# Feature-Preserving MRI Denoising: A Nonparametric Empirical Bayes Approach

Suyash P. Awate* and Ross T. Whitaker

***Abstract*—This paper presents a novel method for Bayesian denoising of magnetic resonance (MR) images that bootstraps itself by inferring the prior, i.e., the uncorrupted-image statistics, from the corrupted input data and the knowledge of the Rician noise model. The proposed method relies on principles from empirical Bayes (EB) estimation. It models the prior in a nonparametric Markov random field (MRF) framework and estimates this prior by optimizing an information-theoretic metric using the expectation-maximization algorithm. The generality and power of nonparametric modeling, coupled with the EB approach for prior estimation, avoids imposing ill-fitting prior models for denoising. The results demonstrate that, unlike typical denoising methods, the proposed method preserves most of the important features in brain MR images. Furthermore, this paper presents a novel Bayesian-inference algorithm on MRFs, namely iterated conditional entropy reduction (ICER). This paper also extends the application of the proposed method for denoising diffusion-weighted MR images. Validation results and quantitative comparisons with the state of the art in MR-image denoising clearly depict the advantages of the proposed method.**

***Index Terms*—Denoising, empirical Bayes (EB), information theory, Markov random fields (MRF), nonparametric statistics.**

## I. 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, can result in images that exhibit significant artifacts e.g., noise, partial voluming, and intensity inhomogeneity (also known as intensity nonuniformity [1] or bias field). For instance, the need for shorter acquisition times for patients in certain clinical studies often undermines the ability to obtain images having both high resolution and high SNR. Another example concerns diffusion-tensor (DT) MRI that has become quite popular over the last decade because of its ability to measure the anisotropic diffusion of water in structured biological tissue. DT MRI differentiates between the anatomical structures of cerebral white matter, which was previously impossible with MRI, *in vivo*, and noninvasively. The effects of Rician noise on DT MRI, however, are severe because of the inherent nature of the process—higher tissue anisotropy produces progressively lower intensities in diffusion-weighted images (DWIs) that, in turn, are more susceptible to Rician noise. The efficacy of higher-level post processing of MR and DT-MR images, e.g., segmentation and tractography, that assume specific models on regions of interest, 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 Rician bias, structure and edge preservation, generality, reliability, automation, and computational cost.

This paper presents a novel method for Bayesian denoising of magnetic resonance (MR) images that bootstraps itself by inferring the prior, i.e., the uncorrupted image statistics, from the corrupted input data and the knowledge of the Rician noise model. The proposed method relies on principles from empirical Bayes (EB) estimation. It models the prior in a nonparametric Markov random field (MRF) framework and estimates this prior by optimizing an information-theoretic metric using the expectation-maximization algorithm. The generality and power of nonparametric modeling, coupled with the EB approach for prior estimation, avoids imposing ill-fitting prior models for denoising. The results demonstrate that, unlike typical denoising methods, the proposed method preserves most of the important features in brain MR images. Furthermore, the paper presents a novel Bayesian-inference algorithm on MRFs, namely iterated conditional entropy reduction (ICER). This paper also extends the application of the proposed method for denoising diffusion-weighted MR images. Validation results and quantitative comparisons with the state of the art in MR-image denoising clearly depict the advantages of the proposed method.

## II. Related Work

This section presents a brief overview of MR-denoising methods based on variational, statistical, and information-theoretic frameworks.

The image-processing literature presents a multitude of denoising methods based on partial differential equations (PDEs),

Manuscript received February 15, 2007; revised May 3, 2007. This work was supported by the National Science Foundation (NSF) under Grant EIA0313268, in part by the NSF CAREER Grant CCR0092065, in part by the "Center for Alzheimer's Care, Imaging and Research," in part by the University of Utah, and in part by the National Institutes of Health under Grant HD046159. *Asterisk indicates corresponding author.*

*S. P. Awate was with the Scientific Computing and Imaging (SCI) Institute and the School of Computing, University of Utah, Salt Lake City, UT 84112 USA. He is now with the Penn Image Computing and Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA.

R. T. Whitaker is with the School of Computing, University of Utah, Salt Lake City, UT 84112 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2007.900319

in variational frameworks, for a wide variety of images and applications [2], including some specifically geared towards MR images [3]–[6]. Such methods, however, impose certain kinds of models on local image structure that are often too simple to capture the complexity of anatomical MR images. Furthermore, these methods entail manual tuning of *critical* free parameters that control the conditions under which the models prefer one sort of structure over another. These factors have been an impediment to the widespread adoption of PDE-based techniques for processing MR images.

Another approach to denoising relies on statistical inference on multiscale representations of images. A prominent example includes methods based on wavelet transforms. Healy *et al.* [7] were among the first to apply wavelet techniques, based on soft thresholding, for denoising MR images. Hilton *et al.* [8] apply a threshold-based scheme for functional MRI data. Nowak [9], operating on the square magnitude MR image, incorporates a Rician noise model in a threshold-based wavelet denoising scheme and, thereby, corrects for the bias introduced by the noise. Pizurica *et al.* [10] rely on the prior knowledge of the correlation of wavelet coefficients that represent significant features across scales. Their method first detect the wavelet coefficients that correspond to these significant features and then empirically estimate the probability density functions (PDFs) of wavelet coefficients conditioned on the significant features. It employs these probabilities in a Bayesian denoising scheme. Typical wavelet-based methods, however, can produce significant artifacts in the processed images that relate to the structure of the underlying wavelet.

Recent approaches on image restoration have employed nonparametric statistical methods. For instance, Awate and Whitaker [11], [12] propose an unsupervised information-theoretic adaptive filter, namely UINTA, that relies on nonparametric MRF models derived from the corrupted images. UINTA restores images by generalizing the mean-shift procedure [13], [14] to incorporate neighborhood information. They show that entropy measures on first-order image statistics are ineffective for denoising and, hence, advocate the use of higher-order/Markov statistics. UINTA, however, does not assume a specific noise model during restoration. Buades *et al.* [15], [16] have, independently, proposed a denoising strategy along similar lines, namely NL-Means, but one relying on principles in nonparametric regression. NL-Means estimates the regression between the corrupted pixel intensity and its neighborhood and, in turn, produces an optimal estimate of the uncorrupted pixel intensity. NL-Means, however, does not incorporate a Rician noise model.

The proposed method aims to address the aforementioned limitations of the current methods by following the EB approach [17], pioneered by Robbins [18], coupled with a nonparametric Markov model for the prior. The EB approach was designed for the situation where multiple independent instances of a Bayesian decision problem (e.g., denoising each voxel intensity) all rely on exactly the same, but unknown, prior PDF (e.g., uncorrupted-signal Markov PDF). Under such circumstances, the EB 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 Bayesian processing automatically adapts to the unknown prior PDFs. The proposed method exploits the EB approach by first inferring a maximum likelihood (ML) estimate of the prior distribution using the corrupted observed image and, subsequently, employing the inferred prior model to compute the posterior.

Weismann *et al.* [19] address optimal information-theoretic denoising of discrete-valued signals in a MRF framework following the EB approach. They propose a discrete universal denoiser, termed DUDE, that relies on inverting the channel transition matrix (noise model) to give a closed-form estimate for signal 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 iteratively through the reduction of a Kullback–Leibler (KL) divergence. The proposed approach also presents a method for practically dealing with the nonstationarity in real MR images. Motta *et al.* [20] have extended DUDE to continuous images by approximating them as discrete images with a large alphabet and exploiting prior information on image data for effective estimation of empirical Markov histograms in the resulting high-dimensional alphabet spaces.

Cordy and Thomas [21] deconvolve low-dimensional PDFs corrupted with independent and identically distributed (i.i.d.) additive Gaussian noise—quite different from the signal-dependent Rician noise—using the expectation-maximization (EM) algorithm [22], [23]. Cordy and Thomas model the uncorrupted-signal PDF as a Gaussian mixture model and use the EM algorithm to estimate only the weights of Gaussians in the mixture—the means and variances of the Gaussians are set manually before EM is applied. They set the means in order to spread the Gaussians uniformly over the entire domain of the PDF. Such a strategy, however, it not likely to be effective for estimating PDFs in high-dimensional domains because of the enormous number 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. Cordy and Thomas also present an overview of research on the PDF-deconvolution problem.

## III. MAP Denoising With MRFs

This section introduces the MRF image model and the associated mathematical notation used in the paper. It briefly describes Besag's method of iterated conditional modes (ICM) [24] for the maximum-a-posteriori (MAP) Bayesian estimation and motivates the variation on ICM proposed in this paper (Section VI).

A *random field* is a family of random variables (RVs) $\mathbf{X} = \{X_t\}_{t \in \mathcal{T}}$, where $\mathcal{T}$ is a set of points defined on a discrete Cartesian grid. For each voxel $t$, the RV $X_t$ takes values denoted by $x_t$. MRFs capture the regularity in data—the dependencies between voxel intensities in images—by specifying only the joint PDFs of the RVs in image neighborhoods. We denote the set of voxels in the neighborhood of a voxel $t$ by $\mathcal{N}_t$. The set $\mathcal{N}_t$ excludes $t$. The random vector $\mathbf{Y}_t = \{X_t\}_{t \in \mathcal{N}_t}$ comprises RVs in $\mathcal{N}_t$. Markovity implies

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

Defining 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 $\langle \mathbf{z} \rangle$.

Given the noisy image $\widetilde{\mathbf{x}}$, the goal is to find the MAP estimate $\mathbf{x}^*$ of the true image $\mathbf{x}$

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}} \; P(\mathbf{x}|\widetilde{\mathbf{x}}). \tag{2}$$

Writing the posterior as

$$P(\mathbf{x}|\widetilde{\mathbf{x}}) = P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \widetilde{\mathbf{x}}) P(\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}|\widetilde{\mathbf{x}}) \tag{3}$$

where $t$ is an arbitrary voxel, motivates us to employ an iterative restoration scheme where, an initial image estimate gets updated, one voxel at a time, in such a way that the posterior probability never decreases. Besag's ICM algorithm gives one such strategy that updates $x_t$ to the mode of the PDF $P(X_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \widetilde{\mathbf{x}})$. Finding modes of PDFs, however, is not always straightforward or computationally efficient. Therefore, we propose a new variation that updates $x_t$ by moving it closer to the local mode of $P(X_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \widetilde{\mathbf{x}})$. The proposed algorithm is similar in spirit to the ICM algorithm, but relies on a gradient ascent on the logarithm of the aforementioned PDF. This guarantees nondecreasing values for $P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \widetilde{\mathbf{x}})$ and, thereby, convergence to a local maximum of the posterior PDF $P(\mathbf{X}|\widetilde{\mathbf{x}})$. Section VI, later, describes the relationship of the proposed algorithm with entropy reduction.

We now simplify (3) to make it computationally tractable. We make the standard assumption from MRF literature that, given the true image $\mathbf{x}$, the RVs in the MRF $\widetilde{\mathbf{X}}$ are conditionally independent [24], [25]. Such an assumption is reasonable for the Rician noise model because the underlying Gaussian noise, in the real and imaginary components of the complex-valued MR image, is i.i.d. Given the noise model $P(\widetilde{X}_t|x_t)$ for the Rician degradation process, conditional independence implies that the conditional probability of the observed image given the true image is

$$P(\widetilde{\mathbf{x}}|\mathbf{x}) = \prod_{t\in\mathcal{T}} P(\widetilde{x}_t|x_t). \tag{4}$$

Subsequently, Bayes rule and Markovity imply

$$\underset{x_t}{\operatorname{argmax}} P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}}, \widetilde{\mathbf{x}}) = \underset{x_t}{\operatorname{argmax}} P(x_t|\mathbf{y}_t) P(\widetilde{x}_t|x_t) \tag{5}$$

where $P(X_t|\mathbf{y}_t)$ is the unknown Markov prior PDF and $P(\widetilde{x}_t|X_t)$ is the likelihood PDF as determined by the Rician noise model. We propose to model the prior PDF nonparametrically and optimally estimate it from the observed corrupted data itself. The next section describes the details of the nonparametric modeling strategy for the Markov prior.

## IV. Nonparametric Models for Markov Priors

This section motivates the nonparametric MRF image model and introduces the associated notation. The proposed method employs this model to represent the prior Markov PDF $P(X_t|\mathbf{y}_t)$ in a Bayesian denoising framework.

Typical parametric-MRF image models assume the functional forms for the Markov PDFs to be known *a priori* [25]. These forms 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 [26]–[28]. As a result, in many instances, the data does not comply well with such parametric MRF models. Unlike typical denoising methods that use strong, and often ill-fitting, parametric Markov priors, the proposed method exploits the power of data-driven nonparametric modeling to estimate a prior PDF that conforms well to the observed data.

The proposed method employs a data-driven method for nonparametrically modeling the Markov PDFs. In order to rely on neighborhood intensities $\{\mathbf{z}_t\}_{t\in\mathcal{T}}$ from the image to produce nonparametric estimates of Markov statistics $P(\mathbf{Z}_t)$, we must assume that intensities in different neighborhoods are drawn from the same PDF. Thus, the Markov PDFs $P(\mathbf{Z}_t)$ must be the same for all $t$. We refer to this as a stationary MRF with the Markov PDF as $P(\mathbf{Z})$. The nonparametric Parzen-window [29] technique gives the probability of a specific intensity neighborhood $\mathbf{z}_t$ as

$$P(\mathbf{Z} = \mathbf{z}_t) \approx \frac{1}{|\mathcal{A}|} \sum_{s\in\mathcal{A}} G(\mathbf{z}_t; \mathbf{z}_s, \sigma) \tag{6}$$

where $G(\cdot)$ is the isotropic Gaussian kernel with mean $\mathbf{z}_s$ and standard deviation $\sigma$ along each dimension. We now describe practical ways for defining Markov neighborhoods $\mathcal{N}_t$ and choosing the Parzen sample $\mathcal{A}$.

The results in this paper are on 2-D MR-image slices and use a $9 \times 9$ voxel neighborhood ($\mathcal{N}_t$). The intensities in the neighborhood are weighted so as to make the neighborhood more isotropic. One such scheme appears in [12]. We handle image boundaries by performing statistical estimation in the cropped feature spaces resulting from the partial neighborhoods.

In practice, neighborhood patterns/statistics are more consistent in proximate regions in the image than between distant regions. In other words, images could be better approximated as realizations of piecewise stationary MRFs. We account for this phenomenon through an effective Parzen-window sampling strategy, namely the local sampling strategy: for each voxel $t$, we draw a unique random sample $\mathcal{A} = \mathcal{A}_t$ from an isotropic Gaussian PDF on the image-coordinate space, with mean at the voxel $t$ and variance $\sigma^2_{\text{spatial}}$. Thus, the sample $\mathcal{A}_t$ is biased and contains more voxels near the voxel $t$ being processed. We have found that the proposed method performs well for any choice of $\sigma_{\text{spatial}}$ that encompasses more than several hundred voxels, together with a correspondingly large value of $|\mathcal{A}_t|$. This behavior is consistent with the previous findings in [12], [30], [31]. All the results in this paper use $\sigma_{\text{spatial}} = 30$ voxels along each cardinal direction and $|\mathcal{A}_t| = 2000$.

## V. Estimating Nonparametric Markov Priors

A Bayesian denoising framework relies on a prior statistical model of the uncorrupted signal. This section describes a novel

method for adaptively inferring the Markov PDF $P(\mathbf{Z})$, that uniquely determines the prior PDF $P(X_t|\mathbf{y}_t)$, based on the input data $\widetilde{\mathbf{x}}$ and the knowledge of the Rician noise model $P(\widetilde{X}_t|x_t)$.

We can, potentially, derive Markov priors from a suitable database of high-SNR brain-MR images, e.g., a set of 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 $\mathbf{x}$. This data, subsequently, gets corrupted by Rician noise. What we observe is only the corrupted data $\widetilde{\mathbf{x}}$—the prior PDF $P(X_t|\mathbf{y}_t)$ or the Markov PDF $P(\mathbf{Z})$ remain unknown. Given sufficiently-many corrupted observations, we can infer the Markov statistics of the corrupted signal accurately [32]. 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, i.e., the prior. In this way, we can *empirically* estimate the unknown prior PDF.

Let us denote the Markov PDF of the corrupted signal by $P_C(\widetilde{\mathbf{Z}})$. For the uncorrupted signal, Parzen-window representation of the Markov PDF is

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

where $\{\mathbf{z}_u\}_{u \in \mathcal{U}}$ and $\sigma$ are the independent (or control) variables for the prior model. This nonparametric model is a general model that can consistently estimate complex Markov PDFs asymptotically as $|\mathcal{U}| \to \infty$ [32]. The goal is to estimate the set $\{\mathbf{z}_u\}_{u \in \mathcal{U}}$ and $\sigma$ based on the knowledge of the observed corrupted-signal Markov statistics $P_C(\widetilde{\mathbf{Z}})$ and the Rician corruption process. The proposed method estimates the independent variables underlying the prior PDF $P_U(\mathbf{Z})$ using an inverse-methods approach that breaks down the problem in the following two parts.

1) *Forward problem*: Given the Rician noise level $\sigma_R$ and an estimate of the independent variables in the prior model, estimate the corrupted-signal statistics $\widehat{P}_C(\widetilde{\mathbf{Z}})$ (Section VIII describes an automatic strategy for estimating $\sigma_R$).
2) *Inverse problem*: Match the estimate of the corrupted-signal Markov PDF $\widehat{P}_C(\widetilde{\mathbf{Z}})$ to the actual Markov PDF corresponding to the corrupted data $P_C(\widetilde{\mathbf{Z}})$ by updating the independent variables underlying the prior model.

The next two sections describe the solutions to the forward and inverse problems in more detail.

### A. Forward Problem: Numerical Solution

This section gives a detailed analysis of the interplay between the Rician noise model and the nonparametric prior model, and presents 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

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

where $\sigma_R$ 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 the proposed Parzen-window prior model, the input to the Rician corruption system is a sum of Gaussians. 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 use a Levenberg–Marquardt curve-fitting technique to fit Gaussians to the corrupted PDFs. 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 $\mu$ and variances $\sigma^2$ of input Gaussians and the noise level $\sigma_R$. We discretize the input parameters at a sufficiently-high resolution and employ bilinear interpolation to read values from the table.

While approximating the system response by a Gaussian, some theoretical issues deserve attention. The Rician PDF $P(\widetilde{X}|x)$ is defined only for nonnegative $x$. However, the Parzen-window prior model with Gaussian kernels extends to negative values too. In this case, the Rician corruption process applies only to the nonnegative part of the Gaussian input (a truncated Gaussian). A large $\sigma$, relative to the magnitude of the input-Gaussian means $\| \mathbf{z}_u \|$, may lead to a poor fit of the system output to a Gaussian. Nevertheless, the central limit theorem significantly alleviates such fears—the theorem states that the sum of arbitrary dependent RVs approaches a Gaussian RV [33]–[35]. This theorem applies to the linear shift-variant Rician corruption process because the Rician noise model $P(\widetilde{X}|x)$ 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. This ensures good fits. Fig. 1 shows that the fitted Gaussians approximate the Rician-corrupted output PDFs reasonably well. For input Gaussians that extend significantly to the negative axis [Fig. 1(a) and (b)], the fit is good, while for the other cases, the fit is close to perfect.

Given the uncorrupted PDF $P_U(\cdot)$ and the Rician noise level $\sigma_R$, the approximate corrupted-signal Markov PDF is

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

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

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

and the entry on the $i$th row of the diagonal covariance matrix $\Psi'_u$ is

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

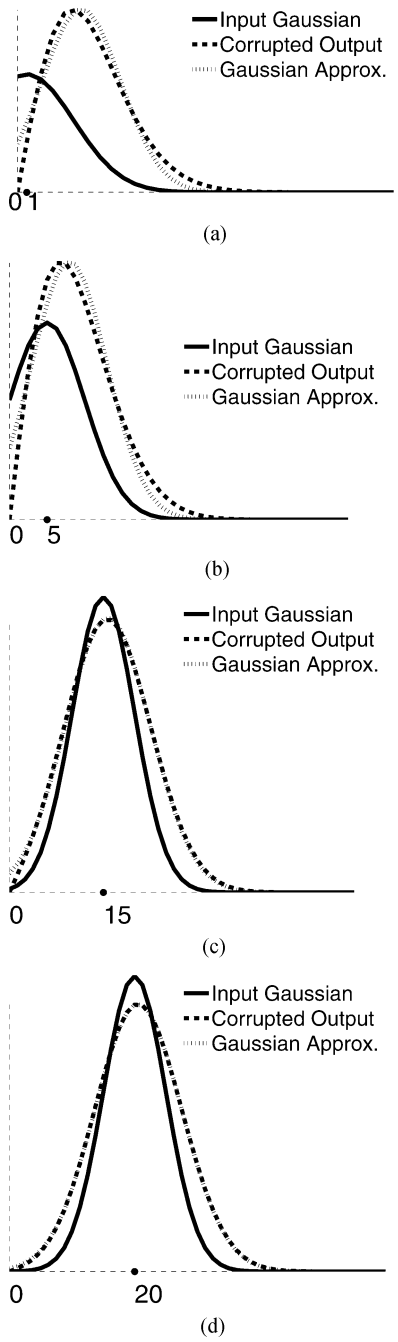



Fig. 1. Rician corruption process in 1-D with Parzen-window bandwidth $\sigma = 5$ and Rician noise level $\sigma_R = 5$. Input Gaussian PDF (solid) is corrupted by Rician noise resulting in the corrupted output PDF (thick dashes). We fit a Gaussian (thin dashes) to this corrupted PDF and measure fitting error by the ratio of the area under the absolute-error curve to the area under the output curve. Graphs show this process for different means of the input Gaussian: (a) $x_u = 1$, fitting error: 14%, (b) $x_u = 5$, fitting error: 13.5%, (c) $x_u = 15$, fitting error: 2.3%, and (d) $x_u = 20$, fitting error: 0.8%.

$\Psi'_u$ is diagonal because of the conditional-independence assumption made previously in Section III.

### B. Inverse Problem: KL-Divergence Optimality

The estimated corrupted-signal PDF $\widehat{P}_C(\widetilde{\mathbf{Z}})$, derived from the uncorrupted-signal model $P_U(\mathbf{Z})$, needs to match the actual Markov PDF $P_C(\widetilde{\mathbf{Z}})$ underlying the observed corrupted data. We propose the KL divergence as a measure of the discrepancy between the two PDFs. Defining $\Theta = \{\mathbf{z}_u\}_{u\in\mathcal{U}}$, the optimal estimate for the model parameters is

$$\begin{aligned}\{\Theta^*, \sigma^*\} &= \underset{\Theta,\sigma}{\operatorname{argmin}}\, \mathrm{KL}\left(P_C \,\|\, \widehat{P}_C\right) \\ &= \underset{\Theta,\sigma}{\operatorname{argmin}}\, E_{P_C}\left[\log \frac{P_C}{\widehat{P}_C}\right] \\ &= \underset{\Theta,\sigma}{\operatorname{argmax}}\, E_{P_C}\left[\log \widehat{P}_C\right] \\ &\approx \underset{\Theta,\sigma}{\operatorname{argmax}} \sum_{t\in\mathcal{T}} \log \widehat{P}_C(\widetilde{\mathbf{z}}_t) \\ &= \underset{\Theta,\sigma}{\operatorname{argmax}} \sum_{t\in\mathcal{T}} \log\left(\sum_{u\in\mathcal{U}} G(\widetilde{\mathbf{z}}_t; \mathbf{z}'_u, \Psi'_u)\right) \quad (12)\end{aligned}$$

where $E_{P_C}$ denotes an expectation computed over a sample drawn from the PDF $P_C(\widetilde{\mathbf{Z}})$.

The optimization problem in (12) is that of ML estimation. ML estimation procedures, however, are well known to need regularization—e.g., Grenander's method-of-sieves [36] regularization—to reduce the chances of the optimization getting stuck on local maxima. We propose to regularize the ML estimation by fixing the value of $\sigma$ beforehand. The enforcement of this regularization is similar in spirit to that used by Geman and Hwang [37] for nonparametric PDF estimation. The proposed method produces an effective optimal estimate for $\sigma$ as follows. It first computes a ML estimate $\widetilde{\sigma}$ for the nonparametric Markov PDF of the corrupted observed sample $\{\widetilde{\mathbf{z}}_t\}_{t\in\mathcal{T}}$. Subsequently, it sets

$$\sigma^* \leftarrow \sqrt{\widetilde{\sigma}^2 - \sigma_R^2} \quad (13)$$

relying on a strategy similar to that used with i.i.d. additive Gaussian noise. Fixing this value for $\sigma$, the proposed method computes an optimal ML estimate for the set $\Theta$ using the EM algorithm. In practice, it suffices usually do not need a highly precise estimate for $\sigma$ because the subsequent-estimation of $\Theta$ (described in the next section) allows a sufficient degree-of-freedom to adapt the PDF estimate to the actual PDF. The results demonstrate that this strategy for choosing $\sigma$ works effectively in practice.

### C. Inverse Problem: Optimization Using EM

This section motivates the use of the EM algorithm [22], [23] for finding the optimal values for $\Theta$ and describes the formulation of the iterative optimization in detail.

The inverse problem in (12) is that of mixture-density parameter estimation—the parameter is the set $\Theta = \{\mathbf{z}_u\}_{u\in\mathcal{U}}$ of the means of Gaussians that define the uncorrupted-signal Markov PDF. The optimization formulation in (12) is a little unwieldy because it contains the logarithm of a sum. The EM approach precisely gets rid of the summation that the logarithm applies to. If we knew which Gaussian component generated each observation, then the probability $\widehat{P}_C(\widetilde{\mathbf{z}}_t)$ is given by evaluating a single Gaussian—the one that generated $\widetilde{\mathbf{z}}_t$. The key idea behind EM is to assume the existence of one hidden RV $L$ along with the observed RV $\widetilde{\mathbf{Z}}$ and their joint PDF $P(\widetilde{\mathbf{Z}}, L)$. The EM algorithm views $(\widetilde{\mathbf{z}}_t, l_t)$ as the complete data. In our case, the

values of $L$, i.e., $\{l_t\}_{t\in\mathcal{T}}$, actually represent the elements in $\mathcal{U}$ and are unknown. Hence, from now on, the values taken by $L$ are denoted by $u$. The PDF $P(L|\widetilde{\mathbf{z}}_t)$ gives the probabilities for different Gaussian components to have generated $\widetilde{\mathbf{z}}_t$. The probability of the observation $\widetilde{\mathbf{z}}_t$ assuming that it came from the $u$th Gaussian is

$$P(\widetilde{\mathbf{z}}_t|u) = G(\widetilde{\mathbf{z}}_t; \mathbf{z}'_u, \Psi'_u) \tag{14}$$

where $\mathbf{z}'_u$ and $\Psi'_u$ are the mean and covariance values, respectively, for the $u$th Gaussian. The EM algorithm iteratively computes the ML estimate of the parameter $\Theta$ as

$$\Theta^* = \operatorname{argmax}_{\Theta} \log P(\widetilde{\mathbf{z}}|\Theta). \tag{15}$$

Each iteration comprises an $E$ (expectation) step and an $M$ (maximization) step [22], [23]. 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 simplification [38], the maximization performed in the $m$th iteration reduces to

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

where $\Theta^{m-1}$ 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 nondecreasing likelihood values $P(\widetilde{\mathbf{z}}|\Theta^m)$ of the observed data and the sequence of estimates converge to a local maximum of the likelihood function.

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

1) Let $\{\widehat{\mathbf{z}}^m_u\}_{u\in\mathcal{U}}$ be the parameter estimate at the $m$th iteration.

2) Use the lookup tables to compute $\widehat{\mathbf{z}'}^m_u$ and $\Psi'^m_u$, $\forall u \in U$, where

$$\widehat{\mathbf{z}'}^m_u(i) = \mathcal{L}_\mu(\widehat{\mathbf{z}}^m_u(i), \sigma, \sigma_R)$$

and

$$\widehat{\Psi'}^m_u(i,i) = \mathcal{L}_\sigma(\widehat{\mathbf{z}}^m_u(i), \sigma, \sigma_R). \tag{17}$$

3) Use the Parzen-window modeling scheme to compute

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T} : P(\widetilde{\mathbf{z}}_t|u) = G(\widetilde{\mathbf{z}}_t; \widehat{\mathbf{z}'}^m_u, \Psi'_u). \tag{18}$$

4) Use Bayes rule to evaluate $P(u|\widetilde{\mathbf{z}}_t), \forall t \in \mathcal{T}, \forall u \in \mathcal{U}$. The initial set of observations $\widehat{\mathbf{z}}^0_u$ is biased because it is derived from the PDF $P(\widetilde{\mathbf{Z}})$ that is close to $P(\mathbf{Z})$. This allows us to ignore the *a priori* probabilities $P(u)$ or consider them equal for all $u$. Thus, the Bayes rule gives

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

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

$$\forall u \in \mathcal{U} : \widehat{\mathbf{z}}^{m+1}_u = \widehat{\mathbf{z}}^m_u + \left(\frac{\partial \widehat{\mathbf{z}}_u}{\partial \widehat{\mathbf{z}'}^m_u}\right)\left(\frac{\sum_{t\in\mathcal{T}} P(u|\widetilde{\mathbf{z}}_t)\widetilde{\mathbf{z}}_t}{\sum_{t\in\mathcal{T}} P(u|\widetilde{\mathbf{z}}_t)} - \widehat{\mathbf{z}'}^m_u\right) \tag{20}$$

where the Jacobian $\partial\widehat{\mathbf{z}}_u/\partial\widehat{\mathbf{z}'}^m_u$ is a diagonal matrix—each component of the neighborhood $\widehat{\mathbf{z}}_u$ is corrupted independently because of the conditional independence assumption on the corruption process—that can be computed numerically using the lookup table $\mathcal{L}_\mu(\cdot)$. For i.i.d. additive Gaussian noise, the Jacobian is the identity matrix. Numerical evidence shows that the derivatives in the Jacobian are always greater than unity, and approach unity for large SNR (where Rician noise behaves very similar to i.i.d. additive Gaussian noise). For low SNR, however, the derivatives can be much larger than unity and this may lead to numerically-large updates. In practice, we replace the Jacobian by identity. This results in a projected-gradient ascent strategy that is still guaranteed to converge.

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

### D. Inverse Problem: Engineering Enhancements for EM

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

To further reduce the computational and space requirements of the algorithm, we 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 $\mathcal{U}$ as a random sample from $\mathcal{T}^\dagger$, with $|\mathcal{U}| = \alpha|\mathcal{T}^\dagger|$. Thus, the computational and space complexities of the EM algorithm are reduced to $O(\alpha\beta^2|\mathcal{T}|^2)$. The results in this paper use $\beta = 0.66$.

## VI. ICER for MRF Inference

The EB approach employs the uncorrupted-signal Markov PDF estimated from the observed corrupted data, as a prior in

the Bayesian denoising of each voxel intensity. At each voxel $t$, the prior PDF $P(X_t|\mathbf{y}_t)$ is

$$P(x_t|\mathbf{y}_t) = \frac{\sum_{u\in\mathcal{U}} G(\mathbf{y}_t;\mathbf{y}_u,\sigma)G(x_t;x_u,\sigma)}{\sum_{u\in\mathcal{U}} G(\mathbf{y}_t;\mathbf{y}_u,\sigma)} \tag{21}$$

and the likelihood PDF $P(\widetilde{x}_t|X_t)$ is

$$P(\widetilde{x}_t|x_t) = \frac{1}{\eta(\widetilde{x}_t,\sigma_R)}\frac{\widetilde{x}_t}{\sigma_R^2}\exp\left(-\frac{\widetilde{x}_t^2+x_t^2}{2\sigma_R^2}\right)I_0\left(\frac{\widetilde{x}_t x_t}{\sigma_R^2}\right) \tag{22}$$

where $\eta(\widetilde{x}_t,\sigma_R)$ is the normalization factor that depends on the observed value $\widetilde{x}_t$ and the noise level $\sigma_R$.

We propose to update voxel intensities $x_t$, to increase the posterior probability $P(x_t|\{x_u\}_{u\in\mathcal{T}\setminus\{t\}},\widetilde{\mathbf{x}})$ in (3), by performing a gradient ascent on the logarithm of the posterior. A gradient ascent on the logarithm of a PDF—i.e., the mean shift [13]—is equivalent to an entropy reduction using the Shannon's entropy measure [11], [12]. Fashing and Tomasi [39] describe the mean shift as a quadratic-bound optimization scheme and present its similarities, and advantages, with respect to a Newton–Raphson scheme. Thus, mean shift, or entropy reduction, is an efficient and practical choice for the optimization. Entropy reduction on this posterior PDF results in the following update rule for all voxel intensities $x_t$:

$$\begin{aligned} x_t &\leftarrow x_t - \lambda\frac{\partial h(X_t|\mathbf{y}_t,\widetilde{x}_t)}{\partial x_t} \\ &= x_t + \lambda\left[\frac{\partial \log P(x_t|\mathbf{y}_t)}{\partial x_t} + \frac{\partial \log P(\widetilde{x}_t|x_t)}{\partial x_t}\right] \\ &= x_t + \lambda\left[\frac{\sum_{u\in\mathcal{U}} G(\mathbf{y}_t;\mathbf{y}_u,\sigma)G(x_t;x_u,\sigma)(x_u-x_t)}{\sum_{u\in\mathcal{U}} G(\mathbf{y}_t;\mathbf{y}_u,\sigma)G(x_t;x_u,\sigma)}\right. \\ &\quad \left. - \frac{\widehat{x}_t^m}{\sigma^2} + \frac{\widetilde{x}_t}{\sigma^2}\frac{I_1\left(\frac{\widetilde{x}_t\widehat{x}_t^m}{\sigma^2}\right)}{I_0\left(\frac{\widetilde{x}_t\widehat{x}_t^m}{\sigma^2}\right)}\right] \end{aligned} \tag{23}$$

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 [40]. These sequence of updates leads to image estimates with nondecreasing posterior probabilities and, hence, guarantee convergence to a local maximum of the posterior PDF. We call this novel algorithm for performing Bayesian estimation on MRFs as the ICER.

## VII. MRI-Denoising Algorithm

The proposed MRI-denoising algorithm produces a MAP image estimate as follows.

1) Represent the prior PDF $P(\mathbf{Z})$ by a Parzen-window sum of isotropic Gaussian kernels with means $\{\mathbf{z}_u\}_{u\in\mathcal{U}}$ and standard deviation $\sigma$. Infer the optimal values of the independent variables underlying the prior PDF (as described in Section V) by minimizing the KL-divergence/mismatch between the observed corrupted-signal Markov PDF and its estimate derived from the prior PDF model.

2) Compute an initial denoised ML image $\widehat{\mathbf{x}}^0 = \{\widehat{x}_t^0\}_{t\in\mathcal{T}}$ where

$$\widehat{x}_t^0 = \text{argmax}_{x_t} P(\widetilde{x}_t|x_t). \tag{24}$$

Compute the mode of each likelihood PDF $P(\widetilde{x}_t|X_t)$ numerically using the iterative mode-seeking mean-shift optimization scheme [39].

3) Given the denoised image estimate $\widehat{\mathbf{x}}^m$ at iteration $m$, the next estimate $\widehat{\mathbf{x}}^{m+1}$ is, $\forall t \in \mathcal{T}$

$$\begin{aligned} \widehat{x}_t^{m+1} = \widehat{x}_t^m + \lambda\left[\frac{\sum_{u\in\mathcal{U}} G(\widehat{\mathbf{y}}_t^m;\mathbf{y}_u,\sigma)G(\widehat{x}_t^m;x_u,\sigma)(x_u-\widehat{x}_t^m)}{\sum_{u\in\mathcal{U}} G(\widehat{\mathbf{y}}_t^m;\mathbf{y}_u,\sigma)G(\widehat{x}_t^m;x_u,\sigma)}\right. \\ \left. - \frac{\widehat{x}_t^m}{\sigma^2} + \frac{\widetilde{x}_t}{\sigma^2}\frac{I_1\left(\frac{\widetilde{x}_t\widehat{x}_t^m}{\sigma^2}\right)}{I_0\left(\frac{\widetilde{x}_t\widehat{x}_t^m}{\sigma^2}\right)}\right] \end{aligned} \tag{25}$$

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

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

The computation for each iteration is $O(|\mathcal{A}_t||\mathcal{T}||\mathcal{N}_t|)$. The results in this paper run the algorithm for one iteration.

## VIII. Results, Validation, and Discussion

This section gives a detailed qualitative and and quantitative analysis of the proposed MRI-denoising algorithm. It compares the performance of the proposed method with several other methods including the state of the art. It also validates the proposed method using simulated MR images with a variety of modalities, noise levels, and intensity-inhomogeneity levels. For experiments with simulated images, we precisely simulate Rician noise by corrupting noiseless real and imaginary MR images (with and without intensity inhomogeneity) by i.i.d. additive Gaussian noise and, subsequently, taking the magnitude of the resulting complex-valued image. All uncorrupted images have intensities ranging between 0 and 100. For an image $\mathring{\mathbf{x}}$, the SNR is $10\log_{10}(\text{Var}(\mathbf{x})/\text{Var}(\mathring{\mathbf{x}}-\mathbf{x}))$, where $\text{Var}(\cdot)$ gives the variance of the intensities in the image. This is the same metric used by Pizurica *et al.* [10]. In this paper, $\mathring{\mathbf{x}}$ is either the noisy image $\widetilde{\mathbf{x}}$ or the denoised image $\widehat{\mathbf{x}}$.

For real MR images, the proposed method entails estimating the Rician noise level. An unbiased estimate of the Rician noise level $\sigma_R$ is given by the square root of half of the expectation of the squared intensity values in the zero-intensity regions of the corrupted image [9]. We can effectively extract the zero-intensity regions from the histogram of image intensities by first locating the peak with the lowest intensity value and then thresholding, conservatively, based on the peak intensity.

Fig. 2 presents the results of denoising a particular slice from volumetric T1-weighted simulated BrainWeb [1] image. Fig.

2(d) shows the difference between the uncorrupted and the corrupted images. The positive bias in the intensity PDF introduced by Rician noise is evident in the lighter background region (higher intensity on the average)—the background corresponds to low signal intensities. The intensities in this difference image do not appear spatially correlated because Rician noise corrupts each voxel independently. The denoised image in Fig. 2(c) shows that the proposed MRI-denoising method reduces the root mean square error (RMSE) by about 40%, or increases SNR by 2 db, as compared to the corrupted image. The image in Fig. 2(e) shows the difference (residual) between the denoised and the uncorrupted images. The low correlation in this images indicates that the proposed method preserves the significant image features. The power spectrum of the difference image in Fig. 2(f) shows the whiteness of the residual. Fig. 2(e) also shows that the proposed method effectively corrects for the positive Rician bias in the corrupted-intensity PDF and thereby enhance inter-tissue contrast—darker background region, as compared to that in Fig. 2(d), implying low error. For the T1-weighted BrainWeb image with 5% noise and 40% intensity inhomogeneity in Fig. 2(b), the average background values are 0.1 for the uncorrupted image, 3.1 for the corrupted image, and 0.03 for the denoised image.

### A. Comparison With Other Markov Priors

This section evaluates the merits in the proposed strategy of modeling the prior nonparametrically and estimating the prior Markov PDF optimally from the data. Fig. 3 compares the performance of the proposed algorithm, qualitatively and quantitatively, with several other methods prevalent in the literature. Each of these methods relies on a Bayesian estimation scheme that employs a likelihood term based on the Rician noise model, but uses different kinds of priors—parametric as well as nonparametric—that are specified by the user beforehand.

First, we compare with methods that use parametric Markov prior PDFs. Such schemes exploit the equivalence between MRFs and Gibbs random fields. The parameters, then, are the parameters of the potential functions in the Gibbs energy. The potential functions, ideally, should adapt to the edges/discontinuities inherent in the image. Some early works along these lines came from Geman and Reynolds [43] and Li [44]. Recently, Basu *et al.* [6] propose a method for denoising diffusion tensor images—also applicable for denoising MR images—that employs an anisotropic smoothness prior based on the work of Perona and Malik [41]. In this paper, we compare the proposed method to two different anisotropic smoothness priors: the anisotropic prior by Perona and Malik [41] and another anisotropic prior based on curvature flow [42]. These methods entail nontrivial parameter tuning and we have manually tuned the free parameters to the best of our ability. Fig. 3(a) and (b) shows results using these schemes that produce lower RMSEs, but at the cost of significant loss in image structure—showing significant correlation in the difference images.

Second, we employ a nonparametric Markov prior—a weaker (more general) prior than the strong parametric anisotropic priors—that adapts to the regularity in the image. This is equivalent to incorporating a fidelity term in UINTA [11], [12] based on the Rician noise model. Unlike this approach, the proposed method optimally estimates the prior before the denoising process starts. Fig. 3(c) shows that there is less loss of structure as compared to the parametric priors. The nonparametric prior reduces the erosion of edges in the image, but is not effective at reducing erosion at the corners/extremities. This performance is, however, at the cost of an increase in RMSE.

Third, we compare with a wavelet-based prior model from Pizurica *et al.* [10]. Wavelet-based denoisers rely on some empirical knowledge of the correlation of wavelet coefficients that represent significant image features across scales. Fig. 3(d) shows that the wavelet denoiser introduces artifacts related to the structure of the underlying wavelet and erodes many important features in the image.

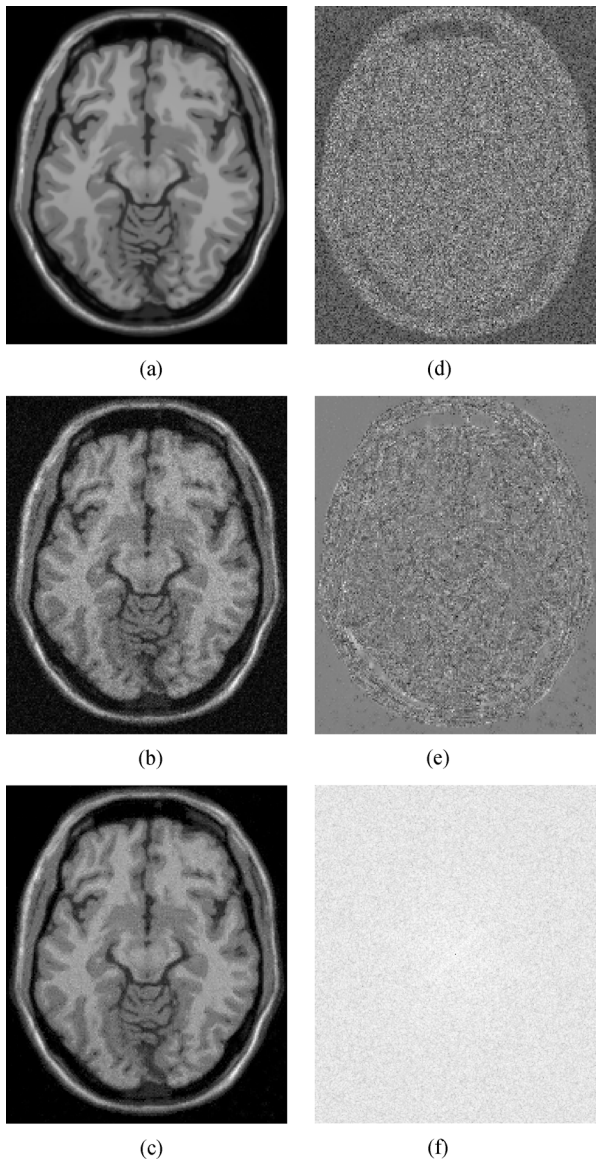



Fig. 2. Results with T1-weighted simulated BrainWeb data (intensity range $0:100$) with the Rician noise level $\sigma_R = 5$ and a 40% intensity inhomogeneity. (a) Noiseless image with a 40% intensity inhomogeneity. (b) Rician-noise corrupted image: $\mathrm{RMSE} = 5.53\%$, $\mathrm{SNR} = 6.6$ db. (c) Denoised image: $\mathrm{RMSE} = 3.3\%, \mathrm{SNR} = 8.6$ db. (d) Difference between the corrupted and uncorrupted images. (e) Difference between the denoised and uncorrupted images. (f) Power spectrum (close to white) of the image in (e).

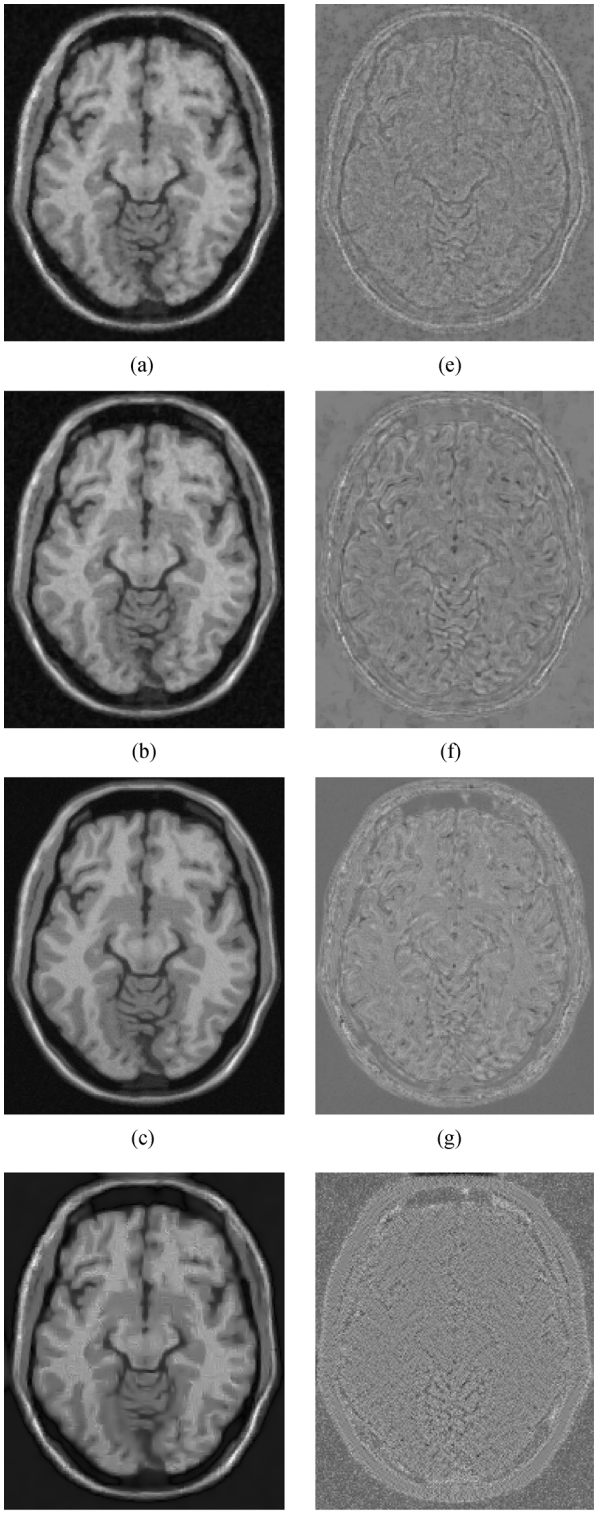



Fig. 3. Effect of different Markov priors on denoising the T1 BrainWeb image in Fig. 2(b) ($\mathrm{RMSE} = 5.53\%$, $\mathrm{SNR} = 6.6$ db). Images denoised using: (a) anisotropic diffusion [41]: RMSE 3.2%, SNR 8.9 db, (b) curvature flow [42]: RMSE 2.8%, SNR 9.3 db, (c) UINTA [12]: RMSE 4%, SNR 7.8 db, and (d) wavelet-based denoiser [10]: RMSE 4.67%, SNR 7.16 db. Images (e)-(h) show the differences between the denoised images in (a)-(d) and the uncorrupted image in Fig. 2(a).

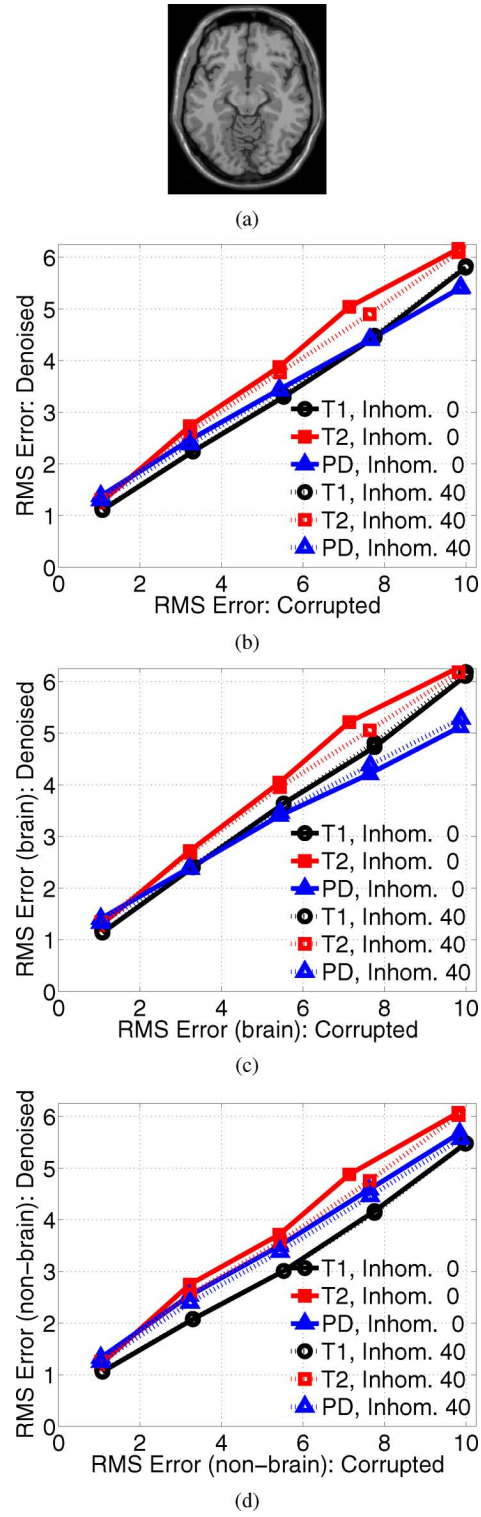
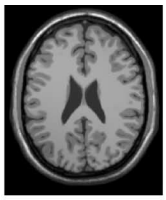
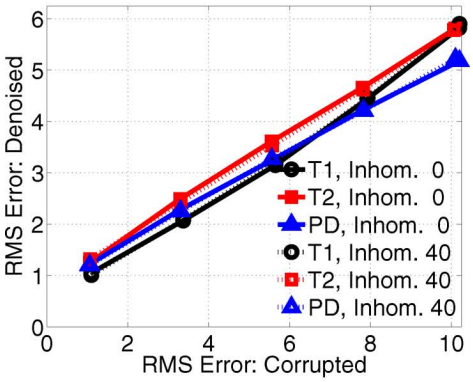



Fig. 4. Denoising results on BrainWeb T1, T2, and PD image slices; only T1 slice shown in (a). Graphs (b)-(d) show RMSEs for noisy (0% and 40% intensity inhomogeneity) and denoised images, on the entire image, the brain region, and the nonbrain region, respectively.

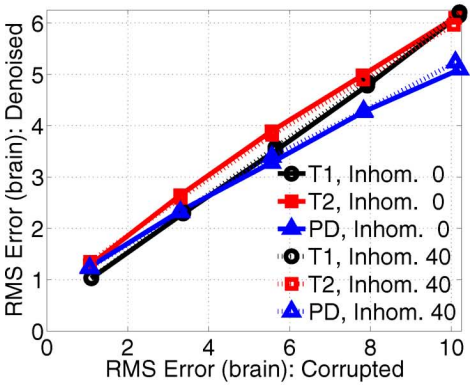
The following picture emerges from the analysis of the aforementioned experiments. Denoising strategies incorporating specific prior image models work best when the data conforms to that model and poorer otherwise. Typically, models imposing stronger/restrictive constraints (e.g., parametric models) give better results when images, or parts of images, conform to those constraints, than with weaker/generic models (e.g., nonparametric models). However, schemes with restrictive models also



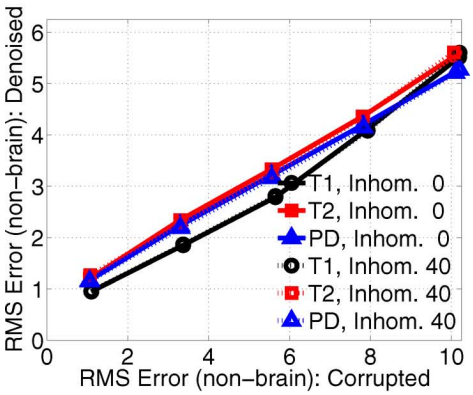
(a)




(b)

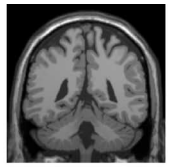



(c)

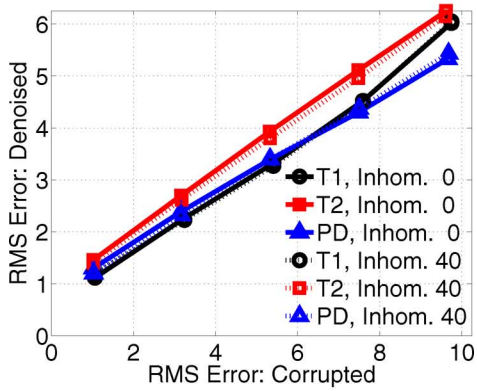



(d)

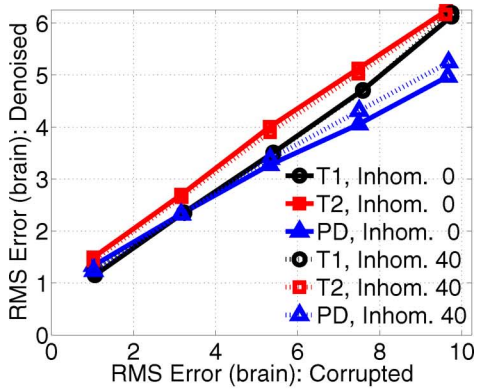
Fig. 5. Denoising results on BrainWeb T1, T2, and PD image slices; only T1 slice shown in (a). Graphs (b)-(d) show RMSEs for noisy (0% and 40% intensity inhomogeneity) and denoised images, on the entire image, the brain region, and the nonbrain region, respectively.



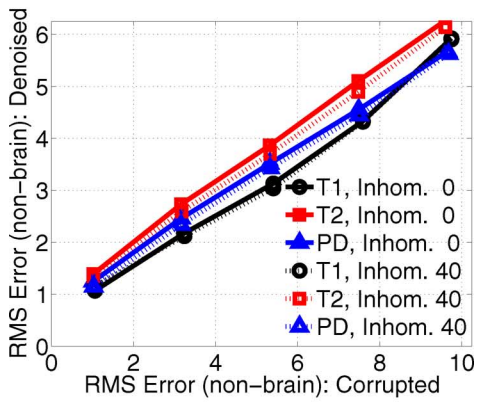
(a)




(b)




(c)




(d)

Fig. 6. Denoising results on BrainWeb T1, T2, and PD image slices; only T1 slice shown in (a). Graphs (b)-(d) show RMSEs for noisy (0% and 40% intensity inhomogeneity) and denoised images, on the entire image, the brain region, and the nonbrain region, respectively.

fare much poorer when the features in the image deviate from the model. In general, selecting a specific Markov prior involves a trade off between achieving low RMSEs and preserving important image structure. The proposed denoising method produces a result that has the lowest degradation of image features, e.g., edges and corners/extremities, as well as a reasonably low RMSE.

### *B. Validation on Simulated MR Images*

This section gives validation results on simulated brain MR images with a wide range of noise and intensity-inhomogeneity

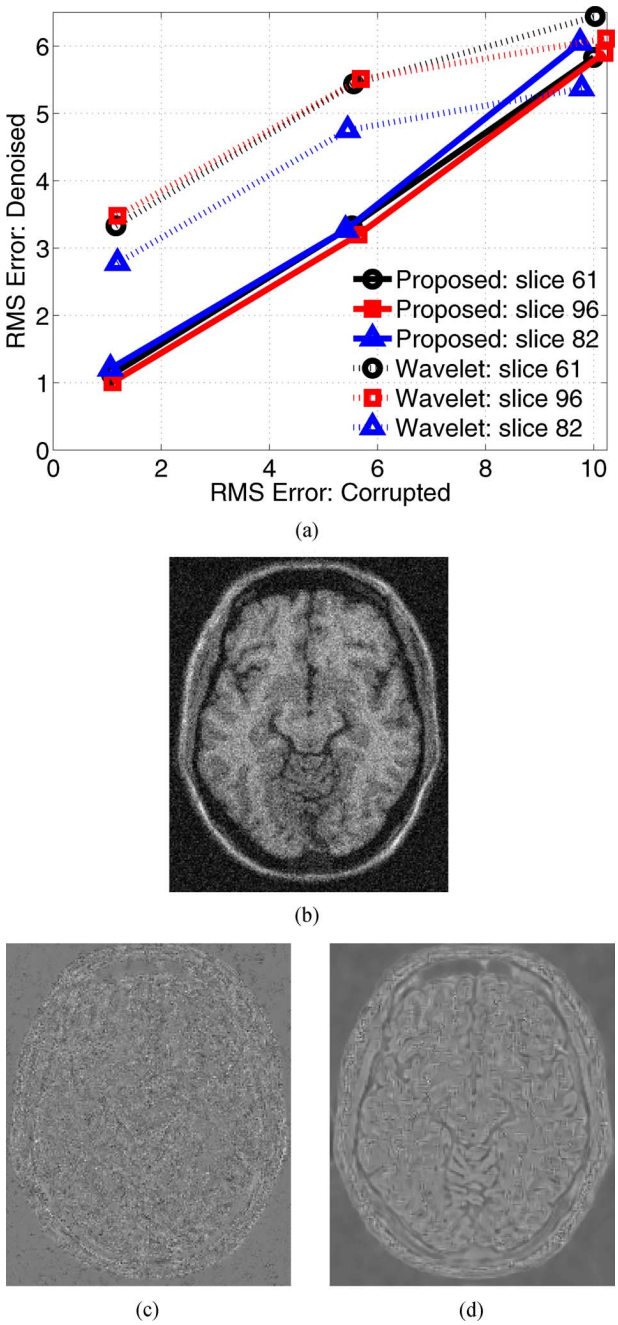



Fig. 7. (a) Quantitative comparison of the proposed method with the state-of-the-art wavelet-based MRI-denoiser by Pizurica *et al.* [10] for the three different slices of T1-weighted BrainWeb data (shown in Figs. 4–6) with varying noise levels and a 40% intensity inhomogeneity. (b) Corrupted T1-weighted data with 9% noise and 40% intensity inhomogeneity. Images (c) and (d) show the difference between the denoised and uncorrupted images for the proposed and wavelet-based [10] methods, respectively, when these methods are applied to the corrupted data in (b).

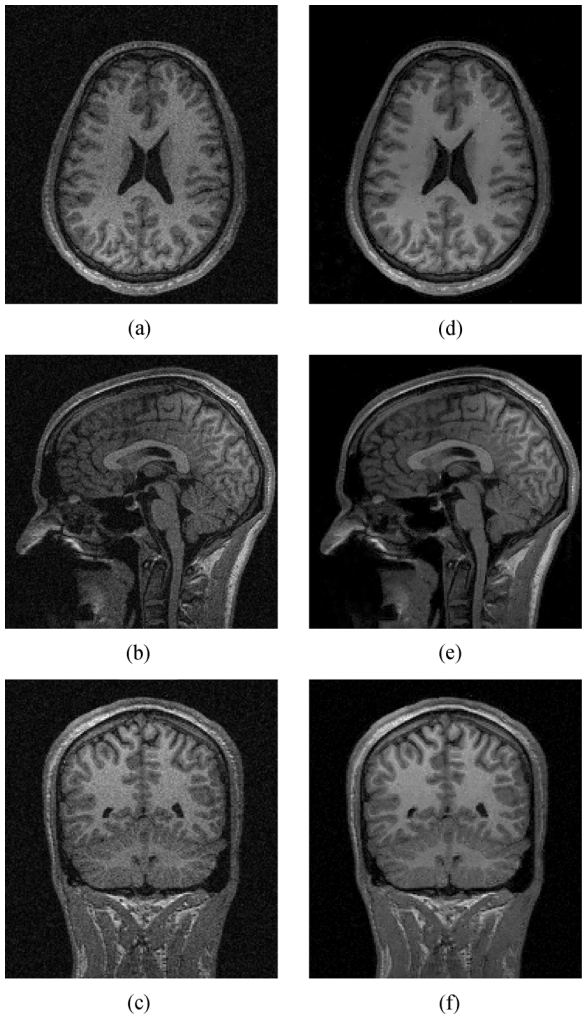



Fig. 8. Denoising real brain-MR images. (a)-(c) Noisy T1-weighted images and (d)-(f) their denoised versions.

levels. We validate using simulated MR images from the BrainWeb [1] database. Figs. 4–6 give the performance of the proposed algorithm on three different slices of the BrainWeb MR data for a wide range of noise and intensity-inhomogeneity levels. We observe that the performance on images with and without intensity inhomogeneity is equivalent. This stems from the ability of data-driven nonparametric models to effectively infer the appropriate Markov statistics for each case and denoise based on the inferred model. We also observe that for a very low Rician noise level, the proposed method does not effectively reduce the RMSE. This may be because of a similar level of variability inherent in the uncorrupted data, and in the estimated uncorrupted-signal Markov PDFs, which makes the prior less informative. As the amount of noise increases, the prior clearly differentiates the structure underlying the data from the noise, thereby yielding better performance. Figs. 4–6 also depicts the RMSE separately in the brain and nonbrain regions. This is because, for most applications, the region of interest is the part of the image comprising the brain. The RMSEs in the nonbrain regions are slightly lower because of the higher regularity (low variability) present in the signal. All graphs are roughly linear with slopes around 0.55.

Fig. 7 shows the qualitative and quantitative comparison of the proposed method with a state-of-the-art wavelet-based MRI-denoising method by Pizurica *et al.* [10]. The experiments employ a T1-weighted BrainWeb MR image with varying noise levels and a 40% intensity inhomogeneity. The proposed method produces lower RMSEs at all noise levels, except the 9% noise level where the RMSEs are similar to the

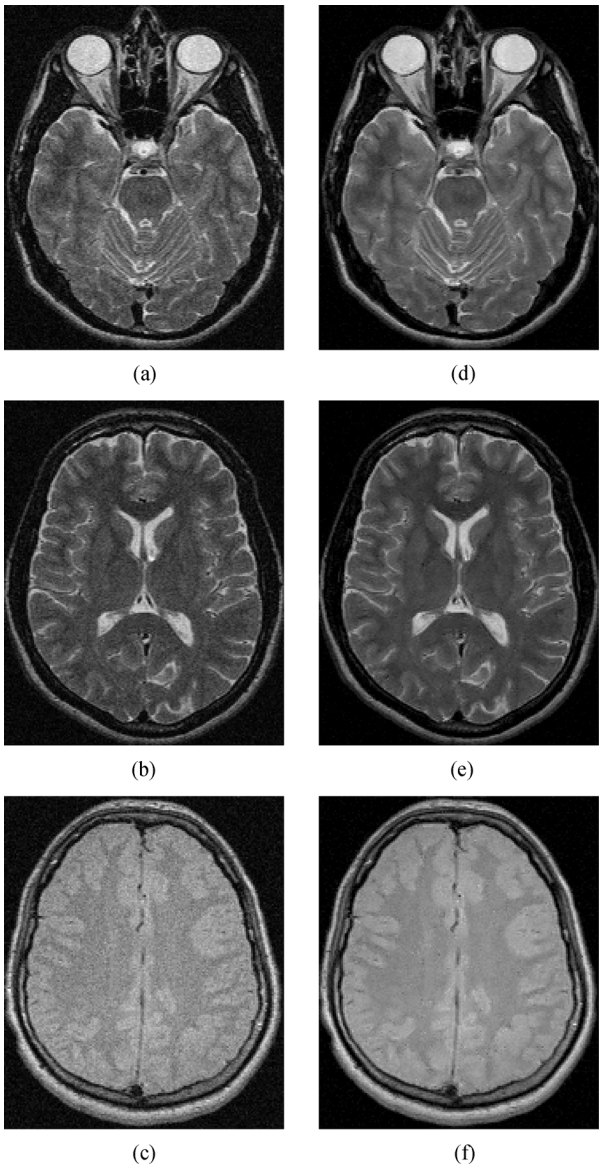



Fig. 9. Denoising real brain-MR images. (a), (b) Noisy T2-weighted images and (d), (e) their denoised versions. (c) Noisy PD-weighted image and (f) its denoised version.

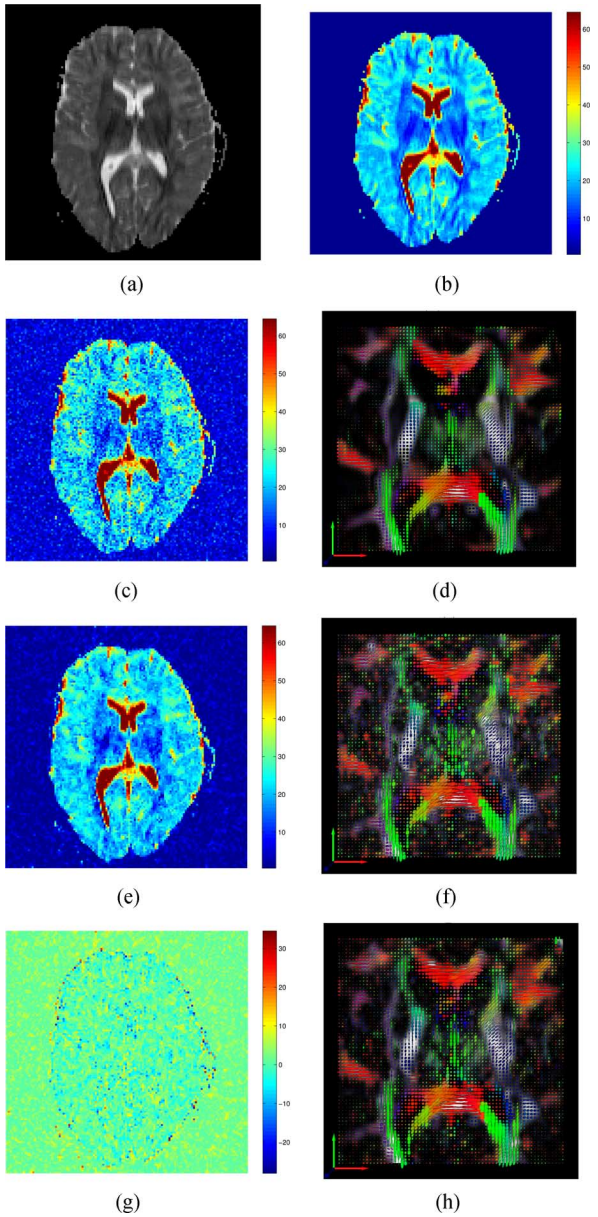



Fig. 10. Denoising a diffusion weighted image (DWI) of the brain (axial slice). (a) Original DWI having high SNR, computed from repeated DWI scans of the same subject (intensity range normalized to 0:100). (b) Original DWI with color-mapped intensities to aid in visualization of subtle image features. (c) Noisy DWI: RMSE 6.1%, $\mathrm{SNR} = 5$ db. (d) Denoised DWI: RMSE 3.6%, $\mathrm{SNR} = 6.6$ db. (e) Difference between the denoised and original image. Diffusion tensor images (zoomed in) (colorcoded: tensor orientations encoded in glyph colors, finite anisotropy encoded in the grayscale background) corresponding to the (f) noiseless, (g) noisy, and (h) denoised DWIs.

wavelet denoiser. However, Fig. 7(c) and (d) shows that the residual for the wavelet-based method is significantly more correlated. Fig. 7(d) also indicates the presence of artifacts in the wavelet-denoised image that relate to the structure of the underlying wavelet.

### C. Results on Real Images From MRI and DWI

Figs. 8 and 9 show the performance of the proposed method on corrupted MR images of adult human brains. The proposed method is able to recover the image features to a significant extent, qualitatively, despite a significant level of intensity inhomogeneity apparent in some images. Note that high levels of intensity inhomogeneity can produce low signal intensities where the effect of the Rician bias can be pronounced.

Fig. 10 shows such an example of the application of the proposed method for denoising DWIs produced by DT MRI. The ground truth [Fig. 10(a) and (f)] is constructed from multiple scans of the same patient followed by tensor reconstruction using linear regression. In comparison to the MR images, we use a smaller $5 \times 5$ voxel weighted neighborhood ($\mathcal{N}_t$) because of the lower spatial resolution of the DWI. The difference image, between the noiseless and denoised, in Fig. 10(e) shows that the proposed method prevents degradation of the key image features. Fig. 10(f)-(h) shows the DT images corresponding to the noiseless, noisy, and denoised DWIs. These results show that the proposed method holds potential for effectively denoising DT images. A detailed analysis of the application of the proposed method for DT denoising, including comparison

with other tensor denoising methods, is an important part of future work.

## IX. Conclusion

This paper presented a novel method for Bayesian denoising of MR images. Following the EB approach, the proposed method employs the uncorrupted-signal Markov PDF, estimated automatically from the observed corrupted data, as a prior in the Bayesian denoising of each voxel intensity. In this way, the proposed Bayesian denoising scheme bootstraps itself by estimating the prior by optimizing an information-theoretic metric using the EM algorithm. The generality and power of nonparametric modeling, coupled with the novel EB scheme for prior estimation, ensures that the proposed method avoids imposing ad hoc prior models for denoising. In this way, the proposed method produces denoised images that preserve the important features, e.g., edges and corners/extremities, and have a reasonably-low RMSE. Furthermore, the paper presents a novel Bayesian-inference algorithm on MRFs, namely ICER that computes an optimal image estimate by performing a gradient ascent on the logarithm of the posterior PDF at each voxel. This results in a mean-shift update that is an efficient bound-optimization technique [39]. The method generalizes in a straightforward manner to multimodal MR images and vector-valued images. A detailed analysis of the method for tensor denoising is an important component of future work.

An intrinsic limitation of the nonparametric prior-PDF model is that its performance degrades for image regions not having sufficiently-many repeated patterns. For instance, the proposed method may find it difficult to denoise features/structures that occur rarely in the image because of theoretically-insufficient data to feed into the nonparametric model. Section IV proposes a practical way to deal with the sparsity of data by generating a Parzen-window sample in the proximity of the voxel being denoised. Nevertheless, the results demonstrate that, for brain-MR images, the proposed algorithm performs well even as the underlying theoretical conditions are relaxed.

## Acknowledgment

The authors would like to thank Prof. G. Shamir and Prof. S. Joshi for valuable discussions and feedback, Prof. J. Gee for the DWI data, and H. Zhang for providing the DT-visualization software.

## References

[1] 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, Jun. 1998.

[2] X. Tai, K. Lie, T. Chan, and S. Osher, Eds., *Image Processing based on Partial Differential Equations*. New York: Springer, 2005.

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

[4] 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. Image Process.*, vol. 12, no. 12, pp. 1579–1590, Dec. 2003.

[5] 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," *Inf. Proc. Med. Imag.*, pp. 148–159, 2003.

[6] S. Basu, P. T. Fletcher, and R. T. Whitaker, "Rician noise removal in diffusion tensor MRI," *Med. Imag. Comput. Comput. Assist. Intervention*, pp. 117–125, 2006.

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

[8] M. Hilton, T. Ogden, D. Hattery, G. Jawerth, and B. Eden, *Wavelet Denoising of Functional MRI Data*, pp. 93–114, 1996.

[9] R. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1408–1419, Oct. 1999.

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

[11] S. P. Awate and R. T. Whitaker, "Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognition*, Jun. 2005, vol. 2, pp. 44–51.

[12] S. P. Awate and R. T. Whitaker, "Unsupervised, information-theoretic, adaptive image filtering for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 364–376, Mar. 2006.

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

[14] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[15] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *IEEE Int. Conf. Comp. Vis. Pattern Recog.*, 2005, vol. 2, pp. 60–65.

[16] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[17] G. Casella, "An introduction to empirical Bayes analysis," *Amer. Statistician*, vol. 39, no. 2, pp. 83–87, 1985.

[18] H. Robbins, "The empirical Bayes approach to statistical decision problems," *Ann. Math. Stat.*, vol. 35, no. 1, pp. 1–20, 1964.

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

[20] G. Motta, E. Ordentlich, I. Ramirez, G. Seroussi, and M. J. Weinberger, "The DUDE framework for continuous tone image denoising," in *Int. Conf. Imag. Proc.*, 2005, pp. 345–348.

[21] C. B. Cordy and D. R. Thomas, "Deconvolution of a distribution function," *J. Amer. Statistical Assoc.*, vol. 92, no. 440, pp. 1459–1465, 1997.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. B, no. 39, pp. 1–38, 1977.

[23] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.

[24] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc.*, vol. 48, pp. 259–302, 1986.

[25] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. New York: Springer, 1995.

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

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

[28] 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.

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

[30] S. P. Awate, "Adaptive Markov Models with Information-Theoretic Methods for Unsupervised Image Restoration and Segmentation," Ph.D., School Comput., Univ. Utah, Salt Lake City, 2006.

[31] S. P. Awate and R. T. Whitaker, "Nonparametric neighborhood statistics for MRI denoising," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2005, vol. 3565, Lect. Notes Comp. Sci., pp. 677–688.

[32] E. Levina, "Statistical issues in texture analysis," Ph.D., Dept. Stat., Univ. California, Berkeley, 1997.

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

[34] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math. J.*, vol. 15, pp. 773–780, 1948.
[35] K. N. Berk, "A central limit theorem for $\mathrm{m}$-dependent random variables with unbounded $\mathrm{m}$," *Ann. Prob.*, vol. 1, no. 2, pp. 352–354, 1973.
[36] U. Grenander, *Abstract Inference*. New York: Wiley, 1975.
[37] S. Geman and C. R. Hwang, "Nonparametric maximum likelihood estimation by method of sieves," *Ann. Stat.*, vol. 10, no. 2, pp. 401–414, 1982.
[38] J. Bilmes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models", Univ. Berkeley, Berkeley, CA, Tech. Rep., 1997.
[39] M. Fashing and C. Tomasi, "Mean shift is a bound optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 471–474, Mar. 2005.
[40] J. Sijbers, A. den Dekker, P. Scheunders, and D. V. Dyck, "Maximum likelihood estimation of Rician distribution parameters," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 357–361, Jun. 1998.
[41] 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, Jul. 1990.
[42] R. Malladi and J. Sethian, "Image processing via level set curvature flow," *Proc. Natl. Acad. Sci. USA*, vol. 92, pp. 7046–7050, 1995.
[43] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
[44] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 6, pp. 576–586, Jun. 1995.