SCI INSTITUTE
TECHNICAL REPORT

Adaptive Markov Models for Information-Theoretic Empirical-Bayesian MRI Denoising

Suyash P. Awate and Ross T. Whitaker

UUSCI-2006-030

Scientific Computing and Imaging Institute
University of Utah
Salt Lake City, UT 84112 USA

October 13, 2006

Abstract:

This paper presents a novel framework for denoising magnetic resonance images. The framework relies on adaptive Markov-random-field (MRF) image models that we infer nonparametrically from the corrupted input data itself. The proposed denoising method produces an optimal reconstruction based on principles in empirical-Bayesian estimation and information theory. Given the corrupted input data and the knowledge of the Rician noise model, the Bayesian-denoising method bootstraps itself by estimating the uncorrupted-signal Markov statistics, by optimizing an information-theoretic metric using the expectation-maximization (EM) algorithm. It then employs the inferred uncorrupted-signal Markov statistics as an adaptive prior in a Bayesian-denoising process at each pixel. Furthermore, it proposes a novel Bayesian-inference algorithm on MRFs incorporating entropy reduction, namely iterated conditional entropy reduction (ICER). The results demonstrate that the method denoises conservatively while ensuring the preservation of most of the important features in brain-MR images. Qualitative and quantitative comparisons with the state of the art clearly depict the advantages of the proposed method.

THE UNIVERSITY OF UTAH

Adaptive Markov Models for Information-Theoretic Empirical-Bayesian MRI Denoising

Suyash P. Awate and Ross T. Whitaker

Abstract

This paper presents a novel framework for denoising magnetic resonance images. The framework relies on *adaptive* Markov-random-field (MRF) image models that we infer *nonparametrically* from the corrupted input data itself. The proposed denoising method produces an optimal reconstruction based on principles in empirical-Bayesian estimation and information theory. Given the corrupted input data and the knowledge of the Rician noise model, the Bayesian-denoising method bootstraps itself by estimating the uncorrupted-signal Markov statistics, by optimizing an information-theoretic metric using the expectation-maximization (EM) algorithm. It then employs the inferred uncorrupted-signal Markov statistics as an *adaptive prior* in a Bayesian-denoising process at each pixel. Furthermore, it proposes a novel Bayesian-inference algorithm on MRFs incorporating entropy reduction, namely *iterated conditional entropy reduction* (ICER). The results demonstrate that the method denoises conservatively while ensuring the preservation of most of the important features in brain-MR images. Qualitative and quantitative comparisons with the state of the art clearly depict the advantages of the proposed method.

1 Introduction

Over the last several decades, magnetic resonance imaging (MRI) technology has benefited from a variety of technological developments resulting in increased resolution, signal-to-noise ratio (SNR), and acquisition speed. However, fundamental trade-offs between resolution, speed, and SNR combined with scientific, clinical, and financial pressures to obtain more data more quickly, result in images that still exhibit significant levels of noise. In particular, the need for shorter acquisition times, such as in dynamic imaging, often undermines the ability to obtain images having both high resolution and high SNR. Furthermore, the

efficacy of higher-level, post processing of MR images, including tissue classification and organ segmentation, that assume specific models of tissue intensity (e.g. homogeneous), are sometimes impaired by even moderate noise levels. Hence, denoising MR images remains an important problem. From a multitude of statistical and variational denoising formulations proposed, no particular one appears as a clear winner in all relevant aspects, including the reduction of randomness and intensity bias, structure and edge preservation, generality, reliability, automation, and computational cost.

This paper presents a novel framework for denoising MR images that relies on the adaptive Markov-random-field (MRF) image model described in [1, 2]. The work in this paper is a significant modification of our previous approach in [3]. The key idea in the modeling approach is to *adapt* or infer the model from the corrupted input data itself and subsequently process the data based on the infer model. The proposed denoising method produces an optimal reconstruction based on principles in empirical-Bayesian estimation [4, 5] and information theory. The method bootstraps itself by estimating the uncorrupted-signal Markov statistics, using an information-theoretic optimality metric, from the corrupted input data and the knowledge of the Rician noise model. It then employs the *inferred* uncorrupted-signal Markov statistics as an *adaptive prior* in a Bayesian denoising process at each pixel. In this way, it avoids the need of imposing ad hoc prior models. Furthermore, it proposes a novel iterative Bayesian-inference algorithm on MRFs that incorporates entropy reduction on posterior PDFs. We call this new approach as *iterated conditional entropy reduction* (ICER). The results demonstrate that the method denoises conservatively while ensuring the preservation of most of the important features in the brain-MR images. Qualitative and quantitative comparisons with the state of the art clearly depict the advantages of the proposed method.

2 Overview of MRI Denoising

A multitude of variational methods based on partial differential equations have been developed for a wide variety of images and applications [6, 7], with some of these having applications to MRI [8, 9, 10]. However, such methods impose certain kinds of models on local image structure that are often too simple to capture the complexity of anatomical MR images. These methods, typically, do not take into account the bias introduced by Rician noise. Furthermore, such methods usually involve manual tuning of *critical* free parameters that control the conditions under which the models prefer one sort of structure over another; this has been an

impediment to the widespread adoption of these techniques.

Another class of methods relies on statistical inference on multiscale representations of images. A prominent example includes methods based on wavelet transforms. Healy *et al.* [11] were among the first to apply soft-thresholding based wavelet techniques for denoising MR images. Hilton *et al.* [12] apply a threshold-based scheme for functional-MRI data. Nowak [13], operating on the square magnitude MR image, includes a Rician noise model in the threshold-based wavelet denoising scheme and thereby corrects for the bias introduced by the noise. Pizurica *et al.* [14] rely on the prior knowledge of the correlation of wavelet coefficients that represent significant features across scales. They first detect the wavelet coefficients that correspond to these significant features and then empirically estimate the PDFs of wavelet coefficients conditioned on the significant features. They employ these probabilities in a Bayesian denoising scheme.

In our previous work [1, 15], we described UINTA which restores images by generalizing the mean-shift to incorporate neighborhood information. UINTA, however, relies neither on the knowledge of a noise model nor a prior model. Some MR-inhomogeneity correction methods are based on the quantification of information content in MR images [16, 17]. They follow from the observation that inhomogeneities increase the entropy of the 1D *gray scale* PDFs. However, entropy measures on first-order image statistics are insufficient for effective denoising; thus this paper extends the information-theoretic strategy to higher-order Markov PDFs.

The proposed method takes the *empirical-Bayes* approach [4, 5, 18], pioneered by Robbins [4, 5], for Bayesian denoising without making any ad hoc assumptions on the prior PDFs. The empirical-Bayes approach is applicable when we encounter multiple independent instances of a Bayesian decision problem (i.e. denoise each pixel) that all rely on exactly the same *fixed, but unknown*, prior PDF (i.e. uncorrupted-signal Markov PDF). In this special case, the empirical-Bayes approach allows accurate data-driven computation of the posterior PDF without the need to impose ad hoc or ill-fitting prior models. In this way, the decision procedure automatically *adapts* to the unknown prior PDFs. Robbins employed the empirical-Bayes approach to first obtain a *maximum likelihood* (ML) estimate of the prior distribution using the observations corrupted by a known noise model, and then employ the estimated prior model to compute the posterior [19]. The strategy in this paper closely follows Robbin's strategy.

Weismann *et al.* [20] address optimal image denoising using Markov statistics and empirical-Bayes approach [20]. Their discrete universal denoiser (DUDE) focuses on *discrete* signal intensities and, subse-

quently, relies on inverting the channel transition matrix (noise model) to give a closed-form estimate for source statistics from the observed statistics. The proposed method addresses continuous-valued signals, which is essential for medical-imaging applications, and thus entails estimating uncorrupted-signal statistics nonparametrically through the reduction of a *Kullback-Leibler* (KL) divergence. Snyder *et al.* [21] also use kernel density estimators for density deconvolution. The proposed approach also presents a method for practically dealing with the non-stationarity of real MRI data.

Cordy and Thomas [22] employ the expectation-maximization (EM) algorithm [23, 24] for *deconvolving* PDFs corrupted with i.i.d. additive Gaussian noise. They model the uncorrupted-signal PDF as a Gaussian mixture model, but use the EM algorithm to estimate only the weights of Gaussians in the mixture—the means and variances of the Gaussians are tuned manually before EM is applied. They constrained the Gaussians to be spread uniformly over the entire domain of the PDF. Such a strategy, however, it not likely to be effective for density estimation in high-dimensional domains because of the enormous numbers of Gaussians needed to cover the space and sparsity of the data in the space—uniformly-distributed Gaussians will tend to oversmooth the PDF structure in high-curvature regions and will be inefficient in the tails of the PDF.

3 Adaptive Markov Image Modeling

This section briefly introduces the adaptive nonparametric image model and the associated mathematical notation used in the paper. For a more detailed discussion of the modeling approach, we request the reader to refer to our previous work [2, 25].

The proposed method relies on a MRF image model. A *random field* [26] is a family of random variables (RVs) $\mathbf{X} = \{X_t\}_{t \in \mathcal{T}}$, for some index set $\mathcal{T}$. For each index t, the RV X_t is defined on some sample-space Ω. If we let $\mathcal{T}$ be a set of points defined on a discrete Cartesian grid and fix $\Omega = \omega$, we have a *realization* of the random field, $\mathbf{X}(\omega) = \mathbf{x}$, called the *digital image*. In this case, $\mathcal{T}$ is the set of grid points in the image.

MRFs capture the regularity in data, i.e. the dependencies between pixel intensities in images, by specifying only the joint PDFs of the RVs in image *neighborhoods*. This is, essentially, the *Markovity* constraint on random fields. Let us denote the set of voxels in the neighborhood of a voxel t by N_t and define a random

vector $\mathbf{Y}_t = \{X_t\}_{t\in\mathcal{N}_t}$ to denote RVs in the neighborhood of pixel t. Markovity implies that

$$P\Big(X_t|\{x_u\}_{u\in\{\mathcal{T}\setminus\{t\}\}}\Big) = P(X_t|\mathbf{y}_t). \tag{1}$$

We define a random vector $\mathbf{Z}_t = (X_t, \mathbf{Y}_t)$. We refer to the PDFs $P(X_t, \mathbf{Y}_t) = P(\mathbf{Z}_t)$ as *Markov PDFs* defined on the *feature space* $< \mathbf{z} >$.

In typical Markov image models, the Markov PDFs or the associated Gibbs PDFs are described parametrically. This means that the functional forms for the PDFs must be known *a priori*. These forms, typically, correspond to a parameterized family of PDFs, e.g. Gaussian. Typically, these parameterized families of PDFs are relatively simple and have limited expressive power to accurately capture the structure and variability in image data [27, 28, 29]. As a result, in many instances, the data does not comply well with such parametric MRF models. This paper proposes a method [1, 2] of modeling the Markov PDFs *nonparametrically* that is driven by the input data itself. In this way the model can capture the properties underlying the data more accurately and is able to *adapt* to the input data.

In order to rely on image samples to produce nonparametric estimates of Markov statistics, we must assume that different neighborhood-intensities in the image are derived from the same PDF. Mathematically, this is known as the *stationarity* property that implies the Markov PDFs $P(\mathbf{Z}_t)$ are the same for all t. We refer to this stationary Markov PDF as $P(\mathbf{Z})$. To represent the Markov PDFs $P(\mathbf{Z})$, we use the nonparametric *Parzen-window* density estimation technique [30, 31] which gives the probability of an observed neighborhood $\mathbf{z}$ as

$$P(\mathbf{z}) \approx \frac{1}{|\mathcal{A}|}\sum_{t\in\mathcal{A}} G_d(\mathbf{z}-\mathbf{z}_t;\sigma), \tag{2}$$

where d is the dimensionality of the feature space, $G_d(\cdot)$ is the isotropic d-dimensional Gaussian function (called *kernel*) with standard deviation σ, and the set $\mathcal{A}$ is a small subset of T.

Taking $\mathcal{A} = \mathcal{T}$ increases the algorithmic complexity of the scheme. Our previous work [1, 2] describes an effective technique of choosing the Parzen-window sample A and an optimal ML-based estimate of the kernel parameter σ. To describe the strategy briefly here, for each pixel p, we select a random sample of pixels $\mathcal{A}_p$ derived from a Gaussian PDF on the image coordinates centered at pixel p and having standard deviation σ_{spatial}. For all of the results in this paper, we use $\sigma_{\text{spatial}} = 30$ voxels along each cardinal

direction and $|\mathcal{A}_p| = 2000$. We use a 9×9 neighborhood (N_t) where the intensities are masked so as to make the neighborhood more isotropic [1]. We handle image boundaries by performing the statistical estimation in the cropped feature spaces resulting from the partial neighborhoods [1].

4 Maximum-a-Posteriori Denoising using Entropy Reduction

The proposed strategy relies on several pieces of technology that interact to provide accurate, practical models of image statistics. For clarity, the discussion begins at a high level and successive sections discuss how each of these pieces is developed from the input data.

Given the noisy image $\tilde{\mathbf{x}}$, our goal is to find the *maximum-a-posteriori* (MAP) estimate $\mathbf{x}^*$ of the true image $\mathbf{x}$:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{argmax}}\ P(\mathbf{x}|\tilde{\mathbf{x}}). \tag{3}$$

Writing the posterior as

$$P(\mathbf{x}|\tilde{\mathbf{x}}) = P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \tilde{\mathbf{x}})P(\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}|\tilde{\mathbf{x}}), \tag{4}$$

where t is an arbitrary pixel, motivates us to employ an iterative restoration scheme where, starting from some initial image estimate, we update the estimate pixel-wise so that the posterior never decreases. Besag's ICM algorithm [32] gives one such strategy that updates x_t to the mode of the PDF $P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \tilde{\mathbf{x}})$. Finding modes of PDFs, however, is not always straightforward or computationally efficient. Therefore, we propose a new algorithm that updates x_t by moving it closer to the *local mode* of $P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \tilde{\mathbf{x}})$. The proposed algorithm is similar in spirit to the ICM algorithm, but relies on entropy reduction on the PDF that updates pixel intensities by performing a gradient ascent on the logarithm of the PDF—hence called *iterated conditional entropy reduction* (ICER). The relationship between reducing Shannon's entropy of Parzen-window PDFs and gradient ascent on the logarithm of the posterior PDF is described in detail in [1, 15]. It follows that by updating intensities x_t to reduce the entropy $h(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \tilde{\mathbf{x}})$ and bringing them closer to their local modes, we can guarantee non-decreasing values for $P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \tilde{\mathbf{x}})$ and, thereby, convergence.

Let us assume for simplicity that, given the true image $\mathbf{x}$, the RVs in the MRF $\tilde{\mathbf{X}}$ are *conditionally independent*. Given the stochastic noise model $P(\tilde{x}_t|x_t)$ for the Rician degradation process, conditional independence implies that the conditional probability of the observed image given the true image is

$$P(\tilde{\mathbf{x}}|\mathbf{x}) = \prod_{t\in\mathcal{T}} P(\tilde{x}_t|x_t). \tag{5}$$

Subsequently, Bayes rule gives [32]

$$\underset{x_t}{\text{argmax}}\ P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \tilde{\mathbf{x}}) = \underset{x_t}{\text{argmax}}\ P(x_t|\mathbf{y}_t)P(\tilde{x}_t|x_t), \tag{6}$$

where $P(x_t|\mathbf{y}_t)$ is the *unknown prior PDF* and $P(\tilde{x}_t|x_t)$ is the *likelihood* as determined from the Rician noise model. We model the prior using nonparametric Parzen-window density estimates with Gaussian kernels. The next section describes a method for adaptively inferring the prior based on the input data and the knowledge of the noise model.

5 Estimating Uncorrupted-Signal Markov Statistics

A Bayesian denoising framework implicitly assumes the existence of a prior statistical model of the uncorrupted signal. We can, potentially, derive such priors from a suitable database of high-SNR brain-MR images (e.g. different images of the same modality and anatomy). This effectively amounts to *training* the denoising system. Effective training data, however, is not easily available for many applications. Alternatively, we can *infer* the uncorrupted signal statistics from the observed data by making suitable assumptions. Let us assume a *fixed*, but *unknown*, Markov model $P(\mathbf{Z})$ for the uncorrupted signal that generates *all* uncorrupted data. This data, subsequently, gets corrupted by Rician noise. What we observe is only the corrupted data—the prior remains unknown. However, the following analysis provides a way of inferring the prior.

Given sufficiently-many corrupted observations, we can infer the Markov statistics of the corrupted signal accurately [1, 2]. With this knowledge of the corrupted-signal Markov statistics and knowing the properties of the corruption process, we can accurately estimate the uncorrupted-signal Markov statistics. In this way, we can *empirically* estimate the unknown prior PDF. This essentially amounts to solving an *inverse problem*, which we discuss in detail in the next section.

5.1 Forward Problem: Numerical Solution

Let us denote the Markov PDF of the corrupted signal by $P_C(\tilde{\mathbf{Z}})$. Let us model the Markov PDF of the uncorrupted signal using Parzen-windowing as:

$$P_{\mathcal{U}}(\mathbf{z}) = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} G(\mathbf{z} - \mathbf{z}_u, \sigma), \tag{7}$$

where $\{\mathbf{z}_u\}_{u \in \mathcal{U}}$ denotes the means of the Gaussians and σ their standard deviation along each dimension. This nonparametric model is a general model capable of representing arbitrary PDFs for large $|\mathcal{U}|$. The goal is to estimate the set $\{\mathbf{z}_u\}_{u \in \mathcal{U}}$ and σ, i.e. the parameters of the model, based on the knowledge of the observed corrupted-signal Markov statistics and the Rician corruption process. The key idea is as follows. An estimate of the uncorrupted-signal model parameters and the Rician noise level gives us an estimate of the corrupted-signal statistics. In the inverse-methods literature, this is the process of solving the so-called *forward problem*. We must match this estimate of the corrupted-signal Markov PDF with the Markov PDF obtained from the corrupted data by suitably updating the prior-model parameters. We use the KL-divergence measure to quantify the goodness of the match. We now analyze the noise model in detail and present a numerical scheme for solving the forward problem.

The Rician noise model corresponds to a linear *shift-variant* system whose *impulse response* for an impulse PDF located at $x \geq 0$ is

$$P(\tilde{x}|x) = \frac{\tilde{x}}{\sigma_R^2} \exp\left(- \frac{\tilde{x}^2 + x^2}{2\sigma_R^2} \right) I_0\left(\frac{\tilde{x}x}{\sigma_R^2} \right), \tag{8}$$

where σ_R is the noise level and $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind. For $x \gg 3\sigma_R$, Rician noise corrupts in a way very similar to additive independent Gaussian noise. For smaller x, though, the effect is more complex. For a Gaussian input PDF $G(x - \mu, \sigma)$, a general analytical formulation of the output PDF makes the denoising framework very cumbersome. To alleviate this problem, we compute the system response numerically and approximate it by a Gaussian. We construct two *lookup tables* $\mathcal{L}_\mu(\cdot)$ and $\mathcal{L}_\sigma(\cdot)$ that provide the means and variances of the output Gaussians $G(x' - \mu', \sigma')$, given the means μ and variances σ^2 of input Gaussians and the noise level σ_R. We discretize the input parameters at a sufficiently-high resolution and employ bilinear interpolation to read values from the table.

We must be aware of some important issues while computing the system response. The Rician PDF $P(\tilde{x}|x)$ is defined only for non-negative x. However, the Parzen-window model with Gaussian kernels extends to negative values too. This model approximates the system poorly in cases where σ values are relatively large as compared to the magnitude of their means $\| \mathbf{z}_u \|$. In such cases, the Rician corruption process that applies only to the non-negative part of the Gaussian input (a truncated Gaussian) and produces an output that may not be fitted well by a Gaussian. However, we can view the situation more positively because of the implications of the *central limit theorem* [33, 34, 35, 36]. This classic theorem [33, 34] states that the PDF for the sum of *independent* RVs asymptotically approaches a Gaussian. In the same vein, there exists a central limit theorem for arbitrary *dependent* RVs too [35, 36] that proves their sum to approach a Gaussian RV. The theorem concerning dependent RVs applies to the Rician corruption process—the functional form of $P(\tilde{X}|x)$ depends on x. In our case, while one of the RVs is a Gaussian (input PDF), the other (Rician PDF) resembles a Gaussian in general and approaches a Gaussian for specific parameter values. These facts help us obtain good fits. Figure 1 shows that the fitted Gaussians approximate the Rician-corrupted output PDFs reasonably well. We observe that for input Gaussians that extend significantly to the negative axis, in Figure 1(a)-(b), the fit is not perfect while for the other cases, the fit is close to perfect. We use a Levenberg-Marquardt curve-fitting technique [37] to fit Gaussians to the output corrupted PDFs.

Given the uncorrupted PDF $P_{\mathcal{U}}(\cdot)$ and the Rician noise level σ_R, we can approximate the corrupted-signal Markov PDF as

$$\hat{P}_C(\tilde{\mathbf{z}}) \approx \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} G(\tilde{\mathbf{z}} - \mathbf{z'}_u, \Psi'_u), \tag{9}$$

where we define the i-th component of the neighborhood-intensity vector $\mathbf{z'}_u$ as

$$\mathbf{z'}_u(i) = \mathcal{L}_\mu(\mathbf{z}_u(i), \sigma, \sigma_R) \tag{10}$$

and the entry on the i-th row of the diagonal covariance matrix Ψ'_u as

$$\Psi'_u(i,i) = \mathcal{L}_\sigma(\mathbf{z}_u(i), \sigma, \sigma_R). \tag{11}$$

5.2 Inverse Problem: KL-Divergence Optimality

We want the corrupted-signal PDF $\hat{P}_C(\tilde{\mathbf{z}})$, derived from the uncorrupted-signal model $R_{\mathcal{U}}(\mathbf{Z})$, to match the Markov PDF $P_C(\tilde{\mathbf{z}})$ estimated from the observed corrupted data. We propose the Kullback-Leibler (KL) divergence as a measure of the discrepancy between the two PDFs. If we define $\Theta = \{\mathbf{z}_u\}_{u\in\mathcal{U}}$, then we want to find

$$\begin{aligned}
\{\Theta^*, \sigma^*\} &= \underset{\Theta,\sigma}{\text{argmin}}\ \text{KL}\left(P_C \parallel \hat{P}_C\right) \\
&= \underset{\Theta,\sigma}{\text{argmin}}\ E_{P_C}\left[\log \frac{P_C}{\hat{P}_C}\right] \\
&= \underset{\Theta,\sigma}{\text{argmin}}\ E_{P_C}\Big[\log P_C - \log \hat{P}_C\Big] \\
&= \underset{\Theta,\sigma}{\text{argmax}}\ E_{P_C}\Big[\log \hat{P}_C\Big] \\
&\approx \underset{\Theta,\sigma}{\text{argmax}}\ \sum_{t\in\mathcal{T}} \log \hat{P}_C(\tilde{\mathbf{z}}_t) \\
&= \underset{\Theta,\sigma}{\text{argmax}}\ \sum_{t\in\mathcal{T}} \log\left(\sum_{u\in\mathcal{U}} G(\tilde{\mathbf{z}}_t - \mathbf{z'}_u, \Psi'_u)\right).
\end{aligned} \tag{12}$$

What we have here is a ML optimization problem. ML estimation procedures, however, are well known to need regularization to reduces the chances of the optimization getting stuck in local maxima and to produce effective estimates, e.g. the classic *method-of-sieves* regularization by Grenander [38]. We propose to regularize the ML estimation by fixing the value of σ beforehand. The enforcement of this regularization is similar in spirit to that used by Geman and Hwang [39] for nonparametric density estimation.

We can produce an effective optimal estimate for σ as follows. We first find a ML-based estimate $\tilde{\sigma}$ for the nonparametric Markov PDF of the corrupted observed sample $\{\tilde{\mathbf{z}}_t\}_{t\in\mathcal{T}}$ (details in [1, 2]). We know that a significant fraction of intensities in the image are much larger than the noise level σ_R where the Rician noise model is close to an additive independent Gaussian noise model. Therefore, we approximate

$$\sigma^* \approx \sqrt{\tilde{\sigma}^2 - \sigma_R^2}. \tag{13}$$

Fixing this σ value, we subsequently obtain an optimal ML estimate for the set Θ relying on the EM algorithm. We have found that this approximation for σ works effectively in practice.

5.3 Optimization using the EM Algorithm

The inverse problem we have here is that of mixture-density parameter estimation—the parameter here is the set $\Theta = \{\mathbf{z}_u\}_{u \in \mathcal{U}}$ of the means of Gaussians that defines the uncorrupted-signal Markov PDF. We propose to solve this using the EM algorithm [23, 24]. The EM algorithm computes a ML parameter estimate when the data is *incomplete*, i.e. a part of the data remains unobserved or *hidden*. We now describe the key idea behind the working of the EM algorithm.

The optimization formulation in (12) is a little unwieldy because it contains the logarithm of a sum. If we knew which Gaussian component generated each observation, then we could obtain the probability $\hat{P}_C(\tilde{\mathbf{z}}_t)$ by evaluating a single Gaussian: the one that generated $\tilde{\mathbf{z}}_t$. The EM approach gets rid of the summation that the logarithm applies to. The key idea behind EM is that it assumes the existence of one *hidden* RV associated with each observation $\tilde{\mathbf{z}}_t$. The PDF of this hidden RV gives the probabilities for different Gaussian components to have generated $\tilde{\mathbf{z}}_t$. Let us call this RV L. The values of L are, however, never observed. The EM algorithm starts by assuming a joint PDF $P(\tilde{\mathbf{Z}}, L)$ of the observed and hidden RVs, i.e. the *complete data*. It defines the probability of the observation $\tilde{\mathbf{z}}_t$ assuming that it came from the l-th Gaussian as

$$P(\tilde{\mathbf{z}}_t|l) = G(\tilde{\mathbf{z}}_t - \mathbf{z'}_l, \Psi'_l), \tag{14}$$

where $\mathbf{z'}_l$ and Ψ'_l are the mean and covariance values, respectively, for the l-th Gaussian. The goal of the EM algorithm is to iteratively find the ML estimate of the parameter Θ as

$$\begin{aligned} \Theta^* &= \underset{\Theta}{\text{argmax}}\ \log P(\tilde{\mathbf{z}}|\Theta) \\ &= \underset{\Theta}{\text{argmax}}\ \log \left(\int_{\mathcal{S}_L} P(\tilde{\mathbf{z}}, l|\Theta) dl \right), \end{aligned} \tag{15}$$

where $\mathcal{S}_L$ is the support of $P(L)$. Each iteration comprises the *E* (expectation) step and the *M* (maximization) step. The E step formulates an expectation of the complete-data likelihood function over the PDF of the hidden RV conditioned on the observed data and current parameter estimate. The M step maximizes this expectation with respect to the parameter. After much simplification [40], the maximization performed in

the m-th iteration reduces to

$$\underset{\Theta}{\text{argmax}} \sum_{u \in \mathcal{U}} \sum_{t \in \mathcal{T}} P(u|\tilde{\mathbf{z}}_t; \Theta^{m-1}) \log P(\tilde{\mathbf{z}}_t|u, \mathbf{z}_u), \tag{16}$$

where Θ^{m-1} is the $(m-1)$-th parameter estimate that is held constant and $\Theta = \{\mathbf{z}_u\}_{u \in \mathcal{U}}$ is the free variable. The parameter updates guarantee no decrease in the likelihood $P(\tilde{\mathbf{z}}|\Theta)$ of the observed data and, hence, the sequence of estimates converge to a local maximum of the likelihood function.

An important element in this entire process of inferring the uncorrupted-signal Markov statistics is the initial choice of the sample $\{\hat{\mathbf{z}}_u^0\}_{u \in \mathcal{U}}$ for the EM algorithm. We initialize $\{\hat{\mathbf{z}}_u^0\}_{u \in \mathcal{U}}$ to comprise a small random fraction of the entire set of observed neighborhood-intensities $\{\tilde{\mathbf{z}}_t\}_{t \in \mathcal{T}}$, spread uniformly over the image domain $\mathcal{T}$. This ensures the representation of all important features in the image and produces an initial estimate close to the global maximum of the likelihood function.

The EM updates, for density estimation using a sum of Gaussians, are as follows.

1. Let $\{\hat{\mathbf{z}}_u^m\}_{u \in \mathcal{U}}$ be the parameter estimate at the m-th iteration.

2. Use the lookup tables to compute $\hat{\mathbf{z}'}_u^m$ and Ψ'^m_u, $\forall u \in U$, where

$$\begin{aligned} \hat{\mathbf{z}'}_u^m(i) &= \mathcal{L}_\mu(\hat{\mathbf{z}}_u^m(i), \sigma, \sigma_R) \text{ and} \\ \hat{\Psi'}_u^m(i,i) &= \mathcal{L}_\sigma(\hat{\mathbf{z}}_u^m(i), \sigma, \sigma_R). \end{aligned} \tag{17}$$

3. Compute

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, P(\tilde{\mathbf{z}}_t|u) = G(\tilde{\mathbf{z}}_t - \hat{\mathbf{z}'}_u^m, \Psi'_u) \tag{18}$$

4. Use Bayes rule to evaluate $P(u|\tilde{\mathbf{z}}_t), \forall t \in \mathcal{T}, \forall u \in \mathcal{U}$. Because we derive the initial set of observations $\hat{\mathbf{z}}_u^0$ from the PDF $P(\tilde{\mathbf{Z}})$ that is close to $P(\mathbf{Z})$, we can ignore the *a priori* probabilities $P(u)$—treat them equal for all u. Thus, we compute

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, P(u|\tilde{\mathbf{z}}_t) \approx \frac{P(\tilde{\mathbf{z}}_t|u)}{\sum_{v \in \mathcal{U}} P(\tilde{\mathbf{z}}_t|v)}. \tag{19}$$

5. Update the current parameter estimate using a gradient-ascent scheme using first-order finite forward differences:

$$\forall u \in \mathcal{U}, \hat{\mathbf{z}}_u^{m+1} = \hat{\mathbf{z}}_u^m + \Lambda \odot \left(\frac{\partial \hat{\mathbf{z}}_u}{\partial \hat{\mathbf{z}'}_u^m} \right) \odot \left(\frac{\sum_{t \in \mathcal{T}} P(u|\tilde{\mathbf{z}}_t) \tilde{\mathbf{z}}_t}{\sum_{t \in \mathcal{T}} P(u|\tilde{\mathbf{z}}_t)} - \hat{\mathbf{z}'}_u^m \right), \tag{20}$$

where the partial derivatives can be computed numerically using the lookup table $\mathcal{L}_\mu(\cdot)$, the operator $\odot$ denotes the element-by-element product of two vectors that produces another vector, and Λ denotes the time-step vector associated with the gradient ascent. We define the i-the component of Λ as

$$\Lambda(i) = \frac{\partial \hat{\mathbf{z}'}_u^m(i)}{\partial \hat{\mathbf{z}}_u(i)} \tag{21}$$

for the following reasons:

- We have numerically found that the magnitudes of the partial-derivative components always lies between 0 and 1 which implies

$$0 \leq \Lambda(i) \leq 1. \tag{22}$$

  Figure 2 shows the variation of e derivative values as a function of the noise level and the input-Gaussian mean and standard deviation.

- We observe that when the input-Gaussian mean $\mathbf{z}_u(i)$ is much larger than the noise-level σ_R and Parzen-window Gaussian variance σ^2, the output-Gaussian mean $\mathbf{z'}_u(i)$ is almost the same as the input-Gaussian mean and, hence, the time step is unity:

$$\begin{aligned} \left(\mathbf{z}_u(i) \gg \sqrt{\sigma_R^2 + \sigma^2} \right) &\Rightarrow \left(\mathbf{z'}_u(i) \rightarrow \mathbf{z}_u(i) \right) \\ &\Rightarrow \left(\frac{\partial \mathbf{z'}_u(i)}{\partial \mathbf{z}_u(i)} \rightarrow 1 \right) \\ &\Rightarrow \left(\Lambda(i) \rightarrow 1 \right). \end{aligned} \tag{23}$$

With this choice of the time step, we can ensure that the EM-based update is always in a direction that increases the likelihood and guarantees convergence. This particular choice of the time-step Λ,

simplifies the update to

$$\forall u \in \mathcal{U}, \hat{\mathbf{z}}_u^{m+1} = \hat{\mathbf{z}}_u^m + \left(\frac{\sum_{t \in \mathcal{T}} P(u|\tilde{\mathbf{z}}_t) \tilde{\mathbf{z}}_t}{\sum_{t \in \mathcal{T}} P(u|\tilde{\mathbf{z}}_t)} - \hat{\mathbf{z}'}_u^m \right). \tag{24}$$

6. If $\sum_{u \in \mathcal{U}} \parallel \hat{\mathbf{z}}_u^{m+1} - \hat{\mathbf{z}}_u^m \parallel_2^2 < \epsilon$, where ϵ is a small threshold, then stop, otherwise go to Step 3.

5.4 Engineering Enhancements for the EM Algorithm

Our initialization strategy gives $|\mathcal{U}| = \alpha|\mathcal{T}|$, where α is a free parameter and $0 < \alpha \leq 1$. Too small an α reduces the ability of the nonparametric PDF to well approximate the uncorrupted-signal Markov PDF. Too large an α increases the number of parameters to be estimated—equal to $|\mathcal{U}|$—thereby increasing the chance of the EM algorithm getting stuck on local maxima. A large α also increases the space requirements of the algorithm: $O(|\mathcal{U}||\mathcal{T}|)$. We have found that, in practice, the algorithm is not very sensitive to the specific choice of α and a choice of $\alpha = 0.33$ works well in practice.

To further reduce the computational and space requirements of the algorithm, we can replace the set $\mathcal{T}$ itself by a uniformly-distributed random sample of observations $\mathcal{T}^\dagger$, with $|\mathcal{T}^\dagger| = \beta|\mathcal{T}|, 0 < \beta \leq 1$, and subsequently choose U as a random sample from $\mathcal{T}^\dagger$, with $|\mathcal{U}| = \alpha|\mathcal{T}^\dagger|$. This makes the computational and space complexity of the EM algorithm both to be $O(\alpha\beta^2|\mathcal{T}|^2)$. The results in this paper use $\alpha = 0.33$ and $\beta = 0.66$.

6 Iterated Conditional Entropy Reduction (ICER)

At each pixel t, the prior PDF is

$$P(x_t|\mathbf{y}_t) = \frac{\sum_{u \in \mathcal{U}} G(\mathbf{y}_t - \mathbf{y}_u, \sigma) G(x_t - x_u, \sigma)}{\sum_{u \in \mathcal{U}} G(\mathbf{y}_t - \mathbf{y}_u, \sigma)} \tag{25}$$

and the likelihood PDF is

$$P(\tilde{x}_t|x_t) = \frac{1}{\eta(\tilde{x}_t, \sigma_R)} \frac{\tilde{x}_t}{\sigma_R^2} \exp\left(-\frac{\tilde{x}_t^2 + x_t^2}{2\sigma_R^2} \right) I_0\left(\frac{\tilde{x}_t x_t}{\sigma_R^2} \right), \tag{26}$$

where $\eta(\tilde{x}_t, \sigma_R)$ is the normalization factor that depends on the observed value $\tilde{x}_t$ and the noise level σ_R. We propose updating pixel intensities x_t, to increase the posterior probability $P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \tilde{\mathbf{x}})$ in (4), by performing a gradient ascent on the logarithm of the posterior. In [1, 15], we showed the equivalence between a gradient ascent on the logarithm of a PDF and entropy reduction using the Shannon's entropy measure. Entropy reduction on this posterior PDF results in the following update rule for all pixel intensities x_t

$$\begin{aligned} x_t &\leftarrow x_t - \lambda \frac{\partial h(x_t|\mathbf{y}_t, \tilde{x}_t)}{\partial x_t} \\ &= x_t + \lambda \left(\frac{\partial \log P(x_t|\mathbf{y}_t)}{\partial x_t} + \frac{\partial \log P(\tilde{x}_t|x_t)}{\partial x_t} \right) \\ &= x_t + \lambda \left(\frac{\sum_{u\in\mathcal{U}} G(\mathbf{y}_t - \mathbf{y}_u, \sigma) G(x_t - x_u, \sigma)(x_u - x_t)}{\sum_{u\in\mathcal{U}} G(\mathbf{y}_t - \mathbf{y}_u, \sigma) G(x_t - x_u, \sigma)} - \frac{\hat{x}_t^m}{\sigma^2} + \frac{\tilde{x}_t}{\sigma^2} \frac{I_1(\tilde{x}_t \hat{x}_t^m/\sigma^2)}{I_0(\tilde{x}_t \hat{x}_t^m/\sigma^2)} \right), \end{aligned} \quad (27)$$

where $I_1(\cdot)$ is the first-order modified Bessel function of the first kind. (The expression for the gradient of the logarithm of the Rician likelihood PDF appears in [41].) These sequence of updates leads to image estimates with non-decreasing posterior probabilities and, hence, guarantee convergence to a local maximum of the posterior PDF. We call this novel proposed algorithm for performing Bayesian estimation on MRFs as the *iterated conditional entropy reduction* (ICER).

7 MRI-Denoising Algorithm

The proposed iterative denoising algorithm requires an initial estimate. We obtain an initial estimate entirely based on the knowledge of the noise model, without any use of Markov prior. Thus, the initialization is a ML estimate of the image. The MRI-denoising algorithm finally produces the MAP image estimate as follows:

1. Infer the prior PDF $P(\mathbf{Z})$ (as described in Section 5) by minimizing the KL divergence, using the EM algorithm, between the observed corrupted-signal Markov PDF and its estimate derived from the prior-PDF model. The prior PDF is represented by a Parzen-window sum of isotropic Gaussian kernels with means $\{\mathbf{z}_u\}_{u\in\mathcal{U}}$ and standard deviation σ.

2. Obtain an initial denoised ML image $\hat{\mathbf{x}}^0 = \{\hat{x}_t^0\}_{t\in\mathcal{T}}$:

$$\forall t \in \mathcal{T}, \hat{x}_t^0 = \underset{x_t}{\text{argmax}}\ P(\tilde{x}_t|x_t). \tag{28}$$

We compute the mode of each likelihood PDF numerically using the iterative mode-seeking *mean-shift procedure* [42, 43].

3. Given the denoised-image estimate $\hat{\mathbf{x}}^m$ at iteration m, obtain the next estimate $\hat{\mathbf{x}}^{m+1}$ as

$$\forall t \in \mathcal{T}, \hat{x}_t^{m+1} = \hat{x}_t^m + \lambda\left(\frac{\sum_{u\in\mathcal{U}} G(\hat{\mathbf{y}}_t^m - \mathbf{y}_u, \sigma)G(\hat{x}_t^m - x_u, \sigma)(x_u - \hat{x}_t^m)}{\sum_{u\in\mathcal{U}} G(\hat{\mathbf{y}}_t^m - \mathbf{y}_u, \sigma)G(\hat{x}_t^m - x_u, \sigma)} - \frac{\hat{x}_t^m}{\sigma^2} + \frac{\tilde{x}_t}{\sigma^2}\frac{I_1(\tilde{x}_t\hat{x}_t^m/\sigma^2)}{I_0(\tilde{x}_t\hat{x}_t^m/\sigma^2)}\right), \tag{29}$$

where all the symbols have the same meaning as in Section 6.

4. If $\| \hat{\mathbf{x}}^{m+1} - \hat{\mathbf{x}}^m \|_2 < \epsilon$, where ϵ is small threshold, then stop, otherwise go to Step 3.

8 Results and Validation

This section gives validation results on synthetic brain-MR images with a wide range of noise and bias values as well as real MR data. The computation for each iteration is $O(|\mathcal{A}_t||\mathcal{T}||\mathcal{N}_t|)$. We have found empirically that, with simulated MR images from the BrainWeb [44] database, ICER produces the largest reduction in RMS errors after a single iteration itself. Subsequent iterations converge at a nearby RMS-error value. All results in this paper employ a single iteration of ICER.

8.1 Validation on Simulated and Real MR Images

Figure 3 presents the results of denoising a particular slice from volumetric T1-weighted simulated Brain-Web data. The proposed MRI-denoising algorithm acts conservatively, reducing the RMS error by about 40%. Figure 3(d) shows the difference between the corrupted and the uncorrupted images. The shift in the intensity PDF introduced by Rician noise is evident in the lighter background region (higher intensity on the average) corresponding to low signal intensities. The intensities in this difference image also possess a very low degree of spatial correlation. Figure 3(e) shows the difference between the denoised and the uncorrupted

images. We see that algorithm reduces the Rician-noise-introduced shift in intensities in the low-intensity background region—fewer bright spots. Empirical analysis shows that denoised image effectively corrects the for the shift in the corrupted-intensity PDF caused by Rician noise—as measured by the average value of the background intensities in the uncorrupted, corrupted, and denoised images. For the case of T1-weighted BrainWeb data with 5% noise and 40% bias in Figure 3 the average background values are: (a) 0.1 for the uncorrupted image, (b) 3.1 for the corrupted image, and (c) 0.03 for the denoised image. The difference images in Figure 3(e) and Figure 7(c) show low magnitudes for errors in the background region. The difference image also possesses low correlation indicating that the proposed algorithm retained the significant image features more-or-less intact. The power spectrum of the difference image in Figure 3(f) shows the *whiteness* [45] of the residual.

Figure 4 gives the performance of the proposed algorithm on three different slices of the BrainWeb MR data for varying noise and bias levels. We observe that the performance on biased and unbiased data is equivalent. This stems from the ability of adaptive-MRF model to effectively infer the appropriate Markov statistics for each case and denoise based on the inferred model. We also observe that for very low Rician noise, i.e. $\sigma_R \approx 1$, the algorithm does not effectively reduce the RMS error. This may be because of a similar level of variability inherent in the data, and in the estimated uncorrupted-signal Markov PDFs, which makes the algorithm not clearly identify the noise. As the amount of noise increases, the proposed method can clearly differentiate the structure underlying the data from the noise. Figure 5 shows the performance of proposed algorithm on real data that depicts a significant inhomogeneity/bias.

Figure 6 compares, qualitatively and quantitatively, the performance of the proposed algorithm with several other recent and popular filtering algorithms. We have manually tuned all the free parameters in these other algorithms in order to give the best possible results. The proposed algorithm does better qualitatively, with an RMS error of 3.3 (RMS error for noisy image is 5.53) as compared to the RMS errors produced by other algorithms of around 4.0 or more. Qualitatively too, the proposed algorithm gives a residual (difference between denoised and uncorrupted image) that is significantly less correlated. The state-of-the-art wavelet-based denoising algorithm [14] also seems to introduce artifacts in the denoised image.

Figure 7 show the qualitative and quantitative comparison of the proposed method with a state-of-the-art wavelet-based MRI-denoising algorithm [14]. We see that the proposed method produces lower RMS errors at all noise levels except with one image at the 9% noise level. Although the RMS error for the proposed

method is a little more for this high-noise case, Figure 7(c) and Figure 7(d) show that the residual for the wavelet-based method is significantly more correlated. This residual also indicates the presence of artifacts in the wavelet-denoised image.

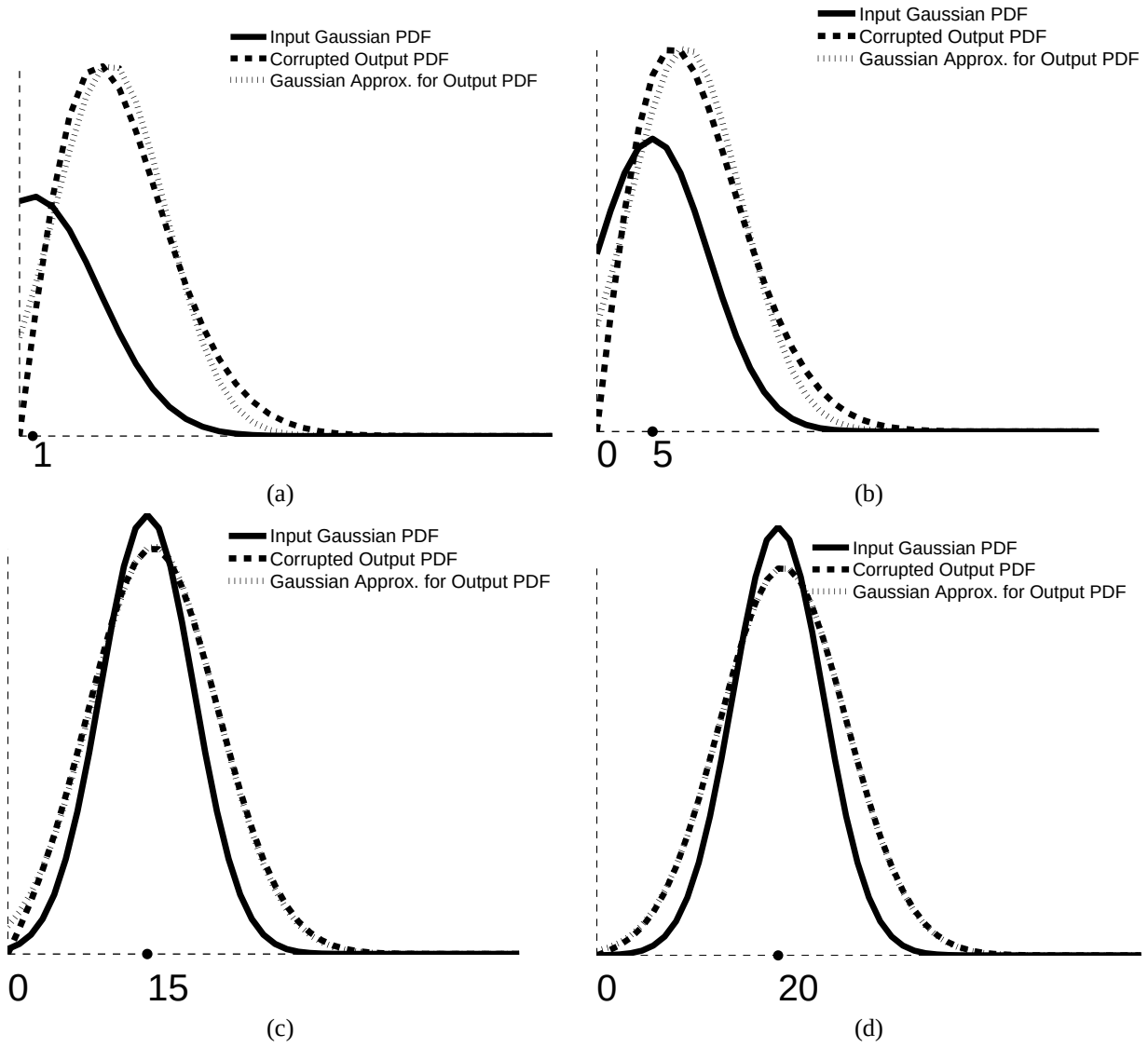
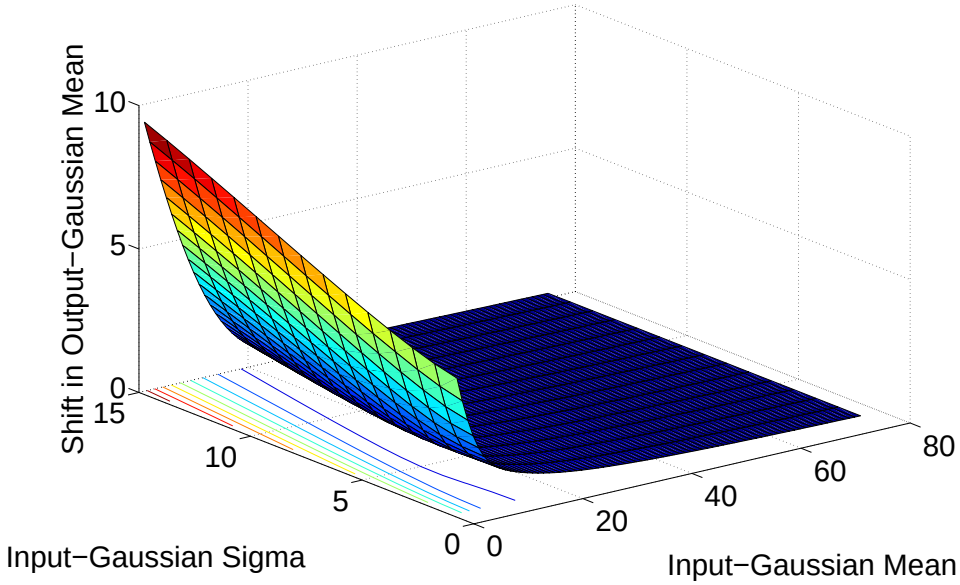
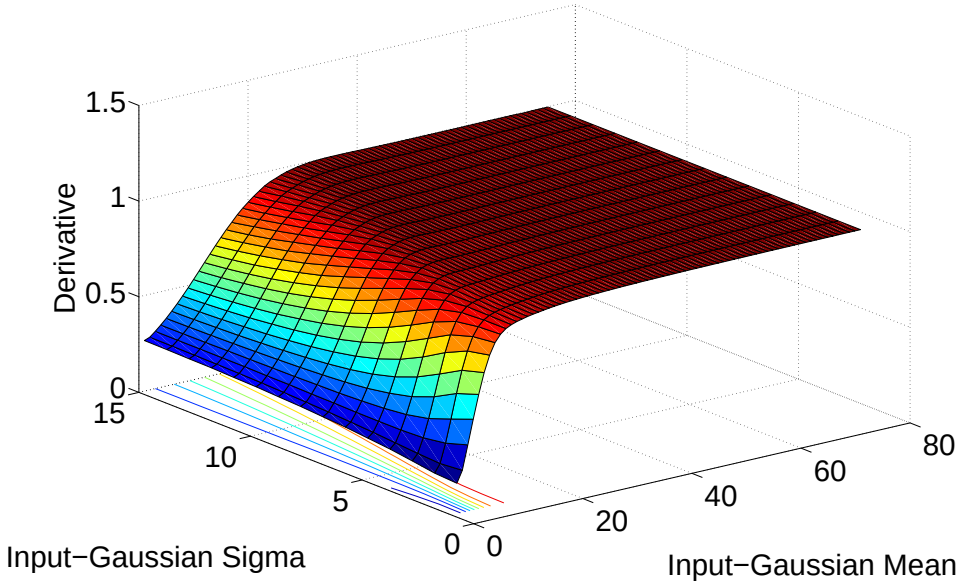


Figure 1: These graphs depict the Rician corruption process in 1D with $\sigma = 5$ and σ_R = 5. The input Gaussian PDF is corrupted by Rician noise resulting in the output corrupted PDF. We fit a Gaussian to this corrupted PDF. The graphs show this process for different means of the input Gaussian: (a) x_u = 1, (b) x_u = 5, (c) x_u = 15, and (d) x_u = 20.

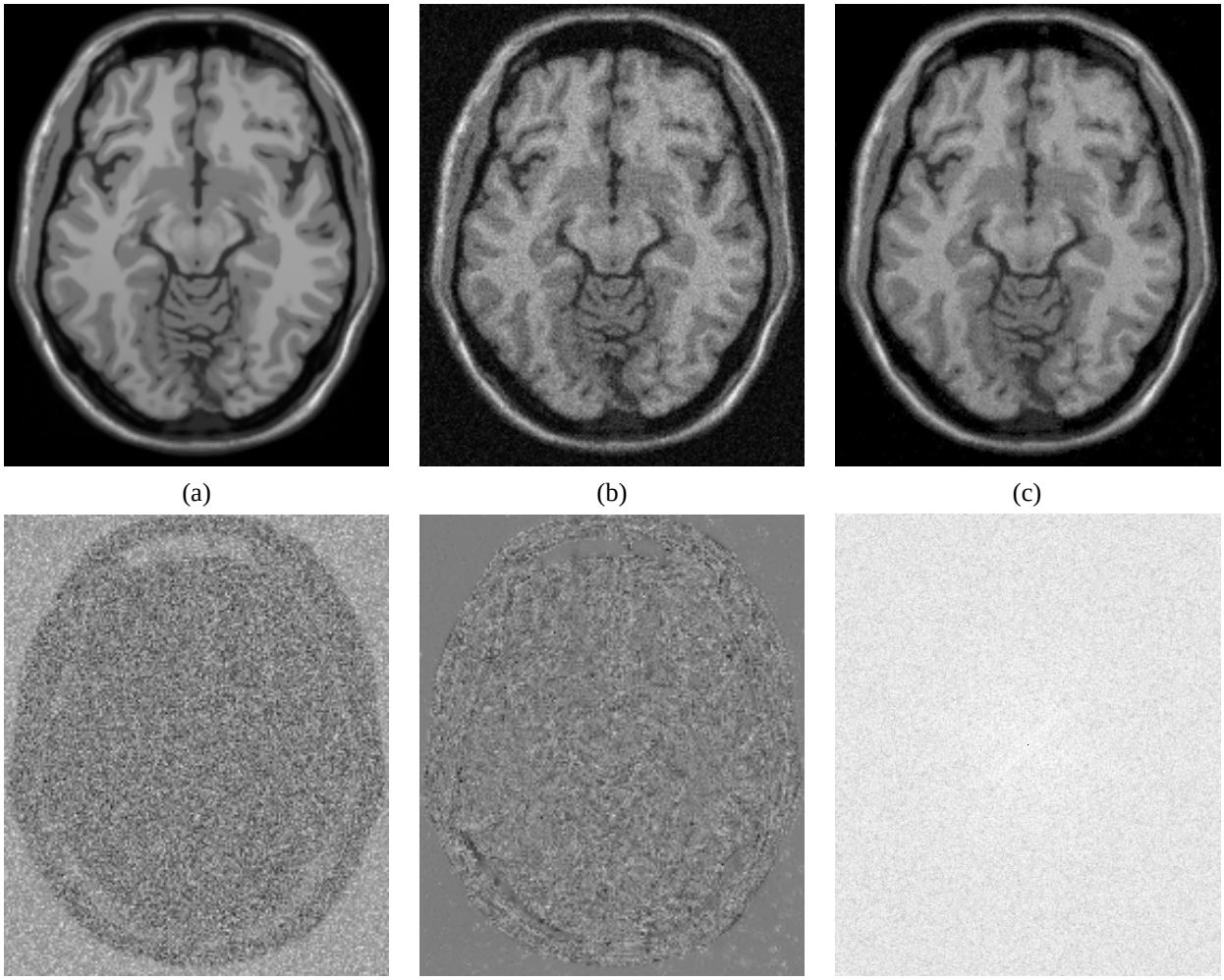


(a)



(b)

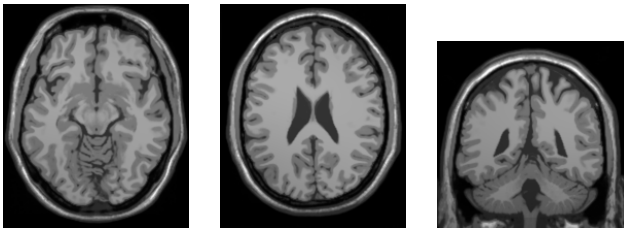
Figure 2: The axes in the horizontal plane depict the values of the mean x_u and standard deviation σ of Gaussian-PDF inputs to the Rician-corruption process. The Rician noise level is $\sigma_R = 5$. The vertical axis depicts the following: (a) The difference between the means of the Gaussian that approximates the corrupted PDF and the input Gaussian, i.e. $x'_u - x_u$. This is exactly the *shift* in intensities introduced by the Rician corruption process. We observe that the shift approaches zero as $x_u \gg \sqrt{\sigma^2 + \sigma_R^2}$. (b) The derivative of the mean of the output Gaussian with respect to the mean of the input Gaussian, i.e. $\partial x'_u / \partial x_u$. We observe that this derivative is always non-negative, less than unity, and approaches unity as $x_u \gg \sqrt{\sigma^2 + \sigma_R^2}$.



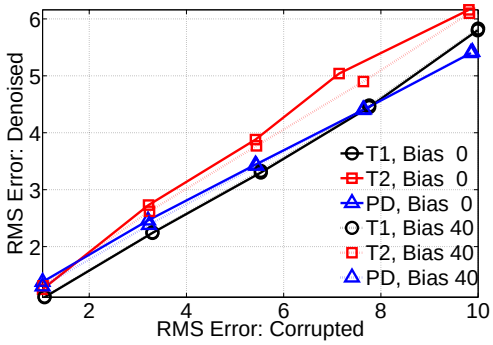
(a) (b) (c)

(d) (e) (f)

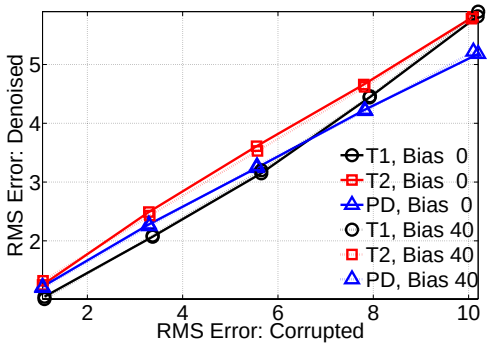
Figure 3: Results with T1-weighted simulated BrainWeb data (intensity range $0:100$) with the Rician noise level $\sigma_R = 5$ and a 40% bias field. (a) Uncorrupted image. (b) Rician-noise corrupted image: RMSE = 5.53. (c) Denoised image: RMSE = 3.3. (d) Difference between the corrupted and uncorrupted images. (e) Difference between the denoised and uncorrupted images. (f) Power spectrum of the image in (e): close to *white*.



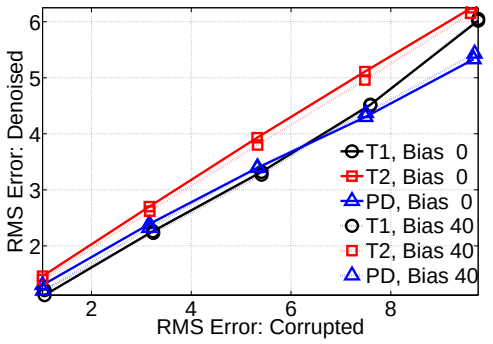
(a) (b) (c)



(d)



(e)

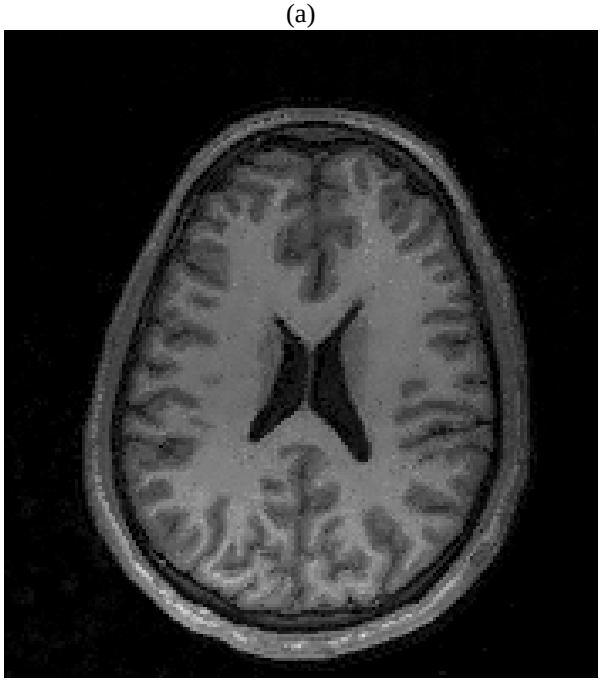


(f)

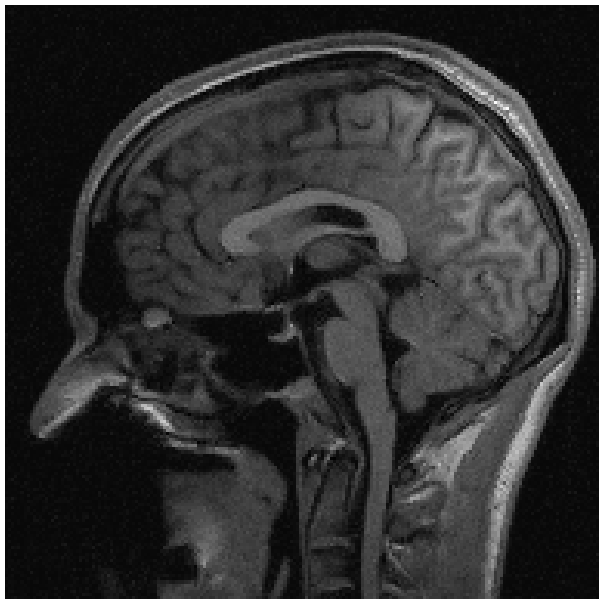
Figure 4: (a)-(c) Three different brain slices from the BrainWeb dataset (only T1 modality shown). (d)-(f) Graphs indicating RMS errors for denoised and noisy images, with 0% and 40% bias fields, for T1, T2, and PD modalities on the three slices above.

(a)

(b)



(c)



(d)

Figure 5: (a),(b) Noisy slices from a real MR volume. (c),(d) Denoised images.

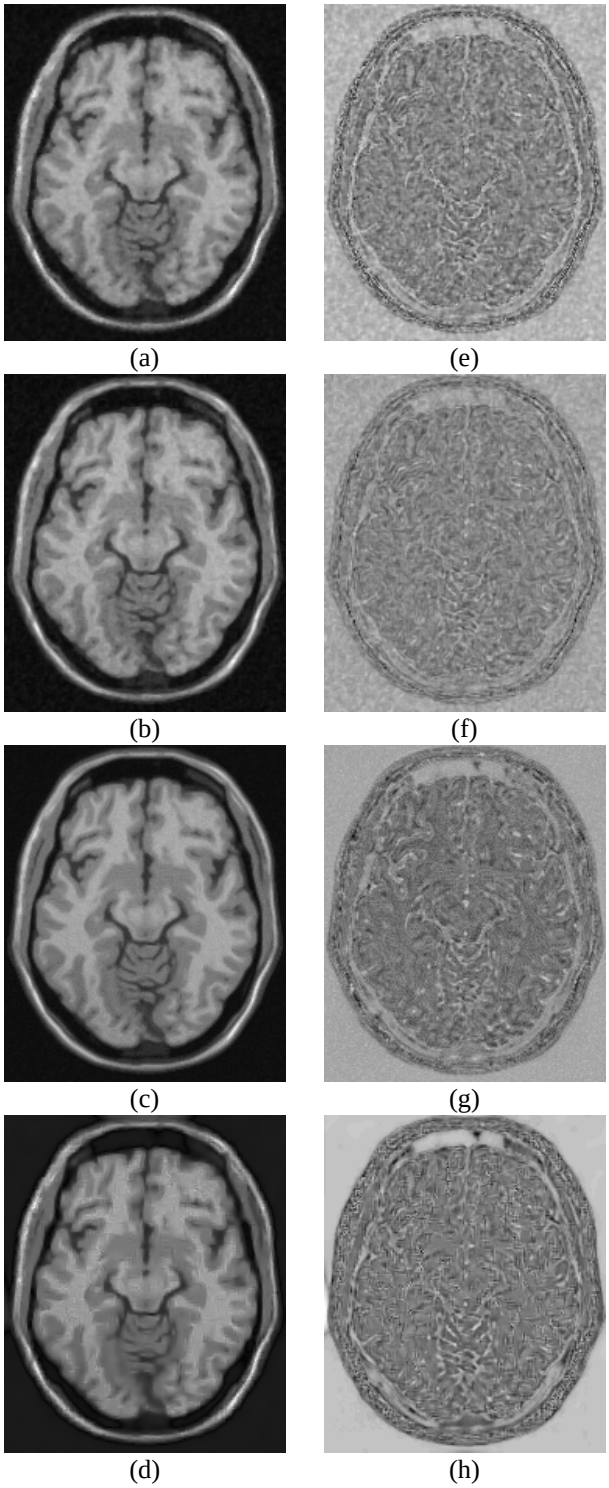
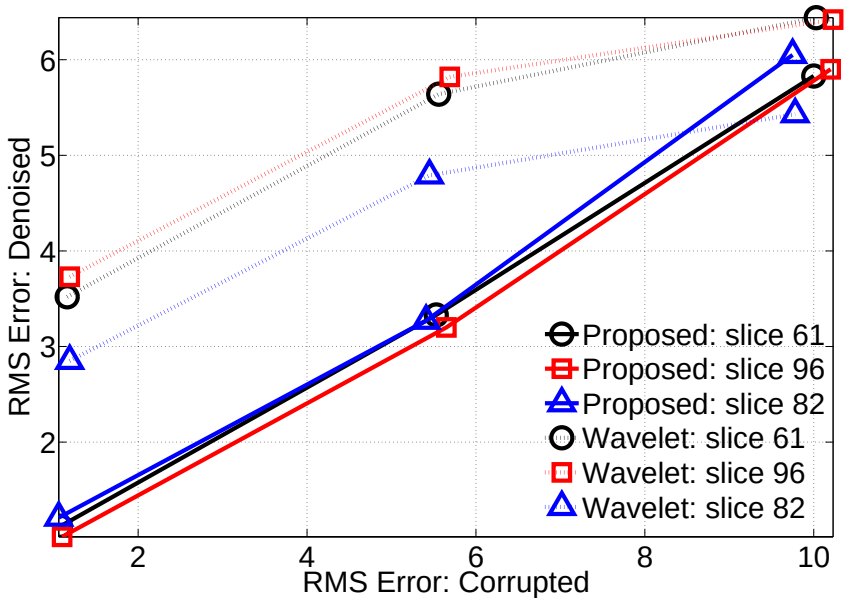
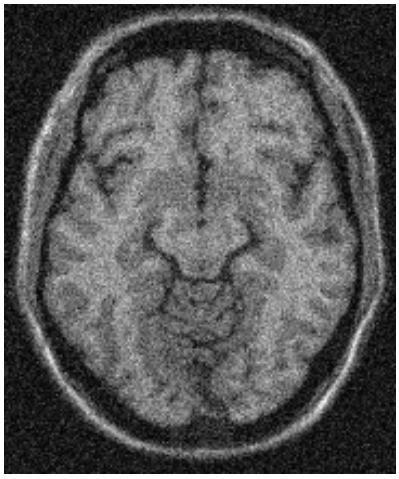


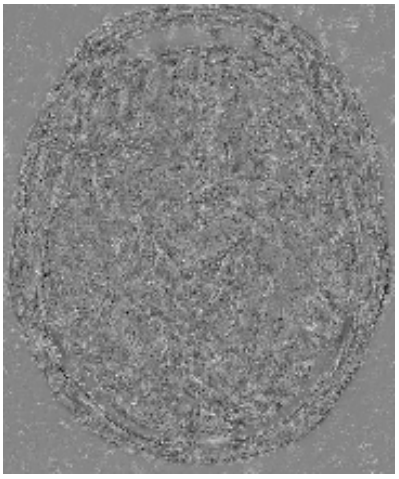
Figure 6: Results with T1-weighted simulated BrainWeb data (intensity range $0 : 100$) with the Rician noise level $\sigma_R = 5$ and a 40% bias field. The noisy image in Figure 3(b) (RMSE = 5.53) denoised using (a) anisotropic diffusion [6]: RMS error 4.03, (b) curvature flow [46]: RMS error 3.93, (c) UINTA [1]: RMS error 4.0, and (d) the state-of-the-art wavelet-based MRI denoiser [14]: RMS error 5.64, (e)-(h) show the differences between the denoised images in (a)-(d) and the uncorrupted image in Figure 3(a).



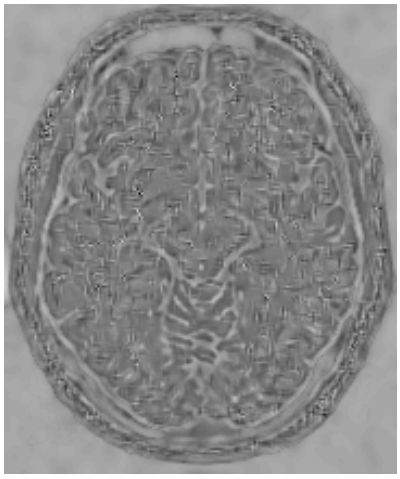
(a)



(b)



(c)



(d)

Figure 7: (a) Quantitative comparison of the proposed method with a state-of-the-art wavelet-based MRI-denoiser [14] for the three different slices of T1 BrainWeb data (shown in Figure 4) with varying noise levels and a 40% bias field. (b) Corrupted T1 data with 9% noise and 40% bias field. (c) and (d) show the difference between the denoised and uncorrupted images for the proposed and wavelet-based [14] methods, respectively, when these methods are applied to the corrupted data in (b).

References

[1] S. P. Awate and R. T. Whitaker, "Unsupervised, Information-Theoretic, Adaptive Image Filtering for Image Restoration," *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, vol. 28, no. 3, pp. 364–376, March 2006.

[2] S. P. Awate, T. Tasdizen, R. T. Whitaker, and N. L. Foster, "Adaptive, nonparametric markov modeling for unsupervised, MRI brain-tissue classification," *Medical Image Analysis*, vol. 10, no. 5, pp. 726–739, 2006.

[3] S. P. Awate and R. T. Whitaker, "Nonparametric Neighborhood Statistics for MRI Denoising," in *Proc. Int. Conf. Information Processing in Medical Imaging (IPMI), Springer, Lect. Notes in Comp. Sci.*, vol. 3565, 2005, pp. 677–688.

[4] H. Robbins, "An empirical bayes approach to statistics," in *Proc. Third Berkeley Symp. Math. Stat. Prob.*, 1964, pp. 157–164.

[5] ——, "The empirical bayes approach to statistical decision problems," *Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 1–20, 1964.

[6] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, July 1990.

[7] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.

[8] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Tr. Med. Imaging*, vol. 11, no. 2, pp. 221–232, 1992.

[9] M. Lysaker, A. Lundervold, and X. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time," *IEEE Trans. Imag. Proc.*, 2003.

[10] A. Fan, W. Wells, J. Fisher, M. Çetin, S. Haker, R. Mulkern, C. Tempany, and A. Willsky, "A unified variational approach to denoising and bias correction in mr." in *Info. Proc. Med. Imag.*, 2003, pp. 148–159.

[11] D. Healy and J. Weaver, "Two applications of wavelet transforms in magnetic resonance imaging," *IEEE Trans. Info. Theory*, vol. 38, no. 2, pp. 840–860, 1992.

[12] M. Hilton, T. Ogden, D. Hattery, G. Jawerth, and B. Eden, "Wavelet denoising of functional MRI data," 1996, pp. 93–114.

[13] R. Nowak, "Wavelet-based rician noise removal for magnetic resonance imaging," *IEEE Trans. Imag. Proc.*, vol. 8, pp. 1408–1419, 1999.

[14] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging." *IEEE Trans. Med. Imaging*, vol. 22, no. 3, pp. 323–331, 2003.

[15] S. P. Awate and R. T. Whitaker, "Higher-Order Image Statistics for Unsupervised, Information-Theoretic, Adaptive, Image Filteringa," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 44–51.

[16] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in mri data," *IEEE Trans. Medical Imaging*, vol. 17, pp. 87–97, 1998.

[17] J. Mangin, "Entropy minimization for automatic correction of intensity nonuniformity," in *IEEE MMBIA*, 2000, pp. 162–169.

[18] G. Casella, "An introduction to empirical bayes analysis," *The American Statistician*, vol. 39, no. 2, pp. 83–87, 1985.

[19] N. Laird, "Nonparametric maximum likelihood estimation of a mixing distribution," *Journal of the American Statistical Association*, vol. 73, no. 364, pp. 805–811, 1978.

[20] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. Weinberger, "Universal discrete denoising: Known channel," *IEEE Trans. Info. Theory*, vol. 51, no. 1, pp. 5–28, 2005.

[21] D. L. Snyder, M. Miller, and T. Schultz, "Constrained probability-density estimation from noisy data," in *Proc. 22nd Annual Conference on Information Sciences and System*, 1988, pp. 170–172.

[22] C. B. Cordy and D. R. Thomas, "Deconvolution of a distribution function," *Journal of the American Statistical Association*, vol. 92, no. 440, pp. 1459–1465, 1997.

[23] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. B, no. 39, pp. 1–38, 1977.

[24] G. J. McLachlan, *The EM Algorithm and Extensions*. John Wiley, 1997.

[25] S. P. Awate, "Adaptive markov models with information-theoretic methods for unsupervised image restoration and segmentation," *Ph.D. Dissertation, School of Computing, University of Utah.*

[26] H. Stark and J. W. Woods, *Probability and random processes with applications to signal processing*. Prentice Hall, 2001.

[27] S. C. Zhu and D. Mumford, "Prior learning and gibbs reaction-diffusion," *IEEE Trans. Pattern Analysis Machine Intell.*, vol. 19, no. 11, pp. 1236–1250, 1997.

[28] J. Huang and D. Mumford, "Statistics of natural images and models." in *Proc. IEEE Comp. Vis. Pattern Recog.*, 1999, pp. 1541–1547.

[29] A. Lee, K. Pedersen, and D. Mumford, "The nonlinear statistics of high-contrast patches in natural images," *Int. J. Comput. Vision*, vol. 54, no. 1-3, pp. 83–103, 2003.

[30] E. Parzen, "On the estimation of a probability density function and the mode," *Annals of Math. Stats.*, vol. 33, pp. 1065–1076, 1962.

[31] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley, 2001.

[32] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, series B*, vol. 48, pp. 259–302, 1986.

[33] K. S. Trivedi, *Probability and Statistics With Reliability, Queuing, and Computer Science Applications*. Wiley, 2001.

[34] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. McGraw-Hill, 2001.

[35] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math J.*, vol. 15, pp. 773–780, 1948.

[36] K. N. Berk, "A central limit theorem for m-dependent random variables with unbounded m," *Annals of Prob.*, vol. 1, no. 2, pp. 352–354, 1973.

[37] S. S. Rao, *Engineering Optimization, Theory and Practice*. Wiley, 1996.

[38] U. Grenander, *Abstract Inference*. Wiley, 1975.

[39] S. Geman and C. R. Hwang, "Nonparametric maximum likelihood estimation by method of sieves," *Annals of Statistics*, vol. 10, no. 2, pp. 401–414, 1982.

[40] J. Bilmes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," University of Berkeley, Tech. Rep., 1997.

[41] S. Basu, "Rician noise removal in diffusion tensor mri," *Masters Thesis, School of Computing, University of Utah*, 2006.

[42] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Info. Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[43] M. Fashing and C. Tomasi, "Mean shift is a bound optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 471–474, 2005.

[44] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom." *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 463–468, 1998.

[45] A. K. Jain, *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.

[46] J. Sethian, *Level Set Methods and Fast Marching Methods*. Cambridge Univ. Press, 1999.