COMPRESSIVE SENSING METHODS FOR MULTI-CONTRAST MAGNETIC RESONANCE IMAGING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

ALPER GÜNGÖR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SEPTEMBER 2017

Approval of the thesis:

COMPRESSIVE SENSING METHODS FOR MULTI-CONTRAST MAGNETIC RESONANCE IMAGING

submitted by **ALPER GÜNGÖR** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** _____

Prof. Dr. Fatoş T. Yarman Vural
Supervisor, **Computer Engineering Dept., METU** _____

Assist. Prof. Dr. Tolga Çukur
Co-supervisor, **Electrical and Electronics Eng. Dept., Bilkent University** _____

Examining Committee Members:

Assoc. Prof. Dr. Murat Manguoğlu
Computer Engineering Department, METU _____

Prof. Dr. Fatoş T. Yarman Vural
Computer Engineering Department, METU _____

Assist. Prof. Dr. Tolga Çukur
Electrical and Electronics Eng. Dept., Bilkent University _____

Assist. Prof. Dr. Emre Akbaş
Computer Engineering Department, METU _____

Assist. Prof. Dr. Emine Ülkü Sarıtaş
Electrical and Electronics Eng. Dept., Bilkent University _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ALPER GÜNGÖR

Signature :

ABSTRACT

COMPRESSIVE SENSING METHODS FOR MULTI-CONTRAST MAGNETIC RESONANCE IMAGING

Alper Güngör

M.S., Department of Computer Engineering

Supervisor : Prof. Dr. Fatoş T. Yarman Vural

Co-Supervisor : Assist. Prof. Dr. Tolga Çukur

September 2017, 116 pages

Compressive sensing (CS) is a signal processing tool that allows reconstruction of sparse signals from highly undersampled data. This study investigates application of CS to magnetic resonance imaging (MRI). In this study, first, an optimization framework for single contrast CS MRI is presented. The method relies on an augmented Lagrangian based method, specifically alternating direction method of multipliers (ADMM). The ADMM framework is used to solve a constrained optimization problem with an objective function consisting of a linear combination of the total variation on the magnitude image and the ℓ_1-norm. Then, a fast implementation is derived for MRI, which requires only two FFT operations per iteration. Second, for better exploitation of sparsity, a joint reconstruction method for multi-contrast CS MRI is presented. This method uses non-convex group-ℓ_p-sparsity as well as joint total variation as objective functions. Finally, a joint dictionary learning based method for finding the

sparsifying transformation along with the image is presented. The sparsifying transformation reconstructed by the method enforces group sparse representation on all contrast images. All the proposed methods are compared quantitatively and qualitatively with previous methods that exist in the literature using both experimental in-vivo and simulated datasets. The effectiveness of the ADMM for single contrast reconstruction is demonstrated over other single contrast methods. Then, the advantages of using joint reconstruction is discussed and demonstrated. Although dictionary learning based method require high computational cost, it presents benefits in terms of image quality is shown.

Keywords: Compressive Sensing, Magnetic Resonance Imaging, Sparsity, Dictionary Learning, Multi-Contrast

ÖZ

MANYETİK REZONANS GÖRÜNTÜLEME İÇİN SIKIŞTIRILMIŞ ALGILAMA YÖNTEMLERİ

Alper Güngör

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. Fatoş T. Yarman Vural

Ortak Tez Yöneticisi : Yrd. Doç. Dr. Tolga Çukur

Eylül 2017 , 116 sayfa

Sıkıştırılmış Algılama (SA) daha az ölçüm kullanılarak seyrek bir sinyalin isabetli bir şekilde gerikazanmak için kullanılan bir sinyal işleme tekniğidir. Bu çalışma SA'nın MRGye uygulanışını incelemektedir. Bu çalışmada, öncelikle, SA MRG için bir optimizasyon çatısı sunulmaktadır. Metot, genişletilmiş Lagrange tabanlı bir metot olan yön değiştiren çarpanlar yöntemine (YDÇY) dayanmaktadır. Bu çatı, hedef fonksiyonları olarak görüntünün büyüklüğü üzerinde toplam değişinti ve ℓ_1-norm fonksiyonlarının lineer kombinasyonunu içeren bir kısıtlı bir optimizasyon problemini çözmek için kullanılmıştır. Ardından, bu çatının MRG için her yinelemede iki FFT işlemi ile çözülmesine olanak sağlayan hızlı uygulaması geliştirilmiştir. İkinci olarak, seyreklikten daha çok yararlanmak için, çoklu-kontrast MRG için bir ortak geri-kazanım metodu sunulmaktadır. Bu metot hedef fonksiyonu olarak konveks olmayan grup-ℓ_p-seyrekliği fonksiyonunun

yanı-sıra ortak toplam değişinti fonksiyonunu kullanmaktadır. Son olarak, seyreklik dönüşümü ile birlikte görüntünün bulunmasına olanak sağlayan bir ortak sözlük öğrenimi tabanlı metot sunulmuştur. Yöntem tarafından bulunan seyreklik dönüşümü, bütün kontrast görüntüler üzerinde grup seyrek temsili sağlamaktadır. Bütün önerilen yöntemler literatürdeki önceki yöntemlerle hem deneysel hem de benzetim veri kümeleri kullanılarak nicelik ve kalite açılarından kıyaslanmıştır. Tekli kontrast için YDÇY'nin etkililiği diğer tekli kontrast yöntemler üzerinden gösterilmiştir. Daha sonra, ortak geri-kazanımın avantajları tartışılmış ve gösterilmiştir. Sözlük öğrenimi tabanlı yöntem yüksek işlem gücü gerektirse de, bu yöntemin görüntü kalitesi açısından avantajları gösterilmiştir.

Anahtar Kelimeler: Sıkıştırılmış Algılama, Manyetik Rezonans Görüntüleme, Seyreklik, Sözlük Öğrenimi, Çoklu Kontrast

To my family

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor Prof. Dr. Fatoş Tünay Yarman Vural and co-advisor Assist. Prof. Dr. Tolga Çukur. I feel very fortunate to work under their guidance. They have always motivated and encouraged me during this research.

I also would like to thank Toygan Kılıç and Assist. Prof. Dr. Emine Ülkü Sarıtaş for supplying multi-contrast experimental data used in this thesis. The data played a crucial role while demonstrating this work.

I would like to thank Dr. H. Emre Güven and Dr. Emre Kopanoğlu for their support, invaluable comments and contributions to this thesis. I have always enjoyed taking suggestions from them. I also would like to thank Dr. Can Barış Top and Mert Kalfa for letting me benefit from their academic experience. They all have been valuable friends. They have also checked and edited this study.

I am deeply grateful to my mother, Hamide Güngör, and my family. They have always been there for me, and supported me in every way they can.

Last but not least, I would like to express my deepest gratitude to my beautiful wife, Zülal Özdemir Güngör. She put up with me from the very beginning until the end of this study. I am very lucky to have her beside me.

I acknowledge support of TÜBİTAK (The Scientific and Technological Research Council of Turkey) within the scope of 2210/A scholarship.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ADMM	Alternating Direction Method of Multipliers
ALM	Augmented Lagrangian Method
BCS	Blind Compressive Sensing
BRIP	Block Restricted Isometry Property
CS	Compressive Sensing
FCSA	Fast Composite Splitting Algorithm
FFT	Fast Fourier Transformation
fMRI	Functional Magnetic Resonance Imaging
FOV	Field of View
GSMRI	Group Sparsity Magnetic Resonance Imaging
JTV	Joint Total Variation
LASSO	Least Absolute Shrinkage and Selection Operator
MAP	Maximum a Posteriori
mmE	Mean Magnitude Error
MOD	Method of Optimal Directions
MP	Matching Pursuit
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
OMP	Orthogonal Matching Pursuit
PD	Proton Density
pdf	Probability Density Function
pSNR	peak Signal to Noise Ratio
RF	Radio Frequency
RGB	Red Green Blue
RIC	Restricted Isometry Constant
RIP	Restricted Isometry Property
RMSE	Root Mean Squared Error
SAR	Synthetic Aperture Radar
SPGL1	Spectral Projected Gradient-L1

SSIM	Structural Similarity Index Measure
SVD	Singular Value Decomposition
TV	Total Variation

CHAPTER 1

INTRODUCTION

Magnetic Resonance Imaging (MRI) is a widely utilized imaging modality due to its relatively high resolution and contrast for imaging soft tissue. In the past 30 years, it has found use in various applications, including diagnosis, treatment monitoring and interventional guidance. For many applications, multiple images with different contrasts are required. However, the imaging process is rather slow, and multi-contrast imaging leads to long scan times, which may become infeasible impractically long scans. A recent approach to accelerated imaging that has attracted broad attention is compressed sensing (CS) [37]. CS is a signal processing technique that is used to reconstruct signals using fewer measurements by exploiting sparsity of the signal of interest in a transform domain [16, 25]. However, due to high computational requirements of compressive sensing reconstruction algorithms, its use in clinical practice has remained limited.

Appropriate selection of the sparsifying transform domain is critical in successful applications of CS. Recent methods in the literature rely on pre-chosen sparsifying transform domains such as finite differences, wavelets, total variation etc. for image reconstruction [10, 22, 23, 34, 37]. While these methods are shown to be moderately successful, the transform domains have to be application specific, and selected heuristically by the end user. In order to avoid this problem, another trend in CS is to use a transform domain that is inferred from data using dictionary learning [51]. Dictionary learning is a signal processing tool which can be used to infer sparse representation of a group of signals in some transform domain. However, it involves solving a highly non-convex problem. Furthermore, exploiting any type of apriori knowledge on the data is very important for finding the best solution. While several applications of dictionary learning to

MRI exist [4, 51, 52, 53], a study on multi-contrast dictionary learning for CS MRI has not been extensively carried out.

In this thesis, we deal with the problem of image reconstruction from undersampled multi-contrast MRI data. We first describe the basic principles of MRI and CS. Next, we propose three techniques for image reconstruction. The first technique is a fast and efficient single contrast image reconstruction technique, and it reconstructs each contrast image separately. We propose a novel objective function, and analyse the convergence of the proposed method. The second technique is an extension of the first technique such that it allows reconstruction of multi-contrast images using joint information across contrasts. The last technique relies on dictionary learning that works for multi-contrast images taking structural similarity of different contrasts into consideration. It learns the structural similarity across contrasts while simultaneously reconstructing images. We describe each method extensively and discuss their fast and efficient implementations. Finally, we compare each method with five methods in the literature and with each other in terms of both qualitative and quantitative metrics.

The novelty of this thesis can be summarized as follows:

- We propose an efficient framework for single contrast MRI reconstruction. We propose an efficient augmented Lagrangian based method CS MRI reconstruction. We use novel single contrast objective functions suitable for brain images and give efficient implementation of the functions within the proposed framework. The algorithm requires using well-known "proximal mapping functions" associated with each objective function. We apply gradient transformation on the magnitude of the image rather than the complex-valued image. This requires deriving the associated proximal mapping functions for these magnitude operations. We rigorously derive a generic method for extending the proximal mapping of any real-valued objective function for complex-valued imaging.

- We propose an efficient framework for joint multi-contrast MRI reconstruction. The algorithm is an augmented Lagrangian based method for

joint reconstruction of multi-contrast CS MRI. We use novel joint objective functions suitable for multi-contrast brain images and give efficient implementations of the functions within the proposed framework. Similar to single-contrast case, the functions are extensions of real-valued functions for magnitude complex-valued imaging. We also use the previously proposed theorem for deriving the associated proximal mapping.

- We propose a joint dictionary learning based image reconstruction method for multi-contrast MRI. The method involves finding both the transform in which the image is sparse and the image itself from only the measurements. Moreover, we use a single transform for multi-contrast images, and assume joint sparsity.

- We compare both previous and proposed algorithms in terms of both qualitative and quantitative metrics. We show that each proposed algorithm outperforms the reference methods in the literature. Moreover, we compare ADMM based proposed methods to each other. In the cases where multi-contrast data is available, exploitation of multi-contrast data results in better reconstruction quality, as well as faster reconstruction. We also compare dictionary learning based methods to the proposed augmented Lagrangian based methods. Although dictionary learning results in higher quality reconstructions, it requires considerably more computation time. The choice between augmented Lagrangian based methods and dictionary learning based methods presents trade-off between computation time and desired image quality.

Rest of this thesis is outlined as follows: First, chapter 2 gives background information on MRI and CS. Then, chapter 3 describes a novel method for individual reconstruction of CS MRI. Chapter 4 extends the method for joint reconstruction of multi-contrast MRI and gives implementation specifics. Chapter 5 describes a novel joint dictionary learning based CS method for multi-contrast MRI. Finally, chapter 6 demonstrates the proposed methods, and compare the methods

to previous reference algorithms.

1.1 Notation

The following notation is used throughout this thesis. We denote vectors with small bold letters such as $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$. Subscripts denote the iteration count, superscripts denote a sub-block within a vector, where $\mathbf{x}_k^{(i)}$ denotes the i-th block within vector $\mathbf{x}$ at iteration k. For multi-block structures such as i-th channel t-th block, we use $\mathbf{x}_k^{(i,t)}$. We use bold capital letters for matrices or transformations such as $\mathbf{A}$, $\mathbf{W}$. $\mathbf{x}[i]$ denotes the i-th element within vector $\mathbf{x}$, while $\mathbf{x}[i,j]$ denotes the element corresponding to i-th row, j-th column of the image stacked as a vector ($(iN_v + j)$-th element of $\mathbf{x}$ where each row has N_v elements). $\mathbf{x}^H$, $\mathbf{A}^H$ denotes the Hermitian transpose (or conjugate transpose) of $\mathbf{x}$ and $\mathbf{A}$, respectively. This operator is expressed in some contexts as $\mathbf{x}^*$ and $\mathbf{A}^*$.

CHAPTER 2

BACKGROUND ON MAGNETIC RESONANCE IMAGING AND COMPRESSED SENSING

We deal with the problem of image recovery for compressed multi-contrast magnetic resonance imaging. In this chapter, we provide a background to the reader, and discuss the basics related to magnetic resonance imaging, compressive sensing, and some of the optimization algorithms used. First, we discuss physics of Magnetic Resonance Imaging (MRI), next we discuss conventional image reconstruction methods for MRI. After that, we discuss the theory behind compressed sensing (CS), and explain some basic applications. We also give information on the key points for a successful CS application, and how to achieve it. We then give an overview on the recent work for optimization algorithms applied to solve CS, and discuss the strong suits of each. Finally, we discuss the theory behind the proposed method, a version of augmented Lagrangian method (ALM) called alternating direction method of multipliers (ADMM).

2.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging is a non-invasive imaging technology that has gained attention due to its high resolution and contrast imaging of soft tissues. The imaging process does not include ionizing radiation as in x-ray and is therefore safe to be repeated for humans. MR scanner measures the magnetic response of the molecules in a field of view. In MRI, an area is first excited then measured. Although the fundamental principles remain the same, different types of excitation and measurements result in images containing different information. For example, while Magnetic Resonance Angiography (MRA) produces images with enhanced blood vessels, functional Magnetic Resonance Imaging

(fMRI) produces blood-oxygen-level dependent contrast images, which can then be used for brain-decoding [3, 30]. Among many applications, we are primarily interested in structural brain imaging.

The imaging process is based on a phenomena called Nuclear Magnetic Resonance (NMR). In NMR, nuclei in a magnetic field absorb and re-emit electromagnetic radiation. Initially, the net magnetic field vector in a nuclei is zero. However, when a very strong magnetic field denoted by $\mathbf{B}_0$ is applied, it polarizes the nuclei such that magnetic dipoles align with the external magnetic field (either in parallel or antiparallel direction to $\mathbf{B}_0$, however there is a slight preference for parallel direction and hence the polarization occurs. Then, a radio frequency (RF) pulse, which is basically a time-varying magnetic field denoted by $\mathbf{B}_1(t)$, is applied to select and excite or **tip** these magnetic dipoles to the transverse plane. Hence, while the net magnetic field in a selected slice points towards the transverse direction, all other spins point towards the direction of $\mathbf{B}_0$. Here, magnetic field in transverse direction denoted by $\mathbf{B_{xy}}$ can be used to determine the tissue composition of the excited area. To gather information on spatial position of the various tissues within the excited slice, spatial encoding is required. Spatial encoding is performed via three coils that create magnetic field gradient in three orthogonal directions. These coils are used to create spatially varying magnetic fields, $G(\vec{\mathbf{r}})$, in all three dimensions, which provides position-encoding in the received signal. Due to the applied gradient, the net magnetic field at each position is slightly rotated in transverse plane. After application of $\mathbf{B}_1(t)$ is completed, the magnetic field in the tipped region slowly re-aligns with the constant magnetic field, $\mathbf{B}_0$. The magnetization in transverse direction, $\mathbf{s_r}(t)$ is measured using a coil. The measured signal has two dimensions, x and y. To simplify analysis, complex numbers are used to denote the coordinate system. Hence, the received signal and the reconstructed image are both complex-valued. The received signal has the form

$$\mathbf{s_r}(t) = \int_R m(\vec{\mathbf{r}}) e^{-i2\pi \vec{k}(t)\cdot\vec{\mathbf{r}}} d\mathbf{r}, \tag{2.1}$$

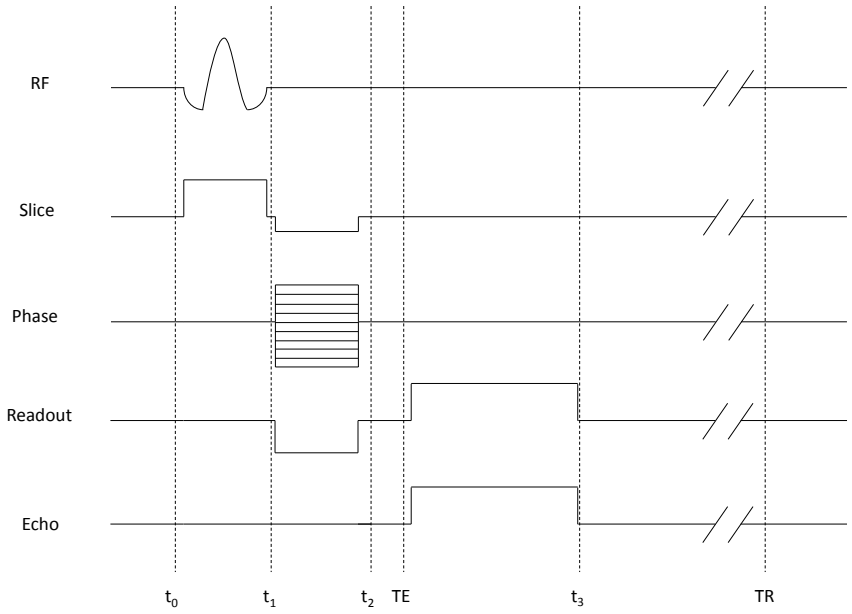
where $\vec{k}(t)$ is proportional to $\int_0^t G(\vec{s})d\vec{s}$, and $m(\vec{\mathbf{r}})$ denotes the magnetic response of the excited volume. Note that $\vec{k}(t)$ provides a different rotation for each position, which provides spatial encoding. Eq. (2.1) basically dictates the imaging

process. To sum up, first a pre-designed RF pulse $\mathbf{B}_1(t)$ along with gradients $G(\vec{\mathbf{r}})$ are used simultaneously to excite or tip an area of interest. Then, the response is measured after waiting for some time. Imaging algorithm simply uses the measured signal to create an image.

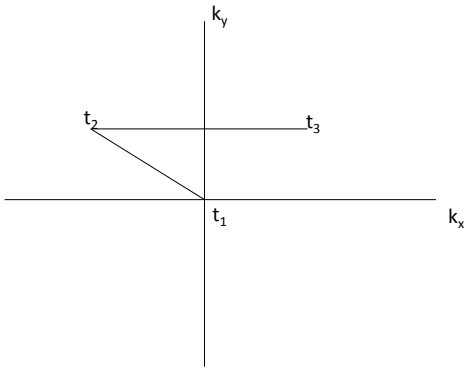
For linear gradient fields, the equation is exactly the Fourier transformation. Hence, the signal becomes 2D spatial Fourier transform of the field of view (FOV) [45]. The frequency domain representation is also called the **k-space**. Depending on the change in all x,y, and z components of gradient G over time, different sampling schemes can be applied. Hence, simply put, one can say that MR scanner samples frequency domain representation of magnetic response of materials inside a FOV [45].

Fig. 2.1 (a) shows the timing of signals for gradient echo sequence Cartesian sampling, (b) depicts the trajectory in the k-space. Before sampling process begins, k-space initial position is set. At time t_0, the integral of the gradients, hence the corresponding wave numbers k_x, k_y, and k_z are all 0. Slice selection gradient is set to a constant value such that the same k_z, wave number of the slice to be sampled in the k-space is sampled. During this time, an RF excitation pulse is applied such that the magnetic field inside the selected/excited slice is tipped 90^o. At this time k-space position is at t_1 and k_z is set. Then at t_1, a phase encoding gradient is applied to select the k_y, frequency value of the line to be sampled in the k-space. After the application the line is selected such that k_y is set. Finally at t_2, frequency encoding gradient is applied. Frequency encoding gradient is first set to a negative value such that its integral is set to negative of the half bandwidth in that direction. While k_x, frequency value to be sampled within the line in the k-space, is increased, the signal is sampled.

The process described above is repeated since all of k-space can not be filled in a single scan. The time waited before sampling process starts is called echo time, **TE**, and the period of the repetition is called repetition time, **TR**. These two parameters affect the reconstructed image. Each tissue is represented by three basic parameters [7], which are longitudinal relaxation time $\mathbf{T_1}$, transverse relaxation time $\mathbf{T_2}$, and proton density (**PD**) $\mathbf{m_0}$. **PD** is the magnitude of



(a)



(b)

Figure 2.1: (a) Gradient echo sequence timing (b) k-space trajectory.

the magnetization in a specific voxel. Initially it is aligned with $\mathbf{B}_0$. After tipping, a voxel starts relaxing in both transverse and longitudinal directions, as dictated by Bloch equations [45]. $\mathbf{T_1}$ and $\mathbf{T_2}$ simply denote the time constants of the relaxation in the associated direction. The measured signal $m(\vec{\mathbf{r}})$ can be represented as:

$$m(\vec{\mathbf{r}}) = \mathbf{m_0}(\vec{\mathbf{r}})e^{-TE/\mathbf{T_1}(\vec{\mathbf{r}})}\left(1 - e^{-TR/\mathbf{T_2}(\vec{\mathbf{r}})}\right). \tag{2.2}$$

Hence, using long **TR** and short **TE**, almost no relaxation occurs and **PD** of the voxel is measured, which results in **PD**-weighted images. Increasing **TE** while using long **TR** results in measuring relaxation in $\mathbf{T_2}$, and hence $\mathbf{T_2}$-weighted images. For short **TE** and short **TR**, the signal is repeated before the longitudinal relaxation is completed. After a few repetition, the signal reaches steady-state and $\mathbf{T_1}$-weighting becomes the dominating imaging parameter, hence results in $\mathbf{T_1}$-weighted images.

Theoretically, infinitely many contrast images can be formed using different values for **TE** and **TR**. However, similar values for the parameters result in very similar images, and include same type of features. In this thesis, we deal with simultaneous reconstruction of different contrast images to reconstruct higher resolution and better quality images. Even though any number of contrasts can be used, we show our work on different contrast images, namely **PD**, $\mathbf{T_1}$, and $\mathbf{T_2}$ -weighted images.

The parameters $\mathbf{T_1}$, and $\mathbf{T_2}$ depend on the strength of the magnetic field $\mathbf{B}_0$. Typical parameters for different tissues for $\mathbf{B}_0 = 1.5T$ are given in Table 2.1. Each parameter set enhances contrast on different type of structures. PD-weighted images are mostly used for imaging underlying anatomical structures. As it can be seen from the table, fat has very low $\mathbf{T_1}$ value and thus fat appears bright, water appears dark in $\mathbf{T_1}$-weighted images. Since the inverse of $\mathbf{T_2}$ decay is imaged in $\mathbf{T_2}$-weighted images, fat appears dark and water appears bright in these images. Also, most pathological structures demonstrate itself as increase in water level, and this contrast allows imaging those structures.

In this thesis, we deal with the multi-contrast MR image reconstruction problem. For more reading on physics of MRI, please refer to [37, 45].

Table2.1: Typical $\mathbf{T_1}$, and $\mathbf{T_2}$ values for different tissues [7]

Tissue	$\mathbf{T_1}$ (ms)	$\mathbf{T_2}$ (ms)	**PD**
CSF	2569	329	1
Gray Matter	833	83	0.86
White Matter	500	70	0.77
Fat	350	70	1

2.2 Image Reconstruction Problem for MRI

The received signal in MRI does not directly provide an image as explained in section 2.1. Hence, an image reconstruction algorithm is required. Given sampled data $\mathbf{y}$, and a forward model $\mathbf{f}_m$ relating image vector $\mathbf{x}$ to $\mathbf{y}$ such that $\mathbf{y} = \mathbf{f}_m(\mathbf{x})$, image reconstruction is the process that is applied to recover image vector $\mathbf{x}$. Here, $\mathbf{f}_m$ can be any linear or non-linear function that relates the data and image. The problem can be cast as:

$$\begin{aligned} &\textbf{recover} && \mathbf{x} \\ &\textbf{subject to} && \mathbf{f}_m(\mathbf{x}) = \mathbf{y} \end{aligned} . \tag{2.3}$$

For MRI, the sampled signal $\mathbf{y}$ corresponds to sampled signal $\mathbf{s_r}(t)$, and $\mathbf{x}$ corresponds to magnetization $m(\vec{\mathbf{r}})$. In this thesis, we deal with 2D images and $\mathbf{x}$ corresponds to 2D images stacked as vectors, while it can easily be extended to be applied in n-dimensional signals.

If a forward model is invertible, then simplest solution would be to set

$$\mathbf{x} = \mathbf{f}_m^{-1}(\mathbf{y}). \tag{2.4}$$

In this thesis, we deal with linear forward models such that $\mathbf{f}_m$ can be represented as a matrix $\mathbf{A}$, hence $\mathbf{f}_m(\mathbf{x}) = \mathbf{A}\mathbf{x}$. The matrix depends on the problem model and is directly related to the excitation pulse for MRI. Number of rows in the matrix $\mathbf{A}$ is equal to the number of samples, and number of columns is equal to the total size of the image. If the matrix $\mathbf{A}$ is full-rank, then a simple matrix inversion operation would give the best estimate to recover $\mathbf{x}$. However, if the matrix $\mathbf{A}$ is tall such that the number of data points is higher than the number

of image pixels, then least squares can be used to recover the image. By solving

$$\underset{\mathbf{x}}{\textbf{minimize}} \ \|\mathbf{Ax} - \mathbf{y}\|_2^2 \ , \tag{2.5}$$

$\mathbf{x}$ is recovered. Also, for fat $\mathbf{A}$ matrix, i.e. the number of data points is less than the number of image pixels, the solution of

$$\begin{array}{ll} \underset{\mathbf{x}}{\textbf{minimize}} & \|\mathbf{x}\|_2^2 \\ \textbf{subject to} & \mathbf{Ax} = \mathbf{y} \end{array} , \tag{2.6}$$

would give the least squares estimate. The solution of this problem is simply the multiplication of $\mathbf{y}$ by the pseudo-inverse of $\mathbf{A}$.

As stated in Section 2.1, MR scanner simply sample Fourier coefficients of the FOV. However, gathering data is a time consuming process. Sub-sampling accelerates imaging process and hence we deal with fat matrices. At each repetition, a trajectory in k-space is sampled. Hence, the forward model $\mathbf{A}$ can be represented using multiplication of two matrices, as a masked Fourier transform, $\mathbf{A} = \mathbf{MU}$, where $\mathbf{U}$ denotes the unitary matrix associated with Fourier transformation of a given image vector $\mathbf{x}$, and $\mathbf{M}$ denotes a row selection matrix. Hence $\mathbf{MUx}$ simply results in selected Fourier coefficients of $\mathbf{x}$. Here note that Fourier matrix is a **unitary** matrix, such that the inverse of the matrix is its conjugate transpose, or its hermitian, $\mathbf{U}^H$.

If a cartesian sampling method is used in MR, then $\mathbf{U}$ operation can be performed using Fast Fourier Transformation (FFT) algorithm, which has a complexity of $O(Nlg(N))$. Hence, if the k-space is fully sampled, then the image would be $\mathbf{U}^{-1}\mathbf{x}$, and this can be calculated using inverse fast fourier transform algorithm with a complexity of $O(Nlg(N))$. For the sub-sampled k-space case, the solution to (2.6) is the image vector.

Define temporary variable $\mathbf{z}$ as $\mathbf{Ux}$. Then, the problem becomes

$$\begin{array}{ll} \underset{\mathbf{x}}{\textbf{minimize}} & \|\mathbf{U}^{-1}\mathbf{z}\|_2^2 \\ \textbf{subject to} & \mathbf{Mz} = \mathbf{y} \end{array} . \tag{2.7}$$

Using Parseval's theorem (or the fact that Fourier transform is a magnitude preserving transformation), we can omit the inverse Fourier matrix in the least

square step. The solution is then simply zero-filling of unknown elements in the k-space, $\mathbf{U}^H\mathbf{M}^H\mathbf{y}$. The reconstructed image is complex-valued. Different information can be extracted from phase and magnitude of the reconstructed images. However, for most cases magnitude images are required, and hence are more privileged.

The above algorithm holds for cartesian sampling and it reconstructs acceptable results for enough samples. However, for highly subsampled data, the image quality is very poor. Other trajectories such as spiral are used to increase image quality. However, the FFT algorithm is not applicable in those cases. For non-cartesian trajectories most algorithms process data in advance and interpolate it onto a cartesian grid to exploit the speed of FFT [55].

For high undersampling factors, the sub-sampling mask has a huge impact on image quality. Even though purely random sub-sampling works for some applications, most natural images are dense in the center of k-space, and MRI is no different. Recent work suggests drawing sub-sampling mask from a gaussian pdf, or even fully sampling a pre-specified radius in the k-space results in higher quality images [37].

Here, we give details about two types of masks: 1D undersampled, and 2D undersampled. For GRE cartesian sampling, it takes negligible amount of time to sample a line in the Fourier transform of the image. Hence, while gathering 2D image data, one of the directions should always be fully sampled, and hence undersampling masks should include lines in one of the directions. Figure 2.2 includes the created sampling masks. As discussed, samples randomly drawn from a uniform probability density function (pdf) does not always result in quality reconstruction. Sampling the Fourier domain densely in low frequencies help improve the image quality. For this reason, when 1D undersampling is employed, the middle region (low frequency) is fully sampled. The rest is drawn from the given pdf in Fig. 2.2 (a). One instantiation is given in Fig. 2.2 (b). While gathering data for 3D image reconstruction, 2D-undersampling can be employed. As in 1D case, the low frequency region is going to be fully sampled in Fig. 2.2 (c) and (d). The rest is drawn from a pdf as shown. One instantiation

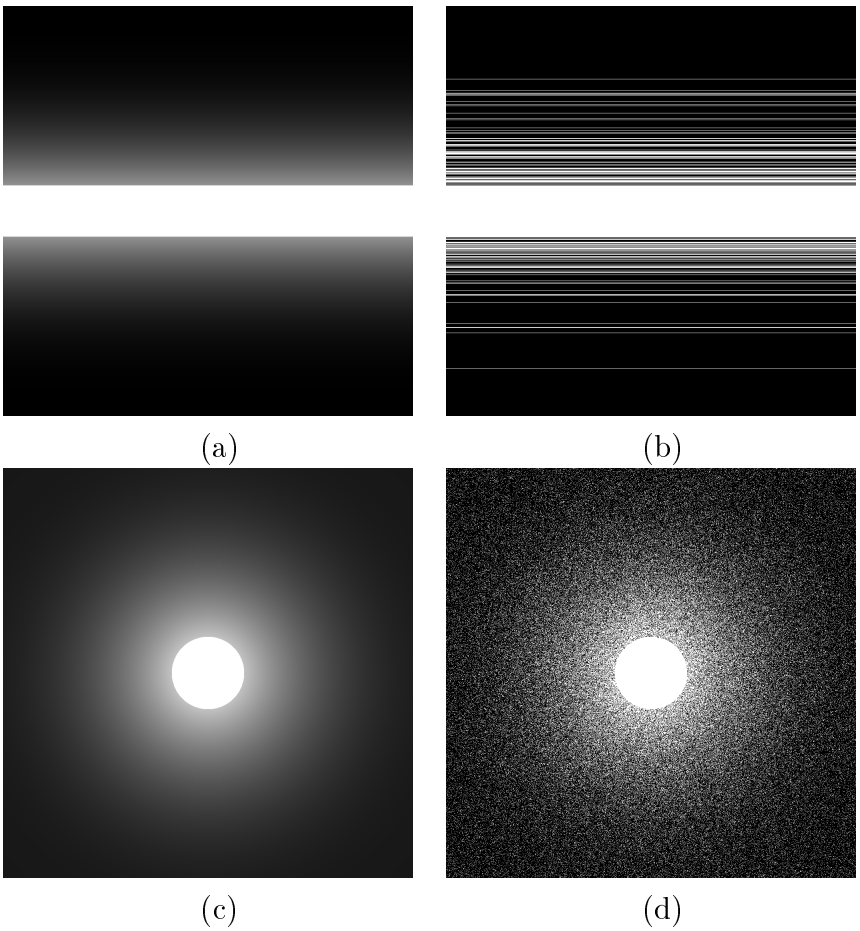


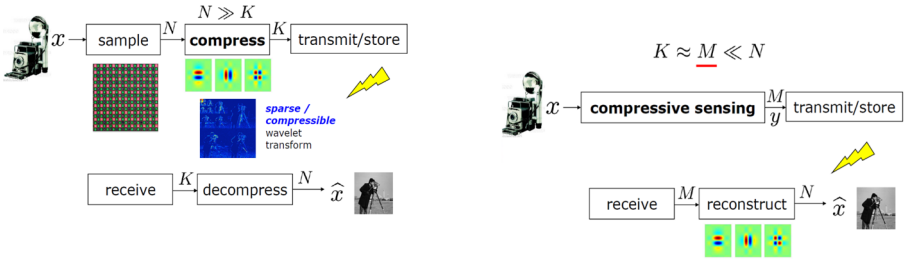
Figure 2.2: (a) 1D undersampling probability density function. (b) A mask created from the PDF in (a). (c) 2D undersampling probability density function. (d) A mask created from the PDF in (c).

of 2D undersampling is given in Fig. 2.2. In figures (b) and (d), white areas are the samples that are gathered, while black areas are omitted. Compressive sensing fills the black areas while denoising the white ones.

In the above formulation that we do not consider noise in the MR imaging scenarios. However, for practical imaging systems, noise should also be considered. In that case, one could change the equality constraint to include noise. For Gaussian noise with zero mean and a standard deviation of ϵ, maximum a posteriori (MAP) estimation corresponds to an ℓ_2-norm bound as $\|\mathbf{Ax} - \mathbf{y}\|_2 \leq \epsilon$. A practical way to retrieve the standard deviation ϵ would be to gather data without any RF excitation, which results in gathering purely noise data. Some algorithms prefer adding the data fidelity term to objective function and use unconstrained optimization with a penalty parameter λ, as $\lambda\|\mathbf{Ax}-\mathbf{y}\|_2^2$. In that case, most algorithms determine λ by solving the problem in multiple times and use the best outcome for parameter selection.

2.3 Compressed Sensing (CS)

To represent a continuous signal $f(x)$ in digital domain, the signal $f(x)$ is sampled at specific time points and digitized into $f[n]$. Nyquist-Shannon's sampling theorem states that if a continuous signal $f(x)$ is sampled at a period twice its highest frequency point, then it can be exactly recovered from its samples. Conventional samplers follow the theorem and gather lots of data. For most applications, the sampled signal is redundant and after sampling the signal is compressed for storage or transmission purposes. Hence the natural question of combining these two processes has been around for many years. Fig. 2.3 (a) depicts the classical approach, (b) depicts the compressive sensing approach. The classical approach includes first sampling a signal, then compressing it. The signal can then be transmitted or stored, and finally the signal can be decompressed at any location to reach the original signal. Compressive sensing approach combines sampling and compressing steps, and arrives at compressive sensing a signal. The signal can then be reconstructed by solving an optimization problem.



(a) (b)

Figure 2.3: (a) Classical Approach (b) Compressive Sensing Approach

Uncompressed Image Compressed Image

Figure 2.4: Uncompressed image (left), Wavelet coefficients of the image (middle), Image after compression (right) [17]

Although it is known that the data is redundant, designing a data gathering process is not straight forward. Fig. 2.4 shows an image with 1 mega-pixels, and the compressed-decompressed image. The image is compressed by keeping top 25k Wavelet coefficients of the image on the left. As it can be seen, the images look almost the same. For most cases, the gathered data is overly redundant. Hence, if there existed some sampling process that could capture top 25k Wavelet coefficients, then the images would look almost the same. However, it is not possible to know a-priori which coefficients constitute the top 25k before hand. This is why compressed sensing requires a non-straight forward data gathering process.

CS is a signal processing technique that can be used to almost exactly recover signals sampled in sub-Nyquist rates. The technique relies on sparsity of the signals of interest in a transform domain. It has been proven that incoherent

measurements of a sparse signal can be recovered using much fewer measurements than Nyquist rate, by solving a non-linear optimization problem [25, 16]. Here, there exists some linear model $\mathbf{A}$ that measures a sparse signal. However, $\mathbf{A}$ is a fat-matrix, and the model $\mathbf{Ax} = \mathbf{y}$ constitutes an under-determined system of equations. Hence, there exists infinitely many solutions. Choosing the "best" solution among these requires a requlatization. One can argue that minimizing the **energy** of the signal may yield the best result, such that $\|\mathbf{x}\|_2^2$ is used as regularization function. However, as shown in [25, 16], this is far from the truth. Undersampling process introduces an inherent noise-like artifacts on the signal. Energy minimization does not help de-noise the signal, and hence the energy-minimized image carries all the noise-like artifacts of the image.

CS theory states that, under some conditions on the forward model $\mathbf{A}$, any sparse signal can be recovered by solving [16, 25]:

$$\begin{aligned} \underset{\mathbf{x}}{\textbf{minimize}} \quad & \|\mathbf{x}\|_0 \\ \textbf{subject to} \quad & \mathbf{Ax} = \mathbf{y} \end{aligned}, \tag{2.8}$$

where $\|\cdot\|_0$ denotes the pseudo-norm called ℓ_0-norm that counts the number of non-zeros in a vector. It has been shown that Eq. (2.8) has a unique solution and will recover the signal $\mathbf{x}$ exactly if the signal is $k-$sparse and $sklog(n)$ measurements are taken, where s is a constant, such as 2 and n is the dimension of signal [16]. However, this is a non-convex optimization problem and this class of problems are known to be NP-Hard [16].

$\mathbf{Ax} = \mathbf{y}$ is an under-determined problem. However, the non-zero positions on $\mathbf{x}$ is known beforehand, then the zero coefficients of $\mathbf{x}$ could be omitted before solving the problem. The corresponding columns in $\mathbf{A}$ could be omitted as well. In that case, the problem becomes an over-determined problem since the number of measurements, $sklog(n)$ stays the same, and the signal can be reconstructed exactly by solving the over-determined system of equations. Here, note that over-determined system of equations may not have a solution. Since we are dealing with the theory here, we assume there exists at least one solution to the over-determined system of equations. However, since the locations of the non-zero coefficients are not known before solution, solving this problem requires combinatorial search on each possible k-sparse combination. Because

the problem is over-determined, it is very likely that most of these combinations will have no solution.

Thus far, we described a way to state that if there exists an under-determined system of equations which the solution is known to be k-sparse, solving for each possible k-sparse combination will yield at least one solution. We have not dealt with the uniqueness of the result, or how to achieve it without solving an NP-Hard problem. There exists another critical contribution of the theory for this very problem. Let us begin with the uniqueness of the result. Assume that $\mathbf{x}$ is a k-sparse vector and the locations of non-zero elements are not known. Then, also let us assume that m measurements are taken using $\mathbf{A}$, where $N >> m$. Hence, this is a highly under-determined system of equations, expressed as $\mathbf{Ax} = \mathbf{y}$. If there are no two $k - sparse$ vectors $(\mathbf{x}_1, \mathbf{x}_2)$ such that

$$\mathbf{Ax}_1 = \mathbf{Ax}_2 = \mathbf{y}, \text{ with } \mathbf{x}_1 \neq \mathbf{x}_2, \tag{2.9}$$

for any $\mathbf{y}$, then the underlying signal $\mathbf{x}$ can be exactly recovered from $\mathbf{y}$ by solving eq. (2.8). Restricted Isometry Property (RIP) is a measure on how well the matrix behaves according to Eq. (2.9). A matrix $\mathbf{A}$ satisfies RIP with restricted isometry constant (RIC) δ, if for every k-sparse vector $\mathbf{x}$ the following inequality holds:

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{Ax}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2. \tag{2.10}$$

If RIP is satisfied for some small RIC, then solving Eq. (2.8) will reconstruct the signal exactly and uniquely. Furthermore, RIC (δ_{2k}) for $2k$-sparse vectors holds the inequality ($\delta_{2k} < \sqrt{2} - 1$), then the image can be recovered using the convex relaxation of Eq. (2.8), ℓ_1-norm regularized version of the problem [17]. In that case, the solution would be:

$$\begin{array}{ll} \underset{\mathbf{x}}{\textbf{minimize}} & \|\mathbf{x}\|_1 \\ \textbf{subject to} & \mathbf{Ax} = \mathbf{y} \end{array}, \tag{2.11}$$

and the error bound on the solution $\mathbf{x}^*$ would be:

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_0\|\mathbf{x} - \mathbf{x}_S\|_1/\sqrt{S}, \text{ and} \tag{2.12}$$

$$\|\mathbf{x}^* - \mathbf{x}\|_1 \leq C_0\|\mathbf{x} - \mathbf{x}_S\|_1, \tag{2.13}$$

where $\mathbf{x}$ is the sampled reference signal, $\mathbf{x}_S$ is the largest S-components of $\mathbf{x}$ with others set to 0, and C_0 is some constant. This theory can easily be extended to handle noise in the image. The best estimate would be the solution of:

$$\begin{array}{ll} \underset{\mathbf{x}}{\textbf{minimize}} & \|\mathbf{x}\|_1 \\ \textbf{subject to} & \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2 \leq \epsilon \end{array}, \tag{2.14}$$

where ϵ is the bound on the noise of the system. Here, the solution $\mathbf{x}^*$ obeys:

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_0 \|\mathbf{x} - \mathbf{x}_S\|_1 / \sqrt{S} + C_1 \epsilon, \tag{2.15}$$

for some constant C_0 and C_1.

While RIP allows the exact reconstruction in computationally feasible times, the verification of this property for a matrix $\mathbf{A}$ is also NP-Hard [16]. There are some structured sampling domains that satisfies this property with very high probability. It is shown that for matrices with elements drawn from independent identically distributed Gaussian satisfies this property with very high probability [16]. Also, random under-sampling in Fourier domain allows the reconstruction with very high probabilities.

There are many extensions of RIP such as BRIP for block compressive sensing [29], DRIP for redundant dictionaries [15], and GRIP for group compressive sensing [6]. Most definitions simply restrict the subset that $\mathbf{x}$ can be in. We will give further details of each respective definition whenever required in the next chapters.

Even though CS theory was first developed for linear forward models with linear sparsifying transforms, the theory is now known to apply to some non-linear functions, such as isotropic total variation [44]. Also, the theory works well under noise and some practical applications such as single pixel camera, MRI and Radar Imaging applications show the effectiveness of the theory [37, 31, 26].

Three critical points are important for a successful CS application [37]:

- Incoherence of undersampling artifacts: The artifacts due to undersampling must be incoherent. This condition can be met if the columns of the sampling matrix are incoherent.

- Non-linear reconstruction: Non-linear reconstruction algorithm such as ℓ_1 or ℓ_0 norm minimization is required to recover image exactly.
- Transform Sparsity: The signal of interest must be sparse or compressible in a transform domain.

Incoherence of undersampling artifacts increase the probability that a sampling matrix satisfies RIP. Since MR images are sampled in Fourier domain, all samples are incoherent and this property is a natural match in MRI. Another way to look at incoherence is that, incoherence provides noise-like artifacts. ℓ_1-norm regularized denoising (also called basis-pursuit) is known to work well. Hence, CS algorithms simply solve a denoising problem to reconstruct the original image, provided that the undersampling artifacts are incoherent noise-like.

Here let us note that compressive sensing is not widely used in clinical MRI applications, due to required high computational cost. For this reason, faster algorithms play a critical role in a practical compressive sensing application. It has been announced by Siemens Healthineers (Erlangen, Germany) that Food and Drug Administration (FDA, Maryland USA) has recently cleared compressive sensing MRI acceleration technology from them [33].

Non-linear reconstruction methods are discussed in detail in section 2.3.1, and transform domain sparsity is discussed in detail in section 5.1.

2.3.1 Optimization Algorithms for CS

Optimization algorithms play a critical role in applications of CS. As described in section 2.3, there are two main approaches to CS. The first approach is pursuing solution to the non-convex problem. Algorithms such as Matching Pursuit (MP) or Orthogonal Matching Pursuit (OMP) directly attacks (2.8), and converges to a sub-optimal point. Here we describe OMP.

The method recovers a single point in the sparse signal at each iteration until a stopping criterion is met. Various adaptations of the algorithm exist, and some successful applications include radar imaging [8, 59]. However, wrong choice

Table2.2: Orthogonal Matching Pursuit Algorithm

1. Set $\mathbf{r} = \mathbf{y}$, $\lambda = \{\}$
2. **repeat**
3. Project $\mathbf{r}$ onto columns of $\mathbf{A}$
4. Set τ to maximum absolute projection, $\max\{| < \mathbf{A}^H, \mathbf{r} > |\}$
5. $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{A}\tau$
6. Add to λ new element τ
7. **until** some stopping criterion is satisfied.

of stopping criterion may result in poor reconstructions [14]. Even though the output of this thesis can be used with approximate solvers, our main focus is on the second type algorithms, and we will not discuss approximate solvers in detail. The algorithm is given in Table 2.2.

First, residual $\mathbf{r}$ is set to $\mathbf{y}$. Then, at each iteration, $\mathbf{r}$ is projected onto the columns of $\mathbf{A}$. The column with the largest absolute value of these projections is added to the solution set. Then the image $\mathbf{x}$ is updated as the least squares solution of the remaining data vector, and the residual $\mathbf{r}$ is also updated. The loop is terminated when a stopping criterion is satisfied. For a review of greedy algorithms, please see [11].

The second type of algorithms solve the convex relaxation of the problem in a known transformation domain $\mathbf{W}$:

$$\begin{aligned} \underset{\mathbf{x}}{\textbf{minimize}} \quad & \|\mathbf{W}\mathbf{x}\|_1 \\ \textbf{subject to} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \end{aligned} . \tag{2.16}$$

The outlined problem in Eq. (2.16) is convex and exact solution can be obtained with different type of algorithms. Various algorithms have been proposed in the literature for the solution of this problem. The method introduced by Çetin and Karl uses conjugate gradient to solve the problem using total variation and reflectivity field sparsity in unconstrained form for synthetic aperture radar (SAR) imaging [18]. [18] uses a quasi-Newton algorithm, while [49] adopts a synthesis based objective function and uses the Spectral Projected Gradient (SPGL1) algorithm [60]. In [38], an adaptive sequential basis selection strategy is employed. [21] adopts a non-convex approach and solves ℓ_p-norm for $p < 1$ for

higher quality reconstruction. Even though non-convex, practical experiments show the reconstructed images are somewhat robust, and have better overall quality.

CS-MRI algorithm shows combination of multiple bases, namely Wavelet and Gradient, result in higher quality images, and uses conjugate gradient to solve the problem [37]. Slow convergence of conjugate gradient disable practical application of the algorithm. Hence, other algorithms are proposed for similar problems. TVCMRI algorithm uses operator splitting while rec-pf algorithm uses variable splitting to solve the problem proposed in CSMRI faster [56, 63]. To accelerate the solution, Fast Composite Splitting Algorithm (FCSA) based on composite splitting denoising and FISTA was proposed [9, 35]. The problem solved by these algorithms are in the form:

$$\underset{\mathbf{x}}{\textbf{minimize}}\ \alpha_1\|\mathbf{W}\mathbf{x}\|_1 + \alpha_2 TV(\mathbf{x}) + \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2, \tag{2.17}$$

where TV is either isotropic (denoted TV_{iso}) or anisotropic (denoted TV_{ani}), as defined below:

$$TV_{iso}(\mathbf{x}) = \sum_n \sqrt{\sum_m \left(\nabla_m \mathbf{x}[n]\right)^2}, \tag{2.18}$$

$$TV_{ani}(\mathbf{x}) = \sum_m \|\nabla_m \mathbf{x}\|_1, \tag{2.19}$$

where ∇_m denotes the gradient operator in $m-th$ dimension. Total variation function enforces sparsity on the gradients, hence reconstructs piece-wise smooth images. The function is applicable to natural images, because natural images tend to have similar intensity within an object. Total variation is minimized for sharp transitions rather than smooth transitions, hence it promotes keeping same intensities within an object. Even though it is easier to solve Eq. (2.19) compared to Eq. (2.18), Eq. (2.18) imposes the sparsity on the gradients to be on exactly same coordinate, hence results in more appropriate reconstruction.

Bayesian image reconstruction algorithms are also applicable to compressed sensing [10, 36]. These algorithms maximize the posteriori probability of image vector $\mathbf{x}$ given the data $\mathbf{y}$, $P(\mathbf{x}|\mathbf{y})$. The process is called Maximum a Posteriori

(MAP) estimation as defined below:

$$\underset{\mathbf{x}}{\textbf{maximize}}\ P(\mathbf{x}|\mathbf{y}), \tag{2.20}$$

equivalently

$$\underset{\mathbf{x}}{\textbf{maximize}}\ \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}. \tag{2.21}$$

Let number of data points be N. Then, for an independent identically distributed Gaussian noise with zero mean and ϵ standard deviation, and a Laplacian probability density prior for $\mathbf{x}$,

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\epsilon^2)^{\frac{N}{2}}} \exp\{-\frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2}{2\epsilon^2}\}, \tag{2.22}$$

$$P(\mathbf{x}) = \frac{1}{2}\exp\{-\|\mathbf{x}\|_1\}, \tag{2.23}$$

$$\underset{\mathbf{x}}{\textbf{maximize}}\ \frac{1}{(2\pi\epsilon^2)^{\frac{N}{2}}} \exp\{-\frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2}{2\epsilon^2}\}\frac{1}{2}\exp\{-\|\mathbf{x}\|_1\}, \tag{2.24}$$

then log-likelihood maximization is defined as:

$$\underset{\mathbf{x}}{\textbf{maximize}}\ -\frac{1}{2\epsilon^2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 - \|\mathbf{x}\|_1, \tag{2.25}$$

$$\underset{\mathbf{x}}{\textbf{minimize}}\ \frac{1}{2\epsilon^2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 + \|\mathbf{x}\|_1. \tag{2.26}$$

As it can be seen, the problem is exactly of the form Eq. (2.17) for $\mathbf{W} = \mathbf{I}$, $\alpha_1 = 2\epsilon^2$ and $\alpha_2 = 0$ [36].

In this thesis, we use an augmented Lagrangian based algorithm. The method is described in detail in section 2.3.2.

2.3.2 Alternating Direction Method of Multipliers (ADMM)

ADMM is an augmented Lagrangian method (ALM) based optimization algorithm that was first introduced in 1970s with roots in 1950s, but was recently re-discovered [12]. Augmented Lagrangian methods solve a constrained optimization problem by adding a term called Lagrangian term that is zero for any feasible vector. This enables the algorithm to converge under far more general conditions. ADMM contains ideas involving dual decomposition, method

of multipliers, proximal methods and variable splitting. In addition to the Lagrangian term, ADMM uses a divide-and-conquer type approach by splitting unconstrained multi-objective optimization problems, augmenting the Lagrangian with a norm-squared error term, and using a non-linear block Gauss-Seidel approach on the resultant terms in the sub-problems. ADMM is advantageous when the sub-problems have analytical solution or have an efficient algorithm to solve. The resulting algorithm converges under mild conditions [12]. A practical compressive sensing application requires fast and reliable solution of the optimization problem. ADMM can provide these properties, and this is the main reason we propose ADMM based algorithms for CS.

The algorithm solves the generic problems of type:

$$\begin{array}{ll} \underset{\mathbf{x},\mathbf{z}}{\textbf{minimize}} & f(\mathbf{x}) + g(\mathbf{z}) \\ \textbf{subject to} & \mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} = \mathbf{c} \end{array}, \tag{2.27}$$

where $\mathbf{G}$ and $\mathbf{H}$ are the constant problem model matrices with a constant vector $\mathbf{c}$, f and g denote the separable functions with respect to $\mathbf{x}$ and $\mathbf{z}$, respectively. Hence, this model allows separation of variables for multiple additive objective functions. ADMM attacks the problem by first writing the Lagrangian function:

$$L_\mu(\mathbf{x}, \mathbf{z}, \mathbf{u}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{u}^T(\mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} - \mathbf{c}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} - \mathbf{c}\|_2^2. \tag{2.28}$$

For a positive step size μ, ADMM consists of the following iterations:

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} L_\mu(\mathbf{x}, \mathbf{z}_k, \mathbf{d}_k), \tag{2.29}$$

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} L_\mu(\mathbf{x}_{k+1}, \mathbf{z}, \mathbf{d}_k), \tag{2.30}$$

$$\mathbf{u}_{k+1} = \mathbf{u} + \frac{\mu}{2}\left(\mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} - \mathbf{c}\right), \tag{2.31}$$

where k is the iteration counter. Hence, the updates are made in an alternating fashion. Let $\mathbf{r} = \mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} - \mathbf{c}$. Then,

$$\mathbf{u}^T\mathbf{r} + \frac{\mu}{2}\|\mathbf{r}\|_2^2 = \mathbf{u}^T\mathbf{r} + \frac{\mu}{2}\mathbf{r}^T\mathbf{r} \tag{2.32}$$

$$= \mathbf{u}^T\mathbf{r} + \frac{\mu}{2}\mathbf{r}^T\mathbf{r} + \frac{1}{\mu}\mathbf{u}^T\mathbf{u} - \frac{1}{\mu}\mathbf{u}^T\mathbf{u} \tag{2.33}$$

$$= \frac{\mu}{2}\|\mathbf{r} + \frac{1}{\mu}\mathbf{u}\|_2^2 - \frac{1}{\mu}\|\mathbf{u}\|_2^2. \tag{2.34}$$

As it can be seen, the second term does not affect the minimization problem with respect to $\mathbf{x}$ or $\mathbf{z}$. Also, define $\mathbf{d} = \frac{1}{\mu}\mathbf{u}$. The solution to equations (2.29) and (2.30) can be made simpler using (2.34) and separability of function f and $\mathbf{z}$:

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z}_k - \mathbf{c} + \mathbf{d}_k\|_2^2 \tag{2.35}$$

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} g(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{x}_{k+1} + \mathbf{H}\mathbf{z} - \mathbf{c} + \mathbf{d}_k\|_2^2 \tag{2.36}$$

$$\mathbf{d}_{k+1} = \mathbf{d} + \mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} - \mathbf{c}. \tag{2.37}$$

The algorithm is proven to converge to an approximate solution very fast. However, it may take time to converge to an exact solution. For applications such as CS, where an approximate solution is good enough, the algorithm is shown to converge. The algorithm is also proven to converge for bi-convex and non-convex problems [12].

2.3.2.1 C-SALSA

One of the first applications of an ADMM-variant for CS is an algorithm called C-SALSA [2, 1]. The original paper shows that ADMM converges faster for problems with only one objective terms. The algorithm solves the problems of type:

$$\begin{array}{ll} \underset{\mathbf{x}}{\textbf{minimize}} & \phi(\mathbf{x}) \\ \textbf{subject to} & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \end{array}, \tag{2.38}$$

for any convex $\phi(x)$. The algorithm first considers the constraint $\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2 \leq \epsilon$ as an objective function using the indicator function of the convex set that has value 0 if the condition is satisfied and ∞ if not, denoted by $\iota(\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon)$. It sets the ADMM variables $f(\mathbf{x}) = 0$, $\mathbf{c} = 0$, $\mathbf{H} = I$ and $\mathbf{G} = -[\mathbf{I}\ \mathbf{A}^T]^T$. The algorithm splits $\mathbf{z}$ vector into two as $\mathbf{z}^{(0)}$ and $\mathbf{z}^{(1)}$ to enforce two objective functions, and sets $g(\mathbf{z}) = \phi(\mathbf{z}^{(1)}) + \iota(\|\mathbf{z}^{(0)} - \mathbf{y}\|_2 \leq \epsilon)$. Hence, the algorithm forces $\mathbf{x} = \mathbf{z}^{(1)}$ and $\mathbf{A}\mathbf{x} = \mathbf{z}^{(0)}$.

The resulting iterations become:

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \|\mathbf{z}_k^{(1)} - \mathbf{x} + \mathbf{d}_k^{(1)}\|_2^2 + \|\mathbf{z}_k^{(0)} - \mathbf{A}\mathbf{x} + \mathbf{d}_k^{(0)}\|_2^2 \tag{2.39}$$

$$\mathbf{z}_{k+1}^{(0)} = \arg\min_{\mathbf{z}^{(0)}} \iota(\|\mathbf{z}^{(0)} - \mathbf{y}\|_2 \leq \epsilon) + \frac{\mu}{2}\|\mathbf{z}^{(0)} - \mathbf{A}\mathbf{x}_{k+1} + \mathbf{d}_k^{(0)}\|_2^2 \tag{2.40}$$

$$\mathbf{z}_{k+1}^{(1)} = \arg\min_{\mathbf{z}^{(1)}} \phi(\mathbf{z}^{(1)}) + \frac{\mu}{2}\|\mathbf{z}^{(1)} - \mathbf{x}_{k+1} + \mathbf{d}_k^{(1)}\|_2^2 \tag{2.41}$$

$$\mathbf{d}_{k+1}^{(0)} = \mathbf{d}^{(0)} - \mathbf{A}\mathbf{x} + \mathbf{z}^{(0)} \tag{2.42}$$

$$\mathbf{d}_{k+1}^{(1)} = \mathbf{d}^{(1)} - \mathbf{x} + \mathbf{z}^{(1)}. \tag{2.43}$$

The problem in (2.39) is a simple least squares problem. The analytical solution can be calculated using,

$$\mathbf{x}_{k+1} = (\mathbf{I} + \mathbf{A}^{\mathbf{H}}\mathbf{A})^{-1}\left[\mathbf{d}_k^{(1)} + \mathbf{z}_k^{(1)} + \mathbf{A}^H\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right)\right]. \tag{2.44}$$

Fast solutions to eq.(2.44) is proposed in [1]. A faster approach is proposed in this work. The problems (2.41) and (2.40) are known as Moreau proximal mapping functions and analytical solutions exist for many cases. Proximal mapping functions are defined as:

$$\text{prox}_{\phi}(\mathbf{v}) = \arg\min_{\mathbf{x}} \ \phi(\mathbf{x}) + \frac{1}{2\tau}\|\mathbf{x} - \mathbf{v}\|_2^2. \tag{2.45}$$

For compressed sensing applications, the function ϕ is chosen as ℓ_1-norm, which results in an associated proximal mapping function known as soft thresholding,

$$soft(\mathbf{v}, \tau) = sign(\mathbf{v}) \cdot \max\{0, |\mathbf{v}| - \tau\}. \tag{2.46}$$

Also, an algorithm by Chambolle is proposed for the proximal mapping of the isotropic total variation function [19].

2.4 Chapter Summary

In this chapter, we overviewed MRI, its physics and typical image reconstruction algorithms. The forward model of MRI is a simple Fourier transform for most applications. Although other data gathering strategies exist, here we use random Fourier sub-sampling for simplicity. Also, we give brief information on compressed sensing, and typical optimization algorithms used in it. We overviewed

the sufficient conditions for sparse recovery, and key aspects in a successful application. Finally, we discussed optimization algorithms, and briefly discussed an augmented Lagrangian based method, ADMM. We gave information on the reasons we use ADMM, and described how it works.

CHAPTER 3

SINGLE-CONTRAST OPTIMIZATION FRAMEWORK USING ADMM (H-ADMM)

In this chapter, we describe a method for MRI image reconstruction. We first start by deriving a solution for single contrast compressive sensing. We then explain the MR specific problems in image reconstruction such as handling phase in images. Then, we describe the algorithm, and how to use it under certain conditions. For the results section, we plan to use this algorithm as baseline for comparison.

3.1 ADMM for Constrained Optimization with a Hybrid Cost Function

In the previous chapter, we have discussed sparsifyable signals in a single transformation domain. However, in most applications including MRI, because the signal comprises of multiple type of features, it can only be sparsified using multiple transformation domains. In that case, one needs to use multiple domains for transformation to either analyse or synthesize the sparse signal. Also, the multiple transformation domains does not necessarily need to include linear transformations. Using analysis formulation has been previously shown to improve the performance of the reconstruction algorithm [37, 63].

To clarify more on the previous paragraph, let us use an example. Assume that the signal of interest consists of periodic impulses such that it is of the form, as in [15]:

$$f(t) = \sum_{j=1}^{\sqrt{n}} \delta(t - j\sqrt{n}) \tag{3.1}$$

The signal is not exactly sparse neither in signal domain nor in Fourier domain.

The signal can only be recovered using both of these sparsifying transformations simultaneously [15]. This example shows that using multiple transformation domains may out-perform using single transformation.

In this study, we use a linear combination of a non-linear transformation (total variation, TV) along with a linear one (ℓ_1-norm). We assume both piece-wise constant features (sparse in the gradient directions) and sparse defects in the image domain. Although MRI is not sparse in the image domain, using such a combination outperforms both only ℓ_1-norm and only TV minimization. Hence, we need an algorithm to solve for multiple separable objective functions. In this section, we describe a generic method for solving constrained optimization problems, and how to apply it to MR imaging.

First, we consider the problem model as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{3.2}$$

where $\mathbf{n}$ is the noise vector. We propose solving:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \alpha_1 \|\mathbf{x}\|_1 + \alpha_{TV} TV(\mathbf{x}) \\ \text{subject to} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \end{aligned} . \tag{3.3}$$

In particular, we use a solution to the constrained optimization problem with a hybrid cost function, by separating the weighted components in the Augmented Lagrangian. Here, we propose a generic framework for problems of type:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \alpha_1 \phi_1(\mathbf{x}) + \cdots + \alpha_m \phi_m(\mathbf{x}) \\ \text{subject to} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \end{aligned} , \tag{3.4}$$

where m is the number of separable objective functions, $\mathbf{x}$ is the image vector, $\mathbf{y}$ is the data vector, $\mathbf{A}$ is the forward model matrix, ϵ is the bound on the noise, $\phi_i(\cdot)$ represents the separable objective functions with α_i as regularization parameters. Here, we use the constrained form of the optimization problem for easier choice of parameter. The data fidelity term $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon$ is actually equal to $\|\mathbf{n}\|_2 \leq \epsilon$, which can simply be set to ℓ_2-norm of $\mathbf{n}$. Hence, we choose $m = 2$, $\phi_1(\cdot) = \|\cdot\|_1$, and $\phi_2(\cdot) = TV(\cdot)$.

ADMM solves problems of type Eq. (2.27). To solve the proposed type of problems using ADMM, variables need to be split and defined accordingly. We em-

ploy splitting scheme used in [1, 2], and define the variables for the transformed problem:

$$\begin{array}{ll} \underset{\mathbf{x},\mathbf{z}}{\textbf{minimize}} & f(\mathbf{x}) + g(\mathbf{z}) \\ \textbf{subject to} & \mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} = \mathbf{c} \end{array}, \tag{3.5}$$

as:

$$\mathbf{z} = \left[\mathbf{z}^{(0)}\ \mathbf{z}^{(1)}\ \cdots\ \mathbf{z}^{(m)}\right]^{\mathbf{T}}, \quad \mathbf{G} = -\left[\mathbf{A}^{\mathbf{T}}\ \mathbf{I}\ \cdots\ \mathbf{I}\right]^{\mathbf{T}}, \tag{3.6}$$

Let us now choose the required problem model variables to solve the proposed problem in eq. (3.4). Now, set $\mathbf{H} = \mathbf{I}, \mathbf{c} = 0, f(\mathbf{x}) = 0$, and

$$g(\mathbf{z}) = \iota_{E(\epsilon,\mathbf{I},\mathbf{y})}\left(\mathbf{z}^{(0)}\right) + \sum_{i=1}^{m} \alpha_i \phi_i\left(\mathbf{z}^{(i)}\right). \tag{3.7}$$

Here, $\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}\left(\mathbf{z}^{(0)}\right)$ is the indicator function for the constraint defined as:

$$\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in E(\epsilon,\mathbf{I},\mathbf{y}) = \{\mathbf{x} | \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\} \\ \infty & \text{otherwise} \end{cases}. \tag{3.8}$$

Hence, the resulting problem ensures that data fidelity is satisfied for $\mathbf{z}^{(0)}$. For the given definitions, let us break down the linear constraints imposed by the equality in ADMM, $\mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} = \mathbf{c}$:

$$\mathbf{A}\mathbf{x} = \mathbf{z}^{(0)} \tag{3.9}$$

$$\mathbf{x} = \mathbf{z}^{(1)} \tag{3.10}$$

$$\cdots \tag{3.11}$$

$$\mathbf{x} = \mathbf{z}^{(m)} \tag{3.12}$$

the equalities force $\mathbf{A}\mathbf{x} = \mathbf{z}^{(0)}$ for data fidelity, while forcing each $\mathbf{z}^{(i)} = \mathbf{x}$ for separable objective functions. These setting ensures the proposed problem can be solved using ADMM framework.

Now let us derive the iterations associated with the proposed method. It consists of iterations:

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \frac{\mu}{2}\left(\|\mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z}_k + \mathbf{d}_k\|_2^2\right), \tag{3.13}$$

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} f_2(\mathbf{z}) + \frac{\mu}{2}\| \left(\mathbf{G}\mathbf{x}_{k+1} + \mathbf{H}\mathbf{z} + \mathbf{d}_k\right) \|_2^2, \tag{3.14}$$

$$\mathbf{d}_{k+1} = \mathbf{d}_k + \mathbf{G}\mathbf{x}_{k+1} + \mathbf{H}\mathbf{z}_{k+1}. \tag{3.15}$$

Here, the solution to eq. (3.4) can be obtained by following these iterations. Now let us further simplify the steps given in the iterations. Since there is no objective function is associated with $\mathbf{x}$, Eq. (3.13) is a simple least squares operation and has an analytical solution as:

$$\mathbf{x}_{k+1} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H(\mathbf{d}_k + \mathbf{z}_k). \tag{3.16}$$

Putting $\mathbf{G} = -\begin{bmatrix}\mathbf{A^T}\ \mathbf{I}\ \cdots\ \mathbf{I}\end{bmatrix}^{\mathbf{T}}$ back,

$$= \left(m\mathbf{I} + \mathbf{A}^H\mathbf{A}\right)^{-1}\left[\sum_{i=1}^{m}\mathbf{d}_k^{(i)} + \mathbf{z}_k^{(i)} + \mathbf{A}^H\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right)\right]. \tag{3.17}$$

This step requires a pre calculated matrix inversion. Efficient solutions to this step will be described in the next section.

Next, we deal with eq. (3.14). This step consists of m separable optimization problems. We split objective functions so that each sub problem can be handled in a very fast way. First, let us write the problem as:

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} \iota_{E(\epsilon,\mathbf{I},\mathbf{y})}\left(\mathbf{z}^{(0)}\right) + \sum_{i=1}^{m}\alpha_i\phi_i\left(\mathbf{z}^{(i)}\right) + \frac{\mu}{2}\|\left(\mathbf{G}\mathbf{x}_{k+1} + \mathbf{H}\mathbf{z} + \mathbf{d}_k\right)\|_2^2 \tag{3.18}$$

$$\mathbf{z}_{k+1}^{(0)} = \arg\min_{\mathbf{z}^{(0)}} \iota_{E(\epsilon,\mathbf{I},\mathbf{y})}\left(\mathbf{z}^{(0)}\right) + \frac{\mu}{2}\|\mathbf{z}^{(0)} - (\mathbf{A}\mathbf{x}_{k+1} - \mathbf{d}_k^{(0)})\|_2^2 \tag{3.19}$$

$$\mathbf{z}_{k+1}^{(i)} = \arg\min_{\mathbf{z}^{(i)}} \alpha_i\phi_i\left(\mathbf{z}^{(i)}\right) + \frac{\mu}{2}\|\mathbf{z}^{(i)} - (\mathbf{x}_{k+1} - \mathbf{d}_k^{(i)})\|_2^2 \text{ for } i = 1\cdots m \tag{3.20}$$

Here, the eqns. (3.19) and (3.20) are called Moreau proximal mapping functions. These functions are studied in the literature and are defined in general form as:

$$\text{prox}_{\tau\phi}(\mathbf{v}) = \arg\min_{\mathbf{x}}\ \phi(\mathbf{x}) + \frac{1}{2\tau}\|\mathbf{x} - \mathbf{v}\|_2^2. \tag{3.21}$$

Proximal mapping functions are a well studied in the literature [12, 48]. The proposed method requires proximal mapping function associated with each objective function. Here, let us assume that each proximal mapping function $\text{prox}_{\tau\phi}(\mathbf{v})$ can be calculated using $\mathbf{\Psi}_{\tau\phi}(\mathbf{v})$. Then, the update steps can be re-written using this definition as:

$$\mathbf{z}_{k+1}^{(0)} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}\left(\mathbf{A}\mathbf{x}_k - \mathbf{d}_k^{(0)}\right) \tag{3.22}$$

$$\mathbf{z}_{k+1}^{(i)} = \mathbf{\Psi}_{\frac{\alpha_i}{\mu}\phi_i}\left(\mathbf{x}_{k+1} - \mathbf{d}_k^{(i)}\right) \text{ for } i = 1\cdots m \tag{3.23}$$

Here, eq. (3.22) is a simple projection onto ℓ_2-norm ball with ϵ radius and center $\mathbf{y}$, defined as:

$$\mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{A}\mathbf{x}_k - \mathbf{d}_k^{(0)}) = \begin{cases} \mathbf{A}\mathbf{x}_k - \mathbf{d}_k^{(0)}, & \text{if } \|\mathbf{A}\mathbf{x}_k - \mathbf{d}_k^{(0)} - \mathbf{y}\|_2 \leq \epsilon \\ \mathbf{y} + \epsilon \frac{\left(\mathbf{A}\mathbf{x}_k - \mathbf{d}_k^{(0)} - \mathbf{y}\right)}{\|\mathbf{A}\mathbf{x}_k - \mathbf{d}_k^{(0)} - \mathbf{y}\|_2}, & \text{if } \|\mathbf{A}\mathbf{x}_k - \mathbf{d}_k^{(0)} - \mathbf{y}\|_2 > \epsilon \end{cases}, \tag{3.24}$$

The final update step of ADMM, eq. (3.15) is a simple algebraic step that calculates sum of vectors, and consists of the following calculations:

$$\mathbf{d}_{k+1}^{(i)} = \mathbf{d}_k^{(i)} - \mathbf{x}_{k+1} + \mathbf{z}_{k+1}^{(i)}, \text{for } i = 1 \cdots m \tag{3.25}$$

$$\mathbf{d}_{k+1}^{(0)} = \mathbf{d}_k^{(0)} - \mathbf{A}\mathbf{x}_{k+1} + \mathbf{z}_{k+1}^{(0)}. \tag{3.26}$$

These last calculations finalize the derivation of the framework. The algorithm is summarized in Table. 3.1 as Algorithm *ADMM with Hybrid cost function (H-ADMM)*. The proposed method is a framework that solves problems of type eq. (3.4).

The specific choice of forward model $\mathbf{A}$, and objective functions $\phi_i(\cdot)$ with their respective regularization coefficients α_i depends on the problem, as discussed earlier. In this study, we use partial Fourier transform as $\mathbf{A}$, which is essentially sub-sampled version of a unitary transformation. This helps greatly in speeding up the algorithm as will be shown in the next sections. Also, in this study, we use two regularization functions ($m = 2$) ℓ_1-norm and total variation for experimental results. Application of these functions to complex imagery will also be discussed in the next sections.

3.2 Fast Solution of Proposed Approach For Unitary Transform Domains

The algorithm in the previous section requires a least squares operation along with single forward ($\mathbf{A}$) and conjugate transpose of forward ($\mathbf{A}^H$) operations. When unitary transforms are used, the cost of least squares operation reduces to single calculation of $\mathbf{A}^H\mathbf{A}$, resulting in a total cost of two forward and two conjugate transpose operators. However, as we will show in this chapter, this cost can be further reduced.

Table3.1: **Algorithm:** *ADMM with Hybrid cost function (H-ADMM)*

1. Set $k = 0$, choose $\mu > 0$, $\mathbf{z}_0^{(i)}$, $\mathbf{d}_0^{(i)}$, α_i for all i
2. **repeat**
3. $\mathbf{x}_{k+1} = (m\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1}\left[\sum_{i=1}^{m} \mathbf{d}_k^{(i)} + \mathbf{z}_k^{(i)} + \mathbf{A}^H\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right)\right]$
4. **for i = 1 $\cdots$ m**
5. $\quad \mathbf{z}_{k+1}^{(i)} = \mathbf{\Psi}_{\phi_i \frac{\alpha_i}{\mu}}\left(\mathbf{x}_{k+1} - \mathbf{d}_k^{(i)}\right)$
6. $\quad \mathbf{d}_{k+1}^{(i)} = \mathbf{d}_k^{(i)} - \mathbf{x_{k+1}} + \mathbf{z}_{k+1}^{(i)}$
7. **endfor**
8. $\mathbf{z}_{k+1}^{(0)} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}\left(\mathbf{A}\mathbf{x} - \mathbf{d}^{(0)}\right)$
9. $\mathbf{d}_{k+1}^{(0)} = \mathbf{d}_k^{(0)} - \mathbf{A}\mathbf{x}_{k+1} + \mathbf{z}_{k+1}^{(0)}$
10. $k \leftarrow k + 1$
11. **until** some stopping criterion is satisfied.

When $\mathbf{A}$ is a masked unitary transform such as Fourier, the algorithm can be carried out using a single forward and inverse operations per iteration. Let us denote the unitary transformation matrix by $\mathbf{U}$, and mask by $\mathbf{M}$, so that $\mathbf{A} = \mathbf{M}\mathbf{U}$. Here $\mathbf{M}$ satisfies $\mathbf{M}\mathbf{M}^H = \mathbf{I}$ while $\mathbf{M}^H\mathbf{M} \neq \mathbf{I}$. Hence $\mathbf{A}\mathbf{A}^H = \mathbf{I}$, while $\mathbf{A}^H\mathbf{A} \neq \mathbf{I}$.

First, let us begin with getting rid of the inversion in eq. (3.17).

$$(m\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1} = \frac{1}{m}\mathbf{I} - \frac{1}{m}\mathbf{I}\mathbf{A}^H(\mathbf{I} + \frac{1}{m}\mathbf{A}\mathbf{I}\mathbf{A}^H)^{-1}\mathbf{A}\frac{1}{m}\mathbf{I} \tag{3.27}$$

$$= \frac{1}{m}\left(\mathbf{I} - \frac{1}{m}\mathbf{A}^H(\mathbf{I} + \frac{1}{m}\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}\right) \tag{3.28}$$

$$= \frac{1}{m}\left(\mathbf{I} - \frac{1}{m}\mathbf{A}^H(\mathbf{I} + \frac{1}{m}\mathbf{I})^{-1}\mathbf{A}\right) \tag{3.29}$$

$$= \frac{1}{m}\left(\mathbf{I} - \frac{1}{m}\frac{m}{m+1}\mathbf{A}^H\mathbf{A}\right) \tag{3.30}$$

$$= \frac{1}{m}\left(\mathbf{I} - \frac{1}{m+1}\mathbf{A}^H\mathbf{A}\right) \tag{3.31}$$

Fast inverse calculation given in eq. (3.31) has been shown in [1]. Eq. (3.31) requires two forward and two inverse operations. Now let us further decrease the cost by defining $\mathbf{q}_k$ as:

$$\mathbf{q}_k = \sum_{i=1}^{m} \mathbf{d}_k^{(i)} + \mathbf{z}_k^{(i)}. \tag{3.32}$$

Then, step 3 of the algorithm table 3.1 can be re-written as:

$$\mathbf{x}_{k+1} = \frac{1}{m}(\mathbf{I} - \frac{1}{m+1}\mathbf{A}^H\mathbf{A})\left[\mathbf{q}_k + \mathbf{A}^H\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right)\right], \tag{3.33}$$

$$= \frac{1}{m}\left[\mathbf{q}_k + \mathbf{A}^H\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right) - \frac{1}{m+1}\left(\mathbf{A}^H\mathbf{A}\mathbf{q}_k + \mathbf{A}^H\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right)\right)\right] \tag{3.34}$$

$$\mathbf{x}_{k+1} = \frac{1}{m}\left[\mathbf{q}_k + \frac{1}{m+1}\mathbf{A}^H\left(m\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right) - \mathbf{A}\mathbf{q}_k\right)\right]. \tag{3.35}$$

Then, $\mathbf{A}\mathbf{x}_{k+1}$ can be calculated as

$$\mathbf{A}\mathbf{x}_{k+1} = \frac{1}{m}\left[\mathbf{A}\mathbf{q}_k + \frac{1}{m+1}\mathbf{A}\mathbf{A}^H\left(m\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right) - \mathbf{A}\mathbf{q}_k\right)\right] \tag{3.36}$$

$$\mathbf{A}\mathbf{x}_{k+1} = \frac{1}{m+1}\left[\mathbf{A}\mathbf{q}_k + \left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right)\right]. \tag{3.37}$$

As shown, $\mathbf{A}\mathbf{x}_{k+1}$ and $\mathbf{x}_{k+1}$ can be calculated using only one forward $\mathbf{A}$ and one inverse $\mathbf{A}^H$ operations, which corresponds to Fourier transformation for our problem. Note that this method now has the same computational cost as Orthogonal Matching Pursuit, a greedy optimization algorithm [8]. The updated algorithm can be found in Table 3.2 This algorithm has been published in [31].

3.3 Handling Phase for Complex Imaging

The choice of regularization function has a huge effect on the reconstructed image. This corresponds to choosing the transform domain in which MR images are sparse. Although compressive sensing mostly deals with linear transformations, non-linear sparsifying transforms such as total variation is known to work well [16]. Also, since MRI produces complex-valued images due to the nature of data gathering process (such as chemical shift, off-resonance effects) phase of the images should be handled carefully.

One can simply choose to handle the real and imaginary parts of the image as different channels, and apply the sparsifying transform on the complex-valued image and use ℓ_1-norm or Total Variation based minimization. A different approach includes taking magnitude of the image prior to applying sparsifying

Table3.2: **Algorithm:** *ADMM with Hybrid cost function for Unitary Transforms*

1. Set $k = 0$, choose $\mu > 0$, $\mathbf{z}_0^{(i)}$, $\mathbf{d}_0^{(i)}$, α_i for all i
2. **repeat**
3. $\mathbf{q}_k = \sum_{i=1}^{m} \mathbf{d}_k^{(i)} + \mathbf{z}_k^{(i)}$, calculate $\mathbf{A}\mathbf{q}_k$
4. $\mathbf{x}_{k+1} = \frac{1}{m}\left[\mathbf{q}_k + \frac{1}{m+1}\mathbf{A}^H\left(m\left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right) - \mathbf{A}\mathbf{q}_k\right)\right]$
5. $\mathbf{A}\mathbf{x}_{k+1} = \frac{1}{m+1}\left[\mathbf{A}\mathbf{q}_k + \left(\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}\right)\right]$.
6. **for i** $= \mathbf{1} \cdots \mathbf{m}$
7. $\quad \mathbf{z}_{k+1}^{(i)} = \mathbf{\Psi}_{\phi_i \frac{\alpha_i}{\mu}}\left(\mathbf{x}_{k+1} - \mathbf{d}_k^{(i)}\right)$
8. $\quad \mathbf{d}_{k+1}^{(i)} = \mathbf{d}_k^{(i)} - \mathbf{x}_{\mathbf{k+1}} + \mathbf{z}_{k+1}^{(i)}$
9. **endfor**
10. $\mathbf{z}_{k+1}^{(0)} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}\left(\mathbf{A}\mathbf{x} - \mathbf{d}^{(0)}\right)$
11. $\mathbf{d}_{k+1}^{(0)} = \mathbf{d}_k^{(0)} - \mathbf{A}\mathbf{x}_{k+1} + \mathbf{z}_{k+1}^{(0)}$
12. $k \leftarrow k + 1$
13. **until** some stopping criterion is satisfied.

transform, since we are mostly interested in magnitude images rather than phase images. In this case, instead of minimizing $\|\mathbf{W}\mathbf{x}\|_1$, we deal with $\|\mathbf{W}|\mathbf{x}|\|_1$, where $|\cdot|$ is the element-wise magnitude operator.

Now, let us go back to proximal mapping functions. The proposed method requires proximal mapping functions associated with each objective function. In most cases, it is easy to compute the proximal mapping function for a real valued function, whereas not straight-forward to compute proximal mapping of a magnitude-input function. For this purpose, let us make the following proposition:

Theorem 1 *If* $\text{prox}_{\phi}(\mathbf{v})$ *returns an all-positive output for an all-positive input, then:*

$$\text{prox}_{|\phi|}(\mathbf{v}) = \exp\{j\angle\mathbf{v}\}\,\text{prox}_{\phi}(|\mathbf{v}|), \tag{3.38}$$

where $\text{prox}_{|\phi|}(\mathbf{v})$ *denotes the proximal mapping function for the magnitude-input version of* $\phi(\cdot)$ *as* $\phi(|\cdot|)$*, and* $\angle\mathbf{v}$ *denotes a vector with entries consisting of complex-angle of each element in* $\mathbf{v}$*.*

Proof: The optimal solution to a optimization problem lies in the domain where

its subgradient is equal to 0. For complex-valued vector $\mathbf{v}$, the solution lies where the derivative with respect to both the real and imaginary parts are equal to zero. Let us write derivatives with respect to both real ($\Re$) and imaginary parts ($\Im$) as:

$$0 = \frac{\partial \phi(|\mathbf{x}|)}{\partial |\mathbf{x}|} \frac{\partial |\mathbf{x}|}{\partial \Re\{\mathbf{x}\}} + \mu \Re\{\mathbf{x} - \mathbf{v}\}, \tag{3.39}$$

$$0 = \frac{\partial \phi(|\mathbf{x}|)}{\partial |\mathbf{x}|} \frac{\partial |\mathbf{x}|}{\partial \Im\{\mathbf{x}\}} + \mu \Im\{\mathbf{x} - \mathbf{v}\}. \tag{3.40}$$

Now, let us add Eqn. (3.40) to Eqn. (3.39) by multiplying it with imaginary j.

$$0 = \frac{\partial \phi(|\mathbf{x}|)}{\partial |\mathbf{x}|} \left(\frac{\partial |\mathbf{x}|}{\partial \Re\{\mathbf{x}\}} + j \frac{\partial |\mathbf{x}|}{\partial \Im\{\mathbf{x}\}} \right) + \mu(\mathbf{x} - \mathbf{v}), \tag{3.41}$$

$$0 = \frac{\partial \phi(|\mathbf{x}|)}{\partial |\mathbf{x}|} \frac{\mathbf{x}}{|\mathbf{x}|} + \mu(\mathbf{x} - \mathbf{v}). \tag{3.42}$$

In Eq. (3.42), The equality must be satisfied for both real and imaginary parts. Angle of all terms except for $\mathbf{v}$ is equal to the angle of $\mathbf{x}$. Hence, for the equality to be satisfied for both real and imaginary parts, the angle of $\mathbf{v}$ is either exactly equal to $\mathbf{x}$, or the negative of it. Knowing this fact, let us move back to original proximal mapping definition in Eq. (3.21). We take the magnitude of the input $\mathbf{x}$ before using it. Hence, the angle of $\mathbf{x}$ does not affect the first term. For the second term to be minimal, $\angle \mathbf{x} = \angle \mathbf{v}$ must be satisfied. Using this fact back, let us go back to Eq. (3.42).

$$0 = \frac{\mathbf{x}}{|\mathbf{x}|} \left(\frac{\partial \phi(|\mathbf{x}|)}{\partial |\mathbf{x}|} + \mu(|\mathbf{x}| - \mathbf{v} \exp\{-j \angle \mathbf{x}\}) \right), \tag{3.43}$$

$$0 = \frac{\mathbf{x}}{|\mathbf{x}|} \left(\frac{\partial \phi(|\mathbf{x}|)}{\partial |\mathbf{x}|} + \mu(|\mathbf{x}| - |\mathbf{v}|) \right), \tag{3.44}$$

$$0 = \frac{\partial \phi(|\mathbf{x}|)}{\partial |\mathbf{x}|} + \mu(|\mathbf{x}| - |\mathbf{v}|). \tag{3.45}$$

Eq. (3.45) is solution to the optimization problem:

$$\arg\min_{|\mathbf{x}|} \phi(|\mathbf{x}|) + \frac{\mu}{2} \| |\mathbf{x}| - |\mathbf{v}| \|_2^2, \tag{3.46}$$

$$\arg\min_{\mathbf{x}} \phi(\mathbf{x}) + \frac{\mu}{2} \| \mathbf{x} - |\mathbf{v}| \|_2^2 \text{ subject to } \mathbf{x}[i] \geq 0, \text{ for all } i \tag{3.47}$$

Here, if proximal mapping of regular $\phi(\cdot)$ function returns an all-positive output for an all-positive input, then the solution to Eq. (3.47) becomes exactly equal to the proximal mapping of $\phi(|\cdot|)$, since the constraint is dropped. Hence,

$$\text{prox}_{|\phi|}(\mathbf{v}) = \exp\{\angle \mathbf{v}\} \, \text{prox}_{\phi}(|\mathbf{v}|), \tag{3.48}$$

for proximal mapping functions that return all-positive output for all-positive inputs. Also, this theorem suggests that one can needs to derive a method for calculation of the real-valued proximal mapping with positivity constraint to find the proximal mapping applied on the magnitude.

We use two functions with magnitude input, Total Variation (TV) and ℓ_1-norm of the magnitude vector. Let us begin with TV. Piecewise constant characteristics (such as natural images) can be better recovered using total variation. However, since the phase image in MRI does not necessarily need to be continuous, we consider applying this function to the magnitude of the image rather than complex image. Chambolle projection algorithm can be used as Moreau proximal mapping function for real-valued inputs [19]. However, there is not a well-known solution for total variation applied on the magnitude image. In this section, we derive the proximal mapping for magnitude TV and show that it can be used with Chambolle Projection algorithm.

Total variation belongs to a family of functions called bounded variation, and is defined in the continuous domain as the integral:

$$TV(\mathbf{x}) = \int_{\Omega} |\nabla \mathbf{x}(\mathbf{r})| d\mathbf{r}. \tag{3.49}$$

TV calculates the magnitude sum of gradients of a signal. The magnitude of the gradient operator ∇ calculates the Euclidian norm of the gradients. This operation is also rotation invariant, since the value of the integral does not depend on the direction that it is integrated on. Hence, using this isotropic definition of the TV, one can recover signals in a rotation invariant way, i.e. although we implement gradient operators in two dimensions as vertical (∇_v) and horizontal (∇_h), TV minimization can recover edges in any rotation.

In discrete setting, TV can explicitly be defined as the sum of gradients:

$$TV\left(|\mathbf{x}|\right) = \sum_{i,j} |\nabla \left(|\mathbf{x}|\right)|[i,j], \tag{3.50}$$

where the gradients are in two direction (horizontal, ∇_h, and vertical, ∇_v):

$$|\nabla \left(|\mathbf{x}|\right)|[i,j] = \sqrt{\left(\nabla_h |\mathbf{x}|\right)^2 + \left(\nabla_v |\mathbf{x}|\right)^2}, \tag{3.51}$$

and

$$(\nabla_h|\mathbf{x}|) = |\mathbf{x}[i+1,j]| - |\mathbf{x}[i,j]|, \tag{3.52}$$

$$(\nabla_v|\mathbf{x}|) = |\mathbf{x}[i,j+1]| - |\mathbf{x}[i,j]|. \tag{3.53}$$

To facilitate TV on the magnitude of the image, let us use Thm. 1. TV denoises a signal by minimizing the gradient of that signal. If the maximum value of a signal were to be increased, then that would result in a higher gradient change, as well as higher data fidelity term ($\|\mathbf{x}-\mathbf{v}\|_2^2$). For the total variation term to stay constant, the signal's mean value can be changed, which would simply result in a higher data fidelity term. Hence, the output of TV proximal mapping function can not be greater than the maximum value of the input signal at any point. The same procedure also holds for minimum value. Hence, the output of TV proximal mapping is bounded by the minimum and maximum of the input signal. Since we know that the minimum value of the input signal is zero, the output is always greater than zero. Hence using Thm. 1,

$$\text{prox}_{|TV|}(\mathbf{v}) = \exp\{\angle\mathbf{v}\}\,\text{prox}_{TV}(|\mathbf{v}|). \tag{3.54}$$

Chambolle projection is an iterative algorithm that approximately solves TV proximal mapping for real valued inputs. It is a fast-converging fixed point method that relies on minimization of the dual problem [19]. For the sake of completeness, the algorithm is also given in table 3.3. Aside from the above definitions, the algorithm also requires transpose of gradient operator, which is divergence defined as:

$$\nabla^T\mathbf{x} = \nabla_h^T\mathbf{x}_h + \nabla_v^T\mathbf{x}_v \tag{3.55}$$

$$\left(\nabla_h^T\mathbf{x}_h\right) = \mathbf{x}_h[i,j] - \mathbf{x}_h[i-1,j], \tag{3.56}$$

$$\left(\nabla_v^T\mathbf{x}_v\right) = \mathbf{x}_v[i,j] - \mathbf{x}_v[i,j-1]. \tag{3.57}$$

We also suggest applying the objective to ℓ_1-norm of the magnitude of the input, $\phi(\mathbf{x}) = \|\mathbf{W}|\mathbf{x}|\|_1$. Natural images are known to be sparse in some transformations such as Wavelet. Also, once we consider applying the transform to magnitude of the input, thm. 1 is not directly valid for calculating proximal mapping

Table3.3: **Algorithm:** *Proximal Mapping of TV (Chambolle Projection)*

1. Set $k = 0$, choose $\tau = 0.25$
2. **repeat**
3. $\mathbf{q}_k[i,j] = \mathbf{p}_k[i,j] + \tau\nabla(\nabla^T\mathbf{p}_k - \mathbf{v}/\lambda)[i,j]$
4. $\mathbf{p}_k[i,j] = \frac{\mathbf{q}_k[i,j]}{\max\{1,|\mathbf{q}_k[i,j]|\}}$
5. $k \leftarrow k+1$
6. **until** number of iterations are reached.

of $\phi(\mathbf{x}) = \|\mathbf{W}|\mathbf{x}|\|_1$, since the real valued counterpart does not always result in all-positive output for all-positive inputs. For this reason, one has to solve eq. (3.47) to calculate the proximal mapping. Due to its higher cost, we drop the transformation domain, and apply the ℓ_1-norm objective directly on the image magnitudes. ℓ_1-norm without any linear transformations satisfy thm. 1, as will be shown. Proximal mapping for magnitude ℓ_1-norm proximal mapping is also previously derived in [43].

$$\text{prox}_{\|\cdot\|_1}(\mathbf{v}) = \exp\{\angle\mathbf{v}\}\,\text{soft}(\mathbf{v}, 1/\mu), \tag{3.58}$$

where soft is the soft thresholding function defined as:

$$\text{soft}(\mathbf{x}, \tau) = \text{sign}(\mathbf{x}) \cdot \max\{|\mathbf{x}| - \tau, 0\}, \tag{3.59}$$

Here, since the sign of the input does not change, the condition of thm. 1 is satisfied, and eq. (3.58) can be used.

Here, we use the isotropic TV definition [54], which, although non-linear, is both proven to be compatible with natural images, and exhibits compressive sensing guarantees [44]. However, in this study we use TV on the magnitude images, which is a non-convex function. This function is a new function and does not necessarily need to exhibit the guarantees. Although further study should be done whether the guarantees hold, our study empirically suggests improved performance for complex-valued imaging.

We derived the necessary equations and formulations to use total variation and ℓ_1-norm with complex-valued imaging. These proximal mappings can be used with the proposed algorithms in sections 3.1 and 3.2.

3.4 Parameter Selection

In the previous sections, we proposed an algorithm to solve a given problem in the form of eq. (3.4). We simply followed the rules of the framework, and created the algorithm. In this section, we discuss how one can choose parameters for the described algorithm.

The proposed method requires choosing four parameters: ϵ, α_1, α_{TV}, and μ. As previously discussed, ϵ can be chosen as the expected noise power. The expected noise power can be measured directly for each patient by simply gathering data in MRI without any RF excitation. Since no slice is selected due to lack of excitation, the gathered data will consist of only noise. Also, this is a very fast process.

Next, we deal with the choice of regularization parameters α_1 and α_{TV}, associated with the objective functions ℓ_1-norm and TV. Since we use the constrained version, the two variables actually reduce to a single free variable. Hence, we simply set the α_1 to $(1 - \alpha_{TV})$. The choice of α_{TV} requires empirical studies. This choice entirely depends on the characteristics of the signal. However, as we will show in the results chapter, once it is set, the same value can be used across many patients as long as the same anatomy is imaged (e.g. brain).

Finally, μ needs to be selected. This is not a problem parameter, but a parameter related to the convergence of the algorithm. It can also be defined as $1/stepsize$, since choosing a greater μ results in a smaller step size, and a smaller μ results in a greater step size. Choosing a large step size results in a fast convergence at the beginning, however it may result in oscillation as the number of iterations increase. Choosing a small step size results in a higher accuracy, however it may take infeasible reconstruction times. There are different studies in the literature on how to select μ [12, 1]. Although convergence analysis requires a fixed μ, it is suggested in [1] to gradually increase μ to move the scale from faster convergence to higher accuracy. Another scheme is given in Chapter 3.4.1 of [12]. This scheme checks how the dual problem behaves, and updates step size parameter accordingly.

First, define residual ($\mathbf{r}$) and dual residual ($\mathbf{s}$) as:

$$\mathbf{r}_k = \mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} - \mathbf{c}, \tag{3.60}$$

$$\mathbf{s}_k = \mu \mathbf{G}^H \mathbf{H}(\mathbf{z}_{k+1} - \mathbf{z}_k). \tag{3.61}$$

For our problem, these can be defined as:

$$\mathbf{r}_k = \begin{bmatrix} \mathbf{A}\mathbf{x}_k - (\mathbf{d}_k^{(0)} + \mathbf{z}_k^{(0)}) \\ \mathbf{x}_k - (\mathbf{d}_k^{(1)} + \mathbf{z}_k^{(1)}) \\ \cdots \\ \mathbf{x}_k - (\mathbf{d}_k^{(m)} + \mathbf{z}_k^{(m)}) \end{bmatrix}, \mathbf{s}_k = -\mu \left(\mathbf{A}^H(\mathbf{z}_{k+1}^{(0)} - \mathbf{z}_k^{(0)}) + \sum_{i=1}^{m} (\mathbf{z}_{k+1}^{(i)} - \mathbf{z}_k^{(i)}) \right) \tag{3.62}$$

The update scheme of μ is given as [12, 32, 61]:

$$\mu_{k+1} = \begin{cases} \tau^{incr}\mu_k & \text{if } \|\mathbf{r}_k\|_2 > \rho \|\mathbf{s}_k\|_2 \\ \mu_k / \tau^{decr} & \text{if } \|\mathbf{s}_k\|_2 > \rho \|\mathbf{r}_k\|_2 \\ \mu_k & \text{otherwise} \end{cases} \tag{3.63}$$

Here, typical choices include $\rho = 10$, $\tau^{incr} = \tau^{decr} = 2$. Although more parameters are introduced to the problem, varying μ optimizes the step size through out iterations, instead of using a fixed pattern.

To illustrate the effect of μ on optimization, let us give brief results on a toy problem. The objective that we propose minimizing is an unconstrained ℓ_1-norm optimization problem:

$$\arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 + \lambda/2 \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2. \tag{3.64}$$

We use a forward model $\mathbf{A}$ with its elements drawn from independent identically distributed random Gaussian distribution of size 60×100. The data $\mathbf{y}$ is generated using a sparse vector of size 100, with 20 non-zero entries, in a noiseless manner. We set $\lambda = 3$, and see how the cost evolves versus iteration. Figure 3.1 shows what we explained in this section. Using larger μ values results in slower convergence as it can be seen for $\mu = 100$. While, for small number of iterations, using $\mu = 21.54$ has a poor convergence, for higher number of iterations this setting results in the most accurate result. Also, while decreasing μ results in a faster convergence speed, as it can be seen using very small μ causes the algorithm to oscillate.

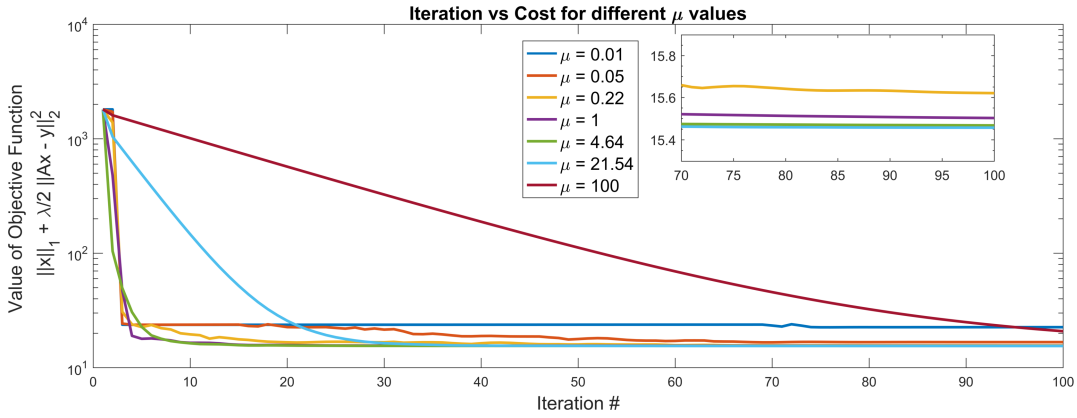


Figure 3.1: Evolution of cost function for different μ values

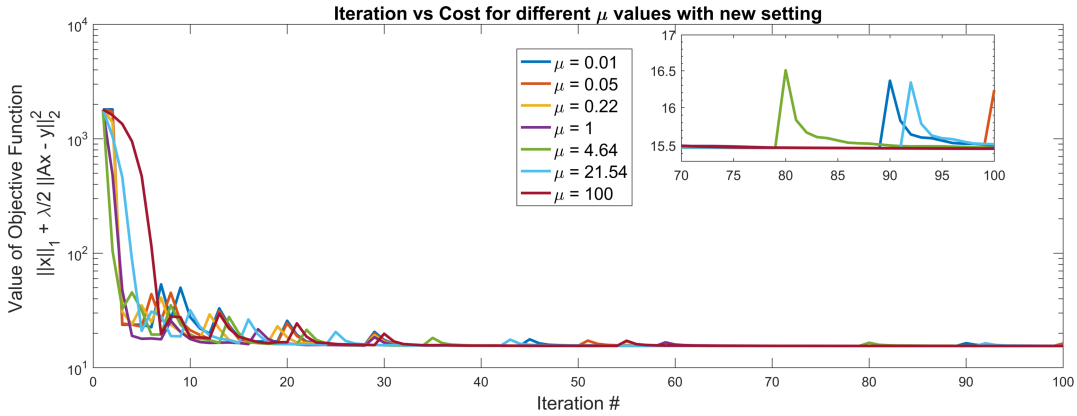


Figure 3.2: Evolution of cost function for different initial μ values with the update scheme in eq. (3.63).

Whereas, when we use the μ evolution scheme, although it may cause oscillation, the algorithm converges to a better point for various initial μ values, as can be seen in Fig. 3.2.

3.5 Chapter Summary

In this chapter, we dealt with the problem of single contrast reconstruction from undersampled data. We first defined the proposed approach, then proposed a framework to solve the problems of the proposed type. The proposed algorithm is an ADMM based method that requires a least squares operation and proximal mappings associated with each objective function. We then derived

a fast implementation method for the least squares operation for data models involving masked unitary transformations. MRI data model fits this definition exactly the reconstruction process can be accelerated. Then, we proposed using magnitude-input of the complex-valued images for TV, instead of the complex-valued input. We also made a proposition and shown that for a wide variety of functions including magnitude-TV, the associated proximal mappings can be computed using the proximal mapping associated with the real-valued TV. Finally, we discussed the parameter selection, and the step size.

CHAPTER 4

MULTI-CONTRAST OPTIMIZATION FRAMEWORK USING ADMM (MC-H-ADMM)

It is shown in previous studies that using multi-contrast data increases the reconstruction quality [34, 10, 29]. Hence, exploiting this additional information across separate contrasts of the same anatomy yields better results. In this chapter, we first describe the previous studies in multi-contrast joint compressive sensing reconstruction, and give a brief information on the theory of block compressive sensing. Then, we propose an algorithm that allows the reconstruction of multiple signals using joint and/or individual objective functions across contrasts. The proposed method describes a splitting scheme to allow solving an optimization problem provided that the proximal mappings associated with the objective functions are known.

4.1 Multi-Contrast Compressed Sensing for MRI

Exploitation of any prior information on the signal increases reconstructed signal quality. First, let us go back to Restricted Isometry Property (RIP) defined in eq. (2.10). Restricted Isometry Constant (δ) is a measure that helps determine whether a k-sparse signal can be reconstructed or not. When we move to multi-contrast reconstruction problems, RIP is no longer the tightest bound on the performance. A block restricted isometry property (BRIP), which is suitable for block or group compressive sensing, is defined as the lowest δ_b that satisfies the below equality for all k-block-sparse $\mathbf{x}$:

$$(1-\delta_b)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta_b)\|\mathbf{x}\|_2^2. \tag{4.1}$$

BRIP restricts the set that $\mathbf{x}$ can be chosen from, hence lowest bound of δ_b is equal to δ of RIP. The theory suggests that in the worst case, block compressive

sensing gives exactly the same result as compressive sensing provided that the signals are indeed block sparse.

Multi-channel signals or images consist of data gathered from the same scene using different parameters. For example, red, green and blue (RGB) channels of an optical camera gather data from the same scene in different wavelengths. A similar phenomenon exists in MRI. Gathering data using different **TE** and **TR** parameters results in different contrast images of the same underlying anatomy. Even though images are not exactly the same, the structural similarity can be exploited. Recent work in the literature shows that reconstructing these multi-channel/multi-contrast images jointly results in better images compared to individual reconstructions [10, 34, 36].

The problem model is depicted in Fig. 4.1. As it can be seen from the Fig. 4.1 (a), each contrast has individual data that can be expressed with $\mathbf{y}^{(i)} = \mathbf{A}^{(i)}\mathbf{x}^{(i)}$. Figure 4.1 (b) shows the overall forward model $\mathbf{A}$, which has a block-diagonal structure with $\mathbf{A}^{(i)}$ as its entries. Both image vector $\mathbf{x}$ and $\mathbf{y}$ are concatenated vectors with elements $\mathbf{x}^{(i)}$, and $\mathbf{y}^{(i)}$, respectively.

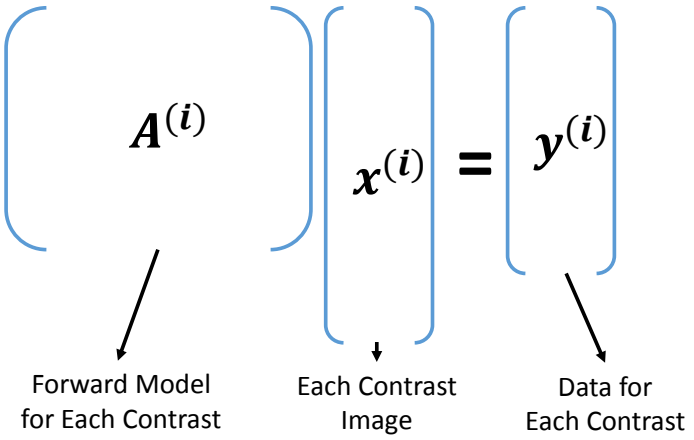
Now that we defined the model, let us move to an algorithm to clarify the design of an algorithm for multi-contrast compressed sensing. [34] solves the problems of type:

$$\underset{\mathbf{x}}{\textbf{minimize}}\, \alpha_1 JTV(\mathbf{x}) + \alpha_2 \|\mathbf{W}\mathbf{x}\|_{2,1} + \sum_i \|\mathbf{A}^{(i)}\mathbf{x}^{(i)} - \mathbf{y}^{(i)}\|_2^2, \tag{4.2}$$

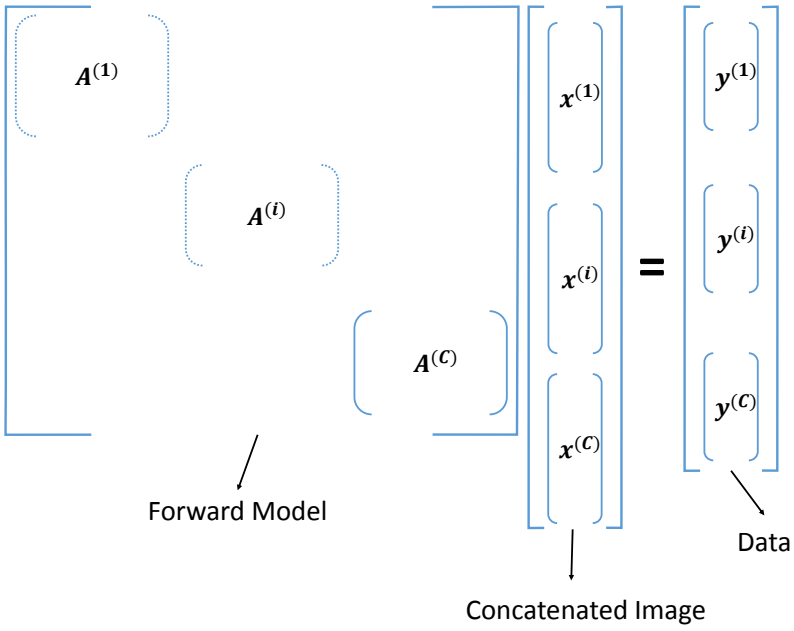
where $\mathbf{x}^{(i)}$ denotes i-th reconstructed channel of the image, $\mathbf{x}$ denotes the concatenated vector of each channel image, $\mathbf{y}^{(i)}$ and $\mathbf{A}^{(i)}$ denotes the data and forward model of channel i, respectively. Recall that $\mathbf{y}^{(i)}$ is the received signal, and $\mathbf{x}^{(i)}$ is the resulting image for each experiment in MRI. The objective functions JTV and $\|\cdot\|_{2,1}$ are defined below. α_1 and α_2 denote the respective weights of the regularized problem. $\|\cdot\|_{2,1}$ denotes the group sparsity function $\ell_{2,1}$-norm defined as:

$$\|\mathbf{x}\|_{2,1} = \sum_n \sqrt{\sum_i \left(\mathbf{x}^{(i)}[n]\right)^2}, \tag{4.3}$$

where n denotes the index of the vector. This formula is also applicable to 2D or 3D images, since we can consider an image concatenated into a vector. Just as



(a)



(b)

Figure 4.1: Forward models for multi-contrast imaging: (a) Linear Forward Model for each Contrast (b) Overall Linear Forward Model

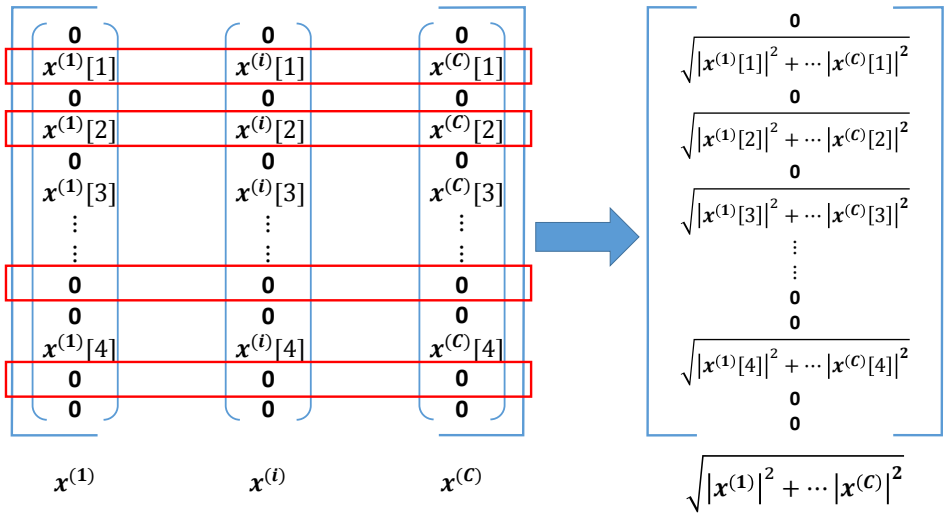


Figure 4.2: Application of group sparsity function to a group of vectors.

ℓ_1-norm, this function imposes sparsity in elements of $\mathbf{x}$. In addition to ℓ_1-norm, $\ell_{2,1}$-norm imposes sparsity on the same coordinates in different channels. This is a reasonable assumption since multi-channel signals have the same underlying anatomical structure. To further illustrate this function, it is depicted in Fig. 4.2. As it can be seen, each image is concatenated horizontally, then row-wise ℓ_2-norm operation is applied. Finally, ℓ_1-norm of the resulting vector is calculated.

JTV denotes the joint total variation function defined as:

$$JTV(\mathbf{x}) = \sum_n \sqrt{\sum_i \sum_m (\nabla_m \mathbf{x}^{(i)}[n])^2}, \tag{4.4}$$

where m is the direction of gradient. This function can also be represented as $\ell_{2,1}$-norm of gradient of an image. Just as the group sparsity constraint, this function is minimized when the change across different channels occur on the same coordinate. Since the underlying anatomy is the same, different channels are expected to have change on the same spot. Note that JTV is not $\ell_{2,1}$-norm of any linear transformation of each channel, since it also imposes the sparsity on the same coordinate for different sparsifying transforms.

Analysis and synthesis are two approaches for regularized image reconstruction. If the regularization is applied on $\mathbf{Wx} = \theta$, then the signal is **analysed** using $\mathbf{W}$

into θ. However, if the regularization is applied on θ, and image is reconstructed using $\mathbf{x} = \mathbf{W}^T\theta$, then the signal is **synthesised** using $\mathbf{W}$ and θ. We will give further information regarding these approaches in chapter 5. Majumdar et al. uses a synthesis based [42] and analysis based [41] approach to reconstruct images jointly. They propose solving the following problem for synthesis:

$$\begin{aligned} &\arg\min_{\theta} \|\theta\|_{2,1} \\ &\text{subject to } \|\mathbf{y}^{(i)} - \mathbf{A}^{(i)}\mathbf{W}^T\theta_i\|_2 \leq \epsilon_i. \end{aligned} \tag{4.5}$$

Here, the signal is synthesised using redundant wavelet transformation dictionary.

In another study, they propose analysis formulation as:

$$\begin{aligned} &\arg\min_{\mathbf{x}} \|\mathbf{W}\mathbf{x}\|_{2,1} \\ &\text{subject to } \|\mathbf{y}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(i)}\|_2 \leq \epsilon_i. \end{aligned} \tag{4.6}$$

Here, like the previous study, the signal is analysed using redundant wavelet transformation dictionary. Both studies work with multi-contrast images, and utilize group sparsity for reconstruction.

Another study by [10] approaches the problem using a Bayesian framework. This framework considers the compressed sensing approach as maximum a posteriori (MAP) estimation of a sparse signal and Laplacian noise model. The method first calculates horizontal and vertical gradients using MAP estimations, then uses least squares to combine the gradients to reconstruct an image. The proposed algorithm solves the problem for real and imaginary parts of the image separately, and combines them. Even though the proposed method produces quality results, it is very slow. However, the paper notes that spatial registration of different channels may be required prior to running the algorithm.

4.2 A Generic Multi-Channel ADMM for Constrained Optimization

As described in previous section, the sparsifying transform has the greatest effect on the reconstructed image quality. Crafting better sparsifying transformations

allows the reconstruction of signals from less data. As Block Restricted Isometry Property (BRIP) dictates in Eq. (4.1), block processing of signals results in better reconstruction results provided that the signals are group sparse. As in single-contrast reconstruction we further improve reconstruction quality using multiple dictionaries. The multiple dictionaries may include both joint functions such as group sparsity ($\|\cdot\|_{2,1}$) and individual functions such as sparsity ($\|\cdot\|_1$). Adding more sparsifying transforms restricts the solution space, leading to improved performance of image reconstruction.

As previously given in section 2.1, using different echo time (**TE**) and repetition time (**TR**) results in different contrast images, since different parameters of a tissue are enhanced with each parameter set. This results in images with same overall features, especially in the edges. Also, if some part of the image is non-zero in the image domain in some contrast, then it is more likely for the same part of the image to be non-zero in other contrasts. Hence, imposing sparsity on a block, rather than each pixel individually increases the reconstruction performance. Also, we impose group version of TV function, Joint Total Variation (JTV) function. This function imposes group sparsity on the gradients, hence it forces edges to be at the same parts of the image.

Group sparsity is applied on a set of vectors. The function first takes ℓ_2-norm in across vector direction, then takes ℓ_1-norm of the resulting single vector. The ℓ_2-norm operation first groups multiple vectors in a single vector. Then, the ℓ_1-norm takes magnitude sum. Using group sparsity as an objective function results in imposing sparsity on the grouped vector. Hence, the each pixel of the reconstructed signals are likely to either be zero in all contrasts, or non-zero in all contrasts. Group sparsity is defined as:

$$\|\mathbf{x}\|_{2,1} = \sum_n \sqrt{\sum_i \left(\mathbf{x}^{(i)}[n]\right)^2}, \tag{4.7}$$

JTV is an extended version of TV across contrasts. As in group sparsity, it first takes ℓ_2-norm across contrasts, then takes ℓ_1-norm of the resulting image. However, contrary to group sparsity, ℓ_2-norm of the gradient vectors are calculated, rather than direct image. This imposes sparsity in the same pixel in both the horizontal and vertical gradient of each contrast. Therefore, if some pixel is

to be non-zero, then it is more than likely that that pixel will correspond to a vertical and/or horizontal edge in all contrasts. JTV is defined as:

$$JTV(\mathbf{x}) = \sum_{n} \sqrt{\sum_{i} \sum_{m} \left(\nabla_m \mathbf{x}^{(i)}[n]\right)^2}, \tag{4.8}$$

Now let us describe the methodology for the suggested algorithm. We consider the problem model:

$$\mathbf{y}^{(i)} = \mathbf{A}^{(i)}\mathbf{x}^{(i)} + \mathbf{n}^{(i)}, \tag{4.9}$$

where i is the channel number, and $\mathbf{n}^{(i)}$ is the noise vector associated with each channel. Here, the proposed method splits and solves for multiple constraint functions.

We develop an algorithm that solves problems of type:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \alpha_1\phi_1\left(\mathbf{x}\right) + \cdots + \alpha_m\phi_m\left(\mathbf{x}\right) \\ \text{subject to} & \|\mathbf{A}^{(i)}\mathbf{x}^{(i)} - \mathbf{y}^{(i)}\|_2 \leq \epsilon_i, \text{ for i = 1 to C,} \end{array} \tag{4.10}$$

where C is the number of channels, $\phi_i(\cdot)$ denotes separable either individual or joint functions for each channel. ϵ_i is chosen related to the power of the noise of each channel as in single channel reconstruction.

Let us employ a splitting scheme as:

$$\mathbf{z} = [(\mathbf{z}^{(0)})^T(\mathbf{z}^{(1)})^T \cdots (\mathbf{z}^{(c)})^T]^T, \mathbf{x} = [(\mathbf{x}^{(1)})^T \cdots (\mathbf{x}^{(C)})^T]^T, \tag{4.11}$$

$$\mathbf{z}^{(i)} = [(\mathbf{z}^{(i,0)})^T(\mathbf{z}^{(i,1)})^T \cdots (\mathbf{z}^{(i,m)})^T]^T. \tag{4.12}$$

Hence, we create a dual variable per channel and objective function. Next, we define the linear transforms in ADMM to solve the defined problem Eq. (4.10):

$$\mathbf{G} = -\begin{bmatrix} \mathbf{G}_1 & \cdots & 0 \\ & \cdots & \\ 0 & \cdots & \mathbf{G}_C \end{bmatrix}, \mathbf{G}_i = -\begin{bmatrix} \mathbf{A^{(i)^T}} \ \mathbf{I} \cdots \mathbf{I} \end{bmatrix}^{\mathbf{T}}, \tag{4.13}$$

Here, sizes of each $\mathbf{G}_i$ may vary, since the number of $\mathbf{I}$'s are equal to number of objective functions associated with that contrast. Hence, this setting allows

adding some specific transformation domains to specific signals aside from the joint functions. We also set, $\mathbf{H} = \mathbf{I}, \mathbf{c} = 0, f_1(\mathbf{x}) = 0$.

The objective function associated with the dual variables are:

$$f_2(\mathbf{z}) = \sum_{i=1}^{C} \iota_{E(\epsilon_i, \mathbf{I}, \mathbf{y}^{(i)})} \left(\mathbf{z}^{(i,0)}\right) + \sum_{t=1}^{m} \alpha_t \phi_t \left(\{\mathbf{z}^{(i,t)}\}_{i=1\cdots C}\right). \tag{4.14}$$

The resulting problem ensures data fidelity is satisfied for each channel, $\mathbf{z}^{(i,0)}$. For the given definitions, let us break down the linear constraints :

$$\mathbf{G}\mathbf{x} + \mathbf{H}\mathbf{z} = \mathbf{c} \tag{4.15}$$

$$\mathbf{A}^{(i)}\mathbf{x}^{(i)} = \mathbf{z}^{(i,0)}, \text{ for all i} \tag{4.16}$$

$$\mathbf{x}^{(i)} = \mathbf{z}^{(i,1)}, \text{ for all i} \tag{4.17}$$

$$\cdots \tag{4.18}$$

$$\mathbf{x}^{(i)} = \mathbf{z}^{(i,m)}, \text{ for all i} \tag{4.19}$$

This setting ensures the solved problem is exactly equal to Eq. (4.10).

Now let us derive the associated steps of the algorithm. The algorithm consists of the generic iterations in Eqns. (3.13), (3.14), and (3.15). Each step can be further simplified. Eq. (3.13) can be written in the following way:

$$\mathbf{x}_{k+1} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H(\mathbf{d}_k + \mathbf{z}_k) \tag{4.20}$$

$$\mathbf{x}_{k+1}^{(i)} = \left(m\mathbf{I} + (\mathbf{A}^{(i)})^H\mathbf{A}^{(i)}\right)^{-1}\left[\sum_{t=1}^{m} \mathbf{d}_k^{(i,t)} + \mathbf{z}_k^{(i,t)} + (\mathbf{A}^{(i)})^H\left(\mathbf{d}_k^{(i,0)} + \mathbf{z}_k^{(i,0)}\right)\right]. \tag{4.21}$$

Hence, $\mathbf{x}$-update step can be handled separately for each contrast. Eq. (3.14) can be re-written for multi-contrast reconstruction as:

$$\begin{aligned}\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} &\sum_{i=1}^{C} \iota_{E(\epsilon_i, \mathbf{I}, \mathbf{y}^{(i)})} \left(\mathbf{z}^{(i,0)}\right) \\ &+ \sum_{t=1}^{m} \alpha_t \phi_t \left(\{\mathbf{z}^{(i,t)}\}_{i=1\cdots C}\right) + \frac{\mu}{2}\| \left(\mathbf{G}\mathbf{x}_{k+1} + \mathbf{H}\mathbf{z} + \mathbf{d}_k\right) \|_2^2\end{aligned} \tag{4.22}$$

$$\mathbf{z}_{k+1}^{(i,0)} = \arg\min_{\mathbf{z}^{(i,0)}} \iota_{E(\epsilon_i, \mathbf{I}, \mathbf{y}^{(i)})} \left(\mathbf{z}^{(i,0)}\right) + \frac{\mu}{2}\| \left(-\mathbf{A}^{(i)}\mathbf{x}_{k+1}^{(i)} + \mathbf{z}^{(i,0)} + \mathbf{d}_k^{(i,0)}\right) \|_2^2 \tag{4.23}$$

$$\mathbf{z}_{k+1}^{(i,t)} = \arg\min_{\mathbf{z}^{(i,t)}} \alpha_t \phi_t \left(\{\mathbf{z}^{(i,t)}\}_{i=1\cdots C}\right) + \frac{\mu}{2}\sum_{i=1}^{C} \|\mathbf{z}^{(i,t)} - (\mathbf{x}_{k+1}^{(i)} - \mathbf{d}_k^{(i,t)})\|_2^2 \text{ for } t = 1\cdots m \tag{4.24}$$

Table4.1: **Algorithm:** *Multi-Contrast ADMM with Hybrid cost function (MC-H-ADMM)*

1. Set $k = 0$, choose $\mu > 0$, $\mathbf{z}_0^{(i,t)}$, $\mathbf{d}_0^{(i,t)}$, α_t for all i, t
2. **repeat**
3. **for i** $= \mathbf{1} \cdots \mathbf{C}$
4. $\quad \mathbf{x}_{k+1}^{(i)} = \left(m\mathbf{I} + (\mathbf{A}^{(i)})^H \mathbf{A}^{(i)}\right)^{-1} \left[\sum_{t=1}^{m} \mathbf{d}_k^{(i,t)} + \mathbf{z}_k^{(i,t)} + (\mathbf{A}^{(i)})^H \left(\mathbf{d}_k^{(i,0)} + \mathbf{z}_k^{(i,0)}\right)\right]$
5. $\quad \mathbf{z}_{k+1}^{(i,0)} = \mathbf{\Psi}_{\iota_{E(\epsilon_i,\mathbf{I},\mathbf{y}^{(i)})}} \left(\mathbf{A}^{(i)}\mathbf{x}_k^{(i)} - \mathbf{d}_k^{(i,0)}\right)$
6. $\quad \mathbf{d}_{k+1}^{(i,0)} = \mathbf{d}_k^{(i,0)} - \mathbf{A}^{(i)}\mathbf{x}_{k+1}^{(i)} + \mathbf{z}_{k+1}^{(i,0)}$
7. **endfor**
8. **for t** $= \mathbf{1} \cdots \mathbf{m}$
9. $\quad \{\mathbf{z}^{(i,t)}\}_{i=1\cdots C} = \mathbf{\Psi}_{\frac{\alpha_t}{\mu}\phi_t} \left(\{\mathbf{x}_{k+1}^{(i)} - \mathbf{d}_k^{(i,t)}\}_{i=1\cdots C}\right)$ for $t = 1 \cdots m$
10. $\quad \mathbf{d}_{k+1}^{(i)} = \mathbf{d}_k^{(i)} - \mathbf{x_{k+1}} + \mathbf{z}_{k+1}^{(i)}$, for all i
11. **endfor**
12. $k \leftarrow k + 1$
13. **until** some stopping criterion is satisfied.

The Eqns. (4.23) and (4.24) are Moreau proximal mapping for joint or individual reconstruction functions. If $\phi_t(\cdot)$ is an individual function applied on each contrast separately, then the proximal mapping can also be calculated separately. However, for joint case, joint proximal mapping functions should be used. We can re-write these functions as:

$$\mathbf{z}_{k+1}^{(i,0)} = \mathbf{\Psi}_{\iota_{E(\epsilon_i,\mathbf{I},\mathbf{y}^{(i)})}} \left(\mathbf{A}^{(i)}\mathbf{x}_k^{(i)} - \mathbf{d}_k^{(i,0)}\right), \text{ for } i = 1 \cdots C \tag{4.25}$$

$$\{\mathbf{z}^{(i,t)}\}_{i=1\cdots C} = \mathbf{\Psi}_{\frac{\alpha_t}{\mu}\phi_t} \left(\{\mathbf{x}_{k+1}^{(i)} - \mathbf{d}_k^{(i,t)}\}_{i=1\cdots C}\right) \text{for } t = 1 \cdots m \tag{4.26}$$

And, Eq. (3.15) can be calculated using:

$$\mathbf{d}_{k+1}^{(i,t)} = \mathbf{d}_k^{(i,t)} - \mathbf{x}_{k+1}^{(i)} + \mathbf{z}_{k+1}^{(i,t)}, \text{for } i = 1 \cdots C, t = 1 \cdots m \tag{4.27}$$

$$\mathbf{d}_{k+1}^{(i,0)} = \mathbf{d}_k^{(i,0)} - \mathbf{A}^{(i)}\mathbf{x}_{k+1}^{(i)} + \mathbf{z}_{k+1}^{(i,0)}, \text{for } i = 1 \cdots C. \tag{4.28}$$

For the sake of completeness, the algorithm is given in Table. 4.1.

Step 4 of the algorithm can be implemented in a faster way using the ideas developed in section 3.2.

Table4.2: **Algorithm:** *Proximal Mapping of JTV (Color Chambolle Projection)*

1. Set $k = 0$, choose $\tau = 0.25$
2. **repeat**
3. $\mathbf{q}_k^{(c)}[i,j] = \mathbf{p}_k^{(c)}[i,j] + \tau\nabla(\nabla^T\mathbf{p}_k^{(c)} - \mathbf{v}^{(c)}/\lambda)[i,j]$, for all c
4. $\mathbf{r}[i,j] = \sqrt{\sum_{c=1}^{C}|\mathbf{q}_k^{(c)}[i,j]|^2}$
5. $\mathbf{p}_k^{(c)}[i,j] = \frac{\mathbf{q}_k^{(c)}[i,j]}{\max\{1,|\mathbf{r}_k[i,j]|\}}$, for all c
6. $k \leftarrow k+1$
7. **until** number of iterations are reached.

4.3 Handling Phase for Multi-Contrast Complex Imaging

In this chapter, we describe how we employ the described algorithm to compressive multi-contrast MRI reconstruction problem. We use a linear combination of JTV and group sparsity functions as objective functions. However, similar to single contrast imaging, handling phase requires further care.

First, let us begin with JTV. Let us define JTV on the magnitude as:

$$|JTV|(\mathbf{x}) = \sum_n \sqrt{\sum_i \sum_m \left(\nabla_m |\mathbf{x}^{(i)}[n]|\right)^2}. \tag{4.29}$$

Thm. 1 states that if proximal mapping of JTV results in an all-positive output for an all-positive input, then we can directly use real-valued proximal mapping function. The principles that hold for single-contrast TV also holds for JTV. If any point in the output were to be below the minimum of the input, then both the value of the JTV function, and the data fidelity term would increase in value. Hence, the proximal mapping of JTV must return all-positive output for an all-positive input. Using thm. 1, we conclude that the proximal mapping for JTV can be expressed as:

$$\text{prox}_{|JTV|}(\mathbf{v}) = \exp\{\angle\mathbf{v}\}\,\text{prox}_{JTV}(|\mathbf{v}|). \tag{4.30}$$

Now, let us deal with the proximal mapping of JTV function. The previously described Chambolle's method does not directly work here. The extended method can be found in Table 4.2 [13].

We also propose imposing group sparsity constraint. We select the groups as the

corresponding pixels across different contrasts. The proximal mapping function of group sparsity function is well known and can be computed using:

$$\text{prox}_{\|\cdot\|_{2,1}/\tau}(\mathbf{v}) = \mathbf{v} \cdot \max\left\{0, 1 - \frac{\tau}{\sqrt{\sum_{i=1}^{C} |\mathbf{v}^{(i)}|^2}})\right\} \tag{4.31}$$

Furthermore, for all positive inputs, the proximal mapping returns all-positive outputs. Hence, by Thm. 1 proximal mapping of complex-valued inputs can be calculated using Eq. (4.31).

Using the outlined proximal mapping functions, one can solve the problem:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \alpha_{G1}\|\mathbf{x}\|_{2,1} + \alpha_{JTV} JTV\,(\mathbf{x}) \\ \text{subject to} & \|\mathbf{A}^{(i)}\mathbf{x}^{(i)} - \mathbf{y}^{(i)}\|_2 \leq \epsilon_i, \text{ for i = 1 to C} \end{array}, \tag{4.32}$$

4.4 Group Lp-norm

We further extend the study to handle ℓ_p-norm based solutions. Compressive sensing deals with sparsity, which can be expressed using ℓ_0-pseudo norm. The solution to problem given in Eq. (2.8) results in the exact recovery of the signal of interest. However, as discussed before, it is a non-convex problem, and can not be solved efficiently. It requires a combinatorial search algorithm to solve it, which is classified as NP-HARD. Let us now define the ℓ_0-norm as:

$$\|\mathbf{x}\|_0 = \sum_{i=1}^{N} |\mathbf{x}[i]|^0, \tag{4.33}$$

where, 0^0 is assumed to be 0. We call this pseudo-norm because this formulation does not satisfy the triangular inequality to be a norm function. Actually, a family of norm functions can be defined using:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{N} |\mathbf{x}[i]|^p\right)^{1/p}, p \geq 1 \tag{4.34}$$

$p \geq 1$ is required because any ℓ_p-norm with $p < 1$ does not satisfy triangular inequality and constitute non-convex functions, which are again classified as NP-HARD. For this very reason, we use the convex relaxation of ℓ_0-pseudo norm,

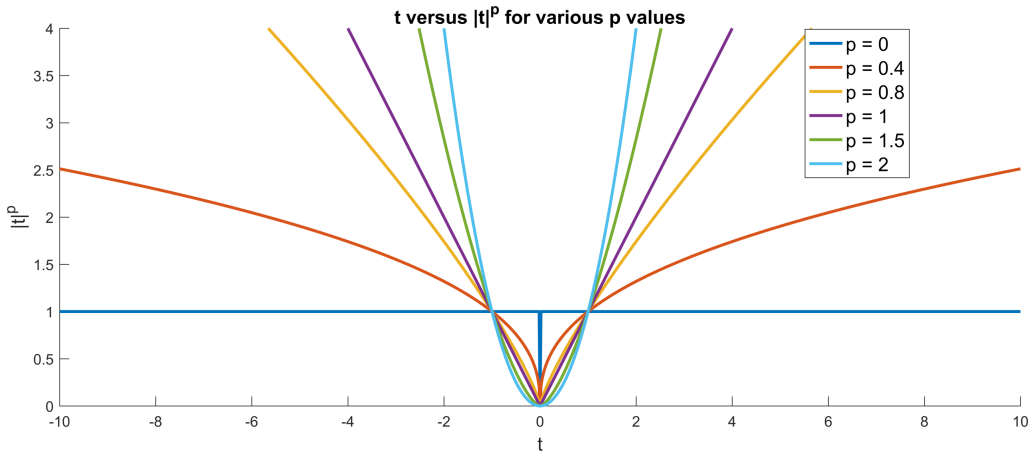


Figure 4.3: ℓ_p-norm function for various p values.

which is ℓ_1-norm, and under certain conditions both optimization problems yield the exact same results.

Although $p < 1$ can not be classified as norms, the problem is still studied in the literature, and some promising results exist [21]. ℓ_p-norm with $p < 1$ is a better relaxation compared to ℓ_1-norm, and gives sparser results. Since in compressed sensing we desire sparser results, we propose using this formulation for image recovery. Also, Fig. 4.3 shows values of $|t|^p$ for various p values. As it can be seen, the desired function p is a horizontal line except for at $t = 0$. As p is increased the function goes further away from the function with $p = 0$. Due to the shapes of the costs, we penalize the positive values of the vector, and hence the minimization problem results in sparser results for $p < 1$, since they penalize small values just as much as large values in the vectors. The closest convex function is at $p = 1$.

In this section, we propose using group p-norm sparsity, and develop a proximal mapping function associated with that function. We define the group p-norm sparsity as:

$$\|\mathbf{x}\|_{2,p} = \sum_n \left(\sum_i \left(\mathbf{x}^{(i)}[n] \right)^2 \right)^{p/2}. \tag{4.35}$$

The proximal mapping associated with this function does not have an analytical solution. Hence, we deal with approximate solution. However, as given in [12],

approximate solutions does not prevent ADMM from converging to a solution. Here, note that ADMM still does not have a convergence guarantee. First, let us define group 2-norm as:

$$\|\mathbf{v}\|_2[n] = \sqrt{\sum_i \left(\mathbf{x}^{(i)}[n]\right)^2}. \tag{4.36}$$

The solution to this problem is given as [20]:

$$\text{prox}_{\|\cdot\|_{2,p}/\tau}(\mathbf{v}) = \mathbf{v} \cdot \max\left\{0, 1 - \tau^{2-p}\|\mathbf{v}\|_2^{p-2}\right\}. \tag{4.37}$$

Here, for $p = 1$ the proximal mapping reduces to regular group sparsity.

4.5 Individual Sparsity versus Group Sparsity

Now, to demonstrate the power of group sparsity over individual sparsity, let us assume a toy problem. Compressive sensing is known to work well with forward model matrices ($\mathbf{A}$) that the elements of which are drawn from independent identically distributed random Gaussian distribution.

We assumed an unknown signal of $n = 100$, with sparsity level $s = 20$, and number of measurements of $m = 50$. We assumed a group of $C = 3$ signals, such that the signals are group sparse. The measurements are calculated using $\mathbf{y} = \mathbf{A}\mathbf{x}^{(i)}$, with no noise. We solved three problems for individual sparsity, group sparsity and group ℓ_p-sparsity. Three individual problems were solved for individual sparsity using a regularization of ℓ_1-norm. For group sparsity and group ℓ_p-sparsity, we solved the previously defined problems. The proposed methods were used to solve all problems, and all problems have converged to their respective optimal solutions.

Individual sparsity problem can be expressed as in Eq. (2.11):

$$\begin{aligned} \underset{\mathbf{x}^{(i)}}{\textbf{minimize}} \quad & \|\mathbf{x}^{(i)}\|_1 \\ \textbf{subject to} \quad & \mathbf{A}^{(i)}\mathbf{x}^{(i)} = \mathbf{y}^{(i)} \end{aligned}, \tag{4.38}$$

while group sparsity can be expressed as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \|\mathbf{x}\|_{2,1} \\ \text{subject to} \quad & \mathbf{A}^{(i)}\mathbf{x}^{(i)} = \mathbf{y}^{(i)}, \text{ for i = 1 to C} \end{aligned}. \tag{4.39}$$

The results helped demonstrate the goal. Fig. 4.4 shows the signal to be reconstructed, and the reconstruction errors for various problems. As it can be seen, for this sparsity level, exploiting joint sparsity improved reconstruction quality. Moreover, using ℓ_p-norm instead of ℓ_1-norm further decreased the reconstruction error for some p values. Since group ℓ_p-norm is a non-convex function, increasing or decreasing p values may not always result in the best reconstruction quality. As it can be seen $p = 0.9$ has the least root mean squared error (RMSE), computed for each pixel of the multi-channel signal as:

$$RMSE(\mathbf{x}[l]) = \sqrt{\frac{1}{N}\sum_{l=1}^{C} |\mathbf{x}_{ref}^{(i)}[l] - \mathbf{x}^{(i)}[l]|^2}. \tag{4.40}$$

We further analysed the results in terms of the convergence of the algorithm. We fixed the μ value across all problems. Fig. 4.5 shows the convergence result for different cost functions. For convergence, we computed peak signal-to-noise ratio (pSNR) as:

$$pSNR(\mathbf{x}) = 20\log\left(\frac{\text{peak intensity of the signal}}{\|\mathbf{x} - \mathbf{x}_{ref}\|_2}\right) \tag{4.41}$$

Figure 4.5 shows the change in mean PSNR for all channels per iteration. As it can be seen, Group-$\ell_{0.9}$-norm enforced sparsity resulted in the highest quality signal.

4.6 Chapter Summary

In this chapter, we dealt with the problem multi-contrast compressive sensing. We first described the theory behind block-compressive sensing. Then, we moved on to MRI and assumed sparsity in a linear combination of two domains, joint TV and group-sparsity ($\ell_{2,1}$-norm). Then, we extended the algorithm in the previous chapter to handle multi-contrast imaging. We described phase handling, and assumed sparsity on magnitude-image rather than complex-valued image. Next, we described a non-convex pseudo-norm, group-ℓ_p-norm (with $p < 1$), which is a better approximation of sparsity, i.e. ℓ_0-norm. However, since it is a non-convex function, global minimum is not guaranteed to be found using the

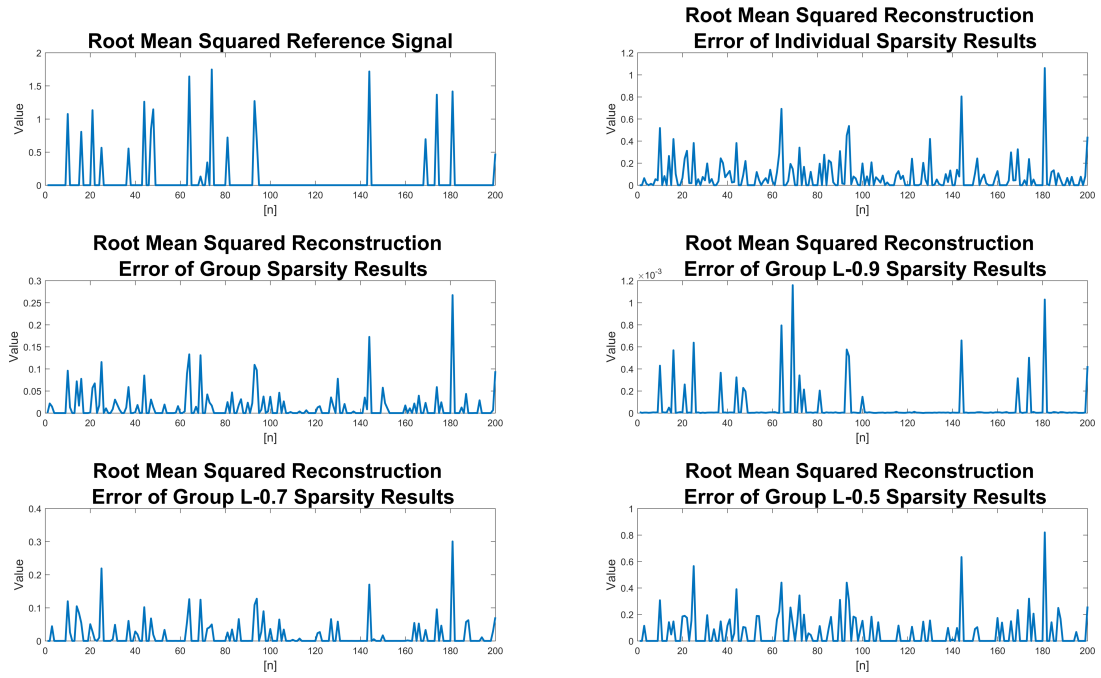


Figure 4.4: Row-wise ℓ_2-norm of the reference signal (top-left), Row-wise reconstruction error of individual sparsity (top-right), Row-wise Reconstruction error of group sparsity (middle-left), Row-wise Reconstruction error of group $\ell_{0.9}$ sparsity (middle-right), Row-wise Reconstruction error of group $\ell_{0.7}$ sparsity (bottom-left), Row-wise Reconstruction error of group $\ell_{0.5}$ sparsity (bottom-right),

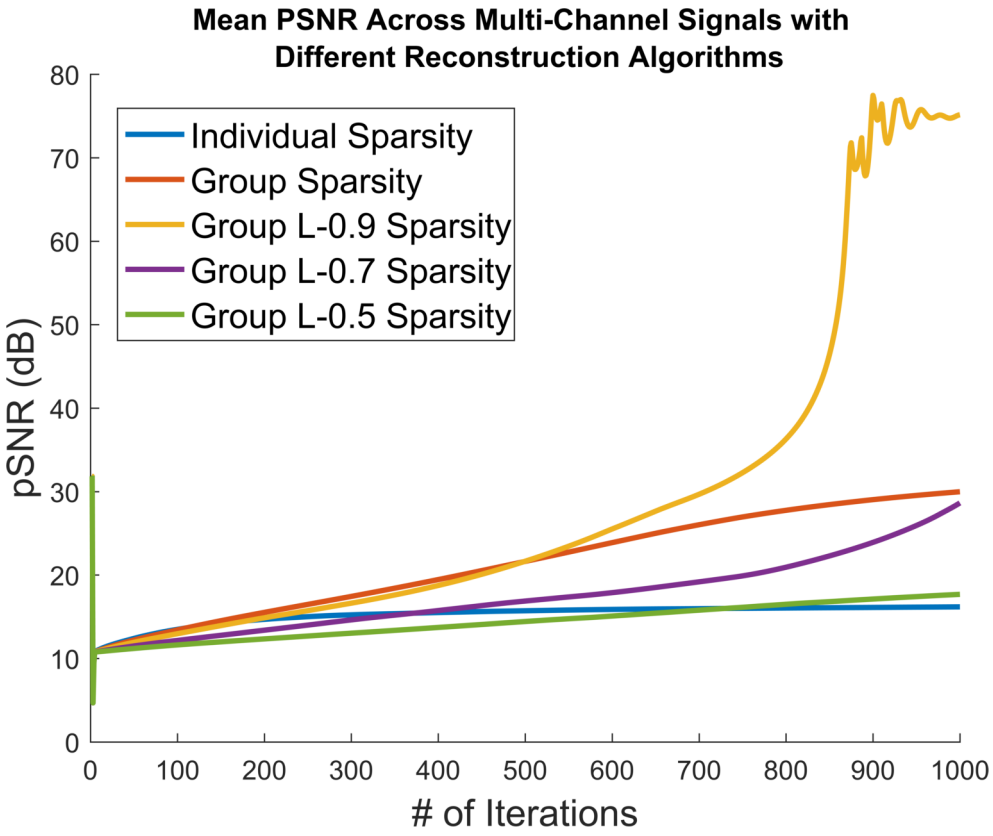


Figure 4.5: Mean of PSNR versus iteration for five reconstruction methods. Each color line in the figure is calculated by taking the mean of PSNR values of all channels.

proposed algorithm. We replaced group-ℓ_p-norm with group sparsity function, and showed the power of group-ℓ_p-norm on a toy problem.

CHAPTER 5

DICTIONARY LEARNING BASED MULTI-CONTRAST OPTIMIZATION FRAMEWORK (MC BCS MRI)

In previous chapters, we dealt with the problem of single contrast and multi-contrast compressive sensing for MRI reconstruction. We manually hand-crafted a transformation domain for which the image is expected to be sparse. We assumed the images are sparse in a redundant dictionary of linear combination of total variation and image domain ℓ_1-norm. However, hand-crafted domains may not yield the best possible performance. For this reason, in this chapter, we deal with a concept called dictionary learning. Dictionary learning is a tool in signal processing that is used to find a transformation domain that yields a sparse representation for a given signal. The idea has many uses such as in the domains of feature extraction, compression, and denoising. From compressive sensing perspective, finding a transformation that yields a sparse representation for a given signal is required for the recovery of the signal. In this chapter, we first give a background on the studies in dictionary learning, then propose a framework for multi-contrast dictionary learning only from gathered data points.

5.1 Dictionary Learning

For many applications, such as image compression or feature extraction, representation of data using fewer number of samples is necessary. However, this requires some information contained in the signal to be either omitted or stored in a way that allows reconstruction of the signal. If the signal of interest is sparse, then the zeros in the signal can simply be omitted. However, most signals such as natural images are not sparse. Linear transformations such as Wavelet or discrete cosine transform are known to sparsify natural images. In

fact, JPEG2000 standard makes use of sparsity of wavelet coefficients of natural images [5]. A different way to look at sparse representation is that, the differentiating information within the signal of interest is embedded into the sparsifying transform. Hence, the problem of feature extraction can also make use of sparsifying transformations. Compressive sensing can reconstruct sparse signals from fewer data. For problems that the signals themselves are not sparse, a sparsifying transformation is required. However, such transformations may not be always so well-defined.

Overcomplete curvelet, wavelet, short time Fourier transform, discrete cosine transform, steerable wavelet are some of the transforms that natural images are known to be sparse on [5]. The main motivation behind dictionary learning or vector quantization problem is to create a better dictionary than a predetermined one for the training input.

Dictionary learning deals with the problem of representing signals using linear bases. Each basis is called an atom and the atoms are combined to produce a dictionary. Let us denote signals $\mathbf{y}$ concatenated as columns into a matrix with $\mathbf{Y}$, and the dictionary with $\mathbf{W}$. Then, dictionary learning recovers a sparsifying transformation $\mathbf{W}$ that satisfies both $\mathbf{Y} = \mathbf{WX}$, and $\|\mathbf{X}\|_0 \leq s$, for some sparsity level s.

Synthesis and Analysis are the two main approaches to dictionary learning. Synthesis approach can be formulated as

$$\begin{aligned} \underset{\mathbf{W},\mathbf{X}}{\textbf{minimize}} \quad & \|\mathbf{WX} - \mathbf{Y}\|_F^2 + Q(\mathbf{W}) \\ \textbf{subject to} \quad & \|\mathbf{X}\|_0 \leq S, \end{aligned} \tag{5.1}$$

where s denotes the level of sparsity. Q denotes the cost function associated with the dictionary, $\mathbf{W}$. As it can be seen, the signal $\mathbf{y}$ is synthesised using dictionary $\mathbf{W}$ and the representation $\mathbf{x}$. Analysis approach is formulated as:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{X}}{\textbf{minimize}} \quad & \|\mathbf{X} - \mathbf{Y}\|_F^2 + Q(\mathbf{W}) \\ \textbf{subject to} \quad & \|\mathbf{WX}\|_0 \leq S. \end{aligned} \tag{5.2}$$

In this approach, the transformation $\mathbf{W}$ analyses signal $\mathbf{x}$ to get the sparse representation. The dictionary here is the same as the transformations in the

previous section. Since the outlined problem is non-convex even for a fixed dictionary due to ℓ_0-norm, many applications change it with ℓ_1-norm [4, 52, 39]. This makes both (5.2) and (5.1) formulations **bi-convex** for $\mathbf{W}$ and $\mathbf{Y}$, i.e. keeping one of these variables fixed, results in a convex problem.

Synthesis based approaches gained much attraction and is widely used [42]. However, analysis based approaches are not so far behind. For under-complete or square transforms, the approaches are exactly equivalent and choosing either one results in the same solution. For the over-complete case however, these approaches differ and the superiority of one is still a debated topic in literature [28].

One of the main problems in dictionary learning is the size of the signal. The required computational time for the algorithms grow quickly as the data size increases. Also, the required number of samples increase along with the data size. For example, if there exists an n-dimensional data, then the required number of samples N is at least greater than n to avoid trivial solutions. In that case, the most trivial solution would be for the dictionary to memorize signals as its atoms, and use 1-sparse signals to represent the signals. The memorized transformation is indeed not what is expected of the learned dictionary. To avoid this, one possible solution is to impose a structure on the dictionary. One example is to allow a dictionary of small size, then apply it on patches of the signal [57]. Another would be to choose $Q(\mathbf{W})$ such that $\mathbf{W}$ becomes a unitary matrix.

For the next sections in this chapter, we denote n as the dimension of data, N as the number of samples, and K as the number of dictionary atoms. Hence, the definitions here are of the sizes:

$$\mathbf{Y} \in R^{n \times N}, \mathbf{W} \in R^{n \times K}, \mathbf{X} \in R^{K \times N}. \tag{5.3}$$

Many methods have been proposed for the solution to the problems defined by Eq. (5.1) and Eq.(5.2). Different cost (Q), and different formulations are employed in these methods as will be described in section 5.2.

5.2 Methods for Dictionary Learning

As mentioned in the previous section, the dictionary learning problem is a non-convex optimization problem and is NP-Hard [24]. Hence, finding exact solution is not a practical goal. Some methods use matching pursuit (MP) or Orthogonal Matching Pursuit (OMP) type algorithms for an approximate solution [4]. Others use the convex relaxation formulation of the problem in (5.1) and solve using basis pursuit algorithms [39].

A previous work by [46] approaches the dictionary learning using maximum likelihood estimation, and maximize the probability that a set of training signals $\mathbf{Y}$ with elements consisting of $\mathbf{y}_i$ are sparse, given a dictionary $\mathbf{W}$, with the assumption that training signals are independent.

$$\underset{\mathbf{W}}{\textbf{maximize}}\ P(\mathbf{Y}|\mathbf{W}) \tag{5.4}$$

$$\underset{\mathbf{W}}{\textbf{maximize}}\ \Pi_i P(\mathbf{y}_i|\mathbf{W}). \tag{5.5}$$

They define a hidden variable as sparse codes, and integrate over it. The final problem yields

$$\underset{\mathbf{W},\mathbf{X}}{\textbf{minimize}}\ \|\mathbf{Y}-\mathbf{W}\mathbf{X}\|_2^2+\lambda\rho(\mathbf{X}), \tag{5.6}$$

where $\rho(\cdot)$ denotes the logarithm of prior assumption on the sparse signal such as log summation of the super-Gaussian distribution.

Method of optimal direcitons (MOD) is a dictionary learning algorithm closely related to k-means clustering algorithm [4]. The algorithm attacks the same problem by dividing it into two steps. The algorithm alternates between minimizing for $\mathbf{W}$ and $\mathbf{X}$. It uses a coordinate-descent like structure for minimization and at each iteration solves for one of the variables. Solving the problem for $\mathbf{X}$ is called **sparse coding** and solving for $\mathbf{W}$ is called **dictionary update**. The algorithm uses Newton's iterations to solve the dictionary update step, and OMP or FOCUSS to solve the sparse coding step. However, the algorithm has very high computational complexity, and is not practically applicable for high dimensionality.

One of the most famous method in dictionary learning is called K-SVD [4]. The method relies on the generalization of a similar algorithm, k-means. The algorithm solves synthesis approach in (5.1) with $Q(\mathbf{W}) = 0$. Similar to MOD, the algorithm alternates between minimizing for $\mathbf{W}$ and $\mathbf{X}$. The improvement brought to dictionary update step is the differentiating factor of K-SVD.

Before going further into K-SVD, let us discuss k-means algorithm. K-means clustering is an algorithm that finds a representation for a given data set using fewer number of points by clustering them. It represents each cluster using only one data point. Hence, it searches for both the data points that represents the clusters, and a mapping between each data point and cluster. For example, if a set of signals $\mathbf{Y}$ is to be represented using only one data point with minimum squared error, then the mean of all data points would represent the data the best. For more data points, the problem can be represented using an optimization problem as:

$$\begin{aligned} &\underset{\mathbf{W},\mathbf{X}}{\textbf{minimize}} \ \|\mathbf{WX} - \mathbf{Y}\|_F^2, \\ &\text{subject to} \ \ \|\mathbf{X}_i\|_0 = 1, \end{aligned} \tag{5.7}$$

where $\mathbf{X}_i$ is the $i-th$ column of the matrix $\mathbf{X}$. Here, K is the number of rows in $\mathbf{X}$, $\mathbf{W}$ is the transformation domain, or the points selected by the algorithm for sparse representation. Hence, it represents the center points of the clusters, while columns of $\mathbf{X}$ corresponds to the cluster for which the data point in $\mathbf{Y}$ belongs to. The constraint imposes each data point to belong to only one cluster. Hence, $\mathbf{X}$ is the described mapping between data points and the cluster. Here, the problem Eq. (5.7) is again, a non-convex problem and only heuristic algorithms exist to solve it. K-means alternates between solving for $\mathbf{W}$ and $\mathbf{X}$.

At this point, we may prefer to represent each data point using more than one data point of $\mathbf{W}$, for example at most s. Then, the problem at hand becomes:

$$\begin{aligned} &\underset{\mathbf{W},\mathbf{X}}{\textbf{minimize}} \ \|\mathbf{WX} - \mathbf{Y}\|_F^2, \\ &\text{subject to} \ \ \|\mathbf{X}_i\|_0 \leq s. \end{aligned} \tag{5.8}$$

Although the sparse representation can be imposed on each training sample separately as in Eq. (5.8), it can also be imposed on the whole matrix, as $\|\mathbf{X}\|_0 \leq$

S, where $S = N \times s$ [52]. This is an extension of k-means algorithm. Now let us go back to K-SVD algorithm. The algorithm alternates between sparse coding and dictionary update steps.

In the sparse coding step, any number of compressive sensing algorithms as well as least absolute shrinkage and selection operator (LASSO) solver algorithms can be used. The sparse coding step is represented as solving Eq. (5.1) for $\mathbf{X}$:

$$\begin{aligned} \underset{\mathbf{X}}{\textbf{minimize}} \quad & \|\mathbf{WX} - \mathbf{Y}\|_F^2 \\ \textbf{subject to} \quad & \|\mathbf{X}\|_0 \leq S, \end{aligned} \tag{5.9}$$

where $\mathbf{W}$ and $\mathbf{Y}$ are constants. Here, if sparse representation of each sample is required separately, then the problem reduces to:

$$\begin{aligned} \underset{\mathbf{X}_i}{\textbf{minimize}} \quad & \|\mathbf{WX}_i - \mathbf{Y}_i\|_2^2 \\ \textbf{subject to} \quad & \|\mathbf{X}_i\|_0 \leq s, \end{aligned} \tag{5.10}$$

for all i.

For dictionary update, the algorithm keeps $\mathbf{X}$ constant, and solves for $\mathbf{W}$:

$$\underset{\mathbf{W}}{\textbf{minimize}} \ \|\mathbf{WX} - \mathbf{Y}\|_F^2, \tag{5.11}$$

$$\underset{\mathbf{W}}{\textbf{minimize}} \ \|E_k - \mathbf{W}_k\mathbf{X}_k^T\|_F^2, \tag{5.12}$$

where $\mathbf{X}_k$ and $\mathbf{W}_k$ denote the elements of $\mathbf{X}$ and $\mathbf{W}$, respectively. E_k is defined as $\mathbf{Y} - \sum_{j \neq k} \mathbf{W}_j\mathbf{X}_j^T$. The problem resembles a previously solved problem using K singular value decomposition (SVD) operations. Hence the algorithm is called K-SVD. At each iteration, SVD operations are performed K times to find the optimal solution for the dictionary update.

The algorithm is considered a *natural generalization* of the algorithm k-means clustering algorithm, and it is preferred due to its simplicity and power of generating sparse coding books. The algorithm K-SVD is used for many applications such as compression, denoising and compressive sensing [4, 27]. The algorithm can be used to learn both over-complete and under-complete dictionaries. K-SVD is summarized as in table 5.1.

In this thesis, we deal with compressed multi-contrast MRI reconstruction. We deal with group sparsity, and hence the learned dictionary must be associated

Table5.1: **Algorithm:** *K-SVD*

1. Choose initial $\mathbf{W}^{(0)} \in R^{n \times K}$ with ℓ_2-normalized columns
2. **repeat**
3. **Sparse Coding Stage:** Solve for approximate solution of Eq. (5.9) or Eq. (5.10) as:

$$\underset{\mathbf{X}}{\textbf{minimize}} \|\mathbf{WX} - \mathbf{Y}\|_F^2$$
$$\textbf{subject to} \|\mathbf{X}\|_0 \leq S$$

4. **Dictionary Update Stage:** For each column $k = 1, \cdots, K$ in $\mathbf{W}^{(J-1)}$

4a. Define the group of examples that use this atom $\mathbf{W}_k = \{i | 1 \leq i \leq N, \mathbf{X}_i^k \neq 0\}$, so that only samples that are associated with this atom is used. (In k-means algorithm, this step is simply calculated as the mean of the associated samples.)

4b. Compute the overall representation error matrix E_k, using $E_k = \mathbf{Y} - \sum_{j \neq k} \mathbf{W}_j \mathbf{X}_j^T$.

4c. Restrict E_k by choosing only the columns corresponding to $\mathbf{W}_k$, and obtain E_k^R, the restricted matrix.

4d. Apply Singular Value Decomposition on the restricted error matrix $E_k^R = U \Delta V^T$. Choose the updated dictionary column $\mathbf{W}_k^{(J)}$ as the first column of U. Update coefficient vector $\mathbf{W}_k^R$ as the first column of $V \times \Delta(1,1)$.

5. $J \leftarrow J + 1$
6. **until** some stopping criterion is satisfied.

with group sparse structures for multi-contrast data. An extension of K-SVD algorithm is related to the work in this thesis [40]. In [40], the algorithm is extended for color images, hence can be applied to multi-channel signals. The K-SVD algorithm is extended to handle 2D RGB images by extending the dictionary size to a 3^{rd}-dimension for color, such that the dictionary is of the size $\mathbf{W} \in R^{n \times n \times 3}$. However, while allowing learning multi-channel dictionaries, the algorithm learns dictionaries that promote gray levelness in the image. They avoid the problem by changing the definition of inner product, and adding a pre-conditioner matrix. The choice of the matrix is empirical and a wrong choice results in poor learned dictionaries.

Another extension to K-SVD algorithm involves using wavelet-like multi-scale dictionaries [47]. The algorithm produces dictionaries in multi-scale. Hence, the dictionary can be applied in a fast manner with scales. Hence, it leads to faster update rules, and allows high dimensional dictionaries to be learned. Thus, multi-scale dictionaries have the potential to outperform single-scale dictionaries. Different pyramid-like structures are considered for fast reconstruction. A similar work involves doubly-sparse dictionaries, which are sparse dictionaries that apply to a signal in a sparsifying transformation domain [50, 57]. Recent work on KSVD is related to learning structured dictionaries for application of the algorithm to high dimensional signals.

One common problem among these algorithms are the problem of initialization. Since the problem is non-convex, initialization from a well-known sparsifying transformation such as DCT may result in convergence to a better minima, in addition to faster convergence [27].

5.3 Dictionary Learning for Compressed Sensing

Compressed sensing requires sparsifying transformation domains. Given a set of signals, dictionary learning can be used to find dictionaries that can sparsify these signals. Better sparsity domains allow the signals to be recovered using fewer number of samples, as dictated by restricted isometry property in

Eq. (2.10). In this section, we give brief information on how these methods can be utilized within the context of compressive sensing.

As discussed in sections 5.1 and 5.2, numerous dictionary learning methods can be used for denoising, feature extraction and compression. All the discussed methods can be incorporated into compressive sensing, since the output of most of these methods are $\mathbf{W}$. For example, one can use K-SVD to find a sparsifying transformation domain. Then, use that domain to recover a signal from fewer number of measurements. Different types of compressive sensing recovery algorithms should be used for different types of K-SVD algorithm, such as analysis or synthesis. However, there may not always exist enough number of samples to train a dictionary learning algorithm. In that case, one simple trick is to use some type of structure to reduce the number of unknowns in the transform domain ($\mathbf{W}$) [52, 50, 57].

Here, we first categorize algorithms into two. Recovery of only the dictionary $\mathbf{W}$ and the sparse codes $\mathbf{B}$ from given signals $\mathbf{x}$ is called "Offline Dictionary Learning", while recovery of all three vectors from given data sets are called "Online Dictionary Learning", or blind compressive sensing (BCS). In offline dictionary learning, the algorithms solve problems of type Eq. (5.1) or Eq. (5.2). Then, the recovered dictionaries are general purpose in the signals of interest, and can be used for various purposes. However, BCS assumes the signal that will be recovered is going to be sparse in some transformation domain, and the transform domain is not known a-priori. Hence, it is called transformation blind compressive sensing.

Offline dictionary learning algorithms are mostly discussed in the previous sections. One can simply recover a transformation domain using the methods given in the previous sections, then use that transformation domain to solve the associated compressed sensing problem. Since this is a two step process, for these type of algorithms to work, some samples from the signals of interest are required. These samples has to be fully sampled. One can again use the tricks discussed above (such as structured dictionaries applied to patches of the image, or multi-scale sparse structures) to decrease number of required samples. How-

ever, at least one sample is required. Also, the choice of the samples may affect the recovered transformation domain. Hence, for example if the transformation is applied on an MRI image, then it should include most features seen in MRI. This is a potential pitfall for dictionary learning algorithms, since the results are affected by choice of the sample set. Acknowledging potential problems one can always choose to use offline algorithms and recover a plausibly good transformation domain prior to application of CS. However, online algorithms approach the problem in a different way.

Online dictionary learning is an inherently harder problem to solve, and can be expressed as:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{X},\mathbf{B}}{\textbf{minimize}} \quad & \|\mathbf{WX}-\mathbf{B}\|_2^2+\nu\|\mathbf{AX}-\mathbf{Y}\|_F^2+Q(\mathbf{W}) \\ \textbf{subject to} \quad & \|\mathbf{B}\|_0 \leq s \end{aligned} . \tag{5.13}$$

Here, the critical difference of this problem is the knowledge of the image. Here, the image itself ($\mathbf{X}$) is not known, since it is not fully sampled. The image in some transformation domain ($\mathbf{A}$) is known, which already poses an underdetermined problem. The discussed problem in Eq. (5.13) reduces to offline problems, if $\mathbf{A}=\mathbf{I}$, and $\nu \to \infty$. Hence, online dictionary learning problem is more generic compared to offline counterpart. Moreover, if only one sample of $\mathbf{X}$ exists, such that it is a vector that can be expressed with $\mathbf{x}$, then the problem reduces to:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{X},\mathbf{B}}{\textbf{minimize}} \quad & \|\mathbf{Wx}-\mathbf{B}\|_2^2+\nu\|\mathbf{Ax}-\mathbf{y}\|_F^2+Q(\mathbf{W}) \\ \textbf{subject to} \quad & \|\mathbf{B}\|_0 \leq s \end{aligned} . \tag{5.14}$$

This problem is inherently harder, since the number of samples is assumed to be 1. In this case, to avoid trivial solutions, some structure on the $\mathbf{W}$ is a must. To avoid this problem, [51] assumes patch-wise sparsity and also imposes unitary-ness on the transformation. They solve the problem:

$$\min_{\mathbf{x},\mathbf{W},\mathbf{B}} \nu\|\mathbf{Ax}-\mathbf{y}\|_2^2+\sum_{j=1}^{N}\|\mathbf{WP}_j\mathbf{x}-b_j\|_2^2+\lambda Q(\mathbf{W}) s.t. \|\mathbf{B}\|_0 \leq s, \|\mathbf{x}\|_2 \leq C, \tag{5.15}$$

where $Q(\mathbf{W})$ penalizes choice of transformation matrix to both avoid trivial solutions, and enforce other wanted qualities. In [51], this function is chosen

as $-log|det\mathbf{W}| + 0.5\|\mathbf{W}\|_F^2$ to avoid trivial solutions, scale ambiguity and also enforce unitary solutions for $\mathbf{W}$ as λ goes to infinity. Here, C is the bound on the energy of the solution, which is usually ignored as it is set to ∞, and $\mathbf{P}_j$ is the patch selection matrix that selects patches from $\mathbf{x}$. Note that this structure is the synthesis based structure.

Here, $\mathbf{P}_j$ can be visualized as in Fig. 5.1. As it can be seen, overlapping patches of size 6×6 are selected in this example. One can also consider random patches, or decrease the amount of overlapping to avoid dealing with large amount of data. In that case, the summation term over all patches would reduce to summation for specific patch indices, j.

In [51], other problem models have also been proposed. These models include sparsity of overall system instead of every single patch, different selection of $Q(\mathbf{W})$ such that it becomes the indicator function for $\mathbf{W}^H\mathbf{W} = \mathbf{I}$, and the one with the sparsity penalty instead of using it as a constraint. These models all have their respective advantages, and disadvantages. However, the overall framework stays the same. The specific advantage of this framework is to recover dictionaries specific to data, hence the training set is also the test set within this context. Hence, the requirement for the training set to include most probable structures within the image family is no longer needed. However, note that this problem is highly non-convex.

Now let us discuss the implementation of the algorithm with $Q(\mathbf{W})$ being the indicator function for unitary $\mathbf{W}$. The algorithm consists of three steps for each variable. It is organized in a block-coordinate descent type structure, hence it iteratively minimizes for each variable and alternates between them. Like K-SVD, the first and second steps are the sparse coding and the dictionary update steps. In sparse coding step, the problem to be solved becomes:

$$\min_{\mathbf{B}} \sum_{j=1}^{N} \|\mathbf{W}\mathbf{P}_j\mathbf{x} - b_j\|_2^2 s.t. \|\mathbf{B}\|_0 \leq s. \tag{5.16}$$

Now, let us define $\mathbf{Z}$ as the concatenated vectors with entries as $\mathbf{W}\mathbf{P}_j\mathbf{x}$. In that case, the problem becomes:

$$\min_{\mathbf{B}} \|\mathbf{Z} - \mathbf{B}\|_F^2 s.t. \|\mathbf{B}\|_0 \leq s. \tag{5.17}$$

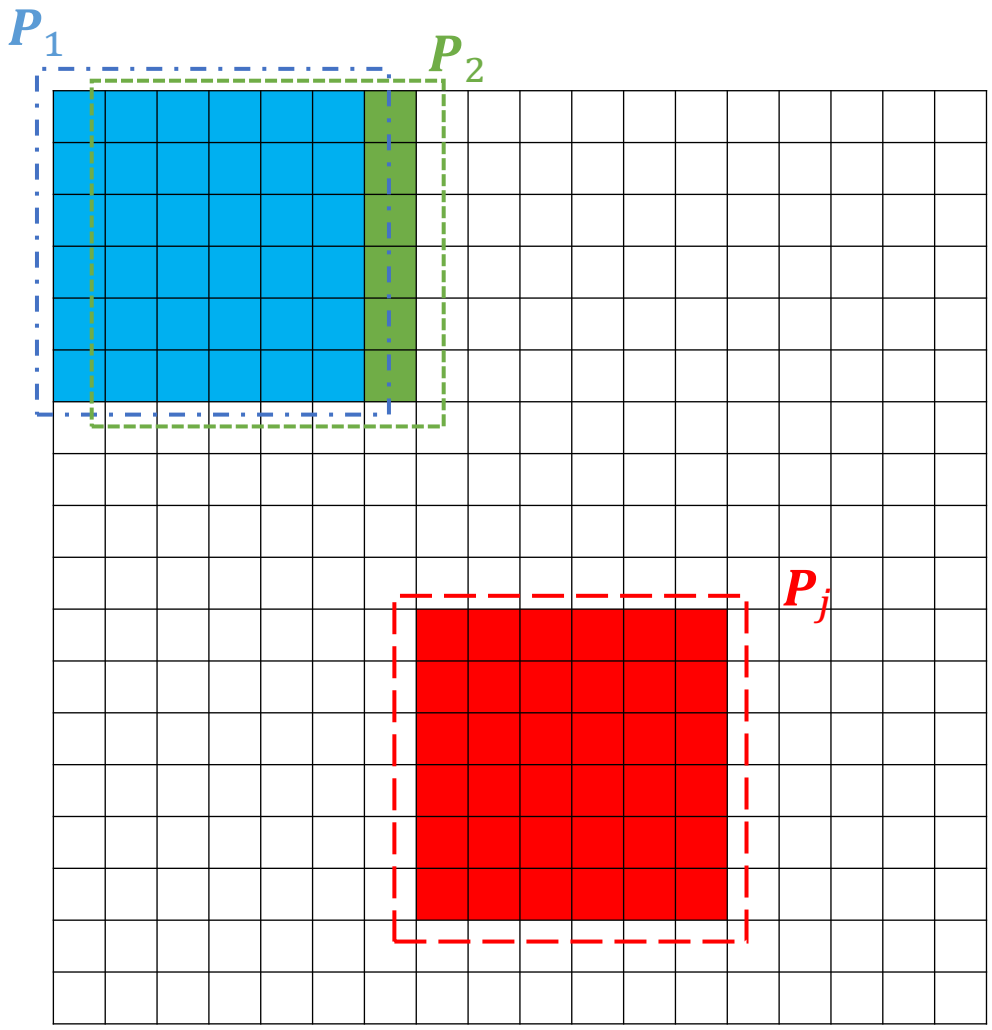


Figure 5.1: Visualization of patch selection matrix $\mathbf{P}_j$, in Eq. (5.15).

This is a projection onto a non-convex set problem. Here, the optimal sparse codes are simply maximum-s hard thresholded version of $\mathbf{Z}$, hence:

$$\mathbf{B}^*[i] = \begin{cases} \mathbf{Z}[i] & \text{if } \mathbf{Z}[i] \text{ is in the top s elements} \\ 0 & \text{otherwise} \end{cases} . \tag{5.18}$$

The next step is the dictionary update step, which is essentially the solution of the problem:

$$\min_{\mathbf{W}} \sum_{j=1}^{N} \|\mathbf{W}\mathbf{P}_j\mathbf{x} - b_j\|_2^2 + \lambda \iota_{\mathbf{W}unitary}(\mathbf{W}) . \tag{5.19}$$

As in sparse coding step, let us define $\mathbf{X}$ as the concatenated vector with $\mathbf{P}_j\mathbf{x}$ as its entries. Then, the problem becomes:

$$\min_{\mathbf{W}} \|\mathbf{W}\mathbf{X} - \mathbf{B}\|_F^2 + s.t.\mathbf{W}^H\mathbf{W} = \mathbf{I}. \tag{5.20}$$

The solution to the outlined problem has been recently derived, and can be implemented using [51]:

$$\mathbf{W}^* = \mathbf{V}\mathbf{U}^H, \tag{5.21}$$

where $\mathbf{X}\mathbf{B}^H$ have a full singular value decomposition of $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$.

The final step includes an image update step, and requires the solution of:

$$\min_{\mathbf{x}} \nu\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \sum_{j=1}^{N} \|\mathbf{W}\mathbf{P}_j\mathbf{x} - b_j\|_2^2 s.t. \|\mathbf{x}\|_2 \leq C. \tag{5.22}$$

In this study, we deal with structures with no energy bound on $\mathbf{x}$, and hence $C = \infty$. The solution is then,

$$\left(\sum_{j=1}^{N} \mathbf{P}_j^T\mathbf{W}^H\mathbf{W}\mathbf{P}_j + \nu\mathbf{A}^H\mathbf{A}\right)\mathbf{x} = \sum_{j=1}^{N} \mathbf{P}_j^T\mathbf{W}^H b_j + \nu\mathbf{A}^H\mathbf{y}. \tag{5.23}$$

In that case, the algorithm including the solution for MRI is given in table 5.2.

5.4 Multi-Contrast Blind Compressive Sensing

In the previous sections, we have worked with two established methods for dictionary learning, one for online and one for offline dictionary learning. Although

Table5.2: **Algorithm:** *Blind Compressive Sensing MRI (BCS MRI) [51]*

1. Initialize $(\mathbf{W}^0, \mathbf{B}^0, \mathbf{x}^0)$.
2. **for t = 1:J**
3. Form the matrix $\mathbf{X}$ by stacking $\mathbf{P}_j\mathbf{x}$ as its columns.
   Compute $L^{-1} = \left(\mathbf{X}\mathbf{X}^H + 0.5\lambda\mathbf{I}\right)^{-1/2}$.
4. **for l = 1:M**
5. Transform Update:
5.a. Set $W = \mathbf{V}\mathbf{U}^H$, where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \mathbf{X}\mathbf{B}^H$.
6. Sparse Coding:
6.a. Update $\mathbf{B}$ according to Eq. (5.18).
7. **end**
8. Image Update:
8.a. Compute $c = \sum_{j=1}^{N} \mathbf{P}_j^T\mathbf{W}^H b_j$, $S = FFT(c)$.
8.b. Compute a_1 as the first column of $c = \sum_{j=1}^{N} \mathbf{P}_j^T\mathbf{W}^H\mathbf{W}\mathbf{P}_j$.
8.c. Set $\gamma = \sqrt{p} \times FFT(a_1)$.
8.d. Set Fourier transform of $\mathbf{x}$ with indices (k_x, k_y) to

$$\begin{cases} \frac{S(k_x,k_y)}{\gamma(k_x,k_y)} & , \text{ if } (k_x,k_y) \notin \mathbf{M} \\ \frac{S(k_x,k_y)+\nu\mathbf{A}^H\mathbf{y}}{\gamma(k_x,k_y)+\nu} & , \text{ if } (k_x,k_y) \in \mathbf{M} \end{cases}.$$

9. **end**

there exists some study on multi-channel imaging in the previous methods such as color-ksvd, the algorithm is specialized for color images, and is an offline dictionary learning technique such that it works in the presence of previous data. However, in this thesis, we deal with compressed sensing of only single measurement. Hence, we move towards more compressed sensing compatible techniques such as transformation blind compressed sensing [51]. Moreover, we deal with the problem of multi-contrast compressed sensing. To the best of our knowledge, there is not an extension of such algorithms for multi-contrast dictionary learning. In this section, we propose an algorithm for online dictionary learning and reconstruction of multi-contrast MRI.

The algorithm in section 5.3 solves problems of type Eq. (5.15). The solved problem can be extended for multi-contrast MRI. The simplest way to extend the optimization problem is to use group ℓ_0 sparsity, instead of ℓ_0 sparsity. We have discussed the advantages of group sparsity over sparsity in section 4.1. We have also compared the results on individual and group sparsity on section 4.5, and shown the advantages of group sparsity over individual sparsity. However, the resulting dictionaries in the previous section are ambiguous in terms of location. For example, interchanging the rows of $\mathbf{W}$ along with the corresponding locations of $\mathbf{B}$ results in the exact same result. However, changing the locations may cause problems for multi-contrast imaging. Hence, the proposed algorithm has to recover a dictionary for all contrasts at once for group sparsity approach to work.

In this study, we propose extending Eq. (5.15) as:

$$\min_{\mathbf{x},\mathbf{W},\mathbf{B}} \sum_i \left(\nu \|\mathbf{A}^{(i)}\mathbf{x}^{(i)} - \mathbf{y}^{(i)}\|_2^2 + \sum_{j=1}^{N} \|\mathbf{W}\mathbf{P}_j\mathbf{x}^{(i)} - b_j^{(i)}\|_2^2 \right) + \lambda Q(\mathbf{W}) s.t. \|\mathbf{B}\|_{2,0} \leq s, \tag{5.24}$$

where $\|\cdot\|_{2,0}$ is defined as the number of non-zero entries in a group of vectors, which corresponds to the sparse codes of different contrasts. Although it has been previously shown that analysis [41] or synthesis [42] based redundant wavelet transforms result in group sparse images, there has been not been a study on learned dictionaries. Here, we optimize for a single transformation for all contrast images. This constrains the solution space to a much smaller space,

since the data size is tripled. We expect this version to have better convergence speed, as well as a better sparsifying transform.

As in [51], let us further derive the required steps for the implementation of the algorithm. We follow similar steps to [51], and update the necessary update step for sparse code update by solving:

$$\min_{\mathbf{B}} \sum_{i} \left(\sum_{j=1}^{N} \|\mathbf{W}\mathbf{P}_j\mathbf{x}^{(i)} - b_j^{(i)}\|_2^2 \right) s.t. \|\mathbf{B}\|_{2,0} \leq s, \tag{5.25}$$

Here, we can construct matrix $\mathbf{Z}$ by concatenating column vectors $\mathbf{W}\mathbf{P}_j\mathbf{x}^{(i)}$ for each contrast. Then, the problem becomes:

$$\min_{\mathbf{B}} \|\mathbf{Z} - \mathbf{B}\|_F^2 s.t. \|\mathbf{B}\|_{2,0} \leq s, \tag{5.26}$$

The optimal solution is hard thresholded $\mathbf{Z}$, as before. The only difference here is to concatenate column vectors for each contrast. Next, we deal with the transform update step, as:

$$\min_{\mathbf{W}} \sum_{i} \left(\sum_{j=1}^{N} \|\mathbf{W}\mathbf{P}_j\mathbf{x}^{(i)} - b_j^{(i)}\|_2^2 \right) s.t. \mathbf{W}^H\mathbf{W} = \mathbf{I}. \tag{5.27}$$

Let us begin by defining $\mathbf{X}$ as the concatenated vectors $\mathbf{P}_j\mathbf{x}^{(i)}$ for all contrasts. Then, the problem becomes:

$$\min_{\mathbf{W}} \|\mathbf{W}\mathbf{X} - \mathbf{B}\|_F^2 s.t. \mathbf{W}^H\mathbf{W} = \mathbf{I}. \tag{5.28}$$

Here, the optimal solution is similar to section 5.3, and can be expressed using $\mathbf{V}\mathbf{U}^H$. Finally, let us deal with the image update step. This step requires solving:

$$\min_{\mathbf{x}} \sum_{i} \left(\nu \|\mathbf{A}^{(i)}\mathbf{x}^{(i)} - \mathbf{y}^{(i)}\|_2^2 + \sum_{j=1}^{N} \|\mathbf{W}\mathbf{P}_j\mathbf{x}^{(i)} - b_j^{(i)}\|_2^2 \right). \tag{5.29}$$

As can be seen, the problem can be divided for each contrast. In that case, the update steps is simply to carry out the solution given in table 5.2 multiple times. For the sake of completeness, the algorithm is summarized in table 5.3.

5.5 Chapter Summary

In this chapter, we first briefly described background information on dictionary learning. We described the concept, and some basic algorithms. Then, we moved

Table5.3: **Algorithm:** *Multi-Contrast Blind Compressive Sensing MRI (MC BCS MRI)*

---

1. Initialize $(\mathbf{W}^0, \mathbf{B}^0, \mathbf{x}^0)$.
2. **for t = 1:J**
3. Form the matrix $\mathbf{X}$ by stacking $\mathbf{P}_j\mathbf{x}^{(i)}$ as its columns for each contrast. Compute $L^{-1} = \left(\mathbf{X}\mathbf{X}^H + 0.5\lambda\mathbf{I}\right)^{-1/2}$.
4. **for l = 1:M**
5. Transform Update:
5.a. Set $W = \mathbf{V}\mathbf{U}^H$, where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \mathbf{X}\mathbf{B}^H$.
6. Sparse Coding:
6.a. Update $\mathbf{B}$ according to Eq. (5.18).
7. **end**
8. **for i = 1: # of contrasts**, Image Update:
8.a. Compute $c = \sum_{j=1}^{N} \mathbf{P}_j^T\mathbf{W}^H\mathbf{B}_j^{(i)}$, $S = FFT(c)$.
8.b. Compute a_1 as the first column of $c = \sum_{j=1}^{N} \mathbf{P}_j^T\mathbf{W}^H\mathbf{W}\mathbf{P}_j$.
8.c. Set $\gamma = \sqrt{p} \times FFT(a_1)$.
8.d. Set Fourier transform of $\mathbf{x}$ with indices (k_x, k_y) to

$$\begin{cases} \frac{S(k_x,k_y)}{\gamma(k_x,k_y)} & , \text{ if } (k_x,k_y) \notin \mathbf{M} \\ \frac{S(k_x,k_y)+\nu(\mathbf{A}^{(i)})^H\mathbf{y}^{(i)}}{\gamma(k_x,k_y)+\nu} & , \text{ if } (k_x,k_y) \in \mathbf{M} \end{cases}.$$

8e. **end**
9. **end**

---

on to dictionary learning for CS. We described online dictionary learning within the context of CS. We then proposed a joint reconstruction algorithm that can recover multi-contrast images simultaneously as well as the single sparsifying transformation applied on all contrast images. We also assumed group sparsity instead of sparsity on different contrast images. This helped improve reconstruction quality. We finally gave details on the implementation of the proposed multi-contrast joint dictionary learning algorithm.

CHAPTER 6

RESULTS OF EXPERIMENTS

In previous chapters, we described methods for compressed multi-contrast MRI image reconstruction. We first described the basics of MRI, and the most basic image reconstruction technique, i.e. inverse Fourier transformation. Then, we moved on to compressed sensing based techniques for single contrast imaging, then multi-contrast imaging. Finally, we moved on to dictionary learning based methods. Each chapter describes an approach and an algorithm related to that approach. Although the strong-suits of each approach are described, these are not demonstrated yet. We have just given a comparison of single contrast versus multi-contrast reconstruction technique in section 4.5. In that section, we have compared the effect of group sparsity and p-norm for a toy problem. In this chapter, we analyse the performance of each algorithm. We compare the methods for individual and joint reconstruction methods, under different imaging masks, SNR and parameter selections. We also demonstrate the effectiveness of each algorithm against conventional methods, as well as the state of the art. In this chapter, we first describe the nature of the experiments, the set up used and then move on to analysis of the methods on various datasets, and undersampling ratios.

6.1 Description of Experiments

In this thesis, we have worked with two simulated, and one experimental datasets, and different undersampling masks. First, let us describe the experimental procedure. As given in section 2.1, MR scanner gathers data in k-space. Hence, undersampling in MRI is a straightforward process of simply selecting some lines or points from k-space. Therefore, in this chapter, we first undersample an image by omitting some points from the k-space, then use one of the proposed

approaches to reconstruct the image. We, then, compare the reference image with the reconstructed image using both qualitative and quantitative metrics. For qualitative metrics, we use visual inspection and emphasize the regions of the image that is better or worse. For quantitative metrics, we use peak signal-to-noise-ratio (pSNR), root mean squared error (RMSE), mean magnitude error (mmE), and structural similarity index measure (SSIM). For a reconstructed image $\mathbf{x}$, and reference image $\mathbf{f}$, the metrics are defined as:

$$\text{pSNR}(\mathbf{x}, \mathbf{f}) = 20 \log \left(\frac{\max(\mathbf{f})}{\|\mathbf{x} - \mathbf{f}\|_2} \right), \tag{6.1}$$

$$\text{RMSE}(\mathbf{x}, \mathbf{f}) = \sqrt{\frac{1}{N} \|\mathbf{x} - \mathbf{f}\|_2^2}, \tag{6.2}$$

$$\text{mmE}(\mathbf{x}, \mathbf{f}) = \frac{1}{N} \|\mathbf{x} - \mathbf{f}\|_1. \tag{6.3}$$

SSIM is an image metric developed for comparing the similarity between two images similar to human eye [62]. It is combination of three metrics for an image, namely: luminance, contrast and the structural terms, defined as:

$$\text{SSIM}(\mathbf{x}, \mathbf{f}) = [I(\mathbf{x}, \mathbf{f})]^{\alpha} [c(\mathbf{x}, \mathbf{f})]^{\beta} [s(\mathbf{x}, \mathbf{f})]^{\gamma}, \tag{6.4}$$

$$I(\mathbf{x}, \mathbf{f}) = \frac{2\mu_x \mu_f + C_1}{\mu_x^2 + \mu_f^2 + C_1}, \tag{6.5}$$

$$c(\mathbf{x}, \mathbf{f}) = \frac{2\sigma_x \sigma_f + C_2}{\sigma_x^2 + \sigma_f^2 + C_2}, \tag{6.6}$$

$$s(\mathbf{x}, \mathbf{f}) = \frac{\sigma_{xf} + C_3}{\sigma_x \sigma_f + C_3}, \tag{6.7}$$

where μ_x and μ_f are the mean values, and σ_x, σ_f are the standard deviations of $\mathbf{x}$ and $\mathbf{f}$, respectively. σ_{xf} is the cross covariance of $\mathbf{x}$ and $\mathbf{f}$. For SSIM, we used the definition given in [62]. We used a Gaussian kernel for calculating the SSIM map using Eq. (6.4). Finally, the average of the SSIM map is taken to calculate SSIM.

After the reconstruction process, we scale the image to 0-255 for viewing purposes. In this case, although the overall image has high quality, the image metrics may fail to reflect this. To avoid this, we first fit a line though the reconstruction and reference image by solving:

$$[a, b] = \arg\min_{a,b} \|a\mathbf{x} + b - \mathbf{f}\|_2^2. \tag{6.8}$$

Here, note that multiplying an image with a constant and adding an offset does not change the viewed image, since the changes are cancelled later while scaling to 0 - 255. We then compare $a\mathbf{x} + b$ with the reference image, $\mathbf{f}$, while using quantitative imaging metrics.

Now let us move on to describing the used masks.

6.1.1 Sampling Masks

Compressive sensing is a natural match for MRI, since incoherence of the sampling domain with the signal domain is required for compressed sensing to work [37]. However, natural images such as MRI are not sparse in image domain itself. MR images are mostly sparse in some transformation domain, or gradient of the image, which break the purely random uniform undersampling scenario that compressed sensing is reported to work with [16]. In fact, drawing samples from a probability density function (pdf), rather than using purely random undersampling has been reported to increase the performance of CS algorithms [37]. In this study, we also use such an approach and create imaging masks accordingly. Sampling low-frequency components of the image in a higher density, while sampling high-frequency components sparsely results in better images. For this reason, it is suggested to use a pdf that decays with a polynomial order of 3, after sampling the low-frequency components fully (for example, one-eighth). We sampled one-eighth of the k-space (i.e. low frequency parts) fully, and draw rest of the samples randomly using the polynomial decay pdf of 3^{rd} order. We used these masks to undersample the image, then added bivariate Gaussian noise to the data.

We have previously described that scanning a full line rather than a point in k-space does not change the required sampling time much. Hence, for 2D imaging, it makes sense to undersample images in only one dimension, and use masks such as the ones given in Fig. 2.2. On the other hand, for 3D imaging, two dimensional undersampling can be employed, which has better incoherence.

Table6.1: Typical **TE**, and **TR** values for different contrasts

Sequence	**TE** (ms)	**TR** (ms)
PD	17	2775
$\mathbf{T_1}$	14	575
$\mathbf{T_2}$	102	2775

6.1.2 Datasets

In this study, we used 3 datasets.

First, we worked with a simulated dataset including eleven types of tissues. Each tissue has its own $\mathbf{T_1}$, $\mathbf{T_2}$ and **PD** values, and images are generated using different **TE** and **TR** values. The detailed data generation process is given in [7], and is beyond the scope of this thesis. In this thesis, we assumed a **TE** and **TR** as given in table 6.1. Since the resulting phantoms are real-valued, we introduced a slowly-varying phase to the overall image.

Next, we worked with SRI 24 atlas [58]. The atlas consists of **PD**, $\mathbf{T_1}$, and $\mathbf{T_2}$ weighted images generated from real patient data. However, as in [7], the images are real-valued, and required slowly-varying phase to overall images. The resulting images are given in figures.

Finally, we worked with experimental data, gathered using 3T scanner (courtesy of UMRAM, Ihsan Dogramaci Bilkent University). This data consists of **PD**, $\mathbf{T_1}$, and $\mathbf{T_2}$ weighted images, all gathered from the same patient. The data was gathered using 32 channels, which are then combined to a single channel image prior to undersampling. No additional phase was introduced to this image, since it is already complex-valued.

6.2 Results of Single Contrast Experiments

In this study, we first demonstrate the proposed individual method over similar techniques. We analyse two aspects of the single contrast framework. First, we compare the algorithm to similar methods in terms of convergence of the

objective function, and change in pSNR. We then compare the method to only ℓ_1-norm and only TV algorithms in terms of image quality, to emphasize the usage of multiple objective functions. Finally, we give the results for multiple masks created using the procedure described in section 6.1.1.

6.2.1 Selection of Regularization Parameters for Single Contrast Framework

First, let us show results for the selection of objective functions for single contrast images. Here, we chose a linear combination of ℓ_1-norm and TV, while other choices were possible. Let us now compare different α_1 and α_{TV} values that result in different linear combinations of regularizations. For simplicity in parameter selection, let us set $\alpha_{TV} = 1 - \alpha_1$.

We now used grid search to see which parameters result in better results in terms of quantitative metrics such as pSNR, SSIM, MME, and RMSE, as well as visual inspection. For the first study, we used the real-valued Aubert-Broche [7] brain phantom, and chose $\mathbf{T_1}$ contrast image for reconstruction, since we dealt with single contrast images. We sub-sampled the image to 33%. Convergence results versus time is depicted in Fig. 6.1. In terms of the quality of the converged point, $\alpha_1 = 0.8$ with $\alpha_{TV} = 0.2$ outperforms other results in terms of all quantitative metrics. This parameter set had the highest PSNR and SSIM, while having the lowest MME and nRMSE. Moreover, it has the fastest convergence. The closest parameter sets are the ones with $\alpha_1 = 1$ and $\alpha_{TV} = 0.6$. While $\alpha_1 = 1$ begins with a similar speed to $\alpha_1 = 0.8$, it diverges from the optimal solution approximately 20 seconds after the beginning of the reconstruction, and converges to a worse point. Other parameter sets result in poorer results in terms of all quantitative metrics.

Fig. 6.2 depicts the imaging results for different parameters, as well as original and zero-filled images. Fig. 6.2 (a) is the reference image used in this experiment, while (b) shows the conventionally reconstructed image from fewer number of samples. (c) – (h) shows the reconstructions using the proposed framework using different regularization weights between objective functions. As it can be seen,

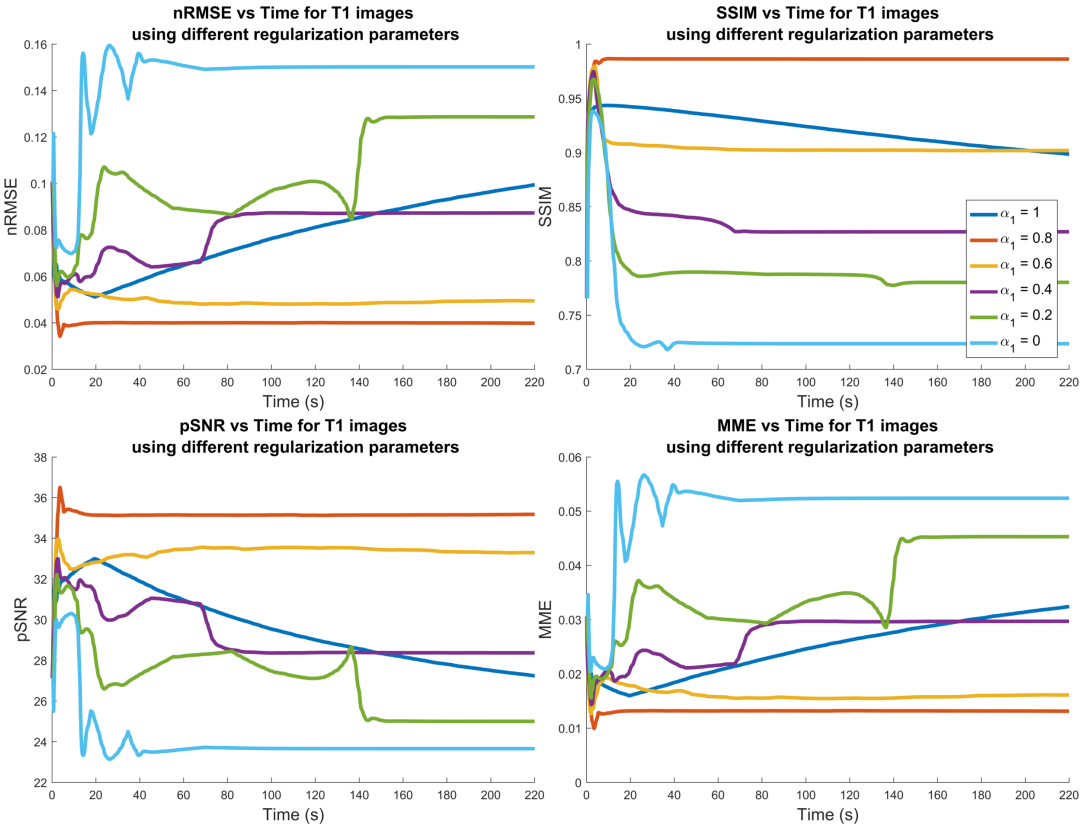


Figure 6.1: Comparison of convergence speed of the single contrast framework with various α_1 selections in terms of SSIM, PSNR, nRMSE, and MME for Aubert-Broche brain phantom $\mathbf{T_1}$-weighted image using 33% of full data.

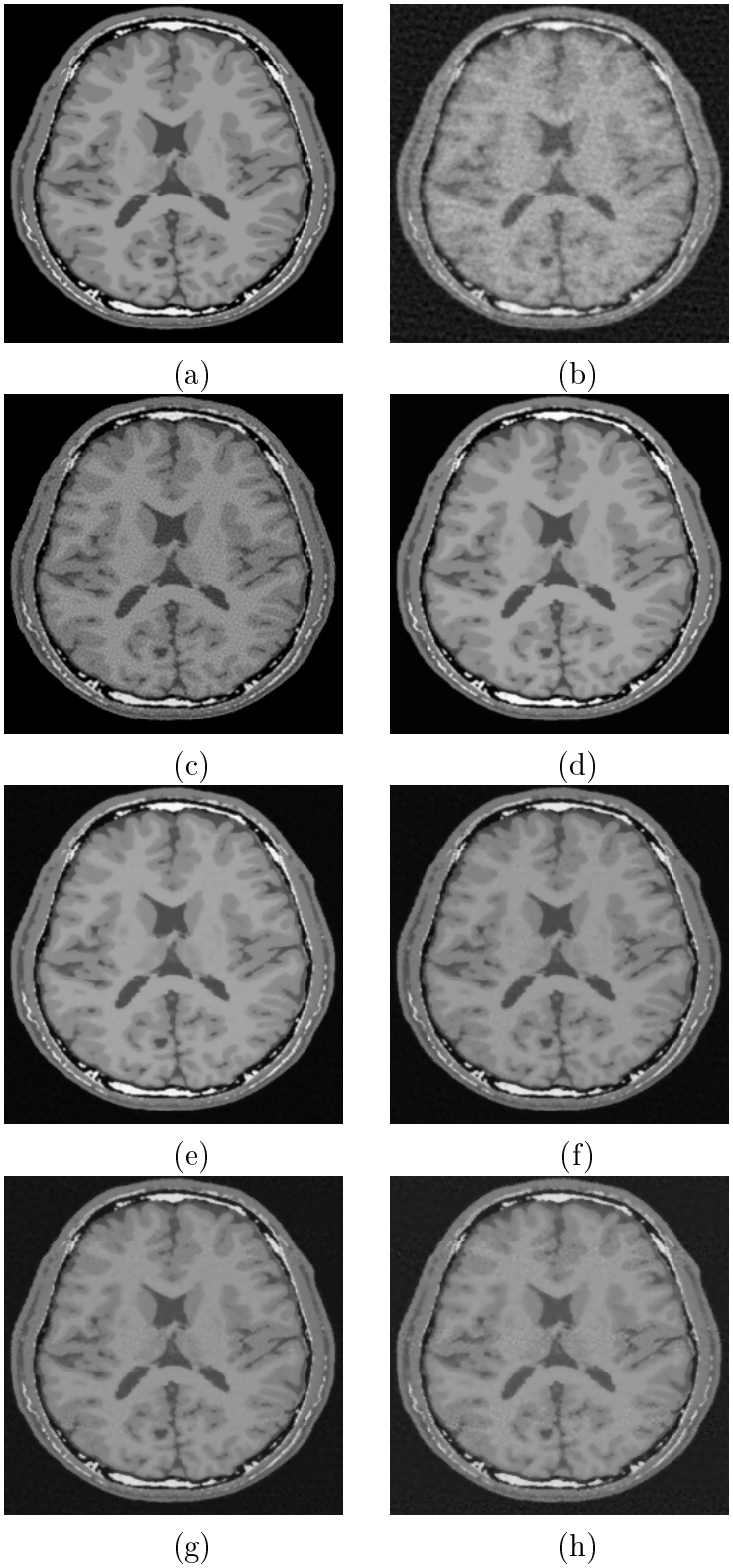


Figure 6.2: (a) Reference (Original) Image (b) Zero-filled Image (c) Image Reconstruction with $\alpha_1 = 1$ (d) Image Reconstruction with $\alpha_1 = 0.8$ (e) Image Reconstruction with $\alpha_1 = 0.6$ (f) Image Reconstruction with $\alpha_1 = 0.4$ (g) Image Reconstruction with $\alpha_1 = 0.2$ (h) Image Reconstruction with $\alpha_1 = 0$, for Aubert-Broche brain phantom $\mathbf{T_1}$-weighted image using 33% of full data.

choosing $\alpha_1 = 1$ results in noise-like artefacts in image (c), due to ℓ_1-norm. The objective function in this image enhances sparsity in the image, although the image itself is not sparse. As α_1 decreases and α_{TV} increases, the reconstructions become more and more blurry. The other end of the spectrum with $\alpha_{TV} = 1$ results in a very blurry image such that the contrast between the image and the background is lower than that of other reconstructions. Better quality images can be obtained by imposing high contrast using ℓ_1-norm while retaining piecewise smoothness with TV. Hence, using both ℓ_1-norm and TV together results in a better reconstruction. Using both qualitative and quantitative results, we can conclude that using a linear combination of ℓ_1-norm and TV results in a better reconstruction. Among these results, the best reconstruction can be seen in image (d).

Next, we analyse these results in terms of robustness. We ran the same experiment 20 times, and compared for different parameters for robustness. In each run, we used different mask and noise patterns. Quantitative results are given in Fig. 6.3. As can be seen, for each case, $\alpha_1 = 0.8$ yielded the best result quantitatively.

6.2.2 Comparison of Single Contrast Image Reconstruction Algorithms

Next, we deal single contrast images. We compare the performance of the single contrast techniques. First, let us briefly describe previous algorithms. Here, all compared algorithms solve the same optimization problem. Each algorithm use a different technique to solve their respective problem. Each algorithm tackles the problems of the form:

$$\underset{\mathbf{x}}{\textbf{minimize}}\ \alpha_1\|\mathbf{W}\mathbf{x}\|_1 + \alpha_2 TV(\mathbf{x}) + \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2, \tag{6.9}$$

where $\mathbf{W}$ is some transformation domain and $TV(\cdot)$ is the isotropic total variation function previously defined as the squared sum of gradients in two different directions. We used the default settings for all algorithms in terms of the choice of the transformation domain, $\mathbf{W}$.

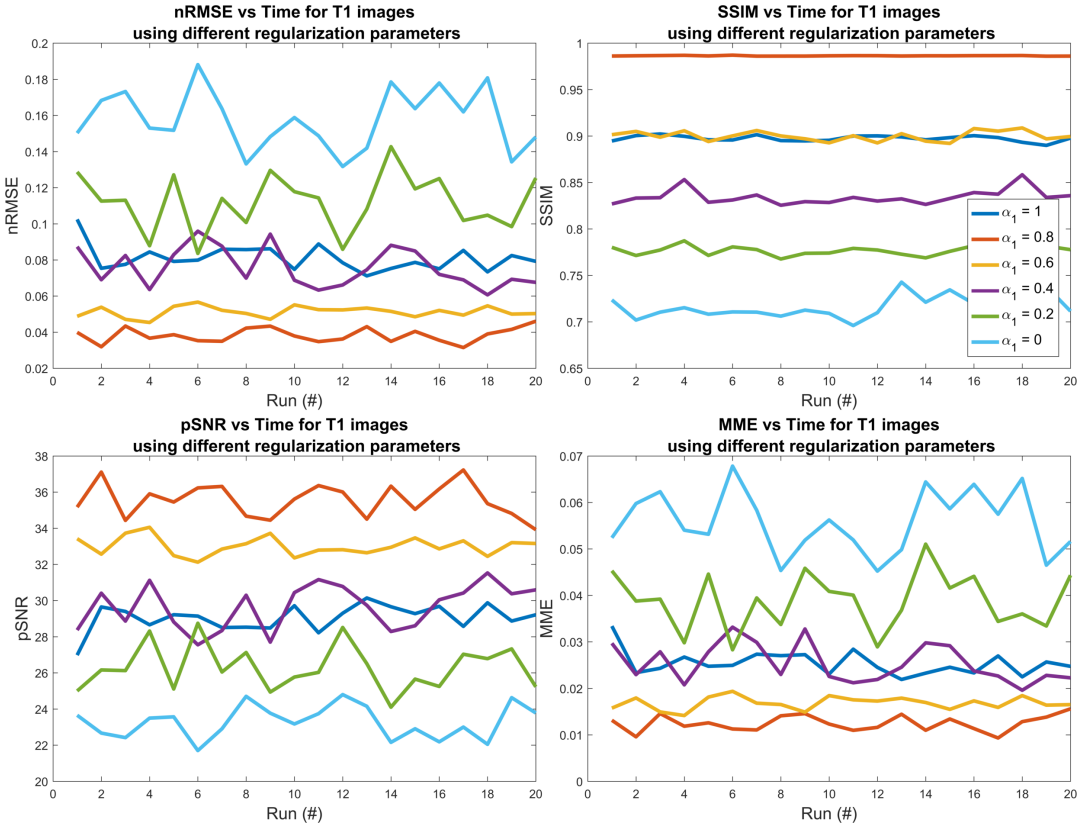


Figure 6.3: Comparison of image quality values of 20 runs for single contrast framework with various α_1 selections in terms of SSIM, PSNR, nRMSE, and MME for Aubert-Broche brain phantom $\mathbf{T_1}$-weighted image using 33% of full data.

SparseMRI is an algorithm proposed by Lustig et al. [37]. The algorithm solves the problem using conjugate gradient. The main drawback of this algorithm is the time required for the algorithm to converge. Conjugate gradient is known to converge with high precision. This helps produce quality images at the expense of high image reconstruction time. This method is also the first application of compressed sensing to MRI.

TVCMRI uses fixed point based algorithm to solve the same problem [56]. It tackles another form of the same problem depicted as:

$$\underset{\theta}{\textbf{minimize}}\ \alpha_1\|\theta\|_1 + \alpha_2 TV(\mathbf{W}^{-1}\mathbf{x}) + \frac{1}{2}\|\mathbf{A}\mathbf{W}^{-1}\theta - \mathbf{y}\|_2^2, \tag{6.10}$$

where the image vector $\mathbf{x} = \mathbf{W}^{-1}\theta$. Here, the outlined problem is exactly same as the one used in [37]. However, the used algorithm is different which affects the time required for convergence and the precision which the algorithm converges to. The algorithm iterates using fixed-point iterations to minimize with respect to each objective function separately.

RecPF algorithm uses ADMM based algorithm to solve the given problem [63]. The algorithm is formulated using ADMM, and it is solved by dividing the problem into simpler sub-problems. The algorithm converges in few number of iterations thanks to ADMM framework. However, the used formulation is different from the proposed method. The proposed single contrast optimization framework takes advantage of the complex nature of the problem, and applies the TV regularization to magnitude of the image, while the depicted problem for the compared algorithms do not. Also, although this method uses ADMM, the specific formulation affects the quality of the reconstruction as well as the convergence speed.

FCSA algorithm uses Composite-Splitting based algorithm to solve the same problem [35]. The algorithm employs a fast version of the composite splitting algorithm, and hence is called fast composite-splitting algorithm (FCSA).

All algorithms are compared for robustness against noise and mask selection, and convergence speed, as well as the quality of the converged signal. We used the codes supplied online for each of these algorithms. All parameters are opti-

Table6.2: Selected parameters for each algorithm

Algorithm	α_{TV}	α_1
SparseMRI	0.0118	0.0100
TVCMRI	0.3550	0.6958
RecPF	0.4187	0.0399
FCSA	0.0100	0.0350
Proposed Single Contrast	0.0175	0.9825

mized for highest SSIM value using grid search. The used parameters for each algorithm are given in table 6.2. There exists scale differences between algorithms on the definitions of TV and $\mathbf{W}$. Hence, each algorithm runs with a different regularization parameter. Each algorithm is optimized for the best SSIM performance.

We compared the algorithms in terms of the reconstruction quality using complex-valued images. As in the previous experiment, we used a 33% acceleration with the same Aubert-Broche brain phantom, with the exception of using complex-valued images as described in section 6.1.2. Figure 6.4 depicts the convergence speed for the given algorithms in terms of SSIM, PSNR, nRMSE and MME. As can be seen from the figure, the proposed method outperforms all other methods quantitatively. While the proposed individual method reaches the lowest result in terms of nRMSE and MME, it results in the highest pSNR and SSIM values. Although RecPF algorithm converges faster, the end-result reached by the proposed method is better.

Robustness analysis for the same phantom experiment is included in Fig. 6.5. The figure depicts comparison of SSIM, PSNR, nRMSE, and MME values for various algorithms in 20 runs. As can be seen, the proposed method outperforms other methods in terms of nRMSE, MME PSNR, and SSIM values. Here, as in the previous case, RecPF scored similar to the proposed individual method. The proposed method consistently outperformed all other methods quantitatively. We have observed consistent results in visual inspection, which we omit here.

Next, we switch to experimental dataset. We used **PD** image from the experimental dataset for a similar analysis on the convergence speed and visual

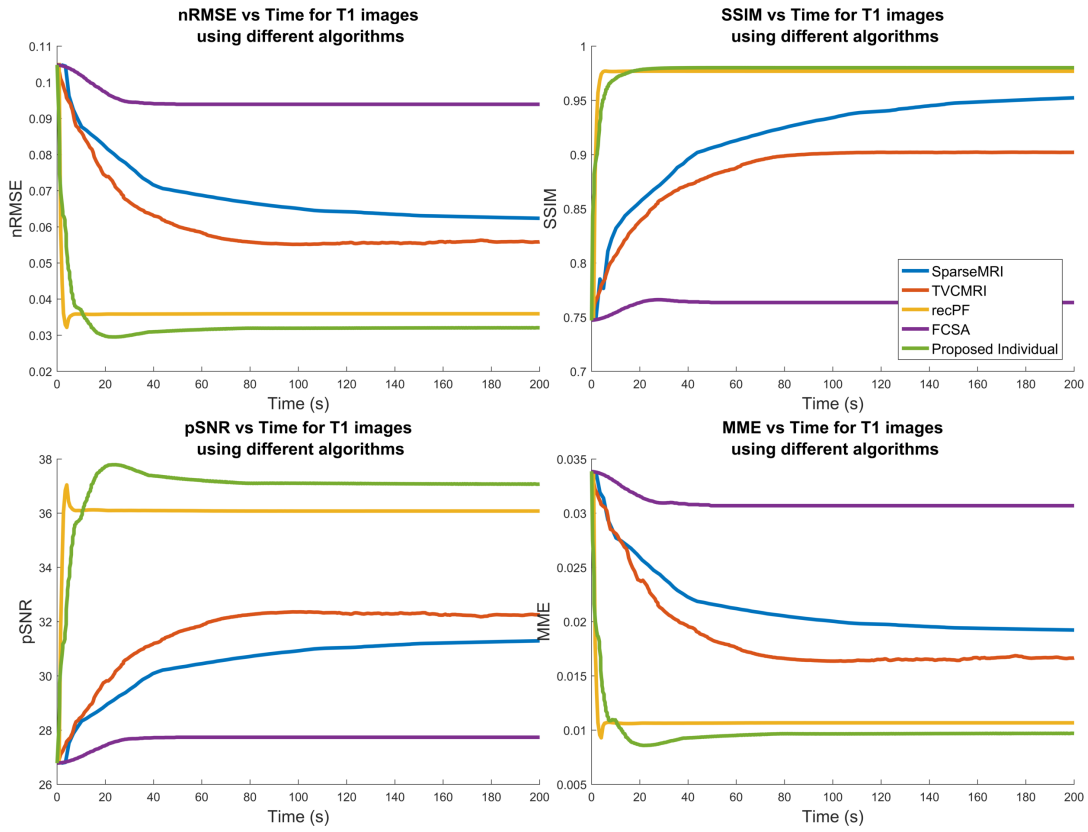


Figure 6.4: Comparison of convergence speed of different algorithms in terms of SSIM, PSNR, nRMSE, and MME for Aubert-Broche brain phantom $\mathbf{T_1}$-weighted image using 33% of full data.

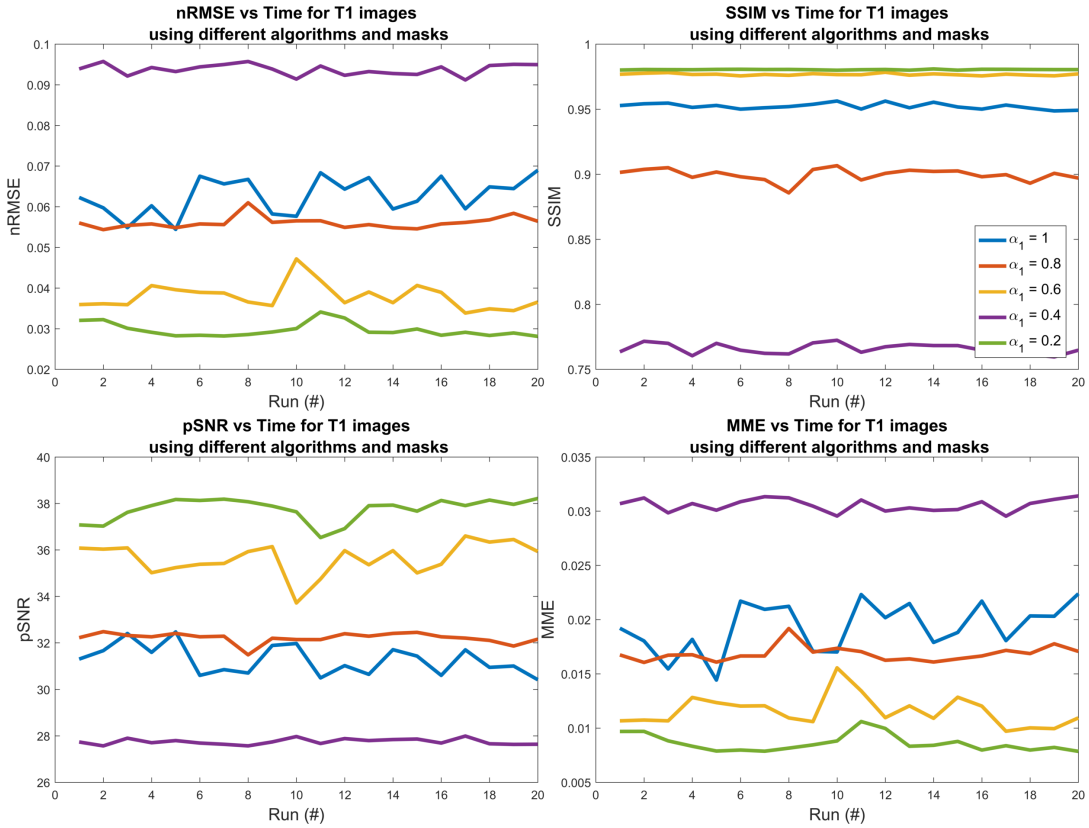


Figure 6.5: Comparison of image quality of 20 runs for various single contrast algorithms in terms of SSIM, PSNR, nRMSE, and MME for Aubert-Broche brain phantom $\mathbf{T_1}$-weighted image using 33% of full data.

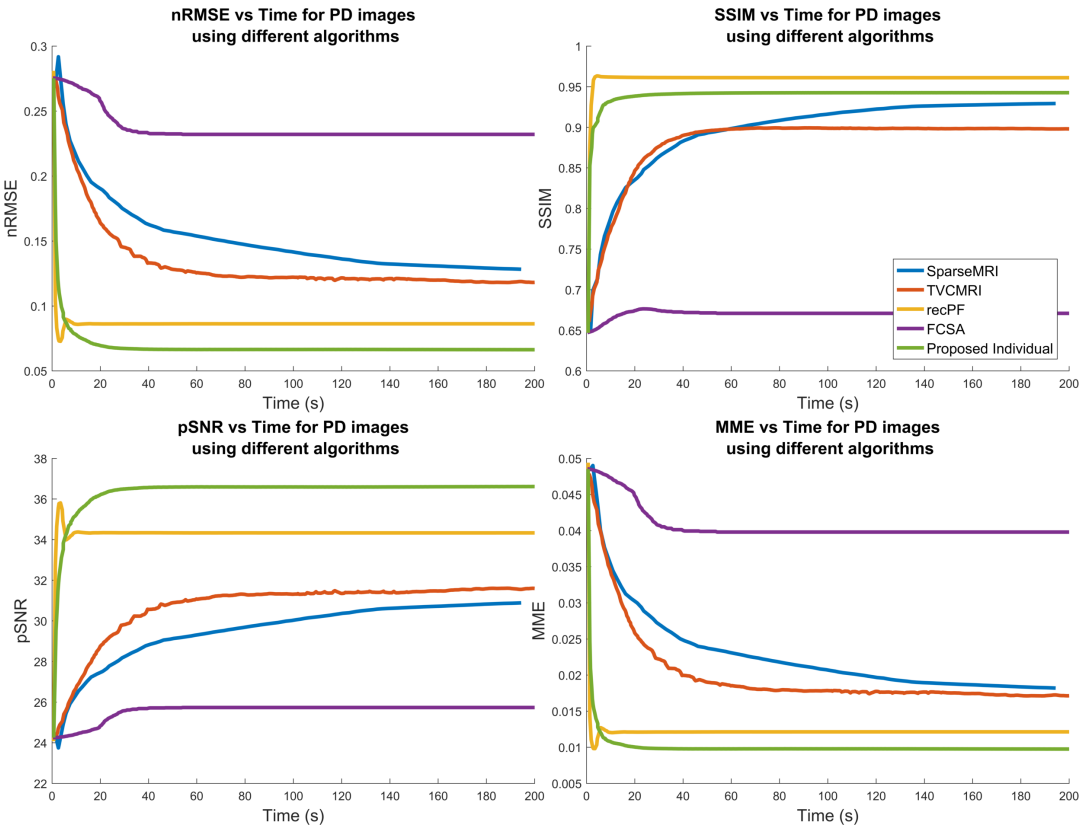


Figure 6.6: Comparison of convergence speed of different algorithms in terms of SSIM, PSNR, nRMSE, and MME for experimental **PD**-weighted image using 25% of full data.

inspection. We subsampled the data 25% using 2D undersampling patterns, and ran the same algorithms 20 times using different masks. Here, since the data is experimentally collected, the data itself contains a noise with SNR 40 dBs. ϵ value in the proposed method is set according to this fact. Convergence analysis is included in Fig. 6.6. The figure depicts comparison of SSIM, PSNR, nRMSE, and MME values for various algorithms versus time. As can be seen, the proposed method outperforms other methods in terms of MME, PSNR, and nRMSE values. RecPF outperforms the proposed method in terms of SSIM. However, visual inspection supports the claim of MME, pSNR, and nRMSE as will be discussed.

Although quantitative results help compare various methods, this is not the only metric available. Visual inspection is still a powerful tool and visual inspection results are given in Fig. 6.7. Fig. 6.7 (a) depicts the reference image, while

others depict reconstructed images using various algorithms. SparseMRI result in part (b) of the same figure suffers from loss in resolution in tissue boundaries. TVCMRI result in part (c) suffers from blocky artefacts due to the use of multi-scale wavelet transformation. FCSA, part (e), has poor visual image quality due to blocky artefacts. The closest algorithm to the proposed individual method, RecPF, has similar overall quality compared to the proposed method. However, RecPF produces oversmooth images as in part (b), and has lower contrast in the zoomed region on the bottom right of the images.

6.3 Results of Multi-Contrast Experiments

Next, we demonstrate the multi-contrast optimization framework. We first find the most suitable parameters for the algorithm. Here, we search for the best p value defined in section 4.4, and α_{JTV}. Next, we compare the algorithm to both individual and joint reconstruction methods in terms of convergence speed of the algorithm, and performance. We demonstrate the differences between individual and joint reconstruction algorithms. We compare the performance of the algorithm in terms of both quantitative metrics such as pSNR, SSIM, nRMSE and mmE, and by visual inspection. Here, we also give monte carlo results to demonstrate the robustness of the proposed method.

6.3.1 Selection of Regularization Parameters for Multi-Contrast Framework

Before comparing multi-contrast algorithms to individual counterparts, let us optimize parameters for the proposed multi-contrast method. Here, we deal with two main parameters: α_{JTV} that determines the regularization between joint total variation function and the group-ℓ_p-norm, and p that determines the coefficient of p-norm in group-ℓ_p-norm. Here, we ran the algorithm for 200 iterations with different parameter sets. We set α_{G1} associated with group-ℓ_p-norm objective to $1 - \alpha_{JTV}$. The final problem we solve is of the form (as in

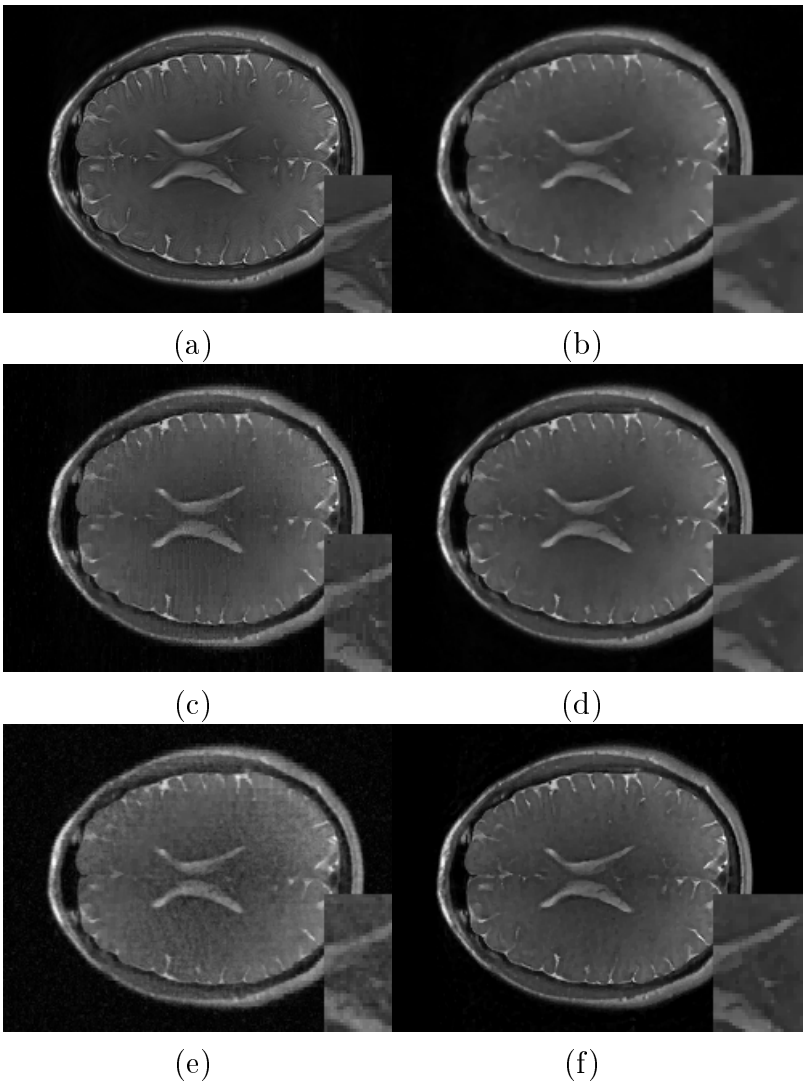


Figure 6.7: (a) Reference (Original) Image (b) Image Reconstruction using SparseMRI (c) Image Reconstruction using TVCMRI (d) Image Reconstruction using RecPF (e) Image Reconstruction using FCSA (f) Image Reconstruction using Proposed Single Contrast Method, for experimental **PD**-weighted image using 25% of full data.

Table6.3: Mean pSNR values (given in dB) for various selections of α_{JTV} and p values, for the multi-contrast optimization framework. The values depict mean pSNR of three contrast images for Aubert-Broche brain phantom images using 16% of full data.

p \ α_{JTV}	0.05	0.09	**0.14**	0.18	0.23	0.27	0.32	0.36	0.41	0.45
0.50	29.33	31.94	31.71	31.38	30.82	30.26	29.46	28.48	27.25	25.81
0.56	29.22	31.94	31.83	31.56	31.01	30.51	29.77	28.87	27.79	26.46
0.61	29.28	31.96	31.92	31.68	31.20	30.77	30.15	29.27	28.27	27.10
0.67	29.34	32.07	32.03	31.80	31.40	31.00	30.47	29.63	28.76	27.75
0.72	29.48	32.12	32.02	31.93	31.60	31.22	30.76	30.03	29.18	28.29
0.78	29.79	32.09	32.14	32.05	31.78	31.44	31.02	30.40	29.60	28.86
0.83	30.12	31.97	32.30	32.19	31.97	31.60	31.25	30.77	29.99	29.28
0.89	30.38	32.16	32.39	32.28	32.09	31.77	31.47	31.05	30.42	29.70
0.94	30.83	32.41	**32.53**	32.34	32.13	31.90	31.61	31.31	30.79	30.12
1.00	31.13	32.34	32.51	32.37	32.20	31.96	31.71	31.50	31.10	30.50

Eq. (4.32)):

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \alpha_{G1}\|\mathbf{x}\|_{2,p} + \alpha_{JTV} JTV(\mathbf{x}) \\ \text{subject to} \quad & \|\mathbf{A}^{(i)}\mathbf{x}^{(i)} - \mathbf{y}^{(i)}\|_2 \leq \epsilon_i, \text{ for i = 1 to C} \end{aligned} . \tag{6.11}$$

We used Aubert-Broche brain phantom, subsampled to 16% of full data. We used 3 contrast images of $\mathbf{PD}$, $\mathbf{T_1}$ and $\mathbf{T_2}$. We ran the proposed framework using only one mask, and calculated mean pSNR values given in table 6.3. Mean PSNR values are calculated as the mean of 3 different pSNR values. The table shows choosing α_{JTV} of 0.14 and p value of 0.94 results in the highest pSNR reconstruction, as opposed to using $p = 1$ as proposed by other methods.

We next compare the proposed method to the ones in the literature.

6.3.2 Comparison of Multi-Contrast Image Reconstruction Algorithms

In this study, we compared the proposed multi-contrast optimization framework to both the proposed individual method, and the previous methods exist in the literature, for joint multi-contrast CS MRI reconstruction. Let us first briefly discuss the compared methods. We compared the methods to two previous methods, namely FCSA for multi-contrast images (FCSA-MT) [34], and Group Sparsity MRI (GSMRI) [42].

FCSA-MT is previously discussed in section 4.1. It uses a Composite-Splitting based algorithm to solve problems of type:

$$\underset{\mathbf{x}}{\textbf{minimize}}\, \alpha_1 JTV(\mathbf{x}) + \alpha_2 \|\mathbf{W}\mathbf{x}\|_{2,1} + \sum_i \|\mathbf{A}^{(i)}\mathbf{x}^{(i)} - \mathbf{y}^{(i)}\|_2^2. \tag{6.12}$$

The algorithm is a natural extension of individual reconstruction algorithm FCSA.

The second algorithm, GSMRI, is previously discussed in sections 4.1 and 5.4. The method relies on solving the problems of the form using synthesis based redundant wavelet dictionaries:

$$\begin{aligned} &\arg\min_{\theta} \|\theta\|_{2,1} \\ &\text{subject to } \|\mathbf{y}^{(i)} - \mathbf{A}^{(i)}\mathbf{W}^T\theta_i\|_2 \leq \epsilon_i. \end{aligned} \tag{6.13}$$

Now let us describe the experiments in this section. We work with two cases. We first demonstrate the advantages of the proposed multi-contrast method using simulated Aubert-Broche dataset. We create 5 contrast images, **PD**, $\mathbf{T_1}$ weighted, $\mathbf{T_2}$ weighted, FLAIR and STIR. We then subsample each image to 16%. Figure 6.8 depicts the resulting images. Individually reconstructed images are noisy due to presence of low data. GSMRI reconstructed blocky images, while FCSA-MT reconstructed images with overall artefacts, as well as noise. The proposed method outperformed all other methods and reconstructed fine images.

We then worked with the experimental data. We used all three contrasts (**PD**, $\mathbf{T_1}$ weighted, $\mathbf{T_2}$ weighted) of the experimental data for joint reconstruction. We then subsampled the data to 50%, 33%, 17%, and 8%. For each case we ran the proposed methods single contrast optimization framework and multi-contrast optimization framework, as well as GSMRI and FCSA-MT. Quantitative performance results are given in table 6.4. The results show that for all given 2D undersampling ratios, proposed multi-contrast algorithm outperformed other algorithms. Another interesting result is that the proposed individual algorithm outperformed previous joint methods that exist in the literature. Joint reconstruction assumptions may not always hold and some noisy contrast image may

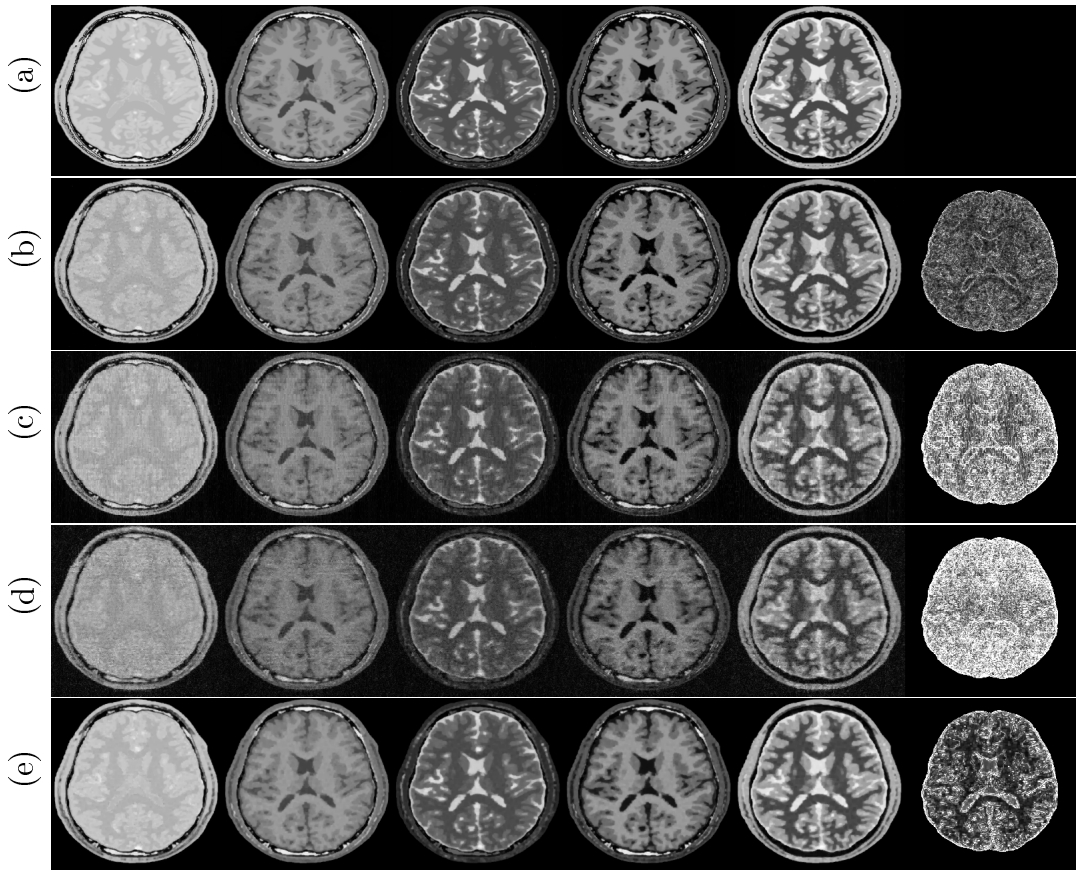


Figure 6.8: (a) Reference (Original) Image, (b) Image Reconstruction using Proposed Single Contrast Method, (c) Image Reconstruction using GSMRI, (d) Image Reconstruction using FCSA-MT, (e) Image Reconstruction using Proposed Multi-Contrast Method. From left to right, **PD**, $\mathbf{T_1}$ weighted, $\mathbf{T_2}$ weighted, FLAIR and STIR images are given. The rightmost image includes 4x magnitude sum of error in all parts.

Table6.4: Mean pSNR, SSIM, nRMSE, and MME values for four algorithms using 50%, 33%, 17%, and 8% of full data for Aubert-Broche brain phantom images.

Method	Ratio	PSNR (dB)	SSIM	nRMSE (%)	MME (%)
Proposed Individual	50%	37.86	0.88	%9.38	%0.87
GSMRI	50%	33.76	0.85	%12.70	%1.33
FCSA MT	50%	34.54	0.77	%12.09	%1.09
Proposed Multi-Contrast	50%	39.19	0.92	%7.16	%0.67
Proposed Individual	33%	34.78	0.84	%13.00	%1.19
GSMRI	33%	31.95	0.80	%15.64	%1.51
FCSA MT	33%	28.93	0.64	%22.66	%2.13
Proposed Multi-Contrast	33%	36.06	0.89	%10.13	%0.89
Proposed Individual	17%	30.99	0.80	%19.08	%1.69
GSMRI	17%	25.24	0.55	%33.78	%3.42
FCSA MT	17%	23.73	0.52	%41.75	%4.16
Proposed Multi-Contrast	17%	31.51	0.83	%16.79	%1.44
Proposed Individual	8%	25.65	0.69	%33.51	%3.07
GSMRI	8%	21.54	0.41	%52.02	%4.94
FCSA MT	8%	19.45	0.42	%66.49	%7.16
Proposed Multi-Contrast	8%	26.38	0.70	%30.43	%2.71

degrade other contrast images. One should be careful while using joint reconstruction.

Next, we demonstrate the convergence speed in terms of the given metrics for 33% of full experimental data. We also provide images for qualitative comparison. Quantitative results are given in fig. 6.9. The figure contains mean nRMSE, SSIM, pSNR and MME values that are calculated by averaging the respective metric of all three contrasts. As can be seen from the figure, the proposed method yielded fast convergence as well as better quantitative metrics, i.e. highest pSNR and SSIM and lowest MME and nRMSE values.

All contrast images are given in fig. 6.10. The figure depicts proposed individual algorithm, GSMRI, FCSA-MT and the proposed multi-contrast algorithm reconstructions. Each line in the result depicts $\mathbf{PD}$, $\mathbf{T_1}$, $\mathbf{T_2}$ weighted images, respectively. The rightmost image includes 4x magnitude sum of error in all lines. As can be seen, proposed individual method introduced an overall noise to the image due to lack of exploitation of joint features. GSMRI resulted in a

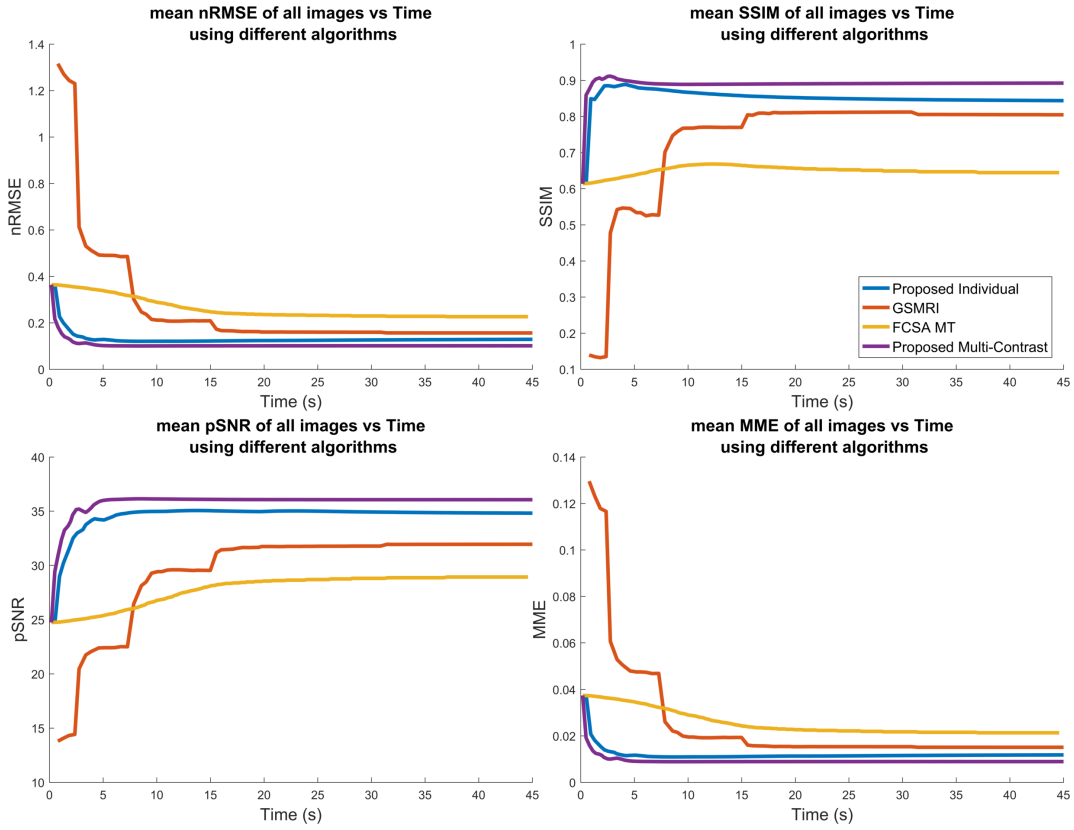


Figure 6.9: Comparison of convergence speed of four CS MRI reconstruction algorithms in terms of SSIM, PSNR, nRMSE, and MME for experimental images using 33% of full data.

blocky reconstruction due to Wavelet, while FCSA-MT resulted in a low resolution reconstruction (edges can not be properly captured). The proposed method performed the best compared to others algorithms.

Next, we compare the algorithms in terms of robustness. Although we have shown that the proposed multi-contrast algorithm works better for various undersampling ratios, let us now show robustness of the algorithm to different noise patterns and undersampling masks for 33% data. We ran the methods 20 times using 20 different masks and noise patterns for each contrast. Then, we compared the mean pSNR, SSIM, nRMSE and MME values of the algorithms. The results for each run is given in Fig. 6.11. As shown in the figure, the proposed multi-contrast method outperformed all methods in all runs.

6.4 Results of Dictionary Learning Based Method Experiments

In this study, we compared the proposed multi-contrast dictionary learning based method to both its single contrast counterpart, and other proposed methods. Let us first begin by finding the best parameters for the algorithm. The algorithm solves problems of type:

$$\min_{\mathbf{x},\mathbf{W},\mathbf{B}} \sum_i \left(\nu \|\mathbf{A}^{(i)}\mathbf{x}^{(i)} - \mathbf{y}^{(i)}\|_2^2 + \sum_{j=1}^{N} \|\mathbf{W}\mathbf{P}_j\mathbf{x}^{(i)} - b_j^{(i)}\|_2^2 \right) + \lambda Q(\mathbf{W}) s.t. \|\mathbf{B}\|_{2,0} \leq s, \tag{6.14}$$

The algorithm includes three parameters: ν (regularization parameter on data fidelity), λ (regularization parameter on dictionary $\mathbf{W}$) and s (sparsity level). λ is used to tune the likeliness of the learned dictionary to a unitary transformation, and ν is used to enforce data fidelity term. For the sake of simplicity, we avoid tuning for ν, and set it to a high value of $1e6/\{size_of_image\}$, which seems to work fine, since we are first dealing with high SNR scenarios. We optimized parameters using the experimental data set. We selected all three contrast data, and ran an exhaustive search algorithm to find the best PSNR values for the proposed dictionary learning method.

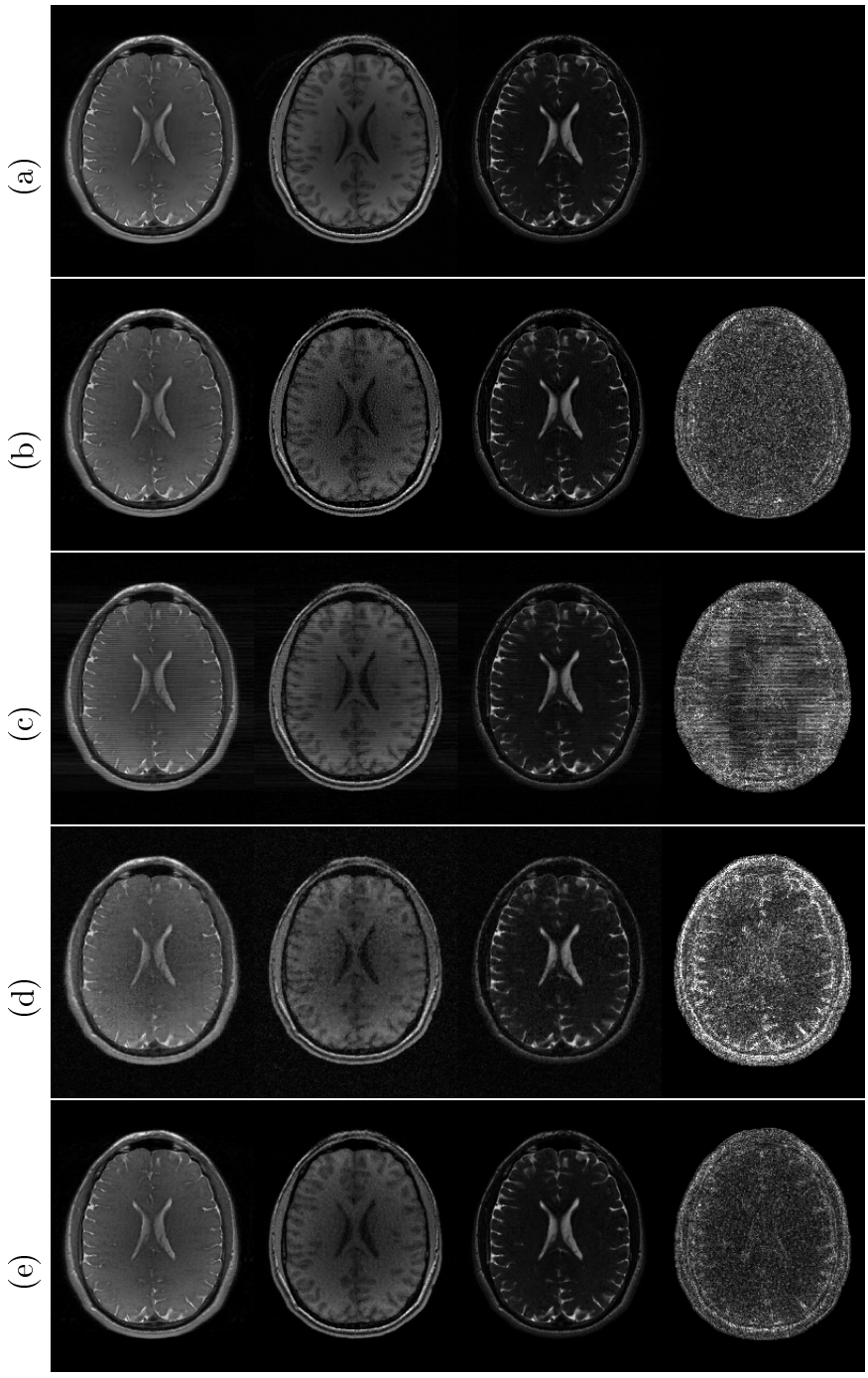


Figure 6.10: (a) Reference (Original) Image, (b) Image Reconstruction using Proposed Single Contrast Method, (c) Image Reconstruction using GSMRI, (d) Image Reconstruction using FCSA-MT, (e) Image Reconstruction using Proposed Multi-Contrast Method. From left to right, $\mathbf{PD}$, $\mathbf{T_1}$, $\mathbf{T_2}$ weighted images are given. The rightmost image includes 4x magnitude sum of error in all parts.

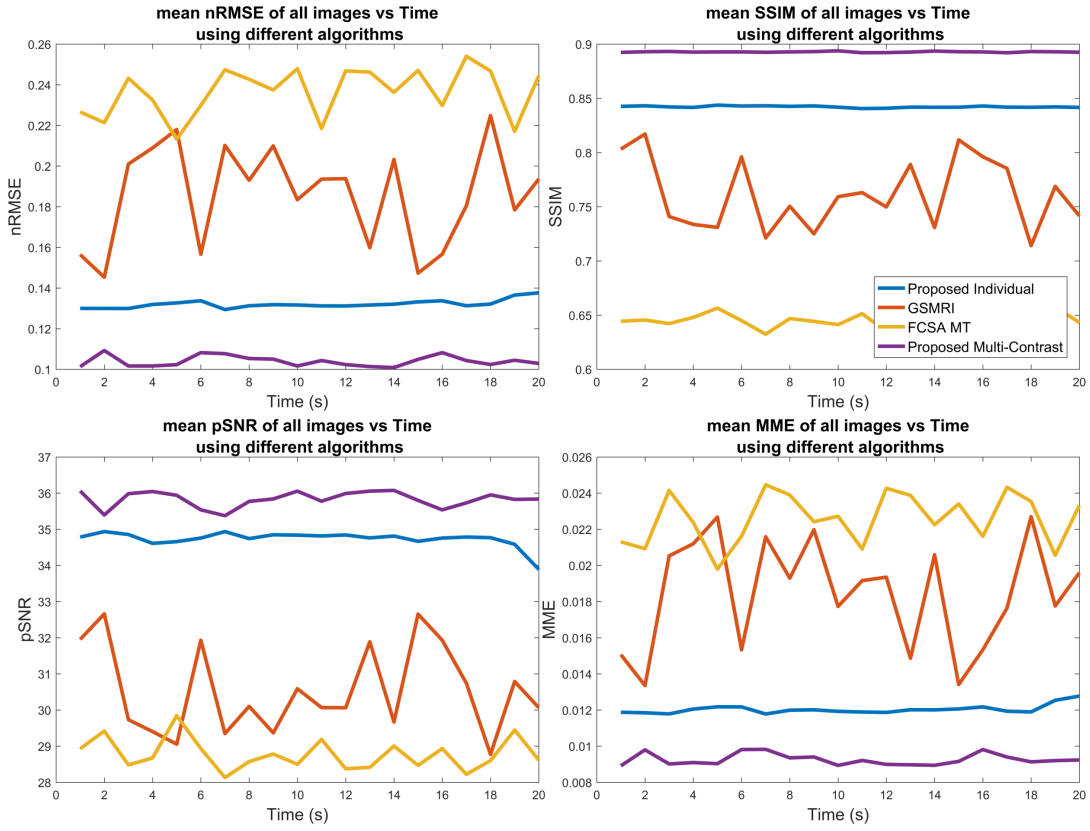


Figure 6.11: Robustness comparison for four algorithmsin terms of SSIM, PSNR, nRMSE, and MME for multi-contrast joint image reconstruction using experimental data.

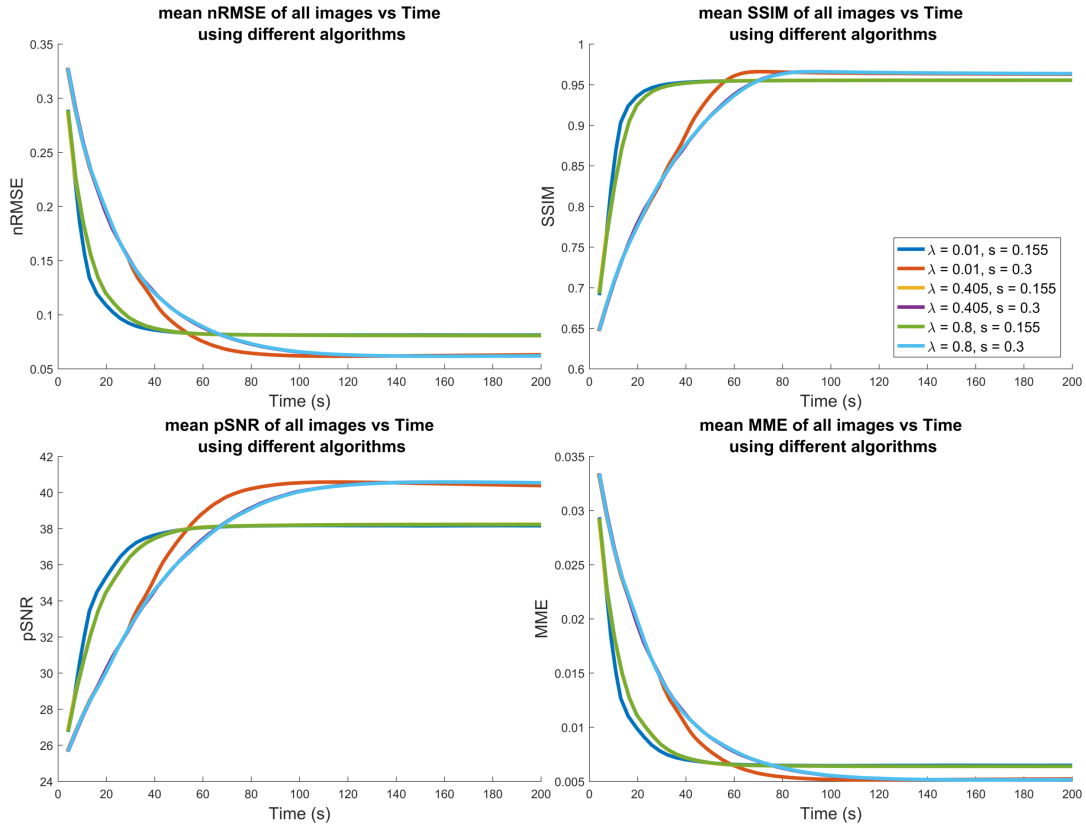


Figure 6.12: Evolution of SSIM, PSNR, nRMSE, and MME for different parameters of the proposed multi-contrast dictionary learning based method using experimental data subsampled to 33%.

Figure 6.12 shows convergence of different parameters for the proposed method. As can be seen, while setting higher value of sparsity level helps produce better images, setting lower λ increases reconstructed image quality. Next, we compare the proposed multi-contrast dictionary learning based method to individual dictionary learning based method.

6.4.1 Comparison of Multi-Contrast Dictionary Learning Based Method to Other Algorithms

In this section, we first compare the proposed multi-contrast dictionary learning framework to its individual counterpart, BCSMRI. The individual algorithm is summarized in section 5.2. We first show the advantage of multi-contrast algorithm to individual algorithm, as well as to other proposed methods.

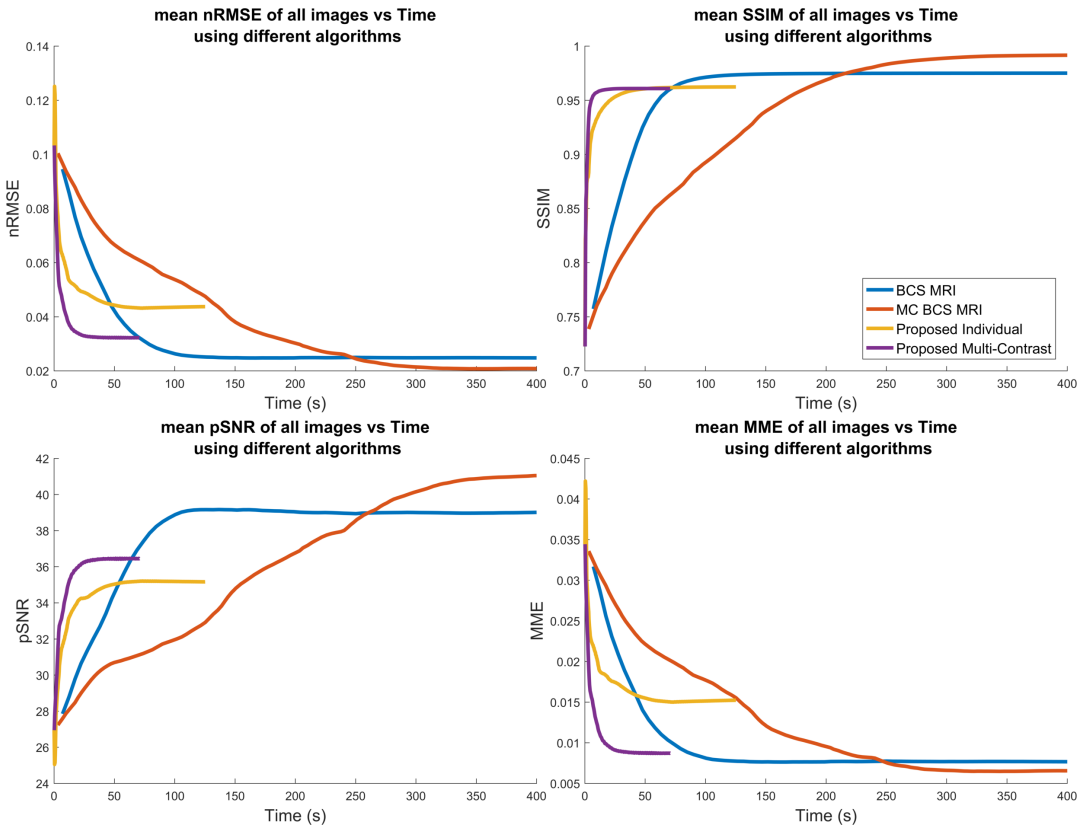


Figure 6.13: Evolution of SSIM, PSNR, nRMSE, and MME for different parameters of the proposed multi-contrast dictionary learning based method using Aubert-Broche dataset data subsampled to 33%.

We ran an experiment on the four algorithms, proposed individual method, proposed multi-contrast method, single contrast dictionary learning based method (BCS MRI) and the proposed multi-contrast dictionary learning based method (MC BCS MRI) using Aubert-Broche dataset and 33% of full data with three contrasts. Quantitative results can be found in fig. 6.13. Here, ADMM based algorithms are cut early because these methods have already converged. The figure shows clear advantage of using multi-contrast dictionary learning method over single contrast. However, as can be seen from the figures, the required reconstruction time is increased compared to other methods. Also, the proposed multi-contrast ADMM framework outperforms individual BCS MRI method.

Next, we investigate the reconstructed image qualities in experimental data for experimental data using 25% data. Fig. 6.14 shows the quantitative results for four algorithms same as above. As can be seen, although the dictionary learn-

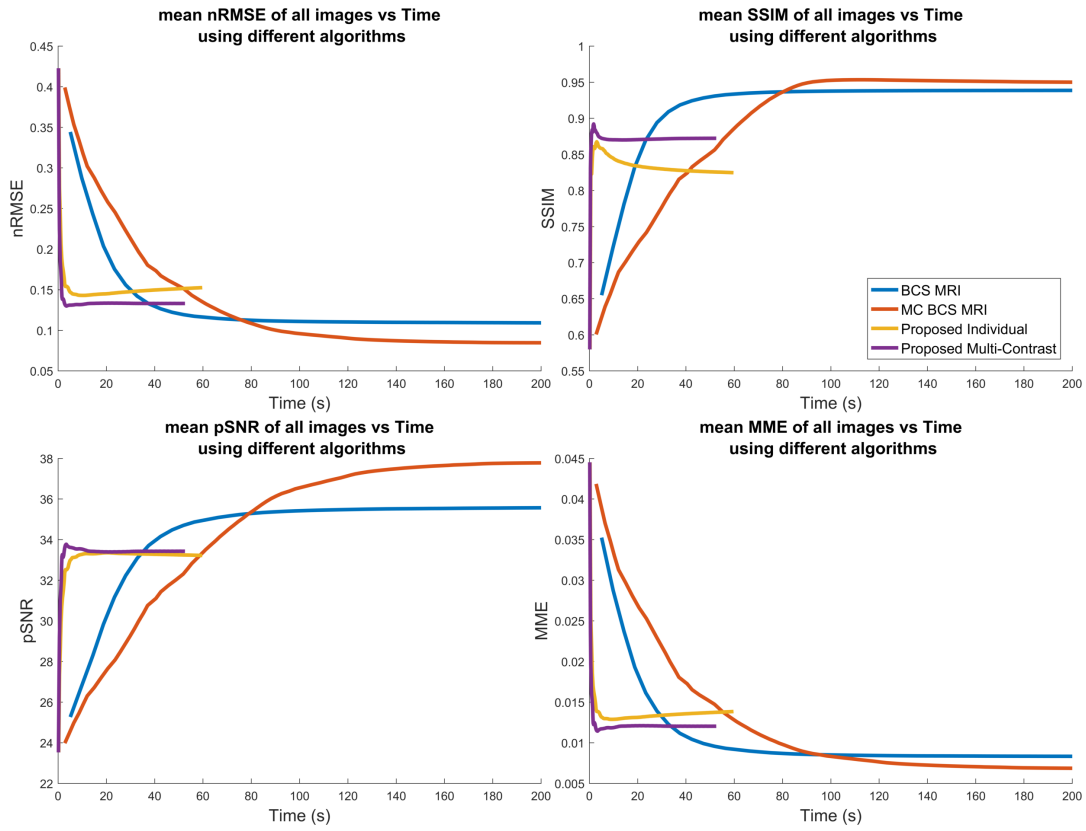


Figure 6.14: Evolution of SSIM, PSNR, nRMSE, and MME for different parameters of the proposed multi-contrast dictionary learning based method using experimental data subsampled to 25%.

ing based methods require high computation time to generate a solution. The methods provide a more accurate solution rather than a fast one. Dictionary learning is more suitable for cases where computation time is not the limiting factor. Also, contrary to previous experiment, individual dictionary learning based method performed better than proposed multi-contrast ADMM framework in the long run.

Finally, we give visual results of all contrast images. Figure 6.15 shows the reconstructed images. As can be seen, individual ADMM reconstruction has an overall noisy image due to lack of data. Multi-contrast ADMM reconstruction is a little better, however the image still is noisy. Individual BCS MRI failed to preserve the contrast on edges. MC BCS MRI has the least overall error.

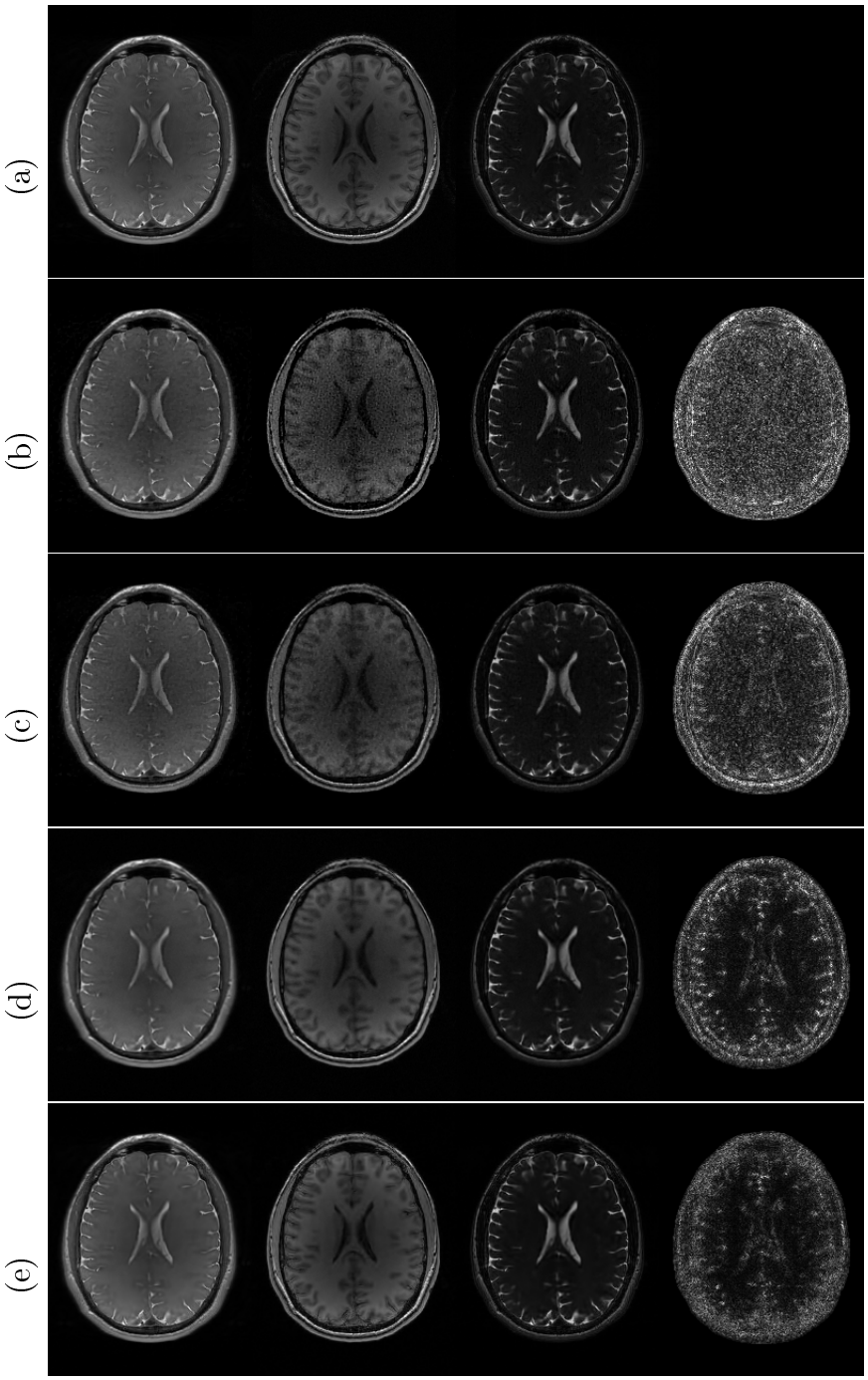


Figure 6.15: (a) Reference (Original) Image, (b) Image Reconstruction using Proposed Single Contrast Method, (c) Image Reconstruction using Proposed Multi-Contrast Method, (d) Image Reconstruction using BCS MRI, (e) Image Reconstruction using MC BCS MRI. From left to right, $\mathbf{PD}$, $\mathbf{T_1}$, $\mathbf{T_2}$ weighted images are given. The rightmost image includes 4x magnitude sum of error in all parts.

CHAPTER 7

CONCLUSIONS

In this thesis, we dealt with the problem of multi-contrast MRI image reconstruction from under-sampled data, and presented three novel methods. We first described the imaging process and compressive sensing theory. Next, we gave brief information on the optimization algorithms used within the context of this thesis. We then discussed current problems regarding the application of CS to MRI.

Two key issues exist in a practical application of CS to MRI. The first is that, CS algorithms require high computation time that undersampling the data for shortening scan time loses its advantage. The second aspect is on the selection of transformation domain on which the data is assumed to be sparse. We discuss the selection of transformation domain, and compare hand-crafted transformation domains to inferred ones using dictionary learning.

In this study, we approached these problems from two different perspectives. We first dealt with hand-crafted transformation domains. For hand-crafted transformation domains, we first proposed an optimization framework based on a variant of augmented Lagrangian method, ADMM. Then, we presented details on how this framework can be employed for CS MRI reconstruction. We used this algorithm for reconstructing single contrast MR images from undersampled data. We proposed a hybrid cost function specific for complex-valued imagery. We rigorously derived the necessary proximal mapping functions for the algorithm, and showed how any objective function can be applied to magnitude images. Next, we derived necessary equations for fast implementation of the algorithm. We discussed how each parameter of the algorithm can be selected.ADMM converges to an approximate solution in a few iterations. Since a fast solution was desired rather than a very accurate solution, ADMM was suitable for the

problem at hand. The method is suitable for not only CS MRI, but also other imaging modalities that include complex valued image reconstruction problems.

After establishing the framework for single contrast images, we moved to multi-contrast CS MRI. We first discussed the theoretical background of block compressive sensing. We rigorously derived an algorithm for multi-contrast imaging. We then discussed how this algorithm can be applied within the context of multi-contrast CS MRI. We introduced an objective function similar to the one proposed in single contrast CS MRI. We then proposed using group-ℓ_p-sparsity with $p < 1$ instead of group-ℓ_1-sparsity for better imposing sparsity on the image. We gave details of efficient implementation of the objective functions using the proposed multi-contrast framework. We then demonstrated the power of group sparsity and group ℓ_p sparsity over individual sparsity on a toy example.

In our final approach, we dealt with inferred transformation domains using dictionary learning. Dictionary learning is a signal processing tool used for finding the most suitable sparsifying transformation on training data. In chapter 5, we first discussed dictionary learning and some basic methods. We then discussed previous applications of dictionary learning to CS. In this study, we used an online dictionary learning method, i.e. simultaneous reconstruction of both the transformation domain and the signal, for image reconstruction. We proposed a dictionary learning based multi-contrast CS MRI algorithm. Although dictionary learning methods are computationally intensive, this method aims for improvement on the selection of transformation domain rather than computation time.

Finally, we tested the three proposed algorithms against each other and ones that exist in the literature. We first tested the performance of single contrast optimization framework for CS MRI, and compared it to similar single contrast reconstruction algorithms in the literature in terms of both quantitative and qualitative metrics. We have shown that the proposed single contrast algorithm outperforms similar reference algorithms in terms of both metrics. We then compared the algorithm for robustness using different masks and noise patterns. We have shown that the algorithm produces consistent results under different

mask and noise patterns.

We then moved to multi-contrast imaging algorithms. We have first shown the advantage of group-ℓ_p-norm over group ℓ_1-norm using MRI data. Then, we compared the proposed algorithm to both other joint reconstruction algorithms, and the proposed single contrast algorithm. The proposed single contrast method outperformed other joint reconstruction algorithms, while the proposed multi-contrast method performed the best in terms of both qualitative and quantitative metrics in experimental MRI data. Hence, using mutual information across channels does not always result in better images. We investigated the method in terms of both convergence speed, and robustness. We have found that using joint objective functions help the algorithm converge in fewer number of iterations as well as to a better solution. Hence, exploiting joint features across contrasts improved image quality for these experiments.

Finally, we tested the dictionary learning based methods. Here, multi-contrast ADMM framework performed similar to individual dictionary learning based CS. However, in all cases the proposed MC BCS MRI algorithm outperformed the other algorithms. The major setback of dictionary learning based methods is the required computation time. Here, we believe that the specific choice of the method is a trade-off between computation time and desired image quality.

Future work of this thesis includes performance testing of the algorithms on more datasets, parallel implementation and comparison of more objective functions. Although we have tested the proposed methods on both simulated and experimental datasets, large scale applicability of the methods are not shown in this thesis due to lack of data. Also, ADMM is a highly parallelizable optimization framework. Although we have not implemented the algorithms in a parallel framework using multiple CPUs or GPUs, such an implementation will surely decrease the required time, which would increase feasibility in the clinical settings. Within this thesis we have used ℓ_1-norm, $\ell_{2,p}$-norm, total variation and joint total variation as hand-crafted transformations. However, other functions such as total generalized variation, or ℓ_1-norm in some transformation domain such as wavelet may result in better image quality. Finally, we have

not compared the sparsifying transformations found using the single contrast dictionary learning based method and multi-contrast dictionary learning based method. The inherent properties of these transformations may lead to better hand-crafted or auto-reconstructed sparsifying transformations.

REFERENCES

[1] M. Afonso, J. M. Bioucas-Dias, and M. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Processing*, 20:681–695, March 2011.

[2] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9):2345–2356, 2010.

[3] A. Afrasiyabi, I. Onal, and F. T. Y. Vural. A sparse temporal mesh model for brain decoding. In *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2016 IEEE 15th International Conference on*, pages 198–206. IEEE, 2016.

[4] M. Aharon, M. Elad, and A. Bruckstein. *rmk*-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, Nov 2006.

[5] M. Aharon, M. Elad, and A. M. Bruckstein. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. *Linear Algebra and its Applications*, 416(1):48 – 67, 2006. Special Issue devoted to the Haifa 2005 conference on matrix theory.

[6] M. E. Ahsen and M. Vidyasagar. Error bounds for compressed sensing algorithms with group sparsity: A unified approach. *Applied and Computational Harmonic Analysis*, 2015.

[7] B. Aubert-Broche, A. C. Evans, and L. Collins. A new improved version of the realistic digital brain phantom. *NeuroImage*, 32(1):138 – 145, 2006.

[8] R. Baraniuk and P. Steeghs. Compressive radar imaging. In *2007 IEEE Radar Conference*, pages 128–133, April 2007.

[9] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 693–696, April 2009.

[10] B. Bilgic, V. K. Goyal, and E. Adalsteinsson. Multi-contrast reconstruction with bayesian compressed sensing. *Magnetic Resonance in Medicine*, 66(6):1601–1615, 2011.

[11] T. Blumensath, M. E. Davies, and G. Rilling. Greedy algorithms for compressed sensing. May 2012.

[12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[13] X. Bresson and T. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problems and Imaging*, 2(4):455–484, 2008.

[14] S. Camlica, A. C. Gurbuz, and O. Arikan. Sar image reconstruction with joint off-grid target and phase error corrections. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4502–4505, July 2015.

[15] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2011.

[16] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.

[17] E. J. Candes and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.

[18] M. Cetin and W. C. Karl. Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization. *IEEE Trans. Image Processing*, 10:623–631, April 2001.

[19] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imag. Vis.*, 20:89–97, 2004.

[20] R. Chartrand and B. Wohlberg. A nonconvex admm algorithm for group sparsity with sparse groups. In *International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2013.

[21] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2008.

[22] T. Cukur, M. Lustig, and D. G. Nishimura. Improving non-contrast-enhanced steady-state free precession angiography with compressed sensing. *Magnetic resonance in medicine*, 61(5):1122–1131, 2009.

[23] T. Cukur, M. Lustig, E. U. Saritas, and D. G. Nishimura. Signal compensation and compressed sensing for magnetization-prepared mr angiography. *IEEE transactions on medical imaging*, 30(5):1017–1027, 2011.

[24] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive Approximation*, 13(1):57–98, 1997.

[25] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[26] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, March 2008.

[27] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, Dec 2006.

[28] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947, 2007.

[29] Y. C. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Transactions on Signal Processing*, 58(6):3042–3054, 2010.

[30] I. O. Ertugrul, M. Ozay, and F. T. Y. Vural. Encoding the local connectivity patterns of fmri for cognitive state classification. *arXiv preprint arXiv:1610.05036*, 2016.

[31] H. E. Güven, A. Güngör, and M. Cetin. An augmented lagrangian method for complex-valued compressed sar imaging. *IEEE Transactions on Computational Imaging*, 2(3):235–250, 2016.

[32] B. S. He, H. Yang, and S. L. Wang. Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 106(2):337–356, 2000.

[33] S. Healthineers. Fda clears compressed sensing mri acceleration technology from siemens healthineers. `http://usa.healthcare.siemens.com/news/fda-clears-mri-technology-02-21-2017.html`. Accessed: 27-08-2017.

[34] J. Huang, C. Chen, and L. Axel. Fast multi-contrast mri reconstruction. *Magnetic Resonance Imaging*, 32(10):1344–1352, 2014.

[35] J. Huang, S. Zhang, and D. Metaxas. Efficient mr image reconstruction for compressed mr imaging. *Medical Image Analysis*, 15(5):670 – 679, 2011. Special Issue on the 2010 Conference on Medical Image Computing and Computer-Assisted Intervention.

[36] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, June 2008.

[37] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing mri. *IEEE Signal Processing Magazine*, 25(2):72–82, March 2008.

[38] Z. m. Wang and W. w. Wang. Fast and adaptive method for sar superresolution imaging based on point scattering model and optimal basis selection. *IEEE Transactions on Image Processing*, 18(7):1477–1486, July 2009.

[39] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 689–696, New York, NY, USA, 2009. ACM.

[40] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, Jan 2008.

[41] A. Majumdar and R. K. Ward. Accelerating multi-echo t2 weighted mr imaging: Analysis prior group-sparse optimization. *Journal of Magnetic Resonance*, 210(1):90 – 97, 2011.

[42] A. Majumdar and R. K. Ward. Joint reconstruction of multiecho mr images using correlated sparsity. *Magnetic Resonance Imaging*, 29(7):899–906, 2011.

[43] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk. Asymptotic analysis of complex lasso via complex approximate message passing (camp). *IEEE Transactions on Information Theory*, 59(7):4290–4308, July 2013.

[44] D. Needell and R. Ward. Near-optimal compressed sensing guarantees for total variation minimization. *IEEE Transactions on Image Processing*, 22(10):3941–3949, Oct 2013.

[45] D. Nishimura. *Principles of Magnetic Resonance Imaging*. Stanford University, 2010.

[46] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311 – 3325, 1997.

[47] B. Ophir, M. Lustig, and M. Elad. Multi-scale dictionary learning using wavelets. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):1014–1024, Sept 2011.

[48] N. Parikh and S. P. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.

[49] V. M. Patel, G. R. Easley, D. M. H. Jr., and R. Chellappa. Compressed synthetic aperture radar. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):244–254, April 2010.

[50] S. Ravishankar and Y. Bresler. Learning doubly sparse transforms for images. *IEEE Transactions on Image Processing*, 22(12):4598–4612, Dec 2013.

[51] S. Ravishankar and Y. Bresler. Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging. *SIAM Journal on Imaging Sciences*, 8(4):2519–2557, 2015.

[52] S. Ravishankar, B. Wen, and Y. Bresler. Online sparsifying transform learning, part i: Algorithms. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):625–636, June 2015.

[53] R. Rubinstein, T. Peleg, and M. Elad. Analysis k-svd: A dictionary-learning algorithm for the analysis sparse model. *IEEE Transactions on Signal Processing*, 61(3):661–677, Feb 2013.

[54] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259 – 268, 1992.

[55] H. Sedarat and D. G. Nishimura. On the optimality of the gridding reconstruction algorithm. *IEEE Transactions on Medical Imaging*, 19(4):306–317, April 2000.

[56] M. Shiqian, Y. Wotao, Z. Yin, and A. Chakraborty. An efficient algorithm for compressed mr imaging using total variation and wavelets. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8.

[57] J. Sulam, B. Ophir, M. Zibulevsky, and M. Elad. Trainlets: Dictionary learning in high dimensions. *IEEE Transactions on Signal Processing*, 64(12):3180–3193, June 2016.

[58] E. S. T Rohlfing, NM Zahr and A. Pfefferbaum. The sri24 multi-channel atlas of normal adult human brain structure. *Human brain mapping*, 31(5):798–819, 2010.

[59] S. Ugur and O. Arıkan. Sar image reconstruction and autofocus by compressed sensing. *Digital Signal Processing*, 22(6):923 – 932, 2012.

[60] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2009.

[61] S. L. WANG and L. Z. LIAO. Decomposition method with a variable parameter for a class of monotone variational inequality problems. *Journal of Optimization Theory and Applications*, 109(2):415–429, 2001.

[62] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[63] J. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010.