Universiteit Utrecht

UMC Utrecht

Department of Mathematics

Master's Thesis
Mathematical Sciences

Reducing geometric distortions of Diffusion-Weighted Imaging using Compressed Sensing

Author:
Hannelore K. van der Burgh

Supervisors:
Dr. Gerard L.G. Sleijpen (UU)
Dr. ir. Astrid L.H.M.W. van Lier (UMCU)

April 17, 2014

Acknowledgements

This thesis is written for the conclusion of my master program Mathematical Sciences at the Utrecht University. This thesis is also the result of my internship at the UMC Utrecht. I would like to thank all people who helped me with this project.

I would like to thank my supervisor at the mathematical department of the Utrecht University, Gerard Sleijpen, who always made time to help me find solutions to problems during the project. Without his mathematical insights and our brainstorm sessions, this thesis could not have been written as it is now.

I would also like to thank my supervisor at the UMC Utrecht, Astrid van Lier, who introduced me to the principles of MRI. Without her, I could not have made the connection between mathematics and the medical research. Her feedback and input during my research and the time of writing this thesis was very useful.

Furthermore, I would like to thank my colleagues from the Radiotherapy department for their interest in my research. In particular, I want to thank Didi, Ellis, Esther, Floris, Jiri, Lauren, Mariska and Nicole for the great atmosphere in our room. You made my internship more pleasant.

Finally, I would like to thank my family for supporting me. Special thanks to my sister, who helped me editing images used in this thesis. You saved me a lot of time.

Abstract

Diffusion-weighted imaging (DWI) is a type of contrast imaging used in Magnetic Resonance Imaging (MRI) that visualizes the amount of diffusion of water molecules in tissue. Tumors are well visible on DWI images. DWI is often acquired with Echo Planar Imaging (EPI) techniques. Unfortunately, these techniques lead to geometric distortions in the diffusion-weighted images. This causes problems in locating the position of the tumor exactly, which is required for radiotherapy.
In this thesis, an approach called Compressed Sensing (CS) was investigated as a technique to reduce the geometric distortions. In theory, the distortions are reduced by obtaining less MR data during scan acquisition (undersampling). By enforcing sparsity of the data in a transform domain, a well reconstructed image can be obtained as the solution of an appropriate minimization problem. The reconstruction algorithm used to solve this problem was cFISTA, a modification of FISTA developed by Beck and Teboulle [1].
Five undersampling strategies were retrospectively used on a DWI patient data set and the best strategy among these five was identified. The reconstruction quality of the whole image and the quality of the tumor reconstruction were assessed using the so-called Structure Similarity Image Measure. A strategy called 'centerincreased' gave the best balance between the average percentage of the MR data required for high quality reconstruction and the variation between the test images, for both the tumor reconstruction and reconstruction of the whole image. High quality reconstructions were obtained for this strategy, when on average only 20% of the MR data was included.
The implemented reconstruction algorithm cFISTA was also used to reconstructed MR data that was acquired in an undersampled fashion. This experiment proved the working of cFISTA on the complex MR data, but a reduction of the geometric distortions could not be demonstrated yet due to practical limitations of the EPI technique implemented on the MR scanner.

Contents

Contents iii

List of Figures iv

List of Tables v

List of Algorithms v

1 Introduction 1

2 Magnetic Resonance Imaging 3
- 2.1 NMR and MRI . . . 3
  - 2.1.1 History . . . 3
  - 2.1.2 Principles of Nuclear Magnetic Resonance . . . 4
  - 2.1.3 Principles of Magnetic Resonance Imaging . . . 5
- 2.2 k-space . . . 7
  - 2.2.1 Trajectories . . . 9
- 2.3 Contrast Imaging . . . 10
  - 2.3.1 Spin-lattice relaxation and T_1-weighting . . . 11
  - 2.3.2 Spin-spin relaxation and T_2-weighting . . . 11
  - 2.3.3 Diffusion-weighted imaging . . . 12
- 2.4 Geometric distortions . . . 13
  - 2.4.1 Echo Planar Imaging . . . 13
  - 2.4.2 Reducing geometric inaccuracy . . . 14

3 Compressed Sensing 16
- 3.1 Undersampling . . . 16
  - 3.1.1 Uniform vs non-uniform sampling . . . 17
  - 3.1.2 Compressed Sensing . . . 19
- 3.2 Reconstruction problem . . . 19
  - 3.2.1 l_1-minimization . . . 20
  - 3.2.2 TV-norm . . . 21
  - 3.2.3 The minimization problem . . . 22
- 3.3 Solving the minimization problem . . . 23
  - 3.3.1 Solving simpler problems . . . 23
  - 3.3.2 ISTA . . . 24
  - 3.3.3 Error estimation and convergence of ISTA . . . 26
  - 3.3.4 Improvement ISTA: FISTA . . . 28
  - 3.3.5 Choices of $g(x)$ . . . 29
  - 3.3.6 mFISTA . . . 34

3.3.7 cFISTA . . . 34
3.3.8 Wavelets . . . 36

4 Image quality measure **39**
4.1 Mean Squared Error . . . 39
4.2 SSIM . . . 41
4.2.1 Luminance . . . 41
4.2.2 Contrast . . . 41
4.2.3 Structure . . . 42
4.2.4 Weighting . . . 42
4.2.5 Implementation . . . 43
4.2.6 Volume . . . 44
4.2.7 Reference image . . . 44
4.3 Power spectral density . . . 45

5 Results **46**
5.1 Undersampling masks . . . 46
5.2 Choice parameters and transforms . . . 50
5.2.1 Wavelets . . . 50
5.2.2 Optimal regularization parameters . . . 52
5.2.3 Size . . . 53
5.2.4 100%-image . . . 54
5.2.5 Stopping criterion . . . 54
5.2.6 Proximal operator orders . . . 55
5.3 Experiments on diffusion-weighted images . . . 56
5.3.1 Comparing patients . . . 57
5.3.2 Comparing strategies . . . 60
5.3.3 Tumor data . . . 61
5.3.4 Prospective undersampling experiment . . . 63
5.3.5 3D . . . 66

6 Discussion **68**
6.1 The optimal strategy . . . 68
6.2 The reconstruction algorithm . . . 70
6.2.1 Stopping criterion . . . 70
6.2.2 Regularization terms . . . 71
6.2.3 Sparsifying transforms . . . 71
6.3 Prospective undersampling . . . 72
6.4 3D . . . 72
6.5 Conclusion . . . 73

7 Recommendations **74**
7.1 Implementation of MR scanner . . . 74
7.2 The reconstruction algorithm . . . 74
7.2.1 Stopping criterion . . . 74
7.2.2 Undersampling strategy . . . 74
7.2.3 Regularization terms . . . 75
7.2.4 Sparsifying transforms . . . 75
7.2.5 Combination SENSE and CS . . . 76
7.3 Applications . . . 76

A Appendix **77**
A.1 ADC formula 77
A.2 Convex and concave 77
A.3 Calculations used for (P_2) 78
A.4 Rewriting of structure term in SSIM formula 78
A.5 Optimal parameters 79

Bibliography **80**

List of Figures

2.1 Spinning nucleus . . . 4
2.2 Magnetic Resonance Imaging system . . . 5
2.3 Acquisition of k-space lines using a gradient sequence . . . 6
2.4 2D k-space . . . 7
2.5 The distribution of amplitudes in k-space and in image space . . . 9
2.6 Cartesian trajectory in k-space . . . 10
2.7 EPI trajectory in k-space . . . 10
2.8 Three images with different contrast imaging . . . 11
2.9 The Stejskal-Tanner sequence . . . 12
2.10 DWI of patient overlayed on a T_2 weighted image . . . 13
2.11 Shifted EPI trajectory . . . 15

3.1 Continuous-time cosine signal at 60 Hz . . . 16
3.2 Naive reconstruction of Shepp-Logan phantom using uniform mask . . . 18
3.3 Naive reconstruction of Shepp-Logan phantom using a non-uniform mask . . . 19
3.4 Schematic representation of ISTA . . . 25
3.5 Magnitude and phase images of Shepp-Logan phantom with constant phase of 0.8 35
3.6 The scaling functions and wavelets of three types of wavelet transforms . . . 37

4.1 Comparison of image fidelity measures for an Einstein image altered with different types of distortions . . . 40
4.2 Blocks shape used as patches . . . 43

5.1 Five undersampling strategies . . . 47
5.2 The 'centerincreased' mask . . . 49
5.3 Wavelet decompositions of three wavelet transforms . . . 50
5.4 The average mSSIM and mean structure values of thresholded images for three wavelet transforms . . . 51
5.5 The block effect of Haar wavelet . . . 52
5.6 Plot of the number of images with a mSSIM ≥ 0.9 . . . 53
5.7 Reconstructions of test slice for optimal parameters . . . 53
5.8 Example of 100%-image . . . 54
5.9 Stopping criteria . . . 55
5.10 Error plot of mSSIM values of ten images of one patient for five undersampling strategies . . . 57
5.11 Two 'random' strategy masks . . . 57
5.12 Percentages and subset sizes of wavelet coefficients needed for good reconstruction of ten patient volumes . . . 58
5.13 100%-image of patient 9 . . . 59

5.14 100%-image and reconstructions of patient 7 60
5.15 Boxplots of percentages required for good reconstruction of ninety test slices . . 60
5.16 Percentages required for good reconstruction of ten tumor volumes 62
5.17 Boxplots of percentages required for good reconstruction of tumor test slices . . . 62
5.18 ADC values 63
5.19 Work-flow 64
5.20 MRIs of fruit with undersampled acquisition 65
5.21 3D mask 66
5.22 Comparing reconstruction slice and reconstruction volume 67

A.1 Contourplots of ‘center’ and ‘increased’ 79

List of Tables

5.1 Order of proximal operators (10 slices) 56
5.2 Power spectral density fraction for 'random' undersampling 40% of the data . . . 58
5.3 Variation and worst case reconstruction percentages 61

A.1 mSSIM values obtained using the optimal α and β combination 79

List of Algorithms

1 Iterative Shrinkage-Thresholding Algorithm (ISTA) 26
2 Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) 28
3 Fast Gradient Projection method (FGP) 32
4 Composite splitting Denoising (CSD) 32
5 Modified FISTA 33
6 Monotone Fast Iterative Shrinkage-Thresholding Algorithm (mFISTA) 34
7 Complex Monotone Fast Iterative Shrinkage-Thresholding Algorithm (cFISTA) . . 35

List of Abbreviations

ADC Apparent Diffusion Coefficient. 13, 69, 77

cFISTA Complex Monotone Fast Iterative Shrinkage-Thresholding Algorithm. viii, 1, 2, 35, 36, 38, 50, 53, 54, 56, 59, 63–68, 70, 72–74, 76

CS Compressed Sensing. 1, 16, 18, 19, 36, 44, 46, 48, 66, 68, 71–73, 76

CSD Composite splitting Denoising. viii, 32, 33, 55, 56, 71

CW-SSIM Complex Wavelet Structural Similarity Image Measure. 44

DWI Diffusion-weighted magnetic resonance imaging. 1, 3, 12, 13, 15, 16, 46, 50, 66, 68, 69, 71–76

EPI Echo Planar Imaging. 1, 3, 4, 10, 13–15, 17, 45, 46, 67, 68, 70, 72–74, 76

FGP Fast Gradient Projection method. viii, 32

FISTA Fast Iterative Shrinkage-Thresholding Algorithm. viii, 1, 28, 29, 32–34, 68

fMRI functional Magnetic Resonance Imaging. 3, 4

FOV field-of-view. 8, 69

ISTA Iterative Shrinkage-Thresholding Algorithm. viii, 1, 24, 26–29, 34, 68

mFISTA Monotone Fast Iterative Shrinkage-Thresholding Algorithm. viii, 34–36

MPG Motion Probing Gradient. 12, 13

MR Magnetic Resonance. 1–3, 5, 11–13, 22, 34, 45, 63–66, 68, 72–74, 76

MRI Magnetic Resonance Imaging. 1, 3–5, 7, 12, 14, 18, 65

MSE Mean Squared Error. 39, 44

mSSIM Mean Structural Similarity Image Measure. 43–45, 50–63, 65–68, 70–73, 79

NMR Nuclear Magnetic Resonance. 1, 3–5

PGSE Pulsed Gradient Spin Echo. 12

RF Radio Frequency. 3, 5, 6, 10, 11

RIP Restricted Isometry Property. 20, 21

ROI region-of-interest. 13

SENSE Sensitivity Encoding. 18, 76

SNR Signal-to-Noise Ratio. 14, 15, 18

SSIM Structural Similarity Image Measure. 41–46, 66, 72, 74

TV Total variation. 21, 22, 30, 32, 59, 71, 75

TwIST Two-step Iterative Shrinkage-Thresholding Algorithm. 28

List of Notations

A	undersampling or measurement matrix with entries $\mathbb{C}^{p\times n}$	19
ΔB_0	field inhomogeneity	14
Λ	dynamical range of allowed image pixel intensities	42
γ	Gyromagnetic ratio	4
L_f	Lipschitz gradient constant of function f	23
b	measurements vector with entries $\mathbb{C}^p$	19
M_0	Longitudinal equilibrium magnetization	4
H	Number of patches in an image	43
p	number of measurements	19
m	number of rows	19
K	Number of slices	44
(P_0)	l_0-minimization problem	20
(P_1)	l_1-minimization problem	20
τ_y	the duration of the phase encoding step	14
τ_{ramp}	the ramp time of the switched gradients	14
$\vec{\mu}$	magnetic dipole moment vector	4
x	Signal or image to be reconstructed	19
x_k	Image patch k of x	41
D_w	Dwell time, so the interval between samplings of the spin echo	14
G_x	gradient in readout direction	6
G_y	gradient in phase encoding direction	6
G_z	slice selection gradient	6
M	the number of phase encoding steps	14
N	the number of readout steps	14

Chapter 1

Introduction

Diffusion-weighted magnetic resonance imaging (DWI) is a type of contrast imaging used in Magnetic Resonance Imaging (MRI). It visualizes the amount of diffusion of water molecules in tissue. This leads to a Magnetic Resonance (MR) image in which for example tumors are well visible. In practice, DWI is often acquired with the Echo Planar Imaging (EPI) technique. Unfortunately, the use of EPI techniques lead to geometric distortions in the DWI image. This in turn causes problems in locating the exact tumor position. The knowledge of the exact tumor location is necessary for radiotherapy.

In this thesis, an approach called Compressed Sensing (CS) is investigated as a technique to reduce geometric distortions in DWI. The data used to reconstruct an image is undersampled, which means that less data is obtained during scan acquisition. Undersampling leads to a reduced scan time and less distortions. By enforcing sparsity of the data in a transform domain, a well reconstructed image can be obtained as the solution of an appropriate minimization problem. After finding a reconstruction algorithm to solve the problem, the goal is to find the optimal undersampling strategy for the least amount of data possible to obtain good reconstructions.

In Chapter 2, an overview of the history and principles of MRI is given. The principles behind MRI and Nuclear Magnetic Resonance (NMR) are discussed and the concept of k-space is described. The working of DWI is explained and the problem of geometric distortions that arise when DWI is acquired with EPI is discussed.

In Chapter 3, the principles of CS are described. First, two possible undersampling strategies are discussed, namely uniform and non-uniform undersampling. Then the minimization problem, used for CS, will be stated. An iterative algorithm (ISTA) that will be used to solve the minimization problem will be derived and explained. ISTA has a slow convergence, and therefore it was improved. The approach by Beck and Teboulle in [1] resulted in the algorithm FISTA.
In this thesis, several modifications and improvements were implemented to this algorithm to finally obtain cFISTA. cFISTA can be used for complex-valued MRI data (Chapter 5).

To quantitatively compare the reconstructions from the undersampled data to the reconstructions without undersampling, a quality measure is required. In Chapter 4, three quality measures are discussed.

In Chapter 5, experiments will be performed on ten diffusion-weighted patient data sets. First, the efficiency of five undersampling strategies will be investigated. Next, choices of the parameters and transforms needed in the minimization problem and the reconstruction algorithm cFISTA will be made. Then the test images will be undersampled using the five undersampling strategies and then again reconstructed using cFISTA. This will lead to a optimal undersampling strategy among these five strategies for diffusion-weighted images of patients.
The working of cFISTA will be tested on actual MR data in Section 5.3.4. Finally, the code will be extended to be applicable to 3D volumes.

Chapter 2

Magnetic Resonance Imaging

In this chapter, an overview of the principles and the history of Magnetic Resonance Imaging (MRI) is given.
In Section 2.1, the principles of MRI and Nuclear Magnetic Resonance (NMR) are explained. In Section 2.2, the concept of k-space is described, which will be used in later chapters. Section 2.3 deals with three types of contrast imaging; the focus will be on diffusion weighted imaging (DWI). Finally, the main problem of this thesis is introduced: the geometric distortions that arise when DWI is acquired with Echo Planar Imaging (EPI).

2.1 NMR and MRI

Magnetic resonance imaging is an imaging technique used for imaging of soft tissues and their metabolic processes in the body [2]. MRI is based on the principle of NMR, which is a spectroscopic technique used to obtain microscopic chemical and physical information about molecules. MRI is noninvasive and does not employ ionizing radiation like X-ray imaging. A MR image represents the relative response of specific nuclei to absorbed radio frequency energy. The image is a function of nuclear density. The image contrast is furthermore influenced by physical factors, including differences in the ability to re-emit the absorbed Radio Frequency (RF) signal (relaxation) and flow phenomena [2].

2.1.1 History

In 1946, Felix Bloch and Eduard Purcell independently discovered the MR phenomenon, for which they receive a Nobel Prize in Physics in 1952. After this discovery, NMR was developed and used for chemical and physical molecular analysis and reaction processes [2]. In 1967, Jasper Johns had measured NMR signals from live animals and he proposed the first human applications [3].
Raymond Damadian found in 1971 that nuclear magnetic relaxation times of tissues and tumors differ, which made the technique useful for detecting diseases [4]. In 1973, the first MRI concepts were published in papers by Paul Lauterbur and Peter Mansfield. Lauterbur provided the first demonstration of MRI on small test tube samples using a back projection technique similar to the technique used in CT [5]. In 1975, Richard Ernst proposed MRI using phase and frequency encoding as well as Fourier transforms [6]. He defined the basis of the current MRI. His technique was used by William Edelstein and others for the demonstration of imaging the body. In 1977, Mansfield developed an imaging technique called EPI and in 1992 functional Magnetic Resonance Imaging (fMRI) was developed [3]. This last technique allowed mapping of the function of various regions of the brain and led to a new application for EPI in mapping regions

of the brain responsible for thought and motor control. Nowadays, fMRI and EPI are not only used for the brain, but also for other parts of the body. In 2003, Lauterbur and Mansfield received the Nobel Prize in Medicine for their discoveries concerning MRI.

2.1.2 Principles of Nuclear Magnetic Resonance

Magnetic resonance imaging started out as a tomographic imaging modality for producing NMR images of a slice through the human body [2]. Since MRI is based on NMR, first the principles of NMR will be discussed.

Certain nuclei possess an inherent angular momentum, or *spin*. A spin possesses an effective current loop which is capable of interacting with the static magnetic field B_0, measured in Tesla (T), as well as producing its own magnetic field. The strength of this loop is denoted by the magnetic dipole moment vector $\vec{\mu}$. The magnetic moments $\vec{\mu}$ in a collection of nuclei will be randomly oriented by the principles of Brownian motion. However, when a static magnetic field B_0 is applied, the spinning nuclei will either become parallel or anti-parallel with the direction of the applied field [2]. Due to thermal energy associated with the absolute temperature T, the number of spins parallel to the magnetic field exceeds the number of spins anti-parallel. This is called *spin excess* and it creates a longitudinal equilibrium magnetization M_0, parallel to the static field:

$$M_0 = \frac{\rho_0 \gamma^2 \bar{h}^2}{4kT} B_0,$$

with ρ_0 the number of protons per unit volume (spin density), k the Boltzmann constant and T the absolute temperature. The gyromagnetic ratio is denoted by γ and is a nucleus specific constant of proportionality. $\bar{h}$ is Planck's constant divided by 2π. This equilibrium value, limited by the spin excess, leads to measurable NMR effects.
The alignment of the magnetic moment with B_0 is not perfect (Figure 2.1). The spin vectors of the nuclei rotate around the axis of B_0 with a precession frequency ω called the *Larmor frequency*.

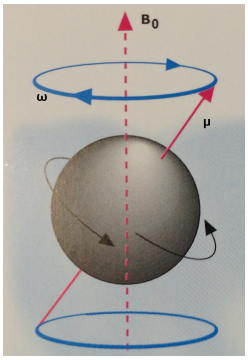


Figure 2.1: Spinning nucleus with magnetic dipole moment $\vec{\mu}$ oriented along external magnetic field B_0, precessing with Larmor frequency ω [3].

This frequency ω depends on the specific physical characteristics of the nucleus involved and the strength of the applied magnetic field [2]:

$$\omega_0 = \gamma B_0,$$

where ω_0 is the Larmor frequency for the applied magnetic field in MHz and γ is the gyromagnetic ratio.

Introducing a RF pulse onto the static magnetic field causes the spinning nuclei to flip from their initial position with magnetization M_0. This will lead to a decrease in the nuclear magnetization $M_z = M_0$, and an increase in the magnetization in the x- and y-direction, M_x and M_y respectively. The transition back to the initial state occurs over a period of time. This means that the amplitude of the received signal decreases over time.

Let $\hat{x}, \hat{y}$ and $\hat{z}$ be the unit vectors in the x-,y- and z-direction respectively. To calculate the nuclear magnetization $M = (M_x, M_y, M_z)$ as a function of time and position, the Bloch equation

$$\frac{dM}{dt} = \gamma M \times B_0\hat{z} + \frac{1}{T_1}(M_0 - M_z)\hat{z} - \frac{1}{T_2}M_{xy} \tag{2.1}$$

is used [2]. The first right-hand term of Equation (2.1) describes the precession of the spins. The second term contains the longitudinal magnetization M_z and denotes the long recovery of the magnetization. This process is called the *spin-lattice relaxation* where T_1 is the spin lattice relaxation time. The last term uses the transverse magnetization $M_{xy} := M_x\hat{x} + M_y\hat{y}$ to describe the transverse magnetization decay, which is the *spin-spin relaxation* with the spin-spin relaxation time T_2. It always holds that T_2 is shorter than T_1 [3]. More details about spin-lattice and spin-spin relaxation are discussed in Section 2.3.

The solution set of Equation (2.1) for each position $r = (x, y, z) = x\hat{x} + y\hat{y} + z\hat{z}$ and time t is given by

$$M_x(r,t) = e^{-t/T_2(r)}\left(M_x(r,0)\cos(\omega_0 t) + M_y(r,0)\sin(\omega_0 t)\right), \tag{2.2a}$$

$$M_y(r,t) = e^{-t/T_2(r)}\left(-M_x(r,0)\sin(\omega_0 t) + M_y(r,0)\cos(\omega_0 t)\right), \tag{2.2b}$$

$$M_z(r,t) = M_z(r,0)e^{-t/T_1(r)} + M_0(1 - e^{-t/T_1(r)}). \tag{2.2c}$$

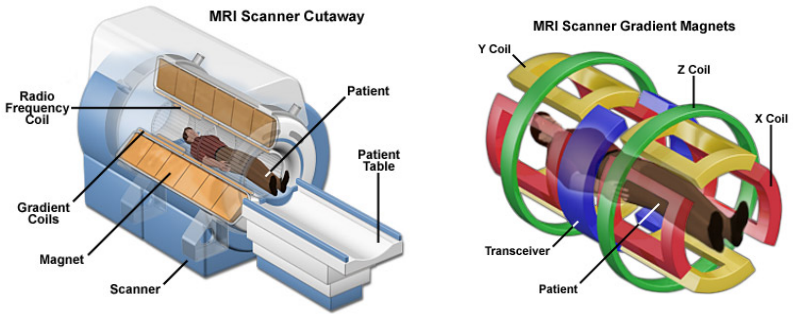
Note that Equation (2.2a) and Equation (2.2b) can be combined to

$$M_{xy}(r,t) = M_{xy}(r,0)e^{-t/T_2(r)}e^{-i\omega t}, \tag{2.3}$$

with $\omega = \omega_0 + \gamma\left(\vec{r}\cdot\vec{G}(t)\right)$. $\vec{G}(t)$ denotes a gradient that will be discussed in Section 2.1.3.

2.1.3 Principles of Magnetic Resonance Imaging

MRI associates a given NMR signal, as in (2.2), with a location in the tissue. The MR imaging system consists of several components (Figure 2.2).



(a) MR scanner.

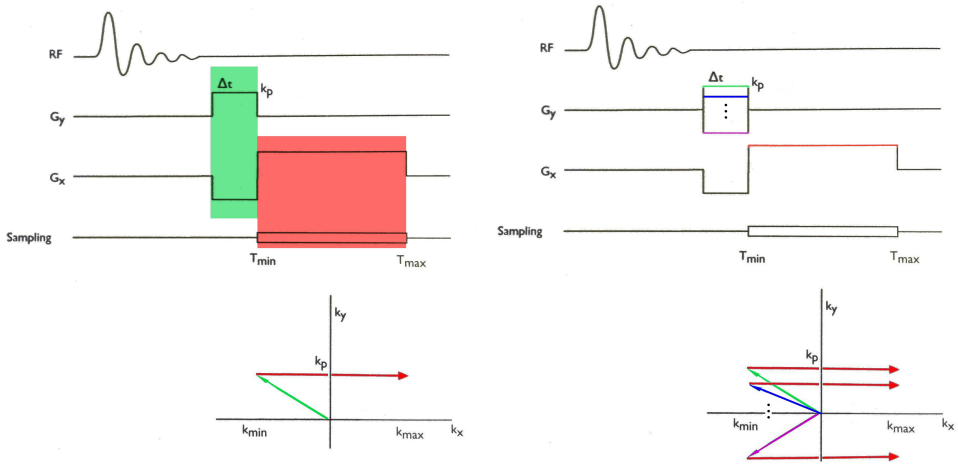
(b) Gradient coils system.

Figure 2.2: Magnetic Resonance Imaging system [7]. (a) A cross-section of a MR scanner. The magnet is used to generate a magnetic field. The radio frequency coil produces pulses to excite the nuclei and detects the signal from the nuclei. (b) The gradient coil system present in the MR scanner. These coils are used to encode the spatial position of the nuclei.

The magnet is used to generate the static magnetic field B_0, which influences the signal and image contrast. A RF coil is used to excite the nuclei and to receive the signal.

A weak magnetic field that linearly changes with position can be introduced on the main static field B_0 and this is used to create a magnetic field gradient. The magnetic field gradient system consists of three sets of orthogonally positioned coils (Figure 2.2b). These coils produce gradients in the x-, y- and z-directions required for spatial selection and spatial encoding.

The Z coil varies the magnetic field intensity in the head-to-feet direction. A patient is exposed to a RF pulse with a narrow range of frequencies, so a narrow bandwidth, which is ideally a block function. Only those nuclei in the thin slice with z-coordinate, say z_0, where the Larmor frequency matches the frequencies of the RF pulse will absorb the RF energy.
This process is a slice selective excitation. It uses the so-called *slice selection gradient* G_z. The slice thickness is controlled by the amplitude of the magnetic field gradient and the bandwidth of the RF pulse. Ideally, a block function in the frequency domain requires a RF pulse shaped like a sinc function in time.
The Y coil varies the intensity of the magnetic field in the anterior-to-posterior (or front-to-back) direction. It leads to the gradient field G_y, which is the so-called *phase encoding gradient*. This gradient is applied after G_z (Figure 2.3) to alter the phases of the spins according to their relative positions along the G_y gradient axis. By varying the strength of the gradient, the phases of the spins will differ in the y-direction.
The X coil varies the intensity of the magnetic field in the left-to-right direction and leads to the gradient field G_x, which is applied after G_y. This gradient is the so-called *readout gradient* and it provides spatial information of the spin density ρ_0 along the x-axis of the image slice.



(a) Acquiring one k-space line

(b) Acquiring several k-space lines

Figure 2.3: (a) A k-space line is acquired with a gradient sequence. Simultaneously applying G_y and G_x leads to a displacement from $(0, 0)$ to $(k_{\min}, k_p)$, as is illustrated by the green line in (a). Then only G_x is applied and between $T_{\min}$ and $T_{\max}$, the signal is sampled. This leads to a displacement from $(k_{\min}, k_p)$ to $(k_{\max}, k_p)$ (red line). (b) Multiple lines are acquired by varying the strength of G_y.

As will be discussed below, the Fourier transform of the detected signal is a projection onto slice. The amplitude of each frequency component is proportional to the summed signal in the

y-direction for that x-position.

The combination of the gradients is the basis for 2D Fourier techniques used in MRI. Each gradient is applied for a period of time, so the mentioned gradients are time-dependent. This results in the combined time varying gradient $\vec{G}(t) = [G_x(t), G_y(t), G_z(t)]$. This gradient leads to a variation at position (x, y, z) at time t in the magnetic field of size

$$xG_x(t) + yG_y(t) + zG_z(t) = \vec{r} \cdot \vec{G}(t).$$

Equation (2.3) can be rewritten as

$$M_{xy}(r,t) = M_{xy}(r,0)e^{-t/T_2(r)}e^{-ir\cdot\int_0^t \gamma\vec{G}(\tau)\,d\tau}, \tag{2.4}$$

where it has been used that the gradient magnetic field $\vec{G}(t)$ is proportional to the frequency ω.

2.2 k-space

During the time the readout gradient G_x is applied, the signal is sampled by acquiring data at N_x points. Repeating this sampling process N_y times leads to data of amplitudes at a $N_x \times N_y$-grid (Figure 2.4).

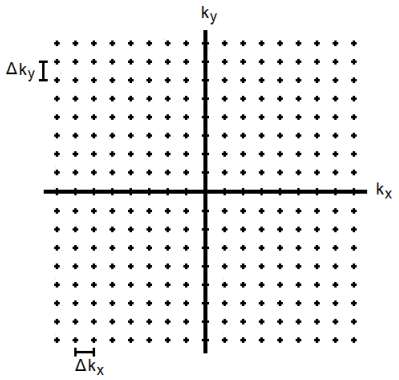


Figure 2.4: 2D k-space, where k_x is the frequency encoding direction and k_y is the phase encoding direction. The step sizes are Δk_x and Δk_y for the frequency encoding and phase encoding respectively.

Every point $k = (k_x, k_y)$ (or $k = (k_x, k_y, k_z)$ in 3D) in k-space is a spatial frequency and corresponds to the frequencies of a sinusoidal wave in the selected slice with wave number k that form the basis of the Fourier transform in the x- and y-directions. The signal emitted by the slice of interest (z_0) is represented in k-space and this is used to reconstruct an image, as will be shown below.

In practice, the spatial frequency k is obtained by

$$k := k(t) = \frac{\gamma}{2\pi}\int_0^t G(\tau)\,d\tau. \tag{2.5}$$

Amplitude information (data) on a line in k-space can be acquired by sampling in presence of the gradients according to (2.5) (Figure 2.3). k-space can be completely sampled, i.e. filling k-space, by varying the strength of the G_y gradient (Figure 2.3b).

In theory, the spatial frequencies could vary from $-\infty$ to ∞, leading to an infinite precision of the image. This is not practical for imaging and thus the process is truncated (between $k_{\min}$ and $k_{\max}$) and discretized (Δk).

$k_{\min}, k_{\max}$ and Δk are determined by the desired spatial resolution and the field-of-view (FOV) of the image. The FOV is the distance across an image, i.e the size of the spatial encoding area of the image. Note that Δk denotes Δk_x for the k_x-direction and Δk_y for the k_y-direction. Furthermore, $k_{\min}$ and $k_{\max}$ could be different for the k_x and k_y directions.

Using the notion of spatial encoding in k-space in Equation (2.5), Equation (2.4) can be rewritten as

$$M_{xy}(r,t) = M_{xy}(r,0)e^{-t/T_2(r)}e^{-i2\pi k(t)\cdot r}.$$

The received signal can be computed by integrating the transverse magnetization over the volume R

$$s(t) = \int_R M_{xy}(r,0)e^{-t/T_2(r)}e^{-i2\pi k(t)\cdot r}\,dr. \tag{2.6}$$

$M_{xy}(r,0)e^{-t/T_2(r)}$ can be written as the effective spin density $\rho_{\text{eff}}(x,y,z)$. The effective spin density is the true spin density corrected for effects like T_2 decay and field inhomogeneity dephasing [2]. The sampled signal can be written as a 2D Fourier transform for discretized k-space:

$$s(k_x,k_y) = \iiint \rho_{\text{eff}}(x,y,z)e^{-i2\pi(k_x x+k_y y)}\,dx\,dy\,dz. \tag{2.7}$$

Hence, s is the representation of ρ_{eff} in k-space, here also referred to as the image of ρ_{eff} in k-space or just as k-space in the chapters to come.
Recall that $t \rightsquigarrow (k_x(t), k_y(t))$ is a curve in k-space (Equation 2.5): $k_x = k_x(t)$ and $k_y = k_y(t)$. The Fourier transform of ρ_{eff} is measured for the points on this trajectory in k-space. For the points in k-space that are not situated on the trajectory, the signal value s is interpolated.

The result of the inverse Fourier transformation with respect to both k_x and k_y is a 2D image:

$$\hat{\rho}(x,y) = \int s(k_x,k_y)e^{i2\pi(k_x x+k_y y)}\,dk_x.$$

Hence, $\hat{\rho}$ is the restriction of $\rho(x,y,z_0)$ and will also be referred to here as the image of s in image space or just as the image.

The low spatial frequencies are located in the center of k-space. These frequencies mainly lead to contrast and large structures [8]. There is little details or edge information. After applying the inverse Fourier transform on a k-space containing only the spatial frequencies in the center, an blurred image is obtained in which the contrast information is preserved (Figure 2.5). To add details in the image, the higher spatial frequencies at the border of k-space must be included. As an example, all but the center part of k-space is selected and an image is obtained by using the inverse Fourier transform (Figure 2.5). Now the edge information and details are preserved and not the contrast information.

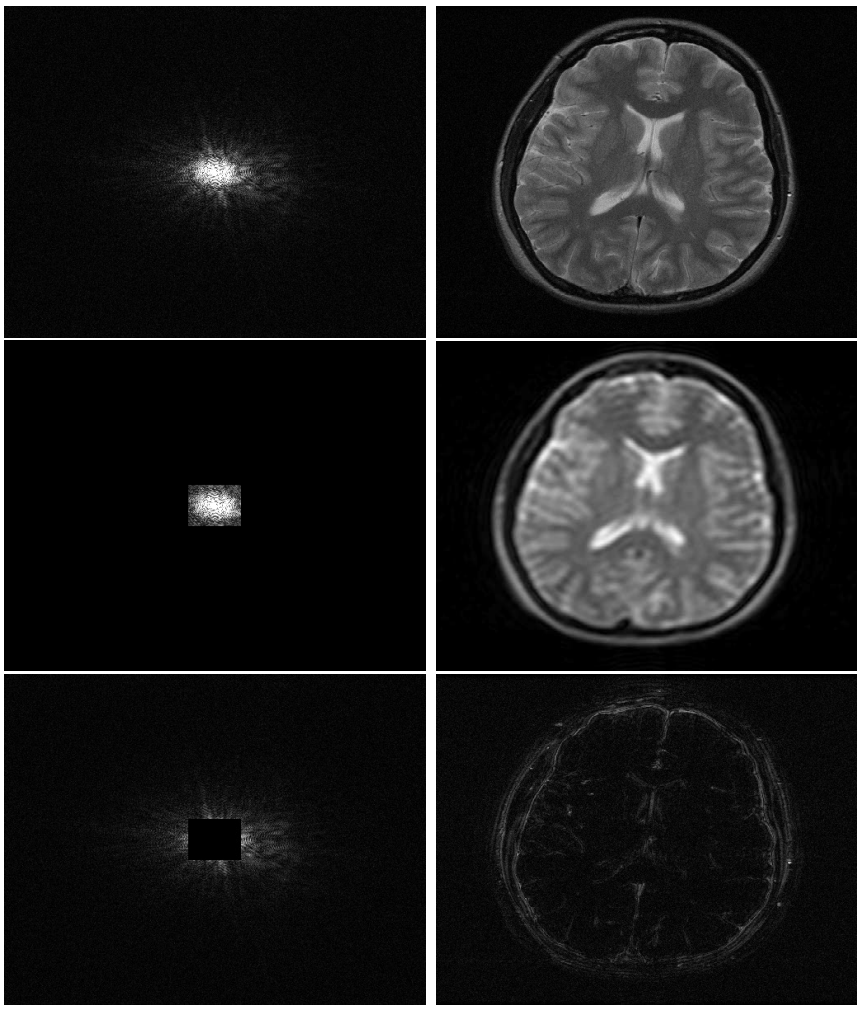


Figure 2.5: An image of the distribution of the amplitudes in k-space (left) and the resulting image of a brain (right) in image space, adapted from [9]. The top row shows the image of the full k-space and the corresponding back transformation in image space. The middle row shows only the center part of k-space. The corresponding image in image space is blurred in comparison with the top row image, but the contrast information is preserved. The bottom row contains an image in k-space where all but the center is selected. The corresponding back transformation in image space contains all details of the original image, but has no information about the contrast.

2.2.1 Trajectories

There are several ways to fill k-space. The standard way of filling k-space is a linear and Cartesian (or sequential) profile order, with one readout line after the other, from the bottom of k-space to the top. The lines are acquired from k_{min} to k_{max}, which gives a Cartesian (or rectilinear) coordinate system. This trajectory will be referred to as *standard Cartesian* (Figure 2.6).

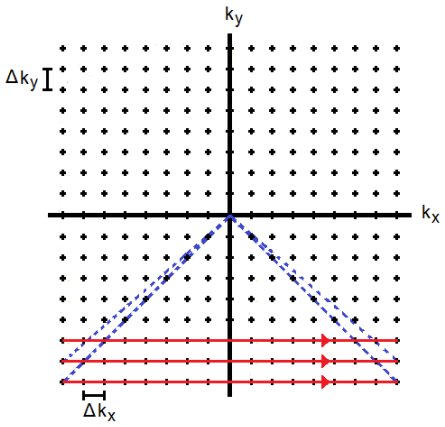


Figure 2.6: Cartesian trajectory in k-space. Applying the G_x and G_y gradients simultaneously leads to a displacement from (0,0) to the beginning of a readout line (blue), The lines are acquired in a sequential order from the bottom to the top of k-space (red). When the end of the readout line is reached, the trajectory goes back to the starting point (blue). Then the cycle can be repeated using different G_y strengths.

k-space can also be sampled in one excitation on the Cartesian grid by using Echo Planar Imaging (EPI)[10, 11, 12] (Figure 2.7).

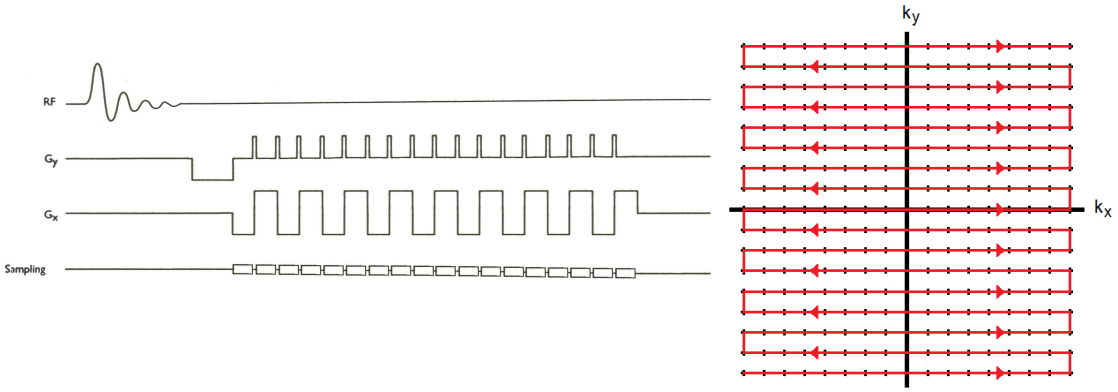


Figure 2.7: (a) The gradient sequence of EPI. The readout gradient is continuously applied with positive and negative alternations. The phase encoding gradients is applied at each echo onset (blip). (b) The EPI trajectory in k-space. k-space is filled in one excitation from the bottom to the top.

The readout gradient G_x is continuously applied, with positive and negative alternations, such that consecutive lines have opposite directions. A phase encoding gradient is applied at each echo onset (blip), giving a rectilinear trajectory.

2.3 Contrast Imaging

At equilibrium, the net magnetization vector M_0 lies along the direction of the applied magnetic field B_0, so $M_0 = M_z$. There is no transverse magnetization, so $M_x = M_y = 0$. By exposing the nuclear spin system to a RF pulse for a short time, the magnetization can be rotated away

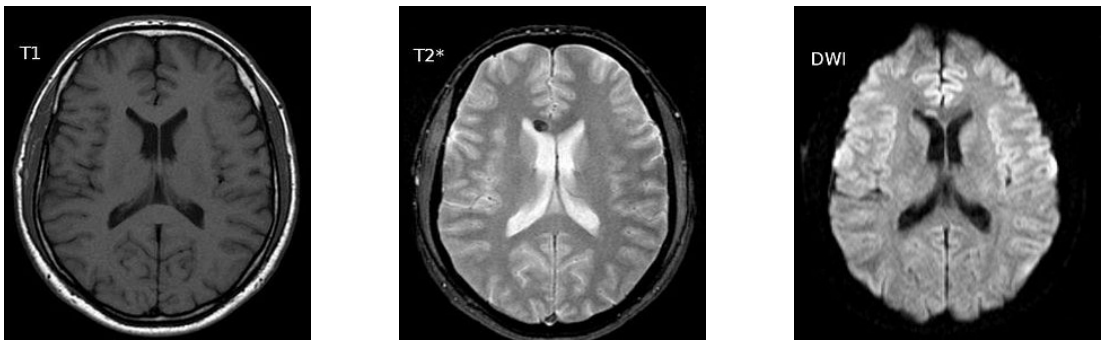
from its alignment along the B_0 axis. If the RF pulse gives enough energy to the system, it is possible to saturate it such that $M_z = 0$.

2.3.1 Spin-lattice relaxation and T_1-weighting

Spin-lattice relaxation is the process in which the spins give the energy, obtained by the RF pulse, to the surrounding lattice. In the Bloch equation (2.1), the spin-lattice relaxation time T_1 is used (Section 2.1.2). T_1 describes how M_z returns to its equilibrium value and can be computed using the Bloch equation solution (2.2c):

$$M_z(t) = M_z(0)e^{-t/T_1} + M_0(1 - e^{-t/T_1}).$$

T_1 is the time it takes to reduce the difference between the longitudinal magnetization M_z and its equilibrium value by a factor e.
T_1-weighted images demonstrate the differences in the T_1 relaxation time of tissues. Tissues with a short T_1 (like fat tissue) appear bright in the images (Figure 2.8a).



(a) T_1-weighted (b) T_2^*-weighted (c) Diffusion-weighted

Figure 2.8: Three MR images of a brain [13], displayed using three different contrast imaging techniques: T_1-weighted image (a), T_2^*-weighted image (b), and diffusion-weighted image (c).

2.3.2 Spin-spin relaxation and T_2-weighting

In addition to the rotation, the net magnetization starts to fan out because each group of spins is experiencing a slightly different magnetic field and rotates at its own Larmor frequency. The fanning out is also referred to as *dephasing*. The longer the elapsed time, the greater the phase difference will become. The time constant which describes the return to equilibrium of the transverse magnetization M_{xy} is called the spin-spin relaxation time T_2 and can be computed with

$$M_{xy}(t) = M_{xy}(0)e^{-t/T_2}. \tag{2.8}$$

T_2 is the time required to reduce the transverse magnetization by factor e and is always less than or equal to T_1. First, the net magnetization in the xy-plane will go to zero and then the longitudinal magnetization grows until M_0 is along the z-axis.

Two factors that contribute to the decay of the transverse magnetization are the molecular interactions and the variations in B_0. The first factor leads to the pure T_2 molecular effect and the second factor leads to an inhomogeneous T_2 effect, denoted by T_2'. This latter factor leads to a reduction in the initial value of the transverse components M_{xy}. A combination of these two factors is what actually results in the decay of the transverse magnetization. This combined

time constant is called T_2^* and is defined as $\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}$. In practice, the formula in (2.8) uses T_2^* instead of T_2, so

$$M_{xy}(t) = M_{xy}(0)e^{-t/T_2^*}.$$

By applying an additional pulse, the dephasing caused by external field inhomogeneities T_2' can be reversed, so the spins are rephased. T_2 cannot be recovered, since it is related to local, random, time-dependent field variations.

T_2-weighted images show the differences in the transverse relaxation of the net magnetization and thus the T_2 relaxation time of tissues. Tissues with a long T_2 (like water or cerebrospinal fluid) appear bright in the images (Figure 2.8b).

2.3.3 Diffusion-weighted imaging

When using diffusion-weighted imaging (DWI), the resulting image contrast is different from that provided by conventional MR techniques [14]. The image contrast is dependent on the molecular Brownian motion of water, which may be altered by disease. So, DWI provides functional information of tissues and thus can be used to detect and characterize pathologic processes. The most commonly applied method for diffusion-weighted contrast is the Pulsed Gradient Spin Echo (PGSE) [15, 16]. In this method, two symmetrical (diffusion) gradient lobes called Motion Probing Gradient (MPG)s are placed on both sides of the 180 ° pulse, which refocuses the spins. The sequence described here is invented by Stejskal and Tanner [17](Figure 2.9).

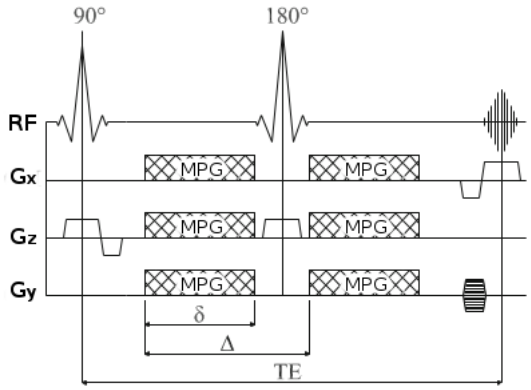


Figure 2.9: The Stejskal-Tanner sequence [16, 17], which is used to make MRI sensitive to diffusion. G_x,G_y and G_z are the gradients in readout, phase encoding and slice selection direction, respectively. δ is the duration of one MPG, Δ is the interval between the start of both MPGs and TE is the echo time, i.e. the time between the 90° and the center of the sampling time.

The degree of signal decrease in DWI depends on the magnitude of diffusion and on the amount of diffusion weighting. The amount of diffusion weighting is defined by the b-value in s/mm^2 and is computed as

$$b = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}),$$

with δ is the duration of one MPG, Δ is the interval between the start of both MPGs (Figure 2.9), G is the strength of the MPG and γ the gyromagnetic ratio.

The signal strength is described by

$$S(b) = S(0)e^{-bD},$$

where $S(b)$ is the signal intensity for a certain b-value, i.e. with MPG application. $S(0)$ is the signal intensity without the diffusion weighting and it is affected by T_2 and ρ. D is the diffusion constant of the tissue.
Tissues with high proton density and restricted diffusion appear bright on diffusion-weighted images (Figure 2.8c). Tissues with restricted diffusion are for example tumors. When tissues have unrestricted diffusion, the MR signal decreases and thus the tissues appear dark on the images. Often, the diffusion is expressed by an effective or Apparent Diffusion Coefficient (ADC). The ADC of a selected region-of-interest (ROI) can be calculated using two or more images with different b-values. Then, the fractioned ADC is computed between each pair of images with b-values b_1 and b_2 using the formula

$$ADC = -\frac{1}{b_2 - b_1} \ln\left(\frac{S[b_2]}{S[b_1]}\right). \tag{2.9}$$

For more details about the derivation of the ADC formula in (2.9), see Appendix A.1.

An ADC-map can be computed by estimation of the regional ADC pixel by pixel, using linear regression. In this map, the T_2 effect is removed. This effect also affects $S(b)$. Therefore, the ADC gives a quantitative measure of the diffusion for the MR data. Using the ADC, diffusion-weighted images of one patient can be compared to another patient. The ADC is also used to compare images of a patient scanned over several different days to monitor the effect of treatment.

2.4 Geometric distortions

For detecting a tumor, a diffusion-weighted image of the esophagus is placed on the T2 weighted image (Figure 2.10). In the center of the image, a tumor is located. It can be noticed that the DWI is slightly shifted in comparison with the T2 weighted image by looking at the tumor. This is immediately the problem when using DWI combined with EPI (Figure 2.7).

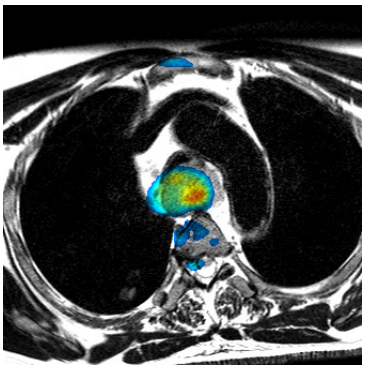


Figure 2.10: A transverse slice of a patient with an esophageal tumor. The diffusion-weighted image (color) is overlaid on a T_2 weighted image (gray). In the center of the image, the tumor is located. The DWI is slightly shifted in comparison with the T_2 weighted image, as can be seen by looking at the tumor.

2.4.1 Echo Planar Imaging

EPI is a fast practicable MRI method, which gives a high Signal-to-Noise Ratio (SNR) per unit time. However, a low pixel bandwidth in the phase encoding direction makes this sequence prone to geometric distortions [18].

In practice, the magnetic field B_0 is not completely homogeneous, so $\tilde{B}_0 = B_0 + \Delta B_0$. This leads to phase evolution between adjacent points in k-space.

The position of a pixel (x, y, z) in the image is determined by the phase evolution between consecutive discretizations of the phase. When k-space is filled using a Cartesian trajectory with multiple excitations, the phase evolution between adjacent points in k-space is defined as stated in [11]:

$$\Delta k_x x = \Delta\phi_x = 2\pi\gamma[G_x x D_w + \Delta B_0(x, y, z) D_w] \tag{2.10a}$$

$$\Delta k_y y = \Delta\phi_y = 2\pi\gamma \Delta G_y y \tau_y. \tag{2.10b}$$

Equation (2.10a) denotes the phase evolution in the readout direction and Equation (2.10b) denotes the phase evolution in the phase direction. ΔB_0 denotes the field inhomogeneity of the signal, γ is the gyromagnetic ratio, G_x and G_y are the gradients in readout and phase encoding direction, where ΔG_y denotes the phase encoding step. τ_y represents the duration of this phase encoding step and D_w denotes the dwell time, so the interval between sampling of the spin echoes. The inverse of the dwell time is called the *bandwidth*.
From (2.10) it can be concluded that the phase evolution in the read direction of k-space is sensitive to the presence of field inhomogeneities, whereas the pixel shift in the phase encoding is zero since there is no field inhomogeneity contribution.

The phase evolution for the EPI trajectory differs from the standard Cartesian trajectory. As defined in [11], EPI gives a phase evolution between adjacent points in k-space for point (x, y, z) as

$$\Delta\phi_x = 2\pi\gamma[G_x x D_w + \Delta B_0(x, y, z) D_w] \tag{2.11a}$$

$$\Delta\phi_y = 2\pi\gamma[G_y y \tau_{ramp} + \Delta B_0(x, y, z)(2\tau_{ramp} + N \cdot D_w)]. \tag{2.11b}$$

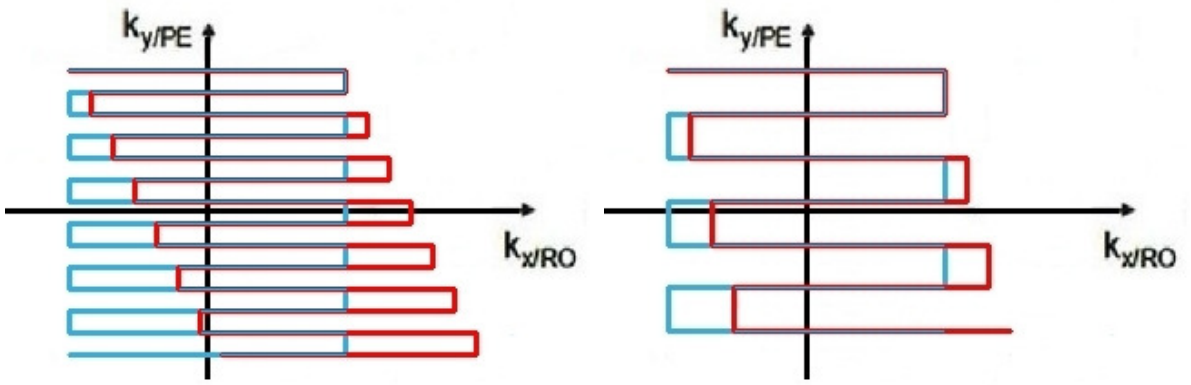
τ_{ramp} is the ramp time of the switched gradients and N is the number of samples on a line in k-space, i.e the number of readout steps.
As defined in [11], this phase evolution leads to a pixel shift as

$$\Delta r_{pe} = \gamma \Delta B_0(x, y, z) M(2\tau_{ramp} + N \cdot D_w). \tag{2.12}$$

Here, M is the number of phase encoding steps.

2.4.2 Reducing geometric inaccuracy

The phase evolution in (2.11a) is the same as the phase evolution in (2.10a). So in the read direction, the pixels are affected by the field inhomogeneity term $\Delta B_0(x, y, z)$. However, the phase evolution formulas (2.10b) and (2.11b) of the phase encoding direction differ. Using the EPI trajectory, the pixels are shifted and the shift is a factor $\frac{2\tau_{ramp}+N\cdot D_w}{D_w}$ bigger than the pixel shift in the read direction (Figure 2.11a). In practice, the pixel shift in the read direction is negligible compared to the shift in the phase encoding direction [11].



(a) Fully sampled k-space. (b) Undersampled k-space.

Figure 2.11: Ideal EPI trajectory in k-space (blue) and EPI trajectory with phase evolution (red). (a) The ideal EPI trajectory through k-space uses the phase evolutions of Equation (2.10). The red trajectory denotes how the phase evolutions of Equation (2.11) change the blue trajectory by some shift. (b) Undersampling k-space in the phase encoding direction leads to a reduced phase evolution and thus a smaller pixel shift.

However, if M is decreased, the pixel shift term also becomes smaller. So there is a smaller pixel shift in the phase encoding direction (Figure 2.11b). Decreasing M means that less lines are scanned in k-space. This principle is called *undersampling*.

By undersampling k-space, the geometric distortions of DWI are reduced because of a higher bandwidth. The frequency difference between successive phase encoding lines is higher. Therefore, the phase error due to magnetic field inhomogeneities has less effect in the total trajectory. Furthermore, a higher bandwidth leads to a shorter time to change the phase of the signal due to magnetic field inhomogeneities. This will produce less image distortions. However, this will be at the cost of a lower SNR.

Hence, to reduce the geometric distortions in DWI combined with EPI, k-space will be undersampled. The best strategies to undersample and to obtain a high quality reconstruction from the undersample k-space will be examined in the next chapters.

Chapter 3

Compressed Sensing

In this chapter, the principles of CS is discussed. In Section 3.1, the difference between uniform and non-uniform sampling will be investigated. In Section 3.2, the objective function of the reconstruction problem will be discussed. The algorithm used for solving the reconstruction problem will be derived in Section 3.3 and improvements on the existing algorithm will be made. Since the algorithm will use wavelet transforms, a short overview of wavelets will be given in Section 3.3.8.

3.1 Undersampling

In Chapter 2, it has been made plausible that geometric distortions in DWI can be reduced by undersampling an image in k-space. However, undersampling leads to other problems, such as *aliasing*.
Aliasing, or foldover, is an effect which causes that different signals cannot be distinguished by looking at their sampled values. A continuous-time cosine signal at 60 Hz is sampled for uniformly distributed (t_i) at 70 Hz (Figure 3.1). However, a cosine signal at 10 Hz can also be drawn through the sampling points $(t_i, f(t_i))$. So when only the sampling points are given, it is not clear which signal lies underneath.

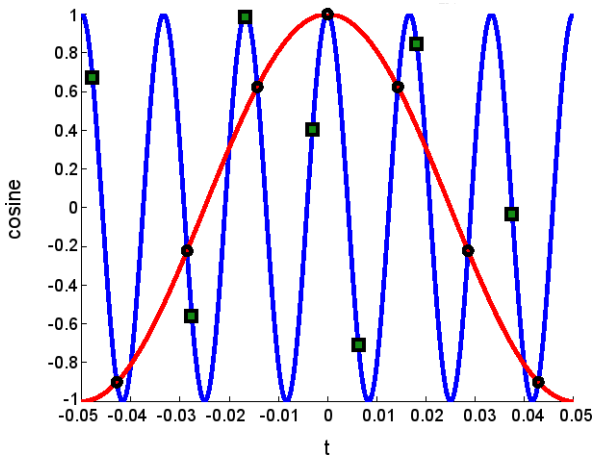


Figure 3.1: Continuous-time cosine signal at 60 Hz (blue), with (t_i) uniformly sampled at 70 Hz (circles), could also be a continuous-time cosine signal at 10 Hz (red). However, for the sampled points with (t_i) non-uniformly distributed (squares), there is only one unique continuous-time cosine signal possible.

The Nyquist-Shannon sampling theorem provides the sampling rate required to avoid aliasing with uniform sampling [19, 20]. The theorem states that the sampling frequency f_s should be at least twice the highest frequency f_h contained in the signal:

$$f_s \geq 2f_h.$$

This theorem gives a sufficient condition for perfect reconstruction of the signal. When k-space is undersampled, the Nyquist criterion is not satisfied and thus aliasing will occur.

However, when the signal is sampled for non-uniformly distributed (t_i) (Figure 3.1), and it is known that the underlying signal is a continuous-time cosine signal, then the cosine signal at 60 Hz is uniquely recovered from the sample points.

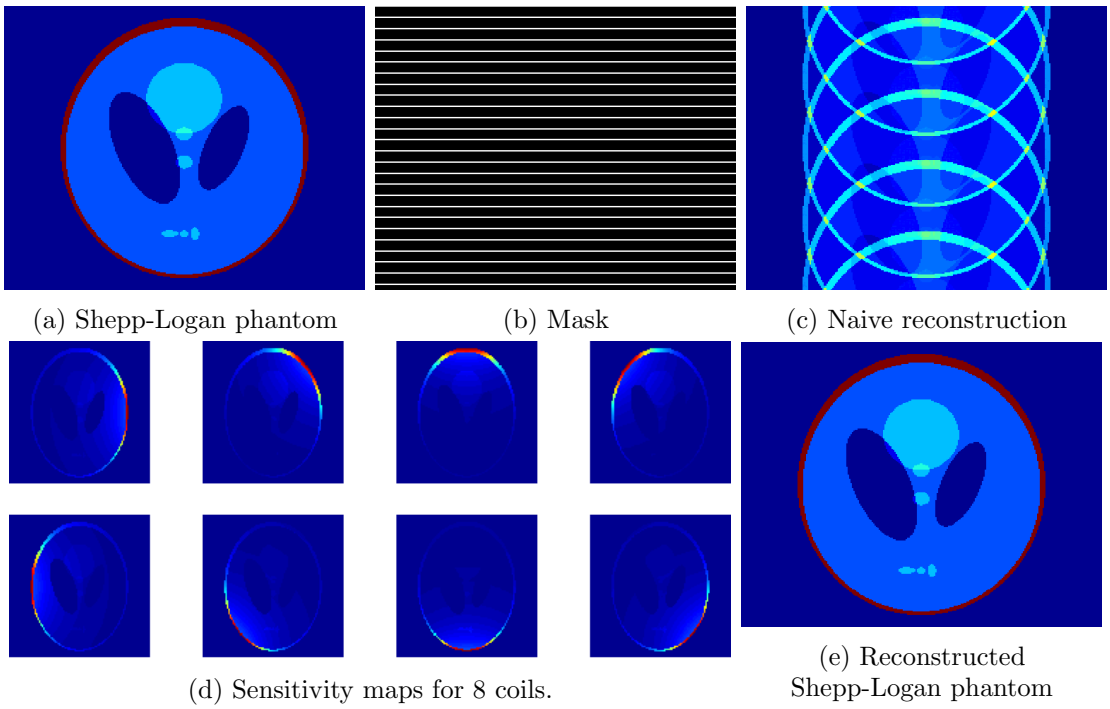
The questions now arising are about the best strategy to undersample k-space and the solution for avoiding aliasing artifacts.

3.1.1 Uniform vs non-uniform sampling

The first distinction that can be made is between uniform or non-uniform sampling. Non-uniform sampling could uniquely define a 1D signal, whereas uniform sampling can only do so if the Nyquist sampling theorem is satisfied (Figure 3.1).

To investigate the difference between these two sampling strategies for images, the Shepp-Logan problem is used in Matlab [21] (The MathWorks Inc., Natick, Massachusetts). The image of the Shepp-Logan phantom is transformed to an image in k-space using a Fourier transform. An undersampling mask is applied to the image in k-space. An undersampling mask indicates the lines that are scanned and the lines that are not, to get a certain percentage of the data in k-space (Figure 3.2b). After the undersampling, an inverse Fourier transform is applied and the result is a reconstructed image in image space. This method just undersamples in k-space and performs a simple Fourier reconstruction without any additional operations. Therefore, here this method is called a *naive reconstruction*.
Note that the undersampling is still only performed in one dimension, instead of the possible two dimensions. The reason for this is the use of the method EPI for data acquisition (Section 2.4.1). Therefore, all samples in one direction (the readout direction) will be taken. The undersampling will take place in phase encoding direction by increasing the bandwidth in this direction.

A naive reconstruction is performed on the Shepp-Logan phantom using a uniform mask for 25% of the image data in k-space (Figure 3.2). A uniform mask chooses lines with equal distance between them. The reconstructed image contains a lot of periodic aliasing (Figure 3.2c).



(a) Shepp-Logan phantom (b) Mask (c) Naive reconstruction

(d) Sensitivity maps for 8 coils.

(e) Reconstructed Shepp-Logan phantom

Figure 3.2: A 256 × 256 Shepp-Logan phantom (a) is undersampled using a uniform mask for 25% of the k-space data (b). The mask indicates the lines that are scanned (white) and the lines that are not (black). A naive reconstruction is obtained by applying an inverse Fourier transform (c). This reconstruction contains periodic aliasing. Using the sensitivity maps for each of the eight coils (d), an unfolding matrix can be computed. This matrix is used to unfold the naive reconstructed image of (c) to obtain an exact reconstruction (e).

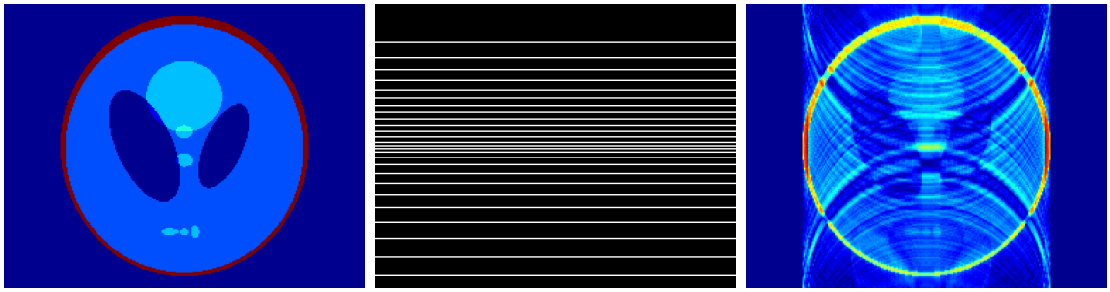
An example where a uniform mask used in practice, is MRI with Sensitivity Encoding (SENSE) [22]. With SENSE, the scan time is reduced at reserved spatial resolution, with the number of k-space samples reduced by a reduction factor R. For example, a reduction factor $R = 2$ leads to a mask where every second line is not scanned. With SENSE, the periodic aliasing is solved by making use of the sensitivity maps of the coils. Each coil produces an image in which a signal superposition occurs with different weights according to the sensitivity of that coil (Figure 3.2d). All coil sensitivities and the superimposed pixels form the sensitivity matrix S, which is used to calculate an unfolding matrix U. Using U, the naive reconstructed image can be unfolded (Figure 3.2e).

Although using SENSE can solve problems with uniform undersampling and thus leads to a scanning time reduction, there are also some disadvantages. In theory, SENSE reconstruction can be applied for reduction factors up to the number of coils used. However, the geometry factor will increase [23]. This factor describes the effect of varying noise enhancement according to the conditioning of the sensitivity-based reconstruction steps. So, using a higher reduction factor leads to more noise and thus a big penalty in the SNR. Furthermore, the performance of SENSE is very dependent of the coil set-up.

Here, a new approach is investigated, which will not be dependent on the coil set-up and also get a good SNR. This technique is called Compressed Sensing (CS).

3.1.2 Compressed Sensing

Using CS, a signal sampled in violation with the traditional Nyquist criterion could be reconstructed. Important for CS is that the signal has some sparse representation domain. To be able to reconstruct the signal, aperiodic artifacts have to be present in the transform domain. These aperiodic artifacts arise when k-space is undersampled non-uniformly (Figure 3.3).



(a) Shepp-Logan phantom (b) Mask (c) Naive reconstruction

Figure 3.3: A 256 × 256 Shepp-Logan phantom (a) is undersampled using a non-uniform mask for 25% of the data in k-space(b). Applying an inverse Fourier transform results in a naive reconstruction containing aperiodic aliasing (c).

Since a non-uniform mask is used, the distance between consecutive lines differs. The inverse Fourier transform, which is used to reconstruct the image in image space from the image in k-space, assumes that the lines are uniformly chosen. Now, the reconstructed image contains aperiodic aliasing, which are non-uniform undersampling artifacts.
To be able to reconstruct the underlying signal, using the aperiodicity of the artifacts, a non-linear method has to be used. The method used for CS reconstruction in Section 3.2 will take both sparsity and data consistency into account by stating an objective function consisting of a l_1-norm, l_2-norm, and Total Variation-norm.

3.2 Reconstruction problem

Without undersampling, there are m measurements. In this case, the signal can be recovered exactly. The main goal is to reconstruct a signal x by using only $p < m$ measurements. In matrix notation:

$$Ax = b, \tag{3.1}$$

where $A \in \mathbb{C}^{p \times n}$ is the measurement matrix, $x \in \mathbb{C}^n$ the signal and $b \in \mathbb{C}^p$ the measurements. When x is an image of size $n_1 \times n_2$, then x is reshaped to be a column vector of length $n = n_1 \cdot n_2$.

In applications, the measurements are often contaminated with some noise, i.e. $\hat{b} = b + \eta$ with η unknown noise. In this case, the system of equations in (3.1) becomes

$$\begin{aligned} Ax &= \hat{b} \\ &= b + \eta. \end{aligned} \tag{3.2}$$

The system of equations in (3.2) is an underdetermined system because there are more unknowns than equations. The goal is to find the underlying signal x that satisfies the data fidelity norm

$$\|Ax - b\|_2.$$

Underdetermined systems, as (3.1) and (3.2), have no unique solution. However, when the extra property that the signal is sparse with respect to a transform basis is given, then there is a probability to find the signal x.

Definition 1 (s-sparse and sparse)**.** *Choose $s \in \mathbb{N}_0$. Then $x = (x(1), \ldots, x(n))^T \in \mathbb{C}^n$ is called* s-sparse *if*

$$\|x\|_0 := \#\{i | x(i) \neq 0\} \leq s.$$

Furthermore, x is called sparse *if x is s-sparse for $s \ll n$.*

If x is sparse, then the underlying solution of (3.1) might be obtained by solving

$$(P_0) : \underset{x}{\operatorname{argmin}} \{\|x\|_0\} \text{ subject to } Ax = b. \tag{3.3}$$

or when there is noise contamination:

$$(P_0) : \underset{x}{\operatorname{argmin}} \{\|x\|_0\} \text{ subject to } \|Ax - b\|_2^2 < \varepsilon \quad (\varepsilon \text{ small}). \tag{3.4}$$

3.2.1 l_1-minimization

The problem (P_0) stated in (3.4) is a hard problem to solve. For this reason, the "l_0-norm" is replaced by a regularization term consisting of a l_1-norm. This leads to the problem

$$(P_1) : \underset{x}{\operatorname{argmin}} \{\|x\|_1\} \text{ subject to } \|Ax - b\|_2^2 < \varepsilon \quad (\varepsilon \text{ small}). \tag{3.5}$$

This problem is also called Basis Pursuit Denoising (BPDN) and is a convex optimization problem. The l_1-norm is used to replace the l_0-norm, because the l_1-norm is a well defined norm that is most similar to the l_0-norm, which is not a well defined norm. Furthermore, the l_1-norm is convex, which will be used in Section 3.3.

To ensure accurate reconstruction, Candès uses in [24] the Restricted Isometry Property (RIP), defined in Definition 2.

Definition 2 (RIP)**.** *Let A be a $m \times n$ matrix $(m < n)$ and take the restricted isometry constant δ_s as*

$$\delta_s := \sup \left\{ \left| 1 - \frac{\|Ax\|_2^2}{\|x\|_2^2} \right| \mid x \text{ is } s\text{-sparse} \right\}.$$

If $\delta_s < 1$, then A is said to have the s-restricted isometry property (RIP).

The combination of RIP and the l_1-minimization problem (P_1) lead to Theorem 1.

Theorem 1. *If the true signal $x^\star$ is s-sparse, $\delta_{2s} < \sqrt{2} - 1$ and $\tilde{x}$ is a minimizer of (P_1), then*

$$\|\tilde{x} - x^\star\|_2 \leq C_s \varepsilon, \text{ where } C_s = \frac{\sqrt{2(1 + \delta_{2s})}}{1 - (\sqrt{2} + 1)\delta_{2s}}.$$

Theorem 1 ensures that all s-sparse x are perfectly recovered via the l_1-minimization of problem (P_1). So the RIP gives a theoretical condition for the recovery of sparse vectors using l_1-minimization.
Certain types of random matrices have a high probability to obey the RIP property [25]. Examples are random matrices with independent and identically distributed entries with mean zero and variance $1/n$, Bernoulli matrices (where the entries are independently chosen from $\{-1, 1\}$

with uniform probability), and partial random Fourier matrices [26]. However, for many matrices it cannot be proved that they have the RIP property. Therefore, a RIPless theory is often used in practice on the measurement matrices [27]. This theory still implies that the solution of (3.6) is also the solution of (3.3).
The matrix A used here is of the latter type of measurement matrices. In the application in mind, the measurements b have a distribution with the most information located in the center of the vector b. A should also have a structure that will take this in account.

To make sure that the vector of interest is sparse, a sparsity transform Ψ is included in the l_1-norm. For example, Ψ can be a wavelet transform. So the l_1-minimization problem becomes

$$(P_1') : \underset{x}{\operatorname{argmin}} \{\|\Psi x\|_1\} \text{ subject to } \|Ax - b\|_2^2 < \varepsilon. \tag{3.6}$$

By taking $y = \Psi x$, problem (P_1') in (3.6) can be written as problem (P_1) in (3.5), that is

$$(\hat{P_1}) : \underset{y}{\operatorname{argmin}} \{\|y\|_1\} \text{ subject to } \|A\Psi^{-1}y - b\|_2^2 < \varepsilon. \tag{3.7}$$

3.2.2 TV-norm

Another regularization term in the form of the Total variation (TV)-norm is introduced to the problem (P_1'). This regularization term is often used in the minimization problem since the TV-norm can also be viewed as a sparsifier, as will be argued below.

The TV-norm for $x \in \mathbb{C}^{n_1 \times n_2}$ is defined as

$$\|x\|_{TV} := \sum_{i=1}^{n_1-1} \sum_{j=1}^{n_2-1} \sqrt{|x_{i,j} - x_{i+1,j}|^2 + |x_{i,j} - x_{i,j+1}|^2} + \sum_{i=1}^{n_1-1} |x_{i,n_2} - x_{i+1,n_2}| + \sum_{j=1}^{n_2-1} |x_{n_1,j} - x_{n_1,j+1}|.$$

For better interpretation of this TV-norm, define the discrete gradient operator $\mathcal{L} : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^{n_1 \times n_2} \times \mathbb{C}^{n_1 \times n_2}$ by

$$\mathcal{L}(x)_{ij} = (\partial_1(x)_{ij}, \partial_2(x)_{ij}), \quad \forall 1 \le i \le n_1, 1 \le j \le n_2, \tag{3.8}$$

where ∂_1 and ∂_2 are the difference operators for vectors in $\mathbb{C}^n = \mathbb{C}^{n_1 \times n_2}$

$$\partial_1(x)_{ij} = x_{i+1,j} - x_{i,j}, \quad \partial_2(x)_{ij} = x_{i,j+1} - x_{i,j} \quad (i = 1, \dots, n_1, j = 1, \dots, n_2), \tag{3.9}$$

using reflective boundary conditions, that is $x_{n_1+1,j} = x_{n_1,j}$ and $x_{i,n_2+1} = x_{i,n_2}$. Note that

$$\mathcal{L}(x)_{n_1,j} = (0, x_{n_1,j+1} - x_{n_1,j}) \text{ and } \mathcal{L}(x)_{i,n_2} = (x_{i+1,n_2} - x_{i,n_2}, 0).$$

Consider a vector space $\mathbb{C}^n \times \mathbb{C}^n$ with norm

$$\|(x, y)\| := \sum_{i=1}^{n} \|(x_i, y_i)\|_2 = \sum_{i=1}^{n} \sqrt{|x_i|^2 + |y_i|^2},$$

with $x = (x_1, \dots, x_n)^T, y = (y_1, \dots, y_n)^T \in \mathbb{C}^n$. This norm is equivalent to the l_1-norm $\|x\|_1 + \|y\|_1$ on $\mathbb{C}^n \times \mathbb{C}^n$, that is

$$\|(x, y)\| \le \|x\|_1 + \|y\|_1 \le \sqrt{2}\|(x, y)\|,$$

using the Minkowski and Cauchy-Schwartz inequalities on $\mathbb{C}^2$.

Then the use of this norm leads to the following representation of the TV-norm

$$\begin{aligned}\|x\|_{TV} &= \|\mathcal{L}(x)\| \\ &= \|(\partial_1(x), \partial_2(x))\|.\end{aligned}$$

If an image is piecewise constant, as is the case for MR images, then the TV-norm produces a sparse x by using the gradients of x. This means that the TV norm enforces sparsity, where it also uses the l_1-norm. So, the TV-norm can be seen as a sparsifier as well.

The TV-norm used in combination with the data consistency term in an objective function like (3.6) allows a denoising effect to happen on the applied image. In this situation, the piecewise smooth parts are smoothed in the resulting image x, while sharp edges are preserved, as an extra property of this norm.

3.2.2.1 3D version of TV-norm

To reconstruct a volume of size $n_1 \times n_2 \times n_3$, the definition of the TV-norm was modified to be applicable for 3D.

The 3D version of the discrete gradient operator $\mathcal{L}$ in (3.8) becomes $\mathcal{L}^{3D} : \mathbb{C}^{n_1 \times n_2 \times n_3} \to \mathbb{C}^{n_1 \times n_2 \times n_3} \times \mathbb{C}^{n_1 \times n_2 \times n_3} \times \mathbb{C}^{n_1 \times n_2 \times n_3}$,

$$\mathcal{L}^{3D}(x)_{ijk} = \left(\partial_1^{3D}(x)_{ijk}, \partial_2^{3D}(x)_{ijk}, \partial_3^{3D}(x)_{ijk}\right), \quad \forall 1 \leq i \leq n_1, 1 \leq j \leq n_2, 1 \leq k \leq n_3, \tag{3.10}$$

where the difference operators are defined as

$$\begin{aligned}&\partial_1^{3D}(x)_{ijk} = x_{i+1,j,k} - x_{i,j,k}, \quad \partial_2^{3D}(x)_{ijk} = x_{i,j+1,k} - x_{i,j,k}, \\ &\partial_3^{3D}(x)_{ijk} = x_{i,j,k+1} - x_{i,j,k} \quad (i = 1, \ldots, n_1, j = 1, \ldots, n_2, k = 1, \ldots, n_3).\end{aligned} \tag{3.11}$$

In Section 5.3.5, these definitions will be used to reconstruct an undersampled 3D volume.

3.2.3 The minimization problem

A combination of both regularization terms $\|\Psi x\|_1$ and $\|x\|_{TV}$ is considered here, with $\theta_1, \theta_2 > 0$, resulting in the following minimization problem

$$(P') : \underset{x}{\operatorname{argmin}} \{\theta_1 \|\Psi x\|_1 + \theta_2 \|x\|_{TV}\} \text{ subject to } \|Ax - b\|_2^2 \leq \varepsilon. \tag{3.12}$$

Using Lagrange multiplier λ, problem (P') can be written as

$$(P) : \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \alpha \|x\|_{TV} + \beta \|\Psi x\|_1 \right\}, \tag{3.13}$$

with $\alpha, \beta \in \mathbb{R}$ the regularization parameters:

$$\alpha = \frac{\theta_2}{2\lambda} \text{ and } \beta = \frac{\theta_1}{2\lambda}.$$

In practice, the values of ε and $\frac{\theta_1}{\theta_2}$ are not known and thus α and β are found experimentally in Section 5.2.2.

3.3 Solving the minimization problem

For solving (P), Beck and Teboulle proposed an iterative method called (F)ISTA [1, 28]. In this section, (F)ISTA and related methods will be derived for solving (P) (Equation (3.13)). First, some simpler problems are discussed, which will lead to a better understanding of problem (P).

3.3.1 Solving simpler problems

Problem (P) in (3.13) can be written as

$$\min_x \{F(x) := f(x) + g(x)\},$$

with $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ and $g(x) = \alpha\|x\|_{TV} + \beta\|\Psi x\|_1$.
$f(x)$ is a smooth convex function that is continuously differentiable with Lipschitz continuous gradient L_f, that is $\|\nabla f(x) - \nabla f(y)\| \leq L_f\|x - y\|\ \forall x, y$ and $g(x)$ is a convex function which is possibly non-smooth. The definition of a convex function is stated in Appendix A.2.
For problem (P) in (3.13), it holds that $\nabla f(x) = A^*(Ax - b)$ and $L_f = \lambda_{max}(A^*A)$.

To get more insight into the algorithm used in [1, 28], first look at a simpler problem:

$$(\tilde{P}) : \operatorname*{argmin}_x \left\{ \frac{1}{2}L\|x - y\|_2^2 + g(x) \right\}, \tag{3.14}$$

where $L > 0$ is a constant and $y := \frac{b}{L}$. In this case, matrix A is equal to $L \cdot I$ (with I the identity matrix). Problem $(\tilde{P})$ can be seen as a quadratic approximation of problem (P) around y.
Define the proximal regularization $\text{prox}_L(g)(y)$ as

$$\text{prox}_L(g)(y) := \operatorname*{argmin}_x \left\{ \frac{1}{2}L\|x - y\|_2^2 + g(x) \right\}. \tag{3.15}$$

If the proximal regularization problem $\text{prox}_L(g)(y)$ is solvable, then it suggests an iterative method as will be discussed below.

As an example how this iterative scheme may look like, consider

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 \text{ and } g(x) = \begin{cases} 0, & \text{if } x \in \mathcal{C} \\ \infty, & \text{if } x \notin \mathcal{C} \end{cases} \tag{3.16}$$

where $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^n$ and $g(x)$ is a convex, discontinuous function. Then $(\tilde{P})$ reads as

$$(P_2) : \operatorname*{argmin}_x \left\{ \frac{1}{2}\|Ax - b\|_2^2 \mid x \in \mathcal{C} \right\}, \tag{3.17}$$

The minimizer is restricted to this closed convex subset. This example will also be of importance in Section 3.3.5.

Next, it will be shown how the quadratic approximation of $\frac{1}{2}\|Ax - b\|_2^2$ $(x \in \mathcal{C})$ leads to a proximal regularization problem as in (3.15).

Let $x_0 \in \mathcal{C}$ be an approximate minimizer and define $x = x_0 + z$ and $r_0 = b - Ax_0$. For $L \geq \|A\|_2^2$, a quadratic upper bound for $\|Ax - b\|_2^2$ can be found that is equal to $\|Ax - b\|_2^2$ in x_0.

$$\begin{aligned}
\|Ax - b\|_2^2 &= \|Az + Ax_0 - b\|_2^2 = \|Az - r_0\|_2^2 \\
&= \|Az\|_2^2 - 2(Az, r_0) + \|r_0\|_2^2 \\
&\leq L\|z\|_2^2 - 2(Az, r_0) + \|r_0\|_2^2, \quad (L \geq \|A\|_2^2) \qquad (3.18) \\
&= L\|z - \frac{1}{L}A^*r_0\|_2^2 - \frac{1}{L}\|A^*r_0\|_2^2 + \|r_0\|_2^2, \quad \text{(Appendix A.3 Equation (A.1))} \\
&= L\|x - x_0 - \frac{1}{L}A^*r_0\|_2^2 - \frac{1}{L}\|A^*r_0\|_2^2 + \|r_0\|_2^2 \\
&= L\|x - y_0\|_2^2 - \frac{1}{L}\|A^*r_0\|_2^2 + \|r_0\|_2^2, \quad \text{with } y_0 = x_0 + \frac{1}{L}A^*r_0.
\end{aligned}$$

Hence,

$$\|Ax - b\|_2^2 \leq L\|x - y_0\|_2^2 + C,$$

with C a constant. So the quadratic function in (3.17) can be bounded from above by the sum of the simple quadratic function of (3.14) and some constant C. If $x = x_0$, then $\|Ax - b\|_2^2 = L\|x - y_0\|_2^2$, as can be easily checked using (3.18).

Computing a minimizer x_1 for $(\tilde{P})$ with $g(x)$ as defined in (3.16) will lead to an improvement of the minimizer x_0. So, if

$$x_1 = \underset{x}{\operatorname{argmin}} \left\{ L\|x - y_0\|_2^2 \;\middle|\; x \in \mathcal{C} \right\} \text{ with } y_0 = x_0 - \frac{1}{L}A^*r_0,$$

then $\|Ax_1 - b\|_2^2 \leq \|Ax_0 - b\|_2^2$. Repeating this process leads to a sequence $\{x_k\}$ of improved minimizers and thus an iterative scheme is derived.
The inequality is always strict, unless $(z, A^*r_0) = (x - x_0, A^*r_0) \leq 0 \; \forall x \in \mathcal{C}$. In that case, one can check that x_0 is the minimizer.

The calculation of $\text{prox}_L(g)(y)$ can require intensive computation, unless $\mathcal{C}$ is simple [1]. For example, half spaces and cubes are simple convex sets for which the projection of x onto these sets (which ensures $x \in \mathcal{C}$) is easy to compute. The set $\mathcal{C}$ used in Section 3.3.5 will be simple and thus the proximal regularization problem will be easy to compute.

3.3.2 ISTA

In this section, an iterative method called ISTA is described, which solves the general minimization problem

$$\underset{x}{\operatorname{argmin}} \left\{ F(x) := f(x) + g(x) \right\}. \qquad (3.19)$$

Suppose x_0 is an approximate solution. The quadratic upper bound of $F(x)$ is found by deriving the approximate quadratic Taylor expansion of f in $x = x_0 + z$ around x_0. For $L \geq L_f$,

$$\begin{aligned}
f(x) &= f(x_0 + z) \\
&= f(x_0) + (z, \nabla f(x_0)) + \frac{1}{2}\nabla^2 f(x_0)\|x_0 + z - x_0\|_2^2 \\
&= f(x_0) + (z, \nabla f(x_0)) + \frac{1}{2}\nabla^2 f(x_0)\|z\|_2^2 \\
&\leq f(x_0) + (z, \nabla f(x_0)) + \frac{1}{2}L\|z\|_2^2, \quad (L \geq \nabla^2 f) \\
&= f(x_0) + (x - x_0, \nabla f(x_0)) + \frac{1}{2}L\|x - x_0\|_2^2.
\end{aligned}$$

If $x = x_0$, then $f(x) = f(x_0) + (x - x_0, \nabla f(x_0)) + \frac{1}{2}L\|x - x_0\|_2^2$.

Hence

$$f(x) + g(x) \leq f(x_0) + (x - x_0, \nabla f(x_0)) + \frac{1}{2}L\|x - x_0\|_2^2 + g(x). \tag{3.20}$$

The upper bound in (3.20) can be seen as an approximate quadratic expansion of $F(x) = f(x) + g(x)$ in x around approximate solution x_0.

The next iterate x_1 is found by solving the minimization problem for the upper bound of $F(x)$. So

$$x_1 = \underset{x}{\operatorname{argmin}} \left\{ (x - x_0, \nabla f(x_0)) + \frac{1}{2}L\|x - x_0\|_2^2 + g(x) \right\}.$$

The objective function is rewritten to make the problem similar to (3.15) (for details, see Equation (A.2) in Appendix A.3). Then:

$$\frac{1}{2}L\|x - x_0\|_2^2 + (x - x_0, \nabla f(x_0)) = \frac{1}{2}L\|x - (x_0 - \frac{1}{L}\nabla f(x_0))\|_2^2 - \frac{1}{2L}\|\nabla f(x_0)\|_2^2.$$

Because $\frac{1}{2L}\|\nabla f(x_0)\|_2^2$ does not depend on x, the next iterate x_1 becomes

$$\begin{aligned} x_1 &= \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}L\|x - (x_0 - \frac{1}{L}\nabla f(x_0))\|_2^2 + g(x) \right\} \\ &= \operatorname{prox}_L(g)(y_1) \text{ with } y_1 := x_0 - \frac{1}{L}\nabla f(x_0). \end{aligned}$$

Hence, the next iterate x_1 can be found by solving the proximal regularization of $y_1 = x_0 - \frac{1}{L}\nabla f(x_0)$. By iteratively forming quadratic approximations around a specific point and by minimizing this approximation, the minimum $x^\star$ of the original problem will be approximately found (Figure 3.4).

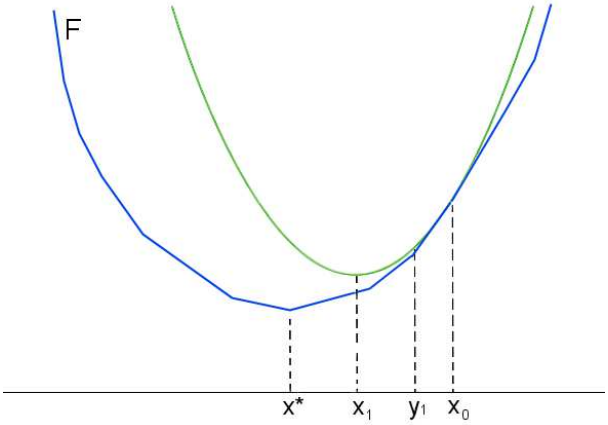


Figure 3.4: A schematic representation of the iterative method used for solving the minimization problem (3.19) with objective function F (blue). The approximate quadratic Taylor expansion (green) is computed in x_0. The gradient step $y_1 = x_0 - \frac{1}{L}\nabla f(x_0)$ controls the width of the parabola. Then the next approximate minimizer x_1 of F is found by minimizing the quadratic function. After several iterations, the solution $x^\star$ of the minimization problem (3.19) will be found.

This results in an iterative method called Iterative Shrinkage-Thresholding Algorithm (ISTA) (Algorithm 1).

Algorithm 1: ISTA

Input: x_0 and $L \geq L_f$
1 **for** $k = 1, 2, \ldots$ **do**
2 $\quad y_k = x_{k-1} - \frac{1}{L}\nabla f(x_{k-1})$
3 $\quad x_k = \text{prox}_L(g)(y_k)$
4 **end**

The name of Algorithm 1 originates from the use of a shrinkage operator to solve the proximal regularization in step 3 of Algorithm 1. For general $g(x)$, the proximal regularization step is solved using other methods, but the name of the algorithm remained ISTA. A shrinkage operator is used when $g(x) = \|\Psi x\|_1$, as in (3.6), which will be discussed in Section 3.3.5.

3.3.3 Error estimation and convergence of ISTA

Before a rate of convergence can be established, first an estimate for the difference in the objective function $F(x)$ of the general problem 3.19 is obtained in (3.22) of Lemma 2 below. This result is needed in (3.24). Recall $F(x) = f(x) + g(x)$.

For the general problem, it has been found useful to compute a lower bound of $F(x)$ for the error estimation. The lower bound simply comes in the form of the approximate linear Taylor expansion of f in $x = x_0 + z$ around x_0:

$$\begin{aligned} f(x) &= f(x_0 + z) \\ &\geq f(x_0) + (x_0 + z - x_0, \nabla f(x_0)) \\ &= f(x_0) + (z, \nabla f(x_0)) \\ &= f(x_0) + (x - x_0, \nabla f(x_0)). \end{aligned} \tag{3.21}$$

Now, an estimate for the difference in the objective function $F(x)$ can be obtained in Lemma 2.

Lemma 2. *If $L \geq L_f$, then*

$$\frac{2}{L}[F(x_1) - F(x)] \leq \|x_0 - x\|_2^2 - \|x_1 - x\|_2^2. \tag{3.22}$$

Proof. In (3.20), a quadratic upper bound $h(x)$ is found for $F(x)$:

$$\begin{aligned} F(x) = f(x) + g(x) &\leq f(x_0) + (x - x_0, \nabla f(x_0)) + \frac{L}{2}\|x - x_0\|_2^2 + g(x) \\ &= h(x) \end{aligned}$$

Taking $x = x_1$ gives $F(x_1) \leq h(x_1)$. Using the lower bound in (3.21) leads to

$$\begin{aligned} F(x) &\geq f(x_0) + (x - x_0, \nabla f(x_0)) + g(x) \\ &= f(x_0) + (x - x_0, \nabla f(x_0)) + \frac{L}{2}\|x - x_0\|_2^2 + g(x) - \frac{L}{2}\|x - x_0\|_2^2 \\ &= h(x) - \frac{L}{2}\|x - x_0\|_2^2. \end{aligned}$$

The difference $h(x_1) - h(x)$ can be bounded from above by $-\frac{L}{2}\|x - x_1\|_2^2$, that is

$$h(x_1) - h(x) \leq -\frac{L}{2}\|x - x_1\|_2^2. \tag{3.23}$$

To compute the difference between the objective functions in x_1 and x, subtract $F(x)$ from $F(x_1)$:

$$\begin{aligned} F(x_1) - F(x) &\leq h(x_1) - h(x) + \frac{L}{2}\|x - x_0\|_2^2 \\ &\leq -\frac{L}{2}\|x - x_1\|_2^2 + \frac{L}{2}\|x - x_0\|_2^2, \text{ (use (3.23))} \\ &= \frac{L}{2}\left(\|x_0 - x\|_2^2 - \|x_1 - x\|_2^2\right), \end{aligned}$$

and (3.22) follows. □

The error estimated in (3.22) will be denoted by $\eta(e)$, that is

$$\eta(e) = \frac{2}{L}[F(x) - F(x^\star)],$$

with $e := x - x^\star$ as the error between x and the minimizer $x^\star$ of $F(x)$. Define also $e_k := x_k - x^\star$ and $f_k := x_{k-1} - x^\star$. Note that $\eta(0) = 0$, since $e = 0 \Leftrightarrow x = x^\star$.

To investigate the convergence of the objective function, take a look at the difference between the errors when using x and x_1:

$$\begin{aligned} \eta(e_1) - \eta(e) &= \frac{2}{L}[F(x_1) - F(x^\star) - (F(x) - F(x^\star))] \\ &= \frac{2}{L}[F(x_1) - F(x))] \\ &\leq \|x_0 - x\|_2^2 - \|x_1 - x\|_2^2, \text{ (use (3.22))} \qquad (3.24) \\ &\leq \|x_0 - x^\star - (x - x^\star)\|_2^2 - \|x_1 - x^\star - (x - x^\star)\|_2^2 \\ &\leq \|f_1 - e\|_2^2 - \|e_1 - e\|_2^2. \qquad (3.25) \end{aligned}$$

Next, the error estimates for two interesting cases, namely $x = x^\star$ and $x = x_0$, are found. These cases lead to $e = 0$ and $e = e_0$, respectively:

$$\begin{cases} \eta_1 := \eta(e_1) & = \eta(e_1) - \eta(0) \leq \|f_1\|_2^2 - \|e_1\|_2^2 \\ \eta_1 - \eta_0 & \leq \|f_1 - e_0\|_2^2 - \|e_1 - e_0\|_2^2. \end{cases} \qquad (3.26)$$

In ISTA, $f_1 = x_0 - x^\star = e_0$. So for ISTA, (3.26) becomes

$$\begin{cases} \eta_1 & \leq \|e_0\|_2^2 - \|e_1\|_2^2 \\ \eta_1 - \eta_0 & \leq -\|e_1 - e_0\|_2^2. \end{cases} \qquad (3.27)$$

In general, it holds that

$$\begin{cases} \eta_k & \leq \|e_{k-1}\|_2^2 - \|e_k\|_2^2 \\ \eta_k - \eta_{k-1} & \leq -\|e_k - e_{k-1}\|_2^2. \end{cases} \qquad (3.28)$$

From the last statement of (3.28) follows that $\eta_k < \eta_{k-1}$.

Using (3.28), Proposition 3 about the convergence of ISTA can be proved.

Proposition 3.

$$F(x_k) - F(x^\star) \leq \frac{L\|e_0\|_2^2}{2k} = \frac{C}{k}, \quad \forall k$$

Proof. The sequence $\{x_k\}$ is generated using ISTA, therefore $f_1 = e_0$ and (3.28) holds true. Then

$$\begin{aligned} k\eta_k &\leq \eta_k + \eta_{k-1} + \ldots + \eta_1 \\ &\leq \|e_{k-1}\|_2^2 - \|e_k\|_2^2 + \|e_{k-2}\|_2^2 - \|e_{k_1}\|_2^2 + \ldots + \|e_0\|_2^2 - \|e_1\|_2^2, \text{ (use (3.28))} \\ &= \|e_0\|_2^2 - \|e_k\|_2 \\ &\leq \|e_0\|_2^2. \end{aligned}$$

So,

$$\begin{aligned} \eta_k &\leq \frac{\|e_0\|_2^2}{k} \\ \frac{2}{L}[F(x_k) - F(x^\star)] &\leq \frac{\|e_0\|_2^2}{k} \\ F(x_k) - F(x^\star) &\leq \frac{L\|e_0\|_2^2}{2k}. \end{aligned}$$

Thus $F(x_k) - F(x^\star) \leq \frac{C}{k}$.

□

So, ISTA has a worst-case complexity result of $\mathcal{O}(1/k)$. The convergence is guaranteed but may be slow.

3.3.4 Improvement ISTA: FISTA

There are several algorithms developed to improve the worst-case complexity result of ISTA, such as the Two-step Iterative Shrinkage-Thresholding Algorithm (TwIST) [29, 30] and the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [1, 28]. In these methods, x_{k+1} depends on both x_k and x_{k-1}, instead of just x_k as in ISTA.
Here, the reconstruction algorithm used for experiments in Chapter 5 is based on FISTA (Algorithm 2).

Algorithm 2: FISTA

Input: x_0 and $L \geq L_f$. Put $t_0 = 1$ and $y_1 = x_0$

for $k = 1, 2, \ldots$ **do**

- $x_g = y_k - \frac{1}{L}\nabla f(y_k)$
- $x_k = \text{prox}_L(g)(x_g)$
- $t_k = \frac{1}{2}\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$
- $y_{k+1} = x_k + \frac{t_{k-1}-1}{t_k}(x_k - x_{k-1})$

end

A sequence $\{y_{k+1}\}$ is generated using both x_k and x_{k-1} to find x_{k+1}. The parameter t_k changes the effect of two previous iterations. It has been chosen such that the relation $t_k^2 = t_{k+1}^2 - t_{k+1}$ holds. This relation is used in the error estimation of FISTA [1].
Beck and Teboulle derived an improved complexity result for FISTA by exploiting (3.25) with better choices for e than $e = 0$ and $e = e_0$ as in ISTA [1]. They proved that with FISTA

$$F(x_k) - F(x^\star) \leq \frac{2L_f\|e_0\|_2^2}{(k+1)^2} = \frac{C}{k^2}.$$

Hence, FISTA has a rate of convergence of at least $\mathcal{O}(1/k^2)$, and thus an improved worst-case complexity result has been found compared to ISTA.

3.3.5 Choices of $g(x)$

In step 3 of Algorithm 2, the proximal regularization of the function g is computed. Three possible functions for g are now discussed:

- $g(x) = \beta\|\Psi x\|_1$,
- $g(x) = \alpha\|x\|_{TV}$,
- $g(x) = \alpha\|x\|_{TV} + \beta\|\Psi x\|_1$.

For each $g(x)$, a method of solving the proximal operator $\text{prox}_L(g)(y)$ is discussed.

Proximal operator for $\|\cdot\|_1$

If $g(x) = \beta\|\Psi x\|_1$, then $\text{prox}_L(g)(y)$ can be easily computed using a soft-thresholding or shrinkage operator $s_\lambda : \mathbb{R}^n \to \mathbb{R}^n$, coordinate-wise defined as

$$s_\lambda(y)_j := \text{sign}(y_j)\max(|y_j| - \lambda, 0), \quad j = 1, \ldots, n. \tag{3.29}$$

So, for $\hat{x} = \Psi x$,

$$\begin{aligned}\text{prox}_L(\beta\|\hat{x}\|_1)(y) &= \underset{\hat{x}}{\text{argmin}} \left\{ \frac{1}{2}L\|\Psi^{-1}\hat{x} - y\|_2^2 + \beta\|\hat{x}\|_1 \right\} \\ &= \underset{\hat{x}}{\text{argmin}} \left\{ \frac{1}{2}L\|\Psi^{-1}(\hat{x} - \Psi y)\|_2^2 + \beta\|\hat{x}\|_1 \right\}. \end{aligned} \tag{3.30}$$

If Ψ is unitary (i.e. the l_2-norm is preserved: $\|\Psi y\|_2 = \|y\|_2$), then Ψ^{-1} is unitary as well and then (3.30) can be written as

$$\underset{\hat{x}}{\text{argmin}} \left\{ \frac{1}{2}L\|\hat{x} - \Psi y\|_2^2 + \beta\|\hat{x}\|_1 \right\},$$

which can be solved using the shrinkage operator

$$\hat{x}^\star = \text{prox}_L(\beta\|\hat{x}\|_1)(y) = s_\beta(\Psi y).$$

Hence, the solution of $\text{prox}_L(\beta\|\Psi x\|_1)(y)$ is found by $x^\star = \Psi^{-1}\hat{x}^\star = \Psi^{-1}s_\beta(\Psi y)$ (Ψ unitary).

Proximal operator for $\|\cdot\|_{TV}$

If $g(x) = \alpha\|x\|_{TV}$, then $\text{prox}_L(g)(y)$ is harder to compute than $\text{prox}_L(\beta\|\Psi x\|_1)(y)$. Chambolle proposed a dual approach and suggested a gradient-based algorithm for solving this dual problem [31].
To construct a dual of the constrained problem

$$\text{prox}_L(\alpha\|x\|_{TV})(y) = \underset{x\in\mathcal{C}}{\text{argmin}} \left\{ \frac{1}{2}L\|x - y\|_2^2 + \alpha\|x\|_{TV} \right\}, \tag{3.31}$$

the $\|\cdot\|$-norm, the discrete gradient operator $\mathcal{L}(x)$ and the difference operators $\partial_x u(i,j)$ and $\partial_y u(i,j)$ are used (Section 3.2.2). $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^n$, as in (3.17).

$\mathcal{L}^T(p,q) := \partial_1^T p + \partial_2^T q$ is the adjoint of $\mathcal{L}(x) = (\partial_1 x, \partial_2 x)$, that is

$$(p, \partial_1 x) + (q, \partial_2 x) = (\partial_1^T p + \partial_2^T q, x), \quad p, q, x \in \mathbb{C}^n.$$

Let $\mathcal{B}$ be the set $\{(p,q) \in \mathbb{R}^n \times \mathbb{R}^n | \|(p_i, q_i)\|_2 \leq 1 \ \ \forall i = 1, \ldots, n\}$. The relations

$$\sqrt{x^2 + y^2} = \max_{p_1, p_2} \left\{ p_1 x + p_2 y \;\middle|\; |p_1|^2 + |p_2|^2 \leq 1 \right\},$$
$$|x| = \max_{p} \left\{ px \;\middle|\; |p| \leq 1 \right\}$$

hold true by using Hölder inequalities. Using these relations, the TV-norm can be rewritten for the constrained case as

$$\|x\|_{TV} = \max \left\{ (\partial_1^T p + \partial_2^T q, x) | (p,q) \in \mathcal{B} \right\} \tag{3.32}$$

Equation (3.32) can be used to write the dual problem of (3.31), as will be done in the proof of Proposition 4 below. The definitions of the used operations differ from the operations by Beck and Teboulle in [28], leading to a different derivation of the dual problem.
Proposition 4 shows that the optimal solution of (3.31) is equal to the optimal solution of the dual problem in (3.33), where a dual of (3.31) is stated. The problem in (3.33) is similar to the problem stated in (3.17), where the minimizer of the problem containing the l_2-norm constrained to a subset $\mathcal{C}$ was approximated.

Proposition 4. *Let $(p^\star, q^\star) \in \mathcal{B}$ be the optimal solution of the problem*

$$\max_{(p,q) \in \mathcal{B}} \left\{ -\frac{1}{2} \| y - \lambda \left(\partial_1^T p + \partial_2^T q \right) \|_2 \right\} \tag{3.33}$$

with $\lambda = \alpha / L$. Then the optimal solution of (3.31) *is given by*

$$x^\star = y - \lambda (\partial_1^T p^\star + \partial_2^T q^\star)$$

Proof. First, notice that

$$\|x - y\|_2^2 = \|x\|_2^2 - 2(x,y) + \|y\|_2^2 \text{ implies } \|x\|_2^2 - 2(x,y) = \|x - y\|_2^2 - \|y\|_2^2. \tag{3.34}$$

Then

$$\begin{aligned} \|x - (y - z)\|_2^2 &= \|x\|_2^2 - 2(x, y - z) + \|y - z\|_2^2 \\ &= \|x\|_2^2 - 2(x,y) + 2(x,z) + \|y - z\|_2^2 \\ &= \|x - y\|_2^2 - \|y\|_2^2 + 2(x,z) + \|y - z\|_2^2 \ \ \text{(use (3.34))}. \end{aligned} \tag{3.35}$$

Rewriting and multiplying (3.35) by a factor $\frac{1}{2}$ leads to

$$\frac{1}{2}\|x - y\|_2^2 + (x,z) = \frac{1}{2}\|x - (y - z)\|_2^2 + \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - z\|_2^2. \tag{3.36}$$

Take $z = \lambda \left(\partial_1^T p + \partial_2^T q \right)$ to relate (3.36) to (3.33) and take the minimum over all $x \in \mathbb{C}^n$,

$$\begin{aligned} \operatorname*{argmin}_{x \in \mathcal{C}} \left\{ \frac{1}{2}\|x - y\|_2^2 + \left(x, \lambda \left(\partial_1^T p + \partial_2^T q \right) \right) \right\} = \operatorname*{argmin}_{x \in \mathcal{C}} \Bigg\{ & \frac{1}{2} \| x - \left(y - \lambda \left(\partial_1^T p + \partial_2^T q \right) \right) \|_2^2 \\ & + \frac{1}{2}\|y\|_2^2 - \frac{1}{2} \| y - \lambda \left(\partial_1^T p + \partial_2^T q \right) \|_2^2 \Bigg\}. \end{aligned} \tag{3.37}$$

Since $\frac{1}{2}\|y\|_2^2$ and $-\frac{1}{2}\|y-\lambda\left(\partial_1^T p+\partial_2^T q\right)\|_2^2$ do not depend on x, these terms are constant in the minimization process and can therefore be omitted in (3.37):

$$\underset{x\in\mathcal{C}}{\operatorname{argmin}}\left\{\frac{1}{2}\|x-y\|_2^2+\left(x,\lambda\left(\partial_1^T p+\partial_2^T q\right)\right)\right\}=\underset{x\in\mathcal{C}}{\operatorname{argmin}}\left\{\frac{1}{2}\|x-\left(y-\lambda\left(\partial_1^T p+\partial_2^T q\right)\right)\|_2^2\right\}, \tag{3.38}$$

which has the solution $x=y-\lambda\left(\partial_1^T p+\partial_2^T q\right)$. The minimum of the left-hand side of (3.38) for this optimal solution is obtained by taking x equal to this optimal solution in the right-hand side of (3.37):

$$\min_{x\in\mathcal{C}}\left\{\frac{1}{2}\|x-y\|_2^2+\left(x,\lambda(\partial_1^T p+\partial_2^T q)\right)\right\}=\frac{1}{2}\|y\|_2^2-\frac{1}{2}\|y-\lambda\left(\partial_1^T p+\partial_2^T q\right)\|_2^2. \tag{3.39}$$

Suppose that $x^\star=\operatorname{prox}_L(g)(y)$, with $g(x)=\alpha\|x\|_{TV}$:

$$x^\star=\underset{x\in\mathcal{C}}{\operatorname{argmin}}\left\{\frac{1}{2}L\|x-y\|_2^2+\alpha\|x\|_{TV}\right\}. \tag{3.40}$$

Then by (3.32), $x^\star$ solves

$$\min_{x\in\mathcal{C}}\max_{(p,q)\in\mathcal{B}}\left\{\frac{1}{2}\|x-y\|_2^2+\left(x,\lambda(\partial_1^T p+\partial_2^T q)\right)\right\}=\max_{(p,q)\in\mathcal{B}}\min_{x\in\mathcal{C}}\left\{\frac{1}{2}\|x-y\|_2^2+\left(x,\lambda(\partial_1^T p+\partial_2^T q)\right)\right\}. \tag{3.41}$$

Equation (3.41) holds since the order of the maximum and minimum can be interchanged for objective functions that are convex in x and concave in (p,q). Using (3.39) and the fact that $\|y\|_2^2$ is constant in the maximization over $(p,q)\in\mathcal{B}$, it holds that

$$\max_{(p,q)\in\mathcal{B}}\min_{x\in\mathcal{C}}\left\{\frac{1}{2}\|x-y\|_2^2+\left(x,\lambda(\partial_1^T p+\partial_2^T q)\right)\right\}=\max_{(p,q)\in\mathcal{B}}\left\{-\frac{1}{2}\|y-\lambda(\partial_1^T p+\partial_2^T q)\|_2^2\right\}. \tag{3.42}$$

From this follows that $x^\star=y-\lambda(\partial_1^T p^\star+\partial_2^T q^\star)\in\mathcal{C}$, where $(p^\star,q^\star)\in\mathcal{B}$ solve the dual problem in (3.42). □

To find this optimal solution $x^\star$ mentioned in Proposition 4, Chambolle proposed the Gradient Projection method [32]. In iteration k of this method, a solution $\tilde{x}=y-\lambda(\partial_1^T p_{k-1}+\partial_2^T q_{k-1})$ is computed. Then, the gradient in $\tilde{x}$ is added to the previous found (p_{k-1},q_{k-1}), resulting into $(\tilde{p}_k,\tilde{q}_k)$. To make sure that $(\tilde{p}_k,\tilde{q}_k)\in\mathcal{B}$, it is projected onto this set by $\mathcal{P}_{\mathcal{B}}(\tilde{p},\tilde{q})=(p,q)$, with $p,q\in\mathbb{R}^n=\mathbb{R}^{n_1\times n_2}$ given by

$$p_{ij}=\begin{cases}\frac{\tilde{p}_{ij}}{\max\{1,\sqrt{|\tilde{p}_{ij}|^2+|\tilde{q}_{ij}|^2}\}}, & i=1,\ldots,n_1-1, j=1,\ldots,n_2-1\\ \frac{\tilde{p}_{in_2}}{\max\{1,|\tilde{p}_{in_2}|\}}, & i=1,\ldots,n_1-1,\end{cases}$$

and

$$q_{ij}=\begin{cases}\frac{\tilde{q}_{ij}}{\max\{1,\sqrt{|\tilde{p}_{ij}|^2+|\tilde{q}_{ij}|^2}\}}, & i=1,\ldots,n_1-1, j=1,\ldots,n_2-1\\ \frac{\tilde{q}_{n_1 j}}{\max\{1,|\tilde{q}_{n_1 j}|\}}, & j=1,\ldots,n_2-1,\end{cases}$$

After a maximum number of iterations S, the optimal solution $x^\star$ of the proximal regularization problem in (3.31) is found.

To improve the Gradient Projection method of Chambolle, Beck and Teboulle have used steps 4 and 5 of FISTA (Algorithm 2) to obtain the Fast Gradient Projection method (FGP) (Algorithm 3) [28].

Algorithm 3: FGP

Input: $\tau < \frac{1}{4}$, $\lambda = \frac{\alpha}{L}$. Put $t_1 = 1$ and $(r_1, s_1) = (p_0, q_0) = (0_{(n_1-1)\times n_2}, 0)_{n_1\times(n_2-1)}$

1 **for** $k = 1, 2, \ldots, S$ **do**

2 $\quad \tilde{x}_k = y - \lambda(\partial_1^T p_{k-1} + \partial_2^T q_{k-1})$

3 $\quad (p_k, q_k) = \mathcal{P}_{\mathcal{B}}\left((r_k, s_k) + \frac{\tau}{\lambda}\mathcal{L}(\tilde{x}_k)\right)$

4 $\quad t_{k+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_k^2}\right)$

5 $\quad (r_{k+1}, s_{k+1}) = (p_k, q_k) + \left(\frac{t_k - 1}{t_{k+1}}\right)((p_k, q_k) - (p_{k-1}, q_{k-1}))$

6 **end**

7 **return** $x^\star = y - \lambda(\partial_1^T p_S + \partial_2^T q_S)$

Chambolle proved convergence of the algorithm (without steps 4 and 5) [32]. It holds that $\|\mathcal{L}(x)\|^2 \leq 8\|x\|$ and this is used to obtain an upper bound for the Lipschitz constant of the gradient of the objective function in (3.33). This leads to $\tau = \frac{1}{8}$, which is the value used in this thesis. The value of τ used to solve the proximal operator of the 3D version of the TV-norm is equal to $\frac{1}{12}$. Algorithm 3 has a rate of convergence of at least $\mathcal{O}(1/k^2)$.

Hence, $\text{prox}_L(g)(y)$ with $g(x) = \alpha\|x\|_{TV}$ can be solved by using FGP.

Proximal operator for $\|\cdot\|_{TV} + \|\cdot\|_1$

Huang *et al.* have proposed a method to compute $g(x) = \alpha\|x\|_{TV} + \beta\|\Psi x\|_1$ [33]. The Composite Splitting Denoising (CSD) method decomposes the problem

$$\min_x \frac{1}{2}\|x - x_g\|_2^2 + \sum_{i=1}^{h} g_i(B_i x), \tag{3.43}$$

with B_i matrices and $x_g = y_k - \frac{1}{L}\nabla f(y_k)$ (Algorithm 2), into h simpler regularization subproblems. Examples of B_i matrices are wavelet transforms and the identity matrix. In [33], B_i is required to be orthogonal. However, only invertibility of the matrices is important. The CSD method is stated in Algorithm 4.

Algorithm 4: CSD

Input: $L \geq L_f, \alpha, \beta, \{z_i^0\}_{i=1,\ldots,h} = x_g$

1 **for** $i = 1$ **to** h **do**

2 $\quad x_i = \text{argmin}_x \frac{L}{2h}\{\|x - z_i^{j-1}\|_2^2 + g_i(B_i x)\}$

3 **end**

4 $x = \frac{1}{h}\sum_{i=1}^{h} x_i$

Huang *et al.* stated a more general version of Algorithm 4 in [33, 34]. Combettes and Pesquet proved that the sequence $\{x_k\}$ generated by this general CSD method will converge to the true solution $x^\star$ of (3.43) [35]. Experiments showed that Algorithm 4 performed just as well as the general version [33, 34] and therefore Algorithm 4 will be used here.

Hence, the original problem with several non-smooth regularization functions is decomposed into easier subproblems, which can be solved individually with their own proximal operator. By fixing either the TV-norm or the l_1-norm, (3.43) can be solved. The solution of the original

problem (3.43) is obtained by taking the average of the solutions of the subproblems. So, the new x value is obtained using a Jacobi type of approach.
However, it may be more efficient to use a Gauss-Seidel type of approach, where each x_i value is immediately used to compute the next x_j ($j > i$). In many practical cases, the Gauss-Seidel method converges faster than the Jacobi method [36]. In Section 5.2.6, the differences between the Jacobi and Gauss-Seidel approaches are investigated for the application in mind.

The CSD method is inserted in FISTA (Algorithm 2) to solve

$$\frac{1}{2}\|x - x_g\|_2^2 + \alpha\|x\|_{TV} + \beta\|\Psi x\|_1,$$

and leads to the modified FISTA algorithm (Algorithm 5).

Algorithm 5: Modified FISTA

Input: x_0 and $L \geq L_f$. Put $t_0 = 1$ and $y_1 = x_0$

1 **for** $k = 1, 2, \ldots$ **do**
2 $\quad x_g = y_k - \frac{1}{L}\nabla f(y_k)$
3 $\quad$ *//start CSD algorithm*
4 $\quad x^1 = \text{prox}_L(2\alpha\|x\|_{TV})(x_g)$
5 $\quad x^2 = \text{prox}_L(2\beta\|\Psi x\|_1)(x_g)$
6 $\quad x_k = \frac{1}{2}(x^1 + x^2)$
7 $\quad$ *//end CSD algorithm*
8 $\quad t_k = \frac{1}{2}\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$
9 $\quad y_{k+1} = x_k + \frac{t_{k-1}-1}{t_k}(x_k - x_{k-1})$
10 **end**

In step 6 of Algorithm 5, the two solutions of the subproblems are combined. In [33], the two solutions are averaged and thus given equal weight. The weight of α and β in the proximal operator is twice as big. The idea behind this is explained next. Step 4 minimizes the problem

$$\frac{1}{2}\|x - x_g\|_2^2 + 2\alpha\|x\|_{TV},$$

and step 5 minimizes the problem

$$\frac{1}{2}\|x - x_g\|_2^2 + 2\beta\|\Psi x\|_1.$$

By averaging these problems, it follows that

$$\begin{aligned}\frac{1}{2}\left(\frac{1}{2}\|x - x_g\|_2^2 + 2\alpha\|x\|_{TV} + \frac{1}{2}\|x - x_g\|_2^2 + 2\beta\|\Psi x\|_1\right) &= \frac{1}{2}\|x - x_g\|_2^2 + \alpha\|x\|_{TV} + \beta\|\Psi x\|_1 \\ &= \text{prox}_L(\alpha\|x\|_{TV} + \beta\|\Psi x\|_1)(x_g),\end{aligned} \tag{3.44}$$

which is the problem to solve. Note that x^1 and x^2 are different and thus combination of these two solutions will differ from (3.44). However, (3.44) gives insight in the combination of the two subproblems.

It is possible to combine the subproblems using different weights: $x_k = px^1 + (1-p)x^2$. The computation steps of the proximal operators are modified for this combination:

$$x^1 = \text{prox}_L\left(\frac{\alpha}{p}\|x\|_{TV}\right)(x_g),$$
$$x^2 = \text{prox}_L\left(\frac{\beta}{1-p}\|\Psi x\|_1\right)(x_g).$$

Here, $p = \frac{1}{2}$ will be used.

3.3.6 mFISTA

Unlike ISTA, FISTA is not a monotone algorithm: the function values are not guaranteed to be non-increasing. When subproblems are not solved exactly, FISTA can diverge. To resolve this, a monotone version of FISTA is introduced, called mFISTA (Algorithm 6)[28].

Algorithm 6: mFISTA

Input: x_0 and $L \geq L_f$. Put $t_0 = 1$ and $y_1 = x_0$

1 **for** $k = 1, 2, \ldots$ **do**
2 $\quad x_g = y_k - \frac{1}{L}\nabla f(y_k)$
3 $\quad x^1 = \text{prox}_L(2\alpha\|x\|_{TV})(x_g)$
4 $\quad x^2 = \text{prox}_L(2\beta\|\Psi x\|_1)(x_g)$
5 $\quad z_k = \frac{1}{2}(x^1 + x^2)$
6 $\quad x_k = \text{argmin}_x\{F(y)|y \in \{z_k, x_{k-1}\}\}$; `// step to ensure monotonicity`
7 $\quad t_k = \frac{1}{2}\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$
8 $\quad y_{k+1} = x_k + \left(\frac{t_{k-1}}{t_k}\right)(z_k - x_k) + \left(\frac{t_{k-1}-1}{t_k}\right)(x_k - x_{k-1})$
9 **end**

The monotonicity step 6 of Algorithm 6 compares the function value of the found result z_k with the function value obtained using the previous iteration x_{k-1}. If $F(z_k) \leq F(x_{k-1})$, then the newest result x_k is equal to z_k, otherwise x_k is equal to the previous iterate x_{k-1}. Note that it is not certain that mFISTA would still work with the introduction of these modifications. However, experiments in Section 5 shows good results.

3.3.7 cFISTA

MR data is complex and therefore the algorithm used for reconstruction has to be applicable to complex values. The proximal regularization $\text{prox}_L(\|\cdot\|_1)(y)$ in Equation (3.30) of Section 3.3.5 is computed using the shrinkage operator in (3.45). However, this operator is not defined for complex-valued vectors. The operator is adapted for the use of complex-valued data by defining $\text{sign}(\rho e^{i\phi}) := e^{i\phi}$. The complex shrinkage operator $s_\lambda^c(y) : \mathbb{C}^n \to \mathbb{C}^n$ is coordinate-wise defined as

$$s_\lambda^c(y)_j := \frac{y_j}{(\max(|y_j| - \lambda, 0) + \lambda)} \cdot \max(|y_j| - \lambda, 0), \quad j = 1, \ldots, n. \tag{3.45}$$

Using this modification, Algorithm 6 becomes applicable to complex-valued problems and will

be called cFISTA (Algorithm 7).

Algorithm 7: cFISTA

Input: $x_0 \in \mathbb{C}^{m\times n}$ and $L \geq L_f$. Put $t_0 = 1$ and $y_1 = x_0$.

for $k = 1, 2, \ldots$ **do**

$x_g = y_k - \frac{1}{L}\nabla f(y_k)$

$x^1 = \text{prox}_L(2\alpha\|x\|_{TV})(x_g)$

$x^2 = \text{prox}_L(2\beta\|\Psi x\|_1)(x_g)$

$z_k = \frac{1}{2}(x^1 + x^2)$

$x_k = \text{argmin}_x\{F(z_k), F(x_{k-1})\}$

$t_k = \frac{1}{2}\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$

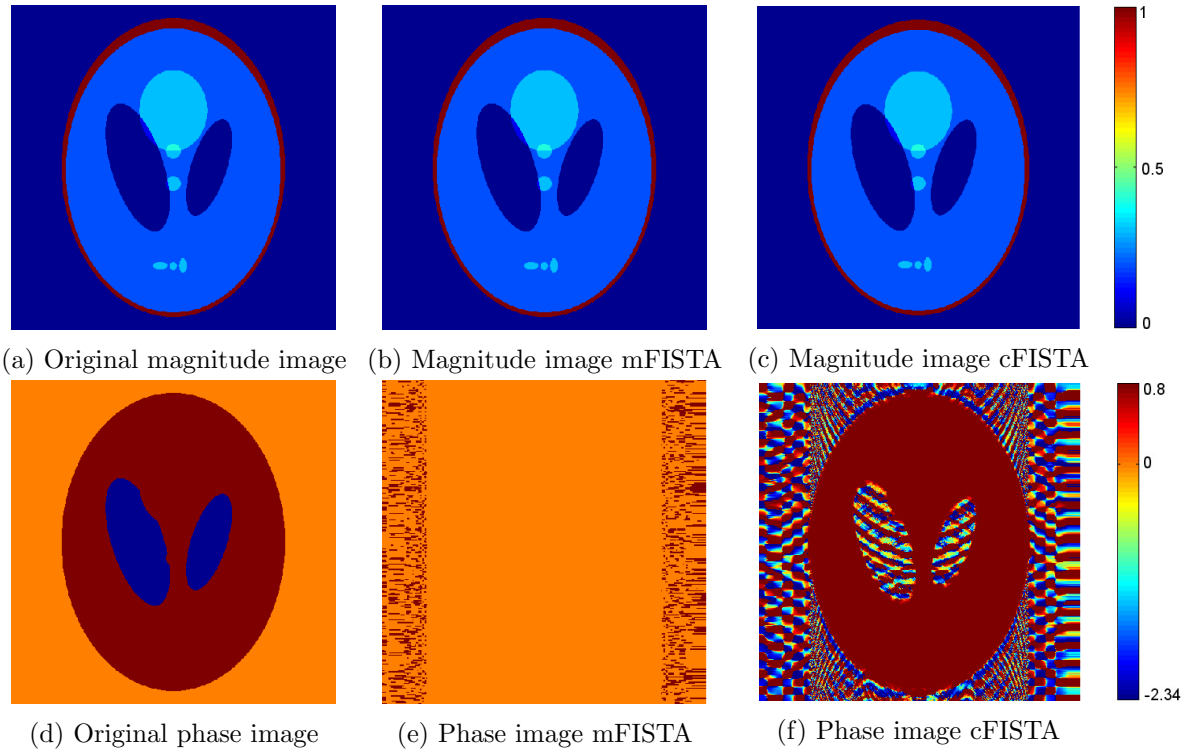
$y_{k+1} = x_k + \left(\frac{t_{k-1}}{t_k}\right)(z_k - x_k) + \left(\frac{t_{k-1}-1}{t_k}\right)(x_k - x_{k-1})$

end

The working of cFISTA is tested on the Shepp-Logan phantom P_{SL} of size 512×512. As an example, a constant phase of 0.8 is applied on P_{SL}, resulting in a new phantom called $P_{SL_{0.8}}$ (Figure 3.5a and Figure 3.5d):

$$P_{SL_{0.8}} = P_{SL} \cdot e^{0.8i}.$$

To observe the differences between mFISTA and cFISTA, the algorithms are applied on $P_{SL_{0.8}}$ to obtain the magnitude images (Figure 3.5b and 3.5c) and phase images (Figure 3.5e and 3.5f from the reconstructions of both algorithms.



(a) Original magnitude image (b) Magnitude image mFISTA (c) Magnitude image cFISTA

(d) Original phase image (e) Phase image mFISTA (f) Phase image cFISTA

Figure 3.5: Shepp-Logan phantom $P_{SL_{0.8}}$ with constant phase of 0.8. The magnitude and phase images of the original $P_{SL_{0.8}}$ (a,d), the reconstructed images using mFISTA (b,e) and the reconstructed images using cFISTA (c,f) are displayed. The phase in 3.5e is equal to zero and thus not restored correctly. The phase in 3.5f is restored correctly.

Instead of a phase of value 0.8, as in the original phase image (Figure 3.5d), the value of the reconstructed phase using mFISTA is equal to zero (Figure 3.5e). Hence, the phase reconstruction fails while using mFISTA. The phase of Shepp-Logan phantom using cFISTA is equal to 0.8 (Figure 3.5f) and is equal to the phase of $P_{SL_{0.8}}$. At the borders of the phantom, it seems that the algorithm has trouble restoring the phase, because it ranges between -2.3416 and 0.8. At the points of these fluctuations, the data is close or equal to zero. The phase cannot be restored correctly, since there is little to no information available about the phase at these points. In practice, a mask is applied to the reconstructed image. This mask is constructed using the magnitude image, from which can be derived where the data is equal to zero and thus the locations where it is not possible to reconstruct the phase correctly.

Hence, cFISTA reconstructs complex-valued data correctly and therefore an implementation of Algorithm 7 in Matlab [21] is used for the experiments in Section 5.

3.3.8 Wavelets

CS relies on the existence of a sparse solution (Section 3.2). To obtain such solution, a sparsifying transform Ψ is used. Here, Ψ is a wavelet transform. This is a tool that divides the data or functions into different frequency components [37]. A wavelet transform exploits a wavelet $\psi(t)$ and a scaling function $\varphi(t)$. Sometimes ψ is also referred to as a mother wavelet and φ as a father wavelet.
For any function h, a basis can be generated by contraction (a) and translations (b):

$$h^{a,b}(t) := \frac{1}{\sqrt{|a|}} h\left(\frac{t-b}{a}\right).$$

Here, the scaling $1/\sqrt{|a|}$ preserves the l_2-norm. In case that h is a wavelet, $h^{a,b}$ is called a daughter wavelet. Discrete translations and contractions are considered only:

$$a := 2^{-j} \text{ and } b := k2^{-j} \quad (j, k \in \mathbb{Z}).$$

This leads to the following definition:

$$h_{j,k}(t) := 2^{j/2} h(2^j t - k).$$

For each j, the scaling function φ generates a set $(\varphi_{j,k})$ that spans a space $\mathcal{V}_j$ of the function "at resolution level j". The scaling function φ has to be selected such that for the function f of interest, there exists a high quality approximation f_j for non-large values of j. The functions f_j from $\mathcal{V}_j$ for j large enough, i.e. for resolution 2^{-j} small enough, are of the form

$$f_j = \sum_k s_{jk} \varphi_{j,k}.$$

A wavelet ψ is associated to the scaling function φ. It is used to split the function f_j into f_{j-1} and the function g_{j-1} with

$$g_{j-1} \in \mathcal{W}_{j-1} = \text{span}\{\psi_{j-1,k} \mid k \in \mathbb{Z}\}$$
$$g_{j-1} = \sum_k d_{j-1k} \psi_{j-1,k}.$$

The wavelet ψ has to be selected such that the number of small coefficients in the collection $\{|d_{j-1k}| \mid k \in \mathbb{Z}\}$ and $\{|s_{j-1k}| \mid k \in \mathbb{Z}\}$ is (much) larger than in the collection $\{|s_{jk}| \mid k \in \mathbb{Z}\}$.

The function f_j is split into the smooth part f_{j-1} and the detail part g_{j-1}. Repeating this procedure l times results in

$$f_j = f_{j-l} + g_{j-l} + \ldots + g_{j-1},$$

where l is the wavelet level of decomposition. ψ is such that $\int \psi(x)\,dx = 0$, whence

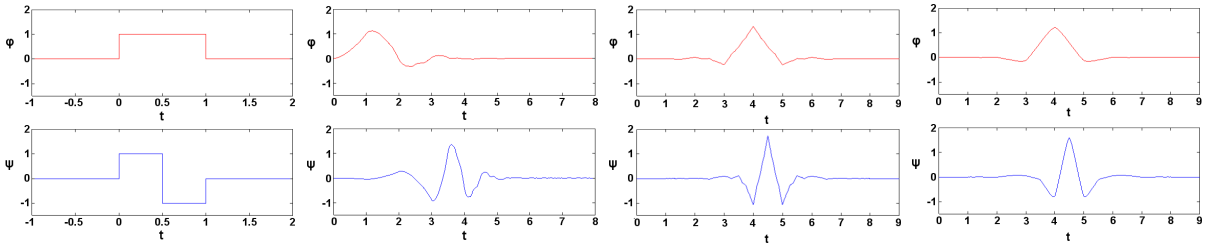
$$\int f_j = \int f_{j-l} \text{ and } \int g_i = 0,\ \forall i = j-1, \ldots, j-l.$$

If

$$\int \psi(t)\,dt = 0, \int t\psi(t)\,dt = 0, \ldots, \int t^r \psi(t)\,dt = 0,$$

then the wavelet is said to have $r+1$ vanishing moments. The number of vanishing moments determines the number of small detail coefficients d_{j-1k} for the smooth function f.

The scaling function φ and wavelet ψ of three wavelets are displayed in Figure 3.6.



(a) Haar wavelet (b) Daub4 (c) CDF 9/7 Decomposition (d) CDF 9/7 Reconstruction

Figure 3.6: The scaling functions (top row) and the corresponding wavelet functions (bottom row) of the Haar wavelet, Daubechies with four vanishing moments (Daub4) and CDF 9/7. CDF 9/7 is a non-orthogonal wavelet and therefore it requires two pairs of scaling and wavelet functions: the decomposition and reconstruction functions. The other two wavelet transforms are orthogonal and thus use the same functions for decomposition and reconstruction.

The map $f_j \rightsquigarrow f_{j-l} + g_{j-l} + \ldots + g_{j-1}$ is the wavelet transform. Usually this transform is computed at the level of coefficients. Then $(s_{jk}) \rightsquigarrow ((s_{j-lk}), (d_{j-lk}) \ldots, (d_{j-1k}))$ is the wavelet transform.

If the $\varphi_{j,k}$ form an orthogonal basis of $\mathcal{V}_j$ (i.e. $\varphi_{0,k} \perp \varphi_{0,m}$ $(k \neq m)$), then f_j can be obtained by projecting f onto $\mathcal{V}_j$:

$$f_j = \sum_k \frac{\langle f, \varphi_{j,k} \rangle}{\langle \varphi, \varphi \rangle} \varphi_{j,k}.$$

If $\psi_{j,k} \perp \psi_{j',m}$ $(\forall j \neq j'$ and k, m, and $\forall j = j', k \neq m)$, then ψ is said to be an orthonormal wavelet.

Wavelets allow good representation of local behavior of functions and can represent data using less terms [38]. The wavelet coefficients are sorted to the scale of the details, from coarse-scale details to fine-scale details. The main difference between the standard Fourier transform and wavelet transforms is that the bases functions of the Fourier transforms are only localized in frequency, whereas wavelets are localized in both time and frequency [39].

In Section 5.2.1, three wavelets are compared on their performance for the reconstruction of diffusion-weighted images using cFISTA, namely the Haar wavelet, Daubechies with four vanishing moments (Daub4) and the Cohen-Daubechies-Feauveau wavelet, CDF 9/7. The wavelet functions and their scaling functions are displayed in Figure 3.6.
The Haar wavelet is a special case of the Daubechies wavelet and has one vanishing moment (Daub1). The Daubechies wavelet family consists of orthogonal wavelets. The more vanishing moments are included in the wavelet, the better piecewise smooth parts of the function f can be approximated.
In some applications, it is more convenient to use a non-orthogonal wavelet transform. An example of such bi-orthogonal wavelet is the CDF 9/7 wavelet, which is used for JPEG 2000 compression and fingerprint compression. Since CDF 9/7 is non-orthogonal, two pairs of scaling and wavelet functions are required: the decomposition and reconstruction functions. These functions have lengths 9 and 7. The CDF 9/7 wavelet is symmetric (Figure 3.6c-3.6d), which has an advantage over asymmetric wavelets with respect to border extension. Symmetric wavelets give a symmetric extension of the boundaries, which will not introduce distortions at the borders [40].

Chapter 4

Image quality measure

In order to determine the quality of the reconstructed images, there is need for a quality measure. In this chapter, such quality measure is defined and discussed. Two signals or images are compared and the quality measure provides a quantitative score that describes the level of similarity between them. For the comparison of images, only real-valued images are used.

4.1 Mean Squared Error

At first glance, the Mean Squared Error (MSE) could be used to determine the quality of the reconstructed image $x \in \mathbb{R}^{m \times n}$ compared to the original image $y \in \mathbb{R}^{m \times n}$. The MSE is defined as

$$\text{MSE}(x, y) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij} - x_{ij})^2, \quad x, y \in \mathbb{R}^{m \times n},$$

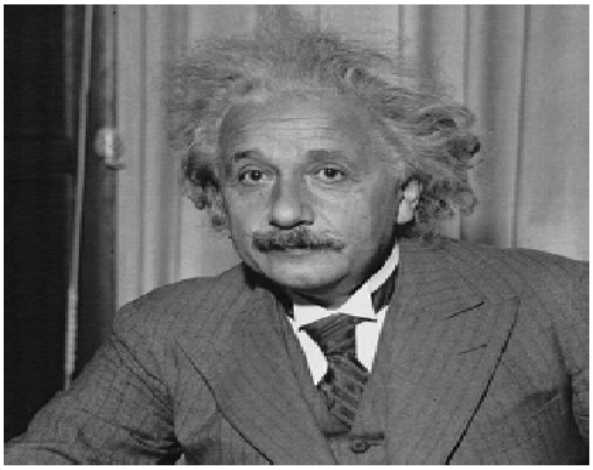
which is the square of the weighted l_2-norm.

MSE is often used because it has nice properties:

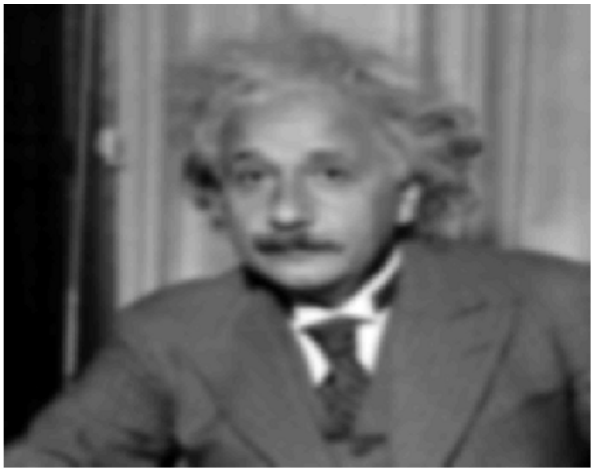
- it is a simple measure, because it is parameter free,
- it is inexpensive to compute,
- the squared error is evaluated at each pixel independent of other pixels,
- it satisfies the nonnegativity, identity and symmetry conditions of the Euclidean metric in $\mathbb{R}^n$.

Because of all the nice properties of MSE, this measure is applied to compare modified images to the original image.

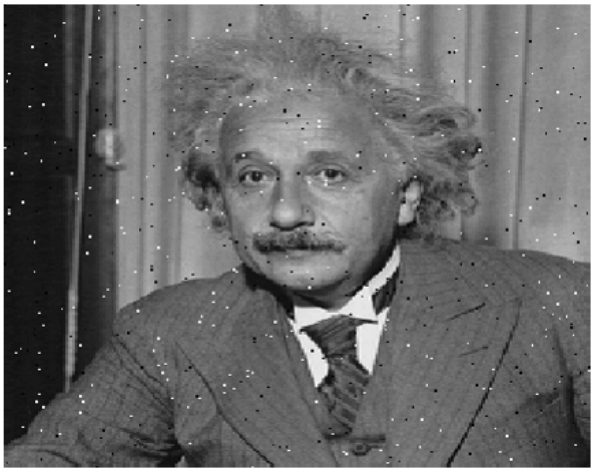
To illustrate the MSE measure, an image of Einstein, altered with different types of distortions such as impulsive noise contamination, blurring and rotation, is used (Figure 4.1). The MSE was computed for each of these resulting images. The MSE for all three images was the same (Figure 4.1b-4.1d). However, these images do not have the same quality from a human visual perspective.



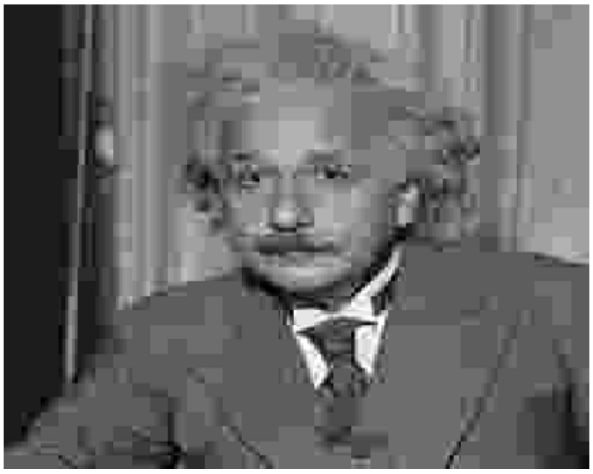
(a) MSE = 0, SSIM = 1



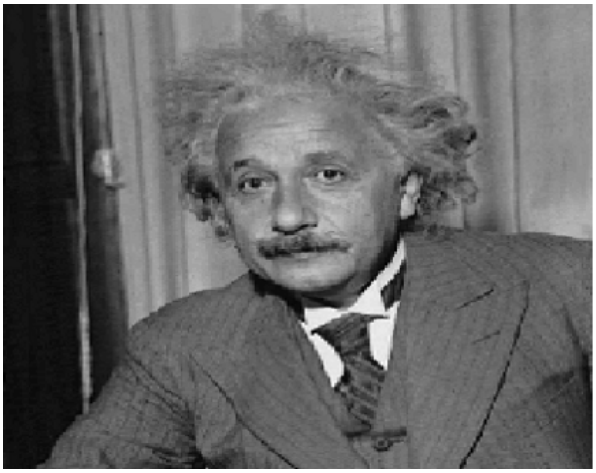
(b) MSE = 0.0022, SSIM = 0.6940



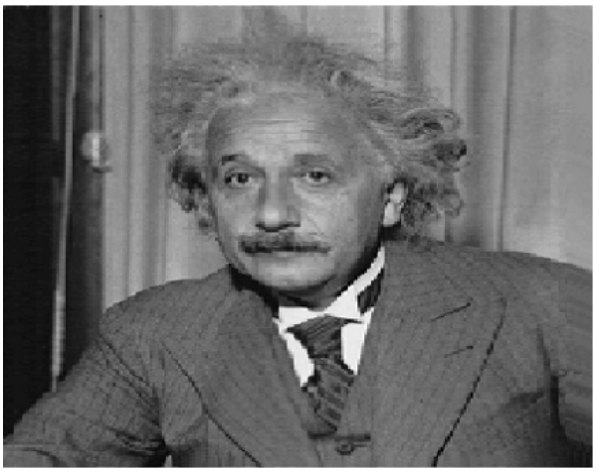
(c) MSE = 0.0022, SSIM = 0.8396



(d) MSE = 0.0022, SSIM = 0.6624



(e) MSE = 0.0135, SSIM = 0.3654



(f) MSE = 0.0168, SSIM = 0.3047

Figure 4.1: Comparison of image fidelity measures for an image of Einstein altered with different types of distortions.
4.1a Reference image. 4.1b Blurring. 4.1c Impulsive noise contamination. 4.1d Recovering from JPEG compression. 4.1e Rotation 5° counter-clockwise. 4.1f Spatial shift to the left.

4.2 SSIM

To be more in line with the human visual perspective, other quality measures have been proposed. One quality measure is called the Structural Similarity Image Measure (SSIM) [41, 42]. In this section, this measure will be explained.

An image $x = \{(i,j) \mid i = 1,\dots,m,\ j = 1,\dots,n\}$ is covered by blocks k of equal size. The image x restricted to such a block is called a *patch*, denoted by x_k. The SSIM computes a so-called SSIM index for each block k in the form of a sliding window that moves pixel-by-pixel across the image. This means that the blocks will overlap. The SSIM index of a patch is stored in the center pixel of the block and the result is a so-called SSIM map.

The SSIM index consists of three terms: local luminance ($\tilde{l}$), local contrast ($\tilde{c}$) and local structures ($\tilde{s}$). The SSIM index is computed in [41, 42] by the formula

$$\begin{aligned} S(x_k, y_k) &= \tilde{l}(x_k, y_k) \cdot \tilde{c}(x_k, y_k) \cdot \tilde{s}(x_k, y_k) \\ &= \left(\frac{2\mu_{x_k}\mu_{y_k} + C_1}{\mu_{x_k}^2 + \mu_{y_k}^2 + C_1}\right)^{\alpha} \cdot \left(\frac{2\sigma_{x_k}\sigma_{y_k} + C_2}{\sigma_{x_k}^2 + \sigma_{y_k}^2 + C_2}\right)^{\beta} \cdot \left(\frac{2\sigma_{x_k y_k} + C_3}{\sigma_{x_k}\sigma_{y_k} + C_3}\right)^{\gamma}. \end{aligned} \tag{4.1}$$

Here, $\mu_{x_k} := \frac{1}{|k|}\sum_{(i,j)\in k} x_{ij}$ is the mean value of the patch k of x with $|k|$ the number of points in the patch. $\sigma_{x_k} := \sqrt{\frac{1}{|k|-1}\sum_{(i,j)\in k} x_{ij} - \mu_{x_k}}$ is the standard deviation of the patch and $\sigma_{x_k y_k} := \frac{1}{|k|-1}\sum_{(i,j)\in k}(x_{ij} - \mu_{x_k})(y_{ij} - \mu_{y_k})$ the correlation coefficient.

Note that if two image patches are the same, then each term is equal to one and thus the SSIM index is also equal to one.
To get more insight into this quality measure, each term of (4.1) will be studied.

4.2.1 Luminance

The first term is the similarity of the luminance of the patches, denoted by $\tilde{l}(x_k, y_k)$. The expression is rewritten to gain more insight in this term.

$$\begin{aligned} \tilde{l}(x_k, y_k) &= \left(\frac{2\mu_{x_k}\mu_{y_k} + C_1}{\mu_{x_k}^2 + \mu_{y_k}^2 + C_1}\right)^{\alpha} \\ &= \left(1 - \frac{(\mu_{x_k} - \mu_{y_k})^2}{\mu_{x_k}^2 + \mu_{y_k}^2 + C_1}\right)^{\alpha} \end{aligned} \tag{4.2}$$

According to (4.2), this term looks at the differences in the signal intensity by taking the difference between local mean samples μ_{x_k} and μ_{y_k} of the images and squaring the result. This is scaled by the mean samples and a constant C_1. The weighting and constants are discussed in Section 4.2.4.

4.2.2 Contrast

The second element is the similarity of the contrast of the patches, denoted by $\tilde{c}(x_k, y_k)$. Rewriting the expression stated in [41, 42] yields:

$$\begin{aligned} \tilde{c}(x_k, y_k) &= \left(\frac{2\sigma_{x_k}\sigma_{y_k} + C_2}{\sigma_{x_k}^2 + \sigma_{y_k}^2 + C_2}\right)^{\beta} \\ &= \left(1 - \frac{(\sigma_{x_k} - \sigma_{y_k})^2}{\sigma_{x_k}^2 + \sigma_{y_k}^2 + C_2}\right)^{\beta} \end{aligned} \tag{4.3}$$

From (4.3) can be concluded that the similarity in contrast can be computed by taking the difference between the local standard deviations σ_{x_k} and σ_{y_k} and squaring the result.

4.2.3 Structure

The third element is the similarity of the structures in the patches, indicated by $\tilde{s}(x_k, y_k)$. The unit vectors $\frac{x_k - \mu_{x_k}}{\sigma_{x_k}}$ and $\frac{y_k - \mu_{y_k}}{\sigma_{y_k}}$ are associated with structures in the patches [41, 42]. By computing the correlation coefficient between the vectors $x_k - \mu_{lpx}$ and $y_k - \mu_{lpy}$, Wang *et al.* find a measure for structural similarity. The linear correlation between the vectors describes the direction and magnitude of the linear relationship between them. The linear correlation coefficient is equal to one in the case of perfect linear correlation between the vectors.

Rewriting this term leads to:

$$\begin{aligned} \tilde{s}(x_k, y_k) &= \left(\frac{\sigma_{x_k y_k} + C_3}{\sigma_{x_k}\sigma_{y_k} + C_3} \right)^{\gamma} \\ &= \left(1 - \frac{\sigma_{x_k}\sigma_{y_k} - \sigma_{x_k y_k}}{\sigma_{x_k}\sigma_{y_k} + C_3} \right)^{\gamma}. \end{aligned} \tag{4.4}$$

By subtracting the correlation coefficient from the product of the standard deviations σ_{x_k} and σ_{y_k}, an estimate for the structure similarity is defined.
To get an even better understanding of the structural similarity term , define $v_i := x_i - \mu_x$ and $w_i := y_i - \mu_y$. Then rewriting the numerator of (4.4) (Appendix A.4):

$$\sigma_{x_k}\sigma_{y_k} - \sigma_{x_k y_k} = \frac{1}{N-1}(\|v\|\|w\| - \|vw\|) \geq 0, \quad \forall v, w. \tag{4.5}$$

The inequality in (4.5) is derived using Cauchy-Schwarz and is an equality if and only v and w are linear dependent. This last statement indicates that the structural similarity term indicates equality between the patches, even when there exist some scaling factor between them. This scaling factor is evaluated in the luminance and contrast terms. Hence, the structure similarity term measures the degree of linear correlation between the patches.

4.2.4 Weighting

The constants C_1, C_2 and C_3 are small positive constants that ensure stability of the three terms in the case the denominators are very close to zero.

C_1 is defined as $C_1 = (K_1\Lambda)^2$ [41, 42]. It is based on the dynamic range of pixel values Λ (i.e. the maximum possible pixel intensity) and another constant $K_1 \ll 1$. In [42], $K_1 = 0.01$.
The constant C_2 is defined similar to C_1, namely $C_2 = (K_2\Lambda)^2$. $K_2 = 0.03$ is chosen in [42].
C_3 is taken as $C_3 = \frac{1}{2}C_2$ [41, 42].

The three terms in (4.1) are given a certain weight using the constants α, β and γ ($0 \leq \alpha, \beta, \gamma \leq 1$). If, for example, the comparability in contrast is more important than in luminance or structures, then β can be chosen larger than α and γ.
When images are compared in Chapter 5, the three terms are taken equally important in the calculation of the SSIM index: $\alpha = \beta = \gamma = 1$.

In Chapter 5, there will also be comparison between images performed where only the structures are of importance. For this situation, the SSIM index with $\alpha = \beta = 0$ and $\gamma = 1$ is used. This situation will be referred to as computing the *mean structures score*. Similar, it holds that *mean contrast score* and *mean luminance score* have the (α, β, γ)-combination of $(0, 1, 0)$ and $(1, 0, 0)$, respectively.

4.2.5 Implementation

On the images, the SSIM method is applied to the local patches using a sliding window approach. The universal image quality index uses square blocks (mentioned above) of size 8×8 as window [43], but this often results in blocking artifacts. As an alternative, an 11×11 normalized, circular-symmetric Gaussian weighting function $w = \{w_{ij} | i = 1, \ldots, 11, j = 1, \ldots, 11\}$ with standard deviation of 1.5 samples is used to create blocks of size 11×11 [41] (Figure 4.2).

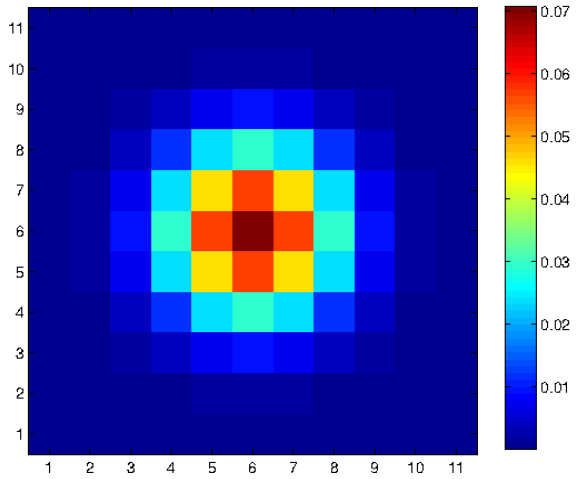


Figure 4.2: An 11×11 normalized, circular-symmetric Gaussian weighting function to create blocks used for the SSIM index computation.

Using this weighting function, the estimates of the local statistics μ_{x_k}, σ_{x_k} and $\sigma_{x_k y_k}$ for block or patch k are now defined as

$$\mu_{x_k} := \frac{1}{|k|} \sum_{(i,j) \in k} w_{ij} x_{ij},$$

$$\sigma_{x_k} := \sqrt{\frac{1}{|k| - 1} \sum_{(i,j) \in k} w_{ij} (x_{ij} - \mu_{x_k})},$$

$$\sigma_{x_k y_k} := \frac{1}{|k| - 1} \sum_{(i,j) \in k} w_{ij} (x_{ij} - \mu_{x_k})(y_{ij} - \mu_{y_k}).$$

To obtain a single overall quality measure of the image, the Mean Structural Similarity Image Measure (mSSIM) index is computed by averaging the SSIM map:

$$\text{mSSIM}(y, x) = \frac{1}{H} \sum_{h=1}^{H} \text{SSIM}_h(y_k, x_k). \tag{4.6}$$

where H is the total number of local patches in an image.

The mSSIM values of the altered images of Einstein are computed (Figure 4.1). According to these values, there exists a quality difference between the altered images (Figure 4.1b-4.1d), in contrast to the results when using MSE as quality measure. This difference corresponds more with the human visual perception. For example, all structures and details in the with noise contaminated image are as visible as in the original image (Figure 4.1c). For the recovering from the JPEG compression holds that the big structures are visible, but the details are not, e.g. the stripes on the suit and tie of Einstein (Figure 4.1d). Therefore, most people would give a higher quality score to the with noise contaminated image (Figure 4.1c) than to the JPEG compression recovery (Figure 4.1d). This scoring is similar to the quality scores obtained by using the SSIM, which suggests that the SSIM corresponds better with the human visual perception.

SSIM index is sensitive to relative translations, scalings and rotations. The mSSIM values computed for a rotated image and a spatially shifted image were low (Figure 4.1e-4.1f), indicating that a change in structures gets a higher penalty. The Complex Wavelet Structural Similarity Image Measure (CW-SSIM) is developed in [42] to be less sensitive in these situations.
However, the sensitivity to translations, rotations and scalings is an advantage. The images should display an exact image of the scanned slice in the body to be useful for radiotherapy. It is not desired that pixels are spatial shifted or shifted by rotation or scaling after a reconstruction using CS. The used quality measure should indeed penalize these transformations in order to test whether reconstruction using CS leads to the same images as the images obtained without undersampling. Hence, the basic SSIM will be used to compare the reconstructed images with reference images.

4.2.6 Volume

The mSSIM is also computed to compare two stacks of images, also referred to as volume, with each other. The total number of local patches will consists of H (i.e. the number of patches in each image) times the number of images K. Then the mSSIM is defined with

$$\begin{aligned} \text{mSSIM}(y, x) &= \frac{1}{H \cdot K} \sum_{k=1}^{K} \sum_{h=1}^{H} \text{SSIM}_{h,k}(y_k, x_k), \\ &= \frac{1}{K} \sum_{k=1}^{K} \left(\frac{1}{H} \sum_{h=1}^{H} \text{SSIM}_{h,k}(y_k, x_k) \right), \\ &= \frac{1}{K} \sum_{k=1}^{K} \text{mSSIM}_{\text{slice } k}. \end{aligned} \tag{4.7}$$

According to (4.7), the mSSIM of a volume can be defined by taking the average of the mSSIM values per image in the stack.

4.2.7 Reference image

To measure the quality of the reconstructed images, these images will be compared with a reference image. When performing experiments retrospectively (i.e. an image is Fourier transformed to k-space and then undersampled), this reference image is known: it is the processed (diffusion-weighted) image. The reconstructed image should be close to the original image and thus the SSIM can be used.

However, the reference scan is not known when the undersampling is performed prospectively (i.e. the raw, unprocessed data is undersampled instead of the processed data). It is not possible to obtain both the undersampled and fully sampled k-space in one scan. Another scan could be made to acquire the full k-space. Since undersampling will theoretically minimize geometric distortions (Chapter 2.4), the image reconstructed from undersampled k-space will differ from the image reconstructed from the fully sampled k-space. The MR data in Section 5.3.4 is not diffusion weighted and aquired with a gradient echo instead of EPI. This type of imaging will not lead to geometric distortions. However, the undersampled and fully sampled raw data differ due to noise and other influences during acquisition. Comparing the reconstruction with the fully sampled reference scan could give some indication of the quality of the reconstruction, but the actual quality cannot be computed.
However, the SSIM can be used to compare the difference in quality between two reconstructions with different undersampling strategies. The mSSIM can be calculated for both reconstructions with the fully sampled reference scan. By comparing the results, an order of the quality of the undersampling strategies could be obtained.

Hence, the SSIM is useful for comparing retrospectively undersampled reconstructions with their reference images, but it gives not the actual quality for prospective experiments. The SSIM could be used to compare undersampling strategies used in prospective experiments. Since the focus of this thesis is on the feasibility of the method, the undersampling is done retrospectively and thus the SSIM gives a indication of the reconstruction quality of the method.

4.3 Power spectral density

Another comparison method that will be used here, is related to k-space. Comparing the coverage of the power spectral densities included in the undersampling gives more insight in the expected quality difference between two reconstructed images. This method will be used in Section 5.3.1.

The power spectral densities [44, 45] are computed using

$$\sum_{i,j}^{m,n} |k_{i,j}^s|^2,$$

where $k_{i,j}$ is the value at position (i, j) in the $m \times n$-image in k-space and s is the number of the used slice. Then the so-called power spectral densities of the k-spaces, undersampled with the masks, are computed using

$$\sum_{i,j}^{m,n} |\text{mask}_h k_{i,j}^s|^2,$$

with h denoting the number of the used mask. To calculated the percentage of the power spectrum taken with undersampling, the latter densities were divided by the power spectral densities of the original slices, that is

$$\frac{\sum_{i,j}^{m,n} |\text{mask}_h k_{i,j}^s|^2}{\sum_{i,j}^{m,n} |k_{i,j}^s|^2}. \tag{4.8}$$

Chapter 5

Results

In this chapter, CS experiments are described for patient DWI data. First, several strategies are formulated to undersample the DWI data in Section 5.1. In Section 5.2, optimal settings of the regularization parameters α and β and sparsifying transform Ψ in the minimization problem (P) (Equation (3.13)) are sought.
After all parameters and transforms were chosen, several experiments were performed on patient data to find an optimal undersampling strategy (Section 5.3). The undersampling strategies were applied on the patient data for several undersampling percentages. The reconstructions were compared using the SSIM metric to find an optimal undersampling strategy and percentage for this type of data.

5.1 Undersampling masks

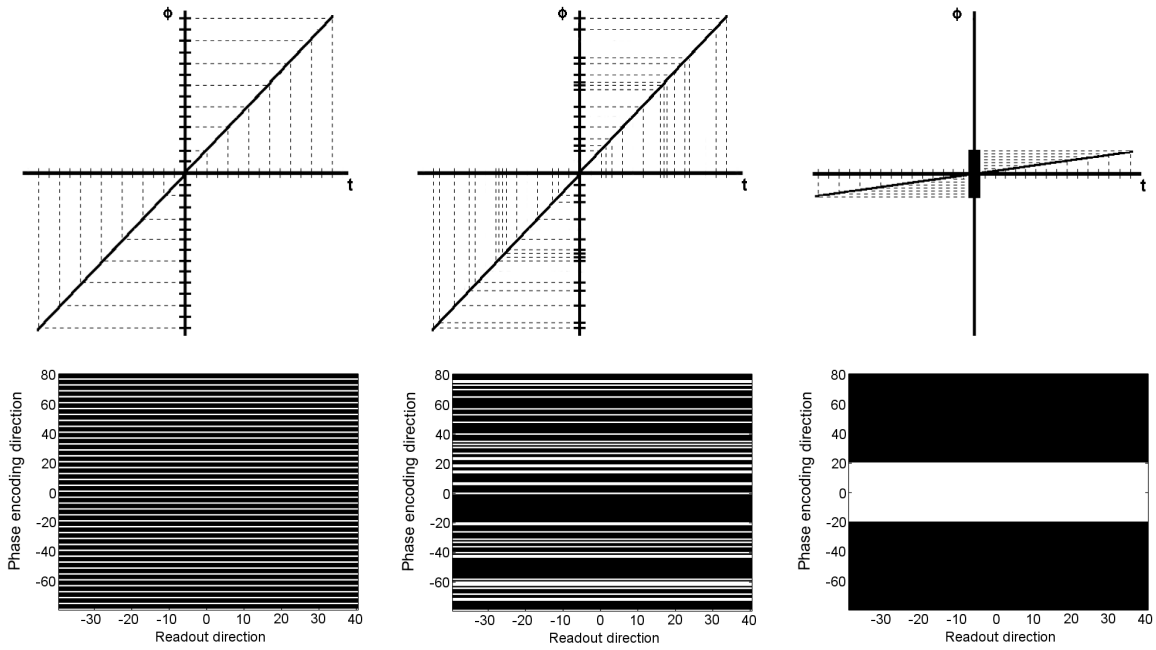
DWI scans are acquired with an EPI trajectory. For this type of trajectory, it is more time efficient to reduce the number of phase encoding steps instead of the number of readout steps. This increased the bandwidth in the phase encoding direction and thus reduced geometrical errors (Section 2.4.1).

In Section 5.3.1, the efficiency of five undersampling strategies is investigated (Figure 5.1). Those strategies were defined in this section. Each strategy took $u_{\text{perc}}\%$ of the full k-space, which led to the selection of $p \in \mathbb{N}$ lines:

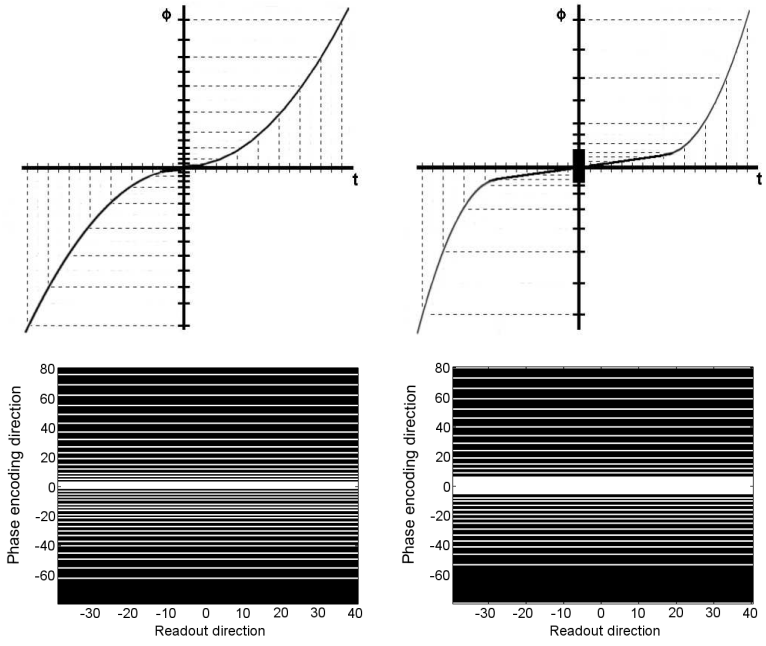
$$p = n * \frac{u_{\text{perc}}}{100},$$

with n the total number of phase encoding steps and u_{perc} the percentage of the data used in the undersampling. The density function $\phi(t)$ described each strategy by selecting the p lines using a uniform distribution of t (Figure 5.1). A random distribution of t was used for the 'random' strategy.

The 'uniform' strategy did not satisfy the conditions of CS (Section 3.1.1), but was included to provide information on the performance of the other undersampling strategies. The strategy 'random' satisfied the conditions of CS, but probably failed to capture the majority of the information in k-space. This information is located in the center of k-space (Section 2.2). The strategies 'center', 'increased' and 'centerincreased' were designed to select more lines in the center of k-space. These strategies corresponded better with the distribution of information in k-space and thus with the distribution of information in the measurement vector b than the 'random' (Section 3.2.1).



(a) uniform (b) random (c) center



(d) increased (e) centerincreased

Figure 5.1: The density plots and the corresponding masks of the five undersampling strategies used in the experiments. A selection of p lines was obtained by using a uniform distribution of t in the density functions ϕ (top figures). An exception was the 'random' strategy, where t is randomly distributed. These distributions were used to determine the lines in the masks. Each of the shown strategies displays the mask used to obtain 25% of the data. The undersampling took place in the phase encoding direction (y-axis). The readout direction was displayed on the x-axis.

Uniform

The first undersampling strategy was the 'uniform' undersampling (Figure 5.1a), which was already discussed in Section 3.1.1. The other strategies should perform at least as well as 'uniform', since uniform undersampling does not lead to aperiodic aliasing. This type of aliasing is required for CS (Section 3.1.2).

Random

The second strategy for undersampling was called 'random' (Figure 5.1b). This strategy was based on the requirements in the CS theory that the measurements should be randomly obtained. For the strategy, p lines in k-space were chosen at random using the function RANDPERM(m, p) of Matlab [21]. An extra condition for this method was that the center line $k = 0$ of k-space was always included. This extra condition made the 'random' strategy more comparable with the other strategies. The highest frequency amplitudes in k-space are located in the center and thus are captured by $k = 0$. The inclusion of $k = 0$ provide information about contrast and large structures of the image (Figure 2.5 in Section 2.2).

Center

The method 'center' was designed as an extreme example of selecting lines in the center of k-space (Figure 5.1c). In this undersampling method, p consecutive lines in the center of k-space were chosen to obtain $u_{\text{perc}}\%$ of the image in k-space. This strategy captured most information about contrast and structures of the image, even when a small number p was chosen (recall Figure 2.5).

Increased

A disadvantage of 'center' was that the periphery of k-space was not sampled (unless a high percentage of the data is included). This meant that no information about details was captured (Figure 2.5). The undersampling strategy 'increased' was designed to include this information (Figure 5.1d). The distance between the consecutive lines, viewed from the center of k-space, was increased. This undersampling strategy used the ideas of variable density, where there was a higher sampling density in the center of k-space and lower density at the borders [46]. Since both the center and the outer regions of k-space were sampled, information about structures and contrast as well as about details was captured. The lines in k-space ranged from $k_{\min} := -\frac{1}{2}n_2+1$ to $k_{\max} := \frac{1}{2}n_2$. The set of the p chosen lines was obtained using

$$k_{\text{sampled}} = \begin{cases} k_{\min} + \left\lfloor \sqrt{|k|} \cdot \frac{k_{\max}}{\sqrt{\frac{1}{2}p}} \right\rceil, & -\frac{1}{2}p \leq k \leq -1, \\ \left\lfloor k^2 \cdot \frac{k_{\max}}{\left(\frac{1}{2}p\right)^2} \right\rceil, & 0 \leq k \leq \frac{1}{2}p. \end{cases}$$

The sampling patterns for the two halves of k-space were not symmetric. This was done deliberately to create a more random-like sampling pattern of the whole k-space.

A disadvantage of this strategy was that the sampling density in the center of k-space was low for small percentages, because the p lines were spread over k-space.

Centerincreased

The last strategy used for undersampling was called 'centerincreased' (Figure 5.1e) and it combined the advantages of 'center' (information about structures for small percentages) and 'increased' (information about details for small percentages). Furthermore, ideas used with keyhole imaging, were at the base of this strategy.
In keyhole imaging, k-space was divided in three parts [47]. The central keyhole was located in the center of k-space, which consisted of $\frac{1}{4}$ of k-space. The other two regions were called the SPARSE regions, where the selected lines were interleaved using some SPARSE factor. A SPARSE factor of 3 was used, which indicates that every third line in the outer regions of k-space was selected [47]. In other words, an equal distance between consecutive lines was obtained, similar to 'uniform'.

The mask for the 'centerincreased' strategy was also divided in three parts: lowerpart, centerpart and upperpart (Figure 5.2).

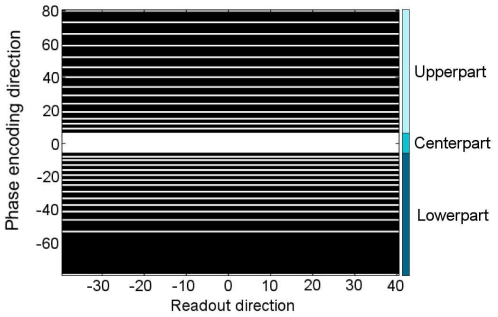


Figure 5.2: The mask of 'centerincreased' consisted of three parts: lowerpart, centerpart and upperpart. In each part, a different number of lines was selected, accordingly to different formulas.

The set k_{sampled} was obtained using

$$k_{\text{sampled}} = \begin{cases} k_{\min} + \left\lfloor \sqrt{|k|} \cdot \frac{\lfloor k_{\max} - \frac{1}{8}p \rceil}{\sqrt{\frac{1}{2}p}} \right\rceil, & -\frac{1}{2}p \leq k < -\frac{1}{8}p, \\ k, & -\frac{1}{8}p \leq k \leq \frac{1}{8}p, \\ \left\lfloor k^{1.5} \cdot \frac{k_{\max}}{(\frac{1}{2}p)^{1.5}} \right\rceil, & \frac{1}{8}p < k \leq \frac{1}{2}p. \end{cases}$$

When $\{k_{\min}, k_{\max}\} \notin k_{\text{sampled}}$, the lowest and highest numbers in k_{sampled} were replaced by $k_{\min}$ and $k_{\max}$. This ensured the given resolution.
In the centerpart, $\frac{1}{4}$ of the p lines were chosen. This part was inspired by the 'center' strategy and the central keyhole in [47].
The lowerpart contained $\frac{3}{8}$ of the p lines of k-space. Instead of taking every third line of k-space in this part as in [47], the lines were sampled in a more random pattern. The lines were chosen using a square root function, so the distance between the successive lines increased, viewed from the center of the mask. This differed from the SPARSE regions in [47], but it was similar to the 'increased' strategy. The upperpart of mask consisted of the last $\frac{3}{8}$ of the wanted k-space lines. Again, the distance between the lines increased when the lines were closer to the border of k-space. These lines were chosen according to a function of the power 1.5. The value of the power could be increased, leading to more distance between lines near the border. The value 1.5 was used here to ensure that a part of the high spatial frequencies in k-space was sampled.

5.2 Choice parameters and transforms

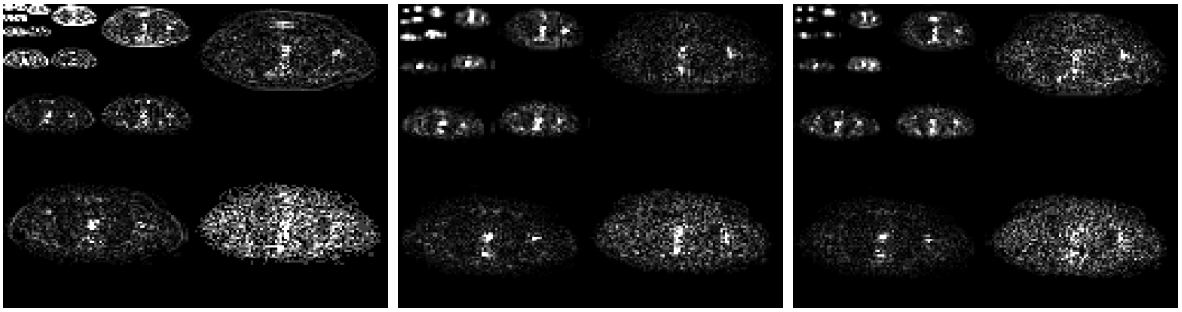
To test the feasibility of the method for diffusion-weighted images, cFISTA was applied retrospectively on ten patient data sets, containing diffusion data with b-values $b = 0\,\mathrm{s/mm^2}, b = 200\,\mathrm{s/mm^2}$ and $b = 800\,\mathrm{s/mm^2}$ (Section 2.3.3). Each patient set contained 50 slices of size 160×160. From the 50 slices, a subset of ten slices from each patient set was used for the experiments. The slices were located in the center of the volume, since the tumor was often located there. The subsets from each of the ten patients were referred to as the 'test slices' of a patient.

Before experiments aimed at finding the optimal strategy could be executed, the parameters and sparsity transforms of the problem (P) (Section 3.2.2) were chosen. First, three wavelet transforms were compared to find the sparsity transform Ψ. Then the regularization parameters α and β of the problem (P) were optimized for the DWI data on which undersampling experiments were executed. Finally, some implementation remarks were made.

5.2.1 Wavelets

In Section 3.3.8, three wavelets were discussed. In this section, the performance of these wavelets was tested on DWI data of the patient set to find the best sparsifying transform among these three for the experiments executed in Section 5.3.

To illustrate the Haar wavelet, Daub4 and CDF 9/7, the wavelet transforms were applied to an image to obtain their decompositions (Figure 5.3).

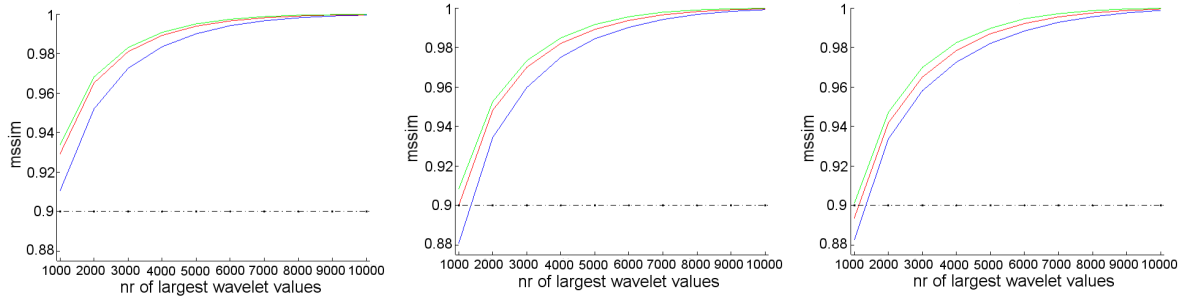


(a) Haar (b) Daubechies (c) CDF 9/7

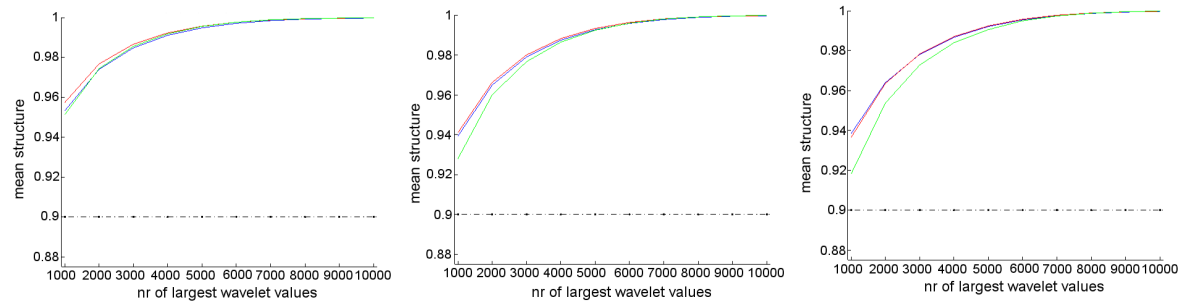
Figure 5.3: Wavelet decompositions of slice 24 of patient 12 ($b = 800\,\mathrm{s/mm^2}$) by the Haar wavelet (a), Daub4 (b) and CDF 9/7 (c). The data is reordered in such a way that the highest frequencies are at the border and the lowest frequencies in the top left corner.

To investigate the sparsifying abilities of the wavelets, the wavelets were applied to an image of each b-value of the ten patients. In the wavelet domain, subsets (of sizes ranging from 1000 to 10000) of the largest wavelet values were selected and the rest of the total number of 20480 values was set to zero. Then, the inverse wavelet transform was applied and the resulting images were compared to the original images by computing the mSSIM and the mean structure score (Chapter 4). The mSSIM measured the quality of the whole reconstructed image and the mean structure score measured the quality of the reconstructed structures (Section 4.2). It is important that the structures were reconstructed well, because structures are used in the process of locating the tumor.

The mSSIM and the mean structure scores were averaged over the ten images of the patients for each b-value (Figure 5.4).



(a) $b = 0\,\mathrm{s/mm^2}$ (b) $b = 200\,\mathrm{s/mm^2}$ (c) $b = 800\,\mathrm{s/mm^2}$

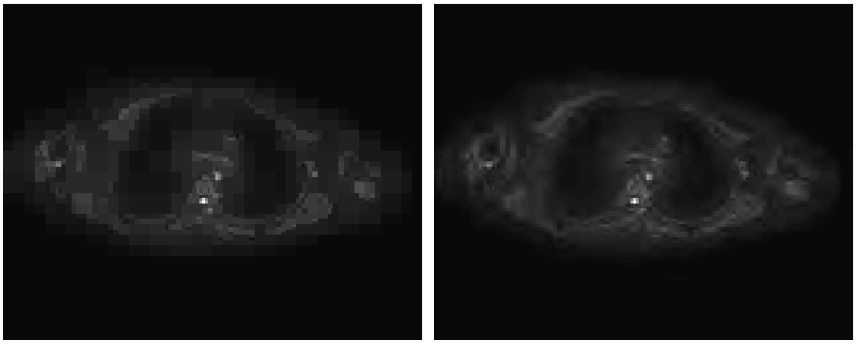


(d) $b = 0\,\mathrm{s/mm^2}$ (e) $b = 200\,\mathrm{s/mm^2}$ (f) $b = 800\,\mathrm{s/mm^2}$

Figure 5.4: The average mSSIM values (a-c) and the mean structure scores (d-f) of the thresholded images for the Haar wavelet (green), Daub4 (red) and CDF 9/7 (blue), using $b = 0\,\mathrm{s/mm^2}$, $b = 200\,\mathrm{s/mm^2}$ and $b = 800\,\mathrm{s/mm^2}$. After applying the wavelets, the images of ten patients were thresholded by selecting several subset sizes of only the largest wavelet values. The mSSIM and mean structure scores were compared with 0.9 (black), the value that indicated a good reconstruction.

On average, all images still had a high mSSIM value using only 9.8% or more of values in the wavelet domain. For the Haar and Daubechies wavelet, there was a good reconstruction (mSSIM$\geq$ 0.9) even when 4.9% of the data was selected. The Haar wavelet was slightly better at preserving the information of the image than the Daubechies wavelet. However, the mean structure score of the Haar wavelet was lower than for the other two wavelets (Figure 5.4d-5.4f). This was especially the case for $b = 200\,\mathrm{s/mm^2}$ and $b = 800\,\mathrm{s/mm^2}$. The Haar wavelet could lead to a coarse, boxlike appearance in the reconstruction, because it had only one vanishing moment (Figure 5.5a).
This implies that the approximations for the piecewise smooth parts of the function were performed using a constant function (Section 3.3.8). If the image resolution is high, then the degree of smoothness becomes large and thus the coarse, boxlike appearance of the Haar wavelet is more visible [48]. In future, the resolution of the test images may be increased and thus it is useful to choose a wavelet that is capable of handling this situation. It was preferred to have better approximations of the function for wavelet-based compression [49]. Daub4 used four vanishing moments and thus led to a better approximation. It allowed a better support width of polynomials on which the degree of smoothness was based [48]. Therefore, the reconstructed image was more smooth and the structures were better visible (Figure 5.5b).

The Haar wavelet and Daub4 were unitary. This meant that only Ψ was used in the calculation of the $\mathrm{prox}_L(\|\cdot\|_1)(y)$ (Section 3.3.5). The CDF 9/7 wavelet was not unitary and thus the calculation of $\mathrm{prox}_L(\|\cdot\|_1)(y)$ is more difficult due to the additional use of Ψ^{-1} next to Ψ.



(a) Haar wavelet (b) Daub4 wavelet

Figure 5.5: An image was reconstructed using the Haar wavelet (a) and Daub4 (b). The Haar wavelet reconstruction had a coarse, boxlike appearance, in which the structures were not well visible. In contrast, the Daub4 reconstruction resulted in a smooth image in which the structures were well visible.

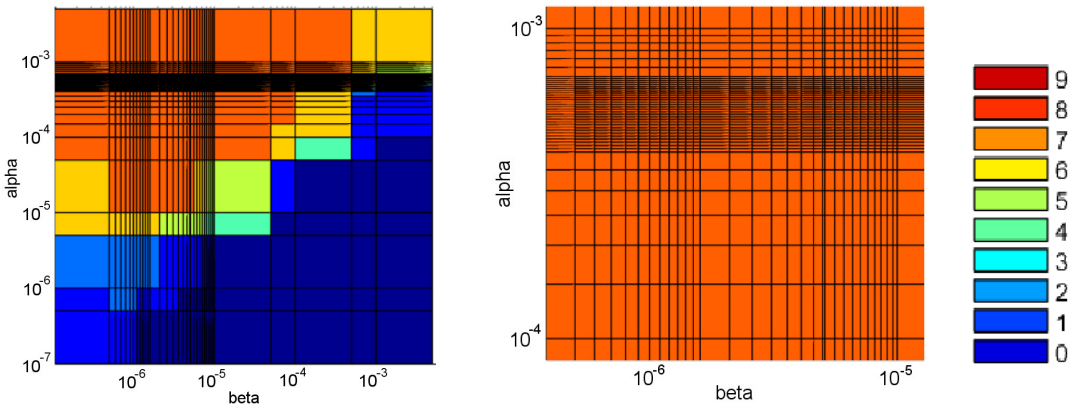
Daub4 resulted in high mSSIM values as well as high mean structure scores. It was unitary and led to smooth approximations. Hence, this wavelet was chosen to be the sparsity transform in the experiments to come.

5.2.2 Optimal regularization parameters

In Section 3.2.2, the parameter α and β were introduced into the problem (P) in Equation (3.13). As was mentioned there, the values of these parameters are found experimentally, since the value of ε was not known.

To find the optimal α and β, a training set was created by taking one slice of nine patients. The slice of the tenth patient was used as a test set. The aim was to find the α and β where the most slices have a mSSIM greater than 0.9, i.e. a good reconstruction.

The nine diffusion-weighted images ($b = 200\,\mathrm{s/mm^2}$) were undersampled to obtain 25% of the data. For hundred (α, β)-combinations, the images were undersampled using 'center', 'increased' or 'centerincreased' and then reconstructed. The reconstructed images were compared to the original image by computing the mSSIM. The number of images with a mSSIM ≥ 0.9 was counted (Figure 5.6 and Figure A.1 in Appendix A.5).



(a) 'centerincreased' (b) zoom of (a)

Figure 5.6: The number of images with a mSSIM ≥ 0.9, displayed as a contour plot. A fine scale of the α and β values was concentrated in the center of the area with the highest number of image with mSSIM ≥ 0.9. An enlargement of this area is shown in (b).

There were several combinations of α and β possible which led to the same number of images with mSSIM ≥ 0.9 in the sampled data set. Using 'centerincreased' and the values of α and β listed in Appendix A.5, the optimal area was refined to find the optimal combination of α and β. This strategy was chosen, because it would be the best undersampling strategy from a theoretical point of view (Section 5.1).
The median of the optimal combinations in the refined area was $(\alpha, \beta) = (5 \cdot 10^{-4}, 1 \cdot 10^{-6})$. This combination was also located in the optimal area of 'center' and 'increased' (Figure A.1). The optimal combination was tested on the test slice of the tenth patient, again using 'centerincreased' and 25% of the data (Figure 5.7).

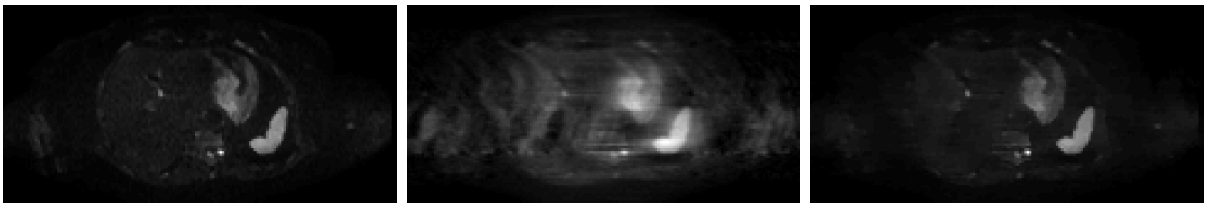


Figure 5.7: The original image of the test patient (left), the naive reconstruction using the undersampling strategy 'centerincreased' with 25% of the data (middle) and the reconstruction using cFISTA with the optimal (α, β)-combination (right). The corresponding mSSIM of the reconstruction is 0.9100.

The mSSIM of the test slice was equal to 0.9100, which indicated a very good reconstruction. The mSSIM calculations of the other b-values and strategies are stated in Table A.1 in Appendix A.5. The average mSSIM value over these nine combinations is equal to 0.9019, so good reconstructions were obtained for this test patient.
The combination $(\alpha, \beta) = (5 \cdot 10^{-4}, 1 \cdot 10^{-6})$ led to good reconstructions. Therefore, these α and β were used for the next experiments.

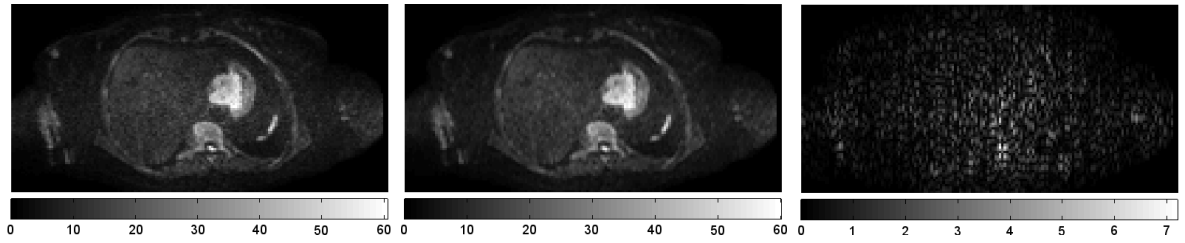
5.2.3 Size

The size of the images (in image space and k-space) should be divisible by 2^w, in order to satisfy the implementation conditions of the wavelet operator. w denoted the maximum decomposition

level of the wavelet. In the implementation of cFISTA, w was chosen to be equal to 5, which led to sizes of the images to be divisible by 32. Note that the images did not have to be square in order to use the implementation of cFISTA.

5.2.4 100%-image

Since cFISTA tried to minimize the TV norm, a denoising effect could be noticed when the algorithm was applied to the original image without undersampling. As an example, this effect was examined for a slice of patient 15 with $b = 200\,{}^{\mathrm{s}}/_{\mathrm{mm}^2}$ and a maximum number of iterations equal to 100 (Figure 5.8).



(a) Original image (b) 100%-image (c) Difference image

Figure 5.8: The original image of a slice of patient 15 (a) and its reconstruction using cFISTA, taking 100% of the data (b). The difference image (c) indicates the suppression of noise in the 100%-image, making this image slightly smoother.

The reconstructed image differed from the original image (Figure 5.8c). The small differences indicated that the reconstructed image contained less noise, making it slightly smoother. The reconstructed image is denoted with the term 100%-*image*. The comparison between the reconstructions with undersampling and the 100%-image gave a better impression of the reconstruction quality. Since a denoising effect always occurred, a reconstructed image would never be the same as the original image. Hence, all mSSIM computations were performed using the 100%-image.

5.2.5 Stopping criterion

Three stopping criteria for the algorithm were investigated. The first criterion compared the value $F(x_k)$ of the objective function value F using the current solution x_k with $F(x_{k-l})$, the value of l iterations back. If $F(x_k) \geq F(x_{k-l})$, then the algorithm was terminated. This criterion is referred to as *obj*.
The second criterion compared the slope between $(k-l, F(x_{k-l}))$ and $(k, F(x_k))$ with a small number ϵ (e.g., ϵ was taken $1 \cdot 10^{-8}$). When the slope was smaller than ϵ, then the algorithm was terminated. This criterion is referred to as *slope*.
The last criterion used a maximum number of iterations as a stopping criterion and is referred to as *max_it*. When this maximum number was reached, the algorithm stopped.

The first two criteria did not lead to satisfactory results of terminating cFISTA: the algorithm was terminated after more iterations than that were required to obtain a mSSIM above 0.9. In some cases, the number of iterations exceed the 5000 iterations. There were cases where the obj-criterion did not terminate the algorithm at all (Figure 5.9). Changing l or ϵ did not satisfactorily solve this problem.

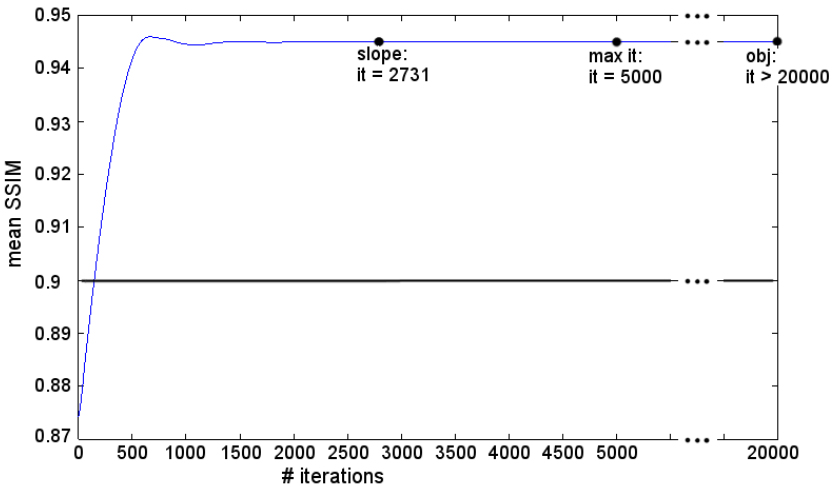


Figure 5.9: The mSSIM was computed for each iteration during the reconstruction of an image of a patient. The number of iterations required for the three stopping criteria before terminating the algorithm are displayed. The criterion based on the comparison of the objective function values is denoted by *obj*. The criterion that uses the slope of the objective function is denoted with *slope* and the third criterion with *max_it*. A good reconstruction is defined as a reconstruction with a mSSIM value above 0.9 (black).

Although the third stopping criterion also led to redundant iterations, it was used here. This criterion led to the same conditions for all experiments. The performance of the strategies with respect to the quality of the reconstructions could be better compared. The maximum number of iterations was set to 5000 iterations.

5.2.6 Proximal operator orders

In Section 3.3.5, it was mentioned that the method to compute a new solution x with the CSD method (Algorithm 4) may not be optimal. So, the performance of several orders of solving the proximal operators was investigated in this section.

Order options

Instead of directly solving

$$\min \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \alpha\|x\|_{TV} + \beta\|\Psi x\|_1 \right\},$$

the problem was split into two parts. $TV \Rightarrow l_1$ denoted the option where first

$$x^{TV} = \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \alpha\|x\|_{TV} \right\}$$

is solved and then used as initial guess for the next problem, where

$$x_{sol} = \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \beta\|\Psi x\|_1 \right\}$$

was solved. The second order of solving is denoted by $l_1 \Rightarrow TV$ and used the same approach as $TV \Rightarrow l_1$, but with the l_1-norm and TV-norm interchanged.

The problem could be directly solved by using the CSD method, where both the proximal operators of the TV-norm and l_1-norm were computed and then combined to get the new value (Section 3.3.5). This order is denoted by csd_{TV+l_1} and it used a Jacobi type of approach.
The fourth and fifth option used a Gauss-Seidel type of approach. The fourth option first computed the proximal operator of the TV-norm using x_g (defined in step 2 of Algorithm 7) giving a solution x_{TV}. Then the solution z_k of the l_1-norm proximal operator was computed using $x_g^{TV} := x_{TV} - \frac{1}{L}\nabla f(x_{TV})$. This order is denoted by $TV \to l_1$. The last option was similar to the fourth option, with the TV-norm and l_1-norm interchanged. This order is denoted by $l_1 \to TV$.

Accuracy orders

The accuracy of the five orders was investigated. Slice 24 of all ten patients was undersampled using the 'centerincreased' mask and 25% of the data and then reconstructed for these orders and the three b-values. The total number of iterations for $TV \Rightarrow l_1$ and $l_1 \Rightarrow TV$ was 10 000 iterations, since both the reconstructions with TV-norm and l_1-norm were performed using 5000 iterations. The other three options used 5000 iterations in total, because the TV-norm and l_1-norm were used in the same reconstruction loop in these options.
The mSSIM values of the reconstructions were computed and averaged (Table 5.1).

Table 5.1: The average mSSIM values obtained using five order combinations of the proximal operators of the TV-norm and l_1-norm for 10 images.

	$b = 0\,\mathrm{s/mm^2}$	$b = 200\,\mathrm{s/mm^2}$	$b = 800\,\mathrm{s/mm^2}$
$l_1 \Rightarrow TV$	0.8946	0.8589	0.8610
$TV \Rightarrow l_1$	0.8909	0.8524	0.8530
$l_1 \to TV$	0.8948	0.8589	0.8610
$TV \to l_1$	0.8947	0.8589	0.8610
csd_{TV+l_1}	0.8947	0.8588	0.8610

The accuracy of $TV \Rightarrow l_1$ was lower than the other options, which had equal results. It should be noted that $l_1 \Rightarrow TV$ used two times more iterations than csd_{TV+l_1}, $TV \to l_1$ and $l_1 \to TV$ to solve the same amount of proximal operators and thus to obtain the same results. Therefore, this order was less favourable. Note that the use of a stopping criterion other than the max_it-criterion could lead to different results. In the current experiment, no significant differences in accuracy between the last three orders was noticed, so each order could be used. For further experiments, the order csd_{TV+l_1} as is stated in Algorithm 7 was used.

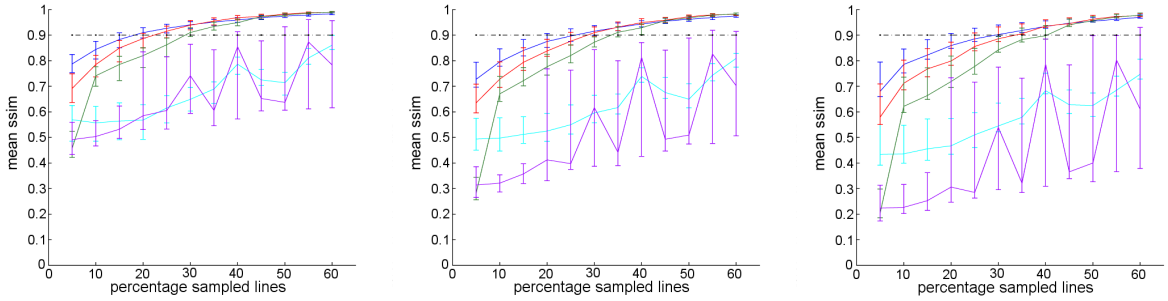
5.3 Experiments on diffusion-weighted images

The five undersampling strategies discussed in Section 5.1 were applied on the test slices of each patient using $5\% - 60\%$ of the image data in k-space, with a step size of 5%. The undersampling masks of a certain percentage were the same for all patients. The undersampled images were then reconstructed using cFISTA, implemented in Matlab.
The goal was to find the best undersampling strategy among the designed strategies in Section 5.1 and the optimal undersampling percentage to get a good reconstruction for each b-value and among all patients.

5.3.1 Comparing patients

In this section, the test sets of patients were compared to each other to investigate the similarities and differences between the patients.

Ten slices of each patient were reconstructed using the undersampling strategies 'center', 'centerincreased', 'increased', 'random' and 'uniform' for twelve undersampling percentages. The corresponding mSSIM values were computed for each reconstruction. In Figure 5.10, the results for the three b-values are shown for patient 13 (as a typical example).



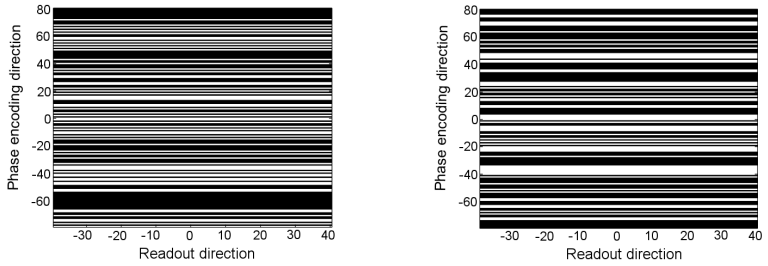
(a) $b = 0\,\mathrm{s/mm^2}$ (b) $b = 200\,\mathrm{s/mm^2}$ (c) $b = 800\,\mathrm{s/mm^2}$

Figure 5.10: Ten slices of patient 13 were reconstructed using the undersampling strategies 'center' (blue), 'centerincreased'(red), 'increased' (green), 'random' (purple) and 'uniform' (cyan) for 12 undersampling percentages. The corresponding mSSIM values were combined in an error plot. The dashed line (black) denotes the mSSIM threshold value of 0.9 to obtain good reconstructions.

The variation between the ten test slices of patient 13 is shown in an error plot (Figure 5.10). The variation between the slices decreased for the strategies 'center', 'centerincreased' and 'increased' when more data was sampled. These three strategies all reached the mSSIM value of 0.9, and thus obtained good reconstructions.

The 'random' strategy did not perform well, independent of the b-value. It performed even worse than 'uniform' in some cases. Recall from Section 5.1 that 'uniform' was included as a lower bound on the performance of the undersampling strategies.

Masks A and B used for 40% of the data were chosen to investigate the poor performance of 'random' (Figure 5.11).



(a) Mask A (b) Mask B

Figure 5.11: Two masks used for undersampling slices with the 'random' strategy to obtain 40% of the data.

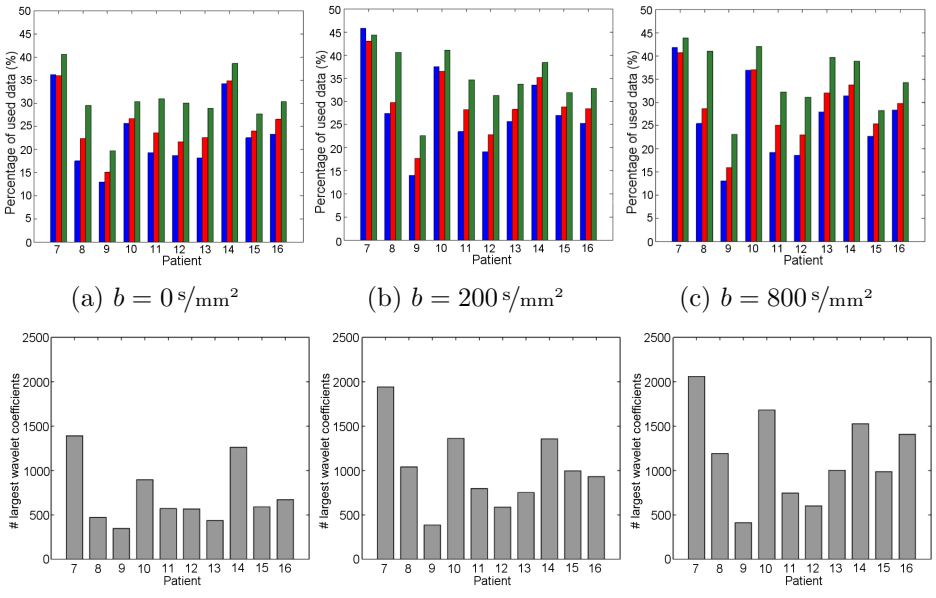
The two masks were applied to slices 20 and 32 of patient 13 with $b = 0\,\mathrm{s/mm^2}$. The power spectral densities (Section 4.3) were computed of both images in k-space (Table 5.2).

Table 5.2: Power spectral density fraction in percent for 'random' undersampling 40% of the data of slices 20 and 32 of patient 13, using masks A and B.

	Slice	
Mask	20	32
A	32.6% mSSIM = 0.6323	28.6% mSSIM = 0.6805
B	71.5% mSSIM = 0.9029	71.0% mSSIM = 0.9134

Comparison of the densities showed that mask A only took approximately 30% of the power spectral density, while mask B took approximately 71% of the density (Table 5.2). Mask B included more information of k-space than mask A, which led to a higher reconstruction quality.

So far, the differences in reconstruction quality between the ten slices for each sampling strategy were examined (Figure 5.10). To compare the reconstructions between patients, the mSSIM of a volume consisting of the ten test slices was computed for each undersampling percentage and for each patient. The 'random' and 'uniform' strategies were omitted, because of their bad reconstruction quality (Figure 5.10). Curves were fitted through the mSSIM values and the intersections of the curves with the threshold value were computed (Figure 5.12a - 5.12c). For the same test slices, the average subset size of largest wavelet coefficients required to obtain a mSSIM ≥ 0.9 was computed to investigate the sparsity per patient (Figure 5.12d - 5.12f).



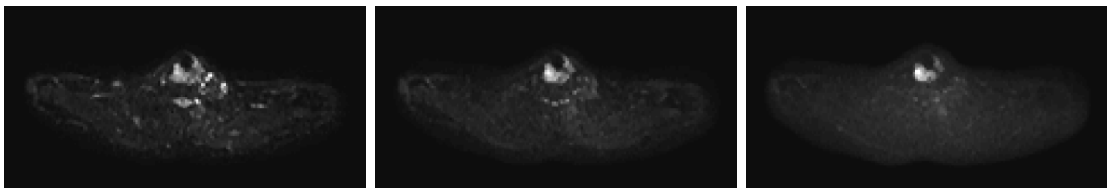
(a) $b = 0\,\mathrm{s/mm^2}$ (b) $b = 200\,\mathrm{s/mm^2}$ (c) $b = 800\,\mathrm{s/mm^2}$

(d) $b = 0\,\mathrm{s/mm^2}$ (e) $b = 200\,\mathrm{s/mm^2}$ (f) $b = 800\,\mathrm{s/mm^2}$

Figure 5.12: Undersampling percentages required to obtain a mSSIM value of 0.9 with three undersampling strategies for a volume consisting of ten slices for ten patients (a-c). The used strategies were 'center' (blue), 'centerincreased' (red) and 'increased' (green).
The average subset size of the largest wavelet coefficients required to obtain a mSSIM ≥ 0.9 for each patient and three b-values (d-f).

There are similarities between the percentages required to obtain a good reconstruction and the number of largest wavelet coefficients required to obtain a good reconstruction. A large number of wavelet coefficients implied that a higher percentage of data was required. A small number of wavelet coefficients often implied that a low percentage of data was required. The average subset size was the smallest for $b = 0\,\mathrm{s/mm^2}$ for all patients.
The test slices of patient 7 performed the worst and the test slices of patient 9 the best among all patients (Figure 5.12). These patients are examined further.

The test slices of patient 9 performed the best of all patients for all combinations. The percentage of data required to obtain a mSSIM value equal to 0.9 was at most 23.1%. This was the case for the combination of $b = 800\,\mathrm{s/mm^2}$ and the 'increased' strategy. Compared to the other patients, there was less data required to obtained the same reconstruction quality.



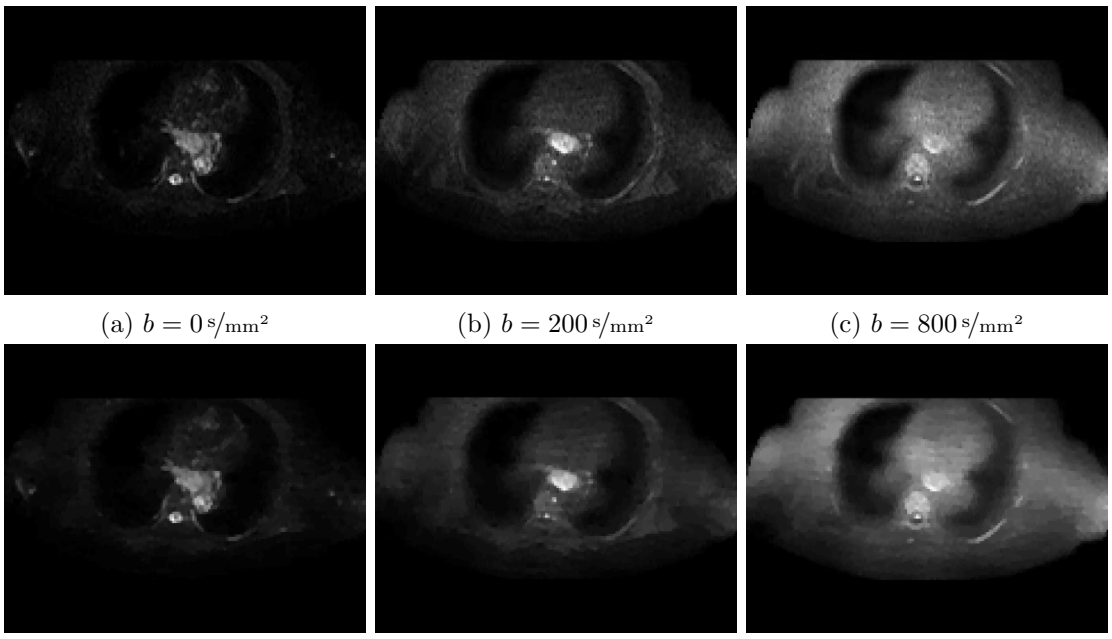
(a) $b = 0\,\mathrm{s/mm^2}$ (b) $b = 200\,\mathrm{s/mm^2}$ (c) $b = 800\,\mathrm{s/mm^2}$

Figure 5.13: 100%-image of slice 30 of patient 9 for the three b-values. The tumor of this patient was located in the proximal part of the esophagus (light part).

Patient 9 had a tumor in the proximal part of the esophagus (Figure 5.13) and therefore the scanned volume was smaller than patients where the tumor is located lower in the esophagus. This implied that the image to be reconstructed was sparser than the images of other patients, making it probably also sparser in the sparsity transform domain. Indeed, a maximum of 400 wavelet coefficients were required to obtain a good reconstruction for patient 9 (Figure 5.12d-5.12f). This was much lower than the average of 950 largest wavelet coefficients required for the other patients.

The test slices of patient 7 performed the worst of all patients for all combinations (Figure 5.12a - 5.12c). The percentage of data needed to obtain a mSSIM value of 0.9 ranged between 36.0% and 45.8%. The subset size of wavelet coefficients ranged between approximately 1400 and 2060.

During the scan of patient 7, one of the coils was not functioning correctly, which resulted in a noise band in the center of the slice images. When cFISTA was used, the noise was suppressed by the TV norm. The noise was already reduced in the 100%-images, but still present (Figure 5.14a-5.14c).
The slices were undersampled using the 'centerincreased' strategy and 30% of the data and then reconstructed (Figure 5.14d-5.14f). These reconstructions were smoother than the 100%-images, so small structures were lost. The large structures were reconstructed well. However, there were too many differences between the 100%-image and reconstructed images due to the noise to obtain a high mSSIM. On the other hand, the light parts of the images seemed to be reconstructed well. Since tumors appear bright on diffusion-weighted images, this meant that the tumor of patient 7 was probably well reconstructed. The quality of the tumor reconstructions is investigated in Section 5.3.3.



(a) $b = 0\,\mathrm{s/mm^2}$ (b) $b = 200\,\mathrm{s/mm^2}$ (c) $b = 800\,\mathrm{s/mm^2}$

(d) $b = 0\,\mathrm{s/mm^2}$ (e) $b = 200\,\mathrm{s/mm^2}$ (f) $b = 800\,\mathrm{s/mm^2}$

Figure 5.14: 100%-image of slice 35 of patient 7 for the three b-values (a-c). The reconstructions of the slice per b-value (d-f). The slice was undersampled using 'centerincreased' and 30% of the data.

5.3.2 Comparing strategies

In Section 5.3.1, ten slices of each of the ten patients were reconstructed and the difference in performance between patients was discussed. In this section, the focus was on the differences in performance between undersampling strategies.

The 'center', 'centerincreased' and 'increased' strategies were compared, since these strategies obtained good quality reconstructions. For each undersampling strategy, ten slices of nine patients were combined to obtain the average mSSIM values for this data set. The data set contained ninety test slices (Figure 5.15). Patient 7 was excluded from this analysis, because this data was contaminated with extra noise due to an incorrectly functioning coil (Section 5.3.1).

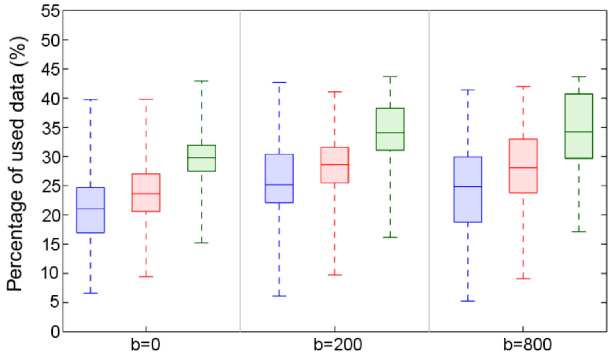


Figure 5.15: Percentages of the required data to obtain a mSSIM value ≥ 0.9 are displayed as boxplots. The data sets consisted of ten slices of nine patients for the undersampling strategies 'center', 'centerincreased' and 'increased'.

For the three b-values, it held that the 'center' strategy required the least data to obtain an average mSSIM equal to 0.9. The average percentage of data required over the b-values was 25.9%. The 'centerincreased' strategy required on average 28.1% of the data to get the same results. The strategy 'increased' required the most data, with an average of 33.4%.

The variation is here defined as the difference in percentages between the reconstructions that required the least and the most amount of data to obtain a mSSIM value of 0.9. The latter reconstruction is here referred to as the *worst case reconstruction*.
The variation and the percentage of the worst case reconstruction was computed for the three undersampling strategies (Table 5.3).

Table 5.3: The average variation between the percentages required for good reconstructions and the worst case reconstruction percentages for three undersampling strategies and the three b-values.

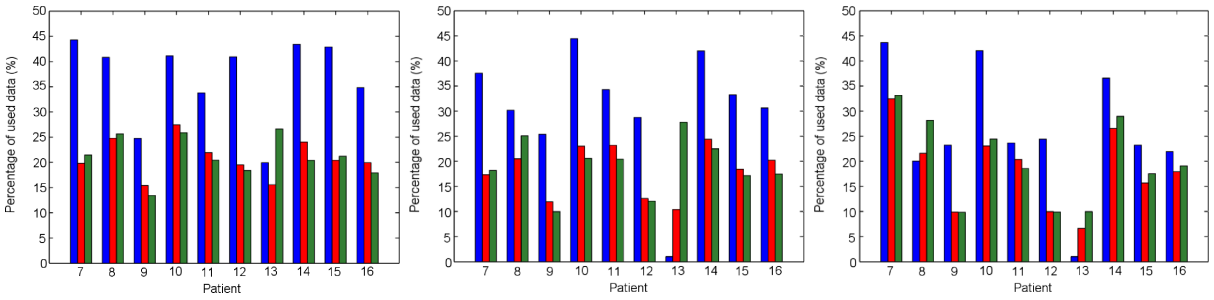
	Variation			Worst case reconstruction		
	$b = 0$	$b = 200$	$b = 800$	$b = 0$	$b = 200$	$b = 800$
center	33.2%	36.5%	36.1%	39.7%	42.7%	41.4%
centerincreased	30.4%	31.4%	32.9%	39.8%	41.1%	42.0%
increased	27.7%	27.6%	26.6%	42.9%	43.7%	43.7%

The largest variation for each b-value was for 'center' (on average 35%). The variation was the lowest for 'increased' (on average 27%), followed by the average variation for 'centerincreased' (32%). The variation between the slices undersampled with 'increased' decreased when higher b-values were used. For the other two strategies, an increase was seen of approximately 3%. The percentages required for the worst case reconstructions were similar for 'center' and 'centerincreased'. This percentage was the lowest for $b = 0\,\mathrm{s/mm^2}$ and $b = 800\,\mathrm{s/mm^2}$ using 'center' (40% and 41% respectively) and for $b = 200\,\mathrm{s/mm^2}$ using 'centerincreased' (41%).

5.3.3 Tumor data

In Section 5.3.1, it was mentioned that a good reconstruction quality of the tumor might be obtained even though the whole image does not have a good quality. In this section, the quality of the tumor reconstruction of the ten patients was investigated. The tumors were delineated by clinicians. By applying the delineation masks onto the reconstructed images and the 100%-images, images of only the tumor were obtained for these type of images. The obtained tumor image of the 100%-image is referred to as the 100%-tumor image.

Next, the 100%-tumor images and the reconstructed tumor images were compared and the mSSIM values were computed (Figure 5.16).



(a) $b = 0$ (b) $b = 200$ (c) $b = 800$

Figure 5.16: Undersampling percentages required to obtain a mSSIM value of 0.9 with three undersampling strategies for a tumor volume consisting of ten slices for ten patients. The used strategies were 'center' (blue), 'centerincreased' (red) and 'increased' (green).

'Center' performed worse than the other two strategies, where patient 13 was an exception. The reason for this will be discussed in Section 6.1.
A data set of ninety test slices was obtained by including ten slices of the tumor of each of the nine patients. Note that patient 7 was still left out of the analysis due to the ill-functioning coil.

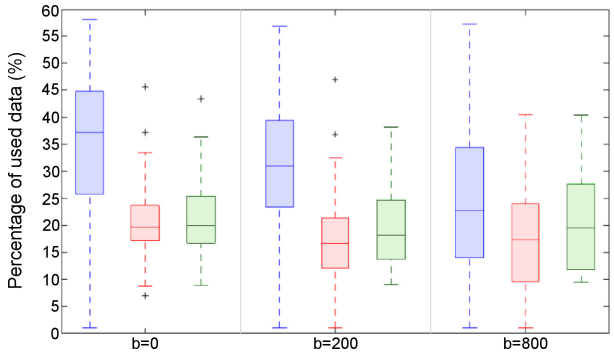


Figure 5.17: Percentages of the used data required to obtain a mSSIM value ≥ 0.9 are displayed as boxplots. The data set consisted of ten slices containing the tumor of nine patients for the undersampling strategies 'center', 'centerincreased' and 'increased'. The outliers, a result of using the boxplot function in Matlab [21], are denoted by a +-sign. A point is called an outlier if its value is smaller than $q_1 - 1.5(q_3 - q_1)$ or larger than $q_3 + 1.5(q_3 - q_1)$, with q_1 and q_3 the 25th and 75th percentiles, respectively.

The 'center' strategy required on average the most data (31%) and 'centerincreased' the least (19%). 'Increased' required on average 20% of the data. The variation between the test slices was the smallest for 'increased' (29%), followed by 'centerincreased' (32%). The variation between the slices for 'center' was the largest with approximately 56%. The percentage of worst case reconstruction was the smallest for each b-value using 'centerincreased'.

The ADC value was used to obtain a quantitative measure to compare the strategies (Section 2.3.3). If the mean ADC values of the undersampling strategies were close to the ADC value of the reference data, then this also indicated a good reconstruction.
As an example, the mean ADC values of patient 11 were computed for the 'center', 'centerincreased', 'increased' and 'random' undersampling strategies and the 12 percentages. These results were compared with the mean ADC value of the reference tumor data set of patient 11 (Figure 5.18).

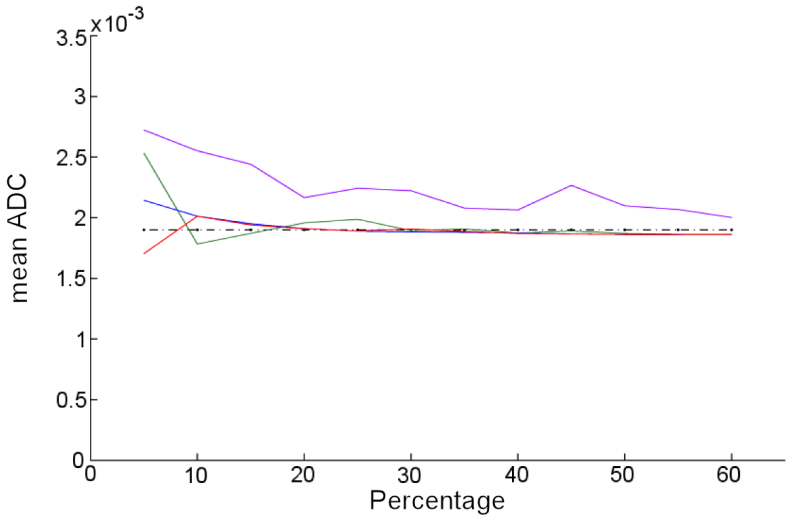


Figure 5.18: The mean ADC value of the reference image (black) of patient 11 is compared with the mean ADC values of the reconstructed images using the undersampling strategies 'center' (blue), 'centerincreased' (red), 'increased' (green) and 'random' (purple) for 12 undersampling percentages.

The mean ADC value of 'center' was approximately the same as the reference data when a minimum of 20% of the data was used during reconstruction. This was less than the percentages required for the whole image reconstruction (25.9%) and the tumor reconstruction (31%) based on the mSSIM (Figure 5.15 and Figure 5.17). The percentage of data required to obtain a similar mean ADC value for 'increased' was approximately 35%. The result of this strategy was comparable with the percentage found for the whole image reconstruction (33.4%), but it was higher than the percentage found for the tumor reconstruction (20%). The mean ADC value of 'centerincreased' was approximately the same when 20% of the data or more was used during reconstruction, which was similar to the percentage found for the tumor reconstruction with this strategy.

5.3.4 Prospective undersampling experiment

The experiments performed above were executed retrospectively, i.e. an image was Fourier transformed to k-space and there undersampled to obtain a naive reconstruction (Figure 5.19). The goal was however to directly acquire the MR data in an undersampled fashion. For this reason, a prospective experiment was also performed (Figure 5.19) to test the performance of cFISTA on raw (i.e. unprocessed) undersampled MR data.

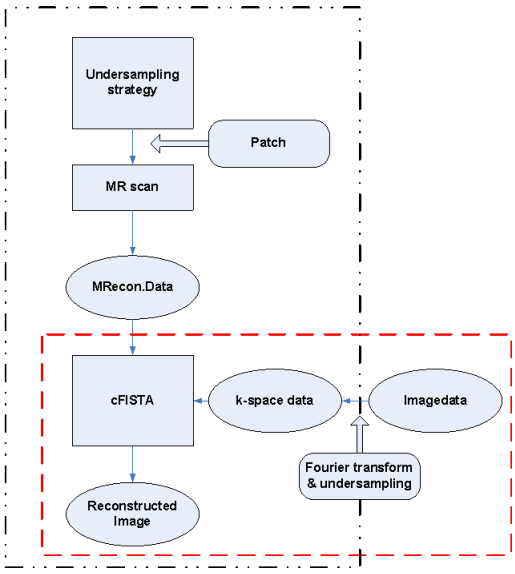
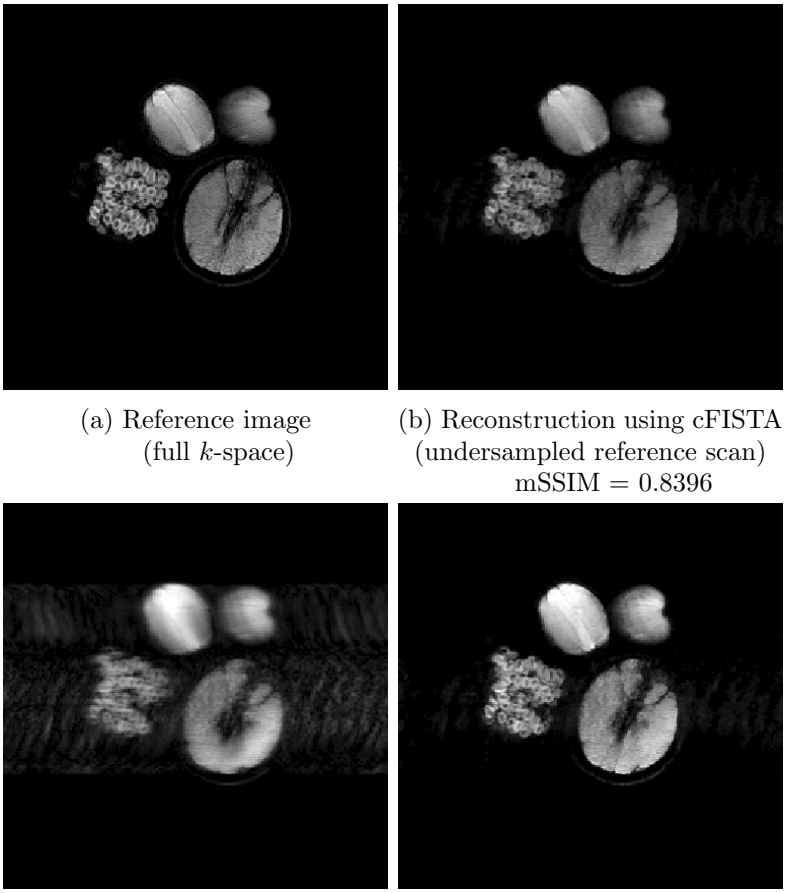


Figure 5.19: Schematic representation of the work-flow of the experiments. The lower box (red) displays the retrospectively performed experiments. From image data, a naive reconstruction is obtain. This reconstruction is used as input of cFISTA. The output of cFISTA is the reconstructed image. The other box displays the prospectively performed experiments. An image in k-space is obtained in a undersampled fashion by performing a MR scan with a special patch that controls the MR scanner. Using this data, a naive reconstruction is obtained using 'ReconFrame'[50]. Then the image is reconstructed using cFISTA.

For this experiment, a kiwi, a pomegranate, an orange and a lemon were scanned on a 1.5 Tesla MR scanner, using a gradient echo. Two scans were performed: a reference scan using the full k-space and a scan using the undersampling strategy 'centerincreased' to obtain 25% of the data. The raw MR data was loaded into Matlab using the software package called 'ReconFrame' [50] (Gyrotools, Zurich, Switzerland).
Subsequently, the undersampled raw data was reconstructed using cFISTA (Figure 5.20). An 8-channel head coil was used for signal reception. Reconstructions for each of the eight coils were performed separately. Then, the eight images were combined using the function 'CombineCoils' in 'ReconFrame'. This function combines the coils by performing a sum-of-squares combination in image space.



(a) Reference image (full k-space)

(b) Reconstruction using cFISTA (undersampled reference scan) mSSIM = 0.8396

(c) Reference image (undersampled k-space)

(d) Reconstruction using cFISTA mSSIM = 0.7930

Figure 5.20: Four MRIs of a kiwi (top left), pomegranate (bottom left), orange (bottom right) and a lemon (top right). An 8-channel head coil was used for signal reception, and the displayed images are the combined reconstructed images of eight coils using sum-of-squares. The raw MR data of two scans was acquired: the full k-space from a reference scan and the undersampled k-space where 'centerincreased' was used to obtain 25% of the raw MR data. The reference image used the full k-space (no undersampling) (a). This MR data set was undersampled ('centerincreased' and 25% of the data) and reconstructed to obtain the image in (b). From the MR data that was acquired in a undersampled fashion, a reference image was obtain by a naive reconstruction (c). The reconstruction using cFISTA for the same data set as the undersampled reference image is displayed in (d). The reconstructions are compared to (a).

The pieces of fruit in the reference scan using undersampled k-space were blurred and there was a band of aliasing artifacts (Figure 5.20c). Recall from Section 5.1 that the undersampling was done in the phase encoding direction. Therefore, the aliasing artifacts only appeared in the phase encoding direction (left-right) and not the readout direction (top-bottom).
After applying cFISTA, most of these artifacts have vanished (Figure 5.20d). Details also have been reconstructed, as can been seen for example in the kiwi (top left piece of fruit). There is a clear transition between the center part of the kiwi and its outer parts, just like in the reference image (Figure 5.20a). The reconstruction was compared to the reference image of the full k-spaceand had a mSSIM equal to 0.7930. However, the comparison is performed between two different scans and therefore the actual quality of the image may differ from this found mSSIM (Section 4.2.7).

To investigate the influence of comparing different scans, the full k-space data of the reference image was also undersampled. The raw MR data was used, so the experiment was again performed prospectively. Using the strategy 'centerincreased', 25% of the raw MR data was obtained, which was then reconstructed using cFISTA (Figure 5.20b). This reconstruction led to a mSSIM value equal to 0.8396. This mSSIM value was higher than the value found for the reconstruction where the MR data was acquired in an undersampled fashion (Figure 5.20d).

This fruit experiment showed the working of the process to reconstruct an image from k-space data, which was acquired in an undersampled fashion. The process is therefore ready to be implemented and used on MR data.

5.3.5 3D

In all experiments discussed before, the reconstructions were performed on 2D images with undersampling only in the phase encoding direction. The theory of CS (Section 3.2) uses the sparsity of the solution, which is in this case an image. However, the total number of pixels is quite small in the used DWI data. The technique of CS may be better exploited when a volume is reconstructed using 3D undersampling.

A 3D volume was undersampled by using a 3D mask (Figure 5.21). The slices of the volume were undersampled using an undersampling strategy (e.g. 'centerincreased') to obtain $u_{\text{perc}}\%$ of that slice. By again using an undersampling strategy, it was determined which slices are undersampled and which slices are not sampled at all. This created a mask that takes $u_{3D\text{perc}}\%$ of the total data.

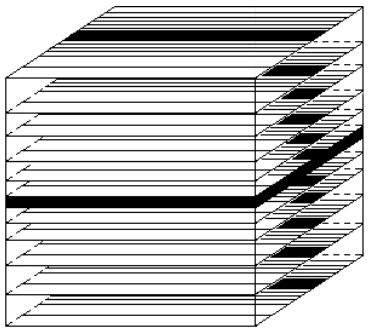
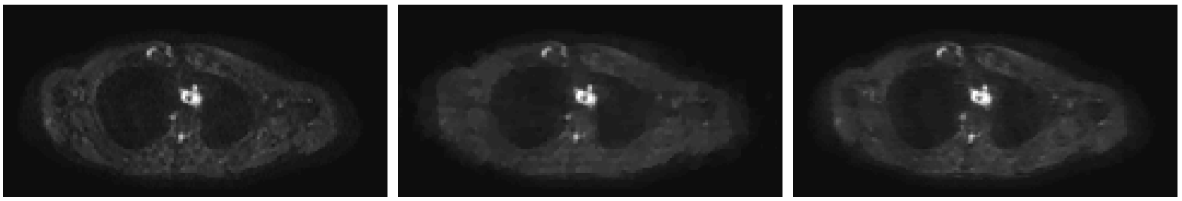


Figure 5.21: 3D mask to obtain 25% of the total data. Slices of the volume are undersampled using strategy 'centerincreased' to obtain 35% of the data in that slice. By again using 'centerincreased', it is determined which slices are undersampled and which slices are not sampled.

To investigate the working of the 3D version of cFISTA (cFISTA 3D), one patient volume of size $160 \times 160 \times 50$ in image space was transformed to a volume in k-space using the 3D Fourier transform. It was undersampled to obtain 25% of the total data. The mask used for the undersampling is displayed in Figure 5.21. Then the volume was reconstructed using cFISTA 3D. The reconstructed volume was compared with the original data set using the SSIM applicable to volumes (Section 4.2.6). The mSSIM was equal to 0.9402 and thus a (very) good reconstruction was obtained.

To compare the accuracy of cFISTA 3D with cFISTA, two reconstructed images were chosen from the volume. One image was reconstructed individually using cFISTA (Figure 5.22b). The image was first undersampled using the strategy 'centerincreased' to obtain 25% data. The

other image was reconstructed within the volume using cFISTA 3D (Figure 5.22c). The mask in Figure 5.21 was used for the undersampling. These images were compared with the original image of the slice used for the other two images (Figure 5.22a).



(a) Reference image 100%

(b) Reconstructed slice 2D 25% of total data mSSIM = 0.9032

(c) Reconstructed slice 3D 25% of total data mSSIM = 0.9379

Figure 5.22: An image from a patient volume is undersampled with 'centerincreased' to obtain 25% data. It is reconstructed using cFISTA (b). The same image is undersampled within a volume. The slices were also undersampled with 'centerincreased' to obtain 25% of the total data. The reconstruction of the volume was performed using cFISTA 3D and the same slice as in (b) was displayed in (c). The reconstructions were compared to the 100%-image (a).

The same undersampling was performed on both slices, namely 'centerincreased' to obtain 25% of the total data. The selected image was not included in the undersampling mask. The mSSIM of the reconstructed volume slice was higher than the mSSIM of the reconstructed slice.

It is not (yet) possible to acquire a 3D volume using an EPI trajectory. Therefore, another method to obtain a 3D volume was performed.
A stack of images was transformed to the Fourier domain, one image at the time. The obtained volume in k-space was undersampled, where the slices were randomly chosen from the stack. The undersampled volume was again reconstructed using cFISTA 3D. Since each slice was separately appended to the k-space volume, this method should be applicable to the current EPI acquisitions. The reconstruction of the volume was compared to the 100%-volume. The mSSIM value of the volume reconstruction was 0.3914, i.e. a poor reconstruction was obtained.

Chapter 6

Discussion

In Chapter 2, the problem of geometric distortions of DWI when using an EPI trajectory has been discussed. It has been made plausible that these distortions could be reduced by undersampling the image in k-space. In Chapter 3, the technique of Compressed Sensing was explained. The minimization problem to be solved was stated (Equation 3.13):

$$(P) : \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \alpha\|x\|_{TV} + \beta\|\Psi x\|_1 \right\}, \tag{6.1}$$

where x was the solution (reconstructed image) and A was a Fourier transform combined with a mask to undersample the image in k-space. Ψ was the sparsifying transform which was chosen to be the Daubechies wavelet with four vanishing moment (Daub4) (Section 5.2.1). The parameters α and β were found using a leave-one-out-method (Section 5.2.2). These parameters determine the influence of two regularization terms $\|\cdot\|_{TV}$ and $\|\cdot\|_1$.
ISTA was explained as an algorithm to solve the minimization problem. This algorithm was improved by Beck and Teboulle [28, 1] and was called FISTA. In this thesis, FISTA was improved and modified such that it was applicable to the complex-valued (MR) data. The resulting algorithm was called cFISTA. In Chapter 5, experiments were executed on diffusion-weighted images of ten patient data sets. The images were undersampled and again reconstructed using cFISTA. A reconstruction was said to be good when it had a mSSIM value ≥ 0.9. In this chapter, the results are discussed.

6.1 The optimal strategy

CS relies on the assumption that the information is more or less evenly distributed over the measurement vector b. This assures that random sampling from the vector leads to sampling sufficient informative data to obtain an exact reconstruction. MR data, however, is not evenly distributed over b. Most information is contained in the center of k-space, i.e. in the center of b.

Therefore, the 'random' strategy did not perform well in all cases (Section 5.3.1). Masks A and B used for 40% of the data illustrated this (Figure 5.11). Mask A included approximately 30% of the power spectral density, while mask B took approximately 71% of the density (Table 5.2). This coincided with a higher mSSIM value for mask B than mask A.
This also explained why the mSSIM value of a reconstruction could be less when the percentage of the used data was increased. This was for example the case when the percentage was increased from 40% to 45% (Figure 5.10). The power spectral density of the 40%-mask decreased to a density of 25% for the 45%-mask.

So, the performance of 'random' is highly dependent on which lines were chosen.

Based on these findings, three other undersampling strategies were designed to have a high sampling density in the center of k-space, namely 'center', 'increased' and 'centerincreased'. These three strategies were compared to find the optimal strategy among them for undersampling the DWI data set (Section 5.3.2). The optimal strategy should on average require a small amount of data with a high reproducibility. Furthermore, the highest percentage required (the worst case reconstruction) should be as small as possible. The used data set contained ten images of nine patients for each of the b-values $b = 0\,\mathrm{s/mm^2}$, $b = 200\,\mathrm{s/mm^2}$ and $b = 800\,\mathrm{s/mm^2}$. The variation between the slices gives an indication of the reproducibility of the strategy.

'Center' required on average the least amount of data for the three b-values (25.8%), but had the largest variation between the ninety test images (35.3%). The opposite held for 'increased', which required on average 33.4% with a variation of 27.3%. 'Centerincreased' required on average approximately 3% more data than 'center', but the differences between all test slices were closer to the variation between the slices for 'increased' (31.6%). The percentage required for the worst case was the lowest for $b = 0\,\mathrm{s/mm^2}$ and $b = 800\,\mathrm{s/mm^2}$ using 'center' and for $b = 200\,\mathrm{s/mm^2}$ using 'centerincreased'.

The strategies were also compared by investigating the differences in the reconstruction of only the tumor (Section 5.3.3). By focussing on the tumor reconstruction, the average percentages of data required to obtain a good reconstruction could be reduced from 28.1% and 33.4% to 19% and 20% for 'centerincreased' and 'increased', respectively. The 'center' strategy required more data to obtain good tumor reconstructions than the amount of data necessary for a good reconstruction of the whole image. The low spatial frequencies were captured very well by 'center' (Figure 2.5). It only includes information about the details when a high percentage of data is used. However, 'center' worked very well for the reconstruction of the tumor of patient 13. The tumor of this patient was very homogeneous and thus could be represented by low spatial frequencies in k-space.

The reproducibility was the same as the case where the complete FOV was investigated: 'increased' gave the smallest variation between the test slices and 'center' the largest. The percentages required on average and for the worst case were the smallest for 'centerincreased' (for all b-values).

Therefore, it can be stated that the sampling strategy 'centerincreased' gave the best balance between the amount of data used and the differences between all slices for $b = 0\,\mathrm{s/mm^2}$, $b = 200\,\mathrm{s/mm^2}$, $b = 800\,\mathrm{s/mm^2}$. The mean ADC value of this strategy was approximately the same as the original ADC value when 20% of the data or more was used. This is the same percentage required for good tumor reconstruction with this strategy. 'Centerincreased' required approximately 30% of the data to obtain a good reconstruction of the whole image.

For good tumor reconstruction using 'centerincreased', 20% of the data was required. For a good reconstruction of the whole image, 30% of the data was required. Hence, the geometric distortions would be reduced by 80% or 70% respectively. The original average pixel shift for the ten patient data sets was eight, i.e. a pixel was shifted 8 pixels compared to its actual position. For the undersampling situations, the average pixel shift became 1.6 and 2.4, respectively.

Unfortunately, there are technical restrictions that prevent the distance between consecutive lines in the EPI trajectory to be too large. A large distance between successive lines might lead to a higher value of the ramp time τ_{ramp} in the pixel shift equation (Equation2.12)

$$\Delta r_{pe} = \gamma \Delta B_0(x, y, z) M(2\tau_{ramp} + N \cdot D_w).$$

Although M (the number of phase encoding steps) is reduced, the pixel shift does not have to be reduced due to the increase in ramp time. This increase could lead to overlap between the G_x and G_y gradients. This means that the spatial encoding of the points sampled at that time will be different. This situation is not desired and could give a lower bound on the amount of undersampling.
Hence, EPI might not be the most optimal technique to use Compressed Sensing.

6.2 The reconstruction algorithm

The reconstruction algorithm cFISTA takes an initial guess, for which a naive reconstruction (i.e. the resulting image after applying the inverse Fourier transform on the undersampled image in k-space) is chosen here. By iteratively solving minimization problems, a reconstruction is found.
The reconstruction process uses, among others, regularization parameters, sparsifying transforms and a stopping criterion (Section 5.2). Some of these elements are discussed in this section.

6.2.1 Stopping criterion

Since cFISTA is an iterative method, a stopping criterion was used to terminate the algorithm. For the ease of discussion, the algorithm is said to be converged when the mSSIM does not increase any more. Recall that a mSSIM above 0.9 defines a good reconstruction.

Three stopping criteria were investigated:

- obj: the criterion that compares the objective function values of two iterants.
- slope: the criterion that computes the slope between two function values and compares it with a constant.
- max_it: a maximum number of iterations is used.

The max_it-criterion was chosen as stopping criterion (Section 5.2.5). There were cases where the obj-criterion did not terminate, even though a good reconstruction was obtained using a small number of iterations. The slope-criterion performed better than obj. However, a good reconstruction was obtained sooner than the stopping criterion indicated.

The maximum number of iterations for all experiments in Chapter 5 was set to 5000 iterations ($\approx$ 300 seconds). However, it is not guaranteed that the algorithm has converged within 5000 iterations. There were cases where the algorithm could have found the optimal solution within a number of iteration much less than 5000. In that case, redundant iterations are performed that only increase the computation time.

So, the used stopping criterion is also not optimal. However, it has an advantage over the other two criteria. By using the same amount of iterations for all experiments, the conditions of the experiments were all the same. Therefore, the performance of the five strategies with respect to the quality of the reconstructed images could be better compared.

6.2.2 Regularization terms

The regularization terms used in (P) (Equation (6.1)) are $\|x\|_{TV}$ and $\|\Psi x\|_1$. Recall that the TV-norm is essentially also a l_1-norm (Section 3.2.2). The images used for experiments are piecewise constant which leads to a very good performance of the TV-norm. The sparsifying abilities of the TV-norm are too efficient in this situation, making the l_1-norm $\|\Psi x\|_1$ a bit redundant.

The large influence of the TV-norm could also be due to the choice of the regularization parameters α and β. These parameters were determined by using a leave-one-out-method (Section 5.2.2). Computing the mSSIM of the reconstructions using several (α, β)-values led to an area of optimal values. The median of this area was $(\alpha, \beta) = (5 \cdot 10^{-4}, 1 \cdot 10^{-6})$, which led to good reconstructions of the test image. Therefore, these values were chosen for the regularization parameters.
However, the values of β in the optimal area had a large range from 10^{-7} to $5 \cdot 10^{-4}$. This also gives an indication of the small influence of the l_1-norm combined with a wavelet transform. When β became too large, the reconstruction quality deteriorated.

Nevertheless, preliminary test results showed (not reported here) that the additional use of this l_1-norm in the objective function led to a faster convergence of the algorithm. Therefore, the combination of the regularization terms was used for the experiments. However, the absence of a good stopping criterion has made it difficult to investigate the gain of including the l_1-norm in the objective function. When a better stopping criterion is defined, several orders of executing the proximal operator terms might give different results. Using the current stopping criterion, no difference was observed (Section 5.2.6).
The balance between the TV-norm and the l_1-norm $\|\Psi x\|_1$ might also be improved when the weights are changed in the step

$$x_k = px^1 + (1-p)x^2$$

of the CSD method (Algorithm 4). Recall that $p = \frac{1}{2}$ was used for the experiments (Section 3.3.5). However, without a better stopping criterion it is difficult to investigate this.

6.2.3 Sparsifying transforms

The sparsifying abilities of Daub4 were investigated on the DWI data (Section 5.2.1). The size of the set containing the largest wavelet coefficients could give an indication of the sparsifying abilities of the wavelets. If a clear transition between large and small wavelet coefficients exists, the amount of large coefficients s could be determined. This could give an indication of the sparsity of the image in the transform domain. By using the number of largest coefficients s and putting the other coefficients to zero, a s-sparse vector can be obtained. Then, an indication of the number of measurements p could be obtained by

$$p = \mathcal{O}\left(s \log\left(\frac{n}{s}\right)\right),$$

as is suggested by results in CS theory [25]. This means that an optimal undersampling percentage might be found.

However, there was not a clear transition between large and small wavelet coefficients for Daub4 and thus the optimal cut-off to determine the largest coefficients could not be found. This leads to the question whether this wavelet transform is the most suitable sparsifying transform.

6.3 Prospective undersampling

In Section 5.3.4, two prospective undersampling experiments were performed. When the full acquired k-space was undersampled, the mSSIM value of the reconstruction compared to the reference image was higher than the value found for the reconstruction where the MR data was acquired in an undersampled fashion.
The reconstruction based on the undersampling of the full k-space included the same noise as the reference image. The reconstruction based on directly undersampled MR data acquisition included different noise. The luminance and structures of the reconstruction in Figure 5.20b were closer to the luminance and structures of the reference image than the reconstruction in Figure 5.20d. This resulted in the difference between the mSSIM values.

It can be concluded that performing cFISTA onto undersampled MR data leads to fine reconstructions, although the SSIM gives a lower mSSIM due to the comparison of different scans. However, the mSSIM could be used to compare the different strategies.

6.4 3D

The undersampling of k-space was only performed in one dimension (the phase encoding direction) due to the EPI trajectory. However, CS reaches more of its potential when the measurement vector b has a larger size. The image size used in the experiments was 128×160, which is small. The resolution of the image could be improved, leading to a larger image size and thus a larger b.

Another method to obtain a larger b is by undersampling in more dimensions. For contrast imaging that does not use an EPI trajectory, the undersampling in more dimensions can be more easily performed. When, for example, single point imaging is used, 2D undersampling can be performed on an image in k-space. This leads to (pseudo-)random undersampling in two dimensions. In these types of imaging, there will probably be less geometric distortions present than in EPI. Nevertheless, CS techniques are advantageous in this case, because it will lead to a relatively large reduction of the scanning time.

An experiment on DWI patient data was performed where the undersampling took place in more than one dimension. (Section 5.3.5). The 3D volume in image space was transformed to a volume in k-space by applying a 3D Fourier transform. After using an undersampling mask (volume), cFISTA 3D could be used. The undersampling for the 3D volume was performed in 2 dimensions, namely the phase encoding direction as well as the slice selection. A 3D inverse Fourier transform was then applied to obtain a naive reconstruction, which was used as input for cFISTA 3D (Section 5.3.5). The reconstruction of the volumes had a better quality than the reconstructions obtained for each slice separately (Figure 5.22).
cFISTA 3D uses information of the neighboring slices for the reconstruction of a slice in the volume. The CS techniques are better used in this case (sparser measurement vector b), leading to a better reconstruction.
Another advantage of 3D reconstruction is that a whole volume is reconstructed during one reconstruction process. To obtain a reconstructed volume using 2D reconstruction, every slice needs to be reconstructed separately. This takes in total more time than the 3D reconstruction. Hence, it might be beneficial to use cFISTA 3D, although more research should be performed.

Since it is not (yet) possible to perform a 3D EPI trajectory, another experiment was performed (Section 5.3.5). A 2D Fourier transform was applied on a stack of slices (i.e. each slice was transformed separately). A 3D inverse Fourier transform was applied to obtain a naive reconstruction. Used cFISTA 3D, the volume was reconstructed. However, the mSSIM value of the volume was 0.3914, so a poor reconstruction was obtained. More research is required to improve this method in order to obtain better reconstructions.

6.5 Conclusion

The main goal of this thesis was to investigate Compressed Sensing as a technique to reduce geometric distortions in DWI. Undersampling leads to a higher bandwidth, which in turn would lead to less distortions.
Five undersampling strategies were designed and experiments on DWI patient data set were performed. The derived algorithm cFISTA could reconstruct undersampled images by solving an appropriate minimization problem based on the theory of CS.
From five undersampling strategies, the optimal strategy for DWI patient data was found: 'centerincreased' leads to good tumor reconstructions where just 20% of the MR data is required. This would theoretically lead to an 80% reduction of geometric distortions.

Chapter 7

Recommendations

Although a reconstruction algorithm is derived and an optimal undersampling strategy has been found, there still remain several problems that should be solved. There are also topics that could be improved. In this chapter, some following research subjects will be mentioned.

7.1 Implementation of MR scanner

The goal was to minimize the geometric distortions of DWI when an EPI trajectory was used. However, it is not (yet) possible to acquire MR data in an undersampled fashion using an EPI trajectory on the current MR scanners. For this reason, the experiments are only performed retrospectively. Since these experiments are performed on already processed data, it could not be verified whether undersampling k-space using the optimal strategy indeed reduces the geometric distortions.
Undersampling k-space does not necessarily lead to a reduction of the geometric distortions due to a possible increase of the ramp time τ_{ramp} (Section 6.1). There might be an optimum, which gives the best balance between τ_{ramp} and the number of sampled lines M. However, it is not possible to investigate the existence of this optimum at this moment due to the technical issues with undersampled EPI on the current scanners.

7.2 The reconstruction algorithm

Parameters and transforms were chosen for the reconstruction algorithm cFISTA during this research. However, some choices could be improved when more research is performed. Furthermore, there still remain some problems that need to be solved.

7.2.1 Stopping criterion

The stopping criterion should be modified to perform better and be more designed for this type of data. For example, monitoring the quality of the reconstruction, using SSIM, during every iteration might give more insight into a way of designing a more appropriate stopping criterion.

7.2.2 Undersampling strategy

It might be more advantageous to use a spiral type of trajectory instead of the EPI trajectory (Section 2.2.1). These trajectories intrinsically result in a sampling density in the center of k-space and thus will probably capture the most information located in the center. Furthermore, this trajectory leads to 2D undersampling.

DWI could be performed using such spiral trajectory [51]. However, it leads to a non-Cartesian coordinate system and thus extra computational steps are required. This means that more post-processing steps are necessary. The undersampling for these trajectories could take place in the form of increasing the speed of sampling (i.e. more sampling in the center of k-space) or by increasing the radius of the spiral. The first type of undersampling might be better for the application in mind.

7.2.3 Regularization terms

The regularization term in the minimization has to be a l_1-norm, since this well-defined norm is most similar to the l_0-norm (which is not a well-defined norm). To improve the quality of the reconstructions, better sparsifiers should be found.

One improved sparsifier might be the use of the TV-norm in combination with only the first (and maybe second) level of the Haar wavelet. The use of the TV-norm leads to the loss of the constant function information. By combining the TV-norm with a wavelet which preserves this constant function, a good sparsifier might be obtained. The challenge here however is to obtain only the constant function information of the wavelet instead of the information for more levels as is normally the case.

7.2.4 Sparsifying transforms

Another possibility to improve the sparsifier is to use other sparsifying transforms Ψ in (P) than the Daubechies wavelet with four vanishing moments (Daub4).
There may be other sparsifying transforms that will have a clear transition between the large and small coefficients and thus perform better for the diffusion-weighted data. Some examples below might be worthwhile to be investigated.

Curvelets

Curvelets are similar to wavelets because they also are parameterized by scale and location. However, curvelets also included orientation, which leads to a good approximation of edges [52, 53] in much less coefficients than are needed for wavelets. The diffusion-weighted images consist of piecewise smooth regions separated by smooth contours. Curvelets could probably approximate these smooth contours better than wavelets, although they may represent point-like features less than wavelets. Curvelets require a rotation operation which is challenging for discrete images sampled on a rectangular Cartesian grid and therefore not appropriate for image compression [54].

Contourlets

Contourlets [55] can also sparsely represent curves, and have a lower computing complexity than curvelets for discrete images. The reason for this is that contourlet transforms are directly defined on discrete rectangular grids. Unlike wavelets, contourlets are not good in representing point-like image features. However, contourlets could probably also approximate the smooth contour of the diffusion-weighted images better than the wavelets. Contourlet functions have less clear directional features than curvelets, resulting in artifacts in compression and denoising processes [54].

A transform which combines the properties of a wavelet and a curvelet or contourlet might lead to the best results in sparsifying the DWI data. This combined transform [56] will have

the wavelet property of representing point-like features as well as the ability to represent curve-like features due to the curvelet or contourlet transform properties. Although curvelets possess more directional features, the use of a contourlet might be favorable, since this is it designed on discrete grids.

7.2.5 Combination SENSE and CS

Nowadays, SENSE is used for DWI-EPI to reduce the amount of acquisition data. It might be useful to combine SENSE with CS to reduce the fraction of required data even further. In [57], CS was combined with SENSE for chemical shift imaging in muscular dystrophy. The coil sensitivities S_i ($i = 1, \ldots, \#\text{coils}$) were included in A (Equation (6.1)), which already consisted of the Fourier transform $\mathcal{F}$ and the undersampling mask M. By including the coil sensitivities and thus combining CS with SENSE, problem (P) (Equation (6.1)) becomes

$$(P_{\text{SENSE}}) : \underset{x}{\operatorname{argmin}} \left\{ \sum_{i=1}^{\#\text{coils}} \frac{1}{2} \|M\mathcal{F}S_i x - b_i\|_2^2 + \alpha \|x\|_{TV} + \beta \|\Psi x\|_1 \right\}. \tag{7.1}$$

More research has to be performed in order to find the optimal parameters, undersampling strategies and percentages of data required from the image in k-space for DWI.

7.3 Applications

cFISTA is an algorithm that solves a minimization problem. The parameters α and β in the problem were optimized for this particular DWI data set. These parameters lead to good reconstructions in other data sets as well. These other data sets included the Shepp-Logan phantom, the image of Lena [58] and a MR phantom data set acquired with a gradient echo sequence (not reported in this thesis). Therefore, the choice of parameters seems to be robust and could be used for other data sets.

There was a large variation of required undersampling percentages observed between the patients (Section 5.3.1). An image displaying a small volume of the patient can be reconstructed using less data than an image displaying a large volume.
The reproducibility of the reconstruction quality of the tumor was high (Section 5.3.3). The tumor in the test data set was located in the esophagus. This means that the tumor size was approximately the same for all patients.
The undersampling and reconstruction can also possibly be performed on other tumor locations, because no additional information about the type or location of the tumor was used during the reconstruction of the image. The variation between the tumors of patients indicates that it is expected that tumors in other site give the same results as tumors in the esophagus when the tumors in other sites are small or homogeneous. However, the variation between patients indicates that large patient volumes or large tumors might lead to different results.
Therefore, further analysis is required to guarantee a good reconstruction when the tumor is for example in the abdomen (large volume).

Appendix A

Appendix

A.1 ADC formula

For b-values b_1 and b_2, the strength is described by

$$S(b_1) = S(0)e^{-b_1 D},$$
$$S(b_2) = S(0)e^{-b_2 D}.$$

Then

$$\frac{S(b_2)}{S(b_1)} = \frac{e^{-b_2 D}}{e^{-b_1 D}},$$
$$-(b_2 - b_1)D = \ln\left(\frac{S(b_2)}{S(b_1)}\right),$$
$$D = -\frac{1}{b_2 - b_1}\ln\left(\frac{S(b_2)}{S(b_1)}\right).$$

So, the ADC is equal to D.

A.2 Convex and concave

Definition 3 (Convex function)**.** *A real-valued function h is convex if*

$$h(tx + (1-t)y) \leq th(x) + (1-t)h(y), \quad x, y \in \mathbb{R}^n, t \in [0,1].$$

Definition 4 (Concave function)**.** *A real-valued function h is concave if*

$$h(tx + (1-t)y) \geq th(x) + (1-t)h(y), \quad x, y \in \mathbb{R}^n, t \in [0,1].$$

A.3 Calculations used for (P_2)

$$\begin{aligned} L\|z\|_2^2 - 2(z, A^*r_0) &= L\|z\|_2^2 - 2L\left(z, \frac{1}{L}A^*r_0\right) + \frac{1}{L}\|A^*r_0\|_2^2 - \frac{1}{L}\|A^*r_0\|_2^2 \\ &= L\left\|z - \frac{1}{L}A^*r_0\right\|_2^2 - \frac{1}{L}\|A^*r_0\|_2^2. \end{aligned} \tag{A.1}$$

$$\begin{aligned} \frac{1}{2}L\|x - y_1\|_2^2 + (x - y_1, \nabla f(y_1)) &= \tfrac{1}{2}L\|x - y_1\|_2^2 + L(x - y_1, \tfrac{1}{L}\nabla f(y_1)) + \tfrac{1}{2L}\|\nabla f(y_1)\|_2^2 - \tfrac{1}{2L}\|\nabla f(y_1)\|_2^2 \\ &= \tfrac{1}{2}L\|x - y_1\|_2^2 + L(x - y_1, \tfrac{1}{L}\nabla f(y_1)) + \tfrac{L}{2}\|\tfrac{1}{L}\nabla f(y_1)\|_2^2 - \tfrac{1}{2L}\|\nabla f(y_1)\|_2^2 \\ &= \frac{1}{2}L\|x - y_1 + \frac{1}{L}\nabla f(y_1)\|_2^2 - \frac{1}{2L}\|\nabla f(y_1)\|_2^2 \\ &= \frac{1}{2}L\|x - (y_1 - \frac{1}{L}\nabla f(y_1))\|_2^2 - \frac{1}{2L}\|\nabla f(y_1)\|_2^2. \end{aligned} \tag{A.2}$$

A.4 Rewriting of structure term in SSIM formula

To get an even better understanding of the structural similarity term, the definitions of $\sigma_{x_k}\sigma_{y_k}$ and $\sigma_{x_k y_k}$ are used. Define $v_i := x_i - \mu_x$ and $w_i := y_i - \mu_y$, then

$$\begin{aligned} \sigma_{x_k}\sigma_{y_k} &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_{x_k})^2} \cdot \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu_{y_k})^2} \\ &= \frac{1}{N-1}\sqrt{\sum_{i=1}^{N}(x_i - \mu_{x_k})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_i - \mu_{y_k})^2} \\ &= \frac{1}{N-1}\sqrt{\sum_{i=1}^{N} v_i^2} \cdot \sqrt{\sum_{i=1}^{N} w_i^2} \\ &= \frac{1}{N-1}\|v\|\|w\| \end{aligned}$$

and

$$\begin{aligned} \sigma_{x_k y_k} &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_{x_k})(y_i - \mu_{y_k}) \\ &= \frac{1}{N-1}\sum_{i=1}^{N} v_i w_i \\ &= \frac{1}{N-1}\|vw\|. \end{aligned}$$

So,

$$\sigma_{x_k}\sigma_{y_k} - \sigma_{x_k y_k} = \frac{1}{N-1}(\|v\|\|w\| - \|vw\|)$$

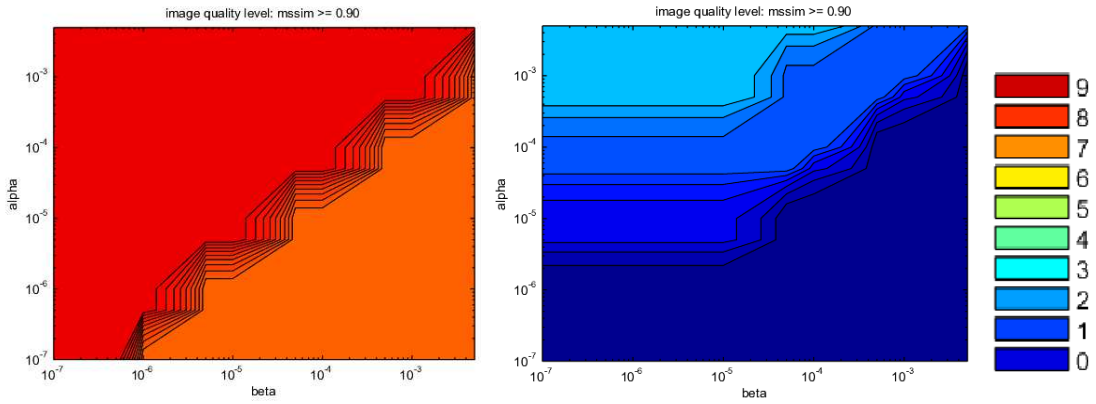
A.5 Optimal parameters

In Section 5.2.2, the following values for α and β are used.

$$\alpha = [1 \cdot 10^{-7}, 5 \cdot 10^{-7}, 1 \cdot 10^{-6}, 5 \cdot 10^{-6}, 1 \cdot 10^{-5}] \cup [5 \cdot 10^{-5} : 5 \cdot 10^{-5} : 4 \cdot 10^{-4}] \cup [4 \cdot 10^{-4} : 1 \cdot 10^{-5} : 7 \cdot 10^{-4}] \cup [7 \cdot 10^{-4} : 5 \cdot 10^{-5} : 1 \cdot 10^{-3}] \cup 5 \cdot 10^{-3}$$

and

$$\beta = 1 \cdot 10^{-7} \cup [5 \cdot 10^{-7} : 1 \cdot 10^{-7} : 1.6 \cdot 10^{-6}] \cup [1.6 \cdot 10^{-6} : 5 \cdot 10^{-7} : 1 \cdot 10^{-5}] \cup [5 \cdot 10^{-6}, 5 \cdot 10^{-5}, 1 \cdot 10^{-4}, 5 \cdot 10^{-4}, 1 \cdot 10^{-3}, 5 \cdot 10^{-3}]$$



(a) 'center' (b) 'increased'

Figure A.1: The number of images with a mSSIM ≥ 0.9, displayed as a contour plot. The undersampling strategies 'center'(left) and 'increased' (right) are used.

Table A.1: mSSIM values obtained using the optimal α and β combination, for the three b-values and three strategies, using 25% of the data.

	center	centerincreased	increased
$b = 0$	0.9128	0.9137	0.8783
$b = 200$	0.9242	0.9100	0.8925
$b = 800$	0.9093	0.8966	0.8801

Bibliography

[1] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM*, vol. 2, no. 1, pp. 183–202, 2009.

[2] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley-Liss, 1 ed., June 1999.

[3] P. M. Systems, *Basic Principles of MR Imaging*. Philips Medical Systems, 1993.

[4] R. Damadian, "Tumor Detection by Nuclear Magnetic Resonance," *Science*, vol. 171, pp. 1151–1153, Mar. 1971.

[5] P. Lauterbur, "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance," *Nature*, vol. 242, 1973.

[6] A. Kumar, D. Welti, and R. R. Ernst, "NMR Fourier Zeugmatography," *Journal of Magnetic Resonance*, vol. 18, Apr. 1975.

[7] K. Coyne, "MRI: a guided tour." `http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri/fullarticle.html`, . Accessed December 2, 2013.

[8] P. Brynolfsson, *Using Radial k-space Sampling and Temporal Filters in MRI to Improve Temporal Resolution*. PhD thesis, Medical Physics, Department of Radiation Sciences, Umea University, Sweden, 2010.

[9] M. Lustig, "SparseMRI Workshop." `http://www.eecs.berkeley.edu/~mlustig/CS.html`. Accessed March 8, 2013.

[10] R. L. DeLaPaz, "Echo-planar Imaging," *RadioGraphics*, vol. 14, no. 5, 1994.

[11] P. Jezzard and R. S. Balaban, "Correction for geometric distortion in echo planar images from B0 field variations.," *Magn Reson Med*, vol. 34, pp. 65–73, July 1995.

[12] J. P. Hornak, *The Basics of MRI*. Interactive Learning Software, 2008.

[13] Hellerhoff. `http://commons.wikimedia.org`, 2013. Accessed March 24, 2014.

[14] P. W. Schaefer, P. E. Grant, and R. G. Gonzalez, "Diffusion-weighted MR imaging of the brain.," *Radiology*, vol. 217, pp. 331–45, Nov. 2000.

[15] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI From Picture To Proton*. Cambridge University Press, 2 ed., 2007.

[16] T. C. Kwee, T. Takahara, R. Ochiai, R. A. J. Nievelstein, and P. R. Luijten, "Diffusion-weighted whole-body imaging with background body signal suppression (DWIBS): features and potential applications in oncology.," *Eur Radiol*, vol. 18, pp. 1937–52, Sept. 2008.

[17] E. Stejskal and J. Tanner, "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient," *Journal of Chemical Physics*, vol. 42, pp. 288–292, Jan. 1965.

[18] K. V. Embleton, H. A. Haroon, D. M. Morris, M. A. L. Ralph, and G. J. Parker, "Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes," *Human Brain Mapping*, vol. 31, pp. 1570–1587, Oct. 2010.

[19] H. Nyquist, "Certain topics in telegraph transmission theory." *Transactions of the A.I.E.E*, vol. 47, 1928; Reprint as classic paper: *Proceedings of the IEEE*, vol. 90, no. 2, Feb. 2002.

[20] C. E. Shannon, "Communication in the presence of noise." *Proceedings of the IRE*, vol. 37, no. 1, 1949; Reprint as classic paper: *Proceedings of the IEEE*, vol. 86, no. 2, Feb. 1998.

[21] MATLAB, *version 7.14.0.739 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2012.

[22] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity Encoding for Fast MRI," *Magn Reson Med*, vol. 42, pp. 952–962, 1999.

[23] M. Weiger, K. P. Pruessmann, C. Leussler, P. Röschmann, and P. Boesiger, "Specific coil design for SENSE: a six-element cardiac array.," *Magn Reson Med*, vol. 45, pp. 495–504, Mar. 2001.

[24] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Math. Acad. Sci. Paris*, vol. 346, no. 9-10, pp. 589–592, 2008.

[25] E. J. Candès and M. Wakin, "People hearing without listening: An introduction to compressive sampling," *IEEE Signal Processing Magazine*, 2007.

[26] M. Fornasier and H. Rauhut, "Compressive Sensing," 2010.

[27] E. J. Candès and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[28] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems.," *IEEE Trans Image Process*, vol. 18, pp. 2419–34, Nov. 2009.

[29] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration.," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992–3004, 2007.

[30] J. Bioucas-Dias and M. A. T. Figueiredo, "Two-Step Algorithms for Linear Inverse Problems with Non-Quadratic Regularization," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1, pp. I–105–I–108, 2007.

[31] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," *Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.

[32] A. Chambolle, "Total Variation Minimization and a Class of Binary MRF Models," in *Energy Minimization Methods in Computer Vision and Pattern Recognition* (A. Rangarajan, B. Vemuri, and A. Yuille, eds.), vol. 3757 of *Lecture Notes in Computer Science*, pp. 136–152, Springer Berlin Heidelberg, 2005.

[33] J. Huang, S. Zhang, H. Li, and D. Metaxas, "Composite splitting algorithms for convex optimization," *Computer Vision and Image Understanding: CVIU*, vol. 115, pp. 1610–1622, Dec. 2011.

[34] J. Huang, S. Zhang, and D. Metaxas, "Efficient MR Image Reconstruction for Compressed MR Imaging." Division of Computer and Information Sciences, Rutgers University, NJ, USA, 2011.

[35] P. L. Combettes and J.-C. Pesquet, "A Proximal Decomposition Method For Solving Convex Variational Inverse Problems," *Inverse Problems*, vol. 24, no. 6, 2008.

[36] S. Venit, "The Convergence Jacobi and Gauss-Seidel Iteration," *Mathematics Magazine*, vol. 48, pp. 163–167, May 1975.

[37] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, 1992.

[38] N. G. Roland, "Fourier and Wavelet Representations of Functions," *Electronic Journal of Undergraduate Mathematics*, vol. 6, pp. 1–12, 2000.

[39] N. Goyal and M. Aggarwal, "Implementation of Image Compression using fast Wavelet Transform using Haar and Daubechies Wavelets," *IOSR Journal of Engineering*, vol. 2, pp. 18–24, Nov. 2012.

[40] S. Rout, "Orthogonal vs. Biorthogonal Wavelets for Image Compression," Master's thesis, Virginia Polytechnic Institute and State University, Aug. 2003.

[41] 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*, vol. 13, no. 4, pp. 600–612, 2004.

[42] Z. Wang and A. C. Bovik, "Mean Squared Error: Love it or leave it?," *IEEE Signal Processing Magazine*, 2009.

[43] Z. Wang and A. C. Bovik, "A Universal Image Quality Index," *IEEE Signal Processing Letters*, vol. 92, pp. 81–84, Mar. 2002.

[44] A. V. Oppenheim and G. C. Verghese, "Signals, Systems, and Inference," 2010.

[45] F. Rieke, D. Warland, R. d. R. van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. MIT Press, 1999.

[46] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging.," *Magn Reson Med*, vol. 58, pp. 1182–95, Dec. 2007.

[47] M. Zaitsev, K. Zilles, and N. J. Shah, "Shared k-space echo planar imaging with keyhole.," *Magn Reson Med*, vol. 45, pp. 109–17, Jan. 2001.

[48] S. Datta, K. Ni, P. Mahanti, and S. Roudenko, "Stability of Efficient Deterministic Compressed Sensing For Images with Chirps and Reed-Muller Sequences," *AMS Mathematics Subject Classification - 94A08,94A20*, 2000.

[49] S. Grgic, M. Grgic, and B. Zovko-Cihlar, "Performance Analysis of Image Compression Using Wavelets," *IEEE Transactions on Industrial Electronics*, vol. 48, pp. 682–695, June 2001.

[50] ReconFrame, *version 3.0.400*. Zurich, Switserland: GyroTools, 2013.

[51] T.-Q. Li, A. M. Takahashi, T. Hindmarsh, and M. E. Moseley, "ADC mapping by means of a single-shot spiral MRI technique with application in acute cerebral ischemia," *Magnetic Resonance in Medicine*, vol. 41, pp. 143–147, Jan. 1999.

[52] E. J. Candès, "What is ... a Curvelet?," *Notices of the AMS*, vol. 50, Dec. 2003.

[53] L. Demanet, *Curvelets, Wave Atoms, and Wave Equations*. PhD thesis, California Institute of Technology, 2006.

[54] J. Ma and G. Plonka, "The Curvelet Transform: A Review of Recent Applications," *IEEE Signal Processing Magazine*, vol. 27, Mar. 2010.

[55] M. N. Do and M. Vetterli, "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation," *IEEE Transactions on Image Processing*, vol. 14, Dec. 2005.

[56] X. Qu, X. Cao, D. Guo, C. Hu, and Z. Chen, "Combined sparsifying transforms for compressed sensing MRI," *Electronic Letters*, vol. 46, Jan. 2010.

[57] K. G. Hollingsworth, D. M. Higgings, M. McCallum, L. Ward, A. Coombs, and V. Straub, "Investigating the Quantitative Fidelity of Prospectively Undersampled Chemical Shift Imaging in Muscular Dystrophy with Compressed Sensing and Parallel Imaging Reconstruction," *Magnetic Resonance in Medicine*, vol. 71, no. 3, 2013.

[58] M. Wakin, "Standard Test Images: Lena." `http://www.ece.rice.edu/~wakin/images/`. Accessed March 11, 2013.