Magnetic Resonance Image Reconstruction with Sparse Sampled Data

David Pérez Benito
University of Eastern Finland
Department of Applied Physics
September 2012

UNIVERSITY OF EASTERN FINLAND, Faculty of Science and Forestry
Magnetic Resonance Image Reconstruction with Sparse Sampled Data
David Pérez Benito
Master of Science Thesis, 75 pages.
Supervisors: Docent, Ville Kolehmainen
Prof., Marko Vauhkonen
September 2012

The inherent slow data acquisition in Magnetic Resonance Imaging (MRI) hampers current applications and the progress in moving target imaging. The rate at which fast sampling strategies assemble data have nearly reached physical and physiological limits. Current techniques attempt to reduce the scanning time by collecting reduced amount of data without degrading the image quality. Typically image reconstruction from highly undersampled data results in an ill-conditioned problem. The common approach is to treat this task as an inverse problem. In this framework, variational formulation is employed. Sparsity of natural images in a suitable basis such as wavelets is exploited as regularization term. On the other hand, the implementation of Total Variation (TV) regularization is particularly beneficial in denoising for piecewise-constant objects such as the Shepp-Logan phantom. This approach with two regularization terms poses a convex optimization problem that is solved with non-linear conjugate gradients numerical optimization. In this thesis work, traditional reconstruction method, inverse Fourier transform, was compared to the proposed approach in three different undersampled simulated data patterns: Cartesian, radial and spiral. It was found that in the presence of both regularization terms, the target image was accurately reconstructed. Noise and undersampling artefacts are removed while preserving fine details in the reconstructed image. The method outperforms the common Nyquist sampling limits and traditional image reconstruction methods. Scanning time is proportional to the number of acquired samples in Cartesian and radial pattern. Therefore based on the minimal visual impact between the reconstructed image and the target image, the present method is of potential application in order to save scanning time in a real case scenario.

Acknowledgments

First of all I want to thank the Excellent Inverse Problem group in Kuopio for the opportunity to write my master thesis with them. I frankly appreciate the friendly treatment that every single member provided me during the year we were together. I would like to thank specially my supervisors Ville Kolehmainen and Marko Vauhkonen for the freedom, confidence, guidance, help and patience with my mistakes during this humble research work. My initial learning expectations have been exceeded by far.

I do not forget all my friends in Finland, thankfully too many to be named here. Thanks all of you for making me happy in the dark, bright, cold and less cold days. I would like to have special words for Gerardo Pendejo del Muro for his help and sharing a daily almost-Spanish chat with me, Guozhi Dong for his exemplary mathematical thinking and Christina Mischke for being close from the first day to the very last minute. Cheers mates!

I have to send my most loving thanks, unfortunately in the distance, to Rosa Torguet and David Sánchez, simply because without you guys I would have been another person.

Kuopio, September 2012
David Pérez

Abbreviations

MRI	Magnetic Resonance Image
NMR	Nuclear Magnetic Resonance
PET	Proton Emission Tomography
CT	Computed Tomography
IFFT	Inverse Fast Fourier Transform
RF	Radio Frequency
CS	Compressive Sensing
DCT	Discrete Cosine Transform
SVD	Singular Value Decomposition
TV	Total Variation
FID	Free Induction Decay
IDFT	Inverse Discrete Fourier Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
SOCP	Second Order Cone Program
NLCG	Non-Linear Conjugate Gradient
SNR	Signal-to-Noise Ratio

Contents

Chapter 1

Introduction

Magnetic Resonance Imaging (MRI) is a versatile tomography imaging method based on Nuclear Magnetic Resonance (NMR) phenomena. Reconstructions of images or three dimensional representation of the spatial density of the target area or volume are achieved. Moreover, the time evolution of living tissues or blood flow can be also obtained. MRI provides an excellent soft tissue contrast and it is used in many types of diagnosis. On the contrary, other techniques such as ultrasound, Proton Emission Tomography (PET) or Computed Tomography (CT) among others, only have a particular benefit for a specific type of examination. In the present days only CT is comparable to MRI in terms of contrast and resolutions of the resulting image. Unlike CT, MRI does not involve ionizing radiation. Furthermore, the extensive variety of modifiable parameters makes MRI a common technique to discern anatomic structures at any plane in the body.

It was not until the 1970s when the echoes of the theoretical background led to results. Raymond Vahan Damadian is conceded as the inventor of the first Magnetic Resonance Scanning Machine in 1969; in fact he patented the concept of NMR for detecting cancer in 1974 [1]. Damadian's initial work flaws were overcome by Paul Lauterbur [2] and Sir Peter Mansfield whose studies deserved the Nobel Prize in Medicine in 2003. In spite of that Damadian claimed that he discovered MRI and the two Nobel-winning scientists only refined his technology. In order to express his unrest he wrote a full page in The New York Times, The Washington Post and Los Angeles Times headlined *"The Shameful Wrong That Must Be Righted" [3].*

The sampled signal is the spatial frequency component of the image to be reconstructed. Under a fully sampled scheme, the common Inverse Fast Fourier Transform (IFFT) is sufficient and efficient in MR image reconstruction. Unfortunately, time required to collect data is a hindrance that nowadays hampers broader applications, for instance, in emergency diagnosis or moving targets such as heart or lungs. During the last 30 years, MRI has improved dramatically in image qual-

ity and imaging speed. The latter has been developed mainly by acceleration of the hardware operation. However it has nearly reached physical limits of the hardware components and physiological limitations for safety issues, mainly nerve stimulation and Radio Frequency (RF) heat deposition [4].

From the 1990s, two main research areas aim to reconstruct the image from short data set. The first approach exploits the redundancy of data domain, known as k-space. The most significant result is parallel imaging which makes use of an array of coils, each coil associated with a short area in the subject to be imaged. Signal is received in different channels and combined together. Image reconstruction is a combination of inverse Fourier transform and particular methods such as SENSE [5–7] or GRAPPA [8]. The second exploits spatial redundancy of medical images. In particular the novel sampling technique Compressed Sensing (CS) reconstructs the image from vastly reduced incoherent measurement [9–15]. Several sparsifying transformations have been used in MRI applications. Remarkable publications employ wavelet domain [16–19] , Discrete Cosine Transform (DCT) [18], Singular Value Decomposition (SVD) [20] or discrete difference [18, 21].

Problems arising from highly undersampled data set are typically treated in the framework of inverse problems. In this sense, regularization is employed to alleviate the ill-conditioned problem where the l_1-norm of the sparsifying transform is the regularization term. In this thesis work, two regularization terms are employed and compared for the piecewise constant image the Shepp-Logan phantom: wavelets and Total Variation (TV). This approach poses a convex optimization problem that is solved with non-linear conjugate gradients numerical optimization.

Chapter 2 offers a brief review of the basic principles of MRI in order to gain some fundamental insight in the signal origin. Chapter 3 surveys the mathematical approach of the image reconstruction enclosed in a sparse image representation. In Chapter 4 methods to achieve the final reconstruction are detailed and finally presented in Chapter 5. Last chapter is dedicated for discussion of the results and conclusions.

Chapter 2

Basics in Magnetic Resonance Imaging

The present chapter outlines a global description of the main theoretical and mathematical MRI background along with a brief comment about the physical constrains that fundamental scanner components impose. In this regard the behavior of an individual proton up to an assemble of particles is explained. Special attention is paid to the origin of the MRI signal, collection of raw data and traditional image reconstruction methods. It is beyond of the scope of this thesis work any meticulous background description since it should be described in a rigorous quantum mechanics formulation. The main principles can be described appealing to classical mechanics. For more detailed MRI scheme, reader is referred to the following articles and books [24–32].

2.1 Nucleus in a magnetic field

MRI is based on Nuclear Magnetic Resonance (NMR). A global description may cover the behavior of a nucleus in the presence of an external magnetic field and the subsequent resonance phenomena resulting from the interaction. The mentioned magnetic field is typically generated by a superconducting magnet of strength varying from 0.1 Tesla (T) to 3 T for clinical applications and up to 10 T magnets for research purposes.

Up to certain scale, matter is compounded of atoms where electrons orbit about a nucleus formed by protons and neutrons. In a classical picture protons and neutrons are known to rotate around its own axis with a degree of freedom termed spin. However, protons carry a positive charge and spin represents a minute spinning charged particle with the consequent circulating electric current. It is common to describe the magnetic field associated with the current with a vector

quantity called magnetic moment $\vec{\mu}$; perpendicularly orientated with respect to the plane of the loop and proportional to the current magnitude. In order to determine uniquely this quantity, both direction and magnitude must be specified. The basic postulates in quantum mechanics in Section 2.1.1 will be of help to elaborate the concept.

In the sequel magnetic field $\vec{B}$ is consider to be constant and it is placed along the z-direction

$$\vec{B} = B_0 \vec{k} \tag{2.1}$$

For a free system, protons are randomly orientated due to the thermal motion. Upon applying an external magnetic field, off-center forces cause the system to rotate toward a state of lower energy

$$E = -\vec{\mu} \cdot \vec{B} \tag{2.2}$$

From basic physics, rotation is linked to angular moment $\vec{J}$. Its time derivative characterizes the relative position between the external magnetic field and angular magnetic moment

$$\frac{d\vec{J}}{dt} = \vec{\mu} \times \vec{B} \tag{2.3}$$

The inner product in equation (2.2) indicates that the system lies on a lower energy state with decreasing angle between the magnetic moment and the magnetic field. Nevertheless a complete alignment is not feasible due to the constraints imposed by quantum mechanics.

2.1.1 Quantum constrictions

Classical mechanics is not applicable at atomic scale and quantum mechanics is required to explain the behavior of the system under study. In order to describe how a proton reacts to an external magnetic field $B_0\vec{k}$, it is necessary to measure the projection of the angular moment on the z-axis. The angular moment in quantum mechanics is not a vector but a vector operator. Its three components do not commute with each other and as a consequence of the Heisenberg uncertainty principle they cannot be accurately and simultaneously measured [33]. The result of a precise measurement of a component of the angular moment is one of the eigenvalues of the operator. Along the z-direction these are found to be

$$J_z = 0, \pm\frac{\hbar}{2}, \pm\hbar, \pm\frac{3\hbar}{2}, \dots$$

which means that feasible projection on z-axis of the spin angular moment are quantized.

Another connection between quantum and classical physics is the relation between the magnetic moment with the spin angular moment

$$\vec{\mu} = \gamma \vec{J} \tag{2.4}$$

where the proportional constant γ is the particle-dependent gyromagnetic ratio, Table 2.1. Now the magnitude of magnetic moment for a proton is

$$|\vec{\mu}| = \gamma \hbar \sqrt{I(I+1)}$$

and its z-component is

$$\mu_z = \gamma m_s \hbar$$

where I is the spin number and m_s is the co-called secondary quantum number. A crucial consequence is that m_s is restricted for a given I to the $2I + 1$ values

$$I(I+1) \geq m_s^2$$

Protons have spin $I = 1/2$. Hence there are two possible orientations corresponding to spin-up and spin-down. Since the z-component is precisely determined, the direction of its transverse component $\vec{\mu}_{xy}$ remains unknown as stated by the Heisenberg uncertainty principle.

Table 2.1: *Gyromagnetic ratios and spin for different elements. 1M = 1 mole/liter. The quoted body abundance varies from tissue to tissue. Values obtained from [25].*

Nucleus	Gyromagnetic ratio	Spin	Abundance in human body
^{1}H	42.58	1/2	88 M
^{23}Na	11.27	3/2	80 mM
^{31}P	17.25	1/2	75 mM
^{17}O	5.77	5/2	16 mM
^{19}F	40.08	1/2	4 μM

The inability of the magnetic moment to align to the external magnetic field is of paramount importance since it makes the proton to experience a torque that in turn evokes a precession around the direction of the field, shown in Figure 2.1.

The last feature subtracted from quantum mechanics is that nuclei with an even mass number and even charge number have zero spin. Thus no all of the chemical elements are suitable for a MR signal.

2.1.2 Equation of motion

Substituting the equation relating the magnetic moment with the angular moment or spin, equation (2.4), into (2.3), leads to the following equation of motion

$$\frac{d\vec{\mu}}{dt} = \gamma \vec{\mu} \times \vec{B} \tag{2.5}$$

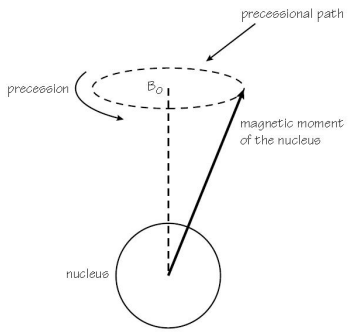

Figure 2.1: *Precession of a proton about an external magnetic field. Figure taken from [25].*

This expression is a rudimentary version of the Bloch equation that regulates the system behavior. It means that $\frac{d\vec{\mu}}{dt}$ is perpendicular to the plane defined by $\vec{\mu}$ and $\vec{B}$ and the magnetic moment precesses about the constant magnetic field in a clockwise circular path described by a cone of angle θ in Figure 2.1. By substituting the external magnetic field (2.1) in the equation of motion (2.5) and solving the homogeneous differential equation, the solution for the nuclear precession in Cartesian components follows

$$\begin{cases} \mu_{xy}(t) = \mu_{xy}(0)\mathrm{e}^{-i\gamma B_0 t} \\ \mu_z(t) = \mu_z(0) \end{cases} \tag{2.6}$$

with angular frequency given by

$$\omega_0 \equiv \gamma B_0 \tag{2.7}$$

The angular frequency ω_0 is known as Larmor frequency and ties magnetic field with precessing frequency for a single spin. In reference to Table (2.1), particles of different type are found to precess at particular angular velocity. The particle-dependent frequency is an important feature in MRI because particles are excited accordingly to their rotation frequency, Section 2.2.2. Then in the excitation process, a certain type of particle can be activated with its particular frequency while leaving the rest unaffected. Hence it favours the discrimination among tissues using their physical and biochemical properties as well as the application of contrast agents and spectroscopy. It is also remarkable that calcium present in bones produce minimal signal and tissues surrounded by bone can be imaged in contrast to CT or ultrasound. Similarly, the higher presence of hydrogen present in water and fat in a human body (Table 2.1) is the main contribution to the signal.

2.1.3 Complex representation

It is particularly convenient to introduce a complex representation for those problems involving rotation. The modulus of a vector describes the distance from the origin to a measurement point and the phase angle select the point on the trajectory. The transverse magnetic moment in Bloch equation (2.5) can be solved with the definition of a complex quantity

$$\mu_+(\mathrm{t}) = \mu_x(t) + i\mu_y(t)$$

Then the solution of the first equation in (2.6) reads

$$\mu_+(\mathrm{t}) = |\mu_+(0)|\mathrm{e}^{-i\phi_0(t)}$$

Previous equation illustrates the rotational movement with constant magnitude $|\mu_+(0)|$ and phase $\phi_0(t)$ that absorbs the angular velocity and any arbitrary initial phase $\phi_0(0)$.

2.1.4 Bulk magnetization

MRI signal in any practical experiment stems from a large collection of nuclei. Then the formulation must be set in a global pattern. In this section the arrangement of the individual magnetic momentums is outlined.

In a homogeneous system, the bulk magnetization is the sum over the individual magnetic moments present in the sample

$$\vec{M} = \frac{1}{V}\sum_{n=1}^{N_s}\vec{\mu}_n \tag{2.8}$$

where N_s is a vague total number of spins in the object being imaged and V is the volume of the sample. In view of above equation, bulk magnetization is nothing but a summation over the individual magnetic moments. Bloch equation (2.5) is linear in $\vec{\mu}$; thus it can be rewritten as a function of the bulk magnetization

$$\frac{d\vec{M}}{dt} = \gamma\vec{M}\times\vec{B} \tag{2.9}$$

In the sequel, the previous equation is referred as Bloch equation .

In a sample exacted to a magnetic field $B_0\vec{k}$, more spins are aligned along the direction of the field than spins aligned against it. Even though the difference is small, a discernible net magnetization results (see (A.6) in Appendix A.4 for further details).

$$|\vec{M}| \simeq \frac{\gamma^2\hbar^2 B_0 N_s}{4KT_s} \tag{2.10}$$

Therefore, the magnitude of the bulk magnetization is directly proportional to the external magnetic field strength B_0 and inversely proportional to the temperature although it does not represent and adjustable parameter in clinical MRI. The net magnetization origins a magnetic field of a magnitude of the order of 10^{-6} T. This field is considerably small but still measurable if it is turned into the xy plane in order to make visible against the main magnetic field of order of Tesla. That is the role played by the excitation RF pulse, Section 2.2.2.

2.2 Radio frequency pulse

Spins of the same kind rotate at the same frequency, equation (2.7), but out of phase. The sum over all microscopic transverse magnetizations of each spin results in null transverse macroscopic magnetization. RF pulses are short bursts of RF energy which set coherence among spins. Due to resonance the system of protons of the same kind will experience an energy transition to higher level when it is perturbed with a RF pulse tunned at the Larmor frequency of the system of spins. The consequence is the emission of another RF wave at the same frequency. This constitutes the basic MRI signal.

For ease of mathematical operation, rotating frame of reference is introduced in this section along with a brief description of the excitation RF pulse. Finally the reader will find a concise comment on non-uniformities of the main magnetic field.

2.2.1 Rotating frame of reference

It is conceptually and mathematically easier to work in a rotating frame of reference at the rate of the magnetization precessing frequency or Larmor frequency, (2.7). Bloch equation (2.9) is restated into the rotating frame $\{x', y', z'\}$ around the z-axis in the fixed, or inertial system $\{x, y, z\}$.

One can infer directly from Figure 2.2 the trigonometric relation between both systems

$$\begin{cases} \vec{i'} = \cos(\omega t)\vec{i} - \sin(\omega t)\vec{j} \\ \vec{j'} = \sin(\omega t)\vec{i} + \cos(\omega t)\vec{j} \\ \vec{k'} = \vec{k} \end{cases} \tag{2.11}$$

where ω is the angular velocity of the rotating frame.

In the laboratory frame of reference the unit vectors are fixed whereas in the primed or rotating frame of reference they are time-dependent due to rotation. On the other hand, it is in order to recall that in both system $\vec{M}'(t)$ and $\vec{M}(t)$ refer to the same vector quantity so there is no need to distinguish them. The rate of

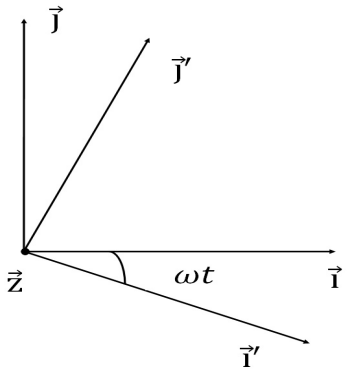

Figure 2.2: *Rotating frame* $\{x', y', z'\}$ *about* z*- axis in the steady frame of reference* $\{x, y, z\}$

change of the bulk magnetization reads

$$\frac{d\vec{M}}{dt} = \frac{dM_{x'}}{dt}\vec{i'} + \frac{dM_{y'}}{dt}\vec{j'} + \frac{dM_{z'}}{dt}\vec{k'} + M_{x'}\frac{d}{dt}\vec{i'} + M_{y'}\frac{d}{dt}\vec{j'} + M_{z'}\frac{d}{dt}\vec{k'} \tag{2.12}$$

Applying the time derivative into equation (2.11) and using the identity $\vec{A}\times\vec{B} = -\vec{B}\times\vec{A}$, the time derivative of the unit vectors reads as follows

$$\begin{cases} di'/dt = \vec{\omega}\times\vec{i'} \\ dj'/dt = \vec{\omega}\times\vec{j'} \\ dk'/dt = \vec{\omega}\times\vec{k'} \end{cases}$$

Substituting into equation (2.12) the time derivative in the stationary frame of reference is

$$\frac{d\vec{M}}{dt} = \frac{d\vec{M'}}{dt} + \vec{\omega}\times\vec{M'}$$

Finally, combining the previous equation with (2.9)

$$\frac{d\vec{M'}}{dt} = \mu\vec{M}\times\vec{B}_{\text{eff}} \tag{2.13}$$

where

$$\vec{B}_{\text{eff}} = \vec{B} + \frac{\omega}{\gamma}$$

is the effective magnetic field in the rotation frame which gather any type of magnetic field exciting the sample. Equation (2.13) manifests that precession will occur around the direction of the effective magnetic field.

2.2.2 Excitation pulse

The inclusion of another magnetic field perpendicular to z'-axis enforces a movement of the bulk magnetization towards $x'y'$ plane. It tracks a spiral down trajectory seen from the laboratory frame, Figure 2.3. The perpendicular magnetic field is denoted by $\vec{B_1}$ and it is described in the following complex notation

$$\vec{B}_1(t) = B_{1x}(t) + iB_{1y}(t) = B_1(t)\mathrm{e}^{-i(\omega_{\mathrm{rf}}t+\phi)}$$

where $B_1(t)$ is the envelope shape, ω_{rf} the pulse frequency and ϕ the initial phase. The previous expression stands for a circularly polarized magnetic field vector in the laboratory frame of reference. For ease of argument the previous circularly polarized field is written as the sum of two linearly polarized fields [34], placed in one of the axis x'

$$\vec{B}_1(t) = B_1(t)\vec{i'}$$

Commonly $B_1(t)$ duration is of the order of micro or milli-seconds whilst its strength is of the order of milli-Tesla.

When the spin system has a unique frequency of $\omega_0 = \gamma B$, and the frequency of the rotation frame of reference is selected at the same frequency ω_0, the system is said to be on resonance. Under this condition Bloch equation has a simple analytical solution, [27]

$$\begin{cases} dM_{i'}/dt = 0 \\ dM_{j'}/dt = M_z^0 \sin\left(\int_0^\tau \gamma B_1(t)\right) \\ dM_{k'}/dt = M_z^0 \cos\left(\int_0^\tau \gamma B_1(t)\right) \end{cases}$$

Now the pulse shape has to be determined. A rectangular pulse applied for a time τ will tip the magnetization of an angle

$$\delta\theta = \gamma B_1 \tau$$

In Section 2.3 it is shown that the signal is higher when the magnetization lies on the xy-plane so that a 90° flipped bulk magnetization is desired.

This review just drew the basic mechanism of excitation but it is oversimplified. There is a great number of B_1 pulses and special attention has to be paid to wave envelope since it explains how and how long it excites the magnetization. For broader description of this theme, see [35].

2.2.3 Off resonance effect

Although essentially right, the previous discussion just sketches an ideal model. Commonly the magnetic field B_0 is neither homogeneous nor constant in strength

along the area to be imaged due to imperfections in the main magnet. Inhomogeneities cause spins to rotate at a different frequency. It leads to a phase shift, artefacts [36] and even loss of the signal. However there exists models [37,38] that partially offset these non-idealities.

Complete homogeneity limitation is rarely fulfilled and typically acceptable value of inhomogeneity is 5ppm (part per million) in a small imaging volume, called isocenter [31]. Moreover, chemical shift effect takes into account the fact that nuclei are attached to different chemical environment in a spin system. It describes that those nuclei attached to molecules may be affected by the weak magnetic field created by the spinning electrons. Chemical shift depends upon the specific location but it is accepted that, for instance, in fat it is about 3.35 ppm shift in Larmor frequency compared to water [27].

In despite of these effects, in this thesis work it is considered a system free of error along the entire derivation.

2.3 MRI Signal

When the RF pulse are tunned at the resonance frequency a large number of protons flip into the transversal plane with respect to the main magnetic field direction. Once the pulse is turned off the individual protons start to flip over and precess toward the magnetic field direction at the Larmor frequency of the excited spins. On basis of Faraday's induction law [40, 41], any coil resonating at the same frequency can be employed as a receiver. The induced voltage in the RF antenna collecting the magnetic flux sweeping through is named Free Induction Decay (FID). This signal is the basic source of raw data that ultimately forms the image. In the sought of this signal expression, it is first necessary to understand the relaxation process after the excitation. For our purpose it is sufficient to assume that the RF pulse is instantaneous and then consider the effect of the main magnetic field together with the relaxation terms. Some assumptions have been introduced to alleviate the complexity of the signal equation.

2.3.1 Relaxation

A phenomenological description of the relaxation terms arising from the interaction of a proton with its neighboring atoms complete the elementary Bloch equation (2.9). Once the excitation pulse is switched off the system returns to the original state of minimum energy,i.e., spin system tends to aligned with the main magnetic field. The distinction of two relaxation phenomena is in order. Both phenomenas are used to enhance tissues contrast in the image.

- *Spin-Lattice interaction.* Alignment of the magnetic moment with the magnetic field is feasible insomuch as the system can exchange energy with its surroundings. The population of the two states along with the energy transfer between protons and lattice lead to the recovery of the longitudinal magnetization following the next equation

$$\frac{dM_{z'}}{dt} = -\frac{M_{z'} - M_z'^0}{T_1} \tag{2.14}$$

  where T_1 is the experimental spin-lattice relaxation time. Atomic level details that lead to the decay are far beyond of the matter of the present discussion. Interested reader is referred to [24]. It is sufficient to know that the longitudinal recovery is characterized by the parameter T_1 and is caused by the exchange of energy from nuclei to their surrounding lattice. The rate at which this occurs is an exponential tissue-dependent process, Table 2.2

- *Spin-Spin interaction.* This important decay mechanism stems from the effect of local fields from neighboring spins. The precession frequency depends directly on magnetic field so that variations in local field lead to complicated rotation frequency and reduction of the net magnetization. The empirical rate of decay is gathered in the T_2 relaxation time.

$$\frac{dM_{x'y'}}{dt} = -\frac{M_{x'y'}}{T_2} \tag{2.15}$$

Table 2.2: *Representative values of relaxation parameters for hydrogen components of different human body tissues at* $B_0 = 1.5T$ *and* $37°C$.

Tissue	T_1 (ms)	T_2 (ms)
Gray matter	950	100
White matter	600	80
Muscle	900	50
Fat	250	60
Blood	1200	100-200

Bringing together all the relevant terms for the precession equation (2.13) along with the relaxation effects, (2.14) and (2.15), Bloch equation (2.9) takes the following form

$$\frac{d\vec{M}}{dt} = \gamma \left(\vec{M} \times \left(\vec{B}_0 + \vec{B}_1 \right) - \frac{M_{x'}\vec{i'} + M_{y'}\vec{j'}}{T_2} - \frac{(M_z - M_z^0)\,\vec{k}}{T_1} \right) \tag{2.16}$$

The equation above is the starting point for both the relaxation and excitation equation. The latter can be obtained by dropping the relaxation factors since it is usually accepted that the length of B_1 is much shorter than the relaxation terms T_1 and T_2. Whereas magnetization magnitude is constant in the excitation process, it is a spiral upwards motion for the relaxation stage, Figure 2.3.

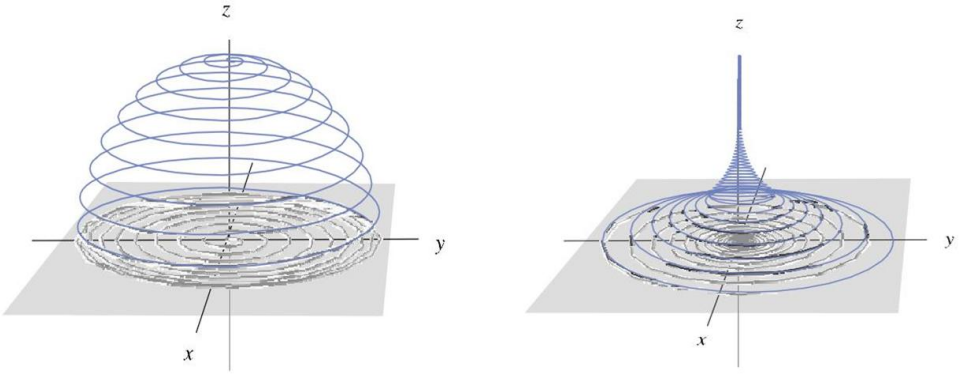

Figure 2.3: *Excitation (left) and relaxation (right) from the laboratory frame of reference. During excitation the magnitude of the bulk magnetization $\vec{M}$ is constant while it decreases along the relaxation process. Image extracted from [29].*

The solution of the relaxation effect is written in the complex notation for the x' and y' axis following the same procedure as it was done in Section 2.1.3.

$$M_+(t) \equiv M_x(t) + iM_y(t) \tag{2.17}$$

Then the solution of equation (2.16) for the transversal component yields

$$M_+(\vec{r}, t) = \mathrm{e}^{-\frac{t}{T_2(\vec{r})}} \mathrm{e}^{-i\omega_0 t} M_+(\vec{r}, 0) \tag{2.18}$$

which is composed of a sinusoidal oscillation modified with a decay factor $\mathrm{e}^{-1/T_2(\vec{r})}$. The spatial dependence of the decay factor introduces the effects of local magnetic field inhomogeneities due to neighboring spins and lack of coherence in the initial longitudinal magnetization.

The longitudinal part takes the form

$$M_z(t) = M_z(0)\mathrm{e}^{-t/T_1(\vec{r})} + M_0(1 - \mathrm{e}^{-t/T_1(\vec{r})}) \tag{2.19}$$

2.3.2 Signal detection

According to Faraday's law the voltage induced in a coil is

$$V(t) = -\frac{\partial \Phi(t)}{\partial t}$$

where $\Phi(t)$ is the magnetic flux through the coil of area S

$$\Phi(t) = \int_{\text{coil}} \vec{B}(\vec{r}, t) \cdot d\vec{S}$$

where $\vec{B}(\vec{r}, t)$ is the magnetic field sweeping the coil. Hence either a time-dependent magnetic flux or a coil of varying area will produce induced voltage in the coil. The induced voltage can also be expressed as a function of the magnetic field per unit of current $\vec{B}_r$ originated by the detection coil and the bulk magnetization $\vec{M}(\vec{r}, t)$. The field $\vec{B}_r(\vec{r})$ is known as received field and the spatial dependence refers to the fact that in general the receiver coils may carry non-uniformities. Detailed derivation of the following equation can be found in Appendix A.

$$V(t) = -\frac{\partial}{\partial t} \int_{\text{coil}} \vec{M}(\vec{r}, t) \cdot \vec{B}_r(\vec{r}) d\vec{r} \tag{2.20}$$

where the components of the bulk magnetization are given by equations (2.18) and (2.19).

Let $S(t)$ denote the signal. The fundamental MR signal is known to be proportional to the detected induced voltage [27] so that from (2.20) it results

$$S(t) \propto -\frac{\partial}{\partial t} \int \left[B_{rx}(\vec{r}) M_x(\vec{r}, t) + B_{ry}(\vec{r}) M_y(\vec{r}, t) + B_{rz}(\vec{r}) M_z(\vec{r}, t) \right] d\vec{r}$$

M_z is considered to vary slowly compared to transversal magnetic moment as can be inferred from equations (2.18), (2.19) and Table 2.2. Using the complex representation (2.17) and taking the time derivative inside the integrand, the signal yields

$$S(t) \propto \int - \left(\frac{1}{T_2(\vec{r})} + i\omega_0 \right) \mathrm{e}^{-t/T_2(\vec{r})} \left[B_{rx}(\vec{r}) \mathrm{Re} \left(M_+ (\vec{r}, 0) \, \mathrm{e}^{-i\omega_0} \right) + B_{ry}(\vec{r}) \mathrm{Im} \left(M_+ (\vec{r}, 0) \, \mathrm{e}^{-i\omega_0} \right) \right] d\vec{r}$$

For static fields at the scale of Tesla the Larmor frequency is typically four orders-of-magnitude higher than values of $1/T_2$ and the term derived from the decay factor can be neglected

$$S(t) \propto \omega_0 \int \mathrm{e}^{-t/T_2(\vec{r})} \left[B_{rx}(\vec{r}) \mathrm{Re} \left(i M_+ (\vec{r}, 0) \, \mathrm{e}^{-i\omega_0} \right) + B_{ry}(\vec{r}) \mathrm{Im} \left(i M_+ (\vec{r}, 0) \, \mathrm{e}^{-i\omega_0} \right) \right] d\vec{r}$$

The previous expression can be further simplified by writing the cartesian components of the received magnetic field in the following polar parametrization

$$B_{rx}(\vec{r}) \equiv B_\perp(\vec{r}) \cos\theta_B \qquad B_{ry}(\vec{r}) \equiv B_\perp(\vec{r}) \sin\theta_B$$

where $B_\perp$ is the magnitude $B_\perp = \sqrt{B_{rx}^2 + B_{ry}^2}$. The same notation is also applied for the components of the bulk magnetization. In addition, the following trigonometric identity $\sin(\alpha + \beta) = \sin\alpha\,\cos\beta + \cos\alpha\,\sin\beta$ leads to the following signal equation

$$S(t) \propto \omega_0 \int \mathrm{e}^{-t/T_2(\vec{r})} M_\perp(\vec{r}, 0) B_\perp(\vec{r}) \sin\big(\omega_0 t + \theta_B)\big) d\vec{r}$$

The rapid oscillation at Larmor frequency ω_0 in the previous signal equation is generally not supported by the common electronic devices and demodulation is required in order to suppress this dependency. During this process signal is multiplied by a sine and cosine of frequency close to Larmor frequency, $\Omega = \omega_0 + \delta\omega$ and lowpass-filtered [42], [43]. Signal is then divided into two channels: real and imaginary [26].

A further assumption is in order. The received magnetic field is assumed to be spatial independent. It means that coils are considered homogeneous and uniform. Furthermore, inhomogeneities at atomic level are neglected; hence the decay factor e^{-t/T_2} is spatially independent. The latter assumption is of minor consequence and it lightens the mathematical model considerably. Taking these terms outside the integral, the complex signal reads

$$S(t) \propto \omega_0 \mathrm{e}^{-t/T_2} B_\perp \mathrm{e}^{-i\theta_B} \int M_\perp(\vec{r}, 0) \mathrm{e}^{i\delta\omega t} d\vec{r}$$

Finally, enclose all terms before the integral in a constant κ.

$$S(t) = \kappa \int M_\perp(\vec{r}, 0) \mathrm{e}^{i\delta\omega t} d\vec{r}$$

This is the so-called Free Induction Decay (FID) signal, basic signal in MRI and mother of many imaging strategies.

2.3.3 Gradients

In order to differentiate the distinct constituents of a target plane, the spin density distribution has to be encoded into the MRI signal. From this capability, images with excellent contrast for soft tissue are achieved. For this idea to be accomplished the signal has to be spatially dependent, i.e., nuclei have to precess at different frequency at different spatial location. Here gradients $\vec{G}(t)$ come into play.

Let $\vec{B}(\vec{r}, t)$ be a explicit time and spatial-varying magnetic field, both dependencies coming from the gradient $G(t)$.

$$\vec{B}(\vec{r}, t) = xG(t)_x\vec{i} + yG(t)_y\vec{j} + (B_0 + zG(t)_z)\vec{k}$$

According to Larmor frequency, equation (2.7), the angular frequency becomes time and spatial dependent with the inclusion of gradients and the subsequent frequency term derived from them $\omega_G(\vec{r}, t)$

$$\omega(\vec{r}, t) = \omega_0 + \omega_G(\vec{r}, t)$$

The spatial information can be encoded in two different ways: frequency and phase encoding. Both can be used simultaneously or individually in the same sequence. The first modality modifies the frequency switching a gradient for a

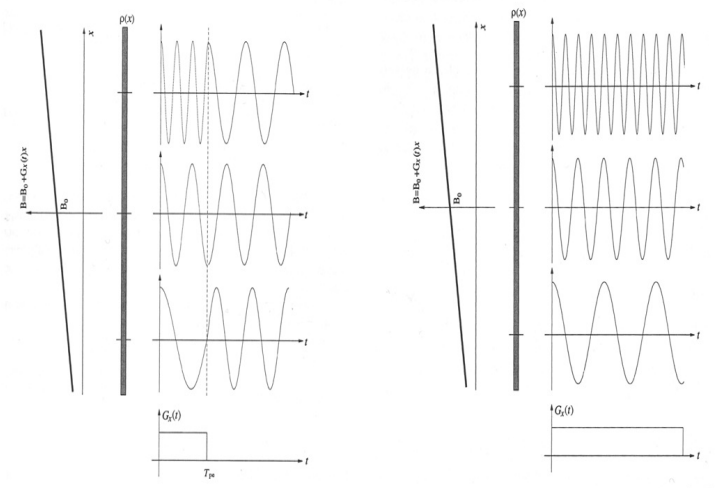

Figure 2.4: *Phase and frequency encoding. The presence of spatially varying magnetic field for a determined time will affect the phase or frequency of the signal. Image selected from [27].*

determined time while the second affects the phase by a short life gradient, Figure 2.4.

$$\phi_{\text{fr}}(\vec{r}, t) = \int_0^t \gamma(B_0 + \vec{r} \cdot \vec{G}(t'))dt'$$
$$\phi_{\text{ph}}(\vec{r}) = \gamma(\vec{G} \cdot \vec{r})T_{pe}$$

The combined effect of frequency $\phi_{\text{fr}}(\vec{r}, t)$ and phase encoding $\phi_{\text{ph}}(\vec{r})$ shift is gathered under a unique term $\phi(\vec{r})$ in the FID signal (2.3.2)

$$S(t) = \kappa \int M_{\perp}(\vec{r}, 0)\mathrm{e}^{i(\delta\omega\, t + \phi(\vec{r}, t))}d\vec{r} \tag{2.21}$$

At this point it is important to notice that the term $M_{\perp}$ is related to the total number of available spins to contribute to the signal, N_s in equation (2.10). This

feature suggests that signal can be written as a function of the spin density $\rho(\vec{r})$. Further assume that demodulation is performed at Larmor frequency, $\delta\omega = 0$ in the equation (2.3.2).The final signal equation takes the following form

$$S(t) = \kappa \int \rho(\vec{r}) \mathrm{e}^{i\phi(\vec{r},t)} d\vec{r} \tag{2.22}$$

Gradient devices

Gradients are physically constrained and not any sampling trajectory is feasible. A concise characterization covers just two of their main properties

- *Maximum gradient strength.* It is the maximum change of field strength per meter along the scanner table. This factor determines the maximum achievable resolution because thin slice selection requires high gradients. Some sampling schemes may also demand high gradient.

- *Rise time.* It is the time required to reach the maximum amplitude. Fast changes are desirable but they are also restricted by regulation due to danger of nerve stimulation [4].

Faults in the gradients can result in geometric distortions in the MR image. Fast scan requires powerful motor to move the gradients and generally more expensive material in addition to high power consumptions. Then sampling density is limited by scanner gradient switching speed and precision properties.

Yet another limiting factor comes from basic circuit analysis

$$U = L\frac{dI}{dt} + RI \tag{2.23}$$

where U is the voltage, I is the current, R is the resistance and L the auto inductance. Voltage and current are limited by design which in turn restraint the rise time and feasibility of k-space patterns.

2.4 Spin Echo sequence

In the present section an explanatory scheme of the microscopic origin of the FID signal is displayed along with an introduction of the way that data is obtained.

A signal is generated and sampled in a pulse sequence, a series of RF pulses and gradients synchronized in time. Among all possible sequences the earliest imaging method, this is, Spin Echo is introduced [39]. For a complete coverage of this and others imaging schemes any MRI book is of help, see [24–32, 35].

Shortly, Spin Echo sequence extends the utility for acquiring information of the FID signal partially overcoming the destructive phase caused by the diversity in rotation frequency of the individual spins. As it was expounded, frequency is directly proportional to the experienced magnetic field causing the precession, equation (2.7). Field inhomogeneities fan out the precessing frequencies at a rate of T_2^*, Figure 2.5. This decay gathers effects due to internal and external changes in the magnetic field. The read-out gradient is a frequency encoding gradient applied during the period when the echo is active to sample data.

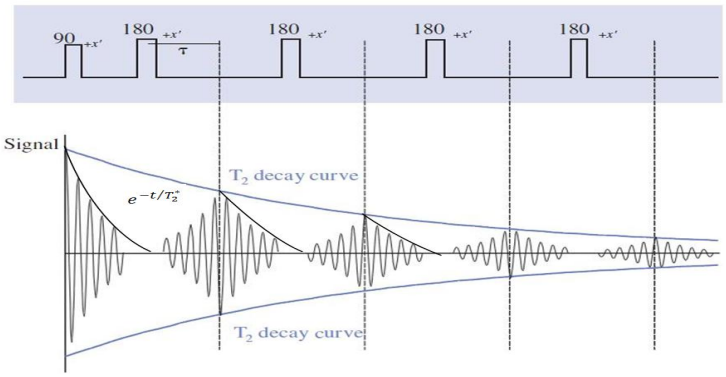

Figure 2.5: *Train Spin Echo sequence. Echoes appear as consequence of π-pulses. FID decays at T_2 rate while individual echos faint out at T_2^*. Image obtained from [29]*

.

Spin Echo is based on the application of an excitation or $\pi/2$-pulse tilting the magnetization into the transversal plane followed by a π-pulse at time T_E, Figure 2.6 that will refocus the spins. After the first instantaneous signal, spins are spread out with the help of the dephasing gradient. The π-pulse will advance the "slow" proton ahead of the "faster" in such a way that right after the π-pulse the proton lagging behind will be the "fast" and both will meet at time $t = 2T_E$. The ensuing signal, known as echo, is ready to be sampled once again. Subsequent π-pulse will produce new echoes of decreasing strength since local field inhomogeneities associated to spin-spin interaction cannot be mitigated and the signal will eventually faint out at a decay rate of T_2.

2.5 Sampling trajectories and k-space

In this section it is made explicit the relation between the image and the Fourier transform in the sense of MRI. Three sampling trajectories in the data domain or k-space are showed.

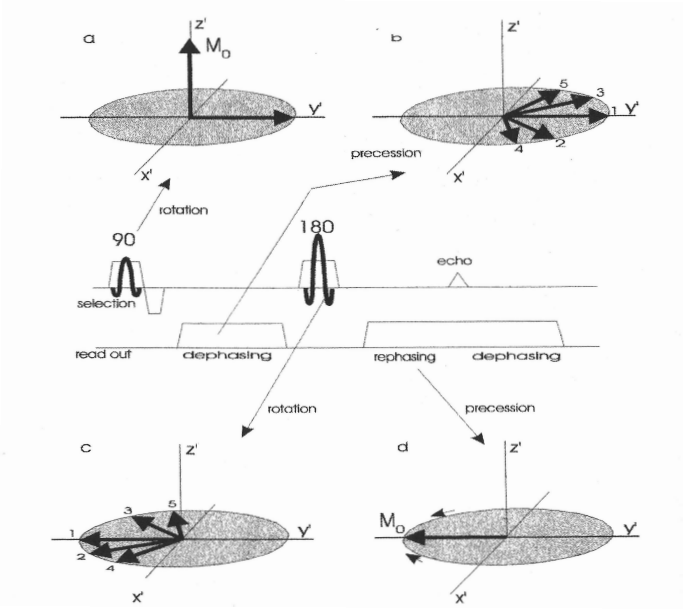

Figure 2.6: *Spin Echo. Effect of the gradients on a group of individual spins. Sketch obtained from [31].*

Let be $\vec{k}(t)$ the following three-dimensional time-dependent vector

$$\vec{k}(t) = \frac{\gamma}{2\pi} \int_0^t \vec{G}(t')dt' \tag{2.24}$$

whose components depend linearly on the gradient value $\vec{G}(t)$. Substituting the previous definition in equation (2.22), the signal equation reads

$$S(\vec{k}) = \int_{-\infty}^{\infty} \rho(\vec{r}) e^{-i2\pi\vec{k}\cdot\vec{r}} d\vec{r} \tag{2.25}$$

The relation between the acquired signal and the image can be emphasized. Given a three-dimensional object, a two-dimensional image can be thought of as a projection of the object into a plane

$$I(x,y) = \int_{-\infty}^{\infty} \rho(x,y,z)dz$$

where limits in the integral must be set at the area of nonzero spin density. Writing explicitly the two-dimensional k-vector in (2.25) the signal equation remains

$$S(k_x,k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)e^{-i2\pi(k_x x+k_y y)}dxdy$$

This equation highlights that the relation between the k-space data and image is the Fourier transformation. From definition (2.24), the k-space can be charted through any direction by modifying the gradients with respect to the time. Hence there is great flexibility to design a sampling trajectory with the unique constraint that any trajectory must departure from the origin of the k-space since at $t = 0$ we have $k(0) = 0$ in (2.24). Inverse Fast Fourier Transfer (IFFT) is traditionally used to reconstruct the image. However a large number of frequencies have to be sampled in order to meet condition imposed by the reconstruction transform. Shannon-Nyquist sampling rate (sampling rate must be at least twice the maximum frequency present in the signal) constitutes the traditional bounding factor. Chapter 3 explains the proposed approach to overcomer this limit.

From the k-space definition it can be seen that it has units of inverse of distance m^{-1} or spatial frequency. Low frequency at the central area represents rough details in the image. Moving away from the center, fine details or resolution are found [44]. However, most of the information lie on the center and sampling strategies are inclined to sample more densely this area. In addition, applications in which time is a crucial factor such as moving target imaging, high frequencies cannot be visited by the sampling strategy and consequently this images typically suffer from low resolution. Generally speaking sampling trajectory is usually a trade-off between speed and image quality.

Advantages of three k-space sweeping strategies, namely, cartesian, radial and spiral are explored. They will be set in the spin echo technique although other signal method such as gradient echo would lead to similar results.

2.5.1 Cartesian sampling

This is the most common technique because a simple inverse Fourier transform of raw data is sufficient to reconstruct the image. This technique collect data with equal density along the entire k-space.

Figure 2.7 shows a multiple line scanning strategy where solid lines indicate the current scanned line and dotted lines represent the subsequent lines, achieved after a certain repetition time long enough to let the longitudinal magnetization to recover. After that a new cycle starts over. Gradient in the z-axis is responsible of the slice selection. Gradient in the y-axis is a phase encoding gradient and it is in charge of moving the vector in the k-space vertically, Figure 2.7. Finally gradient in x-direction is the read-out gradient . It is constant throughout the sequence.

The line selection (point A in Figure 2.7) can be written mathematically as follows

$$\left.\begin{array}{l} k_x = \frac{\gamma}{2\pi} G_x (t - t_0) \\ k_y = \frac{\gamma}{2\pi} n G_y (t - t_0) \end{array}\right\} t_0 < t < T_{acq}/2 + t_0$$

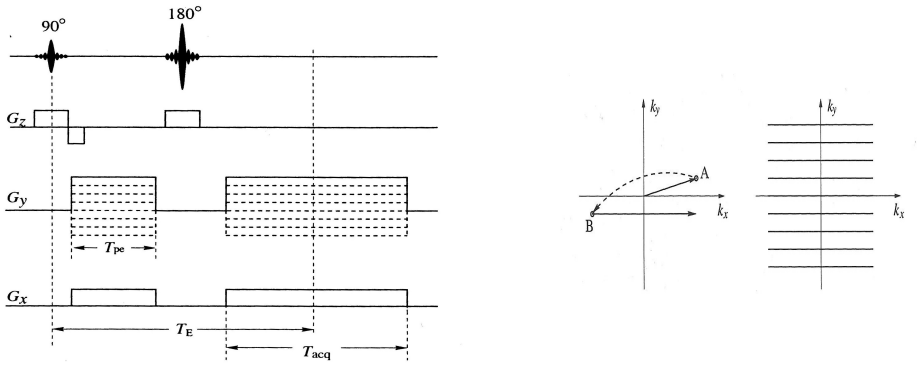

Figure 2.7: *Spin Echo in cartesian sequence and k-space line sweeping. Image obtained from [27]*

where t_0 is the time at which gradients are switched on and n is a multiplicative factor for the amplitude of the gradient in the vertical direction on the basis that all the lines are equally spaced. At time $T_E/2$ a π RF pulse maps the vector to the left side of the k-space. Subsequently a horizontal line travels from left to right sampling the signal. This second part can be described as follows

$$\left.\begin{array}{l} k_x = \frac{\gamma}{2\pi} G_x (t - T_E) \\ k_y = \frac{\gamma}{2\pi} n G_y (t - T_E) \end{array}\right\} |t - T_E| < \frac{T_{acq}}{2}$$

2.5.2 Radial sampling

Historically, radial sampling was one of the first to be implemented in clinical practice. Due to the radially decreasing coverage this technique sacrifices resolution.

Let$\{G\cos(\phi_0), G\sin(\phi_0)\}$ be the values of the gradients magnitudes. It is straightforward to visualize the k-space trajectory as polar coordinates of the spokes of a circle, Figure 2.8. Mathematically

$$\left.\begin{array}{l} k_x = \frac{\gamma}{2\pi} G\cos(\phi_0)(t - t_0) \\ k_y = \frac{\gamma}{2\pi} G\sin(\phi_0)(t - t_0) \end{array}\right\} t_0 < t < \frac{T_{acq}}{2} + t_0$$

The π-pulse mirrors the vector $\vec{k}$ through the origin as indicated in Figure 2.8 and sampling follows with the same k-space values. For the radial strategy both gradients are read-out gradients.

2.5.3 Spiral sampling

Spiral sampling is considered as an ultrafast method because it collects data in a single excitation, Figure 2.9. It has the advantage that rapid values acquired in

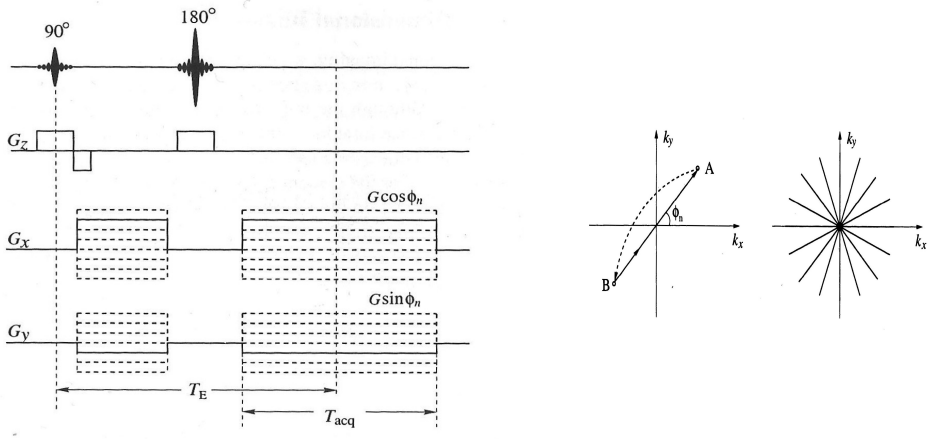

Figure 2.8: *Spin Echo in radial sequence (left) and one single spoke sampling k-space (right). From [27]*

the center of the k-space have minor T_2^* effect, Figure 2.5. This strategy reduces remarkably the sampling time and allows the gradients to move in a more natural way avoiding the rapid changes. As in radial sampling this technique is expected to have low resolution due to the scant sampling in the outer k-space where high frequencies lie. It is also more versatile because different types of spirals can be designed although constraints of the gradients limit the freedom.

A spiral scheme in the k-space is described by:

$$k_x(t) = \omega(t)\cos\omega(t)$$
$$k_y(t) = \omega(t)\sin\omega(t)$$

where $\omega(t)$ is to be determined in the sequence design. One approach is a linear function of time $\omega(t) = \omega_0 t$. From (2.24) it follows that gradient expression is found from the time derivative of the $\vec{k}$ values. Then the required gradient yields

$$G_x(t) = A\cos\omega_0 t - At\omega_0\sin\omega_0 t$$
$$G_y(t) = A\sin\omega_0 t + At\omega_0\cos\omega_0 t$$

One of the limitations on the gradients can be found here. It is known that current supplied to the gradient coil and the gradient magnitude are related through the sensitivity C [31]

$$I = \frac{G}{C} = \frac{2\pi}{\gamma}C^{-1}\frac{d|\vec{k}(t)|}{dt}$$

and turning to equation (2.23) without taking into account the resistance, leads

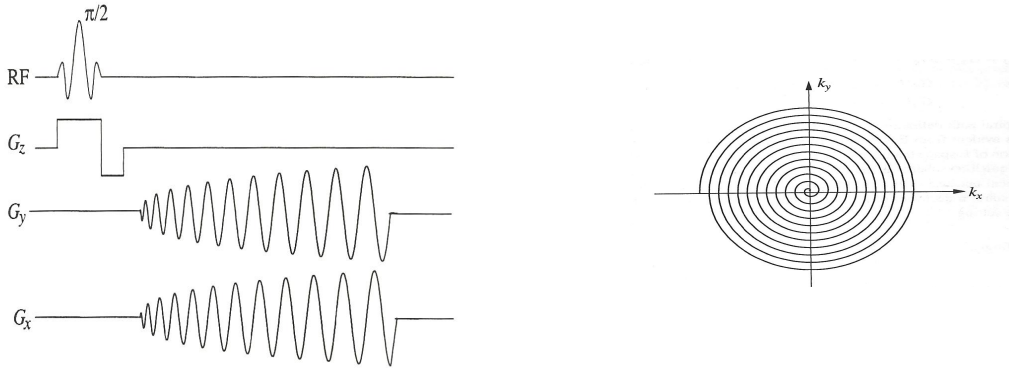

Figure 2.9: *Spiral sequence and k-space charting. Scheme got from [27]*

to the following voltage

$$U = L\frac{dI}{dt} = C^{-1}L\frac{dG(t)}{dt} = \frac{2\pi}{\gamma}C^{-1}L\frac{d^2|\vec{k}(t)|}{dt^2}$$

These two equations point out that the outermost region may be restricted because it requires high intensity values and acceleration is limited by the voltage. In addition, the idea of sampling a T_2^* decay rate signal before it dies out is ambitious in the sense that it requires high speed and acceleration.

2.5.4 Rapid imaging techniques

Common approaches to save time in real MRI scanner exploit the redundancy in Fourier domain. The half-Fourier approach only samples half k-space on the basis of the conjugate symmetry property of the Fourier transform of real-valued functions

$$F(-k) = F^*(k)$$

where $F^*(k)$ is the complex conjugate of $F(k)$. Intuitively this method offers a half-time scanner but its main assumption is not met in the presence of non-idealities such as phase shifted originated from motion artefacts. Moreover noise breaks the Fourier symmetry.

Yet another already implemented method is parallel MR imaging. An array of coils is placed along the scanner and reduced data set in the phase encoding direction is acquired in every coil. Later data from the array of coils is combined. A weighting factor according to the position of the coil related to the imaged body is necessary to reconstruct the image via inverse Fourier transform. In cartesian

sampling the gain in speed is directly proportional to the reduction of phase-encoding steps.

2.6 Discrete sampling

It was shown that signal and image are Fourier pairs, meaning that the image is reconstructed through a simple inverse Fourier Transform (2.25)

$$I(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(k_x,k_y)\mathrm{e}^{i2\pi(k_x x+k_y y)}dk_x dk_y$$

Sampling is of discrete nature and the previous equation needs to be modified. When only a set of point $s_d(k_x,k_y)$ are available, an approximate image $\hat{I}(x,y)$ rather than the target image is obtained. Any type of discrepancy between $I(x,y)$ and $\hat{I}(x,y)$ is called artefact. Assume that $2n$ points of the signal are sampled in a cartesian grid for time T_{acq} along $2m$ horizontal lines, Figure 2.7.
Sampling a continuous signal is equivalent to multiply it by a series of delta spikes function, i.e, a completely localized function of unit value at certain equispaced set of points and zero elsewhere. According to this idea the previous equation reads

$$\hat{I}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{p=-n}^{n-1}\sum_{q=-m}^{m-1} s(k_x,k_y)\delta(k_x-p\Delta k_x)\delta(k_y-q\Delta k_y)\mathrm{e}^{i2\pi(k_x x+k_y y)}dk_x dk_y$$

Under this approach the integrals are just evaluated at point where deltas are non-zero

$$\hat{I}(x,y) = \sum_{p=-n}^{n-1}\sum_{q=-n}^{n-1} s_d(p\Delta k_x,q\Delta k_y)\mathrm{e}^{+i2\pi(p\Delta k_x x+q\Delta k_y y)}$$

This equation is known as the Inverse Discrete Fourier Transform (IDFT).

Although a complex image has little physical meaning it is worth notice that $\hat{I}(x,y)$ is not necessary real. Errors in data acquisition, demodulation or signal processing will lead to a phase shifted k-space. Phase in the reconstructed image will arise from the mentioned situation. Then, commonly MR images are magnitude images, i.e, absolute value of the output function.

It is necessary to introduce a discrete version of the Fourier transform (2.25) in order to move from a mathematical model to the actual numerical implementation. For a NxN digital image I_d, it is straightforward from previous relation that the direct DFT is written as

$$s_d(k_x,k_y) = \sum_{p=-N}^{N-1}\sum_{q=-N}^{N-1} I_d(p\Delta x,q\Delta y)\mathrm{e}^{-i2\pi(p\Delta x k_x+q\Delta y k_y)}$$

where the summation cover all the pixels in the image.

From the above discussion one can discern that the reconstructed image is spatially periodic due to sine function periodicity

$$\hat{I}(x, y) = \hat{I}(x + 1/\Delta k_x, y + 1/\Delta k_y)$$

Thus periodicity is given by the reciprocal of the spacing of the delta functions. It is the well-known Nyquist sampling conditions and violation results in wraparound artefact. Space between adjacent samples depends on the sampling scheme and therefore Nyquist condition is sampling dependent. As it will become evident in Chapter 5, undersampling will result in this type of frequency misleading.

Chapter 3

Problem setting

The k-space sampling strategies introduced in Section 2.5 and the relation of the signal with the image through the Fourier transform, suggest that collected data is equivalent to selecting a subset of frequencies in the Fourier domain. From such data set, the image reconstruction is aimed. Fourier transform is a linear operator, hence data acquisition is represented by matrix multiplication. The $N \times N$ image is stacked as a column-wise array $\mathbf{x} \in \mathbb{C}^{N^2 \times 1}$. It is the array of unknowns. Likewise the measurement vector remains: $\mathbf{y} \in \mathbb{C}^{m \times 1}$. Then the measurements are linearly obtained from Fourier domain of $\mathbf{x}$ through the measurement matrix $F_s = \mathcal{M}F \in \mathbb{C}^{m \times N^2}$, where $F \in \mathbb{C}^{N^2 \times N^2}$ is the fully sampled DFT matrix operator and $\mathcal{M}$ is the undersampling matrix which can be thought as a binary matrix that select only certain frequencies from the whole k-space. The measurement vector yields

$$\mathbf{y} = F_s \mathbf{x} \tag{3.1}$$

Typically the undersampling matrix $\mathcal{M}$ is chosen to meet the Nyquist sampling limit and the image is reconstructed through the common IFFT algorithm. However, time required to collect data in MRI is very slow and there exists high motivation in the medical sector to reduce the scanning time. In the cartesian sampling shown in Section 2.5.1, scanning time is proportional to the number of acquired lines in the read-out direction. Similarly, it is proportional to the number of sampled spokes in the radial sampling, Section 2.5.2. An obvious strategy then to reduce scanning time is to reduce the number of samples. However when data domain is highly undersampled, IFFT results in a image suffering from artefacts.

Reconstructing the image $\mathbf{x}$ from a short data set $\mathbf{y}$ in (3.1) is an ill-conditioned task. Traditionally, regularization is employed to overcome this situation. In the recent years the novel sampling technique Compressed Sensing (CS) has arrived with significant contributions in many fields. This technique is built upon the fact that natural images can be represented by few non-zero coefficients in a suitable basis such as wavelets. Then it is possible to find these coefficients sampling the

image in another basis; sinusoid basis or Fourier transform in the case of MRI. Both techniques exploit the spatial redundancy of the natural images in diverse transformations such as wavelets [16–18] , DCT [18], Singular Value Decomposition (SVD) [20] or discrete difference [18, 21].

In the sequel the Shepp-Logan Phantom, Figure 3.1, is used as the target image. It is a piece-wise image with several ellipses meant to represent different areas with distinguishable spin density. It was used first as a standard head in CT simulation.

Figure 3.1: *Shepp-Logan Phantom*

In the present chapter it is showed the idea of sparsity for the Shepp-Logan phantom as well as the proposed solution to reconstruct the target image from reduced data set withot degrading image quality. In Chapter 5 the benefits of using the l_1-norm of the sparsifying transform as the regularization term are presented. In this thesis work two regularization terms are employed and compared: wavelets and total variation.

3.1 Sparse representation

Let $\mathbf{\Psi}$ be an orthonormal basis decomposition $\mathbf{\Psi} = \{\Psi_1, \Psi_2, \ldots, \Psi_{N^2}\}$ such as Discrete Wavelet Transform (DWT), Sections 3.1.1, or DCT . Then from linear algebra, an image or signal $\bar{\mathbf{x}}$ can be expanded as follows

$$\bar{\mathbf{x}} = \sum_{i=1}^{N^2} \alpha_i \Psi_i \tag{3.2}$$

where $\boldsymbol{\alpha}$ is the vector of coefficients of $\bar{\mathbf{x}}$ in the basis $\boldsymbol{\Psi}$

$$\alpha_j = \langle \bar{\mathbf{x}}, \Psi_j \rangle$$

An image $\bar{\mathbf{x}}$ of length N^2 is said to be sparse in a convenient basis $\boldsymbol{\Psi}$ when it can be represented only by significant fewer s terms, with $s << N^2$. Mathematically, sparsity assures that $\bar{\mathbf{x}}$ is s-sparse when at most s terms of the vector coefficient $\boldsymbol{\alpha}$ are nonzeros, i.e, $||\boldsymbol{\alpha}||_0 \leq s$ where

$$||\boldsymbol{\alpha}||_0 \equiv |\text{supp}(\boldsymbol{\alpha})|$$

The support is defined as the set of points where the vector entries are nonzero: $\text{supp}(\boldsymbol{\alpha}) = \{j : \langle \bar{\mathbf{x}}, \Psi_j \rangle \neq 0\}$. However barely any real-world image is sparse in a strict sense and more flexible definition is in order. It is said that $\bar{\mathbf{x}}$ is compressible in a given basis $\boldsymbol{\Psi}$ when it is well-approximated with s non-zero coefficient.

Intuitively an image is compressible when reconstruction with just the largest set of coefficients $\boldsymbol{\alpha}_s = \{\alpha_1, \alpha_2, \ldots, \alpha_s, 0, \ldots, 0\}$ results in a minor perceptual loss. For an orthonormal basis it is possible to quantify the compressibility through the error committed when just the s largest terms are taken into account:

$$||\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}||_2 = ||\boldsymbol{\alpha} - \boldsymbol{\alpha}_s||_2$$

Figure 3.2 shows decreasing error versus increasing amount of coefficients in a DWT basis and Discrete Cosine Transform for the Shepp-Logan Phantom, Figure 3.1.

3.1.1 Wavelets

The wavelet transform is a multi-scale representation of an image, i.e., wavelet basis provides a frame to represent signals or images with any desirable resolution. Whilst Fourier Transform or DCT have only sine and cosine functions as the basis set, wavelets basis use functions localized in space. It is a desirable feature for sharp signal representation and in particular for the Shepp-Logan phantom because it is a piecewise constant function with steep jumps, especially for the peripheral ellipse, Figure 3.1.

The image $\bar{\mathbf{x}}(x, y)$ can be written as a linear combination of the expansion function $\psi_{m,n}(x, y)$:

$$\bar{\mathbf{x}}(x, y) = \sum_m \sum_n \alpha_{m,n} \psi_{m,n}(x, y)$$

A detailed description on mathematical basis of wavelet function is beyond the purpose of this thesis. Reader is referred to [45] for detailed coverage of this

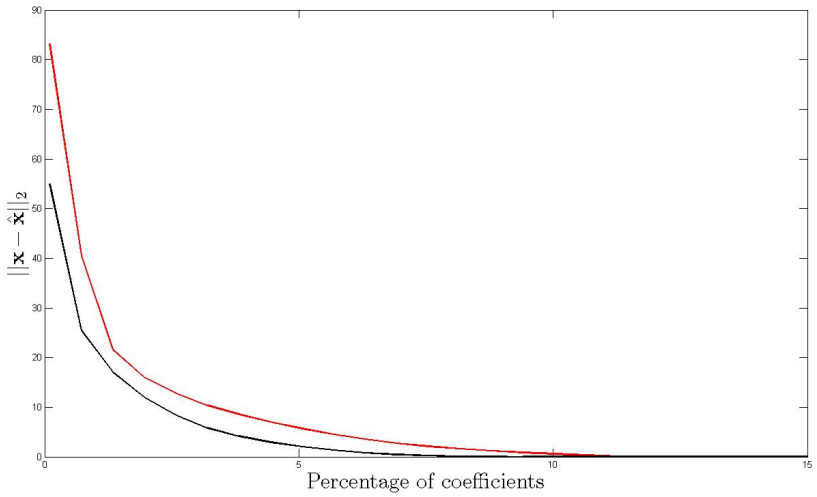

Figure 3.2: *Error in image representation in D4 Daubechies DWT (black) and DCT (red) with increasing percentage of coefficients of the 512x512 phantom image, Figure 3.1.*

theme. It is sufficient for our purposes, however, to know that a wavelet expansion is written as the sum of the so-called mother and father wavelets

$$\phi_{j,m,n}(x,y) = 2^{j/2}\phi(2^j x - m, 2^j y - n)$$
$$\psi^i_{j,m,n}(x,y) = 2^{j/2}\psi^i(2^j x - m, 2^j y - n)$$

where $i = \{H, V, D\}$ stands for the horizontal, vertical and diagonal directional wavelets that maintain the detail in that directions, j is the dilatation coefficient and m and n the translation coefficients of the basis function. Further, the discrete wavelet transform of an image $\bar{\mathbf{x}}(x,y)$ of size N x M is a decomposition operation that produces four output images: approximation, horizontal, vertical, and diagonal detail images. A five-scale phantom image is found in Figure 3.3.

For a discrete function $\bar{\mathbf{x}}(x,y)$ the DWT is

$$W_\phi(j_0,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\bar{\mathbf{x}}(x,y)\phi_{j_0,m,n}(x,y)$$
$$W^i_\psi(j,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\bar{\mathbf{x}}(x,y)\psi^i_{j_0,m,n}(x,y)$$

and the series expansion of the function $\bar{\mathbf{x}}(x, y)$ yields

$$\bar{\mathbf{x}}(x,y) = \frac{1}{\sqrt{MN}} \sum_m \sum_n W_\phi(j_0, m, n)\phi_{j_0,m,n}(x,y) + \\ + \frac{1}{\sqrt{MN}} \sum_{i=H,V,D} \sum_{j=j_0}^{\infty} \sum_m \sum_n W_\psi^i(j, m, n)\psi_{j_0,m,n}^i(x,y)$$

The original function can be written as a linear combination of the wavelet functions. Setting a threshold in the vector of coefficients one controls the details of the image when it is transformed back into the pixel domain. Hence wavelet transformation is a versatile tool in denoising images corrupted by high detailed noise just dropping off high resolution coefficients.

There exists certain freedom in the basis choice and selecting a particular wavelet is mainly an application-dependent problem. From the time being, the wavelet basis employed is the Daubechies D4 although non significant differences in the sparsity were found compared to other orthogonal wavelet basis such as symmlet and Coiflet. All wavelet operations were performed with the WaveLab softwave from Stanford University [46].

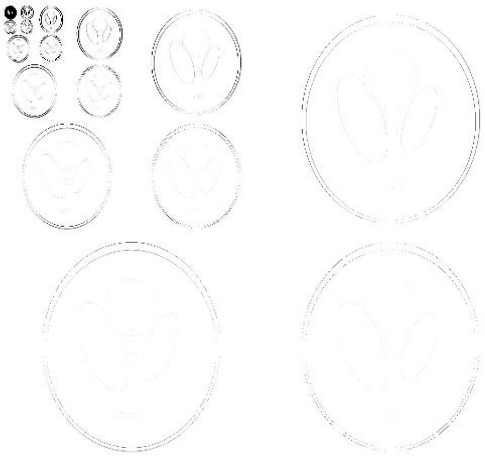

Figure 3.3: *Five-scale Wavelet representation. Horizontal, vertical and diagonal details of the Sheppan-Logan phantom.*

3.1.2 Discrete Difference

As stated above, Shepp-Logan phantom is a piecewise constant image so just a simple discrete finite difference can work as sparse transformation. In this sense it is related to edge detection in digital image processing [44]. The discrete difference is operated over the set of 4-neighbors pixels denoted by $\mathcal{N}_j$ for pixel x_j, Figure 3.4.

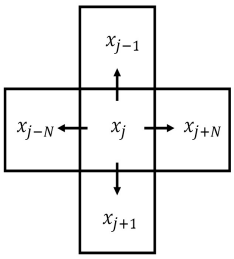

Figure 3.4: *Scheme of the $\mathcal{N}_4$ neighboring pixels of x_j pixel.*

3.2 Image reconstruction via l_1 regularization

For highly undersampling levels, solving the linear system (3.1) for $\mathbf{x}$ is an underdetermined problem in which more unknowns than equations are present. This problem has either no classical solution if $\mathbf{y}$ is not in the span of the columns of the matrix F_s or infinitely many solutions whenever $\mathbf{y} \in \mathcal{R}(F_s)$. On the other hand the existence of a nontrivial null space, $\mathcal{N}(F_s) \neq \{0\}$, implies non-uniqueness of the solution. In this sense there are infinite number of solutions that fit the data because solutions lying on the null space have no effect in data fit $\left(F_s\left(\mathbf{x}_0 + \mathbf{x}_1\right) = F_s\left(\mathbf{x}_0\right) = \mathbf{y} \text{ with } \mathbf{x}_0 \in \mathcal{N}(\mathcal{F}_s) \text{ and } \mathbf{x}_1 \in \mathcal{N}(F_s)^{\perp}\right)$. The problem is said to be ill-conditioned. The process employed to obtain a particular solution is called regularization. In this scheme the inversion is carried-out by imposing additional constraints that bias the solution. Constraints are based on prior available information about the system, the image $\mathbf{x}$ in the present task. Then an approximate solution in some sense of (3.1) is sought. One way to write regularization is as follows

$$\text{minimize} \quad \mathcal{J}(\mathbf{x}) \quad \text{subject to} \quad \mathbf{y} = F_s\mathbf{x} \tag{3.3}$$

where $\mathcal{J}(\mathbf{x})$ gathers the prior information.

In the recent years significant academic research activity has been exploiting the benefits of the sparsity of the signal or image to be reconstructed in a certain basis $\mathbf{\Psi}$. In particular the l_1-norm of the mentioned sparsifying transformation is

employed as regularization term $\mathcal{J}(\mathbf{x})$. In the remaining of the present section, the l_1-norm choice is argued.

3.2.1 l_1 norm of the wavelet transform

Definition: l_p-norm for $p \geqslant 1$ is defined as

$$||\mathbf{x}||_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

For $0 < p < 1$, norms are known as pseudo-norms because they fail to meet norm properties, namely

- Zero vector: $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = 0$
- Absolute homogeneity: $\forall\, t \neq 0, ||t\mathbf{v}|| = |t|||\mathbf{v}||$
- Triangle inequality: $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$

Definition: The ball of l_p-norm is the subset

$$B(r) = \{\mathbf{x} \in \mathbb{R}^{2N^2} : \sum_{i=1}^{2N^2} |x_i|^p < r^p\}$$

Within the regularization scheme (3.3) the objective function $\mathcal{J}(\mathbf{x})$ is a function in l_p-norm and possible solutions are restricted to be in the ball of l_p-norm. The linear set of equations $\mathbf{y} = F_s\mathbf{x}$ forming the constraints in (3.3) determines a feasible set of solutions. It defines a hyperplane where solution is sought. It is equivalent geometrically to grow the l_p ball until it intersects the feasible set. The intersection is the solution.

Let us consider a simple illustrative example in Figure 3.5. Let $\bar{\mathbf{x}} \in \mathbb{R}^2$ be a 1-sparse vector, i.e, just one of their entries is nonzero $\bar{\mathbf{x}} = \{x_1, 0\}$ which means that $\bar{\mathbf{x}}$ is sparse itself and no further sparsity transform is required. Let $A \in \mathbb{R}^{1\times 2}$ be the measurement matrix that takes 1 measurement $\bar{\mathbf{y}} \in \mathbb{R}^1$. Then the set $\bar{\mathbf{y}} = A\bar{\mathbf{x}}$ defines a line. Accordingly to the sparse idea from Section 3.1, the desired solution $\mathbf{x}$ should have the minimum number of coefficients or minimum l_0-norm. However, l_0-minimization problem is NP-hard [47, 48], this is, solution of problem (3.3) can be found but standard convex analysis fail yet to assure uniqueness. Further, the candidate solution cannot be guaranteed to be a global minimizer because comparison throughout all possible sparse subset is computationally intractable. Another l_p-norm suitable for sparse solutions is sought.

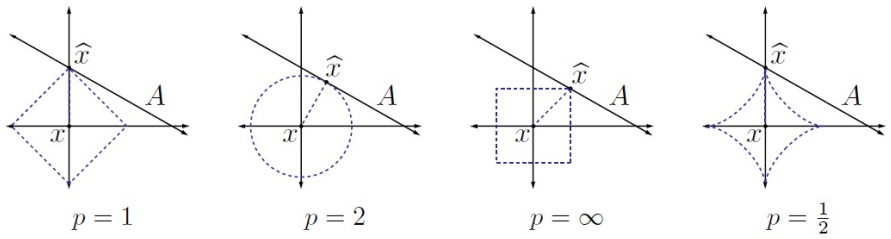

Figure 3.5: *Geometrical visualization of sparsity of l_p balls with $p = 1, 2, \infty$ and $\frac{1}{2}$*

Figure 3.5 shows that norms in the range $0 < p \leq 1$ are appropriated to find a sparse solution while for larger norms the solution is spread more evenly between the two components. Subset formed by balls of smaller p are geometrically sharper; hence they promote sparsity. However, the problem is expected to be solved by means of numerical optimization. In this sense convexity is a desirable characteristic since strict convex functions has a unique solution and with a well-designed algorithm the global minimum can be reached. For non-convex functions, solutions are still possible but optimization becomes daunting.

Definition: A set Ω is convex if $\forall\, \mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and $\forall\, t \in [0, 1]$, the convex combination $\mathbf{x} = t\mathbf{x}_1 + (1 - t)\mathbf{x}_2$ is also in Ω.

Definition: A function $\mathcal{J}(\mathbf{x}) : \Omega \longrightarrow \mathbb{R}$ is convex if $\forall\, \mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and $\forall\, t \in [0, 1]$, the convex combination $\mathbf{x} = t\mathbf{x}_1 + (1 - t)\mathbf{x}_2$ satisfies

$$\mathcal{J}(t\mathbf{x}_1 + (1 - t)\mathbf{x}_2) \leq t\mathcal{J}(\mathbf{x}_1) + (1 - t)\mathcal{J}(\mathbf{x}_2)$$

Unfortunately pseudo-norms $0 < p < 1$ with good properties for sparsity lead to a non-convex optimization problem. Therefore in order to work in a convex set with mild properties for sparsity, the only feasible option is the norm $p = 1$. In the regularization scheme (3.3), the solution reads as follows: among all possible images consistent with the data, select that whose representative coefficients in a basis $\boldsymbol{\Psi}$ has minimum $||\boldsymbol{\alpha}||_1$

$$\text{minimize} \quad ||\boldsymbol{\alpha}||_1 \quad \text{subject to} \quad \mathbf{y} = F_s\mathbf{x} \tag{3.4}$$

This problem is known as basis pursuit [49]. This constrained optimization offers great results in denoising. Minimization of (3.4) is a convex optimization problem and it can be posed as a linear programming for real-valued entries and be solved by using using interior-point methods [15] or simplex method. For complex-valued it is equivalent to a second order cone program (SOCP) [50].

In the presence of noise the data fidelity term must be relaxed. Commonly the equality constraint is replaced by a quadratic discrepancy measure between $\mathbf{y}$ and $F_s\mathbf{x}$

$$\text{minimize} \quad ||\boldsymbol{\alpha}||_1 \quad \text{subject to} \quad ||F_s\mathbf{x} - \mathbf{y}||_2^2 < \epsilon \tag{3.5}$$

where ϵ is the expected noise level. Accordingly to the literature in MRI models, Gaussian noise is commonly accepted [52] although Rician distribution has been also proposed [53]. A traditional way to measure the misfit between data and forward model is the square of the l_2 norm although it can also be written as a function of the l_1 norm.

It is known from constrained optimization theory that the minimization problem (3.5) can be restated in the so-called *augmented Lagrangian method* [54]. With this reformulation, constraints are added to the objective function and the approximation of problem (3.5) becomes

$$\underset{\mathbf{x}}{\text{minimize}} \quad \left\{||F_s\mathbf{x} - \mathbf{y}||_2^2 + \lambda_1||\boldsymbol{\Psi}\mathbf{x}||_1\right\} \tag{3.6}$$

Note that the dependency of the image $\mathbf{x}$ with respect to the vector coefficients $\boldsymbol{\alpha}$ through equaiton 3.1 has been made explicit in the previous equation. For a certain choice of parameter λ_1 this optimization problem will yield the same result as the constrained version [55]. Nevertheless, it is employed a more intuitive approach in which equation (3.6) can be also seen as regularization problem where increasing values of λ_1 penalize constraint violation more severely. Equation (3.7) poses an unconstrained optimization problem that can be solved with the common optimization algorithms. In Section 4.2 we present the chosen non-linear conjugate gradient.

$$||F_s\mathbf{x} - \mathbf{y}||_2^2 + \lambda_1||\boldsymbol{\Psi}\mathbf{x}||_1 \tag{3.7}$$

3.2.2 Total Variation

Closely related to our problem setting is the well-known Total Variation (TV) denoising. The approach known as ROF model is particularly suitable in the presence of Guassian additive noise [59].

The TV is defined as follows. Given a function $f : \Omega \to \mathcal{R}$, where Ω is a bounded open subset of $\mathcal{R}$

$$\text{TV}(f) = \int_\Omega ||\nabla f(x)||_p dx$$

where $||\nabla f(x)||_p$ is the l_p-norm of the gradient of f at x. If Ω is a discrete space and for $p = 1$ the continuous definition leads to

$$\text{TV}(\mathbf{x}) = \sum_k^{2N^2} \sum_{j \sim \mathcal{N}_k} |x_k - x_j|$$

which means that discrete TV is the absolute value of the difference between the $\mathcal{N}_4$ neighboring pixels, Figure 3.2.2 .

For piece-wise constant image as the Shepp-Logan phantom, total variation measures the sum of the magnitudes of the jumps between pixels of the image.

TV is included in (3.7) as the second regularization term. The discrete difference defined in Section 3.1.2 is the sparsifying transform. Finally gathering the data fidelity term, wavelet transform and TV, the following optimization problem results

$$||F_s\mathbf{x} - \mathbf{y}||_2^2 + \lambda_1||\mathbf{\Psi x}||_1 + \lambda_2 TV(\mathbf{x}) \tag{3.8}$$

A note on Compressed Sensing

Compressive Sensing (CS) is intimate related to problem (3.4). Ultimately CS leads to the same optimization problem but the difference lies on the way that data is sampled. In a nutshell, this technique assures that number of samples can be drastically reduced when the image to be reconstructed is sparse in a basis $\mathbf{\Psi}$ other than basis where data is sampled $\mathbf{\Phi}$ and the mutual coherence is low, i.e., basis are dissimilar. For instance Fourier basis and a series of deltas have maximal incoherence. Random matrices are also largely incoherent with any other sparsifying matrix $\mathbf{\Psi}$. It leads to a surprisingly result at the first sight: a sparse image in $\mathbf{\Psi}$ domain can be recovered by acquiring random sampled in another domain $\mathbf{\Phi}$. The level of coherence indicates the number of samples required to recover the image. In MRI, sampling randomly the k-space has little sense because most of the information lies in the center. However this technique is still applicable in MRI. In practice, it has been found that the number of k-space samples should be roughly two to five times the number of sparse coefficients [51]. Previously published papers to this thesis in compressed sensing in the context of MRI are [18, 19, 21–23] among others. For an introductory review in CS, reader is referred to [14] and a more detailed description is offered in [58].

CS is robust in the presence of noise but a higher number of samples is usually required in order to have a solid reconstruct. Note that CS is based on probability theory and reconstruction can be compromised when the undersampling limit is exceeded.

Chapter 4

Methods

The present chapter is devoted to outline how data was obtained accordingly to basics in MRI theory in Chapter 2 and to detail the reconstruction algorithm in the sense regularization discussed in Section 3.2.

4.1 Forward problem, simulated data

The forward problem is straightforward in the problem under study. It was shown in Section 2.5 that under certain assumptions, MRI data can be modeled as the frequency components of the target image. Hence, initial data is obtained from a subset of frequencies in the Fourier domain, F_s. In Section 2.5 three sampling strategies were introduced. Accordingly, three binary k-space masks have been designed to acquire a given percentage of samples. In other words, the mask is placed over the frequency components of the Shepp-Logan phantom; frequencies lying beneath a zero are dropped whilst values beneath a one (black lines in Figure 4.1) are collected. The undersampling level is controlled by increasing the number of zeros in the mask. The three masks have in common a denser sampling in the center area. In particular the rectangular grid sweeps a central square with higher priority and later the mask will also contain equispaced vertical and horizontal lines up to the undersampling level in the external area. Radial mask calculates the number of spokes to sample as a function of the given undersampling level. Finally the square spiral, like the previous masks, starts from the center and it samples the k-space outward with different densities consistent with the undersampling number. The higher presence in the center is based on the fact that most of the information in the Fourier domain is located there, especially for an image with low detail level as the employed phantom, Figure 3.1. This feature is presented in Figure 4.2. For the central column in the k-space, in Figure 4.2 normalized absolute value of the frequencies along the central row are shown in the vertical

axis while horizontal axis represents the relative position with respect to the center of the matrix.

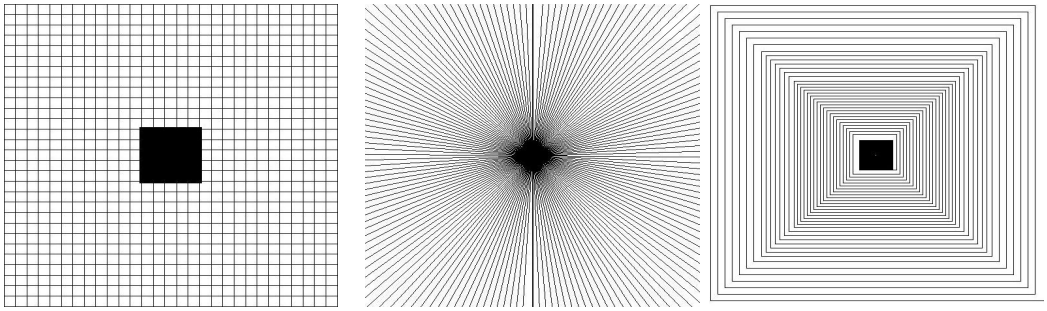

Figure 4.1: *Cartesian, radial and spiral mask with 85% undersampling level*

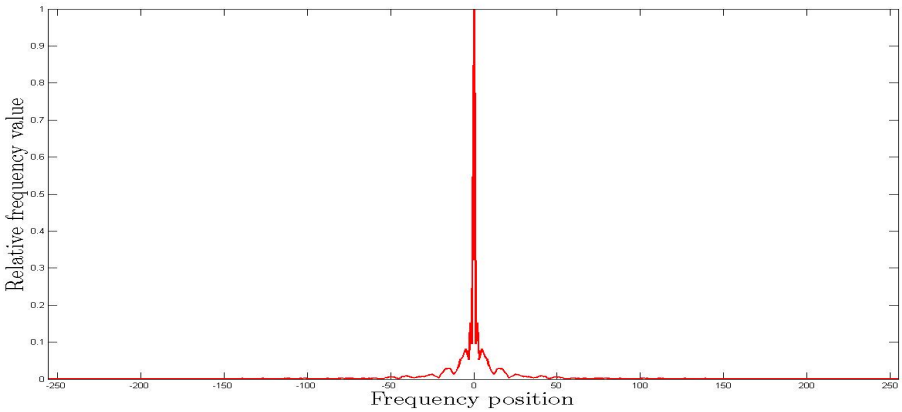

Figure 4.2: *Transverse section of the absolute value of the frequency matrix for the central column of the frequency components of 512 × 512 Shepp-Logan phantom.*

A further comment is in order. Square spirals are of scant benefit in time-saving regards and physically demanding for the hardware components since square corners would require high rising times, Section 2.5. However they are used here with the aim to compare the proposed method to the traditional IFFT. Strategies regarding different sampling than traditional along uniformly sampled rectilinear k-space trajectory requires extra care rather than an straight IFFT. In radial and spiral imaging the k-space trajectories are nonuniform. In a process called regridding, the data is resampled or interpolated onto a uniform rectilinear grid before the application of IFFT. In addition data is corrected for the nonuniform sampling density by an appropriate weighting factor [56, 57].

4.2 Reconstruction algorithm. Non-linear conjugate gradients

Non-linear conjugate gradients (nlcg) is a numerical optimization method whose strategy is to find a sequence of n iterations $\{\mathbf{x}\}_{i=1}^{n}$ in the sense of decreasing an objective function $f(\mathbf{x})$. The proposed algorithm terminates for a determined number of iterations or when the approximation is good enough under certain criteria, for instance certain error threshold.

Minimization is carried-out from an initial guess, selecting the next element in the iteration accordingly to the previous one and moving along a minimizing direction $\mathbf{p}_k$. Parameter t_k stands for the distance to advance in that direction, (4.1) and (4.2). Ideally t_k should be selected to minimize $f(\mathbf{x}_k + t_k\mathbf{p_k})$. Instead the inexact line search is employed. Two methods were tested: backtracking and second order fitting with the objective function. Backtracking ensures that t_k is short enough but not too short to meet Armijo condition, i.e., algorithm always moves in a descent direction

$$\mathbf{while}\big(f(\mathbf{x_k} + t_k\mathbf{p_k}) > f(\mathbf{x_k}) + \alpha t_k(\nabla f(\mathbf{x_k}) \cdot \mathbf{p_k})\big) \tag{4.1a}$$

$$t_k = \beta t_k \tag{4.1b}$$

$$\mathbf{end}(\mathbf{while})$$

for any $\alpha, \beta \in (0, 1)$. It is however a loose way to choose t_k because it just decreases the step length until decrease in the objective function is found. It may get stuck when the function is far from a minimum or it would require significant number of iterations to find the proper step length when the algorithm is close to the minimum. On the other hand it does not prevent steps that are too long relative to the decrease in f. Therefore a proper tuning is needed for β and α.

Second order fitting for the selection of the step length improves the numerical performance. In this approach a small constant initial step length is given. If the step length is found to offer a decreasing value of the objective function, algorithm keeps moving along that direction until the objective function increases. If on the other hand, the initial step length is poor, algorithm decreases the step length until it finds a lower value for the objective function. In both cases three points result: the original, a lower and a higher one. Under the assumption that the objective function between these points can be well approximated as a quadratic function, one can fit a second order function within these three points and finally select the step length that makes the objective function hit the minimum. This strategy obtains more accurate step length but it may require considerable number of calls to the objective function which in turn is more expensive time routine than backtracking.

The characteristic properties of the conjugate direction suggest to select the minimization line as a linear combination of the gradient of the objective function: $\mathbf{p}_{k+1} = -\nabla f_{k+1} + \beta_{k+1}\mathbf{p_k}$ [54]. This property make conjugate gradients suitable for large scale problems since it just requires the calculation of the gradient for the objective function and the cumbersome second order derivations are not needed. Further, it only requires of little storage at each iteration since algorithm parameters are overwritten at every new iteration. Algorithm requires the initial minimization direction which is given by the steepest decent direction: $\mathbf{p}_0 = -\nabla(f_0)$

Several variants are proposed in the literature for scaling factor β_k in minimization line, (4.3). In this thesis work, two methods were tested: Fletcher-Reeves with backtracking and Polak-Ribière with a second order curve fitting. For an objective function strongly convex these two methods are identical but for general non-linear functions the latter method is found to be more robust. Comparison of the numerical performance is illustrated in the Figure 4.3. Polak-Ribière with a second order curve fitting is found to be more robust and offers larger minimization of the objective function for smaller number of iterations. This is the strategy employed in the rest of simulations.

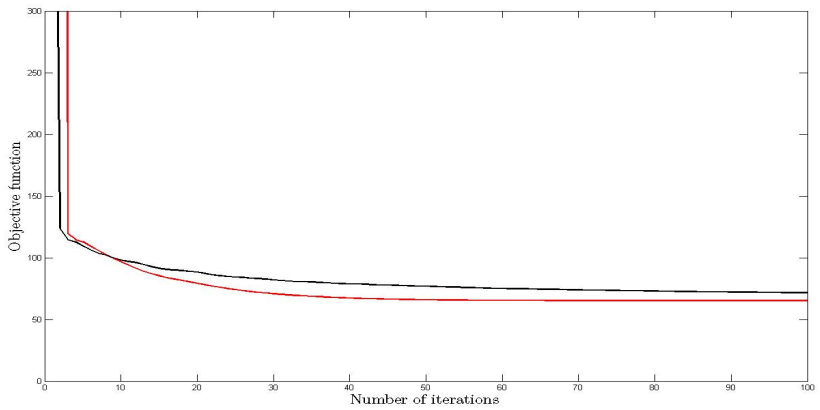

Figure 4.3: *Non-linear conjugate gradient minimization. Objective function value (ordinate) vs the number of iterations (abscissa). Polak-Ribière (red) and Fletcher-Reeves method (black). Experiment performed over a 85% sparse rectangular mask with* $\lambda_1 = 0.005$, $\lambda_2 = 0.002$ *and* $\alpha = 0.3$ *and* $\beta = 0.6$ *for backtracking, (4.1)*

Finally, the implemented algorithm is showed in Table 4.1. Notice that l_1-norm is a non-smooth function, hence it is non-differentiable near the minimum. In general algorithms have an unpredictable behavior in the proximity of a non-smooth point. It can be overcome adding a small smoothing parameter μ.

Table 4.1: *Details of the parameters used in the optimization algorithm*

NON-LINEAR CONJUGATE GRADIENTS. POLAK-RIBIÈRE METHOD $f(\mathbf{x}) = \|\|F_s\mathbf{x} - \mathbf{y}\|\|_2^2 + \lambda_1\|\|\boldsymbol{\alpha}\|\|_1 + \lambda_2 TV(\mathbf{x})$
Input Number of iterations: nItr = [40-110] Sparsity transform: Ψ: Daubechies D4 wavelet basis Undersampled Fourier transform: rectangular, radial or spiral mask Data: $\mathbf{y} \in \mathbb{R}^{2m^2 \times 1} = F_s\mathbf{x} + \eta$ l_1 norm smoothing: $\mu = 10^{-6}$ Wavelet term weight: $\lambda_1 = [0 - 0.1]$ TV- weight: $\lambda_2 = [0 - 0.1]$
Output Numerical approximation $\mathbf{x}$
Initialization: $\mathbf{g_0} = \nabla f$ *Set*: $\mathbf{p_0} = -\mathbf{g_0}$ $k = 0;$ *Iterations:* **while**$(k < nItr)$ $t_k = \text{line search}(\mathbf{x}, \mathbf{p}, f(\mathbf{x_k}))$ $\mathbf{x_{k+1}} = \mathbf{x_k} + t_k\mathbf{p_k}$ (4.2) $\mathbf{g_{k+1}} = \nabla f(\mathbf{x_{k+1}})$ $\beta = \frac{\mathbf{g_{k+1}}^T(\mathbf{g_{k+1}} - \mathbf{g_k})}{\|\|\mathbf{g_k}\|\|_2^2}$ $\beta = \max([\beta, 0])$ $\mathbf{p_{k+1}} = -\mathbf{g_{k+1}} + \beta\mathbf{p_k}$ (4.3) $k = k + 1$ **end**(**while**)

The gradient equation is specified in the Appendix B.

The nlcg is an numerical optimization algorithm for real-valued functions. In order to meet this property every vector is written as a real vector. Hereafter we cope with real-valued functions moving from $\mathbb{C}^{N^2\times 1}$ to $\mathbb{R}^{2N^2\times 1}$ through the mapping ϕ

$$\phi : \mathbf{x} \in \mathbb{C}^{N^2\times 1} \longrightarrow \mathbf{x} \in \mathbb{R}^{2N^2\times 1} \tag{4.4}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N^2} \end{pmatrix} \longrightarrow \begin{pmatrix} \mathrm{Re}(x_1) \\ \mathrm{Re}(x_2) \\ \vdots \\ \mathrm{Re}(x_{N^2}) \\ \mathrm{Im}(x_1) \\ \mathrm{Im}(x_2) \\ \vdots \\ \mathrm{Im}(x_{N^2}) \end{pmatrix}$$

This mapping does not discard the imaginary part of the signal but only rearranges it. While magnitude in the Fourier domain carries intensity, phase dominates intelligibility of the image by placing the discernible objects in the image; translation of position has no effect on the magnitude but it does on phase adding a phase term [60]. Therefore the role of the phase cannot be neglected and previous mapping is of crucial relevance.

Chapter 5

Simulations

In this section the results of the optimization problem described in Section 4.2 are presented. Cartesian, radial and spiral masks were designed to collect a representative percentage of frequencies obeying its characteristic pattern. In the sense frame of compressed sensing a pseudo-random mask with lines scanned only along the read-out direction was examined. For every grid we compare the undersampling level to the relative error

$$\text{Relative Error} = \frac{||\mathbf{x} - \hat{\mathbf{x}}||_2}{||\mathbf{x}||_2} \tag{5.1}$$

5.1 Cartesian

Results are plotted in Figure 5.1. The used parameters are detailed in Table 5.1 :

Table 5.1: *Cartesian reconstruction. Figure 5.1*

Number of iterations: 70
Data: $\mathbf{y} = F_s\mathbf{x} + \eta$ with $\eta \sim \mathcal{N}(0, 0.01)$
Sparsity transform: Ψ: Daubechies D4 wavelet basis
Wavelet weight: $\lambda_1 = 0.001$
TV- weight: $\lambda_2 = 0.01$

Cartesian grid offered very competitive properties. Increasing undersampling level leads to blurring the IFFT image with an increasing presence of ringing artefacts. However noise-like artefacts are totally removed, right column in Figure 5.1 with the proposed method.

CARTESIAN GRID	
Undersampling level (%)	**Relative Error**
65	0.00464
75	0.00993
85	0.01916
95	0.06866

5.2 Radial

From the center of the k-space different number of spokes were traced accordingly to the desired undersampling level . It simulated the way data is collected in real radial sampling, Figure 2.8.

Table 5.2: *Radial reconstruction. Figure 5.2*

Number of iterations: 70
Data: $\mathbf{y} = F_s\mathbf{x} + \eta$ with $\eta \sim \mathcal{N}(0, 0.01)$
Sparsity transform: Ψ: Daubechies D4 wavelet basis
Wavelet term weight: $\lambda_1 = 0.001$
TV- weight: $\lambda_2 = 0.01$

RADIAL GRID	
Undersampling level (%)	**Relative Error**
80	0.00466
85	0.00695
90	0.02348
95	0.21380

Radial undersampling outcomes in “honeycomb” artefacts. It is shown in Figure 5.2 that the proposed regularization recovers with an acceptable low visual impact compared to the original phantom up to 90% undersampling level. However at 95% undersampling level one can infer that data set is excessively small and useless image results. This results can be associated with the minimum sampling in the center area since as it was already stated, most of the information lie in this area. Further it was experienced that proposed method is robust in these two previous sampling cases and not fine tuning for the algorithm parameters was needed.

5.3 Spiral

Spiral mask represents a non-uniform sampling strategy. For the introduced undersampling level, spiral mask will return loops sampled with decreasing density.

Table 5.3: *Spiral reconstruction. Figure 5.3*

Number of iterations: 110
Data: $\mathbf{y} = F_s\mathbf{x} + \eta$ with $\eta \sim \mathcal{N}(0, 0.01)$
Sparsity transform: Ψ: Daubechies D4 wavelet basis
Wavelet term weight: $\lambda_1 = 0.001$
TV- weight: $\lambda_2 = 0.008$

SPIRAL GRID	
Undersampling level (%)	**Relative Error**
65	0.06152
75	0.06168
85	0.06173
95	0.27698

Spiral sampling employed suffers from artefacts at the image detail level. For a low number of iterations recovery is adequate only until 65% undersampling level although proposed approach was able to recover the target image fairly accurately. Throughly inspection detects that the effect of total variation have smoothed artefact into a constant areas with "stain" appearance, Figure 5.4.

5.4 Pseudo-Random Sampling

In a real cartesian sampling scheme lines would be acquire just in the read-out direction. The saving time is proportional to the number of non-sampled lines, Section 2.5.1.

Table 5.4: *Pseudo-random reconstruction. Figure 5.5*

Number of iterations: 50
Data: $\mathbf{y} = F_s\mathbf{x} + \eta$ with $\eta \sim \mathcal{N}(0, 0.05)$
Sparsity transform: Ψ: Daubechies D4 wavelet basis
Wavelet term weight: $\lambda_1 = 0.05$
TV- weight: $\lambda_2 = 0.02$

PSEUDO-RANDOM SAMPLING	
Undersampling level (%)	**Relative Error**
65	0.043792

Pseudo-random cartesian scheme evidence the versatility of this method and offers a new sampling strategy that meets the principles of compressed sensing. Reconstruction in this case required a higher TV penalization to promote denoising and higher wavelet term weight to preserve edges. Fine details are still distinguishable, Figure 5.5.

5.5 High noise condition

Robustness of this methods was experienced at higher noise conditions in 97% sparse Cartesian scheme, only a small central square is sampled. Choice of parameters λ_1 and λ_2 is based on the minimum relative error at certain number of iterations, Figure 5.6. Penalization on TV and wavelet is higher than in the rest of simulations. Coarse regions are recovered but fine details are lost. The reason is that high frequencies were not sampled. In addition data in the external area of the small sampled square is buried in noise. Nevertheless artefacts were reasonably well removed in the coarse areas and borders are still distinguishable but jagged. In these regard the effect of wavelets smooth slightly the borders (left-most image versus central image) in Figure 5.7. On the contrary wavelets without total variation is not able to flatten the artefacts (right image).

Table 5.5: *High noise reconstruction. Figure 5.7*

Number of iterations: 60
Data: $y = F_s x + \eta$ with $\eta \sim \mathcal{N}(0, 0.1)$ 97% undersampling level in Cartesian grid
Sparsity transform: Ψ: Daubechies D4 wavelet basis
Wavelet term weight: $\lambda_1 = \{0; 0.01\}$
TV- weight: $\lambda_2 = \{0; 0.1\}$

HIGH NOISE CONDITIONS. Cartesian sampling			
Undersampling level (%)	TV weight	Wavelet	**Relative Error**
97	0.035	0.01	0.08272
97	0.035	0	0.08135
97	0	0.01	0.08618

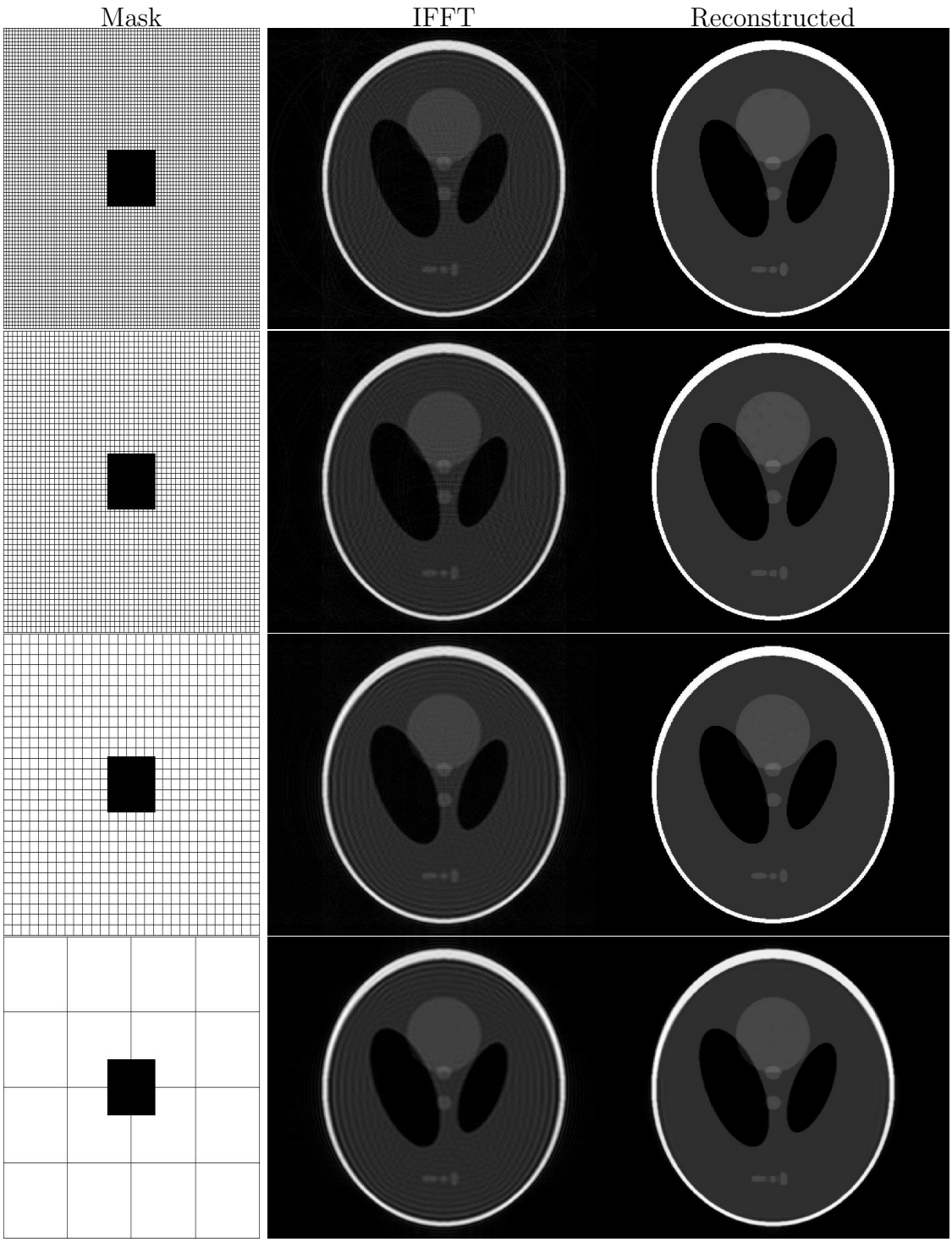

Figure 5.1: *Cartesian grid simulation. Undersampling levels from the top to the bottom : 65%, 75%, 85% and 95%*

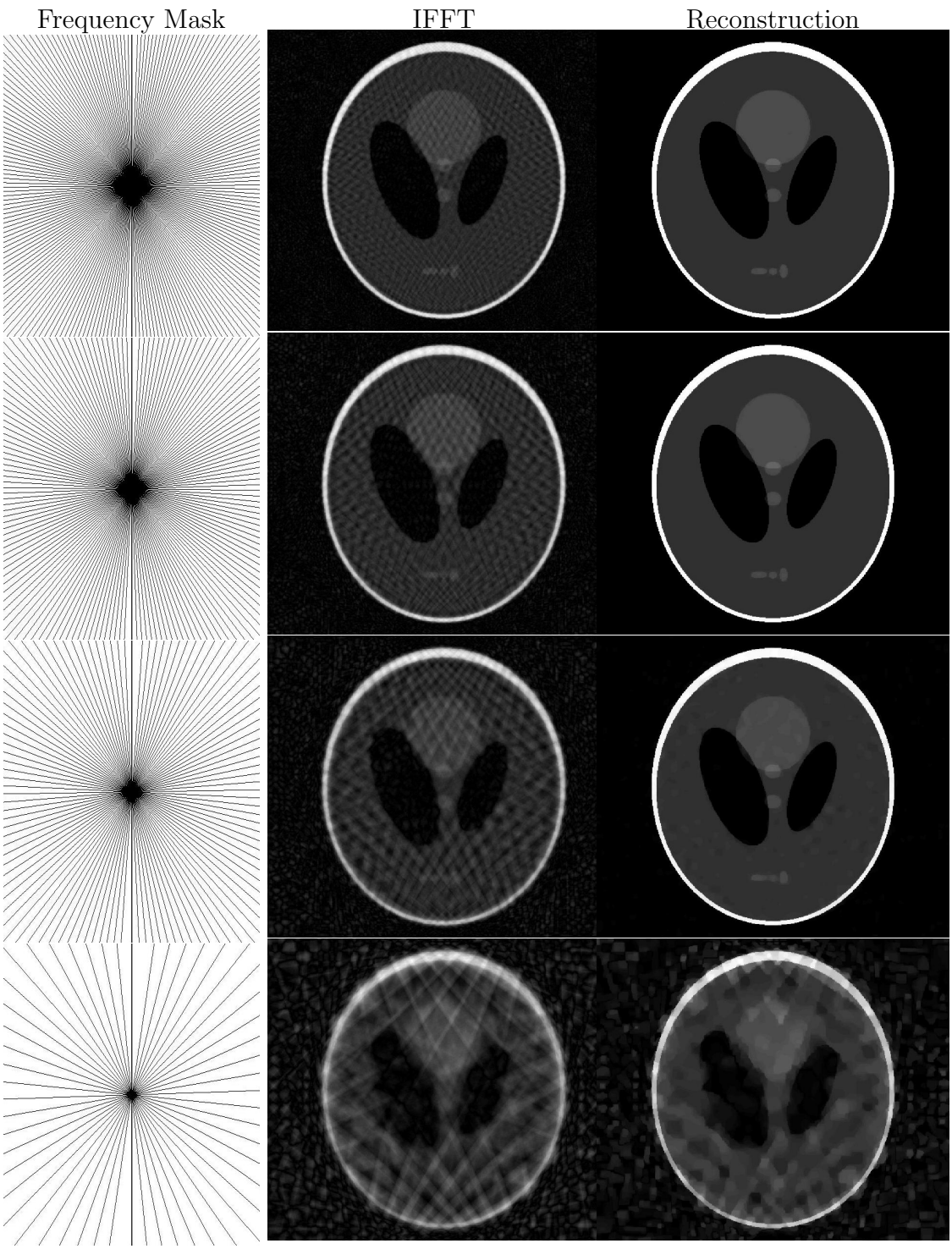

Figure 5.2: *Radial grid simulations. Undersampling levels: 80%, 85%, 90% and 95%*

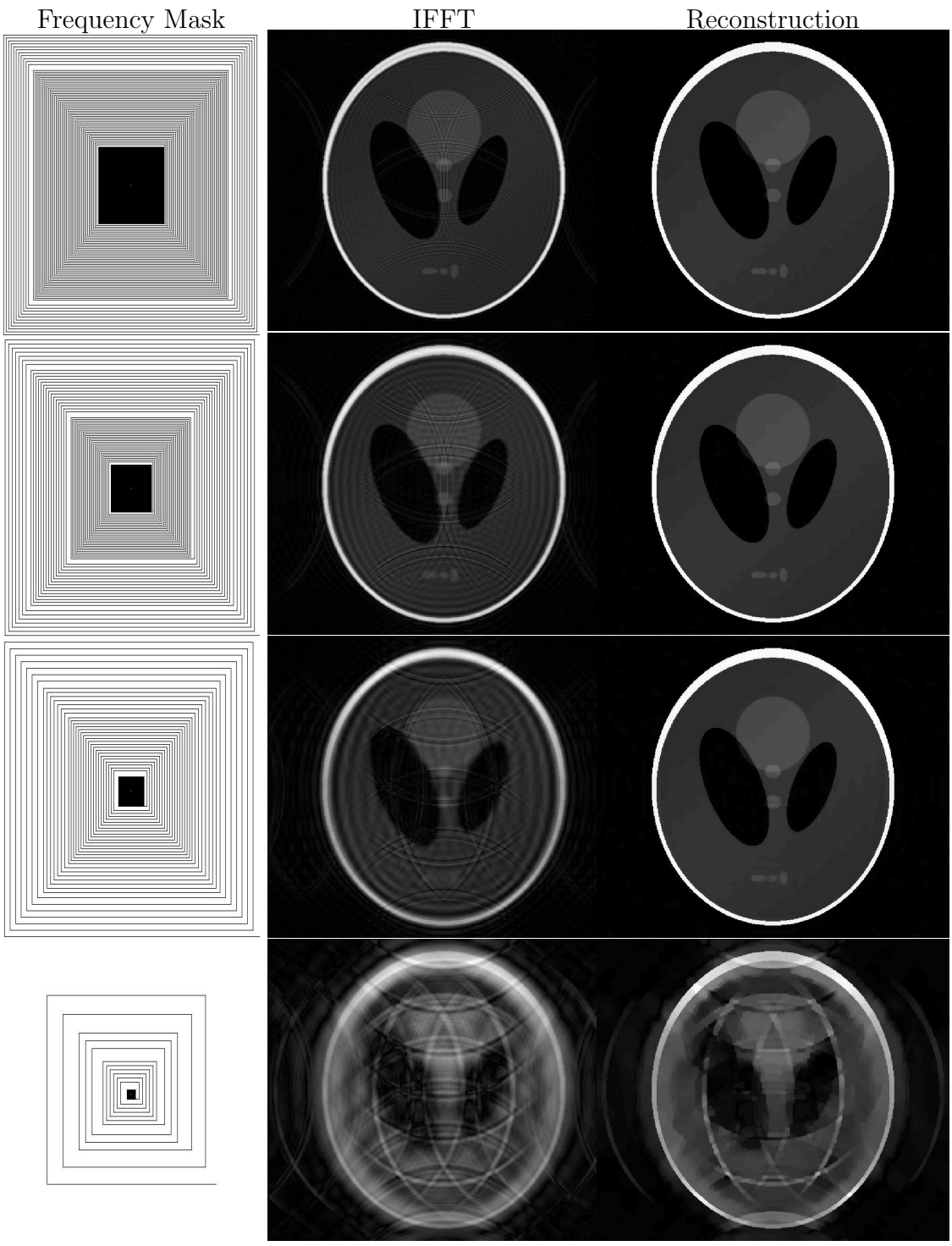

Figure 5.3: *Spiral grid simulation. Undersampling levels: 65%, 75%, 85% and 95%*

Figure 5.4: *Increased detail image of the spiral reconstruction for an undersampling level of 85%*

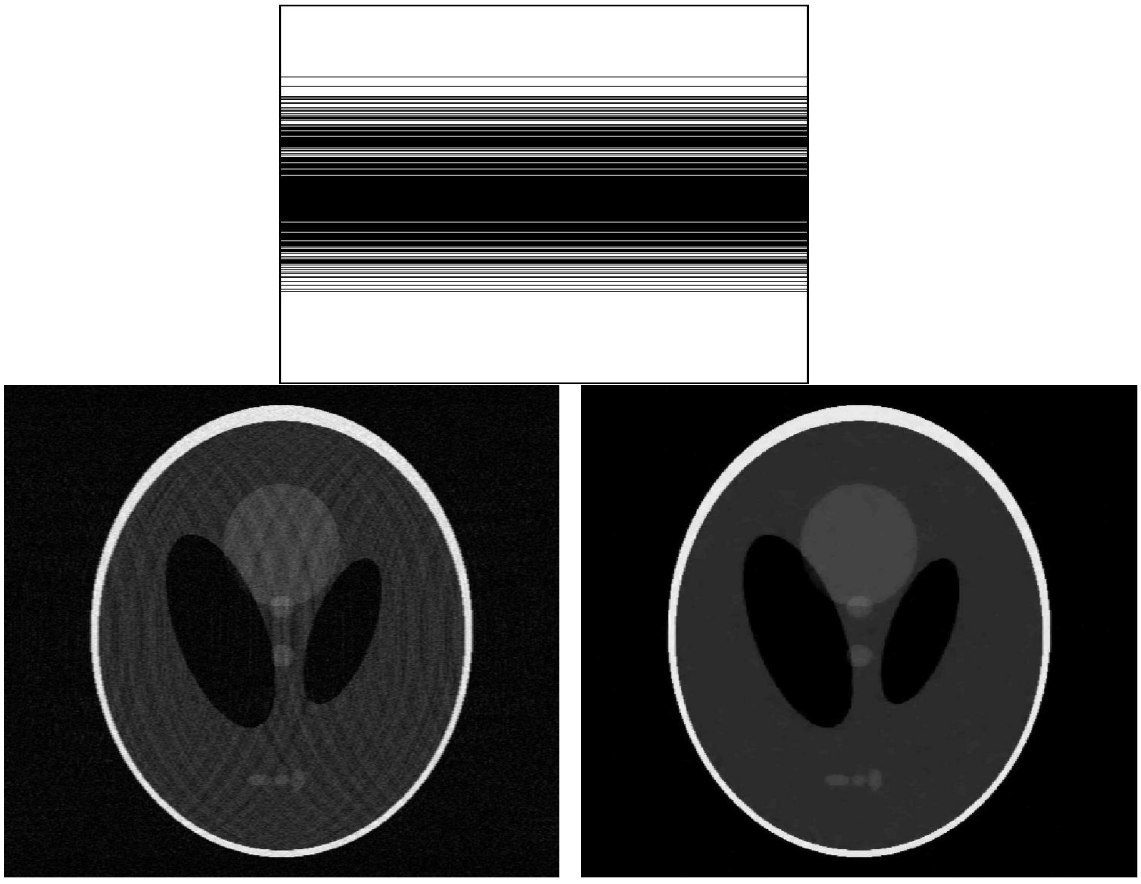

Figure 5.5: *Pseudo-random grid (top). Initial and reconstructed image (bottom row)*

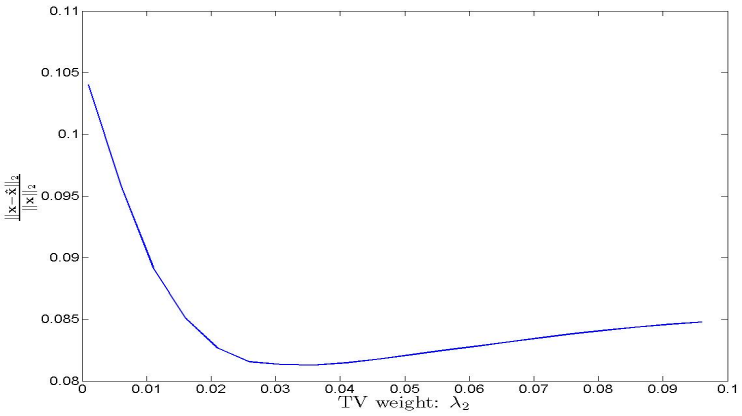

Figure 5.6: *Relative error versus weight of TV.*

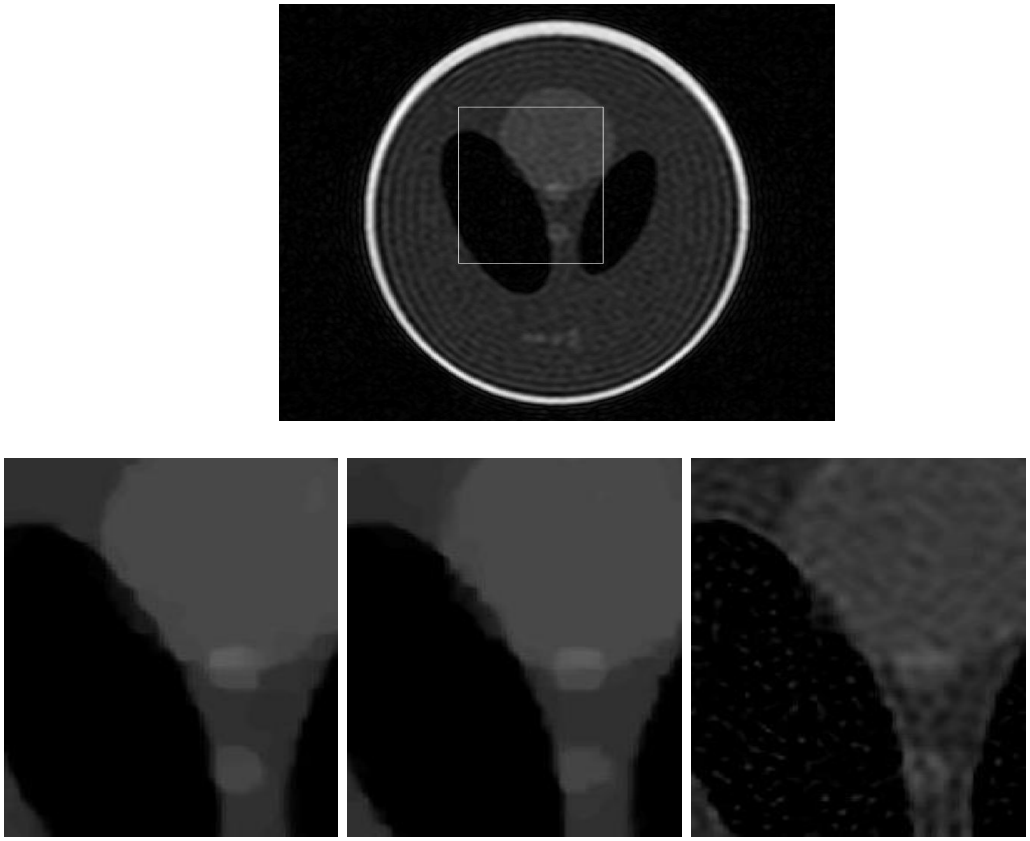

Figure 5.7: *Top image: Inverse FFT. Bottom row: Detailed reconstructed image. Left: both total variation and wavelet were used. Middle: only TV weight. Right: Only wavelet weight*

Chapter 6

Discussion and conclusions

In this thesis we have discussed a method to reduce the number of samples in the MRI data acquisition stage without degrading the image quality. The proposed method was based in wavelet and TV regularization, exploiting in this way the spatial redundancy of the target image, the Shepp-Logan phantom. Data was obtained from three characteristic sampling strategies in MRI, namely, cartesian, radial and square spiral with diverse undersampling level. Reconstructions of this method were compared to the traditional IFFT. It was found that, up to certain undersampling level, target image was recovered with less visual degradation than IFFT whose reconstructions suffered from undersampled artefact such as aliasing and Gibbs ringing. Relative errors between the reconstructed and target image, Tables 5.1, 5.2 and 5.3, indicate that degradation increases with the undersampling level but discrepancy is moderate in the three sampling schemes up to 85% undersampling level. In the same way, visual inspection in Figures 5.1, 5.2 and 5.3 uphold this conclusion.

It is important to recall that the Shepp-Logan phantom consists only of ten ellipses of different sizes, location and contrast. Therefore it is a simple image of little detail and not high frequencies are expected in Fourier domain, Figure 4.2. Then the omission of large areas in the external region of the k-space appears as slight blurring artefact. However, real medical images have in general higher detail level. In addition reliable diagnosis are based on high contrast and resolution to discern between diverse anatomical structures. This method is still exportable to real case although extra care has to be taken into account. Sampling patterns should be then adapted to particular imaged region, physiological characteristics and specific diagnosis. For instance brain image may require less contrast that an abdominal slice due to the higher concentration of organs and fine details. In general, high detailed image and scanning time is a trade-off decision. In particular the discussed technique is likely to have successful implementation in brain imaging where spatial redundancy is relatively high. In the same way more relevant role of

sparsifying transform term is expected for more elaborated images.

The quality of an image is commonly quantified in MRI with the Signal-to-Noise Ratio (SNR). It depends on a great deal of factors which were beyond the theme of this thesis. However further studies may include data modeling and hence asses whether reconstructed images meet quality standards in relation to the undersampling factor.

Time saving in cartesian is proportional to the number of omitted horizontal lines or read-out direction in the k-space, Figure 2.7, and proportional to the number of omitted spokes in the radial case, Figure 2.8. Therefore radial mask is closer to saving-time real scenario whilst employed cartesian mask would require longer sampling time in order to acquire data in the central square. In this sense the pseudo-random mask is more realistic. Nevertheless it is necessary to have an objective parameter to estimate the quality of the image with respect to saved time. Likewise scanning time depends also on the hardware components of the particular scanner. Therefore it is not possible to estimate a concrete saved time factor in the present study because more realistic model and hardware features are required.

Appendix A

Basic Electromagnetism

A.1 Potential Vector

Let $\vec{B}$ be the magnetic field evaluated at certain point determined by the position vector $\vec{r}$ (unprimed coordinates) with respect to an arbitrary center **0** due to a line current element $I'd\vec{l'}$ (primed coordinates), Figure A.1. The integral expression for the magnetic field is computed from the Ampere's force law and it reads

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} \oint_{\text{loop}} \frac{I'd\vec{l'} \times \vec{R}}{R^3}$$

where μ_0 is the vacuum permeability. The previous equation is named Biot-Savat's law. These line currents arise from the individual precessing proton.

The potential vector is defined as

$$\vec{B}(\vec{r}) = \vec{\nabla} \times \vec{A}(\vec{r}) \tag{A.1}$$

The integral equation for the potential vector is obtained from (A.1) and the following vector identities: $\nabla \times \left(\lambda\vec{C}\right) = (\nabla\lambda) \times \vec{C} + \lambda\left(\nabla \times \vec{C}\right)$ and $\nabla\left(\frac{1}{R}\right) = -\nabla'\left(\frac{1}{R}\right) = -\frac{\vec{R}}{R^3}$ where $\vec{R} = \vec{r} - \vec{r'}$ from Figure A.1. The integrand in equation (A.1) yields

$$\frac{d\vec{l'} \times \vec{R}}{R^3} = -d\vec{l'} \times \nabla\left(\frac{1}{R}\right) = \nabla \times \left(\frac{d\vec{l'}}{R}\right) - \left(\frac{\nabla \times d\vec{l'}}{R}\right) = \nabla \times \left(\frac{d\vec{l'}}{R}\right)$$

since $\nabla \times d\vec{l'} = 0$. Back into equation (A.1), the magnetic field remains

$$\vec{B}(\vec{r}) = \nabla \times \left(\frac{\mu_0}{4\pi} \oint_{\text{loop}} \frac{I'd\vec{l'}}{R}\right) \tag{A.2}$$

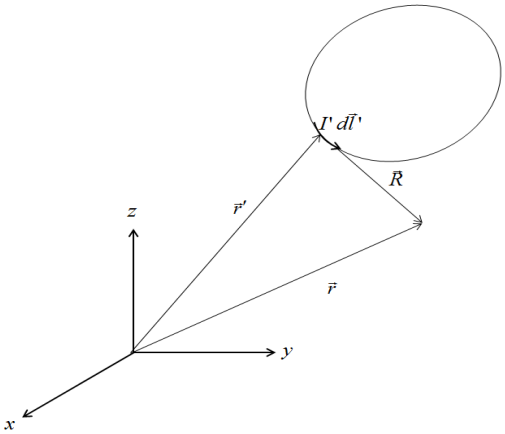

Figure A.1: *Vectors relation for the magnetic field at a certain point given by the vector position $\vec{r}$ due to an element current $Id\vec{l}$*

By comparison of (A.1) and (A.2), the integral expression of the potential vector follows

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \oint_{\text{loop}} \frac{I' d\vec{l'}}{R} \tag{A.3}$$

The collective arrangement of individual protons in a macroscopic sample results in a volume density current, $\vec{J'}(\vec{r'})$. The integral expression including the volume density current is straightforward. From (A.3), the line current I' is replaced by $\vec{J'}(\vec{r'})$ and the integral has to be evaluated for the sample volume V$'$

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \int_{\text{V}'} \frac{\vec{J'}(\vec{r'})}{R} d\vec{r'}$$

The bulk magnetization defined in (2.8) is related to the volume density current $\vec{J'}(\vec{r'})$ as follows [40]

$$\vec{J'}(\vec{r'}) = \nabla \times \vec{M}(\vec{r'})$$

Then the potential vector can be written as

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \int_{\text{V}'} \frac{\nabla \times \vec{M}(\vec{r'})}{R} d\vec{r'} \tag{A.4}$$

A.2 Principle of reciprocity

Let us suppose two loops C_j and C_k carrying with two currents I_j and I_k respectively. Due to Faraday's induction law, both currents induce another current in

the opposite loop. Let Φ_{kj} denote the flux of current k in loop j. The potential vector produced by the circuit C_k at r_j in loop C_j is

$$\Phi_{kj} = \oint_{C_j} \vec{A_k}(\vec{r_j}) \cdot d\vec{l_j} = \frac{\mu_0}{4\pi} \oint_{C_j} \oint_{C_k} \frac{I_k d\vec{l_k} \cdot d\vec{l_j}}{R_{jk}}$$

From the previous equation it can be inferred that the flux through C_j is proportional to I_k. The proportional term $\mathcal{K}_{jk}$ is purely geometric-dependent

$$\mathcal{K}_{jk} = \frac{\mu_0}{4\pi} \oint_{C_j} \oint_{C_k} \frac{d\vec{l_k} \cdot d\vec{l_j}}{R_{jk}}$$

Since $d\vec{l_k} \cdot d\vec{l_j} = d\vec{l_j} \cdot d\vec{l_k}$ and $R_{jk} = R_{kj}$ it then follows that $\mathcal{K}_{jk} = \mathcal{K}_{kj}$. Then the flux through C_j due to a current I_0 in loop C_k equals to the flux through C_k due to the same current in C_j. It is known as the reciprocity principle.

A.3 MRI signal

The net effect of individual currents due to spinning protons is gathered under the vector quantity magnetization, $\vec{M}$. Typically flux is referred to a known coil due to a magnetization source. Thanks to reciprocity principle, that flux can be also obtained from the magnetization generated in a coil and going through the sample. This flux is denoted by Φ_M. Recall that in the following derivation the primed vectors indicating the magnetization source now refer to the coil while the unprimed coordinates refer to the sample. We also have to take into account that magnetization is time-dependent because it arise from the relaxation, Section 2.3.1,

$$\begin{aligned}
\Phi_M(t) &= \oint d\vec{l} \cdot \left(\frac{\mu_0}{4\pi} \int \frac{\vec{\nabla}' \times \vec{M}(\vec{r'}, t)}{R} \right) d\vec{r'} = \\
&= \frac{\mu_0}{4\pi} \int d\vec{r'} \oint d\vec{l} \cdot \left(-\vec{\nabla}' \left(\frac{1}{\vec{R}} \right) \times \vec{M}(\vec{r'}, t) \right) = \\
&= \int d\vec{r'} \vec{M}(\vec{r'}, t) \cdot \left(\vec{\nabla}' \times \frac{\mu_0}{4\pi} \left(\oint \frac{d\vec{l}}{\vec{R}} \right) \right)
\end{aligned}$$

From equation (A.1) and (A.3), the term in parentheses reminds to the magnetic field per unit of current $\vec{B}_r(\vec{r'}) = \vec{B}(\vec{r'})/I = \vec{\nabla}' \times \frac{\mu_0}{4\pi} \left(\oint \frac{d\vec{l}}{\vec{R}} \right)$. Finally the flux can be written as

$$\Phi_M(t) = \int d\vec{r'} \vec{B}_r(\vec{r'}) \cdot \vec{M}(\vec{r'}, t)$$

This is equation (2.20) where the prime has been removed to alliavate notation although it has to take into account that integration and vector are referred to the coil.

A.4 Magnetization

In the presence of an external magnetic field $B_0\vec{k}$, the stored magnetic energy (2.2) is given by

$$E = -\vec{\mu} \cdot \vec{B_0} = -\mu_z B_0 = -\gamma\hbar m_s B_0$$

Protons have two spin states $m_s = \pm 1/2$ which implies two different energy states. The energy level splitting in the presence of a noticeable magnetic field is called Zeeman splitting phenomena

$$\triangle E = E_\uparrow - E_\downarrow = \gamma\hbar B_0$$

The number of protons in either of both states, spin population, is related to their energy difference in the sense that physical systems tend to be in the minimum energy state. In this regard the spin-up state is less energetic and consequently more populated even for small energy gap. Just to picture a number it is estimated that at body temperature in a 1.5 T scanner, for every million protons in the spin-down direction there are a million-and-four in the spin-up direction [27].

Population of states obeys the Boltzmann distribution

$$\frac{N_\uparrow}{N_\downarrow} = \exp\Big(\frac{\triangle E}{KT_s}\Big)$$

where T_s is the absolute temperature. Since $\triangle E \ll KT_s$, the exponential can be approximated as the following Taylor series

$$\exp\Big(\frac{\triangle E}{KT_s}\Big) \simeq 1 + \frac{\gamma\hbar B_0}{KT_s}$$

and the ratio of the population yields

$$\frac{N_\uparrow}{N_\downarrow} \simeq 1 + \frac{\gamma\hbar B_0}{KT_s} \tag{A.5}$$

Taking into account the population in both states, equation (2.8) for bulk magnetization can be reformulated as follows

$$\vec{M} = \frac{1}{V}\Big(\sum_{n=1}^{N_\uparrow}\frac{1}{2}\gamma\hbar - \sum_{n=1}^{N_\downarrow}\frac{1}{2}\gamma\hbar\Big)\vec{k} = \frac{1}{2V}\big(N_\uparrow - N_\downarrow\big)\gamma\hbar\vec{k}$$

With equation (A.5), its magnitude becomes

$$|\vec{M}| \simeq \frac{\gamma^2 \hbar^2 B_0 N_s}{4KT_s} \tag{A.6}$$

This is roughly the measurable magnetization, 2.1.4

Appendix B

Gradient of the objective function

In this appendix the expression of the gradient of the objective function (3.8) in Section 4.2 is presented.

B.1 Basics in complex numbers

Let $\mathbf{z} \in \mathbb{C}^N$ be a complex vector.

- *Absolute value*: The absolute value of a complex vector $\mathbf{z} \in \mathbb{C}^N$ is the following real vector

$$|\mathbf{z}| = \sqrt{\mathbf{z}^* \cdot \mathbf{z}} = \sqrt{\left(\mathrm{Re}\left(z_1\right)^2 + \mathrm{Im}\left(z_1\right)^2\right), \ldots, \left(\mathrm{Re}\left(z_N\right)^2 + \mathrm{Re}\left(z_N\right)^2\right)} \in \mathbb{R}^N \tag{B.1}$$

  where $\mathbf{z}^*$ is the conjugate of $\mathbf{z}$.

- *l_2-norm*: The l_2-norm is computed adding the square of the real and imaginay part of the entries of the complex vector $\mathbf{z}$. It is then equivalent to sum the square of the $2N$ entries of a real vector $\mathbf{x} \in \mathbb{R}^{2N^2}$ formed stacking the real and imaginary part of $\mathbf{z}$ as it was indicated in (4.4)

$$||\mathbf{z}||_2 = \sqrt{\left(\mathrm{Re}\left(z_1\right)^2 + \mathrm{Im}\left(z_1\right)^2\right) + \cdots + \left(\mathrm{Re}\left(z_N\right)^2 + \mathrm{Re}\left(z_N\right)^2\right)} = \left(\sum_{i=1}^{2N} x_i^2\right)^{1/2} \tag{B.2}$$

- *l_1-norm*: The l_1-norm is the sum of the absolute value of the vector entries.

$$||\mathbf{z}||_1 = \sum_{i=1}^{N} |z_i| = \sum_{i=1}^{N} \sqrt{\left(\mathrm{Re}\left(z_i\right)^2 + \mathrm{Im}\left(z_i\right)^2\right)} \tag{B.3}$$

- *Gradient*: The gradient of a scalar function $f \in \mathbb{R}^{N^2}$ is denoted ∇f where ∇ is the differential operator.

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_{N^2}} \end{pmatrix}$$

  Minimization of the objective function (3.8) is performed over the vector of unknowns and consequently, the derivative of the objective function is evaluated with respect to $\mathbf{x}$. Further, the mapping (4.4) transforms the complex quantity $\mathbf{x} \in \mathbb{C}^{N^2}$ into a real-valued vector $\mathbf{x} \in \mathbb{R}^{2N^2}$. As it was stated earlier this is a common strategy to meet conditions for optimization algorithm developed for real functions. Nevertheless it must be clear that $\mathbf{x}$ should be treated as a complex entity in the computation of the norms and derivatives, hence indexes in the following summation are over N^2 rather than $2N^2$.

Along the next three sections we detail the computation of the three elements of the objective function (3.8) in index notation.

B.2 Gradient of data fidelity term

The data fidelity term in equation (3.8) is the square of the l_2-norm

$$||F_s\mathbf{x} - \mathbf{y}||_2^2$$

The j$^{\text{th}}$ element of the m entries of the matrix multiplication $F_s\mathbf{x} - \mathbf{y}$ is

$$(F_s\mathbf{x} - \mathbf{y})_j = \sum_{i=1}^{N^2} F_{s_{ji}} x_i - y_j$$

The l_2-norm of the previous complex number can be written accordingly to equation (B.2) as the sum of the square of the real and imaginary part

$$||F_s\mathbf{x} - \mathbf{y}||_2^2 = \sum_{j=1}^{2m} \left(\sum_{i=1}^{N^2} F_{s_{ji}} x_i - y_j \right)^2$$

where the index j of the first summation now extends to $2m$ because the complex number $F_{s_{ji}} x_i - y_j$ has been split into real and imaginary components.

The gradient of the of the norm remains

$$\nabla\left(||F_s\mathbf{x}-\mathbf{y}||_2^2\right)=\begin{pmatrix}\frac{\partial}{\partial x_1}\left(\sum_{j=1}^{2m}\left(\sum_{i=1}^{N^2}F_{s_{ji}}x_i-y_j\right)^2\right)\\ \frac{\partial}{\partial x_2}\left(\sum_{j=1}^{2m}\left(\sum_{i=1}^{N^2}F_{s_{ji}}x_i-y_j\right)^2\right)\\ \vdots\\ \frac{\partial}{\partial x_{2N^2}}\left(\sum_{j=1}^{2m}\left(\sum_{i=1}^{N^2}F_{s_{ji}}x_i-y_j\right)^2\right)\end{pmatrix}$$

The l$^{\text{th}}$ element of the gradient yields

$$\begin{aligned}&\frac{\partial}{\partial x_l}\left(\sum_{j=1}^{2m}\left(\sum_{i=1}^{N^2}F_{s_{ji}}x_i-y_j\right)^2\right)=\\ &=\sum_{j=1}^{2m}\frac{\partial}{\partial x_l}\left(\sum_{i=1}^{N^2}F_{s_{ji}}x_i-y_j\right)^2=\\ &=2\sum_{j=1}^{2m}\left(\sum_{i=1}^{N^2}F_{s_{ji}}x_i-y_j\right)F_{s_{jl}}\end{aligned}$$

B.3 Gradient wavelet transformation

Let us consider next the gradient of the sparsifying transform $||\boldsymbol{\alpha}||_1 = ||\boldsymbol{\Psi}\mathbf{x}||_1$. The j$^{\text{th}}$ element of the matrix multiplication is

$$\left(\boldsymbol{\Psi}\mathbf{x}\right)_j=\sum_{i=1}^{N^2}\Psi_{ji}x_i$$

From (B.3), the l_1-norm of the wavelet transform results

$$||\boldsymbol{\Psi}\mathbf{x}||_1=\sum_{k=1}^{N^2}\left|\sum_{i=1}^{N^2}\Psi_{ki}x_i\right|=\sum_{k=1}^{N^2}\sum_{i=1}^{N^2}\left(\mathrm{Re}\left(\Psi_{ki}x_i\right)^2+\mathrm{Im}\left(\Psi_{ki}x_i\right)^2\right)^{1/2}$$

The l^{th} element of the gradient is

$$\frac{\partial}{\partial x_l}||\mathbf{\Psi x}||_1 \sim \frac{\partial}{\partial x_l}\sum_{k=1}^{N^2}\left(\sum_{i=1}^{N^2}\left(\Psi_{ki}x_i\right)^*\left(\Psi_{ki}x_i\right)+\mu\right)^{\frac{1}{2}} =$$

$$=\sum_{k=1}^{N^2}\frac{\partial}{\partial x_l}\left(\sum_{i=1}^{N^2}\left(\Psi_{ki}x_i\right)^*\left(\Psi_{ki}x_i\right)+\mu\right)^{\frac{1}{2}} = \sum_{k=1}^{N^2}\frac{\sum_{i=1}^{N^2}\left(\Psi_{ki}x_i\right)\Psi_{kl}^*}{\left(\sum_{i=1}^{N^2}\left(\Psi_{ki}x_i\right)^*\left(\Psi_{ki}x_i\right)+\mu\right)^{\frac{1}{2}}}$$

where it was used the derivative of the square root. Notice the approximation in the derivative above. The square root in the denominator can go to infinity where the function inside the square root is near to zero. In order to avoid this issue, approximation of the square root is commonly used with a small smoothing parameter μ of the order of 10^{-6}.

B.4 Gradient total variation

Following the same idea, the gradient of total variation is computed. From the definition of discrete total variation (3.2.2) and the absolute value (B.1) it yields

$$\sum_{k=1}^{N^2}\sum_{j\sim\mathcal{N}_j}\sqrt{\left(\mathrm{Re}\left(x_k-x_j\right)\right)^2+\left(\mathrm{Im}\left(x_k-x_j\right)\right)^2+\mu}$$

Once again for the l^{th} element of the gradient

$$\frac{\partial}{\partial x_l}\sum_{k=1}^{N^2}\sum_{j\sim\mathcal{N}_j}\sqrt{\left(\mathrm{Re}\left(x_k-x_j\right)\right)^2+\left(\mathrm{Im}\left(x_k-x_j\right)\right)^2} \tag{B.4}$$

In the equation above, the index k sweeps the entire matrix while the index j stands for the 4-neighboring pixels of pixel x_k. Recall from Figure 3.4 that at any pixel x_l, the previous derivative is zero everywhere except for $k = l$ and $k = \{l-N, l-1, l+1, l+N\}$ because pixel x_l belongs to the 4-neighboring of the pixels in the latter group. For $k = l$ (B.4) yields

$$\sum_{j\sim\mathcal{N}_j}\frac{\mathrm{Re}\left(x_k-x_j\right)+\mathrm{Im}\left(x_k-x_j\right)}{\sqrt{\left(\mathrm{Re}\left(x_k-x_j\right)\right)^2+\left(\mathrm{Im}\left(x_k-x_j\right)\right)^2+\mu}} \tag{B.5}$$

This equation is formed by 4 sums. The gradients for the other four pixels at position $k = \{l-N, l-1, l+1, l+N\}$ has only one summand because the other 3

pixels have null derivative with respect to x_l. Since the index k visit the N^2 pixels in the matrix, the derivative for $k = l$ and the other 4 derivatives add up. Finally the TV derivative is

$$2 \sum_{j \sim \mathcal{N}_j} \frac{\mathrm{Re}\,(x_k - x_j) + \mathrm{Im}\,(x_k - x_j)}{\sqrt{(\mathrm{Re}\,(x_k - x_j))^2 + (\mathrm{Im}\,(x_k - x_j))^2 + \mu}}$$

Bibliography

[1] R.V. Damadian. "Tumor Detection by Nuclear Magnetic Resonance". Science, Vol. 171, 1151-1153, March 1971.

[2] P.C. Lauterbur. "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance". Nature, Vol. 242, 190-191, March 1973.

[3] "The shameful wrong that must be righted". October 2003. *http* : *//www.fonar.com/pdf/times*$_a$*d.pdf*

[4] IEC, International Electrotechnical Comission. "Medical electrical equipment Part 2-33: Particular requirements for the safety of magnetic resonance equipment for medical diagnosis". Project Number 60601-2-33. 2002.

[5] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, P. Boesiger. "SENSE: Sensitivity encoding for fast MRI". Magn Reson Med, Vol. 42(5),952-962, 1999.

[6] B. Liu, K. King, M. Steckner, J. Xie, J. Sheng, L. Ying. "Regularized sensitivity encoding (SENSE) reconstruction using Bregman iterations, Magn. Reson. Med., Vol. 61(1), 145152, 2009.

[7] Y. Xiaojing, C. Yunmei, L. Wei, H. Feng. "Fast MR Image Reconstruction for partially parallel imaging with arbitrary k-space trajectories". IEEE, Vol. 30 ,575 - 585, July 2010.

[8] M. A. Griswold, P. M. Jackob, R. M. Heidemann, M. Nittka,V. Jellus, J. Wang, B. Kiefer, A. Haase, "Generalized Autocalibrating Partially Parallel Acquisition (GRAPPA). Magn. Reson. Med., Vol. 47 (6), 12021210, 2002.

[9] E. Candès, J. Romberg, T. Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information". IEEE Trans. Inform. Theory, Vol. 52, 489-509, 2006.

[10] D. Donoho. "Compressed sensing". IEEE Trans. Inform. Theory, Vol. 52, 1289-1306, 2006.

[11] E. Candès, J. Romberg, T. Tao. "Stable signal recovery from incomplete and inaccurate measurements". Comm. Pure Appl. Math., Vol. 59, 1207-1223, 2006.

[12] E. Candès, J. Romberg. "Quantitative robust uncertainty principles and optimally sparse decompositions". Found. Comput. Math., Vol. 6, 227-254, 2006.

[13] R. Baraniuk. "Compressive sensing". IEEE Signal Processing Mag., Vol. 24, 118-120, 2007.

[14] E. Candès, M. Wakin. "An introduction to Compressive Sampling". IEEE Signal Processing magazine, March 2006.

[15] E. Candès, J. Romberg, Caltech "l1-MAGIC: Recovery of Sparse Signals via Convex Programming". October 2005.

[16] M. Guerquin-Kern, D. Van De Ville, C. Vonesch, J.C. Baritaux, K.P. Pruessmann, M. Unser. "Wavelet-Regularized Reconstruction for Rapid MRI". IEEE, 193-196, 2009.

[17] M. Guerquin-Kern, M. Hberlin, K. P. Pruessmann, M. Unser. "A Fast Wavelet-Based Reconstruction Method for Magnetic Resonance Imaging". IEEE Trans. Medi Imag. Vol. 30(9) 1649-1660, 2011.

[18] M. Lustig, D. Donoho, J. Santos, J. Pauly. "Compressed Sensing MRI". IEEE Signal Processing Magazine, Vol 25, 72-82, March 2008.

[19] Z. Liu, B. Nutter, J. Ao, S. Mitra. "Wavelet encoded MR image reconstruction with compressed sensing". Proc. SPIE, Conference Vol. 7961, 79613Z, 2011.

[20] Y. Yu, M. Honh, F. Liu, H. Wang, S. Crozier. "Compressed Sensing MRI Using Singular Value Decomposition Based Sparsity Basis". Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE, 5734 - 5737.

[21] T. Block, M. Uecker, J. Frahm. "Undersampled Radial MRI with Multiple Coils. Iterative Image Reconstruction Using a Total Variation Constraint". Magn. Reson. Med., Vol. 57, 1086-1098, 2007.

[22] J. P. Haldar, D. Hernando, Z. P. Liang. "Compressed-Sensing MRI with Random Encoding". IEEE Trans. on Medical Imaging,Vol. 30, 893 - 903, 2010.

[23] M. Lustig, J. M. Santos, D. L. Donoho, J. M. Pauly. "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity". ISMRM, Seattle, Washington, page 2420, May 2006.

[24] A. Abragam. “The Principles of Nuclear Magnetic Resonance”. Oxford University Press, Oxford 1988

[25] C. P. Slichter. “Principles of Magnetic Resonance”. Springer-Verlag, Berlin 1989

[26] E. M. Haacke, R.W. Brown, M.R. Thompsom, R. Venkatesan. “Magnetic Resonance Imaging. Physical Principles and Sequence Design”. John Wiley and Sons, New York, 1999.

[27] Z.P. Liang, P.C. Lauterbur. “Principles of Magnetic Resonance Imaging. A signal Processing Perspective”. IEEE Press, New York 2000.

[28] M.A. Brown, R.C. Semelka. “MRI Basic Principles and applications”. Wiley-Blackwel, 2010.

[29] D. W. McRobbie, E. A. Moore, M. J. Graves, M. R. Graves. “MRI from picture to proton”. Cambridge University Press 2006.

[30] C. Westbrook. “MRI at a glance”. Blackwell Science Ltd. 2002.

[31] M.T. Vlaardingerbroel, J.A der Boer. “Magnetic Resonance Imaging, theory and applications”. Springer, New York 2010.

[32] G. A. Wright. “Magnetic resonance imaging”. IEEE Signal Processing Mag., Vol. 14, 5666, Jan. 1997.

[33] B.H. Bransden and C.J. Joachain. “Quantum Mechanics”. Pearson Education, Harlow 2000.

[34] E. Hecht. “Optics”. Addison-Wesley, 2002.

[35] M.A Berstein, K.F King, X.J. Zhou. “Handbook of MRI. Pulse Sequence”. Elservier, 2004.

[36] R.M. Henkelman, M. J. Bronskill. “Artifacts is magnetic resonance imaging”. Magn. Reson. Med., Vol. 2, 1-126, 1986.

[37] B. P. Sutton, D. C. Noll, J. A. Fessler. “Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities”. IEEE Trans Med Imaging, Vol. 22, 178-188, Feb. 2003.

[38] L. C. Man, J. M. Pauly, A. Macovski.“ Multifrequency interpolation for fast off-resonance correction”. Magn Reson Med, Vol. 37, 785-792, May 1997.

[39] E.L. Hahn. “Spin Echoes”. American Physical Society, Vol. 80, 580-594, 1950.

[40] R. K. Wangsness. “Electromagnetic Fields”. John Wiley and Sons, New York, 2004.

[41] J.D. Jackson. “Classical Electrodynamics”. John Wiley ans Sons, New York, 1975.

[42] W.S. Hinshaw, A.H. Lent. “An introduction to NMR imaging: From the Bloch equation to the imaging equation”. Proceedings of the IEEE, Vol. 71, 338 - 350, March 1983.

[43] A. Papoulis. “Signal Analysis”. McGraw Hill , New York 1977.

[44] R.C. Gonzalez, R.E. Woods. “Digital Image Processing”. Pearson Prectice Hall, 2008.

[45] S. Mallat. “A Wavelet Tour of Signal Processing”. Academic Press, 1999.

[46] D. Donoho, I. Johnstone, J. Buckheit and S. Chen from Stanford University, J. Scargle at NASA-Ames Research Center. WaveLab: wavelet library. http://stat.stanford.edu/ wavelab/.

[47] S.G Mallat, Z. Zhang. “Matching pursuits with time-frequency dictionaries”. IEEE Trans Signal Process, Vol. 41, 3397-3415, 1993.

[48] B.K. Natarajan. “Sparse approximate solutions to linear systems”. SIAM J Comput, Vol. 24, 227-234, 1995.

[49] S.S. Chen, D.L. Dohono, M. A. Saunders. “Atomic Decomposition by Basis Pursuit”. SIAM J. Sci. Comput.,Vol. 20, 33-61, 1998.

[50] S. Boyd and L. Vandenberghe. ”Convex Optimization”. Cambridge University Press, 2004.

[51] E. Candès, J.K. Romberg. “Signal recovery from random projections”. Proceedings of SPIE Computational Imaging III, 5674. 2005.

[52] R.M. Henkelman. “Measurement of signal intensities in the presence of noise in MR images”. Med Phys. Vol. 12, 232-233, Mar-Apr. 1985.

[53] H. Gudbjartsson, S. Patz. “The Rician Distribution of Noisy MRI Data”. Magn. Reson. Med., Vol. 34, 910-914, December 1995.

[54] J. Nocedal, S.J. Wright.“Numerical optimization”. Springer New York, 2006.

[55] P. C. Hansen. “Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion”. SIAM, Philadelphia, 1998

[56] H. Schomberg, J. Timmer. “The gridding method for image reconstruction by Fourier transform”. IEEE Trans. Medi Imag. Vol. 14, 596 - 607, 1995.

[57] J.A. Fessler. “On NUFFT-based gridding for non-Cartesian MRI”. Journal of Magnetic Resonance, Vol. 188, 191195, 2007.

[58] Y. C. Eldar, G. Kutyniok. “Compressed sensing. Theory and applications”. Cambridge University Press May, 2012.

[59] L. Rudin, S. Osher, and E. Fatemi. “Nonlinear total variation based noise removal algorithms”. Physica D, Vol. 60, 259-268, 1992.

[60] A.V. Oppenheim, J.S. Lim. “The Importance of Phase in Signals”. Proceeding of the IEEE, Vol. 69, 529 - 541 , May 1981.