COMPRESSIVE SENSING IN MAGNETIC RESONANCE IMAGING

by

Emmanouilidou S. Dimitra

A thesis submitted to the faculty of

University Of Crete

in partial fulfillment of the requirements for the degree of

M.Sc. in Computer Science

Computer Science Department

University of Crete

November 2010
UNIVERSITY OF CRETE
DEPARTMENT OF COMPUTER SCIENCE

COMPRESSIVE SENSING IN MAGNETIC RESONANCE IMAGING

by Dimitra S. Emmanouilidou
A thesis submitted to the faculty of University Of Crete
in partial fulfillment of requirements for the degree of
M.Sc. in Computer Science

Committee Approval :

Panagiotis Tsakalides, Professor, Department of Computer Science, University of Crete,
Supervisor

Athanasios Mouchtaris, Assistant Professor, Department of Computer Science, Univer-
sity of Crete

Kostas Marias, Principle Researcher, Institute of Computer Science, FO.R.T.H.

Department Approval :

Angelos Mpilas, Professor, Director of Postgraduate Studies, Department of Computer
Science, University of Crete

Heraklion, November 2010
ACKNOWLEDGMENTS

At first, I would like to thank my supervisor professor mr. Panagiotis Tsakalides for his continuous support and help through the M.Sc. program. Professor Tsakalides was constantly available for guidance and support and he always listened to every question, significant or not. I am really grateful for the values, the encouragement, the guidance and motivation, the trust and the support during the last one and a half years of collaboration.

I would also like to thank professor mr. Ioannis Tollis for his encouragement and guidance through my undergraduate studies. He was the first to get me involved with researching in the world of Biomedical Imaging, throughout the first year of my M.Sc. program. I really appreciate his believing in me, his guidance and all the support.

Mr. Franco Chiarugi has always been there for me for the last three years. He has really inspired me very much and helped me during my first steps in the research community. I truly could not have come this far without his knowledge, guidance, encouragement and great collaboration. My gratitude is all I have to offer.

I would also like to thank Dimitris Mil., Giorgos Tzag., Nikos Pap., Kalliroi Fl., Stefanos Pap., Christos Tzag., Pavlos Ch., Despoina Pav., Sofia Nik., Manolis Pl., Anthony Gr., Manolis Sp., Kiriaki Kais., and everyone else in the laboratories of Bioinformatics and of Telecommunications and Networks for all the support and guidance.

All these years of studying and living in Crete would be unbearable without my dearest friends who were always on my side. Thank you all Lucy Mel., Eva Zach., Nikos Gian., Stella Kop., Ioanna Pan., Kyriaki Myr., Spyridoula Zach., Theonymfi
Mak., Katerina Dem., Aggelina Kour., Stratos Kap., Aris Ef., and all other friends who were always there for me in difficulties and in good times. I have learned so much from each of you. You are all in my heart and made me the person I am today... and I guarantee you that this is a compliment!

Last but not least, I would like to thank the two people that made it all possible: my parents Stefanos Emmanouilidis and Rodopi Gougoulaki. Words are not enough to express how grateful and proud I am for my parents. Thank you for believing in me, supporting me and most of all loving me. You are the best parents and the best people I have ever known. The love and encouragement of my two dearest brothers, Nikos and Manos, has kept me going throughout all the years of studying. I wish all brothers in the world were like you. I really love you guys.
Compressive Sensing

in

Magnetic Resonance Imaging

Συμπιεστική Δειγματοληψία

στην

Απεικόνιση Μαγνητικού Συντονισμού
Abstract

Since the first magnetic nuclear image was produced in 1973, magnetic resonance (MR) imaging has evolved into a clinically indispensable and effective tool in diagnostic medicine. Over the years, the rapid growth in clinical applications has been accompanied by numerous technological advances in MR imaging (MRI). Much of this evolution has been accomplished through advances in localization techniques and new mechanisms of contrast which have greatly improved image quality. Still, MRI could benefit from approaches for scan time reduction, with benefits for patients and health care economics.

The nature of the MRI constitutes a natural fit for Compressive Sensing (CS). Compressive sensing is a novel framework for recovering and reconstructing compressible signals from undersampled data. The theory of CS goes beyond conventional compression schemes where a signal should be sampled first and compressed afterwards, by stating that a successful signal reconstruction can be guaranteed with high probability by solving a convex optimization problem using only a small number of linear combinations of the signal’s values. Successful reconstruction is guaranteed under two assumptions, namely, the signal is sparse or compressible in some basis and the signal measurements are acquired through "random" sampling. The MR modality meets the two assumptions above. Indeed, MR images are either naturally sparse or may be sparsely represented in an appropriate transformed domain. Furthermore, MR acquisition schemes are quite flexible and can be explicitly designed in order to incorporate the notion of randomness.

In this thesis, we study the performance of three compressive sensing algorithms when applied to magnetic resonance signal modalities. Our goal is to present the basic MRI concepts as incorporated into the theory of CS, in a fashion that is comprehensible to a wide range of readers. All methods use the CS theory to recover the undersampled raw MR data and reconstruct the MR image but they differ in the minimization formulation of the reconstruction schemes they employ. The methods are thoroughly analyzed, compared and evaluated in terms of reconstruction quality, algorithmic complexity, and time consumption. The first method, Smoothed $\ell_0$, invokes the theory of CS and uses an $\ell_0$ approximation to solve the reconstruction problem. It is a very fast technique with low complexity. The two other methods exhibit higher complexity but they are able to achieve better reconstruction results: $\ell_1$-magic, a commonly used reconstruction algorithm, solves the optimization problem through Newton steps while Sparse MRI uses a non linear gradient descent technique with backtracking. The algorithms presented herein provide a coherent understanding of the secrets and the ideas behind both CS and MRI theories.
Περίληψη

Από τη πρώτη μαγνητική τυρινική εικόνα που παράγχηκε το 1973, η απεικόνιση μαγνητικού συντονισμού (MS) εξελίχθηκε σε ένα κλινικό απαραίτητο και αποτελεσματικό εργαλείο απεικόνισης στη διαγνωστική ιατρική. Η ραγδαία ανάπτυξη σε κλινικές εφαρμογές συνοδεύτηκε, με το χρόνο, από πολυάριθμες καινοτομίες και προσέγγιση στην απεικόνιση MS. Σε σημαντικούς βαθμούς, αυτή η εξέλιξη έχει επιτευχθεί μέσω αλλαγών στις τεχνικές εντοπισμού και μέσω νέων μηχανισμών αντίδρασης που έχουν βελτιωθεί σε μεγάλο βαθμό την ποιότητα της εικόνας. Η MS απεικόνιση μπορεί να επωφεληθεί από μεθόδους μέτρησης του χρόνου σύρωσης, με ομάδα τόσο για τον ασθενή, όσο και στο κόστος της παροχής των υπηρεσιών υγείας.

Ο μηχανισμός της απεικόνισης MS αποτελεί ένα ταχιστό πεδίο εφαρμογής της νέας θεωρίας της Συμπεπτυστικής Δεγματοληψίας (ΣΔ). Η ΣΔ είναι ένα καινοτόμος μέθοδος για την ανάκτηση και ανακοσμοποίηση αραιών σημάτων, από δεδομένα που έχουν ληφθεί με υπο-δεγματοληψία. Η θεωρία της ΣΔ προγραμμάζει πέρα από τη συνήθη μεθοδολογία της συμπίεσης όπου ένα σήμα πρώτα πρέπει να δεγματοληφθεί κατά Νυμπμις και έπειτα να συμπεπτεί, αποδεικνύοντας πως μια επιτυχής ανακοσμοποίηση ενός αραιού σήματος μπορεί να πραγματοποιηθεί με μεγάλη πιθανότητα, λόγω να ένα πρόβλημα βελτιστοποίησης με περιορισμούς, χρησιμοποιώντας ένα πολύ μικρό αραιό μετρήσεων. Το πλήθος αυτών των μετρήσεων, γραμμικών συνδυασμών των τιμών του σήματος, είναι ανάλογο της αραιότητας του σήματος και μερικότατο από το πλήθος των δειγμάτων που προβλέπει η κατά Νυμπμις θεωρία. Δύο προσεπεθέσεις εγκινούν την επιτυχία αυτής ανακοσμοποίησης: το σήμα πρέπει να είναι αραιό ή συμπίεσιμο σε κάποια βάση και οι μετρήσεις του σήματος πρέπει να αποτελούν μέσα ενός «τυχαίου» τρόπου δεγματοληψίας. 

Η διαδικασία MS πληροί τις προϋποθέσεις αυτές. Πράγματι οι εικόνες MS είναι είτε φυσικά αραιές είτε μπορούν να ανακατασταθούν αραιά σε κάποιο κατάλληλο πεδίο μεταγχηματισμού. Ακόμη, η διαδικασία που λαμβάνεται το σήμα MS μέσω του τομογράφου, είναι αρκετά ευέλικτη και προσφέρει, και μπορεί να σχεδιαστεί με τρόπο που να ενσωματώνει την έννοια της τυχαίας λήψης δειγμάτων.

Σε αυτή τη διατριβή μελετάμε την απόδοση τριών ολογράμμων ΔΣ κατά την εφαρμογή τους σε σήματα μαγνητικών τομογραφιών. Στόχος μας είναι η παρουσίαση των βασικών ιδεών της απεικόνισης ΜΣ και με τρόπο που να ενσωματώνει τη θεωρία της ΣΔ. Όλες οι μέθοδοι χρησιμοποιούν την θεωρία ΣΔ για την ανάκτηση των ακατέργαστων δεδομένων ΜΣ από υπο-δεγματοληψία, με σκοπό την ανακοσμοποίηση της εικόνας ΜΣ. Οι ολογράμμοι διαφέρουν ως προς τη διατύπωση και τον τρόπο που λύνουν το πρόβλημα βελτιστοποίησης. Οι μέθοδοι παρουσιάζονται αναλυτικά, συγκρίνονται και αξιολογούνται με βάση την ποιότητα ανακοσμοποίησης, την ολογραμμική πολυπλοκότητα αλλά και τις απαιτήσεις χρόνου. Η πρώτη μέθοδος, Smoothed –l0, λύνει το πρόβλημα της ανακοσμοποίησης χρησιμοποιώντας μια προσέγγιση της l0 νόρμας. Είναι μια πολύ γρήγορη τεχνική με χαμηλή πολυπλοκότητα. Οι άλλες δύο μέθοδοι έχουν μεγαλύτερη πολυπλοκότητα αλλά επιτυγχάνουν καλύτερη απόδοση ανακοσμοποίησης: η μέθοδος l0-magic, λύνει το πρόβλημα βελτιστοποίησης της l1 νόρμας μέσω της μέθοδου Newton και χρησιμοποιείται ευρέως από τη κοινότητα ΣΔ. Η τεχνική SparseMRI, χρησιμοποιεί ένα μη-γραμμικό βαθμιδωτό ολόγραμμο χαλάδου κλίσης με σπειροειδή δόμηση. Οι ολογράμμοι που παρουσιάζονται συνεισφέρουν στην κατανόηση και στη σύνδεση των μυστικών που κρύβονται πίσω από τις θεωρίες ΣΔ και απεικόνισης ΜΣ.
Contents

Table of Contents

1 Magnetic Resonance (MR) Imaging ............................................. 13
  1.1 Magnetic Fields .............................................................. 14
    1.1.1 Main static magnetic field, $B_0$ .................................. 15
    1.1.2 RF excitation pulse, $B_1$ .......................................... 16
    1.1.3 Gradient magnetic fields ............................................ 20
  1.2 Principles of Magnetic Resonance Imaging .............................. 22
    1.2.1 Slice Selection ....................................................... 22
    1.2.2 Readout or Frequency Encoding ................................... 24
    1.2.3 Phase Encoding ..................................................... 27
  1.3 Anatomy of a pulse sequence and Image Acquisition .................. 29
  1.4 Nuclear Magnetic Resonance Signal ..................................... 34
  1.5 The patient’s experience .................................................. 37

2 Compressive Sensing (CS) in MR Imaging .................................. 39
  2.1 Problem Statement .......................................................... 40
  2.2 Theory and Principles of Compressive Sensing ......................... 41
    2.2.1 Transform operator .................................................. 41
    2.2.2 Sensing operator .................................................... 49
    2.2.3 The restricted isometry property (RIP) .......................... 51
    2.2.4 Randomness: Why is it so important? ............................. 52
    2.2.5 Signal recovery ..................................................... 54
  2.3 History and Highlights of CS in MR Imaging ............................ 56

3 Exhaustive Comparison of three Algorithms ............................... 59
  3.1 Smoothed $\ell$-0 Algorithm .............................................. 60
    3.1.1 The Algorithm Outline ............................................. 62
    3.1.2 Experiments .......................................................... 63
    3.1.3 Results and Findings .............................................. 65
  3.2 $\ell_1$-magic Algorithm ................................................... 72
    3.2.1 Min-TV with equality constraints ................................. 72
    3.2.2 Results and Findings .............................................. 76
## Contents

3.3 Sparse MRI Algorithm ........................................ 85  
3.3.1 Results and findings ....................................... 87  
3.4 Discussion on Results ....................................... 90  

4 Bayesian Compressive Sensing 105  
4.1 Marginal Likelihood Maximization ............................ 107  
4.2 The Algorithm .................................................. 108  
4.3 Results and findings ........................................... 109  

5 Conclusions - Future work 113  

6 Appendix 115  
6.1 Measurement Matrix Φ .......................................... 115  
6.2 Transform Matrix Ψ ............................................ 116  
6.3 Wavelets Analysis: the basic idea ............................. 116  

Chapter 1

Magnetic Resonance (MR) Imaging

Magnetic Resonance Imaging (MRI) is an imaging method, based mainly on the properties of hydrogen nucleus. It has evolved since the beginnings of the 1970s, where the first nuclear magnetic resonance image was produced by Lauterbur [1], and has become the imaging method of choice for a large proportion of medical examinations. Body tissues are made up to 70-90% by water which contains 2 hydrogen nucleus. A possible disease or injury may drastically change the properties and amounts of water in specific tissues, and this makes Magnetic Resonance (MR) a very sensitive diagnostic technique. MR can be used to visualize the anatomy or the pathology of a body area, to investigate organ functionality or even brain activity. The examination takes place in the MR scanner (figure 1.1a), where the patient lies down. Sagittal, coronal or transverse imaging of any body area is possible using an MR scanner(figure 1.1b).
CHAPTER 1. MAGNETIC RESONANCE (MR) IMAGING

(a) An MR scanner.  (b) Head visualization via MRI.

Figure 1.1: All kinds of visualization are possible with the MR technology: sagittal, coronal, transverse or even diagonal aspect, of any body area. From [2] pp. 69

The two basic ideas behind Magnetic Resonance Imaging are the magnetic fields and the magnetization of protons in the body. Three kinds of magnetic fields are involved: the main field of the scanner, the gradients which are used for spatial localization, and the oscillating magnetic field of the RF pulses. In MRI, the signal originates from the hydrogen nucleus, which is a single proton. The response of hydrogen to an applied magnetic field is one of the largest found in nature. An indication of a high hydrogen concentration may indicate fat or tumor, whereas low concentration may indicate a bone area. The fact that the human body is mostly composed of tissues, mainly containing water and fat makes hydrogen nucleus a very natural choice for the MR techniques.

1.1 Magnetic Fields

A relevant property of the hydrogen nucleus (composed of a single proton) is its spin (intrinsic angular moment), $I$. The a proton spin has an associated dipole magnetic moment, $\mu$, collinear with $I$. In a single volume element (ex. of a tissue) corresponding to a pixel in an MR image, there are many protons, each with an associated dipole magnetic moment. The net magnetization $\mathbf{M}$, of the volume element, is the vector sum of the individual dipole moments $\mathbf{M} = M_x \mathbf{i} + M_y \mathbf{j} + M_z \mathbf{k}$, where $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are unit vectors along $x$, $y$, $z$ axes respectively. In the absence of a magnetic field, the spatial orientation of each dipole moment is random and $\mathbf{M}=0$. 
1.1. MAGNETIC FIELDS

1.1.1 Main static magnetic field, $B_0$

The main magnetic field, $B_0$, is constantly present during an MR scan. It has a constant strength (static) and is directed along the $z$ axis. Upon the placement of the body tissue inside this strong magnetic field $B_0$, the magnetic moments induced by proton spins all over the body are aligned with the field: the protons commence to precess in a rotation axis parallel to $B_0$ with a (Larmor) frequency $\omega_0$ proportional to the strength of the magnetic field:

$$\omega_0 = \gamma B_0$$

(1.1)

Frequency $\omega_0$ is measured in megahertz (MHz), the strength, $B_0$, of the magnetic field in Tesla (T) and the gyro-magnetic ratio, $\gamma$, that refers to a property of the nucleus in question (water or fat protons), has the value $\sim 42$ Tesla/MHz for the hydrogen proton. $B_0$ is usually 1.5 to 3 Tesla, i.e. 30000 to 60000 times the strength of the Earth’s magnetic field. In the presence of $B_0$, the tissue is polarized and the magnetization of the protons, $M$, is aligned along $B_0$. This means that the net magnetization of all protons’ spins have grown from zero toward an equilibrium value $M_0$ along $z$ axis. So, in equilibrium we have $M = M_0 = M_0 k$. This induced magnetization is the source of the MR signal. The proton’s experience of the static magnetic field, as described so far, is illustrated in figure 1.2.

![Figure 1.2: (a) A single proton inside the $B_0$ magnetic field. It precesses around z axis. (b) Microscopic(i) and macroscopic(ii) pictures of a collection of protons in the presence of $B_0$. The x,y,z directions correspond to the ones shown in figure 1.1(a). For a better view, the z axis has been drawn vertically. The rate or frequency of precession is proportional to the strength of the magnetic field. From [3] pp.18-20](image)

Inside the magnetic field, a proton precesses or evolves about the magnetic field (figure 1.2(a)). The precessional axis is parallel to the main magnetic field $B_0$. The
z component of the spin vector (projection of the spin onto the z axis) is the component of interest because it does not change in magnitude or direction as the proton precesses. The x and y components vary with the time at a frequency $\omega_0$ proportional to $B_0$. In figure 1.2(b), it is shown a microscopic(i) and macroscopic(ii) picture of a collection of protons in the presence of the external magnetic field. Each proton precesses around the direction of the applied magnetic field. The z components are one of the two values (one positive and one negative), but the x and y components can be any value, positive or negative. The protons will appear to track along two "cones", one with a positive z component and one with a negative z component. Because there are always more protons in the upper cone, there will be a nonzero vector sum $M$, the net magnetization. It will be of constant magnitude and parallel to $B_0$.

### 1.1.2 RF excitation pulse, $B_1$

The main magnetic field has been turned on and the protons’ magnetization lies on z axis. This very strength of the magnetization may reveal the amount of hydrogen concentration inside the body area under consideration. There is only one thing preventing us from separately measuring it: it has a very low strength compared to the strength of $B_0$. The body’s magnetization is of the order of few $\mu$ Tesla, whereas the field’s is about 1-10 Tesla. For this reason, a second magnetic field $B_1 = \cos(\omega_0 t)i + \sin(\omega_0 t)j$, is applied. Despite $B_0$ being static, notice that $B_1$ is time-varying. It is also called 90° RF (Radio-Frequency) excitation pulse and is used to create a magnetic field transverse to $B_0$ and thus tip $M$ into the transverse plane (figure 1.3). $B_1$ rotates about the static field direction $k$ at radian frequency $\omega_0$ and the produced energy (magnetization) is absorbed by the protons. This absorbed energy causes the net magnetization of the protons to precess in a plane transverse of $B_0$, at angular frequency $\omega_0$. Accompanying the rotating dipole magnetic moment of $M$ is a radiated electromagnetic force (emf) which is the signal detected in MRI, known as Free Induction Decay (FID), by a resonant receiver coil, sensitive to the transverse $B_1$ magnetization. However, the MR signal, like all kinds of energy, cannot avoid suffering of relaxation.
1.1. MAGNETIC FIELDS

Figure 1.3: The application of the RF pulse produces the magnetic field $B_1$, which rotates about the direction of $B_0$. Protons absorb the energy of $B_1$ resulting in the alignment of the net magnetization on the rotating x-y frame, transverse to $B_0$. A receiver coil detects the emitted MR signal.

Relaxation is the dynamic physical process in which the system of spins release the energy that they absorbed from the RF pulse and returns to equilibrium, and affects both the longitudinal ($M_z$) magnetization and the transverse ($M_{xy}$) magnetization. Relaxation can be described by Bloch equations and is a fundamental process in MR, as essential as energy absorption, and provides the primary mechanism for image contrast. The signal emitted by the precision of $M$ will eventually decay as a result of the thermal equilibrium and the net magnetization $M$ will return again parallel to $B_0$ resulting in a longitudinal relaxation described by the time constant $T_1$. Supplementary, over time, the vector sum, $M_{xy}$, decreases in magnitude towards zero, since the individual dipole moments no longer add constructively (spin dephasing) causing the transverse relaxation described by time constant $T_2$. Both $T_1$ and $T_2$ times measure the spontaneous energy transfer by an excited proton, but they differ in the final disposition of the energy.

**$T_1$ Relaxation**

The relaxation time $T_1$ is the time required for the z component of $M$ to return to 63% of its original value following an excitation pulse. It is also known as the spin-lattice relaxation time or longitudinal relaxation time. Recall that $M_0$ is parallel to $B_0$ at equilibrium and that energy absorption will rotate $M_0$ into the transverse plane. $T_1$ relaxation provides the mechanism by which the protons give up their energy to return to their original orientation. If a 90° pulse is applied to a sample in equilibrium, $M_0$ will rotate as illustrated in Figure 1.3, and there will be no longitudinal magnetization at the end of the pulse.

As time goes on, a return of the longitudinal magnetization will be observed as
the protons release their energy. This return of magnetization follows an exponential growth process, with $T_1$ being the time constant describing the rate of growth:

$$M(\tau) = M_0(1 - \exp^{-\tau/T_1})$$ (1.2)

where $\tau$ is the time following the RF pulse. After three $T_1$ time periods, $M$ will have returned to 95% of its value prior to the excitation pulse, $M_0$. The term spin lattice refers to the fact that the excited proton (“spin”) transfers its energy to its surroundings (“lattice”) rather than to another spin. The energy no longer contributes to spin excitation.

Since the emitted signal decays due to the above reasons, the RF excitation schemes needs to be applied more than once to refocus the protons and recover the $M_z$ magnetization component. Due to practical reasons, as for example, $B_0$ field inhomogeneity, the time between successive RF pulses is usually insufficient for complete $T_1$ relaxation so that $M$ will not be completely restored to $M_0$. Application of a second RF pulse prior to complete relaxation will rotate $M$ into the transverse plane, but with a smaller magnitude than following the first RF pulse. Notice however that that $T_1$ decreases with decreasing strength of the magnetic field.

A lower frequency $\omega_0$ (lower $B_0$) is closer to the low frequency of an excited proton, meaning that there is a better match of the two frequencies. As spin-lattice relaxation measures energy transfer from the excited proton to its surroundings, the closer the frequencies, the more readily the motion (i.e. rotation) absorbs the energy and the more frequently this energy transfer occurs, allowing the collection of protons to return to its equilibrium configuration sooner. This is also the reason that a larger $B_0$ does not necessarily translate to a greater signal, as saturation is more prevalent due to the longer $T_1$ times.

**T2 Relaxation**

The relaxation time $T_2$ is the time required for the transverse component of $M$ to decay to 37% of its initial value via irreversible processes. It is also known as the spin-spin relaxation time or transverse relaxation time. Recall that $M_0$ is oriented only along the $z$ ($B_0$) axis at equilibrium and that no portion of $M_0$ is in the xy plane. The coherence (uniformity) is entirely longitudinal. Absorption of energy from a 90° RF pulse, as in Figure 1.3, causes $M_0$ to rotate entirely into the xy plane, so that the coherence is in the transverse plane at the end of the pulse. As time elapses, this coherence disappears, while at the same time the protons release their energy and reorient themselves along $B_0$. This disappearing coherence produces the MR signal (FID) as described earlier. As this coherence disappears, the value of $M$ in the xy plane decreases toward zero. $T2$ or $T2^*$ relaxation is the process by which this transverse magnetization is lost. $T2^*$ includes dephasing caused by magnetic field inhomogeneities and susceptibility effects as well: $1/T2^* = 1/T2 + 1/2\gamma \Delta B$, where $\gamma$
1.1. MAGNETIC FIELDS

is the gyromagnetic ratio, and $\Delta B$ is the variation in magnetic field.

At the end of the $90^\circ$ RF pulse, when the protons have absorbed energy and are oriented in the transverse plane, each proton precesses at the same frequency $\omega_0$ and is synchronized at the same point or phase of its precessional cycle. Since a nearby proton of the same type will have the same molecular environment and the same $\omega_0$, it can readily absorb the energy that is being released by its neighbor. Spin-spin relaxation refers to this energy transfer from an excited proton to another nearby proton. The absorbed energy remains as spin excitation rather than being transferred to the surroundings as in T1 relaxation. This proton-proton energy transfer can occur many times as long as the protons are in close proximity and remain at the same $\omega_0$.

Molecular interactions such as vibrations or rotations cause $\omega_0$ to fluctuate. This fluctuation produces a gradual, irreversible loss of phase coherence to the spins as they exchange energy and reduce the magnitude of the transverse magnetization and the generated signal. T2 is the time when the transverse magnetization is 37% of its value immediately after the $90^\circ$ pulse when this irreversible process is the only cause for the loss of coherence. As more time elapses, this transverse coherence completely disappears, only to reform in the longitudinal direction as T1 relaxation occurs. This dephasing time T2 is always less than or equal to T1. Proper design of the pulse sequence may eliminate this dephasing. The other sources (dephasing time due to the main field inhomogeneity and dephasing time due to the magnetic susceptibility differences) contribute to the total transverse relaxation time, $T2^\ast$.

This decay of the transverse magnetization following a $90^\circ$ RF pulse, the FID, follows an exponential process with the time constant of $T2^\ast$ rather than just T2:

$$M_{XY}(t) = M_{XY,\text{max}} \exp\left(-t/T2^\ast\right)$$

(1.3)

where $M_{XY,\text{max}}$ is the transverse magnetization $M_{XY}$ immediately following the excitation pulse. Ways of reversing some sources of proton dephasing include the application of extra refocusing pulses, i.e. additively apply other magnetic fields, to re-align the dipole moments of the protons, or even use other methods depending on the scanner’s technology.

The important difference between T1 and T2 relaxation is in the influence of $B_0$. As mentioned earlier, T1 is very sensitive to $B_0$, with longer T1 times measured for a tissue at higher $B_0$. T2 is relatively insensitive to $B_0$ at the relatively large field strengths currently used in MRI. Only at very low $B_0$ (less than 0.05 Tesla) there will be significant changes in T2. The different T1 and T2 relaxation times of the different kinds of tissues, enables the contrast scaling of an MR image. Chemical substances (contrast agents) introduced to the anatomical or functional region being imaged, can increase the differences between different tissues or between normal and abnormal tissue, by altering the relaxation times. The relative difference of signal
intensities in two adjacent regions of an image may be emphasized or eliminated, with the help of these agents, according to the necessity of the MR application.

So, after completing the process of measuring the emitted MR signal, what is it exactly that we have measured?

1.1.3 Gradient magnetic fields

The application of the $B_0$ and $B_1$ fields, has made it possible for the protons to emit a measurable signal. But what is it exactly that we have measured? As the magnetic moment of protons rotates with frequency $\omega_0$, we measure, via the coil, the signal emitted by the net magnetization i.e. the protons’ density of the whole scanned body-area. Is there a way, however, to spatially resolve the signal and obtain knowledge of the hydrogen concentration of every small tissue area (voxel)? The answer is to use special magnetic fields, known as the Gradient fields.

Although the MR signal is based at the RF frequency $\omega_0$ (Larmor frequency), it can actually contain many different frequencies that encode information about the location of various tissues. MR image reconstruction attempts to visualize the MR signal (as detected by the receiver coil), depicting the spatial distribution of the transverse magnetization. The variation of the MR signal upon position is possible through the use of Gradient fields, which are the third type of the basic magnetic fields, involved in MRI.

Gradients are additional spatially linear varying magnetic fields. They are not constantly present during an MR scan, as opposed to $B_0$, and can be applied in any of the three Cartesian directions. These gradients are small perturbations superimposed on the main magnetic field $B_0$, with a typical imaging gradient producing a total field variation of less than 1%. They are also linear perturbations to $B_0$, so that the exact magnetic field is linearly dependent on the location inside the magnet.

In Figure 1.4 one may see the effect of an $x$ Gradient ($G_x$): it will add to or subtract from the magnitude of the static field at different points along the $x$ axis or $x$ direction. The total field is represented in the figure by the spacing of the "field lines". In (a) only the main field is present, whereas in (b) an $x$ gradient has been added. Protons experiencing the same field have the same frequency (a), but when a gradient is applied (b), the magnetic field produced by the gradient is added to the main field $B_0$, making the frequency to vary upon position.
1.1. MAGNETIC FIELDS

Figure 1.4: Effect of field gradient on nuclei: (a) if only $B_0$ is present, all nuclei process at the same frequency. (b) But if we have $B_0$ plus a gradient $G_x$, the precession frequency now depends upon spatial position. Protons in different x positions rotate with different frequencies, depending on the magnetic field they experience. From [2] pp. 110-111

Notice that, at the center ($x = 0$), the total field experienced by a nuclei is simply $B_0$, so these spins resonate at the Larmor frequency. As we move along the x direction, however, the total either increases or decreases linearly and thus the protons resonate faster or slower. Faster or slower precession is detected as higher or lower frequencies in the MR signal, and so frequency measurements are used to distinguish between MR signals at different positions in space. Gradients are applied in a controllable manner. They can be applied in any direction or orientation and are normally present only for a short time. It is these three sets of gradients that give MR its three dimensional capability.

So, when a gradient is applied, the total field in the gradient’s direction experienced by nuclei will be dependent upon the position in real space. For example, in the presence of a $G_x$ gradient, the magnetic field in position $x_i$ will alter to $B(x_i) = B_0 + x_i \cdot G_x$ and so the precession frequency will depend upon the total component (in the gradient’s direction) of the magnetic field and will become spatially dependent. This implies that the presence of magnetic field gradients requires an expanded version of the Larmor equation of Equation 1.1:

$$\omega_i = \gamma \cdot (B_0 + r_i \cdot G) \quad (1.4)$$

where $\omega_i$ is the frequency of the proton at position $r_i$ and $G$ is a vector representing
the total gradient amplitude and direction. Although the gradients are oriented in any of the three orthogonal directions, the gradient magnetic fields themselves are parallel to the main magnetic field $B_0$. Equation 1.4 states that, in the presence of a gradient field, each proton will resonate at a unique frequency that depends on its exact position within the gradient field. The MR image is simply a frequency and phase map of the protons generated by unique magnetic fields at each point throughout the image. The displayed image consists of digital picture elements (pixels) that represent volume elements (voxels) of tissue. The pixel intensity is proportional to the number of protons contained within the voxel weighted by the $T_1$ and $T_2$ relaxation times for the tissues within the voxel.

### 1.2 Principles of Magnetic Resonance Imaging

Spatial localization of the MR signal requires the use of three orthogonal linear magnetic field gradients. These are generated by gradient coils mounted on a cylindrical former just inside the bore of the magnet. In a standard cylindrical magnet, the direction along the bore is termed the z axis, the left-right direction is termed the x axis and the top-bottom direction is termed the y axis (recall figure 1.1). Each MR image is a 2D representation of a 3D slice of the patient. A consecutive application of different gradients fields allow us to view a 3D volume like a series of 2D images, as obtained by a slice selection process.

#### 1.2.1 Slice Selection

The initial step in MRI is the localization of the RF excitation to a region of space, which is accomplished through the use of frequency-selective excitation in conjunction with a gradient known as the slice selection gradient, $G_{SS}$. We will use the bold notation for the gradient vector ($\mathbf{G}$), and the simple $G$ notation for the gradient’s amplitude. The gradient direction (x, y, or z) determines the slice orientation, whereas the gradient amplitude together with certain RF pulse characteristics determine both the slice thickness and slice position.

A frequency-selective RF pulse has two parts associated with it: a central frequency and a narrow range or bandwidth of frequencies (typically 1-2 kHz). When such a pulse is broadcast in the presence of the slice selection gradient, only a narrow region of tissue achieves the resonance condition (Equation 1.4) and absorbs the RF energy. The duration of the RF pulse and its amplitude determines the amount of resulting proton rotation (e.g., $90^\circ$, $180^\circ$).

The central frequency of the pulse determines the particular location excited by the pulse when the slice selection gradient is present. Different slice positions are
achieved by changing the central frequency. The slice thickness is determined by the gradient amplitude $G_{SS}$ and the bandwidth of frequencies $\Delta \omega_{SS}$ incorporated into the RF pulse:

$$\Delta \omega_{SS} = \gamma \Delta (G_{SS} \ast \text{Thickness}). \quad (1.5)$$

Typically, $\Delta \omega_{SS}$ is fixed so that the slice thickness is changed by modifying the amplitude $G_{SS}$. Thinner slices require larger $G_{SS}$. Once $G_{SS}$ is determined by the slice thickness, the central frequency is calculated using Equation 1.4 to bring the desired location into resonance. Multislice imaging, the most commonly used approach for MRI, uses the same $G_{SS}$ but a unique RF pulse during excitation for each slice. Each RF pulse has the same bandwidth but a different central frequency, thereby exciting a different region of tissue (Figure 1.5).

The slice orientation is determined by the particular physical gradient or gradients defined as the logical slice selection gradient. The slice orientation is defined so that the gradient orientation is perpendicular or normal to the surface of the slice, making every proton within the slice to experience the same total magnetic field regardless of its position. Orthogonal slices are those in which only one of the x, y, or z gradient is used as the slice selection gradient. Oblique slices, those not in one of the principal directions, are obtained by applying more than one physical gradient when the RF pulse is broadcast.

The total gradient amplitude, whether from one, two, or three physical gradients, determines the slice thickness, as shown in Equation 1.5. When images are viewed on the monitor or film, the slice selection direction is always perpendicular to the surface, that is, hidden from the viewer and a 2D slice image may be visualized.

We have shown the way, how the signal emitted by a whole 3D volume, is reduced to the signal emitted by a specific 2D slice of the volume. However, in order to further resolve the emitted signal in the 2D (x-y) plane, more gradient fields have to be applied in the Frequency and Phase direction of the 2D k-space.
24

CHAPTER 1. MAGNETIC RESONANCE (MR) IMAGING

Figure 1.5: Slice selection process. (a) In the presence of a gradient $G_{SS}$, the total magnetic field that a proton experiences and its resulting resonant frequency depend on its position, according to Equation 1.4. Tissue located at position $z_i$ will absorb RF energy broadcast with a center frequency $\omega_i$. Each position will give a unique resonant frequency. The slice thickness $\Delta z$ is determined by the amplitude of $G_{SS}$ and by the bandwidth of transmitted frequencies $\Delta \omega$. (b) The relationships between pixels in the screen and the patient being scanned. The pixel in the screen is just the front face of a three-dimensional voxel within a patient. From [3] pp.45-46

1.2.2 Readout or Frequency Encoding

The signal detection portion of the MRI measurement is known as the readout (RO) or frequency encoding (FE). In an imaging pulse sequence, the MR signal is always detected in the presence of a gradient known as the readout gradient $G_{RO}$, which produces one of the two visual dimensions of the image on the 2D film. A typical pulse sequence uses some form of excitation, such as a 90° slice-selective pulse ($B_1$), to excite a particular region of tissue. Following excitation, the net magnetization within the slice is oriented transverse to $B_0$ and will precess with frequency $\omega_0$. The readout gradient, $G_{RO}$, is then applied perpendicular to the slice direction. Under the influence of this new gradient field, the protons begin to precess at different frequencies depending on their position within it, in accordance with Equation 1.4. Each of these
frequencies is superimposed into the signal. At the desired time, the signal is measured by the receiver coil and digitized for later process (Fourier transformation). The magnitude of $G_{RO}$ ($G_{RO}$) and the frequency that is detected enable the corresponding position of the proton to be determined (Figure 1.6).

![Figure 1.6: Readout process. Following excitation, each proton within the excited volume (slice) precesses at the same frequency. During signal detection, a gradient ($G_{RO}$) is applied, causing a variation in the frequencies for the protons generating the signal. The frequency of precession $\omega_i$ for each proton depends upon its position $x_i$, according to Equation 1.4. Frequencies measured are mapped to the corresponding position. From [3] pp.48](image)

The magnitude $G_{RO}$ is determined by two user-definable parameters: the field of view (the distance across an image, typically in centimeters) in the readout direction, $FOV_{RO}$, and the Nyquist frequency, $\omega_{NQ}$, for the image, often referred to as the receiver bandwidth. This relationship is expressed in Equation 1.6:

$$\Delta \omega_{RO} = 2 \ast \omega_{NQ} = \gamma \Delta(G_{RO} \ast FOV_{RO})$$

(1.6)

where $\Delta \omega_{RO}$ is the total range of frequencies in the image. $G_{RO}$ is chosen so that protons located at the edge of $FOV_{RO}$ precess at the Nyquist frequency for the image (Figure 1.7). Smaller $FOV_{ROS}$ are achieved by increasing $G_{RO}$, keeping the Nyquist frequency and thus the total frequency bandwidth constant.
CHAPTER 1. MAGNETIC RESONANCE (MR) IMAGING

Figure 1.7: In any image, one of the visualized directions is the readout direction and the other is the phase-encoding direction. A proton located at the edge of the FOV in the readout direction precesses at the Nyquist frequency $\omega_{NQ}$ above or below the transmitter frequency $\omega_{TR}$. Changing the FOV of the image changes the spatial resolution (mm per pixel) but not the frequency resolution (Hz per pixel). From [3] pp.49

In an MR image, the resolution may be expressed in one of two ways: spatial resolution or frequency resolution. The spatial resolution, expressed as the voxel size (VOX) with units of mm/pixel, is derived from two user parameters: FOV$_{RO}$ and the number of readout sample points in the acquisition matrix, $N_{RO}$:

$$VOX_{RO} = \frac{FOV_{RO}}{N_{RO}}$$ (1.7)

The frequency resolution, with units of Hz/pixel, is based on $N_{RO}$ and the total bandwidth $\Delta \omega_{RO}$ for the image:

$$\text{Pixel bandwidth} = \frac{\Delta \omega_{RO}}{N_{RO}} = 2 \times \frac{\omega_{NQ}}{N_{RO}}$$ (1.8)

It is possible to improve the frequency resolution for the measurement independent of the spatial resolution by increasing the total sampling time used to measure the signal. This reduces the Nyquist frequency for the image and the background noise contributing to the measurement. In order to maintain the correct spatial resolution within the image, $G_{RO}$ is reduced, in accordance with Equation 1.6.
1.2.3 Phase Encoding

The third direction in an MR image is the phase encoding (PE) direction. It is visualized along with the readout direction in an image (see Figure 1.7). The phase encoding gradient, $G_{PE}$, is perpendicular to both $G_{SS}$ and $G_{RO}$ and is the only gradient that changes amplitude during the data acquisition loop of a standard two-dimensional (2D) imaging sequence. Any signal amplitude variation detected from one acquisition to the next is assumed to be caused by the influence of $G_{PE}$ during the measurement.

The principle of phase encoding is based on the fact that the proton precession is periodic in nature. Prior to the application of $G_{PE}$, a proton within a slice precesses at the base frequency $\omega_0$. In the presence of $G_{PE}$, its precessional frequency increases or decreases according to Equation 1.4. Once $G_{PE}$ is turned off, the proton precession returns to its original frequency, but is ahead or behind in phase relative to its previous state. The amount of induced phase shift depends on the magnitude and duration of $G_{PE}$ that the proton experienced and the proton location.

Protons located at different positions in the phase encoding direction experience different amounts of phase shift for the same $G_{PE}$ pulse (Figure 1.8). A proton located at the edge of the chosen FOV experiences the maximum amount of phase shift from each phase encoding step.

The MR image information is obtained by repeating the slice excitation and signal detection multiple times, each with a different amplitude of $G_{PE}$. The second Fourier transformation in the image converts signal amplitude at each readout frequency from a function of $G_{PE}$ to a function of phase.
Figure 1.8: Concept of phase encoding. Prior to application of $G_{PE}$, all protons precess at the same frequency. When $G_{PE}$ is applied, a proton increases or decreases its precessional frequency, depending on its position, $y_i$. A proton located at $y_i = 0$ ($y_2$) experiences no effect from $G_{PE}$ and no change in frequency or phase ($\phi_2 = 0$). A proton located at $y_3$ precesses faster while $G_{PE}$ is applied. Once $G_{PE}$ is turned off, the proton precesses at its original frequency, but is ahead of the reference frequency (dashed curve); that is, a phase shift $\phi_3$ has been induced in the proton by $G_{PE}$. A proton located at $y_1$ decreases its frequency while $G_{PE}$ is applied. Once $G_{PE}$ is turned off, it precesses at its original frequency but is behind the reference by a phase shift of $\phi_1$. From [3] pp.52-53

The spatial resolution in the phase-encoding direction depends on two user-selectable parameters: the field of view in the phase encoding direction $FOV_{PE}$, and the number of phase encoding steps in the acquisition matrix, $N_{PE}$. The $FOV_{PE}$ is determined by the change in $G_{PE}$ from one step to the next. $N_{PE}$ determines the total number of cycles of phase change produced. The spatial resolution in the phase-encoding direction is expressed as the voxel size and is measured in mm/pixel:

$$VOX_{PE} = \frac{FOV_{PE}}{N_{PE}}$$ (1.9)

Increased resolution is obtained by reducing the $FOV_{PE}$ or by increasing $N_{PE}$. The $FOV$ reduction is accomplished by increasing the gradient amplitude change from one $G_{PE}$ to the next.
Because of the two different physical processes involved, the \( \text{FOV} \) in the phase-encoding direction is not required to be the same as the \( \text{FOV} \) in the readout direction, nor is the voxel size. The ratio of \( \text{VOX}_{RO} \) to \( \text{VOX}_{PE} \) is known as the aspect ratio between the two dimensions. An aspect ratio of 1.0 (100\%) means that the voxel size is the same in both directions, a situation referred to as isotropic resolution. An aspect ratio less than 1.0 (< 100\%) is referred to as anisotropic resolution, with \( \text{VOX}_{PE} \) typically larger than \( \text{VOX}_{RO} \).

After the discussion on the various magnetic fields, and their presence during an MR scan, the following section completes the understanding, by explaining the way each one of them contributed to the final MR image acquired.

### 1.3 Anatomy of a pulse sequence and Image Acquisition

The combination of gradient pulses, RF pulses, data sampling periods, and the timing between each of them that are used to acquire an image is known as a pulse sequence. The functionality of the gradients, as discussed earlier, may be better understood looking at an MR pulse sequence diagram, commonly used for describing the operation of the magnetic fields in MR. Gradient pulses are applied in a controlled manner during an MRI scan, and form an MR pulse sequence, in order to interrogate every possible spatial frequency that may contribute in the image.

In Figure 1.9 one may see how a general pulse sequence works (basic gradient-echo sequence). Each horizontal line corresponds to a different hardware component (e.g. RF transmitter, gradients, analog-to-digital converter (ADC) sampling). The vertical axis represents amplitude, and the elapsed time, during sequence execution, is indicated from left to right along the horizontal axis.
CHAPTER 1. MAGNETIC RESONANCE (MR) IMAGING

Figure 1.9: Basic gradient-echo MR imaging sequence. (a) Vertically the amplitude is shown, and horizontally the time. $G_{SS}$ is the slice-selective gradient; $G_{PE}$ is the phase-encoding and $G_{RO}$ the frequency encoding gradient. For every different amplitude of the $G_{PE}$ gradient, the $G_{RO}$ gradient goes through the steps A-D. Changing the amplitude of $G_{PE}$ results in the change of the $k_{PE}$ coordinate in the k-space. From [2] pp. 109

In figure 1.9, first (top line), an RF pulse is applied simultaneously with a slice-selective gradient $G_{SS}$ (line 2 of the figure). The RF pulse stimulates the MR interactions in tissue which lead to the MR signal. By combining the RF excitation with this gradient, the MR interactions are restricted to a two-dimensional plane, slab or slice. Any physical gradient $G_x$, $G_y$, or $G_z$ or combinations of these can be used for the purpose of $G_{SS}$, allowing us to produce transverse, sagittal or coronal, oblique or double oblique slices.

Next, in line 3, phase encoding is applied in a direction orthogonal to the slice selection and encodes the MR signal in the PE direction. The amplitude of this gradients represents the $k_{PE}$ position in the k-space. When the position is set, the RO gradient is applied (line 4) in the third direction ($k_{RO}$). Finally, line 5 shows the time when the MR signal is measured or acquired. Note that this is during the RO gradient but after the PE. The whole sequence pattern has to be repeated for every 'line' of data, corresponding to a different value of $G_{PE}$ until the data or k-space matrix is filled. A time period, TR, occurs between the application of one RF excitation and the next. This way of sampling the k-space, line per line is called Cartesian sampling trajectory.

We usually choose the RO direction to be the longer one (longer with respect to the anatomical axis of the body part we want to scan), and the PE direction to be the short one. This way the k-space is sampled much faster as in the opposite
scenario more RF pulses would need to be applied in order to cover all different $k_{RO}$ positions. When we set the frequency- (FE) and phase-encoding (PE) matrix, we are controlling the matrix size of the final image and also the size of k-space. For example if we choose a size of 256 for frequency encoding, each MR echo will have 256 sample points, thus requiring 256 columns in the k-space matrix for temporary storage (see figure 1.10). When we set the PE matrix, we control how many echoes have to be acquired and thus how many rows are needed in k-space (more rows result in greater scan time, and this is why we choose the PE anatomical axis to be the shorter one).

![Figure 1.10](image)

**Figure 1.10:** In (a) we see the k-space which is a raw data space. The computer stores the digitized raw data during the scan, in a section of memory which has the same number of rows and columns the final image (b). During a conventional gradient-echo or spin-echo scan, k-space is filled with raw data one line per time TR (c). From [2] pp. 66

Although both k-space and real space have the same matrix size, the pixels do not correspond directly with each other. That means that the information in the bottom left pixel in k-space does not contain the raw information for the bottom left pixel in the image. This is because the reconstruction processing uses the Fourier transform. Instead, data in the middle of k-space contain all the signal-to-noise and contrast
CHAPTER 1. MAGNETIC RESONANCE (MR) IMAGING

information for the image, and data around the outside contain all the information about the image resolution like edges and boundaries (figure 1.11).

![Figure 1.11: By reconstructing only the data from the middle of k-space (a) we get all the signal and contrast information, but it is very blurred (b). If we erase the middle of k-space (c) and just reconstruct the outside data we can see where the tissue boundaries are (d), but the signal-to-noise ratio is very low and we have no contrast information. Clearly we need both parts of k-space to get a useful MR image! From [2] pp. 67](image)

Once all the data are acquired a two-dimensional Fourier transform is applied. This converts the data, already encoded as spatial frequencies, into an image. Reconstruction in MRI is generally simpler than, e.g., in X-ray Computed Tomography; most of the hard work has been done during the acquisition by the gradients. The 2D scanning process as described in figure 1.9 is one of the simplest possible in MR imaging.

There are plenty of different MR pulse designs. When K-space is acquired (or ‘traversed’) one line per excitations pulse, adjacent $k_{PE}$ lines are acquired serially in time and are controlled by the $G_{PE}$ gradient. This kind of filling order is called a Cartesian trajectory. Other trajectory types also exist exploiting properties of the K-space and gradient hardware. Since the center of the k-space contains denser data, as opposed to the peripherals, we may exploit such a knowledge and construct trajectories like in figure 1.12 (b-d).
1.3. ANATOMY OF A PULSE SEQUENCE AND IMAGE ACQUISITION

Dense sampling may be done in the center and sparse sampling in the outer space. The modalities of the hardware in the MR scanner allow the direction of the gradient to change, but in a smooth way. For example a trajectory consisting of random position sampling in the k-space is not possible with the current hardware technology. The trajectory has to follow a quite smooth route in order to be realizable.

It is time to understand the kind of data that an MR scanner acquires with the help of the gradient fields, also referred as the k-space. Nuclear physics may explain the nature of the acquired signal. In the following section, the basic, yet complete theory behind the MR signal acquisition is presented.
1.4 Nuclear Magnetic Resonance Signal

According to standard NMR (Nuclear Magnetic Resonance) theory the contribution of a small voxel $d\mathbf{r}$ (located at position $\mathbf{r}$ in the $\mathfrak{R}$-space), to the NMR signal is:

$$dS(t) = q(r) \exp^{-t/T_2} \exp^{2\pi if t} d\mathbf{r}$$  \hspace{1cm} (1.10)

where $f$ is the Larmor frequency of the nucleus and $q(r)$ represents the local density of the observed nucleus (magnetization distribution of $|M_{xy}|$). The magnetization component $M_{xy}$ that rotates around $B_0$, will eventually decay. This irreversible signal loss, as discussed earlier, is modeled by an exponential, with $T_2$ time constant being the transverse relaxation time. Recall that the $M_z$ component also has a relaxation time, $T_1$. Protons in environment corresponding to different materials (ex for tissues in the head: fat, cerebrospinal fluid, etc) have different relaxation times $T_1$ and $T_2$ and the differences between these parameters are used to produce contrast between materials in an MR image.

The phase detector of an MRI scanner which measures the signal component $dS(t)$, in practice, subtracts a fixed operator-defined reference frequency $F$, and $\Delta f$ is the RF offset of the signal component. If $\Delta f = f - F$, the signal equation becomes:

$$dS(t) = q(r) \exp^{-t/T_2} \exp^{2\pi i \Delta f t} d\mathbf{r}$$  \hspace{1cm} (1.11)

After the application of a generic field gradient $G$, the Larmor frequencies of different voxels spread upon position $\mathbf{r}$, according to:

$$\Delta f = \gamma G \cdot \mathbf{r} = \gamma(xG_x + yG_y + zG_z)$$  \hspace{1cm} (1.12)

In general, a gradient $G$ is time-dependant, and we could say that $\Delta f$ are angular velocities of the magnetization vector (or angular velocities of the signal phase evolution rate). In this sense, the total accumulated signal phase $\Phi(t)$ at time $t$ would be:

$$\Phi(t) = \int_0^t \gamma G(t) \cdot \mathbf{r} dt$$  \hspace{1cm} (1.13)

These very gradient fields will form the so-called K-space:

$$\mathbf{k}(t) = \gamma \int_0^t G(t) dt$$  \hspace{1cm} (1.14)

The receiver coil detects a total signal $S(t)$, which is the integral of the contributions of $dS(t)$ arising from all voxels of the imaged object:

$$S(t) = \int_V q(r) \exp^{-t/T_2} \exp^{2\pi i (\mathbf{k}(t) \cdot \mathbf{r})} d\mathbf{r}$$  \hspace{1cm} (1.15)
Neglecting the decay term \(e^{-t/T_2}\) and assuming it remains approximately equal to the initial value \(= 1\) during the experiment we may rewrite eq. 1.15 as:

\[
S(t) = \int_V q(r) \exp^{2\pi i(k(t) \cdot r)} dr = Q(k(t))
\]  

(1.16)

In other words, \(Q(k(t))\) is the \(K\)-space transform of \(q(r)\) and the signal received by the coil \((S(t))\), is the value of \(Q(k(t))\) along the \(K\)-space path \(k(t)\). The significant points here are that the path \(k(t)\) is time-dependent, and always starts at the origin \((k(0) = 0)\). After the pulse excitation (appliance of field \(B_1\)), we depart from the origin of the K-space and using gradients, we move along any desired K-space path (remember figure 1.9).

Along the way, a record of the \(Q(k)\) values is built up (digital sampling), for a subset of the visited K-space points, describing the \(Q(k(t))\) surface. Once enough data have been collected to approximate \(Q(k(t))\) for any value of \(k(t)\), we may recover the \(R\)-space function \(q(r)\) via the inverse transform. Effective resolution of an \(R\)-space image is encoded into the peripheral K-space regions. Although the signal intensity is low in the peripheral, the outer data is essential for a good resolution (recall figure 1.11).

Signal \(Q(k(t))\) is complex, unlike \(q(r)\), and if the procedure of the scanning could be noiseless, then we could introduce the notion of symmetry \((Q(-k(t)) = Q(k(t)))\), which is not the case in reality. The area of the K-space where \(Q(k(t))\) reaches large values is limited to a central region. At large distances it becomes smaller than the noise (figure 1.13). Note that \(Q(k(t) = 0)\) is the integral over the whole volume of interest \(V\). Applying a constant Gradient \(G\), makes us in the K-space move away from the center at a constant speed in some fixed direction. After a \(\Delta t\) we head for the peripherals. When being in the peripherals, if we change sign of the gradient, then the sign of the velocity will also be inverted and start going backwards! The K-space image is transient: it gets re created after every application of an excitation sequence and then starts decaying back to zero. So the cartographer of the K-space deals with time limitations and with the appropriate choice of K points.
Figure 1.13: Large values of the MR signal are concentrated on the center of the k-space. By scaling large-to-low values using white-to-black intensities, k-space is represented as in (a). If a 3D view was available, we would see something like in (b) where high amplitude signal values are represented by height.

In case that the scanned subject is not placed in the center of the MRI scanner \( r_1 = r + d \) the only thing changing is the balance among real and imaginary K-space components at different points, and not the power of \( Q(k(t)) \) signal. With \( r=(x,y,z) \) being that vector that corresponds to the R-space position, vector \( k(t)=(k_x(t), k_y(t), k_z(t)) \) corresponds to the k-space position. Using the notation \( k \) instead of \( k(t) \), but not forgetting that the gradients may be time-dependant, we may rewrite Eq.(1.16) and express \( Q(k) \) as the inverse Fourier transform of the spin density \( q(r) \):

\[
Q(k) = \int_V q(r) \exp^{2\pi i \cdot k \cdot r} \, dr \tag{1.17}
\]

If we take the case of a 2D plane where we have selected a slice through gradient \( G_{SS} \) and our only changing fields are \( G_{RO} \) of the readout encoding and \( G_{PE} \) of the phase encoding, then position in R-space would be \( r = (x,y) \) and position in K-space would be \( k = (k_x, k_y) \), where

\[
k_x = k_{RO} = \frac{\gamma}{2\pi} \cdot G_{RO} \cdot t_{RO}, \quad \text{and}
\]

\[
k_y = k_{PE} = \frac{\gamma}{2\pi} \cdot G_{PE} \cdot t_{PE}, \tag{1.18}
\]

with \( t_{RO} \) and \( t_{PE} \) corresponding to the cumulative time in which the respective gradient is active. The spectrum measured in MRI of the object being imaged is the K-space and consists of an array or matrix of individual spatial frequencies. Remember that the reconstructed image has the same size as its K-space data matrix, with the readout direction displayed in the horizontal plane, and the phase-encoding direction displayed in the vertical plane. Its dimensions depend on the number of
1.5. THE PATIENT’S EXPERIENCE

readout data points \( N_{RO} \) and the number of phase encoding steps \( N_{PE} \) for the scan. \( N_{RO} \) and \( N_{PE} \) are set by the gradients properties during the designing of the MR pulse. The speed of k-space traversal is limited by physical constraints. In current systems, gradients are limited by maximum amplitude and maximum slew-rate (rate of change). In addition, high gradient amplitudes and rapid switching can produce peripheral nerve stimulation. Since this must be avoided, physiology provides a fundamental limit to gradient system performance.

Summarizing the basic concepts of the MR theory, we have seen that the MR signal is based at a Radio Frequency \( \omega_0 \) which later is fixed by the strength of the main magnetic field). However the MR signal contains many different frequencies that encode information about the location of various tissues. Although both K-space and real space have the same matrix size, the pixels do not correspond directly with each other. Exploiting nuclear physic theory, we are able to express the k-space as being Fourier domain. Intuitively, data in the middle of K-space contain all the signal-to-noise and contrast information for the image, and data around the outside contain all the information about the image resolution (edges and boundaries). After a scanner have digitally acquired the whole grid of k-space data, a 2D Fourier transform allows us to view the scanned body area as a 2D \( \Re \) space image.

But why bothering to improve the acquisition process and acquisition time since MRI is already a very successful, powerful and effective imaging technique, both for prognosis and diagnosis? Let’s see, shortly, the MR scanning procedure from a patient’s point of view, and the answer will then be clear.

1.5 The patient’s experience

Before the procedure starts, patients need to remove all metallic objects, jewelry, watches and credit cards or any magnetic friendly cloth, since the magnets of the scanner produce strong magnetic fields and accidents may occur. Some rare occasions have been report where the patient failed to report the existence of pacemaker, resulting in its strong, fatal magnetization. A patient moved into the scanner may be claustrophobic. There have been cases where the patient became hysterical inside the scanner. Also, large patients may not fit into the bore of the scanner. This results in the patient having to stay still in an uncomfortable position during the whole experiment.

As the patient is exposed to the main static field, the gradient fields and RF fields (the main effect of RF is the heating of tissue), in extreme cases, the gradients can induce peripheral nerve stimulation which may be alarming or annoying (but it is not harmful). Furthermore, the MR scanner can be very noisy during the operation producing high knocking sounds. In some MR imaging sequences the noise level can
exceed safety guidelines and except from irritating, also become harmful. This is why ear-plugs are recommended—given to the patients during scanning.

In cases when a patient needs an MR scanning of a large body area, i.e. a leg, he may be asked to stay still even for a whole hour or more. The slightest movement can produce artifacts in the reconstructed image, meaning that the experiment need to be done from the beginning! The fundamental limits on the gradients amplitude and their slew-rate by physical constraints do not allow great hardware improvements and have caused the research community to search for ways of reducing the amount of acquired data (i.e. the scan time) without degrading the quality of the image produced.

Frequently discussed approaches in the literature of medical imaging for reconstructing an object from polar frequency samples are the so-called filtered backprojection algorithms. In a nutshell one assumes that the Fourier coefficients at all of the unobserved frequencies are zero and thus reconstructing the image if ‘minimal energy’ under the observation constraints. This strategy does not perform very well, and could hardly be used for medical diagnostics as the artifacts produced by under-sampling degrades the image seriously. It seems that a good reconstruction algorithm would have to guess the values of the missing Fourier coefficients i.e. interpolate the missing frequencies. However a possible prediction of missing values based on neighboring coefficients is a very problematic procedure due to the highly oscillating nature of the Fourier transform.

A strategy based on convex optimization, the Compressive Sensing theory, is the answer to the above constraints and inconveniences, as caused by both an algorithmic point of view and the patients’s scope. It is a newly applied theory on the MR field and has already been proved very promising.
Chapter 2

Compressive Sensing (CS) in MR Imaging

CS was first proposed in the literature of Information Theory and Approximation Theory in an abstract general setting. One measures a small number of random linear combinations of the signal values - much smaller than the number of signal samples nominally defining it. The signal is reconstructed with good accuracy from these measurements by a non-linear procedure.

Data redundancy or correlation in the data infers that the data is compressible and can be represented by fewer coefficients in a certain bases. Compression in the conventional sense is not applicable here, in an MRI context since the object is typically not known beforehand and also data may possibly not be acquired in the corresponding basis given the encoding mechanism used in MRI.

The widespread success of compression algorithms with real images raised the following questions: Since the images we intend to acquire will be compressible, with most transform coefficients negligible or unimportant, is it really necessary to acquire all that data in the first place? Can’t we simply measure the compressed information directly from a small number of measurements, and still reconstruct the same image which would arise from the fully sampled set? In MRI we look at a special case of CS, where the sampled linear combinations are simply individual Fourier coefficients (k-space samples). In that setting, CS is claimed to be able to make accurate reconstructions from a small subset of k-space, rather than an entire k-space grid.

A first step towards exploiting data redundancy in this direction was made by David Donoho et al. [7] who set the basis by defining the discrete uncertainty principle. More recent publications [5], [6] outline a general mathematical framework for the reconstruction of compressible signals from undersampled data, a method referred to as Compressive Sensing. In the context of CS, the following notation is considered:
\( m \): signal of interest \( \in \mathbb{R}^N \)

\( \Psi \): sparsifying transform matrix, \( N \times N \) of signal \( m \)

\( \Phi \): measurement matrix, \( M \times N \) applied on \( m \) or \( \Psi^{-1}m \)  \((2.1)\)

\( y \): the measurements acquired \( \in \mathbb{R}^M \)

### 2.1 Problem Statement

The conventional compression scheme, that of transform coding, is founded on the concept that compressible \( N \)-signals are well approximated by \( K \)-sparse representations. In data acquisition systems (for example, digital cameras) transform coding plays a central role: the full \( N \)-sample signal \( m \) is acquired; then, the complete set of transform coefficients \( \{s_i\} \) is computed via \( s = \Psi^{-1}m \); the \( K \) largest coefficients are located and the \( (N - K) \) smallest coefficients are discarded; and the \( K \) values and locations of the largest coefficients are encoded. Unfortunately, this "sample-then-compress" framework suffers from three inherent inefficiencies. First of all, the initial number of samples \( N \) may be large even if the desired \( K \) is small. Even if acquisition follows the Nyquist-Shannon sampling rate - at least two times the signals’ bandwidth \([8]\)- still, the number of samples remains very high. Secondly, the set of all \( N \) transform coefficients \( \{s_i\} \) must be computed even though all but \( K \) of them will be discarded. Finally, the locations of the large coefficients must be encoded, thus introducing an overhead.

**The Compressive Sensing Problem:** Compressive sensing address these inefficiencies by directly acquiring a compressed signal representation without going through the intermediate stage of acquiring \( N \) samples. Consider a general linear measurement process that computes \( M < N \) inner products between \( m \) and a collection of vectors \( \{\phi_j\}_{j=1}^M \) as in \( y_j = \langle m, \phi_j \rangle \). Arrange the measurements \( y_j \) in an \( M \times 1 \) vector \( y \) and the measurement vectors \( \phi_j^T \) as rows in an \( M \times N \) matrix \( \Phi \). Then, by substituting \( m = \Psi s \), \( y \) can be written as

\[
y = \Phi \Psi s \tag{2.2}
\]

where \( \Phi \Psi \) is an \( M \times N \) matrix. The measurement process is not adaptive, meaning that \( \Phi \) is fixed and does not depend on the signal \( m \).

**The goal of Compressive Sensing:** Compressive Sensing tries to recover the full signal \( m \), having only knowledge of the measured data \( y \). The process of measuring the signal of interest, \( m \), is also known as the sensing process. According to conventional sampling, in order not to have any signal loss, the signal should be acquired (sensed) according to Nyquist rate: the number of samples needed to reconstruct a signal without error is dictated by its bandwidth - the length of the shortest
2.2. THEORY AND PRINCIPLES OF COMPRESSIVE SENSING

interval which contains the support of the spectrum of the signal under study.

In the world of Compressive Sensing, it is possible to reconstruct the original signal \( \mathbf{m} \), almost flawlessly, using much fewer measurements than those implied by Nyquist, as long as \( \mathbf{m} \) can be sparsely represented in some domain, and the measurements are acquired in a random sense (as will explained later). Infinitely many vectors \( \mathbf{m} \) can yield the recorded measurements \( \mathbf{y} \) due to the rank deficiency of the matrix \( \Phi \cdot \Psi \). Compressive Sensing calculates the solution of the problem stated below via \( \ell_1 \)-minimization:

\[
\hat{\mathbf{m}} = \arg \min_{\mathbf{m}} \|\mathbf{m}\|_{\ell_1} \quad s.t. \|\Phi \mathbf{m} - \mathbf{y}\|_{\ell_2}^2 \leq \epsilon
\]

(2.3)

\( \epsilon \) may be considered as an error tolerance or as an accepted noise level. Despite the fact that infinitely many vectors \( \mathbf{m} \) can yield the recorded measurements of \( \mathbf{y} = \Phi \mathbf{m} \), the scheme of Eq. 2.3 guarantees a unique solution high very high probability [5]-[10]. However, in order for the scheme in Eq.2.3 to work well, two sufficient conditions need to be satisfied: (a) the signal of interest must be sparse, and (b) the Sensing Operator must obey a restricted isometry property (RIP).

2.2 Theory and Principles of Compressive Sensing

2.2.1 Transform operator

A sparsifying transform is an operator mapping a vector of image data to a sparse vector. In recent years, there has been extensive research in sparse image representation. As a result we currently possess a library of diverse transformations that can sparsify many different type of images. For example, piecewise constant images can be sparsely represented by spatial finite-differences (i.e, computing the differences between neighboring pixels) ; indeed, away from boundaries, the differences vanish. Real-life MR images are of course not piecewise smooth. But in some problems, where boundaries are the most important information (angiograms for example) computing finite- differences results in a sparse representation.

The measured signal, \( \mathbf{x} \) must be sparse itself (contain a large number of zero coefficients) or alternately compressible (contain a large number of negligible coefficients). Consider, here, the case that \( \mathbf{x} \) itself is not sparse but can be sparsely represented via a transform basis, \( \Psi \). Let it be a real-valued, finite-length, one-dimensional, discrete-time signal \( \mathbf{m} \), which can be viewed as an \( N \times 1 \) column vector in \( \mathbb{R}^N \) with elements \( \mathbf{m}[n] \), \( n = 1, 2, \ldots, N \). (We treat an image or higher-dimensional data by vectorizing it into a long one-dimensional vector.) Any signal in \( \mathbb{R}^N \) can be represented in terms of a basis of \( N \times 1 \) vectors \( \{\psi_i\}_{i=1}^N \). For simplicity, assume that the basis is orthonormal. Thus, using the \( N \times N \) basis matrix \( \Psi = [\psi_1|\psi_2|...|\psi_N] \) with the vectors
\{\psi_i\} as columns, a signal \(\mathbf{m}\) can be expressed as

\[
\mathbf{m} = \Psi \mathbf{s} = \sum_{i=1}^{n} s_i \psi_i
\]  \hspace{1cm} (2.4)

where \(\mathbf{s}\) is the \(N \times 1\) column vector of weighting coefficients \(s_i = \langle \mathbf{m}, \psi_i \rangle = \psi_i^T \mathbf{m}\) and \(^T\) denotes transposition (See also Fig. 2.1).

Figure 2.1: Illustrative example of the transform process. Signal \(\mathbf{s}\) is the sparse representation of \(\mathbf{m}\) in the \(\Psi\) domain.

More precisely, if \(\mathbf{s}\) contains just \(K\) non-zero coefficients, then it is called a \(k\)-sparse signal and is strictly sparse. In the case where the rest \(N - K\) coefficients are not exactly zero, but small enough, approaching zero, we have a weak sparseness. Clearly, \(\mathbf{m}\) and \(\mathbf{s}\) are equivalent representations of the signal, with \(\mathbf{m}\) in the time or space domain and \(\mathbf{s}\) in the \(\Psi\) domain. Sparsifying transforms that are preferred among the research community for application of CS in MRI are Discrete Cosine Transform (DCT), Wavelet Transform, Finite Differences (or else Total Variation) since Magnetic Resonance Images such as a brain axial-view image or an angiography are highly sparse in these domains.

Using these sparsifying transforms, MR images may be completely represented using only a very low percentage of high valued coefficients. Fig 2.2(a-b) demonstrates the simple zero-filling reconstruction of an MR image where only a 1, 5, 10, 30 or 50 % of the highest coefficients of the transformed image are kept. The rest coefficients are filled with zeros, and an inverse transform produces the reconstructed images shown. It is clear that MR images are sparse enough themselves or may be sparsely represented using only few non zero coefficients on a transform domain and, thus, fit perfectly in the ”sparsity” requirement of the CS theory.
2.2. THEORY AND PRINCIPLES OF COMPRESSIVE SENSING

Figure 2.2: Transform-domain sparsity of images. (a) Axial brain image; (b) coronal leg image. The DCT, wavelet, and finite-differences transforms were calculated for all the images (Left column). The images were then reconstructed from a subset of 1, 5, 10, 20, 30, 50% of the largest transform coefficients, and compared to the original images - 100%.

A short description of the Discrete Cosine and Wavelet Transform follows, for a better understanding of the sparsifying processes. Readers familiar with the trans-
CHAPTER 2. COMPRESSIVE SENSING (CS) IN MR IMAGING

form, may proceed directly to section 2.2.2

Discrete Cosine Transform: The Discrete Cosine Transform (DCT) is a Fourier-like transform, which was first proposed in 1974 [14]. While the Fourier Transform represents a signal as the mixture of sines and cosines, the Cosine Transform performs only the cosine-series expansion. Like other transforms, the Discrete Cosine Transform (DCT) attempts to decorrelate the image data. After decorrelation each transform coefficient can be encoded independently without losing compression efficiency. This section describes the DCT and some of its important properties. The most common DCT definition of a 1-D sequence of length N is

\[
C(u) = a(u) \sum_{x=0}^{N-1} f(x) \cos\left[\frac{\pi(2x + 1)u}{2N}\right] \tag{2.5}
\]

for \( u = 0, 1, 2, \ldots, N - 1 \). Similarly, the inverse transformation is defined as

\[
f(x) = \sum_{u=0}^{N-1} a(u)C(u) \cos\left[\frac{\pi(2x + 1)u}{2N}\right] \tag{2.6}
\]

for \( x = 0, 1, 2, \ldots, N - 1 \). In both Eq.2.5 and 2.6, \( a(u) \) is defined as

\[
a(u) = \begin{cases} 
\sqrt{1/N}, & \text{for } u = 0 \\
\sqrt{2/N} & \text{for } u \neq 0
\end{cases} \tag{2.7}
\]

It is clear from Eq.2.5 that for \( u = 0 \), \( C(u = 0) = \sqrt{1/N} \sum_{x=0}^{N-1} f(x) \). Thus, the first transform coefficient is the average value of the sample sequence. In literature, this value is referred to as the DC Coefficient. All other transform coefficients are called the AC Coefficients. To better understand the meaning of the equation, let’s ignore the \( f(x) \) and \( a(u) \) components and plot the \( \sum_{x=0}^{N-1} \cos\left[\frac{\pi(2x + 1)u}{2N}\right] \) component alone for \( N = 8 \) and varying values of \( u \). Then Fig 2.3 would be produced.

![Figure 2.3: One dimensional cosine basis function, \( \sum_{x=0}^{N-1} \cos\left[\frac{\pi(2x + 1)u}{2N}\right] \), for varying \( u \) when \( N=8 \).](image-url)
Indeed, the first waveform \((u = 0)\) renders a constant (DC) value, whereas, all other waveforms \((u = 1, 2, ..., 7)\) give waveforms at progressively increasing frequencies. These waveforms are called the cosine basis function. Note that these basis functions are orthogonal. Hence, multiplication of any waveform in Fig 2.3 with another waveform followed by a summation over all sample points yields a zero (scalar) value, whereas multiplication of any waveform in Fig 2.3 with itself followed by a summation yields a constant (scalar) value. Orthogonal waveforms are independent, that is, none of the basis functions can be represented as a combination of other basis functions.

The 2-D DCT is a direct extension of the 1-D case and is given by

\[
C(u, v) = a(u)a(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left[ \frac{\pi(2x + 1)u}{2N} \right] \cos \left[ \frac{\pi(2y + 1)v}{2N} \right],
\]

for \(u, v = 0, 1, 2, ..., N - 1\) and \(a(u)\) and \(a(v)\) as defined in 2.7. The inverse 2-D DCT derives accordingly. The 2-D basis functions can be generated by multiplying the horizontally oriented 1-D basis functions (shown in Fig 2.3) with vertically oriented set of the same functions. The basic operation of the DCT is as follows:

1. The input image is \(N \times M\);
2. \(f(x,y)\) is the intensity of the pixel in row \(x\) and column \(y\);
3. \(C(u,v)\) is the DCT coefficient in row \(u\) and column \(v\) of the DCT matrix.
4. For most images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT.
5. Compression is achieved since the lower right values represent higher frequencies, and are often small - small enough to be neglected with little visible distortion.

The main advantage of DCT image transformation is the removal of redundancy between neighboring pixels. This leads to uncorrelated transform coefficients which can be encoded independently. Furthermore, DCT exhibits excellent energy compaction for highly correlated images and can pack input data into as few coefficients as possible. Other properties such as separability, symmetry or orthogonality are of high importance, as well. The reader may refer to [15, 16] for an insight in DCT.

**Wavelet Transform** For a detailed consideration of the Wavelet Transform process, the reader is advised to refer to Appendix, and can then skip the subsection. Wavelets are the result of collective efforts that recognized common threads between ideas and concepts that had been independently developed and investigated by distinct research communities. While wavelets have been traced all the way back to
1910 [17], for many, the starting point of their modern history coincides with two publications in the late 80s[18, 19]. Wavelets are oscillating, wave-like signals with finite duration and zero mean value and usually are non-symmetrical.

They are represented in a time-scale domain, where scales is somewhat related to the frequency notion of the Fourier Transform. Large scales correspond to the "big picture" of a signal (ex. the big basic shapes of an image), including contrast information, whereas small scales correspond to fine details such as resolution information. In wavelet analysis a fully scalable modulated window is shifted along the signal and we calculate the spectrum for every position. Then we repeat this process, using a longer or shorter version of the specified window. In the end the result would be a collection of time-frequency representations of the signal, all with different resolution. All these different window-signals are the wavelets. They are generated using scaling and translation (time shifting) onto a single basic wavelet function, called the mother wavelet, $\psi(t)$:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right)$$

with $s$ being the scale factor, and $\tau$ the translation factor. Energy normalization across the different scales is expressed with coefficient $s^{1/2}$. The Continuous Wavelet Transform (CWT) is written as in Eq. 6.8

$$C(s, \tau) = \int f(t)\psi^*_s(t)dt,$$  \hspace{1cm} (2.10)

where function $f(t)$ is decomposed into a set of basis functions (i.e. the wavelets), $\psi_{s,\tau}(t)$.

CWT signifies that it operates at every different scale up until an upper bound determined for a specific application. It also uses all possible shifts-translation across the full signal of interest. The steps to follow for a CWT are described below, and shown in Figure 6.2:

1. Take a wavelet and compare it to a section at the start of the original signal.
2. Calculate a number, C, that represents how closely correlated the wavelet is with this section of the signal. The higher C is, the more the similarity. More precisely, if the signal energy and the wavelet energy are equal to one, C may be interpreted as a correlation coefficient. Note that the results will depend on the shape of the wavelet you choose.
3. Shift the wavelet to the right and repeat steps 1 and 2 until you’ve covered the whole signal.
4. Scale (stretch) the wavelet and repeat steps 1 through 3.
5. Repeat steps 1 through 4 for all scales.

Similarities between parts of the signal and wavelets are indicated with the wavelet coefficients. For a given wavelet scale, the higher the coefficients, the more the resemblance with the signal.

![Figure 2.4: Graphical illustration of the general CWT procedure.](image)

Figure 2.4: Graphical illustration of the general CWT procedure. During steps 1-2 a correlation between the wavelet and a section of the signal is calculated. In step 3 the wavelet is shifted along the whole signal and correlation for every segment is calculated as in 1-2. In step 4, the wavelet is scaled and all previous steps are repeated.

However, the Wavelet transform in its continuous form is not practical. It is not possible to calculate the transform using continuous shifting (translation) and scaling. This is the reason that Discrete Wavelet Transform (DWT) has been introduced. DWT uses only a subset of the possible scales and positions, and it has been proven that using dyadic scales and positions (sampling the frequency and the time axis based on powers of two), the resulting wavelet analysis is as accurate as a continuous one, and of course, much more efficient. Note that the output of a DWT is a continuous signal, and the term discrete refers only to the scales and positions. Mallat, 1998 developed an efficient way to implement this scheme, with the use of filters, introducing a fast wavelet transform. Approximations and Details are the two main characteristics. Approximations refer to low-frequency components (i.e. high scales), wherein lies the most of a signal’s identity. The Approximations are the high frequency components (i.e. low scales).

In this DWT analysis, filters are used to create a distinction between the Approximation and the Detail coefficients and the original signal, S, passes through two complementary filters and emerges as two signals (see Figure 6.4). The filtering process will produce two times more coefficients, in total, than the original signal samples. It has been shown that downsampling the filtered signals by a factor of 2, we lose no information, and reduce the total number of produced coefficients down to the original signal length.
Figure 2.5: One stage Filtering on 1D signal. A low-pass filter is applied, and coefficients $c_A$ are produced after dyadic downsampling. In the same time, a high-pass filter is applied, and coefficients $c_D$ are produced after dyadic downsampling.

The high-pass filtered signal contains the smallest details we are interested in and we could stop there. However, the low-pass part still contains some details and therefore we can split it again and again, until we are satisfied with the number of bands we have created. In this way we have created an iterated filter bank.

In order to reconstruct the original signal using the Detail and Approximation coefficients, a reverse procedure must be followed: we need to apply appropriate reconstruction filters, after up-sampling the signal (lengthen the signal components by inserting zeros between samples). The choice of filters is crucial in achieving perfect reconstruction of the original signal. The decomposition and reconstruction filters are the tools that define the very form of the wavelet (Daubechies 1, 2, .., Symlets 1, 2, .. etc.) that analyzes our original signal. If we implement the wavelet transform as an iterated filter bank, we do not have to specify the wavelets explicitly.

The same idea may be also applied for a 2D signal. The standard solution consists of alternating one decomposition by rows and another one by columns, iterating only on the low-pass sub-image. An one-stage analysis outputs approximation coefficients $c_A$, together with three detail coefficient matrices $c_{D_h}$, $c_{D_v}$, and $c_{D_d}$ (horizontal, vertical, and diagonal, respectively). The process (further decomposition) may be repeated again and again, on the coefficients produced by low-pass filtering (see Fig. 6.6).
2.2. THEORY AND PRINCIPLES OF COMPRESSIVE SENSING

Figure 2.6: Multiple Level Decomposition using 2D DWT. Iteratively decompose rows and columns using high and low-pass filtering. Decomposition is done only on the low-pass sub-image of every step.

2.2.2 Sensing operator

A sensing or sampling operator is the modality which allows to retain only a subset of data when measuring a signal, instead of obtaining the whole amount of data. In the MRI context this means that we only keep the subset of the $k$-space data: the data actually acquired by the MR scanner. In a mathematical formulation, the signal of interest is sampled via the $M \times N$ sensing matrix $\Phi$. This means that $M$ linear recordings of the original signal $m$ and we obtain signal $y = \Phi m$. If, further, the signal of interest $m$ can be represented by a sparse signal $s$, when transformed in the $\Psi$ domain (see also Fig. 2.7), then the measurement vector is obtained as:

$$y = \Phi m = \Phi \Psi s$$  (2.11)
CHAPTER 2. COMPRESSION SENSING (CS) IN MR IMAGING

If for example, the rows of $\Phi$ are Dirac delta functions, then the acquired signal would be sampled values of $m$. We are interested in the cases where $M \ll N$, which means that the measurements constitute only a small fraction of the original samples. In MRI the sensing waveforms are sinusoids, as described in Section 1. This modality makes the measured signal $y$ a vector of Fourier coefficients.

In MRI we do not have knowledge of the $\mathbb{R}$-space signal (i.e. the image) itself. Consider the $\mathbb{R}$-space 2-D image $m_{2D}$, $n_x \times n_y$. We can obtain information for $m_{2D}$, sampling its $k$-space representation or in other words, sample the Fourier coefficients of the image. Consider

$$\Omega_{C_x} : \text{the } n_x \times n_x \text{ complete valued Fourier matrix}$$

$$\Omega_{C_y} : \text{the } n_y \times n_y \text{ complete valued Fourier matrix}$$

The k-space data of image $m_{2D}$, according to previous notation would be the Fourier transform of the image:

$$k - space_{2D} = \Omega_{C_y}m_{2D}\Omega_{C_x}^T \Rightarrow$$

(2.12)

The pre and post multiplication with the Fourier matrices of Eq. 2.12 may be replaced via the Kronecker product property with $\Phi_F$. Reader may refer to Appendix for a more detailed derivation. So, our sensing operator is $\Phi_F$. If we sample the k-space with $Phi_F$, then the acquired measurements would be the whole data grid.

$$\text{vector}(y) = \Omega_{C_y} \otimes \Omega_{C_x} \text{vector}(m) \Rightarrow$$

(2.13)

$$\text{vector}(y) = \Phi_F \text{vector}(m)$$

where $\text{vector}(.)$ denotes the vectorization operator that stacks the columns of the argument signal.

$\Phi_F$ is of size $(n_x \times n_y) \times (n_x \times n_y)$. Each row of $\Phi_F$ when multiplied with vector($m$) results in a single $K$-space measurement, and thus, a single $y_i$ coefficient. In order to represent the incomplete measurements of $K$-space, rows of
2.2. THEORY AND PRINCIPLES OF COMPRESSIVE SENSING

Φ are removed. These removed rows correspond to the K-space data that will not be measured. In case of M measurements, (N − M) k-space samples are not measured. This being the case, the complete measurement matrix Φ_F reduced to matrix Φ_k of size (M × (n_x × n_y)), and the k-space measurements are of size (M × 1). The measurements taken are predetermined by the sampling patterns as discussed in section 1.3. If for example we follow an equispaced Cartesian sampling pattern, then equispaced data are left out of sampling, and equispaced rows are left out of the measurement matrix Φ_k. However, this equispaced sampling is not preferable from a Compressive-Sensing point of view: a notion of randomness needs to be introduced in the sensing operator, for maximum efficiency. The reason will be better described in the following sections.

2.2.3 The restricted isometry property (RIP)

The restricted isometry property is a fundamental concern of Compressive Sensing. In simple terms the RIP needs to be satisfied by the sensing operator Φ, so that the operator has similar effects on all k-sparse vectors. More precisely, if a_k represents the largest possible coefficient and b_k the smallest coefficient, then Eq. 2.14 needs to be satisfied for all k-sparse vectors x ∈ ℜ^N, in the case where values of a_k and b_k are similar. The restricted isometry constant δ_k = b_k − a_k / a_k + b_k of Eq. 2.14, can be a means of performance guarantee for Eq.2.3. Small values for the normalized difference of b_k − a_k, can guarantee a better performance.

a_k∥x∥^2 ≤ ∥Φx∥^2 ≤ b_k∥x∥^2 (2.14)

However, the current CS theory is not able to offer a feasible way of computing the isometry constants for a matrix. It can be proved, though, that a randomly generated matrix Φ has a high probability of having small isometry constant[11], [22], [13]. A more concrete metric that could be seen as a loosen alternative to the restricted isometry property is the mutual coherence between base Φ that senses the signal, and base Ψ that represents the signal. The encoding waveform should be incoherent with the sparsifying basis. Mutual coherence measures the largest correlation between any two elements of Φ and Ψ, where correlated elements have large coherence values. For stable and robust recovery low coherence is most desired. The mutual coherence of the bases Φ ∈ ℜ^{n_x × n_y} and Ψ ∈ ℜ^{n_y} is the maximum absolute value for the inner product between elements of the two bases[4]:

µ(Φ, Ψ) = √N \max_{1 \leq k,j \leq n} |⟨φ_k, ψ_j⟩| (2.15)

In the case were no transform bases is used, then mutual coherence measures the coherence among rows of the measurement matrix. µ can also been viewed as a way to determine the number of the necessary measurements. It is easy to see that the range of possible accepted coherence values  ∈ µ(Φ, Ψ) is [1, N^{1/2}]. A suitable range for the number of measurements would be in [O(K log(N)) to O(N)]. It has
been shown [23] that if a signal is $k$-sparse, after projection onto $\Psi$, then $M$ random measurements on the $\Phi$ domain are enough for Eq. 2.3 to have an exact solution, when

$$M \geq c \cdot \mu^2(\Phi, \Psi) \cdot K \cdot \log N$$

(2.16)

for a small constant $c$. A different expression of the number of measurements can be found in [6]. In case of iid Gaussian measurements, $\ell_1$ norm minimization can exactly recover $K$-sparse signals and closely approximate compressible signals with high probability, using only $M = cK \log(N/K)$ measurements.

2.2.4 Randomness: Why is it so important?

Incoherent aliasing interference in the sparse transform domain is an essential ingredient for CS. A very high degree of incoherence is guaranteed when a random sampling process is used. To get intuition for the importance of incoherence and the feasibility of CS in MRI, consider the following simple example of Fig 2.8. A sparse 1D signal (Fig. 2.8a), 256 samples long, is undersampled in k-space (Fig. 2.8b) by a factor of eight. Here, the sparse transform is simply the identity. Equispaced k-space undersampling and reconstruction by zero-filling results in coherent aliasing, a superposition of shifted replicas of the signal as illustrated in Fig. 2.8 c. In this case, there is an inherent ambiguity; it is not possible to distinguish between the original signal and its replicas, as they are all equally likely.
2.2. THEORY AND PRINCIPLES OF COMPRESSION SENSING

Figure 2.8: An intuitive reconstruction of a sparse signal from pseudo-random k-space undersampling. A sparse signal (a) is 8-fold undersampled in k-space (b). Equispaced undersampling results in coherent signal aliasing (c) that cannot be recovered. Pseudo-random undersampling results in incoherent aliasing (c). A recovery process could include detection of the strong components, and substraction of their interference in a repeated way.

Random undersampling results in a very different situation. The zero-filling Fourier reconstruction exhibits incoherent artifacts that actually behave much like additive random noise (Fig. 2.8d). Despite appearances, the artifacts are not noise; rather, undersampling causes leakage of energy away from each individual nonzero coefficient of the original signal. This energy appears in other reconstructed signal coefficients, including those which had been zero in the original signal. It is possible, if all the underlying original signal coefficients are known, to calculate this leakage analytically. This observation enables the signal in Fig. 2.8d to be accurately recovered although it was 8-fold undersampled. An intuitive plausible recovery procedure could the following: Based on thresholding, detect and recover the strong components, calculate the interference caused by them and subtract it. Subtracting the interference of the strong components reduces the total interference level and enables recovery of weaker, previously submerged components. By iteratively repeating this procedure, one can recover the rest of the signal components.

In the original CS papers [5, 6], sampling a completely random subset of k-space was chosen to simplify the mathematical proofs and in particular to guarantee a very high degree of incoherence. Compressive sampling is mainly concerned with
low coherence pairs, and we now give examples of such pairs. As a first example, let \( \Phi \) be the canonical or spike basis \( \phi_k(t) = \delta(t - k) \) and \( \Psi \) be the Fourier basis, \( \phi_j(t) = N^{-1/2} \exp(i2\pi jt/n) \). Since \( \Phi \) is the sensing matrix, this corresponds to the classical sampling scheme in time or space. The time-frequency pair obeys \( \mu(\Phi, \Psi) = 1 \) and, therefore, we have maximal incoherence. Further, spikes and sinusoids are maximally incoherent not just in one dimension but in any dimension, (in two dimensions, three dimensions, etc.).

A second example takes wavelets bases for \( \Psi \) and noiselets [24] for \( \Phi \). The coherence between noiselets and Haar wavelets (see Appendix) is \( \sqrt{2} \) and that between noiselets and Daubechies \( D4 \) and \( D8 \) (see Appendix) wavelets is, respectively, about 2.2 and 2.9 across a wide range of sample sizes \( N \). This extends to higher dimensions as well. (Noiselets are also maximally incoherent with spikes and incoherent with the Fourier basis.) Our interest in noiselets comes from the fact that 1) they are incoherent with systems providing sparse representations of image data and other types of data, and 2) they come with very fast algorithms; the noiselet transform runs in \( O(N) \) time, and just like the Fourier transform, the noiselet matrix does not need to be stored to be applied to a vector. This is of crucial practical importance for numerically efficient CS implementations.

Finally, random matrices are largely incoherent with any fixed basis \( \Psi \). Select an orthobasis \( \Phi \) uniformly at random, which can be done by orthonormalizing \( N \) vectors sampled independently and uniformly on the unit sphere. Then with high probability, the coherence between \( \Phi \) and \( \Psi \) is about \( \sqrt{2 \log N} \). By extension, random waveforms \( (\phi_k(t)) \) with independent identically distributed (i.i.d.) entries, e.g., Gaussian or \( \pm 1 \) binary entries, will also exhibit a very low coherence with any fixed representation \( \Psi \). Note the rather strange implication here; if sensing with incoherent systems is good, then efficient mechanisms ought to acquire correlations with random waveforms, e.g., white noise!

### 2.2.5 Signal recovery

Having acquired \( M \) measurements of the original signal via a sensing mechanism, indicated by \( \Psi \):

\[
y = \Phi \cdot m, \quad \text{or} \quad y = \Phi \cdot \Psi \cdot s
\]  

(2.17)

the original signal can be recovered by \( \ell_1 \)-norm minimization. It is known that signal \( m \) is sparse in the domain indicated by \( \Psi \), and one of the main enablers of CS was the discovery that when the signal being observed is sparse enough, it can be exactly recovered by solving the linear program:

\[
\min_{\hat{s} \in \mathbb{R}^N} \|\hat{s}\|_{\ell_1}, \quad \text{s.t.} \quad \hat{y} = \Phi \cdot \Psi \cdot \hat{s}
\]  

(2.18)
where $\|\cdot\|_{\ell_1}$ denotes the $\ell_1$ norm, which is equal to the sum of the absolute values of the vector entries ($\|\mathbf{s}\|_{\ell_1} = \sum_i |\mathbf{s}_i|$). In words, we seek for a signal $\hat{x}$ among all possible ones ($\mathbf{m} = \Psi \cdot \hat{s}$), consistent to $\mathbf{y} = \Phi \cdot (\Psi \cdot \hat{s})$, and whose sparse representation has minimal $\ell_1$-norm. Minimizing $\ell_1$, subject to linear equality constraints is a convex optimization problem but can easily be recast as a linear program, making available a host of ever more efficient solution algorithms. Linear programming methods may be based either on the classical simplex method of linear programming or the more recent interior point methods [25]. In the case of recovering a signal from noisy data, the same framework as in Eq. 2.18 can be proposed, with relaxed reconstruction constraints:

$$\min_{\hat{s} \in \mathbb{R}^N} \|\hat{s}\|_{\ell_1}, \quad s.t.: \quad \|\Phi \cdot \Psi \cdot \hat{s} - \mathbf{y}\| \leq \epsilon$$

(2.19)

where $\epsilon$ indicates the level of accepted noise level or reconstruction error. The problem of Eq.2.19 is also convex (second-order cone program), and can be solved efficiently [26].

Why minimize via $\ell_1$ and not use $\ell_0$ or $\ell_2$: The general goal of CS is to find a highly sparse decomposition- one with very few nonzero terms. This naturally leads to the optimization problem

$$\min_{\hat{s} \in \mathbb{R}^N} \|\hat{s}\|_{\ell_0}, \quad s.t.: \quad \mathbf{y} = \Phi \cdot \Psi \cdot \hat{s}$$

(2.20)

where $\|\hat{s}\|_{\ell_0}$ indicates the number of nonzero elements in $\mathbf{s}$ and is obviously a sparsity measure. Unfortunately, in general, this problem requires an exhaustive search for all possible locations of the non zero locations (NP-complete) and is also numerically unstable.

If we tried to replace the $\ell_1$ norm with the $\ell_2$ norm, this would make the recovery a least squares problem:

$$\min_{\hat{s} \in \mathbb{R}^n} \|\hat{s}\|_{\ell_2}, \quad s.t.: \quad \mathbf{y} = \Phi \cdot \Psi \cdot \hat{s}$$

(2.21)

However, the $\ell_2$ norm cannot guarantee that the solution of Eq. 2.21 would be sparse and in most cases, solving Eq. 2.21 results in a non-sparse solution with many nonzero elements, as $\ell_2$ measures signal energy and not sparsity. In Fig. 2.9 one may see in (a) the area where a sparse $\mathbb{R}^3$ signal should lie in: close to the coordinate axes. In (b) and (c) the corresponding null spaces of $\ell_2$ and $\ell_1$ minimization. The geometrical shape of the $\ell_1$ null space obviously enforces sparse solutions.
Figure 2.9: (a) The subspaces containing two sparse vectors in $\mathbb{R}^3$ lie close to the coordinate axes. (b) Visualization of the $\ell_2$ minimization that finds the non-sparse point-of-contact $s$ between the $\ell_2$ ball (hyper-sphere, in red) and the translated measurement matrix null space (in green). (c) Visualization of the $\ell_1$ minimization solution that finds the sparse point-of-contact $s$ with high probability thanks to the pointiness of the $\ell_1$ ball. [12]

All in all, optimization based on $\ell_1$ norm enforces sparseness and can exactly recover $k$-sparse signals or approximately reconstruct compressible signals with high probability using only $M \ll N$ measurements, with computational complexity of order $O(N^3)$ [6].

### 2.3 History and Highlights of CS in MR Imaging

Widespread application of the CS principles in Magnetic Resonance Imaging is handicapped by the fact that general anatomical MR images are not necessarily sparse in the image domain directly. One exception relates to angiography, in which image content is typically sparse. In order apply CS to images, which are only sparse in a transform domain, such as the wavelet or the finite differences domain, some prior knowledge about the object is needed so as to choose the right sparsifying transformation, although the discrete cosine transform or the wavelet transform used for JPEG and the JPEG2000 compression seem to be robust sparsifying transformations for many types of images.

Initial work on applying CS onto MRI principles started extensively in 2004. Since theoretical work justified the application of the CS observations, into magnetic resonance imaging concepts, in [5], Candes, Romberg and Tao showed that exact signal recovery may be obtained using only a randomly chosen subset of signal data in the frequency domain, by solving a convex optimization problem (via $\ell_1$ minimization). In effect, authors state that any signal made out of $|T|$ spikes may be recovered by convex programming from almost every set of frequencies of size $O(|T| \cdot \log N)$. Moreover, this is nearly optimal in the sense that any method succeeding with probability $1 - O(N - M)$ would in general require a number of frequency samples at least proportional to $|T| \cdot \log N$. 
Constraints and difficulties of applying CS into such a different field, that of the MRI, where thoroughly traced, recorded and finally surpassed by the research community. Stanford University in California highly contributed in the field, and specifically Michael Lustig, working with D. Donoho, J.M. Pauly and his team. They started in 2004 a series of experiments and publications [32, 33] and continued their extensive work, when in [35, 36] detailed requirements for successful CS were reviewed, describing the natural fit to MRI, by presenting and discussing explicit applications, and thus offering a great starting point for all research communities in the area.

In [34], Lustig et al. demonstrated a 5-fold scan time reduction in MR angiography, by randomly undersampling the phase encodes in a 3D Cartesian scan, as an alternative to the impractical (due to current MRI hardware restriction) case of 2D random sampling.

A few years later, in 2007 Seoul Korea, Jung et al in [53] showed activity in the field of CS MRI, where they employed the CS theory for dynamic MR imaging of time-varying objects to overcome drawbacks of classical approaches using model-based reconstruction methods. They continued with several publications [54, 55, 56, 57] working in the same direction.

New algorithms for MRI reconstruction via $\ell_1$ minimization that were much faster than their predecessors have been developed by Ma et al. [58] in 2008 replacing iterative linear solvers with Fourier domain computations, with substantial time savings. The authors proposed an efficient algorithm that jointly minimizes the $\ell_1$ norm, total variation, and a least squares measure. Their algorithm is based upon an iterative operator-splitting framework, taking advantage of fast wavelet and Fourier transforms. We show that faithful MR images can be reconstructed from a random sampling subset that represents a mere 20 percent of the complete set of measurements.

In 2008, Matthias Seeger and his team of Tübingen Germany [29] employ a Bayesian experimental design to CS theory for recovering (not necessarily) sparse signals. An expectation propagation method is used where the notion of compressibility is encoded in a prior distribution under which signals of low complexity or high transform sparsity have most mass. The authors propose an optimization of the measurement structure so as to recover a signal that is sufficiently likely under the prior, even if the Nyquist-Shannon theorem is violated. They demonstrate the superiority over widely-known CS algorithms. Their work is further continued in [28] where they optimize designment of magnetic resonance imaging sequences through Bayesian scores.

In Los Alamos National Laboratory, USA, Rick Chartrand in 2009 [31] employs
a CS reconstruction method in the context of MRI, by replacing the $\ell_1$ norm of the standard CS formulation problem with the $\ell_p$ norm where $p < 1$, using an iterative algorithm to minimize the concave objective $\|m\|_{\ell_p}$. The algorithm alternates between gradient descent and projection onto the constraint set $y = \Phi m$. Using the algorithm when $p = -1/2$ they managed to achieve a perfect reconstruction using only 9 radial lines in a radial sampling pattern.

Very interesting and recent activity in Xiamen University of China, by Xiaobo Qu and his team [27] presents a new framework, based on smoothed $\ell_0$ norm minimization and the principle of basis pursuit. The authors combine different sparsifying transforms where each transform can efficiently represent specific feature that the other can not. The idea is applied onto overcomplete sparse decomposition for compressed sensing MRI. Great quality improvement is demonstrated even for 15% undersampled measurements of $k$-space, using a random variable density sampling mask.

The field of Compressive Sensing in Magnetic Resonance Imaging is very active. The research communities all over the world are trying their best in this new hot area. It has been only about 6 years from the moment that researchers starting thoroughly to work on CS MRI and experiments and first results have been proved very promising.
Chapter 3

Exhaustive Comparison of three Algorithms

Many different frameworks have been proposed in the literature to solve the CS recovery problem on undersampled signals. The locations of the non zero signal coefficients are not known, nor their amplitudes. Examples of L0-minimization algorithms include the classical Matching Pursuit (MP) [37], Orthogonal Matching Pursuit (OMP) [38], stagewise OMP (StOMP) [39] regularized OMP (ROMP) [40], subspace pursuits [41], CoSaMP [42], SAMP [43] and iterative hard thresholding[44]. The L0-minimization techniques are most promising because they are often based on the greedy approaches that find the solution quickly. Traditional techniques such as MP or OMP, mentioned above, add one variable at a time to the solution leading to a runtime of $O(MNS)$, where $M$ is the number of measurements, $N$ is the signal length and $S$ the sparsity. However, newer techniques such as StOMP, ROMP, subspace-pursuit and IHTs add multiple variables to the solution and fall in the class of near linear-time algorithms with a runtime of $O(MN\text{poly}(\log(MN)))$. Most of these techniques also have conditions on RIP constants $\delta$ under which the exact recovery is guaranteed.

The L1-minimization algorithms include the classical basis pursuit [45] algorithm, the Lasso-modification to LARS [46], random projections onto convex sets [48], homotopy [47], weighted least squares [53], iterative algorithms based on gradient thresholding and other gradient based approaches [35], [58]. The L1-minimization techniques offer tight guarantees on recovery [49] provided the algorithm converges to optimal solution to the problem. These techniques are generally applicable to a broader class of problems of optimizing a convex objective function with L1-penalty. For these techniques, the exact bound on the runtime are either very large or not available.

In the following section three published CS algorithms are presented. My extensive research on the parameters of every algorithm will reveal the way they affect the reconstruction process. The algorithms presented are chosen either for their simplicity ($SL0$, [50]), or for their wide use ($\ell_1$ magic [51]) and innovation (SparseMRI [52]).
3.1 Smoothed $\ell$-0 Algorithm

The reader may refer to "Smoothed $\ell$-0 Algorithm, Jutten C., Mohimani H., Babaie-Zadeh M., 2007". This algorithm has only been introduced in the literature in a general signal acquisition setup. After adapting the algorithm in MR schemes a series of experiments were conducted to better assess and evaluate the reconstruction method, in this thesis.

This algorithm minimizes an $\ell_0$ norm of sparse vector $s$, while replacing the $\ell_0$ norm function by a smooth estimate of the Gaussian family to overcome discontinuities of the $\ell_0$ norm during differentiation. The great achievement of the proposed algorithm is the very low computational cost. The algorithm has not been yet applied on concepts of Magnetic Resonance Imaging (MRI) but the fast convergence triggered our curiosity to be tested on a MR application. Here, the CS aim is to reconstruct images by taking incomplete measurements with the MR scanner. The measurements are acquired in the $K$-space (Fourier domain) while the signal to be reconstructed, is in the $R$-space. Reminder of notation:

- $m : N \times 1$, image of interest ($R$-space).
- $\Phi_k : M \times N$, undersampled Fourier operator.
- $y : M \times 1$, $K$-space measurements.
- $\Psi : N \times N$, linear operator that transforms from pixel representation into a sparse representation.

Problem to solve:

$$\text{minimize} \| \Psi^{-1}m \|_0 \quad \text{s.t.} \quad \Phi_km = y$$

It has been previously stated (Section 2.2.5) that searching the minimum $\ell_0$ norm is an intractable problem as the dimension increases.

In order to avoid great computational complexity, the authors stick to $\ell_0$ minimization and try to overcome the non-differentiation case of $\ell_0$ norm. To simplify the equations, we assume that $m$ image is already sparse, and no sparsifying transform is used. Thus, we formulate the minimization problem as $\|m\|_{\ell_0}$, instead of writing $\|\Psi m\|_{\ell_0}$.

The $\ell_0$ norm is the number of non-zero components in $m$:

$$\|m\|_{\ell_0} = \sum_{i=1}^{N} m_i^0$$

(3.1)
3.1. SMOOTHED $\ell$-0 ALGORITHM

For convenience Eq. (3.1) may be rewritten as:

$$\|m\|_{\ell_0} = \sum_{i=1}^{N} \nu(m_i),$$

(3.2)

where $\nu(m_i) = \begin{cases} 1, & m_i \neq 0 \\ 0, & m_i = 0 \end{cases}$

The discontinuities of $\ell_0$ norm can be clearly understood when looking at Eq. (3.2), urging for a smoother estimation of $\nu$. In order to overcome the problem, the authors invoke function $f(m)$ of the zero-mean Gaussian family because of its differentiability:

$$f_\sigma(m) = e^{-m^2/2\sigma^2}, \quad \text{where}$$

$$\lim_{\sigma \to 0} f_\sigma(m) = \begin{cases} 1, & m_i = 0 \\ 0, & m_i \neq 0 = 1 - \nu(m) \end{cases}$$

(3.3)

Now if we define function $F_\sigma(m)$ as:

$$F_\sigma(m) = \sum_{i=1}^{N} f_\sigma(m) \Rightarrow$$

$$\lim_{\sigma \to 0} F_\sigma(m) = \sum_{i=1}^{N} (1 - \nu(m_i)) = N - \|m\|_{\ell_0} \Rightarrow$$

$$\|m\|_{\ell_0} \approx N - F_\sigma(m)$$

(3.4)

It is known that $\sigma$ controls the width of a Gaussian ”bell curve”, so in our case it is a tradeoff between accuracy (small $\sigma$) and smoothness (large $\sigma$) of the approximation. Minimizing the $\ell_0$ norm becomes equivalent to maximizing function $F_\sigma$ for sufficient small $\sigma$. A small $\sigma$ value will result in a lot of local maxima for $F_\sigma$, and as $\sigma$ grows, the function becomes smoother. It is proven by the authors that there is a sufficient large value for $\sigma$ for which there is no local maxima in $F_\sigma$. The authors propose a way to avoid local maxima: gradually decrease the $\sigma$ value and use a steepest ascent algorithm for maximizing $F_\sigma$ for every $\sigma$ value. The initial value of the current steepest ascent run is the maximizer of $F_\sigma$ obtained for the previous (larger) value of $\sigma$. Since the value of $\sigma$ changes slowly, the steepest ascent algorithm is initialized, before every step, not far from the actual maximum, and consequently, the authors hope that it would not be trapped in the local maxima.
3.1.1 The Algorithm Outline

<table>
<thead>
<tr>
<th>The SL0 Algorithm Outline</th>
</tr>
</thead>
</table>

**Input:** $y, \Phi_k$

**Initialization:**
1. $m = \Phi_k^T(\Phi_k\Phi_k^T)^{-1}y$ (i.e. min $\ell_2$-norm)
2. $\sigma = [\sigma_1, .., \sigma_{\text{min}}]$, where $\sigma_1 = \sigma_{\text{max}}$ and $\sigma_k = \sigma_{k-1} \cdot DF_{\sigma}$, for $k = 1, \ldots, K$:
   1. Let $\sigma = \sigma_k$.
   2. Steepest Ascent (maximize $F_\sigma$)
      - for $j = 1, \ldots, L$:
         (a) Let: $m \leftarrow m + \mu_0 \cdot \sigma_k^2 \nabla F_\sigma(m)$
         (b) Project $m$ back onto the feasible set:
            $m \leftarrow m - \Phi_k^T(\Phi_k\Phi_k^T)^{-1}(\Phi_km - y)$

**Output:** $m$

The step size $\mu_0 \cdot \sigma_k^2$ should be decreasing.

<table>
<thead>
<tr>
<th><strong>Params</strong></th>
<th><strong>Description</strong></th>
<th><strong>High values</strong></th>
<th><strong>Low values</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Steepest Ascent iterations</td>
<td>Local maxima</td>
<td>Increased speed</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Variance sequence ($\sigma^2$) of the model</td>
<td>Fluctuating $F_\sigma$ (local maxima)</td>
<td>Smoother function. Rough $\ell_0$ approx.</td>
</tr>
<tr>
<td>$DF_\sigma$</td>
<td>Defines the $\sigma$ sequence</td>
<td>For low sparseness</td>
<td>Less steps on $\sigma$</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>Scales the step-size sequence</td>
<td>Escaping local maxima.</td>
<td>Affects convergence.</td>
</tr>
</tbody>
</table>

The internal loop of the algorithm (steepest ascent for a fixed $\sigma$) is repeated a fixed and small number of times ($L$). In other words, for increasing the speed, we do not wait for the (internal loop of the) steepest ascent algorithm to converge. This may be justified by gradual decrease in value of $\sigma$, and the fact that for each value, we do not need the exact maximizer of $F(\sigma)$. All we need, is to enter a region near the (absolute) maximizer of $F(\sigma)$ for escaping from its local maximizers.
3.1. SMOOTHED $\ell_0$ ALGORITHM

No stopping criteria is used other than the number $L$ of iterations of each Steepest Ascent run. In an effort to avoid being trapped in local maxima, a high $L$ value is not recommended. Steepest ascent algorithm consists of iterations of the form

$$m \leftarrow m + \mu_k \nabla F_\sigma(m)$$

(3.5)

where the step-size $\mu_k$ should be decreasing, meaning that for smaller $\sigma$ values we should apply smaller $\mu_k$ values, because function $F_\sigma$ is more fluctuating when $\sigma$ is small, meaning that the step $\mu_k$ should be proportional to $\sigma_k^2$, and, thus, the authors propose $\mu_k = \mu \sigma_k^2$.

Any arbitrary solution $m$ of $\Phi_km = y$ could be given as a starting point to the algorithm. However, the best value for the initial estimation of the sparse solution is the minimum $\ell_2$ norm solution of $\Phi_km = y$, which is given by the pseudo-inverse of $\Phi_k : m = \Phi_k^T(\Phi_k \Phi_k^T)^{-1}y$

3.1.2 Experiments

In an effort to observe the behavior of the algorithm and the way that the reconstruction process is affected by the parameters of the algorithm, a series of experiments have been implemented, within an ultimate goal: to propose the best parameter set for MR related modalities.

The famous Shepp-Logan phantom image had been used for the experiments of all algorithms, see Fig. 3.1. This specific phantom image is widely used on experiments among the research community. It has a sparsity factor of 50.93% and closely resembles the features of real MR images. The sampling process of the raw MR data - the $k$-space data- is simulated by the use of a radial trajectory. That is, we 'visit' the $k$-space frequency locations along a radial-line path. MR hardware permits the design of a radial-like trajectory. Furthermore, this kind of sampling may well introduce the desirable "random" notion on the sampling operator. Recall that the central area of $k$-space contains the large values of the MR signal, whereas the energy of the MR signal is low in the peripheral regions- wherein the effective resolution information lies. This specific sampling pattern also provides denser sampling in the center, where needed, and sparser sampling at the high spatial frequencies.
CHAPTER 3. EXHAUSTIVE COMPARISON OF THREE ALGORITHMS

(a) Shepp-Logan phantom image 128x128: the original image, to be reconstructed from undersampled k-space data

(b) Radial sampling pattern of \( k \)-space, surpasses hardware constraints of purely random sampling and yet introduce randomness.

Figure 3.1

Remember, radial lines correspond to the data samples we measure. More lines means more samples. We experimented on different coverage percentages using 6 to 105 radial lines. In Table 3.1 one may see the correspondence of radial lines and sampling coverage, for the specified 128x128 image size.

<table>
<thead>
<tr>
<th>Radial Lines</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>105</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Samples</td>
<td>4.59</td>
<td>7.57</td>
<td>14.89</td>
<td>21.83</td>
<td>28.52</td>
<td>34.79</td>
<td>40.82</td>
<td>46.44</td>
<td>51.86</td>
<td>56.81</td>
<td>63.82</td>
</tr>
</tbody>
</table>

Table 3.1: Different number of radial lines corresponds to different sampling percentage of the \( K \)-space. Note that the coverage percentage is related to the specific image size. The image size, here, is 128 x 128.

In Table 3.2 one may see the values that were used for each input parameter of the \( SL0 \) algorithm. The upper bound of the \( \sigma \) sequence, \( (\sigma_{\text{max}}) \), is calculated automatically by the algorithm, as the signal energy (two times the maximum absolute value of the input signal) \( m \). The lower bound, \( \sigma_{\text{min}} \), should depend on the estimated noise level of the source and on the desired accuracy. Noise may also refer to the energy of the inactive elements of \( m \). The noiseless case refer to the inactive elements of \( m \) being exactly equal to zero, and so the smaller the \( \sigma_k \) value, the better the sparsest solution estimation. Hence, its value may be determined with respect to the desired accuracy.
3.1. SMOOTHED $\ell$-0 ALGORITHM

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decrease factor of $\sigma$ ($DF_\sigma$)</td>
<td>0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9</td>
</tr>
<tr>
<td>$\sigma$ sequence $\sigma$</td>
<td>$[\sigma_{max}, \ldots, \sigma_{min}]$, where $\sigma_K : {10E - 6, 5E - 4, 10E - 3, 5E - 3, 10E - 2, 5E - 2, 0.1}$</td>
</tr>
<tr>
<td>$\mu_0$ value</td>
<td>0.1, 0.5, 1, 1.5, 2, 2.5, 3, 3.5</td>
</tr>
<tr>
<td>$L$ number of iterations</td>
<td>3, 10, 20</td>
</tr>
</tbody>
</table>

Table 3.2: The parameter values used to study and experiment on the reconstruction process of the algorithm.

Reconstruction plots based on the Mean Square Error are presented in the following section.

3.1.3 Results and Findings

We start the analysis with the lower bound value of the $\sigma$ sequence ($\sigma_{min}$). All reconstruction results reveal no particular effect of this parameter. The algorithm reaches the convergence point fast, and an early stop on the $\sigma$ loop does not affect the maximization process because the local-hopefully-global maximum is achieved during the first $\sigma$ iterations. As a result, the $\sigma_{min}$ may be fixed. Considering also the fact that we have no additive noise in the simulations, I chose to fix $\sigma_{min} = 1E - 05$.

The number of iterations $L$, may greatly affect the reconstruction results. Reconstruction error plots, can help decide the optimal value. The mean reconstruction error is calculated for every different $L$ value, over all other parameters (all different $DF_\sigma$ and $\mu_0$ values). The plot below shows the findings, for the phantom image $128 \times 128$, when the sampling percentages vary, i.e. different number for radial lines are used in the sampling trajectory. One may see that no matter the number of radial lines used for the sampling pattern, the MSE is the lowest when only 3 iterations are used in the algorithm. According to these plots, it is clear that the best choice would be to keep $L = 3$. Figure justifies this choice, presenting the maximization of the cost function when 20 iterations are used.
CHAPTER 3. EXHAUSTIVE COMPARISON OF THREE ALGORITHMS

Figure 3.2: On the x axis: number of iterations $L$. On the y axis: Mean Square Error of reconstruction using undersampled measurements. Each line in the plot represents the reconstruction results for each number of Radial Lines used. Every plot line is the mean error over all different $DF_\sigma$ and $\mu_0$ values for the specific Radial Line number.

To conclude, we propose the use of $L = 3$ and $L = 10$. There is no need for extra computational power using $L = 20$. In fact, many "trapped in local maxima" cases were observed when iteration number was set to $L = 20$. An illustrative example is shown in figures 3.3 - 3.4. In Figure 3.3 one may see that the algorithm is trapped in local maxima for the specific $\sigma_i$ iteration, when $L = 20$. In Figure 3.4, the same steps are shown, but this time the number of iterations is $L = 3$. 
3.1. SMOOTHED $\ell$-0 ALGORITHM

Figure 3.3: A screenshot of the algorithm when $L = 20$. Top Left: the $F_\sigma$ cost function. Every line corresponds to a different $\sigma_k$ loop for every $L$-loop iteration. Loop "$k = 1$" is illustrated with the blue line, loop "$k = 2$" with the green line and loop "$k = 3$" with the black line. A steepest ascent loop is performed for every $\sigma_k$ loop. Top Right: the MSE of the reconstructed image, during every step of the algorithm. Bottom: The corresponding reconstructed images at the last step of each $\sigma_k$ loop (each colored line). The algorithm is obviously trapped.
CHAPTER 3. EXHAUSTIVE COMPARISON OF THREE ALGORITHMS

Figure 3.4: A screenshot of the algorithm when $L = 3$. Top Left: the $F_\sigma$ cost function. Every line corresponds to a different $\sigma_k$ loop for every $L$-loop iteration. Loop "$k = 1$" is illustrated with the blue line, loop "$k = 2$" with the green line and loop "$k = 3$" with the black line. A steepest ascent loop is performed for every $\sigma_k$ loop. Top Right: the MSE of the reconstructed image, during every step of the algorithm. Bottom: The corresponding reconstructed images at the last step of each $\sigma_k$ loop (each colored line).

Judging from the previous plots, we consider the best choice to fix the inner iterations to $L = 3$. Now, let us see the behavior of the algorithm, as affected by the $DF_\sigma$ value. Below one may see the corresponding plots for the $MSE$ and the $DF_\sigma$ values, where $\sigma_{\text{min}}$ is fixed to $1E-05$ and iterations $L$ are fixed to 3. The mean plots over all different $\mu_0$ values are shown. They are split for Radial Lines $\in [6, ..., 30]$ and $\in [35, ..., 105]$ for a better view. Looking at the reconstruction results a good range of values for parameter $DF_\sigma$ can be $[0.05, 0.1, 0.3, 0.5]$. 
3.1. SMOOTHED $\ell$-0 ALGORITHM

Figure 3.5: On the x axis: the decrease factor of $\sigma$, $DF_\sigma$. On the y axis: Mean Square Error of reconstruction using undersampled measurements. Each line in the plot represents the reconstruction results for each number of Radial Lines used. Every plot line is the mean error over all different $m_0$ values when $\sigma_{min}$ and $L$ are fixed. (a) Results for Radial Lines $\in [6,..30]$, i.e. sampling percentage of $[4.5,..21.8]$% and (b) results when Radial Lines $\in [35,..,105]$, i.e. sampling percentage of $[25.2, ..., 63.8]$%.

We cannot clearly say that among the proposed $DF_\sigma$ values, any one outperforms the others. For this reason, we choose not to fix $DF_\sigma$ to a specific value, but rather fix it in the range of $[0.05, 0.1, 0.3, 0.5]$ and continue with the $\mu_0$ parameter. The effect of the $\mu_0$ parameter on the algorithm using the fixed parameters $L = 3$, $\sigma_{min} = 1E-05$, and the $DF_\sigma$ ranging in $[0.01, 0.05, 0.1, 0.3, 0.5]$ is shown in plot 3.6. One may see the mean MSE, over the different $DF_\sigma$ values, for all different Radial Lines. The plot is split again to lower and higher sampling and $\mu_0$ ranging in $[0.1, 0.5, 1, 1.3, 2]$ seems a suitable choice.
Figure 3.6: On the x axis: step size parameter $\mu_0$. On the y axis: Mean Square Error of reconstruction using undersampled measurements. Each line in the plot represents the reconstruction results for each number of Radial Lines used. Every plot line is the mean error over a constrained range of $DF_\sigma$ values, and all other parameters fixed. (a) Results for Radial Lines $\in [6,..,30]$, i.e. sampling percentage of $[4.5,..,21.8]$% and (b) results when Radial Lines $\in [35,..,105]$, i.e. sampling percentage of $[25.2, .., 63.8]$%.
3.1. SMOOTHED ℓ-0 ALGORITHM

To conclude with, we propose the following for each parameter of the SL0 algorithm:

- $\sigma_{min}$
  The lower bound of $\sigma$ do not affect the algorithm in non noisy cases. The algorithm converges fast even with few St. Ascent runs and a value of $\sigma_{min} = 1E - 05$ is recommended. However, in cases when the measurement process could introduce noise, higher values should be chosen.

- $DF_\sigma$
  The parameter defines the $\sigma$ sequence and, thus, the initial St. Ascent points. In low sparseness or noisy data a high value should be chosen, in order to produce more $\sigma$ steps. For highly sparse signals, small values are sufficient. Recommended values $DF_\sigma \in [0.01, 0.05, 0.1, 0.3, 0.5]$.

- $\mu_0$
  scales the step-size sequence of the St. ascent algorithm and affects convergence. High values could result in moving far away from the global maximum and being trapped. Low values increase the iterations but can avoid local maxima. Suggested values $\mu_0 \in [0.5, 1, 1.5]$.

- $L$
  The number of inner iterations also affect time and convergence. Low values prevent the algorithm from being trapped but could finish away from the global maximum solution. High values if not trapped in local maxima, achieve better results when $DF_{\sigma \text{sigma}}$ is large, preventing a near location in the next step. Suggested values $L \in [3, 10]$. 
3.2 \( \ell_1 \)-magic Algorithm

The reader may refer to "\( \ell \)-1 Magic Algorithm, Candès E., Romberg J., 2005". A recent series of papers [4]-[10] develops a theory of signal recovery from highly incomplete information. The central results state that a sparse vector \( \mathbf{m} \in \mathbb{R}^N \) can be recovered from a small number of linear measurements \( \mathbf{y} = \Phi \mathbf{m}, \in \mathbb{R}^K, K << N \) by solving a convex optimization program. Here, an optimization problem with linear equality or inequality constraints is solved by reducing it to a sequence of linear equality constrained problems and applying Newton's method.

3.2.1 Min-TV with equality constraints

We begin with the general problem formation. We wish to recover the sparse signal \( \mathbf{m} \), from undersampled measurements, and we need to achieve a solution that is consistent with the measurements \( \mathbf{y} ):

\[
\text{minimize } \| \mathbf{m} \|_{\ell_1} \text{ subject to } \Phi \mathbf{m} = \mathbf{y}
\]  
(3.6)

Remember that the \( \| \cdot \|_{\ell_1} \) norm is the sum of the absolute signal values, i.e. \( \| x \|_{\ell_1} = \sum_{i=1}^N |x_i| \). The \( \ell_1 \) norm enforces sparsity and the constraint \( \Phi \mathbf{m} = \mathbf{y} \) ensures data consistency. The problem of Eq 3.6 can be recast as a Second Order Cone Program. Primal dual algorithms are common in solving SOCP problems, but the authors, instead, solve the problem via a log-barrier method using a series of Newton steps.

An MR oriented approach to the problem states that instead of the image itself, we may say that the gradient of the image is sparse, which leads to the minimization of the Total Variation (TV) norm of \( \mathbf{m} \). The TV norm of a signal is simply the sum of the magnitudes of the discrete gradient values: \( \| \mathbf{m} \|_{TV} = \sum_i \sqrt{|m_i - m_{i-1}|^2} = \sum_i \| D_i \mathbf{m} \|_2 \).

\[
\text{minimize } TV(\mathbf{m}) \text{ subject to } \Phi \mathbf{m} = \mathbf{y}
\]  
(3.7)

If there exists a piecewise constant \( \mathbf{m} \) with sufficiently few edges (i.e. the gradient of \( \mathbf{m} \), \( D_i \mathbf{m} \), has nonzero coefficients only for a small number of indices \( i \)), then problem 3.7 will recover \( \mathbf{m} \) exactly. This problem can be rewritten in the form of problem 3.6 and be recast as a SOCP, as well. The problem of Eq.3.7 is rewritten as follows:

\[
\text{minimize } \sum (t_i) \text{ subject to } \Phi \mathbf{m} = \mathbf{y} \text{ and } \| D_i m_i \|_2 \leq t_i, \ i = 1, ..., N
\]
3.2. $\ell_1$-MAGIC ALGORITHM

\begin{table}[h]
\centering
\begin{tabular}{|l|}
\hline
**TV minimization with equality constraints** \\
\hline
minimize $\ell_1TV(m)$ s.t. $\Phi m = y$. \\
It can be rewritten as an SOCP problem:  \\
minimize $t, m \sum_i t_i$ s.t. $\|D_im\|_2 \leq t_i$ \\
$\Phi m = y$ \\
The inequality functions are:  \\
$f_{ti} = \frac{1}{2}(\|D_im\|_2^2 - t_i^2), \ i = 1, ..., N$ \\
\hline
\end{tabular}
\caption{Table 3.3}
\end{table}

where $m \in \mathbb{R}^N$, $y \in \mathbb{R}^M$, $\Phi$ is an $M \times N$ matrix, and each $f_i = \|D_im\|_2 - t \leq 0$ describes second-order conic constraint. It is obvious now, that it is an optimization problem with generalized constraints. The log-barrier method, follow the generic (but effective) algorithm described in Chapter 11 of [26], and a series of Newton steps is the core. The standard log-barrier method transforms Eq(3.7) into a series of linearly constrained programs:

$$
\text{minimize } m\|D_im\|_2^2 + \frac{1}{\tau_k} \sum_i -\log(-f_i(m)) \text{ subject to } \Phi m = y \quad (3.8)
$$

where $\tau_k > \tau_{k-1}$. The inequality constraints have been incorporated into the functional via a penalty function which is infinite when the constraint is violated (or even met exactly), and smooth elsewhere. The minimization method is based on solving a sequence of linearly constrained minimization problems, using the last point found as the starting point for the next minimization problem. In other words, we compute $m^*$ for a sequence of increasing $\tau$ ($\tau_1, ..., \tau_k$), until $\tau_k \geq N/\epsilon$, which guaranteed that we have an $\epsilon$-optimal solution of the original problem. In our case $\epsilon = \sum \|D_im\|_2 - \sum \|D_im^*\|_2$. Our primal problem is Eq.3.7, with solution $m$ and our dual is Eq.3.8 with solution $m^*$. As $\tau_k$ gets large, the solution $m^k$ to (3.8) approaches the solution $m^*$ to (3.7): it can be shown that $\sum \|D_im^k\|_2 - \sum \|D_im^*\|_2 < N/\tau_k$, i.e. we are within $\tau_k$ of the optimal value after iteration $k$ ($N/\tau_k$ is called the duality gap). The idea here is that each of the smooth subproblems can be solved to fairly high accuracy with just a few iterations of Newton’s method, especially since we can use the solution $z_k$ as a starting point for subproblem $k + 1$. Although any method for linearly constrained minimization can be used, Newton’s method is invoked in order to solve Eq.(3.7) with a step direction of $\Delta m$. At log-barrier iteration $k$, Newton’s method (which is again iterative) proceeds by forming a series of quadratic approximations to (3.8), and minimizing each by solving a system of
equations. The quadratic approximation of the functional in Eq.(3.8): 
\[ f_0(m) = \sum \| D_i m \|_2 + \frac{1}{\tau} \sum_i - \log( - f_i(m)) \] 
around a point \( m \) is given according to Newton method by

\[
f_0(m + \Delta m) \approx m + \langle g_m, \Delta m \rangle + \frac{1}{2} \langle H_m \Delta m, \Delta m \rangle := q(m + \Delta m) \tag{3.9}
\]

where \( g_m \) is the gradient

\[ g_m = D_i + \frac{1}{\tau} \sum_i - f_i(m) \nabla f_i(m) \tag{3.10} \]

and \( H_m \) is Hessian matrix

\[
H_m = \frac{1}{\tau} \sum_i \frac{1}{f_i(m)^2} \nabla f_i(m)(\nabla f_i(m))^T + \frac{1}{\tau} \sum_i - f_i(m) \nabla^2 f_i(m) \tag{3.11}
\]

Given that \( m \) is feasible (that \( \Phi m - y \), in particular), the \( \Delta m \) that minimizes \( q(m + \Delta m) \) subject to \( \Phi m = y \) is the solution to the set of linear equations, the Newton equations

\[
\tau = \begin{pmatrix} H_m & \Phi^T \\ \Phi & 0 \end{pmatrix} \begin{pmatrix} \Delta m \\ u \end{pmatrix} = -\tau g_m \tag{3.12}
\]

The vector \( u \), which can be interpreted as the Lagrange multipliers for the quality constraints in the quadratic minimization problem, is not directly used. With \( \Delta m \) in hand, we have the Newton step direction. The complete log-barrier implementation follows the outline:

1. Inputs: a feasible starting point \( m_0 \), a tolerance \( \text{lbtol} \), and parameters \( \mu \) and an initial \( \tau^1 \). Set \( k = 1 \).
2. Solve (3.8). Compute \( m^* \) by minimizing 3.8 via Newton’s method (followed by the backtracking line search), using \( m^{k-1} \) as an initial point. Call the solution \( m^* k \).
3. Update \( m^k = m^*_k \)
4. If duality gap = \( N/\tau^k < \text{lbtol} \) terminate and return \( m^k \).
5. Else, increase \( \tau^{k+1} = \mu \tau^k \), \( k = k + 1 \) and go to step 2.

The authors choose initial \( \tau^1 \) conservatively; it is set so that the duality gap \( N/\tau^k \) after the first iteration is equal to \( \sum \| D_i m^0 \|_2 \). The cost of computing an extremely accurate minimizer in Step 2, as compared to the cost of computing a good minimizer is only marginally more, i.e. a few Newton steps at most. The choice of parameter \( \mu \)
affects the number of inner and outer iterations required.

<table>
<thead>
<tr>
<th>Params</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>slqtol</td>
<td>Stop criteria of Newton iterations determining the desired precision</td>
</tr>
<tr>
<td>slqmaxiter</td>
<td>Second (backup) stop criteria of Newton method.</td>
</tr>
<tr>
<td>lbtol</td>
<td>Stop criteria of logbarrier algorithm, on the acceptable duality gap.</td>
</tr>
<tr>
<td>µ</td>
<td>Step-size that trades-off the inner and outer iterations.</td>
</tr>
</tbody>
</table>

The plots in Fig. 3.7 clearly show the trade-off in the choice of µ. The progress of the barrier method is shown, using a small example of a $\mathbb{N}^50$ signal, and $\Phi^{5\times100}$ from a randomly generated problem instance. The vertical axis show the duality gap, and the horizontal axis show the cumulative total number of inner iterations (Newton steps). Each stair is associated with one outer iteration, and the width of each stair (horizontal portion) is the number of Newton steps required for that outer iteration. The height (vertical portion) is equal to µ, since the duality gap is reduced by a factor of µ at the end of each outer iteration. The choice of µ has little effect on the total number of Newton steps, provided µ is at least 10 or so.

![Figure 3.7: Progress of barrier method for an SOCP, showing duality gap versus cumulative number of Newton steps. Three plots are shown, corresponding to three values of parameter µ: 2, 50 and 200.](image-url)
3.2.2 Results and Findings

Again, the experiments were done on the Shepp-Loggan phantom image, that can well represent a general MR image content, along with a Radial trajectory, as discussed in $SL0$ algorithm. The different values used in the experiments are shown in the following table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$lbtol$</td>
<td>1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 2E-1, 4e-1</td>
</tr>
<tr>
<td>$\mu$</td>
<td>1.1, 2, 3, 4, 5, 6, 8, 10, 12,14</td>
</tr>
<tr>
<td>$slqtol$</td>
<td>1e-10, 1e-9, 1e-8, 1e-7, 1e-6</td>
</tr>
<tr>
<td>$slqmaxiter$</td>
<td>50, 100, 200, 400, 600, 800</td>
</tr>
</tbody>
</table>

Let us begin with the effect of the $slqtol$ value on the reconstruction results. Among the $slqtol$ values tested, no one clearly outperformed the others, (as expected by the essence/use of the specific parameter in the algorithm, and the specific variable values tested).

To illustrate it, Figure 3.8 below has been constructed. Results only on Radial lines = 10, 20, 30 are shown for a better view, but keep in mind that similar plots were produced for all other Radial Line values. The red color demonstrates the MSE results when Radial lines=10, for every different $lbtol$, $\mu$ and $slqmaxiter$ value. The blue lines correspond to 20 Radial lines, and the magenta to 30 Radial lines. No mean plots were made. The figure is clear enough as is. We choose to fix $slqtol = 1e - 8$, which is also the value proposed by the authors.
3.2. $\ell_1$-MAGIC ALGORITHM

Figure 3.8: Representative plot of $slqtol$ effect, for all different Radial line values. The plot shows results for Radial lines= [10, 20, 30]. On the x axis, the different values of $slqtol$ variable tested. On y axis, the Mean Square reconstruction Error. Red Color: 10 Radial lines. Blue color: 20 radial lines. Magenta: 30 Radial lines. All algorithm runs seem to output the same MSE for any $slqtol$ value.
CHAPTER 3. EXHAUSTIVE COMPARISON OF THREE ALGORITHMS

Let’s continue with the $slqmaxiter$ parameter: Figure 3.9 illustrates the effect of the specific parameter, when $slqtol$ is fixed to $slqtol = 1e \cdot 8$. A mean view is presented, over the different $lbtol$ and $\mu$ values. Every different line correspond to a different Radial Line number and represents the mean MSE plot over all $lbtol$ and $\mu$, versus the $slqmaxiter$.

Judging from the plot of Figure 3.9, when the available data for the reconstruction are below 30 Radial Lines (i.e. a percentage of 10.9%) the $slqmaxiter$ value is recommended to be higher than 200 iterations. In case where enough data is available, even 50 iterations are effective. For the general case, $slqmaxiter = 400$ is proposed: it demands some extra computational time but may result in a slightly better reconstruction error, than using 200 iterations. This being the case, we fix $slqmaxiter = 400$. 
3.2. $\ell_1$-MAGIC ALGORITHM

Figure 3.9: Reconstruction results as affected by $slqmaxiter$. On the x axis, the $slqmaxiter$ variable. On y axis, the Mean Square reconstruction Error. Every different color line corresponds to a specific Radial line value. when $slqtol$ is fixed to $1e^{-8}$. (a) Results for Radial Lines $\in [6,..,30]$ i.e. sampling percentage of $[4.5,..,21.8]$% and (b) results when Radial Lines $\in [35,..,105]$, i.e. sampling percentage of $[25.2,..,63.8]$%.
So, $slqtol$ is fixed to $1e-8$, and $slqmaxiter$ is fixed to 400. Figure 3.10 presents the effect of the $lbtol$ parameter on the reconstruction quality. More specifically, the plot presented below, is a mean MSE plot over all different $\mu$ values, versus the $lbtol$ parameter. Every different color corresponds, again, to a different Radial Line value. Like before, the plot is split for Radial Lines up to 30, and Radial Lines ranging from 30 to 105.

As expected, small tolerance results in lower reconstruction errors. Small $lbtol$ values may increase computational time, but may also increase the accuracy of the reconstructed image. Recommended value here is $lbtol = 1e-5$, but also acceptable if ranging in $\{1e-5, 1e-4, 1e-3\}$.
3.2. $\ell_1$-MAGIC ALGORITHM

Figure 3.10: Reconstruction errors as affected by the varying $lbtol$ parameter, keeping fixed $slqtol = 1e-8$, and $slqmaxiter = 400$. On x axis, the $lbtol$ variable. On y axis, the Mean Square reconstruction Error. Every different color line in the plot corresponds to a specific Radial line value. (a) Results for Radial Lines $\in [6, ..., 30]$, i.e. sampling percentage of $[4.5, ..., 21.8] \%$ and (b) results when Radial Lines $\in [35, ..., 105]$, i.e. sampling percentage of $[25.2, ..., 63.8] \%$. 
The effect of the \( \mu \) parameter is the last one to study. Remember that \( s l q t o l \) is fixed to \( 1e-8 \), \( s l q m a x i t e r \) is fixed to 400 and \( l b t o l \) is fixed to \( 1e-5 \). Figure 3.11 show the MSE of the reconstructed image versus \( \mu \) parameter, when all other parameters are set as described earlier. Every different line in the plot corresponds to a different sampling percentage.

The lower step value \( \mu = 1.1 \) seems to outperform all other tested values, in the cost of a high computational time. Notice that when \( \mu = 1.1 \), the algorithm achieved quite good reconstruction results, even in the case of low \( k \)-space coverage. Recommended value is \( \mu = 1.1 \). One may need to choose among higher values, in case that time is of the essence, and extra computational time is undesirable.
3.2. $\ell_1$-MAGIC ALGORITHM

Figure 3.11: Reconstruction errors as affected by the varying $\mu$ parameter, keeping $ltol = 1e - 5$, $sqtol = 1e - 8$, and $slqmaxiter = 400$. On x axis, the $\mu$ variable. On y axis, the Mean Square reconstruction Error. Every different color line in the plot corresponds to a specific Radial line value. (a) Results for Radial Lines $\in [6, \ldots, 30]$, i.e. sampling percentage of $[4.5, \ldots, 21.8]\%$ and (b) results when Radial Lines $\in [35, \ldots, 105]$, i.e. sampling percentage of $[25.2, \ldots, 63.8]\%$. 
To conclude with, we propose the following parameter values for use,

- **slqtol** First stop criteria of Newton solver algorithm. It determines the desired precision of the algorithm, which terminates when \( \frac{\text{norm}(\Phi_k m - y)}{\text{norm}(y)} < \text{slqtol} \)

  Proposed value: \( \text{slqtol} = 1e^{-8} \).

- **slqmaxiter** Second stop criteria of Newton solver algorithm. Inner iterations stops is any of **slqtol** or **slqmaxiter** criteria are satisfied. It determines the maximum iterations. When enough data is available, even 50 iterations are sufficient. Suggested value : \( \text{slqmaxiter} = 400 \).

- **lbtol** Stop criteria for the logbarrier algorithm. Suggested value: \( \text{lbtol} = 1e^{-5} \) (also acceptable \( \in \{1e^{-5}, 1e^{-4}, 1e^{-3}\} \)).

- **\( \mu \)** Factor by which to increase the barrier constant \( \tau \) at each outer iteration. \( \mu = 1.1 \). If time is essential, \( \mu \) values larger than 2 should be used.

The log barrier algorithm terminates when the duality gap = \( N \tau^k < \text{lbtol} \).

Also, the number of log barrier iterations is completely determined by \( \text{lbtol} \) (barrier iter = \( \frac{\log M}{\log \mu} - \frac{\log \text{lbtol}}{\log \mu} - \log 1 \)).

As parameter \( \tau \) increases, the Newton problem could be considered by many to become more difficult and therefore require more iterations. However, numerical evidence suggests that this is not the case. The Newton problem appears to require a nearly constant number of steps to solve, even as \( \tau \) increases.

The choice of the parameter \( \mu \) involves a trade-off in the number of inner and outer iterations required. If \( \mu \) is small (i.e. near 1) then at each outer iteration \( \tau \) increases by a small factor. As a result, the initial point for the Newton process (the previous iterate \( m \)), is a very good starting point and the number of Newton steps needed to compute the next iterate is small. Thus, for small \( \mu \) we expect a small number of Newton steps, per outer iteration, but of course a large number of outer iterations since each outer iteration reduces the gap by only a small amount.

On the other hand, if \( \mu \) is large we have the opposite situation. After each outer iteration \( \tau \) increases a large amount, so the current iterate is probably not a very good approximation of the next iterate and we expect many more inner iterations.

This ‘aggressive’ updating of \( \tau \) results in fewer outer iterations, since the duality gap is reduced by the large factor \( \mu \) at each outer iteration, but more inner iterations. For \( \mu \) in a fairly large range from around 3 to 100 or so, the two effects nearly cancel, so the total number of Newton steps remains approximately constant.
3.3 Sparse MRI Algorithm

The reader may refer to "Michael Lustig "SPARSE MRI", Ph.D Thesis, Stanford University, August 2008". Lustig et al. addresses the problem of Compressive Sensing in terms of MR image acquisition with a reconstruction approach very similar to the one presented in the previous section ($\ell_1$-magic). Nevertheless the algorithm is presented due to the high contribution of the author to the CS community. Coming from the world of electrical engineering, Lustig studied and recorded many crucial details and secrets of Magnetic Resonance for applications in the hot and recent CS field. He developed algorithms for time-optimal gradient magnetic field design and incorporated phase correction schemes into CS reconstruction problems. He is among the first researchers studying applications of CS into dynamic Heart Imaging and 3D angiography. His algorithm is a faster, simplified version of the $l_1$-magic algorithm, allowing a trade-off between reconstruction sparsity and data consistency.

The authors developed an implementation of an $\ell_1$ penalized non linear conjugate gradient reconstruction. Let $m$ be the image of interest (vector), $\Psi$ denote the linear operator that transforms from pixel representation into a sparse representation, and let $F_u$ be the undersampled Fourier transform. Given $k$-space measurements $y$, the reconstruction is obtained by solving the constraint optimization problem:

$$\min m \quad \lambda_1 \left\| \Psi^{-1} m \right\|_1 + \lambda_{TV} \alpha TV (m)$$

s. t. $\left\| F_u m - y \right\|_2 < \epsilon$ \hspace{1cm} (3.13)

where $\lambda_1$ and $\lambda_{TV}$ trades sparsity between the $\Psi$ and finite-differences ($TV(m)$) domain, and $\epsilon$ controls the fidelity of the reconstruction as compared to the measured data. The reconstructed image is $m$, and the measurement $k$-space data taken by the scanner is $y$.

The constrained convex optimization problem of eq. 3.13 is rearranged to form an unconstrained problem in Lagrangian form:

$$\arg \min m \quad \left\| F_u m - y \right\|_2^2 + \lambda \left\| \Psi^{-1} m \right\|_1$$ \hspace{1cm} (3.14)

Here, $\lambda$ is a regularization parameter between the data consistency and the sparsity. The value can be determined by solving eq. 3.14 for different values, and then choosing $\lambda$, so that $\left\| F_u m - y \right\|_2 \approx \epsilon$. The value of $\lambda$ can be appropriately chosen so that the solution of eq. 3.14 is the same as the one of eq. 3.13.

The proposed method for solving eq. 3.14 is the use of non-linear conjugate gradients and backtracking line-search. The cost function is denoted as $f(m)$ and is the one in eq. 3.14. The conjugate gradient requires the computation of $\nabla f(m)$:

$$\nabla f(m) = 2F_u^*(F_u m - y) + \lambda \nabla \left\| \Psi^{-1} m \right\|_1$$ \hspace{1cm} (3.15)

The $\ell_1$ norm is the sum of the absolute values, but it is not a smooth function, resulting the authors to approximate the absolute value as $|x| = \sqrt{x^*x + \mu}$, with $\mu$
being a smoothing parameter and \( \frac{dx}{dx} \approx \frac{x}{\sqrt{x^2 + \mu}} \). Now, let \( W \) be a diagonal matrix with the diagonal elements \( w_i = \sqrt{(\Psi^{-1}m)_i^2(\Psi^{-1}m)_i + \mu} \), then eq. 3.15 becomes:

\[
\nabla f(m) \approx 2F_*^*(F_*m - y) + \lambda \Psi^{-1}W^{-1}\Psi^{-1}m
\]

(3.16)

and in practice \( \mu \) is chosen to be \([10^{-15}, 10^{-6}]\). The number of CG iterations varies with the problem size.

---

**Figure 3.12:** Outline of Lustig’s non-linear Conjugate-Gradient reconstruction algorithm.
3.3. SPARSE MRI ALGORITHM

3.3.1 Results and findings

As before, the experiments were done on Shepp-Logan phantom image using the Radial Line trajectory. The different parameter values used in the experiments are shown in the following table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration limit</td>
<td>8, 50, 150</td>
</tr>
<tr>
<td>$\lambda_{TV}$</td>
<td>0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03</td>
</tr>
<tr>
<td>$\lambda_{\ell_1}$</td>
<td>0.00, 0.005, 0.01, 0.02, 0.03</td>
</tr>
</tbody>
</table>

Let us begin our assessment with the effect of the Iteration number. The iterations affect the outer loop, the loop of the Gradient Algorithm. In the case were no local minimum exist, we expect to observe a better behavior when a higher number of iterations is used. This is also the case here, since the objective function does have a global minimum, and as Figure 3.13 shows, the more the iterations, the better the convergence of the algorithm.

![Figure 3.13: Representative plot of Iteration effect, for all different Radial line values. On the x axis, the different values of Iteration variable tested. On y axis, the Mean Square reconstruction Error. Results are shown for Radial Lines $\in [6, \ldots, 105]$, i.e.sampling percentage of $[4.5, \ldots, 63.8]\%$. It is a mean plot over all different $\lambda_{\ell_1}$ and $\lambda_{TV}$ values.](image-url)
We choose to fix $\text{Iterations} = 150$ in the expense of a higher computational time, and continue with the $\lambda_{TV}$ variable. $\lambda_{TV}$ denotes the contribution of the Total Variation sparseness that we want to introduce in the model. Figure 3.14 reveals that 0.005 is a good choice for weighting the TV factor for this kind of images.

So, we fix $\lambda_{TV} = 0.005$, $\text{Iterations}$ are already fixed to 150, and continue with the $\lambda_{\ell_1}$ variable. $\lambda_{\ell_1}$ signifies the weight of the $\ell_1$ norm sparsity that contributes in the objective function. Judging from the plot of Figure 3.15 the values 0.005 and 0.01 seem to have a slightly better reconstruction performance than higher values.
3.3. **Sparse MRI Algorithm**

Figure 3.15: Representative plot of $\lambda_\ell_1$ effect, for all different Radial line values. *Iterations* are fixed to 150, and $\lambda_{TV}$ to 0.05. On the x axis, the different values of $\lambda_\ell_1$ variable tested. On y axis, the Mean Square reconstruction Error. Results are shown for Radial Lines $\in [6, .., 105]$ i.e.sampling percentage of [4.5, .., 63.8]%.

To conclude with, we propose the following

- *Iterations* The stopping criteria of Gradient solver algorithm. It determines the steps done towards the minimum solution of the objective function. Proposed value: *Iterations* $> 100$.

- $\lambda_\ell_1$ The weighting factor for the $\ell_1$ norm sparseness of the reconstructed signal. Proposed values: $\lambda_\ell_1 = \{0.005, 0.01\}$

- $\lambda_{TV}$ The weighting factor for the enforcements of the Total Variation sparseness of the reconstructed signal. Proposed value: $\lambda_{TV} = 0.005, 0.01$

The stepsize of the Gradient Conjugate algorithm is completely determined by the rate of gradient change during previous iteration. The number of iterations permitted is the only stopping criterion used, although it is very straightforward to incorporate an extra constraint/criterion and for instance, terminate the algorithm when two consecutive iterations returns the same (almost same) solution. The authors provide a simply written algorithm with high scalability and ability to easily incorporate extra constraints on the minimization problem.
3.4 Discussion on Results

Let us now compare the above algorithms in terms of computational complexity, time demands and reconstruction error. In Fig.3.16 we may see the best reconstruction error, as achieved by the reconstruction process of every one of the three algorithms. For every different percentage coverage of k-space, the best MSE is plotted, as a result of the best set of parameters for the corresponding sampling. A zoom-in view is also incorporated into the plot, for a better understanding. It is clear how the l1-magic outperforms SL0 in terms of the reconstruction error.

Keep in mind, though, that the $\ell_1$-magic algorithm incorporate a sparsifying constraint using the Total Variation (TV) as a sparsifying Transform. The same constraint is also used in the SparseMRI scheme. This is why, a further dotted line plot can be seen in the figure: it represents the reconstruction performance of SL0 algorithm when a TV sparsifying transform has been applied during the reconstruction process. In the end of this section, in Figures 3.18-3.52 the reconstructed images achieved by the three algorithms is shown, with respect to the different sampling percentages used.

![Performance comparison of the algorithms](image)

Figure 3.16: The three algorithm are compared regarding the Mean Square Error of the reconstruction process, as achieved for every different sampling percentage. The Shepp-Logan phantom image is used, together with a radial sampling pattern. For every algorithm, the best MSE is shown, among all sets of parameters, as achieved for the different k-space coverage percentages. Notice the dotted line that corresponds to the SL0 algorithm where, additionally, a TV sparsifying transform was applied.
3.4. DISCUSSION ON RESULTS

Figure 3.17: The comparison results, as presented in Figure 3.16. The first highlighted line, denotes the least radial lines needed for algorithm sparseMRI to achieve a very good reconstruction. The second highlighted line denotes the least radial lines needed for algorithm l1-magic to achieve a very good reconstruction. SL0 algorithm without sparsifying transforms, cannot manage to achieve a good reconstruction even when more than 60% samples are available. This algorithm requires a very high sparsity of the input signal (recall that the sparsity of the shepp-Loggan phantom equals to 50.93%).

<table>
<thead>
<tr>
<th>Radial Lines</th>
<th>% Samples</th>
<th>I1 magic</th>
<th>SL0</th>
<th>SL0 sparse</th>
<th>SparseMRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4.60</td>
<td>0.028111</td>
<td>0.0235</td>
<td>0.0238</td>
<td>0.020996</td>
</tr>
<tr>
<td>7</td>
<td>5.35</td>
<td>0.023378</td>
<td>0.0156</td>
<td>0.0239</td>
<td>0.01503</td>
</tr>
<tr>
<td>8</td>
<td>6.12</td>
<td>0.021383</td>
<td>0.0173</td>
<td>0.023</td>
<td>0.014292</td>
</tr>
<tr>
<td>9</td>
<td>6.86</td>
<td>0.0192</td>
<td>0.0166</td>
<td>0.0214</td>
<td>0.014586</td>
</tr>
<tr>
<td>10</td>
<td>7.58</td>
<td>0.011657</td>
<td>0.0121</td>
<td>0.0169</td>
<td>0.0072484</td>
</tr>
<tr>
<td>11</td>
<td>8.34</td>
<td>0.01143</td>
<td>0.0117</td>
<td>0.0127</td>
<td>0.0075983</td>
</tr>
<tr>
<td>12</td>
<td>9.09</td>
<td>0.010027</td>
<td>0.0119</td>
<td>0.011</td>
<td>0.0059121</td>
</tr>
<tr>
<td>13</td>
<td>9.81</td>
<td>0.005133</td>
<td>0.0084</td>
<td>0.009</td>
<td>0.00065828</td>
</tr>
<tr>
<td>14</td>
<td>10.53</td>
<td>0.0025374</td>
<td>0.0099</td>
<td>0.0083</td>
<td>0.0005835</td>
</tr>
<tr>
<td>15</td>
<td>11.28</td>
<td>0.0010448</td>
<td>0.0095</td>
<td>0.0072</td>
<td>2.24E-05</td>
</tr>
<tr>
<td>16</td>
<td>12.02</td>
<td>0.00034921</td>
<td>0.0068</td>
<td>0.006</td>
<td>1.05E-05</td>
</tr>
<tr>
<td>17</td>
<td>12.71</td>
<td>0.00014827</td>
<td>0.0066</td>
<td>0.005</td>
<td>2.90E-06</td>
</tr>
<tr>
<td>18</td>
<td>13.44</td>
<td>4.95E-06</td>
<td>0.0058</td>
<td>0.0021</td>
<td>1.66E-06</td>
</tr>
<tr>
<td>19</td>
<td>14.16</td>
<td>1.97E-07</td>
<td>0.0054</td>
<td>0.00097</td>
<td>1.43E-06</td>
</tr>
<tr>
<td>20</td>
<td>14.90</td>
<td>4.82E-07</td>
<td>0.0047</td>
<td>0.00083</td>
<td>9.28E-07</td>
</tr>
<tr>
<td>22</td>
<td>16.27</td>
<td>3.00E-17</td>
<td>0.0042</td>
<td>0.00072</td>
<td>7.58E-09</td>
</tr>
<tr>
<td>24</td>
<td>17.69</td>
<td>1.46E-17</td>
<td>0.0040</td>
<td>0.00054</td>
<td>5.58E-11</td>
</tr>
<tr>
<td>26</td>
<td>19.08</td>
<td>5.70E-18</td>
<td>0.0031</td>
<td>0.0002</td>
<td>4.42E-11</td>
</tr>
<tr>
<td>28</td>
<td>20.47</td>
<td>4.23E-18</td>
<td>0.0030</td>
<td>0.00012</td>
<td>8.55E-11</td>
</tr>
<tr>
<td>30</td>
<td>21.84</td>
<td>2.58E-18</td>
<td>0.0024</td>
<td>5.96E-22</td>
<td>4.32E-12</td>
</tr>
<tr>
<td>35</td>
<td>25.18</td>
<td>1.11E-18</td>
<td>0.0020</td>
<td>5.98E-22</td>
<td>2.95E-12</td>
</tr>
<tr>
<td>40</td>
<td>28.53</td>
<td>6.79E-19</td>
<td>0.0014</td>
<td>5.75E-22</td>
<td>1.88E-12</td>
</tr>
<tr>
<td>45</td>
<td>31.67</td>
<td>4.16E-19</td>
<td>0.0014</td>
<td>5.49E-22</td>
<td>1.25E-12</td>
</tr>
<tr>
<td>50</td>
<td>34.80</td>
<td>2.74E-19</td>
<td>0.0009</td>
<td>5.69E-22</td>
<td>1.11E-12</td>
</tr>
<tr>
<td>55</td>
<td>37.81</td>
<td>2.04E-19</td>
<td>0.0008</td>
<td>5.51E-22</td>
<td>7.80E-14</td>
</tr>
<tr>
<td>60</td>
<td>40.83</td>
<td>1.35E-19</td>
<td>0.0007</td>
<td>5.64E-22</td>
<td>6.75E-14</td>
</tr>
<tr>
<td>65</td>
<td>43.65</td>
<td>5.82E-20</td>
<td>0.0006</td>
<td>5.66E-22</td>
<td>5.91E-14</td>
</tr>
<tr>
<td>70</td>
<td>46.45</td>
<td>8.32E-20</td>
<td>0.0006</td>
<td>5.50E-22</td>
<td>4.91E-16</td>
</tr>
<tr>
<td>75</td>
<td>49.18</td>
<td>6.83E-20</td>
<td>0.0004</td>
<td>5.58E-22</td>
<td>4.31E-17</td>
</tr>
<tr>
<td>80</td>
<td>51.87</td>
<td>5.29E-20</td>
<td>0.0003</td>
<td>5.52E-22</td>
<td>3.99E-17</td>
</tr>
<tr>
<td>85</td>
<td>54.38</td>
<td>4.21E-20</td>
<td>0.0003</td>
<td>5.43E-22</td>
<td>3.69E-17</td>
</tr>
<tr>
<td>90</td>
<td>56.82</td>
<td>3.79E-20</td>
<td>0.0002</td>
<td>5.63E-22</td>
<td>3.40E-17</td>
</tr>
<tr>
<td>95</td>
<td>59.25</td>
<td>2.75E-20</td>
<td>0.0002</td>
<td>5.62E-22</td>
<td>3.20E-17</td>
</tr>
<tr>
<td>100</td>
<td>61.68</td>
<td>1.66E-20</td>
<td>0.0001</td>
<td>5.57E-22</td>
<td>2.94E-20</td>
</tr>
<tr>
<td>105</td>
<td>63.83</td>
<td>1.96E-20</td>
<td>0.0001</td>
<td>5.45E-22</td>
<td>2.80E-21</td>
</tr>
</tbody>
</table>
When it comes to time requirements, clearly the SL0 stands out: it is a very fast algorithm. However, in order to achieve an equivalent performance with the other algorithms, more $k$-space samples need to be acquired, resulting in a higher time specification for the patient to remain still. SL0 algorithm needs in average 0.06 to 0.9 seconds. The lower time bound represents cases when $L = 3$ iterations are used, and the upper bound when $L = 20$ iterations were used. On the other hand, the l1-magic needs in average 3 to 100 seconds, depending on the step size $\mu$. There were cases where reconstruction process took up to 8 minutes, when a very small $\mu = 1.1$ was chosen! SparseMRI is a time-consuming algorithm as well. In cases where more than 100 iterations was used even 9 minutes where not enough!

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL0</td>
<td>0.06-0.09 sec</td>
<td>0.01 sec</td>
<td>0.15 sec</td>
</tr>
<tr>
<td>l1-magic</td>
<td>3-100 sec</td>
<td>1 sec</td>
<td>540 sec</td>
</tr>
<tr>
<td>sparseMRI</td>
<td>5-150 sec</td>
<td>1 sec</td>
<td>700 sec</td>
</tr>
</tbody>
</table>

Table 3.4: The average recovery time for every algorithm is shown in column 1. The lower and upper bound represents the minimum and maximum iterations used. In columns 2 and 3, an indication of the minimum and maximum reconstruction times is also available.

The empirical complexity of l1-magic and sparseMRI algorithms with quadratic constraints is $O(N^p)$ with $3 \leq p \leq 4$. SL0 algorithm needs an execution time of the order of $O(N \log N)$, which results from the back and forth FFT projections. Faster implementations of the FFT transform will require even less computational time. However, SL0 requires about two times more measurements, in order to obtain results equivalent to l1-magic, for this kind of images.

Remember that an MR scan may need 20 to 90 minutes to be completed. It is very inconvenient for the patient to stay still inside the scanner bore for so long. Therefore it is far better for the patient to wait a little longer for the reconstruction results, after the scan, rather than wait longer (without dodging) inside the scanner.

The $\ell_1$ reconstruction tends to slightly shrink the magnitude of the reconstructed space coefficients. The resulting reconstructed coefficients are often slightly smaller than in the original signal. This coefficient shrinkage decreases when the reconstruction consistency parameter (i.e. the acceptable error) is small. In CS the contrast in image (the $\Re$-space MR image) plays a major part in the ability to vastly undersample and reconstruct images. High contrast often results in large distinct sparse coefficients, which can be recovered even at very high undersampling (i.e. sampling acceleration on a k-space trajectory). For example a single bright pixel will most likely appear in the reconstruction even with vast undersampling, as revealed by many ex-
3.4. DISCUSSION ON RESULTS

Experimental results.

However features with lower contrast at the same acceleration (undersampling) are deeply submerged by the interference, that they are not recoverable. This means that the most distinct artifacts in CS are not loss of resolution but loss of low contrast features in the image. Therefore, CS is particularly attractive in applications that exhibit high resolution high contrast image features, and where rapid imaging is required.

CS exploits sparsity and compressibility of MR images. These two theories, coupled, can be combined with other acceleration methods that exploit different kinds of redundancy. For example, constraining the image in the minimization problem to be real, effectively implies phase constrains in partial $k$-space with the CS reconstruction. In a similar way, CS can be combined with methods like SENSE \cite{62} reconstruction, by including the coil sensitivity information, such as the case of multiple coil receivers, in the problem formulation. In general any other prior knowledge or information on the image, that can be expressed as a convex constraint, can be incorporated in the reconstruction.

![Figure 3.18](image)

![Figure 3.19](image)
CHAPTER 3. EXHAUSTIVE COMPARISON OF THREE ALGORITHMS

Figure 3.20

Figure 3.21

Figure 3.22
3.4. DISCUSSION ON RESULTS

Table 3.23

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Z/F MSE</th>
<th>SLO MSE</th>
<th>l1-magic MSE</th>
<th>sparseMRI MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 r.l. : 8.34%</td>
<td>0.023774</td>
<td>0.011700</td>
<td>0.011430</td>
<td>0.007598</td>
</tr>
</tbody>
</table>

Figure 3.23

Table 3.24

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Z/F MSE</th>
<th>SLO MSE</th>
<th>l1-magic MSE</th>
<th>sparseMRI MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 r.l. : 9.09%</td>
<td>0.023142</td>
<td>0.011900</td>
<td>0.010027</td>
<td>0.005912</td>
</tr>
</tbody>
</table>

Figure 3.24

Table 3.25

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Z/F MSE</th>
<th>SLO MSE</th>
<th>l1-magic MSE</th>
<th>sparseMRI MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 r.l. : 9.81%</td>
<td>0.022320</td>
<td>0.008400</td>
<td>0.005133</td>
<td>0.000658</td>
</tr>
</tbody>
</table>

Figure 3.25
CHAPTER 3. EXHAUSTIVE COMPARISON OF THREE ALGORITHMS

Figure 3.26

Figure 3.27

Figure 3.28
3.4. DISCUSSION ON RESULTS

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Z/F</th>
<th>SLO</th>
<th>l1-magic</th>
<th>sparseMRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 r.i.</td>
<td>12.71%</td>
<td>0.018361</td>
<td>0.006600</td>
<td>0.000148</td>
</tr>
<tr>
<td>18 r.i.</td>
<td>13.44%</td>
<td>0.019197</td>
<td>0.005800</td>
<td>0.000005</td>
</tr>
<tr>
<td>19 r.i.</td>
<td>14.16%</td>
<td>0.018920</td>
<td>0.005400</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

Figure 3.29

Figure 3.30

Figure 3.31
CHAPTER 3. EXHAUSTIVE COMPARISON OF THREE ALGORITHMS

Figure 3.32

Figure 3.33

Figure 3.34
3.4. DISCUSSION ON RESULTS

Figure 3.35

Figure 3.36

Figure 3.37
### Figure 3.38

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Z/F MSE</th>
<th>SLO MSE</th>
<th>l1-magic MSE</th>
<th>sparseMRI MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 r.L : 25.18%</td>
<td>0.011336</td>
<td>0.002000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

![Image of Figure 3.38](image)

### Figure 3.39

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Z/F MSE</th>
<th>SLO MSE</th>
<th>l1-magic MSE</th>
<th>sparseMRI MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>40 r.L : 28.53%</td>
<td>0.009939</td>
<td>0.001400</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

![Image of Figure 3.39](image)

### Figure 3.40

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Z/F MSE</th>
<th>SLO MSE</th>
<th>l1-magic MSE</th>
<th>sparseMRI MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>45 r.L : 31.67%</td>
<td>0.008913</td>
<td>0.001400</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

![Image of Figure 3.40](image)
3.4. DISCUSSION ON RESULTS

Figure 3.41

Figure 3.42

Figure 3.43
Figure 3.44

Figure 3.45

Figure 3.46
3.4. DISCUSSION ON RESULTS

Figure 3.47

Figure 3.48

Figure 3.49
Chapter 3. Exhaustive Comparison of Three Algorithms

Figure 3.50

Figure 3.51

Figure 3.52
Chapter 4

Bayesian Compressive Sensing

During the previous sections, the Compressive Sensing approach was described, along with the principles that need to be satisfied, revealing a way to incorporate this theory into a Magnetic Resonance Imaging concept. We showed that one may recover an original-unknown signal $m \in \mathbb{R}^N$ using only $K \ll N$ random linear combinations of the original signal values and the knowledge that $m$ is compressible in some basis $\Psi$, with high probability, solving an optimization problem. It would be very interesting to go even further: i.e. indeed we need to estimate the unknown signal $m$, but we also need to estimate our belief or confidence in the resulting reconstruction. Having a confidence estimation on every step of a potential recovery algorithm, we may even adaptively optimize the measurements taken. In order to address these problems, a Bayesian formalism [60], [61] on the Compressive Sensing scheme is needed, one that can also account for possible additive noise during the measurement process, and can estimate the noise variance. High levels of noise (due to hardware noise or patient’s movement) will automatically indicate that the MR experiment needs to be repeated anew.

We will show how bayesian concepts can be incorporated in Compressive Sensing. The minimization problem that Compressive Sensing tries to solve is restated for convenience in Eq. 4.1.

$$\minimize \|w\|_1 \quad \text{s.t.} \quad \|y - \Phi w\|_2^2 \leq \epsilon \quad (4.1)$$

We need to find a vector $w$, as sparse as possible that satisfies the measurements $y$ when projected on the measurement matrix $\Phi$. We may consider that $w$ is a sparse representation of our initial image signal $m$ ($m = \Psi w$), leading to the set of equations below.
CHAPTER 4. BAYESIAN COMPRESSIVE SENSING

If $m$: the image signal of interest
and $w$: its sparse representation, then
\[ w = \Psi^{-1}m \]
\[ y = \Phi w \]
\[ t = y + \epsilon \]
\[ t = \Phi w + \epsilon \]

Errors $\epsilon$ are modeled probabilistically as independent zero-mean Gaussian, with variance $\sigma_n^2$. Errors in equation $t = y + \epsilon$ refer to a possible additive gaussian noise and may even refer to the energy of the very low coefficients of $w$.

\[ p(\epsilon) = \prod N(\epsilon_n|0,\sigma_n^2) \]

This error model implies a multivariate Gaussian likelihood for the target vector $t$:

\[ p(t|w,\sigma_n^2) = (2\pi)^{-N/2}\sigma_n^{-N} \exp\left\{-\frac{\|t - y\|^2}{2\sigma_n^2}\right\} \]

And this likelihood is complemented by a zero-mean Gaussian prior on each element of $w$, over the parameters $a_i = 1/\sigma_{w_i}^2$:

\[ p(w|a) = (2\pi)^{-M/2} \prod_{m=1}^{M} a_i^{1/2} \exp\left(\frac{a_i w_i^2}{2}\right) \]

We assume independence for every $w_i$. Given $a$, the posterior parameter distribution conditioned on the data is:

\[ p(w|t, a, \sigma_n^2) = \frac{p(t|w,\sigma_n^2)p(w|a)}{p(t|a,\sigma_n^2)} \]

The posterior $p(w|t, a, \sigma_n^2)$ is Gaussian $N(\mu, \Sigma)$ with:

\[ \Sigma = (A + \frac{1}{\sigma_n^2} \Phi^T \Phi)^{-1} \]

\[ \mu = \frac{1}{\sigma_n^2} \Sigma \Phi^T t \]

(4.2)

where $A = \text{diag}(a_1, a_2, ..., a_M)$.

We will not extend the Bayesian model to include hyperpriors over $a$, because we are interested in the point estimate of $a$. We may found such a point estimate $a_{\text{MP}}$ via a type-II maximum likelihood procedure: the local maximization of the logarithm of the marginal likelihood with respect to $a$. So we need to maximize following:
4.1 Marginal Likelihood Maximization

We decompose $C$ in Eq.(4.4), considering the dependence of $L(a, \sigma_n^2)$ on a single $a_i$ as:

$$C = \sigma_n^2 I + \Phi A^{-1} \Phi^T = \sigma_n^2 I + \sum_{m \neq i} a_m^{-1} \phi_m \phi_m^T + a_i^{-1} \phi_i \phi_i^T = C_{-i} + a_i^{-1} \phi_i \phi_i^T.$$ 

where $C_{-i}$ is $C$, without the contribution of basis vector $i$. Continuing, we can write the inverse:

$$C^{-1} = C_{-i}^{-1} - \frac{C_{-i}^{-1} \phi_i \phi_i^T C_{-i}^{-1}}{a_i + \phi_i^T C_{-i}^{-1} \phi_i} C_{-i}^{-1} \phi_i$$

Now we can rewrite Eq.(4.3) as:

$$L(a, \sigma_n^2) = -\frac{1}{2} [N \log 2\pi + \log |C| + t^T C^{-1} t] -$$

$$\log a_i + \log(a_i + \phi_i^T C_{-i}^{-1} \phi_i) - \frac{(\phi_i^T C_{-i}^{-1} t)^2}{a_i + \phi_i^T C_{-i}^{-1} \phi_i} =$$

$$L(a_{-i}) + \frac{1}{2} [\log a_i - \log(a_i + s_i) + \frac{q_i^2}{a_i + s_i}] =$$

$$L(a_{-i}) + \ell a_i.$$ 

where

$$s_i \triangleq \phi_i^T C_{-i}^{-1} \phi_i \quad \text{(sparsity factor)}$$

$$q_i \triangleq \phi_i^T C_{-i}^{-1} t = \sigma_n^{-2} \phi_i^T (t - y_{-i}) \quad \text{(quality factor)}$$
CHAPTER 4. BAYESIAN COMPRESSIVE SENSING

Analysis on $\ell a_i$ has shown that there is a unique maximum with respect to $a_i$: 

$$a_i = \frac{s_i^2}{q_i^2 - s_i}, \quad \text{if } q_i^2 > s_i$$

$$a_i = \infty, \quad \text{if } q_i^2 \leq s_i$$

(4.5)

So, iteratively we may compute all $q_i$ and $s_i$ but is more efficient to maintain and update quantities;

$$S_m = \phi_m^T \sigma_n^{-2} I \phi_m - \phi_m^T \sigma_n^{-2} I \Phi \Sigma \Phi^T \sigma_n^{-2} I \phi_m,$$

$$Q_m = \phi_m^T \sigma_n^{-2} I t - \phi_m^T \sigma_n^{-2} I \Phi \Sigma \Phi^T \sigma_n^{-2} I t,$$

(4.6)

where quantities $\Phi$ and $\Sigma$ contain only the basis functions currently in the model. Quantities $S_m$ and $Q_m$ may be computed via an updating formula (described in Appendix of Tipping’s paper). It follows then that:

$$s_m = \frac{a_m S_m}{a_m - S_m}, \quad s_m = \frac{a_m Q_m}{a_m - Q_m}$$

4.2 The Algorithm

1. Initialize $\sigma_n^2$.

2. Initialize a single basis vector $\phi_i$ from Eq.(4.5):

$$a_i = \frac{s_i^2}{q_i^2 - s_i} = \frac{\|\phi_i\|^2}{\|\phi_i^T\|^2 - \sigma_n^2}$$

and set all other $a_m$ to infinity.

3. Compute

$$\Sigma = (A + \frac{1}{\sigma_n^2} \Phi^T \Phi)^{-1}, \text{ and}$$

$$\mu = \frac{1}{\sigma_n^2} \Sigma \Phi^T t \ (\text{which are scalars initially})$$

$$s_m = \frac{a_m S_m}{a_m - S_m}, \text{ where } S_m \text{ as in Eq.(4.6)}$$

$$q_m = \frac{a_m Q_m}{a_m - Q_m}, \text{ where } Q_m \text{ as in Eq.(4.6)}$$

4. Begin iteration:

(a) Select candidate $\phi_i$ from the set of all $M$.

(b) Compute $\theta_i \triangleq q_i^2 - s_i$.

If $\theta_i > 0$ and $a_i < \infty$ ($\phi_i$ is in the model), re-estimate $a_i$.

If $\theta_i > 0$ and $a_i = \infty$, add $\phi_i$ to the model with updated $a_i$. 
4.3. RESULTS AND FINDINGS

If $\theta_i \leq 0$ and $a_i < \infty$, delete $\phi_i$ from the model and set $a_i = \infty$.

(c) Update $\sigma_n^2 = \frac{||t-y||^2}{N-M+\sum_m a_m \Sigma_{mm}}$.

(d) Update $\Sigma, \mu$ using Eq.(4.2), and all $s_m, q_m$ using Eq.(4.5)-(4.6).

(e) If convergence: terminate; otherwise: repeat. Ex. convergence is achieved if the changes in $\log a_i$ of Step (4b) for all basis in the model are smaller than $10^{-6}$ and all other $\theta_i \leq 0$.

4.3 Results and findings

A series of experiments were held in order to reconstruct an MR phantom image from undersampled data, with the use of the BCS algorithm, as presented above. The reader may refer to the Appendix for a more detailed view on the Wavelet transform. Radial and spiral trajectories have been used in the experiments to simulate the measured data.

Figure 4.1: The different sampling trajectories used in the experiments. (a) The random sampling mask with density compensation towards the center of the $k$-space, (b) the radial sampling mask and (c) the spiral sampling mask.

A further comparison with the random sampling pattern (with density compensation towards the center of the $k$-space) has also been realized, although we have already mentioned that such a sampling pattern is not possible to be applied using current MRI scanners and technology. However, this random sampling pattern is widely used among the research community as a point of reference. The sampling patterns discussed above are shown in figure 4.1.
CHAPTER 4. BAYESIAN COMPRESSIVE SENSING

The spiral trajectory seems to be very close to the properties of the random sampling pattern and most importantly, it is feasible on current MR hardware. We insist on this kind of trajectory, as it is very adjustable and varied, as opposed to the radial pattern. Using the sampling masks of figure 4.1, the preliminary reconstruction results of the BCS algorithm, are shown in figure 4.2. The results presented on the figure do not seem to be very satisfying. More than 35% $k$-space sampling is needed for a mere descent reconstruction. But why is this happening? The formulation of the problem assumes that the measurements are taken through a random process and that Gaussian noise could exist in the measurement process.

However, after some research I have found that possible Gaussian noise in the R space image (this noise could also express coefficients very close to zero) is interpreted as Rician distribution noise in the $k$-space. This means that a further investigation on this matter is needed. Cases with sparser phantom images were also investigated. The algorithm seems to deal a lot better with sparser images. However, in real life MR images, it is extremely rare to deal with images sparser than the Shepp Logan phantom, and for this reason no such results are of interest here. Sparsifying transforms need to be incorporated in the reconstruction process, and even a combination

Figure 4.2: In this plot, the MSE of reconstructed images are shown for different $k$-space covering percentage. Solid lines correspond to BCS reconstruction results and the dotted lines to simple zero-filling reconstruction. The blue lines are for spiral sampling, the red for radial sampling and the green lines for random sampling with density compensation.
of sparsifying transforms, each one capturing different image features.

It is remarkable that the algorithm used only a 2-10% percentage out of the available measurements when the random sampling pattern has been used, for the experiments presented in figure 4.2. On the other hand, when radial sampling was used, about 60-80% of the available measurements was used. This means that, given the same amount of measurements the sampling modality to use is of extreme importance for the BCS formalism. Regarding the spiral trajectory, when samples were acquired in longer intervals along a spiral interleaf, as in Figure 4.3 (a), far less samples were needed in order to achieve same quality of reconstructed image with the pattern in Figure 4.3 (b).

![Figure 4.3: Spiral sampling of subfigure (a) tends to perform better when used by the BCS algorithms in term of reconstruction error, than sampling of subfigure (b). All experiments showed that when keeping fixed the sampling percentage of $k$-space, reconstruction process using pattern (a) produced sparser solutions of the same quality, possibly revealing the need of acquiring less samples when (a) is used.](image)

All the above matters need to be thoroughly examined to be of use for the research community. If specific ways of sampling $k$-space seems to need less samples in order to reconstruct images of the same quality, this would yield good news for the patient but more expertise needed by the doctors so as to handle the gradients designs.
Chapter 5

Conclusions - Future work

Different algorithms and approaches were studied in this thesis. The goal is to recover MR images from undersampled measurements of the raw ($k$-space) data. This specific topic, apart from being a challenge to the research community, has substantial importance: to overcome the time requirements during an MR scan procedure. Space limitations inside the magnet bore, together with movement limitations make it a difficult and inconvenient procedure for the patient under examination. Recall that an MR scan can last about 20 to 90 minutes!

Furthermore, the Magnetic Resonance as a diagnostic technique is very expensive. The specialized hardware (strong magnets, coils, etc.) is expensive to buy and to maintain. Expertise on the MR operations is needed by the doctors, a fact that additively contributes to a high cost per examination.

Researchers from both Signal Processing and Medical Imaging fields try their best to overcome these limitations. Algorithmic approaches that can reconstruct images of the same quality when acquiring only a percentage of 12% of the total raw MR data, is quite an achievement, as discussed in this thesis! The techniques in this work were exhaustively studied in the ideal case of non noisy data, however all techniques experimentally have been proven very robust even in the contribution of low noise. Studying signal acquisitions with high noise (due to electronic devices that interfere or due to movement artifacts) is not a priority since high noisy medical examinations are to be repeated by the doctors.

After studying the algorithms presented in this work, as well as the results achieved by each one of them, the $\ell_1$ minimization techniques seem to outperform $\ell_0$ frameworks. However, one could say that the $\ell_0$ techniques are the most promising ones since they are really fast and simple. The empirical complexity of $\ell_1$ minimization algorithms with quadratic constraints is of the order of $O(N^p)$ with $3 \leq p \leq 4$ whereas the $\ell_0$ approach, as proposed in this thesis, require $O(N \log N)$ time, with $N$ being the length of the reconstructed image signal. A more careful choice on the stoping criteria
and the sparsifying transforms is sure to achieve reconstruction quality comparable to the $\ell_1$ techniques.

A fast reconstruction algorithm is of course desirable, but keep in mind that our top priority is to lower the time that the patient has to stay still inside the magnet bore. Possibly some extra minutes of waiting in the lobby (for the reconstruction process to be completed) are not much of the essence as compared to a long wait inside the magnet bore in order to acquire more data samples.

Compressive Sensing exploits the natural sparsity and compressibility of MR images and achieves great results from highly undersampled data. Sampling acquisition schemes should be carefully designed and exploited, taking into account real life hardware limitations of the MR technology. The CS strategy can be combined with other acceleration methods that exploit different kinds of redundancy. For example, constraining the image in the minimization problem to be real, effectively implies phase constraints in partial $k$-space with the CS reconstruction. In a similar way, methods like SENSE [62] can be incorporated into the problem formulation, including the coil sensitivity information such as the case of multiple coil receivers. In general, any other prior knowledge or information on the image or the sampling scheme, that can be expressed as a convex constraint, can be incorporated in the formulation!

Everything mentioned above needs careful and extra examination. The fast $\ell_0$ reconstruction technique needs to be carefully designed in order to achieve reconstruction results of the same quality as the $\ell_1$ minimization techniques. Constraints capturing the features of MR images should be taken into account. The Bayesian formalism presented in Section 4 needs to be further studied. It is a powerful technique that can provide quantitative estimations of our belief of the reconstruction process. It is an approach that may also provide an estimation of the possible noise involved in the measurement process. Incorporating sparsity constraints or even a combination of sparsity constraints on the image can surely improve a lot the performance of the BCS algorithm. The BCS formulation can also provide quality information of the sampling trajectory on the fly. Every new (radial or spiral) 'line' of data, when incorporated in the reconstruction process, can instantly reveal whether this piece of data information was useful for the reconstruction process. This knowledge could be exploited for optimizing the gradient designs for specific kinds of images - body areas scanned. Some particular $k$-space trajectories tend to need far less data samples than others, for specific image contents, in order to reconstruct images of equivalent quality. Thorough examination on all these matters is sure to reveal very interested outcomes for both the Compressive-Sensing and the Magnetic-Resonance communities and provide further directions of study. And so, the research continues...
6.1 Measurement Matrix \( \Phi \)

Consider the \( \mathbb{R} \)-space 2-D image \( m \). In MRI we do not have knowledge of the \( \mathbb{R} \)-space image itself. We can obtain information for \( m \), sampling its \( k \)-space representation or in other words, the 2-D Fourier transform of \( m \).

Here, we recast the 2-D Fourier operator into a matrix multiplication, exploiting properties of the Kronecker product.

If \( m \) is a \( n_y \times n_x \) 2-D image in the \( \mathbb{R} \)-spaces, then:

\[
\Omega_{C_x} : \text{the } n_x \times n_x \text{ complete valued Fourier matrix}
\]
\[
\Omega_{C_y} : \text{the } n_y \times n_y \text{ complete valued Fourier matrix}
\]

Now, the 2-D Fourier transform of image \( m \) would be:

\[
\text{Fourier}_{2D}\{m\} = \Omega_{C_y}m_{2D}\Omega_{C_x}^T
\] (6.1)

We know from (Harville, 1999) that given three matrices \( A, B \) and \( C \)

\[
\text{vector}(ABC) = (C^T \otimes A)\text{vector}(B)
\]

where \( \otimes \) denotes the Kronecker product (6.2)

If we denote \( y \) the 2-D Fourier transform of \( m \), i.e. the \( k \)-space data, and apply Eq 6.3 in our problem statement, we obtain

\[
\text{vector}(y) = \Omega_{C_y} \otimes \Omega_{C_x} \text{vector}(m) \Rightarrow \text{vector}(y) = \Phi_F \text{vector}(m)
\] (6.3)

\( \text{vector}(y^T) \) represents the entire \( k \)-space domain data, and \( \Phi_F \) the sampling operator. Now, let us continue with the \( \Phi_F \) analysis in order to understand the sampling procedure.
\[ \Phi_F = \Omega_{Cy} \otimes \Omega_{Cx} \Rightarrow \]
\[ \Phi_F = (\Omega_{yReal} + i\Omega_{yImag}) \otimes (\Omega_{xReal} + i\Omega_{xImag}) \Rightarrow \]
\[ \Phi_F = (\Omega_{yReal} \otimes \Omega_{xReal}) - (\Omega_{yImag} \otimes \Omega_{xImag}) + i \left[ (\Omega_{yReal} \otimes \Omega_{xImag}) + (\Omega_{yImag} \otimes \Omega_{xReal}) \right] \Rightarrow \]
\[ \Phi_F = \Omega_{Real} + i\Omega_{Imag}. \quad (6.4) \]

\( \Phi_F \) is of size \((n_x \times n_y)\times(n_x \times n_y)\). Each row of \( \Phi_F \) when multiplied with vector\((m)\) results in one \(K-space\) measurement, and thus, a single \(y_i\) coefficient. In order to represent the incomplete measurements of \(K-space\), rows of \( \Phi_F \) are removed. These removed rows correspond to the \(K-space\) measurements left out. The complete measurement matrix \( \Phi_F \) becomes matrix \( \Phi_k \) of size \((M \times (n_x \times n_y))\), and the \(k-space\) measurements are of size \((M \times 1)\):
\[ vector(y) = \Phi_k vector(m) \quad (6.5) \]

From the Compressive Sensing point of view, we need to introduce a level of randomness into the measurement matrix. This is achieved via the sampling patterns as discussed in section 2.

### 6.2 Transform Matrix \( \Psi \)

The idea of Compressive Sensing is best applied in sparse signals, as demonstrated in section CS ... This being the case, the following sparsifying transforms have been introduced into the experiments of this thesis.

Consider the \(\mathbb{R}-space\) 2-D image \(m\) and a sparsifying transform \( \Psi \). Then, signal \(m\) can be recast into its sparse representation \(w\) via the transform matrix \( \Psi \) as follows:
\[ m = \Psi w \quad \text{or} \quad w = \Psi^{-1} m \quad (6.6) \]

Haar Wavelet has been used as a sparsifying Transform. It is the simplest of the wavelet transforms (see Appendix .. for more details).

### 6.3 Wavelets Analysis: the basic idea

Wavelets are oscillating, wave-like signals with finite duration and zero mean value and usually are non-symmetrical. They are represented in a time-scale domain, where scales is somewhat related to the frequency notion of the Fourier Transform. Large scales correspond to the "big picture" of a signal(ex. the big basic shapes of an image),
6.3. WAVELETS ANALYSIS: THE BASIC IDEA

including contrast information, whereas small scales correspond to fine details such as resolution information. In wavelet analysis a fully scalable modulated window is shifted along the signal and we calculate the spectrum for every position. Then we repeat this process, using a longer or shorter version of the specified window. In the end the result would be a collection of time-frequency representations of the signal, all with different resolution. All these different window-signals are the wavelets. They are generated using scaling and translation onto a single basic wavelet function, called the mother wavelet, $\psi(t)$:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right)$$ (6.7)

with $s$ being the scale factor, and $\tau$ the translation factor. Energy normalization across the different scales is expressed with coefficient $s^{1/2}$. See Figure 6.1 for an example.

![Figure 6.1: Different translation examples (b-c) and scaling examples (e-f) of the basic-mother wavelet for Daubechies 9.](image)

The Continuous Wavelet Transform (CWT) is written as in Eq. 6.8

$$C(s, \tau) = \int f(t) \psi_{s,\tau}^*(t) dt$$ (6.8)

, where function $f(t)$ is decomposed into a set of basis functions (i.e. the wavelets), $\psi_{s,\tau}(t)$.

CWT signifies that it operates at every different scale up until an upper bound determined for a specific application. It also uses all possible shifts-translation across the full signal of interest. The steps to follow for a CWT are described below, and shown in Figure 6.2.

1. Take a wavelet and compare it to a section at the start of the original signal.
2. Calculate a number, C, that represents how closely correlated the wavelet is with this section of the signal. The higher C is, the more the similarity. More precisely, if the signal energy and the wavelet energy are equal to one, C may be interpreted as a correlation coefficient. Note that the results will depend on the shape of the wavelet you choose.

3. Shift the wavelet to the right and repeat steps 1 and 2 until you’ve covered the whole signal.

4. Scale (stretch) the wavelet and repeat steps 1 through 3.

5. Repeat steps 1 through 4 for all scales.

![Figure 6.2: Graphical illustration of the general CWT steps.](image)

Similarities between parts of the signal and wavelets are indicated with the wavelet coefficients. For a given wavelet scale, the higher the coefficients, the more the resemblance with the signal. The reader may find Figure 6.3 more convenient to better understand the relations.
6.3. WAVELETS ANALYSIS: THE BASIC IDEA

(a) No rotation

(b) Rotation 1

(c) Rotation 2

(d) Rotation 3

Figure 6.3: An example of a CWT transform, using wavelet Daubechies 7. Time or space, depending on the nature of the signal, is on the x-axis. The CWT coefficients are on the z-axis: high coefficient values are shown with white color and low values are shown with dark colors. The coefficients are calculated for every different wavelet scaling. The 3-D rotation of plot (a) is shown in plots (b-d), to better understand the meaning of the colors.

However, the Wavelet transform in its continuous form is not practical. It is not possible to calculate the transform using continuous shifting (translation) and scaling. This is the reason that Discrete Wavelet Transform (DWT) has been introduced. DWT uses only a subset of the possible scales and positions, and it has been proven that using dyadic scales and positions (sampling the frequency and the time axis based on powers of two), the resulting wavelet analysis is as accurate as a continuous one, and of course, much more efficient. Note that the output of a DWT is a continuous signal, and the term discrete refers only to the scales and positions. Mallat, 1998 developed an efficient way to implement this scheme, with the use of filters, introducing a fast wavelet transform. Approximations and Details are the two main characteristics. Approximations refer to low-frequency components (i.e., high scales), wherein lies the most of a signal’s identity. The Approximations are the high...
frequency components (i.e. low scales).

In this DWT analysis, filters are used to create a distinction between the Approximation and the Detail coefficients and the original signal, \( S \), passes through two complementary filters and emerges as two signals (see Figure 6.4). The filtering process will produce two times more coefficients, in total, than the original signal samples. It has been shown that downsampling the filtered signals by a factor of 2, we lose no information, and reduce the total number of produced coefficients down to the original signal length.

![Figure 6.4: One stage Filtering on 1D signal. A low-pass filter is applied, and coefficients \( cA \) are produced after dyadic downsampling. In the same time, a high-pass filter is applied, and coefficients \( cD \) are produced after dyadic downsampling.](image)

The high-pass filtered signal contains the smallest details we are interested in and we could stop there, However, the low-pass part still contains some details and therefore we can split it again and again, until we are satisfied with the number of bands we have created. In this way we have created an iterated filter bank. The process of splitting the spectrum is graphically displayed in Figure 6.5. The outputs of the different filter stages are the wavelet- and scaling function transform coefficients.

![Figure 6.5: Multiple Level Decomposition on 1D signal. A low-pass filter is applied, and coefficients \( cA_1 \) are produced after dyadic downsampling. In the same time, a high-pass filter is applied, and coefficients \( cD_2 \) are produced after dyadic downsampling. The process may be repeated again and again, for the coefficients produced by the low-pass filtering.](image)

You may happen to observe that the actual lengths of the Detail and Approxima-
tion coefficient vectors are slightly more than half the length of the original signal. This has to do with the filtering process, which is implemented by convolving the signal with a filter. The convolution "smears" the signal, introducing several extra samples into the result.

In order to reconstruct the original signal using the Detail and Approximation coefficients, an inverse procedure must be followed: we need to apply appropriate reconstruction filters, after upsampling the signal (lengthen the signal components by inserting zeros between samples). The choice of filters is crucial in achieving perfect reconstruction of the original signal. The decomposition and reconstruction filters are the tools that define the very form of the wavelet (Daubechies 1, 2..., Symlets 1, 2..., etc.) that analyzes our original signal. If we implement the wavelet transform as an iterated filter bank, we do not have to specify the wavelets explicitly!

The same idea may be also applied for a 2D signal. The standard solution consists of alternating one decomposition by rows and another one by columns, iterating only on the low-pass sub-image. An one-stage analysis outputs approximation coefficients $cA$, together with three detail coefficient matrices $cD_h$, $cD_v$, and $cD_d$ (horizontal, vertical, and diagonal, respectively). Further decomposition may be applied to the $cA$ coefficients, using the same analysis. See Figure 6.6: A low-pass filter (Lo_D) is applied on the rows of the original signal followed by downsampling and a second low-pass filter, so as to produce coefficients $cA_{j+1}$. A low-pass filter (Lo_D) is applied on the rows of the original followed by downsampling and a high-pass filter (Hi_D), so as to produce horizontal coefficients $cD_{j+1}^{(h)}$. A high-pass filter is applied on the rows of the original followed by downsampling and a low-pass filter, so as to produce vertical coefficients $cD_{j+1}^{(v)}$. A high-pass filter is applied on the rows of the original followed by downsampling and a second high-pass filter, so as to produce diagonal coefficients $cD_{j+1}^{(d)}$. The process may be repeated again and again, on the coefficients produced by low-pass filtering.
CHAPTER 6. APPENDIX

Figure 6.6: Multiple Level Decomposition using 2D DWT. Iteratively decompose rows and columns using high and low-pass filtering. Decomposition is done only on the low-pass sub-image of every step.

Using a brain image $128 \times 128$ and perform a 2D DWT, using the Haar Wavelet with one and two stage (level) filtering, one may see the output coefficients in Figure 6.7.
6.3. WAVELETS ANALYSIS: THE BASIC IDEA

(a) One level decomposition.

(b) Three level decomposition.

Figure 6.7: Multiple Level Decomposition using 2D DWT: an illustrative example using a transverse view of a 128 × 128 brain image. The image is iteratively decomposed and downsampled, as explained earlier in Figure 6.6 with one-level (a) and three-level decomposition (b).
However, it is not the scope of this thesis to investigate any further the decomposi-
tion and reconstruction analysis of wavelets.

The Haar Wavelet Transform

The Haar wavelet is now recognized as the first known wavelet and is also the sim-
plest possible wavelet. The technical disadvantage of the Haar wavelet is that it is
not continuous, and therefore not differentiable. This property can, however, be an
advantage for the analysis of signals with sudden transitions. The Haar wavelet has
a mother wavelet function \( \psi(t) \), as described below:

\[
\psi(t) = \begin{cases} 
1, & 0 \leq t < \frac{1}{2} \\
-1, & \frac{1}{2} \leq t < 1 \\
0, & \text{otherwise}
\end{cases}
\]  

(6.9)

It has simple decomposing high-pass filter : \([-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] \) and low-pass filter: \([\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] \). The corresponding reconstruction high-pass and low-pass filters are : \([\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}] \) and 
\([\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] \). We may also express the above in matrix form. Lets see a 1D example, applying Haar Transform, \( H \), on a 2 element vector \( x = (x_1, x_2) \):

\[
\begin{pmatrix}
C_1 \\
C_2
\end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\
1 & -1
\end{pmatrix} \begin{pmatrix} x_1 \\
x_2
\end{pmatrix}
\]  

(6.10)

Note that \( H \) is an orthonormal matrix because its rows are orthogonal to each
other (their dot products are zero) and they are normalized to unit magnitude. In
the 2D case, \( x \) and \( H \), are \( 2 \times 2 \) matrices, and we may transform first the columns
of \( x \), by premultiplying by \( H \), and then the rows of the result by postmultiplying by
\( H^T \), and:

\[
C = H x H^T = H \begin{pmatrix} x_1 \\
x_2 \\
x_3 \\
x_4
\end{pmatrix} H^T
\]

\[
= \frac{1}{2} \begin{pmatrix} x_1 + x_2 + x_3 + x_4 & x_1 - x_2 + x_3 - x_4 \\
x_1 + x_2 - x_3 - x_4 & x_1 - x_2 - x_3 + x_4
\end{pmatrix}
\]  

(6.11)

Explaining Eq 6.11,

- \( x_1 + x_2 + x_3 + x_4 \) corresponds to 2D low-pass filter (low-pass after low-pass) -
  component \( cA \).

- \( x_1 - x_2 + x_3 - x_4 \) corresponds to horizontal high-pass and vertical low-pass filter
  - component \( cD^h \).

- \( x_1 + x_2 - x_3 - x_4 \) corresponds to horizontal low-pass and vertical high-pass filter
  - component \( cD^v \).

- \( x_1 - x_2 - x_3 + x_4 \) corresponds to 2D high-pass filter (high-pass after high-pass)
  - component \( cD^d \).
Bibliography


[27] Qu Xiaobo, Cao Xue, Guo Di, Hu Changwei, ChenZhong, "Compressed sensing MRI with combined sparsifying transforms and smoothed ℓ0 norm minimization" ICASSP, Dallas, TX, USA, 2010 pp. 626 - 629.


[32] Lustig et al. 13th ISMRM 2004:p605


[34] M. Lustig, D.L. Donoho, J.M. Pauly "Rapid MR Imaging with Compressed Sensing and Randomly Under-Sampled 3DFT Trajectories" ISMRM '06


