Title
Restoration of Images in the Presence of Rician Noise and in the Presence of Atmospheric Turbulence

Author
Tong, Melissa

Publication Date
2012

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Restoration of Images in the Presence of Rician Noise and in the Presence of Atmospheric Turbulence

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Melissa Yin Tong

2012

Abstract of the Dissertation

Restoration of Images in the Presence of Rician Noise and in the Presence of Atmospheric Turbulence

by

Melissa Yin Tong

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2012

Professor Luminita Aura Vese, Chair

This thesis is divided into two parts. In the first part, we will discuss the problem of restoring magnetic resonance (MR) images corrupted by blur and Rician noise. We discuss the formation of MR signals and how Rician noise is introduced into these images as a result of the MR acquisition process. Information about the Rician probability distribution and motivation for our proposed variational restoration model is then given. We show the existence of a minimizer and a comparison result. We also perform numerical experiments and comparisons using L^2 and H^1 gradient descent schemes to show the validity of our proposed model. This leads to a related second model that denoises High Angular Resolution Diffusion Imaging (HARDI) data, which is a modality of MR data that is used in reconstructing fiber pathways in the brain. HARDI data is vectorial data of dimension equal to the number of diffusion directions. This data can be used as input to calculate fractional anisotropy (FA) or orientation distribution functions (ODFs) which in turn are used to track fibers in the brain. Having denoised data may lead to more accurate fiber extractions. We test our proposed HARDI denoising model on various data sets, and various metrics are used to gauge improvements after denoising. In the second part of this thesis, we study the problem of restoring images distorted by atmospheric turbulence. Geometric distortions and blur are the two main components of degradations due to atmospheric turbulence, and prior work has been done to address these components separately. We propose a joint variational deblurring and geometric distortion correction model and give preliminary results.

The dissertation of Melissa Yin Tong is approved.

Christopher R. Anderson

Ai-Chi Chien

Stanley J. Osher

Paul M. Thompson

Luminita Aura Vese, Committee Chair

University of California, Los Angeles

2012

To my friends and family.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to acknowledge my advisor Luminita Vese. Thank you for your support and guidance and for all the knowledge that you have imparted to me. Thank you for always being patient, kind and understanding. You have been a great mentor and friend to me.

I would like to acknowledge my collaborators who have given me very useful comments and advice on our joint projects. This research would not have been possible without their help. Many thanks go to Paul Thompson, Liang Zhan, Yunho Kim, Jerome Gilles, Aichi Chien, and Pascal Getreuer. I would like to mention that parts of Chapter 2 have been presented and published in "A Variational Model for the Restoration of MR Images Corrupted by Blur and Rician Noise", Proceedings of ISVC, Part I, LNCS 6938, pp. 686-698, 2011, and parts of Chapter 3 have been presented and published in "Variational Model for Denoising High Angular Resolution Diffusion Imaging", Proceedings of IEEE Int. Symp. Biomed. Imaging (ISBI), pp. 530-533, 2012.

To Paul Thompson, Jonathan Feng, Robert Brown, and Yunho Kim, thank you for the letters of recommendation and support. And to the UCLA mathematics departmental staff, thank you for your helpfulness on many occasions.

To my committee members, Chris Anderson, Aichi Chien, Stanley Osher, and Paul Thompson, thank you for being on my committee– I appreciate greatly the comments, suggestions, and guidance that you have given me.

Thank you to my friends and family who have given me so much support throughout this journey. To my best friend, thank you for everything.

Finally, I would like to acknowledge that this work was supported in part by NSF Expeditions in Computing Award CCF-0926127 and NSF-DMS Award 0714945.

Vita

2007	B.S. (Mathematics and Physics), UCI, Irvine, California.
2009	M.A. (Applied Mathematics), UCLA, Los Angeles, California.
2007–2011	Teaching Assistant, Mathematics Department, UCLA.
2010–2012	Research Assistant, Mathematics Department, UCLA.

Publications

P. Getreuer, M. Tong, and L. Vese, "A Variational Model for the Restoration of MR Images Corrupted by Blur and Rician Noise," Proceedings of ISVC, Part I, LNCS 6938, pp. 686-698, 2011.

M. Tong, Y. Kim, L. Zhan, G. Sapiro, C. Lenglet, B. Mueller, P.M. Thompson, L.A. Vese, "Variational Model for Denoising High Angular Resolution Diffusion Imaging," Proceedings of IEEE Int. Symp. Biomed. Imaging (ISBI), pp. 530-533, 2012.

Part I

Restoration of Magnetic Resonance Images Corrupted by Blur and Rician Noise

CHAPTER 1

Introduction and Background

1.1 Introduction and Motivation

Images arise in many applications in society. From digital art to video surveillance, images are used in a wide range of applications. With the invention of new image acquisition devices, many different types of images corresponding to different types of information may be collected. In addition, advances in computing allow for large amounts of data to be processed. For example, in medical imaging, image acquisition techniques include computerized tomography (CT), X-rays, and magnetic resonance (MR) imaging. These techniques have been invaluable in advancing our understanding of different phenomena in many medical-related fields.

The advent of this wealth of information in the form of images is often coupled with a need to process and analyze this data; this falls under the field of image processing. Image processing techniques include image segmentation, compression, super-resolution, and restoration such as denoising and deblurring.

One type of image acquired in medical imaging is magnetic resonance (MR). MR images are used to visualize the internal features of the body, for example the brain. These images may help detect abnormalities in a patient such as a tumor in the brain or aid doctors in evaluating the success of a brain surgery. Unfortunately, MR images are often corrupted with noise and possibly other types of degradation such as blurring effects. One goal of medical image processing is to design algorithms that decrease this noise, making certain features in MR images easier to see.

The type of noise present in MR images is known to follow a Rician distribution [GP95, Hen85, Ric44, Lat83, BFW06]. This type of noise may severely alter MR images, leading to

incorrect interpretation of important structural information. For example, High Angular Resolution Diffusion Imaging (HARDI) [TWB99, TRW02] is a modality of MR imaging used in reconstructing fibers in the brain, and HARDI data is often contaminated with noise. More specifically, HARDI data is a collection of diffusion weighted MR images (DW-MRI), meaning that for each diffusion direction considered, we have a corresponding MR image. The angular resolution of fiber pathways increases as the number of diffusion directions imaged increases, but the introduction of noise during the data collection process will likely change important characteristics measured from the HARDI data such as fractional anisotropy (FA) [PB96, And01, SLN00] of the fiber pathways. Therefore, it may be beneficial to design algorithms to reduce this noise in hopes of obtaining more accurate information from the data.

During the MR data collection process, a complex valued signal is collected and its magnitude forms the MR signal. Since the real and imaginary parts of this raw complex signal is corrupted by Gaussian distributed noise, the resulting magnitude signal will have Rician distributed noise [GP95, Hen85, Ric44, Lat83, BFW06]. In the next section, we take a closer look at Rician distributed noise and how it arises in MR data.

1.2 MR Signals and Rician Distributed Noise

An underlying clean signal ν is said to be corrupted with Rician distributed noise if the resulting noisy signal r follows a Rician probability distribution, which is given by

$$\mathbf{P}(r;\nu,\sigma) = \frac{r}{\sigma^2}\exp\left(\frac{-(r^2+\nu^2)}{2\sigma^2}\right) I_0\left(\frac{r\nu}{\sigma^2}\right), \tag{1.1}$$

where $r,\nu,\sigma > 0$, I_0 is the zeroth-order modified Bessel function of the first kind, and σ is the standard deviation of the noise. In other words, if at a particular pixel or voxel, the underlying clean signal has value ν, then $Pr(r;\nu,\sigma)$ gives the probability of getting a noisy signal with value r. Rician probability density functions for given ν and $\sigma = 1$ are presented in Fig. 1.1.

In general, a random variable R is said to be *Rician* distributed with standard deviation σ about ν, denoted $R \sim \text{Rician}(\nu,\sigma)$, if it has probability distribution given by (1.1). If we take two independent normal random variables $X \sim N(\nu\cos\theta,\sigma^2)$ and $Y \sim N(\nu\sin\theta,\sigma^2)$ (for any

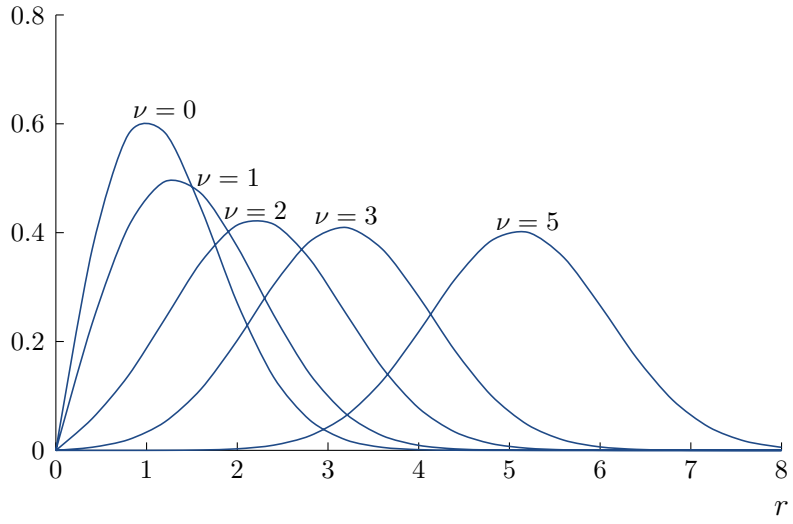


Figure 1.1: Rician probability density function for given ν and $\sigma = 1$.

$\theta \in \mathbb{R}$), then the resulting magnitude of the two random variables $R = \sqrt{X^2 + Y^2}$ has Rician distribution, $R \sim \text{Rician}(\nu, \sigma)$. An application in which Rician distribution arises in this way is the formation of MR signals.

As mentioned earlier, an MR signal is formed by taking the magnitude of the raw complex valued signal collected by an MR imaging machine. If the collection process is perfect, then we start with the machine collected complex valued signal

$$z = z_{real} + iz_{imag} = \nu\cos(\theta) + i\nu\sin(\theta)$$

where ν is the magnitude of z and θ is the phase of z. The resulting MR signal is simply the magnitude ν. However, in real applications, the collection process corrupts z with noise. More specifically, the real and imaginary components of z are corrupted with Gaussian distributed noise with some standard deviation σ [GP95, Hen85, Ric44, Lat83, BFW06]:

$$z = z_{real} + iz_{imag} \to (z_{real} + G(\sigma)) + i(z_{imag} + G(\sigma)).$$

The corrupted MR signal is the magnitude of the corrupted complex valued signal:

$$\text{magnitude} = \sqrt{(z_{real} + G(\sigma))^2 + (z_{imag} + G(\sigma))^2}.$$

As a result, an MR signal is the magnitude of two independent normal random variables, with

distributions $N(\nu\cos\theta, \sigma^2)$ and $N(\nu\sin\theta, \sigma^2)$. It is easy to see from the discussion above that the resulting MR signal is corrupted by Rician distributed noise.

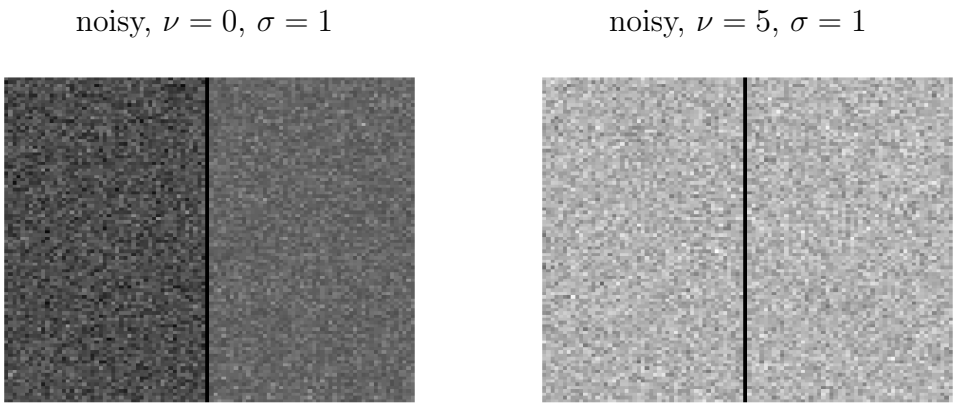


Figure 1.2: Comparison of Gaussian and Rician noise with $\sigma = 1$ on input values $\nu = 0$ (left square) and $\nu = 5$ (right square). Left of each square: additive Gaussian noise; right of each square: Rician noise. For $\nu = 0$, mean of Gaussian: -0.0041, mean of Rician: 1.2550. For $\nu = 5$, mean of Gaussian: 5.0155, mean of Rician: 5.1066.

In Fig. 1.2, we start with two squares of constant intensity $\nu = 0$ and $\nu = 5$. We add Gaussian noise to the left half of the squares and Rician noise to the right half of the squares, where the standard deviation of the noise $\sigma = 1$ in all cases. We note that for input values ν that are small compared to the standard deviation σ, we see a difference between Gaussian and Rician noise. In addition, the images corrupted with Rician noise have a larger mean than the images corrupted with Gaussian noise. We compare the probability distributions for these two cases in Fig. 1.3. When the standard deviation σ of the Rician probability distribution is small compared to ν (e.g. $\nu > 2\sigma$), the Rician probability distribution is approximately Gaussian. This case occurs when the signal to noise ratio (SNR) is high (low noise level). However, when σ is larger or on the order of the value of ν (e.g. $\nu \leq \sigma$), the Gaussian probability distribution is a bad approximation to the Rician one. This occurs when the SNR is low and the noise level is high. Therefore, a Gaussian distribution is a good approximation to the Rician distribution in areas corrupted with small amounts of noise but not in areas corrupted with large amounts

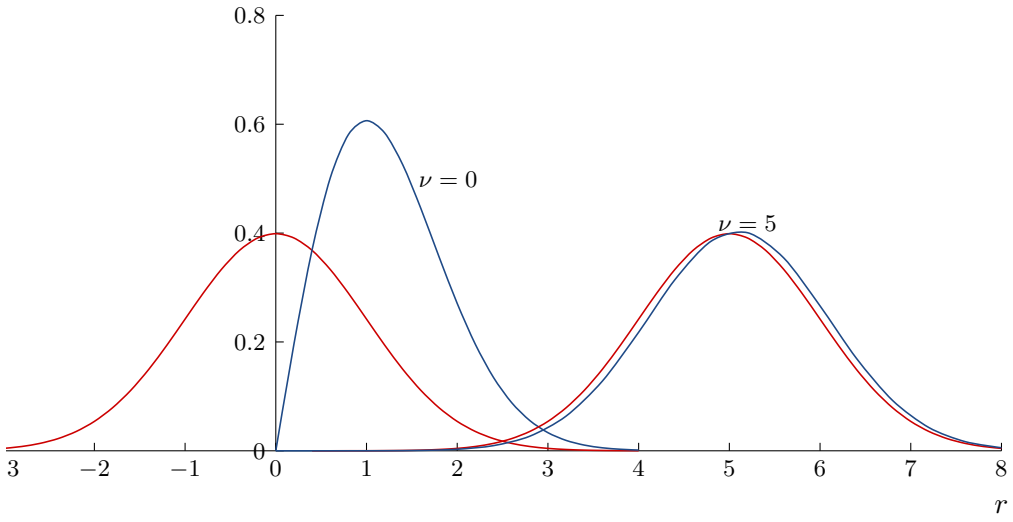


Figure 1.3: Comparison of Rician (blue) and Gaussian (red) probability distribution functions for given ν and standard deviation $\sigma = 1$.

of noise.

In some cases, the standard deviation of the noise σ is unknown. Since σ appears in our denoising algorithm, we must find an estimation for this parameter. As in [SDD98], we approximate σ by using the equation

$$E(X^2) = \nu^2 + 2\sigma^2 \tag{1.2}$$

for the second moment of a Rician distributed variable X. Here, $E(X^2)$ is the expected value of X^2 and ν is the true signal. For regions of a data set where we expect $\nu = 0$ (e.g. background), we can use

$$\sigma \approx \sqrt{E(X^2)/2} \tag{1.3}$$

where $E(X^2)$ is simply the average value of the square of the noisy signal over the region where the true signal is expected to be zero.

1.2.1 Bessel Functions

The Rician probability density (1.1) involves the modified Bessel function I_0, so developing a Rician noise model entails manipulating I_0 and its derivatives. The modified Bessel functions

are the solutions of the modified Bessel differential equation

$$t^2\frac{d^2y}{dt^2} + t\frac{dy}{dt} - (t^2 + n^2)y = 0. \tag{1.4}$$

where $t \in \mathbb{R}$, and n is a constant that is not necessarily an integer. This equation has two linearly independent solutions $I_n(t)$ and $K_n(t)$, which are respectively the modified Bessel functions of the first and second kind of order n.

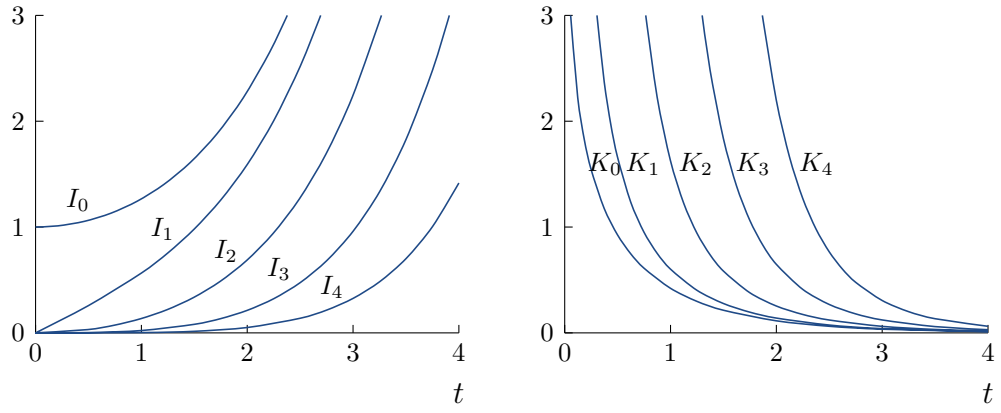


Figure 1.4: Modified Bessel functions shown for $t \geq 0$.

For $t \geq 0$, the functions $I_n(t)$ are exponentially increasing while $K_n(t)$ are exponentially decreasing, see Figure 1.4. For integer n, some properties (from [AS65, Boa83]) of $I_n(t)$ include

$$\begin{aligned}
I_n(t) &= i^{-n}J_n(it) \\
&= \frac{1}{\pi}\int_0^\pi \mathrm{e}^{t\cos\theta}\cos(n\theta)\,d\theta \\
&= (\tfrac{1}{2}t)^n \sum_{k=0}^{\infty} \frac{(\frac{1}{4}t^2)^k}{k!(n+k)!}, \\
\tfrac{d}{dt}I_0(t) &= I_1(t), \\
\tfrac{d}{dt}I_1(t) &= I_0(t) - \frac{1}{t}I_1(t),
\end{aligned}$$

where $J_n(t)$ denotes the nth order Bessel function of the first kind. Furthermore, a property that we will use is that the ratio $I_1(t)/I_0(t)$ is bounded by 1, see Fig. 1.5.

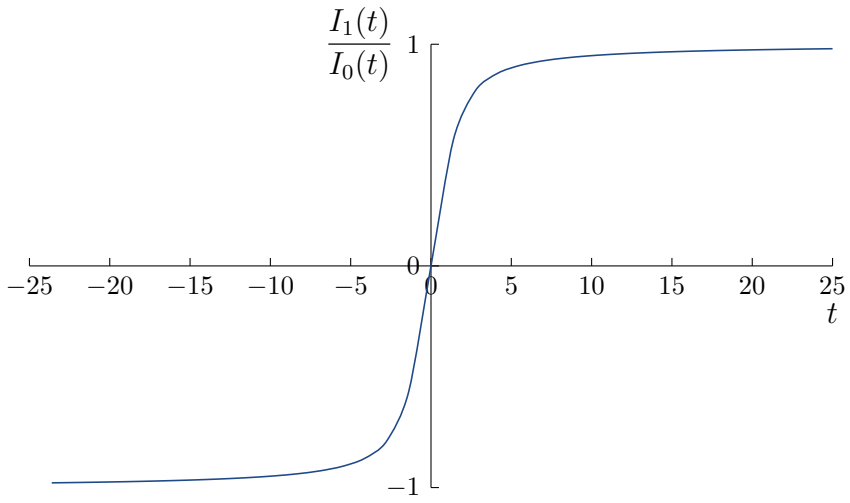


Figure 1.5: The function $I_1(t)/I_0(t)$.

1.3 Bayesian Model Incorporating Rician statistics

Bayesian statistical inference is a powerful technique used in image processing. Following the discussion in [CS05], conceptually, a set of data Q can be thought of as generated by a feature set F. The goal is to find the most likely feature variable from F to have generated a given observed sample of Q using both information about the feature set F and information about how data Q is generated from F.

For our application, consider F to be a set of underlying true images, and Q to be a set of observed data generated by F through a degradation process involving Gaussian blur and Rician noise. More specifically, let f denote an observed degraded image. A Bayesian statistical inference technique called *maximum a posteriori (MAP) estimation* may be used to design a denoising and deblurring method to recover the u from the set F that is most likely to have generated the degraded image f. As in [BFW06], we utilize the MAP estimation technique to denoise images corrupted by Rician distributed noise.

In general, with f the degraded image, and u the underlying true image, the MAP estimate of u is the u that is most likely to have generated the given f:

$$\hat{u} = \arg\max_{u} \mathbf{P}(u|f). \tag{1.5}$$

An application of Bayes' Theorem gives

$$\max_u \mathbf{P}(u|f) = \max_u \Big\{ \mathbf{P}(u)\mathbf{P}(f|u) \Big\} = \min_u \Big\{ -\log \mathbf{P}(u) - \log \mathbf{P}(f|u) \Big\}. \tag{1.6}$$

The first factor $\mathbf{P}(u)$ is called a *prior* on u which is a distribution that incorporates information on what u is likely to be, restricting the possible admissible images. In addition, it acts as a regularization term. The second factor $\mathbf{P}(f|u)$ is the data model, which describes the degradation process that generated f from u.

For example, consider the Rudin-Osher-Fatemi (ROF) restoration model [ROF92, RO94]

$$\min_{u \in BV(\Omega)} \Big\{ \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega (Ku - f)^2 dx \Big\}, \tag{1.7}$$

where $\Omega \subset \mathbb{R}^n$ is open and bounded. The observed image f is assumed to be generated by an unknown restored image u (belonging to the space of functions of bounded variation, $BV(\Omega)$) by the degradation process

$$f = Ku + n, \tag{1.8}$$

where K is a linear blurring operator and n is white Gaussian noise. ROF can be seen to be a MAP estimate using the prior

$$P(u) = \exp(-\alpha \int |Du|) \tag{1.9}$$

and data model

$$P(f(x)|u(x)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(f(x) - Ku(x))^2}{2\sigma^2} \Big). \tag{1.10}$$

Therefore, we have

$$\begin{aligned} -\log \mathbf{P}(f|u) &= -\int_\Omega \log \mathbf{P}\big(f(x)|u(x)\big)\, dx \\ &= -\int_\Omega \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(\frac{-\big(f(x) - Ku(x)\big)^2}{2\sigma^2} \Big) \right] dx \\ &= \frac{1}{2\sigma^2} \int_\Omega \big(f(x) - Ku(x)\big)^2\, dx + \tfrac{|\Omega|}{2} \log(2\pi\sigma^2), \end{aligned}$$

so that with $\lambda = 1/(\sigma^2\alpha)$ we recover (1.7),

$$\begin{aligned} \max_u \mathbf{P}(u|f) &= \min_u \Big\{ -\log \mathbf{P}(u) - \log \mathbf{P}(f|u) \Big\} \\ &= \min_u \Big\{ \alpha \int_\Omega |Du| + \frac{1}{2\sigma^2} \int (Ku - f)^2\, dx \Big\}. \end{aligned}$$

Through this connection between ROF and MAP estimates, ROF can be reformulated to use other degradation models. Specifically, with a degradation model consisting of blurring with blur kernel K and followed with the addition of Rician noise, we employ the Rician probability distribution (1.1) to arrive at data model

$$\mathbf{P}(f|u) = \frac{f(x)}{\sigma^2} \exp\bigg(- \frac{(Ku(x))^2 + f(x)^2)}{2\sigma^2} \bigg) I_0\bigg(\frac{f(x)Ku(x)}{\sigma^2} \bigg) \tag{1.11}$$

and thus

$$-\log \mathbf{P}(f|u) = \int_\Omega \Big[-\log \frac{f}{\sigma^2} - \log I_0\bigg(\frac{f(Ku)}{\sigma^2} \bigg) + \frac{f^2 + (Ku)^2}{2\sigma^2} \Big] dx. \tag{1.12}$$

Using the same prior on u (1.9) as in the ROF model and incorporating everything into (1.6) as before, we arrive at our proposed variational denoising deblurring model incorporating Rician statistics

$$\inf_{u \in BV(\Omega)} \Big\{ F(u) = \int_\Omega |Du| + \lambda \int_\Omega \Big[-\log \frac{f}{\sigma^2} - \log I_0(\frac{f(Ku)}{\sigma^2}) + \frac{f^2 + (Ku)^2}{2\sigma^2} \Big] dx \Big\}. \tag{1.13}$$

In the next section, we give a brief review of the space of functions of bounded variation, which is the space in which we look for our solution.

1.3.1 Review of the Space of Functions of Bounded Variation

Here, we give a brief review of some properties of the space of functions of bounded variation that are relevant to image processing. We refer the reader to [ABM06, AK06, CS05] for more details.

The space of functions of bounded variation is important in image modeling because of its admission of functions with jumps or edges, which are important features that often appear in images. The well-known work by Rudin-Osher-Fatemi [ROF92, RO94] provides an example of a denoising scheme using total variation regularization.

First, we give the definition of total variation and the space of functions of bounded variation. Then we give relevant properties of this space as well as canonical example functions. Lastly, we provide a proof of a property that will be used in the existence proof for our variational Rician denoising model.

Definition. Let $\Omega \in \mathbb{R}^N$ and $u \in L^1(\Omega)$. The total variation (TV) of u is defined as

$$\int_\Omega |Du| = \sup\left\{ \int_\Omega u \text{div}\phi \; dx; \phi = (\phi_1, \phi_2, ..., \phi_N) \in C_0^1(\Omega)^N, |\phi|_{L^\infty(\Omega)} \leq 1 \right\}, \tag{1.14}$$

where $\text{div}\phi = \sum_{i=1}^N \frac{\partial \phi_i}{\partial x_i}(x)$ is the divergence of ϕ, dx is the Lebesgue measure, and $C_0^1(\Omega)^N$ is the space of continuously differentiable functions with compact support in Ω and

$$|\phi|_{L^\infty(\Omega)} = \sup_x \sqrt{\sum_{i=1}^N \phi_i^2(x)}. \tag{1.15}$$

Definition. Let $BV(\Omega)$ denote the space of functions of bounded variation.

$$BV(\Omega) = \left\{ u \in L^1(\Omega); \int_\Omega |Du| < \infty \right\}. \tag{1.16}$$

In other words, a function u belongs to $BV(\Omega)$ if it belongs to $L^1(\Omega)$ and has bounded variation $\int_\Omega |Du| < \infty$.

Two canonical examples are the following. First, if u is differentiable (i.e. $u \in C^1(\Omega)$), then $\int_\Omega |Du| = \int_\Omega |\nabla u(x)| dx$. Second, if we take $A = B(0,1) \in \mathbb{R}^2$ to be the interior of the unit disc centered around 0, and $u = \chi_A$ to be the characteristic function of the subset A defined by

$$\chi_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \in \mathbb{R}^2 \backslash A, \end{cases} \tag{1.17}$$

then $\int_\Omega |Du| = \text{Perimeter}(B(0,1)) = 2\pi$. This corresponds to the jump in value from 0 to 1 at the boundary of the disc.

It can be shown that if $u \in BV(\Omega)$, then its distributional gradient Du is a vector-valued Radon measure. Furthermore, Du can be decomposed into an absolutely continuous part and singular part

$$Du = \nabla u dx + D_s u, \tag{1.18}$$

where dx is the N-dimensional Lebesgue measure, $\nabla u = \frac{d(Du)}{dx}(x) \in L^1(\Omega)$ is the absolutely continuous part, and $D_s u \perp dx$ where $D_s u$ is the singular part. The singular part can be decomposed further into a "jump" part $J_u = (u^+ - u^-) n_u H^{N-1}_{|S_u}$ and "Cantor" part C_u.

$$Du = \nabla u dx + (u^+ - u^-) n_u H^{N-1}_{|S_u} + C_u, \tag{1.19}$$

where H^d denotes the d-dimensional Hausdorff measure. Furthermore,

$$\begin{aligned} u^+(x) &= \inf\left\{t \in [-\infty,\infty];\ \lim_{r\to 0}\frac{dx(\{u>t\}\cap B(x,r))}{r^N}=0\right\} \\ u^-(x) &= \inf\left\{t \in [-\infty,\infty];\ \lim_{r\to 0}\frac{dx(\{u<t\}\cap B(x,r))}{r^N}=0\right\}, \end{aligned} \tag{1.20}$$

and the jump set S_u is given by

$$S_u = \left\{x \in \Omega;\ u^-(x) < u^+(x)\right\}, \tag{1.21}$$

with $n_u(x)$ denoting a normal vector to S_u.

J_u and C_u are mutually singular, which implies that the support of the Cantor part C_u has Hausdorff dimension strictly greater than $N-1$ and strictly less than N. This leads to the decomposition of the total variation

$$\int_\Omega |Du| = \int_\Omega |\nabla u|(x)dx + \int_{S_u} |u^+ - u^-| dH^{N-1} + \int_{\Omega - S_u} |C_u|. \tag{1.22}$$

Revisiting our two canonical examples, we see that with example 1, u has no discontinuities since $u \in C^1(\Omega)$, so the last two terms of the decomposition disappear and we again have $\int_\Omega |Du| = \int_\Omega |\nabla u(x)| dx$. With example 2, $u = \chi_{B(0,1)}$, $|\nabla u| = 0$ a.e. with respect to the Lebesgue measure so the first term of decomposition is 0. The only discontinuity is a jump discontinuity along the boundary of $B(0,1)$ so the last term is 0. With $|u^+ - u^-| = 1$, the second term is simply, again, the perimeter of $B(0,1)$.

The space $BV(\Omega)$ has norm

$$|u|_{BV(\Omega)} = |u|_{L^1(\Omega)} + \int_\Omega |Du| \tag{1.23}$$

but due to the lack of good compactness properties with this topology, the $BV - w*$ (BV weak* topology) is used where convergence is defined by

$$u_j \underset{BV-w*}{\rightharpoonup} u \tag{1.24}$$

if and only if

$$u_j \underset{L^1(\Omega)}{\longrightarrow} u \quad \text{and} \quad \int_\Omega \phi Du_j \to \int_\Omega \phi Du, \quad \forall \phi \in C_0(\Omega)^N. \tag{1.25}$$

A compactness property that follows is that any uniformly bounded sequence $u_j \in BV(\Omega)$ is relatively compact in $L^p(\Omega)$ for $1 \leq p < \frac{N}{N-1}$ for $N \geq 1$. Furthermore, there is a subsequence u_{j_k} and $u \in BV(\Omega)$ that converges in the BV-w* topology to u. In addition, $BV(\Omega)$ is continuously embedded in $L^p(\Omega)$ with $p = +\infty$ if $N = 1$ and $p = \frac{N}{N-1}$ otherwise.

In addition, we have the following two useful properties.

Lower semi-continuity of TV: For a sequence $u_j \in BV(\Omega)$ and $u_j \to u$ strongly in $L^1(\Omega)$, then

$$\int_\Omega |Du| \leq \liminf_{j\to\infty} \int_\Omega |Du_j|. \tag{1.26}$$

Coarea Formula With $u \in BV(\Omega)$, then for a.e. $t \in \mathbb{R}$, the level set $E_t = \{x \in \Omega \,:\, u(x) > t\}$ of u is a set of finite perimeter in Ω and

$$\begin{aligned} Du(\Omega) &= \int_{-\infty}^{\infty} D\chi_{E_t} dt \\ |Du|(\Omega) &= \int_{-\infty}^{\infty} \int_\Omega |D\chi_{E_t}| dt. \end{aligned} \tag{1.27}$$

It will be useful to define the following extension of $BV(\Omega)$ to $L^{\frac{N}{N-1}}(\Omega)$,

$$J(u) = \begin{cases} \int_\Omega |Du| & \text{if } u \in BV(\Omega) \\ +\infty & \text{if } u \in L^{\frac{N}{N-1}}(\Omega) \backslash BV(\Omega). \end{cases} \tag{1.28}$$

The following property that involves this extension is used in the proof of the existence of a minimizer of the Rician denoising model [Cha04, Giu94]

$$J(\max\{u, v\}) + J(\min\{u, v\}) \leq J(u) + J(v) \tag{1.29}$$

where J is defined in (1.28).

Furthermore, for $u \in BV(\Omega)$, we show

$$\int_\Omega |D|u|| \leq \int_\Omega |Du|. \tag{1.30}$$

Indeed, let $u \in BV(\Omega)$, then

$$\begin{aligned} J(|u|) &= J(\max\{u(x), 0\} - \min\{u(x), 0\}) \\ &\leq J(\max\{u(x), 0\}) + J(\min\{u(x), 0\}) \leq J(u) + J(0) \\ &= J(u), \end{aligned} \tag{1.31}$$

where the second line is an application of (1.29).

CHAPTER 2

TV-Regularized Restoration with Rician Noise Model

2.1 Prior Works

The main approaches for denoising MR images fall into one of two categories. In representation based approaches, the goal is to represent the noise free signal as a linear combination of a small number of elements of a fixed basis. This approach uses information about the type of signal to be reconstructed (e.g. MR signals of the brain) to recover a noise free signal from a noisy signal. A second approach is to model explicitly the type of noise that is present in MR images. In this case, information about the noise statistics is used. We will work within this latter framework.

Prior works in MR image denoising include the following. In [WZ06], Wang and Zhou use a TV and wavelet based regularization and assume that the noise found in MR images is Gaussian. In addition, they propose an automatic stopping criterion for the time evolution equation associated with their proposed variational denoising model. This stopping criterion helps avoid an over-smoothed solution. In the works by Descoteaux et al. [DWP] and Wiest-Daessle et al. [WPC], a non-local means filter adapted for Rician denoising is proposed and experiments are performed on diffusion tensor MR images (DT-MRI). The work most related to ours is by Basu et al. [BFW06]. In their work, a maximum a posteriori (MAP) estimate incorporating Rician statistics and a prior based on the Perona-Malik energy functional is used, and experiments are performed on DT-MRIs. For our model, instead of a prior based on the Perona-Malik energy functional, we use a prior based on total variation.

In this chapter, we give first our proposed variational restoration model and the existence of a minimizer and comparison result. Then, we give the numerical implementation to solve

the minimization problem and numerical results. We end with a brief discussion of our results.

2.2 Proposed Variational Restoration Model

Let Ω be an open, bounded, and connected subset of $\mathbb{R}^n$, f and u be functions from $\Omega \to \mathbb{R}$, and K be a linear operator (blur). We propose the following minimization method

$$\inf_{u \in BV(\Omega)} \Big\{ F(u) = \int_\Omega |Du| + \lambda \int_\Omega \Big[-\log \frac{f}{\sigma^2} - \log I_0(\frac{f(Ku)}{\sigma^2}) + \frac{f^2 + (Ku)^2}{2\sigma^2} \Big] dx \Big\}, \tag{2.1}$$

where $\lambda > 0$ is a tuning parameter.

2.2.1 Existence and Comparison Result

Although we deal with a nonconvex data fidelity term in (2.1), we obtain the following existence and comparison results in the purely denoising case ($K = I$).

Inspired by the work of G. Aubert and J.-F. Aujol in [AA08], we show the existence of a minimizer for (2.1) with $K = I$.

Theorem 1. *Assume that* $\inf_\Omega f(x) = \alpha > 0$*, and* $f \in L^\infty(\Omega)$*. Then the minimization problem (2.1) for* $K = I$ *admits at least one solution* $u \in BV(\Omega)$ *satisfying*

$$0 \le u \le \sup_\Omega f. \tag{2.2}$$

Proof. Let $\{u_n\} \subset BV(\Omega)$ be a minimizing sequence for (2.1). Such a sequence exists since $F(u \equiv 1) = C < \infty$. Therefore, without loss of generality, we can assume that $F(u_n) \le C$ for all n.

We denote the integrand of the Rician fidelity term by

$$G(u) = -\log \frac{f}{\sigma^2} - \log I_0(\frac{fu}{\sigma^2}) + \frac{f^2 + u^2}{2\sigma^2}, \tag{2.3}$$

and its derivative by

$$G'(u) = -\frac{I_1(\frac{fu}{\sigma^2})}{I_0(\frac{fu}{\sigma^2})} \frac{f}{\sigma^2} + \frac{u}{\sigma^2}. \tag{2.4}$$

Step 1: We show that without loss of generality, $0 \leq u_n \leq \sup_\Omega f \equiv \beta$ can be assumed for all n.

This is shown by demonstrating that the sequence $\{\phi_{0,\beta}(|u_n|)\}$ satisfies

$$F(\phi_{0,\beta}(|u_n|)) \leq F(u_n), \tag{2.5}$$

where $\phi_{0,\beta} : \mathbb{R} \to [0, \beta]$ is a cut-off function defined by

$$\phi_{0,\beta}(y) = \begin{cases} \beta & \text{for } y > \beta \\ y & \text{for } 0 \leq y \leq \beta \\ 0 & \text{for } y < 0. \end{cases} \tag{2.6}$$

Replacing u_n with $\phi_{0,\beta}(|u_n|)$ results in an equal or more optimal energy. Therefore, $\{\phi_{0,\beta}(|u_n|)\}$ is a minimizing sequence in $[0, \beta]$ that we may use in place of the original minimizing sequence $\{u_n\}$.

First, we can and should consider only non-negative minimizing sequences $\{u_n\}$. Intuitively, one expects to have

$$\int_\Omega |D|u_n|| \leq \int_\Omega |Du_n|, \tag{2.7}$$

which is shown in [Giu94] and see (1.31). Since $I_0(fu/\sigma^2)$ is an even function in u (see [Boa83]), G is also an even function, so

$$\int_\Omega G(|u_n|)dx = \int_\Omega G(u_n)dx. \tag{2.8}$$

Together with (2.7), this gives for each n

$$F(|u_n|) \leq F(u_n). \tag{2.9}$$

Hence, we can consider only non-negative minimizing sequences.

Now we show $F(\phi_{0,\beta}(|u_n|)) \leq F(|u_n|)$. Noting that $|I_1(t)/I_0(t)| < 1$ (Fig. 1.5) for all $t \in \mathbb{R}$, we obtain

$$G'(u) \geq \frac{1}{\sigma^2}(-\beta + u) \tag{2.10}$$

for any value of $u \in \mathbb{R}$. Since $G'(u) > 0$ for $u > \beta$, we can conclude that G is increasing on the interval (β, ∞). From this fact, it is easy to see that

$$G(\min(|u_n(x)|, \beta)) \leq G(|u_n(x)|) \tag{2.11}$$

for any $x \in \Omega$. This gives

$$G(\phi_{0,\beta}(|u_n(x)|)) \leq G(|u_n(x)|), \tag{2.12}$$

and therefore

$$\int_\Omega G(\phi_{0,\beta}(|u_n(x)|))dx \leq \int_\Omega G(|u_n(x)|)dx. \tag{2.13}$$

In addition, we expect

$$\int_\Omega |D\phi_{0,\beta}(|u_n|)| \leq \int_\Omega |D|u_n|| \tag{2.14}$$

to hold for $u_n \in BV(\Omega)$; a proof of a general form of this result can be found in Appendix C of [KDA99]. Combining this with (2.13) and our previous result gives for each n

$$F(\phi_{0,\beta}(|u_n|)) \leq F(|u_n|) \leq F(u_n). \tag{2.15}$$

Step 2: Let $\{u_n\}$ be a minimizing sequence; we can assume that $0 \leq u_n \leq \beta$ thanks to Step 1. Also, as mentioned previously, we can assume that $F(u_n) \leq C < \infty$ for all n. Hence,

$$\int_\Omega |Du_n| + \lambda \int_\Omega G(u_n)dx \leq C. \tag{2.16}$$

In addition,

$$G(u_n) = -\log \frac{f}{\sigma^2} - \log I_0(\frac{fu_n}{\sigma^2}) + \frac{f^2 + u_n^2}{2\sigma^2} \geq -\log(\frac{\beta}{\sigma^2}) - \log I_0(\frac{\beta^2}{\sigma^2}) = C' > -\infty, \tag{2.17}$$

which allows us to conclude that

$$\int_\Omega |Du_n|dx \leq C'' < \infty, \tag{2.18}$$

for all n. Thus, $\{u_n\}$ is a uniformly bounded sequence in $BV(\Omega)$, and there exists a subsequence $\{u_{n_k}\}$ and $u \in BV(\Omega)$ such that $u_{n_k} \to u$ in $BV - w*$ and $u_{n_k} \to u$ in $L^1(\Omega)$ ([EG92]). Hence, we must have $u \in [0, \beta]$, and by the lower semi-continuity property of the total variation and Fatou's lemma, we obtain

$$F(u) \leq \liminf F(u_{n_k}). \tag{2.19}$$

Therefore, we can conclude that u is a solution of (2.1). $\square$

Remark: For the deblurring problem, we can obtain existence of a minimizer by restricting u to $[0, \beta]$. In other words, we consider the modified problem

$$\inf_{u \in BV(\Omega), 0 \le u \le \beta} \int_\Omega |Du| + \lambda \int_\Omega \Big[-\log \frac{f}{\sigma^2} - \log I_0(\frac{f(Ku)}{\sigma^2}) + \frac{f^2 + (Ku)^2}{2\sigma^2} \Big] dx, \tag{2.20}$$

where K is a blurring kernel. With the usual properties on K (a smoothing kernel, such as Gaussian or average kernel), it is easy to conclude that $Ku \in [0, \beta]$. Thus, a proof of existence is obtainable using the ideas in step 2.

Using again techniques from G. Aubert and J.-F. Aujol's work in [AA08], we obtain the following comparison theorem.

Theorem 2. *Let f_1 and f_2 be $L^\infty(\Omega)$ functions such that $0 < \alpha_1 \le f_1 \le \beta_1 < \infty$ and $0 < \alpha_2 \le f_2 \le \beta_2 < \infty$. If we assume $f_1 < f_2$, then $u_1 \le u_2$ where u_1 and u_2 are solutions to (2.1) with $K = I$ corresponding to $f = f_1$ and $f = f_2$ respectively.*

Proof. First, denote

$$J(u) = \begin{cases} \int_\Omega |Du| & \text{if } u \in BV(\Omega) \\ +\infty & \text{if } u \in L^{\frac{n}{n-1}}(\Omega) \backslash BV(\Omega). \end{cases} \tag{2.21}$$

The solutions u_1 and u_2 exist thanks to Theorem 1. Since u_1 and u_2 are minimizers for their respective problems, it should be easy to see that

$$\begin{aligned} J(\min(u_1, u_2)) \; + \; & \int_\Omega \Big[-\log I_0\Big(\frac{f_1 \min(u_1, u_2)}{\sigma^2} \Big) + \frac{(\min(u_1, u_2))^2}{2\sigma^2} \Big] dx \\ \ge \; & J(u_1) + \int_\Omega \Big[-\log I_0\Big(\frac{f_1 u_1}{\sigma^2} \Big) + \frac{u_1^2}{2\sigma^2} \Big] dx, \end{aligned} \tag{2.22}$$

and

$$\begin{aligned} J(\max(u_1, u_2)) \; + \; & \int_\Omega \Big[-\log I_0\Big(\frac{f_2 \max(u_1, u_2)}{\sigma^2} \Big) + \frac{(\max(u_1, u_2))^2}{2\sigma^2} \Big] dx \\ \ge \; & J(u_2) + \int_\Omega \Big[-\log I_0\Big(\frac{f_2 u_2}{\sigma^2} \Big) + \frac{u_2^2}{2\sigma^2} \Big] dx. \end{aligned} \tag{2.23}$$

Adding these two inequalities, we get

$$\begin{aligned} J(\min(u_1,u_2)) &+ J(\max(u_1,u_2)) \\ &+ \int_\Omega \Big[-\log I_0\Big(\frac{f_1\min(u_1,u_2)}{\sigma^2}\Big) - \log I_0\Big(\frac{f_2\max(u_1,u_2)}{\sigma^2}\Big)\Big]dx \\ &+ \int_\Omega \Big[\frac{(\min(u_1,u_2))^2}{2\sigma^2} + \frac{(\max(u_1,u_2))^2}{2\sigma^2}\Big]dx \\ &\geq J(u_1) + \int_\Omega \Big[-\log I_0\Big(\frac{f_1u_1}{\sigma^2}\Big) + \frac{u_1^2}{2\sigma^2}\Big]dx \\ &+ J(u_2) + \int_\Omega \Big[-\log I_0\Big(\frac{f_2u_2}{\sigma^2}\Big) + \frac{u_2^2}{2\sigma^2}\Big]dx. \end{aligned} \tag{2.24}$$

Rearranging terms and using the property $J(\min(u_1,u_2)) + J(\max(u_1,u_2)) \leq J(u_1) + J(u_2)$ (see [Cha04] and [Giu94]), we obtain

$$\begin{aligned} \int_\Omega \Big[\log \frac{I_0\big(\frac{f_2u_2}{\sigma^2}\big)}{I_0\big(\frac{f_1\min(u_1,u_2)}{\sigma^2}\big)} &- \log \frac{I_0\big(\frac{f_2\max(u_1,u_2)}{\sigma^2}\big)}{I_0\big(\frac{f_1u_1}{\sigma^2}\big)} \\ +\frac{(\max(u_1,u_2))^2}{2\sigma^2} &- \frac{u_1^2}{2\sigma^2} + \frac{(\min(u_1,u_2))^2}{2\sigma^2} - \frac{u_2^2}{2\sigma^2}\Big]dx \\ \geq J(u_1) + J(u_2) &- (J(\min(u_1,u_2)) + J(\max(u_1,u_2))) \geq 0. \end{aligned} \tag{2.25}$$

Consider $\Omega_+ = \{x \in \Omega, u_1(x) > u_2(x)\}$, then from above, we arrive at

$$\int_{\Omega_+} -\left(\log \frac{I_0(\frac{f_2u_1}{\sigma^2})}{I_0(\frac{f_1u_1}{\sigma^2})} - \log \frac{I_0(\frac{f_2u_2}{\sigma^2})}{I_0(\frac{f_1u_2}{\sigma^2})}\right)dx \geq 0. \tag{2.26}$$

We have that

$$g(y) = \log \frac{I_0(c_2y)}{I_0(c_1y)} = \log I_0(c_2y) - \log I_0(c_1y) \tag{2.27}$$

is a monotonically increasing function for $y \geq 0$ if $c_2 > c_1 > 0$ since

$$\begin{aligned} g'(y) &= c_2\frac{I_1(c_2y)}{I_0(c_2y)} - c_1\frac{I_1(c_1y)}{I_0(c_1y)} \\ &= c_1\left(\frac{c_2}{c_1}\frac{I_1(c_2y)}{I_0(c_2y)} - \frac{I_1(c_1y)}{I_0(c_1y)}\right) > c_1\left(\frac{I_1(c_2y)}{I_0(c_2y)} - \frac{I_1(c_1y)}{I_0(c_1y)}\right) > 0 \end{aligned} \tag{2.28}$$

for $y > 0$. The last inequality is true because $\frac{I_1(z)}{I_0(z)}$ is a monotonically increasing function for $z \geq 0$ (see Fig. 1.5).

Hence, the integrand must be negative, implying that the Lebesgue measure of Ω_+ is 0, and therefore $u_1 \leq u_2$ on Ω. $\square$

2.3 Numerical Implementation

For numerical purposes, we consider an approximation

$$F_\epsilon(u) = \int_\Omega \sqrt{\epsilon^2 + |\nabla u|^2}\, dx + \lambda \int_\Omega \Big[-\log \frac{f}{\sigma^2} - \log I_0(\frac{f(Ku)}{\sigma^2}) + \frac{f^2 + (Ku)^2}{2\sigma^2} \Big] dx \qquad (2.29)$$

to the energy $F(u)$ to remove the singularity of $|\nabla u| = 0$ encountered in the Euler-Lagrange equation of $F(u)$. Here, ϵ is a small positive number.

In order to solve the proposed minimization problem in practice, we use the L^2 and Sobolev (H^1) gradient descent methods (see [Neu97]). It has been shown in [Neu97, Ren06] that for cases where the derivatives of u appear in the energy functional, the L^2 gradient is not guaranteed to be in L^2, but the Sobolev (H^1) gradient exists in H^1. The use of the Sobolev (H^1) gradient has been shown to give improved results in terms of quality and time in certain image processing applications [SYM07, CMP07, DMN08, BS08, Ric08, JCS09]. For this reason, we implement the Sobolev (H^1) gradient descent method in addition to the more standard L^2 gradient descent method.

In general, gradient descent methods involve the evolution problem

$$\frac{\partial u(x,t)}{\partial t} = -\nabla F_\epsilon(u), \qquad (2.30)$$

where $\nabla F_\epsilon(u)$ is dependent on the function space considered. We consider the L^2 and Sobolev (H^1) spaces and denote the gradients by $\nabla_{L^2} F_\epsilon(u)$ and $\nabla_{H^1} F_\epsilon(u)$ respectively. These gradients are defined by

$$F'_\epsilon(u)v = \langle \nabla_{L^2} F_\epsilon(u), v \rangle_{L^2}, \quad \forall v \in L^2 \qquad (2.31)$$

and

$$F'_\epsilon(u)h = \langle \nabla_{H^1} F_\epsilon(u), h \rangle_{L^2}, \quad \forall h \in H^1, \qquad (2.32)$$

where $F'_\epsilon(u)v$ and $F'_\epsilon(u)h$ are the directional derivatives of F_ϵ at u in the direction of $v \in L^2$ and $h \in H^1$ respectively.

Furthermore, it can be shown in [Neu97] that the L^2 and H^1 gradients are related in the following way

$$\nabla_{H^1} F_\epsilon(u) = (I - \triangle)^{-1} \nabla_{L^2} F_\epsilon(u) \qquad (2.33)$$

or equivalently

$$(I - \triangle)\nabla_{H^1} F_\epsilon(u) = \nabla_{L^2} F_\epsilon(u). \tag{2.34}$$

In the following we give the implementation details for these gradient descent methods.

2.3.1 L^2 Gradient Descent

For our application, the L^2 gradient descent method used to compute the minimizer involves the partial differential equation

$$\begin{aligned}\frac{\partial u}{\partial t} &= -\nabla_{L^2} F_\epsilon(u) \\ &= \lambda\bigg(- \frac{K^*Ku}{\sigma^2} + K^*\bigg(\frac{I_1(\frac{fKu}{\sigma^2})}{I_0(\frac{fKu}{\sigma^2})} \cdot \frac{f}{\sigma^2}\bigg)\bigg) + \nabla \cdot \frac{\nabla u}{\sqrt{\epsilon^2 + |\nabla u|^2}}.\end{aligned}$$

Defining

$$w^n_{i,j,k} := \frac{1}{h^2} \frac{1}{\sqrt{\epsilon^2 + \left(\frac{u^n_{i+1,j,k} - u^n_{i,j,k}}{h}\right)^2 + \left(\frac{u^n_{i,j+1,k} - u^n_{i,j,k}}{h}\right)^2 + \left(\frac{u^n_{i,j,k+1} - u^n_{i,j,k}}{h}\right)^2}},$$

we can describe the discretization we implemented by

$$\begin{aligned}\frac{u^{n+1}_{i,j,k} - u^n_{i,j,k}}{dt} &= \lambda\bigg(- \frac{K^*Ku^n_{i,j,k}}{\sigma^2} + K^*\bigg(\frac{I_1(\frac{fKu^n_{i,j,k}}{\sigma^2})}{I_0(\frac{fKu^n_{i,j,k}}{\sigma^2})} \cdot \frac{f}{\sigma^2}\bigg)\bigg) - \frac{\lambda}{\sigma^2}(u^{n+1}_{i,j,k} - u^n_{i,j,k}) \\ &+ w^n_{i,j,k}(u^n_{i+1,j,k} - u^{n+1}_{i,j,k}) - w^n_{i-1,j,k}(u^{n+1}_{i,j,k} - u^n_{i-1,j,k}) + w^n_{i,j,k}(u^n_{i,j+1,k} - u^{n+1}_{i,j,k}) \\ &- w^n_{i,j-1,k}(u^{n+1}_{i,j,k} - u^n_{i,j-1,k}) + w^n_{i,j,k}(u^n_{i,j,k+1} - u^{n+1}_{i,j,k}) - w^n_{i,j,k-1}(u^{n+1}_{i,j,k} - u^n_{i,j,k-1}),\end{aligned}$$

where $h = 1$ denotes the fixed space step and $dt > 0$ denotes the fixed timestep. We use Neumann boundary condition $\frac{\partial u}{\partial \vec{n}}|_{\partial\Omega} = 0$, where $\vec{n}$ is the unit normal to the boundary $\partial\Omega$, and we use initial condition $u^0 = f$. Note that the second term on the right-hand-side rescales the timestep and is added for numerical purposes. A larger timestep may be used as a result of adding this term.

2.3.2 Sobolev (H^1) Gradient Descent

Formally, the Sobolev (H^1) gradient descent method used to compute the minimizer involves the partial differential equation

$$\frac{\partial u}{\partial t} = (I - c\triangle)^{-1}\left(\lambda\left(- \frac{K^*Ku}{\sigma^2} + K^*\left(\frac{I_1(\frac{fKu}{\sigma^2})}{I_0(\frac{fKu}{\sigma^2})} \cdot \frac{f}{\sigma^2}\right)\right) + \nabla \cdot \frac{\nabla u}{\sqrt{\epsilon^2 + |\nabla u|^2}}\right), \tag{2.35}$$

for some $c > 0$. Note that for $c = 1$, the right-hand-side is equal to the negative of the Sobolev (H^1) gradient. The addition of $c > 0$ may lead to better results than when fixing $c = 1$, and for this reason, we add this parameter.

To understand how to work with (2.35), let us define

$$G = -\nabla_{L^2} F_\epsilon(u) \text{ and } W = -\nabla_{H^1} F_\epsilon(u). \tag{2.36}$$

Then the Sobolev (H^1) gradient descent method is given by

$$\frac{\partial u}{\partial t} = W, \tag{2.37}$$

where W solves

$$G = (I - \triangle)W, \tag{2.38}$$

as a result of (2.34).

Denoting

$$G^n_{i,j,k} = \lambda\left(- \frac{K^*Ku^n_{i,j,k}}{\sigma^2} + K^*\left(\frac{I_1(\frac{fKu^n_{i,j,k}}{\sigma^2})}{I_0(\frac{fKu^n_{i,j,k}}{\sigma^2})} \cdot \frac{f}{\sigma^2}\right)\right)$$
$$+w^n_{i,j,k}(u^n_{i+1,j,k} - u^n_{i,j,k}) - w^n_{i-1,j,k}(u^n_{i,j,k} - u^n_{i-1,j,k}) + w^n_{i,j,k}(u^n_{i,j+1,k} - u^n_{i,j,k})$$
$$-w^n_{i,j-1,k}(u^n_{i,j,k} - u^n_{i,j-1,k}) + w^n_{i,j,k}(u^n_{i,j,k+1} - u^n_{i,j,k}) - w^n_{i,j,k-1}(u^n_{i,j,k} - u^n_{i,j,k-1}).$$

we first solve (2.38) for W using the following semi-implicit scheme

$$W^{l+1}_{i,j,k} - c\Big\{\frac{W^l_{i+1,j,k} - 2W^{l+1}_{i,j,k} + W^l_{i-1,j,k}}{h_w^2} + \frac{W^l_{i,j+1,k} - 2W^{l+1}_{i,j,k} + W^l_{i,j-1,k}}{h_w^2}$$
$$+ \frac{W^l_{i,j,k+1} - 2W^{l+1}_{i,j,k} + W^l_{i,j,k-1}}{h_w^2}\Big\} = G^n_{i,j,k},$$

and denote $W_{i,j,k}$ to be its steady-state solution. Since $\frac{\partial u}{\partial t} = W$,

$$\frac{u_{i,j,k}^{n+1} - u_{i,j,k}^{n}}{dt} = W_{i,j,k},$$

and we arrive at

$$u_{i,j,k}^{n+1} = u_{i,j,k}^{n} + dt \cdot W_{i,j,k}$$

where $dt > 0$ denotes the fixed timestep as before and $h_w = 1$ denotes the fixed space step for (2.38). We use initial conditions $u^0 = f$ and $W^0 = 0$ for the first iteration and W^0 equal to the previous W for all other iterations. For the boundary, we apply Neumann boundary condition $\frac{\partial W}{\partial \vec{n}}|_{\partial\Omega} = 0$.

2.4 Numerical Results

We perform restoration experiments on a synthetic T1 MRI volume (see Fig. 2.1) obtained from BrainWeb (`http://mouldy.bic.mni.mcgill.ca/brainweb`). The L^2 and Sobolev (H^1) gradient descent methods were implemented with MATLAB using the stopping condition

$$|F_\epsilon(u^{n+1}) - F_\epsilon(u^n)| < 10^{-4}|F_\epsilon(u^n)|. \tag{2.39}$$

The data is restored with the same K and σ used to produce the input data. That is, we assume that the level of blur and noise is known.

In addition, we will use root mean square error (RMSE) to quantify the difference between two images u and v

$$\mathrm{RMSE}(u, v) = \sqrt{\frac{1}{N}\sum_{i,j,k}(u_{i,j,k} - v_{i,j,k})^2} \tag{2.40}$$

where N denotes the number of points of the image domain Ω. Preliminary results of this work have been presented and published in [GTV11].

First, we demonstrate the methods when the given image is corrupted with Rician noise but not blurred, $K = I$ and $\sigma = 0.08$. L^2 gradient descent was performed with $\lambda = 0.1$, fixed timestep $dt = 0.1$, and required 33 iterations to satisfy (2.39). Sobolev gradient descent was performed with $\lambda = 0.15$, $c = 1.5$, fixed timestep $dt = 0.05$, and required 16 iterations.

Fig. 2.2 shows the decrease of the energy functionals versus iteration. Since different values of λ were used, energy values are not directly comparable. Fig. 2.3 shows the results, and the experiment computation times and RMSE results are summarized in Table 2.1. For comparison, we give the denoising results using the Rudin-Osher-Fatemi (ROF) model [ROF92, RO94], which was implemented with L^2 gradient descent. We note that implementation of ROF using split Bregman [GO09] reduces the computation time substantially. However, we would also like to mention that since our proposed minimization problem is not convex, split Bregman cannot be applied to our problem. For this denoising experiment, the Sobolev (H^1) gradient descent implementation of our proposed model performs the best in terms of RMSE. Furthermore, we give error histograms for the noisy data and each of the denoised data in Fig. 2.4. The error histograms corresponding to the denoised results using our proposed method are more centered around zero, suggesting a more accurate restoration.

The next two experiments are joint denoising and deblurring problems, where the data is both blurred with a Gaussian kernel and corrupted with Rician noise.

For the first denoising and deblurring experiment, we add a large amount of Rician noise and a small amount of Gaussian blur. In particular, in Fig. 2.5, we show the case where we have added Rician noise with $\sigma = 0.08$ and Gaussian blur K with standard deviation 0.6 voxels. L^2 gradient descent was performed with $\lambda = 0.2$, fixed timestep $dt = 0.1$, and required 21 iterations. Sobolev (H^1) gradient descent was performed with $\lambda = 0.25$, $c = 1.25$, fixed timestep $dt = 0.05$, and required 12 iterations. Fig. 2.6 shows the corresponding error histograms, and Table 2.2 gives a comparison of the computation times and RMSE (middle row). In this case, the Sobolev (H^1) implementation of our model performs better than the L^2 implementation in terms of RMSE.

For the second denoising and deblurring experiment, we show in Fig. 2.7 an experiment with less noise and heavier blur. Here, Rician noise with $\sigma = 0.02$ and Gaussian blur K with standard deviation 1.5 voxels was added to the clean data. L^2 gradient descent was performed with $\lambda = 0.4$, fixed timestep $dt = 0.1$, and required 28 iterations. Sobolev (H^1) gradient descent was performed with $\lambda = 1.1$, $c = 5$, fixed timestep $dt = 0.001$, and required 30 iterations. Fig. 2.8 shows the corresponding error histograms, and Table 2.2 gives a comparison

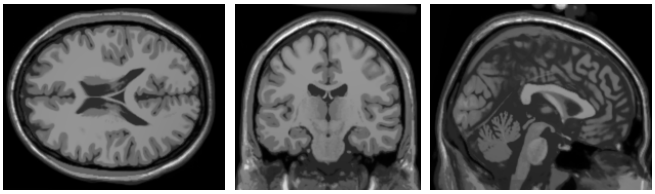


Figure 2.1: Three slices of the clean synthetic T1 MRI volume.

	f	L^2	H^1	ROF
RMSE (whole)	0.093702	0.039935	**0.034541**	0.071503
RMSE (cube)	0.079473	0.029803	**0.027669**	0.031080
iterations	–	33	16	14
time (s)	–	1091	833	82.4

Table 2.1: Comparison of computation times and RMSE values for f, restored u using our model with L^2 gradient descent, restored u using our model with Sobolev (H^1) gradient descent, and restored u using ROF method. Noisy data is generated by applying Rician noise with $\sigma = 0.08$.

of computation times and RMSE (bottom row). For this case, the L^2 implementation of our model performs better than the Sobolev (H^1) implementation in terms of RMSE.

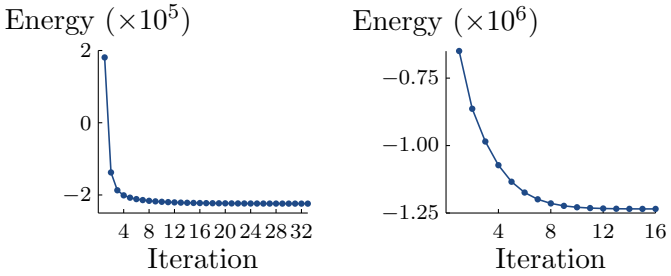


Figure 2.2: Energy versus iteration for denoising experiments. Left: L^2 gradient descent. Right: Sobolev (H^1) gradient descent.

2.5 Discussion

TV-based restoration can be formulated for the Rician noise model as a maximum a posteriori estimate. The Sobolev (H^1) gradient descent implementation of our model seems to perform better with denoising, and the L^2 implementation of our model seems to perform better with

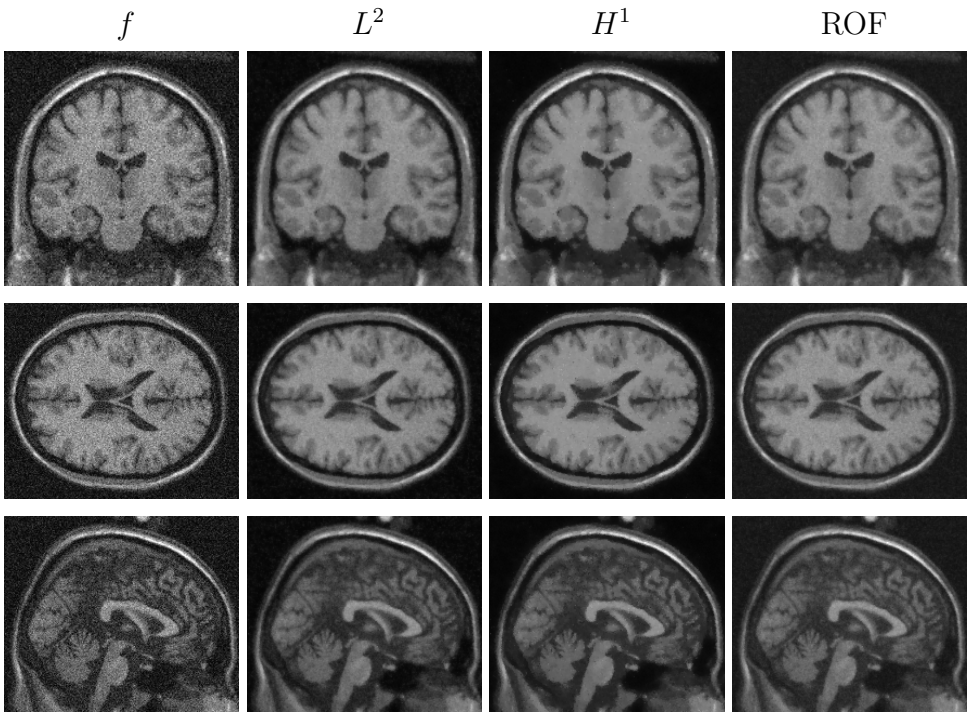


Figure 2.3: Denoising experiment with Rician noise with $\sigma = 0.08$ and no blur. Left to right: noisy data, denoised data using our model with L^2 implementation, denoised data using our model with Sobolev (H^1) implementation, and denoised data using ROF.

Experiment	Method	Time (s)	RMSE
Fig. 2.3	L^2 Gradient Descent	1091	0.039935
	Sobolev Gradient Descent	833	0.034541
Fig. 2.5	L^2 Gradient Descent	791	0.037068
	Sobolev Gradient Descent	929	0.032020
Fig. 2.7	L^2 Gradient Descent	950	0.024279
	Sobolev Gradient Descent	1419	0.029659

Table 2.2: Comparison of computation times and RMSE values.

denoising-deblurring.

Furthermore, there is improvement in denoising using our proposed model over the ROF model [RO94, ROF92], which is due to the fact that the Rician statistics is incorporated in the fidelity term of our variational restoration model. The improvement is most seen in areas

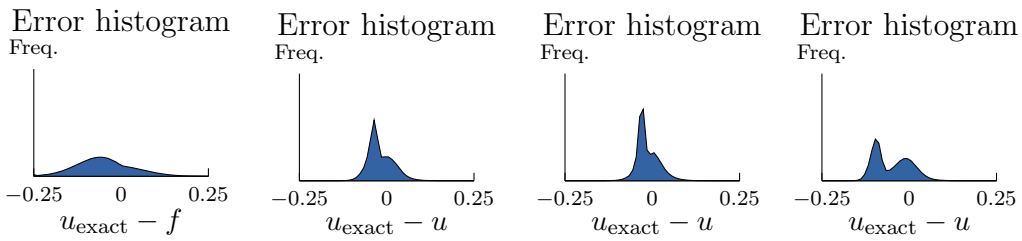


Figure 2.4: Error histograms corresponding to denoising experiment with Rician noise with $\sigma = 0.08$ and no blur (see Fig. 2.3). Left to right: error histograms for $u_{\text{exact}} - f$, $u_{\text{exact}} - u$ with restored u using L^2 implementation, $u_{\text{exact}} - u$ with restored u using H^1 implementation, and $u_{\text{exact}} - u$ with restored u using ROF.

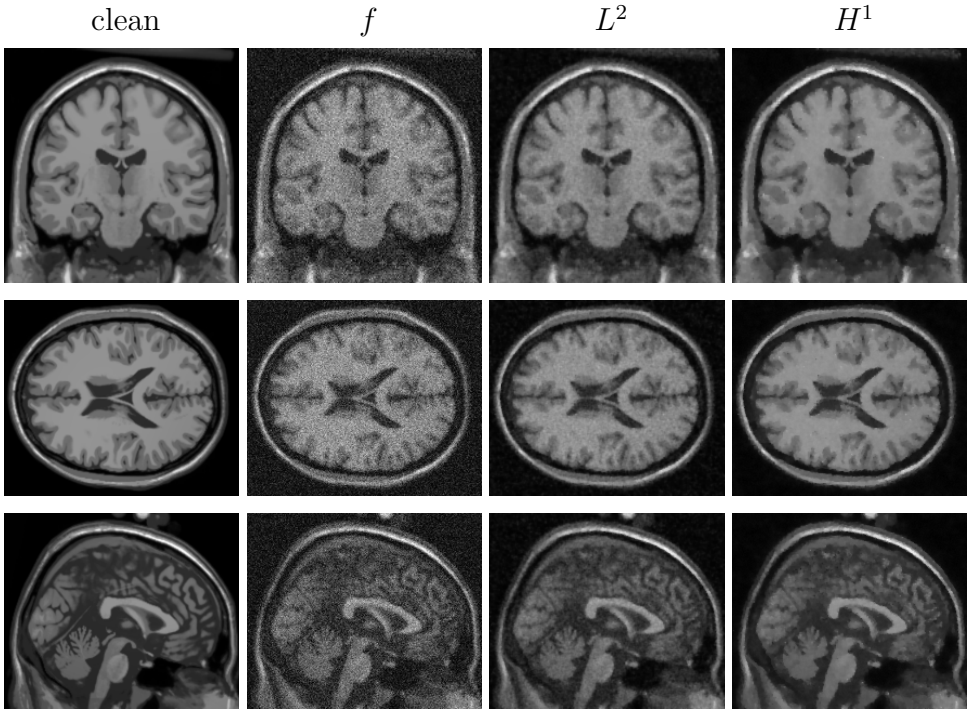


Figure 2.5: Restoration experiment with Rician noise with $\sigma = 0.08$ and Gaussian blur with standard deviation 0.6 voxels. Left to right: clean data, noisy/blurry data, restored data using L^2 implementation, and restored data using Sobolev (H^1) implementation.

where the true intensity values are small compared to the standard deviation of the noise σ. The reason for this is, again, the fact that in these regions, the Rician distribution is no longer well approximated using a Gaussian distribution. As seen in the purely denoising experiment (Fig. 2.3), the dark regions (intensity values close to 0) are better reconstructed using our

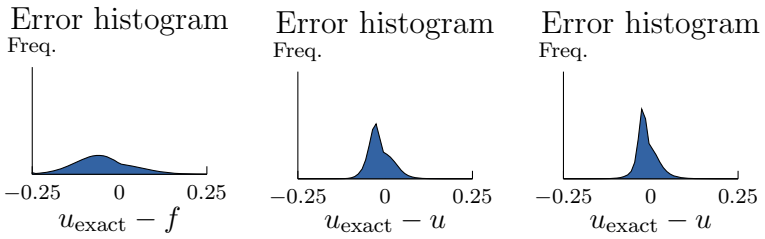


Figure 2.6: Error histograms corresponding to Rician noise with $\sigma = 0.08$ and Gaussian blur with standard deviation 0.6 voxels (see Fig. 2.5). Left to right: error histograms for $u_{\text{exact}} - f$, $u_{\text{exact}} - u$ with restored u using L^2 implementation, and $u_{\text{exact}} - u$ with restored u using H^1 implementation.

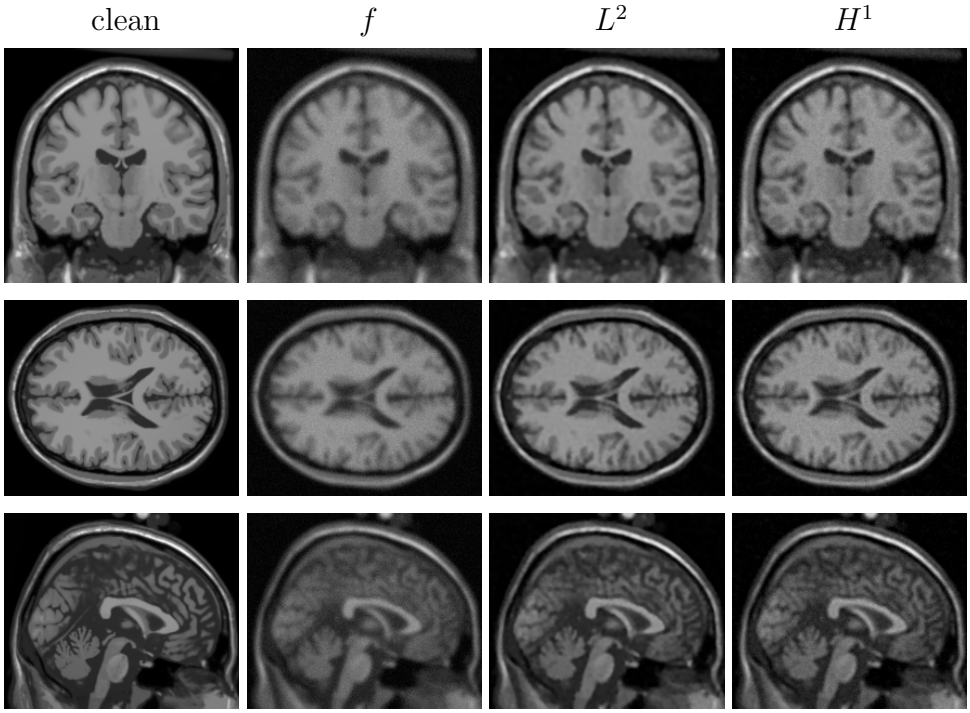


Figure 2.7: Restoration experiment with Rician noise with $\sigma = 0.02$ and Gaussian blur with standard deviation 1.5 voxels. Left to right: clean data, noisy/blurry data, restored data using L^2 implementation, and restored data using Sobolev (H^1) implementation.

proposed model than with ROF. In Fig. 2.4, we see a second bump that is skewed to the left of zero in the error histograms corresponding to all the restored results. This error is due to the difficulty in reconstructing the intensity values that are small compared to σ; the regularization term will inevitably introduce some smoothing effects that lead to a reconstructed value that is

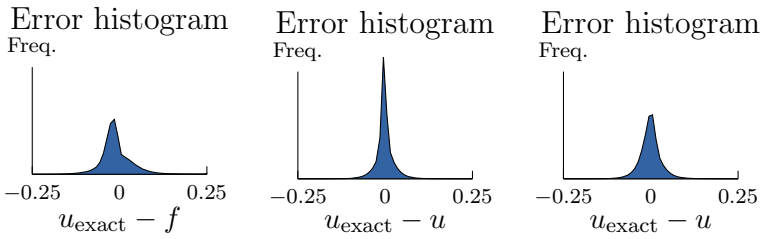


Figure 2.8: Error histograms corresponding to Rician noise $\sigma = 0.02$ and Gaussian blur with standard deviation 1.5 voxels (see Fig. 2.7). Left to right: error histograms for $u_{\text{exact}} - f$, $u_{\text{exact}} - u$ with restored u using L^2 implementation, and $u_{\text{exact}} - u$ with restored u using H^1 implementation.

larger than the true value since Rician noise tends to increase the intensity values on average. However, the restored results using our model with both the L^2 and Sobolev (H^1) gradient descent implementations produce error histograms that are more centered around zero than the error histogram produced by the restored result using ROF. In addition, RMSE values (calculated on the entire 3D volume and in a cube completely inside the brain region) are lower for the restored results using our model than the restored result using ROF.

CHAPTER 3

Restoration of HARDI Images

3.1 Introduction to HARDI Images

High Angular Resolution Diffusion Imaging (HARDI) [TWB99, TRW02] is a modality of Magnetic Resonance (MR) imaging used in reconstructing fibers in the brain. Water diffusion in the brain causes attenuation of MR signals. Since water diffuses preferentially in the direction of fiber pathways, MR imaging presents a method for tracking fibers in the brain.

More specifically, the Stejskal-Tanner equation [ST65] gives a relationship between water diffusion and corresponding MR signal

$$S(x,\theta,\phi) = S_0(x)\exp\left(-b \cdot d(x,\theta,\phi)\right) \tag{3.1}$$

where $S(x,\theta,\phi)$ and $d(x,\theta,\phi)$ are the MR signal and corresponding diffusion at position x and direction on the sphere given by

$$(\cos(\theta)\sin(\phi), \sin(\theta)\sin(\phi), \cos(\phi)).$$

$d(x,\theta,\phi)$ is also known as the spherical apparent diffusion coefficient (sADC). $S_0(x)$ is the MR signal at x when no diffusion gradient is applied, and b is a parameter used in collecting the data.

The true MR signal cannot be larger than the MR signal when no diffusion gradient is applied. That is we have the constraints

$$S(x,\theta,\phi) \leq S_0(x) \quad \text{or} \quad d(x,\theta,\phi) \geq 0. \tag{3.2}$$

HARDI data is a collection of diffusion weighted MR images (DW-MRI), meaning that for each diffusion direction considered, we have a corresponding MR signal. As the number of

diffusion directions increases, the angular resolution increases as well. The number of diffusion directions used in typical HARDI data is ≈ 100, hence the "High Angular Resolution" part of HARDI. For more details about HARDI, we refer to [Fra02, LCD08].

HARDI data is contaminated by Rician noise since it is a collection of diffusion weighted MR images. The level of noise present varies as a function of the collection parameter b, with larger b values leading to more noise but more accurate diffusivity information [CUF09]. Thus, there is a tradeoff between more accurate information and less noise. However, with any choice of b, the introduction of noise will likely cause violations of the constraint (3.2) and change important characteristics measured from the HARDI data such as fractional anisotropy (FA) [PB96, And01, SLN00] of the fiber pathways. Given this knowledge, it may be beneficial to denoise the HARDI data prior to the extraction of fibers.

In [MVO09], the authors propose a variational restoration model to denoise HARDI data that utilizes a Gaussian probability distribution as an approximation to the Rician probability distribution, assuming high signal to noise ratio. This leads to a simple L^2 fidelity term, where the standard deviation of the noise σ does not appear explicitly. In [KTV10, KTT09], an L^1 fidelity term is utilized, which is more suitable than an L^2 fidelity term while being simpler to model than Rician noise. Again, σ does not appear explicitly in the variational denoising model.

Recently, [MRD11, PST11] look for sparse representations of HARDI signals using spherical ridglets and learned dictionaries, respectively. These approaches are similar to modeling HARDI signals using high-order spherical harmonics series [DAF07] in that the goal is to represent the HARDI signal as a linear combination of some basis functions, but differ in that a *sparse* representation is sought.

Of the previous works, the most relevant to ours are [BFW06, MVO09, KTV10, KTT09]. Our proposed model is different from [BFW06, MVO09] in that we incorporate the data acquisition model (3.1) and from [KTV10, KTT09] in that we impose a more accurate noise model, the Rician distributed one.

3.2 Proposed Variational Denoising Model

We propose a variational denoising model motivated by a *maximum a posteriori* (MAP) estimate which attempts to find the most likely sADC, $d = (d_1, ..., d_n)$ to have generated a given observed noisy signal $S = (S_1, ..., S_n)$. Here, n represents the number of diffusion directions and d_i and S_i represent the sADC and noisy signal in direction i on the sphere. This d is the minimizer of an energy composed of two terms:

$$\inf_{d:\Omega\to\mathbb{R}^n} \left\{ F(d) = \int_\Omega |D(d)| + \lambda \int_\Omega \sum_{i=1}^{n} \left(-\log I_0\left(\frac{S_i S_0 e^{-P(d_i)}}{\sigma^2}\right) + \frac{(S_0 e^{-P(d_i)})^2}{2\sigma^2}\right) dx \right\} \tag{3.3}$$

where $\lambda > 0$ is a tuning parameter, σ is the standard deviation of the noise, S_i is the observed noisy HARDI signal measured in the spherical direction i, and d_i is the denoised unknown sADC in the same direction. We denote by $u_i = S_0 \exp(-P(d_i))$ the recovered denoised HARDI signal (here, we use the data acquisition model (3.1) with $b = 1$, which rescales d; the tuning parameter λ accounts for this rescaling). P is a projection operator given by

$$P(z) = \begin{cases} z & \text{if } z \geq 0 \\ 0 & \text{if } z < 0, \end{cases} \tag{3.4}$$

which is introduced to satisfy the constraint $d_i \geq 0$, and $\int_\Omega |D(d)|$ is the vectorial total variation of d defined by

$$\int_\Omega |D(d)|$$

$$= \sup \left\{ \sum_{i=1}^{n} \int_\Omega d_i \mathrm{div}(\phi_i) dx \;\middle|\; \phi = (\phi_1, ..., \phi_n) \text{ and } \phi_i \in C_c^1(\Omega; \mathbb{R}^3),\ ||\phi||_\infty \leq 1 \right\},$$

which reduces to

$$\int_\Omega |D(d)| = \int_\Omega \sqrt{\sum_{i=1}^{n} |\nabla d_i|^2} dx. \tag{3.5}$$

when d is differentiable.

The first term of the energy, $\int_\Omega D(d)$, is a prior on d which provides regularization, making the minimization problem well-posed. As mentioned earlier, this prior is suitable for denoising as it provides smoothing effects while preserving edges.

The second term of the energy is the fidelity term. Up to a constant, it is equal to the sum of $-\log Pr(S_i\ ;\ S_0\exp(-P(d_i)),\sigma)$ over Ω and diffusion directions $i = 1,...n$, where $Pr(\cdot\ ;\ \cdot)$ is the Rician probability density function given in (1.1). This term encourages

$$Pr(S_i\ ;\ S_0\exp(-P(d_i)),\sigma)$$

to be large so that the contribution of

$$-\log Pr(S_i\ ;\ S_0\exp(-P(d_i)),\sigma)$$

is small. This is equivalent to finding d_i such that the probability of generating the noisy S_i given this d_i is large, assuming that the noisy signal S_i is the result of an underlying clean signal $u_i = S_0\exp(-P(d_i))$ corrupted with Rician distributed noise.

3.3 Numerical Implementation

To solve the minimization problem (3.3) numerically, we consider the Euler-Lagrange equations in combination with the L^2 gradient descent method. The resulting evolution equations are

$$\frac{\partial d_i}{\partial t} = \lambda P'(d_i)\left(\frac{u_i^2}{\sigma^2} - \frac{I_1(\frac{S_iu_i}{\sigma^2})}{I_0(\frac{S_iu_i}{\sigma^2})}\frac{S_iu_i}{\sigma^2}\right) + \nabla\cdot\left(\frac{\nabla d_i}{\sqrt{\sum_{i=1}^n|\nabla d_i|^2}}\right) \tag{3.6}$$

for $i = 1,...,n$. We use forward and backward differencing to discretize the second term and take a fully explicit scheme to obtain an iterative method from (3.6). As mentioned earlier, the projection operator P is used to help satisfy the constraint $d_i \geq 0$. In the evolution equations (3.6), we see that the contribution of P appears as a derivative factor, P', multiplying the term in the Euler-Lagrange equation corresponding to the fidelity term of the energy. Since P is the projection operator, then theoretically P' is the Heaviside function. However, for numerical purposes, we will use a smooth approximation to the Heaviside function

$$P'(z) = \begin{cases} 1 & \text{if } z \geq a \\ 0.5\big(1 + \frac{1}{a}z + \frac{1}{\pi}\sin\big(\frac{\pi}{a}z\big)\big) & \text{if } |z| < a \\ 0 & \text{if } z \leq -a, \end{cases} \tag{3.7}$$

where a is a small positive parameter. For places where d_i is positive, then $P'(d_i) = 1$, and we get both the contribution of the fidelity and regularization terms in the evolution equations. However, if d_i is negative (violating the constraint that $d_i \geq 0$), then $P'(d_i) = 0$, and only the regularization term appears in the evolution equations. As mentioned, the regularization term provides smoothing; at voxels where d_i is negative, the value of d_i at these voxels will depend on the values of the surrounding d_i's in space. Considering a specific voxel where $d_i < 0$, its new value will be determined by a "weighted averaging" of the surrounding d_i's in a small neighborhood of this voxel. If the values of the surrounding d_i's are positive, then this "weighted averaging" will influence the negative d_i to increase. This is the motivation for the initial guess that we chose:

$$d_i(x) = \begin{cases} 0.005 & \text{if } S_i(x) > S_0(x) \\ -\log\left(\frac{S_i(x)}{S_0(x)}\right) & \text{if } S_i(x) \leq S_0(x). \end{cases} \tag{3.8}$$

The combination of a positive initial guess and the role of the projection operator P helps satisfy the constraint $d_i \geq 0$. Decreasing the constant 0.005 results in more violations of $d_i \geq 0$ and increasing it may lead to less satisfactory numerical results. As a note, using a sufficiently small timestep in the iterative method will guarantee $d_i \geq 0$ but will lead to a longer computation time.

We implement Neumann boundary conditions $\frac{\partial d_i}{\partial \eta} = 0$ where η is the normal to the boundary. Iterations are stopped when the energy $F(d)$ decreases to a steady state.

Furthermore, for efficiency, we use a cubic rational polynomial approximation to I_1/I_0 (see [GTV11]):

$$\frac{I_1(t)}{I_0(t)} \approx \frac{t^3 + 0.950037t^2 + 2.38944t}{t^3 + 1.48937t^2 + 2.57541t + 4.65314}. \tag{3.9}$$

3.4 Numerical Experiments and Results

We perform numerical experiments on synthetic, phantom, and real data sets to demonstrate the validity of our proposed variational denoising model. With the first two experiments, we consider the root mean square error (RMSE) for vectorial data as well as a visualization of

the data created by the first author of [KTV10] as metrics of evaluation of results. The aims of these initial experiments were to understand how the different components of the model (e.g. projection operator ($P(d)$), initial condition, and tuning parameter λ) affect the denoised result. For the latter experiments, we compute orientation distribution functions (ODFs) and fractional anisotropy (FA) to gain a better idea of the applicability of our denoising model to real applications such as fiber tracking. In the next section, we give a brief summary of the metrics of evaluation that we use.

3.4.1 Metrics of Evaluation

In this section, we briefly explain the metrics of evaluation that we use.

First, we use the root mean square error (RMSE) between two vectorial MR signals u_1 and u_2 given by

$$\text{RMSE}(u_1, u_2) = \sqrt{\frac{1}{n \cdot N} \sum_{i=1}^{N} \sum_{j=1}^{n} (u_1(x_i, s_j) - u_2(x_i, s_j))^2}.$$

$x_i \in \Omega$, $i = 1, ..., N$ are the points of the domain and s_j, $j = 1, ..., n$ represent the diffusion directions.

For our initial visualizations of the vectorial MR signal, we follow [KTV10]. For an example, see Fig. 3.1 for a visualization of exact two dimensional data provided by McGraw et al. [MVO09]. Every third point in the spatial domain is plotted, and 81 vectors correspond to each point. Each vector points in the direction of diffusion and its length is determined by the magnitude of the MR signal when a diffusion gradient is applied in that direction. The tips of the vectors trace out the surface of a morphed sphere. With the special case where all 81 vectors have equal length, we have a sphere with radius equal to the length of the vectors.

We will also look at fractional anisotropy (FA) of fibers and orientation distribution functions (ODFs) and associated Jensen-Shannon divergence (JSD). We would like to mention that FA, ODFs and JSDs were calculated by Liang Zhan at the Laboratory of Neuro Imaging (LONI) at UCLA, and these results were presented and published in the work [TKZ12].

Fractional anisotropy (FA) is a scalar between 0 and 1 that represents the amount of

anisotropy of water diffusion at a particular voxel. FA value of 0 represents isotropic diffusion, meaning equal diffusion in all directions. FA values near 1 represent a more directed diffusion, so that diffusion occurs mostly in one direction only. FA is computed from the eigenvalues, λ_1, λ_2, and λ_3 corresponding to the eigenvectors of the diffusion tensor.

$$\text{FA} = \sqrt{\frac{3}{2}}\sqrt{\frac{(\lambda_1 - \hat{\lambda})^2 + (\lambda_2 - \hat{\lambda})^2 + (\lambda_3 - \hat{\lambda})^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}, \tag{3.10}$$

where $\hat{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$.

We compute the diffusion orientation distribution function (ODF), which is a probability density function measuring the distribution of water diffusion in different directions on the sphere; for an example, see Fig. 3.9. ODF's are calculated using the tensor distribution function algorithm in [LZZ09]. We will also use the Jensen-Shannon divergence (JSD) to measure the difference between the noisy/denoised ODF and the ODF of the ground truth. JSD between two ODFs P and Q are given by

$$JSD(P, Q) = \frac{1}{2}D_{KL}(P, M) + \frac{1}{2}D_{KL}(Q, M), \tag{3.11}$$

where $M = (P+Q)/2$ and $D_{KL}(\cdot, \cdot)$ denotes the Kullback-Leibler information distance between two probability distributions [KL51]. For probability distributions P_1 and P_2 defined on the sphere S^2, D_{KL} is given by

$$D_{KL}(P_1, P_2) = \int_{S^2} \log\Big(\frac{P_1(\theta, \phi)}{P_2(\theta, \phi)}\Big) P_1(\theta, \phi) dS. \tag{3.12}$$

3.4.2 Numerical Results

For our first data set, we consider two dimensional (2D) synthetic data provided by McGraw et al. [MVO09]. It is 16×16 in spatial dimension with 81 diffusion directions (see Fig. 3.1). We add Rician noise with standard deviation $\sigma = 18$ to create the noisy data, denoted S. Since S_0 is not available for this synthetic data, S_0 is taken to be a constant 255, and initial guess is taken to be $d_0 = -\log(S/S_0)$. Recall that the denoised MR signal is given by $u = S_0 \exp(-P(d))$. The visualization of our results is given in Fig. 3.1 and the corresponding error histograms are given in Fig. 3.2. The RMSE of the noisy data is 17.8147 and is reduced to 5.4968 after denoising

with our model. In Table 3.1, we give a comparison of our results with the results from similar denoising experiments found in [MVO09, KTV10]; our model gives a slight improvement in terms of RMSE.

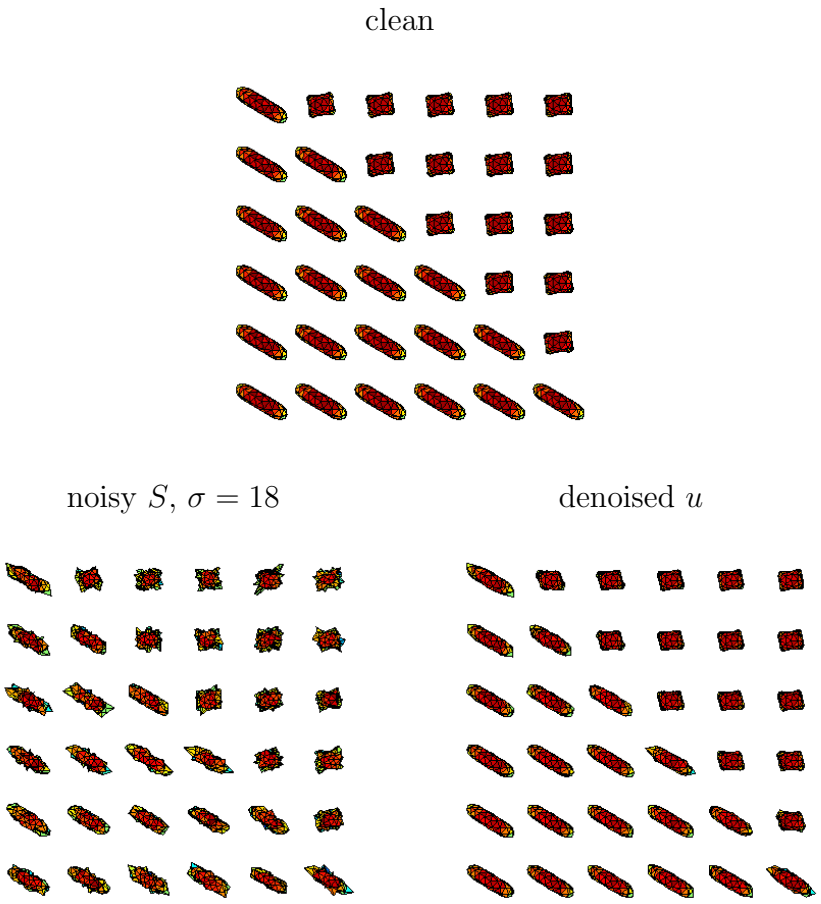


Figure 3.1: Top: exact 2D synthetic data. Left: noisy data, corrupted by Rician noise $\sigma = 18$. Right: restored data. Denoised data satisfies $d \geq 0$.

Second, we consider real HARDI data of the brain. A healthy subject was scanned on a 4 Tesla Bruker Medspec MRI scanner with an optimized diffusion imaging sequence. DWI parameters were: echo and repetition time, TE/TR 92.3/8250 ms, 55×2 mm contiguous slices, field of view: FOV = 23 cm. 41 images were collected: 11 baseline (S_0) images with no diffusion sensitization (i.e., T_2- weighted images) and 30 diffusion-weighted images (b-value: 1159 s/mm^2) with gradient directions evenly distributed on the hemisphere. The reconstruction matrix was 128×128, yielding a 1.8×1.8 mm^2 in-plane resolution. The clean data was created by fitting

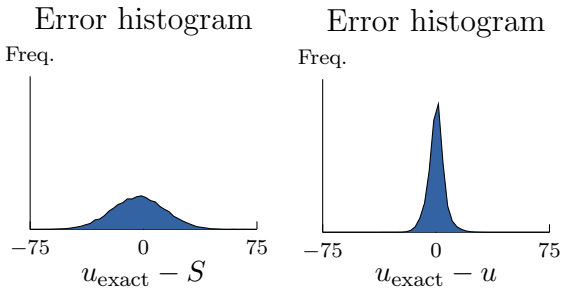


Figure 3.2: Left: error histogram of noisy data, RMSE S: 17.8147. Right: error histogram of denoised data, RMSE u: 5.4968.

	RMSEdenoised/RMSEnoisy
TV+L^1 fidelity denoising in u [KTV10]	5.8991/17.7079 = 0.3331
TV+L^1 fidelity denoising in d [KTV10]	7.6081/17.7079 = 0.4296
FEM in [MVO09]	11.9964/17.7079 = 0.6675
FEM + TV in [MVO09]	7.6367/17.7079 = 0.4313
our model	5.4968/17.8147 = **0.3086**

Table 3.1: Comparison of RMSEdenoised/RMSEnoisy ratios for different methods.

actual HARDI data to a 6th order spherical harmonic expansion, and the resulting data set is considered to be the "ground truth" even though this is not truly correct since it has artifacts and negative d values. The noisy data was generated by adding Rician noise with $\sigma = 15$ to this "ground truth". We take a subset (size $95 \times 128 \times 55$) of the 3D volume, and use all 30 uniformly distributed diffusion directions. Fig. 3.3 gives a visualization of the 29th slice of the clean data, and Fig. 3.4 gives S_0, clean MR signal, and clean diffusion signal d corresponding to the 10th diffusion-sensitized direction. In Fig. 3.5 and Fig. 3.6, we give visualizations of the clean, noisy, and denoised images corresponding to 2 of the 30 diffusion-sensitized directions. For this particular experiment, we used $\lambda = 0.05$. The RMSE of the noisy data is 10.5454 and the RMSE of the denoised data is 8.0845, showing some improvement after denoising. In addition, the number of points where the constraint $d \geq 0$ is not satisfied is reduced from 815 points in the noisy data to 52 points in the denoised data. When we reduce the tuning parameter to $\lambda = 0.001$ to introduce more smoothing, we see that the denoised solution is

oversmoothed (Fig. 3.7 and Fig. 3.8), but the constraint $d \geq 0$ is satisfied everywhere.

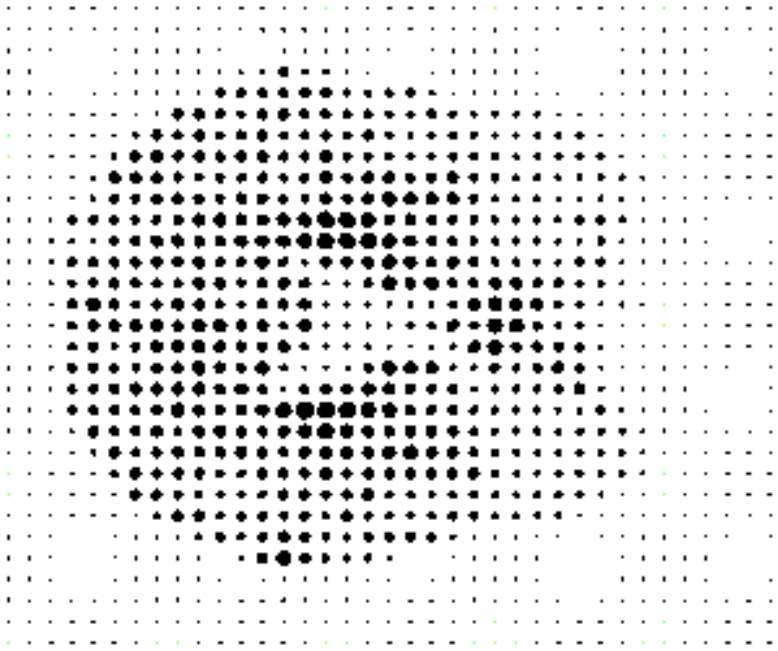


Figure 3.3: Visualization of 29th slice of clean data. Complete volume size: 95x128x55, 30 diffusion directions.

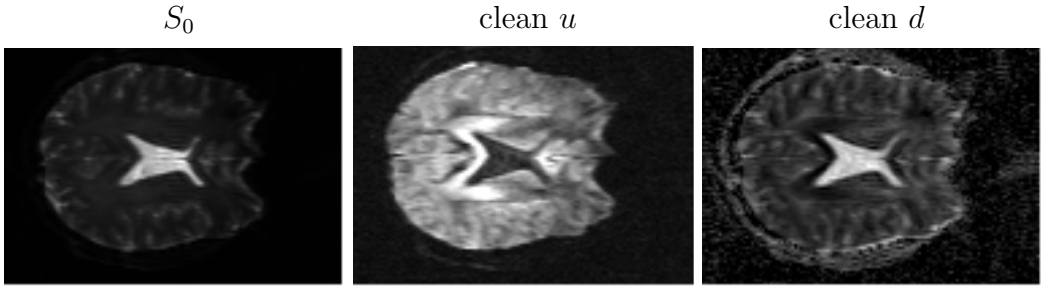


Figure 3.4: Left to right: S_0, clean u, and corresponding clean d of 29th slice, 10th diffusion direction.

For the last three numerical experiments, we look at fractional anisotropy (FA), orientation distribution functions (ODFs), and Jensen-Shannon divergence (JSD). In the last experiment, we will revisit the real HARDI data set and evaluate the performance of our denoising model with ODFs and JSD values.

We consider again 2D synthetic data. The size of the data set is 8×8 with diffusion measured in 94 uniformly distributed directions at each spatial point. S_0 in this case is assumed to be

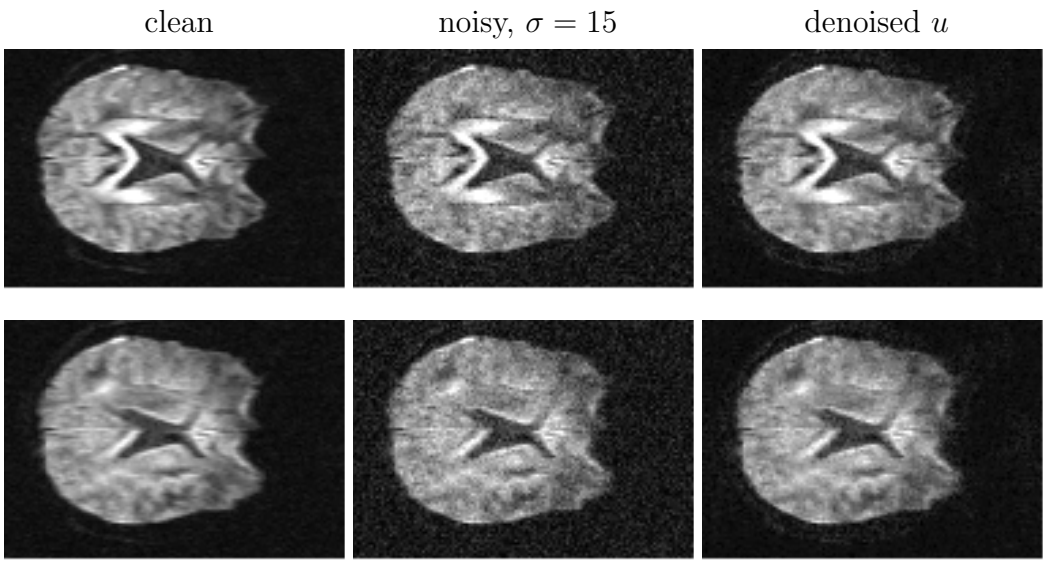


Figure 3.5: Visualization of 2 of 30 diffusion-sensitized MR signals of the real HARDI brain data. Left to right: clean, noisy, denoised data. RMSE of denoised data: 8.0845. RMSE of noisy data: 10.5454. $\lambda = 0.05$, 35 iterations. $d_{\text{noisy}} < 0$ at 815 points, $d_{\text{denoised}} < 0$ at 52 points.

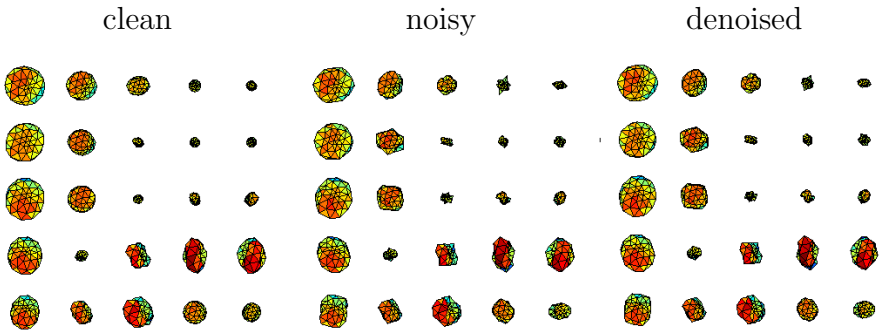


Figure 3.6: Visualization of small portion of clean, noisy, denoised data. RMSE of denoised data: 8.0845. RMSE of noisy data: 10.5454. $\lambda = 0.05$, 35 iterations. $d_{\text{noisy}} < 0$ at 815 points, $d_{\text{denoised}} < 0$ at 52 points.

a constant equal to 1. We create noisy data by adding Rician noise with standard deviation $\sigma = 0.01, 0.02, 0.04, 0.1$ and 0.2. The same σ used to create the noisy data is input as a known parameter in our denoising model. Fig. 3.9 gives the ODFs of the clean, noisy, and denoised data corresponding to the case $\sigma = 0.2$. In all cases, the denoised data gives lower JSD values than those of the corresponding noisy data sets, indicating an increase in similarity to the

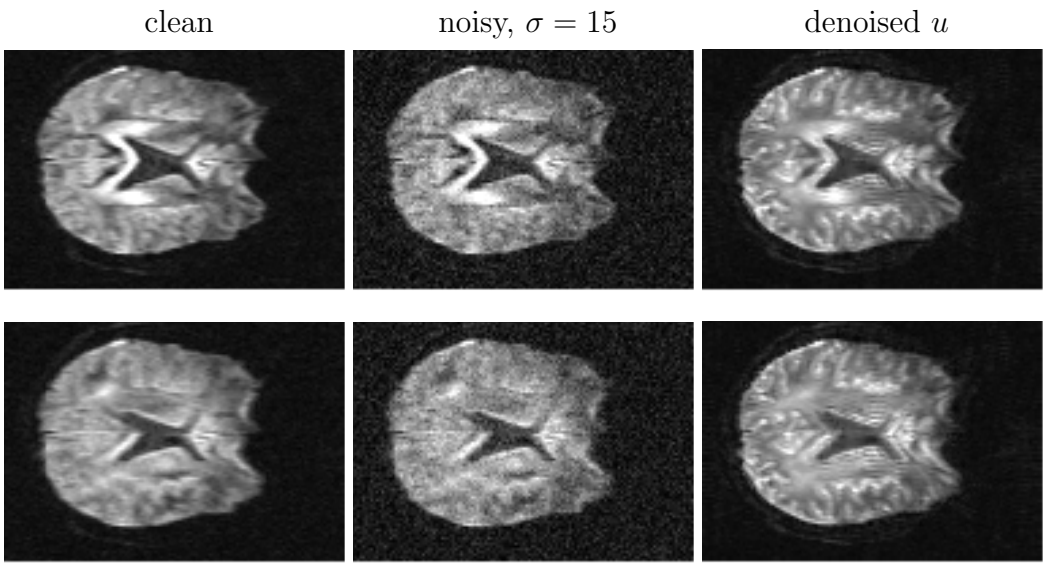


Figure 3.7: Visualization of 2 of 30 diffusion-sensitized MR signals of the real HARDI brain data. Left to right: clean, noisy, denoised data. With $\lambda = 0.001$ and at 100 iterations, the solution is oversmoothed. However, $d_{\text{denoised}} \geq 0$ everywhere.

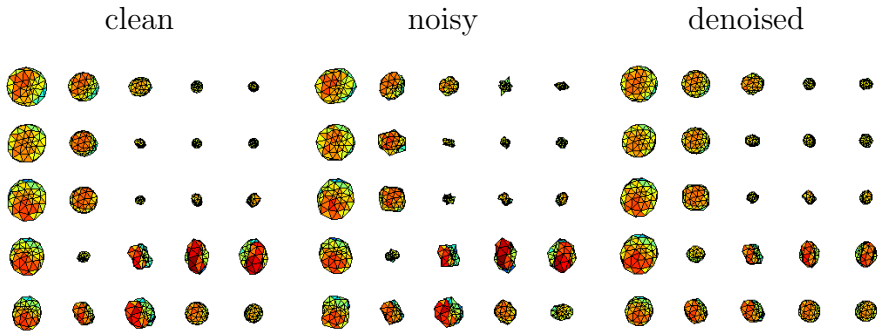


Figure 3.8: Visualization of small portion of clean, noisy, denoised data. With $\lambda = 0.001$ and at 100 iterations, the solution is oversmoothed. However, $d_{\text{denoised}} \geq 0$ everywhere.

ODF of the ground truth after denoising (Table 3.2). Also, RMSE values between the denoised and ground truth data are lower than the RMSE values between noisy and ground truth data (Table 3.2). This suggests that we have a more accurate signal after denoising.

Next, we consider DW-MRI data of a hardware phantom containing synthetic fibers created by Pullens et al. [PRG10] (see Fig. 3.10). The fibers ($\approx$ 10 μm circular diameter) consist of polyester yarns wound into bundles, which are then interdigited on top of each other and

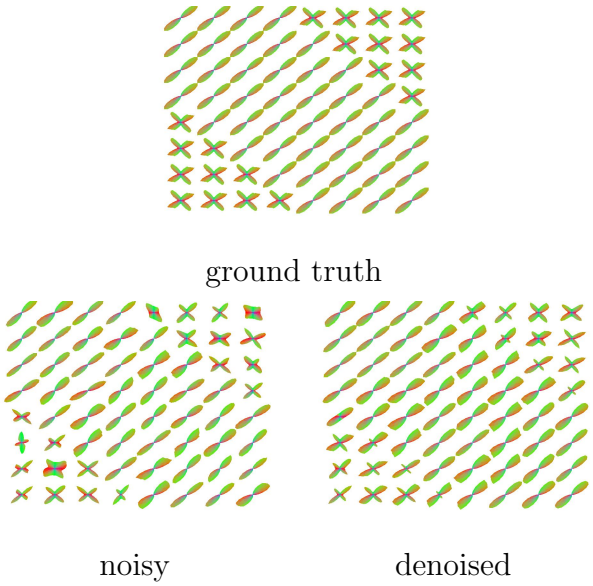


Figure 3.9: Top: ODF of ground truth data. Bottom left: ODF of noisy data generated with Rician noise with $\sigma = 0.2$. Bottom right: ODF of denoised data.

	JSD		RMSE	
σ	noisy	denoised	noisy	denoised
0.01	0.0044	0.0004	0.0100	0.0065
0.02	0.0102	0.0008	0.0201	0.0130
0.04	0.0441	0.0133	0.0396	0.0231
0.1	0.1578	0.0431	0.0986	0.0545
0.2	0.5423	0.1736	0.1950	0.0963

Table 3.2: Jensen-Shannon divergence (JSD) values between ODF of noisy/denoised data and ODF of ground truth data, and RMSE values between noisy/denoised data and ground truth data.

secured with heat shrink tubes. We have two scans of this phantom using a 7T and 3T scanner. For both these scans, the standard deviation of the noise σ is approximated using the method described in Section 1.2.

For the data represented in Fig. 3.11, a 7T scanner was used to collect the DW-MRI data of the phantom. The data collection was performed at the CMRR, University of Minnesota, on a Magnex Scientific MRI scanner driven by a Siemens console, with a head gradient insert

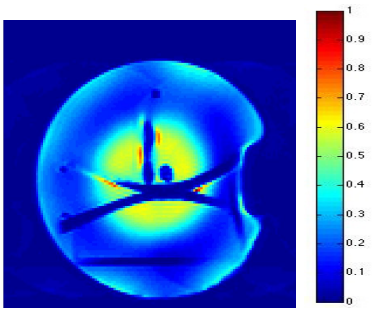
Figure 3.10: Hardware phantom containing synthetic fibers created by Pullens et al. [PRG10].

capable of 80 mT/m in 200 ms. Parameter settings were: 66 slices with FOV = 192 mm $\times$ 192 mm, $1.5 \times 1.5 \times 1.5$ mm^3 voxels, TR/TE=5000/5ms, 128 DWI at b=1000 s/mm^2 and 15 S_0 images. The dataset has an acquisition matrix of size $128 \times 128 \times 84$ and 100 uniformly distributed diffusion directions, but for computational purposes, we consider the 37th to 41st z-slices only and present the T_2 (S_0) and FA images for the 38th z-slice. Given that the fibers are constructed with synthetic materials, the FA should be relatively constant along the fibers. A goal of denoising is to decrease the standard deviation of the FA along a fiber, indicating that the FA along a fiber lies within a smaller range of values and hence is more constant. After denoising using our proposed model, the standard deviation of FA along a fiber drops from 0.1503 to 0.1292. However, the mean FA drops slightly, which may be a result of oversmoothing (see Table 3.3). As a result, the denoised signal implies a less directed diffusion along the fiber.

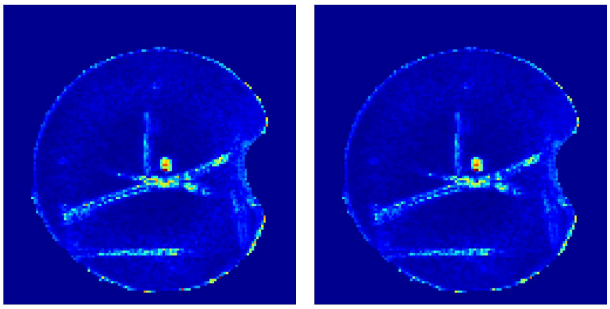
	mean	standard deviation
noisy	0.2392	0.1503
denoised	0.2140	0.1292

Table 3.3: Mean and standard deviation values of FA along a fiber for noisy and denoised phantom data.

For the data represented in Fig. 3.12, a 3T scanner was used to collect the DW-MRI data of the phantom. Data collection was again performed at the CMRR, University of Minnesota, on a



S_0 (T_2)



FA of noisy data FA of denoised data

Figure 3.11: Visualizations of 38th slice of phantom data.

Magnex Scientific MR imaging scanner driven by a Siemens console, with a head gradient insert capable of 80 mT/m in 200 ms. Parameter settings were: 64 slices with FOV = 192×192 mm, $2 \times 2 \times 2$ mm^3 voxels, TR/TE = 7800/82 ms, 128 DWI at $b = 150$ s/mm^2 and 17 S_0 images. For this particular dataset, all raw DWI's were corrected for eddy current distortion using the *eddy_correct* function from the FSL toolbox (http://fsl.fmrib.ox.ac.uk/fsl) [SJB04, WJP09]. Geometric distortions due to magnetic susceptibility were then corrected using a field map collected just prior to the DWI's, using the FSL *prelude* and *fugue* functions. Diffusion Tensor Imaging (DTI) tractography was conducted via the Diffusion Toolkit (http://trackvis.org/dtk/) [WBS07] using the critical angle threshold = 20° and the 2$^{\text{nd}}$ order Runge-Kutta method [BPP00]. The noisy and denoised fiber and fractional anisotropy visualizations of the phantom data are displayed in Fig. 3.12. After denoising, the bottom right of the fiber is visible.

Lastly, we denoise the same real HARDI brain data that we have described earlier (Figures 3.3 and 3.4), but in this case, we use ODFs and corresponding JSDs as metrics of evaluation. We present in Fig. 3.13 the ODFs of the "ground truth", noisy, and denoised 30th z-slice of

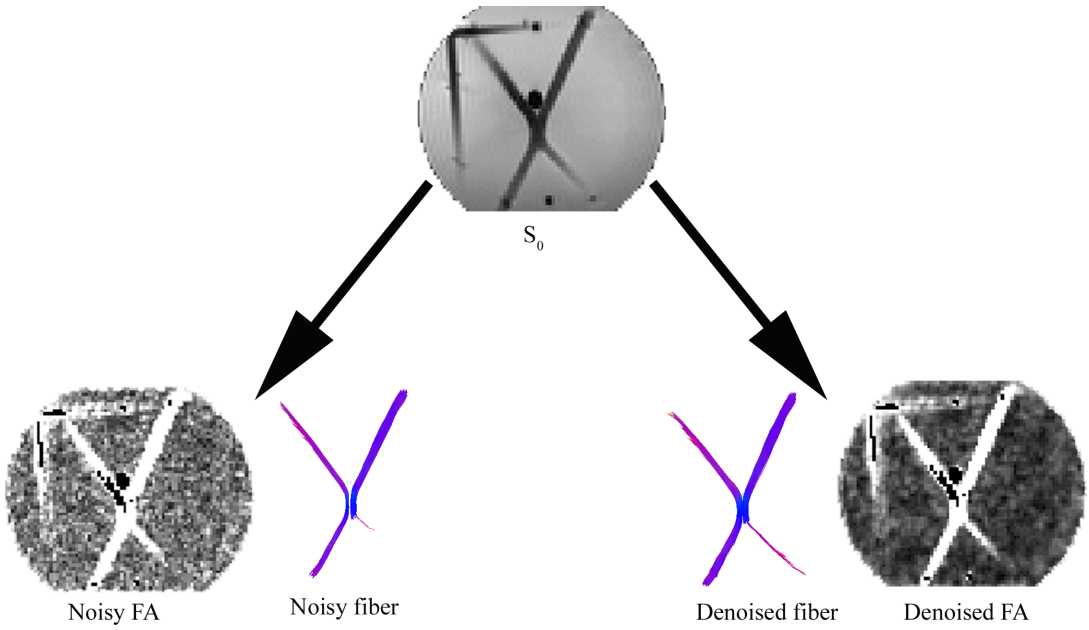


Figure 3.12: Left: noisy fiber and fractional anisotropy visualization. Right: denoised fiber and fractional anisotropy visualization. Note that the bottom right of the fiber is visible in the denoised data but not in the noisy data.

the real HARDI brain data. Fig. 3.14 shows the JSD between the ODF of the noisy data and ODF of the "ground truth" data, and we see that the ODFs of the denoised data and "ground truth" data are more similar than the ODFs of the noisy data and "ground truth" data. This is reflected also in the drop in the mean JSD values from 0.5787 to 0.3297 after denoising.

3.4.3 Extension: Vectorial Version of Rician Denoising Model

We extend the variational denoising model proposed in Chapter 2 to the vectorial case:

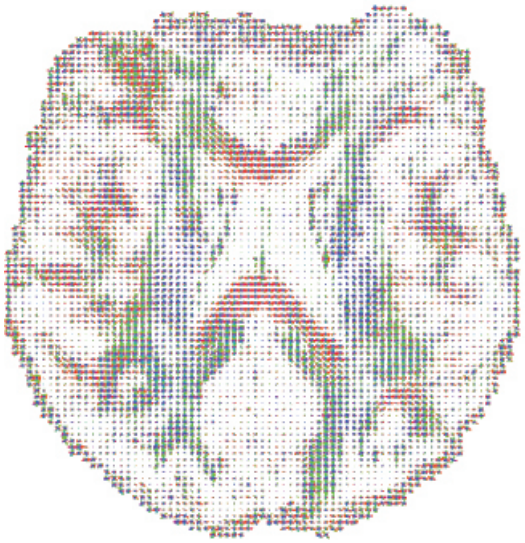
$$\inf_{u\in BV(\Omega)} \Big\{ F(u) = \int_\Omega |Du| + \lambda \int_\Omega \sum_{i=1}^{n} \Big[-\log I_0(\frac{S_i u_i}{\sigma^2}) + \frac{S_i^2 + u_i^2}{2\sigma^2} \Big] dx \Big\}, \tag{3.13}$$

where S_i and u_i denote the noisy MR signal and denoised MR signal in direction i, respectively. σ is again the standard deviation of the noise, and n is the number of diffusion directions.

We perform numerical experiments on the real HARDI brain data introduced above. Fig. 3.15 and Fig. 3.16 give visualizations of the denoising results on the same data set as in Fig. 3.5 and Fig. 3.6. In addition to visual improvements, we see a significant reduction in the RMSE from

ODF of clean data

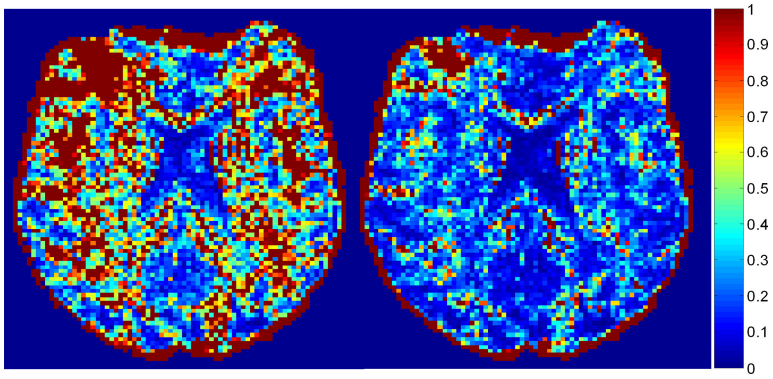
ODF of noisy data



ODF of denoised data

Figure 3.13: ODF visualizations of the 30th z-slice of the real HARDI brain data. Top: ODF of "ground truth" data. Left: ODF of noisy data. Right: ODF of denoised data. The color in this figure indicates the fiber direction: red for left-right, blue for superior-inferior, and green for anterior-posterior.

10.5454 for the noisy data to 6.6475 for the denoised data. Table 3.4 gives a comparison of the perfomances (in terms of RMSE) between the model proposed in [KTV10], our variational denoising model in d (3.3), and the extension of the Rician denoising model to the vectorial



JSD of noisy/clean data JSD of denoised/clean data

Figure 3.14: JSD between ODFs of noisy and "ground truth" data (mean = 0.5787, std. dev. = 0.3445) and between ODFs of denoised and "ground truth" data (mean = 0.3297, std. dev. = 0.3037) for the 30th slice of the real HARDI brain data. In the colorbar, we have values from 0 (blue) to 1 (red).

case (3.13). Though the RMSE and JSD values corresponding to (3.13) are lower than those corresponding to (3.3), there are many places where the constraint $d \geq 0$ is not satisfied when using (3.13).

As in Fig. 3.13, we present the ODFs of the "ground truth", noisy, and denoised (using (3.13)) 30th z-slice of the real HARDI data of the brain. Visual improvements in the ODFs are present in Fig. 3.17 and the decrease in JSD values can be seen in Fig. 3.18.

	RMSEdenoised/RMSEnoisy
TV+L^1 fidelity denoising in d [KTV10]	4.7268/10.5448 = **0.4483**
our model denosing in d + projection	8.2289/10.5454 = 0.7803
our model denosing in u	6.6475/10.5454 = 0.6304

Table 3.4: Comparison of RMSEdenoised/RMSEnoisy ratios for denoising real HARDI brain data using different methods.

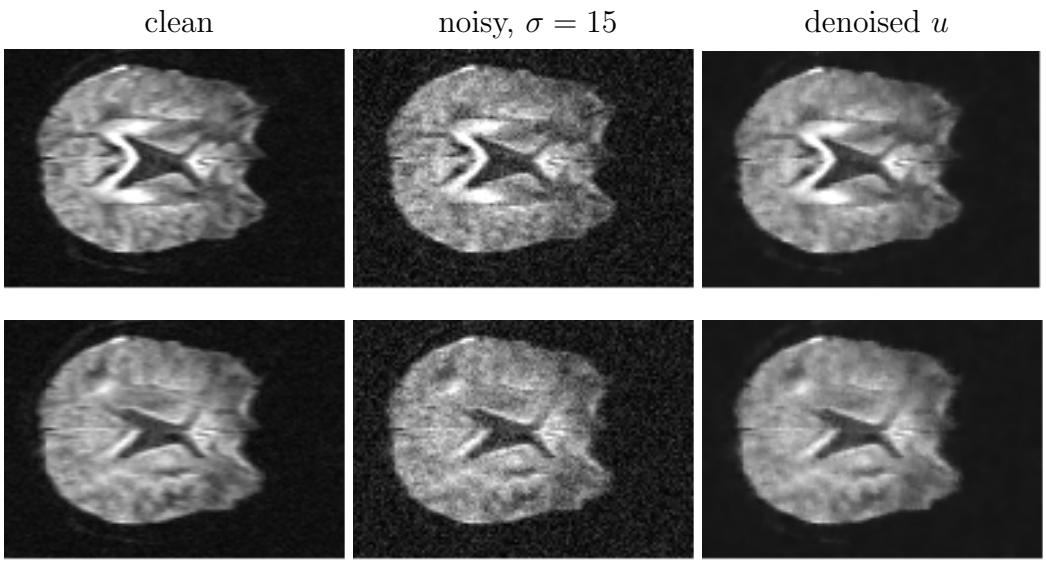


Figure 3.15: Visualization of 2 of 30 diffusion-sensitized MR signals of the real HARDI brain data. Left to right: clean, noisy, denoised data using vectorial version of Rician model (3.13). $d_{\text{denoised}} < 0$ in many places. RMSE of denoised data: 6.6475. RMSE of noisy data: 10.5454.

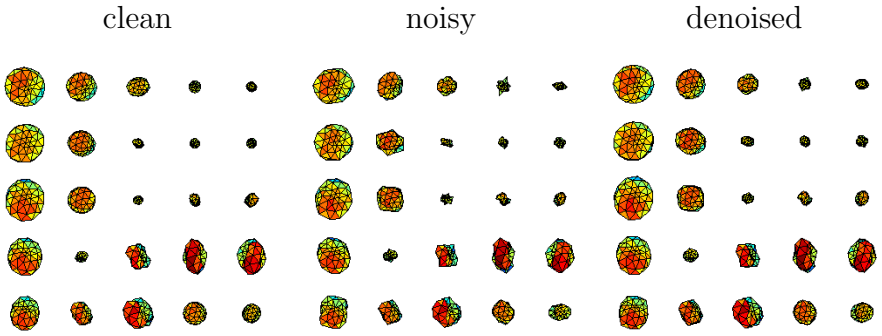
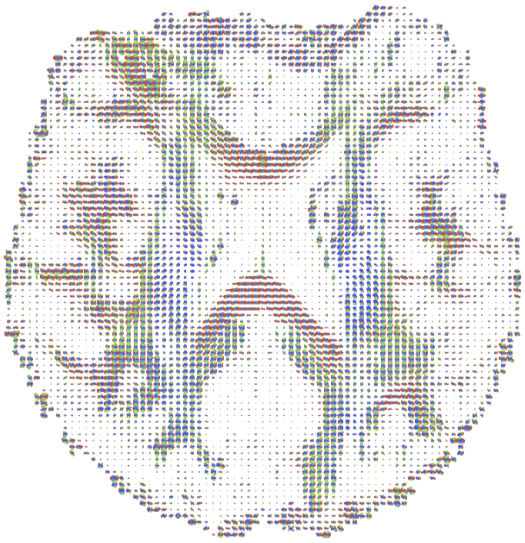


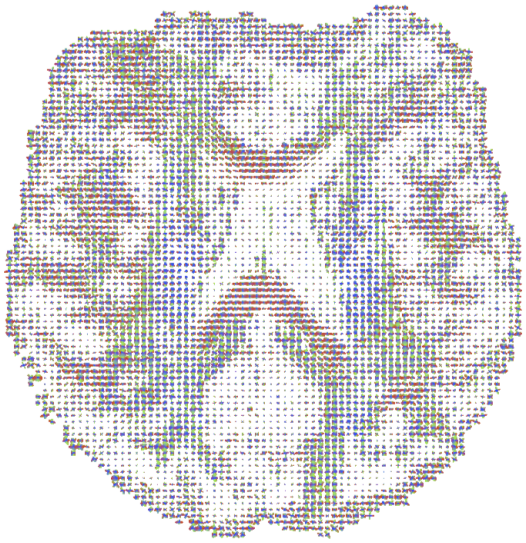
Figure 3.16: Visualization of small portion of clean, noisy, denoised data using vectorial version of Rician model (3.13). $d_{\text{denoised}} < 0$ in many places. RMSE of denoised data: 6.6475. RMSE of noisy data: 10.5454.

3.5 Discussion

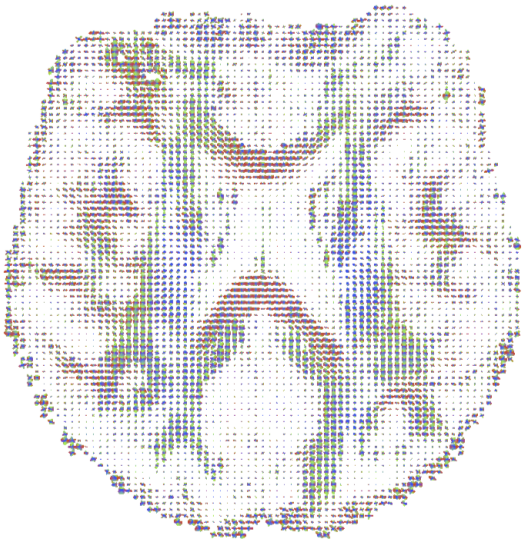
We have performed numerical experiments on multiple types of data and used a variety of metrics to evaluate the performance of our proposed denoising model. We see that there are improvements with both visualizations as well as RMSE and JSD values. The result in Fig. 3.12 is promising for the applicability of the model for fiber tracking. In addition, the use of the



ODF of clean data

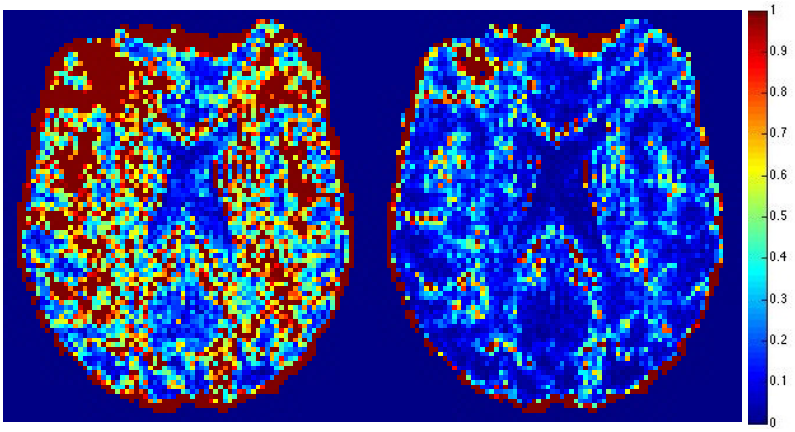


ODF of noisy data



ODF of denoised data

Figure 3.17: ODF visualizations of the 30th z-slice of the real HARDI brain data. Top: ODF of "ground truth" data. Left: ODF of noisy data. Right: ODF of denoised data. The color in this figure indicates the fiber direction: red for left-right, blue for superior-inferior, and green for anterior-posterior.



JSD of noisy JSD of denoised

Figure 3.18: JSD between ODFs of noisy and "ground truth" data (mean = 0.5787, std. dev. = 0.3445) and between ODFs of denoised and "ground truth" data (mean = 0.2523, std. dev. = 0.2750). In the colorbar, we have values 0 (blue) to 1 (red).

projection operator seems to help enforce the constraint $d \geq 0$. The extension of the Rician denoising model to the vectorial case seems to perform better in terms of RMSE and JSD values, but the constraint $d \geq 0$ is not enforced and thus violated in many places. For future work, we may try different regularization terms such as nonlocal regularization or learned priors.

Part II

Restoration of Images Corrupted by Atmospheric Turbulence

CHAPTER 4

Joint Variational Model for Atmospheric Distortion Correction

4.1 Introduction and Background

Atmospherically-distorted images of a static scene arise in long range imaging where the images are distorted by turbulent geometric distortion and blurring effects during their acquisition. A model for this degradation is presented in Frakes et. al. [FMS01, GSF04]

$$f_i(x) = \Phi_i(K(u(x))) + \text{noise}, \tag{4.1}$$

where a static undistorted scene $u(x)$ is distorted with blurring effects modeled by blur kernel K and geometric distortion effects represented by the operator Φ_i to yield respective distorted frames f_i. The goal is to recover the static scene u from a stream of distorted image frames of the static scene.

In the model of turbulence (4.1), the distorted image f_i is generated by first blurring the static image u and then degrading the resulting blurry image with geometric distortions. The authors of [GDF08] study both this case and the case where the static image is first affected with geometric distortions and then with blur (i.e. $f_i(x) = K(\Phi_i(u(x))) + \text{noise}$). In their work, they use temporal filtering in combination with registration to correct for geometric distortions and a blind deconvolution algorithm to correct for blurring effects.

In the work [LMS07], the authors formulate blind image deconvolution as a principal components analysis (PCA) problem, and they perform restoration experiments on atmospheric turbulence-degraded imagery. Whereas in [TLF08], the focus is more on geometric distortion correction. The authors use a Kalman filter to recover the static scene from a series of distorted

frames. They assume a high frame rate and use time-dependent differential equations to model the warping of the frames.

In a different type of approach, the authors of [AVC09] present an improved "lucky-region" fusion (LRF) approach. The LRF approach estimates the local quality of images using an image quality map, which is often based on the gradient of the image. The image quality map selects the best quality regions of each image, and these "lucky-regions" are fused together to give the restored static image.

In a very recent work [LKS12], the authors propose a method that performs joint frame sharpening with the Sobolev gradient method and temporal distortion correction using the Laplace operator. With the reconstructed frames, they apply an approach similar to the lucky-region fusion approach to reconstruct the static image.

In the work [MG12], the authors address specifically the geometric distortions caused by atmospheric turbulence. They start with a reference frame that is a good approximation of the static scene (usually the mean of the input frames) and estimate the optical flows from this reference frame to each of the input frames. Once the optical flows are determined, they are used to determine a new reference frame, where this new reference frame is the solution of a variational problem involving nonlocal TV regularization. Once this new reference frame is found, the process repeats. A geometrically corrected image results after a few iterations of this process.

A recent approach to deblur the effects of atmospheric turbulence is proposed in [GO12], where the authors utilize the Fried kernel [Fri66] in a framelet based deconvolution algorithm. The Fried kernel is an analytical formulation of the atmosphere modulation transfer function (MTF) and depends on parameters of the acquisition system and characteristics of the imaging scene as well as a refractive index structure which reflects the turbulence level in the atmosphere. In their work, a method to estimate this refractive index structure parameter is provided. In the next section, we give a brief review of the Fried kernel.

The last two prior works have been combined to produce very nice results. First, the input frames are used to produce a geometrically corrected image using [MG12]. The geometrically

corrected result is then used as input for the deconvolution algorithm in [GO12]. This two step process yields a deblurred and geometrically corrected image. Our goal is to combine both the deconvolution and geometric correction into one variational restoration model. We give preliminary results for this model.

4.1.1 Review of Fried Kernel

We give here the basic form of the MTF of the Fried kernel to illustrate its behavior as a function of four parameters D, L, λ, and C_n^2. For more details, we refer the reader to [Fri66, GO12]. In two dimensions, letting ω be the frequency modulus, Fried's MTF $M_F(\omega)$ (in the Fourier domain) is given by

$$M_F(\omega) = M_0(\omega) M_{\mathrm{SA}}(\omega) \tag{4.2}$$

where

$$M_0(\omega) = \begin{cases} \frac{2}{\pi}(\arccos(\omega) - \omega\sqrt{1-\omega^2} & \text{for } \omega < 1 \\ 0 & \text{for } \omega > 1, \end{cases} \tag{4.3}$$

and

$$M_{\mathrm{SA}}(\omega) = \exp\left\{ -(2.1X)^{5/3}(\omega^{5/3} - V(Q,X)\omega^2) \right\}. \tag{4.4}$$

$k = \frac{2\pi}{\lambda}$, $r_0 = 2.1\rho_0 = 2.1(1.437(k^2 L C_n^2))^{-3/5}$, $Q = \frac{D}{\sqrt{\lambda L}}$, $X = \frac{D}{r_0}$, and

$$V(Q,X) = \mathrm{A}(Q) + \frac{\mathrm{B}(Q)}{10} \exp\left\{ -\frac{(\log_{10}(X)+1)^3}{3.5} \right\} \tag{4.5}$$

where A and B depend on Q only.

Here, D is the system entrance pupil diameter, L is the path length given by the distance between the sensor and acquired scene, λ is the wavelength on which the imaging system is working, and C_n^2 is the refractive index structure reflecting the turbulence level of the atmosphere (for more information on C_n^2, we refer the reader to [TTV05]). As measured in [TTV05], C_n^2 is typically in the range $[10^{-16}m^{-2/3}, 10^{-12}m^{-2/3}]$ where larger values of C_n^2 correspond to stronger turbulence.

4.1.2 Geometric Distortion Operator and Computation of Its Adjoint

For the notation of the geometric distortion operator and the computation of its adjoint, we follow the notations and method presented in [MG12]. Let $g = g(x_1, x_2)$, $v_1 = v_1(x_1, x_2)$ and $v_2 = v_2(x_1, x_2)$ be functions from $\mathbb{R}^2$ to $\mathbb{R}$. We define the geometric distortion operator Φ by

$$\Phi : g(x_1, x_2) \to g(x_1 + v_1(x_1, x_2), x_2 + v_2(x_1, x_2)). \tag{4.6}$$

For fixed v_1 and v_2, Φ is a linear operator on the space of functions from $\mathbb{R}^2$ to $\mathbb{R}$.

The adjoint of Φ is denoted Φ^T and is defined as the operator such that

$$\langle h, \Phi^T g\rangle = \int h(\Phi^T g)dx = \int (\Phi h)g dx = \langle \Phi h, g\rangle \quad \forall h. \tag{4.7}$$

Numerically, the authors of [MG12] take h in (4.7) to be the 'single spike function'

$$h_y(x) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x, \end{cases} \tag{4.8}$$

and subsequently arrive at the relation

$$(\Phi^T g)(y) = \langle h_y, \Phi^T g\rangle = \langle \Phi h_y, g\rangle. \tag{4.9}$$

$\langle \Phi h_y, g\rangle$ is easy to evaluate since Φh_y is a simple function.

4.2 Proposed Variational Model

We propose two variations of a combined deblurring and geometric distortion correction model. The two variations correspond to the relations

$$\Phi_k(f^k) = Ku + \text{noise} \tag{4.10}$$

and

$$f^k = \Phi_k(Ku) + \text{noise} \tag{4.11}$$

where Φ_k represents the geometric distortion operator corresponding to the kth frame f^k.

4.2.1 Variation 1

The first model that we propose involves the relation

$$\Phi_k(f^k) = Ku + \text{noise}. \tag{4.12}$$

Using this relation, we propose the following minimization problem

$$\begin{aligned} \min_{v_1^k, v_2^k, u} \Big\{ E_1(u, v_1^k, v_2^k) &= \mu \sum_{k=1}^{\text{numFrames}} \int_\Omega (|\nabla v_1^k|^2 + |\nabla v_2^k|^2) dx \\ &+ \lambda \sum_{k=1}^{\text{numFrames}} \int_\Omega (Ku - \Phi_k(f^k))^2 dx + \gamma \int_\Omega |\nabla u| dx \Big\}, \end{aligned} \tag{4.13}$$

where f^k is the kth distorted frame of a true scene, K is the blur kernel, and Φ^k is the linear operator representing geometric distortions of the kth frame given by

$$\Phi_k(f^k(x_1, x_2)) = f^k(x_1 + v_1^k(x_1, x_2), x_2 + v_2^k(x_1, x_2)), \tag{4.14}$$

where each $v_1^k = v_1^k(x_1, x_2)$ and $v_2^k = v_2^k(x_1, x_2)$, and $v^k = (v_1^k, v_2^k)$ represents the turbulence warping from f^k to Ku.

The first term in E_1 is an H^1 regularization on v_1^k and v_2^k, which enforces a smooth turbulence warping. The second term in E_1 acts as a fidelity term that constrains the unknowns u, v_1^k and v_2^k to adhere to (4.12), and the last term in E_1 is simply the total variation (TV) regularization in u [ROF92, RO94], which allows for the restored image u to have edges.

To minimize the energy (4.14), we use Euler-Lagrange equations and alternating minimization in the unknowns. We give the associated gradient descent equations for u, v_1^k and v_2^k, where $k = 1, ..., \text{numFrames}$.

For u:

$$\frac{\partial u}{\partial t} = -\frac{\partial E_1}{\partial u} = \lambda \sum_{k=1}^{\text{numFrames}} \Big\{ -2K^*(Ku - \Phi_k(f^k)) \Big\} + \gamma \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right).$$

For v_1^k:

$$\frac{\partial v_1^k}{\partial t} = -\frac{\partial E_1}{\partial v_1^k} = 2\mu \triangle v_1^k + 2\lambda (Ku - \Phi_k(f^k))(\Phi_k(f_x^k)).$$

For v_2^k:

$$\frac{\partial v_2^k}{\partial t} = -\frac{\partial E_1}{\partial v_2^k} = 2\mu \triangle v_2^k + 2\lambda (Ku - \Phi_k(f^k))(\Phi_k(f_y^k)).$$

The above equations are discretized using finite differences, and a fully explicit scheme is used to update the unknowns. We start with initial guess

$$u_0 = \text{constant} = \underset{(x_1,x_2)\in\Omega, k=1:\text{numFrames}}{\text{mean}} f^k(x_1, x_2)$$

and $v_1^k = v_2^k = 0$. We perform one iteration of gradient descent at each minimization step, using the previous u, v_1^k and v_2^k's in the update process. After updating each of the unknowns, the process is repeated until the energy E_1 reaches a steady state.

4.2.2 Variation 2

Similar to the first model, the second model that we propose involves the relation

$$f^k = \Phi_k(Ku) + \text{noise}. \tag{4.15}$$

With this relation, we propose the following minimization problem

$$\begin{aligned} \min_{v_1^k, v_2^k, u} \Big\{ E_2(u, v_1^k, v_2^k) \;&=\; \mu \sum_{k=1}^{\text{numFrames}} \int_\Omega (|\nabla v_1^k|^2 + |\nabla v_2^k|^2) dx \\ &+\; \lambda \sum_{k=1}^{\text{numFrames}} \int_\Omega (\Phi_k(Ku) - f^k)^2 dx + \gamma \int_\Omega |\nabla u| dx \Big\} \end{aligned} \tag{4.16}$$

where again f^k is the kth distorted frame of the true scene, K is the blur kernel, and Φ^k is the linear operator representing geometric distortions of the kth frame given by

$$\Phi_k((Ku)(x_1, x_2)) = (Ku)(x_1 + v_1^k(x_1, x_2), x_2 + v_2^k(x_1, x_2)), \tag{4.17}$$

where each $v_1^k = v_1^k(x_1, x_2)$ and $v_2^k = v_2^k(x_1, x_2)$, and $v^k = (v_1^k, v_2^k)$ represents the turbulence warping from Ku to f^k.

We use the same regularizations in v_1^k, v_2^k and u as in our first proposed model, and the only difference between this second model and the first is that the fidelity term reflects the relation (4.15).

To minimize (4.17), we again use Euler-Lagrange equations and alternating minimization. We give here the associated gradient descent equations for unknowns u, v_1^k and v_2^k, for $k =$

$1, ..., \text{numFrames}$.

For u:

$$\frac{\partial u}{\partial t} = -\frac{\partial E_2}{\partial u} = \lambda \sum_{k=1}^{\text{numFrames}} \Big\{ -2K^*\Phi_k^T(\Phi_k(Ku) - f^k) \Big\} + \gamma \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right).$$

For v_1^k:

$$\frac{\partial v_1^k}{\partial t} = -\frac{\partial E_2}{\partial v_1^k} = 2\mu \triangle v_1^k - 2\lambda(\Phi_k(Ku) - f^k)(\Phi_k((Ku)_x))$$

For v_2^k:

$$\frac{\partial v_2^k}{\partial t} = -\frac{\partial E_2}{\partial v_2^k} = 2\mu \triangle v_2^k - 2\lambda(\Phi_k(Ku) - f^k)(\Phi_k((Ku)_y))$$

We discretize the above equations using finite differences, and using a fully explicit scheme, we update the unknowns. We start with initial guess

$$u_0 = \text{constant} = \underset{(x_1,x_2)\in\Omega, k=1:\text{numFrames}}{\text{mean}} f^k(x_1, x_2)$$

and $v_1^k = v_2^k = 0$. We perform one iteration of gradient descent at each minimization step, using previous u, v_1^k and v_2^k's in update process. After updating each of the unknowns, we repeat the process until the energy E_2 reaches a steady state.

4.3 Numerical Experiments

In this section, we provide some preliminary numerical results. We compare the results of our proposed variational model with the work found in [MG12] on geometric distortion correction and [GO12] on deconvolution using Fried kernel. In all examples, only 10 frames were used in the reconstructions.

As shown in [Gil07], applying the temporal mean or median filter on the input frames often give a good reference image, with the temporal median producing a less blurred result than the temporal mean. In our results, we include the temporal mean and median of the input frames for comparison.

We begin with the simple case of $K = I$ where the blur kernel is simply the identity, representing the case of geometric distortion only and no blurring effects. The data was generated synthetically, and three sample frames of the geometric distortion are displayed to give the reader a sense of the magnitude of the distortion (top row of Fig. 4.1).

In the second case, we consider joint geometric distortion correction and deblurring, taking K to be the Fried kernel and look at the effect of the refractive index structure C_n^2 on the restored result. The data was collected by NATO SET156 (ex-SET072 RTG40) Group during the 2005 New Mexico's field trials. Three sample frames of each of the data sets are given (middle and bottom rows of Fig. 4.1).

Figure 4.1: Three sample frames of distorted data.

4.3.1 Geometric Distortion Correction, Case: $K = I$

We begin by considering synthetic data that models turbulent geometric distortion without blur. The true image is 256×256, and we consider 10 distorted frames for our reconstruc-

tions. In Fig. 4.2, we present for comparison, the true image, the mean of the input frames, the median of the input frames, the geometric distortion corrected image using the algorithm in [MG12], and our restored results. Both Variation 1 and 2 of the proposed model corrects for the geometric distortion well, but the reconstructed images are not as sharp as the true image. The reconstruction using [MG12] is sharp, but fails to reconstruct parts of the image as well as our proposed model (e.g. the eyes). Furthermore, Variation 2 of our proposed model provides a sharper reconstruction than that of Variation 1. We would like to mention that the reconstructions using [MG12] utilize the nonlocal total variation regularization [GO08], and the reconstructions using our proposed model utilize the local total variation regularization. A more thorough comparison will be made in the future.

Figure 4.2: Geometric distortion correction. Top, left to right: true image, mean of 10 input frames, median of 10 input frames. Bottom, left to right: Variation 1 reconstruction, Variation 2 reconstruction, geometric distortion correction using [MG12].

4.3.2 Joint Deblurring and Geometric Distortion Correction, K Fried Kernel

In this section, we present joint deblurring and geometric distortion correction numerical examples. We take K to be the Fried kernel. We consider two data sets (middle and bottom rows of Fig. 4.1). In both restorations, we use only 10 frames.

With the first example, we performed two restorations; one restoration uses the measured value of $C_n^2 = 1.51 \times 10^{-13}$ and the second uses the estimated $C_n^2 = 2.5 \times 10^{-13}$, which was found using the algorithm in [GO12]. Recall that a larger C_n^2 value creates a stronger blur kernel since it corresponds to higher levels of turbulence. Fig. 4.3 shows our joint deblurring and geometric distortion correction results along with a comparison with the geometrically corrected image using [MG12], and the deblurred results of the geometrically corrected image using [GO12] with Fried kernel corresponding to the two C_n^2 values. The second row of Fig. 4.3 corresponds to $C_n^2 = 1.51 \times 10^{-13}$. The restored images using Variation 1 and Variation 2 of our model are very similar; the result using Variation 2 is slightly sharper (see the lower part of the three bars farthest to the right). When the larger estimated value of $C_n^2 = 2.5 \times 10^{-13}$ is used (see bottom row of Fig. 4.3), no difference can be detected between the restored images using Variation 1 and Variation 2. The restored images using first the geometric distortion correction [MG12] and then deconvolution with [GO12] give a more constant intensity along some of the bars (see the top of the bar second from the right) but overall are similar to the restored images using our proposed models.

With our second example, we performed restoration using the measured $C_n^2 = 1.91 \times 10^{-13}$ for our joint deblurring and geometric distortion correction models. Fig. 4.4 gives our restored images as well as a comparison with the geometrically corrected image using [MG12] and the blind deconvolution of this result using [GO12], where the approximated value of $C_n^2 = 1.7 \times 10^{-13}$. The restored images using our proposed models are similar, with Variation 2 giving a slightly sharper restored image (the top loop of the 'B' in 'ALBEDOS'). The result using [MG12] and [GO12] performs better in certain areas (the 'D' in 'ALBEDOS' looks nicer than in our restored images), but results using our model performs better in other areas (the 'A' and 'L' are more separated in 'ALBEDOS').

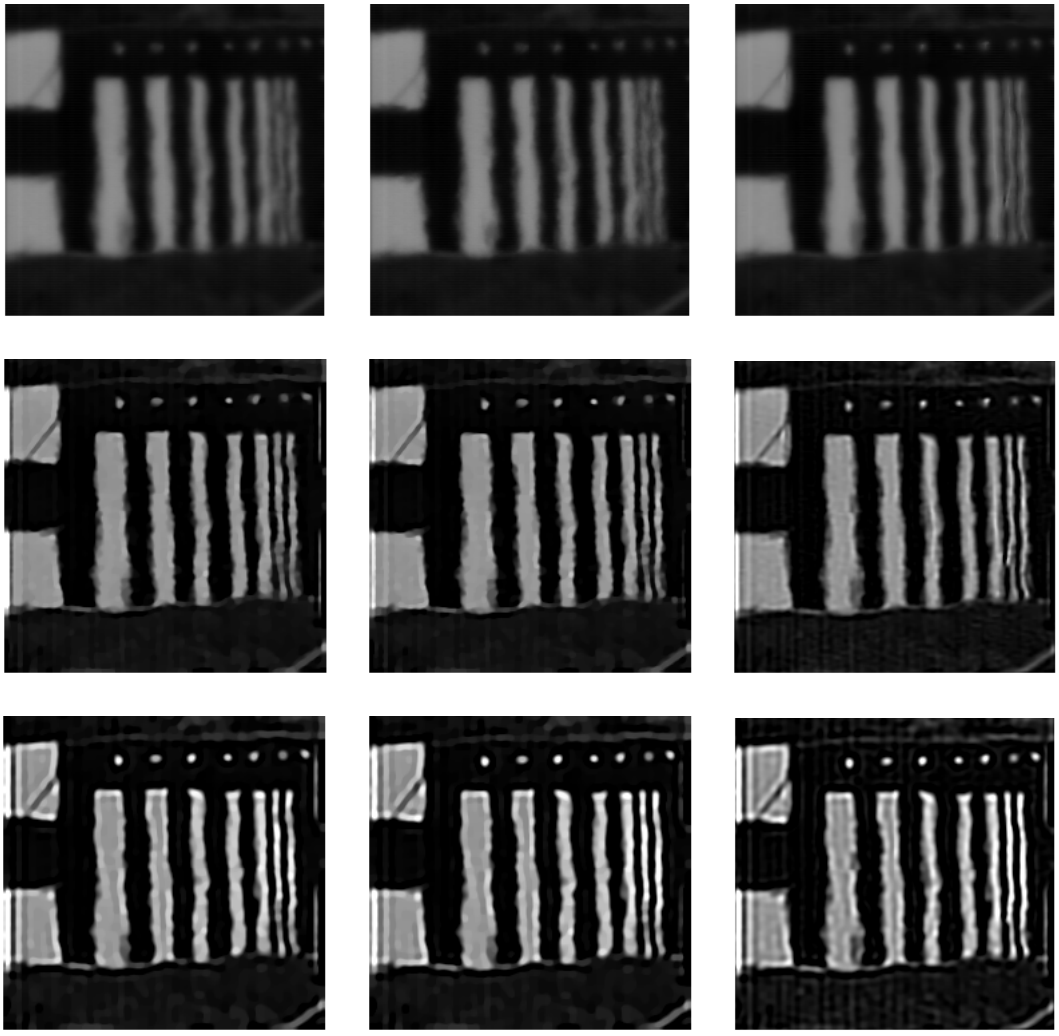


Figure 4.3: Top, left to right: mean of 10 input images, median of 10 input images, geometric distortion correction using [MG12]. Middle, left to right: reconstructions using Variation 1, Variation 2, and framelet non-blind deconvolution of geometrically corrected image (top right) using algorithm in [GO12]; here measured $C_n^2 = 1.51 \times 10^{-13}$ is used. Bottom, left to right: reconstructions using Variation 1, Variation 2, and blind deconvolution of geometrically corrected image (top right) using algorithm in [GO12]; here approximated value of $C_n^2 = 2.5 \times 10^{-13}$ is used.

4.4 Discussion

In the future, we will perform more numerical experiments to gain a better understanding of the performance of our proposed joint deblurring and geometric distortion correction models

Figure 4.4: Top, left to right: mean of 10 input images, median of 10 input images, geometrically restored image using [MG12]. Bottom, left to right: Variation 1 and Variation 2 reconstructions using measured $C_n^2 = 1.91 \times 10^{-13}$, blind Fried deconvolution of geometrically restored (top right) using [GO12] algorithm with estimated $C_n^2 = 1.7 \times 10^{-13}$.

and to see if any improvements can be achieved using a combined model. For a more equal comparison with [MG12], we will implement the nonlocal total variation [GO08] in place of the total variation. In addition, to increase performance speed, we will utilize the Bregman iterative method [OBG05].

References

[AA08] G. Aubert and J.-F. Aujol. "A variational approach to remove multiplicative noise." *SIAM Journal on Imaging Sciences*, 2008.

[ABM06] H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces: Applications to PDEs and optimization.* MPS-SIAM Series on Optimization, 2006.

[AK06] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: Partial differential equations and the calculus of variations.* Springer, 2006.

[And01] A.W. Anderson. "Theoretical analysis of the effects of noise on diffusion tensor imaging." *MRM*, **46**:1174–1188, 2001.

[AS65] M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, pp. 355–379. New York: Dover, 1965.

[AVC09] M. Aubailly, M.A. Vorontsov, G.W. Carhart, and M.T. Valley. "Automated video enhancement from a stream of atmospherically-distorted images: the lucky-region fusion approach." *Proc. of SPIE*, **7463**, 2009.

[BFW06] S. Basu, T. Fletcher, and R. Whitaker. "Rician noise removal in diffusion tensor MRI." *LNCS 4190 MICCAI*, pp. 117–125, 2006.

[Boa83] M. Boas. *Mathematical methods in the physical sciences.* John Wiley and Sons, Inc, 1983.

[BPP00] P.J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. "In vivo fiber tractography using DT-MRI data." *Magnetic Resonance in Medicine*, **44(4)**:625–632, 2000.

[BS08] L. Bar and G. Sapiro. "Generalized Newton method for energy formulations in image processing." *IMA preprint series 2195*, 2008.

[Cha04] A. Chambolle. "An algorithm for mean curvature motion." *Interfaces and Free Boundaries*, **6**:1–24, 2004.

[CMP07] G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, and O. Faugeras. "Generalized gradients: Priors on minimization flows." *IJCV*, **73(3)**:325–344, 2007.

[CS05] T.F. Chan and J. Shen. *Image processing and analysis.* SIAM, 2005.

[CUF09] M. Cihangiroglu, A. Mufit Ulug, Z. Firat, A. Bayram A. Kovanlikaya, and I. Kovanlikaya. "High b-value diffusion-weighted MR imaging of normal brain at 3T." *European Journal of Radiology*, 2009.

[DAF07] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. "Regularized, fast and robust analytical q-ball imaging." *Magnetic Resonance in Medicine*, **58**:497–510, 2007.

[DMN08] G. Dogan, P. Morin, and R. Nochetto. "A variational shape optimization approach for image segmentation with a Mumford-Shah functional." *SIAM Journal on Scientific Computing*, **30(6)**:3028–3049, 2008.

[DWP] M. Descoteaux, N. Wiest-Daessle, S. Prima, C. Barillot, and R. Deriche. "Impact of Rician adapted non-local means filtering on HARDI." *MICCAI 2008, Part II, LNCS 5242*, pp. 122–130.

[EG92] L.C. Evans and R.F. Gariepy. *Measure theory and fine properties of functions*. CRC-Press, 1992.

[FMS01] D. Frakes, J. Monaco, and M. Smith. "Suppression of atmospheric turbulence in video using an adaptive control grid interpolation approach." *IEEE International Conference on Acoustics, Speech, and Signal Processing*, **3**:1881–1884, 2001.

[Fra02] L.R. Frank. "Characterization of anisotropy in high angular resolution diffusion-weighted MRI." *MRM*, **47**:1083–1099, 2002.

[Fri66] D.L. Fried. "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures." *Journal of The Optical Society of America*, **56**:1372–1379, October 1966.

[GDF08] J. Gilles, T. Dagobert, and C.D. Franchis. "Atmospheric turbulence restoration by diffeomorphic image registration and blind deconvolution." *Proceedings of the 10th International Conference on Advanced Concepts for Intelligent Vision Systems*, 2008.

[Gil07] J. Gilles. "Restoration algorithm and system performance evaluation for active imaging systems." In *SPIE Europe Remote Sensing Conference*, Florence, Italy, September 2007.

[Giu94] E. Giusti. *Minimal surfaces and functions of bounded variation*. Birkhauser, 1994.

[GO08] G. Gilboa and S. Osher. "Nonlocal operators with applications to image processing." *Multiscale Model Sim*, **7**:1005–1028, 2008.

[GO09] T. Goldstein and S. Osher. "The split Bregman method for L1 regularized problems." *SIAM Journal on Imaging Sciences*, **2**:323–343, 2009.

[GO12] J. Gilles and S. Osher. "Fried Deconvolution." In *SPIE Defense, Security and Sensing conference*, Baltimore, US, April 2012.

[GP95] H. Gudbjartsson and S. Patz. "The Rician distribution of noisy MRI data." *MRM*, **34**:910–914, 1995.

[GSF04] S. Gepshtein, A. Shteinman, and B. Fishbain. "Restoration of atmospheric turbulent video containing real motion using rank filtering and elastic image registration." *Proceedings of the Eusipco*, January 2004.

[GTV11] P. Getreuer, M. Tong, and L. Vese. "A variational model for the restoration of MR images corrupted by blur and Rician noise." *Proceedings of ISVC, Part I, LNCS 6938*, pp. 686–698, 2011.

[Hen85] R.M. Henkelman. "Measurement of signal intensities in the presence of noise in MR images." *Med. Phys.*, **12**:232–233, 1985.

[JCS09] M. Jung, G. Chung, G. Sundaramoorthi, L.A. Vese, and A.L. Yuille. "Sobolev gradients and joint variational image segmentation, denoising and deblurring." *UCLA C.A.M. Report 09-07*, Feb 2009.

[KDA99] P. Kornprobst, R. Deriche, and G. Aubert. "Image sequence analysis via partial differential equations." *Journal of Mathematical Imaging and Vision*, **11(1)**:5–26, 1999.

[KL51] S. Kullback and R.A. Leibler. "On information and sufficiency." *Annual Math. Stat.*, **22**:79–86, 1951.

[KTT09] Y. Kim, P.M. Thompson, A.W. Toga, L. Vese, and L. Zhan. "Hardi denoising: variational regularization of spherical apparent diffusion coefficients, sADC." *IPMI, LNCS 5636*, pp. 515–527, 2009.

[KTV10] Y. Kim, P.M. Thompson, and L. Vese. "Hardi data denoising using vectorial total variation and logarithmic barrier." *IPI, Special Issue in Medical Image Analysis*, **4**:273–310, 2010.

[Lat83] B.P. Lathi. *Modern digital and analog communication systems.* Holt-Saunders International Edition, 1983.

[LCD08] C. Lenglet, J.S.W. Campbell, M. Descoteaux, G. Haro, P. Savadjiev, D. Wassermann, A. Anwander, R. Deriche, G.B. Pike, G. Sapiro, K. Siddiqi, and P.M. Thompson. "Mathematical methods for diffusion MRI processing." *NeuroImage, Special Issue on Mathematics in Brain Imaging*, 2008.

[LKS12] Y. Lou, S.H. Kang, S. Soatto, and A.L. Bertozzi. "Video stabilization of atmospheric turbulence distrotion." *UCLA C.A.M. Report 12-30*, April 2012.

[LMS07] D. Li, R.M. Mersereau, and S. Simske. "Atmospheric turbulence-degraded image restoration using principal component analysis." *IEEE Geoscience and Remote Sensing Letters*, **4**:340–344, July 2007.

[LZZ09] A.D. Leow, S. Zhu, L. Zhan, K. McMahon, G.I. de Zubicaray, M. Meredith, M.J. Wright, A.W. Toga, and P.M. Thompson. "The tensor distribution function." *MRM*, **61**:205–214, 2009.

[MG12] Y. Mao and J. Gilles. "Non rigid geometric distortions correction - application to atmospheric turbulence stabilization." *to appear in Inverse Problems and Imaging*, 2012.

[MRD11] O. Michailovich, Y. Rathi, and S. Dolui. "Spatially regularized compressed sensing for high angular resolution diffusion imaging." *IEEE Transactions of Medical Imaging*, **30**:1100–1115, 2011.

[MVO09] T. McGraw, B. Vemuri, E. Özarslan, Y. Chen, and T. Mareci. "Variational denoising of diffusion weighted MRI." *IPI*, **3**:625–648, 2009.

[Neu97] J.W. Neuberger. "Sobolev gradients and differential equations." *Springer Lecture Notes in Mathematics*, **1670**, 1997.

[OBG05] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. "An iterative regularization method for total variation-based image restoration." *Multiscale Model Sim*, **4(2)**:460–489, Jan 2005.

[PB96] C. Pierpaoli and P. Basser. "Toward a quantitative assessment of diffusion anisotropy." *MRM*, **36**:893–906, 1996.

[PRG10] P. Pullens, A. Roebroeck, and R. Goebel. "Ground truth hardware phantoms for validation of diffusion-weighted MRI applications." *JMRI*, **3**:482–488, April 2010.

[PST11] V. Patel, Y. Shi, P.M. Thompson, and A.W. Toga. "K-SVD for HARDI denoising." *Proceedings of ISBI*, pp. 1805–1808, 2011.

[Ren06] R.J. Renka. "A simple explanation of the Sobolev gradient method." `http://www.cse.unt.edu/~renka/papers/sobolev.pdf`, 2006.

[Ric44] S.O. Rice. "Mathematical analysis of random noise." *Bell Syst. Tech. J.*, **23**:282–444, 1944.

[Ric08] W.B. Richardson. "Sobolev gradient preconditioning for image processing PDEs." *Commun. Numer. Methods in Engineering*, **24(6)**:493–504, 2008.

[RO94] L. Rudin and S. Osher. "Total variation based image restoration with free local constraints." *Proc. IEEE ICIP*, **1**:31–35, 1994.

[ROF92] L.I. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms." *Physica D*, **60**:259–268, 1992.

[SDD98] J. Sijbers, A.J. Den Dekker, D. Van Dyck, and E. Raman. "Estimation of signal and noise from Rician distributed data." *Proceedings of the International Conference on Signal Processing and Communications*, pp. 140–142, 1998.

[SJB04] S.M. Smith, M.W. Jenkinson, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Drobnjak, D.E. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, and P.M. Matthews. "Advances in functional and structural MR image analysis and implementation as FSL." *NeuroImage*, **23(S1)**:208–219, 2004.

[SLN00] S. Skare, T. Li, B. Nordell, and M. Ingvar. "Noise considerations in the determination of diffusion tensor anisotropy." *MRM*, **18**:659–669, 2000.

[ST65] E.O. Stejskal and J.E. Tanner. "Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient." *JCP*, **42**:288–292, 1965.

[SYM07] G. Sundaramoorthi, A. Yezzi, and A.C. Mennucci. "Sobolev active contours." *IJCV*, **73(3)**:345–366, 2007.

[TKZ12] M. Tong, Y. Kim, L. Zhan, G. Sapiro, C. Lenglet, B. Mueller, P.M. Thompson, and L.A. Vese. "Variational model for denoising high angular resolution diffusion imaging." *Proceedings of IEEE Int. Symp. Biomed. Imaging (ISBI)*, pp. 530–533, 2012.

[TLF08] M. Tahtali, A. Lambert, and D. Fraser. "Self-tuning Kalman filter estimation of atmospheric warp." *Proceedings of SPIE*, Jan 2008.

[TRW02] D.S. Tuch, T.G. Reese, M.R. Wiegell, N. Makris, J.W. Belliveau, and V.J. Wedeen. "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity." *MRM*, pp. 577–582, 2002.

[TTV05] A. Tunick, N. Tikhonov, M. Vorontsov, and G. Carhart. "Characterization of optical turbulence (Cn2) data measured at the ARL A LOT facility." *Army Research Lab, ARL-MR-625*, September 2005.

[TWB99] D.S. Tuch, R.M. Weisskoff, J.W. Belliveau, and V.J. Wedeen. "High angular resolution diffusion imaging of the human brain." *Proc. 7th Annual Meeting of ISMRM*, p. 321, 1999.

[WBS07] R. Wang, T. Benner, A.G. Sorensen, and V.J. Wedeen. "Diffusion toolkit: a software package for diffusion imaging data processing and tractography." *Proceedings of the International Society for Magnetic Resonance in Medicine*, p. 15, 2007.

[WJP09] M.W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S.M. "Bayesian analysis of neuroimaging data in FSL." *NeuroImage*, **45**:S173–186, 2009.

[WPC] N. Wiest-Daessle, S. Prima, P. Coupe, S.P. Morrissey, and C. Barillot. "Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI." *MICCAI 2008, Part II, LNCS 5242*, pp. 171–179.

[WZ06] Y. Wang and H. Zhou. "Total variation wavelet-based medical image denoising." *Hindawi Publishing Corporation: International Journal of Biomedical Imaging*, May 2006.