

Total Variation Based Rician Denoising and Deblurring Model

Pascal Getreuer

Melissa Tong

Luminita Vese

Department of Mathematics
University of California, Los Angeles
E-mail of corresponding author: lvese@math.ucla.edu
UCLA C.A.M. Report 11-67, October 2011

Abstract

Total variation regularization is usually applied with L^2 data fidelity assuming an additive white Gaussian noise model [24, 25]. However, better results are possible when the noise model accurately describes the noise distribution in the given image. Total variation denoising has already been developed with the Laplace noise model [2, 12] (L^1 data fidelity) and the Poisson noise model [19, 26]. In this paper, we develop two variational methods for total variation regularization using the Rician noise model. Our first model uses the Rician probability distribution function directly in the fidelity term, leading to a slightly nonconvex objective function. We give existence and comparison results for this model and investigate the validity of the model on three dimensional synthetic magnetic resonance imaging (MRI) data corrupted with Rician noise and Gaussian blur, both with known standard deviations. In this case, the numerical solutions are obtained using the L^2 and Sobolev H^1 gradient descent methods. In our second model, we develop a variational method for total variation regularization that closely approximates the Rician noise model. A key challenge is to find a convex approximation of the slightly nonconvex objective function. The variational problem is efficiently solved using the split Bregman method [16], and numerical examples are performed on the MRI data described above. Comparison of both models is presented.

1 Introduction

In this paper, we propose two variational models to denoise and deblur magnetic resonance imaging (MRI) data corrupted by Rician noise and Gaussian blur, both with known standard deviations. These variational models consist of total variation (TV) regularization using the Rician noise model (inherent in MR acquisitions), possibly in the presence of blur. Basu, Fletcher and Whitaker [7] applied anisotropic diffusion in diffusion tensor MRI data, using a correction term derived from a maximum a posteriori (MAP) estimate of the Rice distribution. Descoteaux and Wiest-Daessle et al. [13] and Wiest-Daessle, Prima et al. [30] applied a non-local means filter to Rician denoising. Wang and Zhou [28] perform MRI denoising with a combination of total variation and wavelet based regularization and a Gaussian noise model. The difference in this work is that we solve the

MAP estimate problem with the Rician noise model directly using the total variation as a prior, and we propose two formulations that we analyze theoretically and discretize in practice.

We begin this paper by reviewing the connection between the Rudin-Osher-Fatemi model (ROF) and MAP estimates in Section 2. In Section 3, the reformulation of ROF with the Rician noise model gives the basis of our two variational methods. In Section 4, we formulate our first variational method, which incorporates the Rician noise model directly, and give existence and comparison results for the purely denoising case (i.e., Gaussian blur is not present); this is inspired from prior work by Aubert-Aujol [5] on a non-convex variational model for multiplicative noise removal. We compute numerical solutions using the L^2 and Sobolev H^1 gradient descent methods and describe the numerical implementation details. In Section 5, we develop a convex approximation to the objective function and use this approximation in the formulation of our second model; implementation details are given for the application of split Bregman in solving for the numerical solutions of the resulting variational problem. In Section 6, we describe a method for estimating the parameter σ that represents the standard deviation of Rician noise of the observed data. We conclude the paper with numerical examples in Section 7 performed on three dimensional synthetic MRI data corrupted with Rician noise and Gaussian blur and compare the restoration results of the two methods.

2 Rudin-Osher-Fatemi model as a MAP Estimate

Maximum a posteriori (MAP) estimation is a Bayesian statistics technique that can be used to design denoising methods that take into account the distribution of the noise. As in [7], we apply MAP estimation to noise with Rician distribution.

Let f be a known degraded image and u the underlying clean image. The MAP estimate of u is the most likely value of u given f :

$$\hat{u} = \arg \max_u \mathbf{P}(u|f).$$

Applying Bayes' theorem obtains

$$\max_u \mathbf{P}(u|f) = \max_u \left\{ \mathbf{P}(u)\mathbf{P}(f|u) \right\} = \min_u \left\{ -\log \mathbf{P}(u) - \log \mathbf{P}(f|u) \right\}.$$

The first term is called the *prior* on u ; this term acts as a regularization or assumption on what u is likely to be. The second term describes the degradation process that produced f from u .

The Rudin-Osher-Fatemi (ROF) restoration model [24, 25] is

$$\min_{u \in BV(\Omega)} \left\{ \int_{\Omega} |Du| + \frac{\lambda}{2} \int_{\Omega} (Ku - f)^2 dx \right\}, \quad (1)$$

where $\Omega \subset \mathbb{R}^n$ is open and bounded. The observed image f is assumed to be related to the unknown restored image u (of bounded variation, in $BV(\Omega)$) by

$$f = Ku + n, \quad (2)$$

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 $P(u) = \exp(-\alpha \int |Du|)$:

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

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

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

Through this connection between ROF and MAP estimates, ROF can be reformulated to use other degradation models. Suppose that the probability density of $f(x)$ conditioned on $u(x)$ is

$$\mathbf{P}(f(x)|u(x)) = \exp[-H(Ku(x); f(x))]$$

and that the $f(x)$ are mutually independent over x . Then we obtain

$$\min_{u \in BV(\Omega)} \alpha \int_{\Omega} |Du| + \int_{\Omega} H(Ku; f) \, dx. \quad (3)$$

Benning and Burger [8] discuss problems of this form (and more general ones) and develop stability results using the Bregman distance.

3 ROF with Rician Noise

The *Rice* or *Rician* distribution has probability density

$$\mathbf{P}(r; \nu, \sigma) = \frac{r}{\sigma^2} \exp\left(-\frac{(r^2 + \nu^2)}{2\sigma^2}\right) I_0\left(\frac{r\nu}{\sigma^2}\right) \quad (4)$$

where $r, \nu, \sigma > 0$ and I_0 is the modified Bessel function of the first kind with order zero. If $X \sim N(\nu \cos \theta, \sigma^2)$ and $Y \sim N(\nu \sin \theta, \sigma^2)$ are independent normal random variables for any $\theta \in \mathbb{R}$, then $R = \sqrt{X^2 + Y^2}$ has Rician distribution, $R \sim \text{Rician}(\nu, \sigma)$.

Thus we have

$$\begin{aligned} H_{\sigma}(Ku(x); f(x), \sigma) &= -\log \mathbf{P}(f(x)|u(x)) \\ &= \frac{f(x)^2 + (Ku(x))^2}{2\sigma^2} - \log I_0\left(\frac{f(x)Ku(x)}{\sigma^2}\right) + \text{constant}. \end{aligned}$$

3.1 Bessel Functions

The Rician probability density (4) 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

$$t^2 \frac{d^2 y}{dt^2} + t \frac{dy}{dt} - (t^2 + n^2)y = 0. \quad (5)$$

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.

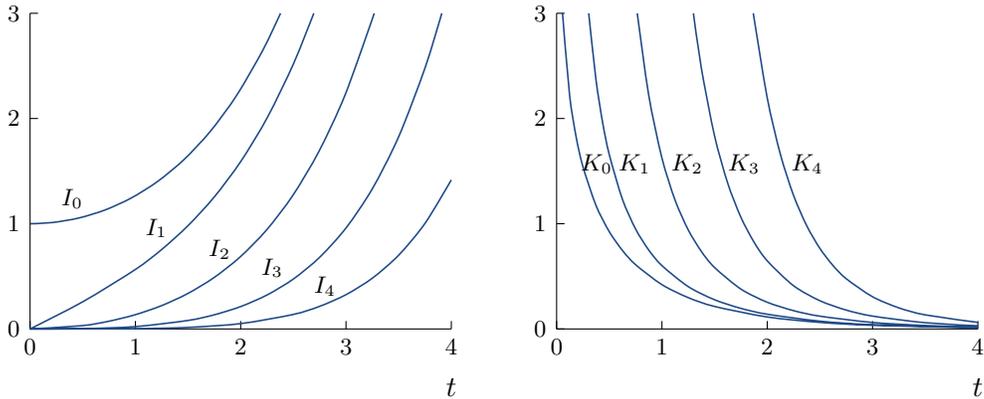


Figure 1: Modified Bessel functions.

The functions $I_n(t)$ are exponentially increasing while $K_n(t)$ are exponentially decreasing, see Figure 1. Some properties (from Abramowitz and Stegun [1]) of $I_n(t)$ are

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

4 TV-Regularized Restoration with the Rician Noise Model

Let Ω be an open, bounded, and connected subset of \mathbb{R}^n . We propose the following minimization method

$$\inf_{u \in BV(\Omega)} \left\{ F(u) = \int_{\Omega} |Du| + \lambda \int_{\Omega} \left[-\log \frac{f}{\sigma^2} - \log I_0\left(\frac{f(Ku)}{\sigma^2}\right) + \frac{f^2 + (Ku)^2}{2\sigma^2} \right] dx \right\}, \quad (6)$$

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

Although we deal with a nonconvex data fidelity term in (6), 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 [5], we show the existence of a minimizer for (6) with $K = I$.

Theorem 1. *Assume that $\inf_{\Omega} f(x) = \alpha > 0$, and $f \in L^{\infty}(\Omega)$. Then the minimization problem (6) for $K = I$ admits at least one solution $u \in BV(\Omega)$ satisfying*

$$0 \leq u \leq \sup_{\Omega} f. \quad (7)$$

Proof. Let $\{u_n\} \subset BV(\Omega)$ be a minimizing sequence for (6). Such a sequence exists since $F(u \equiv 1) = C < \infty$. Therefore, without loss of generality, we can assume that $F(u_n) \leq C$ for all n .

We denote the integrand of the Rician fidelity term by

$$G(u) = -\log \frac{f}{\sigma^2} - \log I_0\left(\frac{fu}{\sigma^2}\right) + \frac{f^2 + u^2}{2\sigma^2}, \quad (8)$$

and its derivative by

$$G'(u) = -\frac{I_1\left(\frac{fu}{\sigma^2}\right)}{I_0\left(\frac{fu}{\sigma^2}\right)} \frac{f}{\sigma^2} + \frac{u}{\sigma^2}. \quad (9)$$

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

$$F(\phi_{0,\beta}(|u_n|)) \leq F(u_n), \quad (10)$$

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

$$\phi_{0,\beta}(y) = \begin{cases} \beta & \text{for } y > \beta \\ y & \text{for } 0 \leq y \leq \beta \\ 0 & \text{for } y < 0. \end{cases} \quad (11)$$

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

$$\int_{\Omega} |D|u_n|| \leq \int_{\Omega} |Du_n|, \quad (12)$$

which is shown in [15]. Since $I_0(fu/\sigma^2)$ is an even function in u (see [9]), G is also an even function, so

$$\int_{\Omega} G(|u_n|)dx = \int_{\Omega} G(u_n)dx. \quad (13)$$

Together with (12), this gives for each n

$$F(|u_n|) \leq F(u_n). \quad (14)$$

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

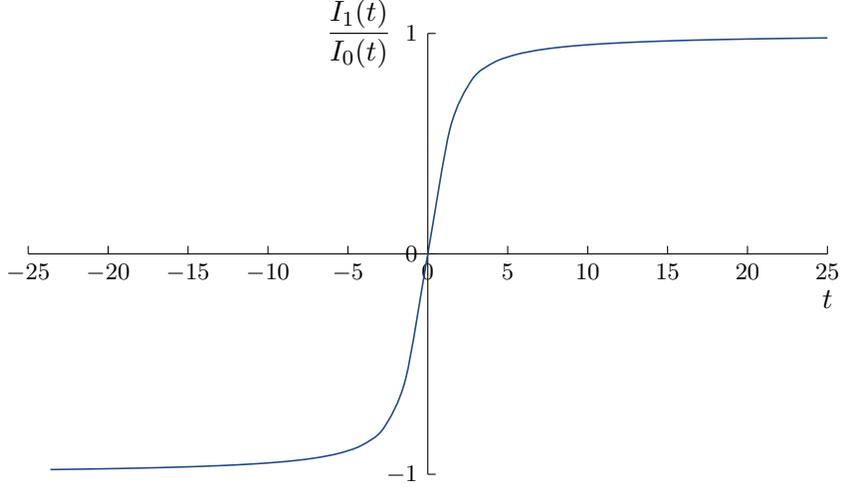


Figure 2: The function $I_1(t)/I_0(t)$.

Now we show $F(\phi_{0,\beta}(|u_n|)) \leq F(|u_n|)$. Noting that $|I_1(t)/I_0(t)| < 1$ (Fig. 2) for all $t \in \mathbb{R}$, we obtain

$$G'(u) \geq \frac{1}{\sigma^2}(-\beta + u) \quad (15)$$

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

$$G(\min(|u_n(x)|, \beta)) \leq G(|u_n(x)|) \quad (16)$$

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

$$G(\phi_{0,\beta}(|u_n(x)|)) \leq G(|u_n(x)|), \quad (17)$$

and therefore

$$\int_{\Omega} G(\phi_{0,\beta}(|u_n(x)|))dx \leq \int_{\Omega} G(|u_n(x)|)dx. \quad (18)$$

In addition, we expect

$$\int_{\Omega} |D\phi_{0,\beta}(|u_n|)| \leq \int_{\Omega} |D|u_n| \quad (19)$$

to hold for $u_n \in BV(\Omega)$; a proof of a general form of this result can be found in Appendix C of [18]. Combining this with (18) and our previous result gives for each n

$$F(\phi_{0,\beta}(|u_n|)) \leq F(|u_n|) \leq F(u_n). \quad (20)$$

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,

$$\int_{\Omega} |Du_n| + \lambda \int_{\Omega} G(u_n) dx \leq C. \quad (21)$$

In addition,

$$G(u_n) = -\log \frac{f}{\sigma^2} - \log I_0\left(\frac{fu_n}{\sigma^2}\right) + \frac{f^2 + u_n^2}{2\sigma^2} \geq -\log\left(\frac{\beta}{\sigma^2}\right) - \log I_0\left(\frac{\beta^2}{\sigma^2}\right) = C' > -\infty, \quad (22)$$

which allows us to conclude that

$$\int_{\Omega} |Du_n| dx \leq C'' < \infty, \quad (23)$$

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 $u_{n_k} \rightarrow u$ in $BV - w*$ and $u_{n_k} \rightarrow u$ in $L^1(\Omega)$ ([14]). Hence, we must have $u \in [0, \beta]$, and by the lower semi-continuity of the total variation and Fatou's lemma, we obtain

$$F(u) \leq \liminf F(u_{n_k}). \quad (24)$$

Therefore, we can conclude that u is a solution of (6). \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

$$\inf_{u \in BV(\Omega), 0 \leq u \leq \beta} \int_{\Omega} |Du| + \lambda \int_{\Omega} \left[-\log \frac{f}{\sigma^2} - \log I_0\left(\frac{f(k * u)}{\sigma^2}\right) + \frac{f^2 + (k * u)^2}{2\sigma^2} \right] dx, \quad (25)$$

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 $k * u \in [0, \beta]$. 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 [5], we obtain the following comparison theorem.

Theorem 2. *Let f_1 and f_2 be $L^\infty(\Omega)$ functions such that $0 < \alpha_1 \leq f_1 \leq \beta_1 < \infty$ and $0 < \alpha_2 \leq f_2 \leq \beta_2 < \infty$. If we assume $f_1 < f_2$, then $u_1 \leq u_2$ where u_1 and u_2 are solutions to (6) with $K = I$ corresponding to $f = f_1$ and $f = f_2$ respectively.*

Proof. First, denote

$$J(u) = \begin{cases} \int_{\Omega} |Du| & \text{if } u \in BV(\Omega) \\ +\infty & \text{if } u \in L^{\frac{n}{n-1}}(\Omega) \setminus BV(\Omega). \end{cases} \quad (26)$$

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

$$J(\min(u_1, u_2)) + \int_{\Omega} \left[-\log I_0\left(\frac{f_1 \min(u_1, u_2)}{\sigma^2}\right) + \frac{(\min(u_1, u_2))^2}{2\sigma^2} \right] dx$$

$$\geq J(u_1) + \int_{\Omega} \left[-\log I_0\left(\frac{f_1 u_1}{\sigma^2}\right) + \frac{u_1^2}{2\sigma^2} \right] dx,$$

and

$$\begin{aligned} J(\max(u_1, u_2)) + \int_{\Omega} \left[-\log I_0\left(\frac{f_2 \max(u_1, u_2)}{\sigma^2}\right) + \frac{(\max(u_1, u_2))^2}{2\sigma^2} \right] dx \\ \geq J(u_2) + \int_{\Omega} \left[-\log I_0\left(\frac{f_2 u_2}{\sigma^2}\right) + \frac{u_2^2}{2\sigma^2} \right] dx. \end{aligned}$$

Adding these two inequalities, we get

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

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 [10] and [15]), we obtain

$$\begin{aligned} \int_{\Omega} \left[\log \frac{I_0\left(\frac{f_2 u_2}{\sigma^2}\right)}{I_0\left(\frac{f_1 \min(u_1, u_2)}{\sigma^2}\right)} - \log \frac{I_0\left(\frac{f_2 \max(u_1, u_2)}{\sigma^2}\right)}{I_0\left(\frac{f_1 u_1}{\sigma^2}\right)} + \frac{(\max(u_1, u_2))^2}{2\sigma^2} - \frac{u_1^2}{2\sigma^2} + \frac{(\min(u_1, u_2))^2}{2\sigma^2} - \frac{u_2^2}{2\sigma^2} \right] dx \\ \geq J(u_1) + J(u_2) - (J(\min(u_1, u_2)) + J(\max(u_1, u_2))) \geq 0. \end{aligned}$$

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

$$\int_{\Omega_+} - \left(\log \frac{I_0\left(\frac{f_2 u_1}{\sigma^2}\right)}{I_0\left(\frac{f_1 u_1}{\sigma^2}\right)} - \log \frac{I_0\left(\frac{f_2 u_2}{\sigma^2}\right)}{I_0\left(\frac{f_1 u_2}{\sigma^2}\right)} \right) dx \geq 0.$$

We have that

$$g(y) = \log \frac{I_0(c_2 y)}{I_0(c_1 y)} = \log I_0(c_2 y) - \log I_0(c_1 y)$$

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

$$g'(y) = c_2 \frac{I_1(c_2 y)}{I_0(c_2 y)} - c_1 \frac{I_1(c_1 y)}{I_0(c_1 y)} = c_1 \left(\frac{c_2}{c_1} \frac{I_1(c_2 y)}{I_0(c_2 y)} - \frac{I_1(c_1 y)}{I_0(c_1 y)} \right) > c_1 \left(\frac{I_1(c_2 y)}{I_0(c_2 y)} - \frac{I_1(c_1 y)}{I_0(c_1 y)} \right) > 0$$

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. 2).

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

For numerical purposes, we consider an approximation

$$F_{\epsilon}(u) = \int_{\Omega} \sqrt{\epsilon^2 + |\nabla u|^2} dx + \lambda \int_{\Omega} \left[-\log \frac{f}{\sigma^2} - \log I_0\left(\frac{f(Ku)}{\sigma^2}\right) + \frac{f^2 + (Ku)^2}{2\sigma^2} \right] dx \quad (27)$$

to $F(u)$ to remove the singularity of $|\nabla u| = 0$ encountered in the Euler-Lagrange equation of $F(u)$.

In order to solve the proposed minimization problem in practice, we use and compare the L^2 and Sobolev (H^1) gradient descent methods (see [21]), as presented below.

In general, gradient descent methods involve the evolution problem

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

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

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

and

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

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 [21] that the L^2 and H^1 gradients are related in the following way

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

In the numerical examples section, we compare the results of these methods with those of the proposed method described below.

4.1 L^2 Gradient Descent

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

$$\frac{\partial u}{\partial t} = \lambda \left(-\frac{K^* K u}{\sigma^2} + K^* \left(\frac{I_1 \left(\frac{f K u}{\sigma^2} \right)}{I_0 \left(\frac{f K u}{\sigma^2} \right)} \cdot \frac{f}{\sigma^2} \right) \right) + \nabla \cdot \frac{\nabla u}{\sqrt{\epsilon^2 + |\nabla u|^2}}.$$

Defining

$$w_{i,j,k}^n := \frac{1}{h^2} \frac{1}{\sqrt{\epsilon^2 + \left(\frac{u_{i+1,j,k}^n - u_{i,j,k}^n}{h} \right)^2 + \left(\frac{u_{i,j+1,k}^n - u_{i,j,k}^n}{h} \right)^2 + \left(\frac{u_{i,j,k+1}^n - u_{i,j,k}^n}{h} \right)^2}},$$

we can describe the discretization we implemented by

$$\begin{aligned} \frac{u_{i,j,k}^{n+1} - u_{i,j,k}^n}{dt} &= \lambda \left(-\frac{K^* K u_{i,j,k}^n}{\sigma^2} + K^* \left(\frac{I_1 \left(\frac{f K u_{i,j,k}^n}{\sigma^2} \right)}{I_0 \left(\frac{f K u_{i,j,k}^n}{\sigma^2} \right)} \cdot \frac{f}{\sigma^2} \right) \right) - \frac{\lambda}{\sigma^2} (u_{i,j,k}^{n+1} - u_{i,j,k}^n) \\ &\quad + w_{i,j,k}^n (u_{i+1,j,k}^n - u_{i,j,k}^{n+1}) - w_{i-1,j,k}^n (u_{i,j,k}^{n+1} - u_{i-1,j,k}^n) + w_{i,j,k}^n (u_{i,j+1,k}^n - u_{i,j,k}^{n+1}) \\ &\quad - w_{i,j-1,k}^n (u_{i,j,k}^{n+1} - u_{i,j-1,k}^n) + w_{i,j,k}^n (u_{i,j,k+1}^n - u_{i,j,k}^{n+1}) - w_{i,j,k-1}^n (u_{i,j,k}^{n+1} - u_{i,j,k-1}^n), \end{aligned}$$

with Neumann boundary condition $\frac{\partial u}{\partial \vec{n}}|_{\partial\Omega} = 0$, where \vec{n} is the unit normal to the boundary $\partial\Omega$. Furthermore, we use the 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.

4.2 Sobolev Gradient Descent

For the gradient descent method using the Sobolev gradient, we consider the PDE

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

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

We implement the method using

$$\begin{aligned} G_{i,j,k}^m &= \lambda \left(-\frac{K^*Ku_{i,j,k}^n}{\sigma^2} + K^* \left(\frac{I_1(\frac{fKu_{i,j,k}^n}{\sigma^2})}{I_0(\frac{fKu_{i,j,k}^n}{\sigma^2})} \cdot \frac{f}{\sigma^2} \right) \right) \\ &\quad + w_{i,j,k}^n (u_{i+1,j,k}^n - u_{i,j,k}^n) - w_{i-1,j,k}^n (u_{i,j,k}^n - u_{i-1,j,k}^n) + w_{i,j,k}^n (u_{i,j+1,k}^n - u_{i,j,k}^n) \\ &\quad - w_{i,j-1,k}^n (u_{i,j,k}^n - u_{i,j-1,k}^n) + w_{i,j,k}^n (u_{i,j,k+1}^n - u_{i,j,k}^n) - w_{i,j,k-1}^n (u_{i,j,k}^n - u_{i,j,k-1}^n). \end{aligned}$$

and

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

where $W_{i,j,k}$ is the steady-state solution of the semi-implicit scheme:

$$\begin{aligned} W_{i,j,k}^{l+1} - c \left\{ \frac{W_{i+1,j,k}^l - 2W_{i,j,k}^{l+1} + W_{i-1,j,k}^l}{h_w^2} + \frac{W_{i,j+1,k}^l - 2W_{i,j,k}^{l+1} + W_{i,j-1,k}^l}{h_w^2} + \frac{W_{i,j,k+1}^l - 2W_{i,j,k}^{l+1} + W_{i,j,k-1}^l}{h_w^2} \right\} \\ = G_{i,j,k}^m. \end{aligned}$$

The scheme becomes

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

with 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 $\frac{\partial W}{\partial \vec{n}}|_{\partial\Omega} = 0$.

5 Convex Approximation of the Rician Noise Model and Implementation with Split Bregman

5.1 Nonconvexity

Our objective functional is nonconvex, but we need a convex objective function for the application of a fast iterative method, the split Bregman method [16], to be described later. Another challenge

is the numerical approximation of the involved Bessel functions. This section investigates the nonconvexity and develops a convex approximation.

For notational convenience, let $z = Ku$ and consider f as a fixed parameter. Then

$$\begin{aligned} H_\sigma(z) &= \frac{f^2 + z^2}{2\sigma^2} - \log I_0\left(\frac{fz}{\sigma^2}\right), \\ H'_\sigma(z) &= \frac{z}{\sigma^2} - \frac{f}{\sigma^2} \frac{I_1\left(\frac{fz}{\sigma^2}\right)}{I_0\left(\frac{fz}{\sigma^2}\right)}, & H'_\sigma(0) &= 0, \\ H''_\sigma(z) &= \frac{1}{\sigma^2} + \frac{f}{z\sigma^2} \frac{I_1\left(\frac{fz}{\sigma^2}\right)}{I_0\left(\frac{fz}{\sigma^2}\right)} + \frac{f^2}{\sigma^4} \left(\frac{I_1\left(\frac{fz}{\sigma^2}\right)}{I_0\left(\frac{fz}{\sigma^2}\right)} \right)^2 - \frac{f^2}{\sigma^4}, & H''_\sigma(0) &= \frac{1}{\sigma^4}(\sigma^2 - \frac{1}{2}f^2). \end{aligned}$$

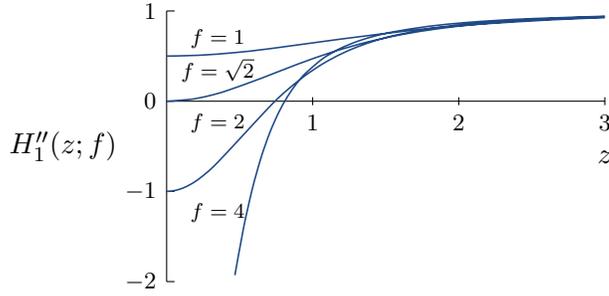


Figure 3: Plot of H''_1 for different values of f .

In our implementation, we approximate I_1/I_0 by a cubic rational implementation,

$$\frac{I_1(t)}{I_0(t)} \approx \frac{t^3 + \alpha_1 t^2 + \alpha_2 t}{t^3 + \alpha_3 t^2 + \alpha_4 t + \alpha_5} \equiv A(t).$$

This approximation is explained in more detail in a later section. By this change, H_σ is first approximated by another (still non-convex) function denoted \tilde{H}_σ . Substituting $A(t)$, we obtain

$$\begin{aligned} \tilde{H}'_\sigma(z) &= \frac{z}{\sigma^2} - \frac{f}{\sigma^2} A\left(\frac{fz}{\sigma^2}\right), & \tilde{H}'_\sigma(0) &= 0, \\ \tilde{H}''_\sigma(z) &= \frac{1}{\sigma^2} - \frac{f^2}{\sigma^4} A'\left(\frac{fz}{\sigma^2}\right), & \tilde{H}''_\sigma(0) &= \frac{1}{\sigma^4}(\sigma^2 - \frac{\alpha_2}{\alpha_5} f^2), \end{aligned}$$

where

$$A'(t) = \frac{(\alpha_3 - \alpha_1)t^4 + 2(\alpha_4 - \alpha_2)t^3 + (3\alpha_5 - \alpha_2\alpha_3 + \alpha_1\alpha_4)t^2 + (2\alpha_1\alpha_5)t + \alpha_2\alpha_5}{(t^3 + \alpha_3 t^2 + \alpha_4 t + \alpha_5)^2}.$$

For the minimization by split Bregman, it is important to substitute $H_\sigma(z)$ by a convex approximation. Since $H_\sigma(z; f) = H_1(\frac{z}{\sigma}; \frac{f}{\sigma})$, we may focus without loss of generality on the case $\sigma = 1$.

Unfortunately, $H''_1(0)$ is negative (and hence also $H''_1(z)$ for z sufficiently small) if $f > \sqrt{2}$. This boundary case and other values of f are shown in Figure 3. It appears that H''_1 increases monotonically, which suggests H_1 is nonconvex if and only if $f > \sqrt{2}$ and similarly \tilde{H}_1 is nonconvex if and only if $f > \sqrt{\frac{\alpha_5}{\alpha_1}} \approx 1.3955$.

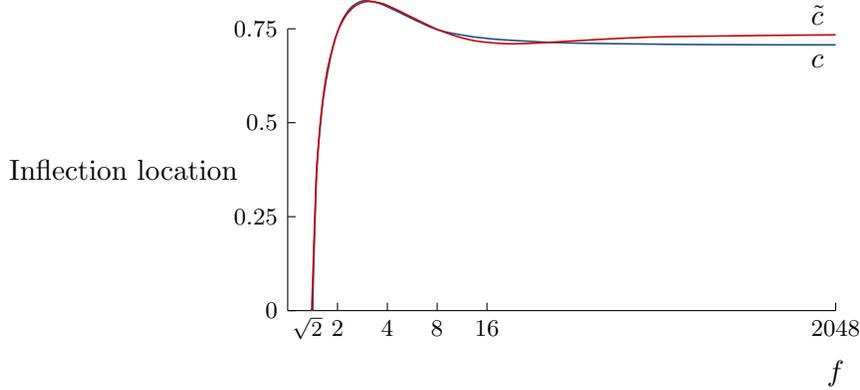


Figure 4: Inflection point c v.s. f .

Let c be the inflection point of H_1 , the point such that $H_1''(c) = 0$. Similarly, let \tilde{c} be the inflection point of \tilde{H}_1 . We can approximate c with numerical root finding (closed-form expression for \tilde{c} is possible as it is a root of a quartic polynomial). As shown in Figure 4, the maximum value of c is approximately 0.8246 at $f = 3$. The maximum value \tilde{c} is slightly less at approximately 0.8224.

Let $\bar{c} \equiv \max_f c \approx 0.8246$ such that c and \tilde{c} are always less than \bar{c} . Then we can make convex approximations as

$$G_1(z) = \begin{cases} H_1(z) & \text{if } z \geq \bar{c}, \\ H_1(\bar{c}) + H_1'(\bar{c})(z - \bar{c}) & \text{if } z \leq \bar{c}, \end{cases} \quad \tilde{G}_1(z) = \begin{cases} \tilde{H}_1(z) & \text{if } z \geq \bar{c}, \\ \tilde{H}_1(\bar{c}) + \tilde{H}_1'(\bar{c})(z - \bar{c}) & \text{if } z \leq \bar{c}. \end{cases}$$

Define $G_\sigma(z; f) = G_1(\frac{z}{\sigma}; \frac{f}{\sigma})$, then we can see that G_σ is convex and differs from H_σ for $z \leq \bar{c}\sigma$,

$$G_\sigma(z) = \begin{cases} H_\sigma(z) & \text{if } z \geq \bar{c}\sigma, \\ H_\sigma(\bar{c}\sigma) + H_\sigma'(\bar{c}\sigma)(z - \bar{c}\sigma) & \text{if } z \leq \bar{c}\sigma, \end{cases} \quad \tilde{G}_\sigma(z) = \begin{cases} \tilde{H}_\sigma(z) & \text{if } z \geq \bar{c}\sigma, \\ \tilde{H}_\sigma(\bar{c}\sigma) + \tilde{H}_\sigma'(\bar{c}\sigma)(z - \bar{c}\sigma) & \text{if } z \leq \bar{c}\sigma. \end{cases}$$

In practice for moderate noise levels, σ is quite small compared to the intensity range. Furthermore, very small intensities in MRI images correspond to void space, so z in this range is less important. Thus it appears reasonable to use \tilde{G}_σ as a convex approximation of H_σ .

5.2 Bessel Approximation

Aside from nonconvexity, the Rician model also has the challenge of numerical approximation of the Bessel functions. In our approach we only need to approximate the ratio $I_1(t)/I_0(t)$. This function is illustrated in Fig. 5.

We approximate I_1/I_0 by a cubic rational approximation,

$$\frac{I_1(t)}{I_0(t)} \approx A(t) := \frac{t^3 + \alpha_1 t^2 + \alpha_2 t}{t^3 + \alpha_3 t^2 + \alpha_4 t + \alpha_5} \quad (28)$$

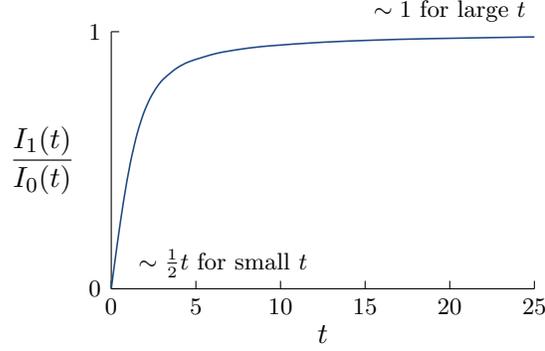


Figure 5: The function $I_1(t)/I_0(t)$.

where the coefficients are

$$\begin{aligned} \alpha_1 &= 0.950037, & \alpha_2 &= 2.38944, \\ \alpha_3 &= 1.48937, & \alpha_4 &= 2.57541, & \alpha_5 &= 4.65314. \end{aligned}$$

The approximation has the property that it exactly satisfies $A(0) = 0$ and $A(+\infty) = 1$. The coefficients $\alpha_1, \dots, \alpha_5$ are selected to minimize the L^∞ error.

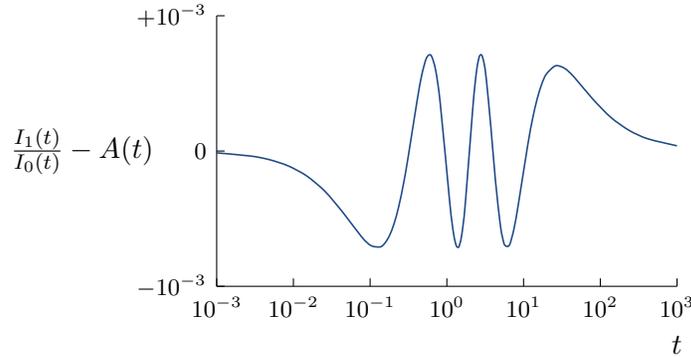


Figure 6: Error with cubic rational approximation (28).

The approximation error of (28) is shown in Figure 6. The exact value of I_1/I_0 is approximated with the highly accurate Bessel software by Amos [3].

Another approximation of I_1/I_0 is

$$\frac{I_1(t)}{I_0(t)} \approx A_2(t) := \begin{cases} \frac{1}{2} \frac{t(1 - t/2.74957)}{1 - t/3.48574} & \text{if } t < 1.64, \\ \frac{0.76272 - t}{0.23610 - t} & \text{if } t \geq 1.64. \end{cases} \quad (29)$$

This approximation has similar accuracy to (28) and the advantage of using lower-degree polynomials. Unfortunately its piecewise definition would further complicate solving the variable subproblem (30) described below, which is why we don't pursue it further here.

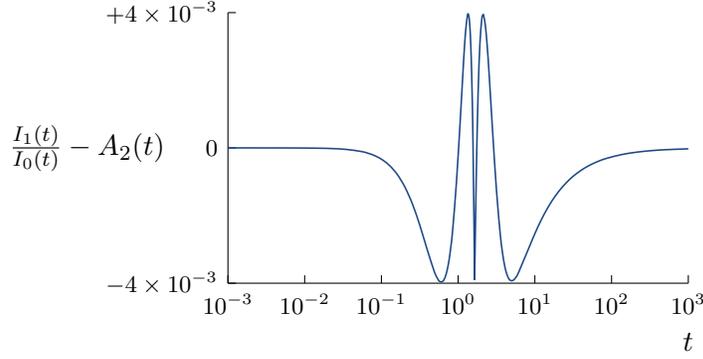


Figure 7: Error with piecewise approximation (29).

5.3 Solving the Variable Subproblem

For the split Bregman method described in the next section for our convex formulation, we will need to solve the variable subproblem

$$\arg \min_{z \geq 0} \tilde{G}_\sigma(z; f) + \frac{\gamma}{2}(z - y)^2, \quad (30)$$

where $\gamma > 0$ is a parameter and \tilde{G}_σ is the convex approximation of H_σ from Sec. 5.1,

$$\tilde{G}'_\sigma(z; f) = \begin{cases} \frac{z}{\sigma^2} - \frac{f}{\sigma^2} A\left(\frac{fz}{\sigma^2}\right) & \text{if } z \geq \bar{c}\sigma, \\ \frac{\bar{c}}{\sigma^2} - \frac{f}{\sigma^2} A\left(\frac{f\bar{c}}{\sigma}\right) & \text{if } z \leq \bar{c}\sigma. \end{cases}$$

The objective is strictly convex, thus it has exactly one minimizer z^* and its derivative

$$\tilde{G}'_\sigma(z^*; f) + \gamma(z^* - y)$$

is strictly monotone increasing. By this monotonicity, we can use the sign of $\tilde{G}'_\sigma(\bar{c}\sigma; f) + \gamma(\bar{c}\sigma - y)$ to conclude

$$\begin{cases} z^* = (y - \frac{1}{\gamma} \tilde{G}'_\sigma(\bar{c}\sigma; f))^+ & \text{if } \tilde{G}'_\sigma(\bar{c}\sigma; f) + \gamma(\bar{c}\sigma - y) \geq 0, \\ z^* > \bar{c}\sigma & \text{if } \tilde{G}'_\sigma(\bar{c}\sigma; f) + \gamma(\bar{c}\sigma - y) < 0, \end{cases} \quad (31)$$

where $(x)^+ := \max\{x, 0\}$. In the latter case, we must then solve

$$\left(\frac{1}{\sigma^2} + \gamma\right)z - \frac{f}{\sigma^2} A\left(\frac{fz}{\sigma^2}\right) = \gamma y, \quad z > \bar{c}\sigma. \quad (32)$$

Equation (32) may be solved by Newton's method,

$$z^{k+1} = z^k + \frac{y - \left(1 + \frac{1}{\gamma\sigma^2}\right)z^k + \frac{f}{\gamma\sigma^2} A\left(\frac{fz^k}{\sigma^2}\right)}{1 + \frac{1}{\gamma\sigma^2} - \frac{f^2}{\gamma\sigma^4} A'\left(\frac{fz^k}{\sigma^2}\right)}. \quad (33)$$

Newton's method is guaranteed to converge by convexity. Alternatively, (32) may be solved directly as the root of a quartic polynomial. Another possibility is to precompute the solution in a two-dimensional look-up table indexed by f and y .

5.4 Implementing Split Bregman

The split Bregman method [16] solves a minimization problem by operator splitting and then applying Bregman iteration [22, 23] to solve the split problem. A splitting of the generalized ROF model (3) is

$$\begin{aligned} \min_{\vec{d}, z, u} \alpha \int_{\Omega} |\vec{d}(x)| dx + \int_{\Omega} H(z(x), f(x)) dx \\ \text{subject to } \vec{d} = \nabla u, z = Ku \end{aligned} \quad (34)$$

For our application we use the convex approximation of the Rician noise model, $H = \tilde{G}_{\sigma}$. The advantage of this splitting is that the first term $\alpha \int |\vec{d}|$ only depends directly on \vec{d} and the second term $\int H(z, f)$ only on z . Variables \vec{d} and z are still related indirectly through the constraints $\vec{d} = \nabla u, z = Ku$.

Bregman iteration is used to solve the split problem. In each iteration, Bregman iteration calls for the solution of the following problem:

$$\begin{aligned} \min_{\vec{d}, z, u} \alpha \int_{\Omega} |\vec{d}| dx + \int_{\Omega} H(z, f) dx \\ + \frac{\gamma_1}{2} \|\vec{d} - \nabla u - \vec{b}_1\|_2^2 + \frac{\gamma_2}{2} \|z - Ku - b_2\|_2^2 \end{aligned} \quad (35)$$

where the additional terms are quadratic penalties enforcing the constraints and \vec{b}_1 and b_2 are variables related to the Bregman iteration algorithm.

The joint minimization over \vec{d}, z, u , is approximated by alternatingly minimizing one variable at a time, that is, fixing z and u and minimizing over \vec{d} , then fixing \vec{d} and u and minimizing over z , and so on. This leads to three variable subproblems:

- **The \vec{d} subproblem**, with z and u fixed, is

$$\min_{\vec{d}} \alpha \int |\vec{d}| dx + \frac{\gamma_1}{2} \|\vec{d} - \nabla u - \vec{b}_1\|_2^2.$$

Its solution decouples over x and is expressed in closed form as a vectorial shrinkage [29]:

$$\vec{d}(x) = \frac{\nabla u(x) + \vec{b}_1(x)}{|\nabla u(x) + \vec{b}_1(x)|} (|\nabla u(x) + \vec{b}_1(x)| - \alpha/\gamma_1)^+.$$

- **The z subproblem**, with \vec{d} and u fixed, is

$$\min_z \int_{\Omega} H(z, f) dx + \frac{\gamma_2}{2} \|z - Ku - b_2\|_2^2.$$

The solution decouples over x to

$$\min_z H(z, f) + \frac{\gamma_2}{2} (z - Ku - b_2)^2.$$

In our case with $H = \tilde{G}_{\sigma}$, the optimal z is given by equations (31) and (33) with $y = Ku + b_2$ as developed in section 5.3.

- **The u subproblem**, with \vec{d} and z fixed, is

$$\min_u \frac{\gamma_1}{2} \|\nabla u - \vec{d} + \vec{b}_1\|_2^2 + \frac{\gamma_2}{2} \|Ku - z + b_2\|_2^2.$$

For pure denoising, K is identity and the optimal u satisfies

$$\frac{\gamma_2}{\gamma_1} u - \Delta u = \frac{\gamma_2}{\gamma_1} (z - b_2) - \operatorname{div}(\vec{d} - \vec{b}_1).$$

In this case, u can be efficiently approximated by Gauss-Sidel iteration as proposed in [16].

Generally, the optimal u satisfies

$$\left(\frac{\gamma_2}{\gamma_1} K^* K - \Delta\right)u = \frac{\gamma_2}{\gamma_1} K^*(z - b_2) - \operatorname{div}(\vec{d} - \vec{b}_1),$$

where K^* is the adjoint of K . If Ku is a convolution, $Ku := \varphi * u$, the equation can be solved in the Fourier domain as

$$\hat{u} = \frac{\frac{\gamma_2}{\gamma_1} \hat{\varphi} \cdot (z - b_2)^\wedge - (\operatorname{div}(\vec{d} - \vec{b}_1))^\wedge}{\frac{\gamma_2}{\gamma_1} \hat{\varphi} \cdot \hat{\varphi} - \hat{\Delta}},$$

where $\hat{\cdot}$ denotes Fourier transform and multiplies and divisions are pointwise. To avoid boundary artifacts, the volume should be doubled along each dimension with its symmetric extension, though this has the drawback of a high cost in memory.

Martucci [20] developed how convolution with symmetric boundary handling can be done through discrete cosine transforms (DCTs). If φ is even in each dimension, then convolution with symmetric boundary handling can be computed as

$$\varphi * f = \mathcal{C}_{2e}^{-1}(\mathcal{C}_{1e}(\varphi) \cdot \mathcal{C}_{2e}(f)),$$

where \mathcal{C}_{1e} and \mathcal{C}_{2e} are the DCT-I and DCT-II transforms of the same period length as defined in [20]. Unlike with the discrete Fourier transform, the data does not need to be padded; symmetric boundaries are implied by the transforms. Noting also that the transformed data is real, the memory cost for 3D convolution is reduced by factor $2 \cdot 2^3 = 16$. So if φ is even in each dimension, a computationally efficient strategy to obtain u is

$$u = \mathcal{C}_{2e}^{-1} \left[\frac{\mathcal{C}_{2e} \left(\frac{\gamma_2}{\gamma_1} \varphi * (z - b_2) - \operatorname{div}(\vec{d} - \vec{b}_1) \right)}{\mathcal{C}_{1e} \left(\frac{\gamma_2}{\gamma_1} \varphi * \varphi - \Delta \right)} \right].$$

The split Bregman algorithm solves the minimization (3) with the following iteration:

$$\left\{ \begin{array}{l} \text{Initialize } u = f, \vec{d} = \vec{b}_1 = 0, z = b_2 = 0 \\ \textbf{while} \text{ "not converged"} \\ \quad \text{Solve the } \vec{d} \text{ subproblem} \\ \quad \text{Solve the } z \text{ subproblem} \\ \quad \text{Solve the } u \text{ subproblem} \\ \quad \vec{b}_1 := \vec{b}_1 + \nabla u - \vec{d} \\ \quad b_2 := b_2 + Ku - z. \end{array} \right. \quad (36)$$

When solving the subproblems, the x th subproblem solution is computed from the current values of all other variables and overwrites the previous value of variable x . Convergence may be checked by testing the difference of u from the previous iterate, for example $\|u^{\text{cur}} - u^{\text{prev}}\|_2 < Tol$.

6 Estimating σ

For practical applications, it is desirable to have few parameters or strategies that set parameters automatically. Here we develop a method to estimate the σ parameter of the Rician noise from the observed data.

In a typical MRI image, there is void space around the object of interest. Let f_1, \dots, f_N be voxel intensities sampled near one of the volume corners. Supposing these voxels are in the void space, their true intensity values are zero. This implies f_1, \dots, f_N are i.i.d. samples of Rayleigh distribution, which has probability density

$$\mathbf{P}(r; \sigma) = \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right). \quad (37)$$

The maximum likelihood estimator of σ is

$$\begin{aligned} \hat{\sigma} &= \arg \max_{\sigma} \mathcal{L}(\sigma | f_1, \dots, f_N) := \prod_n \frac{f_n}{\sigma^2} \exp\left(\frac{-f_n^2}{2\sigma^2}\right) \\ &= \sqrt{\frac{1}{2N} \sum_n f_n^2}. \end{aligned} \quad (38)$$

The maximum value of the likelihood, $\mathcal{L}(\hat{\sigma} | f_1, \dots, f_N)$, quantifies how well the data is modeled as Rayleigh. A small value indicates that the data is unlikely to be Rayleigh distributed. In this way we can possibly detect and avoid corners that are actually not void space.

Our strategy for estimating σ is

1. Extract a window of size $S \times S \times S$ from each of the 8 corners of the volume.
2. Separately for each corner, compute $\hat{\sigma}$ and $\mathcal{L}(\hat{\sigma} | f_1, \dots, f_{S^3})$.
3. For the corner where $\mathcal{L}(\hat{\sigma} | f_1, \dots, f_{S^3})$ is largest, select its $\hat{\sigma}$ as the final estimate.

In simulations, we find that this strategy reliably ignores corners that are not actually void. Thanks to the maximum likelihood estimation, the estimated σ value can be made very accurate by choosing moderately large S .

7 Numerical Examples

We perform restoration experiments on a synthetic T1 MRI volume (see Fig. 8) obtained from BrainWeb.¹ The L^2 and Sobolev gradient methods were implemented with MATLAB, and the proposed split Bregman method was implemented in C. The FFTW library (www.fftw.org) was used to compute DCT transforms in all methods. Furthermore, in each scheme, the data is restored using the same K and σ used in producing the input data (but σ can also be estimated as described above).

¹BrainWeb: Simulated Brain Database, <http://mouldy.bic.mni.mcgill.ca/brainweb>

7.1 Denoising Experiment

First, we show restoration results using L^2 and Sobolev gradient descent methods for images corrupted with Rician noise but not blur. Here, we take $K = I$ and $\sigma = 0.08$ for the standard deviation of the noise. With both methods, we stop at iteration $k + 1$ once the following tolerance is met:

$$\left| \frac{E_{k+1} - E_k}{E_k} \right| < TOL = 10^{-4},$$

where energy $E_k = F_\epsilon(u^k)$ (see (27)) calculated in the discrete sense. For L^2 gradient descent, the data is restored in 33 iterations using parameters $\lambda = 0.1$, $h = 1$, and fixed timestep $dt = 0.1$ (Fig. 9). For Sobolev gradient descent, the data is restored in 16 iterations using parameters $\lambda = 0.15$, $c = 1.5$, $h_w = 1$, and fixed timestep $dt = 0.05$ (Fig. 10). Fig. 11 gives plots of the energy versus iterations. We also show for denoising comparisons with the Rudin-Osher-Fatemi model [24] in Figs. 12 (entire volume) and 13 (cube inside the brain region only). For the ROF model, the data is restored using 14 iterations, $\lambda = 15$, fixed $dt = 0.1$. In both cases, we see that it is better to use the Rician denoising model. Also, the Sobolev gradient descent implementation gives the best results. Table 1 summarizes the denoising results.

7.2 Denoising and Deblurring Experiments

We consider experiments on two data sets corrupted with Rician noise and Gaussian blur. In the first experiment (Fig. 14, Fig. 16, Fig. 18), we have $\sigma = 0.02$ and Gaussian blur with standard deviation of 1.5 voxels. In the second experiment, (Fig. 15, Fig. 17, Fig. 19), we have $\sigma = 0.08$ and Gaussian blur with standard deviation of 0.6 voxels. Implementation details and a table of results (Table 2) is given below.

7.2.1 L^2 Gradient Descent

In the first experiment (Fig. 14), the data is restored with 28 iterations of L^2 gradient descent using parameters $\lambda = 0.4$, $h = 1$, and fixed timestep $dt = 0.1$.

In the second experiment (Fig. 15), the data is restored with 21 iterations of L^2 gradient descent with parameters $\lambda = 0.2$, $h = 1$, and fixed timestep $dt = 0.1$.

7.2.2 Sobolev Gradient Descent

In the first experiment (Fig. 16), the data is restored with 30 iterations of Sobolev gradient descent using parameters $\lambda = 1.1$, $c = 5$, $h_w = 1$, and fixed timestep $dt = 0.001$.

In the second experiment (Fig. 17), the data is restored with 17 iterations of Sobolev gradient descent with parameters $\lambda = 0.25$, $c = 1.25$, $h_w = 1$, and fixed timestep $dt = 0.05$.

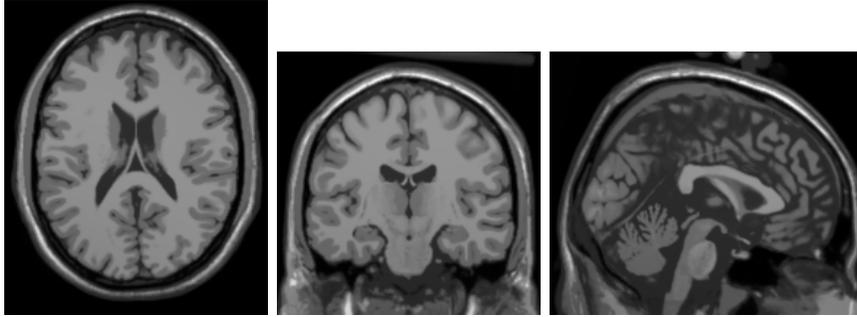


Figure 8: Three slices of the clean synthetic T1 MRI volume.

7.2.3 Split Bregman

In the first experiment (Fig. 18), the data is restored using 40 Bregman iterations using the parameters $\lambda = 10$, $\gamma_1 = \gamma_2 = 2$. The run time is 214 seconds on a recent laptop.²

In the second experiment (Fig. 19), the data is restored using 40 Bregman iterations using the parameters $\lambda = 0.6$, $\gamma_1 = \gamma_2 = 2$.

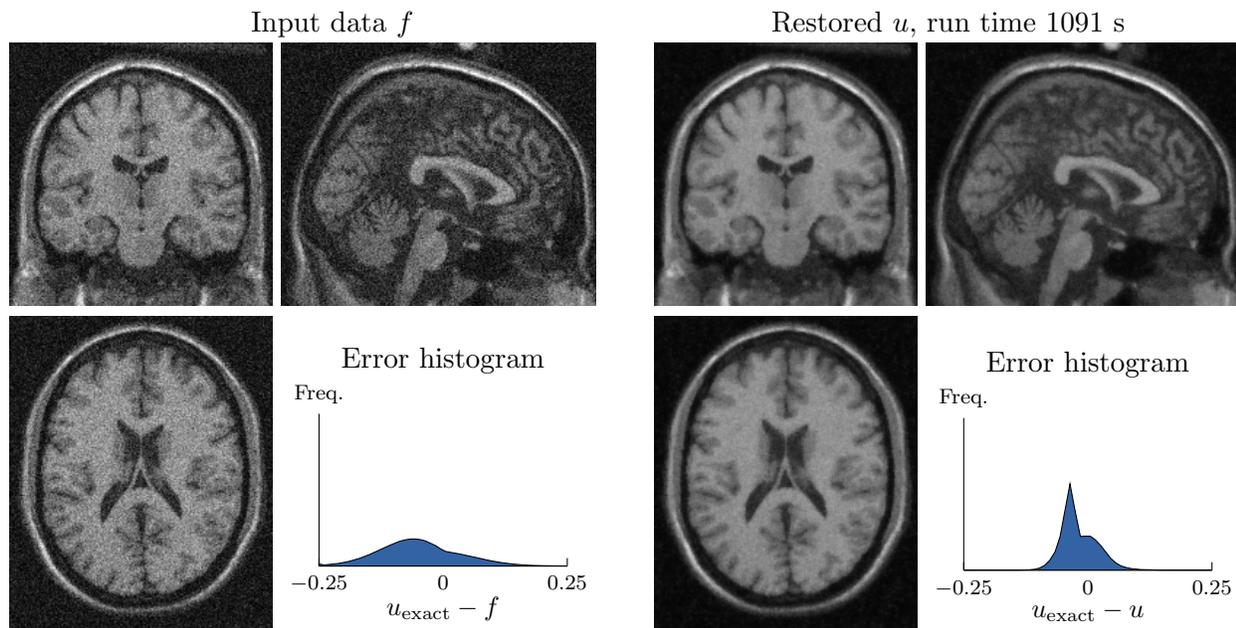


Figure 9: Denoising experiment with $\sigma = 0.08$ using L^2 gradient descent. The RMSE of f is 0.093702 and the RMSE of u is 0.039935.

²Run times are on a 2.40GHz Intel® Core™ 2 Duo T7700 with 2GB RAM.

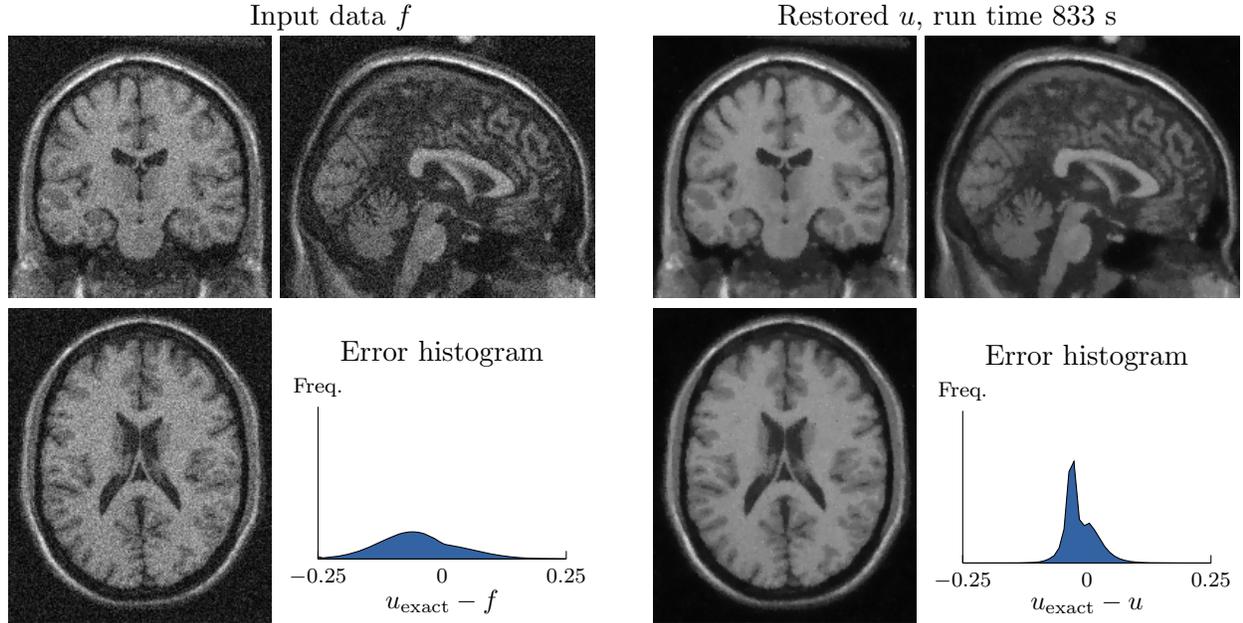


Figure 10: Denoising experiment with $\sigma = 0.08$ using Sobolev gradient descent. The RMSE of f is 0.093702 and the RMSE of u is 0.034541.

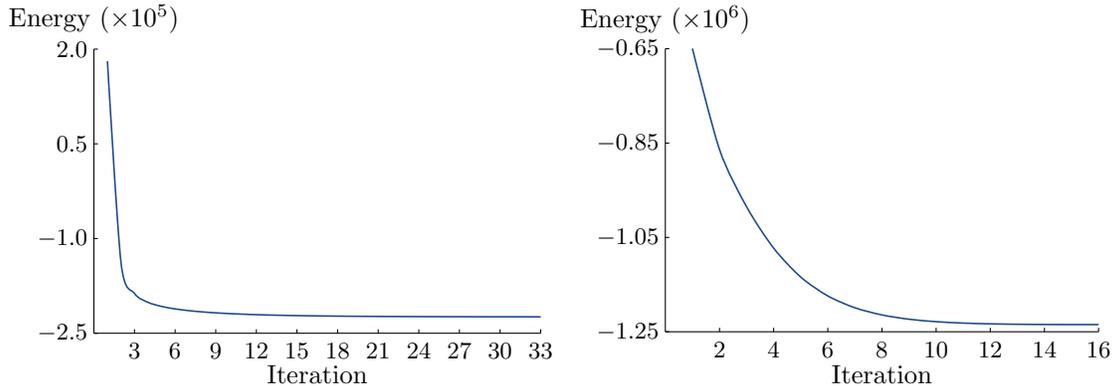


Figure 11: Plot of energy versus iteration for denoising experiments using L^2 and Sobolev gradient descent methods. Left: L^2 gradient descent. Right: Sobolev gradient descent.

8 Conclusions

TV-based restoration can be formulated for the Rician noise model as a maximum a posteriori estimate. Our first proposed variational method incorporates the Rician noise model directly, and we solve for numerical solutions to the resulting minimization problem using the L^2 and Sobolev (H^1) gradient descent methods. As noted, the objective function is slightly nonconvex, which is inconvenient as it limits the choice of applicable minimization methods. To obtain a convex problem, we determined a region in which the objective can be nonconvex and replaced it with a convex approximation. We then solved this convex TV-Rician minimization problem using the

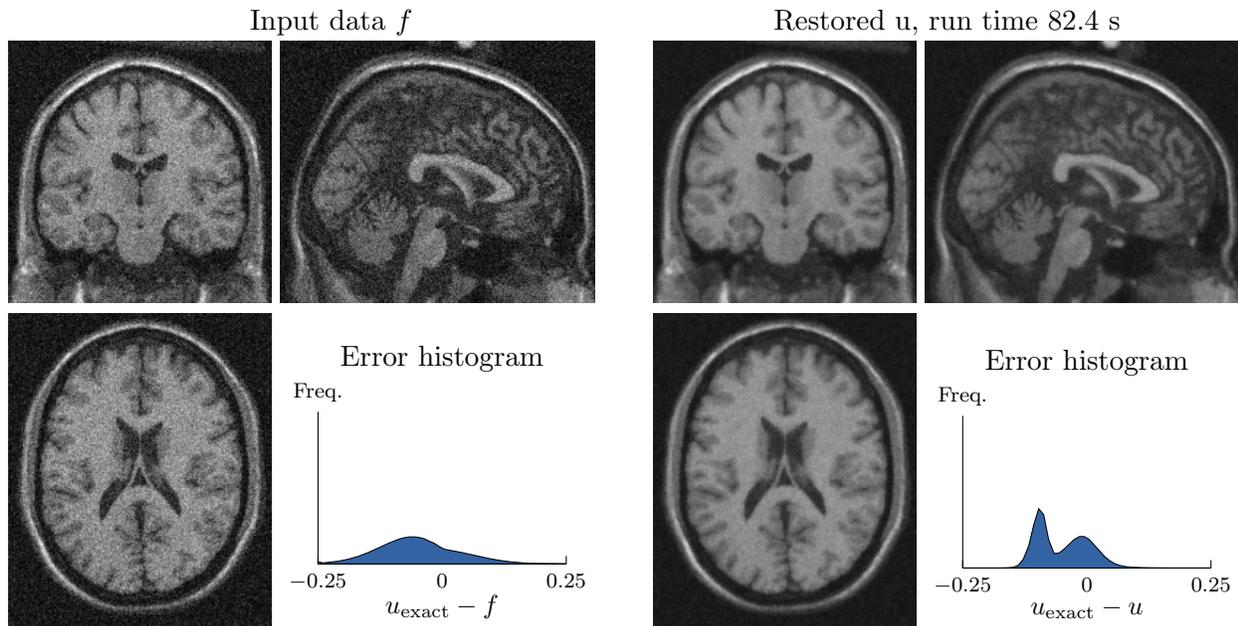


Figure 12: Denoising experiment with $\sigma = 0.08$ using the Rudin-Osher-Fatemi model [24]. The RMSE of f is 0.093702 and the RMSE of u is 0.071503.

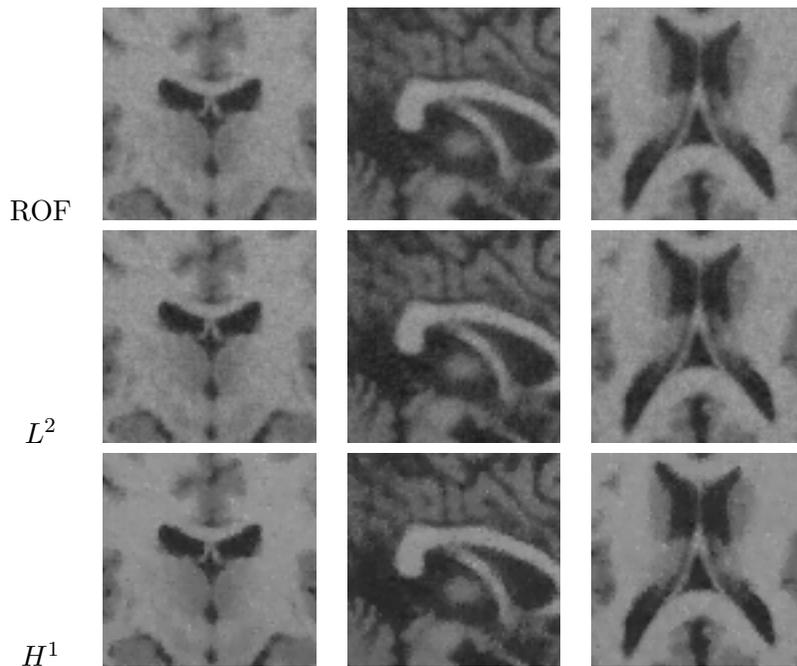


Figure 13: Second denoising comparison on a $80 \times 80 \times 80$ cube inside the brain region only (cube [71:150,51:130,51:130]). RMSE noisy data $f = 0.079473$; RMSE u , L2 gradient descent = 0.029803 (37.5%); RMSE u , H1 gradient descent = 0.027669 (34.8%); RMSE u , ROF = 0.031080 (39.1%).

Method (entire volume)	Time (s)	RMSE	Method (cube brain region)	RMSE
L^2 Gradient Descent	1091	0.039935	L^2 Gradient Descent	0.029803
H^1 Gradient Descent	833	0.034541	H^1 Gradient Descent	0.027669
ROF	82.4	0.071503	ROF	0.031080

Table 1: Summary of results and comparisons in the denoising experiments. Left: entire volume with background. Right: cube inside the brain region.

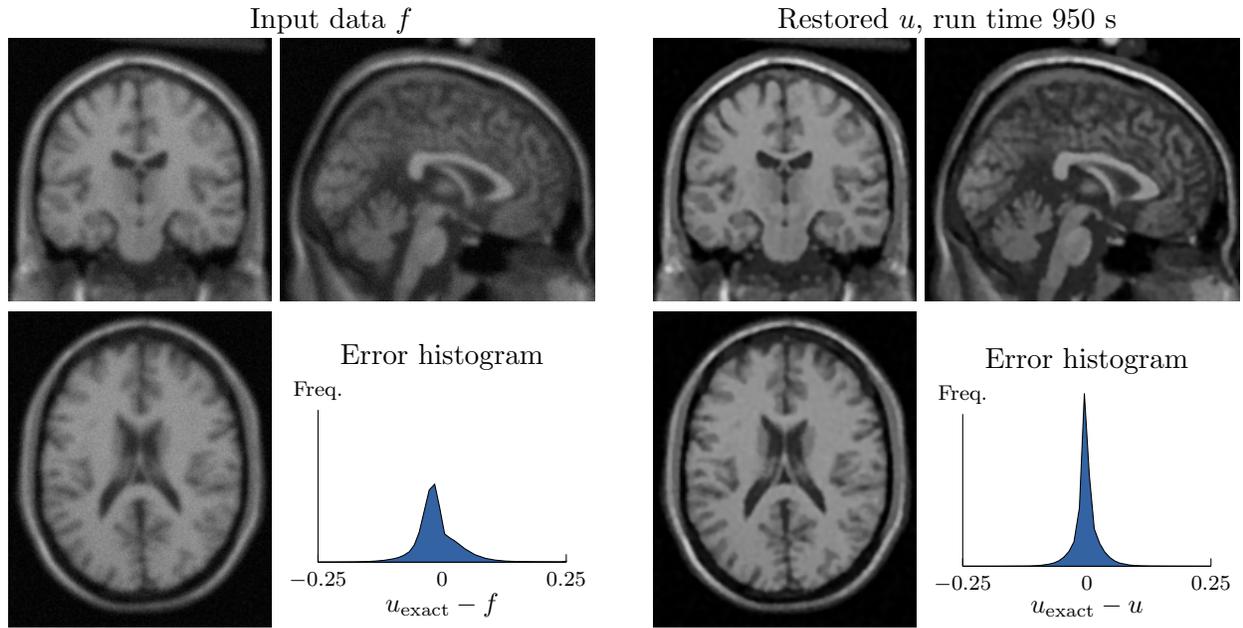


Figure 14: Restoration experiment using L^2 gradient descent. The RMSE of f is 0.046872 and the RMSE of u is 0.024279.

split Bregman method. Numerical experiments are performed on three dimensional synthetic MRI data corrupted with Rician noise and Gaussian blur and restoration results of the two proposed methods are compared.

9 Appendix

For $u \in BV(\Omega)$, we show

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

Using the property $J(\max\{u, v\}) + J(\min\{u, v\}) \leq J(u) + J(v)$ (see [10] and [15]) where $J(u)$ is defined in (26), we obtain for $u \in BV(\Omega)$,

$$J(|u|) = J(\max\{u(x), 0\} - \min\{u(x), 0\}) \leq J(\max\{u(x), 0\}) + J(\min\{u(x), 0\}) \leq J(u) + J(0) = J(u).$$

The result follows easily from the definition of $J(u)$. \square

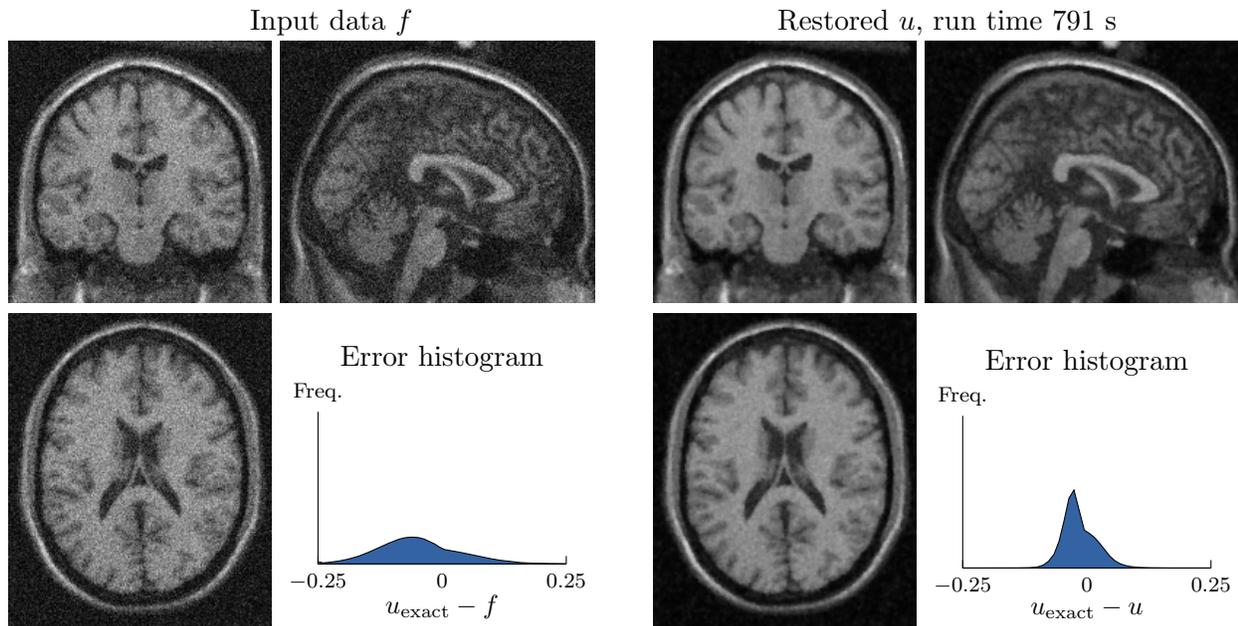


Figure 15: Restoration experiment using L^2 gradient descent. The RMSE of f is 0.094718 and the RMSE of u is 0.037068.

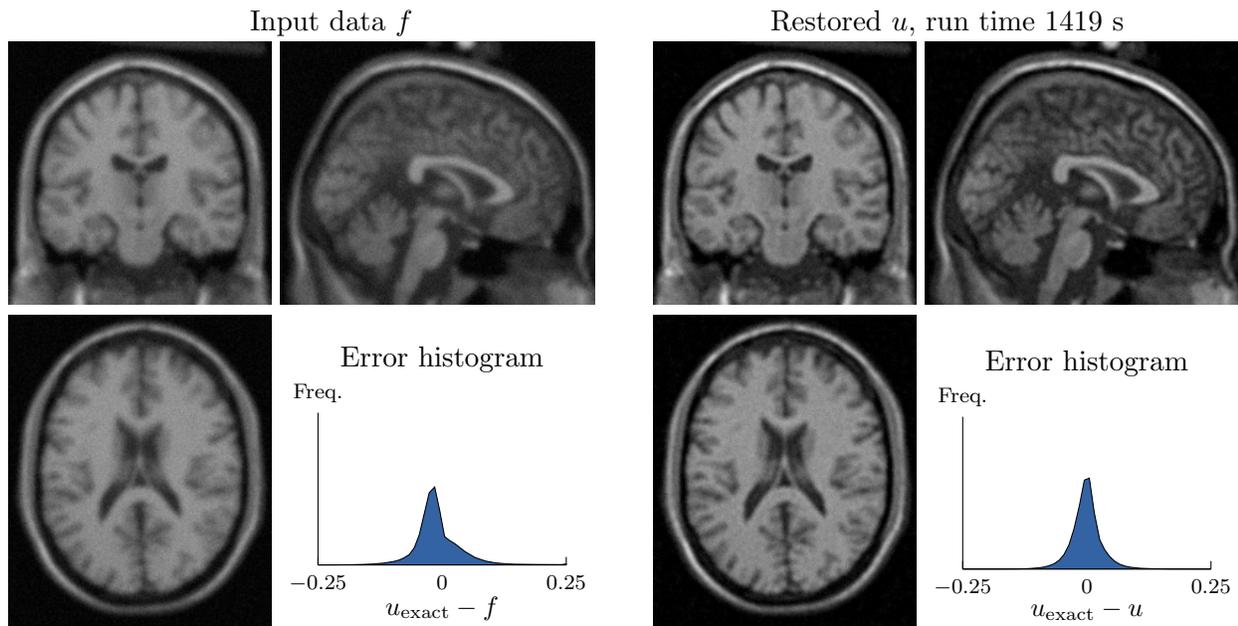


Figure 16: Restoration experiment using Sobolev gradient descent. The RMSE of f is 0.046872 and the RMSE of u is 0.029659.

References

- [1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover, pp. 355-379, 1965.

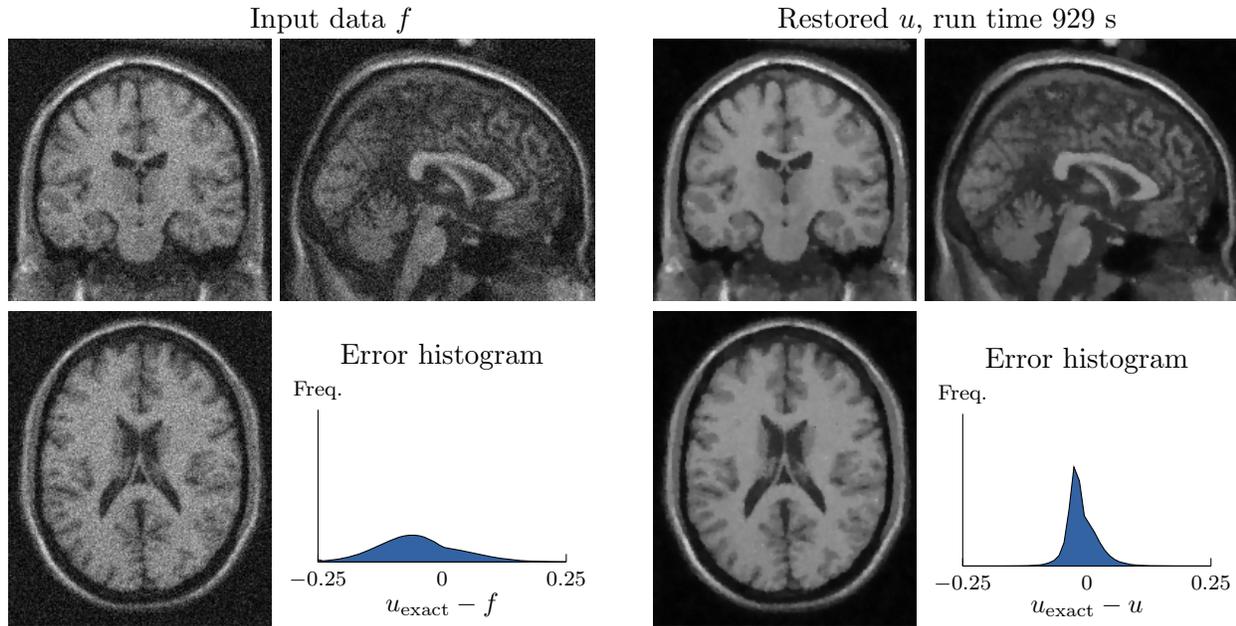


Figure 17: Restoration experiment using Sobolev gradient descent. The RMSE of f is 0.094718 and the RMSE of u is 0.032020.

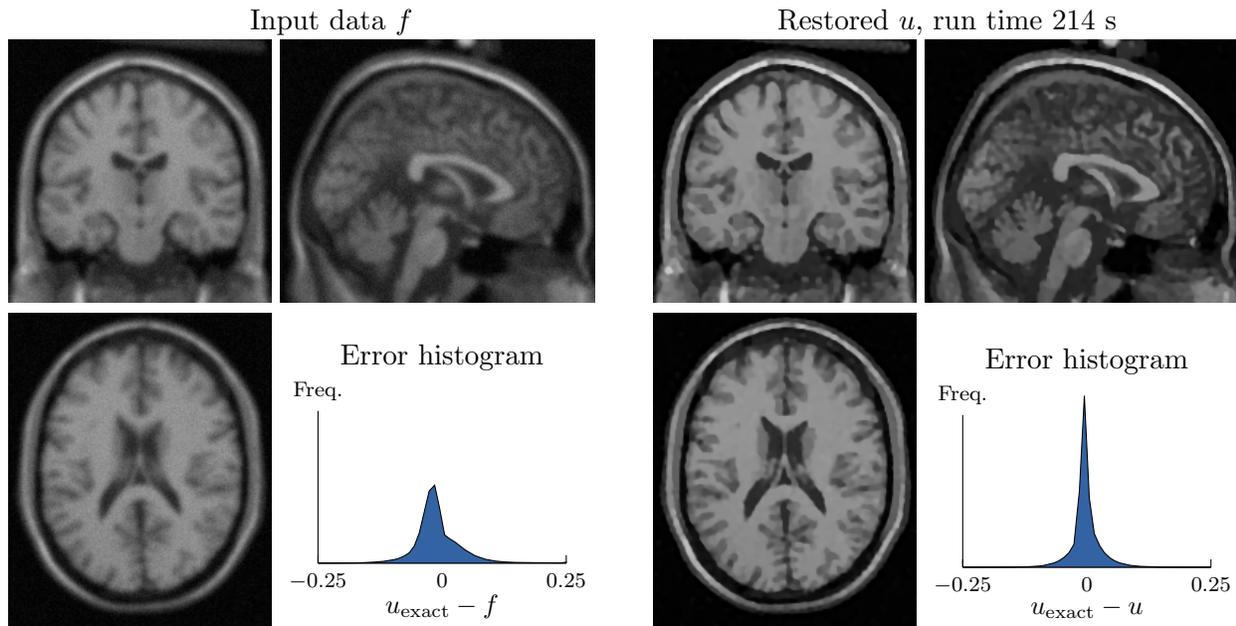


Figure 18: Restoration experiment using split Bregman. The root mean squared error (RMSE) of f is 0.046872 and the RMSE of u is 0.028319.

[2] S. Alliney, *Digital filters as absolute norm regularizers*, IEEE Transactions on Signal Processing, 40(6): 1548-1562, 1992.

[3] D.E. Amos, *A Subroutine Package for Bessel Functions of a Complex Argument and Nonnega-*

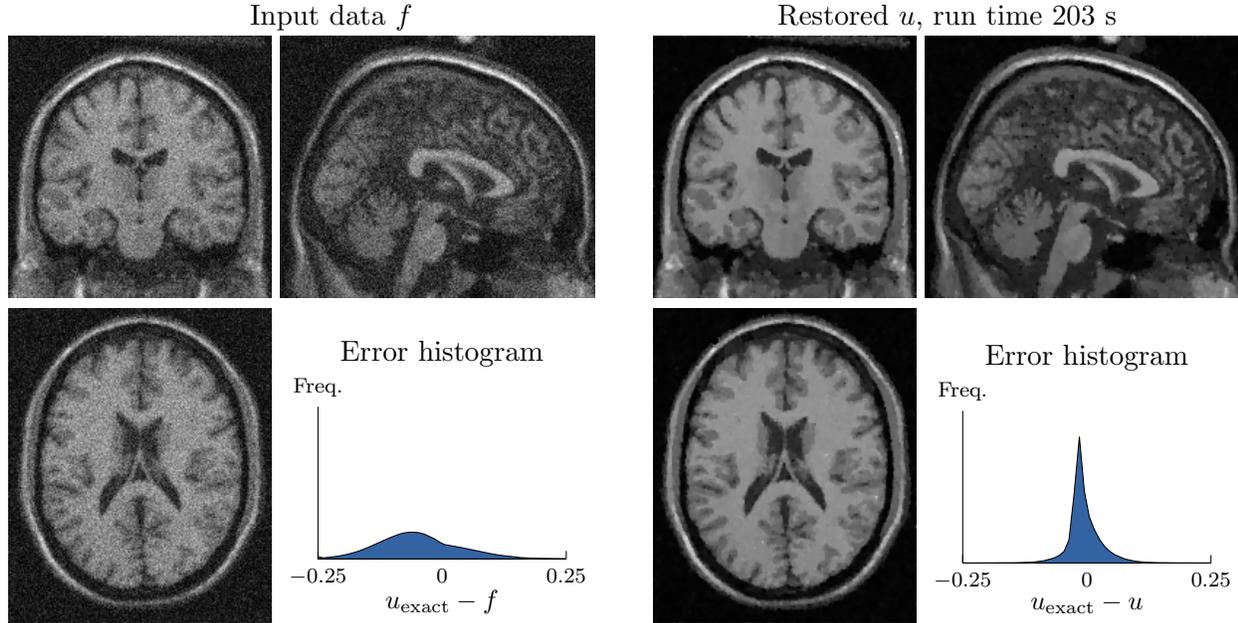


Figure 19: Restoration experiment using split Bregman. The RMSE of f is 0.094718 and the RMSE of u is 0.033471.

Experiment	Method	Image	Time (s)	RMSE
1	L^2 Gradient Descent	Fig. 14	950	0.024279
	Sobolev Gradient Descent	Fig. 16	1419	0.029659
	Split Bregman	Fig. 18	214	0.028319
2	L^2 Gradient Descent	Fig. 15	791	0.037068
	Sobolev Gradient Descent	Fig. 17	929	0.032020
	Split Bregman	Fig. 19	203	0.033471

Table 2: Comparison of time and RMSE values for L^2 gradient descent, Sobolev gradient descent, and proposed split Bregman methods. In experiment 1, input data is formed by Gaussian blur with standard deviation of 1.5 voxels and Rician noise with $\sigma = 0.02$. In experiment 2, input data is formed by Gaussian blur with standard deviation of 0.6 voxels and Rician noise with $\sigma = 0.08$.

tive Order, Sandia National Laboratory Report, SAND85-1018, May, 1985.

- [4] H. Attouch, G. Buttazzo, G. Michaille, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*, MPS-SIAM Series on Optimization, 2006.
- [5] G. Aubert and J.-F. Aujol, *A Variational Approach to Removing Multiplicative Noise*, SIAM J. Appl. Math., 68(4), 925946, 2008.
- [6] G. Aubert, P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, 2nd Ed., Applied Mathematical Sciences, Vol. 147, Springer, 2006.
- [7] S. Basu, T. Fletcher, R. Whitaker, *Rician Noise Removal in Diffusion Tensor MRI*, LNCS 4190, pp. 117-125, 2006.

- [8] M. Benning and M. Burger, *Error Estimates for Variational Models with Non-Gaussian Noise*, UCLA CAM Report 09-40, 2009.
- [9] M. Boas, *Mathematical Methods in the Physical Sciences*, John Wiley and Sons, Inc., 1983.
- [10] A. Chambolle, *An Algorithm for mean curvature motion*, Interfaces and Free Boundaries, 6: 1-24, 2004.
- [11] T. Chan, J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, 2005.
- [12] T.F. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math. 65(5): 1817-1837, 2005.
- [13] M. Descoteaux, N. Wiest-Daessle, S. Prima, C. Barillot, and R. Deriche, *Impact of Rician Adapted Non-Local Means Filtering on HARDI*, LNCS 5242, pp. 122-130, 2008.
- [14] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC-Press, 1992.
- [15] E. Giusti, *Minimal surfaces and functions of bounded variation*, Birkhauser, 1994.
- [16] T. Goldstein and S. Osher, *The Split Bregman Method for $L1$ -Regularized Problems*, SIAM J. Imaging Sci., 2(2), 323343, 2009.
- [17] M. Jung, G. Chung, G. Sundaramoorthi, L.A. Vese, and A.L. Yuille, *Sobolev gradients and joint variational image segmentation, denoising and deblurring*, SPIE Proceedings, Vol. 7246 Computational Imaging VII, C.A. Bouman, E.L. Miller, I. Pollak (editors), 2009.
- [18] P. Kornprobst, R. Deriche, G. Aubert, *Image Sequence Analysis via Partial Differential Equations*, Journal of Mathematical Imaging and Vision, 11(1):5-26, 1999.
- [19] T. Le, R. Chartrand and T. Asaki, *A Variational Approach to Constructing Images Corrupted by Poisson Noise*, JMIV, vol. 27(3): 257-263, 2007.
- [20] S. Martucci, *Symmetric convolution and the discrete sine and cosine transforms*, IEEE Trans. Sig. Processing SP-42: 1038-1051, 1994.
- [21] J.W. Neuberger, *Sobolev gradients and differential equations*, Springer Lecture Notes in Mathematics, Vol. 1670, 1997.
- [22] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An Iterative Regularization Method for Total Variation Based Image Restoration*, SIAM Multiscale Model. Simul., 4(2): 460-489, 2005.
- [23] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman Iterative Algorithms for ℓ_1 -Minimization with Applications to Compressed Sensing*, SIAM J. Imaging Sci., 1(1), 143168, 2008.
- [24] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D 60: 259-268, 1992.
- [25] L.I. Rudin and S. Osher, *Total variation based image restoration with free local constraints*, Proc. IEEE ICIP, Vol. 1: 31-35, 1994.

- [26] A. Sawatzky, C. Brune, J. Müller, and M. Burger, *Total Variation Processing of Images with Poisson Statistics*, LNCS 5702: 533-540, 2009.
- [27] L. Vese, *A Study in the BV Space of a Denoising-Deblurring Variational Problem*, Appl Math Optim 44: 131-161, 2001.
- [28] Y. Wang and H. Zhou, *Total Variation Wavelet-Based Medical Image Denoising*, International Journal of Biomedical Imaging, vol. 2006, Article ID 89095, pp. 1-6, 2006.
- [29] Y. Wang, J. Yang, W. Yin, and Y. Zhang, *A New Alternating Minimization Algorithm for Total Variation Image Reconstruction*, SIAM Journal on Imaging Sciences 1(3): 248-272, 2008.
- [30] 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*, LNCS 5242, pp. 171-179, 2008.