Noise Estimation in Magnitude MR Datasets

Ranjan Maitra and David Faden

***Abstract*—Estimating the noise parameter in magnitude magnetic resonance (MR) images is important in a wide range of applications. We propose an automatic noise estimation method that does not rely on a substantial proportion of voxels being from the background. Specifically, we model the magnitude of the observed signal as a mixture of Rice distributions with common noise parameter. The Expectation-Maximization (EM) algorithm is used to estimate the parameters, including the common noise parameter. The algorithm needs initializing values for which we provide some strategies that work well. The number of components in the mixture model also need to be estimated en route to noise estimation and we provide a novel approach to doing so. Our methodology performs very well on a range of simulation experiments and physical phantom data. Finally, the methodology is demonstrated on four clinical datasets.**

***Index Terms*—mixture model, Bayes Information Criterion, Rayleigh distribution, Rice distribution, model-based clustering, image segmentation, histogram-based estimate**

I. Introduction

Magnetic Resonance (MR) imaging data are very often magnitudes of noise-contaminated complex-valued realizations that are typically well-modeled by the complex Gaussian density [1]. Consequently, the magnitude MR signal at a foreground voxel is Rice-distributed [2], [3] with density

$$r(x;\mu,\sigma) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2+\mu^2}{2\sigma^2}\right)\mathrm{I}_0\left(\frac{x\mu}{\sigma^2}\right), \qquad x > 0 \quad (1)$$

where μ is the underlying true magnitude MR signal, σ is the noise parameter and $\mathrm{I}_0(\cdot)$ is the modified Bessel function of the first kind of zeroth order. The true signal μ at a voxel is determined via the Bloch equation by its underlying physical characteristics [4]. Actually, the Rice distribution is not limited to conventional magnitude MR images, but also arises in MR angiography (MRA) as shown in Andersen and Krisch [5].

As mentioned earlier, the complex data are well-described by a Gaussian distribution [6]. It is the homogeneous standard deviation (SD) of these complex Gaussian densities which translates to the noise parameter σ of the Rice distribution. The parameter σ quantifies the degradation in the MR signal which is disturbed by random noise from a variety of sources such as variation within the magnetic field [7] or random currents in the system under study, or from within the MRI apparatus itself [8], [9].

It is of intrinsic interest to estimate σ in order to judge the quality of acquired images and the imaging setup [10]. Availability of good-quality estimates for σ could potentially help in improving the design of scanners and improve the signal-to-noise ratio (SNR), allowing for shorter image acquisition times and higher contrasts and resolutions [11]. The noise parameter is also important as an input to applications, such as finding contours of the brain [12], in synthetic MRI [13], and in image registration [14], segmentation [15] or restoration [16] algorithms. For more examples, please refer to [17] or [18].

Sijbers *et al* [18] also provide an useful taxonomy of techniques for estimating σ: those based on multiple images and those based on a single image. Estimation in the first case is addressed in [19], which uses the second moment of the Rician distribution. Two geometrically registered images are averaged in k-space. Let $<M_a^2>$ be the spatial average of the corresponding squared magnitude image. Also, let $<M_s^2>$ be the spatial average of one of the squared magnitude images corresponding to the original k-space images. Then $<M_s^2> - <M_a^2>$ provides an unbiased estimator for σ^2. This technique has the advantage of being fully automatic and of gaining precision by using all the data. It is also insensitive to structural errors such as image artifacts [20], but it requires k-space as well as magnitude image data.

At the other end of the spectrum are the single image techniques, many of which are based upon thresholding a histogram of the magnitude image data into background voxels. Essentially these algorithms try to identify the portion of the histogram attributable to background. These background observations can be modeled by a Rayleigh distribution with noise parameter σ. The higher the magnitude of the underlying signal, the more the distribution of the corresponding data is shifted to the right. Hence, the background data are concentrated in the leftmost bins of a histogram. An automated algorithm, provided in [18] selects the number of bins from the left based on a criterion balancing bias and variance and then performs maximum likelihood (ML) estimation for σ assuming the Rayleigh distribution.

The histogram-based algorithm in [18] relies on estimating σ from the background, using the Rayleigh distributional assumption of the voxels, and may be inapplicable in images with little or no background. An alternative, which we propose in Section II of this paper, is to fit Rician mixtures to the distribution of observed magnitude at each voxel, and then to estimate σ using ML estimation. An estimate of the variance of σ provides us with an indication of the stability of our estimate, which we use to select the number of components. Our algorithm estimates both σ and signal levels simultaneously and does not require large areas of background to operate. Results on a detailed series of experiments on computer-generated and phantom datasets performed to evaluate our

R. Maitra is an Associate Professor in the Department of Statistics at Iowa State University, Ames, Iowa, USA and D. Faden is an Engineering Analyst with Google in Mountain View, California.

D. Faden's research was fully supported and R. Maitra's research was partially supported by the National Science Foundation Grant No. CAREER DMS-0437555 and the National Institutes of Health Grant No. DC-000670.

Manuscript received April 22, 2009; accepted May 26, 2009.

algorithm are reported in Section III. We demonstrate application of our methodology on four clinical datasets in Section IV. The paper concludes with some discussion in Section V.

II. Theory & Methods

A. Distribution

Let X_i be the observed magnitude data at the ith voxel, $i = 1, 2, \ldots, n$, where n is the total number of voxels in the image cube. We postulate that each X_i is independently distributed according to the mixture distribution

$$X_i \sim \sum_{j=1}^{J} \pi_j r(x; \mu_j, \sigma) \tag{2}$$

where π_j is the proportion of voxels with underlying signal μ_j and common noise parameter σ. We assume that $\mu_j > 0$ for $j = 1, 2, \ldots, J-1$, while $\mu_J \geq 0$. In case $\mu_J = 0$, the Jth component density is the Rayleigh density given by $r(x; 0, \sigma) = x\sigma^{-2} \exp(-x^2/\sigma^2)$. (In a slight abuse of notation, we will continue, for the sake of convenience, to refer to (2) as a mixture of Ricians, even though one of the components may be Rayleigh-distributed). Further, all μ_js are distinct. One physical interpretation of the mixture model (2) is that there are J (unknown) types of tissue (or material) underlying the image and that every voxel X_i has probability π_j of belonging to the jth type. However, we specifically note that our model is far more general with no restriction on the either the nature or the number of different kinds of tissue or material types or sub-types (called *components*) in the image. We only specify that the observed intensity at a voxel is a composition of the (unobservable) intensity values of these underlying components. Our focus in this paper is exclusively on the estimation of σ given $\vec{X} = \{X_1, X_2, \ldots, X_n\}$ but $\vec{\pi} = \{\pi_j; j = 1, 2, \ldots, J\}$, $\vec{\mu} = \{\mu_j; j = 1, 2, \ldots, J\}$s and the number of components J are unknown nuisance parameters and may need to be estimated in the process. We develop two approaches to estimating J in Section II-C, assuming that it is given in the discussion in the next section.

B. Parameter Estimation Using the EM Algorithm

Let $\vec{\theta} = \{\vec{\mu}, \vec{\pi}, \sigma\}$ be the full set of parameters, assuming J fixed. Chung and Noble [21] have investigated the fitting of a two-component mixture of a Rice and uniform distribution via ML in the context of 3D vessel segmentation of time-of-flight and phase contrast MRA images. Their model had three parameters estimated via a modified expectation-maximization (EM) approach. For the general setting (2), direct parameter estimation can be computationally intractable even for small J. An elegant solution is provided by the EM algorithm [22] which augments the observed magnitude data $\vec{X}$ with unobserved labels $\vec{Z} = \{Z_{i,j}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, J\}$ that correspond to each of the mixture components. $Z_{i,j}$s are indicator variables, with $Z_{i,j} = 1$ indicating that the ith observation has true signal μ_j. Then $\vec{Z}$ and $\vec{X}$ together form the complete data, with complete log likelihood:

$$\ell(\vec{\theta}; \vec{X}, \vec{Z}) = \sum_{i=1}^{n} \sum_{j=1}^{J} Z_{i,j} \left[\log \pi_j + \log r(X_i; \mu_j, \sigma)\right] \tag{3}$$

In the absence of $\vec{Z}$, we replace the terms involving $Z_{i,j}$ in (3) by their conditional expectations given $\vec{X}$ at the current iterates for the parameter values. This forms the *E-Step* of the algorithm. Specifically, letting $\vec{\theta}^{(t)}$ be the parameter estimates at the tth iteration of the EM algorithm, we note that from (3), it is enough to calculate $E_{\vec{\theta}^{(t-1)}}[\vec{Z}|\vec{X}]$ to obtain $E_{\vec{\theta}^{(t-1)}}[\ell(\vec{\theta}; \vec{X}, \vec{Z})|\vec{X}]$. Also, conditional on $\vec{X}$, $\vec{Z}$ has a multinomial distribution with jth cell probability proportional to $\sum_{i=1}^{n} \pi_j r(x_i; \mu_j, \sigma)$. Thus, writing $E_{\vec{\theta}^{(t-1)}}[Z_{i,j}|\vec{X}]$ as $z_{i,j}^{(t)}$ yields in the E-step:

$$z_{i,j}^{(t)} = \frac{\sum_{i=1}^{n} \pi_j^{(t-1)} r(X_i; \mu_j^{(t-1)}, \sigma^{(t-1)})}{\sum_{i=1}^{n} \sum_{q=1}^{J} \pi_q^{(t-1)} r(X_i; \mu_q^{(t-1)}, \sigma^{(t-1)})}.$$

Parameter values maximizing this expected log-likelihood given $\vec{X}$ and current parameter iterates are obtained in the *M-step*. Note that the M-step provides analytical expressions for the updated mixing proportions: $\pi_j^{(t)} = n^{-1} \sum_{i=1}^{n} z_{i,j}^{(t)}, \quad j = 1, 2, \ldots, J$. But unlike for Gaussian mixture components [23], $\vec{\mu}^{(t)}$s and $\sigma^{(t)}$ need to be found by numerical optimization. For this, we use L-BFGS-B [24], a quasi-Newton method capable of handling bounds, which in our case is that all parameters are positive. For this, we need to calculate the gradient vector with components $\frac{\partial}{\partial \vec{\theta}} \ell(\vec{\theta}; \vec{X}, \vec{Z}^{(t)})$ given by:

$$\begin{aligned}
\frac{\partial \ell}{\partial \mu_j} &= \sum_{i=1}^{n} z_{i,j}^{(t)} \left[-\frac{\mu_j}{\sigma^2} + \frac{X_i}{\sigma^2} \frac{\mathrm{I}_1\left(\frac{X_i \mu_j}{\sigma^2}\right)}{\mathrm{I}_0\left(\frac{X_i \mu_j}{\sigma^2}\right)} \right], \quad j = 1, 2 \ldots, J. \\
\frac{\partial \ell}{\partial \pi_s} &= \sum_{i=1}^{n} \left[-\frac{z_{i,J}^{(t)}}{\pi_J} + \frac{z_{i,s}^{(t)}}{\pi_s} \right], \qquad s = 1, 2, \ldots, J-1. \\
\frac{\partial \ell}{\partial \sigma} &= \sum_{i=1}^{n} \sum_{j=1}^{J} z_{i,j}^{(t)} \left[-\frac{2}{\sigma} + \frac{X_i^2 + \mu_j^2}{\sigma^3} - \frac{2 X_i \mu_j}{\sigma^3} \frac{\mathrm{I}_1\left(\frac{X_i \mu_j}{\sigma^2}\right)}{\mathrm{I}_0\left(\frac{X_i \mu_j}{\sigma^2}\right)} \right]
\end{aligned}$$

where $\mathrm{I}_1(t) = \frac{d}{dt} \mathrm{I}_0(t)$ is the modified Bessel function at t of the first kind of first order. $\mathrm{I}_0(\cdot)$ and $\mathrm{I}_1(\cdot)$ are both efficiently obtained via their polynomial approximations (see Page 378 of [25]). Also, $\pi_J = 1 - \sum_{j=1}^{J-1} \pi_j$ is not a free parameter given the other π_js and does not appear in the gradient vector. Further, the general concerns [26] about L-BFGS-B with regard to accuracy and slow convergence in ill-conditioned problems did not bear out in our simulation experiments. Indeed, we noted very substantial improvements in both speed and accuracy in using this over the method of [27]. Thus we advocate using L-BFGS-B in our M-step.

In our implementation, we address separately the cases for when μ_J is positive or zero. Implementation for both cases is similar: when $\mu_J > 0$, it is exactly as above, while for $\mu_J \equiv 0$, we have a mixture of $(J-1)$ Rice and one Rayleigh distribution, all with common noise parameter σ. Hence, there are $(2J-1)$ parameters to be estimated: these are the $(J-1)$ free components in each of $\vec{\mu}$, $\vec{\pi}$ and σ. Parameter estimation proceeds similarly as before, with the additional restriction that $\mu_J \equiv 0$. Once the EM-converged estimates are obtained, the likelihood of (2) is evaluated separately for the cases $\mu_J \equiv 0$ and $\hat{\mu}_J > 0$: the case with the higher value, along with the corresponding parameters $\{\hat{\sigma}, \hat{\vec{\mu}}, \hat{\vec{\pi}}\}$, are the parameter MLEs for given J. For that J, $\hat{\sigma}$, also denoted as $\hat{\sigma}^{(J)}$, is the MLE of the noise parameter of the image.

1) Initialization: The EM algorithm, being iterative, requires initialization, which can tremendously impact performance. Common initialization methods include randomly chosen starting values or using hierarchical clustering to obtain J groupings from which initializing parameters are estimated. However, they can perform poorly in many situations – see *e.g.* [28] who suggested using a multi-staged deterministic initializer that finds a large number of local modes and chooses J representatives from the most widely-separated ones. Though this algorithm was seen to be quite competitive for Gaussian mixtures [28], there was no uniformly clear winner for all cases. Therefore, we adopt a hybrid approach, adapting [28] but also using a more expensive hierarchical scheme when there is evidence of failure to find a true global maximum.

First we describe the adaptation of the initializer in [28] to (2) for given J and n. We propose the following steps:

1) Our first objective is to find a large number (q, say) of local modes. To do so, we choose q evenly-spaced quantiles between zero (which corresponds to a representative of the minimum value) and the maximum. We now find q local modes of the sample, using a computationally efficient implementation [29] of the k-means algorithm, initialized with these quantiles.
2) Our next objective is to obtain a representative from each of the J most widely-separated of the q modes obtained from Step 1. To do so, we apply hierarchical clustering with single linkage to these modes, applying a cut with J clusters. Each of the q local modes are then classified into one of J groups according to this cut.
3) Using the classification thus derived, we set μ_j to be the mean of the q modes assigned to class j. To arrive at initial values for the clustering probabilities π_j, we assign each observation X_i to the closest μ_j and equate π_j to the proportion of observations in the jth class. Finally, σ is chosen to maximize the likelihood given these initializing μs and πs. Parameter values thus obtained are candidate initializers for the EM algorithm.

Comments: A few comments are in order. The first pertains to the choice of q. Obviously $q \geq J$ but $q << n$. Our experiments did not find a set recipe for choosing q that worked well in all situations, so we propose running the initializer with several values of q and then selecting the parameters yielding the highest likelihood from among the resulting candidate initializers. In our experiments, we used q from the set $\{J+2, 2J+2, 3J+2, 4J+2, J^2+2, 2J^2+2, 3J^2+2, J^3+2\}$. While the exact choice of this set was heuristic, the rather large number of candidate qs allows for a greater potential of hitting many local optima and, potentially therefore, the true maxima in Step 3. Finally, for the constrained case $\mu_J \equiv 0$, we modify Step 2 to also include zero and the hierarchical clustering is performed on the q local modes arising from Step 1 and zero.

The J-component Rice mixture has two more free parameters than the $(J-1)$-component Rice mixture, so necessarily has a maximized likelihood value not smaller than the latter. If our estimates do not satisfy this, we get evidence of poor initialization for the J-component model. In such a case, we also try a new set of initial values derived from an expensive hierarchical partitioning approach. Specifically, we obtain a partitioning of the dataset using the parameter estimates $\hat{\vec{\theta}}^{(J-1)}$ for the $(J-1)$-component model and classifying each X_i to the class j with largest $\hat{\pi}_j^{(J-1)} r(X_i; \hat{\mu}_j^{(J-1)}, \hat{\sigma}^{(J-1)})$, but also ensuring that each class is non-empty. This produces intervals of points for each class. We then consider splitting one of the intervals with multiple observations into two partitions using one of the observations as a new end-point of an additional partition. Thus we get a new partition into J groups. Let μ_j and π_j, be the mean and proportion of observations in the jth group respectively, and choose σ to maximize the likelihood given these new μ_j's and π_j's. Thus we get a set of potential parameter initializers. Repeating the process for each split gives us several candidate initial values, from which the best (in terms of highest likelihood) is chosen as the final initializer.

2) Variance of the estimate: A major attraction of likelihood-based parameter estimation methods is the ability to obtain variance estimates. For EM-estimated parameters, Louis [30] provided a convenient approach to calculating the observed information $I_{\vec{X}}$. Thus

$$
\begin{aligned}
I_{\vec{X}} = & - E_{\vec{\theta}} \left[\frac{\partial^2}{\partial\vec{\theta}\partial\vec{\theta}^T} \ell(\vec{\theta}; \vec{X}, \vec{Z}) \mid \vec{X} \right] \Bigg|_{\vec{\theta}=\hat{\vec{\theta}}} \\
& - E_{\hat{\vec{\theta}}} \left[\left\{ \frac{\partial}{\partial\vec{\theta}} \ell(\vec{\theta}; \vec{X}, \vec{Z}) \right\} \left\{ \frac{\partial}{\partial\vec{\theta}} \ell(\vec{\theta}; \vec{X}, \vec{Z} \right\}^T \mid \vec{X} \right] \Bigg|_{\vec{\theta}=\hat{\vec{\theta}}}
\end{aligned}
$$

which can be inverted to form the variance-covariance matrix of $\hat{\vec{\theta}}$. The gradient vector $\frac{\partial}{\partial\vec{\theta}} \ell(\vec{\theta}; \vec{X}, \vec{Z})$ is provided in Section II-B so we now only provide the components of the Hessian $\mathcal{H} = \frac{\partial^2}{\partial\vec{\theta}\partial\vec{\theta}^T} \ell(\vec{\theta}; \vec{X}, \vec{Z})$ needed for obtaining $I_{\vec{X}}$:

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \mu_j^2} = & \sum_{i=1}^{n} -z_{i,j} \left\{ \frac{1}{\sigma^2} + \frac{X_i^2}{\sigma^4} \left[\frac{\mathrm{I}_1(\mu_j X_i/\sigma^2)^2}{\mathrm{I}_0(\mu_j X_i/\sigma^2)^2} - 1 \right] \right. \\
& \left. + \frac{X_i}{\mu_j \sigma^2} \frac{\mathrm{I}_1(\mu_j X_i/\sigma^2)}{\mathrm{I}_0(\mu_j X_i/\sigma^2)} \right\} \\
\frac{\partial^2 \ell}{\partial \mu_j \partial \sigma} = & \sum_{i=1}^{n} 2 z_{i,j} \left\{ \frac{\mu_j}{\sigma^3} + \frac{\mu_j X_i^2}{\sigma^5} \left[\frac{\mathrm{I}_1(\mu_j X_i/\sigma^2)^2}{\mathrm{I}_0(\mu_j X_i/\sigma^2)^2} - 1 \right] \right\} \\
\frac{\partial^2 \ell}{\partial \pi_s^2} = & \sum_{i=1}^{n} \left(-\frac{z_{i,s}}{\pi_s^2} - \frac{z_{i,J}}{\pi_J^2} \right) \\
\frac{\partial \ell}{\partial \pi_r \partial \pi_s} = & \sum_{i=1}^{n} -\frac{z_{i,J}}{\pi_J^2} \qquad \text{when } r \neq s
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \sigma^2} = & \sum_{i=1}^{n} \sum_{j=1}^{J} z_{i,j} \left\{ \frac{2}{\sigma^2} - \frac{3(X_i^2+\mu_j^2)}{\sigma^4} \right. \\
& \left. - \frac{4\mu_j^2 X_i^2}{\sigma^6} \left[\frac{\mathrm{I}_1^2(\mu_j X_i/\sigma^2)}{\mathrm{I}_0^2(\mu_j X_i/\sigma^2)} - 1 \right] + \frac{2\mu_j X_i}{\sigma^4} \frac{\mathrm{I}_1(\mu_j X_i/\sigma^2)}{\mathrm{I}_0(\mu_j X_i/\sigma^2)} \right\}
\end{aligned}
$$

where $r = 1, \ldots J-1$. Elements of $\mathcal{H}$ not specified above are equal to zero. Note that these partial derivatives do not present a substantially additional burden, many of them already having been calculated in the M-Step of Section II-B.

C. Choosing the number of components

Choosing the number of components in finite mixture models or clustering is a long-standing issue with numerous proposed approaches. We refer to [23], [31] for a detailed review of many methods and references on this topic. The most popular approach is, perhaps, the Bayes Information Criterion (BIC) [32] which essentially finds the optimal number (J_{opt}) of groups (from a range $J \in \{1, 2, \dots, J_{\max}\}$) minimizing the negative log likelihood of the J-component model augmented by adding a penalty that is equal to n times the logarithm of the number of parameters in that model. The estimate for σ in the J_{opt}-component mixture model is what we henceforth refer to as our BIC-estimate of σ.

We also propose an alternative approach based on the standard error $SE_{\hat{\sigma}^{(J)}}$. The basic idea here is that as J increases, the model is more adequately specified, decreasing the uncertainty in the parameter estimate. Beyond the true J however, there is again more uncertainty in $\hat{\sigma}^{(J)}$ because (at least some of the) new allocations in the E-step are assigned in error. Thus, once the model has been adequately specified, the $SE_{\hat{\sigma}^{(J)}}$ will again grow for J larger than the true value. Our suggestion therefore, is to look for the first J after which $SE_{\hat{\sigma}^{(J)}}$ rises. We call this the variance approach to estimating J_{opt} and the corresponding $\hat{\sigma}^{(J_{opt})}$ as our variability-based estimate. An advantage of this approach is that it does not need a pre-specified maximum value for the range of the Js. Another advantage is that it takes into account the variability in the parameter estimates in determining J_{opt}.

D. Sampling from the image cube

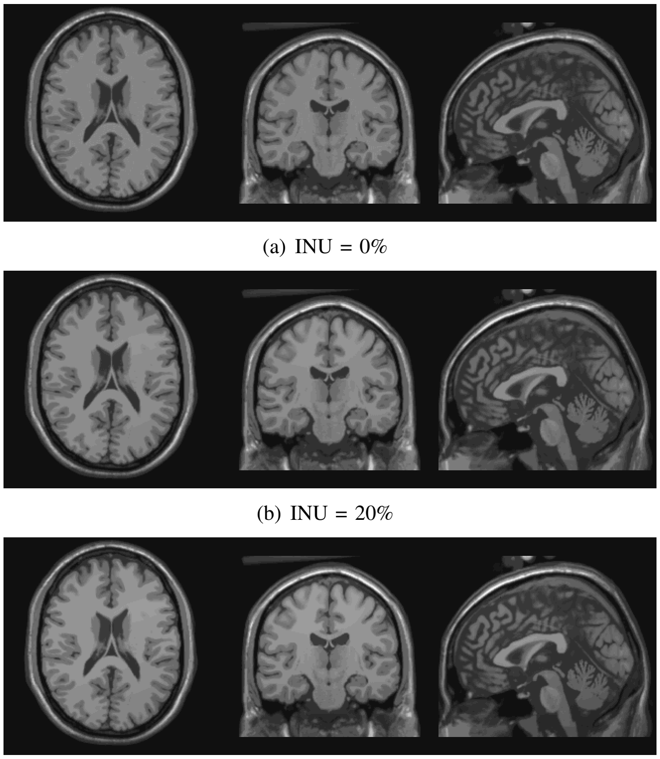
The EM algorithm converges slowly (but surely) and may not be feasible to apply to the entire dataset. Our objective here is simply to obtain an estimate of σ: thus we propose taking a random sample of n voxels. However, we do so by enumerating all the possible m-offset coarse sub-grids of the image cube and randomly selecting one of these sub-grids. Formally, define $\{(u_d, v_e, w_f) : d = 1, 2, \dots, n_u; e = 1, 2, \dots, n_v; f = 1, 2, \dots, n_w\}$ as the voxel coordinates in the original grid of $n_u \times n_v \times n_w$ voxels. Then an m-offset coarse sub-grid is given by $\{(u_{r_u+am}, v_{r_v+bm}, u_{r_w+cm}) : a = 0, 1, \dots, [\frac{n_u}{m}]; b = 0, 1, \dots, [\frac{n_v}{m}]; c = 0, 1, \dots, [\frac{n_w}{m}]\}$, where r_u, r_v, r_w are integers in $[1, m-1]$ and where $[\frac{g}{h}]$ represents the largest integer $\leq \frac{g}{h}$. We propose choosing r_u, r_v and r_w independently at random from $\{1, 2, \dots, m-1\}$ to draw a sub-grid, which we use to estimate σ using our algorithm. Note that we propose selecting a random sub-grid here instead of a simple random sample of n voxels from the entire image cube in order to reduce dependencies that are often introduced along with registration and pre-processing of images.

III. Experiments

The performance of our proposed algorithm was evaluated on a number of realistic computer-generated phantom datasets obtained with known values of σ. We examined performance of our algorithm in obtaining both the BIC-estimated and variability-based-estimated $\hat{\sigma}$. For each dataset, we compared performance with the automatic histogram-based method of [18] using the (default) unit histogram bin width. We also evaluated performance in estimating σ for a twelve-channel 2D physical phantom dataset for which we also stored the complex k-space data. We were thus able to obtain "true" σs from the background voxels of the complex images – these values formed the "ground truth" for our experiments.

A. Experiments on Computer-Simulated Phantom Datasets

1) Experimental Setup: For the phantom, we used the Brainweb interface of [33] with three different proportions of intensity nonuniformity (INU) values to obtain a noiseless version of the ground truth. The INU proportions were set to be at 0%, 20% and 40%, corresponding to the presence of no, modest and substantial bias field in the imaging setup. The noiseless Brainweb image cube was of dimension 181×217×181. We trimmed this image down to 180×216×180 voxels by dropping all voxels with the last index in any dimension. This trimming allowed for uniform sampling over grids with an offset of $m = 12$ pixel coordinates between voxels in each dimension for the BIC- and variability-based estimation methods. Thus, our sample size was reduced to 4,050 well-separated voxels for these two methods. Figures 1 shows axial, coronal and sagittal views of the noiseless Brainweb signal for the three different INU proportions. For the background voxels, we generated independent realizations from the Rayleigh distribution with noise parameter σ: for all other voxels, we generated independent realizations from the Rice distribution with mean given by the true signal at the voxel and the noise parameter σ. We performed experiments with σ equal to be 5, 10, 30, and 50. These σ-values corresponded to average signal-to-noise ratios (SNR) of 5.33, 2.66, 0.89 and 0.53, respectively.

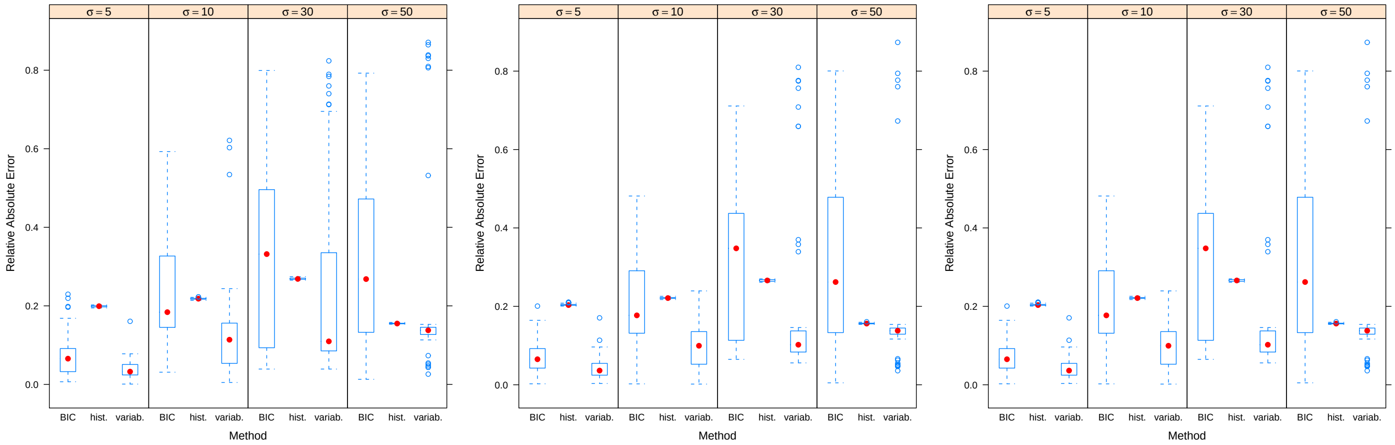


(a) INU = 0%

(b) INU = 20%

(c) INU = 40%

Fig. 1. Axial, coronal and sagittal views of the the noiseless Brainweb image with field intensity nonuniformity (INU) proportions of (a) 0% (b) 20% and (c) 40% used as the true signal in our simulation experiments.



(a) INU = 0%

(b) INU = 20%

(c) INU = 40%

Fig. 2. Relative absolute errors in BIC-, histogram- and variability-based estimated σs for different true values of σ for the Brainweb-simulated data.

TABLE I

SUMMARY OF RELATIVE ABSOLUTE ERRORS OF BIC-, HISTOGRAM- AND VARIABILITY-ESTIMATED σ. FOR EACH TRUE σ, WE REPORT (OVER 50 REPLICATIONS) THE MEAN ABSOLUTE RELATIVE ERROR (**Mean**), THE MEDIAN ABSOLUTE RELATIVE ERROR (**Median**) AND THE AVERAGE RANKING (**Av. Rank**) OBTAINED BY EACH METHOD IN ESTIMATING σ. ALL SUMMARIES ARE IN THREE SIGNIFICANT DIGITS. THE BEST AND SECOND-BEST ESTIMATES FOR EACH SETTING ARE HIGHLIGHTED IN BOLD AND ITALICS, RESPECTIVELY.

		INU Proportion = 0%			**INU Proportion = 20%**			**INU Proportion = 40%**		
True value	**Method**	**Mean**	**Median**	**Av. Rank**	**Mean**	**Median**	**Av. Rank**	**Mean**	**Median**	**Av. Rank**
$\sigma = 5$	BIC	*0.075*	*0.066*	*1.78*	*0.071*	*0.065*	*1.7*	*0.072*	*0.072*	*1.76*
	histogram	0.199	0.199	2.94	0.204	0.203	3.0	0.220	0.217	3
	variability	**0.040**	**0.032**	**1.28**	**0.043**	**0.036**	**1.3**	**0.043**	**0.037**	**1.24**
$\sigma = 10$	BIC	0.239	*0.184*	*2.24*	*0.205*	*0.177*	*2.14*	*0.193*	*0.178*	*2.13*
	histogram	*0.218*	0.218	2.44	0.221	0.221	2.62	0.231	0.232	2.70
	variability	**0.133**	**0.113**	**1.32**	**0.096**	**0.099**	**1.24**	**0.080**	**0.071**	**1.17**
$\sigma = 30$	BIC	0.313	0.332	*2.16*	0.340	0.348	2.41	0.338	0.348	2.38
	histogram	*0.269*	*0.269*	2.20	*0.266*	*0.266*	*2.12*	*0.263*	*0.263*	*2.12*
	variability	**0.241**	**0.110**	**1.64**	**0.202**	**0.102**	**1.47**	**0.207**	**0.104**	**1.50**
$\sigma = 50$	BIC	0.318	0.268	2.26	0.316	0.262	2.27	0.316	0.272	2.30
	histogram	**0.155**	*0.155*	*2.24*	**0.156**	*0.156*	2.32	**0.158**	*0.158*	*2.26*
	variability	*0.230*	**0.138**	**1.50**	0.188	**0.138**	**1.41**	0.216	**0.140**	**1.44**

Thus the noise ranged from the modest to the very substantial. The dataset obtained from each simulation experiment was used to obtain BIC-, variability- and histogram-based [18] estimates of σ. We now report performance evaluations for each of these experiments. Performance was evaluated in terms of relative absolute error, *i.e.* the difference between the absolute value of the estimated $\hat{\sigma}$ and the true value σ. Formally, this is given by $\mid \hat{\sigma} - \sigma \mid / \sigma$. For each setting, we replicated 50 simulated datasets in order to account for simulation variability in our evaluations.

2) Results: Figure 2 provides a graphical display of the distribution of the relative absolute errors in estimating σ using each of the three methods. Descriptive quantitative summaries for these absolute relative errors vis-a-vis estimation method and field INU proportions are provided in Table I. We also ranked each of the estimates provided by the three methods in terms of their closeness to the true σ. The average rank for each estimation method over the 50 replications for each INU setting and true value of σ is also reported in Table I. From the figures and the table, it appears that the BIC-based and the variability-based estimation methods both performed better than the histogram-based method for smaller values of σ. Indeed, for $\sigma = 5$, our variability-based method always outperformed the histogram-based method. For other values of σ, performance was also generally better, except in a few cases. The frequency of these few cases increased with increasing σ. The BIC-based method also performed better than the histogram-based method for $\sigma = 5$, except in a few cases when no bias field was present. Performance of the BIC-based method however decreased markedly with increasing σ: for $\sigma = 30$ and 50, it was at least moderately worse than the histogram-based estimate. The latter appears positively-biased: indeed we over-estimated σ for all 600 experiments performed over the 12 σ-INU combination settings. This positive bias is not surprising since the estimator is based on first automatically identifying the set of background voxels and then calculating the estimator from this set assuming that all observations in it are Rayleigh-distributed. The method in [18] is built on the observed intensity of the voxel only and does not use any spatial information: thus any falsely-identified background voxel would have a positive true signal, biasing upwards $\hat{\sigma}$. In summary however, the overall performance of the variability-based estimator was very competitive. Figure 2 also suggests that there are a few outlying cases where the variability-based estimation method produced substantially high absolute relative errors. An inspection of the cases revealed that there are the cases where σ was grossly under-estimated. Further analysis revealed that this was a consequence of J being over-estimated, with corresponding negatively-biased $\hat{\sigma}$s. These experiments suggest a need for better ways of estimating J. Nevertheless the variability-based estimates of σ performed admirably, generally outperforming both the BIC- and the histogram-based methods. We now investigate performance on a 2D physical phantom.

B. Performance on Physical Phantom Data

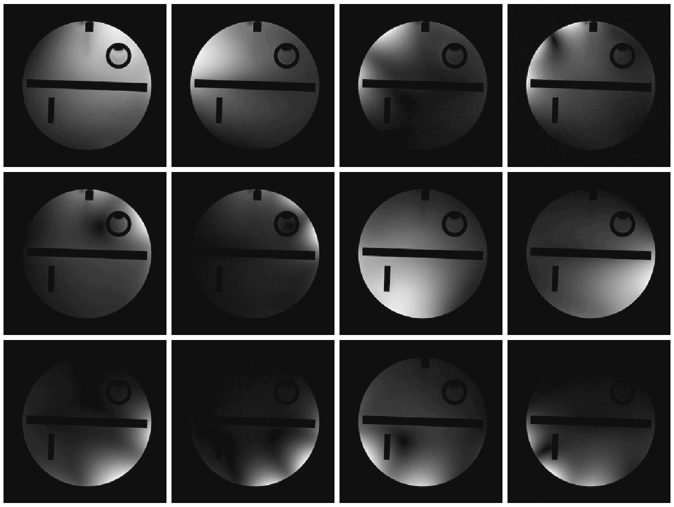


Fig. 3. Magnitude MR images of scanned phantom over 12 channels

1) Experimental Setup: Our next set of evaluations was on a 2D physical phantom scanned in a Siemens 3T Magnetom Trio Scanner using a 12-channel head array coil. A gradient echo (GRE) sequence with parameters: echo time (TE) of 10ms, relaxation time (TR) 180ms and flip angle of 7° was used in this study. The field-of-view (FOV) of the scanned phantom 256mm×256mm and the images were acquired at a resolution of 1mm×1mm. Figure 3 displays the acquired magnitude images for the twelve channels. The complex components of the Fourier-reconstructed k-space data were also stored and available. Background regions were carefully drawn by visual inspection on these complex phantom datasets and σ was estimated for each channel. These twelve σs formed the "ground truth" for this experiment. Finally, we used an offset of $m = 8$ for the BIC- and variability-based estimation methods, reducing the sample size considered for these two methods to $n = 1,024$ pixels.

2) Results: Figure 4 provides a plot of the noise parameter estimates obtained using the three methods for each of the 12 channels of the physical phantom data. The BIC- and variability-based estimates are closer to the "ground truth" than the histogram-based estimates in all but one case. The mean and median absolute relative errors, calculated over the twelve channels were 0.112 and 0.090 for the variability-based method, 0.165 and 0.119 for the BIC-based method and 0.406 and 0.480 for the histogram-based method. Thus, the results on performance of the three methods in on the physical phantom data are in broad agreement with those in Section III-A. Once again, the variability-based method mostly outperformed the histogram-based method. It also moderately outperformed the BIC-based estimation method.

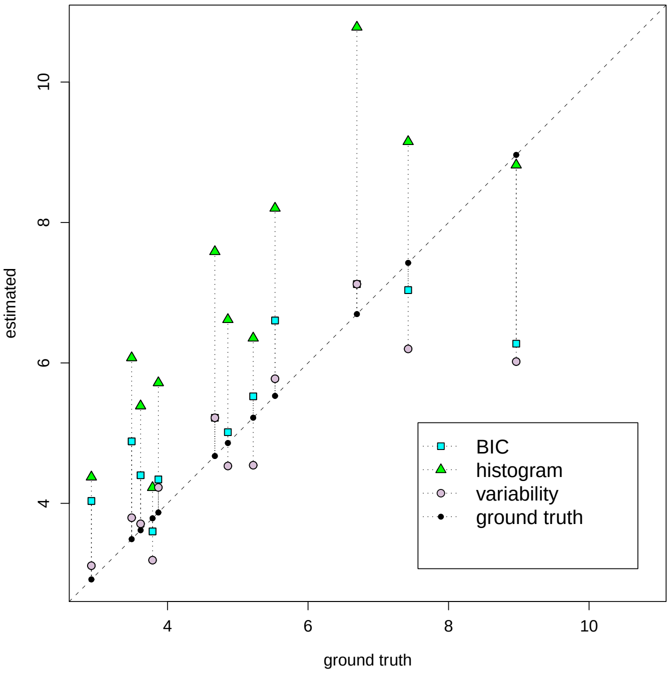


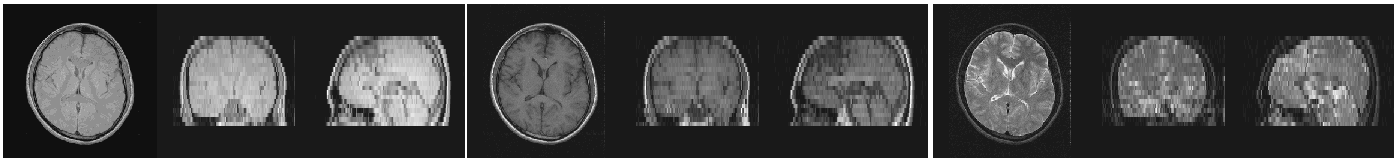
Fig. 4. Estimates of σ obtained using the BIC-, histogram- and variability-based methods plotted against the "ground truth"

In this section, we have demonstrated excellent performance of the suggested estimation methodology on both computer-generated and physical phantom data. We now apply it to four 3D clinical datasets.

IV. Application to Clinical Datasets

A. Description of Datasets

We also report results on applying our noise parameter estimation methodology to four clinical datasets. The first three magnitude MR datasets were obtained on a healthy normal male volunteer using a spin-echo imaging sequence on a GE 1.5T Signa scanner. Proton-density ($\boldsymbol{\rho}$)-weighted, $\mathbf{T}_1$-weighted and $\mathbf{T}_2$-weighted images were obtained at a resolution of 1.15mm×1.15mm×7.25mm in a FOV set to be 294mm×294mm×145mm. Three views of each of these images are presented in Figure 5. For each of these datasets, we also stored the complex-valued images. Once again, background regions were carefully drawn by an expert and the "ground truth" σ calculated as the SD of the complex-valued observations at these background voxels. Our fourth clinical dataset, presented in Figure 6, was on a MR breast scan on a female with suspected malignant lesion. The dataset was obtained on the same Siemens 3T scanner as the phantom, and had TE/TR/flip angle settings of 2.54/4.98/12°. The FOV was 400mm×400mm×220mm and the image was



(a) $\boldsymbol{\rho}$-weighted (b) $\mathbf{T}_1$-weighted (c) $\mathbf{T}_2$-weighted

Fig. 5. Axial (left), coronal (middle) and sagittal views of the the (a) $\boldsymbol{\rho}$-, (b) $\mathbf{T}_1$- and (c) $\mathbf{T}_2$-weighted scans on a normal male volunteer.

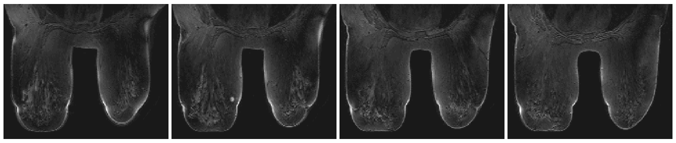


Fig. 6. Axial views (from left to right) at the 72nd, 82nd, 92nd and 102nd slices of the breast image.

acquired at a resolution of 0.8929mm×0.8929mm×1.25mm. A 187.5mm×117.8mm×220mm was cropped to exclude large non-breast regions of chest, air and so on, resulting in an image containing 210×132×176 voxels. Thus, there are few background voxels for this image, thus the histogram-based method may not be particularly applicable. We also do not have access to the complex-valued data at each voxel and thus there is no gold standard for comparisons in this case.

B. Results

Table II summarizes estimates obtained using the three methods for estimating σ. For the $\boldsymbol{\rho}$-weighted MR dataset, it appears that all three methods for estimating σ did not perform particularly well. Our variability-based estimator proved a little better than the BIC-based estimator, over-estimating σ by 36.5% as opposed to 46.4% for the BIC-estimator. Performance of the histogram-based estimator was particularly poor: it over-estimated σ by over 228%. For the $\mathbf{T}_1$-weighted image, the variability-based estimator over-estimated σ by about 7.9%, while the BIC-based and histogram-based estimators had errors of over 59.4% and 136%, respectively. Each of the three estimators had their best performance on the $\mathbf{T}_2$-weighted image, but even here, the histogram-based estimator over-estimated σ by about 52.7%. Both the BIC- and variability-based estimators performed very well, reporting relative errors of under 0.5%. Finally, both the BIC- and variability-based methods estimated σ to be 8.005 for the breast image, while the histogram-based method estimated σ to be 13.366. As mentioned earlier, there is no "ground truth" estimate available here, but the results of the simulation and phantom experiments and the smaller proportion of background voxels in the image provide us with greater assurance on our the variability- and BIC-based estimates.

TABLE II
ESTIMATED σS ON CLINICAL DATASETS OBTAINED USING THE BIC-, HISTOGRAM- AND VARIABILITY-BASED METHODS OF ESTIMATING σ ALONG WITH THEIR "GROUND TRUTH" ESTIMATES (WHERE AVAILABLE).

Dataset	ground truth	BIC	histogram	variability
$\boldsymbol{\rho}$-weighted	0.994	1.455	3.263	1.357
$\mathbf{T}_1$-weighted	0.833	1.328	1.966	0.899
$\mathbf{T}_2$-weighted	0.824	0.828	1.258	0.828
Breast	–	8.005	13.366	8.005

In this section, we have demonstrated application of our σ-estimation methodology to four 3D clinical datasets. Our estimates were the closest to the "ground truth" values when the latter was available, thus providing a measure of surety in the applicability of our methodology to clinical settings.

V. CONCLUSIONS

In this paper, we provide an automated method for estimating the noise parameter in magnitude MR images that is applicable irrespective of whether there is a substantial number of background voxels in the image. Specifically, we model the observed voxel image intensities as a mixture of an unknown number J of Rician distributions with common noise parameter σ. For given J, we use EM to estimate all the parameters in the model given initializing values, strategies to choosing which are also provided. In addition to using BIC to estimate J, we also propose a variability-based approach based on the noise in the estimated σ. Given the EM's computational limitations, we propose choosing at random a coarse sub-grid of the image cube. The EM algorithm is applied to this reduced set of voxels and thus becomes practical to implement. In doing so, we also minimize the effect of local dependencies between observed voxel intensities that may potentially arise in the image as a result of post-processing and image registration. Our methodology supplements the automated histogram estimation method of [18] which relies on identifying background voxels and then using the Rayleigh distribution assumption on these background voxels in order to estimate σ. We report performance on experiments on simulated and physical phantom data, the former in fields with different proportions of bias. Our suggested methodology generally outperformed the others in our experiments, providing evidence of its utility in automatically estimating σ, especially when the presence of large numbers of background voxels is not assured. We also successfully demonstrated application of our methodology to four clinical datasets.

A few points need to be made in this context. First, we note that our algorithm is very computer-intensive with calculations for each J taking as much time as the algorithm in [18]. However, the entire procedure can be parallelized. Further, while not implemented here, the EM algorithm can be substantially sped up using acceleration methods as in [30]. While also not pursued in this paper, we note that the estimates of the signal and associated clustering probabilities provide the ingredients for a segmentation algorithm. The estimation of J

which, although a nuisance parameter, plays an important role in estimating σ. Our experiments indicate that a better choice of J may further improve estimates of σ. One concern with the suggested variability-based approach to estimating J is that it relies entirely on the variability in $\hat{\sigma}$. A more comprehensive approach involving not just $\hat{\sigma}$, but also the other parameters ($\hat{\vec{\pi}}$ and $\hat{\vec{\mu}}$) may possibly help in improving the estimation. Another issue pertains to smoothing and dependent data. We have tried to address this concern by sampling from a sub-grid with offset m (chosen to be 12 in our simulation experiments). It may be possible to explicitly include the dependence structure in our estimation. This is especially true in the context of image segmentation, where the goal is to classify every voxel, unlike the estimation of one parameter (σ), so that a coarser sub-grid may not be possible. Thus, while a promising automated method for noise estimation in magnitude MR images has been developed, a few issues meriting further attention remain.

Acknowledgment

The authors thank J. Sijbers for providing the Matlab code for the the automatic histogram-based estimation method of [18]. We are also very grateful to R. P. Gullapalli and S. R. Roys for the phantom and clinical datasets and to two reviewers whose helpful and insightful comments on an earlier version of this manuscript greatly improved its content.

References

[1] T. Wang and T. Lei, "Statistical analysis of MR imaging and its application in image modeling," in *Proceedings of the IEEE International Conference on Image Processing and Neural Networks*, vol. 1, Apr. 1994, pp. 866–870.

[2] S. O. Rice, "Mathematical analysis of random noise," *Bell System Technical Journal*, vol. 23, p. 282, 1944.

[3] R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Med Phys*, vol. 12, no. 2, pp. 232–233, 1985.

[4] W. S. Hinshaw and A. H. Lent, "An introduction to NMR imaging: From the Bloch equation to the imaging equation," *Proceedings of the IEEE*, vol. 71, no. 3, March 1983.

[5] A. H. Andersen and J. E. Krisch, "Analysis of noise in phase contrast MR imaging," *Med Phys*, vol. 23, no. 6, pp. 857–869, 1996.

[6] J. Sijbers, "Signal and noise estimation from magnetic resonance images," Ph.D. dissertation, University of Antwerp, 1998.

[7] D. Weishaupt, V. D. Köchli, and B. Marincek, *How Does MRI Work?* New York: Springer–Verlag, 2003.

[8] R. C. Smith and R. C. Lange, *Understanding Magnetic Resonance Imaging*. CRC Press LLC, 2000.

[9] M. J. Hennessy, "A three-dimensional physical model of MRI noise based on current noise sources in a conductor," *Journal of Magnetic Resonance*, vol. 147, p. 153169, 2000.

[10] E. R. McVeigh, R. M. Henkelman, and M. J. Bronskill, "Noise and filtration in magnetic resonance imaging," *Med Phys*, vol. 12, no. 5, pp. 586–591, 1985.

[11] R. Bammer, S. Skare, R. Newbould, C. Liu, V. Thijs, S. Ropele, D. B. Clayton, G. Krueger, M. E. Moseley, and G. H. Glover, "Foundations of advanced magnetic resonance imaging," *NeuroRx*, vol. 2, pp. 167–196, April 2005.

[12] M. E. Brummer, R. M. Mersereau, R. L. Eisner, and R. R. J. Lewine, "Automatic detection of brain contours in MRI data sets," *IEEE Transactions on Medical Imaging*, vol. 12, no. 2, June 1993.

[13] I. K. Glad and G. Sebastiani, "A bayesian approach to synthetic magnetic resonance imaging," *Biometrika*, vol. 82, no. 2, pp. 237–250, June 1995.

[14] G. K. Rohdea, A. S. Barnettc, P. J. Bassera, and C. Pierpaoli, "Estimating intensity variance due to noise in registered images: Applications to diffusion tensor MRI," *NeuroImage*, vol. 26, pp. 673–684, July 2005.

[15] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm," *IEEE Trans Med Imaging*, vol. 20, no. 1, pp. 45–47, 2001.

[16] O. A. Ahmed, "New denoising scheme for magnetic resonance spectroscopy signals," *IEEE Trans Med Imaging*, vol. 24, no. 6, pp. 809–816, 2005.

[17] F. d. Pasquale, P. Barone, G. Sebastiani, and J. Stander, "Bayesian analysis of dynamic magnetic resonance breast images," *Journal Of The Royal Statistical Society Series C*, vol. 53, no. 3, pp. 475–493, 2004. [Online]. Available: http://ideas.repec.org/a/bla/jorssc/v53y2004i3p475-493.html

[18] J. Sijbers, D. Poot, A. J. den Dekker, and W. Pintjens, "Automatic estimation of the noise variance from the histogram of a magnetic resonance image," *Phys. Med. Biol.*, vol. 52, pp. 1335–1348, 2007.

[19] J. Sijbers, A. J. den Dekker, J. Van Audekerke, M. Verhoye, and D. Van Dyck, "Estimation of the noise in magnitude MR images," *Magnetic Resonance Imaging*, vol. 16, no. 1, pp. 87–90, 1998.

[20] J. P. D. Wilde, J. Lunt, and K. Straughan, "Information in magnetic resonance images: evaluation of signal, noise and contrast," *Med. Biol. Eng. Comput.*, vol. 35, pp. 259–265, 1997.

[21] A. C. S. Chung and J. A. Noble, "Statistical 3d vessel segmentation using a Rician distribution," in *MICCAI*, 1999, pp. 82–89.

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

[23] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: John Wiley and Sons, Inc., 2000.

[24] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," Northwestern University, Tech. Rep., May 1994. [Online]. Available: www.ece.northwestern.edu/ nocedal/PSfiles/limited.ps.gz

[25] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Ninth Dover printing, tenth GPO printing ed. New York: Dover, 1964.

[26] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "L-BFGS-B – Fortran subroutines for large-scale bound constrained optimization," Northwestern University, Tech. Rep., December 1994.

[27] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[28] R. Maitra, "Initializing partition-optimization algorithms," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 6, pp. 144–157, 2009.

[29] J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," *Applied Statistics*, vol. 28, pp. 100–108, 1979.

[30] T. Louis, "Finding the observed information matrix when using the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 44, no. 2, pp. 226–233, 1982.

[31] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American Statistical Association*, vol. 97, pp. 611–631, 2002.

[32] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, March 1978.

[33] C. Cocosco, V. Kollokian, R. Kwan, and A. Evans, "Brainweb: Online interface to a 3d MRI simulated brain database," *NeuroImage*, vol. 5, no. 4, May 1997.