

A Penalized Likelihood Approach to Magnetic Resonance Image Reconstruction

Vera L. Bulaevskaya

Affiliation: Department of Statistics, Carnegie Mellon University

Address: Department of Statistics, 132 Baker Hall, Carnegie Mellon University,
5000 Forbes Ave., Pittsburgh, PA 15213 Email: vera@stat.cmu.edu

Phone: (412) 268-4248

Fax: (412) 268-7828

Gary W. Oehlert

Affiliation: School of Statistics, University of Minnesota

Summary

Currently, images acquired via Magnetic Resonance Imaging (MRI) and functional Magnetic Resonance Imaging (fMRI) technology are reconstructed using the discrete inverse Fourier transform. While computationally convenient, this approach is not able to filter out noise. This is a serious limitation because the amount of noise in MRI and fMRI can be substantial. In this paper, we propose an alternative approach to reconstruction, based on penalized likelihood methodology. In particular, we focus on non-linear shrinkage estimators and show that this approach achieves a great reduction in Integrated Mean Squared Error (IMSE) of the estimated image with respect to the currently used estimator. This approach is extremely fast and easy to implement computationally. In addition, it can be combined with various alternative approaches to MR image reconstruction and can be easily adapted to other, non-MRI contexts, in which the observed data and the quantities of interest are related via a linear transform.

Key Words: image reconstruction, Magnetic Resonance Imaging, penalized likelihood, shrinkage estimation, Bayes estimation

1 Introduction

Magnetic Resonance Imaging (MRI) is a widely used technique for imaging internal body tissue. One of its main advantages is that it is non-invasive, unlike other imaging methods, such as X-ray radiation or Positron Emission Tomography (PET). However, one of its shortcomings is that the images obtained from MRI technology can be very noisy due to various factors, such as motion-related artifacts and the presence of other signal sources [1]. This is particularly true of functional MRI (fMRI), which is used to study the activation patterns of the brain. In fMRI, a series of MR images of the brain is acquired as the subject performs a particular task or activity. Because many processes in the brain occur very rapidly, the images should be obtained at a high speed in order to detect activation associated with these processes. Thus, in fMRI the spatial resolution of each individual image is sacrificed in favor of the temporal resolution between consecutive images, leading to a larger amount of noise in the data.

For these reasons, a reconstruction technique that controls the variance of the image estimator is desirable in MRI and even more beneficial in fMRI. The current approach to image reconstruction uses the discrete inverse Fourier transform to convert the observed signal data to an image. Although it produces an unbiased estimator of the image and is computationally convenient, this approach does not reduce the variance of the estimated image and is, therefore, highly sensitive to noise. This paper introduces an alternative approach to reconstruction, based on penalized likelihood methodology, that balances

the goals of reducing bias and reducing variance at the same time. The work presented here focuses on a particular form of the penalized likelihood: shrinkage estimation. In particular, we consider non-linear shrinkage estimators because they have been shown to perform better than their linear counterparts when a quadratic loss function is used ([2], [3]). Although in this work the proposed model was specifically considered in the context of MRI, it is applicable to a wider class of problems, as discussed later in the paper.

The following section introduces the data observed in MRI experiments, the currently used modeling approach, and the proposed penalized likelihood model for reconstruction. Section 3 discusses hyperparameter selection for the proposed model. Section 4 evaluates the performance of the proposed approach for simulated MRI data. Finally, Section 5 discusses major conclusions from this work.

2 MRI Data and Reconstruction Approaches

A basic review of MRI technology, as well as that of the standard reconstruction approach, can be found in Bradley and Stark [4], Lange [5], Edelman et al. [6], Haacke et al. [7], and Bushong [8].

When performing MRI, one obtains a number of signals, each with its own frequency t and phase k . By its physical nature, each individual MR signal, denoted by $s(t, k)$, is the 2-dimensional continuous Fourier transform of the image function of interest $m(x, y)$

plus noise ϵ :

$$s(t, k) = \int_X \int_Y m(x, y) e^{2\pi i(tx+ky)} dx dy + \epsilon, \quad (1)$$

where $t = 0, \dots, n - 1$ and $k = 0, \dots, d - 1$. For computational convenience, in MRI n and d are usually taken to be powers of 2, and the most common values are 128 and 256.

2.1 Formulation of the Basic Framework

Let $\{b_{jv}\}$ be an infinite set of basis functions for the image function $m(x, y)$, i.e.,

$$m(x, y) = \sum_{j=1}^{\infty} \sum_{v=1}^{\infty} \beta_{jv} b_{jv}(x, y). \quad (2)$$

Then the expression in (1) can be rewritten as

$$s(t, k) = \sum_{j=1}^{\infty} \sum_{v=1}^{\infty} \beta_{jv} w_{jv}(t, k) + \epsilon, \quad (3)$$

where $w_{jv}(t, k)$ is the continuous 2-dimensional Fourier transform of the $(j, v)^{th}$ basis function $b_{jv}(x, y)$:

$$w_{jv}(t, k) = \int_X \int_Y b_{jv}(x, y) e^{2\pi i(tx+ky)} dx dy. \quad (4)$$

The expression in (3) can be written as

$$\mathbf{s} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (5)$$

where $\mathbf{s}_{nd \times 1} = \{s(t, k)\}$, $\mathbf{W}_{nd \times \infty}$ is a matrix consisting of entries $w_{jv}(t, k)$ as defined in (4), $\boldsymbol{\beta}_{\infty \times 1}$ is the vector of coefficients β_{jv} , and $\boldsymbol{\epsilon}_{nd \times 1}$ is the error vector. If the individual

elements of the error vector $\boldsymbol{\epsilon}$ are assumed to be independently and identically normally distributed each with mean 0 and variance σ^2 , we have

$$\mathbf{s}|\boldsymbol{\beta} \sim N(\mathbf{W}\boldsymbol{\beta}, \sigma^2\mathbf{I}_{nd \times nd}). \quad (6)$$

The problem of obtaining the image function therefore amounts to estimating the coefficient vector $\boldsymbol{\beta}$. Although in theory this vector is infinite-dimensional, in practice the number of basis functions is truncated to some value $l \times h$. Typically, l and h are taken to be equal, but for clarity we will keep the two letters separate. The estimated image function is then given by

$$\hat{m}(x, y) = \sum_{j=1}^l \sum_{v=1}^h \hat{\beta}_{jv} b_{jv}(x, y), \quad (7)$$

or

$$\hat{\mathbf{m}} = \mathbf{B}\hat{\boldsymbol{\beta}}, \quad (8)$$

where $\hat{\mathbf{m}}$ is the vector of the estimated image values at each location of interest (x, y) , \mathbf{B} is a matrix whose $[h(j-1) + v]^{th}$ column is the $(j, v)^{th}$ basis function evaluated at all locations of interest (x, y) , and $\hat{\boldsymbol{\beta}}$ is an $lh \times 1$ vector of estimated coefficient values $\hat{\beta}_{jv}$.

We can now introduce the currently used discrete inverse Fourier transform approach, as well as the proposed penalized likelihood approach to obtaining the coefficient vector estimate $\hat{\boldsymbol{\beta}}$.

2.2 The Discrete Inverse Fourier Transform Approach

Since the observed signal is the continuous 2-dimensional Fourier transform of the image function plus error, the most obvious way to estimate the image function would be to take the inverse continuous Fourier transform of the signal. However, this is impossible because frequencies t and phases k are observed finitely discretely. Therefore, the current approach to the problem is to approximate the image function with the discrete inverse Fourier transform:

$$m(x, y) = \sum_{t=0}^{n-1} \sum_{k=0}^{d-1} s(t, k) e^{-2\pi i(tx+ky)}. \quad (9)$$

In the framework established in Section 2.1, this approach to the solution is equivalent to using the Fourier basis, given by

$$b_{jv}(x, y) = e^{-2\pi i[(j-1)x+(v-1)y]}, \quad j = 1, \dots, n; \quad v = 1, \dots, d, \quad (10)$$

and using the ordinary least squares estimate of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (\overline{\mathbf{W}}^T \mathbf{W})^{-1} \overline{\mathbf{W}}^T \mathbf{s} = \mathbf{s}, \quad (11)$$

where $\overline{\mathbf{W}}$ denotes the conjugate of the complex matrix \mathbf{W} . The second equality in (11) follows from the fact that when the Fourier basis of dimension nd is used, \mathbf{W} is the identity matrix of dimension $nd \times nd$. The least squares estimator is unbiased for the first nd elements of $\boldsymbol{\beta}$, but it does not control the variance, resulting in a highly variable estimate of the image function. This shortcoming of the current estimation approach led us to consider an estimator that is better able to manage the bias/variance tradeoff.

2.3 A Penalized Likelihood Approach

An estimator of the image function that strikes a compromise between achieving a perfect fit and eliminating roughness (or variance) can be obtained by finding the function $m(x, y)$ that minimizes the penalized likelihood

$$\int_X \int_Y [Lm(x, y)]^2 + \sum_{t=0}^{n-1} \sum_{k=0}^{d-1} [s(t, k) - \hat{s}(t, k)]^2, \quad (12)$$

where L is a functional operator and $\hat{s}(t, k)$ is the expected signal given by

$$\hat{s}(t, k) = \int_X \int_Y m(x, y) e^{2\pi i(tx+ky)} dx dy. \quad (13)$$

Wahba and Kimeldorf [9] showed that the minimizer of (12) can be expressed in Bayesian terms. Specifically, selecting the form of the operator L is equivalent to choosing a prior distribution for the coefficient vector $\boldsymbol{\beta}$ introduced in Section 2.1, and the minimizer of (12) is the value of $\hat{\boldsymbol{m}}$ obtained by plugging in the Bayes estimate of $\boldsymbol{\beta}$ into (8). Note that the current approach to reconstruction, the discrete inverse Fourier transform, is a special case of the penalized likelihood approach with improper prior distributions placed on the first n^2 elements of $\boldsymbol{\beta}$ and one-point priors at 0 used for the remaining elements.

In this paper we use the Bayesian formulation as a convenient tool for obtaining an estimate of the image since it is equivalent to the original penalized likelihood formulation [9], while having the advantage of being more mathematically tractable. However, we do not make any further use of the Bayesian framework. In particular, in Section 4.1

we use a frequentist measure, the Integrated Mean Squared Error (IMSE), to evaluate the performance of our estimator.

2.3.1 Proposed Models

Recall that when the Fourier basis is used, the matrix \mathbf{W} of Fourier transforms of the basis functions is the identity matrix. Thus, assuming normal iid errors in the data, from (6) we have

$$\mathbf{s}|\boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \sigma^2 \mathbf{I}_{nd \times nd}). \quad (14)$$

The number of coefficients we will fit in each dimension is equal to the number of observations for each dimension, i.e., $l = n$ and $h = d$.

The Fourier basis functions are complex-valued, implying that the coefficient vector $\boldsymbol{\beta}$ is also complex-valued. As a consequence of this, we can place prior distributions on the real and imaginary parts of $\boldsymbol{\beta}$ separately, or we can model them together. We considered both options. Let

$$\tilde{\mathbf{s}}_{2nd \times 1} = \begin{bmatrix} \mathbf{s}_R \\ \mathbf{s}_I \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{\beta}}_{2nd \times 1} = \begin{bmatrix} \boldsymbol{\beta}_R \\ \boldsymbol{\beta}_I \end{bmatrix}, \quad (15)$$

where the subscripts R and I denote the real and imaginary parts, respectively. Now consider the following two mixture prior distributions and models:

I. The “unconstrained” shrinkage model:

$$\tilde{\mathbf{s}}|\tilde{\boldsymbol{\beta}} \sim N(\tilde{\boldsymbol{\beta}}, \frac{\sigma^2}{2}\mathbf{I}_{2nd \times 2nd}) \quad (16)$$

$$\tilde{\boldsymbol{\beta}}|\mathbf{T}^2 \sim N(\mathbf{0}_{2nd \times 1}, \text{diag } \mathbf{T}^2_{2nd \times 1}),$$

where T_{jvq}^2 , $j = 1, \dots, n$, $v = 1, \dots, d$, $q = 1$ for the real part and $q = 2$ for the imaginary part of the coefficient, are independently distributed and

$$T_{jvq}^2 = \begin{cases} \tau_1^2 & \text{with probability } p \\ \tau_2^2 & \text{with probability } 1 - p, \end{cases} \quad (17)$$

where $\tau_2^2 > \tau_1^2$ for all j, v , and q . Note that the variance of each element of $\tilde{\mathbf{s}}$ is $\sigma^2/2$ because one meaning of the specification in (6) is that the real and imaginary parts of individual elements of the error vector have independent normal distributions, each with mean 0 and variance $\sigma^2/2$. When the prior distribution in (16) is placed on $\tilde{\boldsymbol{\beta}}$, a different amount of shrinkage is allowed for the real and imaginary parts of individual elements of $\boldsymbol{\beta}$. For this reason, we refer to this prior as the “unconstrained” prior.

Using the Integrated Mean Squared Error (IMSE) of the estimated image function $\hat{\mathbf{n}}$ as the criterion, the Bayes estimate of $\tilde{\boldsymbol{\beta}}$ is the posterior mean $E(\tilde{\boldsymbol{\beta}}|\tilde{\mathbf{s}})$, consisting of individual elements

$$\hat{\beta}_{jvq} = E(\tilde{\beta}_{jvq}|\tilde{s}_{jvq}) = \left[\frac{1}{1 + \frac{\sigma^2}{2\tau_1^2}} \frac{1}{1 + g(\tilde{s}_{jvq})} + \frac{1}{1 + \frac{\sigma^2}{2\tau_2^2}} \left(1 - \frac{1}{1 + g(\tilde{s}_{jvq})} \right) \right] \tilde{s}_{jvq}, \quad (18)$$

where \tilde{s}_{jvq} is the $(j, v, q)^{th}$ element of $\tilde{\mathbf{s}}$, and

$$g(\tilde{s}_{jvq}) = \sqrt{\frac{\sigma^2 + 2\tau_1^2}{\sigma^2 + 2\tau_2^2}} \frac{1-p}{p} \exp \left[\tilde{s}_{jvq}^2 \left(\frac{1}{\sigma^2 + 2\tau_1^2} - \frac{1}{\sigma^2 + 2\tau_2^2} \right) \right]. \quad (19)$$

Then the estimate of $\boldsymbol{\beta}$ consists of individual elements

$$\hat{\beta}_{jv} = \hat{\beta}_{jv1} + i \times \hat{\beta}_{jv2}, \quad (20)$$

where $i = \sqrt{-1}$.

II. The “constrained” shrinkage model:

$$\mathbf{s}|\boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \sigma^2 \mathbf{I}_{nd \times nd}) \quad (21)$$

$$\boldsymbol{\beta}|\mathbf{T}^2 \sim N(\mathbf{0}_{nd \times 1}, \text{diag } \mathbf{T}^2_{nd \times nd}),$$

where T_{jv}^2 , $j = 1, \dots, n$, $v = 1, \dots, d$, are independently distributed

$$T_{jv}^2 = \begin{cases} \tau_1^2 & \text{with probability } p \\ \tau_2^2 & \text{with probability } 1 - p, \end{cases} \quad (22)$$

and $\tau_2^2 > \tau_1^2$ for all j and v . In this case, the prior distribution is placed on $\boldsymbol{\beta}$ rather than $\tilde{\boldsymbol{\beta}}$, thus requiring that the same amount of shrinkage be applied to the real and imaginary parts of every element of $\boldsymbol{\beta}$. For this reason, we refer to this prior as the “constrained” prior.

Using the IMSE of $\hat{\mathbf{m}}$ as the criterion, the Bayes estimate of $\boldsymbol{\beta}$ is the posterior mean $E(\boldsymbol{\beta}|\mathbf{s})$, consisting of individual elements

$$\hat{\beta}_{jv} = E(\beta_{jv}|s_{jv}) = \left[\frac{1}{1 + \frac{\sigma^2}{\tau_1^2}} \frac{1}{1 + g(s_{jv})} + \frac{1}{1 + \frac{\sigma^2}{\tau_2^2}} \left(1 - \frac{1}{1 + g(s_{jv})} \right) \right] s_{jv}, \quad (23)$$

where s_{jv} is the $(j, v)^{th}$ element of \mathbf{s} ($j = 1, \dots, n$ and $v = 1, \dots, d$), and

$$g(s_{jv}) = \sqrt{\frac{\sigma^2 + \tau_1^2}{\sigma^2 + \tau_2^2}} \frac{1 - p}{p} \exp \left[\|s_{jv}\|^2 \left(\frac{1}{\sigma^2 + \tau_1^2} - \frac{1}{\sigma^2 + \tau_2^2} \right) \right], \quad (24)$$

with $\|s_{jv}\|^2 = (\text{Re}(s_{jv}))^2 + (\text{Im}(s_{jv}))^2$.

2.3.2 General Remarks

Penalized likelihood has several forms that are well known to a wide statistical audience. For example, when L in (12) is the m^{th} derivative, the resulting function is the smoothing spline of order $2m - 1$. On the other hand, a very familiar example of the Bayesian formulation of penalized likelihood is ridge regression since the coefficient estimator in that case arises from using $N(\mathbf{0}, \tau^2 \mathbf{I})$ prior distribution on the coefficient vector. Like the estimators introduced in Section 2.3.1, the ridge regression estimator is a shrinkage estimator, but in contrast to the proposed estimators, it is linear. Antoniadis and Fan [10] proposed various nonlinear penalty functions in the context of wavelet estimation and showed that the resulting estimators possess good statistical properties, such as being adaptively minimax.

It should be noted that penalized likelihood methodology has been applied previously by MRI researchers. For example, Sutton et al. [11] used a particular form of penalized likelihood, quadratic regularization, to obtain MR images in the presence of magnetic field inhomogeneities, which proved to be successful at correcting for field distortions, both in simulations and in a real dataset. In contrast to our approach, however, their solution is iterative and is, thus, more computationally intensive. Another difference from our approach is that Sutton et al. use linear shrinkage estimators, which are less flexible than nonlinear shrinkage estimators, as discussed later in this section.

MR image reconstruction has received attention in the statistical community. Besag [12], for example, modeled images as Markov random fields in a Bayesian framework. Hu et al. [13] took a very similar approach, using Gibbs random fields. In contrast to these works, we chose to model the raw k -space data produced by MR scanners rather than model the partially reconstructed image space data because this approach is more direct.

Mixture priors have also been given a great deal of consideration in the statistical community. Chipman et al. [14] used the mixture normal priors of the type introduced in (16) and (21) in the context of wavelet coefficient estimation. Clyde et al. [15] took a similar approach to wavelet coefficient estimation by using a mixture of a normal distribution and a delta distribution at 0 as the prior. In addition, Vidakovic [16] proposed the t prior distribution, which can be expressed as a mixture of normal distributions, also in order to estimate wavelet coefficients. All of these authors found that their respective mixture priors performed well in simulations and in applications to real data because they allow for a higher degree of flexibility than do linear shrinkage estimators.

The amount of shrinkage for each coefficient is adaptive in nature as it depends on the magnitude of the coefficient, with the coefficients of smaller magnitudes shrunk heavily toward 0 and those of larger magnitudes resulting in much less shrinkage. This adaptability is the reason that the non-linear shrinkage estimator enjoys a lower Mean Squared Error (MSE) than do the least squares and linear shrinkage estimators. The

reader is referred to Figure 3.12 in [17] for a visual comparison of the MSEs of linear and non-linear shrinkage estimators.

If the variance reduction with respect to the least squares estimator is great enough to compensate for the addition in bias introduced by shrinkage, the resulting IMSE of the shrinkage estimator will be below that of the least squares estimator of the image function. Both the bias and the variance of the shrinkage estimator $\hat{\beta}$ depend on the true coefficient value β . With a prior mean of 0, the bias and the variance are both small when real and imaginary parts of β are near 0 because the estimate will almost always be shrunk toward the correct value near 0. When either the real or the imaginary part of β is far enough from 0, there will be very little shrinkage, so the variance will be very close to that of the least squares estimator, and the bias will be very close to 0, as well. The magnitudes of β that fall between these two extremes are problematic for the shrinkage estimator because in this case it is not clear how much shrinkage, if any, is needed. Hence, both the bias and the variance and, consequently, the MSE, are above those of the least squares estimator for these values. Thus, the non-linear shrinkage estimator performs better than the least squares estimator in terms of the IMSE when most values of β are either close to 0 or quite far away from 0 in magnitude.

3 Choosing Prior Parameters

The values of prior parameters τ_1^2 , τ_2^2 , and p need to be chosen to implement both the “unconstrained” and the “constrained” shrinkage estimators. We will consider the two estimators separately.

3.1 “Unconstrained” Shrinkage Estimator

For the ease of notation, the indices jvq will not be used in this section and will be implicit in the discussion. The “unconstrained” shrinkage estimator in (18) is the discrete inverse Fourier transform, or least squares, estimator multiplied by a shrinkage factor, namely

$$\hat{\beta} = f(\tilde{s}, \tau_1^2, \tau_2^2, p) \cdot \tilde{s}, \quad (25)$$

where

$$f(\tilde{s}, \tau_1^2, \tau_2^2, p) = \frac{1}{1 + \frac{\sigma^2}{2\tau_1^2}} \frac{1}{1 + g(\tilde{s})} + \frac{1}{1 + \frac{\sigma^2}{2\tau_2^2}} \left(1 - \frac{1}{1 + g(\tilde{s})} \right) \quad (26)$$

with $g(\tilde{s})$ as in (19). Instead of selecting the values of the parameters τ_1^2 , τ_2^2 , and p directly, it is simpler to determine the desirable values of the three main characteristics of the shrinkage factor function $f(\tilde{s})$:

- 1) the value of the function when $\tilde{s} = 0$, $f(0)$, which we denote by θ_0 , so

$$\theta_0 \stackrel{def}{=} f(0), \quad (27)$$

2) the asymptotic value of the function when \tilde{s} tends to ∞ , which we denote by θ_∞ ,

so

$$\theta_\infty \stackrel{def}{=} \lim_{\tilde{s} \rightarrow \infty} f(\tilde{s}), \quad (28)$$

3) the rate of change in the function from small to large magnitudes of \tilde{s} , controlled by a parameter, denoted by θ_r and given by

$$\theta_r \stackrel{def}{=} \frac{1}{1 + \frac{\sigma^2}{2\tau_1^2}} \quad (29)$$

(the higher values of θ_r correspond to the lower rates of change in $f(\tilde{s})$).

Note that by definition, θ_0 is a convex combination of θ_r and θ_∞ , implying that $0 < \theta_r < \theta_0 < \theta_\infty < 1$.

To illustrate the effect of these three parameters on the shrinkage factor function, Figure 1 shows $f(\tilde{s})$ for positive values of \tilde{s} (the shrinkage factor for negative values of \tilde{s} is just the mirror image of the plot in Figure 1 about the y -axis) when $\theta_r = 0.05$, $\theta_0 = 0.1$, and $\theta_\infty = 0.999$ and when $\theta_r = 0.09$ while θ_0 and θ_∞ are kept the same (σ^2 is set to 1).

It can be easily shown that

$$\tau_1^2 = \frac{\sigma^2}{2} \cdot \frac{\theta_r}{1 - \theta_r}, \quad (30)$$

$$\tau_2^2 = \frac{\sigma^2}{2} \cdot \frac{\theta_\infty}{1 - \theta_\infty}, \quad (31)$$

and

$$p = \left(\frac{\theta_0 - \theta_r}{\theta_\infty - \theta_0} \cdot \sqrt{\frac{1 - \theta_r}{1 - \theta_\infty}} + 1 \right)^{-1}. \quad (32)$$

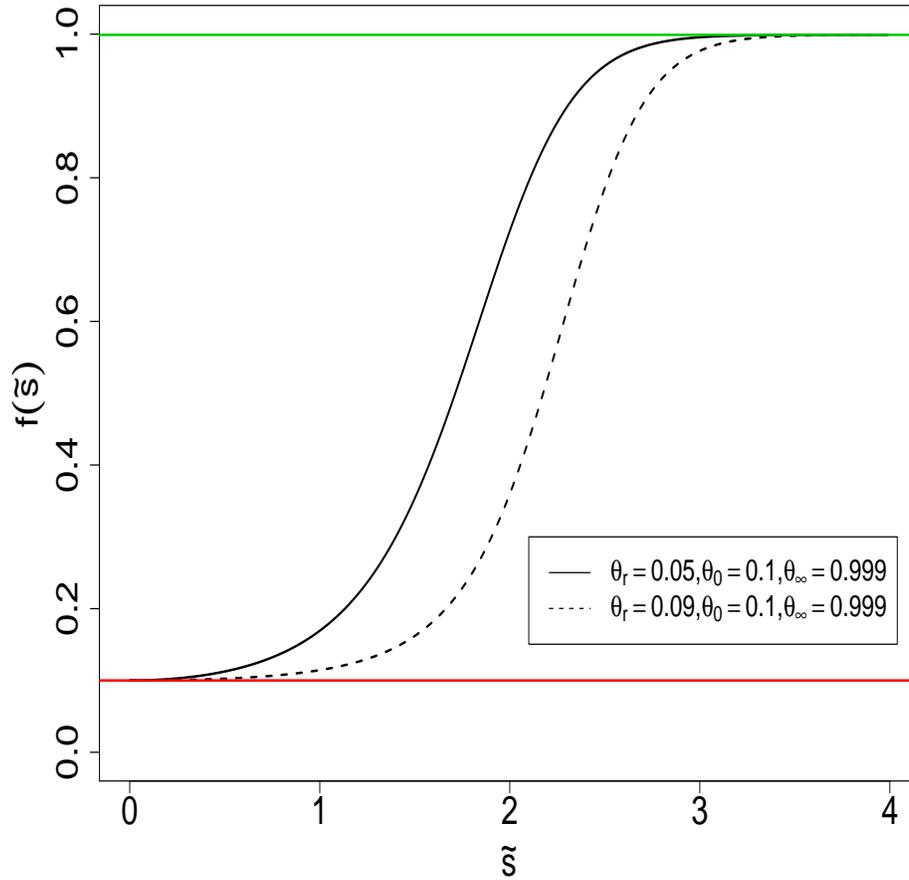


Figure 1: Shrinkage factor $f(\tilde{s})$ as a function of response \tilde{s} with $\theta_r = 0.05$, $\theta_0 = 0.1$, and $\theta_\infty = 0.999$ (black solid line) and $\theta_r = 0.09$, $\theta_0 = 0.1$, and $\theta_\infty = 0.999$ (black dashed line). Red line shows the value of 0.1 of the function at 0 and green line shows the function's asymptotic value of 0.999.

It is desirable for θ_∞ to be close to 1 since high magnitudes of \tilde{s} imply high magnitudes of $\tilde{\beta}$, and, hence, there should be very little, if any, shrinkage at large $|\tilde{s}|$. At the same time, θ_∞ is always strictly below 1, so we arbitrarily set it equal to 0.999. This is a tuning parameter that a user may set; similar results will be obtained with other values of θ_∞ less than but near 1. We will assume $\theta_\infty = 0.999$ for the remainder of this paper and hence, from (31) we have that $\tau_2^2 = 499.5 \times \sigma^2$.

3.1.1 Mean Squared Error profiles

It can be shown that the IMSE of the estimated image function $\hat{\mathbf{m}}$ depends on the estimated coefficient vector $\hat{\beta}$ entirely through the total MSE of $\hat{\beta}$. Thus, to determine the appropriate values of θ_0 and θ_r for the “unconstrained” estimator, it is insightful to study how these two values impact the shape of the MSE profile of the estimator $\hat{\beta}$ as a function of the true coefficient $\tilde{\beta}$.

Since $\hat{\beta} = f(\tilde{s}) \cdot \tilde{s}$, the MSE of $\hat{\beta}$ in (18) is given by

$$\text{MSE}(\hat{\beta}) = \text{E}(\hat{\beta} - \tilde{\beta})^2 = \int_{-\infty}^{\infty} (\hat{\beta} - \tilde{\beta})^2 \frac{1}{\sqrt{\pi}\sigma} \exp\left(-\frac{(\tilde{s} - \tilde{\beta})^2}{\sigma^2}\right) d\tilde{s}. \quad (33)$$

This quantity can be computed using a Gaussian Quadrature approximation, described in [18]. From the expression in (33), it is clear that MSE scaled by σ^2 , i.e., MSE/σ^2 , depends on $\tilde{\beta}$ only through the quantity $\tilde{\beta}/\sigma$. This means that any conclusions about the MSE profile scaled by σ^2 as a function of $\tilde{\beta}/\sigma$ are independent of the value of σ^2 .

Figure 2 shows the effect of changing the value of θ_r on the scaled MSE profile for

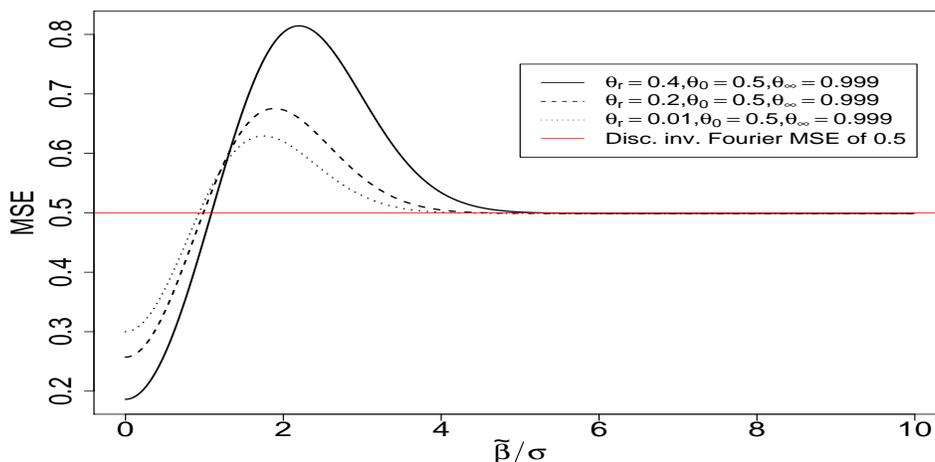


Figure 2: Scaled MSE profile as a function of $\tilde{\beta}/\sigma$ for the discrete inverse Fourier transform estimator (red line) and the “unconstrained” shrinkage estimator with $\theta_\infty = 0.999$, $\theta_0 = 0.5$ and $\theta_r = 0.01$ (dotted line), 0.2 (dashed line), and 0.4 (solid line).

positive values of $\tilde{\beta}/\sigma$ (as in the case of the shrinkage factor, the MSE profile for negative values of $\tilde{\beta}/\sigma$ is just the mirror image of the profile for positive values). Figure 3, on the other hand, shows the effect of changing the value of θ_0 on the scaled MSE profile. Both figures show the scaled MSE of 0.5 for the discrete inverse Fourier transform estimator $\hat{\beta}_{LS} = \tilde{s}$. In both plots, $\theta_\infty = 0.999$.

Both plots show that regardless of the values of θ_r and θ_0 , the scaled MSE profile of the shrinkage estimator of $\tilde{\beta}$ follows the same general shape. Starting below the discrete inverse Fourier transform estimator’s MSE when $\tilde{\beta}/\sigma$ is very close to 0, the shrinkage estimator’s MSE increases and surpasses that of the discrete inverse Fourier estimator as the value of $\tilde{\beta}/\sigma$ increases away from 0, but reaches a maximum and

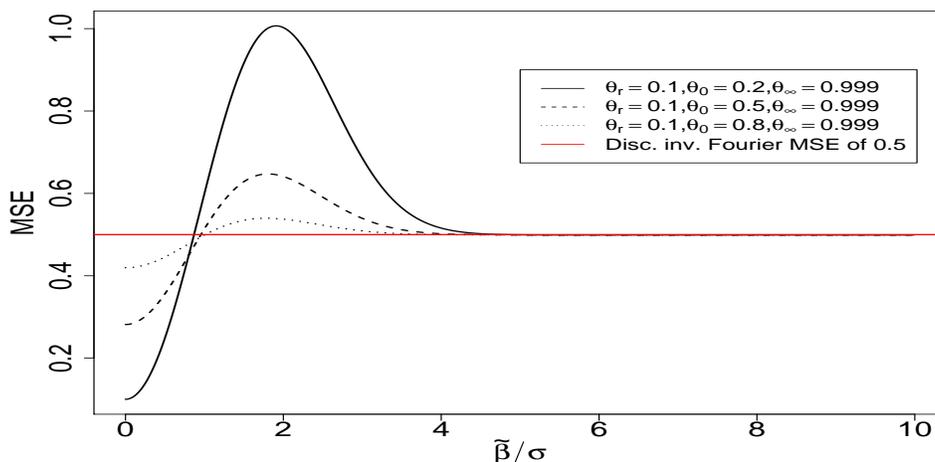


Figure 3: Scaled MSE profile as a function of $\tilde{\beta}/\sigma$ for the discrete inverse Fourier transform estimator (red line) and the “unconstrained” shrinkage estimator with $\theta_\infty = 0.999$, $\theta_r = 0.1$ and $\theta_0 = 0.2$ (solid line), 0.5 (dashed line), and 0.8 (dotted line).

then decreases to just below the discrete inverse Fourier estimator’s MSE, finally slowly growing quadratically and exceeding the discrete inverse Fourier MSE (this last stage is not shown on the plots because this happens at very large values of $\tilde{\beta}/\sigma$).

These observations echo the discussion in Section 2.3.2: the shrinkage estimator performs much better with respect to the discrete inverse Fourier estimator when the true coefficient value is near 0 because in that range the estimated coefficient is shrunk to the correct value of 0 and the shrinkage estimator marginally outperforms the discrete inverse Fourier estimator when the true coefficient value is far enough away from 0 because, correctly, very little shrinkage is applied to the estimated coefficient in this range (but only up to $\tilde{\beta}/\sigma = 32$ when $\theta_\infty = 0.999$). The “medium” values of $\tilde{\beta}/\sigma$

are problematic for the shrinkage estimator as its MSE in this range can dramatically exceed the discrete inverse Fourier MSE. Note that the lower endpoint of this range depends entirely on the values of θ_r and θ_0 .

The values of θ_r and θ_0 together also determine the amount of improvement in MSE at values of $\tilde{\beta}/\sigma$ close to 0: the higher values of θ_r produce a bigger improvement near 0, but at the price of a much worse performance in the “medium” range, while the higher values of θ_0 result in the opposite effect. This is not surprising since the higher value of θ_r corresponds to a slower rate of change in the shrinkage factor, so a great amount of shrinkage toward 0 “lingers” for a greater range of $\tilde{\beta}/\sigma$ values, making the estimator a bigger gamble with greater returns near 0, but greater losses in the “medium” range. The higher value of θ_0 implies less extreme shrinkage at 0 and, hence, everywhere else, making the shrinkage estimator a safer bet: the reduction in MSE is not as dramatic near 0, but at the same time the losses incurred in the “medium” range are minimized.

3.1.2 Distribution of $\tilde{\beta}/\sigma$ values

From the discussion so far, it is clear that the optimal values for θ_r and θ_0 thus completely depend on the distribution of $\tilde{\beta}/\sigma$ values that are typically encountered in MRI. In order to obtain such a distribution, a 3-dimensional image of a normal adult brain was simulated using the MRI Simulator software package developed at the McConnell Brain Imaging Center of the Montreal Neurological Institute at McGill University [19]. The reason simulated data were used for this purpose is because using real (non-simulated)

data makes it impossible to distinguish the true image from noise.

The dimension of the simulated image is $128 \times 128 \times 128$. Thus, 128 2-dimensional slices of size 128×128 can be extracted from this image in each of the three directions. Figure 6 shows nine such slices. Since $n = d = 128$ in this case, each of the 2-dimensional slices in turn produces 2×128^2 values of true $\tilde{\beta}/\sigma$ (a multiplier of 2 is due to the fact that there is a real and an imaginary component for each coefficient). Thus, a total of $2 \times 128^3 = 4,194,304$ values of $\tilde{\beta}/\sigma$ can be obtained in each of the three directions. The results discussed below are very similar in all three directions; hence, the findings for only one direction, top-to-bottom (also known as axial), are illustrated here.

The values of θ_r and θ_0 do not affect the MSE profile of the shrinkage estimator once the shrinkage MSE surpasses that of the discrete inverse Fourier estimator for the second time, which happens when $\tilde{\beta}/\sigma$ equal to 32 in magnitude if $\theta_\infty = 0.999$. Therefore, to determine the values of θ_r and θ_0 we only need to consider the distribution of values of $\tilde{\beta}/\sigma$ less than 32 in magnitude.

Since typical signal-to-noise ratios in MRI fall between 50 and 100 [20], a reasonable range for the values of σ is 0.5% to 2.5% of the median image value for each slice. While it is impossible to consider every value of σ in the above range, it is most informative to study the endpoint cases, i.e., $\sigma = 0.005 \times \tilde{m}$ and $\sigma = 0.025 \times \tilde{m}$, where \tilde{m} is the median value of the image function m for a given slice. If the conclusions are the same or similar for these two cases, then we can assume that they also hold for all the values of σ in the above range.

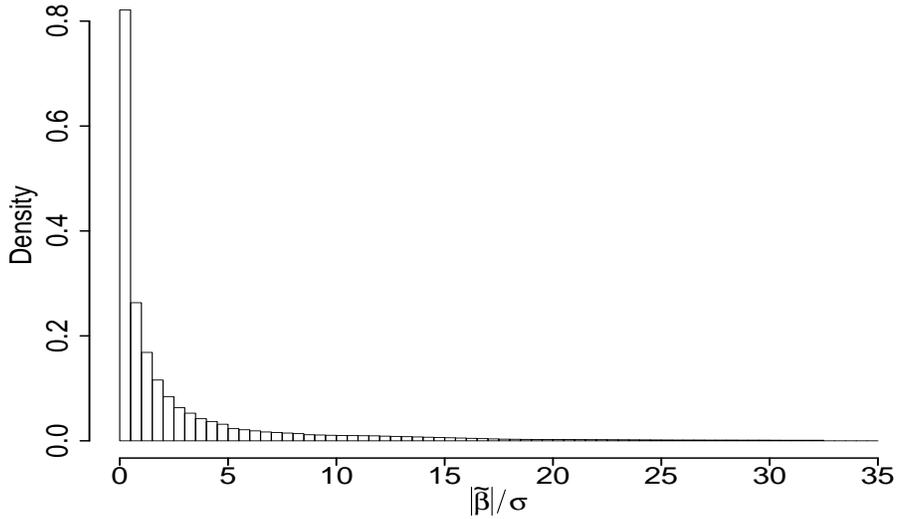


Figure 4: Density histogram of $|\tilde{\beta}|/\sigma$ values below 32 for $\sigma = 0.005 \times \tilde{m}(x)$.

Figure 4 shows the density histogram of absolute values of $\tilde{\beta}/\sigma$ below 32 when $\sigma = 0.005 \times \tilde{m}$. The histogram suggests that values of $\tilde{\beta}/\sigma$ close to 0 are prevalent. As expected, this trend is even more pronounced at the high level of noise, i.e., $\sigma = 0.025 \times \tilde{m}$ (the corresponding histogram is not shown here). Although the histogram provides us with visual insight about the distribution of $\tilde{\beta}/\sigma$, a more formal approach to setting the values of θ_r and θ_0 is needed.

3.1.3 Contour plots of MSE/σ^2

For a given combination of θ_r and θ_0 , we can calculate the average value of MSE/σ^2 of all 4,194,304 values of $\tilde{\beta}/\sigma$ and determine the optimal values of θ_r and θ_0 by comparing the average MSE/σ^2 values yielded by various combinations of the two parameters.

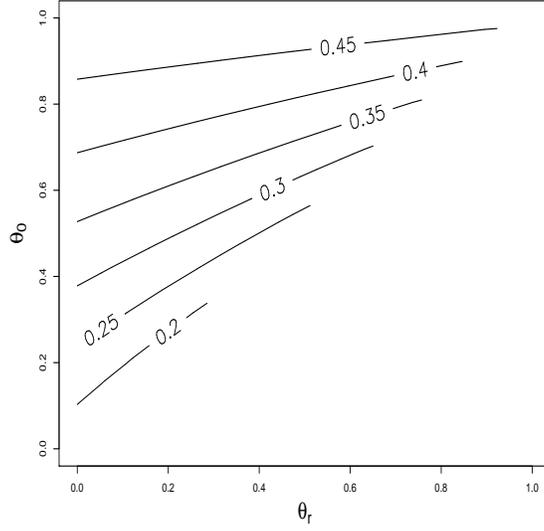


Figure 5: Contour plot of average MSE/σ^2 values of the “unconstrained” shrinkage estimator with $\theta_\infty = 0.999$ and $\sigma = 0.005 \times \tilde{m}$.

Figure 5 shows the contour plot of the average MSE/σ^2 values for all combinations of θ_r and $\theta_0 = 0.005, 0.055, 0.105, \dots, 0.955$, such that $\theta_r < \theta_0$ (a total of 190 combinations) with $\sigma = 0.005 \times \tilde{m}$ and $\theta_\infty = 0.999$. The equivalent contour plot at the high level of noise, $\sigma = 0.025 \times \tilde{m}$, appears almost identical to the one in Figure 5 and is therefore not shown here.

The contour plots show that at both levels of noise, the average MSE/σ^2 of the shrinkage estimator increases toward the discrete inverse Fourier MSE/σ^2 value of 0.5 as θ_r and θ_0 increase toward 1, as expected, because when these two parameters approach 1, the shrinkage estimator tends to the discrete inverse Fourier estimator. For all combinations, the average MSE/σ^2 is below 0.5, suggesting that any amount of

shrinkage is beneficial. However, it is most beneficial at lower values of the two parameters for both noise levels. This is not surprising as both the histogram in Figure 4 and the equivalent histogram for the higher level of noise (not shown) suggest that the values of $\tilde{\beta}/\sigma$ tend to be close to 0. In particular, when $\sigma = 0.005 \times \tilde{m}$, the minimum MSE/σ^2 (0.2) occurs at approximately $0.1 < \theta_0 < 0.3$ and $\theta_r = 1.25 \cdot \theta_0 - 0.125$. When $\sigma = 0.025 \times \tilde{m}$, the minimum MSE/σ^2 (0.15) is produced by $0.15 < \theta_0 < 0.4$ and $\theta_r = 1.4 \cdot \theta_0 - 0.2$. Note that, as expected, shrinkage is more beneficial for the high level of noise. Putting the two ranges of optimal values of θ_r and θ_0 together, we set $\theta_r = 0.18$ and $\theta_0 = 0.25$. Since $\theta_\infty = 0.999$, (30), (31), and (32) imply that the prior variance for each element of $\tilde{\beta}$ has a 2-point distribution given by

$$T_{jvq}^2 = \begin{cases} 0.11 \times \sigma^2 & \text{with probability } 0.27 \\ 499.5 \times \sigma^2 & \text{with probability } 0.73 \end{cases} \quad (34)$$

Thus, 27% of the time the coefficient's prior is a relatively tight $N(0, 0.11 \times \sigma^2)$ distribution, favoring shrinkage, while the remaining 73% of the time the prior is a diffuse $N(0, 499.5 \times \sigma^2)$ distribution, suggesting little or no shrinkage.

3.2 “Constrained” Shrinkage Estimator

The general procedure discussed in Section 3.1 was used with minor adaptations to obtain the prior parameter values in the case of the “constrained” shrinkage estimator. This resulted in the following two-point prior mixture distribution for the variance of

each element of β :

$$T_{jv}^2 = \begin{cases} 0.11 \times \sigma^2 & \text{with probability } 0.21 \\ 999 \times \sigma^2 & \text{with probability } 0.79 \end{cases} \quad (35)$$

Thus, 21% of the time the coefficient's prior is a tight $N(0, 0.11 \times \sigma^2)$ prior distribution, while the remaining 79% of the time the prior is a diffuse $N(0, 999 \times \sigma^2)$ distribution.

4 Performance of the Penalized Likelihood Approach

We can now compare the performance of the proposed penalized likelihood approach to that of the currently used discrete inverse Fourier transform in terms of the IMSE/σ^2 of the estimated image using simulated data. We scale IMSE by σ^2 because this gives a more meaningful comparison across the range of σ^2 values that are typically encountered in MRI.

It can be shown that the scaled IMSE of the discrete inverse Fourier transform approach is given by

$$\frac{\text{IMSE}(\hat{\mathbf{m}})}{\sigma^2} = \frac{n^2}{2} + \frac{1}{\sigma^2} \left(\int_X \int_Y m^2(x, y) dx dy - \sum_{j=1}^n \sum_{v=1}^n \|\alpha_{jv}\|^2 \right), \quad (36)$$

where α_{jv} is the normalized $(j, v)^{th}$ Fourier coefficient equal to

$$\alpha_{jv} = \frac{1}{\|b_{jv}(x, y)\|} \int_X \int_Y b_{jv}(x, y) m(x, y) dx dy, \quad (37)$$

while the scaled IMSE of both shrinkage estimator approaches is given by

$$\frac{\text{IMSE}(\hat{\mathbf{m}})}{\sigma^2} = \sum_{j=1}^n \sum_{v=1}^n \frac{\text{MSE}(\hat{\alpha}_{jv})}{\sigma^2} + \frac{1}{\sigma^2} \left(\int_X \int_Y m^2(x, y) dx dy - \sum_{j=1}^n \sum_{v=1}^n \|\alpha_{jv}\|^2 \right) \quad (38)$$

with α_{jv} as defined in (37).

Note that the second term in the sum on the right-hand side of (36) and (38) is the same. Therefore, the comparison of the IMSEs of the two estimators amounts to the comparison of the first terms in these two expressions. This comparison highlights the fact that the discrete inverse Fourier transform approach has an advantage over the shrinkage estimator only if the amount of noise in the data is very small because in that case the average MSE/σ^2 of the shrinkage estimator of the coefficient is above 0.5. When the amount of noise in the data is not negligible, however, as in the case of MRI and fMRI, the shrinkage estimator of the image yields a better performance in terms of the IMSE, as is illustrated in the next section.

4.1 Simulated MRI Data

The same 3-dimensional simulated brain image as the one introduced in Section 3.1.2 was used. Nine 2-dimensional slices were chosen from the image at one-quarter length, one-half length, and three-quarters length along each of the three directions. These slices are shown in Figure 6.

Since the slices are of the size 128×128 , n is equal to 128 in the first summand in (36). MSEs scaled by σ^2 in the first summand of (38) are obtained using a Gaussian Quadrature approximation, as before, with prior parameters set to values given in (34) and (35). Tables 1 and 2 show the total variance, the total bias squared, and the

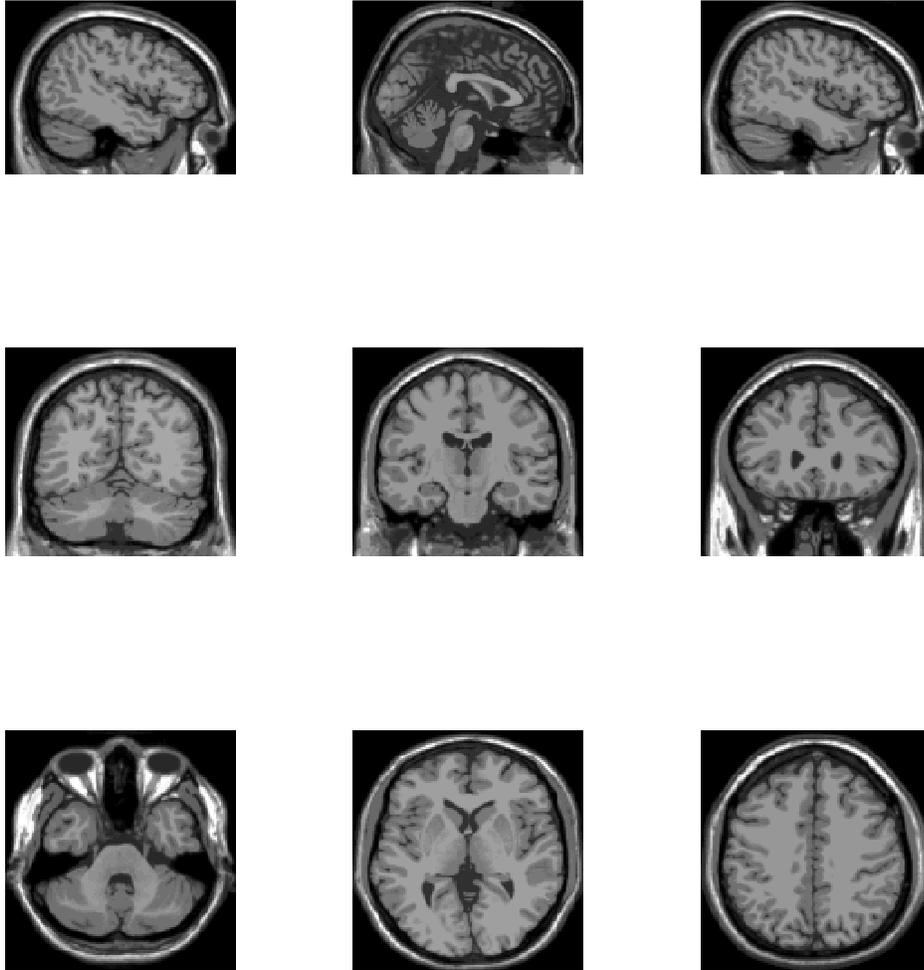


Figure 6: Nine slices of the simulated 3-dimensional image.

Slice	$\sigma = 0.005 \times \tilde{m}$			$\sigma = 0.025 \times \tilde{m}$		
	$\frac{\sum V(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum B^2(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum \text{MSE}(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum V(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum B^2(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum \text{MSE}(\hat{\alpha}_{jv})}{\sigma^2}$
1	1680.58	91.96	1772.54	1532.46	24.10	1556.56
2	1797.82	150.15	1947.97	1548.54	37.92	1586.46
3	1685.44	92.68	1778.12	1531.74	23.99	1555.73
4	1628.94	72.78	1701.72	1518.44	14.71	1533.15
5	1640.86	83.14	1724.00	1519.79	16.55	1536.34
6	1680.68	99.41	1780.09	1528.69	21.91	1550.60
7	1731.71	108.24	1839.95	1545.52	31.57	1577.09
8	1638.14	75.22	1713.36	1514.64	14.92	1529.56
9	1650.50	82.39	1732.89	1519.60	18.63	1538.23

Table 1: Total scaled variance, total scaled bias squared, and total scaled MSE of the normalized estimated coefficients for the “unconstrained” shrinkage estimator.

total MSE of the normalized estimated coefficients (the first term on the right-hand side of (38)), all scaled by σ^2 , for each of the nine slices at $\sigma = 0.005 \times \tilde{m}$, and $\sigma = 0.025 \times \tilde{m}$ for the “unconstrained” and the “constrained” shrinkage estimators, respectively. The values in the first and fourth columns of both tables should be compared to $128^2/2 = 8192$, the total scaled variance for the discrete inverse Fourier transform estimator. Since the discrete inverse Fourier transform estimator is unbiased, its total scaled bias squared and total scaled MSE are 0 and 8192, respectively. Thus,

Slice	$\sigma = 0.005 \times \tilde{m}$			$\sigma = 0.025 \times \tilde{m}$		
	$\frac{\sum V(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum B^2(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum \text{MSE}(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum V(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum B^2(\hat{\alpha}_{jv})}{\sigma^2}$	$\frac{\sum \text{MSE}(\hat{\alpha}_{jv})}{\sigma^2}$
1	1142.10	118.57	1260.67	995.96	31.31	1027.278
2	1250.44	189.02	1439.46	1009.93	49.51	1059.44
3	1144.69	116.77	1261.46	995.48	32.57	1028.05
4	1090.70	91.10	1181.80	985.49	20.03	1005.52
5	1100.84	103.67	1204.51	986.20	22.01	1008.21
6	1137.99	129.17	1267.16	995.33	29.77	1025.10
7	1186.55	133.03	1319.58	1012.16	40.03	1052.19
8	1097.35	94.01	1191.36	980.91	19.46	1000.37
9	1104.33	102.92	1207.25	984.46	25.18	1009.64

Table 2: Total scaled variance, total scaled bias squared, and total scaled MSE of the normalized estimated coefficients for the “constrained” shrinkage estimator.

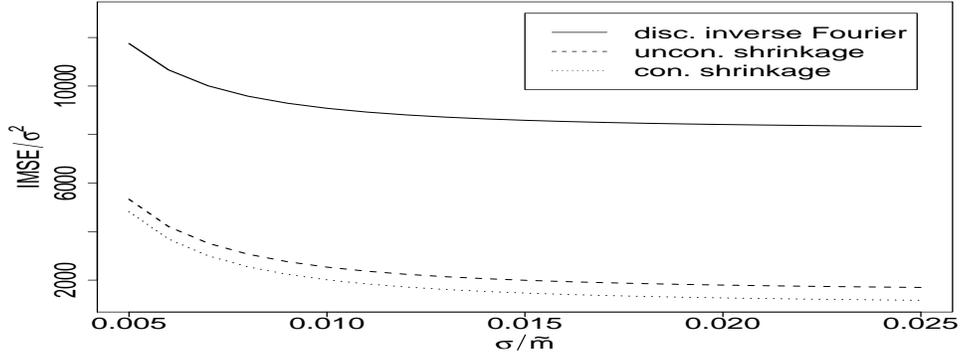


Figure 7: IMSE/σ^2 of the estimated image from one of the nine 2-dimensional slices versus σ/\tilde{m} for the discrete inverse Fourier transform estimator (solid line), the “unconstrained” shrinkage estimator (dashed line), and the “constrained” shrinkage estimator (dotted line).

the values in the second and fifth columns should be compared to 0, while the values in the third and sixth columns should be compared to 8192. As expected, the values in the second and the fifth columns are greater than the analogous value for the discrete inverse Fourier transform estimator. However, the values in the other columns are much lower than the analogous discrete inverse Fourier transform counterparts.

Figure 7 shows the plot of the scaled IMSE for one of the nine slices and each of the three estimators against σ/\tilde{m} ranging from 0.005 to 0.025, as discussed in Section 3.1.2. The equivalent plots for the other 8 slices are very similar and are therefore not shown here. The most striking feature of the plot is that both shrinkage estimators have a much lower IMSE/σ^2 than the discrete inverse Fourier transform estimator for

all values of σ in the range that we have considered. The ratios of the IMSE of the “unconstrained” estimator relative to that of the discrete inverse Fourier transform estimator range from 0.39 to 0.71 when $\sigma = 0.005 \times \tilde{m}$ and from 0.19 to 0.24 when $\sigma = 0.025 \times \tilde{m}$. The equivalent ratios for the “constrained” estimator range from 0.34 to 0.69 when $\sigma = 0.005 \times \tilde{m}$ and from 0.13 to 0.18 when $\sigma = 0.025 \times \tilde{m}$. As the values of σ increase, the ratios decrease, as expected, because the shrinkage estimator controls for both bias and variance and is especially beneficial at higher levels of noise.

The dramatic decrease in the scaled IMSE exhibited by the shrinkage estimators is due to the fact that the values of the coefficients of the Fourier basis functions are favorable to shrinkage toward 0 as a great majority of them are close to 0 when scaled by the value of σ . The increase in bias was greatly compensated by a large decrease in variance relative to the discrete inverse Fourier transform estimator. Specifically, for the hyperparameter values obtained in Section 3.1.3, the scaled MSE of the “unconstrained” shrinkage estimator is below that of the discrete inverse Fourier transform estimator when the value of $\tilde{\beta}/\sigma$ is either below 0.91 or between 4.99 and 32. In our simulated data, at the low level of noise, 56% and 16% of $\tilde{\beta}/\sigma$ values were in the two ranges, respectively. The equivalent percentages at the high level of noise were 82% and 1%. Thus, the “unconstrained” shrinkage estimator’s MSE was below that of the discrete inverse Fourier estimator for a total of 72% of the values at the low level of noise and 83% of the values at the high level of noise, yielding a much lower IMSE for the entire range of noise levels that we considered.

Figure 7 also shows that the “constrained” estimator performs consistently better than the “unconstrained” one, although the difference between these two estimators is not as dramatic as the difference between them and the currently used discrete inverse Fourier estimator. The ratios of the IMSE of the “constrained” estimator to the IMSE of the “unconstrained” one range from 0.87 to 0.97 for $\sigma = 0.005 \times \tilde{m}$ and from 0.67 to 0.75 for $\sigma = 0.025 \times \tilde{m}$. As can be seen from Tables 1 and 2, the better performance of the “constrained” estimator is due to the fact that its total scaled variance is reduced enough relative to the variance of the “unconstrained” one to compensate for its higher total scaled bias squared. The slightly better performance of the “constrained” estimator over its “unconstrained” counterpart at all levels of noise is a consequence of the slightly higher percentages of the scaled coefficients in the “optimal” ranges: for the hyperparameters specified in Section 3.2, 74% and 87% of the values of β/σ were in the ranges where the “constrained” estimator’s scaled MSE is below that of the discrete inverse Fourier estimator for the low and high levels of noise, respectively.

To illustrate the “unconstrained” and “constrained” shrinkage approaches and compare them to the discrete inverse Fourier transform approach, the second image in the bottom row of Figure 6 was used to generate the signal data both at low and high levels of noise. Figure 8 shows the image estimates resulting from each of the three methods at the two noise levels. The plot makes it clear that compared to the estimate obtained via the discrete inverse Fourier method, the two shrinkage methods produce less noisy

estimates that are more true to the original image at both levels of noise. As expected, this difference is more pronounced at the high level of noise. In addition, there does not appear to be a great visual difference between the two shrinkage estimators, which agrees with the finding that their IMSE values are very close.

Figure 9 shows the residuals, i.e., the differences between the true image and the estimated image, at each location of the image. As expected, there is no pattern in the residuals for the discrete inverse Fourier method as the Fourier transform is orthogonal and thus preserves the iid error structure of the assumed model. However, there is a pattern in the residuals for the two shrinkage estimators: the edge of the brain is outlined as the image intensity values produced by the shrinkage estimators at the edges are systematically biased. This is due to the phenomenon referred to as partial voluming: the pixels at the edges have much more highly variable magnetic properties than the pixels in the other parts of the image, leading to a greater amount of noise in the signal coming from these locations. The shrinkage estimator, therefore, is particularly aggressive at reducing the variance at the edges, but does so at the expense of introducing a greater amount of bias.

The edge outline pattern is much less apparent for the “constrained” estimator, suggesting that it may be a better estimator than the “unconstrained” one. This also agrees with the conclusion that the IMSE of the “constrained” estimator is lower than that of the “unconstrained” estimator, although, as reported earlier, the difference is not very large. The distributions of the residuals r for all three estimators and both

Method	Low noise		High noise	
	\bar{r}/σ	s_r/σ	\bar{r}/σ	s_r/σ
Discrete Inverse Fourier Transform	-0.00104	1.0057	-0.00104	1.0057
“Unconstrained” Fourier Shrinkage	-0.00656	0.7202	-0.00262	0.6046
“Constrained” Fourier Shrinkage	-0.00656	0.6973	-0.00262	0.5381

Table 3: Summary statistics for the residual images obtained via discrete inverse Fourier transform, “unconstrained” shrinkage and “constrained” shrinkage methods.

noise levels are all symmetric with the summary statistics shown in Table 3:

4.2 Computational Time and Implementation

The addition of the shrinkage steps to the current reconstruction processing routine, as implemented in the FIASCO software package [21], results in an increase of 1.82 seconds in computational time for a 128×128 image on a 2.4 GHz Intel Pentium 4 machine. The coding of the procedure is very straightforward as it involves multiplying the individual signal observations, either in the raw or already partially processed form, by shrinkage factors. Thus, the method incurs very low computational and coding implementation costs and can, therefore, be very easily added to the existing processing streams for MRI and fMRI data.

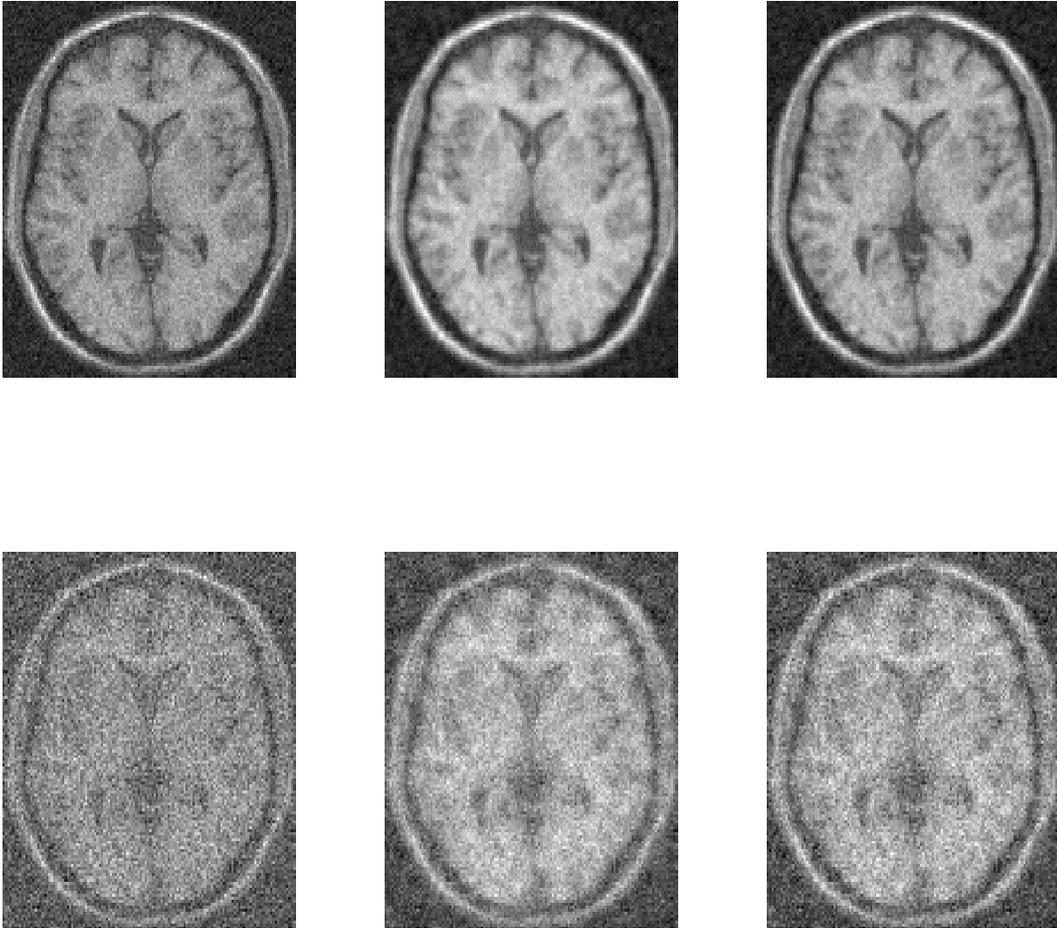


Figure 8: The estimates of the image using the discrete inverse Fourier transform (first column), the “unconstrained” Fourier shrinkage (second column) and the “constrained” Fourier shrinkage (third column) approaches at low (top row) and high (bottom row) levels of noise.

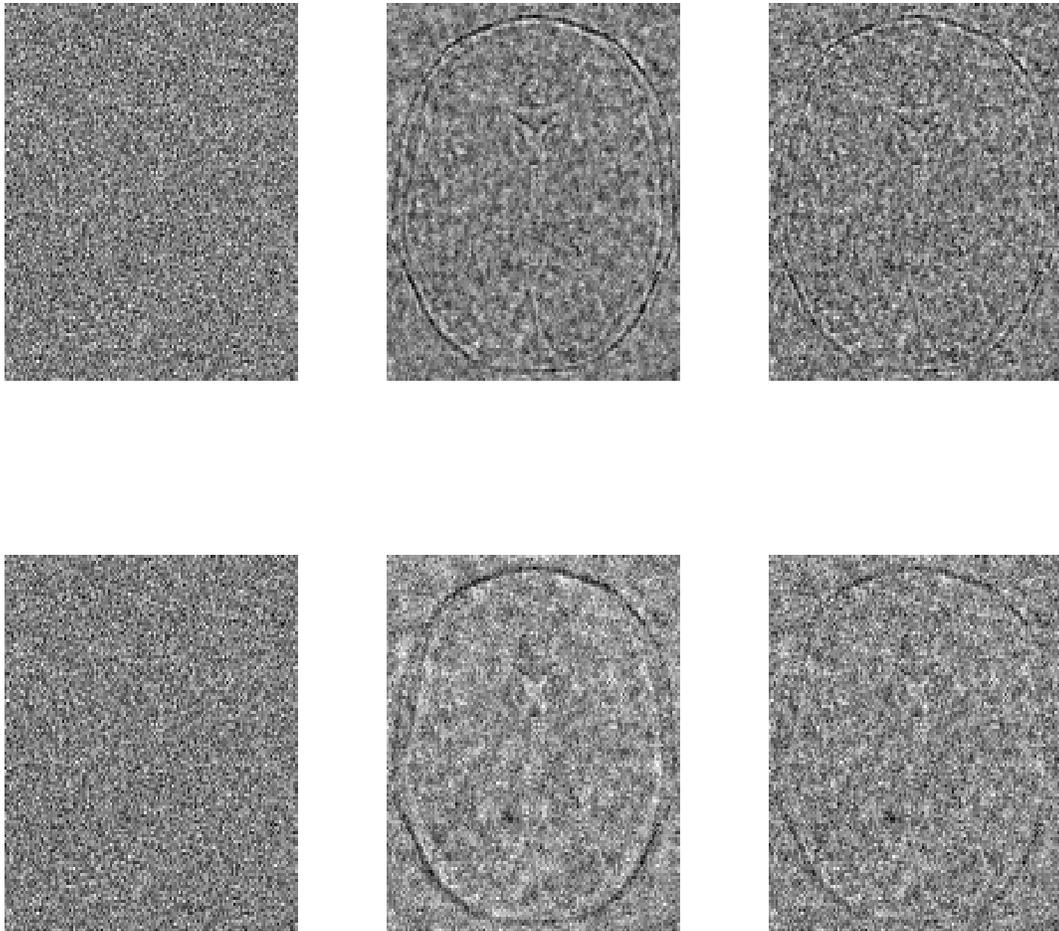


Figure 9: The differences between the estimated image and the true image for the discrete inverse Fourier transform (first column), the “unconstrained” Fourier shrinkage (second column) and the “constrained” Fourier shrinkage (third column) approaches at low (top row) and high (bottom row) levels of noise.

5 Discussion

The work discussed in this paper leads to the conclusion that a special case of penalized likelihood, nonlinear shrinkage estimation, has great potential for image reconstruction in the context of Magnetic Resonance Imaging. In particular, using simulated brain images, we have shown that the shrinkage estimators outperform the currently used discrete inverse Fourier transform approach in terms of the integrated mean squared error. Our “constrained” shrinkage estimator performs the best, although the difference between it and our “unconstrained” estimator does not appear to be substantial. While shrinkage increases bias relative to the discrete inverse Fourier transform estimator, its reduction in variance is much greater, thus yielding a much lower IMSE. As alluded to in Section 2.3.2, the adaptability of the non-linear shrinkage estimator is the key to its better performance relative to the currently used discrete inverse Fourier estimator, as well as the linear shrinkage estimator.

In our simulations, the reductions in IMSE due to shrinkage were especially great at lower levels of noise, and in general, the benefit of shrinkage increases as the signal-to-noise ratio (SNR) decreases. Note that in some cases, SNRs in MRI fall even below the typical range of 50-100 that was cited earlier, making our approach even more beneficial relative to the currently used discrete inverse Fourier transform.

Aside from its desirable statistical properties, our nonlinear shrinkage method has several practical advantages that are important to the MRI and fMRI users and re-

searchers. As discussed in Section 4.2, computationally, it is extremely fast and easy to implement in terms of coding, making it a very easy addition to the existing MRI and fMRI data processing streams. In addition, our method can be easily combined with various recent improvements on the basic model for the MRI signal. For example, Sedarat and Nishimura [22] introduced an optimal gridding reconstruction method to deal with nonuniform sampling in k -space. Raw k -space data can first be processed using their approach of determining the optimal density compensation factors in order to produce an approximation to the Cartesian grid, and then our shrinkage procedure can be applied to the interpolated square-grid data. Schomberg and Timmer [23] addressed the same issue. Their method involves convolving the observed signal space data with a weight function and then deconvolving it in the image space. Their approach can also be easily combined with our method, but in the opposite order: first, the signal space data can be shrunk using our method and then the shrunk data can be put through the convolution and deconvolution steps of the algorithm the authors proposed.

Our approach can also be used in conjunction with the recent work of Rowe [24], in which he models both the phase and the magnitude of the the k -space signal, rather than just the magnitude alone, as is normally done in fMRI. Our method can be combined with models for the image space data, as well. For example, LaConte et al. [25] proposed a method for denoising the images obtained in event-related fMRI experiments, using wavelet transform-based Wiener filtering. The images can first be obtained using our reconstruction technique and then filtered using their approach.

It should be noted that some extensions of the basic MRI model do not lend themselves as easily to being used in conjunction with our approach. One such example is the work by Sutton et al. [11], discussed in Section 2.3.2. Because they model inhomogeneities in the field, the relationship between the signal data and the underlying image is much more mathematically complex than a Fourier transform, so there is no easy way to put their approach in the same framework as ours. Another example is the work by Twieg [26]. His model allows for modulation in the amplitude and phase of the signal in the course of the acquisition, in contrast to the conventional assumption that these signal parameters are constant. As a result, just as in the work by Sutton et al., the relationship between the signal and the image cannot be modeled as a Fourier transform, making it impossible to combine our approach easily with his.

While this incompatibility is a limitation of our approach, it can be argued that the method strikes a compromise between performing a better-than-basic image reconstruction in terms of filtering out the noise in the signal data and computational feasibility. Incorporating some extensions of the basic MR model can come at a price. For example, in contrast to our method's computational time of 1.82 seconds, Sutton et al. report that their iterative procedure requires 4.4 seconds (20 iterations at 0.22 seconds per iteration) to reconstruct a 64×64 image on a 2 GHz Pentium 4 machine (using Matlab software package), while Twieg's method takes 120–140 minutes for a 64×64 image on a 450 MHz Macintosh G4 computer (also using Matlab). Furthermore, the complexity of the programming code needed to implement these authors' procedures is

much greater than in the case of our method. Moreover, the fact that our model does not take into account special problems that arise in MRI and fMRI makes our approach more universal and flexible. It can be used in any context, in which the measurements and the quantities of interest can be related via a linear transform. In neuroimaging, one example is Positron Emission Tomography (PET), in which the data are Radon transforms of the image.

One direction for future work is to develop a computationally feasible and efficient shrinkage method in conjunction with basis sets other than the Fourier basis used in this work. In particular, the Haar wavelet basis may be a good candidate as its step function shape can facilitate edge reconstruction and modeling of the jumps in the image function that occur as a result of motion.

Finally, it is important to note that we used simulated, rather than real, data to illustrate the advantage of the shrinkage estimator approach over the discrete inverse Fourier transform approach. Using simulated data made the comparison of the two approaches possible; however, certain issues that arise in dealing with real data, such as dephasing of the signal caused by motion, could not be addressed as a result. Thus, there is no true “gold standard” for assessing the performance of any approach to reconstruction, and further work is needed in this area. Nonetheless, comparing the performance of our approach to the current method using real MR data is one of the high-priority items on our future work agenda.

Acknowledgments

The authors are grateful to the referees, as well as Bill Eddy, Brian Junker, Joel Greenhouse and Rob Kass for their helpful comments on an earlier version of the manuscript.

References

- [1] Committee on the Mathematics and Physics of Emerging Dynamic Biomedical Imaging of National Research Council at Institute of Medicine, editors. *Mathematics and Physics of Emerging Biomedical Imaging*. National Academy Press: Washington, D.C., 1996.
- [2] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224, 1995.
- [3] T.T. Cai. Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Annals of Statistics*, 27:989–924, 1999.
- [4] G. Bradley and D. Stark. *Magnetic Resonance Imaging*. C.V. Mosby: St. Louis, 2nd edition, 1992.
- [5] N. Lange. Tutorial in biostatistics: Statistical approaches to human brain mapping by functional Magnetic Resonance Imaging. *Statistics in Medicine*, 15:389–428, 1996.
- [6] R. Edelman, J. Hesselink, and M. Zlatkin. *Clinical Magnetic Resonance Imaging*. Saunders: Philadelphia, 2nd edition, 1996.
- [7] M. Haacke. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley: New York, 1999.

- [8] S. Bushong. *Magnetic Resonance Imaging: Physical and Biological Principles*. C.V. Mosby: St. Louis, 2nd edition, 1995.
- [9] G. Wahba and G. Kimeldorf. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Journal of Mathematical Statistics*, 41:495–502, 1970.
- [10] A. Antoniadis and J. Fan. Regularization of wavelet approximations. *Journal of American Statistical Association*, 96:939–967, 2001.
- [11] B.P. Sutton, D.C. Noll, and J.A. Fessler. Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities. *IEEE Transactions on Medical Imaging*, 22:178–188, 2002.
- [12] J. Besag. On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society, series B*, 48:259–302, 1986.
- [13] X. Hu, V. Johnson, W.H. Wong, and Chen C.-T. Bayesian image processing in Magnetic Resonance Imaging. *Magnetic Resonance Imaging*, 9:611–620, 1991.
- [14] A. Chipman, E. Kolaczyk, and R. McCulloch. Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, 92:1413–1421, 1997.
- [15] M. Clyde, G. Parmigiani, and B. Vidakovic. Multiple shrinkage and subset selection in wavelets. *Biometrika*, 85:391–401, 1998.

- [16] B. Vidakovic. Non-linear wavelet shrinkage with Bayes rules and Bayes factors. *Journal of the American Statistical Association*, 93:173–179, 1998.
- [17] V. Bulaevskaya. *A Penalized Likelihood Approach to Magnetic Resonance Image Reconstruction*. PhD thesis, University of Minnesota, 2002.
- [18] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover Publications, Inc., New York, 1970.
- [19] MRI Simulator Package (<http://www.bic.mni.mcgill.ca/brainweb>), 2001. McConnell Brain Imaging Center of the Montreal Neurological Institute, McGill University, Montreal, Canada.
- [20] X. Hu. Personal communication with Xiaoping Hu, Professor of Radiology at the Center for Magnetic Resonance Research at University of Minnesota, Minneapolis, Minnesota, 2001.
- [21] Functional Image Analysis of Computational Olio (FIASCO) Package (<http://www.stat.cmu.edu/~fiasco>). Department of Statistics, Carnegie Mellon University, Pittsburgh, USA.
- [22] H. Sedarat and D.G. Nishimura. On the optimality of the gridding reconstruction algorithm. *IEEE Transactions on Medical Imaging*, 19:306–317, 2000.
- [23] H. Schomberg and J. Timmer. The gridding method for image reconstruction by Fourier transform. *IEEE Transactions on Medical Imaging*, 14:596–607, 1995.

- [24] D. Rowe. Modeling both magnitude and phase of complex fMRI data. *NeuroImage*, 25:1310–1324, 2005.
- [25] S. Laconte, S.-C. Ngan, and X. Hu. Wavelet transform-based Wiener filtering of event-related fMRI data. *Magnetic Resonance in Medicine*, 44:746–757, 2000.
- [26] D. Twieg. Parsing local signal evolution directly from a single-shot MRI signal: A new approach to fMRI. *Magnetic Resonance in Medicine*, 50:1043–1052, 2003.