are described (section 6.6), and the main conclusions are summarized at the end of the chapter (section 6.7).

6.2 Background

6.2.1 SPGR signal model

The basis of VFA SPGR method was already described in subsection 3.4.2. Here, we repeat them for clarity. The VFA SPGR method for T_1 mapping is based on the acquisition of a set of MR images where the images are acquired with different/variable non-zero flip angles (FAs), but constant TR and TE. Since the images are differently T_1-weighted, T_1 mapping can be performed by voxel-wise fitting a prescribed mathematical model to the set of MR images. A very popular model for the noiseless (magnitude) steady-state SPGR signal intensity $\{s_n\}_{n=1}^N$ is the model presented in Eq. (3.10). Each of the signals samples s_n can be hence described by [Teixeira et al., 2017]:

$$s_n(K, T_1) = \frac{K(1 - e^{-\frac{\mathrm{TR}}{T_1}})\sin(\alpha_n)}{1 - e^{-\frac{\mathrm{TR}}{T_1}}\cos(\alpha_n)}, \quad \text{for } n = 1, ..., N, \tag{6.1}$$

where $\{\alpha_n\}_{n=1}^N$ denotes the FAs, with $\alpha_n \neq 180°$ for every n, and N is the total number of FAs used (or images collected). In reality, the signal $\{s_n\}_{n=1}^N$ is always corrupted by noise. Therefore, the measured noisy SPGR samples y_n, for $n = 1, ..., N$, can be considered as realizations of random variables whose

distribution depends parametrically on K and T_1. Given a set of samples $\{y_n\}_{n=1}^N$, inferring the values of the unknown underlying T_1 and K (assuming TR and α_n to be known) is an estimation problem. A wide variety of estimators may be defined. In this work, we focus on the ordinary, i.e., unweighed, NLLS estimator, as well as heuristic linear least squares (LLS) estimators, since these types of estimators are the most popular ones in VFA T_1 mapping. In the succeeding section, we review the most common NLLS optimization algorithms to solve the NLLS estimation problem, as well as the heuristic linearized variants. We discuss their pros and cons, after which we present our novel NLLS method.

6.2.2 NLLS estimation problem

For a given data set $\{y_n\}_{n=1}^N$, the ordinary NLLS estimator is defined as

$$\{\hat{K}, \hat{T}_1\} = \arg \min_{K,T_1} \sum_{n=1}^{N} \big(y_n - s_n(K, T_1)\big)^2. \tag{6.2}$$

Such an optimization problem cannot be solved analytically and hence one has to resort to numerical algorithms. A straightforward, though naive, approach to find $\hat{K}$ and $\hat{T}_1$ would be to perform a grid-search in the domain of the optimization variables. That is, the cost function of Eq. (6.2) is evaluated in a sufficiently dense grid and then, the minimum value is retained. However, the multiple function evaluations make this technique computationally demanding. Instead, optimization algorithms that attempt to find the local minima are commonly employed. Perhaps, the main representative of this class of algorithms is the full Newton algorithm [Nocedal and Wright, 2006]. A downside of this algorithm is that it is considerably time-consuming due to the need of line searches and the computation of the Hessian matrix. To solve NLLS estimation problems, it is often better to exploit the quadratic structure of the cost function, as is done by the celebrated Gauss-Newton (GN) and Levenberg-Marquard (LM) [Levenberg, 1944, Marquardt, 1963] algorithms, which are, by far, the most popular NLLS optimization algorithms. Both algorithms are briefly reviewed below.

6.2.2.1 Gauss-Newton (GN)

Let us define $\boldsymbol{r}(K, T_1) = (r_1(K, T_1), ..., r_N(K, T_1))^T$, where $r_n(K, T_1) = y_n - s_n(K, T_1)$ are the so-called residuals of the NLLS problem. The GN algorithm can be derived by linearly approximating $\boldsymbol{r}(K, T_1)$ around a given estimate (K^k, T_1^k) [Nocedal and Wright, 2006],

$$\boldsymbol{r}(K, T_1) \approx \boldsymbol{r}(K^k, T_1^k) + \boldsymbol{J_r}(K^k, T_1^k)\boldsymbol{\Delta}, \tag{6.3}$$

where $\boldsymbol{J_r}(K^k, T_1^k)$ is the Jacobian matrix of $\boldsymbol{r}(\cdot, \cdot)$, evaluated at (K^k, T_1^k). The vector $\boldsymbol{\Delta} = (K, T_1)^T - (K^k, T_1^k)^T$ is called the step vector. After substituting

Eq. (6.3) into Eq. (6.2), and equating the gradient of the cost function with respect to $\boldsymbol{\Delta}$ to zero, we arrive at the normal equations:

$$\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k)\boldsymbol{\Delta} = -\boldsymbol{J}_{\boldsymbol{r}}(K^k, T_1^k)^T \boldsymbol{r}(K^k, T_1^k), \tag{6.4}$$

with $\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k) = \boldsymbol{J}_{\boldsymbol{r}}(K^k, T_1^k)^T \boldsymbol{J}_{\boldsymbol{r}}(K^k, T_1^k)$. Provided $\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k)$ is non-singular, a condition implicitly assumed in the GN algorithm, the step vector can be obtained analytically for each iteration, which, for the dimensions of the problem at hand (2×2), can be done with a negligible computational effort. The GN algorithm can also be interpreted as an inexact full Newton algorithm where the Hessian matrix is approximated to avoid the computational burden [Nocedal and Wright, 2006]. The GN algorithm works well close to the minimum, however, it may not converge at all if it is not initialized properly [Marquardt, 1963, Dan et al., 2002]. Rigorous conditions to prove local convergence[1] are given in [Chen and Li, 2005].

6.2.2.2 Levenberg-Marquardt (LM)

The LM algorithm [Levenberg, 1944, Marquardt, 1963] can be seen as a generalization of the GN algorithm. Like GN, the LM algorithm can be derived from the normal equations (Eq. (6.4)), but this time $\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k)$ is modified to

$$\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k) + \lambda \operatorname{diag}\left(\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k)\right), \tag{6.5}$$

where λ is a user-controlled parameter which may be updated at each iteration, and $\operatorname{diag}\left(\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k)\right)$ is a diagonal matrix whose entries are the elements on the diagonal of $\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k)$. When $\lambda \to 0$, the calculated step vector $\boldsymbol{\Delta}$ approaches the GN step obtained by solving Eq. (6.4). When λ increases, $\boldsymbol{\Delta}$ approaches a gradient descent step where each component is weighted according to $\operatorname{diag}\left(\tilde{\boldsymbol{J}}_{\boldsymbol{r}}(K^k, T_1^k)\right)$. At initial iterations, λ is set to a non-zero value. The LM algorithm behaves as a modified gradient descent method at early iterations, assuring the desirable descent-property, but mimics GN as it gets closer to the minimum. This is achieved by progressively decreasing λ towards zero. LM is slightly more computationally demanding than GN, but it converges for initializations that are far away from the solution, where GN often fails [Marquardt, 1963]. In this sense, LM shows global convergence properties and is, therefore, the preferred method of choice in common NLLS problems.

6.2.3 Heuristic linearized variants

Both GN and LM make use of the quadratic expression of the NLLS cost function but, as general-purpose NLLS algorithms, they do not consider the particular

[1] In optimization theory's parlance [Chen and Li, 2005, Lanckriet and Sriperumbudur, 2009], local convergence means that convergence is assured if the initial approximation is close enough to a stationary point. In contrast, global convergence implies convergence for any arbitrary initialization. Note that this terminology does not deal with the character of stationary points, e.g., local/global minima, but just refers to convergence to stationary points.

structure of the SPGR signal model described by Eq. (6.1) . Indeed, by dividing both sides of Eq. (6.1) by $\sin(\alpha_n)$, and then rearranging the equation, it can be shown that Eq. (6.1) can be written as [Gupta, 1977]:

$$\frac{s_n}{\sin\alpha_n} = c_2\frac{s_n}{\tan\alpha_n} + c_1 \text{ for } n = 1, ..., N, \tag{6.6}$$

with

$$c_1 = K(1 - c_2), \quad c_2 = \exp(-\mathrm{TR}/T_1), \tag{6.7}$$

and where it has been assumed that $\alpha_n \neq 90°$ and $\alpha_n \neq 270°$ in order for $\tan\alpha_n$ to be well-defined. The linear relation of Eq. (6.6) (w.r.t. $c_1 > 0$ and $c_2 > 0$) is substantially exploited in several LLS-based algorithms, which we briefly review hereunder.

6.2.3.1 DESPOT1

The DESPOT1 algorithm, proposed by Deoni *et al.* [Deoni et al., 2003], aims to estimate c_1 and c_2 in a linear regression framework. Indeed, in the absence of noise, $\frac{s_n}{\tan\alpha_n}$ may be considered as the regressor variables and $\frac{s_n}{\sin\alpha_n}$ as the regressand counterparts. In the noisy case, $\{s_n\}_{n=1}^N$ is not observable and hence DESPOT1 replaces $\{s_n\}_{n=1}^N$ by $\{y_n\}_{n=1}^N$, after which c_1 and c_2 are estimated by linear least squares (LLS) regression. DESPOT1 presents the lowest computational cost since it is an analytical estimator (i.e., it gives the estimates in a closed-form expression). However, as it strongly relies on the linear relation described by Eq. (6.6), and this relation no longer holds when $\{s_n\}_{n=1}^N$ is replaced by $\{y_n\}_{n=1}^N$, a bias is introduced, which becomes more pronounced at low signal-to-noise ratio [Chang et al., 2007].

6.2.3.2 Iterative Re-Weighted Linear Least Squares (IRWLLS)

To increase the accuracy of DESPOT1, Chang *et al.* [Chang et al., 2008] suggested to include a weighting function for each of the samples $n = 1, ..., N$. The linear regression approach is then transformed into a weighted linear regression approach, where the weights are derived based on uncertainty propagation theory. Since the optimal weights depend on the parameters to be estimated, Chang *et al.* proposed an iterative approach where the parameters c_1 and c_2 are estimated using a weighted linear least squares estimator of which the weights are updated iteratively. This iterative re-weighted linear least squares (IRWLLS) method effectively outperforms DESPOT1 in terms of accuracy. However, there is no guarantee that the final estimates share the same optimal properties as a pure NLLS approach. Furthermore, the convergence of the iterative procedure is not guaranteed. Consequently, the algorithm may in fact diverge.

6.3 Method

NOVIFAST is derived directly from the NLLS problem (Eq. (6.2)), and should thus be classified as an NLLS optimization algorithm. Interestingly, in its derivation, the particular structure of the SPGR model is exploited, which resembles the approach that was used in the linearized, but sub-optimal DESPOT1 and IRWLLS algorithms. This way, NOVIFAST combines the best of both worlds: the accuracy/precision of NLLS estimators and the computational speed of heuristic linear algorithms. Since NOVIFAST is not a general-purpose NLLS-based algorithm, we present its complete derivation below. In section 6.3.2, the main features of the algorithm are described and its pseudo-code is presented. Convergence properties are briefly covered in section 6.3.2 and thoroughly studied in subsection B.2.

6.3.1 Derivation of NOVIFAST: exploiting the structure of the SPGR model

With a slight abuse of notation, let us define $s_n(\boldsymbol{c}) = s_n(c_1, c_2)$ as the SPGR model of Eq. (6.1), but parameterized by linear coefficients c_1 and c_2 given by Eq. (6.7). With this change of variables, we can derive the NLLS estimates for $\boldsymbol{c} = (c_1, c_2)^T$ and from these calculate the NLLS estimates $\hat{K}$ and $\hat{T}_1$. The NLLS cost function w.r.t. $\boldsymbol{c}$ is given by

$$\mathcal{L}(\boldsymbol{c}) = \sum_{n=1}^{N} \left(y_n - s_n(\boldsymbol{c})\right)^2, \tag{6.8}$$

and the first-order conditions necessary to locate the stationary points can be shown to be:

$$\frac{\partial \mathcal{L}}{\partial c_1} = 2 \sum_{n=1}^{N} \left(y_n - s_n(\boldsymbol{c})\right) \frac{\sin \alpha_n}{1 - c_2 \cos \alpha_n} = 0 \tag{6.9}$$

$$\frac{\partial \mathcal{L}}{\partial c_2} = -2 \sum_{n=1}^{N} \left(y_n - s_n(\boldsymbol{c})\right) \frac{s_n(\boldsymbol{c}) \cos \alpha_n}{1 - c_2 \cos \alpha_n} = 0. \tag{6.10}$$

The NLLS estimate, by construction, must fulfill this system of non-linear equations. In principle, the complexity to solve Eq. (6.9) and Eq. (6.10) makes this approach unattractive, and probably that is the main reason that existing NLLS optimization algorithms for VFA T_1 mapping do not attempt to follow this line of thinking. However, the particular structure of $s_n(\boldsymbol{c})$, a rational function of sines and cosines, in combination with certain algebraic rules, as those exposed by Dimitrov and Kamenski for rational functions in chemical kinetics [Dimitrov and Kamenski, 1991], yields a surprisingly simple set of non-linear equations, which can be solved iteratively. Indeed, if we substitute $s_n(\boldsymbol{c})$ in Eq. (6.9) by its rational expression, and use the term $1 - c_2 \cos \alpha_n$ as common denominator, which is never zero, we get

$$\sum_{n=1}^{N} \frac{y_n(1 - c_2 \cos \alpha_n) - c_1 \sin \alpha}{\left(1 - c_2 \cos \alpha_n\right)^2} \sin \alpha_n = 0. \tag{6.11}$$

By rearranging and canceling terms, Eq. (6.9) can be written as

$$c_1 \sum_{n=1}^{N} \frac{\sin^2 \alpha_n}{\left(1 - c_2 \cos \alpha_n\right)^2} + c_2 \sum_{n=1}^{N} \frac{y_n \sin \alpha_n \cos \alpha_n}{\left(1 - c_2 \cos \alpha_n\right)^2} = \sum_{n=1}^{N} \frac{y_n \sin \alpha_n}{\left(1 - c_2 \cos \alpha_n\right)^2}. \quad (6.12)$$

Similarly, Eq. (6.10) can be expressed as

$$c_1 \sum_{n=1}^{N} \frac{s_n(\boldsymbol{c}) \sin \alpha_n \cos \alpha_n}{\left(1 - c_2 \cos \alpha_n\right)^2} + c_2 \sum_{n=1}^{N} \frac{y_n s_n(\boldsymbol{c}) \cos^2 \alpha_n}{\left(1 - c_2 \cos \alpha_n\right)^2} = \sum_{n=1}^{N} \frac{y_n s_n(\boldsymbol{c}) \cos \alpha_n}{\left(1 - c_2 \cos \alpha_n\right)^2}. \quad (6.13)$$

If we define the $(N \times 1)$ vectors $\boldsymbol{a}, \tilde{\boldsymbol{a}}, \boldsymbol{b}$ and $\boldsymbol{z}$ as

$$\boldsymbol{a} = \left(\frac{y_1 \cos \alpha_1}{1 - c_2 \cos \alpha_1}, \frac{y_2 \cos \alpha_2}{1 - c_2 \cos \alpha_2}, ..., \frac{y_N \cos \alpha_N}{1 - c_2 \cos \alpha_N} \right)^T, \quad (6.14)$$

$$\tilde{\boldsymbol{a}} = \left(\frac{s_1(\boldsymbol{c}) \cos \alpha_1}{1 - c_2 \cos \alpha_1}, \frac{s_2(\boldsymbol{c}) \cos \alpha_2}{1 - c_2 \cos \alpha_2}, ..., \frac{s_N(\boldsymbol{c}) \cos \alpha_N}{1 - c_2 \cos \alpha_N} \right)^T, \quad (6.15)$$

$$\boldsymbol{b} = \left(\frac{\sin \alpha_1}{1 - c_2 \cos \alpha_1}, \frac{\sin \alpha_2}{1 - c_2 \cos \alpha_2}, ..., \frac{\sin \alpha_N}{1 - c_2 \cos \alpha_N} \right)^T, \quad (6.16)$$

$$\boldsymbol{z} = \left(\frac{y_1}{1 - c_2 \cos \alpha_1}, \frac{y_2}{1 - c_2 \cos \alpha_2}, ..., \frac{y_N}{1 - c_2 \cos \alpha_N} \right)^T, \quad (6.17)$$

the previous system of non-linear equations is concisely expressed as

$$\underbrace{\begin{pmatrix} \langle \boldsymbol{b}, \boldsymbol{b} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ \langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle \end{pmatrix}}_{\boldsymbol{A}(\boldsymbol{c})} \underbrace{\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}}_{\boldsymbol{c}} = \underbrace{\begin{pmatrix} \langle \boldsymbol{b}, \boldsymbol{z} \rangle \\ \langle \boldsymbol{z}, \tilde{\boldsymbol{a}} \rangle \end{pmatrix}}_{\boldsymbol{v}(\boldsymbol{c})}, \quad (6.18)$$

where $\langle \cdot, \cdot \rangle$ is the usual inner-product for vectors in Euclidean spaces. Note that even though we have omitted the dependence on $\boldsymbol{c}$ in the definition of Eq. (6.14) and Eq. (6.16), the 2×2 matrix $\boldsymbol{A}(\boldsymbol{c})$ depends on the linear coefficients, and the 2×2 vector $\boldsymbol{v}(\boldsymbol{c})$ does so as well. Eq. (6.18) is the basis of NOVIFAST. Although the equation is non-linear in $\boldsymbol{c}$, it resembles a purely linear system. If the variation of $\boldsymbol{A}(\boldsymbol{c})$ and $\boldsymbol{v}(\boldsymbol{c})$ w.r.t. $\boldsymbol{c}$ were negligible, a simple 2×2 inversion technique would suffice to find the root of Eq. (6.18). Since this is not the case, a natural approach is to solve it iteratively, thereby still exploiting the semi-linear structure. We propose an iterative technique where, given that $\boldsymbol{A}(\boldsymbol{c})$ and $\boldsymbol{v}(\boldsymbol{c})$ are known, $\boldsymbol{c}$ is solved linearly, after which $\boldsymbol{A}(\boldsymbol{c})$ and $\boldsymbol{v}(\boldsymbol{c})$ are updated with the new guess. The repetition of those two steps constitutes our algorithm, which we dubbed NOVIFAST.

6.3.2 NOVIFAST: algorithm definition

Equations of the form $\boldsymbol{A}(\boldsymbol{c})\boldsymbol{c} = \boldsymbol{v}(\boldsymbol{c})$ often appear in discretization schemes for time differential equations [Weickert and Viergever, 1998]. A common method for its solution is what is called a semi-implicit technique [Weickert and Viergever, 1998]:

an iterate $\boldsymbol{c}^k$ derived from the k-th iteration is used to evaluate $\boldsymbol{A}(\boldsymbol{c})$ and $\boldsymbol{v}(\boldsymbol{c})$, and the following linear equation w.r.t. $\boldsymbol{c}^{k+1}$ is solved:

$$\boldsymbol{A}(\boldsymbol{c}^k)\boldsymbol{c}^{k+1} = \boldsymbol{v}(\boldsymbol{c}^k). \tag{6.19}$$

Cramer's rule [Robinson, 1970] allows us to obtain $\boldsymbol{c}^{k+1}$ explicitly:

$$c_1^{k+1} = \frac{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{z} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ \langle \boldsymbol{z}, \tilde{\boldsymbol{a}} \rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle \end{vmatrix}}{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ \langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle \end{vmatrix}}, \quad c_2^{k+1} = \frac{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b} \rangle & \langle \boldsymbol{b}, \boldsymbol{z} \rangle \\ \langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle & \langle \boldsymbol{z}, \tilde{\boldsymbol{a}} \rangle \end{vmatrix}}{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ \langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle \end{vmatrix}}, \tag{6.20}$$

where $|\cdot|$ denotes the determinant of a matrix, and we have made the assumption that the denominator of both expressions is non-zero for every $\boldsymbol{c}$. In subsection B.2.1 we elaborate on this assumption. We are then ready to define the NOVIFAST algorithm through the following pseudo-code (Algorithm 2):

Algorithm 2 Pseudo-code of NOVIFAST.

Given parameters: TR and flip angles $\{\alpha_n\}_{n=1}^N$
Initial values: K_{ini} and $T_{1\text{ini}}$
$c_2^0 \leftarrow \exp(-\text{TR}/T_{1\text{ini}})$
$c_1^0 \leftarrow K_{\text{ini}}(1 - c_2^0)$
$k \leftarrow 0$
$\boldsymbol{c}^k \leftarrow (c_1^k, c_2^k)^T$
while convergence is not met **do**
    Solve Eq. (6.19) with solutions of Eq. (6.20)
    $k \leftarrow k + 1$
end while
return $\hat{K} = c_1^k/(1 - c_2^k)$ and $\hat{T}_1 = -\text{TR}/\log c_2^k$

Below, we pinpoint some of the most interesting properties of NOVIFAST.

- **Convergence points are roots of Eq. (6.9) and Eq. (6.10)**
  If NOVIFAST converges to some $\boldsymbol{c}^\star$, it is necessarily a root of Eq. (6.9) and Eq. (6.10). To see this, one only needs to take limits when $k \to \infty$ on both sides of Eq. (6.19), and use the fact that both $\boldsymbol{A}(\boldsymbol{c})$ and $\boldsymbol{v}(\boldsymbol{c})$ are continuous functions of $\boldsymbol{c}$. Hence, limit point $\boldsymbol{c}^\star$ meets $\boldsymbol{A}(\boldsymbol{c}^\star)\boldsymbol{c}^\star = \boldsymbol{v}(\boldsymbol{c}^\star)$, and it is a solution of Eq. (6.18), or equivalently a root of Eq. (6.9) and Eq. (6.10). Therefore, the algorithm is well defined.

- **Good convergence properties**
  With NOVIFAST, the only required stopping criterion is to check that the norm of the difference between consecutive iterates is below a fixed tolerance. There is no need to assure the descent property. This is an advantage in terms of computation time when compared to descent-based algorithms such as GM

or LM. Indeed, in a complete metric space, if $\lim_{k\to\infty} ||\boldsymbol{c}^{k+1} - \boldsymbol{c}^k||_2 = 0$, then $\lim_{k\to\infty} \boldsymbol{c}^k = \boldsymbol{c}^\star$, which, as we have seen, is undoubtedly a root of Eq. (6.9) and Eq. (6.10). Conditions to meet $\lim_{k\to\infty} ||\boldsymbol{c}^{k+1} - \boldsymbol{c}^k||_2 = 0$ are thoroughly studied in subsection B.1.2 under the framework of fixed-point algorithms theory. Converge conditions are empirically checked in subsection B.2.2, where it is shown that they hold most likely for realistic clinically achievable signal-to-noise ratios.

- **Low cost per iteration**
  Furthermore, the cost per iteration of NOVIFAST is quite low, since it amounts to calculate two quotients of determinants of just 2×2 matrices. The cost per iteration is similar to that of IRWLLS, and also to the total cost of DESPOT1. Importantly, NOVIFAST aims to get the NLLS estimates (see first bullet point), whereas both IRWLLS and DESPOT are modified heuristic algorithms.

- **Robustness and high convergence rate**
  NOVIFAST is rather insensitive to initial values K_{ini} and $T_{1\text{ini}}$, and convergence is usually reached within 2-4 iterations with the same tolerance criterion as LM or GN algorithms. This makes NOVIFAST an ideal algorithm to be used in practice. Like LM, NOVIFAST shows global convergence properties, but it converges considerably faster. This is in agreement with results provided in [Dimitrov and Kamenski, 1996], and also with those of our previous work [Ramos-Llordén et al., 2016]. In subsection B.1.2, we provide evidence for these claims, and we experimentally check the conditions with an MC analysis in subsection B.2.2.

- **Simplicity**
  NOVIFAST does not need any tuning parameter in contrast to LM, and its implementation is straightforward.

6.3.3 NOVIFAST as an exact, analytical method

It is not difficult to demonstrate that in the noise-free case, that is, when $y_n = s_n$, the NLLS estimates coincide with the ground-truth values, and that they can be obtained analytically. Indeed, in the noiseless case, the ground-truth value $\boldsymbol{c}_{\text{GT}} = (c_{1\text{GT}}, c_{2\text{GT}})^T = (K_{\text{GT}}(1 - c_{2\text{GT}}), \exp(-\text{TR}/T_{1\text{GT}}))^T$ is the global minimum of $\mathcal{L}$ since we have $y_n = s_n \triangleq s_n(\boldsymbol{c}_{\text{GT}})$, thus $\mathcal{L}(\boldsymbol{c}_{\text{GT}}) = 0$. Due to the linear relationship of Eq. (6.6), $\boldsymbol{c}_{\text{GT}}$ can be linearly expressed in terms of just two samples from $\{s_n\}_{n=1}^N$. Since the NLLS estimate is equal to $\boldsymbol{c}_{\text{GT}}$, it can also be expressed this way, and thus $\hat{K}$ and $\hat{T}_1$ can be derived as a closed-form expression. Therefore, in the absence of noise, DESPOT1 becomes an exact method, but it is not clear what happens with iterative algorithms, since LM and GN, by design, do not exploit the particular structure of the SPGR signal. Remarkably, in the absence of noise, and due to the semi-linear structure of Eq. (6.18), NOVIFAST becomes an exact and

analytical method as well, in the sense that it provides the ground-truth values with just one iteration. Indeed, for noiseless data, the following relations can be shown to be true:

$$\langle \boldsymbol{b}, \boldsymbol{z} \rangle = c_{1\mathrm{GT}} \langle \boldsymbol{b}, \boldsymbol{b} \rangle + c_{2\mathrm{GT}} \langle \boldsymbol{b}, \boldsymbol{a} \rangle \tag{6.21}$$

$$\langle \boldsymbol{z}, \tilde{\boldsymbol{a}} \rangle = c_{1\mathrm{GT}} \langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle + c_{2\mathrm{GT}} \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle. \tag{6.22}$$

If we substitute $\langle \boldsymbol{b}, \boldsymbol{z} \rangle$ and $\langle \boldsymbol{z}, \hat{\boldsymbol{a}} \rangle$ in the numerators of Eq. (6.20) by the expressions of Eq. (6.21) and Eq. (6.22), and we make use of determinant properties, it is possible to prove that $c_1^{k+1} = c_{1\mathrm{GT}}$ and $c_2^{k+1} = c_{2\mathrm{GT}}$ for $k = 0, 1, ...$

The interested reader may find the mathematical proof of this result in subsection B.3. In short, NOVIFAST shares with DESPOT1 that just one iteration is needed to provide the ground-truth parameters. This correspondence with DESPOT1 may be useful to better understand the convergence behavior of NOVIFAST in realistic conditions, when noise is present. Of course, we remark that in such real conditions, DESPOT1 is not optimal whereas NOVIFAST is a truly NLLS-based algorithm. For instance, as is shown in subsection B.2.2, NOVIFAST's convergence conditions are very likely to hold at clinically achievable signal-to-noise ratio values, and the number of iterations required to reach convergence decreases with increasing signal-to-noise ratio. This is not surprising since we already know that in the asymptotic case of an infinitely high SNR NOVIFAST *must* be an exact and analytical algorithm. In contrast, there is no theoretical reason why the GN algorithm or the LM algorithm would require fewer iterations to converge when the signal-to-noise ratio increases.

6.4 Experiments

We validated NOVIFAST with Monte Carlo (MC) simulation and in vivo human brain experiments. For the simulation experiments, we generated noisy SPGR magnitude data y_n, $n = 1, ..., N$, as realizations of statistically independent random variables that follow a Rician distribution, where the signal parameter is given by $s_n(K_{\mathrm{GT}}, T_{1\mathrm{GT}})$, with K_{GT} and $T_{1\mathrm{GT}}$ the ground-truth K and T_1 values, respectively, and the noise standard deviation is denoted by σ. We employed the Rician distribution model since it is the most common statistical model for magnitude MR images [Gudbjartsson and Patz, 1995, den Dekker and Sijbers, 2014]. The noise standard deviation σ is parameterized as $\sigma = K_{\mathrm{GT}}/\mathrm{SNR}_{90^\circ}$, with SNR_{90° the maximum signal-to-noise ratio per pixel for an image acquired with an FA of 90° and $\mathrm{TR} > 6 \times T_1$ [Cheng and Wright, 2006]. Such a definition of the signal-to-noise ratio is commonly used in other works on VFA SPGR T_1 mapping [Cheng and Wright, 2006, Chang et al., 2008]. However, for better interpretation, we will additionally utilize the *conventional* signal-to-noise ratio (SNR) that is used in quantitative MRI:

$$\mathrm{SNR} = \frac{1}{N} \sum_{n=1}^{N} \frac{s_n}{\sigma}. \tag{6.23}$$

We illustrate the performance of NOVIFAST with common sequence settings that are reported in the literature. We used a repetition time TR of 5 ms [Cheng and Wright, 2006], and two sets of FAs, namely $\{2°, 9°, 19°\}$ [Cheng and Wright, 2006] and $\{2°, 3°, 4°, 5°, 7°, 9°, 11°, 14°, 17°, 22°\}$ [Deoni et al., 2004, Cheng and Wright, 2006]. We denote the FA sets as $\mathcal{A}_{N=3}$ and $\mathcal{A}_{N=10}$, respectively.

6.4.1 MC simulation experiment with synthetically generated SPGR MR signals

First, experiments with simulated SPGR MR signals were set up to compare the performance of NOVIFAST in terms of convergence and computational efficiency to that of GN, LM, DESPOT1, and IRWLLS. To investigate the convergence properties of GN, LM, and NOVIFAST, ideally, we could check whether the final iterates provided by all of them truly correspond to the global minimum of the NLLS cost function. However, this approach is infeasible since we would need to do so for all possible noisy SPGR signals. Instead, we followed a more modest approach which is common practice when a new VFA T_1 mapping algorithm is introduced. Since convergence failure may introduce a bias in the final estimates, and as this bias is not intrinsic to the NLLS estimator but rather to the optimization algorithm, we could detect convergence pathologies by analyzing those estimates. By using realistic MC simulations, we assessed the statistical performance of GN, LM, and NOVIFAST, as if they were considered as different estimators.

All algorithms were run in Matlab, using both GN and LM implemented within the Matlab function `lsqnonlin`. Although both GN and LM have been described in the domain of K and T_1 , we implement them in the $\boldsymbol{c}$ domain, with an exact, analytical Jacobian. This provides a fair comparison with DESPOT1, IRWLLS, and NOVIFAST, which are genuinely conceived in the $\boldsymbol{c}$ domain. We set the initial value of λ in the LM formulation to the recommended value in `lsqnonlin`. IRWLLS, GN, LM and NOVIFAST were halted with the same tolerance criterion: the relative l_2 norm between consecutive iterates was set to 10^{-6}. The maximum number of iterations was set to 1000.

The experiment setup for the MC simulation was the following.

1. We chose either $\mathcal{A}_{N=3}$ or $\mathcal{A}_{N=10}$ as FA set.

2. We fixed $K_{\mathrm{GT}} = 1$ and generated ground-truth SPGR signals, $\{s_n\}_{n=1}^{N}$, with ten different values of $T_{1\mathrm{GT}}$ logarithmically spaced between 500 ms and 2500 ms.

3. For each of the ten ground-truth signals $\{s_n\}_{n=1}^{N}$, we generated $N_{\mathrm{MC}} = 10^5$ realizations of Rician distributed noisy signals $\{y_n\}_{n=1}^{N}$. For each noisy dataset, $\mathrm{SNR}_{90°}$ was selected among the following list of values: $\{80, 150, 250, 300\}$. Those values are within the range of $\mathrm{SNR}_{90°}$ commonly used in similar MC-based analyses in the literature [Cheng and Wright, 2006, Chang et al., 2008].

4. The five algorithms were run with the same input datasets $\{y_n\}_{n=1}^N$. In order to check the influence of initialization, we chose a constant initialization for all range of $T_{1\mathrm{GT}}$. We studied the robustness of NOVIFAST against the rest of the methods with respect to poor initialization. To that end, we selected two different configurations: 1) $K_{\mathrm{ini}} = 0.5$ and $T_{1\mathrm{ini}} = 1000$ ms, and 2) $K_{\mathrm{ini}} = 0.5$ and $T_{1\mathrm{ini}} = 500$ ms.

The computational time, in milliseconds, amounts the total CPU time of the algorithms' execution for given input data $\{y_n\}_{n=1}^N$. All algorithms were run on an Intel®Xeon®CPU E5-2680 v2 with 25 MB of cache clocked at 2.8 GHz.

6.4.2 MC simulation with synthetic 3D T_1 phantom

Prior to validating NOVIFAST with in vivo SPGR MR images, we conducted an MC-based simulation with a synthetically generated set of SPGR MR images. We used realistic SNR values achievable in practice, realistic ground-truth $T_{1\mathrm{GT}}$ and K_{GT} maps (with a wider range of values than those in the previous experiment, including T_1 beyond that of white/grey matter, and we mimicked clinically realistic noise conditions, e.g., spatially variant noise maps. Specifically, several sets of noise-free 3D SPGR MR images (following Eq. (6.1) with TR = 5 ms and with FAs given by the $\mathcal{A}_{N=10}$ FA set) were created based on ground-truth $T_{1\mathrm{GT}}$ and K_{GT} maps. Those maps were estimated from a simulated IR gradient recalled echo sequence, with similar settings as those given in [Ramos-Llordén et al., 2017]. The size of both 3D maps was $111 \times 93 \times 71$ with an isotropic voxel size of 1.5 mm. Next, noisy Rician distributed images were generated, similarly as in section 6.4.1, but this time with a spatially variant σ. We employed similar noise maps as those reported in [Pieciak et al., 2017], and we scaled them so as to get spatially averaged SNR_{90° values of 400 and 500 (and corresponding SNR values of 12.3 and 15, respectively). Those values are within the range of SNRs encountered in practice [Bouhrara and Spencer, 2017]. For each of those values of SNR_{90°, $N_{\mathrm{MC}} = 10^4$ noisy realizations were generated.

T_1 maps were estimated by applying DESPOT1, IRWLLS, LM and NOVIFAST in a voxel-wise manner, with a mask including only brain tissue voxels[2]. Code parallelization was not performed, and all algorithms were run with the same parameters as in the first experiment. Likewise, they were initialized with constant K and T_1 maps. The initial values were $K_{\mathrm{ini}} = 0.5$ and $T_{1\mathrm{ini}} = 1000$ ms. Algorithms were stopped using the same criterion as in the first experiment. For each algorithm, the total computation time as well as the spatially averaged relative bias, standard deviation (Std.), and RMSE were reported.

[2]Initial experiments showed that the GN algorithm did not converge for most of the voxels in the phantom. Therefore, we decided not to include GN in the MC experiment to avoid a drastic increase in total time, which would have rendered the experiment infeasible for $N_{\mathrm{MC}} = 10^4$ repetitions.

6.4.3 In vivo human brain set of SPGR MR images

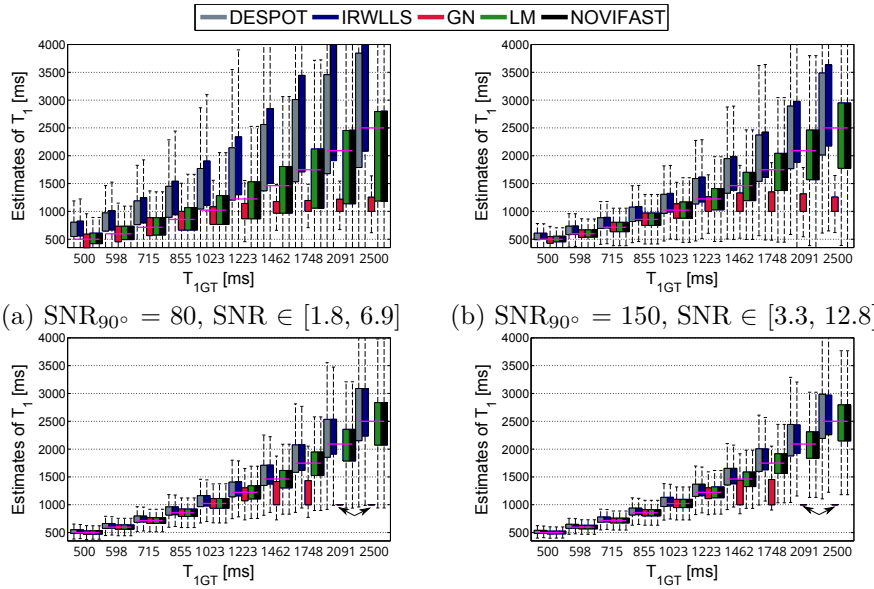
A set of SPGR MR images of a healthy 28-year old male volunteer was acquired with a 3T MRI scanner (MAGNETOM Prisma$^{\text{fit}}$, Siemens) using a 32-channel head coil. For each FA given in the $\mathcal{A}_{N=10}$ set, a $256 \times 256 \times 24$ image (voxel size $1 \times 1 \times 5$ mm^3) was acquired with a 3D FLASH sequence. The sequence parameters were: TR/TE = 7.8/3.48 ms and BW = 320 Hz. B_1 shimming was implemented with SIEMENS's TrueForm mode. Parallel imaging was not applied and magnitude MR images were reconstructed with the adaptive combine method [Walsh et al., 2000]. An SNR = 23 was estimated with the method of [Aja-Fernández et al., 2015b]. The total scan time per FA was about 1 min. In order to show that NOVIFAST does not require a careful initialization, and can be initialized with a constant map, we chose for all the algorithms $T_{1\text{ini}} = 500$ ms and $K_{\text{ini}} = 3 \cdot 10^6$ (being the average value of the estimated K map obtained with DESPOT1). Algorithms were stopped according to the same criteria as used in the first experiment, and identical tuning parameters were used. Similar to the previous experiment, all algorithms were applied voxel-wise with a mask including only brain tissue voxels.

6.5 Results

6.5.1 MC simulation with synthetically generated SPGR MR signals

In this section we only present the results for the $\mathcal{A}_{N=10}$ FA set with initialization $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 1000$ ms. Results for $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 500$ ms, as well as results for the $\mathcal{A}_{N=3}$ FA set (with both initializations), are presented in subsection B.4. In Fig. 6.1, we show box-plots for the T_1 estimates obtained with the five algorithms under test. A box-plot-based visualization allows us to see whether there exist statistically significant differences in the population of the sample estimates of T_1 [Krzywinski and Altman, 2014]. In that box-plot visualization, horizontal magenta lines, representing the values of $T_{1\text{GT}}$, are intentionally marked for ease of interpretation. To illustrate the speed performance, the average computation time is displayed in Fig. 6.2. Results for the second initialization, both regarding the statistical performance and the computation time, are presented in Fig. B.8 and in Fig. B.9.

A first, general observation that we can already point out is that less accurate and precise T_1 estimates are obtained if heuristic linearized estimators, such as DESPOT1 and IRWLLS, are employed. This can be attributed simply to the statistical superiority of the NLLS estimator over modified linear versions. It is true, though, that the difference becomes less noticeable with increasing SNR, and is less drastic for the $\mathcal{A}_{N=3}$ FA data set. The most interesting observations are those related to GN, LM, and NOVIFAST. From Fig. 6.1 it is clear that GN is sensitive to initialization. The local-convergence behavior of GN evidently degrades for $T_{1\text{GT}} > 1000$ ms, yielding a systematic bias which persists over the whole



(a) $\text{SNR}_{90^\circ} = 80$, SNR $\in [1.8, 6.9]$ (b) $\text{SNR}_{90^\circ} = 150$, SNR $\in [3.3, 12.8]$

(c) $\text{SNR}_{90^\circ} = 250$, SNR $\in [5.6, 21.3]$ (d) $\text{SNR}_{90^\circ} = 300$, SNR $\in [6.7, 25]$

Figure 6.1: Box-plots of the T_1 estimates that are obtained with the five SPGR VFA optimization algorithms. Tukey-style whiskers are shown that extend to a maximum of 1.5 × IQR beyond each box, with IQR the interquartile range (corresponding with the length of each box) [Krzywinski and Altman, 2014]. Ground-truth T_1 values are marked with horizontal magenta lines to ease interpretation (Case of $\mathcal{A}_{N=10}$ FA set and fixed initialization of $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 1000$ ms).

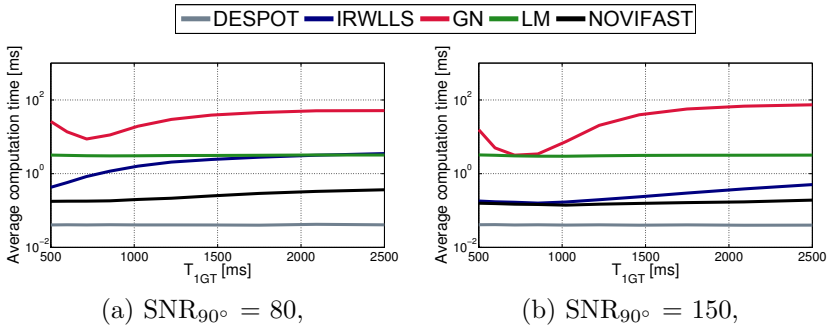
range of SNRs. Observe that for $T_{1\text{GT}} > 1223$ ms, the GN interquartile ranges (IQRs), covering the middle 50% of the sample, do not cover $T_{1\text{GT}}$. It can also be observed that the corresponding IQRs are approximately clustered around 1000 ms, i.e, $T_{1\text{ini}}$, and the variability of the estimates greatly reduces for that regime. Note as well that the length of the GN boxes in Fig. 6.1.(c) and Fig. 6.1.(d) becomes very small for $T_{1\text{GT}} > 1748$ ms (see black arrows pointing the boxes). This behavior is due to convergence failure, since GN is stagnating at the initialization, as we observed in a large number of MC realizations. This is also reflected by the method's computation time in Fig. 6.2, which massively grows for high $T_{1\text{GT}}$, since GN was only stopped when it reached the maximum number of iterations. The LM algorithm and, especially, NOVIFAST present a remarkable insensitivity to initialization, providing very similar results. Results presented in subsection B.4 confirm that the same conclusions hold when the second initialization was employed. Indeed, in that case, NOVIFAST and LM also manifest a substantial robustness to poor initial values in contrast to GN, which starts to fail as soon as the value of $T_{1\text{GT}} = 715$ ms is reached.

Having shown that both LM and NOVIFAST have good global convergence properties, what distinguishes them is, as we have already mentioned, their computational speed. Due to the negligible cost per-iteration and the rapid convergence, the

average computational time of NOVIFAST is overall more than one order lower than LM. Indeed, averaged over SNRs and T_{1GT}, NOVIFAST is 20 times faster. This speed gain increases with SNR, since the number of iterations needed for NOVIFAST to converge asymptomatically decreases to one as the SNR approaches infinity, i.e., NOVIFAST, unlike LM, asymptotically approaches an exact, analytical estimator. By observing Fig. 6.2, it is clear that NOVIFAST's computational time is nearly constant, around 0.16 ms. The reported computational time of NOVIFAST seems similar to that of IRWLLS, but NOVIFAST's statistical performance is considerably higher. Indeed, being a heuristic algorithm, the convergence of IRWLLS in l_2 norm does not imply that the NLLS cost function is effectively minimized. Readers can check that the same conclusions that are drawn in this subsection also hold for the $\mathcal{A}_{N=3}$ FA set (subsection B.4).

6.5.2 MC simulation with synthetic 3D T_1 phantom

The overall accuracy (bias), precision (Std.), RMSE and computational time are shown in Table 6.1 and 6.2 for $\text{SNR}_{90^\circ} = 400$ and $\text{SNR}_{90^\circ} = 500$, respectively. Bias and RMSE maps for one particular mid-axial slice are shown in Fig. 6.3 (for the case $\text{SNR}_{90^\circ} = 500$).



(a) $\text{SNR}_{90^\circ} = 80$, SNR $\in$ [1.8, 6.9]

(b) $\text{SNR}_{90^\circ} = 150$, SNR $\in$ [3.3, 12.8]

(c) $\text{SNR}_{90^\circ} = 250$, SNR $\in$ [5.6, 21.3]

(d) $\text{SNR}_{90^\circ} = 300$, SNR $\in$ [6.7, 25]

Figure 6.2: Total computation time of each of the five optimization algorithms for the MC simulation-based experiment (Case of $\mathcal{A}_{N=10}$ FA set and fixed initialization of $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 1000$ ms).

As expected, LM and NOVIFAST provide nearly identical accuracy, precision, and RMSE. Nevertheless, NOVIFAST provides the NLLS estimated 3D T_1 map in about 23 s, whereas it takes around 10 min for LM to do so. The accuracy of NLLS-based algorithms is drastically higher than that of DESPOT1 and IRWLLS (at least 6×) and the RMSE is lower as well, confirming the statistical superiority of the NLLS estimator over heuristic modifications. Finally, observe that NOVIFAST is consistently faster than IRWLLS.

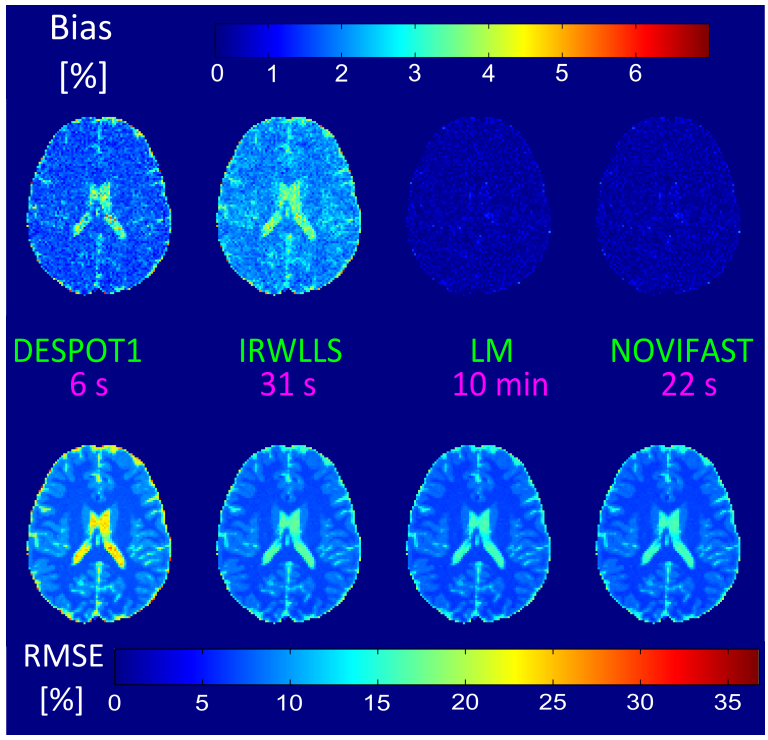


Figure 6.3: Bias and RMSE maps of a mid-axial slice of a synthetic 3D T_1 map ($SNR_{90^\circ} = 500$) obtained with DESPOT1, IRWLLS, LM, and NOVIFAST. The computation times are shown as well.

Table 6.1: Quantitative results obtained with DESPOT1, IRWLLS, LM, and NOVIFAST, when estimating a synthetic 3D T_1 map ($SNR_{90^\circ} = 400$).

	Bias [%]	Std. [%]	RMSE [%]	Time [s]
DESPOT1	2.48	13.45	13.68	6.2
IRWLLS	3.13	11.45	11.87	31.38
LM	0.32	11.03	11.03	640.19
NOVIFAST	0.32	11.03	11.03	23.19

6.5.3 In vivo human brain set of SPGR MR images

Although we are not able to assess the statistical performance of the five methods with real data, due to the overarching issue of lacking a noise-free ground truth, interesting observations can be made about convergence and total computation time. Estimated T_1 maps of two mid-axial slices are shown in Fig. 6.4. A constant initial map is not an impediment for NOVIFAST to estimate a reliable T_1 map, which is not the case for GN (observe the magenta arrow on the lowest part of the brain, where the estimated T_1 value remains constant and equal to 500 ms). Interestingly, several outliers are manifestly clear in the T_1 map that was estimated with LM. Those outliers are located on the boundary between cerebrospinal fluid and grey matter. No outliers were observed in the T_1 maps estimated with DESPOT1, IRWLLS, and NOVIFAST. Differences between the T_1 maps of the latter three methods are, as expected, visually indistinguishable, and statistical claims can only be made based on the previous MC simulations. DESPOT1 and NOVIFAST are the only methods that can provide this high-resolution 3D T_1 map in less than 1 min, where the NOVIFAST speed gain over IRWLLS and LM is about 4× and 40×, respectively.

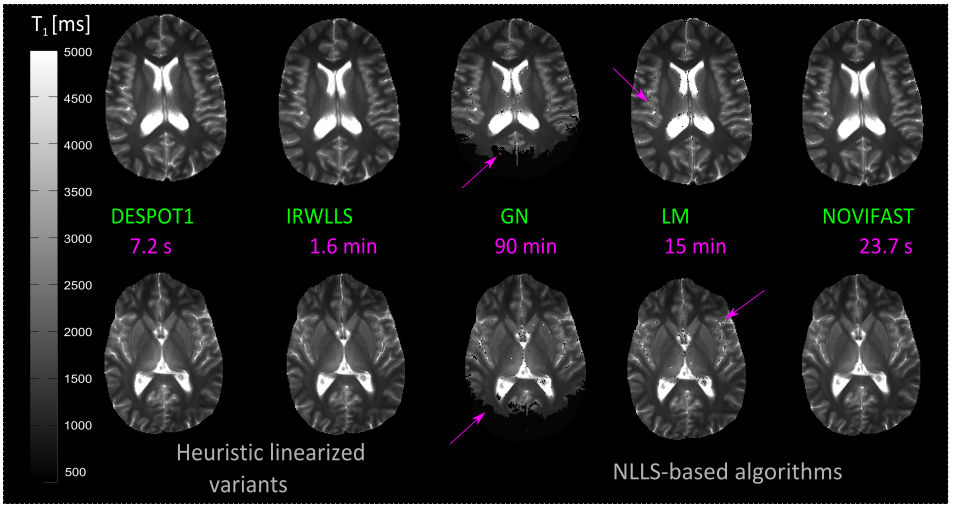


Figure 6.4: Two mid-axial slices of the estimated 3D T_1 map with the five VFA SPGR algorithms for the in vivo MR experiment.

Table 6.2: Quantitative results obtained with DESPOT1, IRWLLS, LM, and NOVIFAST, when estimating a synthetic 3D T_1 map ($SNR_{90^\circ} = 500$).

	Bias [%]	Std. [%]	RMSE [%]	Time [s]
DESPOT1	1.58	10.38	10.5	6.2
IRWLLS	1.99	9.0	9.23	30.15
LM	0.24	8.80	8.81	637.81
NOVIFAST	0.24	8.80	8.80	22.1

It should be noted that the SPGR model of Eq. (6.1) neglects partial volume effects, which may be relevant for highly non-isotropic voxels, as those of the dataset used in this experiment. Furthermore, there may exist incomplete spoiling, which may introduce related T_2 decay effects [Heule et al., 2016]. These remarks are intrinsic to modeling and not to the choice of the algorithm. Indeed, NOVIFAST is always applicable wherever DESPOT1 and IRWLS are.

6.6 Future work

NOVIFAST can be extended in several ways. In this work, it has been presented as an *ordinary* NLLS optimization algorithm, but extension to the *weighted* NLLS estimator is possible (see [Dimitrov and Kamenski, 1991] for more details). Furthermore, NOVIFAST can be applied to other MR data sequences than SPGR. For example, it is well known that the completely balanced TrueFISP sequence can be modeled as a quotient of rational functions, where the (three) linear parameters encode the T_1 and T_2 values [Scheffler and Hennig, 2003]. NOVIFAST can be reformulated for TrueFISP, since the semi-linear structure of Eq. (6.18) also appears when fitting a quotient of rational functions [Dimitrov and Kamenski, 1996]. Finally, NOVIFAST can be embedded in a Maximum-Likelihood (ML) framework, aiming at fully exploiting the statistical knowledge of the data [Sijbers et al., 1998a]. Indeed, it has been shown that an ML estimation problem with non-central χ or Rician distributed data is equivalent to iteratively solving a collection of NLLS subproblems [Varadarajan and Haldar, 2015, Ramos-Llordén et al., 2017]. Each of those NLLS subproblems can be solved with NOVIFAST, greatly boosting the speed of the overall ML estimation procedure.

6.7 Conclusions

In VFA T_1 mapping, it is common to use heuristic linear estimators such as the famous DESPOT1 method [Deoni et al., 2003] or the IRWLLS method of Chang *et al.* [Chang et al., 2008]. They are preferred despite the superiority of NLLS estimators in terms of accuracy and precision, since optimization algorithms for NLLS estimators are much slower than linear estimators, and are also prone to initialization issues. In this work, we reconcile these two separate frameworks by proposing a novel NLLS method, NOVIFAST, which reports the NLLS estimates more than twenty times faster than conventional NLLS optimization algorithms, thereby merging the best of both approaches, i.e., accurate and precise T_1 mapping with a very short computation time. Furthermore, minimal pre-processing is needed for NOVIFAST, since the algorithm can be initialized with constant T_1 maps. We believe that NOVIFAST is a good candidate to be included in every processing pipeline for high-quality, robust, and efficient VFA T_1 mapping.

7

Partial Discreteness: a Novel Prior for Magnetic Resonance Image Reconstruction

Contents

The work in this chapter has been published in:

- **Ramos-Llordén, G.**, den Dekker, A. J., and Sijbers, J. (2017). Partial discreteness: a novel prior for magnetic resonance image reconstruction. *IEEE Trans. Med. Imag.*, 36(5):1041–1053.
- **Ramos-Llordén, G.**, Segers, H., Palestijn, W. J., den Dekker, A. J., and Sijbers, J. (2015). Partially discrete magnetic resonance tomography. In *Proc. IEEE ICIP*, pages 1653–1657.
- **Ramos-Llordén, G.**, Segers, H., Palestijn, W. J., den Dekker, A. J., and Sijbers, J. (2015). Partial discreteness: a new type of prior knowledge for MRI reconstruction. *Proc. Intl. Soc. Mag. Reson. Med.*, 23:3417.

7.1 Introduction

In MRI, the acquired samples, a.k.a. the k-space data, are samples of the Fourier transform of the spatial magnetization distribution (the image) [Fessler, 2010]. From sampling theory, it is well-known that if the k-space is sampled with a Cartesian scheme fulfilling the Nyquist condition, exact recovery of the finite support image is possible [Trzasko and Manduca, 2009]. However, if the number of k-space data points is reduced, the Nyquist condition is violated, making the inverse problem of image reconstruction ill-posed without prior knowledge [Trzasko and Manduca, 2009].

Fortunately, prior knowledge comes in different forms. It can be included by exploiting image properties, such as smoothness, both in the image and Fourier domain [Ying et al., 2004, Knoll et al., 2012], sparsity in a specific image representation [Lustig et al., 2007] (e.g., Fourier [Lingala et al., 2011], Wavelet [Chen and Huang, 2012], [Lai et al., 2016], Curvelet [Candes et al., 2006], Shearlet [Aelterman et al., 2011], structural sparsity, where the sparse coefficients in a given representation domain adhere to certain structural patterns [Pizurica et al., 2011, Chen and Huang, 2014, Panic et al., 2017], redundant dictionaries [Ravishankar and Bresler, 2011, Caballero et al., 2014]), number of discrete gray levels [Segers et al., 2013], minimal Total Variation (TV) [Block et al., 2007], [Liang et al., 2011], [Knoll et al., 2011], limited image support [Haldar, 2014a], [Haldar and Zhuo, 2015], or spatial constraints [Kelm et al., 2009, Velikina and Samsonov, 2014, Haldar et al., 2008]. Additionally, prior knowledge can come from anatomical information derived from a training dataset [Gindi et al., 1993, Cao and Levin, 1995] or can be extracted from reference images [Hanson et al., 1996].

In this work, we introduce a novel type of prior knowledge for MR images that possess, apart from heterogeneous regions, a number of quasi-constant intensity regions. That is, the magnitude range is assumed to be partially discrete. Partial discrete tomography has recently successfully been introduced as a prior in X-ray tomography [Batenburg and Sijbers, 2011, Dabravolski et al., 2014] and electron tomography [Batenburg et al., 2009, Roelandts et al., 2012], however, to the authors' knowledge not yet in MRI. There is a number of MRI applications where the partially discrete assumption may be exploited. Implants MR imaging is a paradigmatic example, where the homogeneous composition of implants naturally leads to partially discrete images [Juanpere et al., 2011]. Contrast-enhanced MRI sequences also produce images that meet the partial discreteness assumption, e.g. contrast-enhanced MR Angiography (MRA) [Lustig et al., 2007]. Finally, the use of specific pulse sequences, such as T2-weighted or short tau inversion recovery (STIR) sequences, may produce hyper-intense regions in brain cyst imaging [Preece et al., 2006].

Specifically, our contribution is the following. We mathematically formalize the *partial discreteness* property and propose a decomposition of every image into its partial discreteness representation and its residual form. The partial discreteness representation is constructed from 1) an auto-learned Gaussian Mixture Model (GMM) [Figueiredo and Jain, 2002] specifically designed to fulfill the partial dis-

creteness assumption and 2) the fusion of the *a posteriori* probability maps with intensity information, both derived from the GMM and the image itself. Because partially discrete images admit an accurate partial discreteness representation, we enforce sparsity in the residual form to promote this type of solution in the reconstruction of under-sampled MR images. In this work, the partial discreteness prior is implemented in a phase-constrained MR reconstruction formulation [McGibney et al., 1993, Samsonov et al., 2004, Bydder and Robson, 2005, Samsonov et al., 2010, Blaimer et al., 2016] with the common assumption of a smoothly varying phase image [Lustig et al., 2007], [Haldar, 2014a].

We illustrate the potential of the partial discreteness prior by showing examples of applications with under-sampled simulated and real k-space data. Thereby, the proposed partially discrete reconstruction method is compared to popular reconstruction methods.

This chapter is organized as follows. In section 7.2, we present the novel partial discreteness prior, which is incorporated in a constrained optimization method in section 7.3. Sections 7.4 and 7.5 illustrate the application of this method to a variety of under-sampling scenarios, in comparison with competitors reconstruction methods. These sections also summarize a sensitivity analysis of the proposed method to various parameters and deviations from assumptions. Finally, future work and conclusions are given in sections 7.6 and 7.7, respectively.

7.2 Partial discreteness

7.2.1 Phase-constraint formulation

Partial discreteness is applicable to magnitude images only. Therefore, the constraint reconstruction problem with prior knowledge (Eq. (2.56)) should be restricted to the set of real and positive images (as in Eq. (2.47)), as follows

$$\min_{\boldsymbol{x}\in\mathbb{R}_{+}^{N}} \Phi(\boldsymbol{x}) \quad s.t. \quad ||\boldsymbol{y}-\tilde{\boldsymbol{A}}\boldsymbol{x}||_2^2 \leq \epsilon, \tag{7.1}$$

with $\tilde{\boldsymbol{A}}$ defined as in subsection 2.7.3.

7.2.2 A Bayesian model for partial discreteness

Consider the following decomposition of a magnitude image into a piece-wise homogeneous part and a texture part:

$$\boldsymbol{x} = \underbrace{\sum_{k=1}^{K} \boldsymbol{x}_{\mathcal{A}_k}}_{\text{piece-wise homogeneous part}} + \underbrace{\boldsymbol{x}_{\bar{\mathcal{A}}}}_{\text{texture part}}, \tag{7.2}$$

where $A = \cup_{k=1}^{K} \mathcal{A}_k$ represents the union of K disjoint homogeneous regions (i.e., pixel sets) of $\boldsymbol{x}$, and $\bar{A}$ is the texture region. For each pixel $n = 1, ..., N$, $\boldsymbol{x}_{\bar{\mathcal{A}}} \in \mathbb{R}_+^N$ and $\boldsymbol{x}_{\mathcal{A}_k} \in \mathbb{R}_+^N$ are defined as

$$[\boldsymbol{x}_{\bar{\mathcal{A}}}]_n = \begin{cases} [\boldsymbol{x}]_n, & \text{if } n \in \bar{\mathcal{A}} \\ 0, & \text{if } n \notin \bar{\mathcal{A}} \end{cases} \tag{7.3}$$

and

$$[\boldsymbol{x}_{\mathcal{A}_k}]_n = \begin{cases} \eta_k, & \text{if } n \in \mathcal{A}_k \\ 0, & \text{if } n \notin \mathcal{A}_k, \end{cases} \tag{7.4}$$

respectively, with η_k the constant intensity of the homogeneous set $\mathcal{A}_k$. In this work, we assume that $|\mathcal{A}_k| \gg 1, \forall k$. While ideally suited to describe partially discrete images, this model is unpractical since the location, cardinality, and intensity of the sets $\{\mathcal{A}_k\}_{k=1}^{K}$ are unknown in practice.

Keeping Eq. (7.2) in mind, in this subsection we approximate Eq. (7.2) with a realizable model, effectively preserving the distinct characteristics of partially discrete images. To that end, a Bayesian framework is proposed which relies on 1) a Gaussian Mixture Model (GMM) that captures the particular intensity properties of Eq. (7.2) and 2) an unsupervised Bayesian probabilistic segmentation. The proposed methodology allows us to identify the different sets and estimate the probabilities that a pixel belongs to each of these sets (probabilistic segmentation) in an unsupervised manner. In what follows, we start with the construction of the GMM. The probabilistic segmentation is described afterwards. With these two main ingredients, we present the Bayesian model for partially discrete images in subsection 7.2.3. With the partial discreteness property formalized, the partial discreteness prior is then defined.

7.2.2.1 The GMM construction

We consider the pixels' intensities of the image $\boldsymbol{x}$, $(x_1, ..., x_N)^T$, as a vector of realizations of a random variable, $X : \Omega \mapsto \mathbb{R}^+$. We assume that each pixel $n = 1, ..., N$ only belongs to one specific, unknown set $\mathcal{A}_k$ or $\bar{\mathcal{A}}$. Assigning a probability to the event that a pixel belongs to a specific set, we define:

$$P(n \in \mathcal{A}_k) = \pi_k, \quad \text{with } k = 1, ..., K, \tag{7.5}$$

$$P(n \in \bar{\mathcal{A}}) = \pi_{\bar{\mathcal{A}}} = 1 - \sum_{k=1}^{K} \pi_k, \tag{7.6}$$

where it is assumed that the probabilities π_k and $\pi_{\bar{\mathcal{A}}}$ are independent of the pixel n, and no a priori spatial information is incorporated. Hence, in what follows, we will denote $P(n \in \mathcal{A})$ as $P(\mathcal{A})$. Each of the pixels' intensities, $x_1, ..., x_N$, is assumed to have been generated by one (randomly selected and unknown) element of a set of $K+1$ random sources. The source-conditional distributions of the random variable X are characterized by the conditional PDFs $p_{X|\mathcal{A}_k}(x)$ and $p_{X|\bar{\mathcal{A}}}(x)$ [Papoulis and

Pillai, 2002]. It is reasonable to assume that the individual conditional PDFs for the homogeneous sets all belong to the same location-scale family. There are several distributions that belongs to the class of location-scale family. In this work, we chose Gaussian distributions. Gaussian PDFs are well-behaving functions and easy to handle, which simplify posterior calculation and analysis. The conditional PDF of set $\mathcal{A}_k$ thus take the form of

$$p_{X|\mathcal{A}_k}(x;\eta_k,\sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\eta_k)^2}{2\sigma_k^2}}, \tag{7.7}$$

where the dispersion of X around the specific discrete value η_k is represented by the standard deviation σ_k. A small value of σ_k reflects the typical low dispersion of such homogeneous regions. Of course, that does not mean that no intensity variations are allowed, as these always occur in real scenarios.

The conditional PDF $p_{X|\bar{\mathcal{A}}}(x;\boldsymbol{\theta}_{\bar{\mathcal{A}}})$ models the random variable X in the texture part. Texture modeling through statistical distributions is beyond the scope of this work. The interested reader is referred to [Zhu et al., 1998, Huang and Mumford, 1999, Awate et al., 2006]. We deem that the shape of $p_{X|\bar{\mathcal{A}}}(x;\boldsymbol{\theta}_{\bar{\mathcal{A}}})$ can be well modeled by a mixture of Gaussian PDFs, and therefore $\boldsymbol{\theta}_{\bar{\mathcal{A}}}$ are the GMM parameters. The use of a GMM for $p_{X|\bar{\mathcal{A}}}(x;\boldsymbol{\theta}_{\bar{\mathcal{A}}})$ should be seen as a way to describe arbitrarily complex distributions [Permuter et al., 2003] and not as an attempt to model quasi-discrete components, as in the case of $p_{X|\mathcal{A}_k}(x;\eta_k,\sigma_k)$. Simple application of the law of total probability [Papoulis and Pillai, 2002] with events $\{n \in \mathcal{A}_k\}_{k=1}^K$ and $\{n \in \bar{\mathcal{A}}\}$, yields the final GMM:

$$p_X(x;\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k p_{X|\mathcal{A}_k}(x;\eta_k,\sigma_k) + \pi_{\bar{\mathcal{A}}} p_{X|\bar{\mathcal{A}}}(x;\boldsymbol{\theta}_{\bar{\mathcal{A}}}), \tag{7.8}$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathcal{A}}, \boldsymbol{\theta}_{\bar{\mathcal{A}}})^T$ and where $\boldsymbol{\theta}_{\mathcal{A}} = \{\pi_k, \eta_k, \sigma_k\}_{k=1}^K$.

7.2.2.2 Bayesian probabilistic segmentation

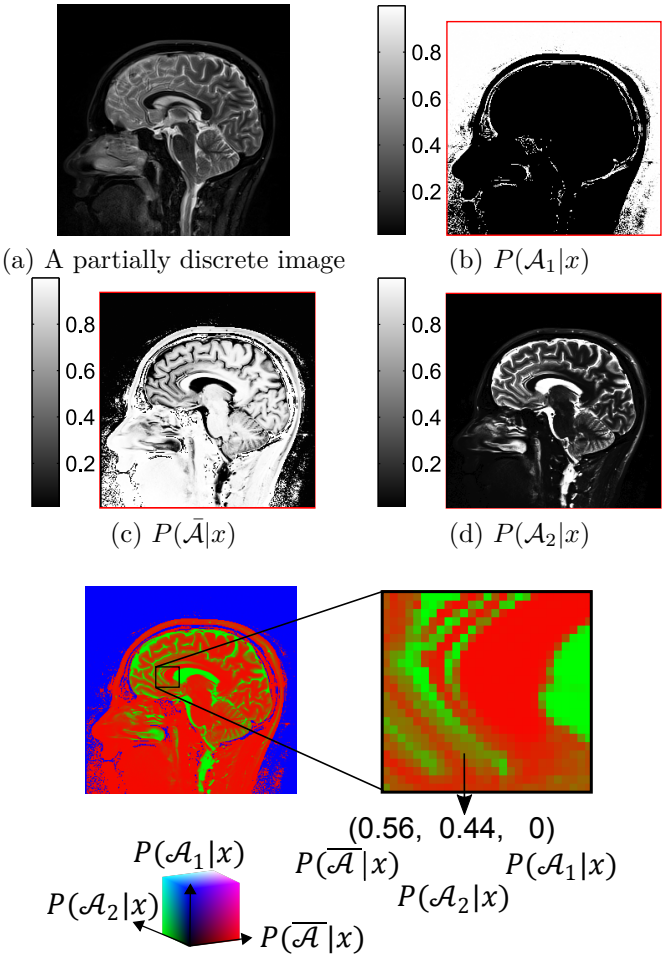
Bayes' theorem [Papoulis and Pillai, 2002] now allows us to derive the *a posteriori* probabilities $P(\mathcal{A}_k|x)$ and $P(\bar{\mathcal{A}}|x)$ as

$$P(\mathcal{A}_k|x) = \frac{\pi_k p_{X|\mathcal{A}_k}(x;\eta_k,\sigma_k)}{p_X(x;\boldsymbol{\theta})}, \tag{7.9}$$

$$P(\bar{\mathcal{A}}|x) = \frac{\pi_{\bar{\mathcal{A}}} p_{X|\bar{\mathcal{A}}}(x;\boldsymbol{\theta}_{\bar{\mathcal{A}}})}{p_X(x;\boldsymbol{\theta})}, \tag{7.10}$$

respectively. These *a posteriori* probabilities denote the probabilities of a pixel belonging to each of the $K+1$ sets given that its intensity is equal to x. If Eqs. (7.9-7.10) are evaluated for each pixel $n = 1, ..., N$, we obtain what is dubbed throughout this work, probability maps, that is, $\boldsymbol{p}_k \in [0,1]^N$ and $\boldsymbol{p}_{\bar{\mathcal{A}}} \in [0,1]^N$ which are pixel-wise defined as

$$[\boldsymbol{p}_k]_n = \frac{\pi_k p_{X|\mathcal{A}_k}(x_n;\eta_k,\sigma_k)}{\sum_{k=1}^{K} \pi_k p_{X|\mathcal{A}_k}(x_n;\eta_k,\sigma_k) + \pi_{\bar{\mathcal{A}}} p_{X|\bar{\mathcal{A}}}(x_n;\boldsymbol{\theta}_{\bar{\mathcal{A}}})} \tag{7.11}$$



(a) A partially discrete image (b) $P(\mathcal{A}_1|x)$

(c) $P(\bar{\mathcal{A}}|x)$ (d) $P(\mathcal{A}_2|x)$

(e) RGB visualization of probability maps

Figure 7.1: Illustration of probability maps for a partially discrete image.

and

$$[\boldsymbol{p}_{\bar{\mathcal{A}}}]_n = \frac{\pi_{\bar{\mathcal{A}}} p_{X|\bar{\mathcal{A}}}(x_n; \boldsymbol{\theta}_{\bar{\mathcal{A}}})}{\sum_{k=1}^{K} \pi_k p_{X|\mathcal{A}_k}(x_n; \eta_k, \sigma_k) + \pi_{\bar{\mathcal{A}}} p_{X|\bar{\mathcal{A}}}(x_n; \boldsymbol{\theta}_{\bar{\mathcal{A}}})}. \tag{7.12}$$

Given a pixel, its *a posteriori* probability of belonging to a specific set can be determined from the corresponding probability map. This Bayesian framework frames a probabilistic segmentation scheme. Indeed, it is not possible to strictly assign pixels to specific sets but, instead, the probability of this assignment can be inferred. Another relevant point of the Bayesian segmentation is that not just local but *global* information is considered. This information is derived through $p_X(x; \boldsymbol{\theta})$. An illustration of the probabilistic segmentation associated to a partially discrete image is shown in Fig.7.1. These maps constitute the core of the partial

discreteness representation that we propose in the next subsection 7.2.3.

7.2.3 The partial discreteness prior

Given a partially discrete image $\boldsymbol{x}$, and its associated GMM $p_X(x;\boldsymbol{\theta})$, the partial discreteness representation of $\boldsymbol{x}$, denoted as $\mathcal{P}(\boldsymbol{x}) \in \mathbb{R}_+^N$, is defined as:

$$\mathcal{P}(\boldsymbol{x}) = \sum_{k=1}^{K} \eta_k \boldsymbol{p}_k + \boldsymbol{x}_\rho \circ \boldsymbol{p}_{\bar{\mathcal{A}}}, \tag{7.13}$$

where $\circ$ denotes the Hadamard product and $\boldsymbol{x}_\rho$ is a spatially filtered version of $\boldsymbol{x}$ with a circularly symmetric Gaussian filter with standard deviation ρ in the image domain. For partially discrete images, the partial discreteness representation essentially behaves as Eq. (7.2).

Behavior in homogeneous regions
In a homogeneous region $\mathcal{A}_{k'}$, the pixels' intensities minimally vary with respect to the mean $\eta_{k'}$. When the probability maps for the homogeneous regions are evaluated, all except the one associated to $\mathcal{A}_{k'}$ approximate zero. This is due to the fact that their conditional PDFs $\{p_{X|\mathcal{A}_k}(x;\eta_k,\sigma_k)\}_{k=1}^K$ are highly concentrated around their mean and when evaluated far from their mode they rapidly fall off to zero. As a consequence, each of the $P(\mathcal{A}_k|x)$ with $k \neq k'$ vanishes as well. Furthermore, it is assumed that the lack of intensity dispersion that characterizes the homogeneous regions is not captured by $p_{X|\bar{\mathcal{A}}}(x;\boldsymbol{\theta}_{\bar{\mathcal{A}}})$, which is the case if $p_{X|\bar{\mathcal{A}}}(x;\boldsymbol{\theta}_{\bar{\mathcal{A}}})$ is nearly zero for $x \in \eta_k$. Therefore, exclusively the remaining probability map of $\mathcal{A}_{k'}$, $\boldsymbol{p}_{k'}$, is approximately one. Hence, the partial discreteness representation becomes

$$\mathcal{P}(\boldsymbol{x}) \approx \eta_{k'} \boldsymbol{p}_{k'} \approx \boldsymbol{x}_{\mathcal{A}_{k'}}, \tag{7.14}$$

as desired.

Behavior in texture regions
In texture regions, the characteristic intensity variability is solely represented by the conditional PDF $p_{X|\bar{\mathcal{A}}}(x;\boldsymbol{\theta}_{\bar{\mathcal{A}}})$. As a result, we get

$$\mathcal{P}(\boldsymbol{x}) \approx \boldsymbol{x}_\rho \circ \boldsymbol{p}_{\bar{\mathcal{A}}} \approx \boldsymbol{x}_\rho. \tag{7.15}$$

Note that $\mathcal{P}(\boldsymbol{x})$ does not exactly approach $\boldsymbol{x}$ but a Gaussian filtered version of $\boldsymbol{x}$. The use of the Gaussian filter should be seen as a way to make partial discreteness a stable representation under very high spatial frequency perturbations that do not correspond to the original texture $\boldsymbol{x}$. In our experiments, the value $\rho = 2$ was consistently used. With this value, the corresponding cut-off frequency is high enough to preserve structural details. Note that in the limit case, which corresponds to $\rho = 0$, $\mathcal{P}(\boldsymbol{x}) \approx \boldsymbol{x}$, since the Gaussian kernel degenerates into a delta function.

Behavior in the frontier between regions
In the frontier between regions, none of the *a posteriori* probabilities has a prevailing effect (see zoomed image of Fig. 7.1.(e)). Indeed, $\mathcal{P}(\boldsymbol{x})$ is a mixture of intensities. Particularly, the intensity along a given profile which crosses two regions is a (convex) combination of two values. If the two regions are a homogeneous region $\mathcal{A}_k$ and the texture region $\bar{\mathcal{A}}$, such values are the mean η_k and $\boldsymbol{x}_\rho$, the filtered texture. The closer we are to $\mathcal{A}_k$, the higher $\boldsymbol{p}_k = 1 - \boldsymbol{p}_{\bar{\mathcal{A}}}$ is. Thus, $\mathcal{P}(\boldsymbol{x})$ approaches η_k. The nearer we are to $\bar{\mathcal{A}}$, the larger $\boldsymbol{p}_{\bar{\mathcal{A}}}$ is and $\mathcal{P}(\boldsymbol{x})$ then approaches $\boldsymbol{x}_\rho$. If the interface divides two homogeneous regions, let's say $\mathcal{A}_k$ and $\mathcal{A}_{k'}$, then $\mathcal{P}(\boldsymbol{x})$ is a convex combination of the two corresponding mean values, that is, η_k and $\eta_{k'}$.

In summary, $\mathcal{P}(\boldsymbol{x}) \approx \boldsymbol{x}$ for partially discrete images. Therefore, instead of the strict but unpractical model of Eq. (7.2), we can fairly justify the employment of the partial discreteness representation for the kind of images targeted in this work. Based on this representation, the partial discreteness prior for the optimization problem (7.1) can be defined. We first note that a partial discreteness representation $\mathcal{P}(\boldsymbol{x})$ can be assigned to every image $\boldsymbol{x}$. In practice, this implies that the number of homogeneous regions K is given and we have estimates of the GMM parameters, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_{\mathcal{A}}, \hat{\boldsymbol{\theta}}_{\bar{\mathcal{A}}})^T$, (details about the GMM learning procedure are presented in subsection 7.3.2). Then, the estimated probability maps can be constructed pixel-wise from Eqs. (7.11-7.12) by replacing the given GMM parameters by the GMM estimates.

A partial discreteness representation is now obtained by substituting the estimated probability maps and $\{\hat{\eta}_k\}_{k=1}^K$ in Eq. (7.13). The thus obtained partial discreteness representation $\mathcal{P}(\boldsymbol{x})$ is unique for each $\boldsymbol{x}$. More interestingly, every image $\boldsymbol{x}$ can be represented through its partial discreteness representation, including images that are not strictly partially discrete. Indeed, every image $\boldsymbol{x} \in \mathbb{R}_+^N$ can be decomposed as

$$\boldsymbol{x} = \mathcal{P}(\boldsymbol{x}) + (\mathcal{I} - \mathcal{P})(\boldsymbol{x}), \tag{7.16}$$

where $\mathcal{I} : \mathbb{R}^N \mapsto \mathbb{R}^N$ is the identity operator. The usefulness of this decomposition is that the second term, the residual form, serves as a measure of the partial discreteness error for all types of images. On the one hand, a partially discrete image is well represented by its partial discreteness representation, so its residual form, $(\mathcal{I} - \mathcal{P})(\boldsymbol{x})$, can be assumed to be small. On the other hand, images that do not obey the partial discreteness assumption possess a non-negligible residual form. Obviously, the prior $\Phi(\cdot)$ should be defined in agreement with this reasoning.

For partially discrete images, $(\mathcal{I} - \mathcal{P})(\boldsymbol{x})$ is almost zero except along edges. Since edges generally represent only a small fraction of the partially discrete image, we can justify that the residual form is sparse. Searching for the $\boldsymbol{x}$ that has the sparsest residual form implies to make the l_0 (pseudo) norm of the residual form minimal. As this combinatorial optimization problem is intractable, it is common practice to resort to other sparsity-inducing norms, such as the l_1 norm, the l_p $(0 < p < 1)$ semi-norms or other non-convex functionals such as log penalty functions [Trzasko and Manduca, 2009]. To illustrate partial discreteness, we use the l_1 norm in our

experiments. Hence, we define our partial discreteness prior $\Phi(\cdot)$ as

$$\Phi(\boldsymbol{x}) = ||(\mathcal{I} - \mathcal{P})(\boldsymbol{x})||_1. \tag{7.17}$$

7.3 Method

In this section, we propose a new reconstruction method that incorporates the partial discreteness (PD) prior term Eq.(7.17). The method will be denoted by the acronym PD.

7.3.1 Split Bregman reconstruction algorithm

In most of the MRI reconstruction algorithms the inequality constrained problem (Eq. (7.1)) is transformed into an unconstrained problem of the form

$$\min_{\boldsymbol{x} \in \mathbb{R}_+^N} J(\boldsymbol{x}, \lambda) \tag{7.18}$$

with

$$J(\boldsymbol{x}, \lambda) = \Phi(\boldsymbol{x}) + \frac{\lambda}{2}||\tilde{\boldsymbol{A}}\boldsymbol{x} - \boldsymbol{y}||_2^2. \tag{7.19}$$

Solving this new unconstrained problem is equivalent to solving (Eq. (7.1)) if and only if λ is selected according to the Karush-Kuhn-Tucker (KKT) conditions [Trzasko and Manduca, 2009]. Otherwise, the unconstrained solution, that is, the solution of (Eq. (7.18)), may not meet the data fidelity condition. Within the KKT approach, the optimal λ is called the KKT multiplier [Bertsekas, 1975]. Unfortunately, the analytic determination of the KKT multiplier is rather difficult or, in most of the cases, even impossible [Trzasko and Manduca, 2009]. As an alternative to the KKT technique, several iterative optimization algorithms replace the original problem (Eq. (7.1)) by a sequence of unconstrained minimization problems (as (Eq. (7.18))) where the cost function $J(\cdot, \lambda)$ is augmented to account for the constraints [Nocedal and Wright, 2006]. These subproblems are iteratively solved in combination with an update of some of the parameters included now in the augmented cost function. This class of algorithms includes penalty-based methods [Bertsekas, 1975], Augmented Lagrangian (AL) methods [Wu and Tai, 2010,Birgin and Martínez, 2012], and Split Bregman methods [Osher et al., 2005,Cai et al., 2009, Goldstein and Osher, 2009]. Ideally, the sequence of solutions of each subproblem asymptotically approaches the original KKT solution, i.e., it solves problem (Eq. (7.1)). In this work, to illustrate its potential, the Split Bregman method is chosen for the following practical reasons: it is numerically more stable than penalty-based methods [Goldstein and Osher, 2009] and it is simpler than AL methods. It should be noted that partial discreteness is not limited to a specific optimization algorithm. More recent optimization algorithms can be used as well, such as SpaRSA (Sparse Reconstruction by Separable Approximation) [Wright et al., 2009].

With the Split Bregman method, $J(\cdot, \lambda)$ is modified with the so-called Bregman distance [Bregman, 1967]. We refer the reader to [Goldstein and Osher, 2009] for a more detailed explanation. After some algebra, it can be demonstrated [Goldstein and Osher, 2009] that this method adopts the following simplified recursive scheme:

$$\boldsymbol{x}^{(t+1)} = \arg \min_{\boldsymbol{x} \in \mathbb{R}_+^N} \Phi(\boldsymbol{x}) + \frac{\lambda}{2} ||\tilde{\boldsymbol{A}}\boldsymbol{x} - \boldsymbol{b}^{(t)}||_2^2, \tag{7.20}$$

$$\boldsymbol{b}^{(t+1)} = \boldsymbol{b}^{(t)} + \boldsymbol{y} - \tilde{\boldsymbol{A}}\boldsymbol{x}^{(t+1)}. \tag{7.21}$$

The parameters update is done in Eq. (7.21), through the modified data vector $\boldsymbol{b}$ [Goldstein and Osher, 2009]. Each of the unconstrained minimization problems (Eq. (7.20)) is solved with a modified version of the Majorize-Minimize (MM)-based algorithm proposed by Muckley et al. [Muckley et al., 2015], with the non-linear conjugate gradient method. Exact convergence properties for these subproblems can be demonstrated [Muckley et al., 2015]. Further details on the MM-based algorithm, details on how to impose the real positivity constraint, as well as the analytical derivation of the gradient of $\Phi(\boldsymbol{x})$ are provided in subsection C.1.1.

7.3.2 GMM learning

Estimating the GMM parameters, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_{\mathcal{A}}, \hat{\boldsymbol{\theta}}_{\bar{\mathcal{A}}})^T$, is often called GMM training or learning. To learn the GMM, a set of samples drawn from the GMM, and a learning criterion are required.

7.3.2.1 Training data

If a dataset of fully-sampled reconstructed partially discrete images from the same object is a priori available, it can be used to train the GMM. However, in this work, we focus on the automatic application of the algorithm, and the GMM is learned in situ from an image reconstructed from under-sampled k-space data. Specifically, in our experiments, the GMM is trained with the magnitude of a low-resolution image $\boldsymbol{x_{LR}}$ calculated at the beginning of the algorithm. Such image is obtained as follows: an N_{Hamm}-point symmetric Hamming window is applied to the under-sampled k-space data $\boldsymbol{y}$ and then, an inverse Fourier transform of the windowed k-space data is calculated. It is assumed that the center of the k-space is fully sampled. Additionally, the GMM can be retrained every T_{GMM} iterations in order to keep track of the variations in the GMM along the reconstruction process.

7.3.2.2 GMM learning criterion

To train the GMM described by Eq. (7.8), we use the algorithm proposed by Figueiredo and Jain [Figueiredo and Jain, 2002]. In this algorithm, the Minimum Message Length criterion is implemented. It estimates the parameters of each component as well as the optimal number of components (i.e., the total number of

classes) of the GMM. Another benefit of this learning algorithm is that a careful initialization is not required, as opposed to ML expectation-maximization based methods [Figueiredo and Jain, 2002].

7.3.2.3 The selection of K

Once the GMM described by Eq. (7.8) is learned, the number of homogeneous regions K should be selected and the corresponding conditional PDFs for the homogeneous regions, $\{p_{X|\mathcal{A}_k}(x;\hat{\eta_k},\hat{\sigma_k})\}_{k=1}^{K}$, should be detected. In our work, we advocate for a manual selection of K, prior to the application of the PD algorithm. In this approach, K is selected by visual inspection of the image obtained by basic zero-filled (ZF) reconstruction. For most images we investigated, at least two homogeneous regions can be discerned: a background and a hyper-intense region. The corresponding conditional PDFs are easily identified by selecting those with the lowest and highest estimated mean, respectively. The remaining conditional PDFs are arranged in ascending order with respect to their standard deviation. Selecting $K \geq 3$ homogeneous regions may enhance the performance of the PD if the partially discrete image at hand is indeed constituted of more homogeneous regions than just the background and a hyper-intense region. This was illustrated in a simulation experiment that was set up to evaluate the sensitivity of the PD method to the selection of K and which is described in subsection C.2.1 If $K = 3$, in addition to the conditional PDFs associated with the background and hyper-intense region, the conditional PDF with the lowest standard deviation is chosen. If $K = 4$, also the second conditional PDF is selected and so on. In the experiments that we performed to compare PD with state of the art reconstruction methods, we consistently set K equal to 2, which can be considered as a conservative choice.

7.3.3 Parameters selection

For convex reconstruction problems with noiseless under-sampled data $\boldsymbol{y}$, independent of the λ selection in Eq. (7.20), the iterative solutions of the Split Bregman method asymptotically satisfy the data fidelity condition ($\tilde{\boldsymbol{A}}\boldsymbol{x} = \boldsymbol{y}$) and monotonically decrease the prior term [Goldstein and Osher, 2009]. For non-convex prior terms, such as partial discreteness, and with noisy data $\boldsymbol{y}$, as in Eq. (7.1), convergence to the global minimum cannot be guaranteed. Fortunately, in this situation, the Split Bregman algorithm has been experimentally observed to converge, even though theoretical proof is still lacking [Liu et al., 2013, Chartrand, 2009, Li et al., 2012a, Wang et al., 2014, Estellers et al., 2012].

To achieve a good performance, a careful initialization $\boldsymbol{x}^{(0)}$ and an adequate selection of λ are of great importance. Our choice for $\boldsymbol{x}^{(0)}$ is the magnitude of the low-resolution reconstructed image $\boldsymbol{x_{LR}}$. This image was used as well for training the GMM. Our magnitude image $|\boldsymbol{x_{LR}}|$ lacks details and texture is hardly preserved, but it has important advantages for partial discreteness: 1) artefacts are not so strongly manifested as in other $\boldsymbol{x}^{(0)}$ choices (e.g., Tikhonov regularization on $\boldsymbol{x}$), and 2) the background area is easily discernible. Consequently, the initial partial

discreteness representation, $\mathcal{P}(|\boldsymbol{x_{LR}}|)$, does not undesirably magnify artefacts, and more important, the background is already accurately represented by $\mathcal{P}(|\boldsymbol{x_{LR}}|)$. The low resolution image $\boldsymbol{x_{LR}}$ also serves to estimate the phase $\psi_{\boldsymbol{x}}$. From this image, the principal complex argument is voxel-wise calculated [Lustig et al., 2007]. With the estimated phase image, $\hat{\psi}_{\boldsymbol{x}} \in (-\pi, \pi]^N$, $\tilde{\boldsymbol{A}}$ is defined.

The basis for defining the weighting λ parameter is the following simple rule: the better the type of image we are reconstructing adheres to a partially discrete image, the lower λ should be. Furthermore, the design of λ should take into account that the prior term $\Phi(\cdot)$ is often several orders of magnitude larger than the data fidelity l_2 norm. Empirically, we have corroborated that a satisfactory formula in our experiments was the following:

$$\lambda = 2(1-r) \cdot 10^3, \tag{7.22}$$

with $0 \leq r < 1$ a value which we term the partial discreteness degree. In practice, the more constant we expect the hyper-intense regions to be, the closer r should be to 1. The closer r is to 0, the less relevant the partial discreteness prior becomes.

The pseudo code of the proposed PD image reconstruction algorithm is presented in Fig. 3.

Algorithm 3 Pseudo-code of the PD algorithm.

$\hat{\psi}_{\boldsymbol{x}} \leftarrow$ Phase-estimation$\{\boldsymbol{x_{LR}}\}$
$\hat{\boldsymbol{\Psi}} = \text{diag}(e^{i\hat{\psi}_{\boldsymbol{x}}})$
$\tilde{\boldsymbol{A}} = \boldsymbol{A}\hat{\boldsymbol{\Psi}}$
$\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_{\mathcal{A}}, \hat{\boldsymbol{\theta}}_{\bar{\mathcal{A}}}] \leftarrow$ GMM-learning$\{|\boldsymbol{x_{LR}}|\}$
Define $\mathcal{P}(\cdot)$ with $\hat{\boldsymbol{\theta}}$ as in subsection 7.2.3
set $t = 0$, $\boldsymbol{x}^{(0)} = |\boldsymbol{x_{LR}}|$, $\boldsymbol{b}^{(0)} = \boldsymbol{y}$
while $t < T_{\max}$ and $||\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}||_2 \geq$ Tol **do**
    $\boldsymbol{x}^{(t+1)} = \arg\min_{\boldsymbol{x} \in \mathbb{R}_+^N} \Phi(\boldsymbol{x}) + \frac{\lambda}{2}||\tilde{\boldsymbol{A}}\boldsymbol{x} - \boldsymbol{b}^{(t)}||_2^2$
    $\boldsymbol{b}^{(t+1)} = \boldsymbol{b}^{(t)} + \boldsymbol{y} - \tilde{\boldsymbol{A}}\boldsymbol{x}^{(t+1)}$
    if $t = vT_{\text{GMM}}$ for any $v \in \mathbb{N}^+$ **then**
        $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_{\mathcal{A}}, \hat{\boldsymbol{\theta}}_{\bar{\mathcal{A}}}] \leftarrow$ GMM-learning$\{\boldsymbol{x}^{(t+1)}\}$
        Define $\mathcal{P}(\cdot)$ with $\hat{\boldsymbol{\theta}}$ as in subsection 7.2.3
    end if
    $t \leftarrow t + 1$
end while
$\hat{\boldsymbol{x}} = \boldsymbol{x}^{(t)}$

7.3.4 Multi-coil extension

The proposed method can be extended to be applicable to multi-coil acquisitions. In that case, the algebraic linear model (Eq. (2.43)) should be extended. Let $\boldsymbol{y}_r \in \mathbb{C}^M$

be the k-space data acquired by the r-th coil, with $r = 1, ..., R$. The relation with the reconstructed image $\boldsymbol{x}_r$ is again

$$\boldsymbol{y}_r = \boldsymbol{A}\boldsymbol{x}_r + \boldsymbol{n}_r. \tag{7.23}$$

Each of $\boldsymbol{x}_r \in \mathbb{C}^N$ is related to the true magnitude partially discrete image $\boldsymbol{x} \in \mathbb{R}_+^N$ through the coil sensitivities, $\boldsymbol{c}_r \in \mathbb{C}^N$, as $\boldsymbol{x}_r = \boldsymbol{C}_r \boldsymbol{\Psi} \boldsymbol{x}$ with $\boldsymbol{C}_r = \text{diag}(\boldsymbol{c}_r)$. If we call $\boldsymbol{y}^T = (\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, ..., \boldsymbol{y}_R^T)$ and $\tilde{\boldsymbol{A}} = \boldsymbol{A}_{\boldsymbol{Block}} \boldsymbol{\Psi}$ with

$$\boldsymbol{A}_{\boldsymbol{Block}} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A} \end{pmatrix} \begin{pmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \\ \vdots \\ \boldsymbol{C}_R \end{pmatrix}, \tag{7.24}$$

partial discreteness can be applied as in subsection 7.3.1.

7.3.5 Relation to Non-Local Total Variation (NLTV)

Because partially discrete images are composed of several homogeneous regions with very low-intensity variation, it may be tempting to reconstruct them with spatially-adapted versions of TV-based methods, such as Non-Local (NL) TV algorithms [Gilboa and Osher, 2007, Gilboa and Osher, 2008, Liang et al., 2011]. Here, we provide an insightful comparison between NLTV and our partial discreteness prior. While NLTV accounts for the low-intensity variation using concepts rooted in calculus of variations and measure theory [Gilboa and Osher, 2007], partial discreteness uses tools from Bayesian inference and unsupervised learning and clustering.

Aside from being different in nature, both approaches handle prior information very differently. On one hand, partial discreteness promotes quasi-constant images in an explicit fashion, through the partial discreteness representation, i.e., an image. On the other hand, NLTV does it indirectly, through the non-local gradient [Gilboa and Osher, 2007,Sawatzky, 2011]. Indeed, in NLTV algorithms, the spatial gradient of $\boldsymbol{x}$, included in the original TV measure, is replaced by the non-local gradient of $\boldsymbol{x}$: a vector $\nabla_m^w \boldsymbol{x} \in \mathbb{R}^N$ which at pixel m is defined as [Liang et al., 2011]

$$[\nabla_m^w \boldsymbol{x}]_n = \sqrt{w_{mn}}(x_m - x_n) \quad \text{with} \quad n = 1, ..., N, \tag{7.25}$$

where $\boldsymbol{W} \triangleq \{w_{mn}\} \in \mathbb{R}_+^{N \times N}$ is the graph matrix. In order to define the NLTV term, first the l_2 norm of each $\nabla_m^w \boldsymbol{x}$ with $m = 1, ..., N$ is taken and then, the l_1 norm of the resulting vector is calculated [Liang et al., 2011]. The graph matrix $\boldsymbol{W}$ plays the role of adaptive mechanism and weights the intensity differences according to the image spatial content. A large weight w_{mn} is assigned to two similar pixels, therefore penalizing deviation in intensity. Intensity deviations of pairs of dissimilar pixels must not contribute to the NLTV term, hence, zero values are selected in this case. The interested reader is referred to [Liang et al., 2011] for a more specific interpretation. At this point, a pertinent observation can be made. NLTV still

needs a learning mechanism to account for the image structure and thus to define $\boldsymbol{W}$. This means that the performance of NLTV can not be separately assessed from the learning process. In fact, since the graph matrix $\boldsymbol{W}$ is the tool which most leverages NLTV performance, it should be very carefully designed. Yet, the selection of an appropriate similarity metric for weighting the pixels is still arguable and application-dependent [Gilboa and Osher, 2007, Sawatzky, 2011], not to mention the computational complexity of exhaustive searches in patch-based similarity algorithms [Sawatzky, 2011]. Fortunately, the problem with the similarity metrics is circumvented with partial discreteness. The learning mechanism is naturally embedded into the GMM, which carries its own sophisticated learning/clustering procedure, based on Bayesian inference. The special image features of partially discrete images are condensed in the partial discreteness representation, derived from the unified framework presented in subsection 7.2.2.

We emphasize the importance of the GMM in the construction of the partial discreteness prior. Its employment in MRI is not new. For instance, it has been used for unsupervised segmentation [Van Leemput et al., 1999b] and bias field correction [Van Leemput et al., 1999a, Mai et al., 2011]. On top of that, it has been exploited as prior knowledge in model-based image restoration problems, where it has shown excellent results [Yang et al., 2015]. Certainly, in the Compressed Sensing (CS) field, it has been recently proved [Renna et al., 2014] that exact signal reconstruction drawn from a GMM is achievable with a substantially lower number of measurements than commonly required with other sparse recovery methods.

7.4 Experiments

In this section, we describe the experiments that were carried out to evaluate the performance of the proposed PD algorithm. First, PD was compared against commonly used reconstruction methods, conducting experiments on simulated as well as real k-space data. Next, dedicated simulation experiments on a digital phantom were performed to test the sensitivity of the algorithm to (i) the pre-selected number of homogeneous regions, K, (ii) phase profiles that do not satisfy the assumption of a slowly varying phase, and (iii) intensity inhomogeneity or bias field [Fan et al., 2003, Vovk et al., 2007]. In all experiments, the starting point was a fully sampled data set which was retrospectively under-sampled. Throughout, validation was not only based on visual assessment but also on quantitative results. We employed the following measures to evaluate the reconstruction quality: the Peak Signal-to-Noise Ratio (PSNR) [Ravishankar and Bresler, 2011], the Quantitative Index based on Local Variance (QILV) [Aja-Fernández et al., 2006], the High Frequency Error Norm (HFEN) [Ravishankar and Bresler, 2011] and the Feature Similarity Index Metric (FSIM) [Zhang et al., 2011].

7.4.1 Comparison of PD with popular reconstruction methods

Comparative experiments were conducted on simulated as well as real k-space data. Different types of under-sampling scenarios were considered, namely, structural and random patterns. In this way, we show that PD does not require any specific assumption on the type of sampling, in contrast to common CS-based reconstruction methods [Haldar et al., 2011]. PD was compared against the basic ZF reconstruction as well as three conventional reconstruction methods dedicated to the reconstruction of under-sampled data, namely:

1. SparseMRI, proposed by Lustig [Lustig et al., 2007], which implements the concept of CS including, as prior terms, TV and l_1-wavelet sparsity. The Matlab code is publicly available [Lustig,].

2. CS+NLTV, proposed by Lian [Liang et al., 2011], which incorporates NLTV in the CS framework.

3. LORAKS proposed by Haldar [Haldar, 2014a]. LORAKS assumes images to have small finite support compared to the Field Of View (FOV) and/or a slowly varying phase. The Matlab code is publicly available [Haldar, ,Haldar, 2014b].

SparseMRI was applied with the built-in parameters, except for the TV weight and the number of iterations. Those values were extensively varied until no further artefacts reduction could be achieved. LORAKS was implemented with the built-in parameters. Experiments with different settings did not provide any remarkable difference. We followed the guidelines presented in the original work [Liang et al., 2011] to implement CS+NLTV. Parameter settings were chosen according to the recipe in the original work. Regarding PD, the length of the Hamming window N_{Hamm} was set in all experiments to 50. PD was stopped with the following parameters: $T_{\max} = 8$ and $\text{Tol} = 10^{-4}$. For simplicity and speed, we did not retrain the GMM. Regarding the weighting parameter λ, the corresponding r value is mentioned. The number of discrete classes was in all experiments set to $K = 2$.

7.4.1.1 Simulated k-space data

For the simulation experiment, a 256×256 actual magnitude brain MR image (Fig. 7.2.(a)) was used as ground-truth. The image was acquired with an Inversion Recovery (IR) pulse sequence on a 3T Siemens scanner with a 32-channel. A smoothly varying phase was simulated by Legendre polynomials up to the second degree. Legendre polynomials were used because of their demonstrated suitability for simulating slow-varying intensity profiles [Styner et al., 2000]. The resulting complex image was polluted with additive complex-valued white Gaussian noise with uncorrelated real and imaginary parts of equal standard deviation σ. The

value of σ was chosen such that the signal-to-noise ratio (SNR), defined as

$$\mathrm{SNR} = \frac{\overline{\boldsymbol{x}}}{\sigma}, \tag{7.26}$$

with $\overline{\boldsymbol{x}}$ the spatial mean of the magnitude image $\boldsymbol{x}$, was equal to 10. From the noisy, complex-valued image, k-space samples on a Cartesian grid were generated using the linear model of Eq. (2.43). Single-coil data were mimicked with structural under-sampling: a pseudo-radial sampling mask (Fig. 7.2.(b)) with 70% of non-acquired k-space points. This procedure was repeated $N_{\mathrm{Runs}} = 20$ times with different noise realizations. We chose a moderate value of the *partial discreteness degree*, r, i.e., $r = 0.1$, which was observed to give reconstructed images with good details preservation.

7.4.1.2 In vivo human knee k-space data

In this experiment, in vivo 3D fully sampled k-space data of a human knee were employed. The k-space data, available at [Lustig and Vasanawala,], were acquired with a 3D Fast Spin Echo (FSE) pulse sequence on a 3T scanner with a eight-channel coil. From the 3D k-space data, 2D slice reconstruction was accomplished. We applied an inverse Fourier transform [Haldar, 2014a] along the third dimension, and then, one and the same mid-sagittal 2D k-space slice was extracted per each coil. This approach is valid since the 3D Fourier transform is a separable operator [Jain, 1989]. The corresponding k-space slices (320×320) were under-sampled with a random phase-encoding mask of 50% of missing k-space lines (Fig. 7.3.(b)).

For ease of comparison, and because publicly available Matlab codes of SparseMRI and LORAKS only permit single-coil reconstruction, each of the eight 2D k-space data sets was reconstructed independently. To create a final image per method, the eight reconstructed images were combined using the Sum of Squares (SoS) method [Larsson et al., 2003]. Since the FSE sequence produces a highly bright and constant area in the cartilage region, it is reasonable to assume the partial discreteness holds even more than in the simulation experiment. Therefore, for PD, we increased the r value to 0.9.

7.4.1.3 In vivo human brain k-space data

To finish the experiments section, we validate PD in a truly multi-coil reconstruction scheme, where coil sensitivities estimation is required. In vivo 3D fully sampled k-space data were used again, this time, from a human brain. K-space data were acquired with a 3T scanner with eight coils. To allow for 2D slice reconstruction, we followed the same routine as in the knee experiment. Specifically, we applied an inverse Fourier transform along the second dimension and then, one and the same mid-axial 2D k-space slice was extracted per each coil. The resulting k-space slices (230×180) were under-sampled with a 2D variable density random mask of 75% of missing k-space points (Fig. 7.4.(b)).

Coil sensitivities, $\{\boldsymbol{c}_r\}_{r=1}^{R=8}$, were estimated with the SoS method [Larsson et al., 2003] from the undersampled data. Next, the initial low-reconstructed image $\boldsymbol{x_{LR}}$ was obtained as $\boldsymbol{x_{LR}} = \boldsymbol{A}_{\mathbf{Block}}^H \boldsymbol{y}_{N_{\text{Hamm}}}$ where $\boldsymbol{A}_{\mathbf{Block}}^H$ denotes the Hermitian transpose of $\boldsymbol{A_{Block}}$ and $\boldsymbol{y}_{N_{\text{Hamm}}}$ are the under-sampled k-space data filtered with a Hamming window ($N_{\text{Hamm}} = 50$). The phase $\psi_{\boldsymbol{x}}$ was estimated from $\boldsymbol{x_{LR}}$ as well. From $\hat{\psi}_{\boldsymbol{x}}$, $\boldsymbol{\Psi}$ was determined after which $\tilde{\boldsymbol{A}}$ was redefined as proposed in subsection 7.3.4. Finally, PD was applied with the same Split Bregman method. As in the brain simulation experiment, $r = 0.1$ was selected. We also compared PD with the ZF reconstruction, calculated as $\boldsymbol{A}_{\mathbf{Block}}^H \boldsymbol{y}$, and with the CS+NLTV method. In both cases, the coil estimation procedure was done in exactly the same manner as for PD.

7.4.2 Sensitivity analysis of the PD method

This subsection summarizes the main results of a sensitivity analysis of the proposed PD method to various parameters and deviations from assumptions. The analysis is based on dedicated simulation experiments, which are extensively discussed in subsection C.2 of the Appendix C.

7.4.2.1 Sensitivity to the pre-selected number of homogeneous regions K

To test the sensitivity to the pre-selected number of homogeneous regions K, a dedicated simulation experiment was conducted on a digital phantom image. This simulation experiment is described in subsection C.2.1.

7.4.2.2 Sensitivity to non-slowly varying phase

As described in subsection 7.3.3, the required estimate of the image phase is obtained from a low-resolution image $\boldsymbol{x_{LR}}$. The rationale for this procedure is that real-life phase images are often slowly varying. Note that a slowly varying or smooth phase is a common assumption in MR image reconstruction methods [Lustig et al., 2007]. Nevertheless, phase images may also have significantly higher spatial frequency content [Zhao et al., 2012], especially when gradient echo instead of spin echo imaging sequences are used [Feng et al., 2013]. To study the sensitivity of the PD algorithm to non-slowly varying phase, we carried out a simulation experiment (cfr. subsection C.2.2), in which the performance of PD for three different phase profiles was evaluated.

7.4.2.3 Sensitivity to bias fields

Bias fields are undesired low-frequency signals induced by inhomogeneities in the magnetic fields of the MRI system [Juntu et al., 2005]. The presence of a bias field may challenge the main assumption underlying the PD method, namely that the image to be reconstructed contains regions of quasi-constant intensity. To

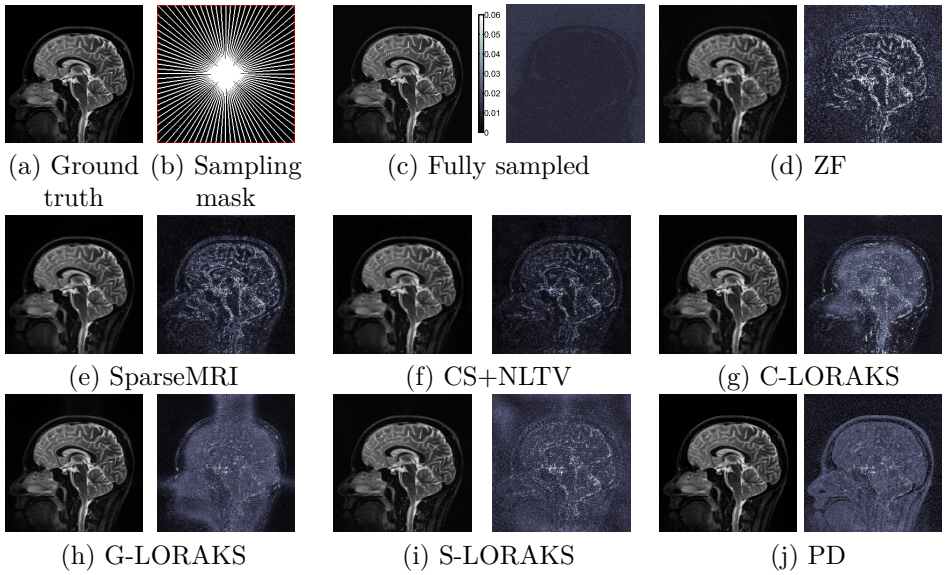
evaluate the performance of the PD method in the presence of a bias field, dedicated simulation experiments were conducted on a digital phantom. These simulation experiments are described in subsection C.2.3. Bias fields with different degrees of variation were considered.

7.5 Results

7.5.1 Comparison of PD with popular reconstruction methods

7.5.1.1 Simulated k-space data

The expected value of the magnitude of the reconstructed images, that is, the sample mean over the N_{Runs} realizations are shown in Fig. 7.2. Root-Mean-Squared Error (RMSE) maps are presented as well. Numerical results are provided in Table 7.1.



(a) Ground truth (b) Sampling mask (c) Fully sampled (d) ZF

(e) SparseMRI (f) CS+NLTV (g) C-LORAKS

(h) G-LORAKS (i) S-LORAKS (j) PD

Figure 7.2: Visual results for the experiment with simulated k-space data. The sample mean of the magnitude of the reconstructed images are shown in companion of the RMSE maps. To highlight small errors, the colorbar range of the RMSE maps was adapted to $[0, T]$, where T is 40% of the maximum value which was found in all RMSE maps (i.e., considering altogether).

From Fig. 7.2, it can be observed that SparseMRI and CS+NLTV, though successful in removing noise, do not recover a high resolution image. Their RMSE maps reveal substantial structural errors, mainly located at edges. This observation agrees with what was already pointed out in [Haldar, 2014a] for SparseMRI. As

Table 7.1: Quantitative results for the experiment with simulated k-space data.

	PSNR [dB]	QILV	HFEN	FSIM
Fully sampled	35.3	0.998	0.070	0.990
ZF	28.7	0.747	0.235	0.903
SparseMRI	**34.6**	0.907	0.153	0.938
CS+NLTV	33.4	0.921	0.141	0.934
C-LORAKS	33.2	0.930	0.121	0.952
G-LORAKS	31.1	0.953	0.134	0.940
S-LORAKS	33.5	0.976	0.112	0.953
PD	34.5	**0.986**	**0.090**	**0.962**

expected, edges are slightly better preserved with CS+NLTV [Liang et al., 2011]. All LORAKS versions and specially PD restore images with substantially higher resolution than ZF, SparseMRI and CS+NLTV. Indeed, their RMSE maps exhibit a noisy pattern with a very moderate structural degradation effect. This is expected if reconstruction methods succeed in recovering high-resolution missing k-space data (see the fully sampled case).

Among all LORAKS versions, S-LORAKS seems to perform best. However, still some (small) structural details errors can be observed. These are considerably attenuated with PD though. Numerical results are in agreement with visual findings. PD obtains the best result (disregarding the fully sampled scenario), in terms of QILV, HFEN, and FSIM. PD is closely followed by S-LORAKS, while there is a significant difference compared to CS+NLTV and SparseMRI. The highest PSNR for SparseMRI might be attributed to its noise removal capability, having a relevant effect specially in the background. With an Intel Core i7-4770K 3.5 GHz (32 GB RAM) processor, the average time for LORAKS reconstruction was about 10 min. The Singular Value Decomposition (SVD) incorporated in the method is probably the reason of such a computational burden. As already mentioned in subsection 7.3.5, the calculation of $\boldsymbol{W}$ for the CS+NLTV method is computationally quite expensive. With the parameters chosen as those recommended in the original work [Liang et al., 2011], the average computation time for CS+NLTV was about 8 min. PD was able to reconstruct images within roughly 4 min. The main computational effort is in the GMM learning. Note that when the GMM parameters are known, each of the subproblems of the Split Bregman method can be rapidly implemented with the MM algorithm presented in section C.1. Clearly, the fastest algorithm is SparseMRI with a computation time sometimes below 1 min. It should be noted that all algorithms were implemented in Matlab. A C++ implementation would greatly speed up all algorithms. This holds especially for the graph matrix calculation of CS+NLTV and the GMM learning for PD.

7.5.1.2 In vivo human knee k-space data

Reconstructed SoS images as well as the absolute error maps (with the magnitude of the fully sampled image as ground truth), are shown in Fig. 7.3, while quantitative

results are reported in Table 7.2. We can draw similar conclusions about the

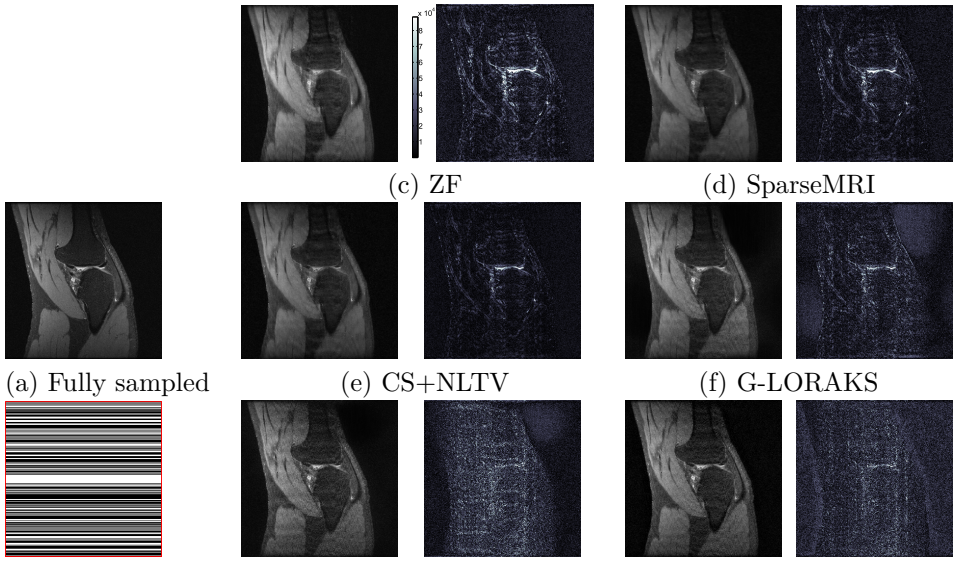


Figure 7.3: Visual results for the experiment with real k-space data of the knee. The SoS reconstructed images are shown in companion of the absolute errors maps. To highlight small errors, the colorbar range of the absolute error maps was adapted to $[0, T]$, where T is 40% of the maximum value which was found in all absolute error maps (i.e., considering altogether).

Table 7.2: Quantitative results for the experiment with in vivo human knee k-space data.

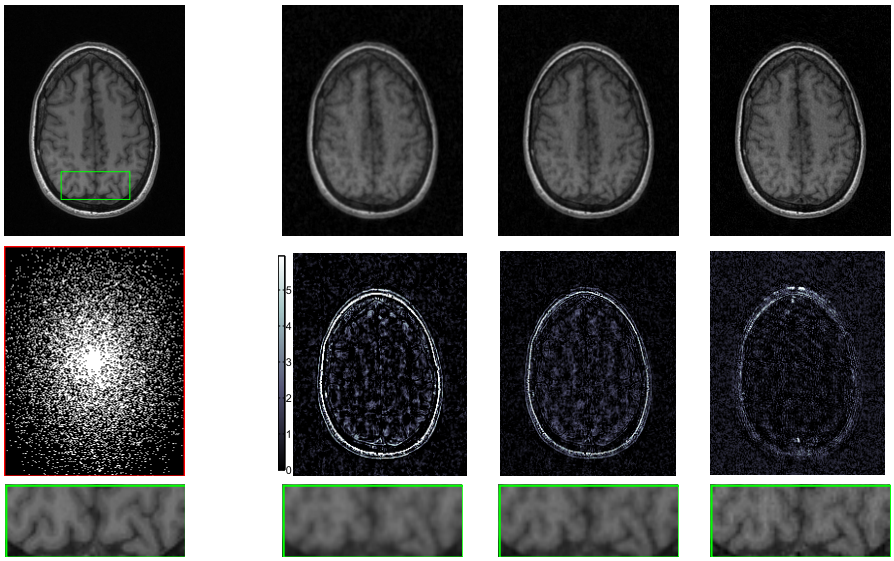
	PSNR [dB]	QILV	HFEN	FSIM
ZF	29.84	0.454	$2.96 \cdot 10^5$	0.887
SparseMRI	30.29	0.620	$2.72 \cdot 10^5$	0.860
CS+NLTV	**30.33**	0.710	$2.67 \cdot 10^5$	0.888
G-LORAKS	30.15	0.798	$2.48 \cdot 10^5$	**0.893**
S-LORAKS	27.78	0.847	$2.45 \cdot 10^5$	0.837
PD	27.92	**0.857**	$\mathbf{2.41 \cdot 10^5}$	0.883

performance of PD in comparison to the rest of the methods. The artefacts manifested due to the random sampling are partially removed by SparseMRI and CS+NLTV, but both methods failed in recovering highly detailed clinical relevant areas, for example, the contours in the cartilage region. PD restored a higher detailed image as can be seen as well by looking at the errors map. Concerning the LORAKS versions, we first notice that C-LORAKS failed to converge (results not shown). G-LORAKS was able to recover a better defined cartilage but still it is largely outperformed by S-LORAKS and PD. Reconstructed images with S-LORAKS are slightly less accurate than those obtained by PD, as is manifest

in the metrics of Table 7.2. Certainly, PD scores best for metrics which are specially conceived to assess small details preservation, that is, QILV and HFEN. The best FSIM case for G-LORAKS may be understood if we notice that G-LORAKS provides a good balance between artefacts suppression and structural details preservation. Finally, the highest PSNR for CS+NLTV could be largely based on its ability for artefacts removal. The computational time of all methods, for each coil, were very similar to the times reported in the simulation experiment and hence they are not repeated here.

7.5.1.3 In vivo human brain k-space data

Undoubtedly, it can be seen from Fig. 7.4 that the magnitude of the reconstructed image with PD possesses higher resolution than the image restored with CS+NLTV and specially ZF. A closer look at zoomed images (Fig. 7.4.(j-l)) reveals that the interfaces between white/gray matter are better preserved with PD compared to CS+NLTV. Absolute error maps also demonstrate that higher structural errors are more widely manifested with CS+NLTV than with PD. Quantitative results in Table 7.3 further suggest the superiority of PD in detail preservation and resolution enhancement. PD ranked best for three of the four metrics. Nevertheless, it is



(a) Fully sampled (b) ZF (c) CS+NLTV (d) PD

Figure 7.4: Visual results for the experiment with real k-space data of the brain. The magnitude of the multi-coil reconstructed images are shown (top row) in companion of the absolute error maps (middle row (b-d)). To highlight small errors, the colorbar range of the absolute error maps was adapted to $[0, T]$, where T is 40% of the maximum value which was found in all absolute error maps (i.e., considering altogether). Zoomed region for each case are shown in the bottom row.

also clear that NLTV outperforms PD in noise suppression. This is not surprising

Table 7.3: Quantitative results for the experiment with in vivo human brain k-space data.

	PSNR [dB]	QILV	HFEN	FSIM
ZF	28.76	0.547	17.99	0.888
CS+NLTV	31.06	0.844	11.70	**0.930**
PD	**31.50**	**0.995**	**8.340**	0.926

since the NLTV prior term, incorporated in the CS+NLTV method, has proved to be a very effective denoising mechanism [Gilboa and Osher, 2008]. Indeed, the first application of the TV measure in the image processing field was noise removal [Rudin et al., 1992]. It is interesting to remark that our partial discreteness prior does not attempt to remove noise but to recover the unknown partially discrete image using a priori information, which we carefully modeled from its special structure. Neither the GMM nor the Bayesian probabilistic segmentation were designed to tackle noise. However, showing its flexibility, our partial discreteness prior can easily accommodate a simple regularization term in companion of the partial discreteness representation to cope with noise while still exploiting all the potential of this novel image representation. For the reader's interest, we point out that the *a posteriori* probability maps derived from the GMM have a broad range of applications for MR reconstruction, tissue-selective filtering being one of them. A reduced list of these applications and some extensions of the GMM are given at the end of the conclusion section.

7.5.2 Sensitivity analysis of the PD method

7.5.2.1 Sensitivity to the pre-selected number of homogeneous regions K

The results of the experiment show a clear gain in reconstruction quality if at least one homogeneous region (e.g., background) is chosen. A further substantial improvement was achieved by also incorporating the hyper-intense region ($K = 2$). The optimal value of K for the phantom image was found to be equal to 4, which demonstrates that for particular partially discrete images, the performance of PD can be further improved by selecting K higher than 2. Obviously, the optimal value of K will depend on the image to be reconstructed. Note that in the experiments that we performed to compare PD with state-of-the-art reconstruction methods (see subsection 7.4.1), K was consistently set equal to 2, which can be considered as a conservative choice.

7.5.2.2 Sensitivity to non-slowly varying phase

As expected, the performance of PD degrades for highly-varying phase profiles, which indicates the importance of the smooth phase assumption. However, it was also found that PD is robust to moderate phase variations.

7.5.2.3 Sensitivity to bias fields

The results of the simulation experiments show that, in general, the performance of PD deteriorates when the degree of variation of the bias field increases. In addition, since not all homogeneous regions may be equally affected, a bias field may also influence the optimal choice of K, suggesting that in the presence of a severe bias field, a conservative choice of K is advisable, as is further motivated in subsection C.2.3. In that subsection, we also elaborate on the possibility of improving the robustness of the PD method by including a bias field correction technique.

7.6 Future work

Accelerating the reconstruction of individual MR images with undersampled k-space data effectively reduces the total scanning time of quantitative MRI (qMRI) modalities, such as relaxometry. In this sense, partial discreteness is of undeniable help to shorten the long acquisition time of T_1 and T_2 mapping. However, it becomes natural to wonder whether the information provided by the relaxation model can be integrated into the reconstruction method, aiming at exploiting the coupled information in the dataset of T_1 or T_2-weighted images. This idea will be touched lightly at the end of the conclusions section (section 7.7). In this section, we give an outlook and discuss the possible ways to include relaxometry information in the PD method, thereby leading to unprecedentedly high acceleration ratios for MRI relaxometry acquisitions.

Essentially, there are two methodologies so as to include relaxation model information. We can include relaxometry information by exploiting inter-image redundancy along the parametric dimension, that is, the dimension given by the number of images. Alternatively, the relaxation model can be included explicitly in the reconstruction framework. We will call the methods that relate to the first methodology *implicit* reconstruction approaches, whereas the second class of methods will be called simply *relaxation model-based* reconstruction techniques.

7.6.1 PD as an implicit relaxation reconstruction method

Inter-image redundancy is the key characteristic to be captured along the parametric dimension. The parametric dimension could be represented by time, e.g., inversion times for an IR T_1-weighted images dataset, or echo times for a multi-echo T_2 weighted image sequence, but could be of different nature as well, e.g., flip angles for an SPGR T_1-weighted data set. Inter-image redundancy has been employed by using principal component analysis [Petzschner et al., 2011], imposing smoothness on the relaxation signal model [Velikina et al., 2013], or, more recently, by encouraging spatiotemporal low-rankness [Zhao et al., 2015,Peng et al., 2016,Lee et al., 2016].

An extension of the GMM of partial discreteness to a multivariate GMM can simply accommodate the desirable inter-image redundancy along the second dimension.

In the original formulation, the intensity of the pixel was modeled as a random variable $X : \Omega \mapsto \mathbb{R}^+$. Now, we are given a set of images $\boldsymbol{x}^s$ with $s = 1, ..., S$, and where S is the number of T_1 or T_2-weighted images. We model the intensity curve, that is the intensity values from the relaxation model that can be obtained by focusing one given pixels, as a random vector $\boldsymbol{X} = (X^1, ..., X^S)^T$. Hence, for a fixed pixel n, the vector of intensity values is a realization of the random vector $\boldsymbol{X}$.

We then assume that $\boldsymbol{X}$ follows a multivariate GMM model, where each component is a S-dimensional Gaussian distribution. The covariance matrix of each of the components of the GMM now captures the intensity of a particular class, and its (temporal) evolution along the parametric dimension. Constraints on the covariance matrix of each GMM component may be enforced so as to promote highly correlated consecutive variables, e.g., X^s and X^{s+1}.

Finally, we note that the multivariate GMM model extension can be applied beyond relaxometry, allowing a joint reconstruction of a set of MR images, for example, in accelerated dynamic MRI [Otazo et al., 2015], diffusion-weighted MRI or 3D imaging [Setsompop et al., 2012].

7.6.2 PD as a relaxation model-based reconstruction method

Implicit relaxation methods exploit the redundancy along the inter-image dimension, without including the relaxation model in the reconstruction process. Hence, the estimation of spatial T_1 and T_2 maps is performed only after the reconstruction of the whole undersampled k-space data set. That is, implicit relaxation methods are yet two-step approaches.

With model-based reconstruction approaches, on the other hand, the optimization problem is completely reformulated, since the relaxation model is included in the cost function, and now spatial T_1 or T_2 maps are the variables to be reconstructed, (i.e., estimated). Relaxation model-based reconstruction methods have been recently proposed for T_2 mapping [Huang et al., 2012] and T_1 mapping [Block et al., 2009, Wang et al., 2017]. All of these methods incorporate some kind of prior knowledge of the relaxometry maps, being either smoothness by means of Tikhonov regularization [Block et al., 2009], or joint sparsity in the wavelet domains of steady-state magnetization, equilibrium, and T_1 maps (those three maps are estimated simultaneously) [Wang et al., 2017].

The extension of the PD method to a relaxation model-based reconstruction method is conceptually simpler than if we do so in an implicit scheme (see subsection 7.6.1). The main drawback is that the resulting optimization problem is substantially more complicated to solve. Indeed, the partial discreteness prior term $\boldsymbol{\Phi}(\cdot)$ can be trivially extended to account for all of the S T_1 or T_2-weighted images, by applying the partial discreteness representation to each of the N images that constitute the collected dataset. However, the data fidelity term of Eq. (7.1) should be modified to account for the relaxation model, giving a more complicated cost function. Since the new the data fidelity term is a non-quadratic function, the MM technique

of [Muckley et al., 2015] cannot be applied directly, so other optimization algorithms should be conceived.

Obviously, clever options that exploit the partial discreteness representation in a unified way are possible, leading to higher acceleration rates. For example, the partial discreteness prior can be directly applied to the relaxometry maps, instead of the individual relaxometry-weighted images. Finally, it is important to mention that, unlike two-step approaches, model-based approaches have the benefit that, since they constitute a single approach, they can be formally seen as T_1 or T_2 estimators. Note that with a two-step approach this is not possible. Note that the statistical properties of these estimators can be studied in a systematic way using the theory presented in chapter 4.

7.7 Conclusions

In this work, we have presented a novel prior, partial discreteness, for the reconstruction of MR images with quasi-constant intensity regions as well as heterogeneous regions. We have shown that every image can be additively decomposed into its partial discreteness representation and its residual form. The partial discreteness representation, which is based on a GMM, embodies the basic features of partially discrete images: constant intensity in homogeneous regions and texture in heterogeneous regions. Exploiting this partial discreteness representation in MR image reconstruction, by enforcing sparsity on the residual form, we have been able to reconstruct highly detailed images from under-sampled data with structural and random under-sampling schemes, namely, pseudo-radial, random phase-encoding and pseudo-random variable density sampling. In this work, partial discreteness has been implemented in a phase-constrained formulation where the phase map was estimated from a low-resolution image. Hence, we have implicitly made the common assumption of smoothly varying phase images [Lustig et al., 2007, Haldar, 2014a]. Furthermore, it seems that no special assumptions on the type of sampling pattern seem to be required for partial discreteness.

Experiments performed on both simulated and real k-space data have shown that the newly proposed reconstruction method PD performs competitively with and often better than state-of-the-art reconstruction methods such as SparseMRI, LORAKS and CS+NLTV. The results suggest that PD allows better texture preservation than SparseMRI (CS with TV prior), avoiding the staircasing effect, and even CS+NLTV. This is because, in the partial discreteness representation, which is the core of PD, edges are modeled not solely based on intensity or its gradient but merely based on Bayesian (*a posteriori*) probabilities of GMM classes. Enforcing sparsity on the residual form instead of the (non-local) gradient, as (non-local) TV promotes, is less restrictive. Furthermore, PD outperforms LORAKS in terms of computation time, while providing images as highly detailed as LORAKS does. Finally, we note that the probabilistic image presentation in PD inherently leads to tissue classification embedded in a Bayesian framework (provided K is chosen carefully and tissues can be assumed to be homogeneous). *A posteriori* probability maps

derived from the GMM may be used for tissue segmentation [Van Leemput et al., 1999b], texture analysis or tissue-selective filtering schemes [Vegas-Sánchez-Ferrero et al., 2010, Ramos-Llordén et al., 2015b].

Appendices

A

Appendix to Chapter 5

In section A.1 we provide details on the motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$. In subsection A.2, we prove Eq. (5.14) and Eq. (5.15), which gives a complete description of the MM framework used in P.2. In section A.3, we show graphs of estimated motion parameters for the in vivo experiments and one of the simulation experiments described in subsection 5.4.1.2. Finally, section A.4 discusses an extension of the joint MLE to account for intra-image motion.

A.1 Details on motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$

In this section, an explicit expression of the motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ is derived. Furthermore, we sketch the proof for its unitarity property, i.e., $\boldsymbol{H}_{\boldsymbol{\theta}_n}^H \boldsymbol{H}_{\boldsymbol{\theta}_n} = \boldsymbol{H}_{\boldsymbol{\theta}_n} \boldsymbol{H}_{\boldsymbol{\theta}_n}^H = \boldsymbol{I}$, with $\boldsymbol{I}$ the identity matrix. Specific details can be found in [Condat et al., 2008], [Jain, 1989] and especially in [Larkin et al., 1997].

Let $\boldsymbol{r}_m^n$ be a spatial point related to the reference-system point $\boldsymbol{r}_m$ through a rigid transformation matrix $\boldsymbol{M}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{4\times 4}$:

$$\begin{pmatrix} \boldsymbol{r}_m^n \\ 1 \end{pmatrix} = \boldsymbol{M}_{\boldsymbol{\theta}_n} \begin{pmatrix} \boldsymbol{r}_m \\ 1 \end{pmatrix}. \tag{A.1}$$

The rigid transformation matrix $\boldsymbol{M}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{4\times 4}$, which includes 3D rotation and translation, can then be written as [Poot, 2010, Goldstein et al., 2014]:

$$\boldsymbol{M}_{\boldsymbol{\theta}_n} = \begin{pmatrix} \boldsymbol{R}(\alpha_n, \beta_n, \gamma_n) & \boldsymbol{t}_n \\ \boldsymbol{0}^T & 1 \end{pmatrix}, \tag{A.2}$$

with $\boldsymbol{t}_n = (t_{x_n}, t_{y_n}, t_{z_n})^T$ a vector of translation parameters, $\mathbf{0}^T$ a 1×3 zero vector, and $\boldsymbol{R}(\alpha_n, \beta_n, \gamma_n) \in \mathbb{R}^{3\times3}$ the product of three elementary rotation matrices $\big(\boldsymbol{R}_x(\alpha_n)$, $\boldsymbol{R}_y(\beta_n)$ and $\boldsymbol{R}_z(\gamma_n)\big)$ describing rotations around the x, y and z axis, with angles α_n, β_n and γ_n, respectively. With such parametrization, we get

$$\boldsymbol{r}_m^n = \boldsymbol{R}_x(\alpha_n)\boldsymbol{R}_y(\beta_n)\boldsymbol{R}_z(\gamma_n)\boldsymbol{r}_m + \boldsymbol{t}_n. \tag{A.3}$$

Let $f(\cdot)$ be spatially-continuous function (a relaxation model in our problem). Then, to calculate $f(\boldsymbol{r}_m^n)$, the following spatial transformations on $f(\cdot)$ are consecutively applied:

1. 3D translation: $f_{\text{T}}(\boldsymbol{r}) = f(\boldsymbol{r} + \boldsymbol{t}_n)$
2. Rotation around x axis $f_{\text{Rot-x}}(\boldsymbol{r}) = f_{\text{T}}(\boldsymbol{R}_x\boldsymbol{r})$
3. Rotation around y axis $f_{\text{Rot-y}}(\boldsymbol{r}) = f_{\text{Rot-x}}(\boldsymbol{R}_y\boldsymbol{r})$
4. Rotation around z axis $f_{\text{Rot-z}}(\boldsymbol{r}) = f_{\text{Rot-y}}(\boldsymbol{R}_z\boldsymbol{r})$

Indeed, by evaluating $f_{\text{Rot-z}}(\cdot)$ at $\boldsymbol{r}_m$, we get $f(\boldsymbol{r}_m^n)$.

In a discrete domain, each of the previous four operations is represented by linear operators, hence matrices, that we denote as $\boldsymbol{H}_{\text{T}}$, $\boldsymbol{H}_{\text{Rot-x}}$, $\boldsymbol{H}_{\text{Rot-y}}$, and $\boldsymbol{H}_{\text{Rot-z}}$, respectively. Note that we have omitted the dependence on the motion parameter for the sake of notational convenience. As a consequence, the motion operator, $\boldsymbol{H}_{\boldsymbol{\theta}_n}$, can be written as

$$\boldsymbol{H}_{\boldsymbol{\theta}_n} = \boldsymbol{H}_{\text{Rot-z}}\boldsymbol{H}_{\text{Rot-y}}\boldsymbol{H}_{\text{Rot-x}}\boldsymbol{H}_{\text{T}}, \tag{A.4}$$

and its Hermitian transpose as

$$\boldsymbol{H}_{\boldsymbol{\theta}_n}^H = \boldsymbol{H}_{\text{T}}^H\boldsymbol{H}_{\text{Rot-x}}^H\boldsymbol{H}_{\text{Rot-y}}^H\boldsymbol{H}_{\text{Rot-z}}^H. \tag{A.5}$$

It is clear that if $\boldsymbol{H}_{\boldsymbol{T}}$, $\boldsymbol{H}_{\text{Rot-x}}$, $\boldsymbol{H}_{\text{Rot-y}}$ and $\boldsymbol{H}_{\text{Rot-z}}$ are unitary, $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ is unitary as well.

A.1.1 Sketch of Proof 1: $\boldsymbol{H}_{\text{T}}$ is unitary

The translation operator $\boldsymbol{H}_{\text{T}}$ consists of 1) a 3D FFT, 2) a voxel-wise multiplication with a purely complex exponential whose phase depends linearly on the translation parameters, and 3) an inverse 3D FFT [Larkin et al., 1997]. By noting that the multidimensional FFT is a unitary operator [Jain, 1989], the translation operator can be succinctly written as

$$\boldsymbol{H}_{\text{T}} = \boldsymbol{F}_{\text{3D}}^H\boldsymbol{\Delta}\boldsymbol{F}_{\text{3D}}, \tag{A.6}$$

where $\boldsymbol{F}_{\text{3D}}$ is the 3D unitary Discrete Fourier Transform (DFT) matrix and $\boldsymbol{\Delta}$ is a diagonal matrix whose entries are purely complex exponentials. It is known that $\boldsymbol{\Delta}$ is a unitary matrix if and only if the modulus of each diagonal entries is one. Since this is always true for purely complex exponentials, it demonstrates that $\boldsymbol{H}_{\text{T}}$ is unitary.

A.1.2 Sketch of Proof 2: $\boldsymbol{H}_{\text{Rot-x}}$, $\boldsymbol{H}_{\text{Rot-y}}$, and $\boldsymbol{H}_{\text{Rot-z}}$ are unitary.

For brevity, we present the proof only for $\boldsymbol{H}_{\text{Rot-x}}$. The proof for $\boldsymbol{H}_{\text{Rot-y}}$ and $\boldsymbol{H}_{\text{Rot-z}}$ is completely similar. Because $\boldsymbol{R}_x(\alpha_n)$ can be decomposed as the product of three one-dimensional shear matrices [Larkin et al., 1997], it is possible to write

$$\boldsymbol{H}_{\text{Rot-x}} = \boldsymbol{S}_{\boldsymbol{x}}\boldsymbol{S}_{\boldsymbol{y}}\boldsymbol{S}_{\boldsymbol{x}}, \tag{A.7}$$

where $\boldsymbol{S}_{\boldsymbol{x}}$ and $\boldsymbol{S}_{\boldsymbol{y}}$ are fractional delay filters [Condat et al., 2008], which model the shearings in the x and y dimension, respectively. Note that these filters can be implemented efficiently with FFT [Larkin et al., 1997]. If both $\boldsymbol{S}_{\boldsymbol{x}}$ and $\boldsymbol{S}_{\boldsymbol{y}}$ are unitary, $\boldsymbol{H}_{\text{Rot-x}}$ is unitary as well. Indeed, $\boldsymbol{S}_{\boldsymbol{x}}$ has essentially the same diagonal expression as Eq. (A.6), where the role of the 3D DFT matrices is fulfilled by a (unitary) Fourier matrix which applies an FFT only along the x direction. The phase of the complex exponential in the diagonal matrix now depends linearly on the shearing parameter [Larkin et al., 1997], which is a real value. Therefore, the associated diagonal matrix is unitary. The unitarity property of $\boldsymbol{S}_{\boldsymbol{x}}$ follows immediately. The proof for $\boldsymbol{S}_{\boldsymbol{y}}$ is equivalent, with the exception that the unitary Fourier matrix now represents an FFT along the y direction. We can prove then that $\boldsymbol{S}_{\boldsymbol{y}}$ is unitary and thus $\boldsymbol{H}_{\text{Rot-x}}$ is unitary. As already mentioned, the proof for $\boldsymbol{H}_{\text{Rot-y}}$ and $\boldsymbol{H}_{\text{Rot-z}}$ are analogous. Combining Proof 1 and Proof 2, the unitary property of $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ is demonstrated.

A.2 Separable Quadratic Surrogate (SQS) function derivation for the joint MLE

In order to get the final version of the joint MLE algorithm, a necessary step was to obtain a surrogate function for

$$G_n(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = ||\boldsymbol{W}_n^{1/2}\big(\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) - \breve{\boldsymbol{s}}_n^k\big)||_2^2 + C_n(k) \tag{A.8}$$

with $\boldsymbol{W}_n = \text{diag}\{\frac{1}{2\boldsymbol{\sigma}_n^2}\}$.

The choice we made in this work was a SQS function [Muckley et al., 2015], that when applied to Eq. (A.8), yields Eq. (5.14) and Eq. (5.15). Here, we present the proof of these equations. To that end, we build on results presented in [Muckley et al., 2015]. In that work, a SQS function was applied to a generic quadratic form $\frac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2$. Such SQS function had the following expression:

$$\frac{1}{2}||\boldsymbol{x} - \big(\boldsymbol{x}^k - \boldsymbol{D}_{\boldsymbol{f}}^{-\mathbf{1}}\boldsymbol{A}^H\big(\boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{y}\big)\big)||_{\boldsymbol{D}_{\boldsymbol{f}}}^2 + \xi, \tag{A.9}$$

with ξ a constant independent of $\boldsymbol{x}$ and where the matrix $\boldsymbol{D}_{\boldsymbol{f}}$ is defined in such way that it satisfies $\boldsymbol{D}_{\boldsymbol{f}} \succeq \boldsymbol{A}^H\boldsymbol{A}$, that is, $\boldsymbol{D}_{\boldsymbol{f}} - \boldsymbol{A}^H\boldsymbol{A}$ is a positive-semidefinite matrix. We can easily identify the terms of the quadratic form at hand, i.e.,

$G_n(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$, with the terms of $\frac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2$, and hence easily define our SQS function, as

$$G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = ||\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) - \boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)||_{\boldsymbol{D_f}}^2 + C_n^*(k), \tag{A.10}$$

with

$$\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = \boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) - \boldsymbol{D_f^{-1}}\boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) - \boldsymbol{y}), \tag{A.11}$$

$C_n^*(k)$ a constant independent of $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$, and where $\boldsymbol{A} = \boldsymbol{W}_n^{1/2}\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}$ and $\boldsymbol{y} = \boldsymbol{W}_n^{1/2}\breve{\boldsymbol{s}}_n^k$. After some algebra, we obtain

$$\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = \boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) + \boldsymbol{D_f^{-1}}\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n\big(\breve{\boldsymbol{s}}_n^k - \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)\big). \tag{A.12}$$

Before giving an expression for $\boldsymbol{D_f}$ satisfying $\boldsymbol{D_f} \succeq \boldsymbol{A}^H\boldsymbol{A}$, first we recognize that

$$\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n^{1/2^H}\boldsymbol{W}_n^{1/2}\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} = \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}. \tag{A.13}$$

Furthermore, it is easy to show that the diagonal matrix $\boldsymbol{W}_n$ fulfills $\boldsymbol{W}_n \preceq (\sigma_n^*)^{-1}\boldsymbol{I}$ with $(\sigma_n^*)^{-1}$ being the maximum value along its diagonal, which is,

$$(\sigma_n^*)^{-1} \triangleq \frac{1}{2\min\limits_m [\boldsymbol{\sigma}_n]_m^2}. \tag{A.14}$$

Thus, if $\boldsymbol{W}_n \preceq (\sigma_n^*)^{-1}\boldsymbol{I}$, it follows that

$$\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} \preceq (\sigma_n^*)^{-1}\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} = (\sigma_n^*)^{-1}\boldsymbol{I}, \tag{A.15}$$

since the motion operator is unitary. Therefore, by defining $\boldsymbol{D_f}$ as $\boldsymbol{D_f} \triangleq (\sigma_n^*)^{-1}\boldsymbol{I}$, $\boldsymbol{D_f} \succeq \boldsymbol{A}^H\boldsymbol{A}$ holds. Note that previous $\boldsymbol{D_f}$ definition depends on n and hence the NLLS problem (Eq. (A.10)) is weighted differently along dimension n. To provide an unweighted NLLS problem, that is, the version we have presented in 5.3.2, we set $\sigma^* = 2\min\limits_{n,m} [\boldsymbol{\sigma}_n]_m^2$. Clearly $\boldsymbol{W}_n \preceq (\sigma^*)^{-1}\boldsymbol{I}$ for all n. Hence, we redefine $\boldsymbol{D_f}$ as $\boldsymbol{D_f} \triangleq (\sigma^*)^{-1}\boldsymbol{I}$, and trivially we get $\boldsymbol{D_f} \succeq \boldsymbol{A}^H\boldsymbol{A}$ as desired. By substituting $\boldsymbol{D_f}$ into Eq. (A.12), we arrive at the final expression which is shown in Eq. (5.15).

A.3 Graphs of motion

To illustrate the quality of the motion estimation, in this section, we show the graphs, as a function of n, of the ground-truth and estimated motion parameters (for the three methods). In Fig. A.1, we show those graphs for one of the rotational motion (R-m) simulations of subsection 5.4.1.2.

Fig. A.2 shows the graphs for the ground-truth based real experiments whereas Fig. A.3 and Fig. A.4 present the curves of the motion estimates for the in vivo axial and the in vivo sagittal human brain data experiments, respectively.

A.4 Extension of the joint MLE to account for intra-image motion

As mentioned in the discussion section (section 5.6), the joint MLE can be extended to include intra-image motion, in particular, motion between the acquisition of the different slices of a multi-slice image. A brief outlook to such an extension is given here. The implementation of the extended algorithm should take into account the following considerations. First, given an inversion time TI_n, the z-th noiseless and motion-corrupted 2D slice T_1-weighted $\tilde{\boldsymbol{f}}_{n,z}(\boldsymbol{\theta}_{n,z}, \boldsymbol{\kappa}, \boldsymbol{T_1})$ is related to the unobserved 3D image, $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$, through the motion parameters $\boldsymbol{\theta}_{n,z}$. Note that the number of motion parameters scales with $M_z \times N$, where M_z is the number of slices. Second, the mapping between a 2D slice T_1-weighted image and the noiseless unobserved 3D image also requires a slice-selective profile filter, which can be included as a matrix $\boldsymbol{\delta}_z$ [Gholipour et al., 2010], just after the motion operator, that is, $\tilde{\boldsymbol{f}}_{n,z}(\boldsymbol{\theta}_{n,z}, \boldsymbol{\kappa}, \boldsymbol{T_1}) = \boldsymbol{\delta}_z \boldsymbol{H}_{\boldsymbol{\theta}_{n,z}} \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$. The final details of the derivation of the MM algorithm are beyond the scope of this work and therefore not presented here.

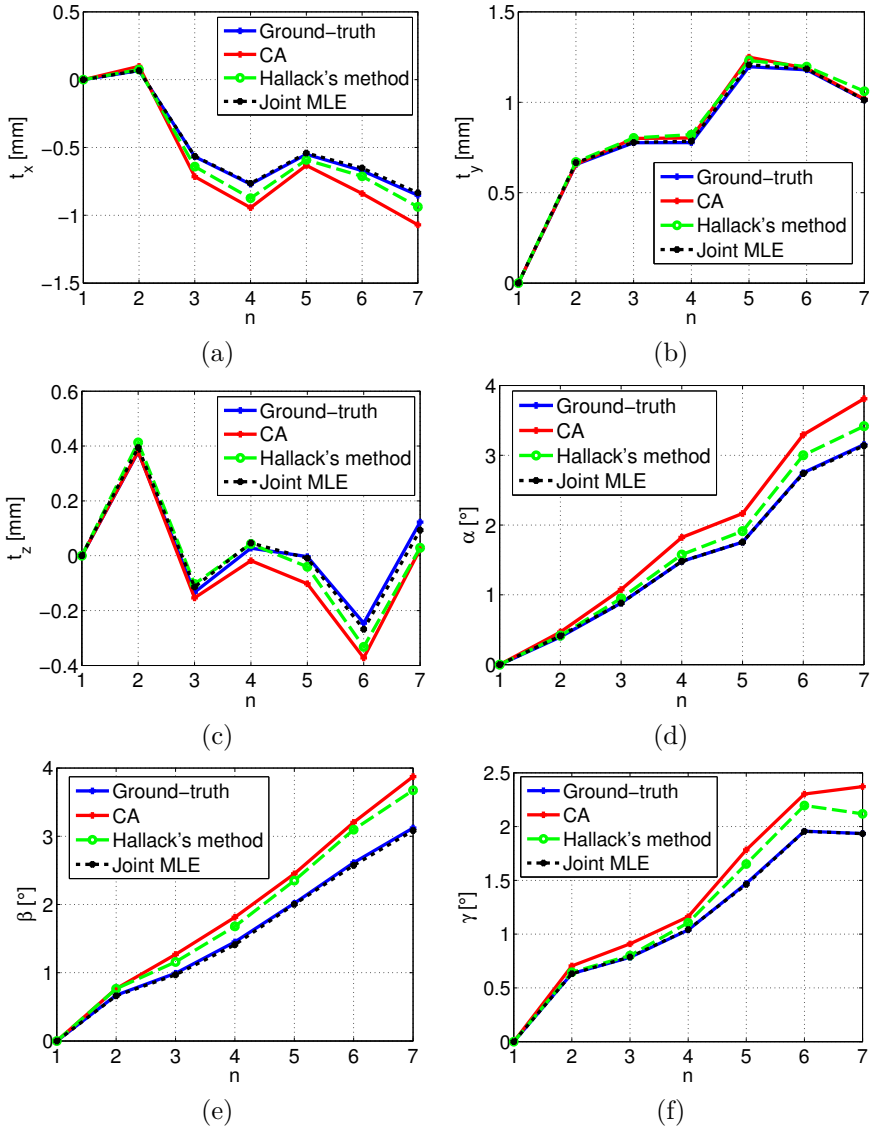


Figure A.1: Graphs of the ground-truth and estimated motion parameters for one realization of the simulation experiment with rotational motion: (a) t_x, (b) t_y, (c) t_z, (d) α, (e) β, (f) γ.

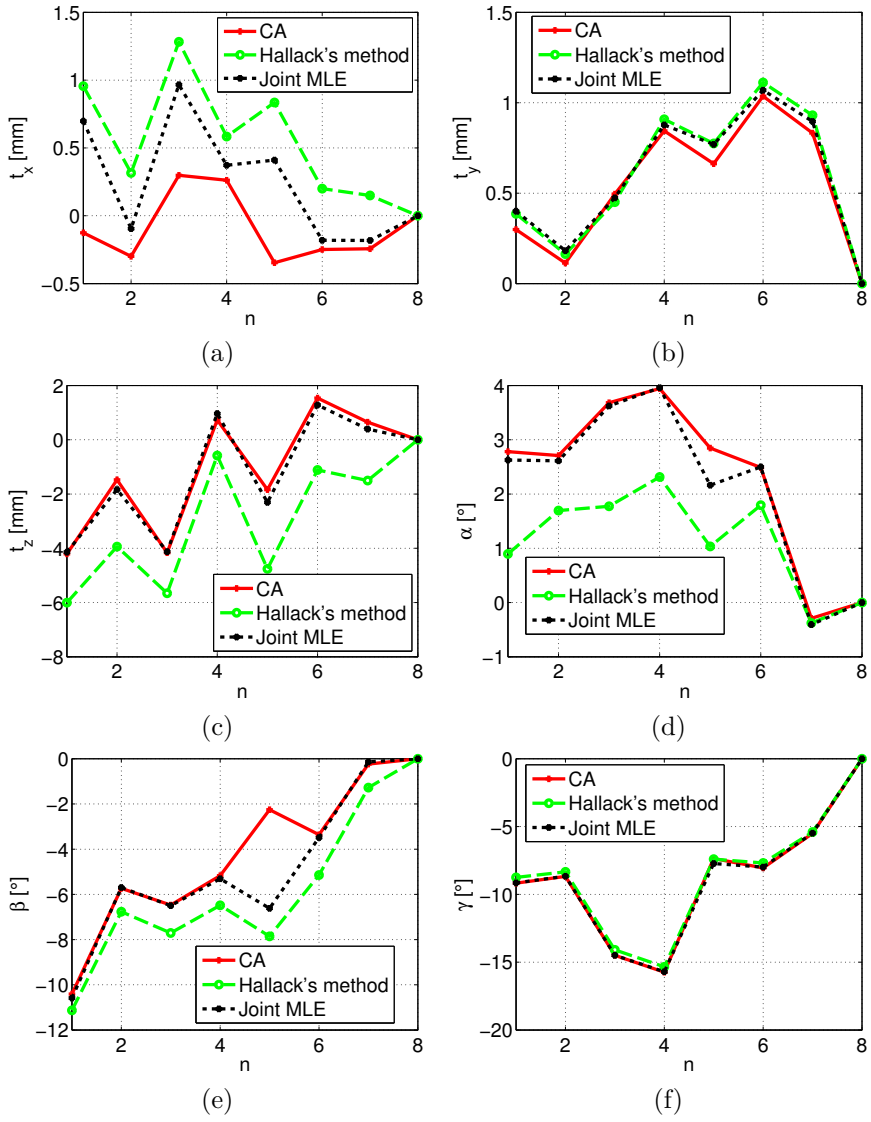


Figure A.2: Graphs of the estimated motion parameters for the ground-truth watermelon experiment: (a) t_x, (b) t_y, (c) t_z, (d) α, (e) β, (f) γ.

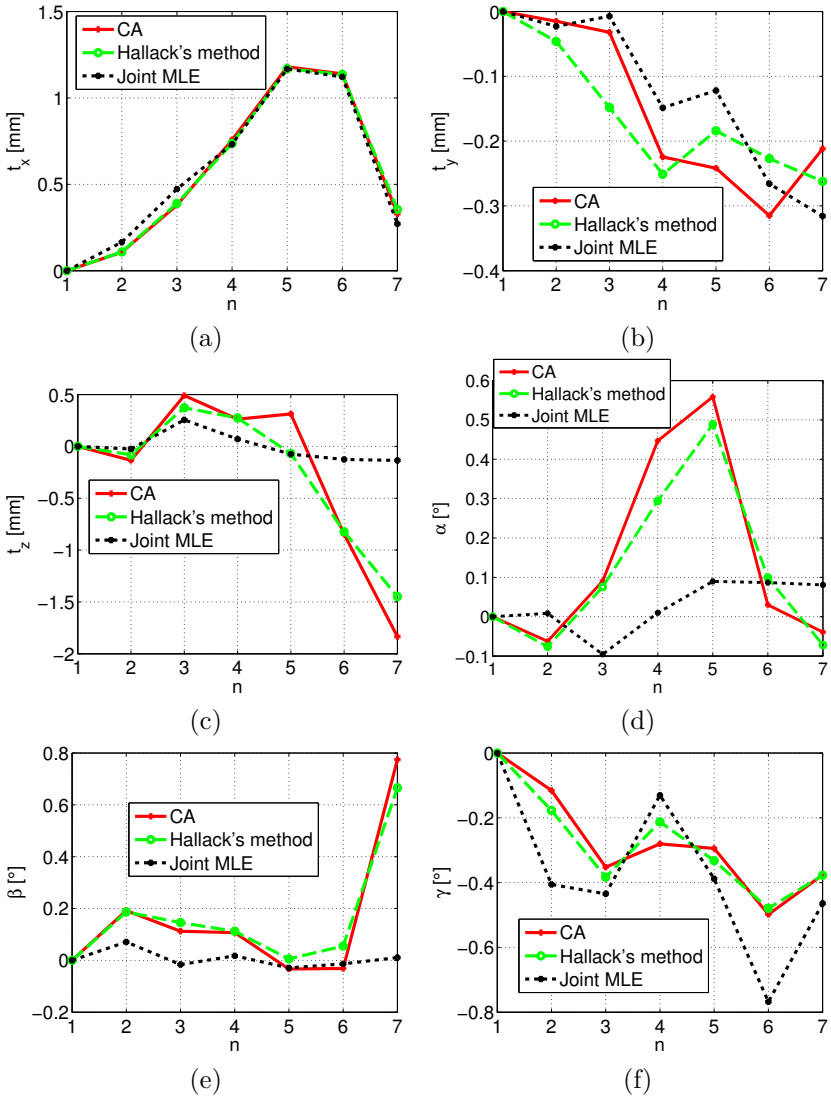


Figure A.3: Graphs of the estimated motion parameters for the in vivo axial human brain data experiment: (a) t_x, (b) t_y, (c) t_z, (d) α, (e) β, (f) γ.

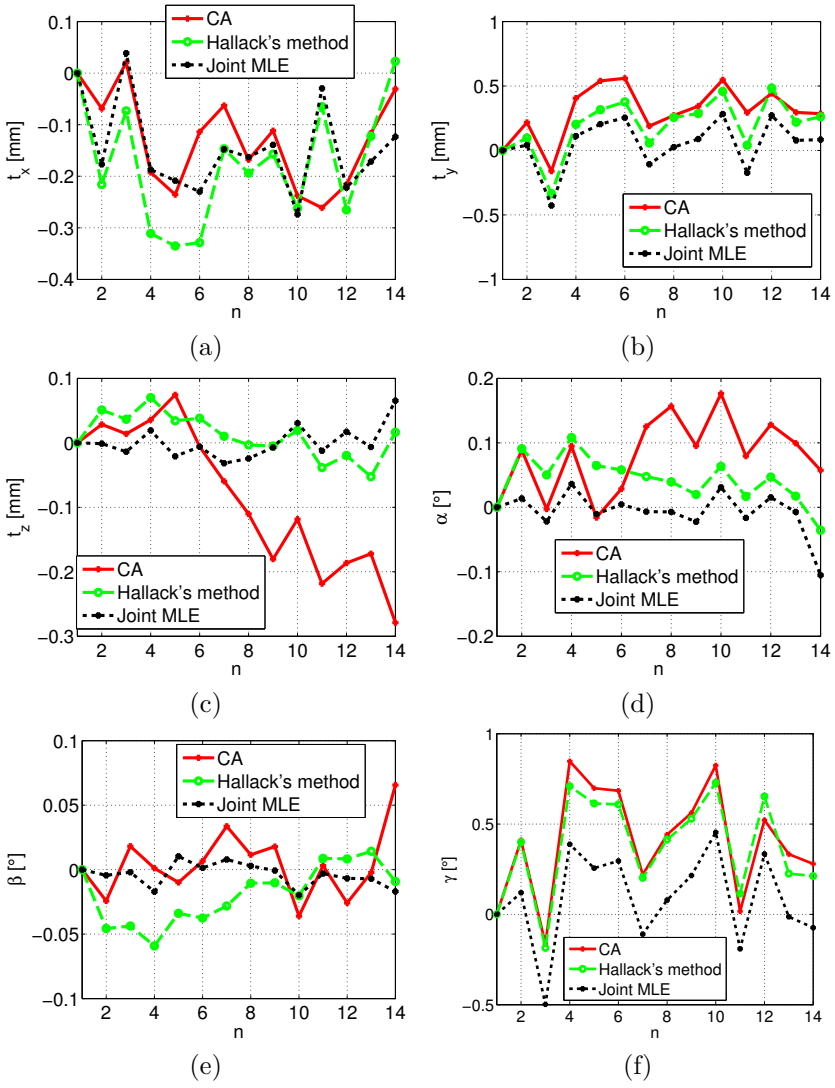


Figure A.4: Graphs of the estimated motion parameters for in vivo sagittal human brain data experiment. (a) t_x, (b) t_y, (c) t_z, (d) α, (e) β, (f) γ.

B

Appendix to Chapter 6

B.1 Theoretical study of NOVIFAST's global convergence

B.1.1 Independence of c_1^{k+1} and c_2^{k+1} with respect to c_1^k

In this section we demonstrate that the iterative procedure of Eq. (6.20) is independent of c_1^k, which is assumed to be nonzero. We first observe that in Eq. (6.20) the dependency of c_1^{k+1} on c_1^k is only via $\tilde{\boldsymbol{a}}$, through $s_n(\boldsymbol{c}^k)$ (see Eq. (6.15)). Moreover, since c_1^k enters linearly in $\tilde{\boldsymbol{a}}$, $\tilde{\boldsymbol{a}}/c_1^k$ is independent of c_1^k. With the previous considerations, due to the linear property of inner products we can rewrite Eq. (6.20) as

$$c_1^{k+1} = \frac{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{z} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ c_1^k \langle \boldsymbol{z}, \frac{\tilde{\boldsymbol{a}}}{c_1^k} \rangle & c_1^k \langle \boldsymbol{a}, \frac{\tilde{\boldsymbol{a}}}{c_1^k} \rangle \end{vmatrix}}{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ c_1^k \langle \boldsymbol{b}, \frac{\tilde{\boldsymbol{a}}}{c_1^k} \rangle & c_1^k \langle \boldsymbol{a}, \frac{\tilde{\boldsymbol{a}}}{c_1^k} \rangle \end{vmatrix}}, c_2^{k+1} = \frac{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b} \rangle & \langle \boldsymbol{b}, \boldsymbol{z} \rangle \\ c_1^k \langle \boldsymbol{b}, \frac{\tilde{\boldsymbol{a}}}{c_1^k} \rangle & c_1^k \langle \boldsymbol{z}, \frac{\tilde{\boldsymbol{a}}}{c_1^k} \rangle \end{vmatrix}}{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ c_1^k \langle \boldsymbol{b}, \frac{\tilde{\boldsymbol{a}}}{c_1^k} \rangle & c_1^k \langle \boldsymbol{a}, \frac{\tilde{\boldsymbol{a}}}{c_1^k} \rangle \end{vmatrix}}. \tag{B.1}$$

Let us now take into account the following lemma regarding the properties of determinants [Zhang, 1999]:

Lemma 1 *If we multiply a column or a row of a matrix by a number, the determinant of that matrix will be multiplied by the same number.*

If we apply Lemma 1, we can factor out c_1^k from the second row of the determinants, with the result that c_1^k in the numerator cancels with the same factor in the denominator. Since $\hat{\boldsymbol{a}}/c_1^k$ does not depend on c_1^k, both c_1^{k+1} and c_2^{k+1} are independent of c_1^k.

B.1.2 NOVIFAST as a fixed-point algorithm

NOVIFAST can be studied under the umbrella of fixed-point algorithms theory [Kirk and Brailey Sims, 2001, Burden and Faires, 2010]. This framework allows us to study the converge properties of NOVIFAST in a systematic way. Let us start with some observations. As demonstrated in subsection B.1.1, the iterative procedure of Eq. (6.20), which is part of NOVIFAST, does not depend on c_1^k. In other words, we have that $c_1^{k+1} = g_1(c_2^k)$ and $c_2^{k+1} = g_2(c_2^k)$, where the one-dimensional functions $g_1(\cdot)$ and $g_2(\cdot)$ are, respectively, the first and second component of the vector-valued function $\boldsymbol{g}(\cdot)$, defined as $\boldsymbol{g}(\boldsymbol{c}) = \boldsymbol{A}^{-1}(\boldsymbol{c})\boldsymbol{v}(\boldsymbol{c})$. An iterative procedure of the form of $c_2^{k+1} = g_2(c_2^k)$ is called a *fixed-point iteration* [Burden and Faires, 2010, Kirk and Brailey Sims, 2001]. Similarly, a point c_2^* such that $c_2^* = g_2(c_2^*)$ is called a *fixed-point* of $g_2(\cdot)$. Accordingly, a point $\boldsymbol{c}^* = (c_1^*, c_2^*)^T$ with $c_1^* = g_1(c_2^*)$ is a solution of Eq. (6.18) if and only if c_2^* is a fixed-point of $g_2(\cdot)$. We also have that if $\lim_{k\to\infty} |c_2^{k+1} - c_2^k| = 0$, then, provided $g_1(\cdot)$ has a bounded derivative, $\lim_{k\to\infty} |c_1^{k+1} - c_1^k| = 0$, and trivially $\lim_{k\to\infty} ||\boldsymbol{c}^{k+1} - \boldsymbol{c}^k||_2 = 0$. That is, NOVIFAST does converge if $\lim_{k\to\infty} |c_2^{k+1} - c_2^k| = 0$. Therefore, to study NOVIFAST's convergence we should focus on analyzing the fixed-point iteration $c_2^{k+1} = g_2(c_2^k)$.

To do so, let us assume that $|dg_2(c_2)/dc_2| \leq L$ with $L < 1$ for all $c_2 \in C$, where $dg_2(c_2)/dc_2$ denotes the derivative of $g_2(\cdot)$ with respect to c_2, and C is the closed interval $[\exp(-\mathrm{TR}/T_{1\mathrm{min}}), \exp(-\mathrm{TR}/T_{1\mathrm{max}})]$ for appropriate $T_{1\mathrm{min}}$ and $T_{1\mathrm{max}}$. Throughout the rest of the document, we dubbed the condition $|dg_2(c_2)/dc_2| \leq L$ with $L < 1$ for all $c_2 \in C$ as *the bound condition*. If the bound condition holds, it can be demonstrated by means of the mean value theorem that [Burden and Faires, 2010]

$$|g_2(c_2) - g_2(c_2')| \leq L|c_2 - c_2'| \quad \text{for every } c_2 \text{ and } c_2' \in C. \tag{B.2}$$

Regarding NOVIFAST's convergence, Eq. (B.2) allows us to show that [Burden and Faires, 2010]

$$|c_2^{k+1} - c_2^k| \leq L|c_2^k - c_2^{k-1}|, \quad k = 1, 2, ... \tag{B.3}$$

provided all iterates $c_2^k \in C$ for $k = 1, 2, ...$ In words, the absolute difference between consecutive iterates always decreases, and the amount of that decrease is governed by L. Clearly, this implies that if the bound condition holds, then $\lim_{k\to\infty} |c_2^{k+1} - c_2^k| = 0$. We can also demonstrate the existence of a unique c_2^*, and hence a single $\boldsymbol{c}^*$. Indeed, suppose there would exist two fixed-points: c_2^* and $c_2^{*\prime}$. Then, $|c_2^* - c_2^{*\prime}| = |g(c_2^*) - g(c_2^{*\prime})| \leq L|c_2^* - c_2^{*\prime}|$. Hence, $(1-L)|c_2^* - c_2^{*\prime}| \leq 0$. Since $L < 1$, the last inequality only holds if $|c_2^* - c_2^{*\prime}| = 0$, which implies $c_2^* = c_2^{*\prime}$.

A remarkable observation is that, for the problem at hand, the bound condition, with $T_{1\mathrm{min}} = 150$ ms and $T_{1\mathrm{max}} = 4500$ ms, holds with overwhelming probability

for realistic clinically achievable signal-to-noise ratios (SNRs), for typical ground-truth T_1 values of white and gray matter, and for common flip angles and TR values in the literature. This first shows that the existence of a unique local maximum is guaranteed with very high probability (minimally 95% for realistic values of the SNR, see subsection B.2.2 and, in particular, Fig. B.1). It also substantiates our claims about the good convergence properties of NOVIFAST, since $\lim_{k\to\infty} ||\boldsymbol{c}^{k+1} - \boldsymbol{c}^k||_2 = 0$ is guaranteed with high probability as well. Furthermore, the value of L consistently decreases with increasing SNR, which implies that the velocity of NOVIFAST's convergence increases. This fact is related to the important observation that NOVIFAST becomes an exact and analytical algorithm when applied to noiseless data, as demonstrated in subsection 6.3.3. Before ending this theoretical study, we would like to emphasize that the bound condition is just a sufficient condition of convergence for NOVIFAST, but by no means a necessary condition. Indeed, by checking $\lim_{k\to\infty} ||\boldsymbol{c}^{k+1} - \boldsymbol{c}^k||_2 = 0$ manually, for every initialization $T_{1\text{ini}} \in [T_{1\text{min}}, T_{1\text{max}}]$, we obtained an overwhelming high probability of convergence. The empirical analysis is presented in next section.

B.2 Empirical validation of NOVIFAST's theoretical properties

In this section we present the empirical simulation study that was conducted to check the invertibility of the matrix $\boldsymbol{A}(\boldsymbol{c})$, appearing in Eq. (6.18), and the global convergence properties of NOVIFAST.

As in the experimental part of section 7.4, y_n with $n = 1, ..., N$, are realizations of independent random variables that follow a Rician distribution, where the signal parameter is given by $s_n(K_{\text{GT}}; T_{1\text{GT}})$, and the noise standard deviation is $\sigma = K_{\text{GT}}/\text{SNR}_{90^\circ}$. Likewise, TR = 5 ms and the chosen FA sets were the $\mathcal{A}_{N=3}$ and $\mathcal{A}_{N=10}$ set. We tried to cover a much wider range of SNR_{90° values than in the experiments of section 7.4. To that end, SNR_{90° was varied among the following list of values: $\{80, 100, 150, 250, 300, 500, 1000, 2000, 3000, 10000\}$. An observation relating to the fact that the data is Rice distributed, which we will exploit in the experimental analysis, is the following: y_n can also be generated as $y_n = K_{\text{GT}} y'_n$, with y'_n being realizations of independent random variables that follow a Rician distribution, where the signal parameter is given by $s_n(1; T_{1\text{GT}})$, and the noise standard deviation is $\text{SNR}^{-1}_{90^\circ}$. The proof is immediate by virtue of the fundamental theorem of transformation of random variables [Mood et al., 1963, Papoulis and Pillai, 2002].

B.2.1 Invertibility of matrix $\boldsymbol{A}(\mathbf{c})$

A necessary condition for NOVIFAST to be well defined is that the matrix $\boldsymbol{A}(\boldsymbol{c})$ is non-singular, or equivalently, that its determinant $\det(\boldsymbol{A}(\boldsymbol{c}))$, is non-zero (see Eq. (6.19) and Eq. (6.20)). In the first section of this document, it was shown that since c_1^k enters linearly in $s_n(\boldsymbol{c}^k)$, it can be factored out of the determinant of

the denominator of Eq. (6.20). Analogously, since $y_n = K_{\mathrm{GT}} y'_n$, K_{GT} can also be factored out.

Since $\det(\boldsymbol{A}(\boldsymbol{c}))$ can then be written as $K_{\mathrm{GT}} c_1 \det(\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c}))$, where $\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c})$ does not depend on K_{GT} and c_1, $\det(\boldsymbol{A}(\boldsymbol{c})) = 0$ if and only if $\det(\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c})) = 0$. The statistics of the random variable $\det(\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c}))$ (being a function of y'_n) depend on $T_{1\mathrm{GT}}$, c_2, $\{\alpha_n\}_{n=1}^N$, TR, and SNR_{90°. In a simulation experiment, we empirically estimated the probability of the event $\{|\det(\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c}))| \leq \epsilon\}$, with $|\cdot|$ the absolute value operator and ϵ a fixed positive constant ($\epsilon = 1$ in our experiments), for a dense grid of $T_{1\mathrm{GT}}$, c_2, and for the $\mathcal{A}_{N=3}$ and $\mathcal{A}_{N=10}$ FA set. We generated $N_{\mathrm{MC}} = 10^5$ realizations of y'_n with $T_{1\mathrm{GT}} \in [500, 2500]$ ms. We could not find any single case where $\#\{|\det(\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c}))| \leq \epsilon\}/N_{\mathrm{MC}}$ was non-zero, where $\#\{|\det(\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c}))| \leq \epsilon\}$ denotes the number of successful outcomes of the event $\{|\det(\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c}))| \leq \epsilon\}$. Since the case $\det(\boldsymbol{A}_{\mathrm{Norm}}(\boldsymbol{c})) = 0$ is contained in that event, we then concluded that the probability of $\det(\boldsymbol{A}(\boldsymbol{c})) = 0$ was zero for the experiment conducted.

B.2.2 Global convergence properties

In subsection B.1.2, it was mentioned that the bound condition holds with high probability for realistic clinically achievable SNRs. Moreover, convergence of NOVIFAST, in the sense that $\lim_{k\to\infty} ||\boldsymbol{c}^{k+1} - \boldsymbol{c}^k||_2 = 0$, was guaranteed with overwhelmingly high probabilities. We devote this section to presenting the MC analysis and its main conclusions. The experimental set up was similar to the experiment that was conducted to check the invertibility. The function $g_2(\cdot)$, apart from being independent of c_1, is also independent of K_{GT} (K_{GT} can also be factored out from the numerator, and hence it cancels with K_{GT} in the denominator). Therefore, the statistics of the random variable $dg_2(c_2)/dc_2$ depend on same parameters as $\det_{\mathrm{Norm}}(\boldsymbol{A})$. Hence, for every $T_{1\mathrm{GT}}$ and SNR_{90°, and fixed $\{\alpha_n\}_{n=1}^N$ and TR (same values as the invertibility-based experiment), we calculated the ratio $\#\{|dg_2/dc_2(c_2)| \leq L \text{ with } L < 1 \quad \text{for all } c_2 \in C\}/N_{\mathrm{MC}}$, being an estimate of the probability that the bound condition holds. This probability is shown in Fig. B.1. Clearly, for $\mathrm{SNR}_{90^\circ} \geq 250$, a perfectly achievable value in clinical settings, the bound probability was close to 100% for all tested values of $T_{1\mathrm{GT}}$. This is true for the two FA sets used in the experiment.

To study the probability of NOVIFAST convergence in a more general sense, we were interested in checking if $\lim_{k\to\infty} ||\boldsymbol{c}^{k+1} - \boldsymbol{c}^k||_2 = 0$ for every initialization $T_{1\mathrm{ini}} \in [T_{1\mathrm{min}}, T_{1\mathrm{max}}]$. To do so, we adopted the following rule: NOVIFAST converges if 1) the relative norm between two consecutive iterations is below 10^{-6} and 2) the iteration number is smaller than 1000. If any of these two conditions was not met, we concluded that NOVIFAST failed to converge. We then counted the number of successful outcomes, and divided this number by N_{MC}. The resulting probability estimate is shown in Fig. B.2 and Fig. B.3 for the $\mathcal{A}_{N=3}$ and $\mathcal{A}_{N=10}$ FA set, respectively. As we anticipated, this probability is higher than the *bound condition* probability, since the latter is just a sufficient condition. Observe that for $\mathrm{SNR}_{90^\circ} \geq 250$, the probability of convergence was 100% for all values of $T_{1\mathrm{GT}}$, independently of the preselected initialization $T_{1\mathrm{ini}}$. Even for extremely low SNR_{90°,

difficult to find in actual MR scanners, NOVIFAST converges with overwhelmingly high probability.

B.3 Mathematical proof of the exactness of NOVIFAST in the noiseless case

We divide the mathematical proof in two parts. First, we demonstrate the relations given in Eq. (6.21) and Eq. (6.22), that is,

$$\langle \boldsymbol{b}, \boldsymbol{z} \rangle = c_{1\mathrm{GT}} \langle \boldsymbol{b}, \boldsymbol{b} \rangle + c_{2\mathrm{GT}} \langle \boldsymbol{b}, \boldsymbol{a} \rangle \tag{B.4}$$

$$\langle \boldsymbol{z}, \tilde{\boldsymbol{a}} \rangle = c_{1\mathrm{GT}} \langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle + c_{2\mathrm{GT}} \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle. \tag{B.5}$$

Next, we make use of these relations to show that NOVIFAST is exact if we work with noiseless data and it provides ground-truth values with just one iteration.

B.3.1 Proof of Eq. (B.4) and Eq. (B.5)

Eq. (B.4) and Eq. (B.5) hold when we work with noiseless data, i.e., mathematically, when $y_n = s_n \triangleq s_n(\boldsymbol{c}_{\mathrm{GT}})$, with

$$s_n(\boldsymbol{c}_{\mathrm{GT}}) = \frac{c_{1\mathrm{GT}} \sin(\alpha_n)}{1 - c_{2\mathrm{GT}} \cos(\alpha_n)}. \tag{B.6}$$

We shall start with the demonstration of Eq. (B.4). By multiplying the numerator and denominator of b_n by $1 - c_{2\mathrm{GT}} \cos(\alpha_n)$, we obtain

$$\begin{aligned} \langle \boldsymbol{b}, \boldsymbol{b} \rangle = \sum_{n=1}^{N} \frac{\sin^2(\alpha_n)\,(1 - c_{2\mathrm{GT}} \cos(\alpha_n))}{(1 - c_2 \cos(\alpha_n))^2 (1 - c_{2\mathrm{GT}} \cos(\alpha_n))} = \\ \sum_{n=1}^{N} \frac{\sin^2(\alpha_n)}{(1 - c_2 \cos(\alpha_n))^2 (1 - c_{2\mathrm{GT}} \cos(\alpha_n))} - \\ c_{2\mathrm{GT}} \sum_{n=1}^{N} \frac{\sin^2(\alpha_n) \cos(\alpha_n)}{(1 - c_2 \cos(\alpha_n))^2 (1 - c_{2\mathrm{GT}} \cos(\alpha_n))}. \end{aligned} \tag{B.7}$$

Let us first focus on the first term on the right side of Eq. (B.7). If we multiply the numerator and denominator of this term by $c_{1\mathrm{GT}}$, we can easily identify the expression $b_n z_n$ inside the sum. Therefore, the first term on the right side of Eq. (B.7) is equal to $\langle \boldsymbol{b}, \boldsymbol{z} \rangle / c_{1\mathrm{GT}}$. If we repeat the procedure with the second term on the right side of Eq. (B.7), we can similarly identify the expression $a_n b_n$ inside the sum, hence we get $c_{2\mathrm{GT}} \langle \boldsymbol{b}, \boldsymbol{a} \rangle / c_{1\mathrm{GT}}$. By rearranging Eq. (B.7) we then arrive at Eq. (B.4).

Regarding the proof of Eq. (B.5) we have that

$$\langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle = \sum_{n=1}^{N} \frac{s_n(\boldsymbol{c}) \sin(\alpha_n) \cos(\alpha_n)(1 - c_{2\mathrm{GT}} \cos(\alpha_n))}{(1 - c_2 \cos(\alpha_n))^2 (1 - c_{2\mathrm{GT}} \cos(\alpha_n))} =$$
$$\sum_{n=1}^{N} \frac{s_n(\boldsymbol{c}) \sin(\alpha_n) \cos(\alpha_n)}{(1 - c_2 \cos(\alpha_n))^2 (1 - c_{2\mathrm{GT}} \cos(\alpha_n))} -$$
$$c_{2\mathrm{GT}} \sum_{n=1}^{N} \frac{s_n(\boldsymbol{c}) \sin(\alpha_n) \cos^2(\alpha_n)}{(1 - c_2 \cos(\alpha_n))^2 (1 - c_{2\mathrm{GT}} \cos(\alpha_n))}. \tag{B.8}$$

Let us again first focus on the first term on the right side. Let us multiply the numerator and denominator by $c_{1\mathrm{GT}}$. By doing so, we can easily identify the expression $\tilde{a}_n s_n(\boldsymbol{c}_{\mathrm{GT}})/(1 - c_2 \cos(\alpha_n))$ inside the sum. We also know that for noiseless data $y_n = s_n(\boldsymbol{c}_{\mathrm{GT}})$, and hence $\tilde{a}_n s_n(\boldsymbol{c}_{\mathrm{GT}})/(1 - c_2 \cos(\alpha_n)) = z_n \tilde{a}_n$. Consequently, the first term on the right side of Eq. (B.8) is identical to $\langle \boldsymbol{z}, \tilde{\boldsymbol{a}} \rangle / c_{1\mathrm{GT}}$. Regarding the second term, if we multiply both the numerator and denominator by $c_{1\mathrm{GT}}$, we identify the expression $\tilde{a}_n s_n(\boldsymbol{c}_{\mathrm{GT}}) \cos(\alpha_n)/(1 - c_2 \cos(\alpha_n))$, which in the noiseless case yields $a_n \tilde{a}_n$, and we finally get $c_{2\mathrm{GT}} \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle / c_{1\mathrm{GT}}$. Eq. (B.5) can then be proved after rearranging terms.

B.3.2 Proof of NOVIFAST's exactness

NOVIFAST's exactness means that for noiseless data, $c_1^{k+1} = c_{1\mathrm{GT}}$ and $c_2^{k+1} = c_{2\mathrm{GT}}$ for $k = 0, 1, ...$ In order to prove this surprising result, we will need, apart from Eq. (B.4), Eq. (B.5), and Lemma 1, a new Lemma regarding yet another property of the determinant of a matrix [Zhang, 1999]:

Lemma 2 *Adding a scalar multiple of one column to another column in a matrix does not change the value of the determinant of that matrix.*

Equipped with Lemma 2, let us now substitute Eq. (B.4) and Eq. (B.5) into Eq. (6.20). We will just show the derivation of $c_1^{k+1} = c_{1\mathrm{GT}}$, as it is identical for c_2^{k+1}:

$$c_1^{k+1} = \frac{\begin{vmatrix} c_{1\mathrm{GT}} \langle \boldsymbol{b}, \boldsymbol{b} \rangle + c_{2\mathrm{GT}} \langle \boldsymbol{b}, \boldsymbol{a} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ c_{1\mathrm{GT}} \langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle + c_{2\mathrm{GT}} \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle \end{vmatrix}}{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b} \rangle & \langle \boldsymbol{b}, \boldsymbol{a} \rangle \\ \langle \boldsymbol{b}, \tilde{\boldsymbol{a}} \rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}} \rangle \end{vmatrix}}. \tag{B.9}$$

By Lemma 2, Eq. (B.9) is equivalent to

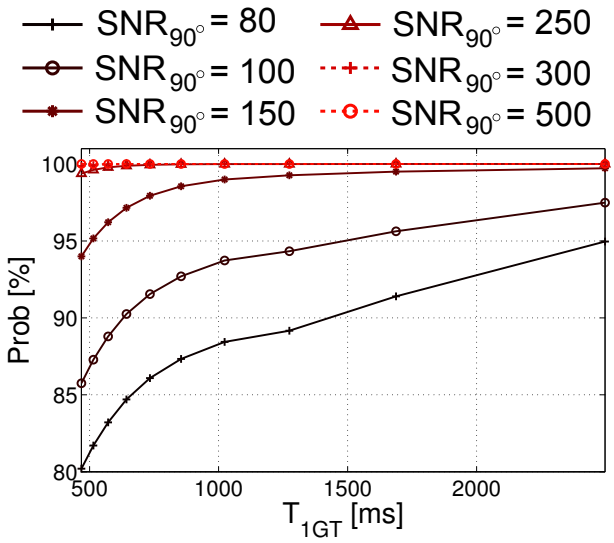
$$c_1^{k+1} = \frac{\begin{vmatrix} c_{1\mathrm{GT}}\langle \boldsymbol{b}, \boldsymbol{b}\rangle & \langle \boldsymbol{b}, \boldsymbol{a}\rangle \\ c_{1\mathrm{GT}}\langle \boldsymbol{b}, \tilde{\boldsymbol{a}}\rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}}\rangle \end{vmatrix}}{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b}\rangle & \langle \boldsymbol{b}, \boldsymbol{a}\rangle \\ \langle \boldsymbol{b}, \tilde{\boldsymbol{a}}\rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}}\rangle \end{vmatrix}}, \tag{B.10}$$

which in virtue of Lemma 1 is equal to

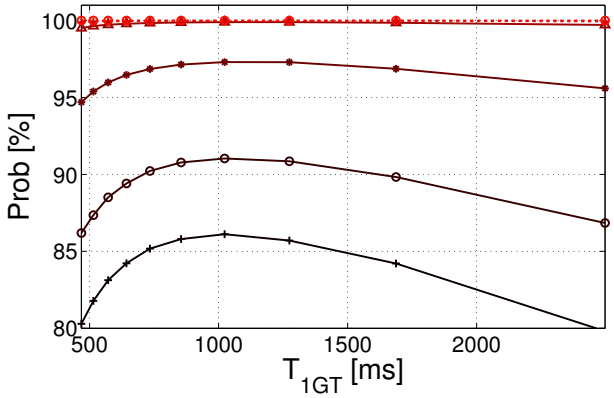
$$c_1^{k+1} = c_{1\mathrm{GT}} \frac{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b}\rangle & \langle \boldsymbol{b}, \boldsymbol{a}\rangle \\ \langle \boldsymbol{b}, \tilde{\boldsymbol{a}}\rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}}\rangle \end{vmatrix}}{\begin{vmatrix} \langle \boldsymbol{b}, \boldsymbol{b}\rangle & \langle \boldsymbol{b}, \boldsymbol{a}\rangle \\ \langle \boldsymbol{b}, \tilde{\boldsymbol{a}}\rangle & \langle \boldsymbol{a}, \tilde{\boldsymbol{a}}\rangle \end{vmatrix}} = c_{1\mathrm{GT}}. \tag{B.11}$$

B.4 Additional results of the experiment "MC simulation with synthetically generated SPGR MR signals"

In this section, we include additional results for the experiment of subsection 6.4.1. Specifically, statistical results and computation times for the case of the $\mathcal{A}_{N=3}$ FA set with the initialization $K_{\mathrm{ini}} = 0.5$ and $T_{1\mathrm{ini}} = 1000$ ms are shown in Fig. B.4 and Fig. B.5, respectively. Results for $K_{\mathrm{ini}} = 0.5$ and $T_{1\mathrm{ini}} = 500$ ms are presented in Fig. B.6 and Fig. B.7. Finally, in Fig. B.8 and Fig. B.9, statistical results and computation time are displayed for the case of the $\mathcal{A}_{N=10}$ FA set and initialization $K_{\mathrm{ini}} = 0.5$ and $T_{1\mathrm{ini}} = 500$ ms.

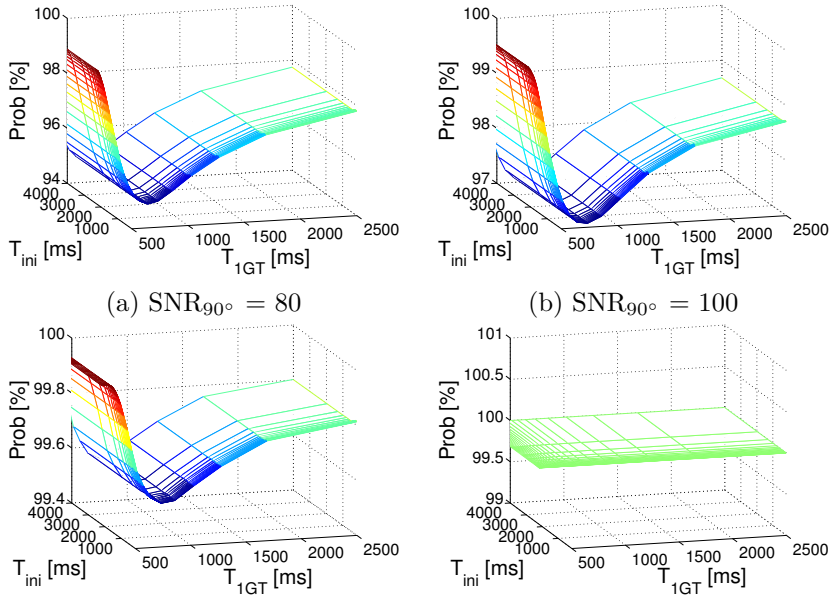


(a) Case of $\mathcal{A}_{N=3}$ FA set



(b) Case of $\mathcal{A}_{N=10}$ FA set

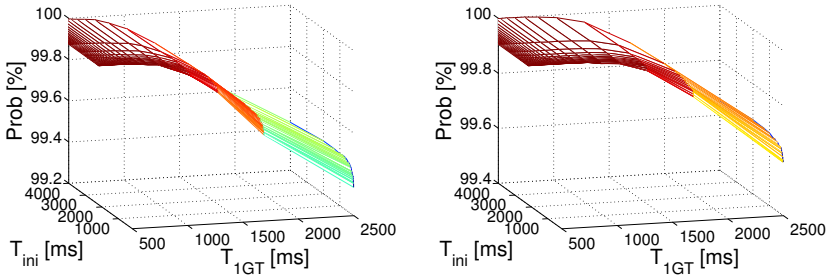
Figure B.1: Probability that the *bound condition* holds for $g_2(\cdot)$ as a function of $T_{1\mathrm{GT}}$ and SNR_{90°. Results with $\mathrm{SNR}_{90^\circ} > 500$ are not shown since the probability was 100%.



(a) $\mathrm{SNR}_{90^\circ} = 80$ (b) $\mathrm{SNR}_{90^\circ} = 100$

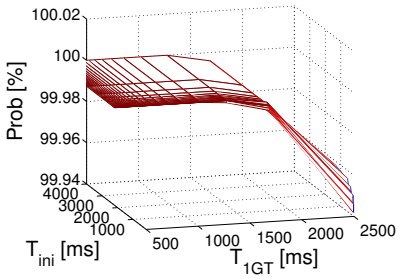
(c) $\mathrm{SNR}_{90^\circ} = 150$ (d) $\mathrm{SNR}_{90^\circ} = 250$

Figure B.2: Probability of NOVIFAST convergence as a function of $T_{1\mathrm{GT}}$ and $T_{1\mathrm{ini}}$ for the case of the $\mathcal{A}_{N=3}$ FA set. Results with $\mathrm{SNR}_{90^\circ} > 250$ are not shown since the probability was 100%.

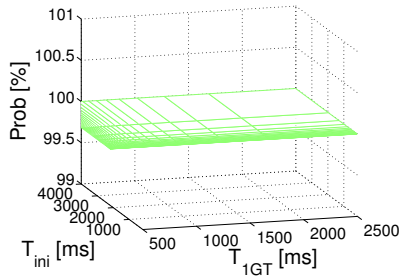


(a) $\mathrm{SNR}_{90^\circ} = 80$

(b) $\mathrm{SNR}_{90^\circ} = 100$

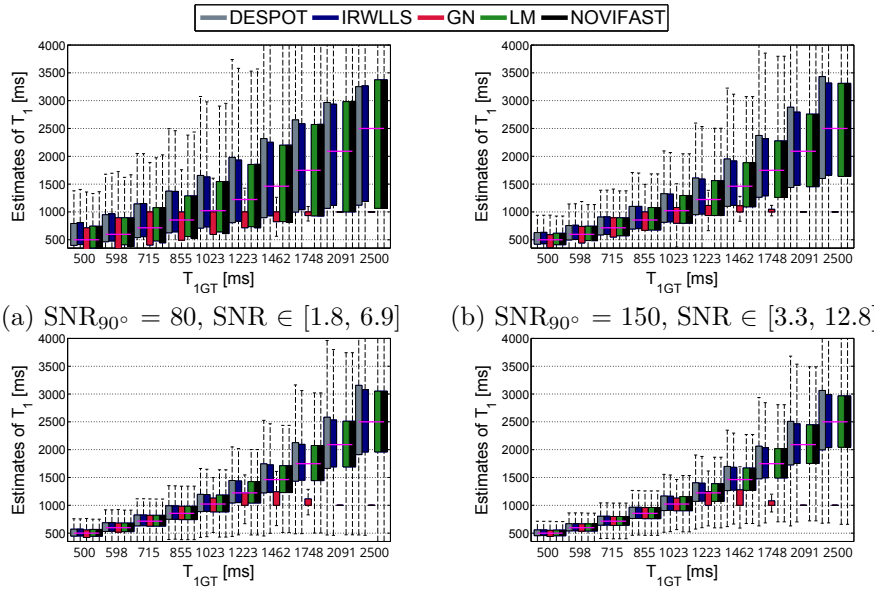


(c) $\mathrm{SNR}_{90^\circ} = 150$



(d) $\mathrm{SNR}_{90^\circ} = 250$

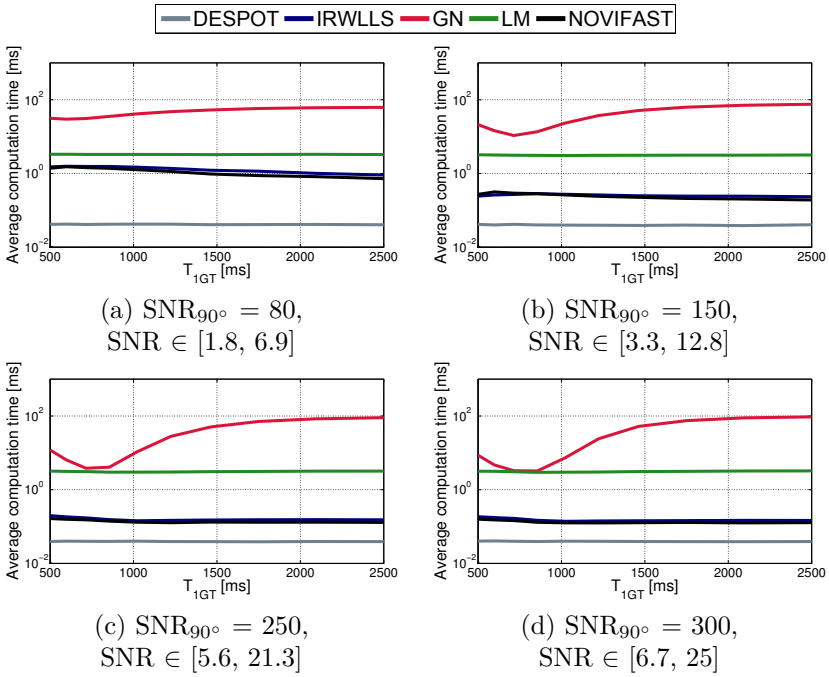
Figure B.3: Probability of NOVIFAST convergence as a function of $T_{1\mathrm{GT}}$ and $T_{1\mathrm{ini}}$ for the case of the $\mathcal{A}_{N=10}$ FA set. Results with $\mathrm{SNR}_{90^\circ} > 250$ are not shown since the probability was 100%.



(a) $\mathrm{SNR}_{90^\circ} = 80$, SNR $\in [1.8, 6.9]$ (b) $\mathrm{SNR}_{90^\circ} = 150$, SNR $\in [3.3, 12.8]$

(c) $\mathrm{SNR}_{90^\circ} = 250$, SNR $\in [5.6, 21.3]$ (d) $\mathrm{SNR}_{90^\circ} = 300$, SNR $\in [6.7, 25]$

Figure B.4: Box-plots of the T_1 estimates that are obtained with the five SPGR VFA optimization algorithm. Tukey-style whiskers are shown that extend to a maximum of 1.5 x IQR beyond each box, with IQR the interquartile range (corresponding with the length of each box) [Krzywinski and Altman, 2014]. Ground-truth T_1 values are marked with horizontal magenta lines to ease interpretation (Case of $\mathcal{A}_{N=3}$ FA data set and fixed initialization of $K_{\mathrm{ini}} = 0.5$ and $T_{1\mathrm{ini}} = 1000$ ms).



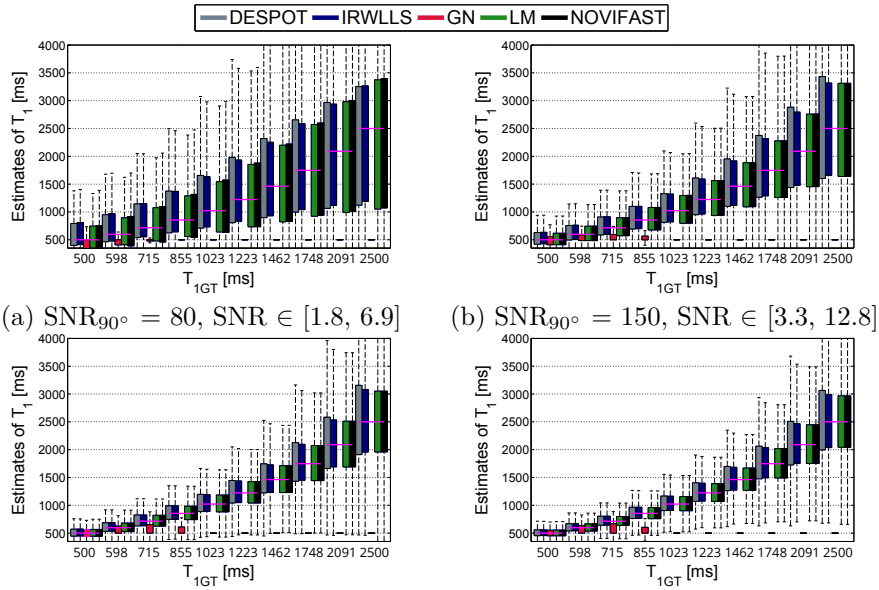
(a) $\text{SNR}_{90^\circ} = 80$, SNR $\in [1.8, 6.9]$

(b) $\text{SNR}_{90^\circ} = 150$, SNR $\in [3.3, 12.8]$

(c) $\text{SNR}_{90^\circ} = 250$, SNR $\in [5.6, 21.3]$

(d) $\text{SNR}_{90^\circ} = 300$, SNR $\in [6.7, 25]$

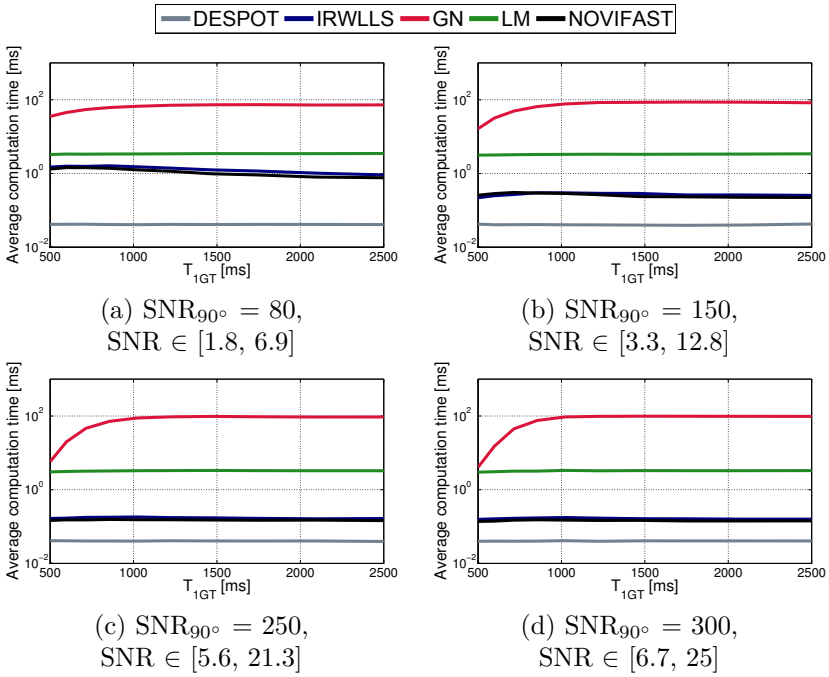
Figure B.5: Total computation time of each of the five optimization algorithms for the MC simulation-based experiment (Case of $\mathcal{A}_{N=3}$ FA set and fixed initialization of $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 1000$ ms).



(a) $\text{SNR}_{90^\circ} = 80$, SNR $\in [1.8, 6.9]$ (b) $\text{SNR}_{90^\circ} = 150$, SNR $\in [3.3, 12.8]$

(c) $\text{SNR}_{90^\circ} = 250$, SNR $\in [5.6, 21.3]$ (d) $\text{SNR}_{90^\circ} = 300$, SNR $\in [6.7, 25]$

Figure B.6: Box-plots of the T_1 estimates that are obtained with the five SPGR VFA optimization algorithm. Tukey-style whiskers are shown that extend to a maximum of 1.5 x IQR beyond each box, with IQR the interquartile range (corresponding with the length of each box) [Krzywinski and Altman, 2014]. Ground-truth T_1 values are marked with horizontal magenta lines to ease interpretation (Case of $\mathcal{A}_{N=3}$ FA set with Rician distributed data and fixed initialization of $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 500$ ms).



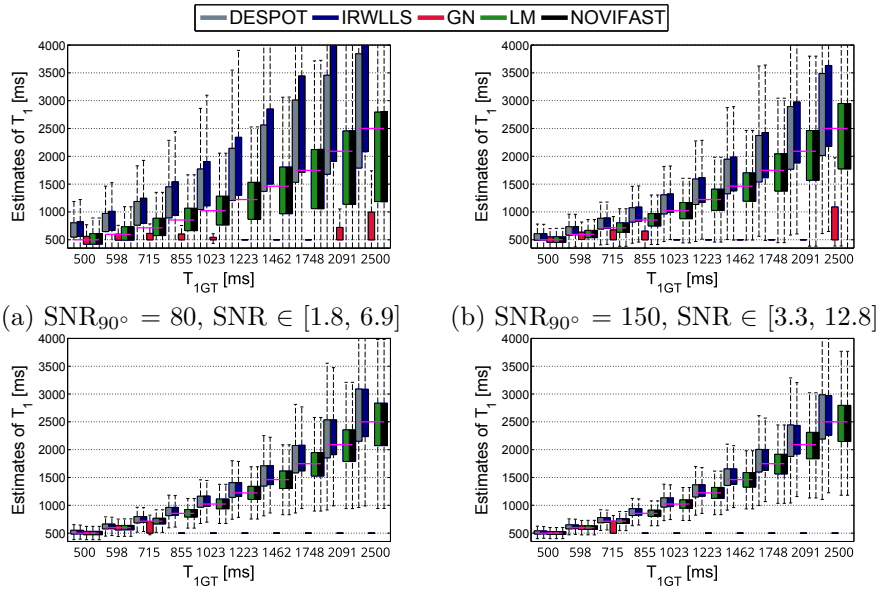
(a) $\text{SNR}_{90^\circ} = 80$, SNR $\in [1.8, 6.9]$

(b) $\text{SNR}_{90^\circ} = 150$, SNR $\in [3.3, 12.8]$

(c) $\text{SNR}_{90^\circ} = 250$, SNR $\in [5.6, 21.3]$

(d) $\text{SNR}_{90^\circ} = 300$, SNR $\in [6.7, 25]$

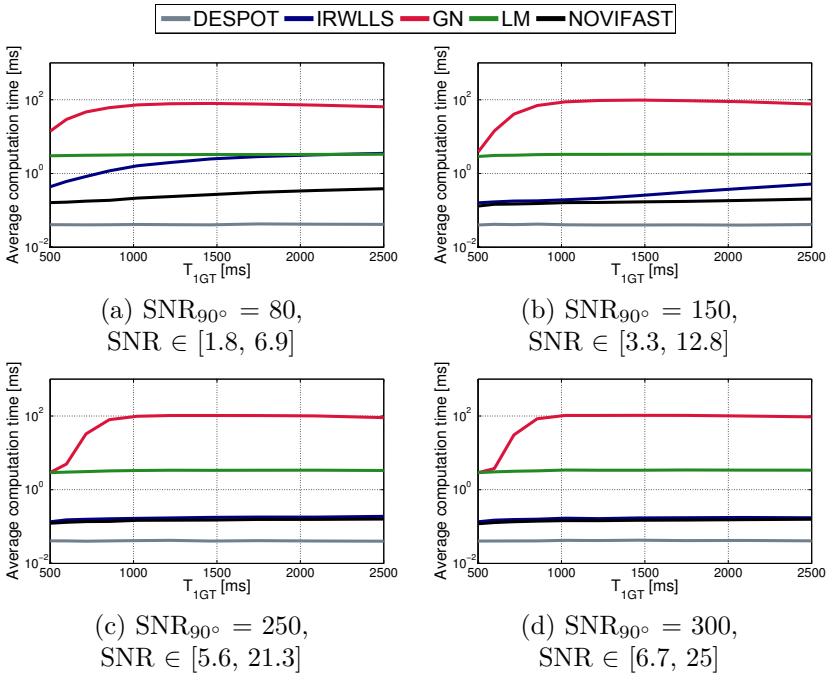
Figure B.7: Total computation time of each of the five optimization algorithms for the MC simulation-based experiment (Case of $\mathcal{A}_{N=3}$ FA set and fixed initialization of $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 500$ ms).



(a) $\text{SNR}_{90^\circ} = 80$, SNR $\in [1.8, 6.9]$ (b) $\text{SNR}_{90^\circ} = 150$, SNR $\in [3.3, 12.8]$

(c) $\text{SNR}_{90^\circ} = 250$, SNR $\in [5.6, 21.3]$ (d) $\text{SNR}_{90^\circ} = 300$, SNR $\in [6.7, 25]$

Figure B.8: Box-plots of the T_1 estimates that are obtained with the five SPGR VFA optimization algorithm. Tukey-style whiskers are shown that extend to a maximum of 1.5 x IQR beyond each box, with IQR the interquartile range (corresponding with the length of each box) [Krzywinski and Altman, 2014]. Ground-truth T_1 values are marked with horizontal magenta lines to ease interpretation (Case of $\mathcal{A}_{N=10}$ FA set and fixed initialization of $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 500$ ms).



(a) $\text{SNR}_{90^\circ} = 80$, SNR $\in [1.8, 6.9]$

(b) $\text{SNR}_{90^\circ} = 150$, SNR $\in [3.3, 12.8]$

(c) $\text{SNR}_{90^\circ} = 250$, SNR $\in [5.6, 21.3]$

(d) $\text{SNR}_{90^\circ} = 300$, SNR $\in [6.7, 25]$

Figure B.9: Total computation time of each of the five optimization algorithms for the MC simulation-based experiment (Case of $\mathcal{A}_{N=10}$ FA set and fixed initialization of $K_{\text{ini}} = 0.5$ and $T_{1\text{ini}} = 500$ ms).

C

Appendix to Chapter 7

In section C.1, we provide the implementation details of the minimization problem included in the Split Bregman method. Furthermore, in section C.2, we describe the simulation experiments that were conducted to perform the sensitivity analysis presented in subsection 7.4.2. Whereas subsection 7.4.2 only summarizes the main results of these experiments, a more detailed description and extensive discussion can be found in this appendix.

C.1 Implementation details of Eq. (7.20)

For a fixed iteration t of the Split Bregman algorithm, we have to solve the following minimization problem (subindices have been removed to simplify the notation):

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathbb{R}_+^N} \Phi(\boldsymbol{x}) + \frac{\lambda}{2}||\tilde{\boldsymbol{A}}\boldsymbol{x} - \boldsymbol{v}||_2^2, \tag{C.1}$$

with $\boldsymbol{v} = \boldsymbol{b}^{(t)}$. To impose the real-valuedness condition, we define $\boldsymbol{v}_{Re} = \mathrm{Re}(\boldsymbol{v})$, $\boldsymbol{v}_{Im} = \mathrm{Im}(\boldsymbol{v})$, $\tilde{\boldsymbol{A}}_{Re} = \mathrm{Re}(\tilde{\boldsymbol{A}})$, and $\tilde{\boldsymbol{A}}_{Im} = \mathrm{Im}(\tilde{\boldsymbol{A}})$, and then solve

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathbb{R}_+^N} \Phi(\boldsymbol{x}) + \frac{\lambda}{2}\left|\left|\begin{pmatrix}\tilde{\boldsymbol{A}}_{Re}\\ \tilde{\boldsymbol{A}}_{Im}\end{pmatrix}\boldsymbol{x} - \begin{pmatrix}\boldsymbol{v}_{Re}\\ \boldsymbol{v}_{Im}\end{pmatrix}\right|\right|_2^2. \tag{C.2}$$

When the problem is formulated in this way, any (gradient based) optimization algorithm will produce a real-valued solution since all variables are real. Furthermore, to impose the positivity constraint, a projection on the positive orthant

as $[\boldsymbol{x}]^+ = (\max(x_1, 0), ..., \max(x_N, 0))^T$ can be used in every iteration [Bertsekas, 1982].

Targeting at an efficient implementation of the PD method, we would like to exploit the fact that matrix $\tilde{\boldsymbol{A}}$ is, in the single-coil case, a cascade of an undersampling mask, a unitary Fourier matrix (in case of Cartesian sampling), and a diagonal matrix that includes the phase. Indeed, multiplications of the form $\tilde{\boldsymbol{A}}\boldsymbol{x}$ can be efficiently implemented with the FFT algorithm. However, if we use the decomposition of $\tilde{\boldsymbol{A}}$ in its real and imaginary part, it is not trivial how to relate the resulting matrices with the FFT. In order to exploit the FFT advantages and still impose the real positivity constraint, we used a modified version of the Majorize-Minimize (MM) algorithm recently proposed by Muckley *et al.* [Muckley et al., 2015]. We constructed a collection of surrogate functions, i.e., majorizers of the function

$$s(\boldsymbol{x}) = \Phi(\boldsymbol{x}) + \frac{\lambda}{2}\left|\left|\tilde{\boldsymbol{A}}\boldsymbol{x} - \boldsymbol{v}\right|\right|_2^2, \tag{C.3}$$

and minimized them, which implies the minimization of $s(\boldsymbol{x})$ [Muckley et al., 2015]. Surrogate functions for $f(\boldsymbol{x}) = \frac{1}{2}\left|\left|\tilde{\boldsymbol{A}}\boldsymbol{x} - \boldsymbol{v}\right|\right|_2^2$ are presented in Eq. (5) of Muckley *et al.* [Muckley et al., 2015] as

$$g(\boldsymbol{x}, \boldsymbol{x}^k) = \frac{1}{2}\left|\left|\boldsymbol{x} - (\boldsymbol{x}^k - \boldsymbol{D}_f^{-1}\tilde{\boldsymbol{A}}^H(\tilde{\boldsymbol{A}}\boldsymbol{x}^k - \boldsymbol{v}))\right|\right|_{\boldsymbol{D}_f}^2 + \xi, \tag{C.4}$$

where $\boldsymbol{D}_f \triangleq L\boldsymbol{I}$, with L the maximum eigenvalue of $\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{A}}$, and where ξ is a constant. If $g(\boldsymbol{x}, \boldsymbol{x}^k)$ is a surrogate function of $f(\boldsymbol{x})$, i.e., if $f(\boldsymbol{x}^k) = g(\boldsymbol{x}^k, \boldsymbol{x}^k)$ and $f(\boldsymbol{x}) \leq g(\boldsymbol{x}, \boldsymbol{x}^k)$ for all $\boldsymbol{x}$, clearly $\lambda g(\boldsymbol{x}, \boldsymbol{x}^k) + \Phi(\boldsymbol{x})$ is a surrogate function for $s(\boldsymbol{x})$. The benefit of this MM approach is that now we do not need to separate matrix $\tilde{\boldsymbol{A}}$ into its real and imaginary part. Indeed,

$$\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x}\in\mathbb{R}_+^N} \lambda g(\boldsymbol{x}, \boldsymbol{x}^k) + \Phi(\boldsymbol{x}) = \arg\min_{\boldsymbol{x}\in\mathbb{R}_+^N} \frac{\lambda}{2}||\boldsymbol{x} - (\boldsymbol{\Delta}_{Re}^k + j\boldsymbol{\Delta}_{Im}^k)||_{\boldsymbol{D}_f}^2 + \Phi(\boldsymbol{x}), \tag{C.5}$$

with $\boldsymbol{\Delta}_{Re}^k$ and $\boldsymbol{\Delta}_{Im}^k$ the real and imaginary parts of $\boldsymbol{\Delta}^k$, with

$$\boldsymbol{\Delta}^k = \boldsymbol{x}^k - \boldsymbol{D}_f^{-1}\tilde{\boldsymbol{A}}^H(\tilde{\boldsymbol{A}}\boldsymbol{x}^k - \boldsymbol{v}). \tag{C.6}$$

The minimization problem now becomes

$$\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x}\in\mathbb{R}_+^N} \frac{\lambda}{2}||\boldsymbol{x} - \boldsymbol{\Delta}_{Re}^k||_{\boldsymbol{D}_f}^2 + \frac{\lambda}{2}||\boldsymbol{\Delta}_{Im}^k||_{\boldsymbol{D}_f}^2 + \Phi(\boldsymbol{x}). \tag{C.7}$$

Because a constant does not affect the minimization, Eq. (C.7) simplifies to

$$\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x}\in\mathbb{R}_+^N} \gamma_k(\boldsymbol{x}), \tag{C.8}$$

with $\gamma_k(\boldsymbol{x}) = \frac{\lambda}{2}||\boldsymbol{x} - \boldsymbol{\Delta}_{Re}^k||_{\boldsymbol{D}_f}^2 + \Phi(\boldsymbol{x})$. Since the calculation of the gradient of $\gamma_k(\boldsymbol{x})$ does not involve any complex number, any method that updates the previous iteration with a gradient step will produce a real-valued solution. In fact,

$$\nabla\gamma_k(\boldsymbol{x}) = \lambda L(\boldsymbol{x} - \boldsymbol{\Delta}_{Re}^k) + \nabla\Phi(\boldsymbol{x}). \tag{C.9}$$

In summary, our optimization method involves solving problem Eq. (C.8). For this purpose, we chose the Fletcher-Reeves nonlinear conjugate gradient (NCG) method [Nocedal and Wright, 2006], with a step size obtained using the Armijo (line search) rule [Armijo, 1966, Nocedal and Wright, 2006]. Positivity was imposed (in each iteration) with $[\boldsymbol{x}]^+ = (\max(x_1, 0), ..., \max(x_N, 0))^T$ [Bertsekas, 1982].

The solution $\boldsymbol{x}^{k+1}$ serves to update $\boldsymbol{\Delta}_{Re}^k$ and hence to update $\gamma_k(\boldsymbol{x})$, which defines a new optimization problem. This problem is solved again with the NCG method, from which $\gamma_k(\boldsymbol{x})$ is updated. This process is repeated until the number of iterations exceeds a given threshold. This process constitutes the whole algorithm to solve the problem of Eq. (C.1). Since we cannot prove that the problem is convex, due to the analytical form of $\mathcal{P}(\boldsymbol{x})$, convergence to a global minimum cannot be assured. However, the combination of steps described above guarantees that $s(\boldsymbol{x})$ decreases with each iteration and the algorithm provided satisfactory solutions in our experiments.

The maximum eigenvalue L of $\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{A}}$ Here, we prove that $L = 1$ as in the work of Muckley *et al.* [Muckley et al., 2015]. If we work with Cartesian sampling, $\tilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Psi}$ can be written as $\tilde{\boldsymbol{A}} = \boldsymbol{S}\boldsymbol{F}\boldsymbol{\Psi}$ where $\boldsymbol{F} \in \boldsymbol{C}^{N\times N}$ is a unitary Fourier matrix and $\boldsymbol{S} \in [0, 1]^{M\times N}$, with $M < N$, is the under-sampling matrix. For our purposes, $\boldsymbol{S}$ can be defined as an identity matrix where $M - N$ rows have been removed. Those rows are indexed depending on the sampling pattern. First, it is trivial to see that $\boldsymbol{R} = \boldsymbol{S}^H\boldsymbol{S}$ is an $N \times N$ diagonal matrix with only 1s and 0s in its diagonal. Second, we have that $\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{A}} = \boldsymbol{\Psi}^H\boldsymbol{F}^H\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{F}\boldsymbol{\Psi} = \boldsymbol{\Psi}^H\boldsymbol{F}^H\boldsymbol{R}\boldsymbol{F}\boldsymbol{\Psi}$. By calling $\boldsymbol{P}^H = \boldsymbol{F}\boldsymbol{\Psi}$, we have that $\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{A}} = \boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^H$ is already in its diagonal form and the eigenvalues are those contained in the diagonal of $\boldsymbol{R}$. Finally, because eigenvalues only take the value 1 or 0, we have that $L = 1$.

Multi-coil case In the multi-coil case, it can be demonstrated [Muckley et al., 2015] that the diagonal matrix $\boldsymbol{D}_f$ can be written as

$$\boldsymbol{D}_f \triangleq L\boldsymbol{\Psi}^H\boldsymbol{C}\boldsymbol{\Psi}, \tag{C.10}$$

where L is the maximum eigenvalue of the block-matrix conformed by matrix $\boldsymbol{A}$, which is again 1. The diagonal matrix $\boldsymbol{C} \in \mathbb{R}_+^{N\times N}$ is defined as

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1^H & \boldsymbol{C}_2^H & \cdots & \boldsymbol{C}_R^H \end{pmatrix} \begin{pmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \\ \vdots \\ \boldsymbol{C}_R \end{pmatrix}, \tag{C.11}$$

where $\{\boldsymbol{C}_r\}_{r=1}^R$ are the diagonal matrices that contain the coil sensitivities (subsection 7.3.4). The gradient in Eq. (C.9) is now replaced by

$$\nabla\gamma_k(\boldsymbol{x}) = \lambda\boldsymbol{D}_f(\boldsymbol{x} - \boldsymbol{\Delta}_{Re}^k) + \nabla\Phi(\boldsymbol{x}). \tag{C.12}$$

Note that $\nabla\gamma_k(\boldsymbol{x})$ is again real since $\boldsymbol{D}_f$ is real. In fact, it is easy to show that $\boldsymbol{D}_f = \boldsymbol{C}$ because $\boldsymbol{C}$ and $\boldsymbol{\Psi}$ commute (they are diagonal) and $\boldsymbol{\Psi}$ is unitary.

C.1.1 Calculation of gradient

Remembering that $\Phi(\boldsymbol{x}) = ||\boldsymbol{x} - \mathcal{P}(\boldsymbol{x})||_1$, its (weak) gradient is given by

$$\nabla\Phi(\boldsymbol{x}) = (\boldsymbol{I} - \boldsymbol{J}_{\mathcal{P}}(\boldsymbol{x}))^T \operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x})), \tag{C.13}$$

where $\operatorname{sgn}(\cdot)$ is the sign function, i.e., $\operatorname{sgn}(x) = 1$ if $x > 0$, $\operatorname{sgn}(x) = -1$ if $x < 0$ and $\operatorname{sgn}(x) = 0$ if $x = 0$; $\boldsymbol{I} \in \mathbb{R}^{N\times N}$ is the identity matrix and $\boldsymbol{J}_{\mathcal{P}}(\boldsymbol{x}) \in \mathbb{R}^{N\times N}$ is the Jacobian of $\mathcal{P}(\boldsymbol{x})$. The expression of $\mathcal{P}(\boldsymbol{x})$ is

$$\mathcal{P}(\boldsymbol{x}) = \sum_{k=1}^{K} \eta_k \boldsymbol{p}_k + \boldsymbol{p}_{\bar{\mathcal{A}}} \circ \boldsymbol{x}_\rho, \tag{C.14}$$

where $\boldsymbol{x}_\rho \triangleq \mathcal{G}_\rho * \boldsymbol{x}$ with $\mathcal{G}_\rho$ the Gaussian kernel and $*$ the convolution operator. To facilitate the Jacobian calculation, note that because the convolution is a linear operation, if we work purely with vectors and matrices, there exist a matrix $\boldsymbol{G}_\rho \in \mathbb{R}^{N\times N}$ [Jain, 1989] that relates $\boldsymbol{x}_\rho$ to $\boldsymbol{x}$ as $\boldsymbol{x}_\rho = \boldsymbol{G}_\rho\boldsymbol{x}$ and $\mathcal{P}(\boldsymbol{x})$ can be written as

$$\mathcal{P}(\boldsymbol{x}) = \sum_{k=1}^{K} \eta_k \boldsymbol{p}_k + \operatorname{diag}(\boldsymbol{p}_{\bar{\mathcal{A}}})\, \boldsymbol{G}_\rho \boldsymbol{x}. \tag{C.15}$$

Because $\boldsymbol{p}_k$ and $\boldsymbol{p}_{\bar{\mathcal{A}}}$ depend pointwise on $\boldsymbol{x}$, we have

$$\boldsymbol{J}_{\mathcal{P}}(\boldsymbol{x}) = \sum_{k=1}^{K} \eta_k \operatorname{diag}\left(\frac{\partial \boldsymbol{p}_k(x_n)}{\partial x_n}\right) + \operatorname{diag}\left(\frac{\partial \boldsymbol{p}_{\bar{\mathcal{A}}}(x_n)}{\partial x_n}\right) \boldsymbol{G}_\rho\boldsymbol{x} + \boldsymbol{G}_\rho^T \operatorname{diag}(\boldsymbol{p}_{\bar{\mathcal{A}}}), \tag{C.16}$$

or

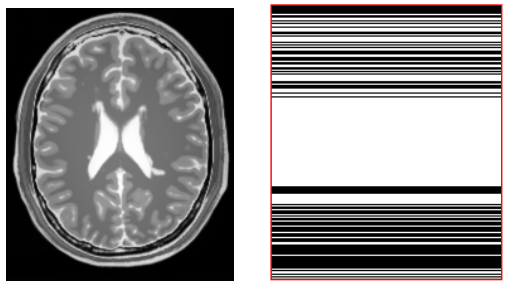
$$\boldsymbol{J}_{\mathcal{P}}(\boldsymbol{x}) = \sum_{k=1}^{K} \eta_k \operatorname{diag}\left(\frac{\partial \boldsymbol{p}_k(x_n)}{\partial x_n}\right) + \operatorname{diag}\left(\frac{\partial \boldsymbol{p}_{\bar{\mathcal{A}}}(x_n)}{\partial x_n}[\boldsymbol{x}_\rho]_n\right) + \boldsymbol{G}_\rho^T \operatorname{diag}(\boldsymbol{p}_{\bar{\mathcal{A}}}). \tag{C.17}$$

The partial derivatives of the probability maps can be obtained by applying the quotient rule for derivatives to the posteriori probabilities of Eqs. (7.9-7.10), i.e.,

$$\frac{\partial \boldsymbol{p}_k(x_n)}{\partial x_n} = \frac{p'_{X|\mathcal{A}_k}(x_n)p_X(x_n) - p_{X|\mathcal{A}_k}(x_n)p'_X(x_n)}{p_X^2(x_n)}, \tag{C.18}$$

$$\frac{\partial \boldsymbol{p}_{\bar{\mathcal{A}}}(x_n)}{\partial x_n} = \frac{p'_{X|\bar{\mathcal{A}}}(x_n)p_X(x_n) - p_{X|\bar{\mathcal{A}}}(x_n)p'_X(x_n)}{p_X^2(x_n)}, \tag{C.19}$$

with $p'_X(x) \triangleq \frac{\partial p_X(x)}{\partial x}$, $p'_{X|\bar{\mathcal{A}}_k}(x) \triangleq \frac{\partial p_{X|\bar{\mathcal{A}}_k}(x)}{\partial x}$, and $p'_{X|\bar{\mathcal{A}}}(x) \triangleq \frac{\partial p_{X|\bar{\mathcal{A}}}(x)}{\partial x}$. Obviously, the derivatives of the Gaussian probability functions can easily be obtained exactly.



(a) Ground-truth (b) Sampling mask

Figure C.1: Magnitude image used as ground-truth (a) and sampling mask (b) with 40% of undersampling.

Finally, it is interesting to note that we do not need to calculate $\boldsymbol{G}_\rho^T$ because

$$\begin{aligned}
&\nabla\Phi(\boldsymbol{x}) = (\boldsymbol{I} - \boldsymbol{J}_{\mathcal{P}}(\boldsymbol{x}))^T \operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x})) = \\
&\operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x})) - \sum_{k=1}^{K} \eta_k \operatorname{diag}\left(\frac{\partial \boldsymbol{p}_k(x_n)}{\partial x_n}\right) \operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x})) - \\
&\operatorname{diag}\left(\frac{\partial \boldsymbol{p}_{\bar{\mathcal{A}}}(x_n)}{\partial x_n}[\boldsymbol{x}_\rho]_n\right) \operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x})) - \operatorname{diag}(\boldsymbol{p}_{\bar{\mathcal{A}}}) \operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x}))_\rho, \qquad \text{(C.20)}
\end{aligned}$$

with $\operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x}))_\rho \triangleq \boldsymbol{G}_\rho \operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x})) = \mathcal{G}_\rho * \operatorname{sgn}(\boldsymbol{x} - \mathcal{P}(\boldsymbol{x}))$.

C.2 Details of the sensitivity analysis of PD

C.2.1 Sensitivity of the PD method to the pre-selected number of homogeneous regions K: experiment setup and results

The digital phantom was constructed as follows. A magnitude image (217 × 181) (Fig. C.1.(a)) was downloaded from *BrainWeb*[1] [Cocosco et al., 1997]. This magnitude image was considered as the ground-truth image. A smoothly varying phase was simulated by Legendre polynomials up to the second degree.

Following the same procedure as in the first simulation experiment (7.4.1.1), complex valued Gaussian noise was added to create $N_{\text{Runs}} = 20$ complex noisy images. Again, the SNR, defined by Eq. (7.26), was chosen equal to 10. From these noisy complex images, under-sampled k-space data were generated with the random phase-encoding mask which is shown in Fig. C.1.(b). The GMM was learned from one realization of $|\boldsymbol{x}_{\boldsymbol{LR}}|$. The optimal number of components (or, classes) was found to be equal to eight. The *partial discreteness degree*, r, was set to 0.99. The

[1] Available at http://brainweb.bic.mni.mcgill.ca/brainweb/

size of the symmetric Hamming window filter was $N_{\text{Hamm}} = 50$. The rest of the PD parameters were the same as in the previous experiments. The corresponding *a posteriori* probability maps related to $|\boldsymbol{x}_{\boldsymbol{LR}}|$ are displayed in Fig. C.2. K was then varied from zero (no homogeneous region) to eight, the total number of components. Graphs of the reconstruction quality measures PSNR, QILV, HFEN and FSIM as a function on K are shown in Fig. C.3. In addition, we also present in Fig. C.4 the expected value of the reconstructed images for different values of K with their corresponding RMSE maps. The first important observation is that the worst

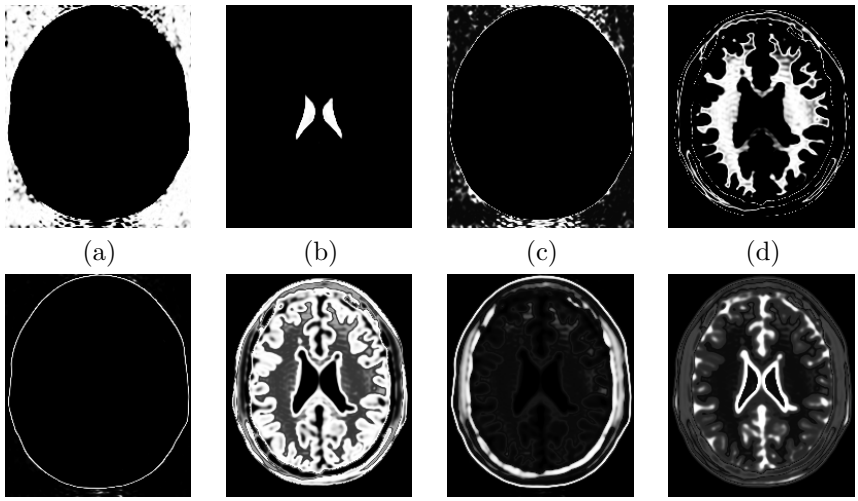
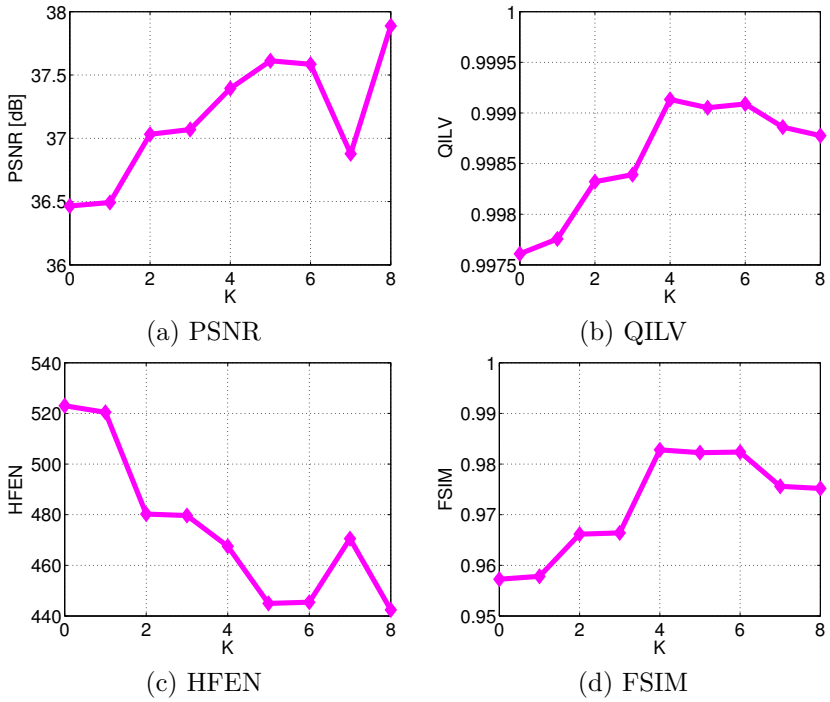


Figure C.2: *A posteriori* probability maps derived from the GMM (total number of components is equal to eight) which is learned from the initial low resolution image, $|\boldsymbol{x}_{\boldsymbol{LR}}|$. In (a) and (b), the probability maps of the classes with the lowest and highest mean are shown. The remaining probability maps are shown in (c-h) in ascending order with respect of the standard deviation.

result is achieved with $K = 0$. This shows that using the partial discreteness prior is beneficial, and reconstruction quality is gained if at least one region is assumed to be constant in intensity. Furthermore, there is a substantial improvement in terms of all four metrics if the hyper-intense region is also incorporated in the PD prior. This region corresponds to the CSF in the ventricles, see the *a posteriori* probability map displayed in Fig. C.2.(b). The graphs of QILV and FISM show an optimum at $K = 4$. This result is not surprising. It follows from the probability map displayed in Fig. C.2.(d) that the fourth conditional PDF corresponds to the white matter region. The fact that artifacts are most severely present in this region may explain why the assumption of a homogeneous white matter region produces the observed gain in reconstruction quality. See also the reconstructed images in Fig. C.4. Choosing K higher than four is not preferable in terms of QILV and FSIM. A possible explanation is that the additional estimated conditional PDFs are related to regions that can hardly be considered as homogeneous, observe for example the *a posteriori* probability maps in Fig. C.2.(e-h). Hence, it is preferable



(a) PSNR (b) QILV

(c) HFEN (d) FSIM

Figure C.3: Performance of PD depending on pre-selected number of homogeneous regions, K.

to consider them as components of the estimated conditional PDF $p_{X|\bar{A}}(x; \hat{\boldsymbol{\theta}}_{\bar{A}})$ for the texture part. Graphs of PSRN and HFEN follow an expected trend up to $K = 4$, showing that reconstruction quality improves when K increases. Above $K = 4$, the behavior of both graphs fluctuates. The source of this fluctuation is unknown to us and we prefer not to speculate about possible explanations.

C.2.2 Sensitivity of the PD method to non-slowly varying phase: experiment setup and results

The three phase profiles that were added to the ground-truth image (same magnitude image as in the previous experiment) are shown in Fig. C.5. Phase profile A mimics a slow-varying phase image. It was constructed using Legendre polynomials up to the second degree. Phase profile B and phase profile C were created emulating the same protocol as in [Haldar, 2014a], see for example Fig. 6 in [Haldar, 2014a]. Following the same procedure as in the previous experiment, complex valued Gaussian noise was added to create $N_{\text{Runs}} = 20$ complex noisy images. Again, the SNR, defined by 7.26, was chosen equal to 10. From these noisy complex images, under-sampled k-space data were produced with the same random phase-encoding mask as in the preceding experiment (see Fig. C.1(b)). For each phase profile, the low resolution image $\boldsymbol{x_{LR}}$ was reconstructed using different sizes of the Hamming window, namely,

$N_{\text{Hamm}} = \{20, 50, 100, 200\}$. In this experiment, we also selected $r = 0.99$. The number of homogeneous regions was set to $K = 4$. Graphs of the reconstruction quality measures PSNR, QILV, HFEN and FSIM as a function on N_{Hamm} are presented in Fig. C.6, whereas the expected value of the reconstructed images, that is, the sample mean over the N_{Runs} realizations, are shown in Fig. C.7. These results show that, not surprisingly, the performance of the PD method decreases for more rapidly varying phase profiles. Nevertheless, the results also show that PD is quite robust to moderate phase variations such as those present in phase profile B. Interestingly, increasing N_{Hamm} does not always lead to an improvement in reconstruction quality, except for the phase profile C case. A plausible reason is that for phase profiles A and B, most of the phase information is preserved with moderate values of N_{Hamm}, and increasing this number may permit artifacts to be preserved, hence negatively affecting the initial reconstruction. Considering the four reconstruction quality metrics, it seems that $N_{\text{Hamm}} = 50$ is an optimal value for profile A and B, since reconstruction quality degrades for higher values. This result motivated the use of $N_{\text{Hamm}} = 50$ in all other experiments described in this work.

C.2.3 Sensitivity of the PD method to bias field: experiment setup and results

Multiplicative linear bias fields were simulated as $C(X + Y) + B$, with X and Y the spatial coordinates of the image. The bias fields were constructed such that their lowest value is always located in the top-left corner and their highest value in the bottom-right corner of the image. Furthermore, the parameters B and C were varied in such a way that the bias fields always have value '1' in the center of the brain (corresponding with the green point in Fig. C.8.(a)), and a predefined decreased/increased value $[10 - 70]\%$ at the top/bottom red points in Fig. C.8. Two examples of such linear bias fields are shown in Fig. C.8.(a) and Fig. C.8.(c). In order to generate the biased partially discrete image, that is, the partially discrete image corrupted by a bias field, we used the same unbiased ground-truth magnitude image that was used in the previous experiment with phase profile A (see subsection C.2.2). This complex image was point-wise multiplied by the simulated linear bias field to finally get the biased partially discrete image. Examples of the magnitude of the thus obtained biased images are shown in Fig. C.8.(b) and Fig. C.8.(d).

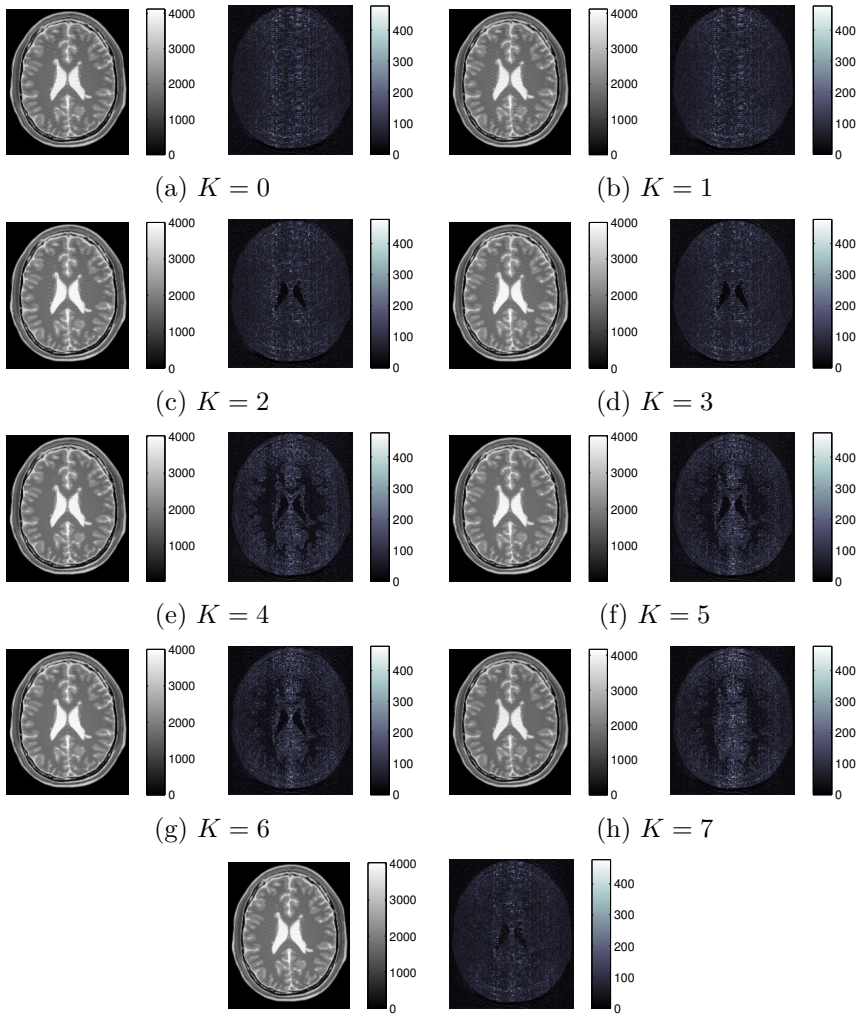
For each of the degrees of variation used in this experiment, which range from 0% (no bias field) to 70% (severe bias field), in increments of 10%, N_{Runs} complex noisy images were created in the same fashion as in the previous experiment. Next, under-sampled k-space data were generated, using again the sampling mask shown in Fig. C.1.(b). The number of homogeneous regions was set to $K = 4$ and, for each of the degrees of variation, the GMM was learned from $|\boldsymbol{x_{LR}}|$. The remaining parameters were the same as those from the previous experiments. Graphs of the reconstruction quality measures PSNR, QILV, HFEN, and FSIM as a function of the degree of bias field variation are shown in Fig. C.6. Moreover, reconstructed images and RMSE maps are presented in Fig. C.10.

It follows from these graphs and error maps that the reconstruction quality decreases with an increasing degree of bias field variation. This effect can easily be understood by realizing that the presence of a bias field challenges the partial discreteness assumption. That is, homogeneous regions become less homogeneous due to the presence of a bias field.

Note that, since not all homogeneous regions may be equally affected, the presence of a bias field may also influence the optimal choice of K. Indeed, the background and hyper-intense region (being a localized homogeneous region) are generally less affected by a bias field than the regions that correspond with $K > 2$ (such as the white matter area in this experiment), suggesting that in the presence of a severe bias field a conservative choice of K is advisable. The effectiveness of such a strategy was illustrated in the experiment with the in-vivo human k-space data that was discussed in subsection 7.5.1.2. In that experiment, each of the coil images $\boldsymbol{x}_r$, $r = 1, ..., 8$, was reconstructed independently, that is, without knowledge of the coil sensitivities. That implies that the partially discrete images we tried to restore were severely affected by bias field, whose source, in this case, was the sensitivity profile of each coil. In that experiment, we chose $K = 2$, since the then treated homogeneous regions (background and hyper-intense region) were not substantially affected by image inhomogeneities. Instead, most of the bias field variation was incorporated in the texture part. The results of the experiment show the good performance of the PD method, despite of the presence of a bias field.

It is also worthwhile to mention that, in a multi-coil acquisition system, if the goal is to reconstruct the original partially discrete $\boldsymbol{x}$ instead of each $\boldsymbol{x}_r$ separately, the image intensity inhomogeneities which are present in the coil images are captured in the forward model, via the profiles of the coil sensitivities (see subsection 7.3.4). If such coil sensitivities are reasonably well estimated, it can be expected that $\boldsymbol{x}$ will have a very low degree of intensity inhomogeneities, which can be well-handled with PD. This was the case in the experiment with in vivo human brain k-space data (see subsection 7.5.1.3), where it was shown that PD succeeded in reconstructing a partially discrete image from under-sampling multi-coil k-space data. Although out of the scope of this work, the robustness of the PD method to bias fields can be improved by including a bias-field correction technique. Prior estimation of the bias field would allow to include its effect into the forward model, in a similar way as the coil sensitivities in the multi-coil model (see subsection 7.3.4).

Furthermore, as already mentioned in subsection 7.3.5, the employment of the GMM has already been proved useful to deal with bias field correction in MR images, where both the original Gaussian component parameters, that is, mean and standard deviation, are jointly estimated, together with a slowly-varying bias field [Van Leemput et al., 1999a, Mai et al., 2011].



(a) $K = 0$ (b) $K = 1$

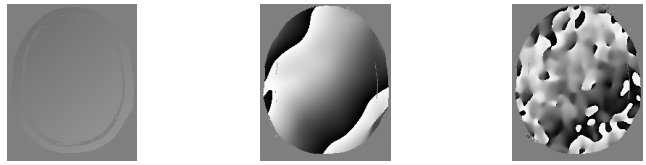
(c) $K = 2$ (d) $K = 3$

(e) $K = 4$ (f) $K = 5$

(g) $K = 6$ (h) $K = 7$

(i) $K = 8$

Figure C.4: Visual results for the experiment with different values of K. The sample mean of the magnitude of the reconstructed images are shown in companion of the RMSE maps. The colorbar range of the RMSE maps was adapted to $[0, T]$, where T is 40% of the maximum value which was found in all RMSE maps.



(a) Phase profile A (b) Phase profile B (c) Phase profile C

Figure C.5: Different profiles of simulated phase images (intensity between $-\pi$ and π).

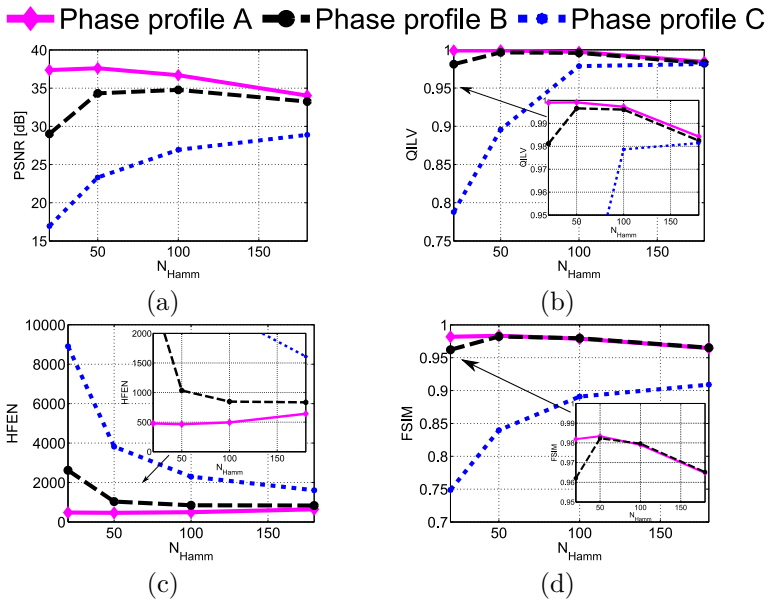
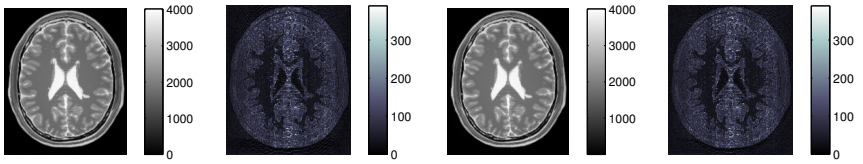
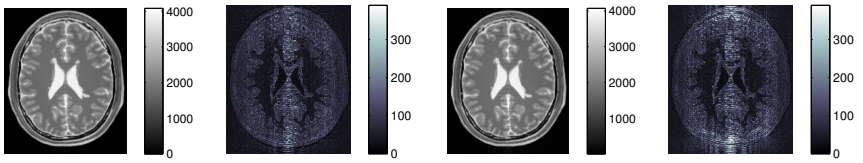


Figure C.6: Reconstruction quality of PD for different profiles of phase images, in terms of (a) PSNR, (b) QILV, (c) HFEN and (d) FSIM. The results are shown as a function of the Hamming filter window size, N_{Hamm}.



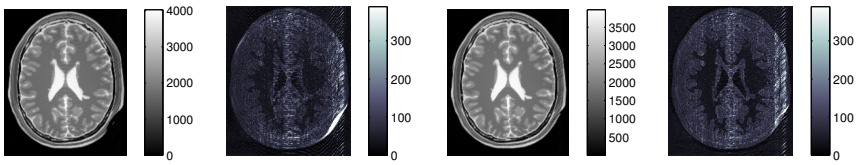
(a) Phase profile A, $N_{\text{Hamm}} = 20$

(b) Phase profile A, $N_{\text{Hamm}} = 50$



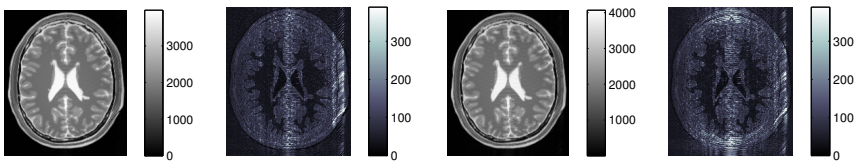
(c) Phase profile A, $N_{\text{Hamm}} = 100$

(d) Phase profile A, $N_{\text{Hamm}} = 200$



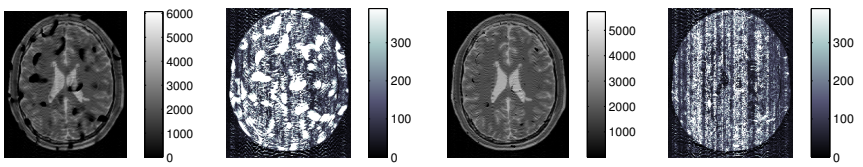
(e) Phase profile B, $N_{\text{Hamm}} = 20$

(f) Phase profile B, $N_{\text{Hamm}} = 50$



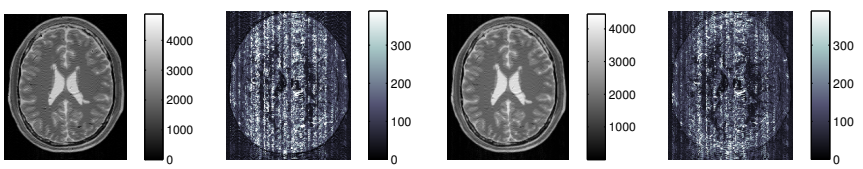
(g) Phase profile B, $N_{\text{Hamm}} = 100$

(h) Phase profile B, $N_{\text{Hamm}} = 200$



(i) Phase profile C, $N_{\text{Hamm}} = 20$

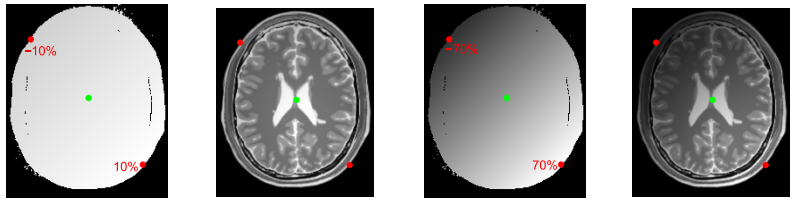
(j) Phase profile C, $N_{\text{Hamm}} = 50$



(k) Phase profile C, $N_{\text{Hamm}} = 100$

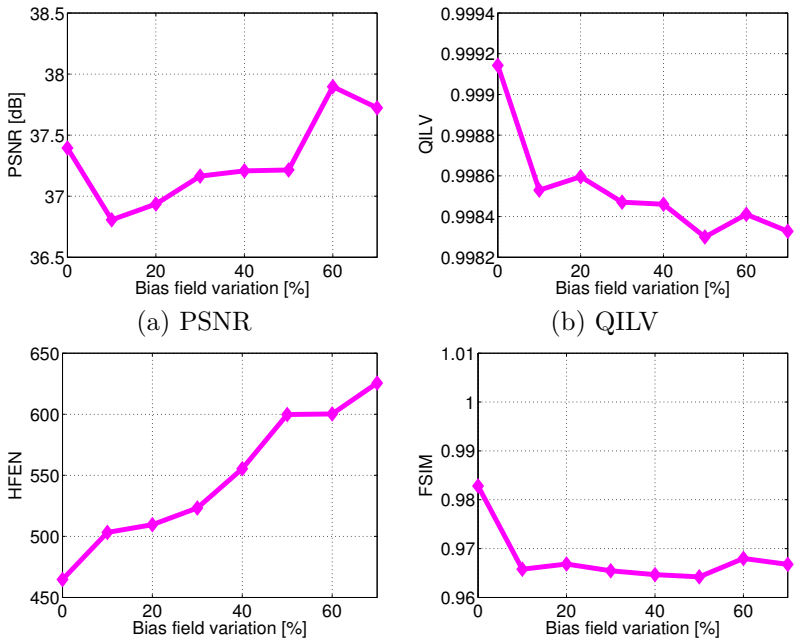
(l) Phase profile C, $N_{\text{Hamm}} = 200$

Figure C.7: Visual results for the experiment with different phase profiles and different Hamming window size N_{Hamm}. The sample mean of the magnitude of the reconstructed images are shown in companion of the RMSE maps. To highlight small errors, the colorbar range of the RMSE maps was adapted to $[0, T]$, where T is 10% of the maximum value which was found in all RMSE maps.



(a) Bias field of 10% (b) Biased partially discrete image (c) Bias field of 70% (d) Biased partially discrete image

Figure C.8: Examples of multiplicative linear bias fields that were used to check the influence of bias field on the PD performance are displayed in (a) and (c). The corresponding biased partially discrete images are shown in (d) and (d).



(a) PSNR (b) QILV

(c) HFEN (d) FSIM

Figure C.9: Performance of PD depending on the bias field variation.

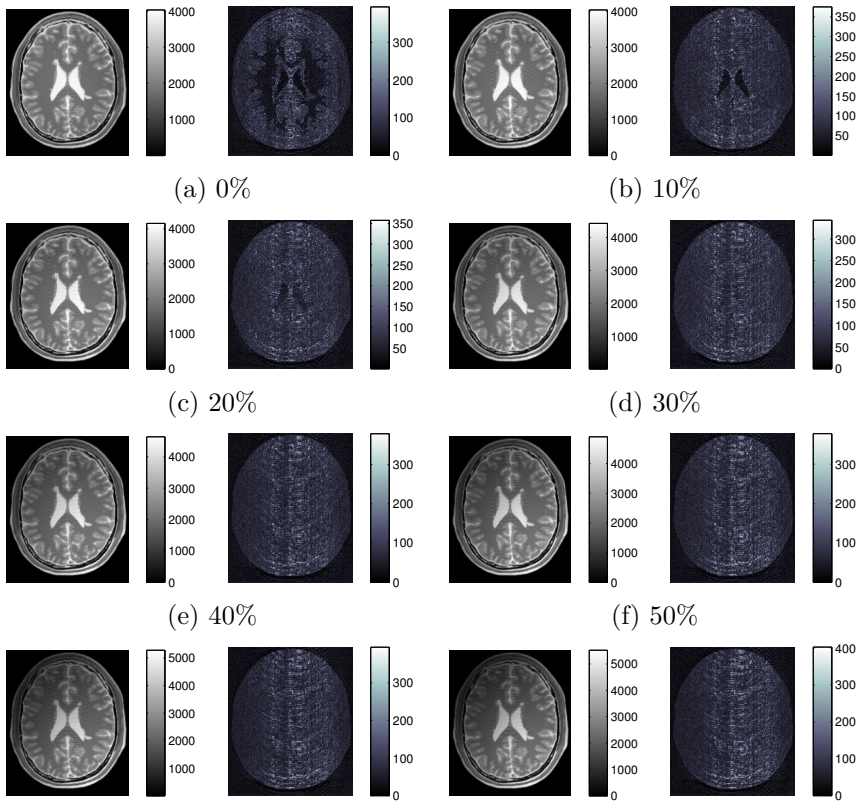


Figure C.10: Visual results for the experiment with different degrees of bias field variation. The sample mean of the magnitude of the reconstructed images images are shown in companion of the RMSE maps. Since, for each degree of variation, the ground-truth image intensity scale differ, the colorbar ranges of the RMSE maps are different.

Conclusion

MRI relaxometry is a very promising qMRI modality which aims at providing new insights into a very wide spectrum of human diseases, and also gives a better understanding of physiological processes of the human body. This PhD thesis has tried to emphasize the importance of this modality for current modern radiology. However, there is a large number of challenges in the current field of MRI relaxometry, and those should be carefully addressed before blindly relying on image biomarkers that are derived from T_1 or T_2 mapping. As is common in other qMRI modalities, overcoming these challenges is, by no means, an easy task. In this thesis, we have stated that, perhaps, the main reason why MRI relaxometry is not used routinely in the clinic is the lack of sufficient accuracy and reproducibility. The long acquisition protocol of T_1 and T_2 mapping is referred to as one of the main causes as well. Since qMRI is definitely an interdisciplinary field, where physicists, radiologists, engineers, and mathematicians can contribute, the resulting approaches for solving current problems are radically different.

In this PhD thesis, we have tried to address some of the problems that appear in MRI relaxometry by using an engineering approach, adopting a signal processing perspective. Indeed, while inaccurate physical modeling is often attributed as the main cause of lack of accuracy and reproducibility, little effort has been made on designing signal/image processing algorithms particularly tailored to MRI relaxometry. This is particularly surprising especially when MRI relaxometry is compared to even younger qMRI modalities, such as diffusion MRI (dMRI), where several groups have set the standards for accurate and precise qMRI estimation using a plethora of theoretically-grounded parameter estimation techniques and signal processing algorithms. As in dMRI, an important part of the errors that end up in unacceptable MRI relaxometry-based biomarkers can be attributed to the use of naïve and inadequate signal processing algorithms, and also standard, suboptimal statistical parameter methods. Further, the long protocol time of MRI relaxometry may be reduced by using sophisticated algorithms based on MRI image reconstruction theory, thereby avoiding further hardware development.

In chapter 5, we demonstrated that our joint Maximum Likelihood Estimator for simultaneous motion and T_1 estimation in T_1 mapping provides more accurate T_1 maps than those that are obtained with the conventional approach, in which image registration is performed prior to the T_1 estimation step. There are fundamentally two reasons for the better performance of our framework. First, when the relaxation model is integrated into the image registration problem, resulting in a model-based registration approach, the added temporal information leverages the motion

estimation step since we do not merely rely on intensity information but on the underlying model that is behind. Second, and importantly, accurate and precise T_1 mapping is doubtful when conventional two-step approaches are in play, even when the image registration has performed relatively well. The reason is that the image interpolation included in the image registration step will modify the statistical assumptions of the data. Deriving the resulting data distribution function is often impossible. Hence, optimal T_1 estimation is no longer possible. The alternative approach that we proposed is to include the relaxation model and the noise model into a global statistical model, which accounts for the data statistics. By doing so, the estimated T_1 map in MLE sense not only benefits from model-based approach registration but is also statistically optimal. As correctly noticed by other scientists, it can be argued that the unified approach ends up in a computationally quite demanding optimization problem, thereby encouraging practitioners to still adhere to the inaccurate conventional two-step technique. Nevertheless, the MLE framework was designed by paying attention to implementation efficiency, exploiting advanced methods of optimization theory, such as Majorize-Minimize algorithms, to transform a large optimization problem into a feasible optimization problem. These techniques may be applied even beyond MRI relaxometry. On top of that, the framework admits extensions to more complex motion and relaxation models, which allows modifications by other users as a "turn-the-crank" procedure.

Careful design of optimization algorithms for relaxometry usually pays off in terms of computational speed and statistical optimality, as shown in chapter 6. The optimization algorithm that we developed for VFA T_1 mapping, NOVIFAST, was proved to give the NLLS estimates, ML estimates in case of Gaussian noise, with an improvement in speed of, at least, twenty times compared to general purpose NLLS optimizers such as Levenberg-Marquardt and Gauss-Newton. This contribution exemplifies and highlights one of the main messages of this PhD thesis: it is often better to develop tailored and specific signal processing algorithms for a given qMRI modality, relaxometry in this case, rather than using general-purpose algorithms as black box procedures. Regarding NOVIFAST, we followed a fundamentally different approach compared to typical optimization toolboxes, such as those of Matlab, and we directly sought for the NLLS estimates by solving the non-linear equations that result from the first-order optimality conditions. Nothing seems exceptional to that point. However, during the derivation, equipped with very basic algebra rules, we made use of the specific structure that the SPGR model possesses, a quotient of rational functions, to end up with a semi-linear system of non-linear equations that could be solved tremendously fast in a very low number of iterations. With this work, our main message was to show that, with NOVIFAST, there are no reasons for VFA T_1 map practitioners to stay away from NLLS estimators due to their computational burden and initialization issues, and resort to fast, linear but suboptimal algorithms, such as DESPOT and IRWLS. As quoted in chapter 6, NOVIFAST can bring the best of both worlds: high accuracy in a very low computation time.

The last contribution of this work (chapter 7) has been a novel type of prior knowledge for accelerated MR image reconstruction with undersampled k-space data. The long acquisition time for reconstruction of an individual MR image is

the main handicap for MRI to become someday a real-time imaging modality. The acquisition time issue is even accentuated in qMRI, e.g., in relaxometry, due to the need to collect several MR images. The reconstruction method that we proposed in chapter 7, termed partial discreteness, compensates the lack of enough k-space data measurements with a priori information based on the spatial variation of the MR image intensity. In particular, partial discreteness exploits the fact that, in specific regions of MR images, the intensity pattern presents a much lower variation compared to those regions where the texture is really prominent. Paradigmatic cases are the background region as well as hyperintense areas that result when contrast-enhanced pulse sequences are in play. This information is modeled by using techniques from the machine-learning community, e.g., a Gaussian Mixture Model (GMM) to capture the salient features of what we called partially discrete images (those that adhere to our prior information). The proposed method competed very well with state-of-the-art reconstruction methods, with a radically different way to model prior information. With this work, we aimed at presenting a fresh approach for MRI reconstruction as well as making MRI reconstruction practitioners aware of the incredible potential that machine learning can bring in, a field that now is in vogue across not only scientists but also technologists.

Partial discreteness presents itself as a technique to accelerate the acquisition of individual MR images, and hence, is of undeniable utility for reducing the long protocol time of MRI relaxometry studies. We would like to note that, in the original paper, as well as in subsection 7.6, we mentioned the possibility to extend partial discreteness by incorporating information encoded in the fourth (temporal) dimension. Particularly, the GMM can be generalized to a multivariate mixture model, fully exploiting the temporal change of intensity that is based on the relaxometry model at hand. Similarly, the relaxation model can be explicitly included in the reconstruction method, thereby estimating the T_1 or T_2 spatial maps directly. Both extensions were left as future work, and they would have represented a truly integrated k-space reconstruction approach for MRI relaxometry. Hopefully, the ingredients that were included in the original partial discreteness method (chapter 7), as well as the small given outlook to perform such a desirable unified extension will be helpful in pushing the limits of MRI relaxometry.

List of Acronyms

This section lists (in alphabetic order) all the acronyms that have been used in this dissertation.

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
4D	four-dimensional
AL	Augmented Langrangian
BFGS	Broyden-Fletcher-Goldfarb-Shanno
CA	conventional approach
cBCD	cyclic block-coordinate descent
CRLB	Cramér Rao lower bound
CS	compressed sensing
CSF	cerebrospinal fluid
CT	computed tomography
DESPOT1	driven equilibrium single pulse observation
DFT	discrete Fourier transform
DTFT	discrete time Fourier transform
EPI	echo planar imaging
ETL	echo train length
FA	flip angle
FFT	fast Fourier transform
FID	free induction decay
FOV	field of view
FSE	fast spin echo
FSIM	feature similarity index metric
FWHM	full with at the half maximum

GE	gradient echo
GMM	Gaussian mixture model
GT	ground truth
GN	Gauss-Newton
GRAPPA	generalized autocalibrating partially parallel acquisitions
HFEN	high frequency error norm
IDFT	inverse discrete Fourier transform
IDTFT	inverse discrete time Fourier transform
IQR	interquartile range
IR	inversion recovery
IRWLLS	iterative re-weighted LLS
KKT	Karush-Kuhn-Tucker
LLS	linear least squares
LM	Levenberg-Marquardt
LORAKS	Low-rank modeling of local k-space neighborhoods
LR	low resolution
LS	least squares
MC	Monte Carlo
MI	mutual information
ML	maximum likelihood
MLE	maximum likelihood estimator
MM	majorize-minimize
MOLLI	modified look-locker IR
NOVIFAST	NOnlinear VarIable Flip Angle data baSed T1
MR	magnetic resonance
MRA	MR angiography
MRI	magnetic resonance imaging
MS	multiple sclerosis
MSE	mean squared error
NCG	nonlinear conjugate gradient
NLLS	nonlinear least squares
NLTV	nonlocal TV
NMR	nuclear magnetic resonance
NP-hard	nondeterministic polynomial time hard
PD	partial discreteness
PDF	probability density function
pMRI	parallel MRI
PSF	point spread function
PSNR	peak SNR

QILV	quantitative index based on local variance
qMRI	quantitative MRI
RF	radio frequency
RMSE	root MSE
ROI	region of interest
RW	random walk
SAR	specific absorption rate
SE	spin echo
SENSE	sensitivity encoding
SMASH	simultaneous acquisition of spatial harmonics
SMF	spatially matched filter
SNR	signal-to-noise ratio
SoS	sum of squares
SPGR	spoiled gradient recalled
SQS	separable quadratic surrogate
STI	short tau IR
SVD	singular value decomposition
TE	echo time
TI	inversion time
TR	repetition time
TSE	turbo spin echo
TV	total variation
UMVUE	uniformly minimum variance unbiased estimator
VFA	variable flip angle
ZF	zero filled

List of Symbols

An extensive, but not exhaustive, list of symbols used in this dissertation is included in this section. Symbols that are not listed are explained in the text or are expected to be understood directly from the context. In addition to a brief description, the equation where the symbols are first referred to is included. The symbols that have not been used in equations are left unlabeled.

Roman alphabet

$\mathbf{1}_x$, $\mathbf{1}_y$, $\mathbf{1}_z$	Eq. (1.10)	unitary vector along axis x, y, and z, respectively
$\boldsymbol{B_0}$	Eq. (1.3)	strong static external magnetic field
$\boldsymbol{B_1}(t)$	Eq. (1.9)	RF magnetic field pulse
$B_1^e(t)$	Eq. (1.9)	envelope of $\boldsymbol{B_1}(t)$
$\boldsymbol{B^R}(\boldsymbol{r})$	Eq. (1.25)	coil sensitivity field map
$\boldsymbol{B}_{Rot}$	Eq. (1.13)	external magnetic field in the RF-rotating frame
$\mathbb{C}^N$		complex-valued N-dimensional space
$\mathbb{C}^{N\times M}$		complex-valued $N \times M$-dimensional matrix space
$\boldsymbol{C_X}$	Eq. (2.26)	covariance matrix of random vector $\boldsymbol{X}$
diag$(\cdot)$		diagonal; diag$(\boldsymbol{x})$ is the matrix with vector $\boldsymbol{x}$ in the
	Eq. (2.46)	diagonal and zeros everywhere else,
	Eq. (6.5)	and diag$(\boldsymbol{A})$, where $\boldsymbol{A}$ is a matrix, is another matrix whose diagonal contains the diagonal entries of $\boldsymbol{A}$.
$D_{M,N}(\cdot,\cdot)$	Eq. (2.9)	two-dimensional Dirichlet kernel of order M and N
$\mathbb{E}\{\cdot\}$		expectation operator
e	Eq. (1.21)	mathematical constant that is the base of the natural logarithm (2.71828...)
$\hbar$	Eq. (1.1)	reduced Planck's constant ($1.05 \cdot 10^{-34} J \cdot s$)
$\boldsymbol{I}(\boldsymbol{\theta})$	Eq. (4.7)	Fisher information matrix of parameter $\boldsymbol{\theta}$
I	Eq. (1.1)	nuclear spin quantum number
$I_0(\cdot)$	Eq. (2.32)	zeroth order modified Bessel function of the first kind
$I_{L/2-1}(\cdot)$	Eq. (2.35)	$(L/2-1)$th order modified Bessel function of the first kind
$\boldsymbol{J}$	Eq. (1.1)	spin vector

K_b	Eq. (2.22)	Boltzmann constant
$\boldsymbol{k}(t)$	Eq. (1.35)	k-space trajectory
$\boldsymbol{M}$	Eq. (1.4)	net nuclear magnetization vector
$\boldsymbol{M_0}$	Eq. (1.4)	net nuclear magnetization vector at equilibrium
$\boldsymbol{M}_{Rot}$	Eq. (1.12)	net nuclear magnetization vector in the RF-rotating frame of reference
$\boldsymbol{M}_{Rot}^{t=0_+}$	Eq. (1.18)	$\boldsymbol{M}_{Rot}$ immediately after the excitation with $\boldsymbol{B_1}(t)$
$p_X(\cdot)$	Eq. (2.25)	PDF of random variable X
$p_{\boldsymbol{X}}(\cdot)$	Eq. (2.26)	PDF of random vector $\boldsymbol{X}$
$p_{X\|\mathcal{A}}(\cdot)$	Eq. (7.7)	PDF of random variable X, conditioned to event $\mathcal{A}$
R_c	Eq. (2.22)	coil resistance
R_s	Eq. (2.22)	induced resistance by conductive losses in the scanning object
$\boldsymbol{R}_p(\theta)$	Eq. (1.5)	rotation matrix along axis p with angle θ
$\mathbb{R}^{N\times M}$		real-valued $N \times M$-dimensional matrix space
$\mathbb{R}^N$		real-valued N-dimensional space
$\mathbb{R}_+^N$		$\mathbb{R}^N$ restricted to non-negative values
$\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$	Eq. (1.24)	real and imaginary parts operators
$\mathrm{rank}(\cdot)$	Eq. (2.57)	rank of a matrix
$\mathrm{sgn}(\cdot)$	Eq. (C.13)	sign function
T_1	Eq. (1.4)	spin-lattice or longitudinal relaxation time
T_2	Eq. (1.4)	spin-spin or transversal relaxation time
T_2^*	Eq. (1.22)	transversal relaxation time under the presence of magnetic inhomogeneities
W_x and W_y	Eq. (2.5)	FOV of a 2D image along axis x, and y, respectively
$\mathbb{Z}$	Eq. (2.1)	set of integer numbers

Greek alphabet

α	Eq. (1.17)	flip angle
γ	Eq. (1.2)	gyromagnetic ratio
Δk_x and Δk_y	Eq. (2.2)	sampling intervals of the k-space data along axis x and y, respectively
Δz		slice thickness
ω_L	Eq. (1.3)	Larmor angular frequency
ω_{RF}	Eq. (1.9)	RF pulse angular frequency
Ω	Eq. (1.25)	volume enclosed by the RF coil
$\boldsymbol{\mu}$	Eq. (1.2)	magnetic dipole moment
π	Eq. (1.35)	ratio of a circle's circumference to its diameter (3.14159...)
σ	Eq. (2.22)	standard deviation of thermal noise in MRI
τ_p	Eq. (1.15)	duration of the RF pulse

Other symbols

$\langle \cdot, \cdot \rangle$	Eq. (1.31)	inner product for vectors in $\mathbb{R}^N$
$\preccurlyeq$ (resp. $\succcurlyeq$)	Eq. (4.7)	lower (resp. greater) or equal to; $\boldsymbol{A} \preccurlyeq \boldsymbol{B}$ (resp. $\boldsymbol{A} \succcurlyeq \boldsymbol{B}$), where $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices, means that the matrix $\boldsymbol{A} - \boldsymbol{B}$ is semidefinite negative (resp. positive)
$\triangleq$	Eq. (A.14)	equal by definition
$\lvert \cdot \rvert$	Eq. (2.5)	$\lvert x \rvert$, where x is a real value, means the absolute value,
	Eq. (2.18)	whereas it represents the modulus of x when x is a complex value.
	Eq. (2.26)	We also use $\lvert \boldsymbol{A} \rvert$ to denote the determinant of the matrix $\boldsymbol{A}$
$\lvert\lvert \cdot \rvert\rvert_p$	Eq. (1.3)	l_p norm with $p > 0$, $\lvert\lvert \boldsymbol{x} \rvert\rvert_p = (\sum_n \lvert [\boldsymbol{x}]_n \rvert^p)^{\frac{1}{p}}$
$\lvert\lvert \cdot \rvert\rvert_0$	Eq. (2.53)	l_0 pseudo-norm $\lvert\lvert \boldsymbol{x} \rvert\rvert_0 = \#\{n : [\boldsymbol{x}]_n \neq 0\}$
$\lvert\lvert \cdot \rvert\rvert_{\boldsymbol{W}}$	Eq. (A.9)	weighted l_2 norm, $\lvert\lvert \boldsymbol{x} \rvert\rvert_{\boldsymbol{W}} = \lvert\lvert \boldsymbol{W}^{\frac{1}{2}} \boldsymbol{x} \rvert\rvert_2$
$\nabla(\cdot)$	Eq. (2.51)	gradient operator
$\circ$	Eq. (5.1)	Hadamard or point-wise product; $\boldsymbol{v} = \boldsymbol{a} \circ \boldsymbol{b}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors with N components, is another vector $\boldsymbol{v}$ defined as $[\boldsymbol{v}]_n = [\boldsymbol{a}]_n [\boldsymbol{b}]_n$, $n = 1, ..., N$

Curriculum Vitae

Gabriel Ramos Llordén was born on September 28, 1988, in Benavente, Spain. He holds an "Ingeniero de Telecomunicación" degree (five-year degree) as well as an MSc degree in Information and Communication Technology (ICT) research, with a major in biomedical engineering and signal processing, both degrees awarded by the University of Valladolid, Valladolid, Spain, in 2012 and 2013, respectively. In September 2013, he moved to Antwerp, Belgium, to pursue a PhD at imec-Vision Lab, Department of Physics, University of Antwerp, Antwerp, under the supervision of Prof. Dr. Jan Sijbers and Prof. Dr. Ir. Arnold J. den Dekker. Gabriel has developed novel signal processing algorithms for improved MR relaxometry data analysis as well as novel image reconstruction techniques to accelerate MRI acquisition. He was also an invited scholar at Harvard Medical School, Boston, USA, from January 2017 till April 2017, where he worked in diffusion MRI. Most of the results of Gabriel's research are collected in this dissertation.

Journal papers

1. **Ramos-Llordén, G.**, Vegas-Sánchez-Ferrero, G., Björk, M., Vanhevel, F., Parizel, P. M., San José Estépar, R., den Dekker, A. J., and Sijbers, J. NOVIFAST: a fast algorithm for accurate and precise VFA MRI T_1 mapping. *IEEE Trans. Med. Imag.,* in review.

2. **Ramos-Llordén, G.**, den Dekker, A. J., and Sijbers, J. (2017). Partial discreteness: a novel prior for magnetic resonance image reconstruction. *IEEE Trans. Med. Imag.*, 36(5):1041–1053.

3. **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Jeurissen, B., Vanhevel, F., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2017). A unified maximum likelihood framework for simultaneous motion and T_1 estimation in quantitative MR T_1 mapping. *IEEE Trans. Med. Imag.*, 36(2):433–466.

4. **Ramos-Llordén, G.**, Vegas-Sánchez-Ferrero, G., Martín-Fernández, M., Alberola-López, C., and Aja-Fernández, S. (2015). Anisotropic Diffusion Filter with Memory based on Speckle Statistics for Ultrasound Images. *IEEE Trans. Image Process.*, 24(1):345-358.

Book chapters

1. **Ramos-Llordén, G.**, Vegas-Sánchez-Ferrero, G., and Aja-Fernández, S. (2018). Chapter 5: Introduction to speckle filtering, *in Handbook of Speckle Filtering and Tracking in Cardiovascular Ultrasound Imaging and Video*, editors: Loizou, C. P., Pattichis, C. S., and D'Hooge, J. ISBN: 978-1-78561-290-9.

2. Aja-Fernández, S., **Ramos-Llordén, G.**, and Vegas-Sánchez-Ferrero, G. (2018). Chapter 8: Non-Linear Despeckle Filtering, *in Handbook of Speckle Filtering and Tracking in Cardiovascular Ultrasound Imaging and Video*, editors: Loizou, C. P., Pattichis, C. S., and D'Hooge, J. ISBN: 978-1-78561-290-9.

Conference papers

1. Bladt, P., Van Steenkiste, G., **Ramos-Llordén, G.**, den Dekker, A. J., and Sijbers, J. (2016). Multi-voxel algorithm for quantitative bi-exponential MRI T_1 estimation. In *Proc. SPIE 9784, Medical Imaging 2016: Image Processing, 978402*

2. **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2015). Simultaneous motion correction and T_1 estimation in quantitative T_1 mapping: an ML restoration approach. *In Proc. IEEE ICIP*, pages 3160–3164.

3. **Ramos-Llordén, G.**, Segers, H., Palestijn, W. J., den Dekker, A. J., and Sijbers, J. (2015). Partially discrete magnetic resonance tomography. In *Proc. IEEE ICIP*, pages 1653–1657.

4. Vegas-Sánchez-Ferrero, G., **Ramos-Llordén, G.**, de Luis-García, R., Tristán-Vega, A., and Aja-Fernández, S. (2013). Anisotropic Diffusion Filtering for Correlated Multiple-Coil MRI. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan.

5. **Ramos-Llordén, G.**, Vegas-Sánchez-Ferrero, G., Aja-Fernández, S., Martín-Fernández, M., and Alberola-López, C. (2013). Fast Anisotropic Speckle filter for Ultrasound Medical Images. In *13th Mediterranean Conf. Medical and Biological Engineering and Computing,* Sevilla, Spain.

6. **Ramos-Llordén, G.**, Vegas-Sánchez-Ferrero, G., Aja-Fernández, S., Martín-Fernández, M., and Alberola-López, C. (2012). Filtro de Difusión Anisótropo con Memoria basado en Modelos Probabilísticos para Imágenes Intravasculares y Cardíacas. In *Conferencia de la Asociación Española de Ingeniería Biomédica,* San Sebastián, Spain.

Conference abstracts

1. **Ramos-Llordén, G.**, Beirinckx, Q., den Dekker, A. J., and Sijbers, J. (2018). Accurate and precise MRI relaxometry: the often disregarded but critical role of statistical parameter estimation. In *26th annual meeting of the ISMRM*, Paris, France.

2. Jeurissen, B., **Ramos-Llordén, G.**, Vanhevel, F., Parizel, P. M., and Sijbers, J. (2018). Super-resolution for spherical deconvolution of multi-shell diffusion MRI data. In *26th annual meeting of the ISMRM*, Paris, France.

3. **Ramos-Llordén, G.**, Beirinckx, Q., den Dekker, A. J., and Sijbers, J. (2018). An educational presentation on accurate and precise MRI relaxometry: the often disregarded but critical role of statistical parameter estimation. In *10th annual of the ISMRM Benelux chapter*, Antwerp, Belgium.

4. **Ramos-Llordén, G.**, den Dekker, A. J., Bladt, P., Cuyt, A., and Sijbers, J. (2017). Statistically optimal separation of multi-component MR signals with a Majorize-Minimize approach: application to MWF estimation. In *34th annual scientific meeting of the ESMRMB*, Barcelona, Spain.

5. **Ramos-Llordén, G.**, den Dekker, A. J., Björk, M., Verhoye, M., and Sijbers, J. (2016). NOVIFAST: A fast non-linear least squares method for accurate and precise estimation of T_1 from SPGR signals. *Proc. Intl. Soc. Mag. Reson. Med.*, 24:2820.

6. **Ramos-Llordén, G.**, den Dekker, A. J., Björk, M., Verhoye, M., and Sijbers, J. (2016). NOVIFAST: A fast non-linear least squares method for accurate and precise estimation of T_1 from SPGR signals. In *8th meeting of the ISMRM Benelux chapter*, Eindhoven, The Netherlands.

7. **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M. and Sijbers, J. (2015). Simultaneous group-wise registration and maximum likelihood T_1 estimation for T_1 mapping. *Proc. Intl. Soc. Mag. Reson. Med.*, 23:447.

8. **Ramos-Llordén, G.**, Segers, H., Palestijn, W. J., den Dekker, A. J., and Sijbers, J. (2015). Partial discreteness: a new type of prior knowledge for MRI reconstruction. *Proc. Intl. Soc. Mag. Reson. Med.*, 23:3417.

9. **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M. and Sijbers, J. (2015). Simultaneous group-wise registration and maximum likelihood T_1 estimation for T_1 mapping. In *7th meeting of the ISMRM Benelux chapter*, Gent, Belgium.

10. **Ramos-Llordén, G.**, den Dekker, A. J., and Sijbers, J. (2014). Joint Motion Correction and Estimation for T_1 mapping: proof of concept. *Proceedings of the workshop "Medical Imaging Summer School"*, https://doi.org/10.13140/rg.2.1.1859.0167 .

11. **Ramos-Llordén, G.** and Sijbers, J. (2013). Misalignment correction for T1 maps using a maximum likelihood estimator approach. *Front. Neuroinform.*

Conf. Abstract: Imaging the brain at different scales: How to integrate multi-scale structural information?, Antwerp, Belgium.

Awards

1. *Winner of the Abroad Internship Challenge (5000 Euro) awarded by the ISMRM Benelux Chapter (10th Annual Meeting)*, 26th January 2018

2. *FWO travel grant (ESMRMB 2017, Barcelona, Spain)* for the work: **Ramos-Llordén, G.**, den Dekker, A. J., Bladt, P., Cuyt, A., and Sijbers, J. (2017). Statistically optimal separation of multi-component MR signals with a Majorize-Minimize approach: application to MWF estimation. In *34th annual scientific meeting of the ESMRMB*, Barcelona, Spain.

3. *FWO travel grant for long stay abroad from January 2017 till April 2017*, Laboratory of Mathematics in Imaging, (LMI), Brigham and Women's Hospital, Harvard Medical School, Boston, USA.

4. *FWO travel grant (ISMRM 2016, Singapore, Singapore)* for the work: **Ramos-Llordén, G.**, den Dekker, A. J., Björk, M., Verhoye, M., and Sijbers, J. (2016). NOVIFAST: A fast non-linear least squares method for accurate and precise estimation of T_1 from SPGR signals. *Proc. Intl. Soc. Mag. Reson. Med.*, 24:2820.

5. *Educational Stipend Award (ISMRM 2016, Singapore, Singapore)* for the work: **Ramos-Llordén, G.**, den Dekker, A. J., Björk, M., Verhoye, M., and Sijbers, J. (2016). NOVIFAST: A fast non-linear least squares method for accurate and precise estimation of T_1 from SPGR signals. *Proc. Intl. Soc. Mag. Reson. Med.*, 24:2820.

6. *Magna Cum Laude Merit Award (ISMRM 2015, Toronto, Canada)* for the work: **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2015). Simultaneous group-wise registration and maximum likelihood T_1 estimation for T_1 mapping. *Proc. Intl. Soc. Mag. Reson. Med.*, 23:447.

7. *FWO travel grant (ISMRM 2015, Toronto, Canada)* for the work: **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2015). Simultaneous group-wise registration and maximum likelihood T_1 estimation for T_1 mapping. *Proc. Intl. Soc. Mag. Reson. Med.*, 23:447.

8. *Educational Stipend Award (ISMRM 2015, Toronto, Canada)* for the work: **Ramos-Llordén, G.**, den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2015). Simultaneous group-wise registration and maximum likelihood T_1 estimation for T_1 mapping. *Proc. Intl. Soc. Mag. Reson. Med.*, 23:447.

9. *2012 National Award for the best Spanish MSc thesis in Engineering and Medicine* for the work: **Ramos-Llordén, G**, Vegas-Sánchez-Ferrero, G., Martín-Fernández, M., Alberola-López, C., and Aja-Fernández, S. (2015). Anisotropic Diffusion Filter with Memory based on Speckle Statistics for Ultrasound Images. *IEEE Trans. Image Process.*, 24(1):345-358.

Research stays

1. Laboratory of Mathematics in Imaging, (LMI), Brigham and Women's Hospital, Harvard Medical School, Boston, USA, from 15 January 2017 to 15 April 2017, supervised by Prof. Dr. Carl-Fredrik Westin and Prof. Dr. Yogesh Rathi.

Invited talks at research groups

1. *Statistical reconstruction methods for optimal analysis of relaxometry data and k-space under-sampling*, Laboratory of Mathematics in Imaging, (LMI), Brigham and Women's Hospital, Harvard Medical School, Boston, USA, April 2017

Teaching and supervision

1. Digital signal and image processing (Digitale signaal-en beeldverwerking), Exercises and Laboratory Lessons, M.Sc. in Physics, University of Antwerp, (Supervisor: Prof. Dr. Jan Sijbers)
2. Co-supervisor of Piet Bladt (thesis MSc in Physics, University of Antwerp): *Quantitative multi-component T_1 mapping* (Supervisor: Prof. Dr. Jan Sijbers)

Relevant courses and workshops

1. Weekend Educational Program organized by the International Society for Magnetic Resonance in Medicine, Toronto, Singapore and Hawaii, 2015-2017
2. Workshop on processing multi-shell diffusion MRI data using Mrtrix3, Antwerp, Belgium, September 23, 2015
3. Sparse Stochastic Processes: A unifying statistical framework for modern image processing, by Michael Unser, ICIP 2015, Quebec, Canada.
4. Medical Imaging Summer School 2014, endorsed by MICCAI Society and SPRINGER, Favignana, Sicily, Italy, July 2014

Scientific reviewer for the following journals

1. Neuroimage
2. Medical Image Analysis
3. IEEE Transactions on Medical Imaging
4. Journal of Magnetic Resonance Imaging, Elsevier
5. IEEE Signal Processing Letters
6. Physica Medica: European Journal of Medical Physics
7. Pattern Recognition Letters, Elsevier
8. IEEE Journal of Biomedical and Health Informatics
9. Ultrasonics, Elsevier
10. Anais da Academia Brasileira de Ciências
11. Journal of Applied Remote Sensing
12. IETE Technical Review

Scientific reviewer for the following conferences

1. Joint annual meeting ISMRM-ESMRMB 2018, Paris, France

Bibliography

General electric healthcare: SIGNA TM premier. http://www3.gehealthcare.com/en/products/categories/magnetic_resonance_imaging/3-0t/signa_premier.

Philips magnetic resonance: Ingenia 3.0 t CX MR system. https://www.usa.philips.com/healthcare/product/HC781271/ingenia-30t-cx-mr-system.

Siemens healthineers: Magnetom prisma the 3t powerpack for exploration.: Technical details. https://www.healthcare.siemens.com/magnetic-resonance-imaging/3t-mri-scanner/magnetom-prisma/technical-details.

Siemens healthineers: Magnetom trio, a tim system: Technical details. https://www.healthcare.siemens.com/magnetic-resonance-imaging/for-installed-base-business-only-do-not-publish/magnetom-trio-tim/technical-details.

Addy, N. O., Wu, H. H., and Nishimura, D. G. (2012). Simple method for MR gradient system characterization and k-space trajectory estimation. *Magn. Reson. Med.*, 68(1):120–129.

Aelterman, J., Luong, H. Q., Goossens, B., Pižurica, A., and Philips, W. (2011). Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data. *Signal Processing*, 91(12):2731–2742.

Ahn, C. B., Kim, J. H., and Cho, Z. H. (1986). High-speed spiral-scan echo planar NMR imaging-i. *IEEE Trans. Med. Imag.*, 5(1):2–7.

Aja-Fernández, S., Alberola-López, C., and Westin, C.-F. (2008). Noise and signal estimation in magnitude MRI and rician distributed images: a LMMSE approach. *IEEE Trans. Image Process.*, 17(8):1383–1398.

Aja-Fernández, S., Estépar, R., Alberola-López, C., and Westin, C.-F. (2006). Image quality assessment based on local variance. In *28th Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pages 4815–4818.

Aja-Fernández, S., Pieciak, T., and Vegas-Sánchez-Ferrero, G. (2015a). Spatially variant noise estimation in MRI: A homomorphic approach. *Med. Image Anal.*, 20(1):184 – 197.

Aja-Fernández, S., Pieciak, T., and Vegas-Sánchez-Ferrero, G. (2015b). Spatially variant noise estimation in MRI: A homomorphic approach. *Med. Image Anal.*, 20(1):184–197.

Aja-Fernández, S. and Tristán-Vega, A. (2013). A review on statistical noise models for magnetic resonance imaging. Technical report, Univ. Valladolid, Valladolid, Spain. Tech. Rep. LPI2013-01.

Aja-Fernández, S. and Vegas-Sánchez-Ferrero, G. (2016a). *Statistical Analysis of Noise in MRI: Modeling, Filtering and Estimation*. 1st ed. Springer International Publishing AG Springer Science.

Aja-Fernández, S. and Vegas-Sánchez-Ferrero, G. (2016b). *Statistical Analysis of Noise in MRI: Modeling, Filtering and Estimation*. 1st ed. Springer International Publishing AG Springer Science.

Aja-Fernández, S., Vegas-Sánchez-Ferrero, G., and Tristán-Vega, A. (2014). Noise estimation in parallel MRI: GRAPPA and SENSE. *Magn. Reson. Imaging*, 32(3):281–290.

Allison, M. J., Ramani, S., and Fessler, J. A. (2012). Regularized MR coil sensitivity estimation using augmented lagrangian methods. In *Proc. IEEE Int. Symp. Biomed. Imag.*, pages 394–397. IEEE.

Allison, M. J., Ramani, S., and Fessler, J. A. (2013). Accelerated regularized estimation of MR coil sensitivities using augmented lagrangian methods. *IEEE Trans. Med. Imaging*, 32(3):556–564.

Altbach, M. I., Outwater, E. K., Trouard, T. P., Krupinski, E. A., Theilmann, R. J., Stopeck, A. T., Kono, M., and Gmitro, A. F. (2002). Radial fast spin-echo method for T2-weighted imaging and T2 mapping of the liver. *J. Magn. Reson. Imaging*, 16(2):179–189.

Armijo, L. (1966). Minimization of functions having lipschitz continuous first partial derivatives. *Pacific J. Math.*, 16(1):1–3.

Awate, S. P., Tasdizen, T., Foster, N., and Whitaker, R. T. (2006). Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification. *Med. Image Anal.*, 10(5):726 – 739.

Barbosa, S., Blumhardt, L., Roberts, N., Lock, T., and Edwards, R. (1994). Magnetic resonance relaxation time mapping in multiple sclerosis: Normal appearing white matter and the invisible lesion load. *Magn. Reson. Imag.*, 12(1):33 – 42.

Barral, J. K., Gudmundson, E., Stikov, N., Etezadi-Amoli, M., Stoica, P., and Nishimura, D. G. (2010). A robust methodology for in vivo T1 mapping. *Magn. Reson. Med.*, 64(4):1057–1067.

Batenburg, K. J., Bals, S., Sijbers, J., Kübel, C., Midgley, P., Hernandez, J., Kaiser, U., Encina, E., Coronado, E., and Van Tendeloo, G. (2009). 3D imaging of nanomaterials by discrete tomography. *Ultramicroscopy*, 109(6):730–740.

Batenburg, K. J. and Sijbers, J. (2011). DART: a practical reconstruction algorithm for discrete tomography. *IEEE Trans. Image Process.*, 20(9):2542–2553.

Beatty, P. J. (2006). *Reconstruction Method for Fast Magnetic Resonance Imaging*. PhD thesis, Stanford University.

Beatty, P. J., Nishimura, D. G., and Pauly, J. M. (2005). Rapid gridding reconstruction with a minimal oversampling ratio. *IEEE Trans. Med. Imag.*, 24(6):799–808.

Bernstein, M. A., King, K. F., and Zhou, X. J. (2004). *Handbook of MRI pulse sequences*. 1st ed. Burlington, MA, USA; Elsevier.

Bertsekas, D. (1975). Necessary and sufficient conditions for a penalty method to be exact. *Math. Program*, 9(1):87–99.

Bertsekas, D. P. (1982). Projected Newton methods for optimization problems with simple constraints. *SIAM J. Control Optim.*, 20(2):221–246.

Bertsimas, D. and Nohadani, O. (2010). Robust optimization with simulated annealing. *J. Global Optim.*, 48(2):323–334.

Besson, J., Corrigan, F. M., Foreman, E. I., Eastwood, L. M., Smith, F. W., and Ashcroft, G. W. (1985). NMR imaging in dementia. In *Modern Approaches to the Dementias*, volume 20, pages 71–79. Karger Publishers.

Birgin, E. J. G. and Martínez, J. M. (2012). Augmented Lagrangian method with nonmonotone penalty parameters for constrained optimization. *Comput. Optim. Appl.*, 51(3):941–965.

Björk, M., Zachariah, D., Kullberg, J., and Stoica, P. (2016). A multicomponent T2 relaxometry algorithm for myelin water imaging of the brain. *Magn. Reson. Med.*, 75(1):390–402.

Blaimer, M., Breuer, F., Mueller, M., Heidemann, R. M., Griswold, M. A., and Jakob, P. M. (2004). SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method. *Top. Magn. Reson. Imaging*, 15(4):223–236.

Blaimer, M., Heim, M., Neumann, D., Jakob, P. M., Kannengiesser, S., and Breuer, F. A. (2016). Comparison of phase-constrained parallel MRI approaches: Analogies and differences. *Magn. Reson. Med.*, 75(3):1086–1099.

Bloch, F. (1946). Nuclear induction. *Phys. Rev.*, 70(7-8):460.

Block, K. T., Uecker, M., and Frahm, J. (2007). Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magn. Reson. Med.*, 57(6):1086–1098.

Block, K. T., Uecker, M., and Frahm, J. (2009). Model-based iterative reconstruction for radial fast spin-echo MRI. *IEEE Trans. Med. Imag.*, 28(11):1759–1769.

Bojorquez, J. Z., Bricq, S., Acquitter, C., Brunotte, F., Walker, P. M., and Lalande, A. (2017). What are normal relaxation times of tissues at 3 T? *Magn. Reson. Imag.*, 35:69–80.

Bojorquez, J. Z., Bricq, S., Brunotte, F., Walker, P. M., and Lalande, A. (2016). A novel alternative to classify tissues from T1 and T2 relaxation times for prostate MRI. *Magn. Reson. Mater. Phy.*, 29(5):777–788.

Bottomley, P. A., Redington, R. W., Edelstein, W. A., and Schenck, J. F. (1985). Estimating radiofrequency power deposition in body NMR imaging. *Magn. Reson. Med.*, 2(4):336–349.

Bouhrara, M. and Spencer, R. G. (2017). Rapid simultaneous high-resolution mapping of myelin water fraction and relaxation times in human brain using BMC-mcDESPOT. *NeuroImage*, 147:800–811.

Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. and Math. Phys.*, 7:200–217.

Bron, E. E., van Tiel, J., Smit, H., Poot, D. H. J., Niessen, W. J., Krestin, G. P., Weinans, H., Oei, E. H., Kotek, G., and Klein, S. (2013). Image registration improves human knee cartilage T1 mapping with delayed gadolinium-enhanced MRI of cartilage (dGEMRIC). *Eur. Radiol.*, 23(1):246–252.

Brown, R., Wang, Y., Spincemaille, P., and Lee, R. F. (2007). On the noise correlation matrix for multiple radio frequency coils. *Magn. Reson. Med.*, 58(2):218–224.

Burden, R. L. and Faires, J. D. (2010). *Numerical Analysis*. 9th ed. Boston, MA: Cengage Learning.

Bydder, M. and Robson, M. D. (2005). Partial Fourier partially parallel imaging. *Magn. Reson. Med.*, 53(6):1393–1401.

Caballero, J., Price, A. N., Rueckert, D., and Hajnal, J. V. (2014). Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Trans. Med. Imag.*, 33(4):979–994.

Cai, J.-F., Osher, S., and Shen, Z. (2009). Split Bregman methods and frame based image restoration. *Multiscale Model Simul.*, 8(2):337–369.

Campbell-Washburn, A. E., Xue, H., Lederman, R. J., Faranesh, A. Z., and Hansen, M. S. (2016). Real-time distortion correction of spiral and echo planar images using the gradient system impulse response function. *Magn. Reson. Med.*, 75(6):2278–2285.

Candes, E., Demanet, L., Donoho, D., and Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Model Simul.*, 5(3):861–899.

Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30.

Cao, Y. and Levin, D. N. (1995). Using an image database to constrain the acquisition and reconstruction of MR images of the human head. *IEEE Trans. Med. Imag.*, 14(2):350–361.

Caravan, P., Ellison, J. J., McMurry, T. J., and Lauffer, R. B. (1999). Gadolinium (iii) chelates as MRI contrast agents: structure, dynamics, and applications. *Chem. Rev.*, 99(9):2293–2352.

Caravan, P., Farrar, C. T., Frullano, L., and Uppal, R. (2009). Influence of molecular parameters and increasing magnetic field strength on relaxivity of gadolinium-and manganese-based T1 contrast agents. *Contrast Media Mol. Imaging*, 4(2):89–100.

Carr, H. Y. and Purcell, E. M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.*, 94(3):630–638.

Chang, L.-C., Koay, C., Deoni, S., and Pierpaoli, C. (2007). Comparison of linear and non-linear fitting methods for estimating T1 from SPGR signals. *Proc. Intl. Soc. Mag. Reson. Med.*, 15:1775.

Chang, L.-C., Koay, C. G., Basser, P. J., and Pierpaoli, C. (2008). Linear least-squares method for unbiased estimation of T1 from SPGR signals. *Magn. Reson. Med.*, 60(2):496–501.

Chartrand, R. (2009). Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *Proc. IEEE Int. Symp. Biomed. Imag.*, pages 262–265.

Chavez, S. and Stanisz, G. J. (2012). A novel method for simultaneous 3D B1 and T1 mapping: the method of slopes (MoS). *NMR in Biomedicine*, 25(9):1043–1055.

Chavhan, G. B., Babyn, P. S., Thomas, B., Shroff, M. M., and Haacke, E. M. (2009). Principles, techniques, and applications of T2*-based MR imaging and its special applications. *Radiographics*, 29(5):1433–1449.

Chen, C. and Huang, J. (2012). Compressive sensing MRI with wavelet tree sparsity. In *Adv. Neural Inf. Process Syst.*, pages 1115–1123.

Chen, C. and Huang, J. (2014). The benefit of tree sparsity in accelerated MRI. *Med. Image Anal.*, 18(6):834 – 842.

Chen, J. and Li, W. (2005). Convergence of Gauss-Newton's method and uniqueness of the solution. *Appl. Math. Comput.*, 170(1):686 – 705.

Cheng, H.-L. M. (2007). T1 measurement of flowing blood and arterial input function determination for quantitative 3D T1-weighted DCE-MRI. *J. Magn. Reson. Imaging*, 25(5):1073–1078.

Cheng, H.-L. M. and Wright, G. A. (2006). Rapid high-resolution T1 mapping by variable flip angles: Accurate and precise measurements in the presence of radiofrequency field inhomogeneity. *Magn. Reson. Med.*, 55(3):566–574.

Cheung, K. F. (1993). A multidimensional extension of Papoulis generalized sampling expansion with the application in minimum density sampling. In *Advanced topics in Shannon sampling and interpolation theory*, pages 85–119. Springer.

Christensen, K. A., Grant, D. M., Schulman, E. M., and Walling, C. (1974). Optimal determination of relaxation times of fourier transform nuclear magnetic resonance. determination of spin-lattice relaxation times in chemically polarized species. *J. Phys. Chem.*, 78(19):1971–1977.

Clare, S. and Jezzard, P. (2001). Rapid T1 mapping using multislice echo planar imaging. *Magn. Reson. Med.*, 45(4):630–634.

Cocosco, C., Kollokian, V., Kwan, R.-S., and Evans, A. (1997). Brainweb: Online interface to a 3D MRI simulated brain database. *Neuroimage*, 5(4).

Condat, L., Van De Ville, D., and Forster-Heinlein, B. (2008). Reversible, fast, and high-quality grid conversions. *IEEE Trans. Image Process.*, 17(5):679–693.

Conlon, P., Trimble, M. R., Rogers, D., and Callicott, C. (1988). Magnetic resonance imaging in epilepsy: a controlled study. *Epilepsy Res.*, 2(1):37–43.

Constantinides, C., Atalar, E., and McVeigh, E. (1997). Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magn. Reson. Med.*, 38(5):852–857.

Cordero-Grande, L., Teixeira, R. P. A. G., Hughes, E. J., Hutter, J., Price, A. N., and Hajnal, J. (2016). Sensitivity encoding for aligned multishot magnetic resonance reconstruction. *IEEE Trans. Comput. Imaging*, 2(3):266–280.

Cowan, B. (1997). *Nuclear Magnetic Resonance and Relaxation*. 1st ed. New York City, NY, USA: Cambridge University Press.

Dabravolski, A., Batenburg, K. J., and Sijbers, J. (2014). A multiresolution approach to discrete tomography using DART. *PLoS ONE*, 9(9):e106090.

Damadian, R. (1971). Tumor detection by nuclear magnetic resonance. *Science*, 171(3976):1151–1153.

Dan, H., Yamashita, N., and Fukushima, M. (2002). Convergence properties of the inexact Levenberg-Marquardt method under local error bound conditions. *Optim. Methods Softw.*, 17(4):605–626.

De Bazelaire, C. M., Duhamel, G. D., Rofsky, N. M., and Alsop, D. C. (2004). MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0 T: preliminary results. *Radiology*, 230(3):652–659.

den Dekker, A. J. and Sijbers, J. (2014). Data distributions in magnetic resonance images: A review. *Phys. Medica*, 30(7):725–741.

Deoni, S. C., Peters, T. M., and Rutt, B. K. (2004). Determination of optimal angles for variable nutation proton magnetic spin-lattice, T1, and spin-spin, T2, relaxation times measurement. *Magn. Reson. Med.*, 51(1):194–199.

Deoni, S. C., Rutt, B. K., and Peters, T. M. (2003). Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. *Magn. Reson. Med.*, 49(3):515–526.

Deoni, S. C. L. (2010). Quantitative relaxometry of the brain. *Top. Magn. Reson. Imaging*, 21(2):101.

Detre, J., Leigh, J., Williams, D., and Koretsky, A. (1992). Perfusion imaging. *Magn. Reson. Med.*, 23(1):37–45.

DeVore, M. D., Lanterman, A. D., and O'Sullivan, J. A. (2000). ATR performance of a Rician model for SAR images. In *AeroSense 2000*, pages 34–45. International Society for Optics and Photonics.

DeWitt, L. D., Kistler, J. P., Miller, D. C., Richardson, E. P., and Buonanno, F. S. (1987). NMR-neuropathologic correlation in stroke. *Stroke*, 18(2):342–351.

Dietrich, O., Freiermuth, M., Willerding, L., Reiser, M. F., and Peller, M. (2015). Flip angle–optimized fast dynamic T1 mapping with a 3D gradient echo sequence. *Magn. Reson. Med.*, 73(3):1158–1163.

Dimitrov, S. and Kamenski, D. (1991). A parameter estimation method for rational functions. *Comput. Chem. Eng.*, 15(9):657–662.

Dimitrov, S. and Kamenski, D. (1996). Parameter estimation in complicated rational functions. *Computers & Chemistry*, 20(3):331 – 337.

Drain, L. E. (1949). A direct method of measuring nuclear spin-lattice relaxation times. *Proceedings of the Physical Society. Section A*, 62(5):301.

Drayer, B. P., Olanow, W., Burger, P., Johnson, G. A., Herfkens, R., and Riederer, S. (1986). Parkinson plus syndrome: diagnosis using high field MR imaging of brain iron. *Radiology*, 159(2):493–498.

Duchi, J. (2016). *Chapter 8: Fisher Information, in Lecture Notes for Statistics 311/Electrical Engineering 377*. Stanford University.

Ebmeier, K. P., Besson, J. A. O., Crawford, J. R., Palin, A. N., Gemmel, H. G., Sharp, P. F., Cherryman, G. R., and Smith, F. W. (1987). Nuclear magnetic resonance imaging and single photon emission tomography with radio-iodine labelled compounds in the diagnosis of dementia. *Acta Psychiat. Scand.*, 75(5):549–556.

Edelstein, W. A., Hutchison, J. M. S., Johnson, G., and Redpath, T. (1980). Spin warp NMR imaging and applications to human whole-body imaging. *Phys. Med. Biol.*, 25(4):751.

Eglese, R. (1990). Simulated annealing: a tool for operational research. *Eur. J. Oper. Res.*, 46(3):271–281.

Englund, E., Brun, A., Györffy-Wagner, Z., Larsson, E.-M., and Persson, B. (1986). Relaxation times in relation to grade of malignancy and tissue necrosis in astrocytic gliomas. *Magn. Reson. Imag.*, 4(5):425–429.

Erkinjuntti, T., Ketonen, L., Sulkava, R., Sipponen, J., Vuorialho, M., and Iivanainen, M. (1987). Do white matter changes on MRI and CT differentiate vascular dementia from Alzheimer's disease? *J. Neurol. Neurosurg. Psychiatry*, 50(1):37–42.

Estellers, V., Zosso, D., Lai, R., Osher, S., Thiran, J.-P., and Bresson, X. (2012). Efficient algorithm for level set method preserving distance function. *IEEE Trans. Image Process.*, 21(12):4722–4734.

Fan, A., Wells, W. M., Fisher, J. W., Cetin, M., Haker, S., Mulkern, R., Tempany, C., and Willsky, A. S. (2003). A unified variational approach to denoising and bias correction in MR. In *Inf. Process Med. Imaging*, volume 2732, pages 148–159. Springer.

Fan, J., Farmen, M., and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *J. R. Stat. Soc. Series B. Stat. Methodol.*, 60(3):591–608.

Feng, L., Grimm, R., Block, K. T., Chandarana, H., Kim, S., Xu, J., Axel, L., Sodickson, D. K., and Otazo, R. (2014). Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magn. Reson. Med.*, 72(3):707–717.

Feng, W., Neelavalli, J., and Haacke, E. M. (2013). Catalytic multiecho phase unwrapping scheme (CAMPUS) in multiecho gradient echo imaging: Removing phase wraps on a voxel-by-voxel basis. *Magn. Reson. Med.*, 70(1):117–126.

Feng, X., Salerno, M., Kramer, C. M., and Meyer, C. H. (2016). Non-cartesian balanced steady-state free precession pulse sequences for real-time cardiac MRI. *Magn. Reson. Med.*, 75(4):1546–1555.

Fennessy, F. M., Fedorov, A., Gupta, S. N., Schmidt, E. J., Tempany, C. M., and Mulkern, R. V. (2012). Practical considerations in T1 mapping of prostate for dynamic contrast enhancement pharmacokinetic analyses. *Magn. Reson. Imag.*, 30(9):1224–1233.

Fessler, J. (2010). Model-based image reconstruction for MRI. *IEEE Signal Process. Mag.*, 27(4):81–89.

Fessler, J. A. and Kim, D. (2011). Axial block coordinate descent (ABCD) algorithm for X-ray CT image reconstruction. In *Proc. Intl. Mtg. on Fully 3D image Recon. in Rad. and Nuc. Med.*, pages 262–5.

Fessler, J. A. and Sutton, B. P. (2003). Nonuniform fast fourier transforms using min-max interpolation. *IEEE Trans. Sig. Process.*, 51(2):560–574.

Figueiredo, M. A., Nowak, R. D., and Wright, S. J. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Topics Signal Process.*, 1(4):586–597.

Figueiredo, M. A. T. and Jain, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(3):381–396.

Foerster, B. U., Tomasi, D., and Caparelli, E. C. (2005). Magnetic field shift due to mechanical vibration in functional magnetic resonance imaging. *Magn. Reson. Med.*, 54(5):1261–1267.

Fogtmann, M., Seshamani, S., Kroenke, C., Cheng, X., Chapman, T., Wilm, J., Rousseau, F., and Studholme, C. (2014). A unified approach to diffusion direction sensitive slice registration and 3-D DTI reconstruction from moving fetal brain anatomy. *IEEE Trans. Med. Imag.*, 33(2):272–289.

Foucart, S. and Rauhut, H. (2010). *A mathematical introduction to compressive sensing*, volume 1. 1st ed. New York City, NY, USA: Springer New York Heidelberg.

Fram, E. K., Herfkens, R., Johnson, G., Glover, G., Karis, J., Shimakawa, A., Perkins, T., and Pelc, N. (1987). Rapid calculation of T1 using variable flip angle gradient refocused imaging. *Magn. Reson. Imaging*, 5(3):201–208.

Gholipour, A., Estroff, J. A., and Warfield, S. K. (2010). Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. *IEEE Trans. Med. Imag*, 29(10):1739–1758.

Gilboa, G. and Osher, S. (2007). Nonlocal linear image regularization and supervised segmentation. *Multiscale Model Simul.*, 6(2):595–630.

Gilboa, G. and Osher, S. (2008). Nonlocal operators with applications to image processing. *Multiscale Model Simul.*, 7(3):1005–1028.

Gindi, G., Lee, M., Rangarajan, A., and Zubal, I. G. (1993). Bayesian reconstruction of functional images using anatomical information as priors. *IEEE Trans. Med. Imag.*, 12(4):670–680.

Glover, G. H. and Pauly, J. M. (1992). Projection reconstruction techniques for reduction of motion effects in MRI. *Magn. Reson. Med.*, 28(2):275–289.

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *J. Econom.*, 60(1-2):65–99.

Gold, G. E., Han, E., Stainsby, J., Wright, G., Brittain, J., and Beaulieu, C. (2004). Musculoskeletal MRI at 3.0 T: relaxation times and image contrast. *Am. J. Roentgenol.*, 183(2):343–351.

Goldstein, H., Poole Jr., C. P., and Safko, J. L. (2014). *Classical Mechanics*. Harlow, Essex, England: Pearson Higher Ed.

Goldstein, T. and Osher, S. (2009). The split Bregman method for L1-regularized problems. *SIAM J. Imaging. Sci.*, 2(2):323–343.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., and Haase, A. (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.*, 47(6):1202–1210.

Gudbjartsson, H. and Patz, S. (1995). The Rician distribution of noisy MRI data. *Magn. Reson. Med.*, 34(6):910–914.

Gupta, R. (1977). A new look at the method of variable nutation angle for the measurement of spin-lattice relaxation times using fourier transform NMR. *J. Magn. Reson.*, 25(1):231–235.

Hahn, E. L. (1949). An accurate nuclear magnetic resonance method for measuring spin-lattice relaxation times. *Phys. Rev.*, 76(1):145.

Hahn, E. L. (1950). Spin echoes. *Phys. Rev.*, 80:580–594.

Haldar, J. P. LORAKS: Implementation and examples. `http://mr.usc.edu/download/LORAKS`.

Haldar, J. P. (2014a). Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. *IEEE Trans. Med. Imag.*, 33(3):668–681.

Haldar, J. P. (2014b). Low-rank modeling of local k-space neighborhoods (LORAKS): Implementation and examples for reproducible research. Technical Report USC-SIPI-414, University of Southern California, Los Angeles, CA.

Haldar, J. P., Hernando, D., and Liang, Z. P. (2011). Compressed-sensing MRI with random encoding. *IEEE Trans. Med. Imag.*, 30(4):893–903.

Haldar, J. P., Hernando, D., Song, S.-K., and Liang, Z.-P. (2008). Anatomically constrained reconstruction from noisy data. *Magn. Reson. Med.*, 59(4):810–818.

Haldar, J. P. and Zhuo, J. (2015). P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data. *Magn. Reson. Med.*

Hallack, A., Chappel, M. A., Gooding, M., and Schnabel, J. (2014). A new similarity metric for groupwise registration of variable flip angle sequences for improved T_{10} estimation in DCE-MRI. In *Lect. Notes Comput. Sc., Biomedical Image Registration*, volume 8545, pages 154–163.

Han, E., Gold, G., Stainsby, J., Wright, G., Beaulieu, C., and Brittain, J. (2003). In-vivo T1 and T2 measurements of muskuloskeletal tissue at 3T and 1.5 T. *Proc. Intl. Soc. Mag. Reson. Med.*, 11:450.

Han, M., Chiba, K., Banerjee, S., Carballido-Gamio, J., and Krug, R. (2015). Variable flip angle three-dimensional fast spin-echo sequence combined with outer volume suppression for imaging trabecular bone structure of the proximal femur. *J. Magn. Reson. Imaging*, 41(5):1300–1310.

Hanson, J. M., Liang, Z.-P., Wiener, E. C., and Lauterbur, P. C. (1996). Fast dynamic imaging using two reference images. *Magn. Reson. Med.*, 36(1):172–175.

Hayes, C. E. and Roemer, P. B. (1990). Noise correlations in data simultaneously acquired from multiple surface coil arrays. *Magn. Reson. Med.*, 16(2):181–191.

Haykin, S. (2001). *Communication Systems*. 4th ed. New York City, NY, USA: McGraw-Hill & sons, Inc.

Hennig, J. (1988). Multiecho imaging sequences with low refocusing flip angles. *J. Magn. Reson.*, 78(3):397–407.

Heule, R., Ganter, C., and Bieri, O. (2016). Variable flip angle T1 mapping in the human brain with reduced T2 sensitivity using fast radiofrequency-spoiled gradient echo imaging. *Magn. Reson. Med.*, 75(4):1413–1422.

Heye, T., Yang, S.-R., Bock, M., Brost, S., Weigand, K., Longerich, T., Kauczor, H.-U., and Hosch, W. (2012). MR relaxometry of the liver: significant elevation of T1 relaxation time in patients with liver cirrhosis. *Eur. Radiol.*, 22(6):1224–1232.

Hodel, J., Outteryck, O., Bocher, A.-L., Zéphir, H., Lambert, O., Benadjaoud, M. A., Chechin, D., Pruvo, J.-P., Vermersch, P., and Leclerc, X. (2014). Comparison of 3D double inversion recovery and 2D STIR FLAIR MR sequences for the imaging of optic neuritis: pilot study. *Eur. Radiol.*, 24(12):3069–3075.

Homer, J. and Beevers, M. S. (1985). Driven-equilibrium single-pulse observation of T1 relaxation. a reevaluation of a rapid "new" method for determining NMR spin-lattice relaxation times. *J. Magn. Reson.*, 63(2):287 – 297.

Hong, M., Razaviyayn, M., Luo, Z.-Q., and Pang, J.-S. (2016). A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing. *IEEE Signal Process. Mag.*, 33(1):57–77.

Hoult, D. and Lauterbur, P. C. (1979). The sensitivity of the zeugmatographic experiment involving human samples. *J. Magn. Reson.*, 34(2):425 – 433.

Huang, C., Graff, C. G., Clarkson, E. W., Bilgin, A., and Altbach, M. I. (2012). T2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing. *Magn. Reson. Med.*, 67(5):1355–1366.

Huang, J. and Mumford, D. (1999). Statistics of natural images and models. In *Proc. IEEE CVPR.*, volume 1, pages 541–547.

Huizinga, W., Poot, D. H. J., Guyader, J.-M., Klaassen, R., Coolen, B. F., van Kranenburg, M., van Geuns, R. J. M., Uitterdijk, A., Polfliet, M., Vandemeulebroucke, J., Leemans, A., Niessen, W. J., and Klein, S. (2016). PCA-based groupwise image registration for quantitative MRI. *Med. Image Anal.*, 29:65 – 78.

Hunter, D. R. and Lange, K. (2004). A tutorial on MM algorithms. *Am. Stat.*, 58(1):30–37.

Ibrahim, T. S., Lee, R., Baertlein, B. A., and Robitaille, P. M. L. (2001). B1 field homogeneity and SAR calculations for the birdcage coil. *Phys. Med. Biol.*, 46(2):609.

Irarrazabal, P. and Nishimura, D. G. (1995). Fast three dimensional magnetic resonance imaging. *Magn. Reson. Med.*, 33(5):656–662.

Jablonowski, R., Nordlund, D., Kanski, M., Ubachs, J., Koul, S., Heiberg, E., Engblom, H., Erlinge, D., Arheden, H., and Carlsson, M. (2013). Infarct quantification using 3D inversion recovery and 2D phase sensitive inversion recovery; validation in patients and ex vivo. *BMC Cardiovasc. Disord.*, 13(1):110.

Jackson, G. D., Berkovic, S. F., Duncan, J. S., and Connelly, A. (1993). Optimizing the diagnosis of hippocampal sclerosis using MR imaging. *Am. J. Neuroradiol.*, 14(3):753–762.

Jain, A. K. (1989). *Fundamentals of digital image processing*. Englewood Cliffs, New Yersey, USA: Prentice-Hall, Inc.

Jara, H., Wehrli, F. W., Chung, H., and Ford, J. C. (1993). High-resolution variable flip angle 3D MR imaging of trabecular microstructure in vivo. *Magn. Reson. Med.*, 29(4):528–539.

Jellis, C. L. and Kwon, D. H. (2014). Myocardial T1 mapping: modalities and clinical applications. *Cardiovasc. Diagn. Ther.*, 4(2):126.

Jin, J. (1998). *Electromagnetic analysis and design in magnetic resonance imaging*, volume 1. CRC press.

Jin, K. H., Lee, D., and Ye, J. C. (2016). A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank hankel matrix. *IEEE Trans. Comput. Imaging*, 2(4):480–495.

Juanpere, S., Perez, E., Huc, O., Motos, N., Pont, J., and Pedraza, S. (2011). Imaging of breast implants: a pictorial review. *Insights Imaging*, 2(6):653–670.

Juntu, J., Sijbers, J., Van Dyck, D., and Gielen, J. (2005). *Bias Field Correction for MRI Images*, pages 543–551. Springer Berlin Heidelberg, Berlin, Heidelberg.

Just, M. and Thelen, M. (1988). Tissue characterization with T1, T2, and proton density values: results in 160 patients with brain tumors. *Radiology*, 169(3):779–785.

Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. 1st ed. Upper Saddle River, NJ: Prentice Hall, Inc.

Kellman, P. and Hansen, M. S. (2014). T1-mapping in the heart: accuracy and precision. *J. Cardiovasc. Magn. Reso.*, 16(1):2.

Kelm, B. M., Menze, B. H., Nix, O., Zechmann, C. M., and Hamprecht, F. A. (2009). Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge. *IEEE Trans. Med. Imag.*, 28(10):1534–1547.

Kemper, V. G., De Martino, F., Yacoub, E., and Goebel, R. (2016). Variable flip angle 3D-GRASE for high resolution fMRI at 7 tesla. *Magn. Reson. Med.*, 76(3):897–904.

Kershaw, L. E. and Buckley, D. L. (2006). Precision in measurements of perfusion and microvascular permeability with T1-weighted dynamic contrast-enhanced MRI. *Magn. Reson. Med.*, 56(5):986–992.

Kim, P. K., Hong, Y. J., Im, D. J., Suh, Y. J., Park, C. H., Kim, J. Y., Chang, S., Lee, H.-J., Hur, J., Kim, Y. J., et al. (2017a). Myocardial T1 and T2 mapping: Techniques and clinical applications. *Korean J. Radiol.*, 18(1):113–131.

Kim, T. H. and Haldar, J. P. (2015). SMS-LORAKS: Calibrationless simultaneous multislice MRI using low-rank matrix modeling. In *Proc. IEEE Int. Symp. Biomed. Imag.*, pages 323–326. IEEE.

Kim, T. H., Setsompop, K., and Haldar, J. P. (2017b). LORAKS makes better SENSE: Phase-constrained partial fourier SENSE reconstruction without phase calibration. *Magn. Reson. Med.*, 77(3):1021–1035.

Kimmich, R. (1997). *NMR: Tomography, Diffusometry, Relaxometry*. 1st ed. Heidelberg, Germany: Springer-Verlab Berlin Heidelberg.

Kirk, W. A. and Brailey Sims, E. (2001). *Handbook of metric fixed point theory*.

Knoll, F., Bredies, K., Pock, T., and Stollberger, R. (2011). Second order total generalized variation (TGV) for MRI. *Magn. Reson. Med.*, 65(2):480–491.

Knoll, F., Clason, C., Bredies, K., Uecker, M., and Stollberger, R. (2012). Parallel imaging with nonlinear reconstruction using variational penalties. *Magn. Reson. Med.*, 67(1):34–41.

Koay, C. G., Özarslan, E., and Basser, P. J. (2009). A signal transformational framework for breaking the noise floor and its applications in MRI. *J. Magn. Reson.*, 197(2):108–119.

Krzywinski, M. and Altman, N. (2014). Points of significance: visualizing samples with box plots. *Nat. Methods*, 11(2):119–120.

Kumar, A., Welti, D., and Ernst, R. R. (1975). NMR fourier zeugmatography. *J. Magn. Reson.*, 18(1):69 – 83.

Kuperman, V. (2000). *Magnetic resonance imaging: physical principles and applications*. 1st ed. San Diego, CA, USA: Academic Press.

Kurki, T. and Komu, M. (1995). Spin-lattice relaxation and magnetization transfer in intracranial tumors in vivo: effects of Gd-DTPA on relaxation parameters. *Magn. Reson. Imag.*, 13(3):379–385.

Lacomis, D., Osbakken, M., and Gross, G. (1986). Spin-lattice relaxation (T1) times of cerebral white matter in multiple sclerosis. *Magn. Reson. Med.*, 3(2):194–202.

Lai, Z., Qu, X., Liu, Y., Guo, D., Ye, J., Zhan, Z., and Chen, Z. (2016). Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform. *Med. Image Anal.*, 27:93 – 104.

Lanckriet, G. R. and Sriperumbudur, B. K. (2009). On the convergence of the concave-convex procedure. In Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A., editors, *Advances in Neural Information Processing Systems 22*, pages 1759–1767. Curran Associates, Inc.

Lansberg, M. G., Thijs, V. N., O'Brien, M. W., Ali, J. O., De Crespigny, A. J., Tong, D. C., Moseley, M. E., and Albers, G. W. (2001). Evolution of apparent diffusion coefficient, diffusion-weighted, and T2-weighted signal intensity of acute stroke. *Am. J. Neuroradiol.*, 22(4):637–644.

Larkin, K. G., Oldfield, M. A., and Klemm, H. (1997). Fast Fourier method for the accurate rotation of sampled images. *Opt. Commun.*, 139(1-3):99 – 106.

Larsson, E. G., Erdogmus, D., Yan, R., Principe, J. C., and Fitzsimmons, J. R. (2003). SNR-optimality of sum-of-squares reconstruction for phased-array magnetic resonance imaging. *J. Magn. Reson.*, 163(1):121–123.

Larsson, H., Frederiksen, J., Kjaer, L., Henriksen, O., and Olesen, J. (1988). In vivo determination of T1 and T2 in the brain of patients with severe but stable multiple sclerosis. *Magn. Reson. Med.*, 7(1):43–55.

Larsson, H. B. W., Frederiksen, J., Petersen, J., Nordenbo, A., Zeeberg, I., O, H., and J, O. (1989). Assessment of demyelination, edema, and gliosis by in vivo determination of T1 and T2 in the brain of patients with acute attack of multiple sclerosis. *Magn. Reson. Med.*, 11(3):337–348.

Lauterbur, P. C. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242:190–191.

Lee, D., Jin, K. H., Kim, E. Y., Park, S.-H., and Ye, J. C. (2016). Acceleration of MR parameter mapping using annihilating filter-based low rank hankel matrix (ALOHA). *Magn. Reson. Med.*, 76(6):1848–1864.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.*, 2(2):164–168.

Li, F., Zeng, T., and Zhang, G. (2012a). Lagrangian multipliers and split Bregman methods for minimization problems constrained on s^{n-1}. *J. Vis. Commun. Image Represent.*, 23(7):1041–1050.

Li, W., Griswold, M., and Yu, X. (2012b). Fast cardiac T1 mapping in mice using a model-based compressed sensing method. *Magn. Reson. Med.*, 68(4):1127–1134.

Liang, D., Wang, H., Chang, Y., and Ying, L. (2011). Sensitivity encoding reconstruction with nonlocal total variation regularization. *Magn. Reson. Med.*, 65(5):1384–1392.

Liang, Z.-P. and Lauterbur, P. C. (2000). *Principles of magnetic resonance imaging: a signal processing perspective*. 1st ed. Washington, USA: SPIE Optical Engineering Press.

Liao, J.-R., Pauly, J. M., Brosnan, T. J., and Pelc, N. J. (1997). Reduction of motion artifacts in cine MRI using variable-density spiral trajectories. *Magn. Reson. Med.*, 37(4):569–575.

Lingala, S. G., Hu, Y., DiBella, E., and Jacob, M. (2011). Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. *IEEE Trans. Med. Imag.*, 30(5):1042–1054.

Listerud, J., Mitchell, J., Bagley, L., and Grossman, R. (1996). OIL FLAIR: Optimized interleaved fluid-attenuated inversion recovery in 2D fast spin echo. *Magn. Reson. Med.*, 36(2):320–325.

Liu, Q., Wang, S., Yang, K., Luo, J., Zhu, Y., and Liang, D. (2013). Highly undersampled magnetic resonance image reconstruction using two-level bregman method with dictionary updating. *IEEE Trans. Med. Imag.*, 32(7):1290–1301.

Lustig, M. Sparse MRI. `http://www.eecs.berkeley.edu/~mlustig/Software.html`.

Lustig, M., Donoho, D., and Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6):1182–1195.

Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. (2008). Compressed sensing MRI. *IEEE Signal Process. Mag.*, 25(2):72–82.

Lustig, M. and Vasanawala, S. MR data for compressed sensing. `http://mridata.org/`.

Mai, Z., Hanel, R., Batenburg, K. J., Verhoye, M., Scheunders, P., and Sijbers, J. (2011). Bias field reduction by localized Lloyd-Max quantization. *Magn. Reson. Imag.*, 29(4):536–545.

Manjón, J. V., Coupé, P., and Buades, A. (2015). MRI noise estimation and denoising using non-local PCA. *Med. Image Anal.*, 22(1):35–47.

Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55–L58.

Marks, R. J. I. (1993). *Advanced topics in Shannon sampling and interpolation theory*. 1st ed. New York City, NY, USA: Springer-Verlag New York, Inc.

Marks, R. J. I. (2009). *Handbook of Fourier analysis & its applications*. 1st ed. New York City, NY, USA; Oxford University Press.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11(2):431–441.

Martin, W. W., Roberts, T. E., Frank, Q. Y., and Allen, P. S. (1998). Increased basal ganglia iron in striatonigral degeneration: in vivo estimation with magnetic resonance. *Can. J. Neurol. Sci*, 25(1):44–47.

Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imag*, 22(1):120–128.

Mattson, J. and Simon, M. (1996). *The pioneers of NMR and magnetic resonance in medicine: the story of MRI*. 1st ed. Ramat Gan, Israel : Bar-Ilan University Press, 1996.

Maximov, I. I., Farrher, E., Grinberg, F., and Shah, N. J. (2012). Spatially variable rician noise in magnetic resonance imaging. *Med. Image Anal.*, 16(2):536–548.

McGibney, G., Smith, M. R., Nichols, S. T., and Crawley, A. (1993). Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI. *Magn. Reson. Med.*, 30(1):51–59.

McKenzie, C. A., Williams, A., Prasad, P. V., and Burstein, D. (2006). 24. *J. Magn. Reson. Imaging*, 24(4):928–933.

Messroghli, D. R., Radjenovic, A., Kozerke, S., Higgins, D. M., Sivananthan, M. U., and Ridgway, J. P. (2004). Modified look-locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. *Magn. Reson. Med.*, 52(1):141–146.

Meyer, C. H., Hu, B. S., Nishimura, D. G., and Macovski, A. (1992). Fast spiral coronary artery imaging. *Magn. Reson. Med.*, 28(2):202–213.

Miller, D., Johnson, G., Tofts, P., MacManus, D., and McDonald, W. (1989). Precise relaxation time measurements of normal-appearing white matter in inflammatory central nervous system disease. *Magn. Reson. Med.*, 11(3):331–336.

Mood, A. M., Graybill, E. A., and Boes, D. C. (1963). *Introduction to the Theory of Statistics*. McGraw-Hill.

Muckley, M. J., Noll, D. C., and Fessler, J. A. (2015). Fast parallel MR image reconstruction via B1-based, adaptive restart, iterative soft thresholding algorithms (BARISTA). *IEEE Trans. Med. Imag*, 34(2):578–588.

Namer, I., Waydelich, R., Armspach, J., Hirsch, E., Marescaux, C., and Grucker, D. (1998). Contribution of T2 relaxation time mapping in the evaluation of cryptogenic temporal lobe epilepsy. *Neuroimage*, 7(4):304–313.

Naruse, S., Horikawa, Y., Tanaka, C., Hirakawa, K., Nishikawa, H., and Yoshizaki, K. (1986). Significance of proton relaxation time measurement in brain edema, cerebral infarction and brain tumors. *Magn. Reson. Imag.*, 4(4):293–304.

Nataraj, G., Nielsen, J. F., and Fessler, J. A. (2017). Optimizing MR scan design for model-based T1, T2 estimation from steady-state sequences. *IEEE Trans. Med. Imag.*, 36(2):467–477.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. 4th ed. NY: McGraw-Hill.

Okujava, M., Schulz, R., Ebner, A., and Woermann, F. G. (2002). Measurement of temporal lobe T2 relaxation times using a routine diagnostic MR imaging protocol in epilepsy. *Epilepsy Res.*, 48(1):131–142.

Ordidge, R. J., Gibbs, P., Chapman, B., Stehling, M. K., and Mansfield, P. (1990). High-speed multislice T1 mapping using inversion-recovery echo-planar imaging. *Magn. Reson. Med.*, 16(2):238–245.

Ormerod, I. E. C .and Miller, D. H., McDonald, W. I., Boulay, E. P. G. H. d., Rudge, P., Kendall, B. E., Moseley, I. F., Johnson, G., Tofts, P. S., Halliday, A. M., Bronstein, A. M., Scaravilli, F., Harding, A. E., Barnes, D., and Zilkha, K. (1987). The role of NMR imaging in the assessment of multiple sclerosis and isolated neurological lesions: a quantitative study. *Brain*, 110(6):1579–1616.

Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. (2005). An iterative regularization method for total variation-based image restoration. *Multiscale Model Simul.*, 4(2):460–489.

Oshio, K. and Feinberg, D. A. (1991). GRASE (Gradient-and Spin-Echo) imaging: A novel fast MRI technique. *Magn. Reson. Med.*, 20(2):344–349.

Otazo, R., Candès, E., and Sodickson, D. K. (2015). Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magn. Reson. Med.*, 73(3):1125–1136.

Panic, M., Aelterman, J., Crnojevic, V., and Pizurica, A. (2017). Sparse recovery in magnetic resonance imaging with a markov random field prior. *IEEE Trans. Med. Imag.*, 36(10):2104–2115.

Papoulis, A. (1977). Generalized sampling expansion. *IEEE Trans. Circuits Syst.*, 24(11):652–654.

Papoulis, A. and Pillai, S. U. (2002). *Probability, Random Variables and Stochastic Processes*. 4th ed. NY: McGraw-Hill.

Park, H. W., Cho, M. H., and Cho, Z. H. (1985). Time-multiplexed multislice inversion-recovery techniques for NMR imaging. *Magn. Reson. Med.*, 2(6):534–539.

Pauly, J. M. (2005). Chapter 5: Reconstruction of non-cartesian data. Technical report, Stanford University.

Peeters, F., Annet, L., Hermoye, L., and Van Beers, B. E. (2004). Inflow correction of hepatic perfusion measurements using T1-weighted, fast gradient-echo, contrast-enhanced MRI. *Magn. Reson. Med.*, 51(4):710–717.

Peng, X., Ying, L., Liu, Y., Yuan, J., Liu, X., and Liang, D. (2016). Accelerated exponential parameterization of T2 relaxation with model-driven low rank and sparsity priors (MORASA). *Magn. Reson. Med.*, 76(6):1865–1878.

Permuter, H., Francos, J., and Jermyn, H. (2003). Gaussian mixture models of texture and colour for image database retrieval. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, volume 3, pages III–569–72.

Peters, D. C., Derbyshire, J. A., and McVeigh, E. R. (2003). Centering the projection reconstruction trajectory: reducing gradient delay errors. *Magn. Reson. Med.*, 50(1):1–6.

Peters, D. C., Rohatgi, P., Botnar, R. M., Yeon, S. B., Kissinger, K. V., and Manning, W. J. (2006). Characterizing radial undersampling artifacts for cardiac applications. *Magn. Reson. Med.*, 55(2):396–403.

Petzschner, F. H., Ponce, I. P., Blaimer, M., Jakob, P. M., and Breuer, F. A. (2011). Fast MR parameter mapping using k-t principal component analysis. *Magn. Reson. Med.*, 66(3):706–716.

Pieciak, T., Aja-Fernández, S., and Vegas-Sánchez-Ferrero, G. (2017). Non-stationary Rician noise estimation in parallel MRI using a single image: a variance-stabilizing approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(10):2015–2029.

Pinsky, M. A. (2002). *Introduction to Fourier analysis and wavelets*, volume 102. 1st ed. Pacific Grove, CA, USA: Brooks/Cole.

Pizurica, A., Aelterman, J., Bai, F., Vanloocke, S., and Quang, Hiep Luong, B. G. W. P. (2011). On structured sparsity and selected applications in tomographic imaging. *Proc.SPIE*, 8138:8138 – 8138 – 12.

Pluim, J. P. W., Maintz, J. B., and Viergever, M. A. (2003). Mutual-information-based registration of medical images: a survey. *IEEE Trans. Med. Imag*, 22(8):986–1004.

Poon, C. S. and Henkelman, R. M. (1992). Practical T2 quantitation for clinical applications. *J. Magn. Reson. Imaging*, 2(5):541–553.

Poot, D. H. J. (2010). *Advances in the reconstruction and statistical processing of magnetic resonance images*. PhD thesis, University of Antwerp.

Poot, D. H. J. and Klein, S. (2015). Detecting statistically significant differences in quantitative MRI experiments, applied to diffusion tensor imaging. *IEEE Trans. Med. Imag*, 34(5):1164–1176.

Preece, M. T., Osborn, A. G., Chin, S. S., and Smirniotopoulos, J. G. (2006). Intracranial neurenteric cysts: Imaging and pathology spectrum. *Am. J. Neuroradiol.*, 27(6):1211–1216.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes in C++: the art of scientific computing*. New York, NJ, USA: Cambridge Univ. Press.

Pruessmann, K. P., Weiger, M., Scheidegger, B. M., and Boesiger, P. (1999). SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.*, 42(5):952–962.

Purcell, E. M., Torrey, H. C., and Pound, R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, 69(1-2):37.

Qian, Y. and Boada, F. E. (2008). Acquisition-weighted stack of spirals for fast high-resolution three-dimensional ultra-short echo time MR imaging. *Magn. Reson. Med.*, 60(1):135–145.

Rabi, I. I., Zacharias, J. R., Millman, S., and Kusch, P. (1938). A new method of measuring nuclear magnetic moment. *Phys. Rev.*, 53(4):318–318.

Rajan, J., Poot, D. H. J., Juntu, J., and Sijbers, J. (2010). Noise measurement from magnitude MRI using local estimates of variance and skewness. *Phys. Med. Biol.*, 55(16):N441.

Rakow-Penner, R., Daniel, B., Yu, H., Sawyer-Glover, A., and Glover, G. H. (2006). Relaxation times of breast tissue at 1.5 T and 3 T measured using IDEAL. *J. Magn. Reson. Imaging*, 23(1):87–91.

Ramos-Llordén, G., den Dekker, A. J., Björk, M., Verhoye, M., and Sijbers, J. (2016). NOVIFAST: A fast non-linear least squares method for accurate and precise estimation of T1 from SPGR signals. *Proc. Intl. Soc. Mag. Reson. Med.*, 24:2820.

Ramos-Llordén, G., den Dekker, A. J., Van Steenkiste, G., Jeurissen, B., Vanhevel, F., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2017). A unified maximum likelihood framework for simultaneous motion and T1 estimation in quantitative MR T1 mapping. *IEEE Trans. Med. Imag.*, 36(2):433–446.

Ramos-Llordén, G., den Dekker, A. J., Van Steenkiste, G., Van Audekerke, J., Verhoye, M., and Sijbers, J. (2015a). Simultaneous motion correction and T1 estimation in quantitative T1 mapping: An ML restoration approach. In *Proc. IEEE ICIP*, pages 3160–3164.

Ramos-Llordén, G., Vegas-Sánchez-Ferrero, G., Martín-Fernández, M., Alberola-López, C., and Aja-Fernández, S. (2015b). Anisotropic diffusion filter with memory based on speckle statistics for ultrasound images. *IEEE Trans. Image Process.*, 24(1):345–358.

Ravishankar, S. and Bresler, Y. (2011). MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imag.*, 30(5):1028–1041.

Renna, F., Calderbank, R., Carin, L., and Rodrigues, M. R. D. (2014). Reconstruction of signals drawn from a Gaussian mixture via noisy compressive measurements. *IEEE Trans. Signal Process.*, 62(9):2265–2277.

Robinson, S. M. (1970). A short proof of Cramer's rule. *Mathematics Magazine*, 43(2):94–95.

Roelandts, T., Batenburg, K. J., Biermans, E., Kübel, C., Bals, S., and Sijbers, J. (2012). Accurate segmentation of dense nanoparticles by partially discrete electron tomography. *Ultramicroscopy*, 114:96–105.

Roemer, P. B., Edelstein, W. A., Hayes, C. E., Souza, S. P., and Mueller, O. (1990). The NMR phased array. *Magn. Reson. Med.*, 16(2):192–225.

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D.*, 60(1):259–268.

Sacolick, L. I., Wiesinger, F., Hancu, I., and Vogel, M. W. (2010). B1 mapping by bloch-siegert shift. *Magn. Reson. Med.*, 63(5):1315–1322.

Samsonov, A. A., Kholmovski, E. G., Parker, D. L., and Johnson, C. R. (2004). POCSENSE: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging. *Magn. Reson. Med.*, 52(6):1397–1406.

Samsonov, A. A., Velikina, J., Jung, Y., Kholmovski, E. G., Johnson, C. R., and Block, W. F. (2010). POCS-enhanced correction of motion artifacts in parallel MRI. *Magn. Reson. Med.*, 63(4):1104–1110.

Sawatzky, A. (2011). *(Nonlocal) Total Variation in Medical Imaging*. PhD thesis, WWU Münster, Germany.

Scheffler, K. and Hennig, J. (2003). Is TrueFISP a gradient-echo or a spin-echo sequence? *Magn. Reson. Med.*, 49(2):395–397.

Schmitt, F., Stehling, M. K., and Turner, R. (1998). *Echo-planar imaging: theory, technique and application*. 1st ed. Heidelberg,Germany: Springer-Verlag Berlin Heidelberg.

Segers, H., Palenstijn, W. J., Batenburg, K. J., and Sijbers, J. (2013). Discrete tomography in MRI: A simulation study. *Fundam. Inform.*, 125(3-4):223–237.

Setsompop, K., Gagoski, A. B., Polimeni, R. J., Witzel, T., Wedeen, J. V., and Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Med.*, 67(5):1210–1224.

Shin, W., Gu, H., and Yang, Y. (2009). Fast high-resolution T1 mapping using inversion-recovery look-locker echo-planar imaging at steady state: Optimization for accuracy and reliability. *Magn. Reson. Med.*, 61(4):899–906.

Sijbers, J. and den Dekker, A. J. (2004). Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magn. Reson. Med.*, 51(3):586–594.

Sijbers, J., den Dekker, A. J., Raman, E., and Van Dyck, D. (1999). Parameter estimation from magnitude MR images. *Int. J. Imag. Syst. Tech.*, 10(2):109–114.

Sijbers, J., den Dekker, A. J., Scheunders, P., and Van Dyck, D. (1998a). Maximum-likelihood estimation of Rician distribution parameters. *IEEE Trans. Med. Imaging*, 17(3):357–361.

Sijbers, J., den Dekker, A. J., Van Audekerke, J., Verhoye, M., and Van Dyck, D. (1998b). Estimation of noise from magnitude MR images. *Magn. Reson. Imaging*, 16(1):87–90.

Sijbers, J., den Dekker, A. J., Verhoye, M., Raman, E. R., and Van Dyck, D. (1998c). Optimal estimation of T2 maps from magnitude MR images. In *Proc. SPIE'98: Medical Imaging*, volume 3338, pages 384–390.

Sijbers, J., Poot, D. H. J., den Dekker, A. J., and Pintjens, W. (2007). Automatic estimation of the noise variance from the histogram of a magnetic resonance image. *Phys. Med. Biol.*, 52(5):1335.

Smith, S. A., Edden, R. A., Farrell, J. A., Barker, P. B., and Van Zijl, P. (2008). Measurement of T1 and T2 in the cervical spinal cord at 3 tesla. *Magn. Reson. Med.*, 60(1):213–219.

Sodickson, D. K. and Manning, W. J. (1997). Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn. Reson. Med.*, 38(4):591–603.

Stanisz, G. J., Odrobina, E. E., Pun, J., Escaravage, M., Graham, S. J., Bronskill, M. J., and Henkelman, R. M. (2005). T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magn. Reson. Med.*, 54(3):507–512.

Stikov, N., Boudreau, M., Levesque, I. R., Tardif, C. L., Barral, J. K., and Pike, G. B. (2015). On the accuracy of T1 mapping: searching for common ground. *Magn. Reson. Med.*, 73(2):514–522.

Stollberger, R. and Wach, P. (1996). Imaging of the active b1 field in vivo. *Magn. Reson. Med.*, 35(2):246–251.

Studler, U., White, L. M., Andreisek, G., Luu, S., Cheng, H.-L. M., and Sussman, M. S. (2010). Impact of motion on T1 mapping acquired with inversion recovery fast spin echo and rapid spoiled gradient recalled-echo pulse sequences for delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) in volunteers. *J. Magn. Reson. Imaging*, 32(2):394–398.

Styner, M., Brechbühler, C., Szckely, G., and Gerig, G. (2000). Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Trans. Med. Imag.*, 19(3):153–165.

Szumowski, J., Durkan, M. G., Foss, E. W., Brown, D. S., Schwarz, E., and Crawford, D. C. (2012). Signal polarity restoration in a 3D inversion recovery sequence used with delayed gadolinium-enhanced magnetic resonance imaging of cartilage (dGEMRIC). *J. Magn. Reson. Imaging*, 36(5):1248–1255.

Tan, H. and Meyer, C. H. (2009). Estimation of k-space trajectories in spiral MRI. *Magn. Reson. Med.*, 61(6):1396–1404.

Teixeira, R. P. A., Malik, S. J., and Hajnal, J. V. (doi:10.1002/mrm.26670, 2017). Joint system relaxometry (JSR) and Crámer-Rao lower bound optimization of sequence parameters: A framework for enhanced precision of DESPOT T1 and T2 estimation. *Magn. Reson. Med.*

Tofts, P. (2004). *Quantitative MRI of the brain: measuring changes caused by disease*. Chichester, England: John Wiley & Sons.

Trzasko, J. and Manduca, A. (2009). Highly undersampled magnetic resonance image reconstruction via homotopic l_0-minimization. *IEEE Trans. Med. Imag.*, 28(1):106–121.

Trzasko, J. D., Mostardi, P. M., Riederer, S. J., and Manduca, A. (2013). Estimating T1 from multichannel variable flip angle SPGR sequences. *Magn. Reson. Med.*, 69(6):1787–1794.

van den Bos, A. (2007). *Parameter Estimation for Scientists and Engineers*. Hoboken, New Yersey, USA: John Wiley & Sons.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999a). Automated model-based bias field correction of MR images of the brain. *IEEE Trans. Med. Imag.*, 18(10):885–896.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999b). Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imag.*, 18(10):897–908.

Van Steenkiste, G. (2016). *Super-resolution estimation of quantitative MRI parameters*. PhD thesis, University of Antwerp.

Van Steenkiste, G., Poot, D. H. J., Jeurissen, B., Den Dekker, A. J., Vanhevel, F., Parizel, P. M., and Sijbers, J. (2017). Super-resolution T1 estimation: Quantitative high resolution T1 mapping from a set of low resolution T1-weighted images with different slice orientations. *Magn. Reson. Med.*, 77(5):1818–1830.

Varadarajan, D. and Haldar, J. P. (2015). A majorize-minimize framework for rician and non-central chi MR images. *IEEE Trans. Med. Imaging*, 34(10):2191–2202.

Vegas-Sánchez-Ferrero, G., Aja-Fernández, S., Martín-Fernández, M., Frangi, A. F., and Palencia, C. (2010). Probabilistic-driven oriented speckle reducing anisotropic diffusion with application to cardiac ultrasonic images. In *Proc. 13th Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, pages 518–525. Springer.

Velikina, J. V., Alexander, A. L., and Samsonov, A. (2013). Accelerating MR parameter mapping using sparsity-promoting regularization in parametric dimension. *Magn. Reson. Med.*, 70(5):1263–1273.

Velikina, J. V. and Samsonov, A. A. (2014). Reconstruction of dynamic image series from undersampled MRI datausing data-driven model consistency condition (MOCCO). *Magn. Reson. Med.*, 74(5):1279–1290.

Vemuri, P., Kholmovski, E., Parker, D., and Chapman, B. (2005). Coil sensitivity estimation for optimal SNR reconstruction and intensity inhomogeneity correction in phased array MR imaging. In *Inf. Process Med. Imaging*, volume 19, pages 603–614. Springer.

Veraart, J. (2013). *Optimal estimation of diffusion MRI parameters*. PhD thesis, University of Antwerp.

Veraart, J., Fieremans, E., and Novikov, D. S. (2016). Diffusion MRI noise mapping using random matrix theory. *Magn. Reson. Med.*, 76(5):1582–1593.

von Knobelsdorff-Brenkenhoff, F., Prothmann, M., Dieringer, M. A., Wassmuth, R., Greiser, A., Schwenke, C., Niendorf, T., and Schulz-Menger, J. (2013). Myocardial T1 and T2 mapping at 3 T: reference values, influencing factors and implications. *J. Cardiovasc. Magn. Reso.*, 15(1):53.

Vovk, U., Pernus, F., and Likar, B. (2007). A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans. Med. Imag.*, 26(3):405–421.

Vymazal, J., Righini, A., Brooks, R. A., Canesi, M., Mariani, C., Leonardi, M., and Pezzoli, G. (1999). T1 and T2 in the brain of healthy subjects, patients with parkinson disease, and patients with multiple system atrophy: Relation to iron content. *Radiology*, 211(2):489–495.

Walsh, D. O., Gmitro, A. F., and Marcellin, M. W. (2000). Adaptive reconstruction of phased array MR imagery. *Magn. Reson. Med.*, 43(5):682–690.

Wang, F., Xu, Z., and Xu, H.-K. (2014). Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems. *arXiv preprint arXiv:1410.8625*.

Wang, X., Roeloffs, V., Klosowski, J., Tan, Z., Voit, D., Uecker, M., and Frahm, J. (doi:10.1002/mrm.26670, 2017). Model-based T1 mapping with sparsity constraints using single-shot inversion-recovery radial flash. *Magn. Reson. Med.*

Wansapura, J. P., Holland, S. K., Dunn, R. S., and Ball, W. S. (1999). NMR relaxation times in the human brain at 3.0 tesla. *J. Magn. Reson. Imaging*, 9(4):531–538.

Warntjes, J. B. M., Dahlqvist, O., and Lundberg, P. (2007). Novel method for rapid, simultaneous T1, T2*, and proton density quantification. *Magn. Reson. Med.*, 57(3):528–537.

Weickert, J.and Romeny, B. and Viergever, M. (1998). Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Trans. Image Process.*, 7(3):398–410.

Weigel, M., Zaitsev, M., and Hennig, J. (2007). Inversion recovery prepared turbo spin echo sequences with reduced SAR using smooth transitions between pseudo steady states. *Magn. Reson. Med.*, 57(3):631–637.

Welch, E. B., Manduca, A., Grimm, R. C., Ward, H. A., and Jack Jr, C. R. (2002). Spherical navigator echoes for full 3D rigid body motion measurement in MRI. *Magn. Reson. Med.*, 47(1):32–41.

Woermann, F. G., Barker, G. J., Birnie, K. D., Meencke, H. J., and Duncan, J. S. (1998). Regional changes in hippocampal T2 relaxation and volume: a quantitative magnetic resonance imaging study of hippocampal sclerosis. *J. Neurol. Neurosurg. Psychiatr.*, 65(5):656–664.

Wright, K. L., Hamilton, J. I., Griswold, M. A., Gulani, V., and Seiberlich, N. (2014). Non-cartesian parallel imaging reconstruction. *J. Magn. Reson. Imaging*, 40(5):1022–1040.

Wright, P. J., Mougin, O. E., Totman, J. J., Peters, A. M., Brookes, M. J., Coxon, R., Morris, P. E., Clemence, M., Francis, S. T., Bowtell, R. W., and Gowland, P. A. (2008). Water proton T1 measurements in brain tissue at 7, 3, and 1.5T using IR-EPI, IR-TSE, and MPRAGE: results and optimization. *Magn. Reson. Mater. Phy.*, 21(1):121–130.

Wright, S. J., Nowak, R. D., and Figueiredo, M. A. (2009). Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, 57(7):2479–2493.

Wu, C. and Tai, X.-C. (2010). Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Imaging. Sci.*, 3(3):300–339.

Xue, H., Greiser, A., Zuehlsdorff, S., Jolly, M.-P., Guehring, J., Arai, A. E., and Kellman, P. (2013). Phase-sensitive inversion recovery for myocardial T1 mapping with motion correction and parametric fitting. *Magn. Reson. Med.*, 69(5):1408–1420.

Yang, J., Liao, X., Yuan, X., Llull, P., Brady, D. J., Sapiro, G., and Carin, L. (2015). Compressive sensing by learning a Gaussian mixture model from measurements. *IEEE Trans. Image Process.*, 24(1):106–119.

Yarnykh, V. L. (2007). Actual flip-angle imaging in the pulsed steady state: A method for rapid three-dimensional mapping of the transmitted radiofrequency field. *Magn. Reson. Med.*, 57(1):192–200.

Ye, F. Q., Allen, P. S., and Martin, W. R. (1996). Basal ganglia iron content in Parkinson's disease measured with magnetic resonance. *Mov. Disord.*, 11(3):243–249.

Ying, L. and Liang, Z. P. (2010). Parallel MRI using phased array coils. *IEEE Signal Process. Mag.*, 27(4):90–98.

Ying, L., Xu, D., and Liang, Z.-P. (2004). On Tikhonov regularization for image reconstruction in parallel MRI. In *26th Conf. Proc. IEEE Eng. Med. Biol. Soc.*, volume 1, pages 1056–1059.

Yudilevich, E. and Stark, H. (1987a). Interpolation from samples on a linear spiral scan. *IEEE Trans. Med. Imag.*, 6(3):193–200.

Yudilevich, E. and Stark, H. (1987b). Spiral sampling in magnetic resonance imaging-the effect of inhomogeneities. *IEEE Trans. Med. Imag.*, 6(4):337–345.

Zhang, F. (1999). *Matrix Theory: Basics results and techniques*. 1st ed. New York, NY: Springer-Verlag New York Inc.

Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2011). FSIM: a feature similarity index for image quality assessment. *IEEE Trans. Image Process.*, 20(8):2378–2386.

Zhao, B., Lu, W., Hitchens, T. K., Lam, F., Ho, C., and Liang, Z.-P. (2015). Accelerated MR parameter mapping with low-rank and sparsity constraints. *Magn. Reson. Med.*, 74(2):489–498.

Zhao, F., Noll, D. C., Nielsen, J.-F., and Fessler, J. A. (2012). Separate magnitude and phase regularization via compressed sensing. *IEEE Trans. Med. Imag.*, 31(9):1713–1723.

Zhu, S. C., Wu, Y., and Mumford, D. (1998). Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. *Int. J. Comput. Vision*, 27(2):107–126.