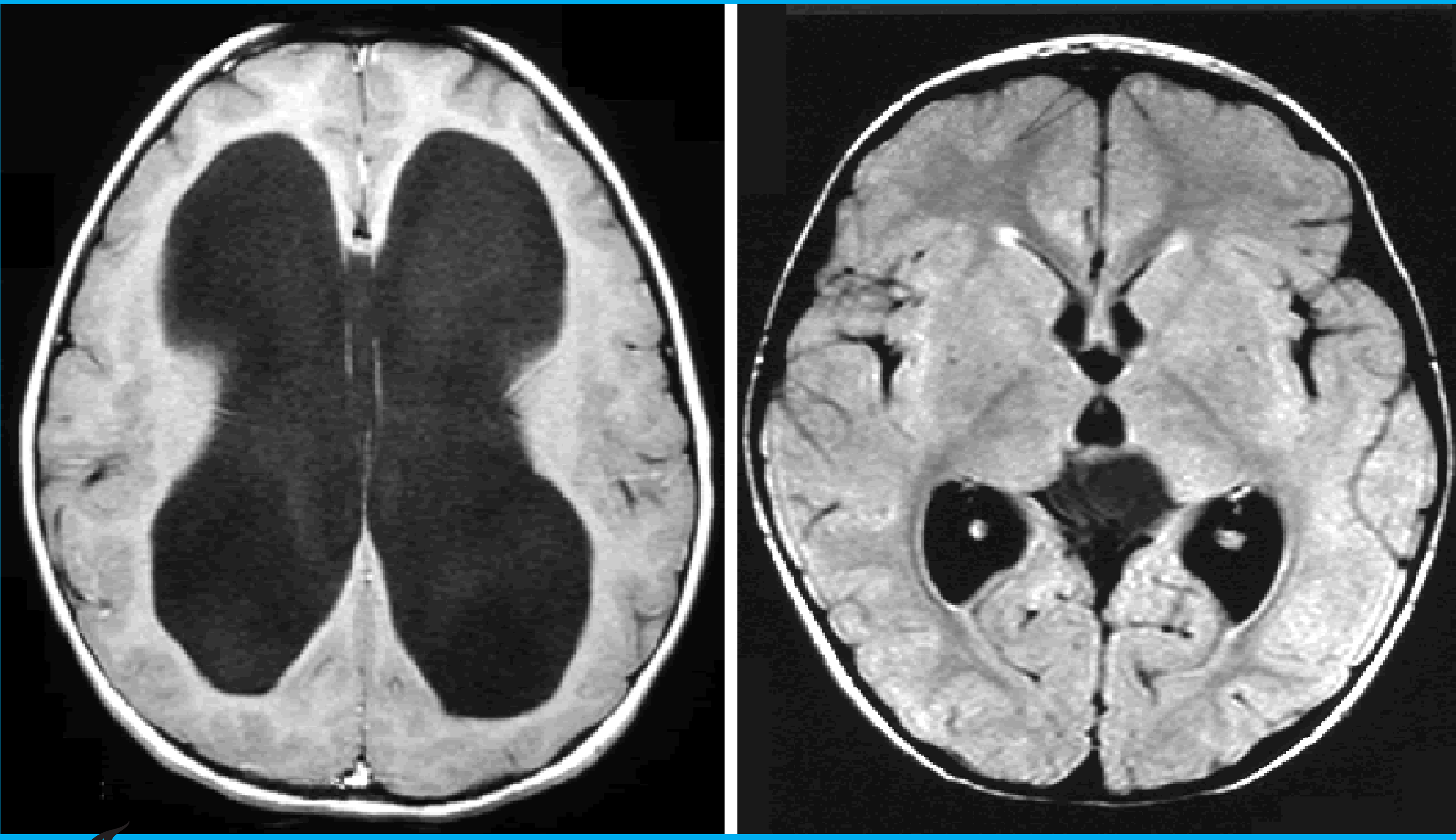
Image Reconstruction in Low-Field MRI

A Super-Resolution Approach

Merel Lisanne de Leeuw den Bouter



TUDelft

Image Reconstruction in Low-Field MRI

A Super-Resolution Approach

by

Merel Lisanne de Leeuw den Bouter

Literature review carried out as part of the master thesis

Student number:	4183576	
Project duration:	September 1, 2016 – Spring 2017	
Thesis committee:	Dr. ir. M.B. van Gijzen,	TU Delft, supervisor
	Dr. ir. R. F. Remis,	TU Delft
	Prof. dr. ir. G. Jongbloed,	TU Delft
	Prof. dr. A. G. Webb,	LUMC
	Prof. dr. ir. S. J. Schiff,	PSU

TU Delft

Contents

1

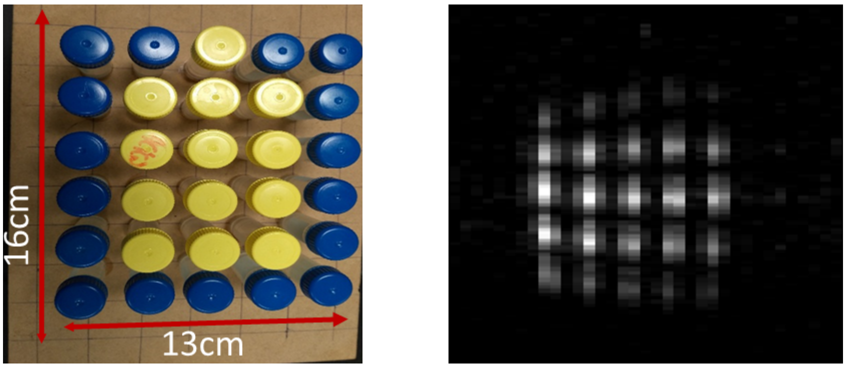
Introduction

Magnetic resonance imaging (MRI) is a powerful technique that reveals the internal structure and function of the human body in a non-invasive way. In clinical MRI scanners, superconducting magnets are used to generate very high magnetic fields. The size and strict infrastructure demands of these magnets make it impossible to use them in many environments. Additionally, these MRI scanners are costly to purchase, site and maintain [1].

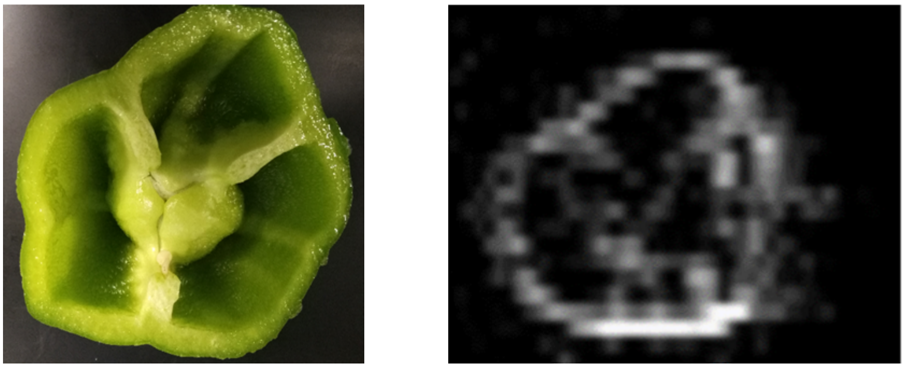
Teams at the Leiden University Medical Center (LUMC) and at Pennsylvania State University (PSU) are working on developing low-cost, portable MRI scanners with a magnetic field in the milliTesla range with the goal of diagnosing children in developing countries with hydrocephalus ('water on the brain'), a debilitating disease that is usually left untreated in the third world. The PSU prototype is shown in Figure 1.1.

Figure 1.1: The PSU low-field MRI scanner prototype.

Some images produced using this prototype are shown in Figure 1.2. The main features of the imaged objects can be discerned, but in order to use this device in a clinical setting, more detailed images are required. The focus of this research is to improve the quality of images generated by low-field MRI scanners using super-resolution reconstruction (SRR). SRR enables us to use several low-resolution images in order to produce one image with a higher resolution. The first super-resolution algorithms were introduced by Gerchberg [2] in 1974 and De Santis and Gori [3] in 1975. The application of super-resolution reconstruction was first mentioned in [4] in 1997. Since then, a large number of articles adapting the super-resolution concept from video-processing to MRI data have been reported and encouraging results have been demonstrated [5]. The approach taken in this research leads to a system of equations of the form $Ax = y$, where x is the unknown quantity. Due to the ill-posedness of this problem, regularization is needed. Another issue is that the higher the desired resolution of the target image, the more pixels, and the larger the matrix. Therefore, for high-resolution problems, matrix inversion is computationally infeasible. Therefore, iterative methods have to be employed to get to the solution. The methods of choice are conjugate gradient methods.



(a) Water bottles. (b) MR image of the water bottles.



(c) Bell pepper. (d) MR image of the bell pepper.

Figure 1.2: The prototype was used to create MR images of an array of water bottles and a bell pepper.

The structure of this report is as follows. Chapter 2 describes conventional MRI: the hardware components, how signals are generated and how they are used to generate an image. In chapter 3, the MRI scanners at LUMC and PSU are discussed and a more general model describing the signal is introduced. Chapter 4 details the super-resolution reconstruction technique that will be used in this research. Chapter 5 contains a description of the conjugate gradient methods that will be used to solve the least-squares problem resulting from the super-resolution reconstruction model. In chapter 6, some preliminary results are shown. In chapter 7, the research goals of this project are formulated.

2

Conventional MRI

This chapter is based on [6]. All images were taken from [6] as well.

2.1. Hardware components

There are three main components in an MR scanner: a main magnet, a magnetic field gradient system and an RF system.

2.1.1. The main magnet

The purpose of the main magnet is to generate a strong and uniform static magnetic field, which is referred to as the B_0-field. In clinical MR scanners, superconducting magnets are used to produce magnetic field strengths between 1.5 and 3 T. However, this research focuses on low-field MRI, in which magnetic fields of less than 0.15 T are generated by a resistive magnet. Resistive magnets can be used at room-temperature, as opposed to superconducting magnets that have to be cooled with liquid helium. This makes low-field MRI less costly. However, one of the main issues with low-field MRI is the lower signal-to-noise ratio.

2.1.2. The magnetic field gradient system

Three orthogonal gradient coils make up the magnetic field gradient system. A schematic representation of two orthogonal gradient coils is shown in Figure 2.1. Gradient coils are used to generate time-varying magnetic fields which vary linearly in each of the spatial dimensions. One of the main specifications of the gradient system is the gradient strength. The higher the gradient strength, the better. Another important feature is the rate at which the maximum gradient strength can be obtained, known as the rise time. Better gradient systems tend to have shorter rise times. In most clinical imaging systems, the maximum gradient strength that can be attained is 10 mT/m and rise times of approximately 1 ms from 0 to 10 mT/m are considered good.

2.1.3. The RF system

The radio frequency (RF) system consists of two components: a transmitter coil and a receiver coil. The transmitter coil generates a rotating magnetic field, called the B_1-field, which excites the electron spins in the object that is to be imaged. The precessing magnetization is converted into an electrical signal by the receiver coil. Both coils are called RF coils because they resonate at radio frequency. The RF system is required to generate a uniform B_1-field and to have a high detection sensitivity.

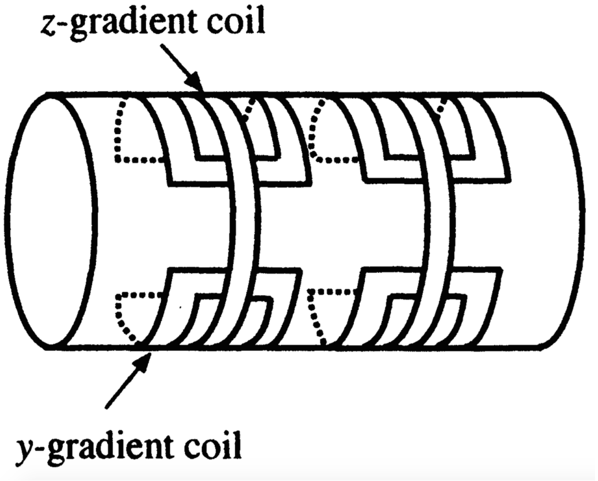


Figure 2.1: Schematic representation of the y-coil and the z-coil that are used to generate the y- and z-gradient.

2.2. Signal generation and detection

2.2.1. Spin

Any physical object consists of atoms, which in turn consist of nuclei and the electrons orbiting around them. Nuclei with odd atomic weights, such as the hydrogen atom, have an angular momentum $\vec{J}$ called spin. Spin can be visualized as a physical rotation, similar to the rotation of a top around its own axis. Because nuclei are electrically charged, each nuclear spin generates its own microscopic magnetic field. This can be represented by the magnetic moment $\vec{\mu}$. $\vec{J}$ and $\vec{\mu}$ are related to each other in the following way:

$$\vec{\mu} = \gamma \vec{J}, \tag{2.1}$$

where γ is the gyromagnetic ratio, the value of which depends on the nucleus in question. For hydrogen atoms, $\gamma = 2.675 \cdot 10^8$ rad(sT)$^{-1}$. In the absence of an external magnetic field, the direction of μ is random, as can be seen in Figure 2.2a. To generate a net magnetic field from the object to be imaged, the spin vectors have to be aligned. This can be done by applying a strong external magnetic field. This field is assumed to be applied in the z-direction:

$$\vec{B} = B_0 \vec{k}, \tag{2.2}$$

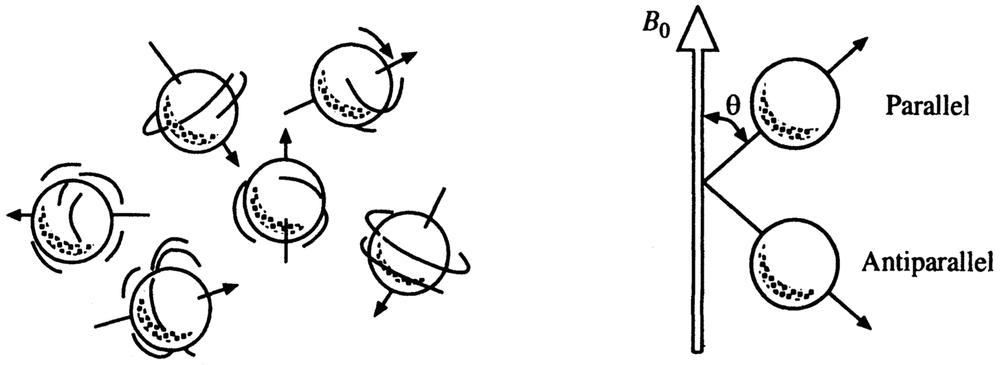
where $\vec{k}$ is the unit vector in the z-direction. For a spin-$\frac{1}{2}$ system, the spin will align either parallel (spin-up) or antiparallel (spin-down) to the magnetic field, as shown in Figure 2.2b. The angular frequency of nuclear precession is described by the Larmor equation

$$\omega_0 = \gamma B_0. \tag{2.3}$$

ω_0 is known as the Larmor frequency.

2.2.2. Bulk magnetization

According to quantum theory, spin-up states have an energy of $E_\uparrow = -\frac{1}{2}\gamma\hbar B_0$ while spin-down states have an energy of $E_\downarrow = \frac{1}{2}\gamma\hbar B_0$, where $\hbar = 1.055 \cdot 10^{-34}$ m^{2}kg/s is the reduced Planck constant. So a parallel alignment is a lower energy state than an antiparallel alignment. Therefore, spins will be more likely to assume the parallel alignment, yielding an observable macroscopic magnetization $\vec{M}$, which is pointed in the same direction as $\vec{B}_0$ and whose magnitude is directly proportional to B_0.



(a) When no external magnetic field is present, the nuclear magnetic moments point in random directions.

(b) In the presence of an external magnetic field, the nuclear magnetic moments align themselves parallel or antiparallel to $\vec{B}_0$.

Figure 2.2: Nuclear magnetic moments in the absence and presence of an external magnetic field.

2.2.3. RF excitations

Planck's law dictates that the energy carried by electromagnetic radiation with frequency ω_{rf} is

$$E_{rf} = \hbar\omega_{rf}. \tag{2.4}$$

In order to make the spins transition from one energy state to another, E_{rf} must be equal to the energy difference between the spin states. Therefore, the following must hold:

$$\hbar\omega_{rf} = E_{\downarrow} - E_{\uparrow} = \gamma\hbar B_0 = \hbar\omega_0, \tag{2.5}$$

which leads to

$$\omega_{rf} = \omega_0. \tag{2.6}$$

Equation (2.6) is the resonance condition.

B_1-field The B_1 field, also known as RF pulse, is a short-lived magnetic field that oscillates in the radio-frequency range. The B_1-field is much weaker than the B_0-field. The B_1-field typically has the following form:

$$\vec{B}_1(t) = 2B_1^e(t)\cos(\omega_{rf}t + \phi)\vec{i}, \tag{2.7}$$

where $B_1^e(t)$ is a the pulse envelope function, ω_{rf} is the excitation carrier frequency and ϕ is the initial phase angle. Two popular envelope functions are the rectangular pulse and the sinc pulse. The B_1-field is linearly polarized because the oscillations occur linearly along the x-axis. Equation (2.7) can be rewritten as

$$\begin{aligned}\vec{B}_1(t) = {} & B_1^e(t)[\cos(\omega_{rf}t + \phi)\vec{i} - \sin(\omega_{rf}t + \phi)\vec{j}] + \\ & B_1^e(t)[\cos(\omega_{rf}t + \phi)\vec{i} + \sin(\omega_{rf}t + \phi)\vec{j}].\end{aligned} \tag{2.8}$$

In (2.8),the first bracketed term rotates clockwise, while the second rotates counterclockwise. The spins rotate clockwise, which means that if the B_1-field has a frequency near the Larmor frequency, the effect of the counterclockwise component is negligible. So the effective B_1-field is described by

$$\vec{B}_1(t) = B_1^e(t)[\cos(\omega_{rf}t + \phi)\vec{i} - \sin(\omega_{rf}t + \phi)\vec{j}] \tag{2.9}$$

with an x-component

$$B_{1,x} = B_1^e(t)\cos(\omega_{rf}t + \phi) \tag{2.10}$$

and a y-component

$$B_{1,y} = -B_1^e(t)\sin(\omega_{rf}t + \phi). \tag{2.11}$$

For brevity, $B_1(t)$ can be described using complex notation:

$$B_1(t) = B_{1,x} + iB_{1,y} = B_1^e(t)e^{-i(\omega_{rf}t + \phi)}. \tag{2.12}$$

If the initial phase ϕ is a constant, it has no significant effect. Therefore, it will be assumed to be 0 from now on.

The Bloch equation The behavior of the net magnetization $\vec{M}$ is governed by the Bloch equation, which, in the context of MRI, has the following form:

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \vec{B} - \frac{M_x \vec{i} + M_y \vec{j}}{T_2} - \frac{(M_z - M_z^{eq})\vec{k}}{T_1}, \tag{2.13}$$

where M_z^{eq} is the thermal equilibrium value for $\vec{M}$ in the presence of $\vec{B}_0$ only. T_1 and T_2 are time constants that characterize the relaxation process of a spin system after it has been excited. In order to describe the excitation effect of an RF pulse, a reference frame that rotates with an angular frequency ω is introduced. In this case, $\omega = \omega_0$, the Larmor frequency. The axes are denoted by x', y' and z' and the unit vectors are $\vec{i}'$, $\vec{j}'$ and $\vec{k}'$. The following transformation relates the rotating frame to the stationary frame:

$$\begin{cases} \vec{i}' = \cos(\omega t)\vec{i} - \sin\omega t\vec{j} \\ \vec{j}' = \sin(\omega t)\vec{i} - \cos\omega t\vec{j} \\ \vec{k}' = \vec{k} \end{cases} \tag{2.14}$$

It can be shown that in this rotating frame, the Bloch equation is given by

$$\frac{d\vec{M}_{rot}}{dt} = \gamma \vec{M}_{rot} \times \vec{B}_{eff} - \frac{M_{x'} \vec{i}' + M_{y'} \vec{j}'}{T_2} - \frac{(M_z - M_z^{eq})\vec{k}'}{T_1}, \tag{2.15}$$

where $\vec{B}_{eff}$, the effective magnetic field that the $\vec{M}$ experiences, is given by

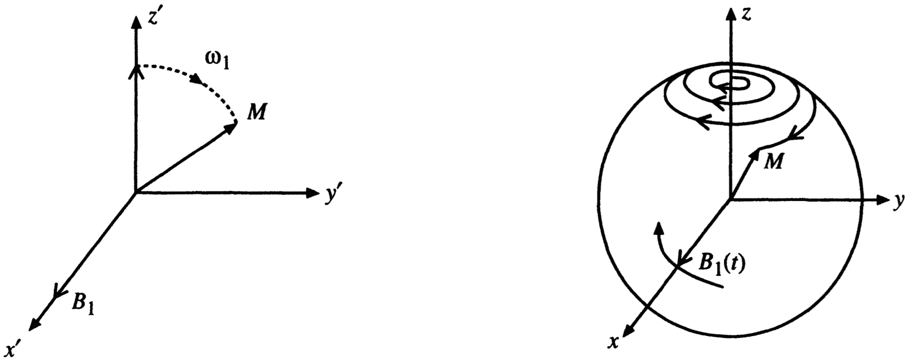
$$\vec{B}_{eff} = \vec{B}_{rot} + \frac{\vec{\omega}}{\gamma}. \tag{2.16}$$

2.2.4. Relaxation

From the Bloch equation it can be shown that the bulk magnetization vector $\vec{M}$ precesses about the x'-axis with angular velocity

$$\vec{\omega}_1 = -\gamma \vec{B}_1, \tag{2.17}$$

as shown in Figure 2.3a.



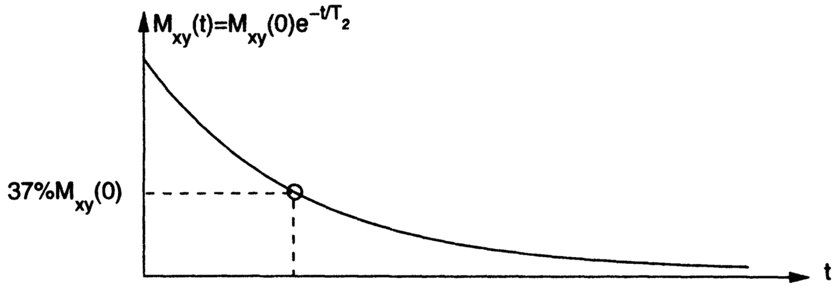
(a) Rotating frame. (b) Stationary frame.

Figure 2.3: Motion of $\vec{M}$ in the presence of a B_1-field.

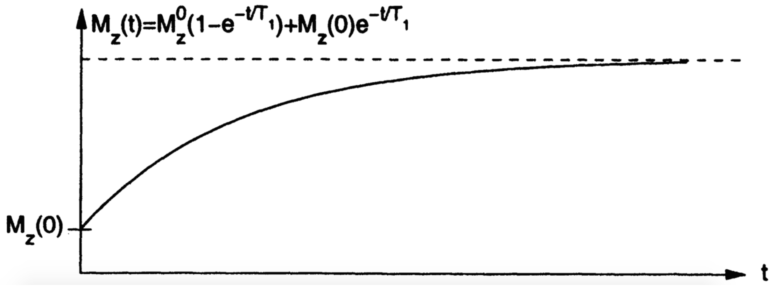
The precession of $\vec{M}$ about the B_1-field is called forced precession, as opposed to the precession of $\vec{M}$ about B_0, which is known as free precession. Forced precession causes the bulk magnetization to tip away from the z'-axis, yielding a measurable transverse component $\vec{M}_{x'y'}$. After the RF pulse has been applied, $\vec{M}$ will return to precessing around the z'-axis. During this return, two relaxation processes occur: longitudinal relaxation and transverse relaxation. Longitudinal relaxation is the recovery of M_z, while transverse relaxation is the destruction of the transverse magnetization M_{xy}. After the RF pulse, $\vec{B}_{eff} = 0$. Therefore, the solution to the Bloch equation looks as follows:

$$\begin{cases} M_{x'y'}(t) = M_{x'y'}(0_+)e^{-t/T_2} \\ M_{z'}(t) = M_z^{eq}\left(1 - e^{-t/T_1}\right) + M_{z'}(0_+)e^{-t/T_1} \end{cases} \tag{2.18}$$

where $M_{x'y'}(0^+)$ and $M_{z'}(0^+)$ are the magnetizations in the transverse plane and along the z'-axis, respectively, immediately after the RF pulse. If all magnetization has been tipped into the transverse plane, $M_{z'}(0^+) = 0$. Figure 2.4 shows plots of $M_{x'y'}$ and $M_{z'}$. $M_{x'y'}(t)$ is characterized by T_2, while $M_{z'}$ is characterized by T_1.



(a) $M_{x'y'}$ after an RF pulse as a function of time.



(b) M_z' after an RF pulse as a function of time.

Figure 2.4: Relaxation curves.

Transforming Equations (2.18) back to the laboratory system yields

$$\begin{cases} M_{xy}(t) = M_{xy}(0_+)e^{-t/T_2}e^{-i\omega_0 t} \\ M_z(t) = M_z^{eq}\left(1 - e^{-t/T_1}\right) + M_z(0_+)e^{-t/T_1} \end{cases} \tag{2.19}$$

where $M_{xy}(0_+) = M_{x'y'}(0_+)e^{-i\omega_0\tau_p}$, with τ_p the duration of the RF pulse. The trajectory of $\vec{M}$ can be seen in Figure 2.5. It should be noted that while $\vec{M}$ approaches the z-axis, its magnitude is not preserved, due to the relaxation process. This is different from the behavior of $\vec{M}$ as it moves away from the z-axis during the excitation period.

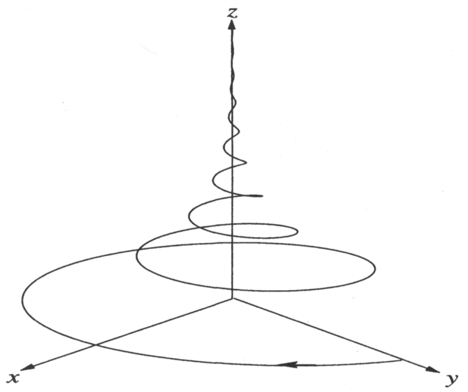


Figure 2.5: The trajectory of $\vec{M}$ in the laboratory frame during the relaxation period.

2.2.5. Signal detection

By Faraday's law of induction, the voltage induced in the receiver coil is described by

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

where $\vec{B}_r(\mathbf{r})$is referred to as the receive field, since it can be interpreted as the magnetic flux density generated by the receive coil carrying a unit current. Note that $\vec{B}_r(\mathbf{r})$ is a weighting vector. After some manipulations, Equation (2.20) can be written as

$$V(t) = \int_{\text{object}} \omega(\mathbf{r})|\vec{B}_{r,xy}(\mathbf{r})||M_{xy}(\mathbf{r},0)|e^{-t/T_2(\mathbf{r})}\cos\left[-\omega(\mathbf{r})t+\phi_e(\mathbf{r})-\phi_r(\mathbf{r})+\frac{\pi}{2}\right]d\mathbf{r}, \tag{2.21}$$

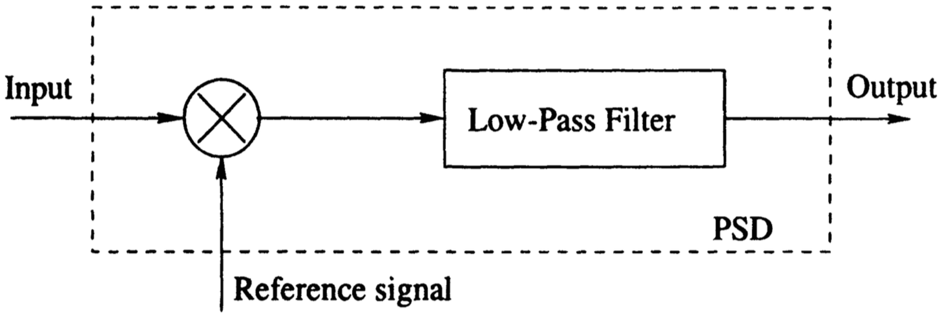
where $\phi_e(\mathbf{r})$ is the initial phase shift induced by RF excitation and $\phi_r(\mathbf{r})$ is the phase of the reception field. Because the Larmor frequency is high, the voltage $V(t)$ is a high-frequency signal, which can cause unnecessary problems in later processing stages. Therefore, in practice, $V(t)$ is moved to a low-frequency band using a phase-sensitive detection method: $V(t)$ is multiplied by the reference sinusoidal signal $2\cos(\omega_0 t)$ and then filtered using a low-pass-filter, effectively removing the high-frequency component. The result will be called $S_R(t)$. $S_R(t)$ can be shown to be described by the following expression:

$$S_R(t) = \int_{\text{object}} \omega(\mathbf{r})|\vec{B}_{r,xy}(\mathbf{r})||M_{xy}(\mathbf{r},0)|e^{-t/T_2(\mathbf{r})}\cos\left[-\Delta\omega(\mathbf{r})t+\phi_e(\mathbf{r})-\phi_r(\mathbf{r})+\frac{\pi}{2}\right]d\mathbf{r}, \tag{2.22}$$

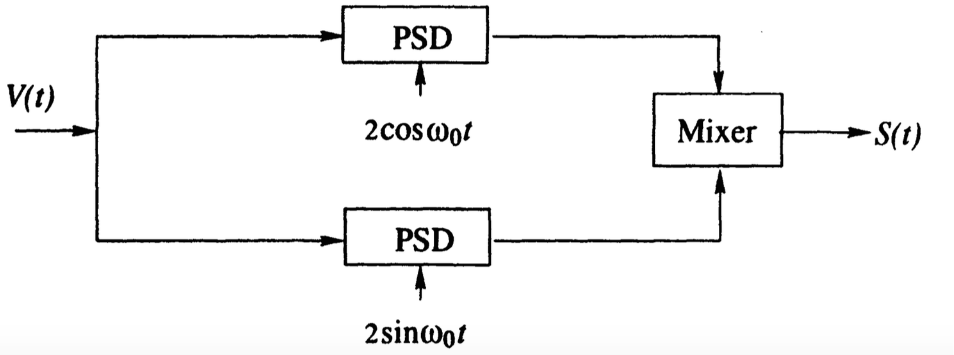
where $\Delta\omega(\mathbf{r}) = \omega(\mathbf{r}) - \omega_0$ is the spatially dependent resonance frequency in the rotating frame. However, now it is not clear whether the precession is clockwise or counterclockwise. Therefore, $V(t)$ is multiplied by $2\sin(\omega_0 t)$ and passed through a low-pass-filter again, yielding $S_I(t)$:

$$S_I(t) = \int_{\text{object}} \omega(\mathbf{r})|\vec{B}_{r,xy}(\mathbf{r})||M_{xy}(\mathbf{r},0)|e^{-t/T_2(\mathbf{r})}\sin\left[-\Delta\omega(\mathbf{r})t+\phi_e(\mathbf{r})-\phi_r(\mathbf{r})+\frac{\pi}{2}\right]d\mathbf{r}. \tag{2.23}$$

In this way, the rotating magnetization is detected with two orthogonal detectors. This procedure is known as quadrature detection and is shown schematically in Figure 2.6.



(a) Phase-sensitive detection



(b) Quadrature detection.

Figure 2.6: Schematic representation of quadrature detection.

The two outputs of the system, $S_R(t)$ and $S_I(t)$ are often put in complex form:

$$S(t) = S_R(t) + iS_I(t). \tag{2.24}$$

Then

$$S(t) = \int_{\text{object}} \omega(\mathbf{r})|\vec{B}_{r,xy}(\mathbf{r})||M_{xy}(\mathbf{r},0)|e^{-t/T_2(\mathbf{r})}e^{-i\left(\Delta\omega(\mathbf{r})t-\phi_e(\mathbf{r})+\phi_r(\mathbf{r})-\frac{\pi}{2}\right)}\,d\mathbf{r}. \tag{2.25}$$

Using

$$\begin{cases} |B_{r,xy}(\mathbf{r})|e^{-i\phi_r(\mathbf{r})} = B^*_{r,xy}(\mathbf{r}) \\ |M_{xy}(\mathbf{r},0)|e^{-i\phi_e(\mathbf{r})} = M_{xy}(\mathbf{r},0) \end{cases} \tag{2.26}$$

and omitting the scaling constant $e^{i\pi/2}$, Equation (2.25) can be rewritten as

$$S(t) = \int_{\text{object}} \omega(\mathbf{r})\vec{B}^*_{r,xy}(\mathbf{r})M_{xy}(\mathbf{r},0)e^{-t/T_2(\mathbf{r})}e^{-i\Delta\omega(\mathbf{r})t}\,d\mathbf{r}. \tag{2.27}$$

Now we assume that $\Delta\omega(\mathbf{r}) << \omega_0$ and assuming that the reception field $B_{r,xy}$ is homogeneous. Then, using $\Delta\omega(\mathbf{r}) = \gamma\Delta B(\mathbf{r})$ and leaving out the constant terms, Equation (2.27) can be simplified to

$$S(t) = \int_{\text{object}} M_{xy}(\mathbf{r},0)e^{-t/T_2(\mathbf{r})}e^{-i\gamma\Delta B(\mathbf{r})t}\,d\mathbf{r}. \tag{2.28}$$

2.3. Signal characteristics

2.3.1. Free induction decays

A single RF pulse applied to a nuclear spin system leads to free induction decay (FID) in the spin system. When the magnetic field the sample is exposed to is perfectly homogeneous, the FID signal is characterized by a T_2 decay. However, when the magnetic field is inhomogeneous, differences in precessional frequency arise, causing the FID signal to decay at a much faster rate. A group of nuclear spins with the same precessional frequency is called an isochromat. The time constant T_2^* is used to characterize the signal decay in the presence of field inhomogeneity.

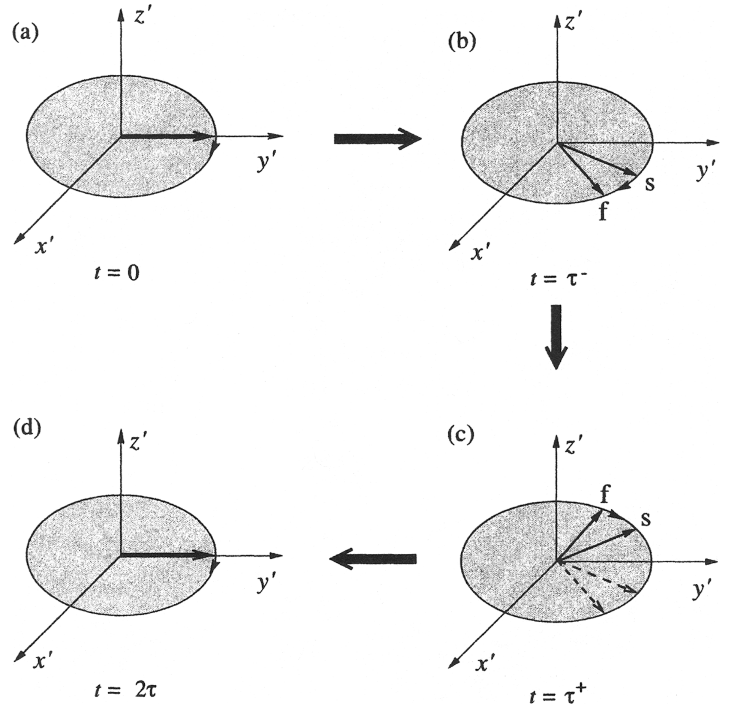


Figure 2.7: Refocusing the bulk magnetic moment of 2 isochromats in a spin-echo experiment.

2.3.2. RF echoes

An echo signal consists of a dephasing period and a refocusing phase. After the RF pulse is applied along the x'-direction (called the 90° pulse), inhomogeneities in the magnetic field will lead to different isochromats. The isochromats will progressively lose phase coherence as the free precession continues. In order to regain phase coherence, after a time τ a pulse is applied along the y'-direction that flips the different magnetic moments over to the other side of the transverse plane (the 180° pulse). This will cause the bulk magnetic moment of isochromats with higher precessional speed to lag behind the ones with lower precessional speed. A time τ after the y'-pulse, the vectors will all have the same phase again. This process is illustrated in Figure 2.7 for two isochromats with precessional frequencies ω_s and ω_f. The strength of the signal during this pulsing sequence is shown in Figure 2.8. The decay of the signal is characterized by T_2^*, while the decay in maximum amplitude is characterized by T_2.

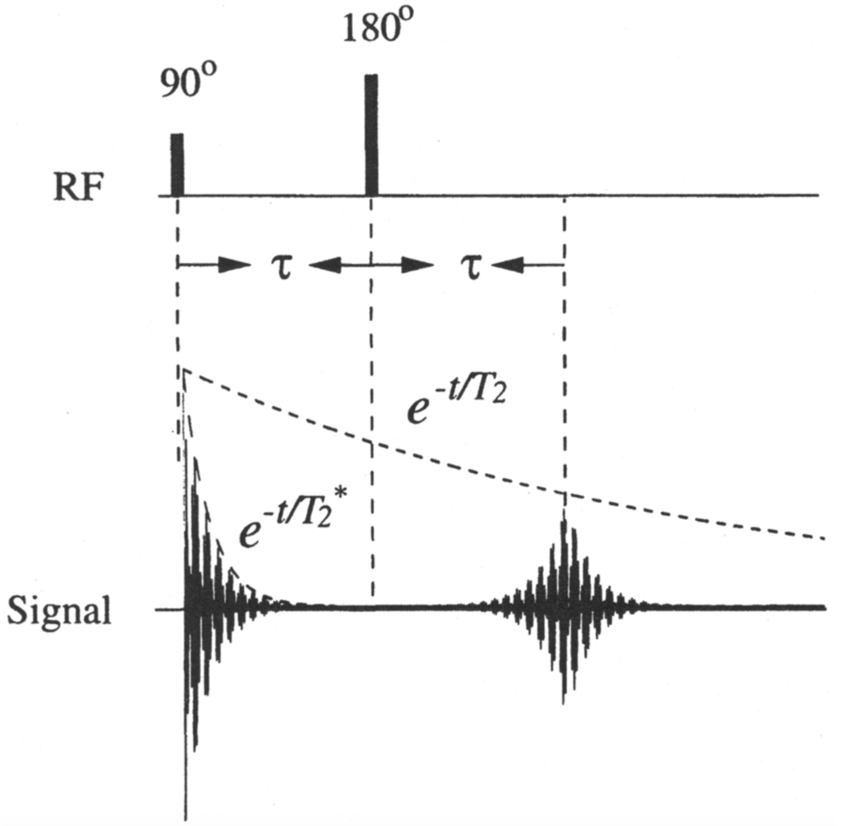


Figure 2.8: Formation of a spin echo signal.

Instead of just one spin echo, a sequence of spin echoes can be generated by repeatedly applying 180° pulses at times τ, 3τ, 5τ,...

2.4. Image contrast

Protons in different materials have different longitudinal and transverse relaxation times, T_1 and T_2. The differences between these parameters can be used to produce contrast in MR images [7]. This is done by tuning the repetition time T_R (the time between two successive B_1- pulses) and the echo time $T_E := 2\tau$ correctly. If T_R is long, all the transverse magnetization will have turned into longitudinal magnetization again. However, if T_R is shorter, a new B_1-pulse is applied before the longitudinal magnetization is restored in all materials, meaning that the transverse magnetization in materials with a long T_1 will be smaller after this pulse than in materials with a short T_1, leading to strong signals from materials with a short T_1 and weak signals from materials with a long T_1: T_1-weighting.

If T_E is short, no significant decay in transverse relaxation will have taken place. However, a long T_E causes materials with a short T_2 to lose their transverse magnetization, while materials with a long T_2 will maintain it. This causes T_2-weighting: materials with a long T_2 yield stronger signals than those with a short T_2.

If T_R is long and T_E is short, only the spin-density will determine the contrast in the image. An overview of how the different kinds of weighting can be achieved is given in Table 2.1.

Table 2.1: Image contrast for different choices of T_E and T_R.

Contrast	T_E	T_R
T_1-weighting	Short	Appropriate
T_2-weighting	Appropriate	Long
Spin density-weighting	Short	Long

2.5. Signal localization

2.5.1. Slice selection

In order to make a 2D image of a 3D object, a slice has to be selected. Then, only the spins in that slice of the object will be excited. To selectively excite spins, a gradient field and a shaped RF pulse are necessary. An RF pulse can only be frequency-selective, which means that the spin resonance frequency has to be made position-dependent. This is achieved by augmenting the B_0-field with a linear gradient field during the excitation period. This gradient field is called a slice-selection gradient. The images are made in the xy-plane, which means that the gradient will be applied in the z-direction, yielding the total magnetic field strength $B(z) = B_0 + G_z z$. Now, to select a slice of thickness Δz centered around z_0, the following spatial selection function is required:

$$p_s(z) = \begin{cases} 1, & |z - z_0| < \frac{\Delta z}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{2.29}$$

It can be shown that the corresponding pulse envelope function has to satisfy

$$B_1^e(t) \propto \operatorname{sinc}\left[\pi \frac{\gamma}{2\pi} G_z \Delta z \left(t - \frac{\tau_p}{2}\right)\right], \tag{2.30}$$

where τ_p is the duration of the pulse.

2.5.2. Spatial information encoding

After the RF pulse, the free precession period allows for spatial information to be encoded into the signal. There are two ways to encode spatial information: frequency encoding and phase encoding.

Frequency encoding When the precession frequency of an activated MR signal is made to be linearly dependent on its spatial origin, frequency encoding is used. This can be achieved by applying a linear gradient field to the magnetic field after the RF pulse. For instance, a gradient G_x in the x-direction leads to a magnetic field

$$\vec{B} = (B_0 + G_x x)\vec{k}, \tag{2.31}$$

which leads to the Larmor frequency being a linear function of x:

$$\omega(x) = \omega_0 + \gamma G_x x. \tag{2.32}$$

Assuming an object with spin distribution $\rho(x, y)$ and omitting the transverse relaxation effect, the FID signal generated locally from spins in an infinitesimal interval dx at point x is

$$dS(x, y, t) \propto \rho(x, y)\ dx\ dy\ e^{-i\gamma(B_0 + G_x x)t}. \tag{2.33}$$

Then, the signal generated by the entire object, neglecting the proportionality constant, is

$$S(t) = \int\int_{\text{object}} dS(x, y, t) = \int\int_{\text{object}} \rho(x, y) e^{-i\gamma(B_0 + G_x x)t}\ dx\ dy \tag{2.34}$$

$$= \left[\int\int_{\text{object}} \rho(x, y) e^{-i\gamma G_x x t}\ dx\ dy\right] e^{-i\omega_0 t}. \tag{2.35}$$

Removal of the carrier signal $e^{-i\omega_0 t}$ (demodulation) yields

$$S(t) = \int\int_{\text{object}} \rho(x, y) e^{-i\gamma G_x x t}\ dx\ dy. \tag{2.36}$$

Phase encoding After an RF pulse, a gradient G_y is turned on in the y-direction for a short time T_{pe}, and then it is turned off. Then the local signal $dS(x, y, t)$ is described by

$$dS(x,y,t) = \begin{cases} \rho(x,y)\ dx\ dy\ e^{-i\gamma(B_0+G_y y)t}, & 0 \le t \le T_{pe} \\ \rho(x,y)\ dx\ dy\ e^{-i\gamma G_y y T_{pe}} e^{-i\gamma B_0 t}, & T_{pe} \le t. \end{cases} \tag{2.37}$$

During the interval $0 \le t \le T_{pe}$, the preparatory period, the signal is frequency-encoded. Therefore, signals from different y-positions will have different phase angles after a time T_{pe}. At time T_{pe}, the signal will have an initial phase angle

$$\phi(y) = -\gamma G_y y T_{pe}. \tag{2.38}$$

$\phi(y)$ is linearly dependent on the position y, so the signal is phase-encoded. Now,

$$S(t) = \int\int_{\text{object}} dS(x,y,t) = \left[\int\int_{\text{object}} \rho(x,y) e^{-i\gamma G_y y T_{pe}}\ dx\ dy\right] e^{-i\omega_0 t}. \tag{2.39}$$

Again, the carrier signal $e^{-i\omega_0 t}$ will be removed after demodulation.

k-space representation Combining a frequency-encoding gradient G_x in the x-direction and a phase-encoding gradient G_y in the y-direction yields (after demodulation)

$$S(t) = \int\int_{object} \rho(x,y) e^{-i(\gamma G_x x t + \gamma G_y y T_{pe})}\ dx\ dy. \tag{2.40}$$

Using the following substitutions:

$$\begin{cases} k_x &= \gamma G_x t \\ k_y &= \gamma G_y T_{pe}, \end{cases} \tag{2.41}$$

the signal is described by

$$S(k_x, k_y) = \int\int_{object} \rho(x,y) e^{-i(k_x x + k_y y)}\ dx\ dy. \tag{2.42}$$

So the substitutions given by Equation (2.41) allow us to use k-space for signal representation. A conventional strategy to produce sufficient data to cover k-space is to generate a set of "identical" signals $\{S_n(t)\}$ and then encode each one properly so that k-space is covered by multiple lines. This can be done by changing the phase-encoding gradient, because a change in gradient strength leads to a different line in k-space.

2.6. Image reconstruction

In practice, the spin density ρ depends not only on x and y, but also on z: $\rho(x,y,z)$. The desired image function will be denoted by $I(x,y)$. $I(x,y)$ is related to the spin density $\rho(x,y,z)$ in the following way:

$$I(x,y) = \int_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} \rho(x,y,z)\ dz. \tag{2.43}$$

Therefore, $\rho(x,y)$ is replaced by $I(x,y)$ in Equation (2.42):

$$S(k_x,k_y) = \int\int_{object} I(x,y) e^{-i2\pi(k_x x + k_y y)}\ dx\ dy. \tag{2.44}$$

Now k_x and k_y are described by

$$k_x = \frac{\gamma}{2\pi} G_x t \tag{2.45}$$

$$k_y = \frac{\gamma}{2\pi} G_y T_{pe}, \tag{2.46}$$

in order to cast Equation (2.44) as a Fourier transform. The aim is to obtain $I(x,y)$ given $S(k_x,k_y)$. Now, assuming that k-space is sampled uniformly, the set of k-space points at which measured data is collected is given by

$$\mathscr{D} = \{(k_{n,x},k_{m,y}) : k_{n,x} = n\Delta k_x, k_{m,y} = m\Delta k_y,\ n,m \in \mathbb{Z}\}. \tag{2.47}$$

The field of view (FOV) is defined as the region $|x| < \frac{W_x}{2}$, $|y| < \frac{W_y}{2}$ where W_x and W_y are finite numbers such that

$$I(x,y) = 0, \quad |x| > \frac{W_x}{2}, |y| > \frac{W_y}{2}. \tag{2.48}$$

Then, if the following inequalities hold:

$$\Delta k_x < \frac{1}{W_x}, \quad \Delta k_y < \frac{1}{W_y}, \tag{2.49}$$

$I(x,y)$ can be shown to be related to $S(x,y)$ in the following way:

$$\begin{aligned} I(x,y) &= \Delta k_x \Delta k_y \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} S(n\Delta k_x, m\Delta k_y) e^{i2\pi(n\Delta k_x x + m\Delta k_y y)}, \\ &|x| < \frac{1}{\Delta k_x}, |y| < \frac{1}{\Delta k_y}. \end{aligned} \tag{2.50}$$

In practice, k-space is sampled a finite number of times. So there are $N, M \in \mathbb{N}$ such that $\mathscr{D}$ is described by

$$\mathscr{D} = \left\{(k_{n,x},k_{m,y}) : k_{n,x} = n\Delta k_x, k_{m,y} = m\Delta k_y,\ -\frac{N}{2} \le n < \frac{N}{2}, -\frac{M}{2} \le m < \frac{M}{2}\right\}. \tag{2.51}$$

Then, the Fourier reconstruction formula is

$$\begin{aligned} I(x,y) &= \Delta k_x \Delta k_y \sum_{n=-N/2}^{N/2-1} \sum_{m=-M/2}^{M/2-1} S(n\Delta k_x, m\Delta k_y) e^{i2\pi(n\Delta k_x x + m\Delta k_y y)}, \\ &|x| < \frac{1}{\Delta k_x}, |y| < \frac{1}{\Delta k_y}. \end{aligned} \tag{2.52}$$

$I(x,y)$ can be obtained easily using a Fast Fourier transform algorithm. Due to the truncation of the Fourier series, the Fourier reconstruction is not identical to the true image function.

3

Low-field MRI

As opposed to conventional MRI, this research focuses on low-field MRI, in which magnetic fields in the order of milliTeslas are generated by a resistive magnet. Resistive magnets can be used at room-temperature, as opposed to superconducting magnets that have to be cooled with liquid helium. This makes low-field MRI less costly. However, one of the main issues with low-field MRI is the lower signal-to-noise ratio [6]. This is due to the increase in SNR with higher B_0 [8].

3.1. The PSU prototype

3.1.1. Magnetic field

Instead of a static and large B_0, the prototype MRI scanner uses a pulsed magnetic field B_p combined with a very small static field B_0. One reason for this choice is that generating a larger static B_0-field generates more heat in the coils, so the higher B_0, the more cooling is necessary, and the more expensive the MRI scanner. The B_p-pulses are called prepolarizing pulses. During a B_p-pulse, the nuclear magnetization will tend towards an enhanced equilibrium value, aligned with the higher effective magnetic field $B_{eff} = B_0 + B_p$. Then, the pulse is ramped down slowly, such that the enhanced magnetization will realign with B_0 without precessing. This enhanced magnetization will decay to the B_0 equilibrium value according to the T_1-value of the sample. If a pulsing sequence is conducted before the magnetization has completely died down, the signal will be enhanced. Therefore, prepolarizing leads to a higher signal-to-noise ratio. The pulse sequence that is employed in the low-field MR scanner is shown in Figure 3.1.

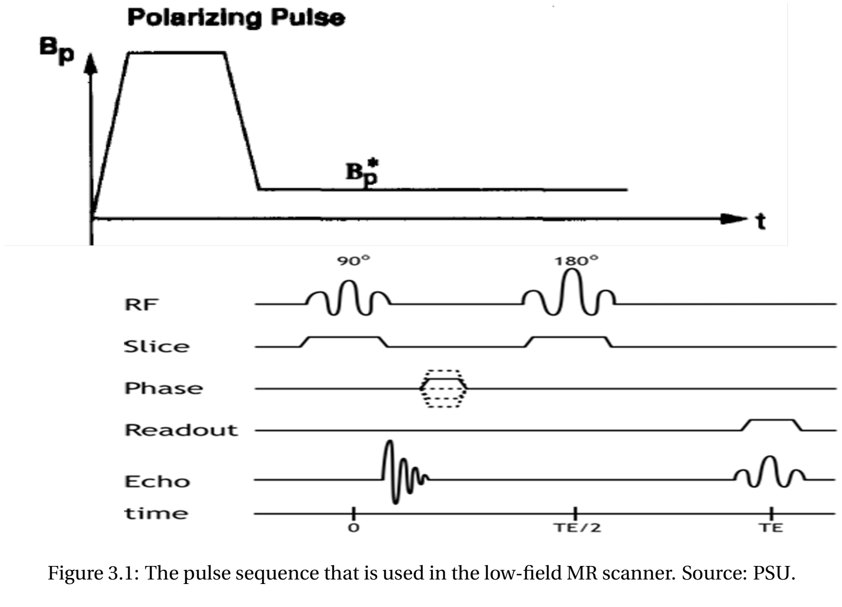


Figure 3.1: The pulse sequence that is used in the low-field MR scanner. Source: PSU.

3.1.2. Field inhomogeneities

In conventional MRI, a Fourier Transform is used to obtain the image given the signal. However, this approach is only possible because of the homogeneity of the B_0-field. In low-field MRI, there is much more inhomogeneity in the magnetic field. Figure 3.2 shows a plot of very coarse measurements of B_0 in the prototype as a function of x and z, where $y = 0$. x and z are defined as in Figure 3.3.

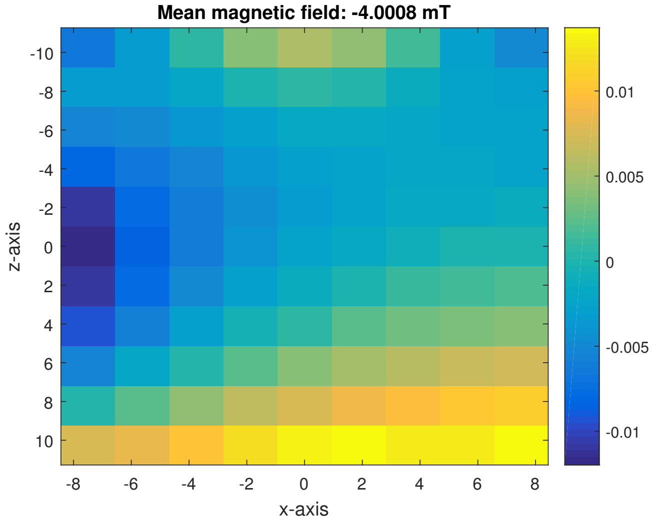


Figure 3.2: Measurements of the B_0-field in the xz-plane, at $y = 0$. The units on the axes are centimeters.

From Figure 3.2, it is clear that the maximum deviation from the mean field is about 0.3%. This does not seem like much, but in conventional MRI the maximum deviation is about a thousand times as small. The PSU team does use the Fourier Transform to obtain the final image.

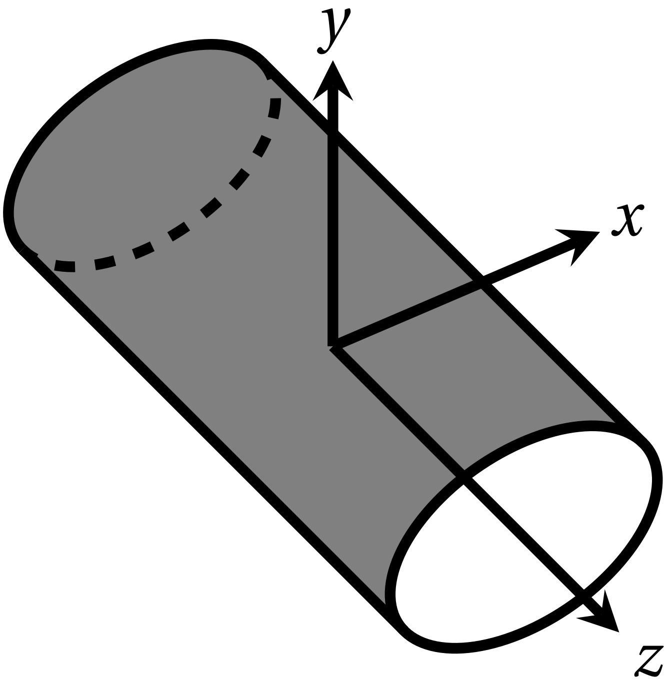


Figure 3.3: x, y and z-direction in the MR scanner.

3.2. The LUMC prototype

A picture of the magnet created by the team at LUMC is shown in Figure 3.4.

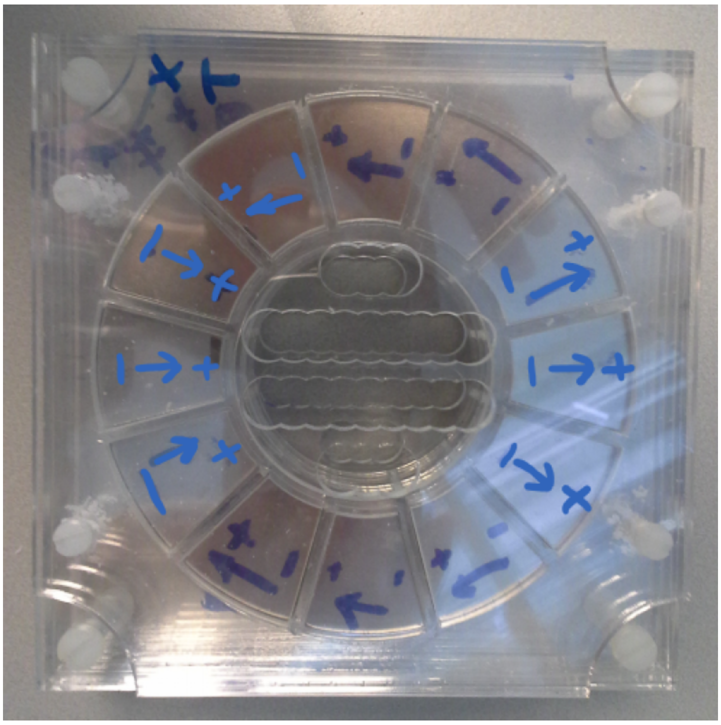
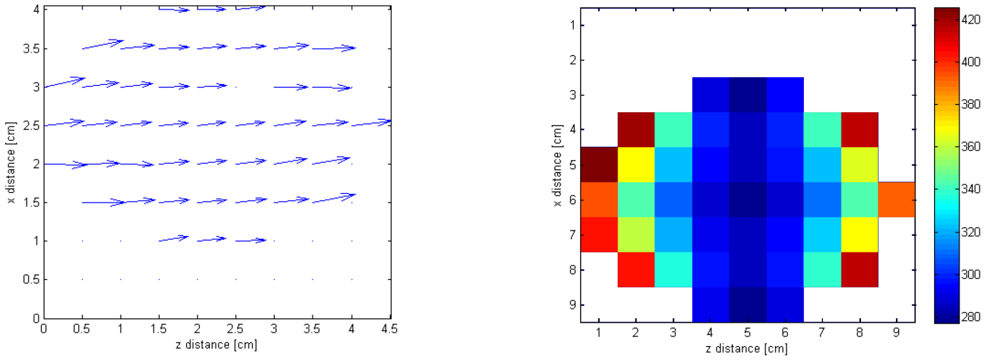


Figure 3.4: The magnet constructed by the team at LUMC. The direction of the magnetization in each magnet is shown using arrows. Source: [9].

They use a configuration of twelve permanent magnets, arranged in a circle, such that the resulting magnetic field (the B_0-field) is oriented mainly in the z-direction. Here, too, the z-axis is oriented along the axis of the cylinder. In [9], the B_0-field was measured at $y = 0$. The result is shown in Figure 3.5.



(a) Direction of the B_0-field.

(b) Magnitude of the B_0-field.

Figure 3.5: B_0-field measurements.

To excite the spins, an RF pulse has to be sent out. In this prototype, the inhomogeneities in the B_0-field are used to allow for spatial encoding, instead of gradient coils. To obtain a sufficient amount of data, the idea is to rotate either the object or the device and measure the signal after each rotation.

3.3. General measurement model

Taking away the assumption of homogeneity, the signal is described by (2.27). Defining $\zeta(\mathbf{r}) = \phi_e(\mathbf{r}) - \phi_r(\mathbf{r})$, Equation (2.27) can be rewritten:

$$S(t) = \int_{\text{object}} \omega(\mathbf{r})|\vec{B}_{r,xy}(\mathbf{r})||M_{xy}(\mathbf{r},0)|e^{-t/T_2(\mathbf{r})}e^{i\zeta(\mathbf{r})}e^{-i\gamma\Delta B(\mathbf{r})t}\, d\mathbf{r}. \tag{3.1}$$

Now, denoting the signal sensitivity response pattern of the coil by $c(\mathbf{r})$, the general forward model for the signal is:

$$S(t) = \int_{\text{object}} \omega(\mathbf{r})c(\mathbf{r})|\vec{B}_{r,xy}(\mathbf{r})||M_{xy}(\mathbf{r},0)|e^{-t/T_2(\mathbf{r})}e^{i\zeta(\mathbf{r})}e^{-i\phi(\mathbf{r},t)}\, d\mathbf{r}, \tag{3.2}$$

where

$$\phi(\mathbf{r},t) = \gamma\Delta B(\mathbf{r})t. \tag{3.3}$$

Here, $M_{xy}(\mathbf{r},0)$ is the desired image. For ease of notation, we define

$$x(\mathbf{r}) := M_{xy}(\mathbf{r},0). \tag{3.4}$$

The measurements recorded in an MRI scan consist of noisy samples of the MRI signal described by Equation (3.2):

$$y_i = S(t_i) + v_i, \quad i = 1,...,L, \tag{3.5}$$

where y_i denotes the ith sample of the signal, measured at time t_i. v_i is the measurement error in the ith sample. v_i are modeled by additive, zero-mean, white gaussian noise. Using Equation (3.2) to describe the measurement model, the image reconstruction problem is to estimate $x(\mathbf{r})$ from a measurement vector $\mathbf{y}$. $x(\mathbf{r})$ is a continuous function. In order to estimate it using a finite set of measurements, $x(\mathbf{r})$ is approximated using a finite series expansion:

$$x(\mathbf{r}) = \sum_{j=1}^{N} x_j b(\mathbf{r}-\mathbf{r}_j), \tag{3.6}$$

where $b(\cdot)$ denotes the object basis function and $\mathbf{r}_j$ is the center of the jth basis function. Usually, rectangular basis functions are used. In that case, N is the number of pixels. Substituting Equation (3.6) into Equation (3.2) yields

$$S(t_i) = \sum_{j=1}^{N} a_{ij}x_j, \tag{3.7}$$

where

$$a_{ij} = \int_{\text{object}} b(\mathbf{r}-\mathbf{r}_j)c(\mathbf{r})\omega(\mathbf{r})|\vec{B}_{r,xy}(\mathbf{r})|e^{-t_i/T_2(\mathbf{r})}e^{\iota\zeta(\mathbf{r})}e^{-\iota\Delta\phi(\mathbf{r},t)}\, d\mathbf{r}. \tag{3.8}$$

Usually, the basis functions are highly localized, allowing 'center of pixel' approximations to be used:

$$a_{ij} = c(\mathbf{r}_j)\omega(\mathbf{r}_j)|\vec{B}_{r,xy}(\mathbf{r}_j)|e^{-t_i/T_2(\mathbf{r}_j)}e^{\iota\zeta(\mathbf{r})}e^{-\iota\Delta\phi(\mathbf{r}_j,t_i)}, \quad \phi(\mathbf{r}_j,t_i) = \gamma\Delta B(\mathbf{r}_j)t_i. \tag{3.9}$$

Combining Equations (3.5), (3.6) and (3.8) allows us to represent the system of equations in matrix-vector form:

$$\mathbf{y} = A\mathbf{x} + \mathbf{v}. \tag{3.10}$$

4

Super-resolution reconstruction

SRR uses multiple low-resolution (LR) images of the same object to form a high-resolution (HR) image. Typically, the different LR images represent different "looks" at the same object. That is, LR images are down-sampled and shifted versions of the HR image. The shift has to be a subpixel shift, because integer shifts yield no new information, as shown in Figure 4.1. Another way of acquiring different LR images would be to rotate the object.

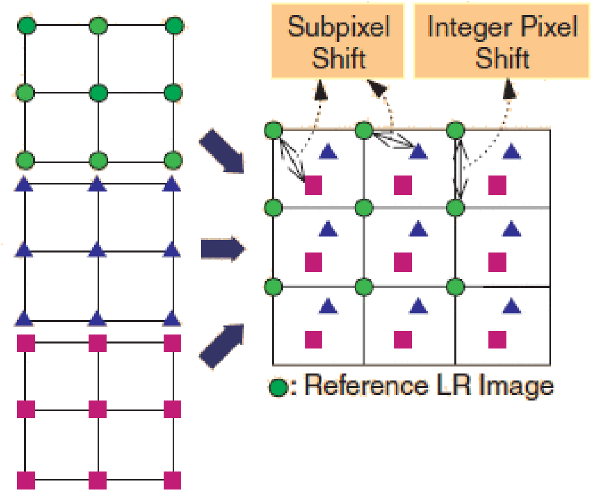


Figure 4.1: Integer shifts do not provide any new information. Subpixel shifts are needed to make SRR possible. Source: [10].

4.1. Data acquisition model

Let X be the desired HR image with dimensions $m \times n$. Then X is an $m \times n$ matrix consisting of the grayscale values of each pixel. In order to apply super-resolution reconstruction, X must be put in vector form:

$$\mathbf{x} = \mathrm{vec}(X) = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}, \tag{4.1}$$

where $\mathbf{x}_i$ is the ith column of X. This notation will be used throughout the whole report, so images in matrix form will be denoted by capital letters and images in vector form will be denoted by the same letter, but in bold lowercase.

Let $\{Y_k\}_{k=1}^{N}$ be a sequence of LR images. The data acquisition model used in SRR is as follows [10], [5]. First, the HR image X undergoes a geometric deformation, such as a subpixel shift or a rotation. This geometric deformation is described by an operator D_k. Then, during the imaging process, each of the geometrically deformed images is subject to a blurring operator B_k and a down-sampling operator denoted by D_k. Also, noise is added to the system: V_k. The noise is usually assumed to be independent, identically distributed Gaussian noise. A pictorial overview of the acquisition model is shown in Figure 4.2.

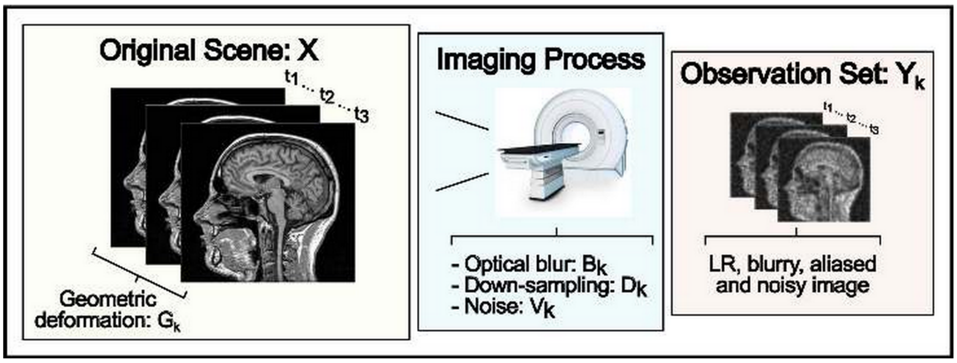


Figure 4.2: The general data acquisition model used in super-resolution reconstruction. Source: [5].

G_k, B_k and D_k can be cast in matrix-form when $\mathbf{x}$, $\mathbf{y}_k$ and $\mathbf{v}_k$ are used instead of X, Y_k and V_k, yielding the following system of equations:

$$\mathbf{y}_k = D_k B_k G_k \mathbf{x} + \mathbf{v}_k, \quad k = 1, ..., N. \tag{4.2}$$

Clearly, the term $D_k B_k G_k$ can be replaced by a single matrix A_k:

$$\mathbf{y}_k = A_k \mathbf{x} + \mathbf{v}_k, \quad k = 1, ..., N. \tag{4.3}$$

By vertically concatenating the vectors $\mathbf{y}_k$, the matrices A_K and the vectors $\mathbf{v}_k$ as follows:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}, \quad A = \begin{pmatrix} A_1 \\ \vdots \\ A_N \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_N \end{pmatrix}, \tag{4.4}$$

a single system of equations is obtained:

$$\mathbf{y} = A\mathbf{x} + \mathbf{v}. \tag{4.5}$$

4.1.1. Geometric deformation

Geometric deformation, represented by G_k, is of fundamental importance in SRR, because it yields different views of the same object, providing additional information. As mentioned before, the deformation has to be of subpixel nature, which requires a very accurate estimation of the deformation. It can be difficult to attain such a high precision, which makes registration a bottleneck in reconstruction process. By introducing predetermined motion, this difficulty can be eliminated. When imaging a static object, the deformation is artificially created by shifting or rotating the field of view by a known value [5].

4.1.2. Blurring

The operator B_k represents the blurring caused by the imaging process. Commonly, blurring is assumed to be spatially-invariant, so B_k simplifies to B. The point-spread function (PSF) P is a function that describes the blurring of one pixel over its surrounding pixels. Usually, blurring is a local phenomenon, so one pixel influences only a very limited amount of other pixels around it. Outside a certain radius, its influence is essentially zero. Therefore, P is a very small matrix compared to the image matrices. Let Z be an image and let Z_b be the blurred version of Z. If the PSF is assumed to be spatially invariant, the blurring of Z can be represented by a two-dimensional convolution:

$$Z_b = P ** Z. \tag{4.6}$$

Most articles concerning SRR in MRI suggest that the PSF is well approximated by a Gaussian function. In the 2D case, the (unscaled) elements of the PSF are described by

$$p_{ij} = \exp\left(-\frac{1}{2} \begin{bmatrix} i-k \\ j-l \end{bmatrix}^T \begin{bmatrix} s_1^2 & \rho^2 \\ \rho^2 & s_2^2 \end{bmatrix} \begin{bmatrix} i-k \\ j-l \end{bmatrix}\right), \tag{4.7}$$

where s_1 and s_2 determine the width of the PSF and ρ determines its orientation. (k,l) is the central element of P. All elements of a PSF have to sum to 1, so P has to be scaled accordingly. The Gaussian function decays exponentially away from the center and it is reasonable to truncate the values of the PSF when they have decayed by a sufficiently large factor ($10^4 - 10^8$) [11]. If $\rho = 0$, the PSF is symmetrical along the horizontal and vertical axes and the PSF takes the simpler form

$$p_{ij} = \exp\left(-\frac{1}{2}\left(\frac{i-k}{s_1}\right)^2 - \frac{1}{2}\left(\frac{j-1}{s_2}\right)^2\right). \tag{4.8}$$

The blurring matrix B_k can be obtained from P. The blurred image is equal to the convolution of P and Z, so pixel (i,j) of Z_b can be computed by rotating P by 180 degrees and matching it with pixels in the image Z by placing the center of P over pixel (i,j) of Z. Corresponding components are multiplied and the results are summed to compute element (i,j) of Z_b. This can be carried out for all elements in Z_b, leading to the blurring matrix B.

4.2. Minimization problem

4.2.1. Least-squares solution

SRR aims to solve Equation (4.5) for $\mathbf{x}$. However, due to noise (and possibly an insufficient number of LR images), the problem is ill-posed. In order to obtain an $\mathbf{x}$ that describes the data well, $||\mathbf{y} - A\mathbf{x}||_2^2$ has to be small. $||\cdot||_2$ denotes the ℓ_2-norm. If A^*A is nonsingular, the least-squares solution, so the solution that minimizes $||\mathbf{y} - A\mathbf{x}||_2^2$, is

$$\mathbf{x}_{\text{ls}} = (A^*A)^{-1}A^*\mathbf{y}. \tag{4.9}$$

$\mathbf{x}_{\text{ls}}$ is equal to

$$\mathbf{x}_{\text{ls}} = (A^*A)^{-1}A^*\mathbf{y} = \mathbf{x} + (A^*A)^{-1}A^*\mathbf{v}. \tag{4.10}$$

The last term is called the inverted noise. We can rewrite A^*A using its singular value decomposition:

$$A^*A = U\Sigma G^*, \tag{4.11}$$

where U and G are orthogonal matrices and $\Sigma = \text{diag}(\sigma_i)$ is a diagonal matrix whose elements σ_i are non-negative and appear in nonincreasing order. The σ_i are the singular values of A^*A. If A^*A is nonsingular, its inverse can be written as

$$(A^*A)^{-1} = G\Sigma^{-1}U^*. \tag{4.12}$$

So the inverted noise is equal to

$$(A^*A)^{-1}A^*\mathbf{v} = G\Sigma^{-1}U^*A^*\mathbf{y} = \sum_i \frac{\mathbf{u}_i^*A^*\mathbf{v}}{\sigma_i}\mathbf{g}_i, \tag{4.13}$$

where $\mathbf{u}_i$ and $\mathbf{g}_i$ are the ith column vectors of U and G respectively. If A^*A is almost singular, its smallest singular values will be very close to 0, causing the inverted noise to become large, contaminating the reconstructed image.

4.2.2. Tikhonov regularization

In order to obtain a solution that is not corrupted by noise, a Tikhonov regularization term can be added to the problem, which leads to the following minimization problem:

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2 + \frac{1}{2}\lambda||F\mathbf{x}||_2^2, \tag{4.14}$$

where λ is the regularization parameter and F is an operator. The term $||F\mathbf{x}||_2^2$ allows us to enforce prior information that is available about $\mathbf{x}$. For instance, F can be chosen to be a first-order difference matrix, because in MR images it is very likely that neighboring pixels have the same value. The value of λ determines to what extent this regularization term is taken into account. A small λ will cause more emphasis to be placed on obtaining a solution that fits the data well, i.e. make sure that the term $\frac{1}{2}||\mathbf{y}-A\mathbf{x}||_2^2$ is small, while a large λ will ensure that the prior information is enforced, so $\frac{1}{2}||F\mathbf{x}||_2^2$ will be small.

It is not immediately clear which value of λ leads to the best solution. One way of choosing λ is by employing the L-curve criterion [12]. The L-curve is a log-log plot of $||\mathbf{y}-A\mathbf{x}||_2$ versus $||F\mathbf{x}||_2$ for a number of values of λ. This plot often has an L-shape and the best regularization parameter is supposed to lie in the corner of the L, balancing out the two terms.

Equation (4.14) is a convex problem, which means that taking the gradient and setting it equal to 0 yields a condition for the global optimal solution:

$$(A^*A+\lambda F^*F)\mathbf{x}=A^*\mathbf{y}, \tag{4.15}$$

leading to

$$\mathbf{x}=(A^*A+\lambda F^*F)^{-1}A^*\mathbf{y}. \tag{4.16}$$

4.2.3. Total variation regularization

Tikhonov regularization is not the only possible way of regularizing the problem. Another popular choice is the ℓ_1-term where F is a first-order difference matrix, defined such that

$$||F\mathbf{x}||_1=\sum_{k=1}^{n}\sum_{l=2}^{m}|X_{l,k}-X_{l-1,k}|+\sum_{l=1}^{m}\sum_{k=2}^{n}|X_{l,k}-X_{l,k-1}|. \tag{4.17}$$

So the minimization problem becomes

$$\min_{\mathbf{x}}\frac{1}{2}||\mathbf{y}-A\mathbf{x}||_2^2+\frac{1}{2}\lambda||F\mathbf{x}||_1. \tag{4.18}$$

The $||F\mathbf{x}||_1$ term penalizes jumps between neighboring pixels. However, jumps are not penalized as much as in the Tikhonov case, because the ℓ_2-norm makes the regularization term grow quadratically with the difference in value between neighboring pixels, while using the ℓ_1-norm ensures that the penalization grows only linearly. Penalizing large jumps very harshly is undesirable, because MR images tend to contain large jumps between different tissues. The form of regularization used in Equation (4.18) (with F as defined in Equation (4.17)) is called total variation regularization.

4.2.4. Edge-preserving regularization

When Tikhonov regularization is used, the penalty term grows quadratically with the difference between pixels. The good thing about this is that neighboring pixels are very likely to have the same value, which is desirable in MR images. However, jumps between pixels of different tissues will be severely penalized, leading to overly smooth images. In that case, total variation minimization yields a better result. Edge-preserving regularization combines the advantages of both types of regularization, such that up to a certain value, discontinuities are penalized in a quadratic way, while larger jumps are penalized in a linear way. The rest of this chapter is based on [13].

MAP estimation Maximum a posteriori (MAP) estimation is another technique commonly used to estimate the original image X given a degraded image Y (or in our case, several degraded images). A MAP technique maximizes the conditional probability of $\mathbf{x}$ when $\mathbf{y}$ is given:

$$\hat{\mathbf{x}}=\arg\max_{\mathbf{x}}\,[\log P(\mathbf{x}|\mathbf{y})]. \tag{4.19}$$

Using Bayes' formula and eliminating constant terms, this can be rewritten as

$$\hat{\mathbf{x}}=\arg\min_{\mathbf{x}}\,[-\log P(\mathbf{y}|\mathbf{x})-\log P(\mathbf{x})]. \tag{4.20}$$

Because the noise is assumed to be independent, identically distributed Gaussian, the probability density of $\mathbf{y}$ given $\mathbf{x}$ is

$$P(\mathbf{y}|\mathbf{x})=\frac{1}{(2\pi\sigma^2)^{mn/2}}\exp\left(\frac{-||\mathbf{y}-A\mathbf{x}||^2}{2\sigma^2}\right) \tag{4.21}$$

where σ is the standard deviation of the noise.

Huber-Markov random field model We suppose that our image X can be modeled as a Markov random field, meaning that the conditional distribution of a pixel value, given all the other pixel values, is only dependent on its neighbors. So, defining the set of neighbors of pixel s by ∂s, we have [14]

$$P(x_s|x_{r\neq s}) = P(x_s|x_{\partial s}). \tag{4.22}$$

Next, we assume that $P(\mathbf{x})$ is a Gibbs distribution:

$$P(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{1}{\mu}\sum_{c\in\mathscr{C}}\rho(\mathbf{h}_c^*\mathbf{x})\right), \tag{4.23}$$

where Z is a normalizing constant, μ is the temperature parameter and $\mathbf{h}_c$ is the coefficient vector for the group of pixels c. $\rho(\cdot)$ is a function that has to satisfy the following properties: convexity, symmetry and $\rho(t) << t^2$ for large $|t|$. The Huber function, defined as

$$\rho_T(t) = \begin{cases} t^2, & |t| \leq T \\ T^2 + 2T(|t| - T), & |t| > T \end{cases} \tag{4.24}$$

satisfies these properties. In Figure 4.3, the Huber function is plotted as a function of t:

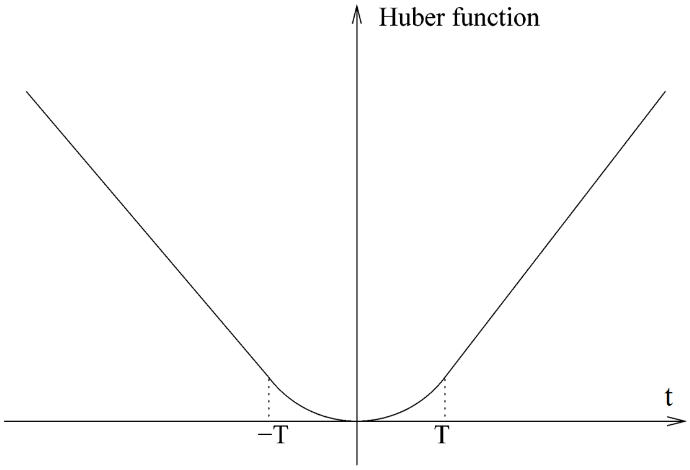


Figure 4.3: The Huber function.

In [13], second-order approximations are used to measure image roughness:

$$\mathbf{h}_{k,l,0}^*\mathbf{x} = X_{k,l+1} - 2X_{k,l} + X_{k,l-1} \tag{4.25}$$

$$\mathbf{h}_{k,l,1}^*\mathbf{x} = \frac{1}{2}(X_{k-1,l+1} - 2X_{k,l} + X_{k+1,l-1}) \tag{4.26}$$

$$\mathbf{h}_{k,l,2}^*\mathbf{x} = X_{k-1,l} - 2X_{k,l} + X_{k+1,l} \tag{4.27}$$

$$\mathbf{h}_{k,l,3}^*\mathbf{x} = \frac{1}{2}(X_{k-1,l-1} - 2X_{k,l} + X_{k-1,l-1}), \tag{4.28}$$

leading to the following expression for $-\log P(\mathbf{x})$:

$$-\log P(\mathbf{x}) = -\log Z + \frac{1}{\mu}\sum_k\sum_l\sum_{m=0}^{3}\rho_T(\mathbf{h}_{k,l,m}^*\mathbf{x}). \tag{4.29}$$

Now, the functional to be minimized is given by

$$M_T(\mathbf{x}) = ||\mathbf{y} - A\mathbf{x}||^2 + \frac{2\sigma^2}{\mu}\sum_k\sum_l\sum_{m=0}^{3}\rho_T(\mathbf{h}_{k,l,m}^*\mathbf{x}). \tag{4.30}$$

Majorization Majorization is an iterative technique that minimizes a function by minimizing its majorizing function. A function $g(\theta)$ is said to majorize $f(\theta)$ at θ_i if

$$f(\theta_i) = g(\theta_i) \tag{4.31}$$

$$f(\theta) \leq g(\theta) \text{ for all } \theta. \tag{4.32}$$

In order to minimize Equation (4.30), a series of functionals $N_T^i(\mathbf{x})$ is defined such that

$$M_T(\mathbf{x}^i) = N_T^{i+1}(\mathbf{x}^i) \tag{4.33}$$

$$M_T(\mathbf{x}) \leq N_T^{i+1}(\mathbf{x}) \quad \text{for all } \mathbf{x}, \tag{4.34}$$

where $\mathbf{x}^i$ is the value at which $N_T^i(\mathbf{x})$ attains its minimum. This algorithm can be shown to converge. To majorize the Huber function in Equation (4.24), we define

$$\tilde{N}_T^{i+1}(t) = \begin{cases} t^2, & |t| \leq T \\ \frac{T}{|t^i|} t^2 + T|t^i| - T^2, & |t| > T. \end{cases} \tag{4.35}$$

Here, t^i is the minimizing point of $\tilde{N}_T^i(t)$. Note that $\tilde{N}_T^i(t)$ is a quadratic function. Now let H_m denote the operator corresponding to $\sum_k \sum_l \mathbf{h}_{k,l,m}^* \mathbf{x}$. The constant term $T|t^i| - T^2$ can be ignored, leaving the modified function

$$\hat{N}_T^{i+1}(\mathbf{x}) = \sum_{m=0}^{3} (H_m \mathbf{x})^* \Gamma_m H_m \mathbf{x}, \tag{4.36}$$

where Γ_m are diagonal matrices whose elements are equal to either 1 (if $\mathbf{h}_{k,l,m}^* \mathbf{x} \leq T$) or $\frac{T}{|\mathbf{h}_{k,l,m}^* \mathbf{x}|}$ (if $\mathbf{h}_{k,l,m}^* \mathbf{x} > T$). Then, the functional $N_T^{i+1}(\mathbf{x})$ is defined as

$$N_T^{i+1}(\mathbf{x}) = ||\mathbf{y} - A\mathbf{x}||^2 + \lambda \sum_{m=0}^{3} (H_m \mathbf{x})^* \Gamma_m H_m \mathbf{x}, \tag{4.37}$$

where $\lambda = \frac{2\sigma^2}{\mu}$ can be seen as a regularizing constant. Taking the gradient of Equation (4.37) yields an optimality criterion for $\mathbf{x}^{i+1}$:

$$\left(A^T A + \lambda \sum_{m=0}^{3} H_m^* \Gamma_m H_m \right) \mathbf{x}^{i+1} = A^T \mathbf{y}. \tag{4.38}$$

Of course, we do not know the values contained in Γ_m beforehand. In order to obtain the correct matrix, fixed-point iteration is used.

5

Conjugate gradient methods

As seen in the previous chapter,

$$(A^*A + \lambda F^*F)\mathbf{x} = A^*\mathbf{y}, \tag{5.1}$$

provides us with an optimality criterion for the minimization problem (4.14). The solution to (5.1) is

$$\mathbf{x} = (A^*A + \lambda F^*F)^{-1}A^*\mathbf{y}. \tag{5.2}$$

When X is an image with $n \times m$ pixels, $A^*A + \lambda F^*F$ is an $n^2m^2 \times n^2m^2$ matrix. That means that when X is a high-resolution image, $A^*A + \lambda F^*F$ becomes very large, making it computationally infeasible to execute a matrix inversion. Therefore, we will rely on iterative solvers to obtain the solution to Equation (5.1). The methods of choice will be conjugate gradient methods. The remainder of this chapter is based on [15] and [16].

5.1. General problem statement

5.1.1. Tikhonov regularization

The minimization problem (4.14) is a special form of

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||^2_{C^{-1}} + \frac{1}{2}\lambda||\mathbf{x}||^2_R, \tag{5.3}$$

where $R = F^*F$ and C is the covariance matrix of the noise. Now, the optimality criterion is represented by the normal equations:

$$(A^*C^{-1}A + \lambda R)\mathbf{x} = A^*C^{-1}\mathbf{y}. \tag{5.4}$$

Note that minimization problem (5.3) can be formulated as the following constrained minimization problem:

$$\min_{\mathbf{r},\mathbf{x}} \frac{1}{2}||\mathbf{r}||^2_C + \frac{1}{2}\lambda||\mathbf{x}||^2_R \tag{5.5}$$
$$\text{subject to } \mathbf{r} = C^{-1}(\mathbf{y} - A\mathbf{x}).$$

Using the technique of Lagrange multipliers, we find that

$$\mathbf{r} = C^{-1}(\mathbf{y} - A\mathbf{x}), \quad \lambda R\mathbf{x} = A^*\mathbf{r}. \tag{5.6}$$

If λR is invertible, $\mathbf{x}$ can be eliminated, yielding

$$\left(\frac{1}{\lambda}AR^{-1}A^* + \lambda C\right)\mathbf{r} = \mathbf{y} \tag{5.7}$$

and $\mathbf{x}$ can be obtained from $\mathbf{r}$:

$$\mathbf{x} = \frac{1}{\lambda}R^{-1}A^*\mathbf{r}. \tag{5.8}$$

In order to be able to work with $\lambda = 0$, we can define $\mathbf{z} := \frac{1}{\lambda}\mathbf{r}$, yielding the system

$$(AR^{-1}A^* + \lambda C)\mathbf{z} = \mathbf{y}, \quad \mathbf{x} = R^{-1}A^*\mathbf{z}. \tag{5.9}$$

Since minimization problems (5.3) and (5.6) are equivalent, Equations (5.4) and (5.9) are equivalent too.

5.1.2. ℓ_p-norm regularization

Now, consider the form

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||^2_{C^{-1}} + \frac{1}{2}\lambda||F\mathbf{x}||^p_p, \tag{5.10}$$

where $||\cdot||_p$ denotes the ℓ_p-norm. Equation (5.10) is not differentiable for all values of p. However, that does not mean that it cannot be solved using the same methods as the ℓ_2-norm case. Additionally, a fixed-point iteration scheme is necessary. Note that $||F\mathbf{x}||^p_p$ can be rewritten as:

$$||F\mathbf{x}||^p_p = \sum_i |(F\mathbf{x})_i|^p = \sum_i |(F\mathbf{x})_i|^{p-2}(F\mathbf{x})^2_i = \mathbf{x}^* F^* W F\mathbf{x}, \tag{5.11}$$

where $W = \text{diag}(|(F\mathbf{x})_i|^{p-2})$ and $(F\mathbf{x})_i$ denotes the ith element of the vector $F\mathbf{x}$. So minimization problem (5.10) can be cast in the form of (5.3) by setting $R = F^* W F$. Now, if an estimate $\mathbf{x}_k$ is available, a better estimate $\mathbf{x}_{k+1}$ can be found by solving (5.10) with $R = F^* W F$ and $W = \text{diag}(|(F\mathbf{x}^k)_i|^{p-2})$.

5.2. Standard conjugate gradient

The conjugate gradient method was developed by Hestenes and Stiefel [17] in 1952. It is a Krylov subspace method for solving systems of the form $K\mathbf{u} = \mathbf{f}$, where K is a square Hermitian positive definite matrix. Starting with an initial estimate $\mathbf{u}_0$ of the solution $\mathbf{u}$, new estimates $\mathbf{u}_1, \mathbf{u}_2, \ldots$ of $\mathbf{u}$ are determined, with $\mathbf{u}_{k+1}$ being closer to $\mathbf{u}$ than $\mathbf{u}_k$. The search directions are denoted by $\mathbf{p}_k$. At each step, the residual $\mathbf{s}_k$ is computed:

$$\mathbf{s}_k = \mathbf{f} - K\mathbf{u}_k. \tag{5.12}$$

In each iteration, the conjugate gradient method computes $\mathbf{u}_k$ such that

$$||\mathbf{u} - \mathbf{u}_k||_K = \min_{\substack{\mathbf{v}\in\mathbf{u}_0+ \\ \text{span}\{\mathbf{p}_0,\ldots,\mathbf{p}_{k-1}\}}} ||\mathbf{u} - \mathbf{v}||_K. \tag{5.13}$$

Given an iterate $\mathbf{u}_k$, the gradient descent method would use $\mathbf{s}_k$ as a search vector. The conjugate gradient method uses search vectors that are conjugate with respect to K (so $\mathbf{p}_k K \mathbf{p}_l = 0$ for $k \neq l$). Then, $P := \{\mathbf{p}_0, \ldots, \mathbf{p}_{n-1}\}$ forms a basis for $\mathbb{R}^n$, which means that $\mathbf{u}$ can be written as a weighted sum of the $\mathbf{p}_k$:

$$\mathbf{u} = \sum_{i=0}^{n-1} \alpha_i \mathbf{p}_i. \tag{5.14}$$

Reasoning from Equation (5.14), $\mathbf{u}_{k+1}$ is calculated as follows:

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \alpha_k \mathbf{p}_k. \tag{5.15}$$

Then the residual $\mathbf{s}_{k+1}$ is

$$\mathbf{s}_{k+1} = \mathbf{f} - K\mathbf{u}_{k+1} = \mathbf{f} - K(\mathbf{u}_k + \alpha_k \mathbf{p}_k) = \mathbf{s}_k - \alpha_k K \mathbf{p}_k. \tag{5.16}$$

Given $\{\mathbf{p}_0, \ldots, \mathbf{p}_k\}$, $\mathbf{p}_{k+1}$ can be calculated using a Gram-Schmidt process:

$$\mathbf{p}_{k+1} = \mathbf{s}_{k+1} - \sum_{i\leq k} \frac{\mathbf{p}_i^* K \mathbf{s}_k}{\mathbf{p}_i^* K \mathbf{p}_i} \mathbf{p}_i. \tag{5.17}$$

However, this would involve the storage of all previous search directions. It can be shown that $\mathbf{s}_{k+1}$ is orthogonal to $\mathbf{p}_i$ for all $i \leq k$, which enables us to store only $\mathbf{s}_k$, $\mathbf{p}_k$ and $\mathbf{u}_k$ and still calculate $\mathbf{s}_{k+1}$, $\mathbf{p}_{k+1}$ and $\mathbf{u}_{k+1}$. $\mathbf{p}_{k+1}$ is calculated as follows:

$$\mathbf{p}_{k+1} = \mathbf{s}_{k+1} + \beta_k \mathbf{p}_k \tag{5.18}$$

where β_k is chosen such that $\mathbf{p}_{k+1}$ is conjugate to $\mathbf{p}_k$. Multiplying Equation (5.14) on the left by $\mathbf{p}_k^* K$ yields

$$\alpha_k = \frac{\mathbf{p}_k^* \mathbf{f}}{\mathbf{p}_k^* K \mathbf{p}_k}. \tag{5.19}$$

Substituting Equations (5.12) and (5.18) into Equation (5.19) leads to an expression for α_k:

$$\alpha_k = \frac{\mathbf{p}_k^*\mathbf{f}}{\mathbf{p}_k^* K\mathbf{p}_k} = \frac{\mathbf{p}_k^*(\mathbf{s}_k + K\mathbf{u}_k)}{\mathbf{p}_k^* K\mathbf{p}_k} = \frac{\mathbf{p}_k^*\mathbf{s}_k}{\mathbf{p}_k^* K\mathbf{p}_k} = \frac{\mathbf{s}_k^*\mathbf{s}_k}{\mathbf{p}_k^* K\mathbf{p}_k}. \tag{5.20}$$

In order to calculate β_k, Equation (5.18) is multiplied on the right by $K\mathbf{p}_k$, yielding

$$\beta_k = -\frac{\mathbf{s}_{k+1}^* K\mathbf{p}_k}{\mathbf{p}_k^* K\mathbf{p}_k}. \tag{5.21}$$

From Equation (5.16), we see that $K\mathbf{p}_k = \frac{1}{\alpha_k}(\mathbf{s}_k - \mathbf{s}_{k+1})$, so

$$\beta_k = -\frac{\mathbf{s}_{k+1}^* K\mathbf{p}_k}{\mathbf{p}_k^* K\mathbf{p}_k} = -\frac{\mathbf{s}_{k+1}^* \frac{1}{\alpha_k}(\mathbf{s}_k - \mathbf{s}_{k+1})}{\mathbf{p}_k^* \frac{1}{\alpha_k}(\mathbf{s}_k - \mathbf{s}_{k+1})} = \frac{\mathbf{s}_{k+1}^*\mathbf{s}_{k+1}}{\mathbf{s}_k^*\mathbf{s}_k}. \tag{5.22}$$

Algorithm **??** is the complete conjugate gradient algorithm.

Algorithm 1 STANDARD CG

```
Require: $K \in \mathbb{C}^{N\times N}, \mathbf{u}_0, \mathbf{f} \in \mathbb{C}^N$;
Ensure: Approximate solution $\mathbf{u}_k$ such that $\|\mathbf{f} - K\mathbf{u}_k\| \le TOL$.
$\mathbf{s}_0 = \mathbf{f} - K\mathbf{u}_0$; $\mathbf{p}_0 = \mathbf{s}_0$; $\gamma_0 = \mathbf{s}_0^*\mathbf{s}_0$;
while $\sqrt{\gamma_k} > TOL$ and $k < k_{max}$ do
    $\xi_k = \mathbf{p}_k^* K\mathbf{p}_k$
    $\alpha_k = \frac{\gamma_k}{\xi_k}$
    $\mathbf{u}_{k+1} = \mathbf{u}_k + \alpha_k\mathbf{p}_k$
    $\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_k K\mathbf{p}_k$
    $\gamma_{k+1} = \mathbf{s}_{k+1}^*\mathbf{s}_{k+1}$
    $\beta_k = \frac{\gamma_{k+1}}{\gamma_k}$
    $\mathbf{p}_{k+1} = \mathbf{s}_{k+1} + \beta_k\mathbf{p}_k$
    $k = k + 1$
end while
```

Clearly, $\text{span}\{\mathbf{p}_0, \mathbf{p}_1, ..., \mathbf{p}_{\mathbf{k-1}}\}$ is equal to the Krylov subspace $\mathcal{K}_k(K, \mathbf{p}_0) := \text{span}\{\mathbf{p}_0, K\mathbf{p}_0, ..., K^{k-1}\mathbf{p}_0\}$. So the conjugate gradient method minimizes $||\mathbf{u} - \mathbf{u}_k||_K$ over $\mathbf{u}_0 + \mathcal{K}_k(K, \mathbf{p}_0)$.

5.3. CGLS

The conjugate gradient method for least squares (CGLS) is obtained by applying the standard conjugate gradient method to the normal equations $K^*K\mathbf{u} = K^*\mathbf{f}$. Additionally, some modifications are made to enhance stability. This method minimizes the residual in every step, because

$$||\mathbf{u} - \mathbf{u}_k||_{K^*K} = (\mathbf{u} - \mathbf{u}_k)^* K^* K(\mathbf{x} - \mathbf{u}_k) = (\mathbf{f} - K\mathbf{u}_k)^*(\mathbf{f} - K\mathbf{u}_k) = \mathbf{s}_k^*\mathbf{s}_k. \tag{5.23}$$

Something similar can be done using the normal Equations (5.4). By replacing K by $A^*C^{-1}A + \lambda R$, $\mathbf{u}$ by $\mathbf{x}$ and $\mathbf{f}$ by $A^*C^{-1}y$, the standard conjugate gradient method in Algorithm 1 can be applied to this problem. By defining

$$r_k := C^{-1}(\mathbf{y} - A\mathbf{x}_k), \tag{5.24}$$

a recursion for the residual is introduced:

$$\begin{aligned} \mathbf{s}_{k+1} &= A^*C^{-1}\mathbf{y} - (A^*C^{-1}A + \lambda R)\mathbf{x}_{k+1} = A^*C^{-1}(\mathbf{y} - A\mathbf{x}_{k+1}) - \lambda R\mathbf{x}_{k+1} \\ &= A^*\mathbf{r}_{k+1} - \lambda R\mathbf{x}_{k+1}. \end{aligned} \tag{5.25}$$

Defining $\mathbf{q}_k = A\mathbf{p}_k$, ξ_k is calculated as

$$\xi_k = \mathbf{q}_k^* C^{-1}\mathbf{q}_k + \lambda\mathbf{p}_k^* R\mathbf{p}_k. \tag{5.26}$$

Algorithm 2 shows CGLS tailored specifically to Equation (5.4).

Algorithm 2 CGLS

```
Require: $A \in \mathbb{C}^{M\times N}, C \in \mathbb{C}^{M\times M}, R \in \mathbb{C}^{N\times N}, \mathbf{x}_0, \in \mathbb{C}^N, \mathbf{y} \in \mathbb{C}^M, \lambda \in \mathbb{R} \geq 0;$
Ensure: Approximate solution $\mathbf{x}_k$ such that $\|(A^*\mathbf{r}_k - \lambda R)\mathbf{x}_k\| \leqslant TOL$.
$\mathbf{r}_0 = C^{-1}(\mathbf{y} - A\mathbf{x}_0)$; $\mathbf{s}_0 = A^*\mathbf{r}_0 - \lambda R\mathbf{x}_0$; $\mathbf{p}_0 = \mathbf{s}_0$; $\mathbf{q}_0 = A\mathbf{p}_0$; $\gamma_0 = \mathbf{s}_0^*\mathbf{s}_0$; $k = 0$;
while $\sqrt{\gamma_k} > TOL$ and $k < k_{max}$ do
    $\xi_k = \mathbf{q}_k^* C^{-1}\mathbf{q}_k + \lambda \mathbf{p}_k^* R\mathbf{p}_k$
    $\alpha_k = \frac{\gamma_k}{\xi_k}$
    $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$; $R\mathbf{x}_{k+1} = R\mathbf{x}_k + \alpha_k R\mathbf{p}_k$
    $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k C^{-1}\mathbf{q}_k$
    $\mathbf{s}_{k+1} = A^*\mathbf{r}_{k+1} - \lambda R\mathbf{x}_{k+1}$
    $\gamma_{k+1} = \mathbf{s}_{k+1}^*\mathbf{s}_{k+1}$
    $\beta_k = \frac{\gamma_{k+1}}{\gamma_k}$
    $\mathbf{p}_{k+1} = \mathbf{s}_{k+1} + \beta_k \mathbf{p}_k$
    $\mathbf{q}_{k+1} = A\mathbf{p}_k$
    $k = k + 1$
end while
```

In each iteration, $\mathbf{e}_k := \mathbf{x} - \mathbf{x}_k$ is minimized over the Krylov subspace $\mathscr{K}_k(A^*C^{-1}A + \lambda R, \mathbf{p}_0)$ in the $(A^*C^{-1}A + \lambda R)$-norm. So

$$
\begin{aligned}
||\mathbf{e}_k||^2_{A^*C^{-1}A+\lambda R} &= (\mathbf{x} - \mathbf{x}_k)^*(A^*C^{-1}A + \lambda R)(\mathbf{x} - \mathbf{x}_k) && (5.27)\\
&= (\mathbf{r} - \mathbf{r}_k)^* C(\mathbf{r} - \mathbf{r}_k) + \lambda \mathbf{e}_k^* R\mathbf{e}_k = ||\mathbf{r} - \mathbf{r}_k||^2_C + \lambda ||\mathbf{e}_k||^2_R && (5.28)
\end{aligned}
$$

is minimized in every iteration. Note the correspondence with the original constrained minimization problem (5.6).

5.4. CGNE

CGNE is obtained by setting $\mathbf{x} = A^*\mathbf{z}$ and applying standard CG to the normal equations

$$
AA^*\mathbf{z} = \mathbf{y}. \tag{5.29}
$$

For underdetermined problems, solving this equation is computationally less expensive than solving $A^*A\mathbf{x} = A^*\mathbf{y}$, because in that case AA^* is of lower dimension than A^*A. Another important advantage is that CGNE minimizes the error in the 2-norm:

$$
||\mathbf{z} - \mathbf{z}_k||^2_{AA^*} = (A(\mathbf{z} - \mathbf{z}_k))^*(A(\mathbf{z} - \mathbf{z}_k)) = (\mathbf{x} - \mathbf{x}_k)^*(\mathbf{x} - \mathbf{x}_k). \tag{5.30}
$$

Unfortunately, this method only works for consistent problems, rendering it useless for most problems in practice due to the presence of noise. However, this issue is solved by regularizing, because regularization leads to a well-posed problem. The standard CG method can be applied to Equation (5.9) by replacing K by $AR^{-1}A^* + \lambda C$, $\mathbf{u}$ by $\mathbf{z}$ and $\mathbf{f}$ by $\mathbf{y}$. Next, we set $\mathbf{x}_k = R^{-1}A^*\mathbf{z}_k$ and $\mathbf{r}_k = \lambda \mathbf{z}_k$ and we define $\mathbf{q}_k := A^*\mathbf{p}_k$, yielding the CGNE algorithm, given as Algorithm 3. In every iteration, the CGNE method minimizes

$$
\begin{aligned}
(\mathbf{z} - \mathbf{z}_k)^*(AR^{-1}A^* + \lambda C)(\mathbf{z} - \mathbf{z}_k) &= (\mathbf{x} - \mathbf{x}_k)^* R^{-1}(\mathbf{x} - \mathbf{x}_k) + \frac{1}{\lambda}(\mathbf{r} - \mathbf{r}_k)^* C(\mathbf{r} - \mathbf{r}_k)\\
&= ||\mathbf{e}_k||^2_{R^{-1}} + \frac{1}{\lambda}||\mathbf{r} - \mathbf{r}_k||^2_C. && (5.31)
\end{aligned}
$$

over the Krylov subspace $\mathscr{K}_k(AR^{-1}A^* + \lambda C, \mathbf{p}_0)$. As far as we know, the CGNE method has not been applied to minimization problem (5.3) before in this way.

Algorithm 3 CGNE

```
Require: $A \in \mathbb{C}^{M\times N}, C \in \mathbb{C}^{M\times M}, R \in \mathbb{C}^{N\times N}, \mathbf{x}_0, \in \mathbb{C}^N, \mathbf{y} \in \mathbb{C}^M, \lambda \in \mathbb{R} > 0;$
Ensure: Approximate solution $\mathbf{x}_k$ such that $\|\mathbf{y} - A\mathbf{x}_k - C\mathbf{r}_k\| \leqslant TOL$.
$\mathbf{s}_0 = \mathbf{y}$, $\mathbf{p}_0 = \mathbf{s}_0$; $\mathbf{q}_0 = A^*\mathbf{p}_0$, $\gamma_0 = \mathbf{s}_0^*\mathbf{s}_0$, $k = 0$
while $\sqrt{\gamma_k} > TOL$ and $k < k_{max}$ do
    $\xi_k = \mathbf{q}_k^* R^{-1}\mathbf{q}_k + \lambda \mathbf{p}_k^* C\mathbf{p}_k$
    $\alpha_k = \frac{\gamma_k}{\xi_k}$
    $\mathbf{r}_{k+1} = \mathbf{r}_k + \lambda\alpha_k\mathbf{p}_k;$
    $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k R^{-1}\mathbf{q}_k;$
    $\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_k(AR^{-1}\mathbf{q}_k + \lambda C\mathbf{p}_k)$
    $\gamma_{k+1} = \mathbf{s}_{k+1}^*\mathbf{s}_{k+1}$
    $\beta_k = \frac{\gamma_{k+1}}{\gamma_k}$
    $\mathbf{p}_{k+1} = \mathbf{s}_{k+1} + \beta_k\mathbf{p}_k$
    $\mathbf{q}_{k+1} = A^*\mathbf{p}_{k+1}$
    $k = k + 1$
end while
```

6

Preliminary results

Some simulations were carried out in order to test the super-resolution reconstruction method described in chapter 4. The MATLAB Shlepp-Logan phantom was chosen as the object to be imaged. The high-resolution image and the low-resolution images were chosen to be 128 × 128 pixels and 32 × 32 pixels respectively. The desired HR image is shown in Figure 6.1.

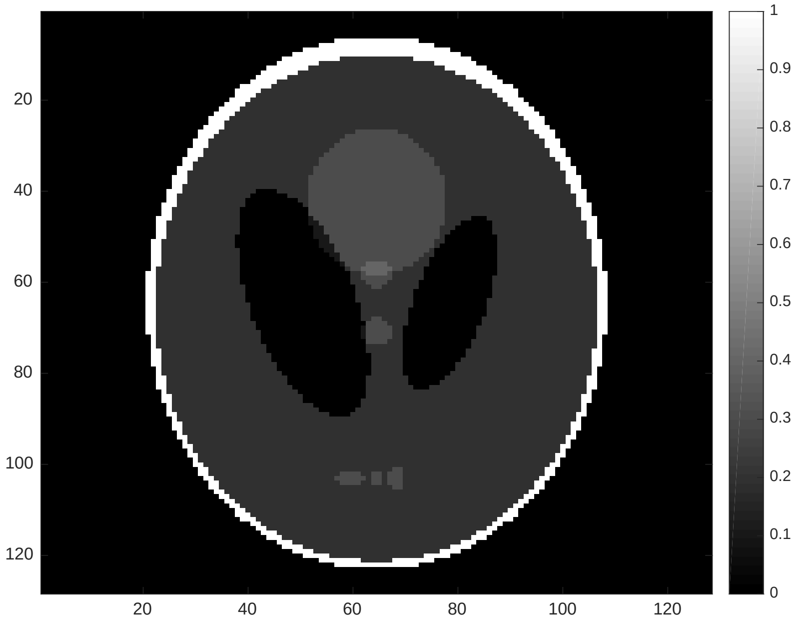


Figure 6.1: Desired HR image (128 × 128 pixels) of the MATLAB Shlepp-Logan phantom.

The grayscale values of the HR image were stored in a matrix X. The first LR image was created without shifting the object. The other 3 LR images were created using three different subpixel shifts: $(0.25, 0.25)$, $(0.5, 0.3)$ and $(0.8, 0.6)$. When $\bar{a} > 0, \bar{b} > 0$, a shift of $(\bar{a}, \bar{b})$ means that the LR grid is shifted to the right by $\bar{a}$ times the size of the LR pixel in the x-direction and towards the bottom by $\bar{b}$ times the size of the LR pixel in the y-direction. So the object itself is shifted to the left and towards the top. Denoting the down-sampling factor in the x-direction by L_1 and in the y-direction by L_2, the image is shifted by $(\bar{a}L_1, \bar{b}L_2)$ HR pixels. Defining $a := \bar{a}L_1$ and $b := \bar{b}L_2$, we can decompose a and b into an integer part and a remainder part:

$$a = \lfloor a \rfloor + a_{rem}, \quad b = \lfloor b \rfloor + b_{rem}. \tag{6.1}$$

Therefore, the value of pixel (i, j) in the shifted image is equal to

$$\begin{aligned}
&(1-a_{rem})(1-b_{rem})X_{i-\lfloor a\rfloor,j-\lfloor b\rfloor} &(6.2)\\
&+a_{rem}(1-b_{rem})X_{i-\lfloor a\rfloor-1,j-\lfloor b\rfloor} &(6.3)\\
&+(1-a_{rem})b_{rem}X_{i-\lfloor a\rfloor,j-\lfloor b\rfloor-1} &(6.4)\\
&+a_{rem}b_{rem}X_{i-\lfloor a\rfloor-1,j-\lfloor b\rfloor-1}. &(6.5)
\end{aligned}$$

For each of the shifts, this information was stored in matrix G_k such that $G_k\mathbf{x}$ yielded the desired shifted version of the HR image. A Gaussian function with $s_1 = s_2 = 1$ and $\rho = 0$ was used as the point spread function, leading to the blurring matrix B. The down-sampling operator D_k simply takes the average over $L_1L_2 = 16$ pixels. The noise was independent, with standard deviation $\sigma = 0.01$. The LR images are shown in Figure 6.2.

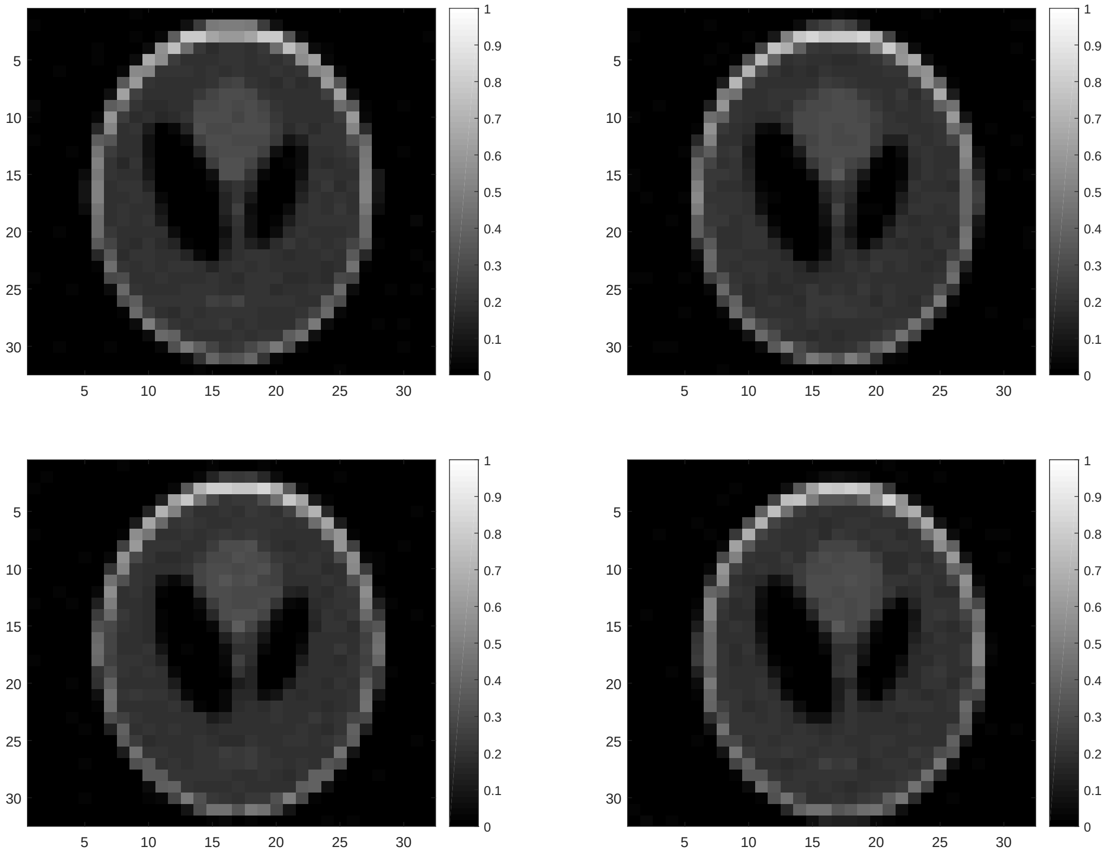


Figure 6.2: Four low-resolution images (32 × 32 pixels).

Using these four $\mathbf{y}_k$, minimization problem (4.14) (with Tikhonov regularization) was solved using the CGLS method described in chapter 5. λ was chosen to be 0.005. The resulting HR image is shown in Figure 6.3.

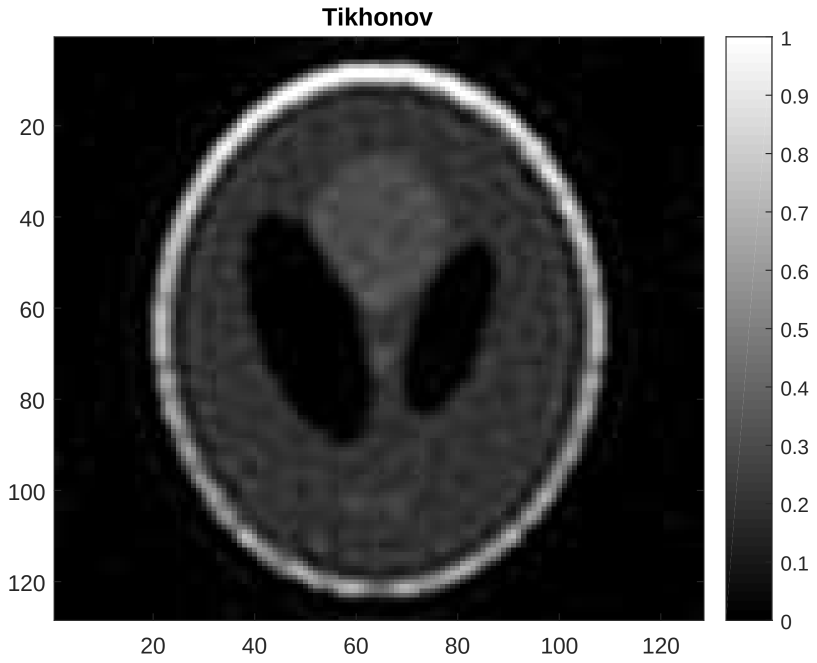
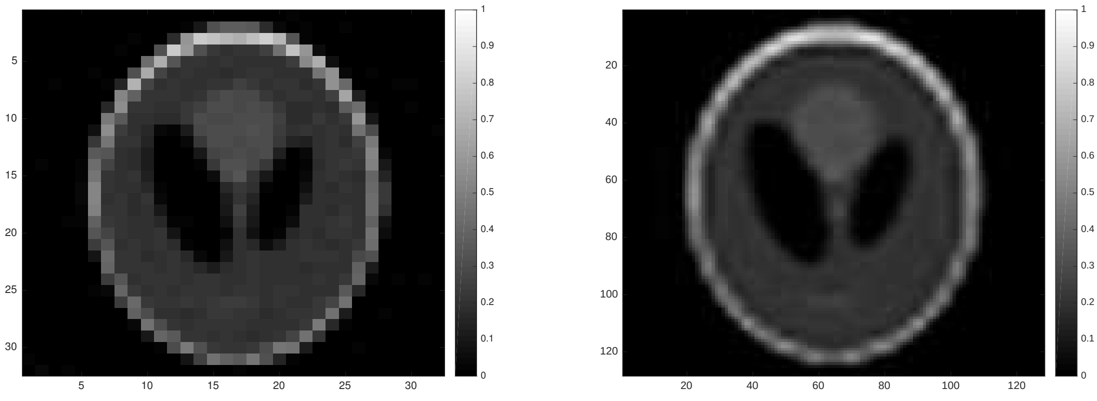


Figure 6.3: The high-resolution result obtained after solving minimization problem (4.14) (with Tikhonov regularization, $\lambda = 0.005$).

In Figure 6.4, the HR image and one of the LR images are plotted side by side, allowing for a better comparison between the quality of the images.



(a) LR image.

(b) HR image.

Figure 6.4: The HR image is plotted next to one of the four LR images used to reconstruct it.

Additionally, the minimization problem (4.18) (with total variation regularization) was solved using fixed-point iteration and CGLS. The resulting HR image is plotted in Figure 6.5.

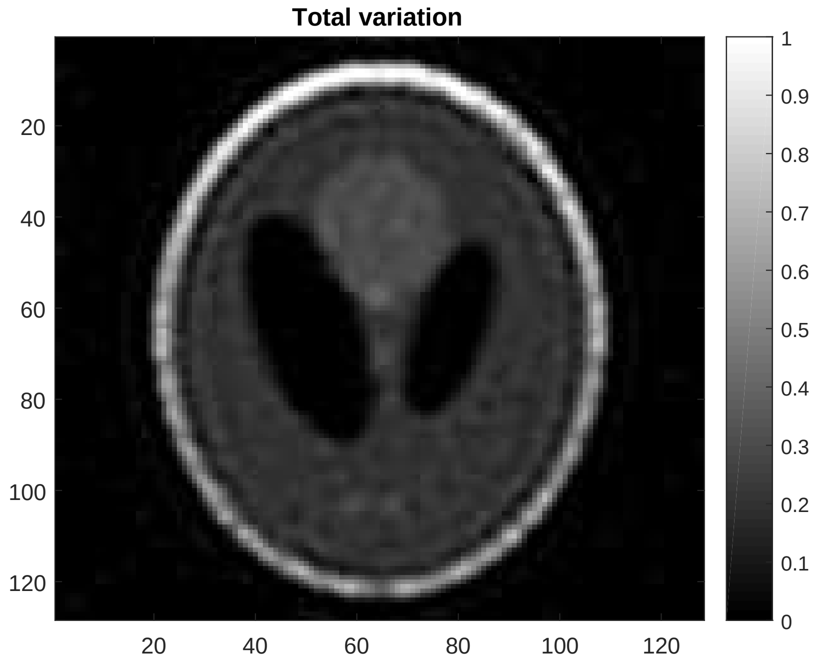


Figure 6.5: The high-resolution result obtained after solving minimization problem (4.18) (with total variation regularization, $\lambda = .3$).

The image resulting from using edge-preserving regularization is shown in Figure 6.6.

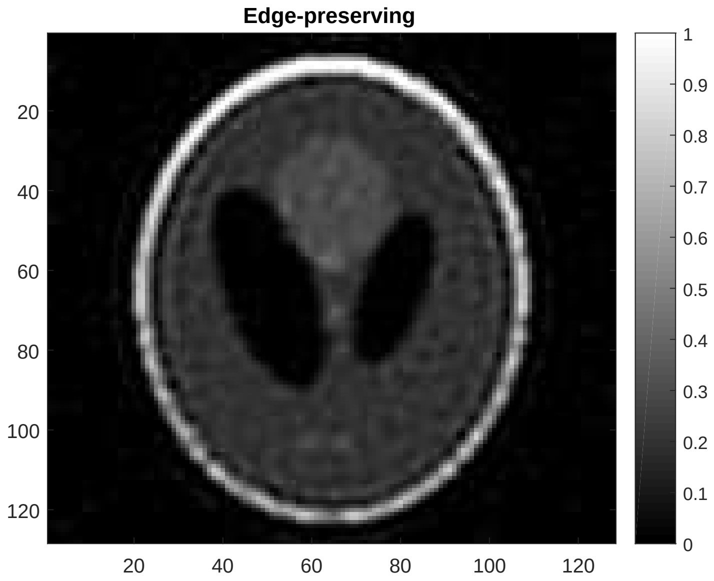
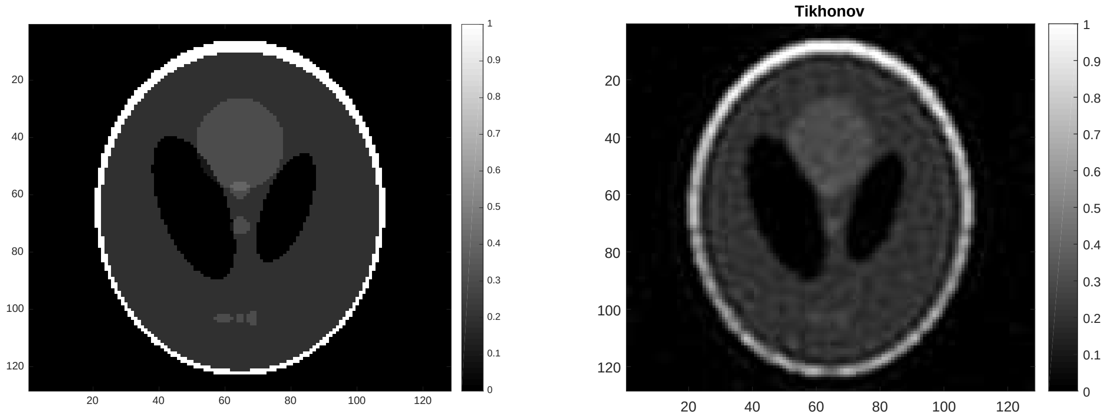


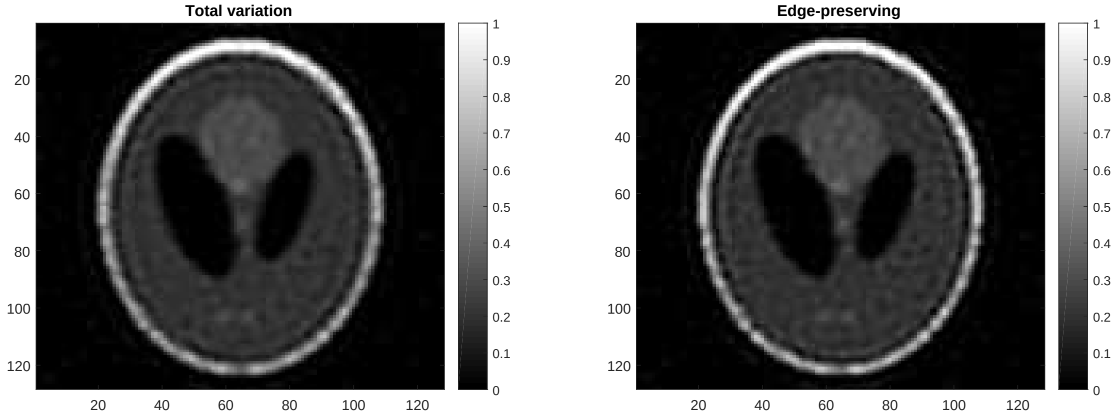
Figure 6.6: The high-resolution result obtained after solving minimization problem (4.30) (with edge-preserving regularization) using $T = 0.05$ and $\lambda = 0.003$.

Figure 6.7 shows the phantom image and the three images resulting from the three different kinds of regularization. The three resulting images look very similar. The image obtained after using edge-preserving regularization seems to be slightly better than the other two. However, the choice of regularization parameter (as well as T) obviously influences the outcome. It is unclear which value should be chosen for these parameters. Employing the L-curve criterion did not resolve this problem, because no L-shape could be detected.



(a) Original image.

(b) Tikhonov regularization.



(c) Total variation regularization.

(d) Edge-preserving regularization.

Figure 6.7: The results obtained by using three different kinds of regularization.

The pixels must have values between 0 and 1, but all the resulting minimizers had values outside of this range (between -0.3 and 1.3). The final images were obtained by projection onto the interval $[0,1]$. The ℓ_1-norm and the ℓ_2-norm of the error of each of the three solutions are shown in Table 6.1. The values are very similar, but edge-preserving regularization yields the smallest errors, total variation the largest.

Table 6.1: ℓ_1 and ℓ_2 norm of the errors of the three obtained solutions.

Regularization	Error in ℓ_1 norm	Error in ℓ_2 norm
Tikhonov	671.1918	12.6632
Total variation	682.5282	12.8118
Edge-preserving	648.0891	12.0353

7

Research goals

7.1. Main goal

The main goal of this project is to research whether super-resolution reconstruction can be used to improve the quality of the images produced using low-field MRI to such an extent that this method can be used to detect hydrocephalus.

7.2. Research questions

The following research questions were formulated in order to structure this research:

1. Can super-resolution reconstruction yield high-resolution MR images when applied to simulated data?
2. Can super-resolution reconstruction yield high-resolution MR images when applied to real data?
3. What is a suitable matrix A in Equation (4.5), describing the transition from the high-resolution object to the low-resolution images?
4. Which method(s) should be used to solve the minimization problem arising in super-resolution reconstruction?
5. Which type of regularization yields the best results?

7.3. Methodology

1. To test the super-resolution reconstruction technique, it will be applied to data obtained from a simulated Halbach array. Because the dataset is obtained using a simulation, we will be able to compare the obtained high-resolution image to the actual phantom, enabling us to analyze the quality of the produced image. The low-resolution images will have 32×32 or 64×64 pixels, the high-resolution image 128×128 pixels.
2. When the team at Pennsylvania State University starts generating images with their low-field MRI prototype, the super-resolution reconstruction technique will be employed to construct high-resolution images building on a number of their low-resolution images. The low-resolution images will be created using a predetermined shift, so that the operator governing the geometric deformation is known. Because the objects they will be imaging are simple objects containing a lot of water (water yields a strong signal in MRI), we will know what the images are supposed to look like approximately, allowing us to have a general idea of how good our results are.
3. By using predetermined shifts, we will be able to construct the geometric deformation precisely. The downsampling operator simply takes averages over a number of pixels. So the only unknown is the blurring operator. Literature shows that the point spread function can be modeled as a Gaussian function. Possibly, the parameters determining the shape of the Gaussian function can be determined using maximum likelihood estimation.

4. The CGLS method as described in chapter 5 is the conventional method of choice for solving problems of the form (5.3). We will look into the CGNE method as well. The advantage of CGNE is that due to its use of AA^* (as opposed to A^*A in the CGLS method), under-determined problems can be solved faster. This is especially useful in MRI applications, due to the long acquisition times currently necessary in order to obtain a sufficient amount of information about the object being imaged. If we are able to generate good solutions to the minimization problem in the under-determined case using the CGNE method, this could mean that fewer measurements have to be taken, allowing for shorter acquisition times. In order to test this, both methods will be applied to overdetermined, underdetermined and critical systems.

5. Tikhonov and total variation regularization will be used to solve the problem. Their results will be compared. Other kinds of regularization, such as the edge-preserving kind, will be looked into as well. Instead of using regularization, another option would be to suppose that the HR image is a sparse combination of vectors in a dictionary based on known images of hydrocephalic and healthy brains, similar to the approach taken in [18].

Bibliography

[1] M. Sarracanie, C.D. LaPierre, N. Salameh, D.E.J. Waddington, T. Witzel, and M.S. Rosen. Low-cost high-performance MRI. *Nature scientific reports*, 5, 2015.

[2] R.W. Gerchberg. Super-resolution through error energy reduction. *Journal of Modern Optics*, 21(9):709–720, 1974.

[3] P. De Santis and F. Gori. On an iterative method for super-resolution. *Journal of Modern Optics*, 22(8):691–695, 1975.

[4] D. Fiat. Method of enhancing an MRI signal, September 25 1997. US Patent 6,294,914.

[5] E. Van Reeth, I.W.K. Tham, C.H. Tan, and C.L. Poh. Super-resolution in magnetic resonance imaging: A review. *Concepts in Magnetic Resonance Part A*, 40A(6):306–325, 2012.

[6] Z.-P. Liang Liang and P.C. Lauterbur. *Principles of magnetic resonance imaging: a signal processing perspective.* The Institute of Electrical and Electronics Engineers Press, 2000.

[7] G.A. Wright. Magnetic resonance imaging. *IEEE Signal Processing Magazine*, 97:56–66, 1997.

[8] C. Kegler, H.C. Seton, and J.M.S. Hutchison. Prepolarized fast spin-echo pulse sequence for low-field MRI. *Magnetic resonance in medicine*, 57(6):1180–1184, 2007.

[9] M.S. Wijchers. Image reconstruction in MRI: The possibilities of portable, low-cost MRI scanners. Master's thesis, Delft University of Technology, 2016.

[10] S.C. Park, M.K. Park, and M.G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.

[11] P.C. Hansen, J.G. Nagy, and D.P. O'Leary. *Deblurring images: matrices, spectra, and filtering*, volume 3. Siam, 2006.

[12] P.C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM review*, 34(4):561–580, 1992.

[13] R. Pan and S.J. Reeves. Efficient Huber-Markov edge-preserving image restoration. *IEEE Transactions on Image Processing*, 15(12):3728–3735, 2006.

[14] C.A. Bouman. *Model Based Image Processing.* 2013.

[15] M. Van Gijzen. CG methods for weighted and regularised least squares problems, 2016. Personal notes.

[16] C. Vuik and D.J.P. Lahaye. Scientific computing, 2015. Lecture notes for the course Scientific Computing.

[17] M.R. Hestenes and E. Stiefel. *Methods of conjugate gradients for solving linear systems*, volume 49. NBS, 1952.

[18] J. Yang, J. Wright, T.S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.