Novel Compressed Sensing Algorithms with Applications to Magnetic Resonance Imaging

by

Yue Hu

Submitted in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Mathews Jacob

Department of Electrical and Computer Engineering
Arts, Sciences and Engineering
Edmund A. Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, New York

2013

Biographical Sketch

Yue Hu was born in Harbin, China. She attended the Harbin Institute of Technology, and graduated with a Bachelor of Science degree in Electrical Engineering in 2008. She began doctoral studies in the department of Electrical and Computer Engineering at the University of Rochester in 2008 and received the Master of Science degree from the University of Rochester in 2011. She was the teaching assistant for two ECE courses: Digital Logic Design and Introduction to Digital Communications. She was awarded the IEEE-ISBI young investigator travel award in 2011 and 2012. She pursued her research in image processing under the direction of Professor Mathews Jacob.

The following publications are a result of work conducted during her doctoral study:

1. **Y. Hu** and M. Jacob, "Higher degree total variation (HDTV) regularization for image recovery", *IEEE Trans. Image Processing*, Vol 21, No 5, pp 2559-2571, May 2012.

2. **Y. Hu**, S. G. Lingala, M. Jacob, "A fast majorize-minimize algorithm for the recovery of sparse and low rank matrices", *IEEE Trans. Image Processing*, Vol 21, No 2, pp 742-753, Feb 2012.

3. S. G. Lingala, **Y. Hu**, E. DiBella, M. Jacob, “Accelerated dynamic MRI exploiting sparsity and low-rank structure”, *IEEE Trans. Medical Imaging*, Vol 30, No 5, pp 1042-1054, May 2011.

4. **Y. Hu** and M. Jacob, “A fast majorize minimize algorithm for higher degree total variation regularization”, *13th IEEE Symposium on Biomedical Imaging*, 2013.

5. W. E. Kwok, **Y. Hu**, Z. You, M. Jacob, “Accelerated high-resolution MR angiography of fingers with compressed sensing”, *International Society Magnetic Resonance in Medicine (ISMRM)*, April 2013.

6. **Y. Hu** and M. Jacob, “Improved higher degree total variation regularization”, *12th IEEE Symposium on Biomedical Imaging*, 2012.

7. **Y. Hu** and M. Jacob, “Image recovery using improved total variation regularization”, *11th IEEE Symposium on Biomedical Imaging*, 2011.

8. **Y. Hu** and Mathews Jacob, “Higher degree total variation (HDTV) algorithms for biomedical inverse problems”, *invited paper in the IV conference on Wavelets and Sparsity, SPIE*, 2011.

9. S. G. Lingala, **Y. Hu**, E. Dibella, M. Jacob, “Highly accelerated myocardial perfusion MRI using k-t SLR with parallel imaging”, *International Society Magnetic Resonance in Medicine (ISMRM)*, May 2011.

10. **Y. Hu**, S. G. Lingala, M. Jacob, “High resolution structural free breathing cardiac MRI enabled by k-t SLR”, *International Society Magnetic Resonance in Medicine (ISMRM)*, May 2011.

11. S. G. Lingala, **Y. Hu**, M. Jacob. "Real-time cardiac MRI using low-rank and sparsity penalties", *10th IEEE Symposium on Biomedical Imaging*, 2010.

Acknowledgements

First and foremost, I would like to thank my advisor Professor Mathews Jacob. This thesis would not have been finished without his patient guidance and generous help during my five years of Ph.D. study. He is a very gentle person who has great enthusiasm about research. When I have questions or get frustrated, he is always there to offer the most patient answers and encouragement. I could not have asked for a better advisor.

I am also grateful to my committee members: Professors Marvin Doyley, Nathan Cahill, Mark Bocko, and Azadeh Vosoughi, who provided encouraging and constructive comments on my thesis.

I would like to thank my labmates in the Computational Biomedial Imaging Group: Sajan Goud, Zhili Yang, Xuan Zhou, Greg Ongie, Merry Mani, Satyananda Kashyap, Sampada Bhave, Chen Cui, Bryan Senchuk, Ipshita Bhattacharya, Arvind Balachandrasekaran, Sunrita Poddar, Yasir Baqqal, for the intense discussions and the sleepless nights before the conference deadlines. The lab would not be such a rich atmosphere to study and explore new ideas without them. It was such a unforgettable journey for me.

There is no way for me to do my research work without the love and support from my parents. They are always there to encourage and support me by all means. The Ph.D. study was a difficult journey, and they gave me the possibility to pursue my dreams.

Abstract

Magnetic Resonance Imaging (MRI) is a widely used non-invasive clinical imaging modality. Unlike other medical imaging tools, such as X-rays or computed tomography (CT), the advantage of MRI is that it uses non-ionizing radiation. In addition, MRI can provide images with multiple contrast by using different pulse sequences and protocols. However, acquisition speed, which remains the main challenge for MRI, limits its clinical application. Clinicians have to compromise between spatial resolution, SNR, and scan time, which leads to sub-optimal performance.

The acquisition speed of MRI can be improved by collecting fewer data samples. However, according to the Nyquist sampling theory, undersampling in k-space will lead to aliasing artifacts in the recovered image. The recent mathematical theory of compressed sensing has been developed to exploit the property of sparsity for signals/images. It states that if an image is sparse, it can be accurately reconstructed using a subset of the k-space data under certain conditions.

Generally, the reconstruction is formulated as an optimization problem. The sparsity of the image is enforced by using a sparsifying transform. Total variation (TV) is one of the commonly used methods, which enforces the sparsity of the image gradients and provides good image quality. However, TV introduces patchy or painting-like artifacts in the reconstructed images. We introduce novel

regularization penalties involving higher degree image derivatives to overcome the practical problems associated with the classical TV scheme. Motivated by novel reinterpretations of the classical TV regularizer, we derive two families of functionals, which we term as isotropic and anisotropic higher degree total variation (HDTV) penalties, respectively. The numerical comparisons of the proposed scheme with classical TV penalty, current second order methods, and wavelet algorithms demonstrate the performance improvement. Specifically, the proposed algorithms minimize the staircase and ringing artifacts that are common with TV schemes and wavelet algorithms, while better preserving the singularities.

Higher dimensional MRI is also challenging due to the above mentioned tradeoffs. We propose a three-dimensional (3D) version of HDTV (3D-HDTV) to recover 3D datasets. One of the challenges associated with the HDTV framework is the high computational complexity of the algorithm. We introduce a novel computationally efficient algorithm for HDTV regularized image recovery problems. We find that this new algorithm improves the convergence rate by a factor of ten compared to the previously used method. We demonstrate the utility of 3D-HDTV regularization in the context of compressed sensing, denoising, and deblurring of 3D MR dataset and fluorescence microscope images. We show that 3D-HDTV outperforms 3D-TV schemes in terms of the signal to noise ratio (SNR) of the reconstructed images and its ability to preserve ridge-like details in the 3D datasets.

To address speed limitations in dynamic MR imaging, which is an important scheme in multi-dimensional MRI, we combine the properties of low rank and sparsity of the dataset to introduce a novel algorithm to recover dynamic MR datasets from undersampled k-t space data. We pose the reconstruction as an

optimization problem, where we minimize a linear combination of data consistency error, non-convex spectral penalty, and non-convex sparsity penalty. The problem is solved using an iterative, three step, alternating minimization scheme. Our results on brain perfusion data show a significant improvement in SNR and image quality compared to classical dynamic imaging algorithms.

Contributors and Funding Sources

This work was supervised by a dissertation committee consisting of Professor Mathews Jacob (advisor) from the Department of Electrical and Computer Engineering at the University of Iowa, Professor Marvin Doyley and Professor Mark Bocko from the Department of Electrical and Computer Engineering at the University of Rochester, and Professor Nathan Cahill of the Center for Applied and Computational Mathematics at Rochester Institute of Technology. All work for the dissertation was completed independently by the student. The work is supported by the Natonal Science Foundation under awards CCF-0844812 and CCF-1116067.

Table of Contents

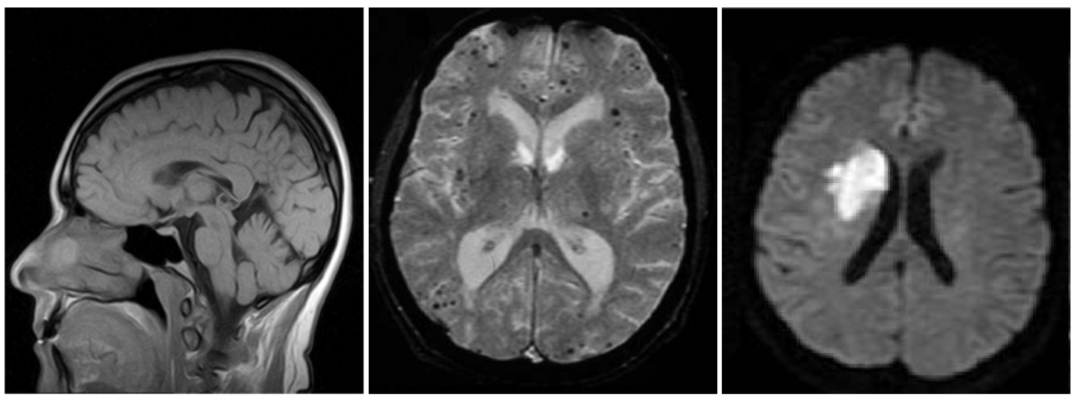
List of Tables

List of Figures

1 Introduction

Magnetic resonance imaging (MRI) is a widely used non-invasive clinical imaging modality. In contrast to other medical imaging techniques such as X-rays, computed tomography (CT), and positron emission tomography (PET), which involve exposure to ionizing radiation, a distinct advantage of MRI is that it uses non-ionizing radiation. Since MRI poses minimal risk, it is preferable for longitudinal research and functional studies [1]. Another popular non-ionizing imaging technique is ultrasound, which uses high frequency sound waves to acquire images. Both ultrasound and MRI can provide images with high resolution. However, the sound waves used in ultrasound could be sub-optimal in imaging regions with bone and air. MRI, on the other hand utilizes the nuclear magnetic resonance properties of hydrogen atoms inside the body, and hence can efficiently image a wide range of tissues.

Another advantage of MRI is that it is capable of providing images with multiple contrast. Specifically, by changing the pulse sequences and the protocols, images with different tissue contrast can be obtained depending upon the specific application. For example, T_1-weighted MRI provides good contrast of gray matter and white matter in the brain, T_2-weighted MRI improves the visualization of myocardial edema, T_2^*-weighted MRI is very useful for the detection of cerebral microbleeds [2], diffusion MRI helps diagnosis of vascular strokes and enables detection of neuronal fiber tracts, and the metabolite distributions can be achieved from chemical shift imaging (CSI). Fig. 1.1 shows a series of MR images with multiple contrast due to different clinical applications [2; 3].



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

Figure 1.1: Brain MR images with different contrast. (a) is a T_1 weighted brain MR image, where the brain gray matter and white matter are distinguished clearly. (b) shows a T_2^* weighted brain image. The microbleeds point can be detected from the image. A diffusion weighted brain image is presented in (c). The obvious high intensity area indicates the cerebral infarction which causes acute stroke.

1.1 Motivation

The main challenge of MRI that limits its clinical application is the relatively slow acquisition speed. The achievable resolution is constrained by signal to noise ratio (SNR) and the scan time. In practical applications, clinicians are often forced to compromise between spatial resolution, SNR, and scan time, often resulting in sub-optimal performance. Fig. 1.2 illustrates the three-way tradeoff among the three factors in MRI. In many applications, the SNR is often sufficient, thus the major tradeoff is between resolution and scan time.

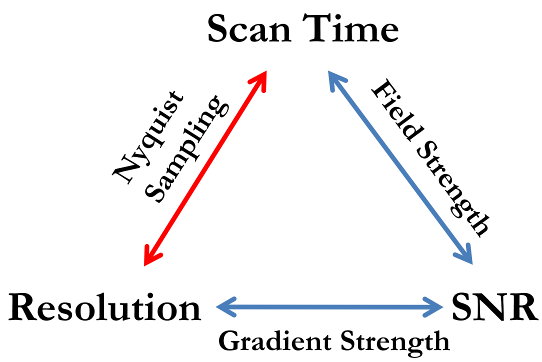


Figure 1.2: MRI is a compromise between three tradeoff factors: the scan time, spatial resolution, and the SNR.

1. **Tradeoff between SNR and scan time:** SNR is one of the most important parameters of image quality. Theoretically, increasing field strength leads to a linearly increased SNR [4]. However, as the main magnetic field ($\mathbf{B_0}$ field) strength increases, the scanner becomes considerably more expensive and the scan time gets longer. In addition, a higher $\mathbf{B_0}$ field strength leads to increment of both $\mathbf{B_0}$ field inhomogeneity and RF magnetic field ($\mathbf{B_1}$ field) inhomogeneity. As a result, it is often difficult to get a good compromise between the two factors.

2. **Tradeoff between SNR and resolution:** If the scan time remains unchanged, the resolution can be increased by either using higher gradient field or less signal averaging. Both approaches lead to the reduction of SNR. Moreover, increased gradient field strength can cause peripheral nerve stimulation [5; 6; 7], which is undesirable. In fact, US food and drug administration (FDA) has strict limits on the gradients, which restrains the compromise between SNR and resolution.

3. **Tradeoff between scan time and resolution:** It is known that MRI scanners acquire samples of the images in the Fourier domain, which is termed as k-space. According to Nyquist theorem, the extension of the k-space region determines the spatial resolution, while the density of the samples in k-space determines the field of view (FOV). Fig. 1.3 shows the relationship between the image domain and the k-space. An effective way to shorten the scan time is to undersample k-space. However, undersampling violates the Nyquist criterion, which will lead to aliasing artifacts in the recovered image.

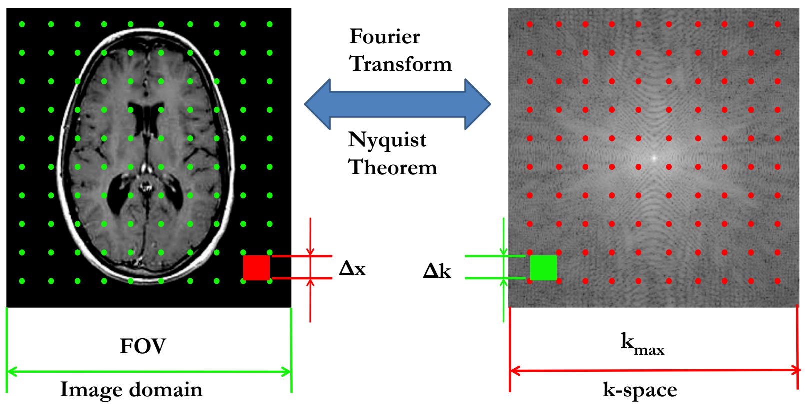


Figure 1.3: The relation between image domain and k-space. The two domains can form a Fourier transform pair. The sampling parameters are inversely proportional. Specifically, FOV $= 1/\Delta$k, $\Delta x = 1/k_{\max}$. The coverage of the k-space samples ($k_{\max}$) determines the spatial resolution of the image. The sampling density in k-space determines the image FOV.

While static MRI schemes are widely used, the acquisition of higher dimensional MRI (e.g. dynamic MRI) is still challenging due to the above mentioned tradeoffs. Instead of imaging a static object, dynamic MRI acquires a series of images of a dynamically evolving object at different time points [8]. There are several dynamic MRI applications such as cardiac, perfusion, gastro-intestinal, and vocal tract imaging. For specific application, researchers have developed customized solutions. For example, in cardiac MRI, the periodicity of the heartbeats is exploited to enable data sharing in k-space. Subsets of the k-space are filled within each heartbeat and the image is reconstructed by combining the k-space samples from different heartbeats. This is possible when the heart is beating periodically, where the motion is captured by electrocardiogram (ECG) gating. Another physiological motion, i.e. respiratory motion, is minimized by the subjects holding their breath. Fig. 1.4 (a) shows the ECG gated breath-holding cardiac MRI, with the k-space sampling pattern in (c) and the reconstructed heart image in (d). Good reconstruction is only possible when the ECG gating is perfect and patients are holding their breaths. However the assumptions of data sharing (periodic heartbeats and breath-holds) are often not met in many clinical scenarios. For instance, patients with arrhythmia have high variability in their heart rates. Pediatric patients and other patients suffering from asthma, dyspneic respiration or congestive heart failure cannot comply with the strict breath-hold demands. This results in inconsistent data sharing, shown in Fig. 1.4 (b), and manifests as artifacts in the reconstruction as demonstrated in (e).

Recent research has been focused on accelerated MRI techniques, which are aimed to address the compromise between the scan time and the image quality by acquiring fewer data samples. The concept of sparsity has been of great interest in accelerated MR acquisition [9]. This is motivated from the field of image compression. Sparsity means that there are relatively few non-zero coefficients in the signal domain or the transform domain. The commonly used image compression method of JPEG/JPEG-2000 relies on the sparsity of the data in discrete cosine or wavelet transform domain. An encoding technique is applied to store the few non-zero coefficients for the later decoding to retrieve the original image. Most MR images are sparse, such as MR angiography (in image domain), and brain MR images (in wavelet domain or finite differences representation) [9]. Fig. 1.5

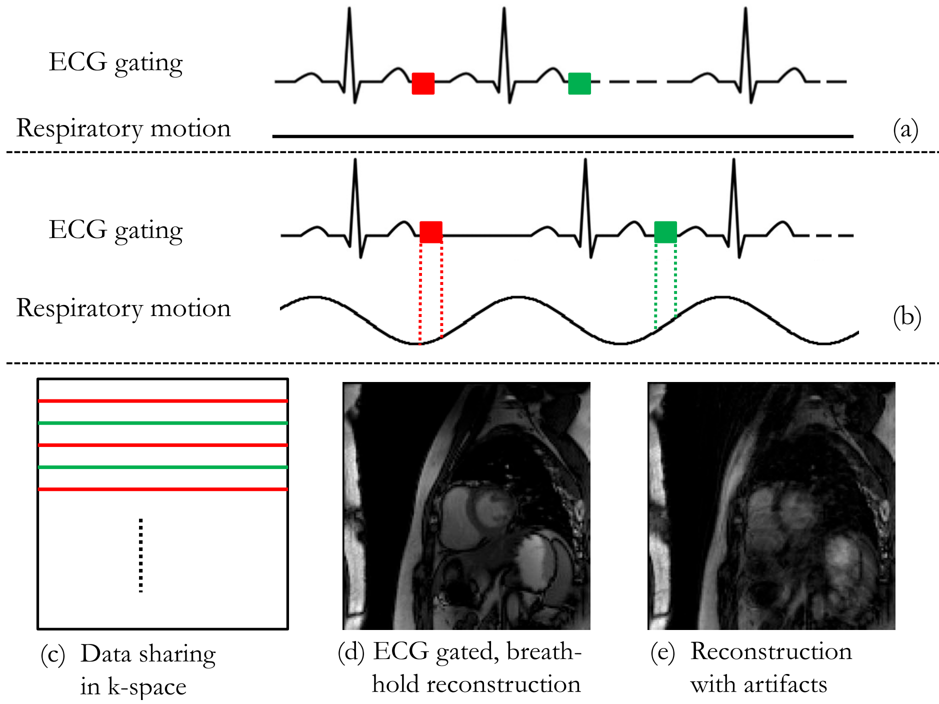
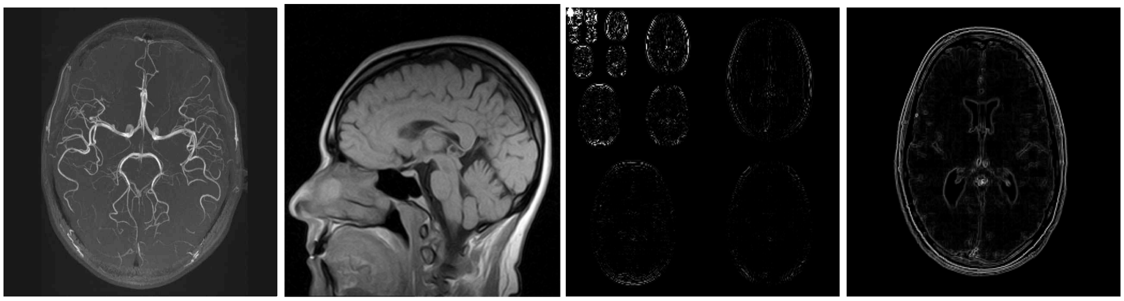


Figure 1.4: Classical dynamic cardiac MRI. (a) depicts the ECG gated breath-holding imaging protocol. For a frame of the cardiac image, the k-space data are collected during different cardiac cycle. As shown in (c), each line of the k-space samples corresponds to one heartbeat. The corresponding reconstructed image is presented in (d). However, the conditions of data sharing (periodic heartbeats and breath-holds) usually are not satisfied, shown in (b), which leads to artifacts in the recovered image, as (e) shows.

illustrates the sparsity of MR images.

The more sparse a signal is, the more it can be compressed, thereby raising the question: **if an image is sparse in a certain transform domain, can it be exactly recovered from fewer k-space samples?** The recent mathematical theory of compressed sensing has been developed to address this problem. According to this theory, if the image is sparse, it can be reconstructed with a subset of the k-space data under some mild conditions. According to an interpretation of compressed sensing in the context of MRI [9], there are three key factors that are required to ensure a perfect reconstruction of the sparse signal from its Fourier



(a) MR angiography (b) brain MR image (c) Wavelet domain (d) Finite difference

Figure 1.5: Sparsity of MR images. (a) shows a sparse MR angiography image, where only the few pixels indicating the blood vessels are with high intensity. Some MR images are sparse in the transform domain. For example, (b) presents a brain MR image, which is sparse in wavelet domain, as shown in (c); or in finite differences representation, illustrated in (d).

samples.

- The signal that is being recovered has to be sparse in a known transform domain.
- The aliasing artifacts due to undersampling in k-space are incoherent (noise-like).
- The reconstruction is performed using a non-linear algorithm, which simultaneously enforces the sparsity and the data consistency.

Specifically, incoherence means that the aliasing artifacts due to k-space undersampling behave much like additive random noise. The strong components in the signal stand out from the interference, which can be detected and recovered using a non-linear algorithm (e.g. thresholding). Fig. 1.6 illustrates the basic concept of compressed sensing [9].

Generally, the image reconstruction is formed as an optimization problem, where sparsity of the image is enforced by using a sparsifying transform. There

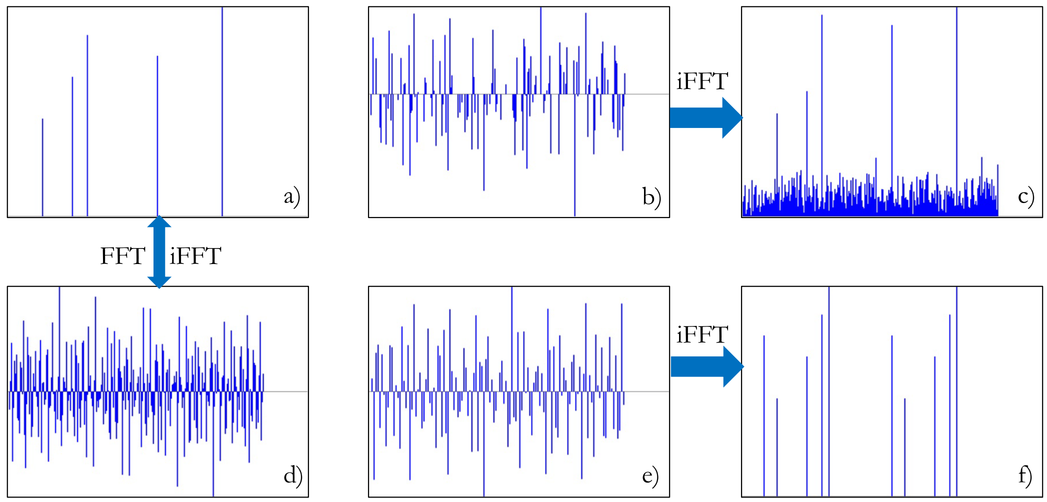
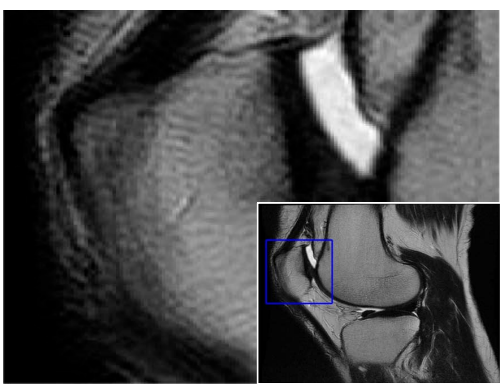


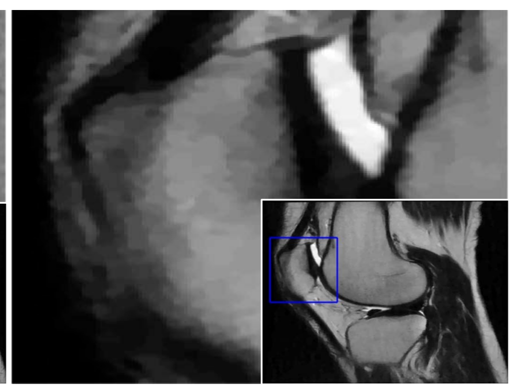
Figure 1.6: The basic concept of compressed sensing. A sparse signal (a) is two-fold undersampled in k-space (d). Equispaced undersampling (e) leads to coherent aliasing (f), from which the signal can not be recovered. Psuedo-random undersampling (b) leads to incoherent aliasing (c). The strong components in the signal can be detected and recovered.

are a number of sparsifying transforms introduced in compressed sensing. Wavelet is one of the most commonly used transforms, which decomposes the image at different scales within three directions (vertical, horizontal, diagonal). The success of wavelet transforms mainly lies in the good performance of capturing the singularities in one-dimensional (1D) signals. Two-dimensional (2D) wavelets are a separable extension of 1D wavelets. Hence, 2D wavelet transforms only provide sparse representation for discontinuities at point-like features, but are sub-optimal in the case of line discontinuities, exhibited as edges or smooth contours in an image. In this context, curvelet [10], contourlet [11], and ridgelet [12] transforms (often denoted by “x-let”) are designed to provide more sparse representations of images with smooth contours. In spite of the good performance in preserving the ridges/curves of the image, x-let transforms often result in curves-like artifacts in the reconstruction. Moreover, the computational complexity and redundancy of x-let transforms are relatively high in practical settings.

In the recent years, total variation, which enforces the sparsity of the image gradients, is emerging as a popular method. The advantages of total variation include its simplicity, rotation invariance, and capability to preserve edges and provide good image quality [13; 14; 15]. Numerous experiments have shown that total variation reconstruction is comparable to more sophisticated schemes such as wavelet and x-let reconstruction [16; 17]. However, total variation has some limitations that restricts its performance in practical applications. The main challenge is that total variation often results in patchy or painting-like artifacts in the reconstructed image that are visually unappealing. Fig. 1.7 shows a TV reconstruction of a knee MR image, from its random selected samples in k-space.



(a) Original knee MR image



(b) Total variation reconstruction

Figure 1.7: Total variation reconstruction of a knee MR image from its undersampled Fourier samples.

In this thesis, we aim to introduce novel methods improving on total variation in order to decrease the samples required in MRI, while preserving image quality. In addition, we combine the sparsity and other important properties of the MR images to accelerate dynamic MRI. The main focus of the thesis is presented in the following section.

1.2 Main focus

The overall goal of the proposal is to develop novel methods to improve the image quality in MRI, while significantly reducing the scan time. We adopt compressed sensing based algorithms, where the sparsity of the image is enforced using a sparsifying transform. Total variation is a widely used method, which tries to find the sparsest gradient of the image. It is simple to implement, and has good performance in preserving edges of the image. However, the total variation method is limited by the staircasing artifacts it introduces into reconstructions. Therefore, we aim to improve the standard total variation (TV) method with so-called higher degree total variation (HDTV) to overcome the current problems with TV reconstruction. In addition, we use both the sparsity property and the low rank property to further improve the performance of the scheme.

The main contributions of this thesis are:

- We formulate the image reconstruction as an optimization problem, where the sparsity of the image gradients is enforced by using the total variation (TV) penalty. We extend the standard TV penalty to higher degree TV (HDTV), using the steerability of the higher degree derivatives. Depending upon two different reinterpretations of standard TV, we introduce two families of HDTV penalties: isotropic HDTV and rotation invariant anisotropic HDTV.

- In order to solve the optimization problem with HDTV penalties, we propose a novel majorize-minimize (MM) algorithm, which involves two steps. In the first step, the objective cost function is majorized using a surrogate function. By solving the surrogate function successively, the minimum of the cost function is obtained. In addition, we use a continuation scheme to accelerate the convergence.

- We generalize the HDTV based compressed sensing scheme into other inverse problems, namely denoising and deblurring. We evaluate the performance of the HDTV penalized algorithm on natural images and MR images in these three applications. The results show that the HDTV based schemes consistently provide better reconstructions, compared to other commonly used sparsifying transforms such as wavelets and curvelets.

- We extend the HDTV regularized method to a 3D-HDTV version to reconstruct high-dimensional MR data and other biological datasets. We demonstrate the utility of 3D-HDTV regularization in the context of compressed sensing, denoising, and deblurring of 3D MR datasets and fluorescence microscope images.

- Motivated by the observation that dynamic MR images are both sparse and low-rank, we combine the TV penalty and the low rank penalty to form a novel image reconstruction scheme, termed as TV-sparsity and Low Rank (TV-SLR) algorithm. We derive an MM algorithm to efficiently solve the problem. Moreover, a continuation scheme is applied to increase the convergence speed.

- We demonstrate the utility of the TV-SLR scheme in 2D image recovery and in dynamic contrast enhanced (DCE) MRI. Results show that TV-SLR achieves high quality image reconstructions with fewer samples than standard methods.

1.3 Thesis outline

In the following, the organization of the thesis is presented.

Chapter 2, entitled "Background", provides a brief overview of MRI and compressed sensing theory .

Chapter 3, entitled "Higher Degree Total Variation (HDTV) Regularization", presents the improved total variation (TV) regularization algorithm: the HDTV regularized algorithm. It introduces the two families of HDTV regularization and the majorize-minimize (MM) algorithm to solve the resulting optimization problem. Some preliminary results on both MR images and natural images are illustrated at the end.

Chapter 4, entitled "Fast Majorize Minimize Three-Dimensional Higher Degree Total Variation (3D-HDTV)", proposes a 3D-HDTV regularized scheme. This chapter describes the formulation of 3D-HDTV and introduces an improved fast MM algorithm to solve the 3D-HDTV regularized problem. The results on 3D MR datasets and 3D fluorescence microscope images demonstrate the effectiveness of the proposed algorithm in improving SNR and image quality.

Chapter 5, entitled "TV Sparsity and Low Rank (TV-SLR) Algorithm", introduces a combined penalties algorithm to recover matrices with sparsity and low rank properties. A brief introduction of low rank matrix recovery is given at the beginning. The TV sparsity and low rank regularized algorithm is then developed to accelerate dynamic MRI. Results on MIT logo image, a simple sparse and low rank matrix, as well as dynamic contrast enhanced (DCE) MRI are presented at the end of this chapter.

Chapter 6, entitled "Conclusion", provides conclusions and further directions for this research.

2 Background

MRI is an important non-ionizing medical imaging technique. It is capable of providing images of soft tissues with good contrast. While static MRI schemes are widely used in clinical practice, higher dimensional MRI is still challenging due to long data acquisition time. The recent theory of compressed sensing enables significant reduction of scan time in MRI. By enforcing the sparsity of MR images, the number of samples required is decreased. Classical methods include exploring the sparsity of the image gradient (total variation), the sparsity in wavelet domain, or in other multi-scale transform domains. Each of these transforms performs well only for a certain class of images. For example, wavelet transforms perform well at representing point singularities, and total variation is effective when the images are piecewise constant. However, none of the transforms are widely applicable. In order to address this problem, we focus on developing novel sparsifying transforms, so as to accelerate MRI without degrading the image quality. This chapter provides background on MRI and compressed sensing, which will facilitate the easy understanding of the rest of the thesis.

2.1 Principles of MRI

2.1.1 Nuclear Magnetic Resonance

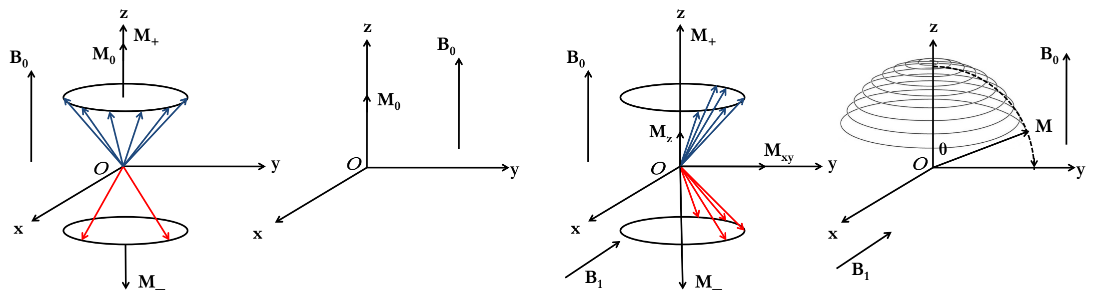
Magnetic resonance imaging (MRI) is based on the phenomenon of nuclear magnetic resonance (NMR). The nuclear spins have the intrinsic angular momentum, which gives rise to a nuclear magnetic moment $\boldsymbol{\mu_I}$. The overall magnetization of a system $\mathbf{M}$ is defined as the vector sum of all the nuclear magnetic moment, i.e. $\mathbf{M} = \sum \boldsymbol{\mu_I}$. Without an external magnetic field, the moments of the spins have random directions. Hence, the net magnetization is zero. In the presence of an external static magnetic field $\mathbf{B_0}$, the spins are polarized and exhibit a net magnetization $\mathbf{M_0}$ that is aligned with the magnetic field $\mathbf{B_0}$, as shown in Fig. 2.1 (a). The polarization and hence the magnetization increases with the main magnetic field strength $\mathbf{B_0}$. The motion of the spins in the presence of $\mathbf{B_0}$ is termed as precession, with the precession frequency (also denoted as Larmor frequency) [1]:

$$\omega_0 = \gamma B_0 \tag{2.1}$$

where γ is the gyromagnetic ratio. In in-vivo MRI, the signal is produced by the spins of protons in water molecules in the body [18].

When an RF excitation field ($\mathbf{B_1}$), modulated at the Larmor frequency, is applied in the transverse plane, the magnetization ($\mathbf{M_0}$) will be tipped away from the equilibrium state. There are two components of the deflected magnetization, i.e longitudinal magnetization (M_z) and transverse magnetization (M_{xy}). After $\mathbf{B_1}$ is removed, the magnetization will gradually return to the original position through the process termed as relaxation, as presented in Fig. 2.1 (b). The detectable signal that is produced by the relaxation of M_{xy} is the MR signal.

We will describe the formulation of the two-dimensional (2D) MR image in the next section.



(a) The polarization of protons (b) Nuclear magnetic resonance

Figure 2.1: In the presence of an external magnetic field $\mathbf{B_0}$, the protons will be polarized and generate a magnetization $\mathbf{M_0}$, as shown in (a). Applying a RF excitation field $\mathbf{B_1}$ along x direction, the magnetization tips away from $\mathbf{M_0}$, producing two components, i.e. longitudinal magnetization M_z and transverse magnetization M_{xy}, as shown in (b).

2.1.2 Two-dimensional Magnetic Resonance Imaging

Signal equation

When an RF excitation pulse is applied, all the protons in the magnetic field are excited. In order to image a slice of the body, a linear gradient field along z direction, i.e. G_z, is used. According to the linear relation between resonant frequency and the magnetic field strength, as Eq. (2.1) presents, the resonant frequencies vary linearly, shown in Fig. 2.2. Therefore, when the band-limited RF field $\mathbf{B_1}$ is applied, only the protons at a slice of the body will be excited with the corresponding resonant frequency. The bandwidth of the RF pulse determines the thickness of the excited slice. Fig. 2.2 illustrates the relationship between the RF pulse bandwidth and the slice thickness.

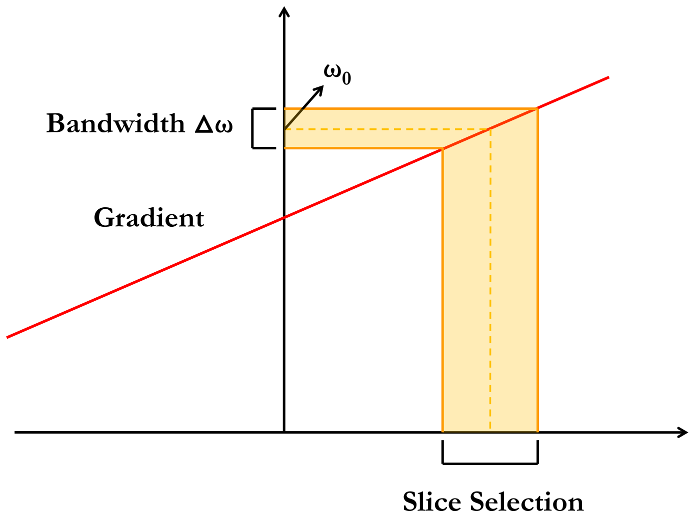


Figure 2.2: Illustration of slice selection. The linear gradient field G_z is exerted to change the resonant frequencies linearly. When the RF pulse centered at ω_0 with bandwidth of $\Delta\omega$ is applied, the protons with the corresponding resonant frequency at a slice of the body are excited. A higher gradient or a larger bandwidth results in thicker selected slice.

Similarly, the gradient field along x direction, G_x, and along y direction, G_y, are applied to localize the MR signal. Hence, the magnetic field $\mathbf{B}$ experienced by protons at a specific spatial location (x, y) and time point t is determined by both the static magnetic field $\mathbf{B_0}$ and the time-varying gradient field $\mathbf{G}$ in two directions [19]:

$$B(t) = B_0 + G_x(t)x + G_y(t)y \tag{2.2}$$

By ignoring the relaxation and the field map effect in the Larmor frequency equation (2.1), we can obtain:

$$M_{xy}(x, y, t) = m(x, y)e^{-j\gamma \int_0^t (G_x(t)x + G_y(t)y)dt} \tag{2.3}$$

The MR signal detected at a specific time point is the summation of MR signal of all voxels:

$$S(t) = \int_x \int_y m(x, y)e^{-j\gamma \int_0^t (G_x(t)x + G_y(t)y)dt} dx\, dy \tag{2.4}$$

We define the k-space location at the time point t as:

$$k_x(t) = \frac{\gamma}{2\pi}\int_0^t G_x(\tau)d\tau \tag{2.5}$$

$$k_y(t) = \frac{\gamma}{2\pi}\int_0^t G_y(\tau)d\tau \tag{2.6}$$

The MR signal equation can thus be expressed as:

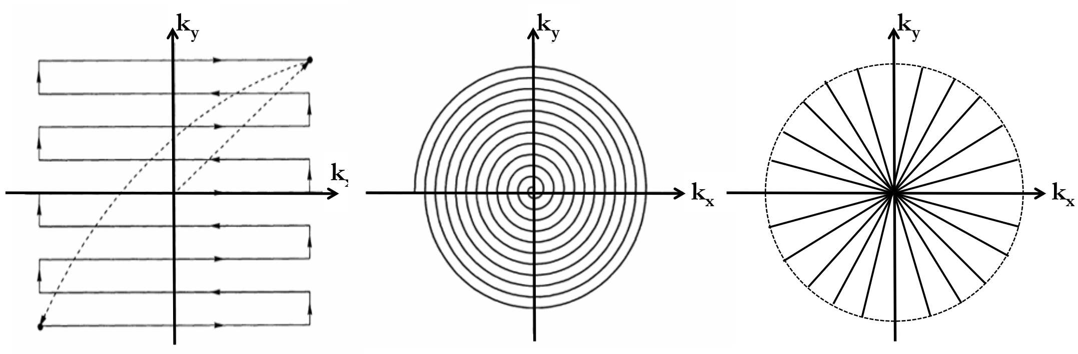
$$S(k_x, k_y) = \int_x \int_y m(x, y)e^{-j2\pi k_x x}e^{-j2\pi k_y y}dx\ dy \tag{2.7}$$

This equation indicates that the received k-space signal $S(k_x, k_y)$ and the image $m(x, y)$ form a Fourier transform pair, where the k-space trajectory is controlled by the gradients.

K-space trajectories

Currently, the most popular k-space trajectory is cartesian acquisition, where the k-space is sampled line by line in order to obtain the whole coverage of k-space domain and reconstruct the image using Fourier transform. This sampling method is slow because it only samples one line per excitation. The time interval between successive excitation pulses is termed as repetition time (TR). Thus, in order to obtain the k-space data of a 256×256 image, the scan time is about 256×TR. Many faster sampling schemes, such as echo-planar imaging (EPI) [20], radial sampling and spiral sampling schemes [21], were introduced to accelerate MRI. The primary advantage of these schemes is that more k-space points are covered during one TR by using time-varying gradient fields. Fig. 2.3 illustrates some k-space sampling trajectories [1].

When the samples in k-space are obtained, the image can be recovered using inverse Fourier transform according to Eq. (2.7). In order to reconstruct the image



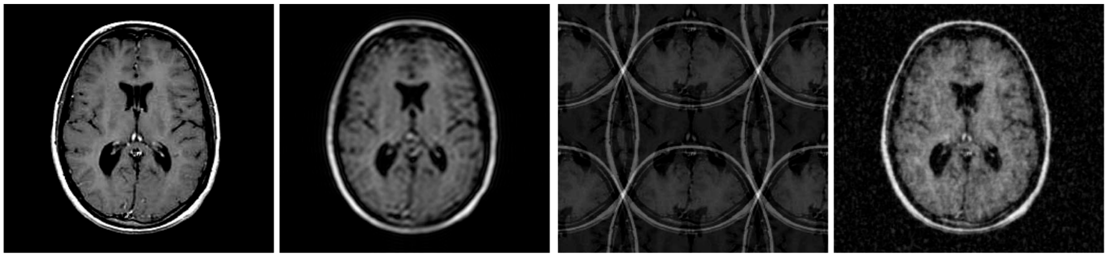
(a) Echo planar imaging (b) Spiral sampling (c) Radial sampling

Figure 2.3: Different k-space sampling trajectories

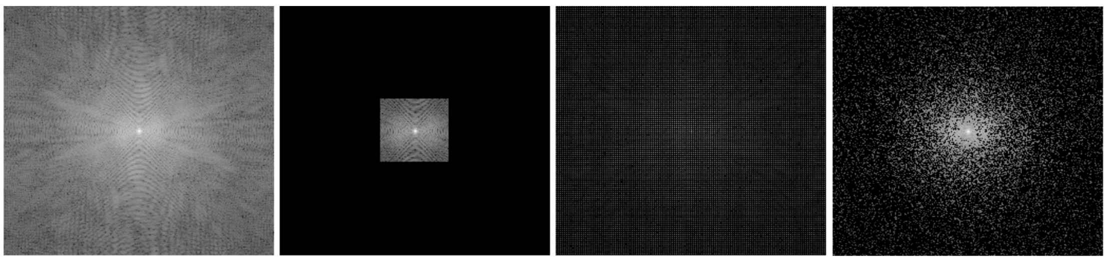
perfectly, the sampling trajectory has to satisfy the Nyquist criterion. When the k-space is undersampled, there are aliasing artifacts exhibited in the reconstruction. Fig. 2.4 illustrates the artifacts due to undersampling.

2.1.3 Dynamic MRI

Dynamic MRI acquires a series of images of a dynamically evolving object at different time points to show the structure and function of the object. The application of dynamic MRI includes cardiac, perfusion, gastro-intestinal, and vocal tract imaging. Dynamic MRI collects more information than static MRI, which is helpful in detection of certain type of diseases (e.g cardiovascular diseases). However, obtaining dynamic MR images with high spatial and temporal resolution in a short period of time is challenging. There are many image reconstruction schemes developed to speed-up the data acquisition of dynamic MRI without degrading the image quality. Generally, the methods fall into three categories: methods using only temporal correlations, using correlations in k-space, and using both correlations.



(a) Brain MR image (b) Low resolution image (c) Aliasing artifacts (d) Incoherent artifacts



(e) K-space samples (f) Center of k-space (g) Equispaced undersampling (h) Random undersampling

Figure 2.4: (a) shows a brain MR image, with the corresponding k-space samples plotted in (e). If only the center of the k-space is sampled (f), the reconstructed image will have lower spatial resolution, as shown in (b). If the k-space is uniformly undersampled (g), the recovered image is contaminated by coherent aliasing artifacts (c), and consequently there is no way to distinguish the original image. However, if the k-space is undersampled in a more irregular pattern, exhibited in (h), incoherent artifacts appear in the MR image (d), which is much slighter than aliasing artifacts in (c), and the original image can be identified.

1. **Methods using temporal correlations:** In the first category, the most popular methods are keyhole imaging [22] and sliding window scheme [23]. Keyhole imaging obtains a full 2D k-space data beforehand as the reference data. During the dynamic data acquisition, the center of k-space is updated at each time frame and the outer of k-space remains the same as in the reference data. Fig. 2.5 shows the concept of keyhole reconstruction. This method is only applicable to dynamic images where the contrast changes are predominant. Imaging a motion involved object using keyhole method will result in artifacts in the re-

construction. In sliding window scheme, the samples are acquired in multiple undersampled subsets of k-space. The image is then recovered by combining the data collected. After the first image is recovered using a full 2D data set collected from multiple subsets, the next image in the series is reconstructed by updating the oldest subset with the most recently acquired subset. The common disadvantage of these two schemes is that a reference frame of 2D k-space data needs to be acquired before the acquisition of the dynamic data. In addition, these methods reconstruct the k-space samples independently, regardless of the highly correlated neighbors.

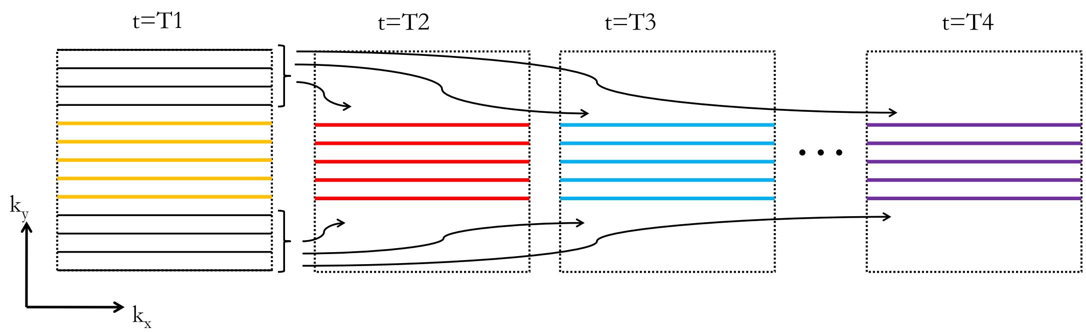


Figure 2.5: Illustration of keyhole reconstruction. The center of k-space is acquired in each time frame, while the periphery remains the same.

2. Methods using k-space correlations: The second type of schemes exploit the correlation between k-space samples when the data is acquired using multiple coils, also known as parallel imaging. In parallel imaging, instead of one RF coil, multiple receive coils are used to share spatial encoding. Since the k-space data from multiple coils are correlated, the image can be recovered by combining the data from different coils. Several parallel imaging techniques are proposed in the recent [24; 25; 26; 27; 28]. Parallel imaging only focuses on the correlation of the samples in k-space. Better performance is expected if both spatial and temporal correlations are taken into consideration.

3. Methods using spatio-temporal correlations: The third type of strategy incorporates both spatial and temporal correlations. This approach is motivated by the fact that the Fourier transform of spatio-temporal basis, i.e. $x - f$ space, is usually very sparse, as depicted in Fig. 2.6. Since the cardiac motion is approximately periodic, most of the energy is concentrated at the harmonics of the cardiac frequencies. This enables the use of compressed sensing based approaches, which rely on enforcing the sparsity in the $x - f$ space. However, the classical sparsifying transforms do not always guarantee a sparse representation of the signal, resulting in blurring or over smoothing artifacts.

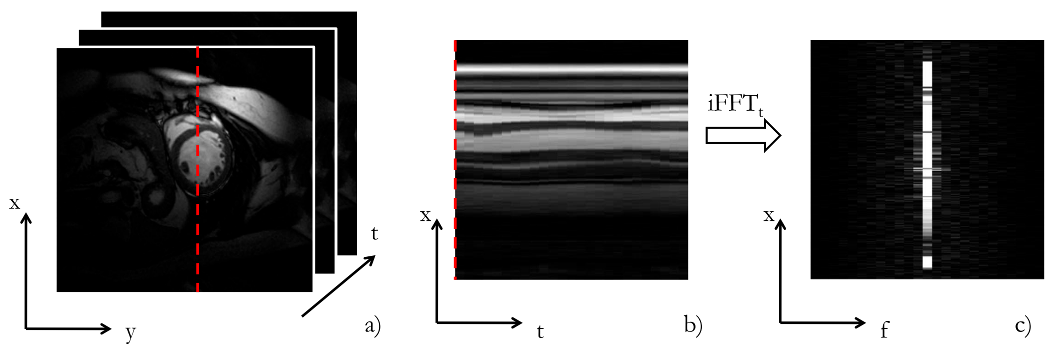


Figure 2.6: $x - f$ representation of a dynamic cardiac dataset. (a) shows a stack of dynamic cardiac images at different time frames. The time profile for one specific line is illustrated in (b). Taking inverse Fourier transform along the temporal axis, the $x - f$ space domain is shown in (c), which is obviously sparse.

2.2 Compressed Sensing

The theory of compressed sensing is motivated by the sparse representations of images used in the field of image compression. The main question in compressed sensing is: if the data is known to be sparse, can it be recovered using fewer measurement samples? In the recent years, this research area has received tremendous interest. Most MR images are sparse in a pre-specified transform domain. In compressed sensing based MRI, the reconstruction is formulated as an optimization problem by enforcing both the data consistency and the sparsity of the image. The solution of the optimization problem is achieved by using non-linear methods.

Compressed sensing (CS) theory was first proposed by Candes, Romberg, and Tao [29], and D. Donoho [30]. Generally speaking, compressed sensing (CS) is a technique which reconstructs a sparse signal from a limited number of its linear measurements. Recently, Lustig has applied this technique to MRI [9]. He adapted the compressed sensing (CS) theory in the context of recovering a sparse image from its undersampled Fourier samples. We will now mathematically describe the essentials of compressed sensing (CS) based schemes.

Suppose $f \in \mathbb{R}^N$ is a sparse signal, Φ is the sparsifying transform (e.g wavelet, finite differences). $\mathcal{A}$ is an arbitrary linear operator such that $\mathcal{A}f + \mathbf{n} = b$, where b is the observed noisy linear measurements. In the context of MRI, $\mathcal{A}$ usually refers to the undersampled Fourier transform, and $\mathbf{n}$ is often modeled as a Gaussian white noise with standard deviation σ. The compressed sensing (CS) method tries to find the signal $\hat{f}$ that is sparsest in the Φ transform domain and satisfies a data consistency requirement,

$$\hat{f}(\mathbf{r}) = \arg\min_{f} \|\Phi f\|_{l_0}, \text{ such that } \|\mathcal{A}(f) - b\|^2 = \sigma^2. \tag{2.8}$$

where the objective function $\|\cdot\|_{l_0}$ is the l_0 norm, which indicates the number of non-zero coefficients in the sparse signal. The sparsity is enforced by the minimization of $\|\Phi f\|_{l_0}$. The constraint $\|\mathcal{A}(f) - b\|^2 = \sigma^2$ promotes data consistency. The problem is often reformulated using Lagrange's multipliers as

$$\hat{f}(\mathbf{r}) = \arg\min_{f} \left(\|\Phi f\|_{l_0} + \lambda \cdot (\|\mathcal{A}(f) - b\|^2 - \sigma^2) \right). \tag{2.9}$$

However, the l_0 reconstruction problem (2.9) is numerically infeasible. Candes et al. and Dohono [31; 29; 30] have theoretically proved that l_1 minimization is equivalent to l_0 minimization on signal recovery if the restricted isometric property (RIP) is satisfied [32]. The l_1 optimization problem is thus presented as:

$$\hat{f}(\mathbf{r}) = \underbrace{\arg\min_{f} \ \|\mathcal{A}(f) - b\|^2 + \lambda \left\|\Phi f\right\|_{l_1}}_{\mathcal{C}(f)}. \tag{2.10}$$

where the l_1 norm is defined as $\|f\|_{l_1} = \sum_i |f_i|$. The restricted isometric property (RIP) guarantees the accuracy of CS reconstruction if the sparsifying transform Φ and the undersampled Fourier transform $\mathcal{A}$ satisfies certain conditions. When the signal is sparse in its domain, i.e. $\Phi = I$, suppose that there is a constant δ_s of the operator $\mathcal{A}$, such that

$$(1 - \delta_s)\|f\|_{l_2}^2 \leq \|\mathcal{A}f\|_{l_2}^2 \leq (1 + \delta_s)\|f\|_{l_2}^2 \tag{2.11}$$

holds for all sparse vectors f with s non-zero coefficients [33], $\mathcal{A}$ is considered to satisfy RIP. Essentially, the aim of RIP is to define an incoherent sampling scheme, so that the operator $\mathcal{A}$ behaves almost like an orthogonal matrix when the data f is sparse. Fig. 2.7 illustrates RIP intuitively. f is a sparse signal, the random undersampling results in incoherent aliasing artifact, where the energy approximates the original signal energy. However, the equispaced undersampling leads to a coherent aliasing, which violates the RIP. When the sparsifying transform Φ is not identity, RIP requires the matrix $E = \mathcal{A}\Phi^{-1}$ to satisfy the condition (2.11).

2.2.1 Sparsifying Transforms

Most MR images exhibit the property of sparsity. Some images such as MR angiography are sparse in the original pixel domain. However, most MR images are

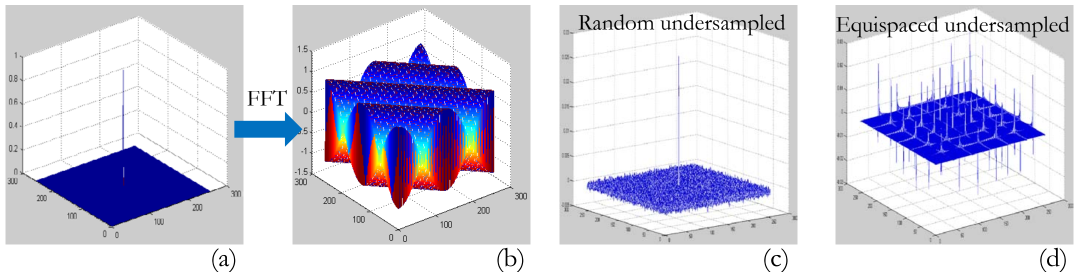


Figure 2.7: f is a sparse signal (a), the random undersampling results in incoherent aliasing artifact (c), where $\|\mathcal{A}f\|_{l_2}^2$ remains approximately the same. However, the equispaced undersampling leads to a coherent aliasing (d), which violates the RIP.

implicitly sparse in other domains. For example, most brain MR images, which are approximately piecewise constant, have sparse finite differences. Another commonly used sparsifying transform are wavelet transforms [34], which also plays an important role in conventional image compression. 1-D wavelet transforms are especially successful in capturing the discontinuites in a 1-D signal. However, since classical 2-D wavelet transforms are obtained by applying separable 1-D transforms in different dimensions, they are limited in their ability to capture the point-like features rather than the contours of the edge. Alternative directional multi-resolution transforms such as curvelet [10], contourlet [11], ridgelet [12], which are referred to as “x-let” transforms, achieve very good performance by capturing curvilinear features of an image [35]. However, the computational complexity of these transforms is high, which is undesirable in practical applications.

In the recent years, the total variation based methods have become very popular. Total variation promotes the sparsity of the image gradient. Hence, it is good at preserving the edges of the image. Moreover, it is simple to implement and the computational time is very fast, compared to the x-let sparsifying transforms. However, the limitation of TV reconstruction is that the reconstructed images often exhibit patchy or painting-like artifacts.

2.2.2 Optimization Algorithms

There are a large number of optimization algorithms introduced to solve the problem in Eq. (2.10). Generally, these algorithms fall in three categories: discrete algorithms, convex algorithms, and majorize-minimize (MM) algorithms, as illustrated in Fig. 2.2.2.

$$\text{Optimization algorithms}\begin{cases}\text{Discrete Methods}\begin{cases}\text{MP}\\ \text{OMP}\\ \cdots\end{cases}\\ \text{Convex Programs}\begin{cases}\text{LP}\\ \text{SDP}\\ \cdots\end{cases}\\ \textbf{Majorize-Minimize}\begin{cases}\text{ISTA}\\ \text{FISTA}\\ \cdots\end{cases}\end{cases}$$

Figure 2.8: Reconstruction algorithms for regularized optimization problem

Discrete Methods

Most discrete reconstruction algorithms are greedy algorithms, which are a family of heuristic methods that compute a local optimal solution at each stage in order to find the global optimal answer at the end. The idea of greedy algorithms can be traced back to Mallat et.al. [36] in 1993, where he put forward the matching pursuit (MP) method. The algorithm decomposes a signal into a linear combination of waveforms, which are chosen from a redundant dictionary of functions. The waveforms are then selected to match the signal optimally. This method is proved to have good approximation property [37] and converges for any signal in the dictionary space. Other greedy methods based on MP are also developed [38; 39; 40]. However, there are two primary drawbacks of MP related methods [41]. Firstly, a prior sparsity level parameter is required, which is unknown under most practical circumstances. Secondly, the algorithms are not robust to noise. These limitations have motivated the development of other types of optimization algorithms.

Convex Programs

Convex optimization studies the problem of finding the minimum of a convex function over a convex set. One of the most common types of convex optimization problems is one with linear constraints and a linear objective, which is called a linear program (LP) [42]. With the development of convex optimization, semidefinite programs (SDP) [43] become widely used in compressed sensing [43]. However, SDP based algorithms are computationally inefficient for large scale problems. Beyond these computation concern, SDP is not applicable to optimization problems with combined penalties.

Majorize-Minimize Methods

Because of the limitations of the previously mentioned algorithms, we focus on majorize-minimize (MM) algorithms. The main advantage of MM algorithms is that they replace the original difficult regularized optimization problem by a sequence of easier quadratic surrogate problems. The surrogate criteria, denoted by $\mathcal{S}^{(m)}(f)$, majorize the original objective function $\mathcal{C}(f)$, and are dependent on the current iterate $f^{(m)}$:

$$\mathcal{C}(f) \leq \mathcal{S}^{(m)}(f), \forall f; \quad \mathcal{S}^{(m)}(f^{(m)}) = \mathcal{C}\left(f^{(m)}\right). \tag{2.12}$$

Thus, the m^{th} iteration of the MM algorithm involves the following two steps (i) evaluate the majorizing criterion $\mathcal{S}^{(m)}(f)$ that satisfy (2.12), and (ii) solve for $f^{(m+1)}(\mathbf{r}) = \arg\min_f \mathcal{S}^{(m)}(f)$ using an appropriate quadratic solver (e.g. conjugate gradients (CG) algorithm). Fig. 2.9 shows the basic concept of MM algorithm.

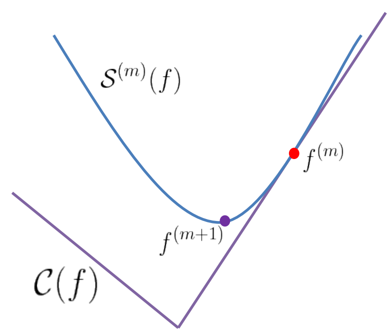


Figure 2.9: Illustration of MM algorithm. The goal is to minimize the cost function $\mathcal{C}(f)$. Using a surrogate function $\mathcal{S}^{(m)}(f)$ to maximize $\mathcal{C}(f)$, the minimization of the surrogate function is used to find the next iteration. By successively minimizing the surrogate function, the minima of $\mathcal{C}(f)$ can be obtained.

One of the special cases of MM algorithm is the Expectation Maximization (EM) algorithm, where there are two steps: E (expectation) step and M (maximization) step. In the E step, the conditional expectation of the data log likelihood is computed, which basically creates a minorizing surrogate function. In the M step, minorizing surrogate function is maximized. Therefore, EM algorithm is essentially an example of MM algorithm [44]. There are other extensions of

MM algorithms such as iterative shrinkage thresholding algorithm (ISTA) and fast ISTA (FISTA) [45; 46].

3 Higher Degree Total Variation (HDTV) regularization

The reconstruction of images from their noisy and ill-conditioned linear measurements is an important problem. The standard approach is to pose the image recovery as an optimization problem, where the criterion is a linear combination of data consistency error and a regularized penalty. Total variation smoothness prior is a popular regularization penalty. The advantages of TV include simplicity, invariance to shift and rotations, and capability of preserving edges and providing good image quality. However, the limitation of TV is that it introduces patchy or painting-like artifacts. To overcome this problem, in this chapter we introduce a higher degree TV penalty using the steerability of higher degree image derivatives. We propose an iteratively reweighted majorize minimize algorithm to solve the HDTV regularized recovery problem efficiently. We demonstrate the utility of the proposed algorithm in compressed sensing, denoising, and deblurring. The results show that HDTV significantly reduces the amount of patchy artifacts and preserves ridge-like image features, compared to standard TV regularization and other current regularization methods.

3.1 Introduction

Rudin et al., have introduced the total variation scheme firstly, where the penalty is the L_1 norm of the gradient magnitude of the signal [13; 14]:

$$\mathcal{J}_1(f) = \int_{\mathbf{r}} \underbrace{\sqrt{\left(\frac{\partial f(\mathbf{r})}{\partial x}\right)^2 + \left(\frac{\partial f(\mathbf{r})}{\partial y}\right)^2}}_{|\nabla f(\mathbf{r})|} \, d\mathbf{r}, \tag{3.1}$$

Inspite of the isotropic definition, the TV regularizer results in anisotropic 1-D smoothing. Specifically, the Euler-Lagrange equation of (2.10) is given by [47]:

$$2\,\mathcal{A}^* \left(\mathcal{A}(f) - b\right) - \lambda \, \frac{f_{\theta^\perp,2}}{|\nabla f|} = 0. \tag{3.2}$$

Here, $f_{\theta^\perp,2}$ is the second derivative of f in the direction orthogonal to the gradient (edge) and $\mathcal{A}^*$ is the adjoint of the operator $\mathcal{A}$. Here, θ is the direction of the gradient. The second term in the above equation corresponds to smoothing along the edge (orthogonal to the gradient). Note that the smoothing across the edge (in the direction of the gradient) is completely attenuated. This one dimensional smoothing property ensures the preservation of sharp image edges in TV regularized reconstructions. Total variation based algorithms are widely used in remote sensing [48], biomedical imaging [49], astronomy [50], and radar imaging.

Inspite of its desirable properties, the TV regularizer has some limitations that restrict its performance in practical applications. The main challenge is its poor approximation property. Steidl et. al have shown that one dimensional TV denoising is equivalent to approximating the noisy signal by a non-uniform spline of degree zero [51; 52]. Since the approximation ability of this representation is poor, TV regularization often results in patchy or painting-like reconstructions that are visually unappealing.

Wavelet representations overcome similar problems by using basis functions that behave like higher order derivative operators [53; 54]. Inspired by the success of such wavelet schemes, we propose to replace the conventional gradient operator by higher order differentials to improve the approximation order. While this extension is straightforward for one dimensional signals, one can construct several penalties involving higher order multidimensional partial derivatives; several such penalties were recently introduced in the context of denoising [52; 55; 56; 57; 58; 59]. Some of these choices may be inappropriate for regularizing inverse problems. For example, the ℓ_1 norm of the Laplacian, which is introduced for denoising [57; 58; 52; 59], has a large kernel [60]; the use of this functional to regularize ill-conditioned inverse problems may still result in ill-posed problems. We are interested in deriving functionals that inherit the desirable properties of the standard TV regularizer.

Based on the steerability of the directional derivatives, we re-interpretate the classical TV regularizer to derive two families of multidimensional higher degree total variation (HDTV) penalties. We term them as **(a)** isotropic and **(b)** anisotropic penalties, respectively. We first interpret the TV functional as the L_1-L_2 penalty of the directional derivatives of the image, along all possible directions. We use this re-interpretation to derive the isotropic HDTV penalty. These functionals have analytical expressions, thanks to the rotation steerability of directional derivatives. L_1-L_2 mixed norms are often used to enhance joint sparsity [61; 62]; the use of this norm encourages all the directional derivatives at any specified voxel to be simultaneously zero or non-zero. Since the simultaneous attenuation of the directional derivatives encourages isotropic smoothing, we term this family as isotropic HDTV penalties. We also generalize this class by considering the L_1-L_2 norm of the rotated versions of general differential operators; we observe that this class of generalized functionals contains some of the current higher degree TV-like penalties [55; 56]. We use the re-interpretation of the TV functional as the separable L_1-L_1 norm of the directional derivatives to derive the class of anisotropic HDTV penalties. Since these penalties are fully separable, the presence of a strong edge/ridge singularity at a specified orientation will not attenuate the smoothing along the edge/ridge. Thus, this property encourages anisotropic smoothing, even-though it is invariant to rotations unlike classical anisotropic TV penalty [13].

3.2 Isotropic HDTV regularization

We now reinterpret the TV functional as a group separable L_1-L_2 norm of directional derivatives of the specified image. This interpretation enables us to generalize the standard TV scheme to higher degree derivatives.

3.2.1 Steerability of directional derivatives

We denote the derivative of a function along the direction specified by the unit vector $\mathbf{u}_\theta = (\cos(\theta), \sin(\theta))$ as

$$f_{\theta,1}(\mathbf{r}) = \frac{\partial}{\partial \gamma} f(\mathbf{r} + \gamma\, \mathbf{u}_\theta). \tag{3.3}$$

Specifically, we have $f_0 = \partial f/\partial x$ and $f_{\frac{\pi}{2}} = \partial f/\partial y$. Directional derivatives are rotation steerable [63; 64]; i.e., the derivative along any direction can be expressed as the linear combination:

$$f_{\theta,1}(\mathbf{r}) = f_0(\mathbf{r}) \cos(\theta) + f_{\frac{\pi}{2}}(\mathbf{r}) \sin(\theta). \tag{3.4}$$

This expression is compactly represented in the vector form:

$$f_{\theta,1}(\mathbf{r}) = \underbrace{[\cos(\theta), \sin(\theta)]}_{\mathbf{s}_1^H(\theta)} \underbrace{\begin{bmatrix} \partial f(\mathbf{r})/\partial x \\ \partial f(\mathbf{r})/\partial y \end{bmatrix}}_{\mathbf{g}_1(\mathbf{r})}. \tag{3.5}$$

Similarly, the n^{th} order directional derivative, specified by $f_{\theta,n}(\mathbf{r}) = \frac{\partial^n}{\partial \gamma^n} f(\mathbf{r}+\gamma\, \mathbf{u}_\theta)$, is rotation steerable as $f_{\theta,n}(\mathbf{r}) = \mathbf{s}_n^H(\theta)\, \mathbf{g}_n(\mathbf{r})$. Here, $\mathbf{g}_n(\mathbf{r})$ is the vector of n^{th} order partial derivatives, while $\mathbf{s}_n(\theta)$ is the vector of trigonometric polynomials. In the second degree case, we have $f_{\theta,2}(\mathbf{r}) = \mathbf{s}_2^H(\theta)\, \mathbf{g}_2(\mathbf{r})$, where

$$\mathbf{s}_2(\theta) = [\ \sin(\theta)^2 \quad 2\sin(\theta)\cos(\theta) \quad \cos(\theta)^2\ ]^T \tag{3.6}$$

$$\mathbf{g}_2(\mathbf{r}) = [\ \tfrac{\partial^2 f(\mathbf{r})}{\partial x^2} \quad \tfrac{\partial^2 f(\mathbf{r})}{\partial x \partial y} \quad \tfrac{\partial^2 f(\mathbf{r})}{\partial y^2}\ ]^T. \tag{3.7}$$

3.2.2 Reinterpretation of TV regularization

Proposition 1. *The gradient based regularizer, specified by*

$$\mathcal{J}_1(f) = \int_{\mathbf{r}} |\nabla f(\mathbf{r})|\, d\mathbf{r}, \tag{3.8}$$

can be expressed as a group separable penalty of the directional derivatives of f:

$$\frac{1}{\sqrt{2}} \int_{\mathbf{r}} |\nabla f(\mathbf{r})|\, d\mathbf{r} = \int_{\mathbb{R}^2} \underbrace{\sqrt{\frac{1}{2\pi} \int_0^{2\pi} |f_{\theta,1}(\mathbf{r})|^2 d\theta}}_{\|f_{\theta,1}(\mathbf{r})\|_{L_2[0,2\pi]}}\, d\mathbf{r}. \tag{3.9}$$

Proof. Using the steerability of first degree directional derivatives, we simplify the

$L_2[0, 2\pi]$ norm of the directional derivatives as

$$\begin{aligned}
\|f_{\theta,1}(\mathbf{r})\|^2 &= \frac{1}{2\pi}\int_0^{2\pi} \left|\mathbf{s}_1^H(\theta)\mathbf{g}_1(\mathbf{r})\right|^2 d\theta \\
&= \mathbf{g}_1^H(\mathbf{r}) \left(\frac{1}{2\pi}\int_0^{2\pi} \mathbf{s}_1(\theta)\, \mathbf{s}_1^H(\theta)\, d\theta\right) \mathbf{g}_1(\mathbf{r}) \\
&= \mathbf{g}_1^H(\mathbf{r}) \underbrace{\frac{1}{2\pi}\int_0^{2\pi} \begin{bmatrix} \cos^2(\theta) & \sin(2\theta)/2 \\ \sin(2\theta)/2 & \sin^2(\theta) \end{bmatrix} d\theta}_{\frac{1}{2}\mathbf{I}}\ \mathbf{g}_1(\mathbf{r}) \\
&= \frac{1}{2}\, \mathbf{g}_1^H(\mathbf{r})\mathbf{g}_1(\mathbf{r}) = |\nabla f|^2/2.
\end{aligned}$$

Substituting this relation in (3.9), we obtain the equivalence. □

3.2.3 Isotropic Higher Degree TV (I-HDTV)

Based on the above reinterpretation, we introduce the isotropic n^{th} degree TV regularizer as

$$\mathcal{J}_n(f) = \int_{\mathbb{R}^2} \|f_{\theta,n}(\mathbf{r})\|_{L_2[0,2\pi]}\, d\mathbf{r}. \tag{3.10}$$

Since we are summing the square magnitude of the directional derivatives of the function along all directions and orientations, this penalty is invariant to rotations and translations and is also convex. Note that (3.10) is the L_1-L_2 mixed norm of the directional derivatives. Such mixed norms are often used in compressed sensing to exploit the joint sparsity of the coefficients [61; 62]. Specifically, it encourages the coefficients that are grouped by the L_2 norms to be zero or non-zero at the same time. Thus, the presence of a strong directional derivative at a specified orientation will encourage the preservation of small directional derivatives along other directions at that specific pixel. Ideally, the strong directional derivatives need to be preserved, while the small ones at other directions need to be attenuated to encourage smoothing along line-like features, thus enhancing them, similar to the standard TV scheme. Note that eventhough the standard TV has an L_1-L_2 interpretation, it exhibits anisotropic smoothing since it can also be interpreted as a fully separable L_1-L_1 penalty, as discussed in the next section. This dual interpretation is unique to the first order TV case and does not extend to higher order derivatives. Since, (3.10) do not inherit the anisotropic smoothing properties of the classical TV regularizer, we term these class of functionals as isotropic HDTV penalty.

Since the only functions for which all directional derivatives vanish are polynomials of degree $n-1$, the kernel associated with $\mathcal{J}_n(f)$ is small. Hence, the regularization of ill-conditioned inverse problems using such penalties will be well-posed. The L_1 norm preserves the directional derivatives in regions with high directional energy (specified by $\|f_{\theta,n}\|_{L_2[0,2\pi]}$), thus preserving the edges/ridges in the image. The use of higher degree derivatives in the criterion will enable the representation of the signal as piecewise polynomials, thus providing representations with improved approximation properties.

We now use the steerability of the directional derivatives to derive analytical expressions for the isotropic HDTV regularizer. Specifically, the $L_2[0,2\pi]$ norm of the n^{th} degree directional derivatives are given by

$$\begin{aligned} \|f_{\theta,n}(\mathbf{r})\|_{L_2} &= \sqrt{\frac{1}{2\pi}\int_0^{2\pi} |f_{\theta,n}(\mathbf{r})|^2 \, d\theta} \\ &= \sqrt{\mathbf{g}_n(\mathbf{r})^H \underbrace{\left(\frac{1}{2\pi}\int_0^{2\pi} \mathbf{s}_n(\theta)\mathbf{s}_n(\theta)^H \, d\theta\right)}_{\mathbf{C}_n} \mathbf{g}_n(\mathbf{r})} \\ &= \sqrt{\mathbf{g}_n(\mathbf{r})^H \mathbf{C}_n \mathbf{g}_n(\mathbf{r})}. \end{aligned} \quad (3.11)$$

Here $\mathbf{C}_n$ is a matrix with entries $c_{i,j} = \frac{1}{2\pi}\int_0^{2\pi} s_i(\theta)s_j(\theta)d\theta$; $i,j = 0,...,n$. Substituting (3.11) into (3.10), we obtain the n^{th} degree HDTV penalty as

$$\mathcal{J}_n(f) = \int_{\mathbb{R}^2} \sqrt{\mathbf{g}_n(\mathbf{r})^H \, \mathbf{C}_n \, \mathbf{g}_n(\mathbf{r})} \, d\mathbf{r} \tag{3.12}$$

We will now consider the special cases of second and third degree HDTV to illustrate the above expression.

Isotropic second degree TV

The coefficients in the steerability relation are given by $\mathbf{s}_2(\theta) = [\cos^2(\theta), \sin(2\theta), \sin^2(\theta)]^T$. Thus, the symmetric matrix $\mathbf{C}_2$ in the second degree TV functional is specified by

$$\mathbf{C}_2 \;=\; \frac{1}{2\pi} \int_0^{2\pi} \mathbf{s}_2(\theta) \mathbf{s}_2(\theta)^H d\theta = \frac{1}{8} \begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix}. \tag{3.13}$$

Substituting back in (3.12), we obtain

$$\mathcal{J}_2(f) = \int_{\mathbb{R}^2} \sqrt{\left(3\left|f_{xx}\right|^2 + 3\left|f_{yy}\right|^2 + 4\left|f_{xy}\right|^2 + 2\Re\left(f_{xx} f_{yy}\right)\right)/8} \; d\mathbf{r}. \tag{3.14}$$

Here, $\Re(f)$ denotes the real part of f.

Isotropic third degree TV

Using the steerability relation of the 3^{rd} order derivative operator, we obtain $\mathcal{J}_3(f) = \int \sqrt{q(\mathbf{r})} d\mathbf{r}/4\sqrt{2}$, where

$$q(\mathbf{r}) = 5\left(\left|f_{xxx}\right|^2 + \left|f_{yyy}\right|^2\right) + 6\Re\left(f_{xxx} f_{xyy} + f_{yyy} f_{xxy}\right) + \; 9\left(\left|f_{xxy}\right|^2 + \left|f_{xyy}\right|^2\right) \tag{3.15}$$

3.2.4 Majorize-Minimize (MM) algorithm for I-HDTV

The HDTV recovery scheme is thus specified by

$$\hat{f}(\mathbf{r}) = \arg\min_{f} \ \|\mathcal{A}(f) - b\|^2 + \lambda \int_{\mathbb{R}^2} \sqrt{\mathbf{g}_n(\mathbf{r})^H\, \mathbf{C}_n\, \mathbf{g}_n(\mathbf{r})}\, d\mathbf{r}. \tag{3.16}$$

We extend the classical iterative reweighted MM formulation to (3.16) to obtain

$$f^{(m+1)}(\mathbf{r}) \ = \ \arg\min_{f} \ \|\mathcal{A}(f) - b\|^2 + \lambda \int_{\mathbb{R}^2} \mathbf{g}_n(\mathbf{r})^H \mathbf{D}_n^{(m)}(\mathbf{r})\, \mathbf{g}_n(\mathbf{r})\, d\mathbf{r}. \tag{3.17}$$

The entries of the weighting matrix $\mathbf{D}_n^{(m)}$ are spatially modulated by the scalar weighting term $\phi_n^{(m)}(\mathbf{r})$:

$$\mathbf{D}_n^{(m)}(\mathbf{r}) = \underbrace{\frac{1}{2\sqrt{{\mathbf{g}_n^{(m)}}^H(\mathbf{r})\mathbf{C}_n^{(m)}\mathbf{g}_n^{(m)}(\mathbf{r})}}}_{\phi_n^{(m)}(\mathbf{r})} \mathbf{C}_n. \tag{3.18}$$

The spatial weights $\phi_n^{(m)}(\mathbf{r})$ are inversely proportional to the directional energy ($\|f_{\theta,n}^{(m)}(\mathbf{r})\|_{L_2[0,2\pi]}$) at the specified location $\mathbf{r}$. The modulation of the quadratic functional by these weights suppresses the regularization in spatial regions with strong n^{th} order singularities, thus enabling the preservation of edges/ridges. Since (3.17) is a quadratic criterion, we solve it efficiently using the conjugate gradient algorithm. The gradient of (3.17) has an analytical expression:

$$\nabla\mathcal{C}^{(m)} = 2\mathcal{A}^*(\mathcal{A}(f) - b) + 2\lambda\, \boldsymbol{\partial}_n(\mathbf{r})^H * \mathbf{D}_n^{(m)}(\mathbf{r})\mathbf{g}_n(\mathbf{r}). \tag{3.19}$$

Here, $\boldsymbol{\partial}_n(\mathbf{r})$ is the vector of n^{th} degree differential operators and $\mathbf{D}_n^{(m)}(\mathbf{r})$ is a spatially varying diagonal matrix, which is obtained by multiplying $\mathbf{C}_n$ with $\phi_n^{(m)}(\mathbf{r})$. We now illustrate the algorithm in the context of first and second degree TV schemes.

Standard TV: isotropic MM algorithm

In the standard TV case, we have $\phi_1^{(m)}(\mathbf{r}) = 1/2|\nabla f^{(m)}(\mathbf{r})|$. Hence, the above expression simplifies to

$$\begin{aligned}
\nabla \mathcal{C}^{(m)} &= 2\mathcal{A}^*\left(\mathcal{A}(f) - b\right) + 2\lambda \underbrace{\left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right]}_{\boldsymbol{\partial}_1(\mathbf{r})^H} \overbrace{\frac{1}{2|\nabla f^{(m)}|}\, \frac{1}{2}\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{C}_1}}^{\mathbf{D}_1^{(m)}(\mathbf{r})} \underbrace{\begin{bmatrix} f_x(\mathbf{r}) \\ f_y(\mathbf{r}) \end{bmatrix}}_{\mathbf{g}_1(\mathbf{r})} \\
&= 2\mathcal{A}^*\left(\mathcal{A}(f) - b\right) + \lambda \nabla \cdot \left(\frac{\nabla f(\mathbf{r})}{2|\nabla f^{(m)}(\mathbf{r})|}\right) \qquad (3.20)
\end{aligned}$$

Since the smoothing at each location is attenuated by $1/|\nabla f^{(m)}|$, the above scheme enables the preservation of singularities in the image. We illustrate the matrix $\mathbf{D}_1^{(m)}(\mathbf{r})$ in Fig. 3.3.3, when $f^{(m)}$ is a Gaussian blurred disk image. Note that $\mathbf{D}_1^{(m)}(\mathbf{r})$ is diagonal and both the diagonal entries are exactly the same, irrespective of the orientation of the edge.

Special case: second order TV

In the second order TV case, the above expression simplifies to

$$2\mathcal{A}^*\left(\mathcal{A}(f) - b\right) + 2\lambda \underbrace{\begin{bmatrix} \frac{\partial^2}{\partial x^2} \\ \frac{\partial^2}{\partial x \partial y} \\ \frac{\partial^2}{\partial y^2} \end{bmatrix}^H}_{\boldsymbol{\partial}_2(\mathbf{r})^H} \overbrace{\phi_2^{(m)}(\mathbf{r})\, \frac{1}{8} \underbrace{\begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix}}_{\mathbf{C}_2}}^{\mathbf{D}_2^{(m)}(\mathbf{r})} \underbrace{\begin{bmatrix} f_{xx}(\mathbf{r}) \\ f_{xy}(\mathbf{r}) \\ f_{yy}(\mathbf{r}) \end{bmatrix}}_{\mathbf{g}_2(\mathbf{r})} \qquad (3.21)$$

3.3 Rotation Invariant Anisotropic HDTV regularization

We now consider an alternate interpretation of the standard TV penalty, which allows us to develop a different class of rotation-invariant, anisotropic HDTV penalties.

3.3.1 Reinterpretation of TV regularization

Proposition 2. *The standard TV regularizer can be interpreted as a separable penalty of the directional derivatives of* f*:*

$$\int_{\mathbf{r}} |\nabla f(\mathbf{r})| \, d\mathbf{r} = \int_{\mathbb{R}^2} \underbrace{\frac{1}{4}\int_0^{2\pi} |f_{\theta,1}(\mathbf{r})| d\theta}_{\|f_{\theta,1}(\mathbf{r})\|_{L_1[0,2\pi]}} \, d\mathbf{r}. \tag{3.22}$$

Proof. The directional derivative at any specified location can be expressed as

$$f_\theta(\mathbf{r}) = |\nabla f(\mathbf{r})| \cos(\theta - \phi), \tag{3.23}$$

where, ϕ denotes the orientation of the gradient. Thus, we have

$$\|f_{\theta,1}(\mathbf{r})\|_{L_1} \;\; = \;\; |\nabla f(\mathbf{r})| \; \underbrace{\frac{1}{4}\int_0^{2\pi} |\cos(\theta - \phi)| \, d\theta}_{1} = |\nabla f(\mathbf{r})|. \tag{3.24}$$

Substituting this relation in (3.22), we obtain the equivalence. □

3.3.2 Anisotropic Higher Degree TV (A-HDTV)

We will now use the above interpretation to obtain a new family of higher degree total variation penalties:

$$\mathcal{G}_n(f) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_0^{2\pi} |f_{\theta,n}(\mathbf{r})| \, d\theta \, d\mathbf{r}. \tag{3.25}$$

Note that this penalty is completely separable unlike the $L_1 - L_2$ mixed norms that we considered in the earlier section. The separable formulation ensures that the presence of a strong directional derivative in a specified orientation will not prevent the attenuation of directional derivatives along other orientations. Thus, this criterion will ensure that an edge-like singularity along a specified orientation will not attenuate the smoothing in the direction orthogonal to the edge. This interpretation explains the anisotropic smoothing properties exhibited by the standard TV regularizer [47]. We expect this class of penalties to provide reconstructions with improved contour regularity and reduced blob-like artifacts, compared to the isotropic extension considered in the previous section. Since we consider all angles, this anisotropic penalty is rotation invariant; it enhances the edges along all the orientations in contrast to the classical anisotropic TV penalty [58].

The proof of proposition 2 can be extended to interpret the standard TV penalty as the $L_1 - L_p; p \geq 1$ penalty of oriented derivatives. Clearly, high values of p are less desirable since they give more isotropic results. Thus, $p = 1$ is the convex choice in this class, which provides the best anisotropic behavior. Unfortunately, the proposed anisotropic HDTV penalty ($p = 1$) does not have analytical expressions similar to the isotropic case ($p = 2$). However, we now show that we still can develop an MM algorithm that is conceptually similar to and shares the computational efficiency of the isotropic MM algorithm.

3.3.3 Majorize-Minimze (MM) algorithm for A-HDTV

We majorize the anisotropic HDTV criterion in (3.25) as

$$\mathcal{G}_n(f) \leq \mathcal{G}_n(f^{(m)}) + \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_0^{2\pi} \phi_n^{(m)}(\mathbf{r}, \theta) \left| f_{\theta,n}(\mathbf{r}) \right|^2 \, d\theta \, d\mathbf{r}, \tag{3.26}$$

where

$$\phi_n^{(m)}(\mathbf{r},\theta) = \frac{1}{2\sqrt{|f_{\theta,n}^{(m)}(\mathbf{r})|^2 + \epsilon}}. \tag{3.27}$$

Here, ϵ is an arbitrarily small constant to ensure that the above expression is well-defined. We discuss the choice of ϵ in the next section. In each outer iteration, we assume the modulation term $\phi_n^{(m)}(\mathbf{r},\theta)$ to be fixed, which depends on the current iterate $f^{(m)}(\mathbf{r})$. We now use the steerability of $f_{\theta,n}(\mathbf{r})$ to expand the second term in (3.26) as

$$\mathcal{G}_n^{(m)}(f) = \int_{\mathbb{R}^2} \mathbf{g}_n(\mathbf{r})^H \underbrace{\frac{1}{2\pi}\int_0^{2\pi} \mathbf{s}_n(\theta)\,\phi_n^{(m)}(\mathbf{r},\theta)\,\mathbf{s}_n^H(\theta)\,d\theta}_{\mathbf{B}_n^{(m)}(\mathbf{r})}\,\mathbf{g}_n(\mathbf{r})d\mathbf{r} \tag{3.28}$$

Here, $\mathbf{B}_n^{(m)}(\mathbf{r})$ is the spatially varying weighting matrix. Similar to the non-separable TV, we solve for the minimum of the equivalent majorized cost function:

$$f^{(m+1)}(\mathbf{r}) = \arg\min_f \ \|\mathcal{A}(f) - b\|^2 + \lambda \int_{\mathbb{R}^2} \mathbf{g}_n(\mathbf{r})^H \mathbf{B}_n^{(m)}(\mathbf{r})\,\mathbf{g}_n(\mathbf{r})\ d\mathbf{r}. \tag{3.29}$$

Note that this expression is similar to (3.17). However, the matrix $\mathbf{B}_n^{(m)}(r)$ is very different from $\mathbf{D}_n^{(m)}(r)$. $\mathbf{D}_n^{(m)}(\mathbf{r})$ is obtained by uniformly weighting all entries of $\mathbf{C}$ by $\phi^{(m)}(\mathbf{r})$. In contrast, the entries of $\mathbf{B}_n^{(m)}(\mathbf{r})$ are dependent on the directional weights $\phi^{(m)}(\mathbf{r},\theta)$. This weighting ensures anisotropic smoothing at each iteration of the MM algorithm. We propose to minimize the above quadratic expression using conjugate gradients algorithm. The gradient of the criterion is given by

$$\nabla\mathcal{C}^{(m)} = 2\mathcal{A}^*\left(\mathcal{A}(f) - b\right) + 2\lambda\boldsymbol{\partial}_n(\mathbf{r})^H\left(\mathbf{B}_n^{(m)}(\mathbf{r})\,\mathbf{g}_n(\mathbf{r})\right), \tag{3.30}$$

where matrix $\mathbf{B}_n^{(m)}(\mathbf{r})$ is re-evaluated at each iteration. We now illustrate the properties of this matrix in the context of first and second order derivatives.

Standard TV: Anisotropic MM algorithm

The matrix $\mathbf{B}_1^{(m)}$ is specified by

$$\mathbf{B}_1^{(m)}(\mathbf{r}) = \frac{1}{2\pi}\begin{bmatrix} \int_0^{2\pi}\phi_1^{(m)}(\mathbf{r},\theta)\cos(\theta)^2 d\theta & \int_0^{2\pi}\phi_1^{(m)}(\mathbf{r},\theta)\sin(2\theta)/2d\theta \\ \int_0^{2\pi}\phi_1^{(m)}(\mathbf{r},\theta)\sin(2\theta)/2d\theta & \int_0^{2\pi}\phi_1^{(m)}(\mathbf{r},\theta)\sin(\theta)^2 d\theta \end{bmatrix} \tag{3.31}$$

Unlike the isotropic case, the trigonometric functions within the integrals are weighted by $\phi_n^{(m)}(\mathbf{r},\theta)$. In the first order case, this algorithm gives exactly the same solution as the isotropic algorithm, since the criterion is the same (see Proposition 2).

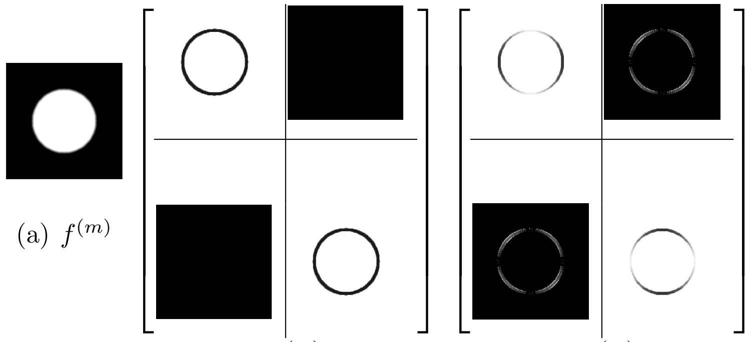
To illustrate the algorithm, we consider a voxel, where the gradient is in the horizontal direction; i.e, $f_y = 0$. We thus have $\phi_n^{(m)}(\mathbf{r},\theta) = 1/2\sqrt{\epsilon + |\nabla f|^2\cos^2(\theta)}$. Substituting in (3.31) and assuming that $\epsilon/|\nabla f^{(m)}| = 10^{-15}$, we get

$$\mathbf{B}_1^{(m)}(\mathbf{r}) = \frac{1}{2|\nabla f^{(m)}|}\begin{bmatrix} 1.27 & 0 \\ 0 & 43.5 \end{bmatrix}. \tag{3.32}$$

Note that the weighting of f_{yy} is approximately 34 times stronger than that of f_{xx}, which makes it very different from (3.20). We show the matrix $\mathbf{B}_1^{(m)}(\mathbf{r})$ corresponding to a Gaussian blurred disk in Fig. 3.3.3 (c). Note that the diagonal entries are not the same as in the case of $\mathbf{D}_1^{(m)}(\mathbf{r})$. The weights depend on the orientation of the edge, resulting in continued smoothing along the edges.

Anisotropic 2^{nd} degree TV

Consider a point along an image ridge, corresponding to $f_{xx}^{(m)}(\mathbf{r}) = 1; f_{yy}^{(m)}(\mathbf{r}) = \frac{1}{100}; f_{xy}^{(m)}(\mathbf{r}) = 0$. Thus, we have $\phi_2^{(m)}(\mathbf{r},\theta) = 1/2|\cos^2(\theta) + 1/100\sin^2(\theta)|$. Substi-



(a) $f^{(m)}$ (b) $\mathbf{D}_1^{(m)}(\mathbf{r})$ (c) $\mathbf{B}_1^{(m)}(\mathbf{r})$

Figure 3.1: Illustration of the weighting matrices $\mathbf{D}_1^{(m)}$ and $\mathbf{B}_1^{(m)}$, when $f^{(m)}(\mathbf{r})$ is a simple blurred disk image as in (a). We show the entries of the matrices as images. Note that $\mathbf{D}_1$ is obtained by uniformly weighting $\mathbf{C}_1$ by $1/|\nabla f^{(m)}|$. The gradients are heavily weighted in most regions, except on the edges. In contrast, the entries of $\mathbf{B}_1^{(m)}(\mathbf{r})$ depends on the orientation of the edge. Note that the weights for the horizontal derivatives are not attenuated and horizontal weights are attenuated in the top and the middle of the disk.

tuting in (3.28), we get

$$\mathbf{B}_2^{(m)}(\mathbf{r}) = \frac{1}{8}\begin{bmatrix} 0.99 & 0 & 0.83 \\ 0 & 2.52 & 0 \\ 0.83 & 0 & 17.35 \end{bmatrix} \tag{3.33}$$

Compared to $\mathbf{D}_2^{(m)}(\mathbf{r})$ in (3.21), we find that this matrix weights the partial derivative along y much more heavily (≈ 17 fold) than the one along x. This results in increased smoothing along the ridge, thus preserving it.

3.4 Numerical Implementation

3.4.1 Discretization of derivatives

To realize efficient numerical algorithms, TV schemes approximate the signal derivatives with simple finite difference filters. While longer filters provide improved approximation, their use in TV regularization will result in ringing and oscillations (similar to wavelet filters). The finite difference filter can be interpreted as the samples of the first derivative of a Bspline of degree 1, evaluated at the samples $k+\delta, \delta \in (0,1)$. This convolution can also be interpreted as evaluating the derivative of $f(x)$ at the locations $k+\delta; k \in \mathbb{Z}$.

Extending this approach to higher order derivatives, we get:

$$g_n(k) = \beta_n^n(k+\delta) * f[k], \tag{3.34}$$

where $\beta_n^n(x)$ is the n^{th} derivative of $\beta^n(x)$: the Bspline of degree n. We compared the performance of the corresponding HDTV algorithms for different values of δ and found that setting $\delta = 1/4$ provides the best results. Hence, we use this parameter for the rest of the experiments in this chapter. We extend this definition to multidimensional derivatives using derivatives of tensor product of Bspline functions. Since tensor product Bspline functions are not isotropic in the strict sense, its partial derivatives are not steerable. However, Bspline windows become more isotropic as the order of the Bspline increases, provided the Bspline orders are the same along different orientations. Thus, derivatives of higher degree Bspline windows are approximately steerable. Hence, we propose to use Bsplines of same degree along x and y dimensions to enhance the steerability of the derivative operator.

$$g_{n_1,n_2}[k_1,k_2] = \underbrace{\beta_{n_1}^d(k_1+\delta) \otimes \beta_{n_2}^d(k_2+\delta)}_{\varphi(k_1,k_2)} * f[k_1,k_2], \tag{3.35}$$

where $d = n_1+n_2$ is the degree of the Bspline. These discrete derivative operators are only approximately steerable.

3.4.2 Evaluation of $\mathbf{B}_n^{(m)}(\mathbf{r})$ in (3.28)

We compute $\mathbf{B}_n^{(m)}$ by discretizing $\phi_n^{(m)}(\mathbf{r}, \theta)$ on a uniform grid and evaluating the Reimann sum. Since the computation of the matrix at a specified voxel is independent of its neighbors, these computations can be efficiently parallelized. Our experiments using different number of samples show that 50-100 angles in the range $0 - \pi$ are sufficient for a good approximation. We also use the symmetry of the directional derivatives to accelerate the computation.

3.4.3 Choice of the parameters to improve convergence

The convergence rate of the algorithm is dependent on the parameter ϵ. Low values of ϵ result in the matrices $\mathbf{B}_2$ and $\mathbf{D}_2$ being ill-defined. Since this results in poorly conditioned quadratic subproblems, the corresponding conjugate gradient algorithms will converge slowly. In contrast, the solution of the quadratic subproblems are poor approximations to the solution of the original non-quadratic problem, when large values of ϵ are used. To overcome this tradeoff, we rely on a continuation strategy. Specifically, we initialize ϵ with a large value and gradually decrease it to a small value. In this work, we initialize ϵ with 10^{-3} and decrease it by $\epsilon_{inc} = 0.5$ in each outer iteration. We observe that this approach significantly improves the convergence, while retaining the accuracy of the final result. The pseudocode for the corresponding isotropic and anisotropic algorithms are shown below.

We typically use twenty outer iterations (MaxOuterIterations=20) and a maximum of fifty CG steps per outer iteration to solve for (3.17) or (3.29). The CG algorithm is terminated when the relative change in the cost function is less than a specified threshold. We observe that we need many CG steps for initial outer iterations, while the number of CG steps are far smaller (2-3) for the later outer iterations. The total number of CG steps needed is dependent on the conditioning of the problem. In general, we need around 200-500 CG steps for the entire algorithm to converge.

Algorithm 3.4.1: $\text{ISOTROPICHDTV}(A, b, \lambda)$

$i \leftarrow 1$

$\epsilon \leftarrow \epsilon_{\text{init}}, f^{(1)} \leftarrow A^H(b)$

while i < MaxContinuationIterations

do
- $m \leftarrow 1$
- **while** m < MaxInnerIterations
- **do**
  - Compute $\mathbf{g}_n(\mathbf{r})$ using (3.35)
  - $f_{\theta,n}(\mathbf{r}) = \mathbf{s}_n^H(\theta)\,\mathbf{g}_n(\mathbf{r})$
  - Set $\phi_n^{(m)}(\mathbf{r}) = \frac{1}{2\sqrt{\epsilon + \|f_{\theta,n}^{(m)}(\mathbf{r})\|^2}}$
  - Compute $\mathbf{D}_n^{(m)}(\mathbf{r})$ using (3.18)
  - Update $f^{(m+1)}(\mathbf{r})$ using (3.17)
  - $m \leftarrow m + 1$
- $\epsilon \leftarrow \epsilon * \epsilon_{\text{incfactor}}$
- $i \leftarrow i + 1$

return (f)

Algorithm 3.4.2: ANISOTROPICHDTV(A, b, λ)

$i \leftarrow 1$

$\epsilon \leftarrow \epsilon_{\text{init}}, f^{(1)} \leftarrow A^H(b)$

while i < MaxContinuationIterations

do $\begin{cases} m \leftarrow 1 \\ \textbf{while } \text{m} < \text{MaxInnerIterations} \\ \textbf{do } \begin{cases} \text{Compute } \ \mathbf{g}_n(\mathbf{r}) \ \text{ using (3.35)} \\ f_{\theta,n}(\mathbf{r}) = \mathbf{s}_n^H(\theta)\,\mathbf{g}_n(\mathbf{r}) \\ \text{Set } \ \phi_n^{(m)}(\mathbf{r},\theta) = \frac{1}{2\sqrt{|f_{\theta,n}^{(m)}(\mathbf{r})|^2+\epsilon}} \\ \text{Compute } \ \mathbf{B}_n^{(m)}(\mathbf{r}) \ \text{ using (3.28)} \\ \text{Update } \ f^{(m+1)}(\mathbf{r}) \ \text{ using (3.29)} \\ m \leftarrow m+1 \end{cases} \\ \epsilon \leftarrow \epsilon * \epsilon_{\text{incfactor}} \\ i \leftarrow i+1 \end{cases}$

return (f)

3.5 Results on Image Recovery Problems

We determine the utility of the isotropic and rotation invariant anisotropic HDTV schemes in the context of three challenging applications: compressed sensing, deblurring and image denoising. In all the cases, we choose the regularization parameter λ such that $\|\mathcal{A}(\hat{f}) - b\|^2 \approx \sigma^2$. We compute the signal to noise ratio (SNR) of the reconstructions as

$$\text{SNR} = -10\log_{10}\left(\frac{\|f_{\text{orig}} - \hat{f}\|_F^2}{\|f_{\text{orig}}\|_F^2}\right), \tag{3.36}$$

where $\hat{f}$ is the reconstructed image; f_{orig} is the original image; $\|\cdot\|_F$ is the Frobenius norm.

We compare the proposed isotropic (IHDTV2) and anisotropic HDTV (AHDTV2) methods with the following state of the art methods **(a)** standard TV, **(b)** Lysaker's second degree anisotropic TV [65], specified by $\mathcal{J}_{Ls1}(f) = \int_\Omega (|f_{xx}| + |f_{yy}|)\, d\mathbf{r}$ **(c)** sparse Laplacian regularization [57], **(d)** sparse wavelet regularization, and **(e)** sparse curvelet regularization.There are two flavors of second degree regularization terms in Lysaker's method, which are given by

$$\mathcal{J}_{Ls1}(f) = \int_\Omega (|f_{xx}| + |f_{yy}|)\, d\mathbf{r}, \tag{3.37}$$

and the Frobenius norm of the Hessian matrix:

$$\mathcal{J}_{Ls2}(f) = \int_\Omega \sqrt{|f_{xx}|^2 + |f_{yy}|^2 + |f_{xy}|^2 + |f_{yx}|^2} d\mathbf{r}. \tag{3.38}$$

We observe that both of the above regularizers give similar results, which is consistent with the observations of Lysaker et al., [65]. We use the definition in (3.37) for our comparisons. We approximate the partial derivatives in standard TV, sparse Laplacian, and Lysaker's method using finite differences, which is the standard practice [57; 65]. We used iterative reweighted algorithms were used to implement all of the above methods. Existing MATLAB toolboxes for curvelet [66] and SURE-let [67] shrinkages were used in the context of denoising.

To ensure fair comparisons between different methods, we optimize the regularization parameter in each case to obtain $\|A(\hat{f}_\lambda) - b\|^2 \approx \sigma^2$. Here f_λ is the reconstructed image with λ as the regularization parameter. We determine the optimal regularization parameter for each noise level, image, and algorithm using a simple bisection algorithm. The ground truth and the standard deviation of the noise process are available to us, since we simulate the image formation.

The determination of the optimal regularization parameter is challenging when the ground truth and the noise level are unknown. The choice of regularization functionals in such settings is a very actively researched area; there are several strategies including the L-curve method [68], cross-validation [69], and Stein's unbiased risk estimator (SURE) [70; 71]. We plan to use one of these methods to determine the optimal parameters in practical applications, when the ground truth and standard deviation of the noise process is unknown.

3.5.1 HDTV in Compressed Sensing

The recovery of images from their undersampled Fourier samples is an important problem in MRI [49; 72]. This approach is often used to reduce the acquisition time in time-critical scans, reduce motion artifacts, and improve spatio-temporal resolution. In the experiments in this thesis, we assume the measurements to be acquired using variable density random Fourier encoding; this sampling pattern is realized in 3-D MR imaging using random phase-encodes and choosing the readout axis to be orthogonal to the image plane [49; 72] (see Fig. 3.3 (b) for the pattern in one k-space plane). We consider four MR images (brain in both sagittal view and axial view, wrist and angiography) and two natural images (Lena and Peppers) to illustrate the algorithm. The natural images are used to illustrate the utility of HDTV scheme in recovering the smoothly varying image regions.

The reconstructions of sagittal brain MR image at accelerations of A=4.35 and A=2 are shown in Fig. 3.2. In (b) to (h), we show the reconstructions using different methods at an acceleration of 4.35. We observe that standard TV reconstruction results in patchy artifacts with some loss in fine details, while the Laplacian method provides reconstructions with blob or point like artifacts.

All reconstructions are blurred at such high accelerations. However, the differences between the methods are more evident. The reconstructions using Lysaker's penalty provides better results than TV, but results in the loss of details. We observe that the A-HDTV2 algorithm provides better recovery of the image features, compared with the I-HDTV2 scheme (see dotted blue arrows in the images in (f) to (h)). By smoothing along the line-like features, anisotropic penalty preserves these characteristics more effectively. The A-HDTV3 method (top row) preserves some of the details that are lost in second degree case (see green arrow), but results in increased blurring and lower SNR. The reconstructions at an acceleration of A=2 are shown in the bottom row. We observe that the A-HDTV2 method preserves the fine features better than the other methods (see blue arrows in (j) to (l)). In this setting, the A-HDTV2 scheme provides a 1.36 dB improvement in SNR over standard TV and 2.73 dB improvement over Lysaker's method.

Fig. 3.3 compares the reconstructions of Lena image at acceleration of A = 4.35. It is seen that the standard TV reconstruction results in painting-like staircase artifacts in the smooth facial regions. In addition, it results in the loss of detail in the hair regions. The Lysaker's method is not patchy compared with TV, while it tends to blur the facial and eye area (see blue arrows), resulting in a lower SNR. The proposed A-HDTV2 method provides reasonably good reconstructions in these regions, resulting in an improvement of around 0.85dB over the TV scheme. We observe that the Laplacian penalty results in excessive amplification of point-like features.Note that A-HDTV3 method provides visually similar results as the second degree counterparts. However, the computational cost of this method is more significant than the second degree methods. See VI.C for details.

The SNRs of the recovered images at various acceleration factors and signal to noise ratios are shown in Table. 3.1. We observe that the A-HDTV2 method provides the best overall SNR for most of the cases. The TV scheme provides reconstructions that are 0.06dB better than A-HDTV2 for the axial brain MRI data in high noise setting. Note that this image is more or less piecewise constant. We observe that the Lysaker and Laplacian reconstructions have lower SNR than classical TV in the compressed sensing setting.

Acceleration	**2.00**		**2.85**		**4.35**		**2.00**		**2.85**		**4.35**	
Noise level	**20dB**	**40dB**	**20dB**	**40dB**	**20dB**	**40dB**	**20dB**	**40dB**	**20dB**	**40dB**	**20dB**	**40dB**
	MR Angiography (512×512)						**wrist MRI** (256×256)					
TV	23.47	33.42	22.33	30.88	20.99	28.42	19.55	28.71	18.46	25.20	17.30	22.20
I-HDTV2	24.21	34.66	22.58	31.62	20.81	28.15	20.32	29.68	18.91	25.54	17.33	21.88
A-HDTV2	**24.52**	**35.51**	**23.07**	**32.58**	**21.39**	**29.74**	**20.45**	**30.30**	**19.18**	**26.36**	**17.66**	**22.33**
Lysaker	23.01	32.28	21.93	28.75	19.82	26.40	19.86	27.80	18.52	22.82	16.60	18.54
Laplacian	21.90	31.66	20.38	26.25	18.30	24.21	19.06	25.13	17.22	19.84	15.48	16.14
	brain MRI axial view (256×256)						**Lena** (256×256)					
TV	21.71	30.32	20.57	27.75	19.64	**25.32**	19.88	27.70	18.99	25.25	17.97	22.52
I-HDTV2	21.65	30.65	20.49	27.49	19.30	24.57	20.23	29.00	18.87	**26.36**	17.83	23.24
A-HDTV2	**21.91**	**31.29**	**20.85**	**28.17**	**19.83**	25.25	**20.38**	**29.11**	**19.20**	26.03	**18.00**	**23.37**
Lysaker	21.63	29.87	20.42	26.15	19.23	22.59	20.17	27.36	18.51	23.59	16.59	19.30
Laplacian	20.70	26.91	19.81	23.29	18.11	20.28	19.55	24.75	17.49	20.72	15.92	17.20
	brain MRI sagittal view(256×256)						**Pepper** (256×256)					
TV	20.79	29.31	19.49	26.61	18.29	24.23	20.20	30.36	19.22	27.54	18.31	25.20
I-HDTV2	21.56	30.41	20.27	27.12	18.70	23.96	20.43	31.45	19.40	28.20	18.33	25.20
A-HDTV2	**21.94**	**30.67**	**20.30**	**27.55**	**18.87**	**24.58**	**20.61**	**31.65**	**19.61**	**28.96**	**18.31**	**25.87**
Lysaker	21.09	27.94	19.09	22.88	16.85	20.31	20.50	30.80	19.24	26.69	17.65	21.59
Laplacian	20.18	25.35	17.89	20.03	15.26	16.36	19.55	28.19	17.99	23.31	16.49	19.73

Table 3.1: Comparison of compressed sensing algorithms

3.5.2 HDTV in Deblurring

Deblurring or deconvolution is an important problem in many areas, including microscopy [73; 74], astronomy [75], and motion correction [76]. We illustrate the utility of the proposed second degree HDTV methods in this inverse problem and compare the results with the standard TV regularized recovery algorithm. We consider Gaussian blurring kernels with different standard deviations ($\sigma_{\text{gau}} =$

0.5 and $\sigma_{\text{gau}} = 1$) and two different noise levels (25dB and 45dB). Quantitative comparisons for two microscopy images and two natural images (Barbara and Lena) are shown in Table 3.2. Note that the anisotropic second degree HDTV provides the best reconstructions, except in one case. The comparison of the algorithms on a monkey kidney cell image is shown in Fig. 3.4. Note that the standard TV results in quite patchy results, while the proposed anisotropic second degree HDTV scheme provides more accurate reconstructions.

σ_{gau}	**0.5**	**0.5**	**1**	**1**	**0.5**	**0.5**	**1**	**1**
noise level	**25dB**	**45dB**	**25dB**	**45dB**	**25dB**	**45dB**	**25dB**	**45dB**
	monkey kidney cell				**Lena**			
TV	21.37	37.83	16.25	21.28	24.41	37.87	20.58	24.42
I-HDTV2	21.66	37.76	16.43	22.31	24.41	37.99	20.03	24.71
A-HDTV2	**21.8**	**37.84**	**16.45**	**22.46**	**24.58**	**38.05**	20.61	**24.76**
	mongoose skin cell				**Barbara**			
TV	21.79	37.65	17.26	21.57	23.56	37.81	**20.65**	23.6
I-HDTV2	22.05	37.72	17.44	22.38	23.63	37.85	20.33	23.83
A-HDTV2	**22.05**	**37.77**	**17.45**	**22.38**	**23.8**	**37.96**	20.45	**23.87**

Table 3.2: Comparison of deblurring algorithms

3.5.3 HDTV in Denoising

The removal of noise from images is a very common problem in image processing. This area has witnessed extensive research with several algorithms that offer very good performance. As discussed previously, most of the second degree TV extensions were originally introduced for denoising. We compare the denoising performance of the HDTV schemes with standard TV, Laplacian, Lysaker's anisotropic second order TV, and sparse Laplacian method in Fig. 3.5. The TV scheme exhibits patchy results, while the Laplacian method results in blob like artifacts in this high noise setting. The Lysaker scheme provides less patchy reconstructions, compared to the TV scheme. However, it is observed to result in blurry reconstructions. The A-HDTV2 scheme is capable of recovering the fine image features in the facial regions and the details of the camera, compared to the Lysaker's penalty, resulting in around 1 dB improvement over the other schemes. We also observe that the A-HDTV3 scheme preserves more details than the A-HDTV2 scheme.

The quantitative comparisons of the denoising performance of the algorithms on six test images, corrupted by Gaussian white noise are shown in Table 3.3. We study the denoising performance for images with different signal to noise ratios; the standard deviation of the noise process is controlled to obtain an input SNR of 5 dB to 30 dB. In the denoising setting, the current second order methods (Lysaker and Laplacian) provides images with better signal to noise ratio than the standard TV scheme; this observation is is consistent with the extensive denoising literature [65; 77]. We observe that the denoising performance of the proposed penalties (anisotropic and isotropic HDTV schemes) are better, or at least comparable, with the state of the art methods. Note that the proposed anisotropic second degree HDTV provides the best SNR in most cases. These experiments show that the proposed anisotropic HDTV penalties also work well in the de-noising setting.

3.5.4 Convergence rate

We compare the convergence rate of the different algorithms in Fig. 3.7. All the methods were implemented using the iterative reweighted least squares algorithm, implemented in MATLAB on a Linux workstation with two Core 2 quad-core pro-

input SNR(dB)	5	15	30	5	15	30	5	15	30
	Lena			Lung carcinoma cell			Pepper		
TV	15.72	20.47	30.01	13.03	18.50	29.92	16.34	21.66	31.37
I-HDTV2	16.46	22.34	33.67	14.34	20.85	33.51	16.46	23.18	34.41
A-HDTV2	**16.53**	**22.48**	**33.77**	**14.41**	**20.91**	**33.61**	**17.13**	**23.59**	**34.57**
Curvelet	16.20	21.43	30.80	14.28	20.44	31.15	16.99	22.98	31.93
SURE-let	16.29	21.12	31.43	14.37	20.44	31.76	16.85	22.72	32.58
Lysaker	16.17	20.71	30.13	13.80	19.29	30.93	16.59	21.86	31.78
Laplacian	15.11	19.45	29.29	13.25	18.56	29.66	16.13	20.90	31.49
	House			Medicago cell			Cameraman		
TV	18.40	23.83	32.21	14.01	18.89	29.17	**16.68**	21.56	31.47
I-HDTV2	18.85	24.11	34.29	15.41	21.29	32.83	16.46	22.67	34.40
A-HDTV2	**18.97**	24.46	**34.49**	**15.50**	**21.44**	**32.89**	16.59	**22.91**	**34.56**
Curvelet	18.47	**24.55**	33.26	15.06	20.59	30.53	16.38	21.80	32.19
SURE-let	18.72	24.19	33.11	10.81	19.59	31.30	16.60	21.92	32.47
Lysaker	18.33	23.52	32.31	14.92	19.86	29.70	16.23	21.14	31.37
Laplacian	18.15	22.39	31.38	14.53	19.88	29.25	15.51	20.15	30.69

Table 3.3: Comparison of denoising algorithms

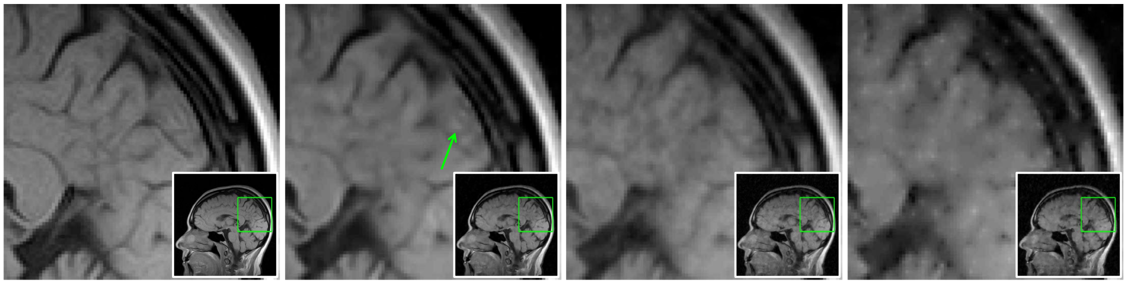
cessors. The change in SNR vs computation time is plotted in Fig. 3.7. We plot the SNR of the A-HDTV scheme and standard TV as a function of the CPU time at different acceleration factors. For acceleration factor of 2, A-HDTV2 method (red line in (a)) needs about 280 iterations for convergence (75 seconds), compared with around 150 iterations (29 seconds) using the TV algorithm. In contrast, the A-HDTV2 method requires around 500 CG steps (127 seconds), compared to approximately 300 CG steps (61 seconds) for standard TV (blue line in (b)). Thus, we see that the A-HDTV2 scheme only results in a moderate increase in computational cost over standard TV, implemented using the lagged diffusivity/IRLS algorithm. Several fast TV algorithms were introduced in the recent past [78;

79], which may provide faster convergence than the IRLS implementation.

3.6 Contributions

Yue Hu, Mathews Jacob. Improved Recovery Using Improved Total Variation Regularization. IEEE *International Symposium on Biomedical Imaging: From Nano to Macro*. Pages: 1154-1157, Chicago, USA, 2011.

Yue Hu, Mathews Jacob. Higher Degree Total Variation (HDTV) Regularization for Image Recovery. Vol 21, No 5, pp 2559-2571, May 2012.

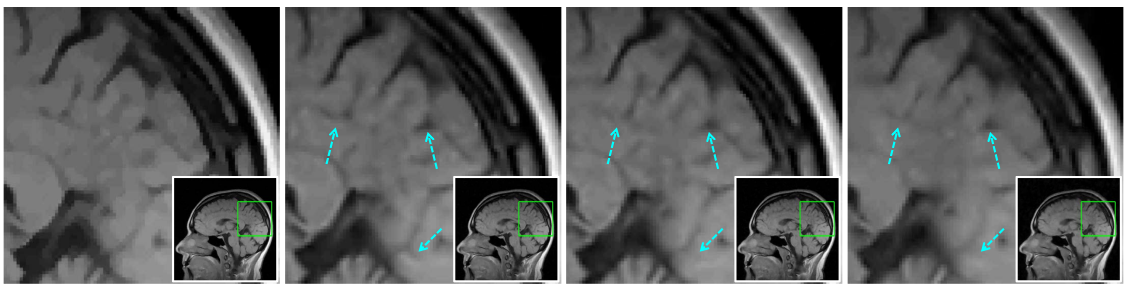


(a) actual

(b) A-HDTV3 (A=4.35): 23.17dB

(c) Curvelet(A=4.35): 17.55dB

(d) Laplacian(A=4.35): 16.36dB

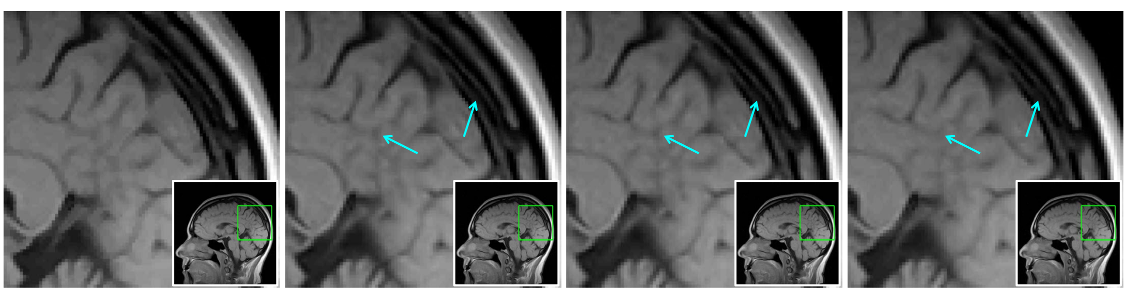


(e) TV(A=4.35): 24.23dB

(f) A-HDTV2(A=4.35): 24.58dB

(g) I-HDTV2 (A=4.35): 23.96dB

(h) Lysaker(A=4.35): 20.31dB



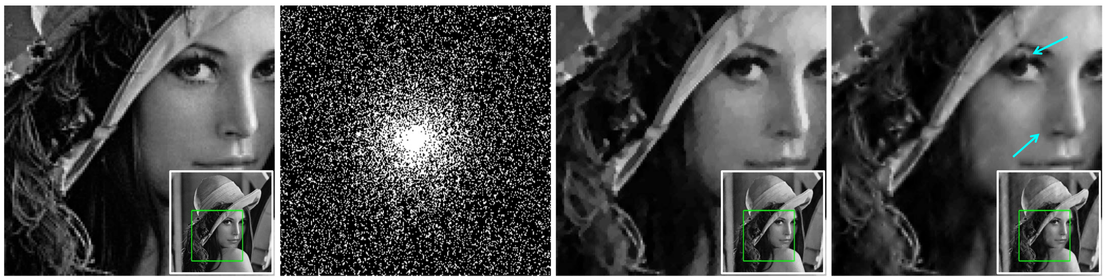
(i) TV (A=2): 29.31dB

(j) A-HDTV2 (A=2): 30.67dB

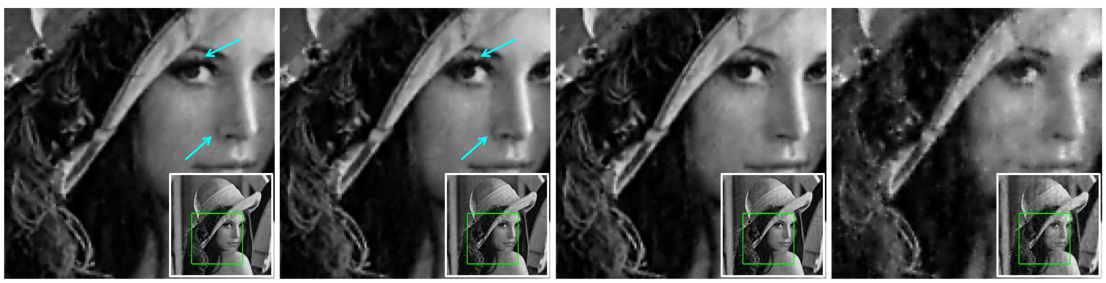
(k) I-HDTV2 (A=2): 30.41dB

(l) Lysaker (A=2): 27.94dB

Figure 3.2: Compressed sensing recovery of brain sagittal MRI from noisy and undersampled Fourier data. (a) is the original image. (b) through (h) are reconstructions at acceleration of 4.35 using A-HDTV3, curvelet, Laplacian, TV, A-HDTV2, I-HDTV2, and Lysaker's method, respectively. The A-HDTV2 (f) scheme provides the best preservation of image features, compared to the competing methods. The reconstructions at an acceleration of 2 using TV, A-HDTV2, I-HDTV2, and Lysaker methods are presented in the bottom row ((i) through (l)). Compared with Lysaker's scheme in (l), both HDTV2 schemes can preserve the details in image.

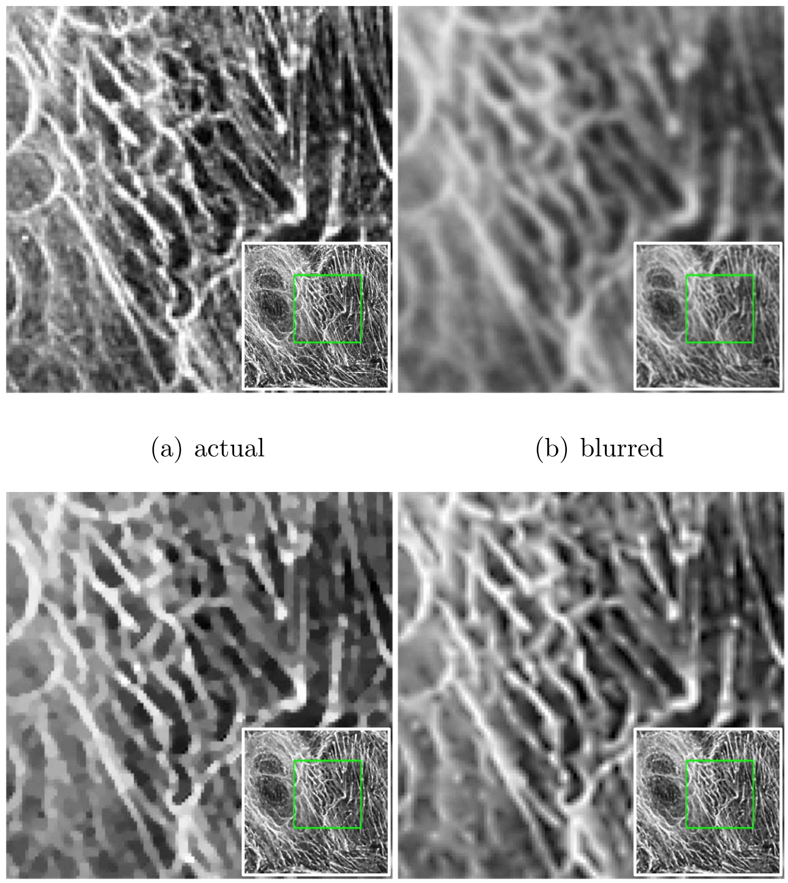


(a) actual (b) sampling pattern (c) TV : 22.52dB (d) Lysaker : 19.30dB



(e) A-HDTV2 : 23.37dB (f) I-HDTV2 : 23.24dB (g) A-HDTV3 : 22.88dB (h) Laplacian : 17.20dB

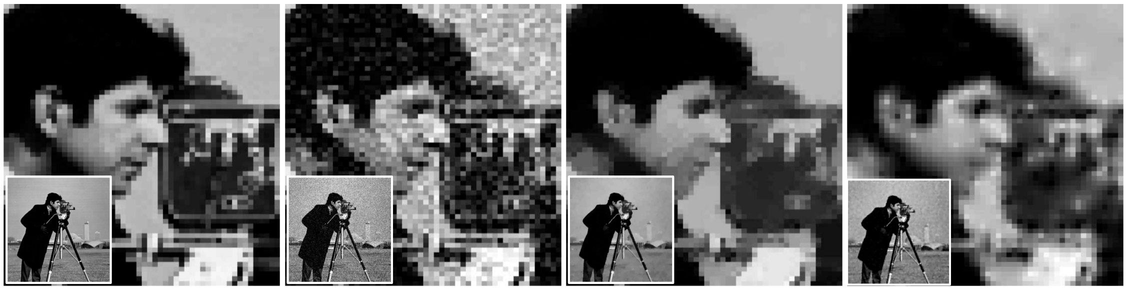
Figure 3.3: Compressed sensing recovery of Lena image: we recover the image from its noisy and under-sampled Fourier measurements. (a) is the actual image and (b) is the sampling pattern in the Fourier domain, corresponding to an acceleration of 4.35. We added white complex noise to the measurements such that the signal to noise ratio of the measurements is 40 dB. The reconstructions using the different methods are shown in (c)-(h). Note that the TV scheme results in staircase artifacts in the facial regions and results in loss of detail in the hair regions. The existing second order methods (Lysaker's anisotropic penalty and Laplacian), which were originally introduced for denoising, results in poor SNR in the CS setting. Note from the regions marked by blue arrows in (d) to (f) that the Lysaker scheme results in more blurring of image features. The HDTV schemes provide better preservation of details and smooth image regions, thus improving the SNR.



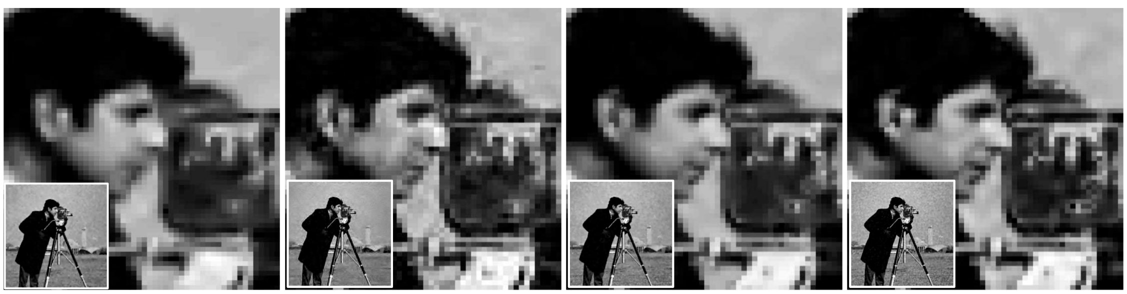
(a) actual (b) blurred

(c) TV: 16.25dB (d) A-HDTV2: 16.45dB

Figure 3.4: Deblurring of a microscopy monkey kidney cell image shown in (a). (b) is the blurred image using Gaussian blurring kernel with standard deviation of 1 and 25dB of noise. (c) shows the deblurred image using standard TV. (d) is the deblurred image using the proposed anisotropic second degree HDTV scheme. The performance improvement with the A-HDTV2 scheme is clear, compared to the standard TV scheme.

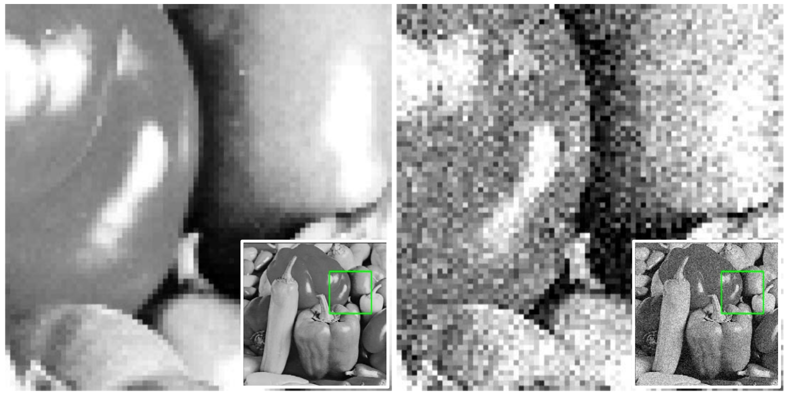


(a) original (b) noisy image: SNR=15dB (c) TV: 21.56dB (d) Laplacian: 20.15dB

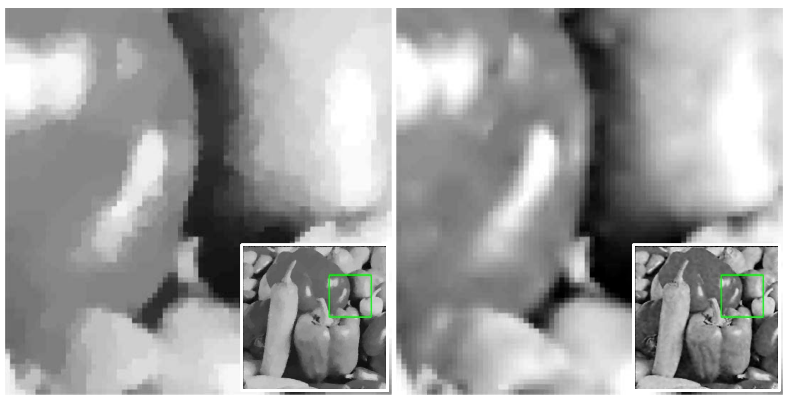


(e) Lysaker: 21.14dB (f) SURElet: 21.92dB (g) A-HDTV2: 22.9dB (h) A-HDTV3: 22.99dB

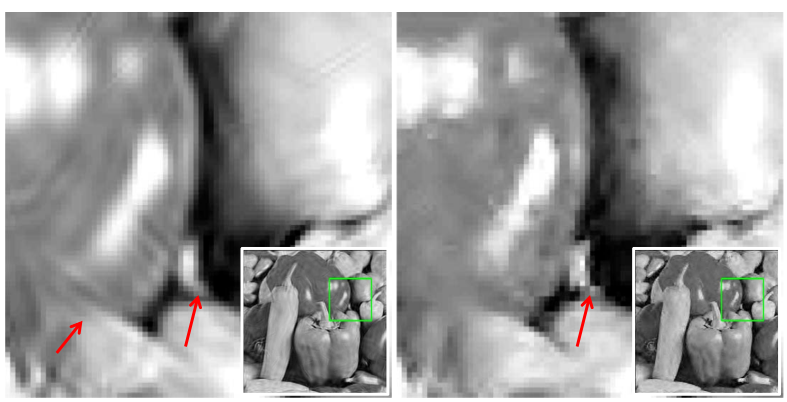
Figure 3.5: Denoising of the cameraman image: (a) is the original image (b) is the noisy image that is obtained by adding Gaussian white noise to (a). We chose the variance of the noise process such that the signal to noise ratio of the noisy image is 15 dB. (c) through (h) are the denoised images using different algorithms. We observe that the TV reconstructions are very patchy, while the Laplacian method results in blob-like artifacts. We observe that the Lysaker method results in blurry reconstructions. The SURElet reconstructions exhibit considerable ringing artifacts. In contrast, the AHDTV reconstructions are smooth and are observed to preserve the fine features. Note that the A-HDTV3 scheme provides an improvement in image quality over A-HDTV2 in this case, eventhough the improvement in SNR is only 0.09 dB.



(a) actual (b) noise level: 15dB

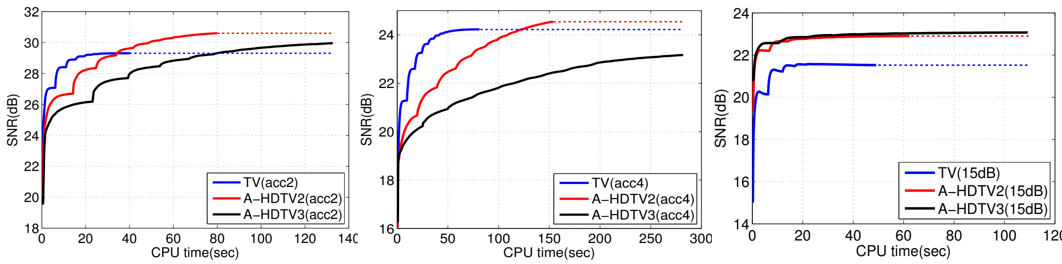


(c) TV: 21.66dB (d) A-HDTV2: 23.5dB



(e) curvelet: 22.98dB (f) SURElet: 22.72dB

Figure 3.6: Denoising of Pepper image (a). The input image with 15dB Gaussian white noise is shown in (b). (c) through (f) illustrate the denoising results of standard TV, proposed anisotropic second degree HDTV, curvelet and SURE-let methods. It is seen that the A-HDTV2 scheme minimizes the staircase and ringing artifacts that are seen with TV and x-let schemes.



(a) Compressed sensing: Acc=2 (b) Compressed sensing: Acc=4.35 (c) Denoising: input SNR=15dB

Figure 3.7: SNR vs CPU time of different algorithms in different settings. The blue, red, and black curves correspond to standard TV, 2nd degree AHDTV, and 3rd degree AHDTV respectively. We extend the actual plots (shown in solid lines) by dotted lines to facilitate easy comparison of the final SNR. The algorithms are terminated when the relative change in the cost function is too small. We observe that the second degree method takes roughly double the time taken by the standard TV scheme in most cases, while improving the SNR by 1-2 dB. In contrast, the SNR improvement offered by the 3rd degree AHDTV is not very significant, considering the increase in computational complexity.

4 Fast Majorize Minimize Three-Dimensional Higher Degree Total Variation (3D-HDTV)

The total variation (TV) image regularization penalty is widely used in many image recovery problems, such as denoising, compressed sensing, deblurring, and others [48; 49; 50]. The TV norm has also been extended to three-dimensional (3D) image recovery problems. However, TV reconstruction of 3D data, such as magnetic resonance angiography (MRA), often results in lost of the small details. In this chapter, we propose a 3D version of HDTV (3D-HDTV) to recover 3D datasets while preserving the line-like features. One of the challenges associated with the previous HDTV framework, which impeded the effective implementation of 3D-HDTV, was the high computational complexity of the algorithm. In this chapter, we introduce a new computationally efficient algorithm for HDTV regularized image recovery problems. We find that this new algorithm improves the convergence rate by factor of ten compared to the previous scheme, making the framework comparable in run time to the state-of-the-art TV methods. Using the new fast algorithm, we demonstrate the utility of 3D-HDTV regularization in the context of denoising, compressed sensing, and deblurring of 3D MRI and fluorescence microscope datasets. We show that 3D-HDTV routinely outperforms 3D-TV in terms of the SNR of reconstructed images and in its ability to preserve ridge-like details in 3D datasets.

4.1 Introduction

4.1.1 Generalized 2-D HDTV regularization penalty

The standard TV regularization penalty is the L_1 norm of the image gradient, specified as $\mathrm{TV}(f) = \int_\Omega |\nabla f| d\mathbf{r}$. We define the generalized HDTV regularization in 2D as

$$\mathrm{HDTV}_n(f) = \int_\Omega \left(\frac{1}{2\pi} \int_0^{2\pi} |\mathcal{D}_{\theta,n}\, f(\mathbf{r})| d\theta \right) d\mathbf{r}, \tag{4.1}$$

where $\mathcal{D}_{\theta,n}$ is the rotated version of an n^{th} degree derivative operator along the unit vector $\mathbf{u}_\theta = (\cos\theta, \sin\theta)$. If $\mathcal{D}_{\theta,n} f = f_{\theta,n}$, the n^{th} degree directional derivative, then the generalized HDTV penalty simplifies to the HDTV regularization penalty introduced in [80]. Thus, (4.1) is essentially a generalization of the functional in [80].

4.1.2 3D-HDTV Penalties

The HDTV penalty may easily be extended to 3D images by penalizing the L_1-L_1 of all directional derivatives in $\mathbb{R}^3$. Specifically, for f a continuously differentiable complex-valued image defined on $\Omega \subset \mathbb{R}^3$, we define the nth degree HDTV penalty in 3D to be

$$\mathrm{HDTV}_n(f) = \int_\Omega \int_{\mathbb{S}^2} |f_{\mathbf{u},n}(\mathbf{r})| \, d\mathbf{u} \, d\mathbf{r}, \tag{4.2}$$

where $\mathbb{S}^2 = \{\mathbf{u} \in \mathbb{R}^3 : |\mathbf{u}| = 1\}$ and $f_{\mathbf{u},n}(\mathbf{r})$ is the nth degree directional derivative defined as

$$f_{\mathbf{u},n}(\mathbf{r}) = \left. \frac{\partial^n}{\partial \gamma^n} f(\mathbf{r} + \gamma \mathbf{u}) \right|_{\gamma=0}; \ \mathbf{u} \in \mathbb{S}^2. \tag{4.3}$$

We may unify both the 2D and 3D HDTV penalties into a general d-dimensional penalty

$$\mathrm{HDTV}_n(f) = \int_\Omega \int_{\mathbb{S}^{d-1}} |f_{\mathbf{u},n}(\mathbf{r})| \, d\mathbf{u} \, d\mathbf{r}, \tag{4.4}$$

where the directional derivative $f_{\mathbf{u},n}$ is defined similarly as in (4.3) but with $\mathbf{u} \in \mathbb{S}^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : |\mathbf{u}| = 1\}$. The new algorithm we derive below for solving HDTV regularized inverse problems applies to this general case.

4.1.3 Steerability of Directional Derivatives

Note that the first degree directional derivatives $f_{\mathbf{u},1}$ have the equivalent expression

$$f_{\mathbf{u},1}(\mathbf{r}) = \mathbf{u}^T \nabla f(\mathbf{r}). \tag{4.5}$$

By recursively applying (4.5) we may also express higher degree directional derivatives $f_{\mathbf{u},n}(\mathbf{r})$ as a separable vector product:

$$f_{\mathbf{u},n}(\mathbf{r}) = \mathbf{s}_n^T(\mathbf{u}) \mathcal{D}_n f(\mathbf{r}), \tag{4.6}$$

where, $\mathbf{s}_n(\mathbf{u})$ is vector of polynomials in the components of $\mathbf{u}$ and $\mathcal{D}_n f(\mathbf{r})$ is the vector of all nth degree partial derivatives of f. For example, in 2nd degree case ($n = 2$) in 3D, we may choose

$$\mathbf{s}_2(\mathbf{u}) = \left[u_x^2, u_y^2, u_z^2, 2u_x u_y, 2u_y u_z, 2u_x u_z \right]^T; \tag{4.7}$$

$$\mathcal{D}_2 f(\mathbf{r}) = \Big[f_{xx}(\mathbf{r}), f_{yy}(\mathbf{r}), f_{zz}(\mathbf{r}), f_{xy}(\mathbf{r}), f_{yz}(\mathbf{r}), f_{xz}(\mathbf{r})\Big]^T. \tag{4.8}$$

Equation (4.6) demonstrates that nth order directional derivatives are *steerable* operators, in the sense that they can written as a finite linear combination of a few fixed operators with coefficients dependent only on the orientations $\mathbf{u} \in \mathbb{S}^{d-1}$. This useful property affords us to make many simplifications in the derivation and numerical implementation of our new algorithm.

4.2 Fast MM Algorithm for HDTV Regularized Inverse Problems

The recovery of a d-dimensional image $f : \Omega \to \mathbb{C}$, $\Omega \subset \mathbb{R}^d$, from its degraded measurements $\mathbf{b} = \mathcal{A}(f) + n$ using HDTV regularization requires us to minimize the following cost function:

$$\mathcal{C}(f) = \|\mathcal{A}(f) - \mathbf{b}\|^2 + \lambda \int_\Omega \int_{\mathbb{S}^{d-1}} |f_{\mathbf{u},n}(\mathbf{r})| \, d\mathbf{u} \, d\mathbf{r}. \tag{4.9}$$

Since the absolute function $|\cdot|$ is not continuously differentiable, we approximate it by the Huber function:

$$\varphi_\beta(x) = \begin{cases} |x| - 1/2\beta & \text{if} \quad |x| \geq \frac{1}{\beta} \\ \beta\,|x|^2/2 & \text{else}\ . \end{cases} \tag{4.10}$$

The approximate cost function is thus specified by

$$\mathcal{C}_\beta(f) = \|\mathcal{A}(f) - \mathbf{b}\|^2 + \lambda \int_\Omega \int_{\mathbb{S}^{d-1}} \varphi_\beta\left(|f_{\mathbf{u},n}(\mathbf{r})|\right) \, d\mathbf{u} \, d\mathbf{r}. \tag{4.11}$$

Note that this approximation tends to the original HDTV penalty when $\beta \to \infty$.

To realize computationally efficient solutions, we majorize the Huber function in the above expression by the quadratic function [81]:

$$\varphi_\beta\left(|f_{\mathbf{u},n}(\mathbf{r})|\right) = \min_{g(\mathbf{u},\mathbf{r})} \left\{ \frac{\beta}{2} |f_{\mathbf{u},n}(\mathbf{r}) - g(\mathbf{u},\mathbf{r})|^2 + \psi\left(|g(\mathbf{u},\mathbf{r})|\right) \right\} \tag{4.12}$$

where $g : \mathbb{S}^{d-1} \times \mathbb{R}^d \to \mathbb{C}$ is an auxiliary function. Using convex conjugates, it can be shown the quadratic function majorizes the original penalty when $\psi\left(|g(\mathbf{u}, \mathbf{r})|\right) = |g(\mathbf{u}, \mathbf{r})|$; see proof in [80]. With this majorization, the cost function (4.12) can now be expressed as

$$\mathcal{C}_\beta(f) = \min_g \|\mathcal{A}(f) - \mathbf{b}\|^2 + \lambda \int_\Omega \int_{\mathbb{S}^{d-1}} \left\{ \frac{\beta}{2} |f_{\mathbf{u},n}(\mathbf{r}) - g(\mathbf{u}, \mathbf{r})|^2 + |g(\mathbf{u}, \mathbf{r})| \right\} d\mathbf{u}\, d\mathbf{r} \tag{4.13}$$

Note that the optimization algorithm now involves the minimization of the right hand side of the above expression with respect to both functions $f : \Omega \to \mathbb{C}$ and $g : \mathbb{S}^{d-1} \times \Omega \to \mathbb{C}$. We rely on an alternating minimization algorithm to solve for the two functions. Specifically, we alternate between the minimization with respect to f and g as shown below.

4.2.1 Step one: Minimization with respect to g, assuming f fixed

Assuming f to be fixed, we minimize the cost function in (4.13) with respect to g. Since the integrand in (4.13) is positive, the minimization at each $(\mathbf{u}, \mathbf{r})$ can be decoupled as

$$\min_{g(\mathbf{u},\mathbf{r})} \frac{\beta}{2} |f_{\mathbf{u},n}(\mathbf{r}) - g(\mathbf{u}, \mathbf{r})|^2 + |g(\mathbf{u}, \mathbf{r})|, \tag{4.14}$$

whose solution is given by the shrinkage formula,

$$g(\mathbf{u}, \mathbf{r}) = \max\left(|f_{\mathbf{u},n}(\mathbf{r})| - \frac{1}{\beta}, 0 \right) \frac{f_{\mathbf{u},n}(\mathbf{r})}{|f_{\mathbf{u},n}(\mathbf{r})|}; \quad \forall \mathbf{r} \in \mathbb{R}^d; \mathbf{u} \in \mathbb{S}^{d-1}. \tag{4.15}$$

The above equation implies that $g(\mathbf{u}, \mathbf{r})$ has to be evaluated for all unit directions $\mathbf{u}$. While the computation of this term is not computationally expensive, the need to store this term will make the algorithm very memory demanding. However, we will see in the next subsection that the rest of the algorithm does not need $g(\mathbf{u}, \mathbf{r})$ for all $\mathbf{u} \in \mathbb{S}^{d-1}$, but only its projection to the space spanned by $\mathbf{s}_n(\mathbf{u})$, specified by $\mathbf{q}_n(\mathbf{r}) = \int_{\mathbb{S}^{d-1}} \mathbf{s}(\mathbf{u}) g(\mathbf{u}, \mathbf{r}) d\mathbf{u}$. This simplification results in an algorithm with considerably less memory demand.

4.2.2 Step two: Minimization with respect to f, assuming g to be fixed

Assuming that g is fixed, we now minimize (4.13) with respect to f. This can be reformulated as

$$\min_f \mathcal{C}_g(f), \tag{4.16}$$

where

$$\mathcal{C}_g(f) = \|\mathcal{A}(f) - \mathbf{b}\|^2 + \lambda \int_\Omega \int_{\mathbb{S}^{d-1}} \left\{ \frac{\beta}{2} \left(\|f_{\mathbf{u},n}(\mathbf{r})\|^2 - 2 \langle f_{\mathbf{u},n}(\mathbf{r}), g(\mathbf{u}, \mathbf{r}) \rangle + \|g(\mathbf{u}, \mathbf{r})\|^2 \right) \right\} d\mathbf{u} d\mathbf{r}. \tag{4.17}$$

Ignoring the constant term $\int_\Omega \int_{\mathbb{S}^{d-1}} \|g(\mathbf{u}, \mathbf{r})\|^2 \, d\mathbf{u} \, d\mathbf{r}$ in the above expression, we obtain:

$$\begin{aligned}
\mathcal{C}_g(f) &= \|\mathcal{A}(f) - \mathbf{b}\|^2 + \frac{\lambda\beta}{2} \int_\Omega \int_{\mathbb{S}^{d-1}} \left\{ \|f_{\mathbf{u},n}(\mathbf{r})\|^2 - 2 \langle f_{\mathbf{u},n}(\mathbf{r}), g(\mathbf{u}, \mathbf{r}) \rangle \right\} d\mathbf{u} d\mathbf{r} \\
&= \|\mathcal{A}(f) - \mathbf{b}\|^2 + \\
&\frac{\lambda\beta}{2} \int_\Omega \left\{ \mathcal{D}_n f(\mathbf{r})^H \underbrace{\left(\int_{\mathbb{S}^{d-1}} \mathbf{s}_n(\mathbf{u}) \mathbf{s}_n^T(\mathbf{u}) d\mathbf{u} \right)}_{\mathbf{C}_n} \mathcal{D}_n f(\mathbf{r}) - 2 \left\langle \mathcal{D}_n f(\mathbf{r}), \underbrace{\int_{\mathbb{S}^{d-1}} \mathbf{s}_n(\mathbf{u}) g(\mathbf{u}, \mathbf{r}) d\mathbf{u}}_{\mathbf{q}_n(\mathbf{r})} \right\rangle \right\} d\mathbf{r}
\end{aligned} \tag{4.18}$$

In the last step, we used the steerability relationship of the directional derivatives from (4.6) to simplify the expression. Note that the criterion $\mathcal{C}_g(f)$ does not depend on $g(\mathbf{u},\mathbf{r})$, but only its projections to the space spanned by $\mathbf{s}_n(\mathbf{u})$, specified by $\mathbf{q}_n(\mathbf{r})$. Since we do not have to store the variable $g(\mathbf{u},\mathbf{r})$ for all $\mathbf{u} \in \mathbb{S}^{d-1}$, this simplification considerably reduces the memory demand of the algorithm. In addition, the above expression is independent of the directional derivatives $f_{\mathbf{u},n}(\mathbf{r})$; it is only dependent on the partial derivatives of f, thanks to the steerability of the directional derivatives in terms of the partial derivatives. From the Euler-Lagrange equation for $\mathcal{C}_g(f)$ we obtain:

$$\left(2\mathcal{A}^T\mathcal{A} + \lambda\beta\ \mathcal{D}_n^T\mathbf{C}_n\mathcal{D}_n\right) f = 2\mathcal{A}^T\mathbf{b} + \lambda\beta\ \mathcal{D}_n^T\mathbf{q}_n \tag{4.19}$$

The operator $\mathcal{D}_n^T\mathbf{C}_n\mathcal{D}_n$ has a simple expression in the discrete Fourier domain. The following are the discrete Fourier domain expressions in 1st and 2nd degree case:

1st degree operator

$$\mathcal{F}[\mathcal{D}_1^T\mathbf{C}_1\mathcal{D}_1] = \omega_x^2 + \omega_y^2 \tag{4.20}$$

2nd degree operator

$$\mathcal{F}[\mathcal{D}_2^T\mathbf{C}_2\mathcal{D}_2] = \frac{1}{8}\left\{3\omega_{xx}^2 + 3\omega_{yy}^2 + 4\omega_{xy} + 2\omega_{xx}\omega_{yy}\right\} \tag{4.21}$$

Here, $\mathcal{F}$ denotes the discrete Fourier transform operator. These equations will get modified by the discrete approximation of the derivatives (e.g. finite differences).

- Fourier sampling: If the measurements are Fourier samples on a Cartesian grid (i.e, $\mathcal{A} = \mathbf{S}\mathcal{F}$), (4.19) can be simplified by evaluating the discrete Fourier transform of both sides. Here, $\mathbf{S}$ is the sampling operator that picks the appropriate Fourier samples. Computing the discrete Fourier transform of both sides of (4.19), we get

$$\left(2\mathbf{S}^T\mathbf{S} + \lambda\beta\ \mathcal{F}[\mathcal{D}_n^T\mathbf{C}\mathcal{D}_n]\right) f = 2\mathbf{S}^T\mathbf{b} + \lambda\beta\mathcal{F}[\mathcal{D}_n^T\mathbf{q}_n]. \tag{4.22}$$

Thus, we obtain the analytical expression for f as:

$$f = \mathcal{F}^{-1}\left\{\frac{2\ \mathbf{S}^T\mathbf{S}\mathbf{b} + \lambda\beta\ \mathcal{F}[\mathcal{D}_n^T\mathbf{q}_n]}{2\ \mathbf{S} + \lambda\beta\mathcal{F}[\mathcal{D}_n^T\mathbf{C}\mathcal{D}_n]}\right\}. \tag{4.23}$$

When the Fourier samples are not on the Cartesian grid (for example, in parallel imaging), where the one step solution is not applicable, we could still solve the minimization problem using a preconditioned conjugate gradient algorithm.

- Deconvolution: Convolution can be considered as a multiplication in the Fourier domain. Taking the Fourier transform on both sides, (4.19) can be solved as:

$$f = \mathcal{F}^{-1}\left\{\frac{2\hat{H}^*\mathbf{b} + \lambda\beta\ \mathcal{F}[\mathcal{D}_n^T\mathbf{q}_n]}{2|\hat{H}|^2 + \lambda\beta\mathcal{F}[\mathcal{D}_n^T\mathbf{C}\mathcal{D}_n]}\right\}. \tag{4.24}$$

Here H is the transfer function of the convolution filter.

4.3 Implementation Details

4.3.1 Discretization of the derivative operators

The TV and HDTV penalties are essentially defined in a continuous domain. In practice, a standard scheme is to approximate the derivatives with finite difference operators. For example, the derivative of a 2D signal along the x dimension is approximated as $q[k,l] = f[k+1,l] - f[k,l] = \Delta_1 * f$. This approximation can be viewed as the convolution of f by $\Delta_1[k] = \varphi(k+\frac{1}{2})$, where $\varphi(x) = \partial\beta_1(x)/\partial x$ and

$\beta_1(x)$ is the first degree cardinal B-spline. [53]. However, this approximation does not possess rotation steerability, i.e, the directional derivative can not be expressed as the linear combination of the finite differences along x and y directions.

To obtain discrete operators that are approximately rotation steerable, in the 2D case we approximate the nth order partial derivatives, $\partial^{n_1,n_2} := \partial_x^{n_1}\partial_y^{n_2}$ for all $n_1+n_2=n$, as the convolution of the signal with the tensor product of derivatives of one-dimensional B-spline functions:

$$\partial^{n_1,n_2} f[k_1,k_2] = \left[\beta_n^{(n_1)}(k_1+\delta) \otimes \beta_n^{(n_2)}(k_2+\delta)\right] * f[k_1,k_2], \quad \forall\ k_1,k_2 \in \mathbb{N} \tag{4.25}$$

where $\beta_n^{(m)}(x)$ denotes the mth order derivative of a nth degree B-spline. In order to obtain filters with small spacial support, we choose the shift δ according to the rule

$$\delta = \begin{cases} \frac{1}{2} & \text{if} \quad n \text{ is odd} \\ 0 & \text{else} \end{cases} \tag{4.26}$$

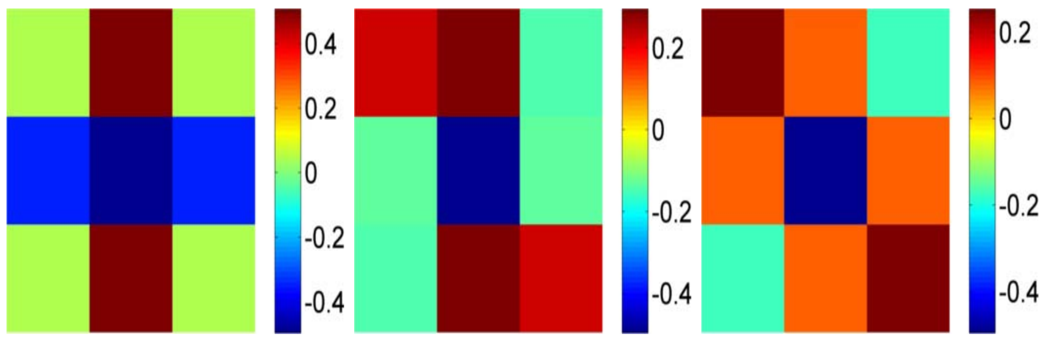
The shift δ implies that we are evaluating the image derivatives at the intersection of the voxels and not at the voxel midpoints. This scheme will result in filters that are spatially supported in a $(n+1)\times(n+1)$ pixel window.

Likewise, in the 3D case we approximate the nth order partial derivatives, $\partial^{n_1,n_2,n_3} := \partial_x^{n_1}\partial_y^{n_2}\partial_z^{n_3}$ for all $n_1+n_2+n_3=n$, as

$$\partial^{n_1,n_2,n_3} f[k_1,k_2,k_3] = \left[\beta_n^{(n_1)}(k_1+\delta) \otimes \beta_n^{(n_2)}(k_2+\delta) \otimes \beta_n^{(n_3)}(k_3+\delta)\right] * f[k_1,k_2,k_3], \tag{4.27}$$

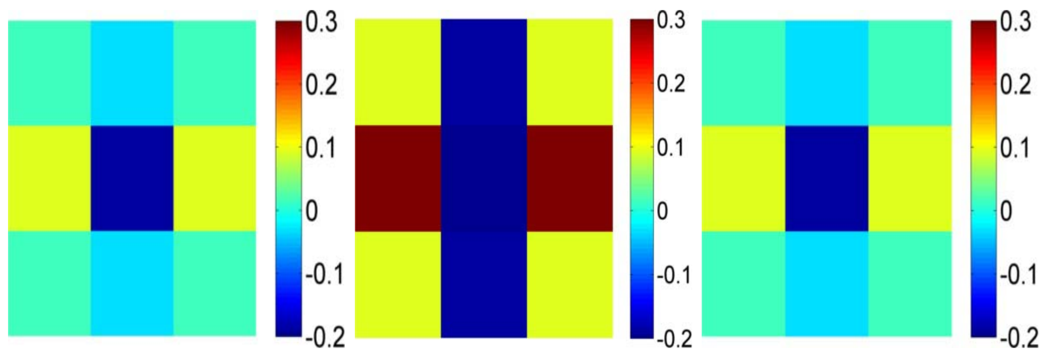
for all $k_1,k_2,k_3 \in \mathbb{N}$ with the same rule for choosing δ.

While the tensor product of B-spline functions are not strictly rotation steerable, B-splines approximate Gaussian functions as their degree increases, and the tensor product of Gaussians is exactly steerable. Thus, the approximation of derivatives we define above is approximately rotation steerable; see Fig. 4.1. We also observe in practice that using B-spline approximations to derivatives results in better image reconstructions than simple finite difference approximations.

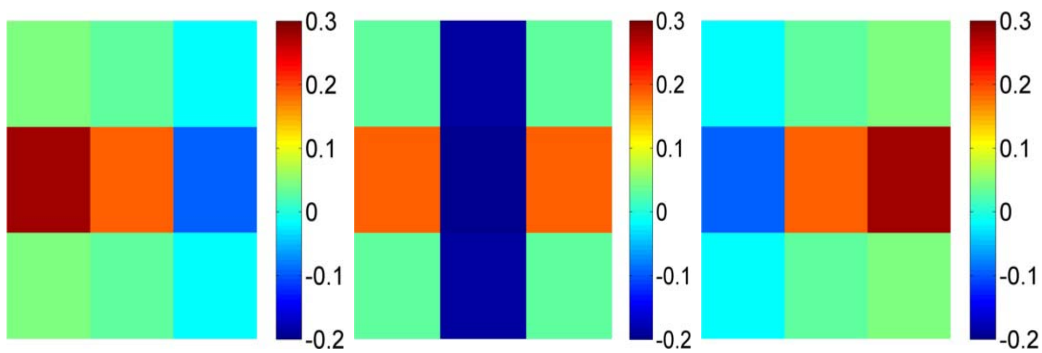


(a) 2D: 0° (b) 2D: 30° (c) 2D: 45°

Figure 4.1: 2D operators

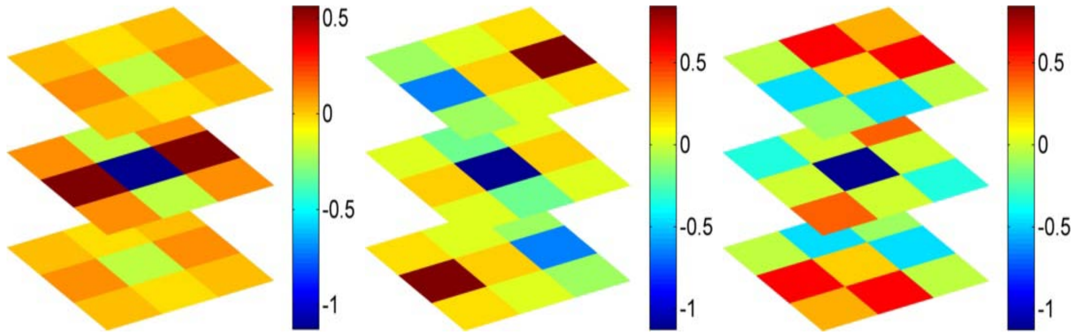


(a) 3D:0°(z=1) (b) 3D:0°(z=2) (c) 3D:0°(z=3)



(d) 3D:45°(z=1) (e) 3D:45°(z=2) (f) 3D:45°(z=3)

Figure 4.2: 3D operators for different slices along z directions



(a) 3D:$\theta = 0°$ (b) 3D:$\theta = 45°$ (c) 3D:$\theta = \phi = 45°$

Figure 4.3: 3D operators

4.3.2 Numerical Integration over $\mathbb{S}^2$

Our algorithm requires us to compute the quantities

$$\mathbf{C}_n = \int_{\mathbb{S}^{d-1}} \mathbf{s}_n(\mathbf{u})\mathbf{s}_n^T(\mathbf{u})d\mathbf{u} \quad \text{and} \quad \mathbf{q}_n(\mathbf{r}) = \int_{\mathbb{S}^{d-1}} \mathbf{s}_n(\mathbf{u})g(\mathbf{u},\mathbf{r})d\mathbf{u}, \tag{4.28}$$

which involve an integration over the unit sphere S^{d-1} in d-dimensions. We approximate these integrals with Riemann sums by uniformly sampling points in S^{d-1}. In the 2D case, this can easily be achieved by parameterizing $\mathbf{u}$ as $\mathbf{u}(\theta) = (\cos(\theta), \sin(\theta))$, then discretizing the parameter θ as $\theta_k = k\frac{2\pi}{K}$, for $k = 1, ..., K$, where K is the specified number of sample points.

However, in the 3D case, if we discretize the the usual parameterization of $\mathbb{S}^2$ given by

$$\mathbf{u}(\theta,\phi) = (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)), \quad \text{for } \phi \in [0, 2\pi], \theta \in [0, \pi] \tag{4.29}$$

by uniformly discretizing θ and ϕ, the samples we obtain are heavily biased towards the poles of the sphere, providing a poor approximation of the integral. Instead, we make use of the ISOI software package [82], [83] based on the HEALPix spherical sampling method [84] to deterministically generate uniformly spaced samples of

$\mathbb{S}^2$. We find in practice that $K \approx 35$ samples are sufficient to approximate the integrals in (4.28). Note that these sample points may be computed in advance and stored in memory to reduce the computational overhead.

4.3.3 Algorithm Overview

Algorithm 4.3.1: 3D-HDTV$(\mathcal{A}, b, \lambda)$

$i \leftarrow 1$

$\beta \leftarrow \beta_{\text{init}}, f^{(1)} \leftarrow \mathcal{A}^*(b)$

while $i <$ MaxOuterIterations

do {
- $ii \leftarrow 1$
- **while** $ii <$ MaxInnerIterations
- **do** {
  - Compute partial derivatives using (4.27)
  - Compute directional derivatives $f_{\mathbf{u},n}(\mathbf{r})$ using (4.6)
  - Update $g(\mathbf{u}, \mathbf{r})$ based on $f_{\mathbf{u},n}(\mathbf{r})$ using (4.15)
  - Determine $f^{(i+1)}(\mathbf{r})$ based on $g(\mathbf{u}, \mathbf{r})$ using (4.23) or (4.24)
  - $ii \leftarrow ii + 1$
- $\beta \leftarrow \beta * \beta_{\text{incfactor}}$
- $i \leftarrow i + 1$

return (f)

The pseudocode for the 3D HDTV fast MM algorithms is shown in 4.3.1. We typically use 10 outer iterations (MaxOuterIterations = 10) and a maximum 10 MM iterations (MaxInnerIterations = 10) per outer iterations. The MM algorithm is terminated when the relative change in the cost function is less than a specified threshold.

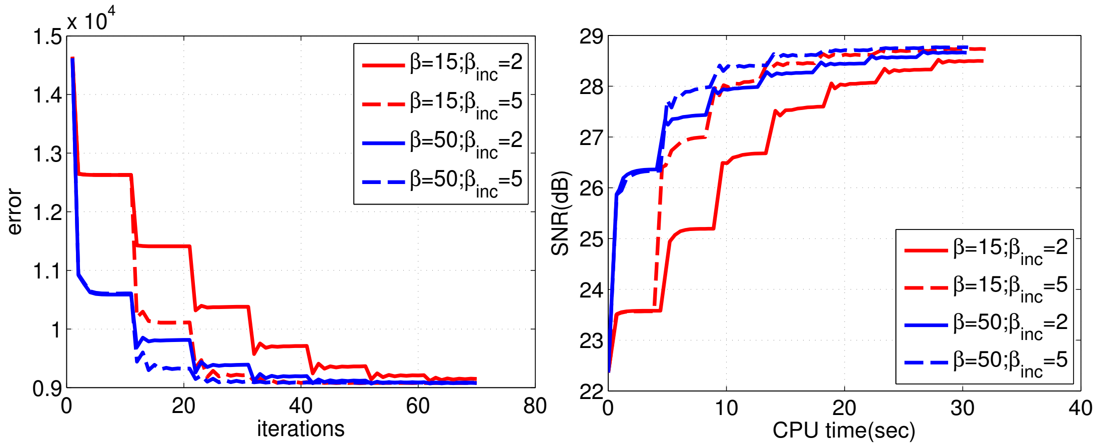
4.4 Results on Image Recovery Problems

4.4.1 Convergence of the fast HDTV algorithm

We study the effect of the parameter β and the increment rate of β, i.e. β_{inc}, on the convergence and the accuracy of the algorithm. We consider the reconstruction of a MRI brain image with the acceleration factor of 1.65 using the fast HDTV algorithm. We plot the cost as a function of the number of iterations and the SNR as a function of the CPU time in Fig. 4.4. It is observed that with different combinations of starting values of β and increment rate β_{inc}, the convergence rates of the algorithms are approximately the same and the SNRs of the reconstructed image are around the same value. However, when we choose the parameters as $\beta = 15$ and $\beta_{\text{inc}} = 2$, which are the smallest among the parameters chosen in the experiments, the SNR of the recovered image is comparatively lower than the others. This implies that in order to enforce full convergence the final value of β needs to be sufficiently large.

4.4.2 Comparison of the fast HDTV algorithm with iteratively reweighted HDTV algorithm

In this experiment, we compare the proposed fast HDTV algorithm with the IRMM algorithm in the context of the recovery of a brain MR image with acceleration factor of 4 in Fig. 4.5. Here we plot the SNR as a function of the CPU time using TV and 2nd degree HDTV with the IRMM algorithm and the proposed algorithm, respectively. We observe that the proposed algorithm (blue curve) takes around 20 seconds to converge compared to 120 seconds by IRMM algorithm (blue dotted curve) using TV penalty, and 30 seconds (red curve) com-



(a) cost function to iterations

(b) SNR to CPU time

Figure 4.4: Performance of the continuation scheme. We plot the cost as a function of the number of iterations in (a) and SNR as a function of CPU time in (b). We study four different combinations of the parameters β and β_{inc}. It is shown in (a) that the convergence rates of different combinations are almost the same. We also observe in (b) that the SNRs of the reconstructed images in four settings are similar except that when the final value of β is not large enough ($\beta = 15$, $\beta_{\text{inc}} = 2$) the SNR is comparatively lower than the others.

pared to 300 seconds (red dotted curve) using 2nd degree HDTV regularization. Thus, we see that the proposed algorithm accelerates the problem significantly (10-fold) compared to IRMM method.

4.4.3 Three-Dimensional HDTV using fast MM algorithm

We determine the utility of the 3D HDTV schemes in the context of compressed sensing, deconvolution and denoising. In each case we optimize the regularization parameters to obtain the optimized SNR to ensure fair comparisons between different schemes. The signal to noise ratio (SNR) of the reconstruction is computed as:

$$\mathrm{SNR} = -10\log_{10}\left(\frac{\|\mathrm{f}_{\text{orig}} - \hat{\mathrm{f}}\|_F^2}{\|\mathrm{f}_{\text{orig}}\|_f^2}\right), \tag{4.30}$$

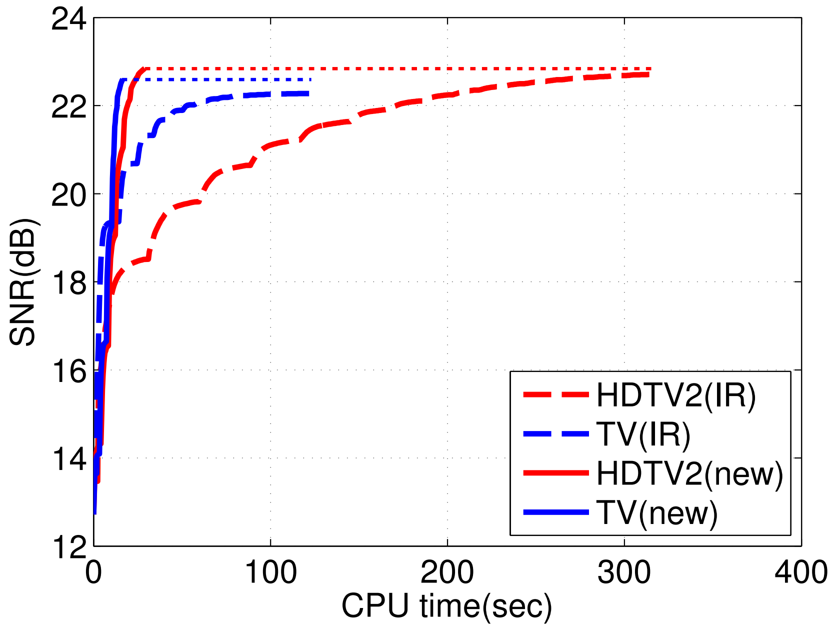


Figure 4.5: IRMM algorithm versus proposed fast HDTV algorithm in different settings. The blue, blue dotted, red, red dotted curves correspond to TV using proposed algorithm, TV using IRMM, HDTV2 using proposed algorithm, HDTV2 using IRMM algorithm, respectively. We extend (solid lines) the original plot by dotted lines for easier comparisons of the final SNRs. We see that the proposed algorithm takes 1/6 of the time taken by IRMM for standard TV, and 1/10 of the time taken by IRMM for HDTV2.

where $\hat{f}$ is the reconstructed image, f_{orig} is the original image, and $\|\cdot\|_F$ is the Frobenius norm.

We compare the 3D-HDTV method with 3D-TV. In the case of 3D-TV we approximate the partial derivatives as finite differences, and as B-splines operators in the case of 3D-HDTV, as discussed above. Fast MM algorithms were used to implement all of the above methods.

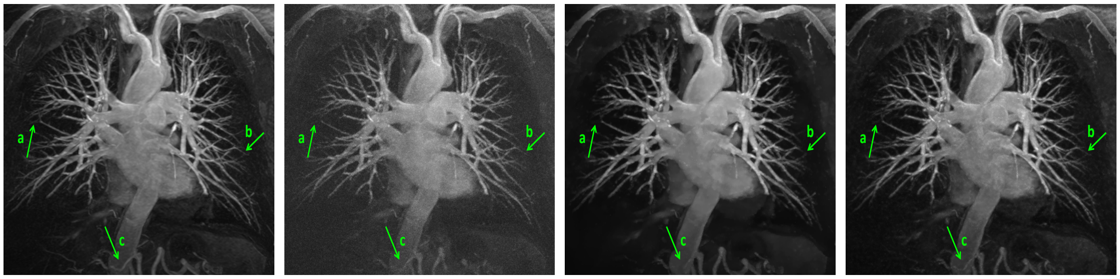
Compressed Sensing

We consider two 3D MR datasets to demonstrate the utility of the algorithm. In the experiments in this chapter, we assume the measurements to be acquired using variable density random Fourier encoding; the sampling pattern is realized in 3D MRI using random phase encodes and choosing the readout axis to be orthogonal to the image plane. The reconstructions in maximum intensity projection (MIP) of a 3D MR angiography of cardiopulmonary vasculature ($512\times512\times76$) are shown in Fig. 4.6 [85]. The acceleration is 1.5 and 5dB of Gaussian noise with standard deviation of $\sigma = 0.53$ is added. We observe that there is a 0.4dB improvement in 3D-HDTV over standard 3D-TV. We also see that 3D-HDTV preserves more line details compared with standard 3D-TV. We have zoomed the three marked regions in Fig. 4.7. It is observed that 3D-HDTV provides more accurate and natural-looking reconstructed image, while 3D-TV has some patchy artifacts that blur some of the details in the image.

The reconstructions of a cardiac MR dataset are shown in Fig. 4.8. The acceleration is 2 and 15dB of Gaussian noise with standard deviation 0.22 is added. Compared with 3D-TV, the reconstructions using 3D-HDTV scheme improves the SNR by 0.4dB. In addition, the 3D-HDTV method provides more accurate reconstruction, which overcomes the blurring in the regions pointed by the arrows.

Deconvolution

Deconvolution is an important problem in image processing. We compare the deconvolution performance of the 3D-HDTV with 3D-TV. Fig. 4.9 shows the decovolution results of a 3D fluorescence microscope dataset ($1024 \times 1024 \times 17$). The original image is blurred with a Gaussian filter with standard deviation of



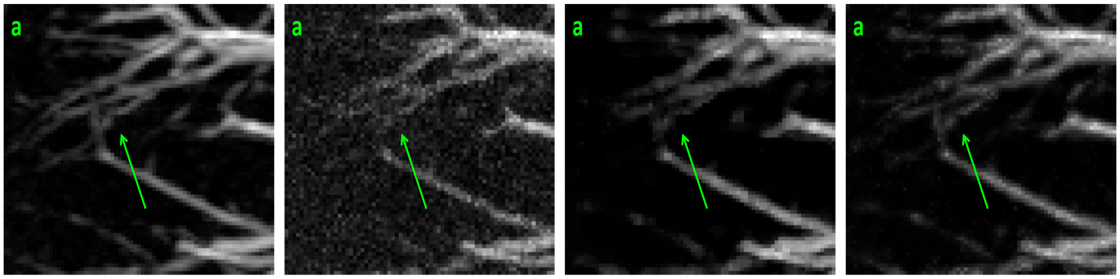
(a) actual (b) inverse FFT 3.47dB (c) 3DTV 13.69dB (d) 3DHDTV 14.05dB

Figure 4.6: Compressed sensing recovery of MR angiography data from noisy and undersampled Fourier data (acceleration of 1.5 with 5dB additive Gaussian noise). (a) through (d) are the maximum density projection image of the dataset. (a) is the original image. (b) is the direct inverse Fourier reconstruction. (c) is the reconstruction using 3D-TV method; (d) is the 3D-HDTV reconstruction image. We observe that 3D-HDTV method preserves more details that are lost in 3D-TV reconstruction. The arrows point out the three regions that are zoomed in Fig. 4.7.

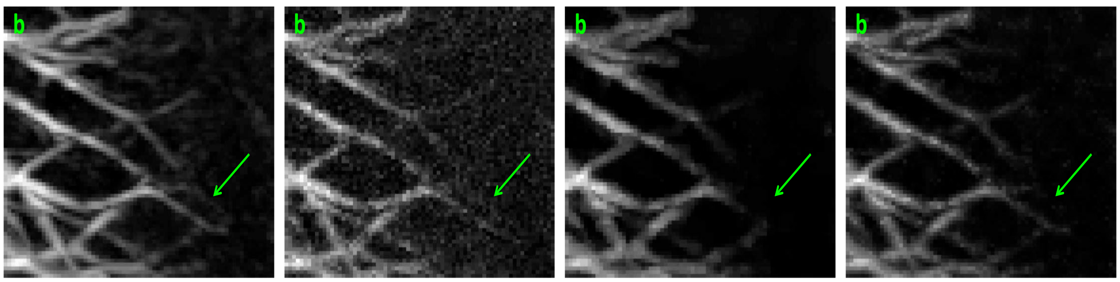
1 ($5 \times 5 \times 5$), with additive Gaussian noise of standard deviation of 0.01 added. The results show that 3D-HDTV scheme is capable of recovering the fine image features of the cell image, resulting in a 0.3dB improvement in SNR over 3D-TV.

Denoising

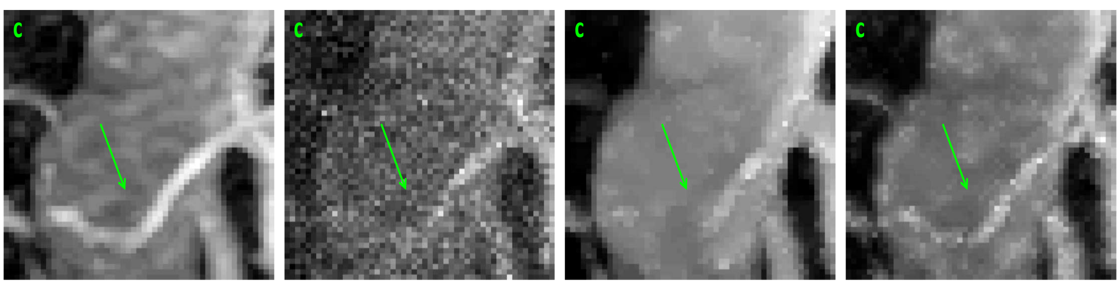
The removal of noise from microscope images is a common problem in image processing. We compare the 3D-HDTV and 3D-TV using the fast MM algorithm in the context of image denoising. In Fig. 4.10, we are showing the denoising performance of the algorithms using a 3D fluorescence microscope dataset ($1024 \times 1024 \times 19$), corrupted by 15dB Gaussian white noise. We observe that 3D-TV results patchy image, while 3D-HDTV preserves the features of the image better.



(a) actual (b) inverse FFT (c) 3DTV (d) 3DHDTV



(e) actual (f) inverse FFT (g) 3DTV (h) 3DHDTV

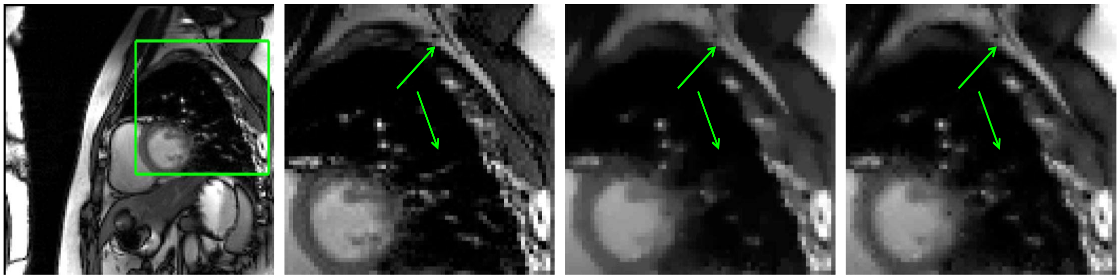


(i) actual (j) inverse FFT (k) 3DTV (l) 3DHDTV

Figure 4.7: The zoomed images of the three regions pointed in Fig. 4.6. The three rows indicate the three different regions reconstructed by inverse Fourier transform, 3D-TV, 3D-HDTV, separately. We observe that 3D-HDTV preserves more line-like features compared with 3D-TV (see green arrows).

4.5 Contributions

Yue Hu, Mathews Jacob. A fast majorize minimize algorithm for higher degree total variation regularization. IEEE *International Symposium on Biomedical Imag-*

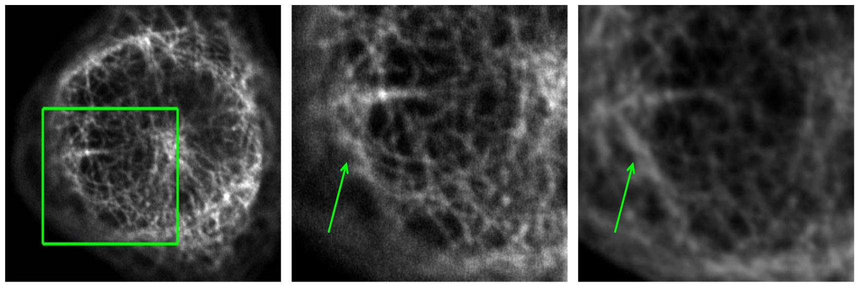


(a) actual (b) zoomed region (c) 3DTV 18.17dB (d) 3DHDTV 18.56dB

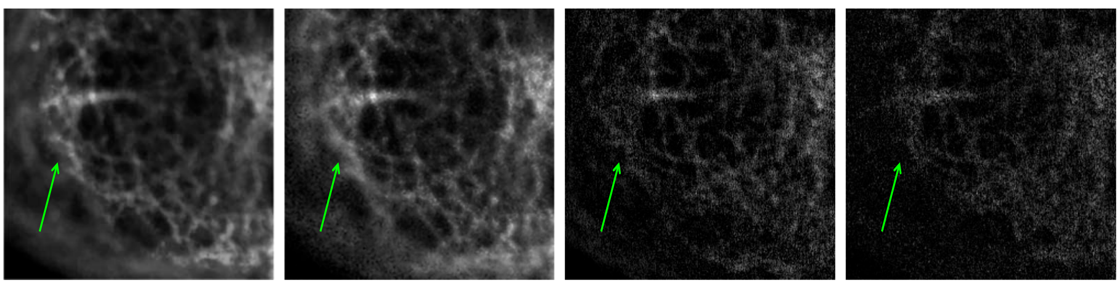
Figure 4.8: Compressed sensing recovery of cardiac MR dataset from noisy and undersampled Fourier data. (acceleration of 2 with 15dB additive Gaussian noise.) (a) is the actual image. (b) is the zoomed region indicated in the green box in (a). (c) and (d) are the reconstructions by 3D-TV and 3D-HDTV schemes separately. It is shown that 3D-HDTV provides images with more details preserved (indicated in green arrows), compared with 3D-TV. There is a 0.4dB improvement in SNR of 3D-HDTV over 3D-TV.

ing: From Nano to Macro. San Francisco, USA, 2013.

Yue Hu, Mathews Jacob. Improved higher degree total variation regularization. IEEE *International Symposium on Biomedical Imaging: From Nano to Macro.* Barcelona, Spain, 2012.

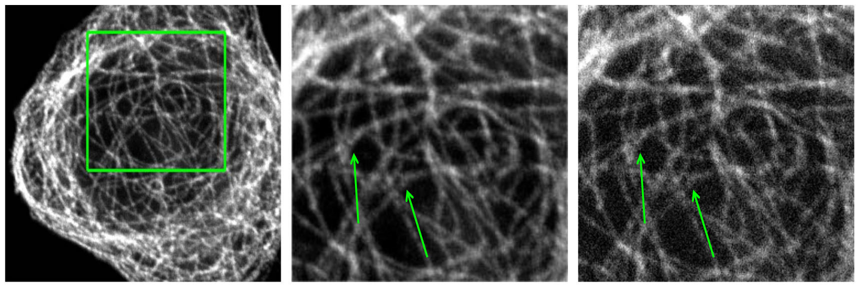


(a) actual (b) zoomed region (c) blurred image

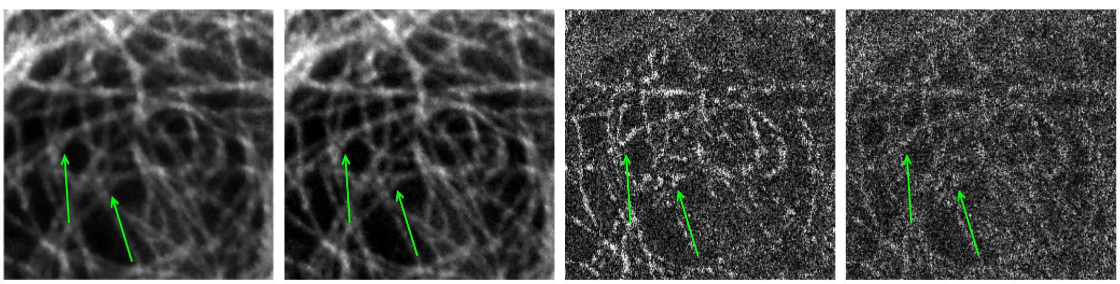


(d) 3D-TV 15.5dB (e) 3D-HDTV 15.88dB (f) 3D-TV error image (g) 3D-HDTV error image

Figure 4.9: Deconvolution of a 3D fluorescence microscope dataset. (a) is the original image. (b) is the zoomed region indicated in the green box in (a). (c) is the blurred image using a Gaussian filter with standard deviation of 1 and size of $5 \times 5 \times 5$ with additive Gaussian noise added ($\sigma = 0.01$). (d) and (e) are deblurred images using 3D-TV and 3D-HDTV separately. (f) and (g) are the error images of 3D-TV and 3D-HDTV recovery. We observe that the 3D-TV recovery is very patchy and some small details are lost. While 3D-HDTV preserves the line-like features (pointed with green arrow) with a 0.38dB improvement in SNR. The error images show that 3D-HDTV captures more details and provides more accurate results.



(a) actual (b) zoomed region (c) noisy image



(d) 3DTV 17.76dB (e) 3DHDTV 17.99dB (f) 3DTV error image (g) 3DHDTV error image

Figure 4.10: Denoising of 3D fluorescence microscope data. (a) is the actual image. (b) is the zoomed region indicated in the green box of (a). (c) is the noisy image corrupted by a 15dB Gaussian white noise. (d) and (e) are the denoised images using 3D-TV and 3D-HDTV schemes separately. (f) and (g) are the error images of the 3D-TV and 3D-HDTV results. It is observed that 3D-TV gives more patchy and blurred results compared with 3D-HDTV, which improves the SNR of the denoised image by 0.23dB. The error images show that 3D-HDTV provides more accurate results and preserves more details compared with 3D-TV (see green arrow).

5 TV Sparsity and Low Rank (TV-SLR) Algorithm

Dynamic MRI plays an important role in many MRI applications including cardiac, brain, breast, and perfusion imaging. Recently, it is demonstrated that dynamic imaging dataset could be modeled as a sparse and low-rank Casoratti matrix. It is also known that the sparsity and low rank properties are somewhat complementary. In this context, another contribution of the proposal is to develop an efficient algorithm for the recovery of sparse and low rank matrices, especially dynamic MRI reconstruction. In this chapter, we introduce an efficient majorize-minimize (MM) based combined TV Sparsity penalty and Low Rank penalty (TV-SLR) algorithm for sparse and low rank matrix recovery. To demonstrate the utility of the algorithm, an arbitrary low-rank and sparse matrix (MIT logo image) is recovered under different number of measurements. It is seen that compared to only low-rank penalty or TV sparsity penalty, the combined TV-SLR algorithm is able to perfectly reconstruct the matrix using far less number of measurements. Followed by this experiment, we consider the recovery of a dynamic MRI dataset and the number of measurements required for accurate recovery is significantly reduced, which enables accelerated MRI.

5.1 Introduction

5.1.1 Dynamic imaging using matrix recovery schemes

Our motivation in developing this algorithm is to use it in dynamic imaging and video restoration. We denote the spatio-temporal signal as $\gamma(\mathbf{x}, t)$, where $\mathbf{x}$ is the spatial location and t denotes time. We denote the sparse and noisy measurements to be related to $\boldsymbol{\gamma}$ as $\mathbf{b} = \mathcal{A}(\gamma) + \mathbf{n}$, where $\mathcal{A}$ is the measurement operator and $\mathbf{n}$ is the noise process. The vectors corresponding to the temporal profiles of the voxels are often highly correlated or linearly dependent. The spatio-temporal signal $\gamma(\mathbf{x}, t)$ can be re-arranged as a Casoratti matrix to exploit the correlations [86; 87; 88; 89]:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma\left(\mathbf{x}_0, t_0\right) & \dots & \gamma\left(\mathbf{x}_0, t_{n-1}\right) \\ \vdots & & \\ \gamma\left(\mathbf{x}_{m-1}, t_0\right) & \dots & \gamma\left(\mathbf{x}_{m-1}, t_{n-1}\right) \end{bmatrix} \tag{5.1}$$

The i^{th} row of $\mathbf{\Gamma}$ corresponds to the temporal intensity variations of the voxel $\mathbf{x}_i$. Similarly, the j^{th} column of $\mathbf{\Gamma}$ represents the image at the time point t_j. Since the rows of this $m \times n$ matrix are linearly dependent, the rank of $\mathbf{\Gamma}$ is given by $r < \min(m, n)$. We will refer to the dynamic imaging dataset either as $\gamma(\mathbf{x}, t)$ or as $\mathbf{\Gamma}$ in the remaining sections. The low-rank structure of dynamic imaging datasets was used to recover them from undersampled Fourier measurements by several authors[86; 87; 88; 90; 89]. These schemes either rely on simpler two-step algorithms, which are relatively inefficient at high acceleration factors, or greedy low-rank decomposition schemes. In contrast to these methods, the proposed scheme is computationally efficient, accurate, highly flexible, and is capable of using multiple non-convex spectral and sparsity priors.

5.1.2 Matrix recovery using nuclear norm minimization

Current theoretical results indicate that a matrix $\mathbf{\Gamma} \in \mathbb{R}^{m\times n}$ of rank $r; r \leq \min(m,n)$ can be perfectly recovered from its linear measurements $\mathbf{b} = \mathcal{A}\left(\mathbf{\Gamma}\right)$ [91; 92]. This recovery can be formulated as the constrained optimization problem:

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma}} \|\mathcal{A}\left(\mathbf{\Gamma}\right) - \mathbf{b}\|^2 \text{ such that } \operatorname{rank}\left(\mathbf{\Gamma}\right) \leq r. \tag{5.2}$$

To realize computationally efficient algorithms, the above problem is often reformulated as an unconstrained convex optimization scheme

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma}} \underbrace{\|\mathcal{A}\left(\mathbf{\Gamma}\right) - \mathbf{b}\|^2 + \lambda\left\|\mathbf{\Gamma}\right\|_*}_{\mathcal{C}(\mathbf{\Gamma})}, \tag{5.3}$$

where $\|\mathbf{\Gamma}\|_*$ is the nuclear norm of the matrix $\mathbf{\Gamma} = \sum_{i=1}^{\min(m,n)} \sigma_i \mathbf{u}_i\, \mathbf{v}_i^H$. This penalty is the convex relaxation of the rank and is defined as the sum of the singular values of $\mathbf{\Gamma}$: $\|\mathbf{\Gamma}\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i$.

5.1.3 Matrix recovery using iterative singular value thresholding

The common approaches to solve for (5.3) involve different flavors of iterative singular value thresholding (IST) [93; 94; 95]. These schemes majorize the data-consistency term in (5.3) with a quadratic expression:

$$\|\mathcal{A}\left(\mathbf{\Gamma}\right) - \mathbf{b}\|^2 \leq \tau\|\mathbf{\Gamma} - \mathbf{Z}_n\|^2 + c_n.$$

Here, τ is a constant such that $\tau\mathcal{I} \geq \mathcal{A}^t\mathcal{A}$, c_n is a constant that is independent of $\mathbf{\Gamma}$, and $\mathbf{Z}_n = \mathbf{\Gamma}_n - \mathcal{A}^t(\mathcal{A}(\mathbf{\Gamma}_n) - \mathbf{b})/\tau$. Here, $\mathcal{I}$ is the identity operator. Thus, we have

$$\mathcal{C}(\mathbf{\Gamma}) \leq \underbrace{\tau\|\mathbf{\Gamma} - \mathbf{Z}_n\|^2 + \lambda\|\mathbf{\Gamma}\|_*}_{\mathcal{C}_{\text{maj}}(\mathbf{\Gamma})} + c_n \tag{5.4}$$

The minimization of the above expression is termed as the proximal mapping of $\mathbf{Z}_n$, associated with the nuclear norm penalty [96]. This proximal mapping has an analytical solution [93]:

$$\mathbf{\Gamma}_{n+1} = \sum_{i=1}^{\min(m,n)} (\sigma_i - \lambda/2\tau)_+ \, \mathbf{u}_i \, \mathbf{v}_i^H, \tag{5.5}$$

where, $\mathbf{u}_i$, $\mathbf{v}_i$ are the singular vectors and σ_i are the singular values of $\mathbf{Z}_n$. The thresholding function in (5.5) is defined as

$$(\sigma)_+ = \begin{cases} \sigma & \text{if} \quad \sigma \geq 0 \\ 0 & \text{else} \; . \end{cases} \tag{5.6}$$

Unfortunately, it is not straightforward to adapt this algorithm to optimization schemes with multiple non-differentiable penalty terms (e.g. spectral and sparsity penalties), as discussed previously.

5.2 Combined TV Sparsity and Low Rank Regularized Algorithm

We introduce the problem formulation and the algorithm in this section. The details of the numerical implementation are covered in Section 5.3.

5.2.1 Matrix recovery using sparsity and spectral penalties

To exploit the low-rank and sparsity of the matrix in the transform domain (specified by $\mathbf{R}$ and $\mathbf{C}$), we formulate the matrix recovery as the constrained optimization scheme:

$$\begin{aligned}\mathbf{\Gamma}^* &= \arg\min_{\mathbf{\Gamma}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 \\ &\quad \text{such that } \left\{\text{rank}(\mathbf{\Gamma}) \leq \text{r}, \|\mathbf{R}^{\text{H}}\mathbf{\Gamma}\mathbf{C}\|_{\ell_0} < \text{K}\right\}\end{aligned} \tag{5.7}$$

We rewrite the above constrained optimization problem using Lagrange's multipliers and relax the penalties to obtain

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 + \lambda_1\, \phi(\mathbf{\Gamma}) + \lambda_2\, \psi(\mathbf{\Gamma}). \tag{5.8}$$

Here, the spectral penalty ϕ is the relaxation of the rank constraint. We choose it as the class of Schatten p matrix penalties ($\mu(\sigma) = |\sigma|^p$), specified by

$$\phi(\mathbf{\Gamma}) = \sum_{i=1}^{\min(m,n)} \sigma_i(\mathbf{\Gamma})^{p_1}. \tag{5.9}$$

Similarly, we specify the sparsity penalty as $\psi(\mathbf{\Gamma}) = \|\mathbf{R}^H\mathbf{\Gamma}\mathbf{C}\|_{\ell_{p_2}}^{p_2}$, which is the ℓ_{p_2} norm of the matrix entries, specified by:

$$\|\mathbf{\Gamma}\|_{\ell_p}^p = \sum_{i,j} |\mathbf{\Gamma}_{i,j}|^p. \tag{5.10}$$

When $p_1, p_2 \geq 1$, the cost function (5.8) is convex and hence has a unique minimum. We now generalize the sparsity penalty to account for non-separable convex and non-convex total variation-like penalties:

$$\psi(\gamma) = \int_{\mathbb{R}^3} \|\nabla\gamma(\mathbf{x}, t)\|_2^{p_2}\, d\mathbf{x}dt, \tag{5.11}$$

which are widely used in imaging applications [97; 98]. The above penalty is often implemented using finite difference operators. Rewriting the above expression in terms of the matrix $\mathbf{\Gamma}$, we get

$$\psi(\mathbf{\Gamma}) = \varphi(\mathbf{P}) = \|\mathbf{P}\|_{\ell_{p_2}}^{p_2}, \tag{5.12}$$

where, $\mathbf{P} = \sqrt{\sum_{i=1}^{q} |\mathbf{R}_i^H \mathbf{\Gamma} \mathbf{C}_i|^2}$. Here $\mathbf{R}_i, \mathbf{C}_i, i = 1, .., q$ are matrices that operate on the rows and columns of $\mathbf{\Gamma}$, respectively. The non-separable gradient penalty in (5.11) is thus obtained when $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ correspond to the finite differences of $\gamma(\mathbf{x}, t)$ along x, y and t respectively; $\mathbf{R}_i$ and $\mathbf{C}_i$ are the corresponding finite difference matrices.

Gao et al., have recently used a linear combination of sparse and a low-rank matrices [99] to model dynamic imaging dataset and recover it from undersampled measurements. They chose the regularization parameters such that the low-rank component is the static background signal. The dynamic components are assumed to be sparse in a pre-selected basis/frame, which is enforced by a convex sparsity prior. The use of a sparse model to capture the dynamic components is conceptually similar to classical compressed sensing dynamic imaging schemes [100; 101; 102]. We have shown that the basis functions estimated from the data itself (using low-rank recovery) are more effective in representing the data compared to preselected basis functions, especially when significant respiratory motion is present [88]. We plan to compare the proposed scheme with the model in [99] and other state of the art dynamic imaging schemes in the future.

5.2.2 Algorithm formulation

We now derive a fast MM algorithm to solve (5.8). Specifically, we majorize the penalty terms by quadratic functions of $\mathbf{\Gamma}$:

$$\phi(\mathbf{\Gamma}) = \min_{\mathbf{W}} \frac{\beta_1}{2} \|\mathbf{\Gamma} - \mathbf{W}\|_F^2 + \eta(\mathbf{W}), \tag{5.13}$$

$$\psi(\mathbf{\Gamma}) = \min_{\{\mathbf{Q}_i, i=1,..,q\}} \frac{\beta_2}{2} \sum_{i=1}^{q} \|\mathbf{R}_i \mathbf{\Gamma} \mathbf{C}_i - \mathbf{Q}_i\|_F^2 + \theta\left(\sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2}\right) \tag{5.14}$$

Here, $\mathbf{W}$ and $\mathbf{Q}_i, i = 1, .., q$, are auxiliary matrix variables and $\|\mathbf{\Gamma}\|_F$ is the Frobenius norm of $\mathbf{\Gamma}$. By definition, $\eta(\mathbf{W})$ and $\theta(\sqrt{\sum_i |\mathbf{Q}_i|^2})$ are matrix functions that are dependent on $\phi(\mathbf{W})$ and $\varphi(\mathbf{P})$ respectively.

Analytical expressions for η and θ can be derived in many cases as shown below. However, we find in Section 5.3 that analytical expressions for η and θ are not required for efficient implementation. Using the above majorizations, we simplify the original cost function in (5.8) as

$$(\mathbf{\Gamma}, \mathbf{W}, \{\mathbf{Q}_i\})_{\text{opt}} = \arg \min_{\mathbf{\Gamma}, \mathbf{W}, \{\mathbf{Q}_i\}} \mathcal{C}\left(\mathbf{\Gamma}, \mathbf{W}, \mathbf{Q}_i\right) \tag{5.15}$$

where

$$\begin{aligned} \mathcal{C} \;=\;& \|\mathcal{A}\left(\mathbf{\Gamma}\right) - \mathbf{y}\|^2 + \frac{\lambda_1 \beta_1}{2} \|\mathbf{\Gamma} - \mathbf{W}\|_F^2 + \frac{\lambda_2 \beta_2}{2} \sum_{i=1}^{q} \|\mathbf{R}_i \mathbf{\Gamma} \mathbf{C}_i - \mathbf{Q}_i\|_F^2 \\ & + \lambda_1 \, \eta\left(\mathbf{W}\right) + \lambda_2 \, \theta \left(\sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2} \right) \end{aligned} \tag{5.16}$$

We propose to use an iterative alternating minimization scheme to minimize the above criterion. Specifically, we alternatively minimize (5.16) with respect to each of the variables, assuming others to be fixed. We denote the n^{th} iterate of these variables as $\mathbf{\Gamma}_n$, $\mathbf{W}_n$, and $\mathbf{Q}_{i,n}; i = 1, .., q$, respectively. One iteration of this scheme is described below.

1. Derive $\mathbf{\Gamma}_{n+1}$, assuming $\mathbf{W} = \mathbf{W}_n, \mathbf{Q}_i = \mathbf{Q}_{i,n}$:

$$\begin{aligned} \mathbf{\Gamma}_{n+1} \;=\;& \arg \min_{\mathbf{\Gamma}} \|\mathcal{A}\left(\mathbf{\Gamma}\right) - \mathbf{y}\|^2 + \frac{\lambda_1 \beta_1}{2} \|\mathbf{\Gamma} - \mathbf{W}_n\|_F^2 \\ & + \frac{\lambda_2 \beta_2}{2} \sum_{i=1}^{q} \|\mathbf{R}_i \mathbf{\Gamma} \mathbf{C}_i - \mathbf{Q}_{i,n}\|_F^2 \end{aligned} \tag{5.17}$$

Since this expression is quadratic in $\mathbf{\Gamma}$, we derive the analytical solutions for many measurement operators in Section 5.3.

2. Derive $\mathbf{W}_{n+1}$, assuming $\mathbf{\Gamma} = \mathbf{\Gamma}_{n+1}$:

$$\mathbf{W}_{n+1} = \arg\min_{\mathbf{W}} \frac{\beta_1}{2} \|\mathbf{\Gamma}_{n+1} - \mathbf{W}\|_F^2 + \eta\left(\mathbf{W}\right) \tag{5.18}$$

The optimal $\mathbf{W}$ is thus obtained as the proximal mapping of $\mathbf{\Gamma}_{n+1}$, corresponding to the spectral penalty η. We derive analytical expressions for this step for the widely used nuclear norm and Schatten-p functionals in Section 5.3.

3. Derive $\mathbf{Q}_{i,n+1}$, assuming $\mathbf{\Gamma} = \mathbf{\Gamma}_{n+1}$:

$$\mathbf{Q}_{i,n+1} \quad = \quad \arg\min_{\mathbf{Q}_i} \frac{\beta_2}{2} \sum_{i=1}^{q} \|\mathbf{R}_i \mathbf{\Gamma}_{n+1} \mathbf{C}_i - \mathbf{Q}_i\|_F^2 + \theta\left(\sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2}\right) \tag{5.19}$$

The optimal $\{\mathbf{Q}, i = 1, .., q\}$ is thus the proximal mapping of $\{\mathbf{R}_i \mathbf{\Gamma}_{n+1} \mathbf{C}_i; i = 1, .., q\}$, associated with the matrix penalty θ. Since θ is non-separable, the corresponding shrinkage involves the simultaneous processing of the component matrices $\mathbf{R}_i \mathbf{\Gamma}_{n+1} \mathbf{C}_i; i = 1, .., q$. This step also has analytical expressions, as shown in section 5.3.

5.2.3 Expression of $\eta(\mathbf{W})$

We now focus on determining the function η, such that the majorization of the spectral penalty term in (5.13) holds. Since analytical expressions for η and θ are not essential to realize an efficient algorithm, readers may skip this section and go directly to Section 5.3.

We reorder the terms in (5.13) to obtain

$$\underbrace{\frac{\|\mathbf{\Gamma}\|_F^2}{2} - \frac{\phi(\mathbf{\Gamma})}{\beta_1}}_{g(\mathbf{\Gamma})} = \max_{\mathbf{W}} \left[\langle \mathbf{\Gamma}, \mathbf{W} \rangle - \underbrace{\left(\|\mathbf{W}\|_F^2/2 + \eta(\mathbf{W})/\beta_1\right)}_{f(\mathbf{W})} \right] \tag{5.20}$$

Here, $\langle \mathbf{W}, \mathbf{\Gamma} \rangle = \text{trace}\left(\mathbf{W}^T\mathbf{\Gamma}\right)$ is the innerproduct of two matrices. From the theory in [103], the above relation is satisfied if $g(\mathbf{\Gamma})$ is a convex function and $f = g^*$ is the convex dual of g:

$$g^*(\mathbf{W}) = \max_{\mathbf{\Gamma}} \left(\langle \mathbf{W}, \mathbf{\Gamma} \rangle - g(\mathbf{\Gamma})\right). \tag{5.21}$$

Note that ϕ need not be convex for the above relation to hold. This majorization is valid if $g(\mathbf{\Gamma})$ is convex, which is possible even when ϕ is concave. Thanks to the property of unitarily invariant functions, the dual of a specified matrix function $g(\mathbf{\Gamma}) = \sum \mu_g(\sigma_i(\mathbf{\Gamma}))$ is obtained as

$$f(\mathbf{W}) = \sum \mu_g^*(\sigma_i(\mathbf{W})). \tag{5.22}$$

Thus, $\mu_f(\cdot) = \mu_g^*(\cdot)$ is the convex conjugate of $\mu_g(\cdot)$. From the above relations, we have $\eta(\mathbf{W}) = \sum \mu_\eta(\sigma_i(\mathbf{W}))$, where $\mu_\eta(x) = \beta_1 \left(\mu_f(x) - x^2/2\right)$.

We now approximate the non-differentiable ϕ penalties by continuously differentiable Huber functionals. These approximations are required to ensure that $g(\mathbf{\Gamma})$ is convex. In addition, differentiability of ϕ also provides additional simplifications.

1. Nuclear norm: We approximate the nuclear norm penalty $\|\mathbf{\Gamma}\|_* = \sum_i \sigma_i(\mathbf{\Gamma})$ as $\phi_{\beta_1}(\mathbf{\Gamma}) = \sum_i \mu_{\phi_{\beta_1}}(\sigma_i(\mathbf{\Gamma}))$. Here, $\mu_{\phi_{\beta_1}}(x)$ is the standard scalar Huber function

$$\mu_{\phi_{\beta_1}}(x) = \begin{cases} |x| - 1/2\beta_1 & \text{if} \quad x \geq \frac{1}{\beta_1} \\ \beta_1 |x|^2/2 & \text{else}. \end{cases} \tag{5.23}$$

Note that $\phi_{\beta_1}(\mathbf{\Gamma}) \to \|\mathbf{\Gamma}\|_*$ as $\beta_1 \to \infty$. With this choice, the corresponding $g(\mathbf{\Gamma}) = \sum \mu_g(\sigma_i(\mathbf{\Gamma}))$ is given by

$$\mu_g(x) = \begin{cases} \frac{1}{2}\left(x - \frac{1}{\beta_1}\right)^2 & \text{if} \quad x \geq \frac{1}{\beta} \\ 0 & \text{else} \end{cases} \tag{5.24}$$

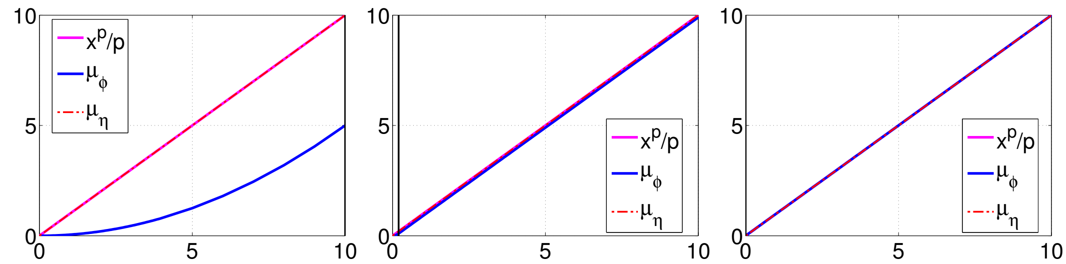
Note that g is convex for any β_1. Using the property of convex conjugate functions described earlier, we find in the Appendix that $\mu_\eta = w$. Thus, we have $\eta(\mathbf{W}) = \|\mathbf{W}\|_*$, $\forall \beta_1$.

2. Schatten p norm: We approximate the Schatten p matrix norm by the corresponding Huber matrix function:

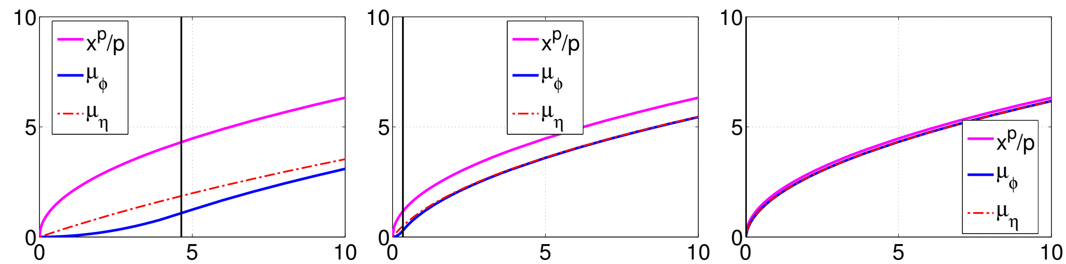
$$\mu_{\phi_{\beta_1}}(x) = \begin{cases} \frac{x^p}{p} - 1/\left(2\alpha\beta_1^\alpha\right) & \text{if} \quad x \geq \beta_1^{1/(p-2)} \\ \beta_1 x^2/2 & \text{else} \end{cases} \tag{5.25}$$

   Here, $\alpha = p/(2-p)$. The threshold specified by $\beta_1^{1/(p-2)}$ and the constant $1/\left(2\alpha\beta_1^\alpha\right)$ is chosen such that $\mu_{\phi_{\beta_1}}$ is continuously differentiable and μ_g is convex. The above formula is essentially an extension of the generalization proposed by [97] to matrix functionals. It is difficult to derive analytical expressions for $\mu_\eta(w)$ for arbitrary $p < 1$. However, we can numerically solve for $x = \partial_x \mu_g$ and evaluate $\mu_\eta(w)$ for specific values of w, as shown in Fig. 5.1. We show in the next section that analytical expressions for the proximal mapping, specified by (5.18), can be derived even if analytical expressions for η are not available.

We plot μ_ϕ, μ_η and x^p/p for $p = 1$ and $p = 0.5$ for different values of β_1 in Fig. 5.1. Note that $\mu_\eta(x) = |x|, \forall\beta_1$, when $p = 1$. However, $\mu_\eta(x)$ is different from $|x|^p/p$ when $p < 1$. This implies that the variable splitting interpretation of the majorize minimize algorithm breaks down when $p < 1$, as explained in Section 5.3.



(a) $p = 1; \beta_1 = 0.1$ (b) $p = 1; \beta_1 = 5$ (c) $p = 1; \beta_1 = 1000$



(d) $p = 0.5; \beta_1 = 0.1$ (e) $p = 0.5; \beta_1 = 5$ (f) p=0.5;β_1=1000

Figure 5.1: Huber approximation of the spectral penalty and the corresponding η function. Note that the approximation of the original spectral penalty by the Huber function improves with increasing values of β. Clearly, large values of β are required to approximate Schatten p-norms; $p < 1$. It is observed that $\eta(x) = |x|, \forall \beta$ when $p = 1$. Hence, the variable splitting interpretation (see Section 5.3.5) is equivalent to the MM scheme. However, this equivalence breaks down when $p < 1$. Specifically, $\eta(\mathbf{\Gamma}) \to \|\mathbf{\Gamma}\|_p^p$ only when $\beta \to \infty$.

5.2.4 Expression for θ

The Huber approximation of the total variation norm ($p_2 = 1$) was considered in [104], where they showed that

$$\theta\left(\sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2}\right) = \sqrt{\sum_{i=1}^{q} |\mathbf{Q}_i|^2}; \quad \forall \beta_2 \tag{5.26}$$

Analytical expressions of θ cannot be obtained when $p_2 < 1$. However, we derive the analytical expression for the shrinkage in Section 5.3, which will enable the efficient implementation.

5.3 Numerical Implementation of Three Subproblems

We now focus on the numerical implementation of the three main subproblems. Specifically, we show that all of the three steps can be either solved analytically or using efficient algorithms for most penalties and measurement operators of practical interest. This enables us to realize a computationally efficient algorithm. We also introduce a continuation scheme to accelerate the convergence of the algorithm.

5.3.1 Quadratic subproblem, specified by (5.17)

Since the subproblem (5.17) is entirely quadratic, we rewrite it as a Tikhnonov regularized image recovery problem:

$$\boldsymbol{\gamma}_{n+1} = \arg\min_{\boldsymbol{\gamma}} \|\mathcal{A}(\boldsymbol{\gamma}) - \mathbf{y}\|^2 + \frac{\lambda_1\beta_1}{2}\|\boldsymbol{\gamma} - \mathbf{w}_n\|^2 + \frac{\lambda_2\beta_2}{2}\sum_{i=1}^{q}\|\mathcal{G}_i(\boldsymbol{\gamma}) - \mathbf{q}_{n,i}\|^2 \quad (5.27)$$

Here, $\boldsymbol{\gamma} \leftrightarrow \boldsymbol{\Gamma}$ and $\boldsymbol{q}_i \leftrightarrow \mathbf{Q}_i$ are the 3-D datasets, corresponding to the corresponding Casoratti matrices. Similarly, $\mathcal{G}_i$ is the linear operator such that $\mathcal{G}_i(\boldsymbol{\gamma}) \leftrightarrow \mathbf{R}_i^*\boldsymbol{\Gamma}\mathbf{C}_i$. We obtain the Euler-Lagrange equation of this variational problem as

$$\left(\mathcal{A}^*\mathcal{A} + \lambda_1\beta_1\mathcal{I} + \lambda_2\beta_2\sum_{i=1}^{q}\mathcal{G}_i^*\mathcal{G}_i\right)\boldsymbol{\gamma}_{n+1} = \mathcal{A}^*\mathbf{y} + \lambda_1\beta_1\,\mathbf{w}_n + \lambda_2\beta_2\sum_{i=1}^{q}\mathcal{G}_i^*\mathbf{q}_{n,i} \quad (5.28)$$

Here $\mathcal{I}$ is the identity operator. Note that the variables in the right hand side of (5.28) are fixed. Thus, this step involves the solution to a linear system of equations. In the general setting, this system of equations can be solved efficiently using iterative algorithms such as conjugate gradient (CG). A few CG steps are often sufficient for good convergence since the algorithm is initialized by the previous iterate $\boldsymbol{\gamma}_n$. We now show that analytical solutions of (5.27) do exist for many measurement operators. When the TV penalty is used, the above equation can be rewritten as

$$\left(\mathcal{A}^*\mathcal{A} + \lambda_1\beta_1\,\mathcal{I} + \lambda_2\beta_2\,\triangle\right)\boldsymbol{\gamma}_{n+1} = \mathcal{A}^*\mathbf{y} + \lambda_1\beta_1\,\mathbf{w}_n + \lambda_2\beta_2\,p_n. \tag{5.29}$$

Here, $\triangle\boldsymbol{\gamma}_{n+1}$ is the 3-D Laplacian of $\boldsymbol{\gamma}_{n+1}$ and $p_n = \nabla\cdot\mathbf{q}_n$ is the divergence of the vector field $\mathbf{q}_n$.

- Fourier sampling: An analytical expression can be derived for (5.27), when the measurements are Fourier samples of $\boldsymbol{\gamma}$ on a Cartesian grid. Specifically, we assume that the index set corresponding to the measured samples to be indicated by Λ and the corresponding measurements to be b_i; i.e., ($b_i = \hat{\gamma}\left(\boldsymbol{\omega}_i\right)$, where $\hat{\gamma}$ is the discrete Fourier transform of γ). We split the frequency samples, specified by ω, into two sets Λ and Λ^c and evaluate the discrete Fourier transform of both sides of (5.29) to obtain:

$$\hat{\gamma}_{n+1}\left(\boldsymbol{\omega}_i\right) = \begin{cases} \frac{b_i+\lambda_1\beta_1\,\hat{w}_n(\boldsymbol{\omega}_i)+\lambda_2\beta_2\,\hat{p}_n(\boldsymbol{\omega}_i)}{1+\,\lambda_1\beta_1+\,\lambda_2\beta_2\|\boldsymbol{\omega}_i\|^2} & \text{if} \quad \boldsymbol{\omega}_i\in\Lambda \\ \frac{\lambda_1\beta_1\,\hat{w}_n(\boldsymbol{\omega}_i)+\lambda_2\beta_2\,\hat{p}_n(\boldsymbol{\omega}_i)}{\lambda_1\beta_1+\,\lambda_2\beta_2\|\boldsymbol{\omega}_i\|^2} & \text{else}\ . \end{cases} \tag{5.30}$$

  Here, $p_n = \nabla\cdot\mathbf{q}_n$.

- Deconvolution: Convolution can be posed as a multiplication in the Fourier domain. Considering the Fourier transform of the matrix, (5.29) can be solved in the Fourier domain as

$$\hat{\gamma}_{n+1}\left(\boldsymbol{\omega}\right) = \frac{\hat{H}\left(\boldsymbol{\omega}\right)^*\hat{b}\left(\boldsymbol{\omega}\right)+\lambda_1\beta_1\,\hat{w}_n\left(\boldsymbol{\omega}\right)+\lambda_2\beta_2\,\hat{p}_n\left(\boldsymbol{\omega}\right)}{|\hat{H}\left(\boldsymbol{\omega}\right)|^2+\,\lambda_1\beta_1+\,\lambda_2\beta_2\|\boldsymbol{\omega}\|^2}. \tag{5.31}$$

  Here, $H(\boldsymbol{\omega})$ is the transfer function of the blurring filter, $\boldsymbol{\omega}$ is the frequency, $\hat{b}(\boldsymbol{\omega})$ is the Fourier transform of the measured blurred image dataset and $\hat{p}_n(\boldsymbol{\omega})$ is the Fourier transform of $\nabla\cdot\mathbf{q}_n$.

5.3.2 Subproblem 2, specified by (5.18)

We will now focus on (5.18) and derive the analytical expression for $\mathbf{W}_{n+1}$:

$$\begin{aligned}\mathbf{W}_{n+1} &= \arg\min_{\mathbf{W}} \overbrace{\beta\|\mathbf{\Gamma}_{n+1}\|^2/2 + \beta\|\mathbf{W}\|_F^2/2 - \beta\left\langle \mathbf{\Gamma}_{n+1}, \mathbf{W}\right\rangle}^{\frac{\beta}{2}\|\mathbf{\Gamma}_n - \mathbf{W}\|_F^2} + \eta\left(\mathbf{W}\right) \\ &= \arg\max_{\mathbf{W}} \beta\left\langle \mathbf{\Gamma}_{n+1}, \mathbf{W}\right\rangle - \beta \underbrace{\left(\|\mathbf{W}\|_F^2/2 + \eta(\mathbf{W})/\beta\right)}_{f(\mathbf{W})} \end{aligned} \tag{5.32}$$

The minimizer of this expression satisfies

$$\mathbf{\Gamma}_{n+1} = \beta\, \nabla f(\mathbf{W}_{n+1}). \tag{5.33}$$

We used the differentiability of ϕ, and hence f to obtain the above expression. This is valid for the Huber approximations of the spectral penalties. Since f and g are convex conjugates, ∇f and ∇g are inverse functions [96; 105]. Thus, we obtain the optimal $\mathbf{W}$ that solves (5.32) as

$$\mathbf{W}_{n+1} = \nabla f^{-1}(\mathbf{\Gamma}_{n+1}) = \nabla g(\mathbf{\Gamma}_{n+1}) = \mathbf{\Gamma}_{n+1} - \partial\phi(\mathbf{\Gamma}_{n+1})/\beta. \tag{5.34}$$

We used the relation $g(\mathbf{\Gamma}) = \|\mathbf{\Gamma}\|_F^2/2 - \phi(\mathbf{\Gamma})/\beta$ in the second step. Thus, analytical expressions for η are not required to derive the shrinkage step, thanks to the property of convex conjugate functions. We now derive the shrinkage steps for specific spectral penalties.

- <u>Special case: nuclear norm</u>

$$\partial\mu_\phi(x) = \begin{cases} \text{sign}(x) & \text{if} \quad |x| \geq \frac{1}{\beta_1} \\ \beta_1 x & \text{else} . \end{cases} \tag{5.35}$$

We assume $\mathbf{\Gamma} = \sum_{i=1}^{\min(m,n)} \sigma_i \mathbf{u}_i\, \mathbf{v}_i^H$ to be the singular value decomposition of $\mathbf{\Gamma}$. Substituting in (5.34), we get

$$\mathbf{W}^* = \sum_{i=1}^{\min(m,n)} \left(\sigma_i - 1/\beta_1\right)_+ \mathbf{u}_i\, \mathbf{v}_i^H, \tag{5.36}$$

where, $\mathbf{u}_i, \mathbf{v}_i, \sigma_i$ are the singular vectors and singular values of $\mathbf{\Gamma}$.

- <u>Schatten p-norms</u> Following the same steps, we obtain the shrinkage step for Schatten p-norms as ,

$$\mathbf{W}^* = \partial g(\mathbf{\Gamma}) = \sum_{i=1}^{\min(m,n)} \left(\sigma_i - \sigma_i^{(p_1-1)}/\beta_1\right)_+ \mathbf{u}_i\, \mathbf{v}_i^H. \tag{5.37}$$

5.3.3 Solving the subproblem 3, specified by (5.19)

Problems similar to (5.19) has been addressed in the context of iterative algorithms for total variation minimization [104] and its non-convex variants [97]. The generalized shrinkage rule to derive $\mathbf{Q}_i; i = 1, .., q$ is specified by

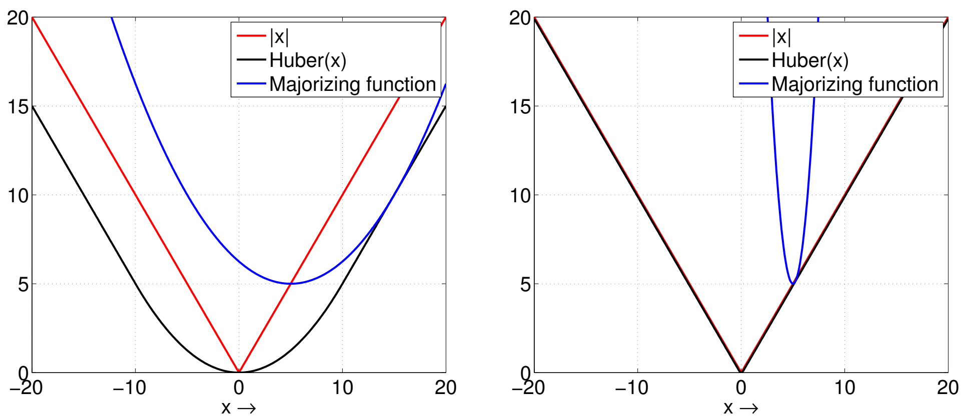
$$\mathbf{Q}_{i,n+1} = \frac{\left(\mathbf{P} - \mathbf{P}^{(p_2-1)}/\beta_2\right)}{\mathbf{P}} \mathbf{R}_i \mathbf{\Gamma} \mathbf{C}_i, \tag{5.38}$$

where $\mathbf{P} = \sqrt{\sum_{i=1}^{q} |\mathbf{R}_i \mathbf{\Gamma}_{n+1} \mathbf{C}_i|^2}$. $\mathbf{P}^p$ is the matrix whose elements are the p^{th} power of the entries of $\mathbf{P}$.

5.3.4 Continuation to improve the convergence

The three-step alternating minimization algorithm involves a tradeoff between convergence and accuracy. Specifically, when $\beta_1 = \beta_2 = 0$, (5.16) simplifies to three decoupled problems in $\mathbf{\Gamma}, \mathbf{W}$ and $\{\mathbf{Q}_i, i = 1, ..q\}$. Since all of these problems have analytical solutions, the entire algorithm converges in a single step to the minimum norm solution, which is a poor approximation of (5.8); this is expected since the corresponding Huber function is a poor approximation to the original cost function. In contrast, the approximation is exact when $\beta_1 = \beta_2 = \infty$. However, it is easy to see that the algorithm fails to converge in this case.

The above mentioned tradeoff between convergence and accuracy can be understood in terms of **(i)** the ability of the Huber function to approximate the original penalty and **(ii)** the proximity of the majorizing function to the Huber approximation. We illustrate this issue in Fig. 5.2 in the context of the nuclear norm penalty. Note that for small values of β, the Huber approximation $\mu_\phi(\sigma)$ of $|\sigma|$ is poor. However, the corresponding quadratic majorizing function $\beta(x-w)^2 + \mu_\eta(\sigma)$ closely approximates μ_ϕ. Hence, the MM scheme converges fast to the minimum of the approximate penalty. In contrast, when $\beta \to \infty$, the Huber function approximates the spectral penalty well, resulting in good accuracy of the final solution. However, the convergence is poor in this case since the approximation of the Huber function by the majorizing quadratic function is poor.



(a) $\beta = 0.1$ (b) $\beta = 5$

Figure 5.2: Effect of β_1 on convergence and accuracy: We demonstrate the approximation of $|x|$ by the corresponding Huber penalty. Note that for $\beta = 0.1$, the majorizing function approximates the Huber function well, resulting in fast convergence to the minimum of the corresponding penalty. However, in this case the approximation of $|x|$ by the Huber function is poor resulting in poor accuracy. In contrast, the approximation of $|x|$ by the Huber function is good when $\beta = 5$. In this case, the majorizing function of the Huber function is a poor approximation. Specifically, it is too narrow, resulting in slow convergence.

To overcome the above mentioned tradeoff, we introduce a continuation scheme.

Specifically, we initialize β_1 and β_2 with small values and progressively increase them, until convergence. The algorithm converges very fast for small values of β_1 and β_2, as discussed before. We use the solution at each step to initialize the next step. For each choice of continuation parameters, we iterate the algorithm to convergence (i.e., until the relative change in the cost function in (5.8) is less than a pre-specified threshold). In all the experiments considered in this chapter, we initialize the continuation parameters as $\beta_1 = \beta_2 = 5$ and increase them by a factor of five for each iteration of the outer loop.

5.3.5 Interpretation as a variable splitting scheme

The majorize-minimize scheme to solve for the spectrally regularized matrix recovery may be interpreted as a variable splitting (VS) strategy [88], similar to such schemes in compressed sensing [106; 107]:

$$\begin{aligned} \mathbf{\Gamma}^* &= \arg\min_{\mathbf{\Gamma},\mathbf{W}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 + \lambda_1\, \phi(\mathbf{W}) \\ & \qquad\qquad \text{s.t. } \mathbf{\Gamma} = \mathbf{W} \end{aligned} \tag{5.39}$$

Here, $\mathbf{W}$ is an auxiliary variable and the above constrained optimization problem is equivalent to (5.8), when $\lambda_2 = 0$. VS algorithms convert the above constrained optimization problem to an unconstrained problem by introducing an additional quadratic penalty:

$$(\mathbf{\Gamma}, \mathbf{W})^* = \arg\min_{\mathbf{\Gamma},\mathbf{W}} \|\mathcal{A}(\mathbf{\Gamma}) - \mathbf{b}\|^2 + \lambda_1\, \phi(\mathbf{W}) + \lambda_1 \frac{\beta_1}{2} \|\mathbf{\Gamma} - \mathbf{W}\|_F^2 \tag{5.40}$$

This unconstrained problem is equivalent to (5.39), when $\beta_1 \to \infty$. Note that (5.40) is very similar to (5.16) with $\lambda_2 = 0$, except that the ϕ is used instead of η. The VS scheme and the MM scheme are exactly the same for $p = 1$, since

$\eta(\mathbf{W}) = \|\mathbf{W}\|_*, \forall\beta_1$. However, $\eta(\mathbf{W}) \neq \|\mathbf{W}\|_p^p$ when $p < 1$ (see Fig. 5.1). Hence, the two schemes are not equivalent for general Schatten p-norms.

The standard practice in VS compressed sensing schemes is to alternatively minimize the criterion with respect to each of the unknowns, assuming the other variable to be fixed [106; 107]. Thus, we obtain $\mathbf{W}_{n+1}$ as

$$\mathbf{W}_{n+1} = \arg\min_{\mathbf{W}} \frac{\beta_1}{2}\|\mathbf{\Gamma}_{n+1} - \mathbf{W}\|_F^2 + \phi\left(\mathbf{W}\right)$$

When $\phi(\mathbf{\Gamma}) = \|\mathbf{\Gamma}\|_*$ [92], this proximal mapping can be efficiently implemented using singular value soft-thresholding. However, analytical closed form expressions for the above proximal mapping do not exist when non-convex spectral penalties (e.g. Schatten p norms; $\phi(\mathbf{\Gamma}) = \|\mathbf{\Gamma}\|_p^p$) are used. To realize efficient algorithms, Ehler introduced approximate hard thresholding shrinkage rules for non-convex ℓ_p penalties [108]. Adapting these approximate rules to non-convex matrix penalties, we get

$$\mathbf{W}_{n+1} = \sum_{i=1}^{\min(m,n)} \chi(\sigma_i)\mathbf{u}_i\,\mathbf{v}_i^H, \tag{5.41}$$

The thresholding function in (5.41) is defined as

$$\chi(\sigma_i) = \begin{cases} \sigma_i - \frac{2p\,c_p}{\beta_1} & \text{if} \quad \sigma_i > \frac{2c_p}{\beta_1} \\ 0 & \text{else}\ . \end{cases} \tag{5.42}$$

where $c_p = 2^{p-2}\left(\frac{(2-p)^{2-p}}{(1-p)^{1-p}}\right)$ and $\mathbf{u}_i, \mathbf{v}_i$ and σ_i are the singular vectors and values of $\mathbf{\Gamma}_{n+1}$, respectively. A similar approximate rule can be used for the non-convex TV penalty. In contrast to these approximate rules, analytical shrinkage formulas can be derived for most spectral penalties $\eta(\mathbf{\Gamma})$ in the MM framework (see (5.34)). We compare the MM algorithm and the VS scheme with the approximate shrinkage rule in the results section; we observe that the proposed MM scheme provides faster convergence, thanks to the exact shrinkage rule.

The MM scheme in the non-convex ($p_1 < 1; p_2 < 1$) case may be alternatively interpreted as an approximate VS algorithm, where (5.37) and (5.38) are used to approximate the corresponding proximal mappings. Note that these approximations are not reported before and is only inspired by the MM framework. The above approximate variable splitting scheme (with approximate MM-inspired proximal mappings) can be further accelerated using augmented-Lagrangian (AL) or split-Bregman (SB) methods [109; 110] as we have shown in [111]. Unlike conventional AL or SB methods (originally developed for convex penalties), the proposed non-convex schemes still require continuation since the approximation is only exact as $\beta \to \infty$ (see Fig. 5.1.)

5.4 Results

We will demonstrate the utility of the combined non-convex penalty in reliably recovering a sparse and low-rank matrix in Section 5.4.1. The convergence of the algorithm and the utility of continuation will be studied in Section 5.4.3. In Section 5.4.2, we demonstrate the utility of the combined non-convex penalty in recovering dynamic contrast enhanced MR images from their undersampled Fourier measurements. The dynamic MRI dataset is only approximately low-rank and sparse.

5.4.1 Recovery of a Low-rank and Sparse Matrix

We first demonstrate the benefits in using the combination of two non-convex penalties, compared to widely used nuclear norm scheme. We consider the recovery of the MIT logo from its sparse measurements to illustrate the algorithm. This image (size of 46 × 81 with 3726 pixels) is ideal for our study since it is low-rank (rank = 5) and also has sparse gradients. We use random measurement matrices and vary the number of measurements M from 100 to 1500 as in [112]. The matrix was then recovered from these measurements using (5.8) with six different parameter settings:

1. nuclear norm penalty alone ($p_1 = 1; \lambda_2 = 0$).

2. non-convex spectral penalty alone ($p_1 = 0.5; \lambda_2 = 0$).

3. standard TV penalty alone ($\lambda_1 = 0; p_2 = 1$).

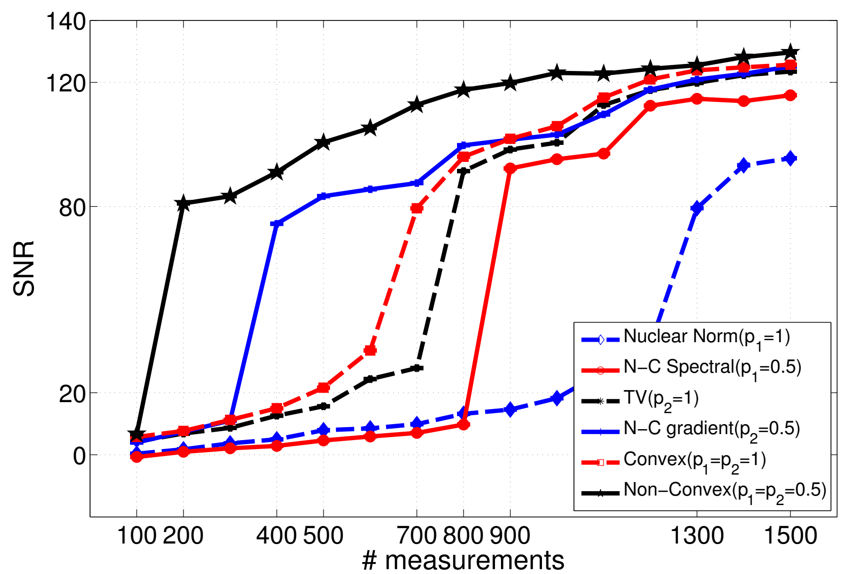


Figure 5.3: Utility of the combination of non-convex penalties. We plot the SNR as a function of the number of measurements on the MIT logo, recovered using the six algorithms. Note that the SNR increases abruptly when the number of measurements exceeds a specified threshold. It is seen that the algorithms using the non-convex spectral (red solid curve) and non-convex TV penalty (blue solid curve) alone reduce the number of measurements required to recover the image considerably over their convex counterparts (blue dotted and black dotted curves, respectively). We also observe that the combination of the convex (red dotted) and non-convex (black solid) penalties work much better than the individual penalties.

4. non-convex gradient penalty alone ($\lambda_1 = 0; p_2 = 0.5$).

5. combination of both convex penalties ($p_1 = p_2 = 1;$).

6. combination of both non-convex penalties ($p_1 = p_2 = 0.5$).

We repeat each experiment for ten different random measurement ensembles and evaluate the average signal to noise ratio (SNR), specified as

$$\mathrm{SNR} = 20 \log \left(\frac{\|\mathbf{\Gamma}_{\mathrm{orig}}\|_F}{\|\mathbf{\Gamma}_{\mathrm{rec}} - \mathbf{\Gamma}_{\mathrm{orig}}\|_F} \right). \tag{5.43}$$

For each setting, we optimize the regularization parameters (λ_1 and λ_2) with respect to SNR. Fig. 5.3 shows the SNR of the recovered image as a function of the number of measurements. It is seen that the SNR rises abruptly when the number of measurements exceeds a specified threshold. An SNR of 80 dB corresponds to almost perfect reconstruction.

We observe that the algorithm with the conventional nuclear norm scheme can perfectly recover the image, if the number of measurements is greater than 1300; these findings are consistent with the results in [112]. In contrast, the non-convex spectral penalty alone requires only 900 measurements. Similarly, the algorithm with the non-convex TV penalty alone requires only 400 measurements to perfectly recover the image, compared to 800 with standard TV. These results demonstrate the benefit in using non-convex penalties over convex schemes. We did not encounter any local minima issues. We believe that the continuation strategy, where the cost function is initialized as a quadratic criterion and gradually made non-convex, minimizes the local minima problems.

It is seen that the combined convex penalty (TV and nuclear norm) requires approximately 700 measurements, compared to 1300 with nuclear norm alone and

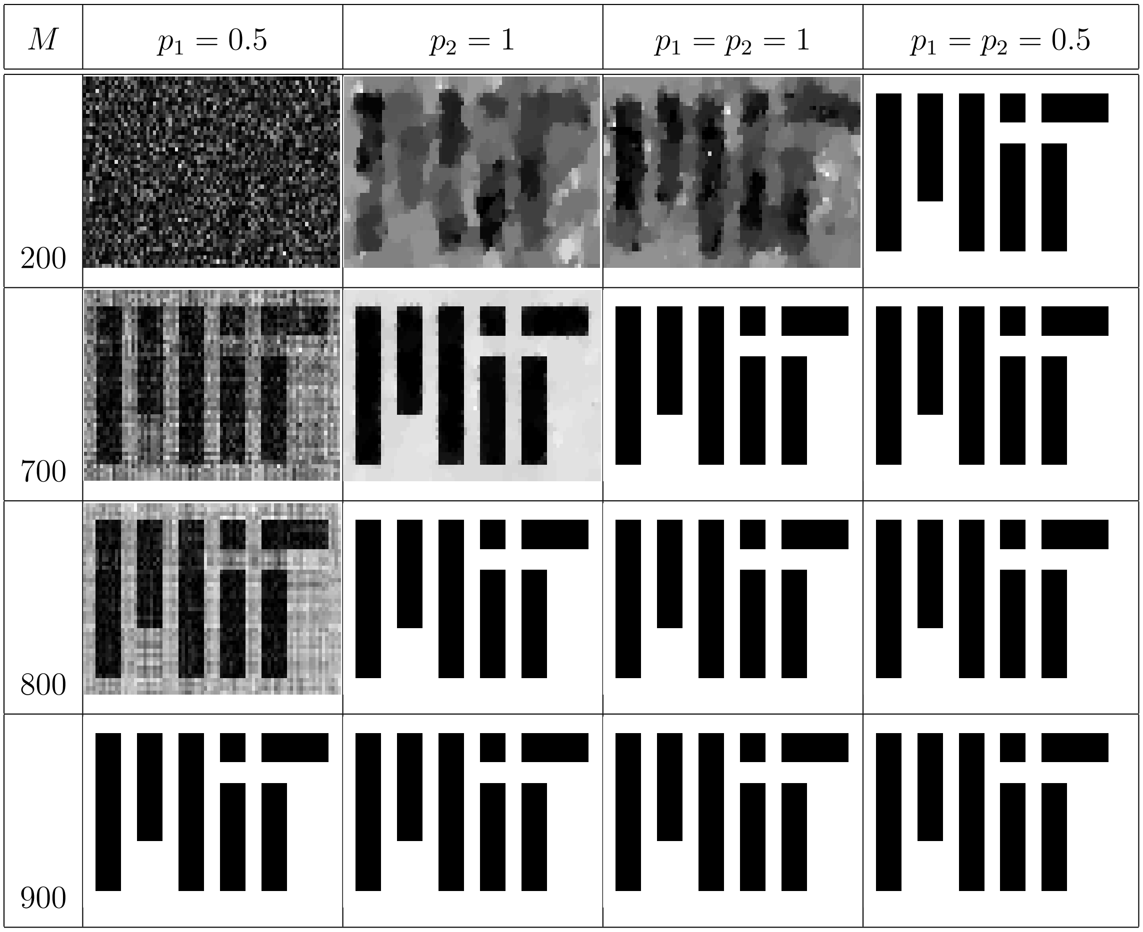


Figure 5.4: Sample recovered images for four different number of measurements using: Non-convex spectral penalty ($p_1 = 0.5$), standard TV ($p_2 = 1$), combination of convex penalties ($p_1 = p_2 = 1$), combination of non-convex penalties ($p_1 = p_2 = 0.5$). The rows correspond to different number of measurements. The results show that almost perfect reconstruction can be obtained when the number of measurements is larger than 200 by using the combination of both non-convex penalties ($p_1 = p_2 = 0.5$). Note that this is 6 times lower than using only the nuclear norm penalty (1300 measurements) and 4 times lower than standard TV (800 measurements). Similarly, we obtain a four fold improvement over non-convex spectral penalty alone.

800 with TV alone. Similarly, the combination of the non-convex penalties requires only 200 measurements, compared to 900 with non-convex spectral penalty alone and 400 with non-convex gradient penalty alone. These experiments demonstrate a significant reduction in the number of measurements required to recover a

matrix, when sparsity and spectral penalties are combined. As described earlier, sparsity and low-rank properties are complementary; since the degrees of freedom of matrices that are simultaneously sparse and low-rank are small, the joint penalty is capable of significantly reducing the number of measurements. Sample images of recovered matrices for four different number of measurements are shown in Fig. 5.4.

5.4.2 Accelerated Dynamic Contrast Enhanced (DCE) MRI

In this section, we illustrate the utility of the proposed algorithm in accelerating dynamic contrast enhanced (DCE) MRI. DCE MRI tracks the dynamic variations in the image intensity, resulting from the passage of a tracer bolus. Specifically, the paramagnetic tracer within the vasculature results in spin dephasing, hence resulting in decreased signal intensity. DCE MRI has shown great potential in diagnosing malignant lesions in the brain, breast and other organs. High temporal resolution is required to accurately estimate the kinetic parameters, while high spatial resolution is required to visualize the lesion morphology. In addition, accelerated imaging can enable the simultaneous acquisition of two echoes (T1 weighted and T2* weighted), thus enabling the accurate quantification of microvascular density and vascular permeability; these parameters are highly correlated with malignancy and have been suggested as surrogate markers for angiogenesis. Several acceleration schemes have been proposed to accelerate DCE MRI (e.g. methods that assume the contrast dynamics to mostly contain low spatial frequencies [22; 113], approaches that exploit the signal structure in the Fourier space [114], and parallel imaging schemes [115; 116]). The accelerations offered by these schemes are modest (2-3 fold), leaving room for further improvement.

We demonstrate the utility of the proposed non-convex scheme in significantly accelerating DCE MRI. The dynamic MRI measurements correspond to the samples of the signal in Fourier ($k-t$) space, corrupted by noise:

$$\mathbf{b}_i \quad = \quad \int_{\mathbf{x}} \gamma(\mathbf{x}, t_i) \exp\left(-j\mathbf{k}_i^T\mathbf{x}\right) d\mathbf{x} + \mathbf{n}_i; \quad i = 1, .., s.$$

Here, $(\mathbf{k}_i, t_i)$ indicates the i^{th} sampling location. We denote the set of sampling locations as $\Xi = \{(\mathbf{k}_i, t_i), i = 1, .., s\}$. The fully sampled 3-D dataset of a single slice is shown in Fig. 5.6.(a); the data corresponds to sixty time points, separated by TR=2 sec; the matrix size is $128 \times 128 \times 60$. We retrospectively resample each slice of the data in the Fourier domain using a uniform radial trajectory. The trajectory is rotated by a random angle for each frame to obtain an incoherent pattern (see Fig. 5.6.(b)). The number of lines per slice is chosen depending on the specified acceleration. For example, 20 k-space lines approximately corresponds to the acceleration factor of $A = 7$. We recover the dynamic imaging dataset from its undersampled Fourier measurements using the proposed scheme. We use a few steps of conjugate gradients algorithm to solve for (5.17) at each iteration, since the samples are not on the 3-D Cartesian grid. We use the previous iterate as an initial guess, thus the CG algorithm converges to the solution of (5.17) in a few steps. The recovery of the DCE MRI dataset using the MATLAB implementation of the proposed algorithm takes approximately eight minutes on an Intel quad core processor with an NVDIA Tesla graphical processing unit (GPU). The computationally expensive components of the algorithm are implemented using Jacket [117].

The SNR of the recovered 3-D dataset as a function of the acceleration is plotted in Fig. 5.5. We observe that the best SNR is obtained when both the non-convex penalties are used, which is around 1.5 dB better than nuclear norm alone and 6 dB superior than TV alone. It is also seen that the combined non-convex penalty (solid black curve) gives reconstructions that are approximately 1-2 dB better than its convex counterpart (dotted red curve), especially at higher accelerations. These experiments demonstrate the utility of the combination of non-convex penalties in challenging practical applications. We show the slice corresponding to the peak of the perfusion contrast, recovered using TV, nuclear norm and the combined non-convex penalties, in Fig. 5.6.(c)-(e). The corresponding error images are shown in Fig. 5.6.(f)-(h), respectively. Here, we consider $A = 7$, which corresponds to 20 k-space lines/frame. We plot the average intensity variations of the recovered images from 5 pixels in the tumor region (green dot) and 5 pixels of the healthy tissue (red dot) in Fig. 5.6.(i)-(k), respectively. Note that the curve from the tumor region has a larger dip and a larger width compared to that of the healthy tissue. This is due to the higher microvessel density and the increased tortuosity of the vessels in the tumor regions. We observe that the combination of non-convex penalties gives good fit to the measured data. The seven fold acceleration without significant degradation in image quality is quite remarkable, especially since we are only assuming a single channel acquisition; we expect to further improve the signal quality and/or acceleration using 12 or 32 channel head arrays that are now available.

5.4.3 Discussion on Convergence Rate

We now study the effect of the parameters β_1 and β_2 on the convergence of the algorithm and the accuracy of the solution. We consider the recovery of the MIT logo from $M = 1000$ measurements using the combined non-convex penalties ($p_1 = p_2 = 0.5$). We plot the evolution of the original cost function in (5.8) and

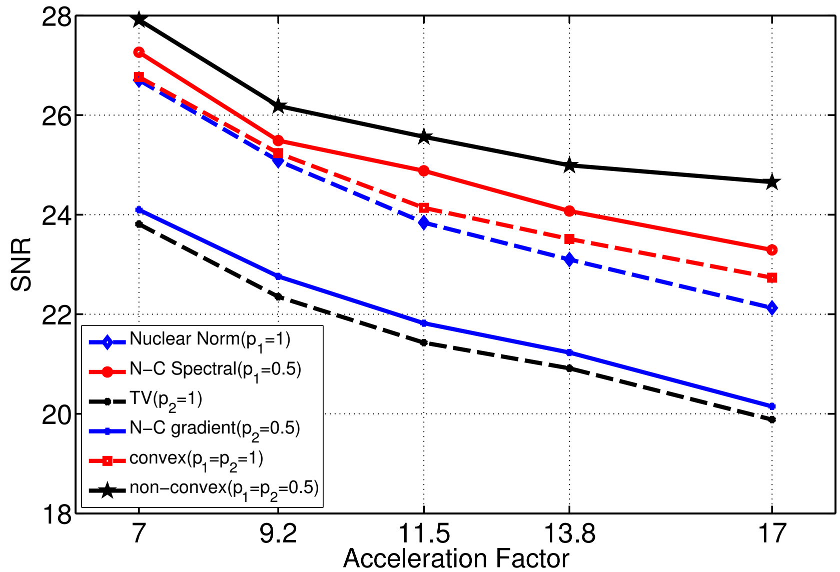


Figure 5.5: SNR of the dynamic MRI dataset as a function to acceleration factor. We reconstruct the dataset from its Fourier samples using six methods: Nuclear Norm ($p_1 = 1$); non-convex spectral penalty ($p_1 = 0.5$); TV ($p_2 = 1$); non-convex gradient penalty ($p_2 = 0.5$), combined convex penalties ($p_1 = p_2 = 1$) and combined non-convex penalties ($p_1 = p_2 = 0.5$). Note that the combined non-convex penalty provides significant gains in SNR at almost all acceleration factors.

SNR with respect to the number of iterations In Fig. 5.7.(a) and (b), respectively. It is observed that lower values of β_1 and β_2 result in fast convergence, but yield solutions with higher cost and lower SNR. This is expected since (5.16) is a poor approximation to (5.8). In contrast, higher values of β_1 and β_2 approximate the original cost function well, but result in slow convergence. We observe that the proposed continuation scheme, where β_1 and β_2 are initialized with small values and are gradually increased, offers the best compromise. In this specific example, the continuation scheme converged in 599 iterations. In contrast, the schemes with fixed values of β_1 and β_2 require far more number of iterations. The images recovered after 500 iterations using different parameter choices are shown in Fig.

5.8.

5.5 Contributions

Yue Hu, Sajan Lingala and Mathews Jacob. High Resolution Structural Free-breathing Cardiac MRI using k-t SLR. In *Proc. Int. Soc. for Magnetic Resonance Imaging (ISMRM)*, Montreal, Canada, 2011.

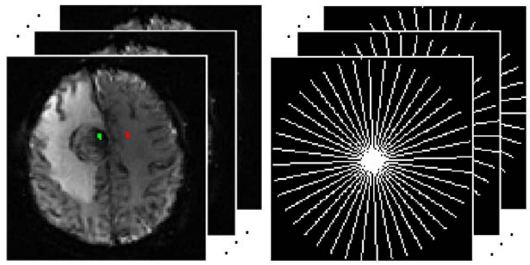
Yue Hu, Sajan Goud and Mathews Jacob. A Fast Majorize-Minimize Algorithm for the Recovery of Sparse and Low Rank Matrices. *IEEE Transaction on Image Processing*. Vol. 21, No.2, pp 742-753, Feb 2012.

Sajan Goud, Yue Hu and Mathews Jacob. Accelerated dynamic MRI exploiting sparsity and low-rank structure. *IEEE Transactions on Medical Imaging*. Vol. 30, Issue 5, Pages 1042-1054, May 2011.

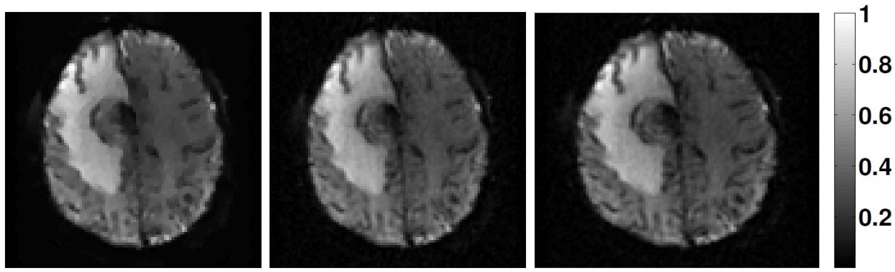
Sajan Goud, Yue Hu, Edward DiBella, Mathews Jacob. Accelerated first pass cardiac perfusion MRI using improved k-t SLR. IEEE *International Symposium on Biomedical Imaging: From Nano to Macro*. Pages: 1280-1283, Chicago, USA, 2011.

Sajan Goud, Yue Hu, Mathews Jacob. Real-time cardiac MRI using low-rank and sparsity penalties. IEEE *International Symposium on Biomedical Imaging: From Nano to Macro*. Pages: 988-991, Rotterdam, 2010.

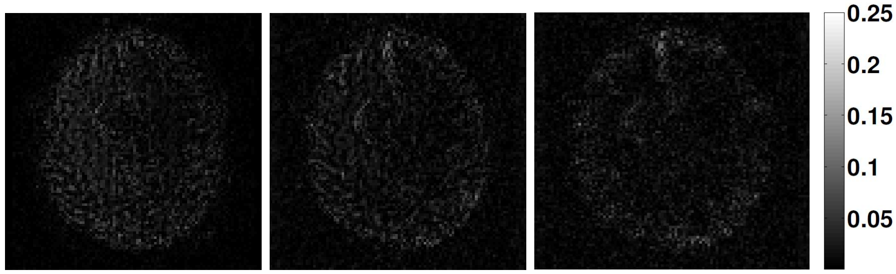
Sajan Lingala, Yue Hu, Mathews Jacob. Highly Accelerated Myocardial Perfuison MRI using k-t SLR with Parallel Imaging. In *Proc. Int. Soc. for Magnetic Resonance Imaging (ISMRM)*, Montreal, Canada, 2011.



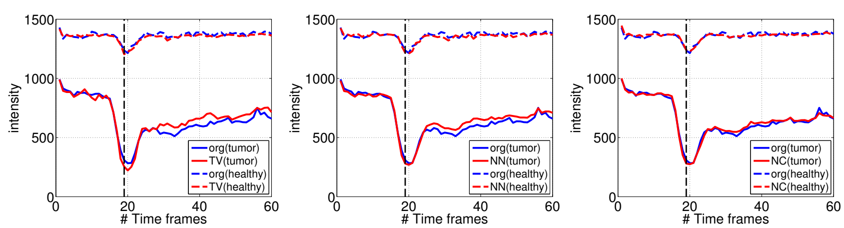
(a) actual (b) radial trajectory



(c) TV (d) Nuclear Norm (e) Combination

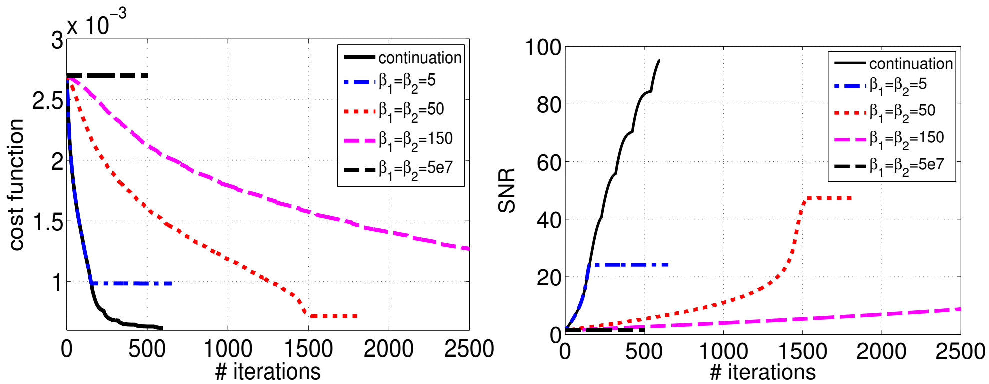


(f) TV (g) Nuclear Norm (h) Combination



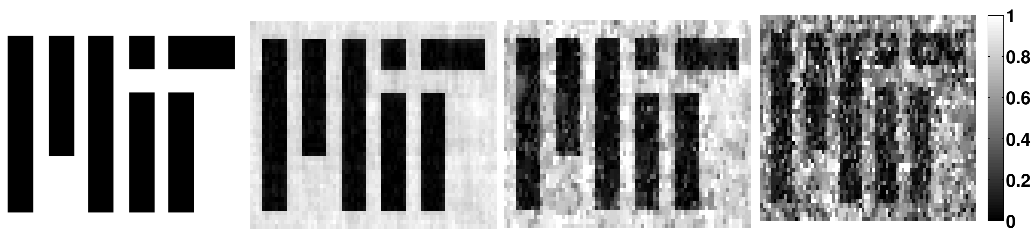
(i) TV (j) Nuclear Norm (k) Combination

Figure 5.6: Reconstruction of dynamic MRI data from undersampled Fourier samples. We consider 20 radial lines/image, which corresponds to an acceleration factor of $A = 7$. The images corresponding to the peak of the bolus (dotted line in the bottom row), which are recovered using TV ($p_2 = 1$), nuclear norm ($p_1 = 1$), and combination of non-convex penalties ($p_1 = p_2 = 0.5$) are shown in (c)-(e). The corresponding error images are shown in (f)-(h), respectively. (e), (h) and (k) correspond to the reconstructions using the proposed combination of non-convex penalties. The SNRs of the reconstructions using TV, Nuclear Norm, and combination of non-convex penalties are 23.81dB, 26.77dB, and 27.92dB, respectively.



(a) cost function to iterations (b) SNR to iterations

Figure 5.7: Utility of the continuation scheme: We plot the cost and SNR as a function of the number of iterations. We observe that lower values of the parameters ($\beta_1 = \beta_2 = 5$) result in a very fast convergence, but yield a solution with higher cost and lower SNR. Higher values of the parameters improve the accuracy at the expense of the number of iterations. Note that the continuation strategy, where the parameters are initialized with $\beta_1 = \beta_2 = 5$ and increased by a factor of 5 within the outer loop, results in fast convergence and solutions with good SNR. We terminate the algorithm when the cost does not change, where the convergence is achieved with $\beta_1 = \beta_2 = 5e7$. We observe that the algorithm fails to converge if it is initialized with these parameters.



(a) continuation (b) $\beta_1, \beta_2 = 5$ (c) $\beta_1, \beta_2 = 50$ (d) $\beta_1, \beta_2 = 150$

Figure 5.8: Utility of continuation schemes in matrix recovery. We reconstruct the MIT logo with continuation and different fixed values of β. We show the reconstructions using (a) continuation scheme (initialzed with $\beta_1 = \beta_2 = 5$ and gradually increased by a factor of 5) SNR=100.64; (b) $\beta_1 = \beta_2 = 5$, SNR=22.41; (c) $\beta_1 = \beta_2 = 50$, SNR=12.14; (d) $\beta_1 = \beta_2 = 150$, SNR=5.98 after 500 iterations. Note that the continuation scheme provides almost perfect recovery, while the other methods result in artifacts.

6 Conclusion

MRI is an important imaging modality. Compared to some other imaging methods, one of the advantages of MRI is that it uses non-ionizing radiation, and thus it poses minimal risk on human body. By changing the pulse sequences and the protocols of MR scanning, images with different contrasts can be obtained. However, the main challenge of MRI is the acquisition speed, which limits the clinical application of MRI. One way to improve the speed of MR acquisition is to collect less measurement samples in k-space. According to the recently developed mathematical theory of compressed sensing, images that are sparse under a certain transform can be accurately reconstructed using a subset of the k-space samples instead of the full dataset.

This thesis proposes novel approaches in the compressed sensing framework to recover 2D and multi-dimensional MR images in order to accelerate MRI while preserving the image quality. We developed novel regularized image reconstruction algorithms to utilize the sparsity and the low rank properties of the dataset. The main contributions of this work include:

1. The development of two families of higher degree total variation regular-

ization penalties, which are essentially non-quadratic norms of directional image derivatives. We observe that the anisotropic HDTV penalties, which rely on separable L_1 norms, provide better preservation of elongated image features and better SNR than isotropic penalties that use non-separable L_1-L_2 mixed norms. We exploited the steerability of directional derivatives to derive efficient majorize-minimize algorithms to solve the resulting optimization problems. Comparisons of the proposed regularization functionals with classical TV penalty, current second degree functionals, and sparse wavelet schemes in a range of practical applications demonstrated the significant improvement in performance.

2. The development of a three-dimensional higher degree total variation (3D-HDTV) penalty. We introduce a fast majorize-minimize algorithm that can be solved efficiently using an alternating minimization method, which improves the convergence speed compared to the previously used scheme. We implement the 3D-HDTV method using the proposed algorithm on compressed sensing, image denoising, and deblurring. The results demonstrate the improvement in SNR and image qualities compared with standard 3D-TV and some other current methods.

3. The development of a novel majorize-minimize algorithm to recover sparse and low-rank matrices from its noisy and undersampled measurements. We majorize the non-convex spectral and sparsity penalties in the cost function using quadratic matrix functions, resulting in an iterative three-step alternating minimization scheme. Since each of the steps in the algorithm has computationally efficient implementations, the algorithm provides fast convergence. We verified the utility of the combination of non-convex spectral

and sparsity penalties to significantly reduce the number of measurements required for perfect recovery in dynamic MRI data, compared to current matrix recovery schemes.

In current research for MRI, exploiting the novel image regularizations has become significant for accelerating and improving the image quality of MRI. In this thesis, we focus on the regularizations in two aspects: a) sparsity regularization, specifically, total variation based penalties, which enforce the sparse image gradients, and b) low-rank regularization, which enforces the low rank property of a matrix/image. In this work, we have exploited the combined penalties of TV and low rank, which obtained good performance. We have also proposed higher degree TV to improve the standard TV. The future directions of this work include using HDTV and low rank regularizations in MR data reconstruction to further improve the performance of the algorithms. However, there are a number of other prior information that can be used as the regularization in order to improve MRI, which we still need to further explore.

The famous American major league baseball catcher Yogi Berra said: “It is difficult to make predictions, especially about the future.” There could be one day, when an MR scanning can be finished in seconds, and the data reconstruction can be obtained in real-time. Though it is a long way to pursue this objective, as Peter Drucker put it, “The best way to predict the future is to create it.” I would like to continue contributing in creating the bright future of MRI research.

Bibliography

[1] Z. P. Liang and P. C. Lauterbur. *Principles of Magnetic Resonance Imaging: a signal processing perspective*. Wiley-IEEE Press, 1 edition, 1999.

[2] David J. Werring, Duncan W. Frazer, Lucy J. Coward, N. A. Losseff, H. Watt, L. Cipolotti, M. M. Brown, and H. Rolf Jäger. Cognitive dysfunction in patients with cerebral microbleeds on T2*-weighted gradient-echo MRI. *Brain*, 127(10):2265–2275, July 2004.

[3] S. Cruz-Flores. Ischemic stroke in emergency medicine. *Medscape Reference*, 2011.

[4] T. W. Redpath. Signal-to-noise ratio in MRI. *The British Journal of Radiology*, 71:704–707, 1998.

[5] A Hoffmann, S. C. Faber, K. J. Werhahn, L. Jager, and M. Reiser. First recordings of peripheral nerve activation caused by fast magnetic field gradients. *Magnetic Resonance in Medicine*, 43:534–539, 2000.

[6] C. L. G. Ham, J. M. L. Engels, G. T. can de Wiel, and A. Machielsen. Peripheral nerve stimulation during MRI: effects of high gradient amplitudes and switching rates. *Journal of Magnetic Resonance Imaging*, 7:933–937, 1997.

[7] W. A. H. Rushton. The effect upon the threshold for nervous excitation of the length of the nerve exposed and the angle between current and nerve. *Journal of Physiol.*, 63(4):357–377, Sep 1927.

[8] J. Tsao, P. Boesiger, and K. P. Pruessmann. k-t blast and k-t sense: Dynamic mri with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine*, 50(5):1031–1042, Oct 2003.

[9] M. Lustig, D. Donoho, and J. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, Dec 2007.

[10] E. Cand'es and D. Donoho. *Curves and Surface Fitting*. Curvelets - a suprisingly effective nonadaptive representation for objects with edges. Vanderbilt University Press, 2000.

[11] M. N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12):2091–2106, Dec 2005.

[12] M. N. Do and M. Vetterli. The finite ridgelet transform for image representation. *IEEE Transactions on Image Processing*, 12:16–28, 2003.

[13] S. Osher and S. Esedoglu. Decomposition of images by the anisotropic rudin-osher-fatemi model. *Comm. Pure Appl. Math*, 57(12):1609–1626, 2004.

[14] L Rudin, S Osher, and E Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, Jan 1992.

[15] Tony Chan, S Esedoglu, F Park, and A Yip. Recent developments in total variation image restoration. *Mathematical Models of Computer Vision*, 2004.

[16] M. Guerquin-Kern, D. Van De Ville, C. Vonesch, J.C. Baritaux, KP Pruessmann, and M. Unser. Wavelet-regularized reconstruction for rapid mri. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 193–196. IEEE, 2009.

[17] M. Figueiredo, J. Bioucas-Dias, J. Oliveira, and R. Nowak. On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavalet denoising. In *IEEE International Conference on Image Processing*, page 2633, 2006.

[18] D. G. Nishimura. Principles of magnetic resonance imaging. Master's thesis, Stanford University, Feb 2010.

[19] S. A. Huettel, A. W. Song, and G. McCarthy. *Functiona Magnetic Resonance Imaing*. Sinauer Associates, Inc, Sunderland, Massachusetts USA, 2004.

[20] R. Rzedzian, B. Chapman, P. Mansfield, D. Guifoyle, B. Chapman, R. E. Coupland, A. Chrispin, and P. Small. Real-time nuclear magnetic resonance: clinical imaging in paediatrics. *Lancet*, 12:1281–1282, 1983.

[21] C. H. Meyer, B. S. Hu, D. G. Nishimura, and A. Macovski. Fast spiral coronary artery imaging. *Magnetic Resonance in Medicine*, 28(2):202–213, 1992.

[22] J J van Vaals, M E Brummer, and J A den Boer. ”keyhole” method for accelerating imaging of contrast agent uptake. *J Magn Reson Imaging*, 3(4):671–675, Jul-Aug 1993.

[23] J. A. dArcy, D. J. Collins, I. J. Rowland, A. R. Padhani, and M. O. Leach. Applications of sliding window reconstruction with cartesian sampling for dynamic contrast enhanced mri. *NMR in Biomedicine*, 15:174–183, 2002.

[24] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[25] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603, Oct 1997.

[26] P. M. Jakob, M. A. Griswold, and R. R. Edelman et.al. AUTO-SMASH, a self-calibrating technique for SMASH imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 7(1):42–54, 1998.

[27] R. M. Heidemann, M. A. Griswold, A. Haase, and P. M. Jakob. VD-AUTO-SMASH imaging. *Magnetic Resonance in Medicine*, 12:1066–1074, 2001.

[28] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Gellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47:1202–1210, 2002.

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

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

[31] E. J. Candes and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies. *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[32] E. Cand‘es and T. Tao. Decoding by linear programming. *IEEE Transaction on Information Theory*, 8(6):737–763, 2005.

[33] J. P. Haldar, D. Hernando, and Z. P. Liang. Compressed-sensing MRI with random encoding. *IEEE Transaction on Medical Imaging*, 30(4):893–903, Apr 2011.

[34] M. N. Do. *Directional Multiresolution Image Representations*. PhD thesis, Swiss Federal Institute of Technology Lausanne, 2001.

[35] Jianwei Ma. Improved iterative curvelet thresholding for compressed sensing and measurement. *IEEE Transactions on Instrumentation and Measurement*, 60(1):126–136, Jan 2001.

[36] Stephane G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, Dec 1993.

[37] R. Gribonval and P. Vandergheynst. On the exponential convergence of matching pursuits in quasi-incoherent dictionaries. *IEEE Transactions on Information Theory*, 52(1):255–261, Jan 2006.

[38] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, 1:40–44, Nov 1993.

[39] Deanna Needell and Roman Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics*, 9(3):317–334, 2007.

[40] D. Needell and J. A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *ACM Technical Report*, 1, July 2008.

[41] H Zhang, G Li, and H Meng. An efficient greeddy algorithm for sparse recovery in noisy environment. *preprint*, 2009.

[42] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Mar 2004.

[43] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions via nuclear norm minimization. *Preprint*, 06 2007.

[44] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 54(1):30–37, 2004.

[45] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57(11):1413–1457, 2004.

[46] A. Beck and M. Teboulle. A fast iterative shrinkage thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

[47] D. Tschumperle and R. Deriche. Vector-valued image regularization with pdes: A common framework for different applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(4):506–517, 2005.

[48] J. Ma and L. Dimet. Deblurring from highly incomplete measurements for remote sensing. *IEEE Trans. Geoscience and Remote Sensing*, 47(3):792–802, 2009.

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

[50] C Vogel and M Oman. Fast, robust total variation-based reconstruction of noisy, blurredimages. *IEEE Transactions on Image Processing*, Jan 1998.

[51] G Steidl, S Didas, and J Neumann. Relations between higher order tv regularization and support vector regression. *Scale-Space and PDE Methods in Computer Vision*, Jan 2005.

[52] G Steidl, S Didas, and J Neumann. Splines in higher order TV regularization. *International Journal of Computer Vision*, Jan 2006.

[53] M. Unser and T. Blu. Wavelet theory demystified. *Signal Processing, IEEE Transactions on*, 51(2):470–483, 2003.

[54] S Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, Jan 1999.

[55] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

[56] F. Knoll, K. Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (tgv) for mri. *Magnetic Resonance in Medicine*, 2010.

[57] Tony Chan, A Marqina, and P.Mulet. Higher-order total variation-based image restoration. *SIAM J. Sci. Computing*, 22(2):503–516, Jul 2000.

[58] T. Chan, S. Esedoglu, F. Park, and A. Yip. Recent developments in total variation image restoration. *Mathematical Models of Computer Vision*, 2005.

[59] Y You and M Kaveh. Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing*, 9(10):1723 – 1730, Oct 2000.

[60] J. Kybic, T.Blu, and M. Unser. Generalized sampling: a variational approach, part i. *IEEE Trans. Signal Processing*, 50:1965–1976, 2002.

[61] M. Fornasier and H. Rauhut. Recovery algorithms for vector valued data with joint sparsity constraints. *Arxiv preprint math/0608124*, 2006.

[62] M. Fornasier, R. Ramlau, and G. Teschke. The application of joint sparsity and total variation minimization algorithms to a real-life art restoration problem. *Advances in Computational Mathematics*, 31(1):157–184, 2009.

[63] E Simoncelli and W Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. *International Conference on Image Processing*, Jan 1995.

[64] WT Freeman and EH Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.

[65] Marius Lysaker, Arvid Lundervold, and Xue-Cheng Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, Dec 2003.

[66] E Candes, L Demanet, D Donoho, and L Ying. Fast discrete curvelet transforms. *SIAM Multiscale Model. Simul*, Jan 2006.

[67] T. Blu and F. Luisier. The sure-let approach to image denoising. *Image Processing, IEEE Transactions on*, 16(11):2778–2786, 2007.

[68] P. C. Hansen. The L-curve and its use in the numerical treatment of inverse problems, 2000.

[69] Mark A Lukas. Robust generalized cross-validation for choosing the regularization parameter. *Inverse Problems*, 22(5):1883–1902, October 2006.

[70] S. Ramani, T. Blu, and M. Unser. Monte-carlo sure: A black-box optimization of regularization parameters. *IEEE Transactions on Image Processing*, 17(9):1540–1554, 2008.

[71] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A SURE approach for digital signal/image deconvolution problems. *IEEE Transactions on Signal Processing*, 57(12):4616–4632, December 2009.

[72] Sathish Ramani and Jeffrey A. Fessler. Regularized parallel mri reconstruction using an alternating direction method of multipliers. *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011.

[73] P.J. Verveer, MJ Gemkow, and T.M. Jovin. A comparison of image restoration approaches applied to three-dimensional confocal and wide-field fluorescence microscopy. *Journal of microscopy*, 193(1):50–61, 1999.

[74] J.G. McNally, T. Karpova, J. Cooper, and J.A. Conchello. Three-dimensional imaging by deconvolution microscopy. *Methods*, 19(3):373–385, 1999.

[75] JL Starck, E. Pantin, and F. Murtagh. Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific*, 114(800):1051–1069, 2002.

[76] S. Cho and S. Lee. Fast motion deblurring. In *ACM SIGGRAPH Asia 2009 papers*, pages 1–8. ACM, 2009.

[77] M Figueiredo, J Dias, J Oliveira, and R Nowak. On total variation denoising: A new majorization-minimization algorithm and an experimental comparisonwith wavalet denoising. *IEEE International Conference on Image Processing, 2006*, pages 2633–2636, 2006.

[78] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sciences*, 1(3):248–272, 2008.

[79] Chunlin Wu and Xue-Cheng Tai. Augmented Lagragian method, dual methods and split Bregman iteration for ROF, vectorial TV, and higher order models. *SIAM J. Imaging Sciences*, 3(3):300–339, 2010.

[80] Y Hu and M Jacob. Higher degree total variation (hdtv) regularization for image recovery. *IEEE Transactions on Image Processing*, 21(5):2559–2571, May 2012.

[81] J Cai, E Candes, and Z Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim*, 20:1956–1982, 2010.

[82] Julie C Mitchell. Sampling rotation groups by successive orthogonal images. *SIAM Journal on Scientific Computing*, 30(1):525–547, 2008.

[83] Anna Yershova, Swati Jain, Steven M LaValle, and Julie C Mitchell. Generating uniform incremental grids on so (3) using the hopf fibration. *The International journal of robotics research*, 29(7):801–812, 2010.

[84] Krzysztof M Gorski, Eric Hivon, AJ Banday, Benjamin D Wandelt, Frode K Hansen, Mstvos Reinecke, and Matthia Bartelmann. Healpix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2):759, 2005.

[85] http://physionet.org/physiobank/database/images/.

[86] Z.P. Liang. Spatiotemporal imaging with partially separable functions. In *Proceedings of the ISBI*, pages 181–182, 2007.

[87] C. Brinegar, Y.J.L. Wu, L.M. Foley, T.K. Hitchens, Q. Ye, C. Ho, and Z.P. Liang. Real-time cardiac MRI without triggering, gating, or breath holding. In *IEEE EMBC*, pages 3381–3384, 2008.

[88] Sajan Goud, Yue Hu, and Mathews Jacob. Real-time cardiac MRI using low-rank and sparsity penalties. In *Proceedings of the ISBI*, 2010.

[89] H Pedersen, S Kozerke, S Ringgaard, K Nehrke, and W Y Kim. k-t pca: temporally constrained k-t blast reconstruction using principal component analysis. *Magn Reson Med*, 62(3):706–716, Sep 2009.

[90] J.P. Haldar and D. Hernando. Rank-constrained solutions to linear matrix equations using powerfactorization. *IEEE Signal Processing Letters*, 19:584–587, 2009.

[91] B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.

[92] E. J. Candes and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[93] J.F. Cai, E.J. Candes, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim*, 20:1956–1982, 2010.

[94] Shiqian Ma, Donald Goldfarb, and Lifeng Chen. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 05 2009.

[95] Kim Toh and Sangwoon Yun. An accelerated proximal gradient method for nuclear norm regularized least squares. *Preprint*, 2009.

[96] R. T Rockafellar and R.J-B Wets. *Variational Analysis*. Springer Verlag, 2004.

[97] R Chatrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. *Proceedings of the ISBI*, (262-265), 2009.

[98] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[99] H. Gao, J-F. Cai, Z. Shen, and H. Zhao. Robust principal component analysis based four-dimensional computed tomography. *Physics in Medicine and Biology*, 56:3181–3198, 2011.

[100] H Jung, J Park, J Yoo, and J C Ye. Radial k-t focuss for high-resolution cardiac cine MRI. *Magn Reson Med*, Oct 2009.

[101] M. Lustig, JM Santos, DL Donoho, and JM Pauly. kt SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity. In *Proceedings of the 13th Annual Meeting of ISMRM, Seattle*, page 2420. Citeseer, 2006.

[102] U. Gamper, P. Boesiger, and S. Kozerke. Compressed sensing in dynamic mri. *Magnetic Resonance in Medicine*, 59(2):365–373, 2008.

[103] A. S. Lewis. The convex analysis of unitarily invariant matrix functions. *Journal of Convex Analysis*, 2(1):173–183, 1995.

[104] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sciences*, 1(3):248–272, 2008.

[105] Legendre transformation.

[106] M.V. Afonso, J.M. Bioucas-Dias, and M.A.T. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Processing*, 19(9):2345–2356, 2010.

[107] M.V. Afonso, J.M. Bioucas-Dias, and M.A.T. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *Arxiv preprint arXiv:0910.4887*, 2009.

[108] M. Ehler and S. Geisel. Arbitrary shrinkage rules for approximation schemes with sparsity constraints. *Schloss Dagstuhl Seminar Proceedings - Structured Decompositions and Efficient Algorithms*, (08492), 2009.

[109] E. Esser. Applications of lagrangian-based alternating direction methods and connections to split bregman. *CAM report*, 9:31, 2009.

[110] T. Goldstein and S. Osher. The split bregman algorithm for l1 regularized problems. *UCLA CAM Report*, pages 08–29, 2008.

[111] S. Goud, Y. Hu, E. DiBella, and M. Jacob. Accelerated first pass cardiac perfusion mri using improved k-t slr. In *ISBI*, 2011.

[112] E.J. Candes and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[113] R A Jones, O Haraldseth, T B Müller, P A Rinck, and A N Oksendal. K-space substitution: a novel dynamic imaging technique. *Magn Reson Med*, 29(6):830–834, Jun 1993.

[114] B Madore. Using UNFOLD to remove artifacts in parallel imaging and in partial-Fourier imaging. *Magn Reson Med*, 48(3):493–501, Sep 2002.

[115] M Han, B L Daniel, and B A Hargreaves. Accelerated bilateral dynamic contrast-enhanced 3D spiral breast MRI using tsense. *J Magn Reson Imaging*, 28(6):1425–1434, Dec 2008.

[116] H K Song and L Dougherty. Dynamic MRI with projection reconstruction and KWIC processing for simultaneous high spatial and temporal resolution. *Magn Reson Med*, 52(4):815–824, Oct 2004.

[117] Jacket: accelerating matlab using gpu.

Appendix A

We derive the conjugate of the function μ_g, specified by (5.24) in this section. Specifically, μ_g^* is defined as

$$\mu_g^*(w) = \max_x (wx - \mu_g(x)) = \max\left(\max_{x;x<\frac{1}{\beta}} wx, \max_{x;x>\frac{1}{\beta}} \left(wx - \frac{1}{2}(x - 1/\beta)^2\right)\right)$$

The maximum value of the first term $(wx; x < 1/\beta)$ is given by w/β. The second term inside the bracket is true if $x = w + 1/\beta$, when the value of the function is given by $w^2/2 + w/\beta$. Since the second term is always greater than the first, we obtain

$$\mu_g^*(w) = \frac{w^2}{2} + \frac{w}{\beta} \tag{1}$$

Since $\mu_\eta(w) = \beta\left(\mu_g(w) - w^2/2\right)$, we have $\mu_\eta(w) = w$.