BAYESIAN REGULARIZATION OF DIFFUSION TENSOR IMAGES USING HIERARCHICAL MCMC AND LOOPY BELIEF PROPAGATION

Siming Wei† *Jing Hua*‡ *Jiajun Bu*† *Chun Chen*† *Yizhou Yu*◇†

† College of Computer Science, Zhejiang University
‡ Wayne State University
◇ University of Illinois at Urbana-Champaign

ABSTRACT

Based on the theory of Markov Random Fields, a Bayesian regularization model for diffusion tensor images (DTI) is proposed in this paper. The low-degree parameterization of diffusion tensors in our model makes it less computationally intensive to obtain a maximum *a posteriori* (MAP) estimation. An approximate solution to the problem is achieved efficiently using hierarchical Markov Chain Monte Carlo (HMCMC), and a loopy belief propagation algorithm is applied to a coarse grid to obtain a good initial solution for hierarchical MCMC. Experiments on synthetic and real data demonstrate the effectiveness of our methods.

Index Terms— Diffusion Tensor Images, Image Restoration, Bayesian Models, Markov Chain Monte Carlo

1. INTRODUCTION

Diffusion tensor imaging (DTI) enables the indirect inference of white matter microstructures by reconstructing local diffusion displacement probability density functions for water molecules from local measurements. When following a 3D Gaussian distribution, this probability density function can be described with a diffusion tensor which is a 3×3 symmetric positive semidefinite matrix whose eigenvalues and eigenvectors characterize underlying fiber orientation and anisotropy. A diffusion tensor is inherently related to the covariance matrix of the Gaussian distribution, and can be reconstructed from diffusion coefficients measured locally along several gradient directions. Due to inherent noise of DTI measurements, the reconstructed tensors are inaccurate, giving rise to possibly erroneous results in the derived white matter fiber orientation which further affects the accuracy of fiber tracking. There is an extensive literature on different regularization techniques [1, 8, 10].

We would like to focus on the statistically sound Bayesian framework for tensor field regularization. Prior distributions in the Bayesian framework are often modeled using Markov Random Fields (MRFs) with pairwise interactions. Bayesian regularization by means of the maximum *a posteriori* (MAP) estimation is well-known in the statistical literature, initiated in the 1980s by Geman and Geman [4]. Bayesian regularization of the primary directions of diffusion tensors has been developed in Poupon *et al.* [10]. Similar work for full diffusion tensors following a multivariate Gaussian distribution was proposed in Martin-Fernandez *et al.* [7]. The generalization of such work to Markov random tensor fields with more generic distributions has been presented in Frandsen *et al.* [3]. It should be noted that the MAP estimation in previous work was achieved using either locally iterative optimization which can be easily trapped in locally optimal solutions or generic Markov Chain Monte Carlo sampling [5] which needs a large number of iterations to converge.

In this paper, we introduce more advanced techniques for computing the MAP estimation of the tensor field. We first develop a Bayesian model based on a low-degree parameterization of diffusion tensors. This model makes MAP estimation less computationally intensive. We further develop a hierarchical MCMC technique that dynamically partitions the original state space into a hierarchy of nested smaller state spaces with an increasing resolution. It is able to converge to better approximate solutions than conventional MCMC in a relatively small number of iterations. Loopy belief propagation (LBP) [9, 6] is applied to a coarse grid of diffusion tensors to quickly obtain a good initial solution for hierarchical MCMC. LBP is an efficient deterministic technique for MRF based optimization. Experimental results confirm that our revised Bayesian model as well as our MAP estimation techniques are both efficient and effective.

2. BASIC MODEL

Let $\mathcal{W}$ be a finite set of voxels in the white matter. We denote the diffusion tensor field reconstructed from measured diffusion coefficients by $\Sigma = \{\Sigma_w : w \in \mathcal{W}\}$, where Σ_w is a 3×3 positive semidefinite matrix at voxel w. Let $\lambda_{w1} \geq \lambda_{w2} \geq \lambda_{w3} \geq 0$ be the eigenvalues of Σ_w with corresponding orthonormal eigenvectors u_{w1}, u_{w2}, and u_{w3}. The fractional anisotropy (FA) index is defined as: $FA_w = \sqrt{\frac{\frac{1}{2}\sum_{i=1}^{3}(\lambda_{wi}-\bar{\lambda}_w)^2}{\frac{1}{3}\sum_{i=1}^{3}\lambda_{wi}^2}}$, and the normalized diffusion tensor at w is: $\bar{\Sigma}_w = \Sigma_w/\bar{\lambda}_w$, where $\bar{\lambda}_w = \frac{1}{3}(\lambda_{w1}+\lambda_{w2}+\lambda_{w3})$. Since most white matter voxels do not have neural fiber crossings, the regularized tensors should most likely be of "cigar" type, which means $\lambda_{w2} = \lambda_{w3}$. For this type of tensors, we define $\sigma_w = \lambda_{w2}/\lambda_{w1}$ as the *eigenratio* of $\bar{\Sigma}_w$. It is obvious that there is a one-to-one mapping between eigenratios and FA values, and both of their ranges are (0,1]. These two quantities are closely related since the larger the eigenratio is, the smaller the FA value is. Moreover, a normalized cigar-type tensor at w can be uniquely determined by the primary direction of the tensor $Md_w = u_{w1}$ and the eigenratio σ_w, i.e. $\bar{\Sigma}_w = \bar{\Sigma}_w(Md_w, \sigma_w)$. These two variables are crucial to fiber tracking. We further define the primary direction field as $Md = \{Md_w : w \in \mathcal{W}\}$ and eigenratio field as $\sigma = \{\sigma_w : w \in \mathcal{W}\}$.

The diffusion function at w is denoted as f_w. For a given direction u on the unit sphere, $f_w(u) = \bar{\lambda}u'\bar{\Sigma}_w(Md_w, \sigma_w)u$. Let $\mathcal{F} = \{f_w : w \in \mathcal{W}\}$ be the field of diffusion functions. Also denote the set of measured diffusion coefficients as $F = \{F_w(u_i) : i = 1, \ldots, k, w \in \mathcal{W}\}$, where $u_1, \ldots, u_k$ are the directions in which the signal intensity is measured. F is determined by the equation $S_w(u_i) = S_{w0} * \exp(-bF_w(u_i))$ and estimated by the least square approximation. Here S_{w0} is the signal intensity without gradient, $S_w(u_i)$ is the measured intensity on direction u_i and b is the diffusion-encoding strength factor. To reduce the noise level of a

given data set, we aim to achieve the MAP estimation $p(\mathcal{F}|F)$. According to the Bayes' theorem, we have $p(\mathcal{F}|F) \propto p(F|\mathcal{F})p(\mathcal{F})$. Therefore, performing regularization is equivalent to solving the following optimization problem

$$\arg\max_{Md,\,\sigma} p(F|\mathcal{F})p(\mathcal{F}). \tag{1}$$

Our prior and likelihood models are revised versions of those in Frandsen *et. al* [3]. The prior distribution $P(\mathcal{F})$ is defined to be

$$p(\mathcal{F}) = \frac{1}{Z_\alpha}\exp\left(-\alpha\sum_{w\sim w'} g(\|\bar{\Sigma}_w(Md_w,\sigma_w) - \bar{\Sigma}_{w'}(Md_{w'},\sigma_{w'})\|)\right), \tag{2}$$

where Z_α is a normalizing constant and $\alpha > 0$. $\|\bullet\|$ represents the Frobenius norm of a matrix, and $\sim$ means w and w' are direct neighbors. To choose an appropriate function g, outlier suppression should be taken into account. So we set $g(x) = c - c*\exp(-x^2/K)$ with constant parameters c and K.

A well-known assumption is that the raw signal intensity follows a Racian-distribution. Thus the noise on measured diffusion coefficients at each voxel is independently and normally distributed([3]). The covariance of these distributions may be varying at different voxels. We define such a covariance as $h_w = \frac{exp(2b\bar{\lambda}_w)+1}{(b*SNR_0)^2}$, where SNR_0 is the signal-to-noise ratio of the signal intensity without any gradient. This is a revised version of the function $h(f_w(\mu))$ in [3]. Their choice of h may result in the preference of spherical (isotropic) tensors during MAP estimation while ours using the average eigenvalue does not have such a bias. Therefore, at each voxel w and direction μ, we have:

$$p(F_w(\mu)|\mathcal{F}) = \frac{1}{\sqrt{2\pi h_w}}\exp\left(-\frac{(F_w(\mu)-f_w(\mu))^2}{2h_w}\right) \tag{3}$$

Then the likelihood $p(F|\mathcal{F})$ is formulated as

$$p(F|\mathcal{F}) = \prod_{w\in\mathcal{W}}\left(\frac{1}{\sqrt{2\pi h_w}}\right)^k \exp\left(-\frac{1}{2}\sum_{i=1}^{k}\frac{(F_w(u_i)-f_w(u_i))^2}{h_w}\right), \tag{4}$$

Now we are ready to solve the optimization problem (1) following the above prior and likelihood models. Set

$$E_1^w(\bar{\Sigma}_w) = \sum_{i=1}^{k}\left(\ln(2\pi h_w) + \frac{(F_w(u_i) - \bar{\lambda}u_i'\bar{\Sigma}_w u_i)^2}{h_w}\right),$$

and

$$E_2^{ww'}(\bar{\Sigma}_w,\bar{\Sigma}_{w'}) = 2\alpha g(\|\bar{\Sigma}_w - \bar{\Sigma}_{w'}\|).$$

Since $\bar{\Sigma}_w$ is uniquely determined by the primary direction and eigenratio, E_1^w is a function of both Md_w and σ_w. Similarly, $E_2^{ww'}$ is a function of Md_w, $Md_{w'}$, σ_w and $\sigma_{w'}$. We also define $E_{Total} = \sum_{w\in\mathcal{W}} E_1^w + \sum_{w\sim w'} E_2^{ww'}$. Therefore, E_{Total} is a function of Md and σ at all voxels. The optimization problem in (1) becomes

$$\arg\min_{Md,\,\sigma} E_{Total} \tag{5}$$

3. MULTILEVEL TENSOR REGULARIZATION

Markov Chain Monte Carlo has been traditionally adopted for solving the optimization problem in (4). Each step of this method needs to sample a normalized cigar tensor from a proposal distribution. Sampling positive semidefinite matrices in the continuous tensor space is time-consuming, and MCMC has a slow convergence rate.

Define the resolution index R_A of a set A with norm $\|\bullet\|$ as

$$R_A = \min\{\|a-a'\| : a \neq a' \text{ and } a,a' \in A\}. \tag{6}$$

We say a set is of high resolution when its resolution index is small. A higher resolution of the tensor state space gives rise to slower convergence of the Markov chain while performing MCMC sampling. Our effort is to overcome this obstacle by means of a multilevel technique for MCMC. As stated in Section 2, a diffusion tensor can be uniquely defined by its primary direction and eigenratio. The key idea is that a tensor space at a low level is generated by a low-resolution primary direction space and a low-resolution eigenratio space. At level l, we denote the state space of primary directions as M^l and that of eigenratios as ER^l. Then the optimization problem at this level is formulated as

$$\arg\min_{Md,\,\sigma} E_{Total}, \quad \forall w\in\mathcal{W},\ Md_w \in M_w^l,\ \sigma_w \in ER_w^l. \tag{7}$$

Our multilevel coarse-to-fine regularization is summarized as follows:

1) At $Level = 1$, let ICOS be an icosahedron with vertices on a unit sphere, and two of its vertices lie on the z axis of the coordinate system. At any voxel w, M_w^l is set to be $\{p : p \text{ is a vertex of ICOS and } z(p) > 0\}$, and ER_w^l is defined as $\{\frac{1}{8},\frac{2}{8},\frac{3}{8},\frac{4}{8},\frac{5}{8},\frac{6}{8},\frac{7}{8}\}$. The resolution indices of this level are $R_{M_w^1} \approx 1.05$ and $R_{ER_w^1} = 0.125$. Perform the loopy belief propagation algorithm detailed in Section 4 to obtain a good initial primary direction and eigenratio at every voxel.

2) Level update: $Level = Level + 1$. Suppose the new level is l. At voxel w, suppose the current state is (Md_w,σ_w). Define a circle on the unit sphere to be

$$C_w = \{u : u \in \text{ unit sphere } S^2 \text{ and } u \in B(Md_w, s*R_{M_w^{l-1}})\},$$

where $B(Md_w, s*R_{M_w^{l-1}})$ is a ball centered at Md_w with radius $s*R_{M_w^{l-1}}$. Here s is a scaling factor. It can be proven mathematically that $s \geq \sqrt{3}/3$ is necessary. M_w^l and ER_w^l are defined as follows. Determine a set PT_w which consists of six points uniformly distributed on the circle C_w. $M_w^l = Md_w \cup PT_w$. ER_w^l is a set of seven points which divide $[\sigma_w - \frac{1}{2}R_{ER_w^{l-1}}, \sigma_w + \frac{1}{2}R_{ER_w^{l-1}}]$ into eight uniform segments. Thus, the tensor state space for the current level l is

$$T^l = \{\bar{\Sigma}_w(Md_w,\sigma_w) | Md_w \in M_w^l,\ \sigma_w \in ER_w^l, w\in\mathcal{W}\}.$$

3) Traverse the voxels at the current level in a sequential order. When updating tensor $\bar{\Sigma}_w$, randomly select a normalized diffusion tensor $\bar{\Sigma}_w'$ in the constructed tensor space, and use it to replace $\bar{\Sigma}_w$ with probability β, where

$$\beta = \exp(\min\{E(\bar{\Sigma}_w) - E(\bar{\Sigma}_w'), 0\}),$$

$$E(\bar{\Sigma}_w) = E_1^w(\bar{\Sigma}_w) + \sum_{w\sim w'} E_2^{ww'}(\bar{\Sigma}_w,\bar{\Sigma}_{w'}).$$

4) If the stopping criteria (which could be a time limit, a number of sweeps, etc.) for the current level have been satisfied, goto 2) unless the highest level has been reached; otherwise, goto 3).

As seen in the above steps, the state space at every level always contains very few candidates, which accelerates the moving speed of Markov chain, and saves both time and memory. In practice, we typically use 6-10 levels in the above multilevel regularization.

4. LOOPY BELIEF PROPAGATION

Quickly obtaining a good initial solution is of great importance to multilevel regularization in Section 3. We chose to achieve this goal through belief propagation [9, 6, 2] which is widely used for MAP estimation to MRF problems. This algorithm is iterative and delivers messages among neighboring nodes in parallel during each iteration. Each message is a vector whose dimension is equal to the number of distinct states in a discrete state space. For a graph without any loops, belief propagation guarantees the optimal solution in a finite number of iterations. For a graph with loops, the same algorithm, which is now called loopy belief propagation (LBP), converges to a good approximation of the optimal solution [6].

Parallel update of messages in BP leads to excessive memory usage since all messages from one iteration need to be saved for the next iteration. To improve the space complexity, we reduce both the number of candidates in the state space and the number of messages. Since our prior distribution prefers similar primary directions at neighboring voxels, we take advantage of this to establish our coarsened LBP model. To reduce the number of messages, we divide the voxel set into $2 \times 2 \times 2$ blocks and only maintain messages among blocks. We denote the block set by $\mathcal{G}$. For any block $g \in \mathcal{G}$,

$$E^g_{G1} = \sum_{w \in g} E^w_1 + \sum_{\substack{w \sim w' \\ w,w' \in g}} E^{ww'}_2, \quad E^{gg'}_{G2} = \sum_{\substack{w \sim w' \\ w \in g \\ w' \in g'}} E^{ww'}_2. \quad (8)$$

It is straightforward to prove that

$$E_{Total} = \sum_{g \in \mathcal{G}} E^g_{G1} + \sum_{g \sim g'} E^{gg'}_{G2}. \quad (9)$$

To reduce the size of the state space, we only optimize the primary directions of the tensors using LBP. Voxels in the same block have the same primary direction, but they still keep their original eigenratios computed from the raw data. From (9), we formulate the following new optimization problem,

$$\arg\min_{Md_g} \Big(\sum_{g \in \mathcal{G}} E^g_{G1}(Md_g) + \sum_{g \sim g'} E^{gg'}_{G2}(Md_g, Md_{g'}) \Big), \quad (10)$$

which can be solved approximately by LBP. An additional technique for saving both the computation time and space cost by half is coloring the blocks with black and white. Adjacent blocks have different colors. The messages from black and white blocks are updated alternatively in consecutive iterations.

Denote the blockwise messages by MG. The block-based LBP is summarized as follows:

1) Devide the dataset into $2 \times 2 \times 2$ blocks, and color the blocks with black and white. Initialize all messages $MG^0_{gg'} = 0$ for $g \sim g'$.

2) Update $MG^t_{gg'}$ iteratively from $t = 1$ to T as follows. The update is only performed on black blocks when t is even and white blocks when t is odd.

$$MG^t_{gg'} = \min_{Md_g} (E^g_{G1}(Md_g) + E^{gg'}_{G2}(Md_g, Md_{g'}) + \sum_{k \neq g', k \sim g} MG^{t-1}_{kg}).$$

3) Compute the optimal state Md^*_g for each block g:

$$Md^*_g = \arg\min_{Md_g} \{ E^g_{G1}(Md_g) + \sum_{k \sim g} MG^{\text{current}}_{kg} \},$$

where MG^{current}_{kg} is the latest version of the message propagated on the edge between k and g.

4) For each voxel $w \in g$, set Md^*_g as its initial primary direction and choose an initial eigenratio closest to the raw tensor's eigenratio from $\{\frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}\}$. Start the multilevel MCMC described in Section 3.

5. RESULTS

Our first experiment is based on synthetic data. Figure 1 contains three images corresponding to the initial, noisy and regularized helical tensor field whose anisotropic elements are colored in red. Figure 1(b) shows the tensor field after Gaussian noise has been added onto both the primary direction and the smaller eigenvalue of the normalized cigar-type tensors. The covariance of the noises were set to 0.05 radian and 0.1, respectively. We further generated synthetic DTIs of this noise-corrupted tensor field for their use in our Bayesian regularization model. Figure 1(c) demonstrates that our regularization method can remove more than 97% of the noise, and thus, is very effective.

We have also applied our regularization algorithm on noisy real DTI data. The resolution of the DTI is 256x256x40, and the diffusion-encoding strength factor $b = 1000$. In our regularization model, we set $\alpha = 3.0$, $c = 1$ and $K = 3$ (see eq. (2)). To verify our method's performance, we compare the results of tractography of well-known fibers from the original noisy data and the processed one using our regularization method. Figure 3(a) and (b) show the fiber tracts that pass through a Region of Interest (ROI) defined on the center sagittal slice of the corpus callosum. In the regularized DTI data, the extracted fiber bundles correctly pass through the corpus callosum ROI laterally making a U-shaped structure, and finally end at the cortex dorsoventral along both sides of the hemispherical cleft, as shown in Figure 3(b). In the original noisy data, the same fiber tracking procedure, however, largely fails in determination of the correct tracts, as shown in Figure 3(a). Figure 3(c) and (d) show the cingulum fiber tracts extracted from the noisy and regularized DTI data, respectively. By defining ROIs, the cingulum fiber tracts can be cleanly extracted from the regularized data as shown in Figure 3(d). However, in the original noisy data, there are many short spurious fibers along the entire tract as shown in Figure 3(c). These comparisons clearly demonstrate the effectiveness of our regularization method.

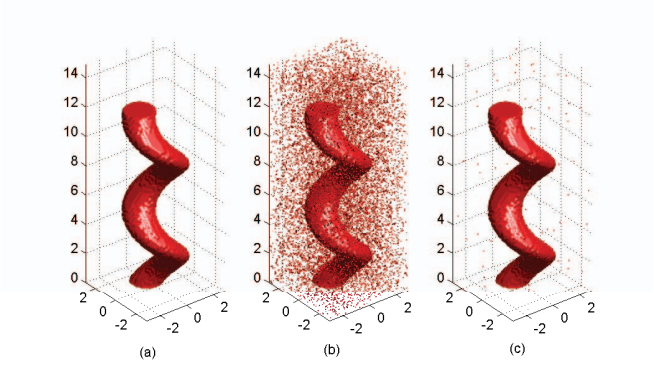

Fig. 1. (a) A synthetic tensor field with anisotropic elements shown in red. (b) Noise-corrupted version of the tensor field in (a). (c) Regularized version of the tensor field in (b).

We further investigated the running time and convergence behavior of our regularization method on the aforementioned real DTI data. A comparison of convergence behavior between conventional MCMC and our Hierarchy MCMC (HMCMC) is shown in Figure 2. Their initial tensor fields are installed by the same LBP algorithm. We maintain six levels in our HMCMC, and the number of sweeps at each level is respectively set to 100,150,50,50,20,20. The curve for HMCMC shows the objective function (minus log probability den-

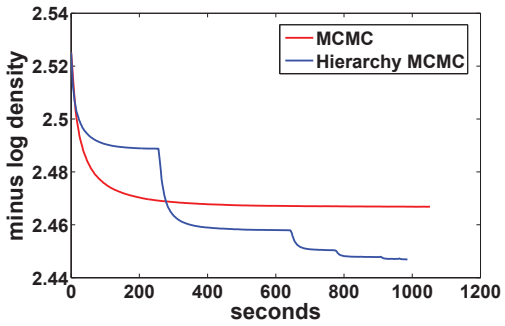

Fig. 2. Convergence behavior of MCMC and Hierarchy MCMC. Minus log probability density is shown as a function of running time on an Intel Pentium D 3.0GHz processor.

# LBP iterations	5	10	15	20
LBP run time(s)	9	18	28	37
# MCMC sweeps	20	35	41	43
MCMC run time(s)	54	93	109	114
saves(s)	45	75	81	77

Table 1. Performance comparison between LBP and pure MCMC on an Intel Pentium D 3.0GHz processor.

sity) drops rapidly at the first few sweeps of each level, and it eventually converges to a better approximate solution than conventional MCMC.

Table 1 justifies the use of LBP to generate an initial solution. It compares the performance between LBP and pure MCMC at the first level. In pure MCMC, we initialize the tensor at each voxel w with the one in the discrete state space that minimizes E_1^w. In this comparison, the number of iterations (sweeps) and the time to reach the same value of the objective function are given in the same row. We can see that running LBP 15 iterations saves more than 80 seconds.

6. CONCLUSION

In this paper, we have presented a Beyesian regularization model for DTIs solved using MAP. This model introduces a low-degree parameterization of diffusion tensors that make MAP estimation less expensive. The hierarchical MCMC algorithm we presented is a much improved version of MCMC and it is able to converge to better approximate solutions with lower energies. We also use LBP on a coarse grid to install a good initial solution for hierarchical MCMC. Experiments demonstrated the effectiveness of our methods.

Acknowledgments

We would like to thank the reviewers for their valuable comments. This work was partially supported by National Science Foundation (IIS 09-14631) and National Natural Science Foundation of China (60728204/F020404).

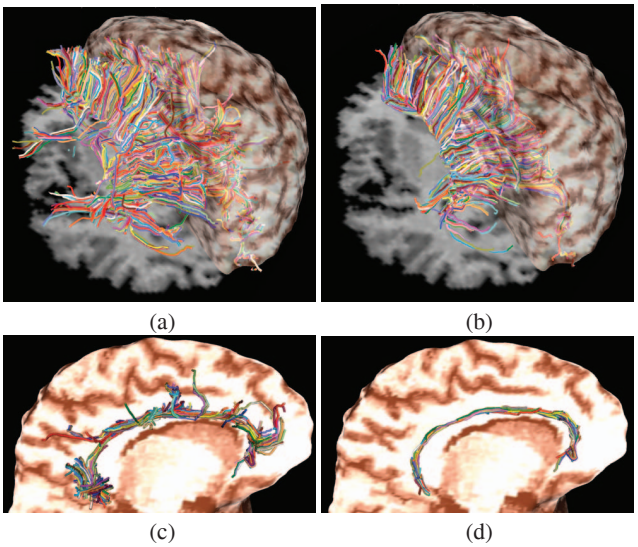

Fig. 3. The results of fiber tracking on the original noisy DTI dataset and the regularized one.(a) The tracked fiber tracts originated from a ROI defined on the center image slice of the corpus callosum in the noisy DTI dataset; (b) The tracked fiber tracts from the same ROI as in (a) in the regularized data; (c) The result of tracking the cingulum fibers in the noisy data; and (d) The result of tracking the cingulum fiber tracts in the regularized data.

References

[1] A.W. Anderson. Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magnetic Resonance in Medicine*, 46:1174–1188, 2001.

[2] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1), 2006.

[3] J. Frandsen, A. Hobolth, L. Østergaard, P. Vestergaard-Poulsen, and E.B. Vedel Jensen. Bayesian regularization of diffusion tensor images. *Biostatistics*, 8(4):784–799, 2007.

[4] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intellligence*, 6:721–741, 1984.

[5] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, 1996.

[6] Y. Weiss J.S. Yedidia, W.T. Freeman. Understanding belief propagation and its generalizations. Technical report, Mitsubishi Electric Research Laboratories, MERL-TR-2001-22, 2002.

[7] M. Martín-Fernández, C.-F. Westin, and C. Alberola-López. 3d bayesian regularization of diffusion tensor mri using multivariate gaussian markov random fields. In *MICCAI, Lecture Notes in Computer Science*, volume 3216, pages 351–359, 2004.

[8] G.J.M. Parker, J.A. Schnabel, M.R. Symms, D.J. Werring, and G.J. Barker. Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 11:702–710, 2000.

[9] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, 1988.

[10] C. Poupon, C.A. Clark, V. Frouin, J. Regis, I. Bloch, D. Le Bihan, and J.-F. Mangin. Regularization of diffusion-based directional maps for the tracking of brain white matter fascicles. *NeuroImage*, 12:184–195, 2000.