A robust variational approach for simultaneous smoothing and estimation of DTI

Meizhu Liu a,⁎, Baba C. Vemuri b, Rachid Deriche c

a Siemens Corporate Research & Technology, Princeton, NJ, 08540, USA
b Dept. of CISE, University of Florida, Gainesville, FL, 32611, USA
c Project-Team ATHENA, Inria Sophia Antipolis - Méditerranée, 06902, France

A R T I C L E   I N F O

Article history:
Accepted 7 November 2012
Available online 17 November 2012

Keywords:
DTI estimation
Variational principle
Non-local means
Total Kullback–Leibler divergence
Diffusion MRI
Limited memory quasi-Newton method

A B S T R A C T

Estimating diffusion tensors is an essential step in many applications — such as diffusion tensor image (DTI) registration, segmentation and fiber tractography. Most of the methods proposed in the literature for this task are not simultaneously statistically robust and feature preserving techniques. In this paper, we propose a novel and robust variational framework for simultaneous smoothing and estimation of diffusion tensors from diffusion MRI. Our variational principle makes use of a recently introduced total Kullback–Leibler (tKL) divergence for DTI regularization. tKL is a statistically robust dissimilarity measure for diffusion tensors, and regularization by using tKL ensures the symmetric positive definiteness of tensors automatically. Further, the regularization is weighted by a non-local factor adapted from the conventional non-local means filters. Finally, for the data fidelity, we use the nonlinear least-squares term derived from the Stejskal–Tanner model. We present experimental results depicting the positive performance of our method in comparison to competing methods on synthetic and real data examples.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Diffusion weighted magnetic resonance imaging (MRI) is a very popular imaging technique that has been widely applied (Jones, 2010) in recent times. It uses diffusion sensitizing gradients to non-invasively capture the anisotropic properties of the tissue being imaged. Diffusion tensor imaging (DTI) approximates the diffusivity function by a symmetric positive definite tensor of order two (Basser et al., 1994). DTI is an MRI modality that provides information about the movement of water molecules in a tissue. DTI describes the diffusion direction of water molecules in the brain which is associated with the direction of fiber tracts in the white matter. When this movement is hindered by membranes and macromolecules, water diffusion becomes anisotropic. Therefore, in highly structured tissues such as nerve fibers, this anisotropy can be used to characterize the local structure of the tissue. Consequently, many applications are based on the estimated diffusion tensor fields, such as registration (Gur and Sochen, 2007; Jia et al., 2011; Wang et al., 2011; Yang et al., 2008; Yeo et al., 2009), segmentation (Descoteaux et al., 2008; Goh and Vidal, 2008; Hasan et al., 2007; Lenglet et al., 2006; Liu et al., 2007; Motwani et al., 2010; Savadjiev et al., 2008; Vemuri et al., 2011; Wang and Vemuri, 2005), atlas construction (Assemlal et al., 2011; Barmpoutis and Vemuri, 2009; Mori et al., 2008; Xie et al., 2010), anatomy modeling (Faugeras et al., 2004), fiber tract related applications (Burgela et al., 2006; Durrleman et al., 2011; Lenglet et al., 2009; Mori and van Zijl, 2002; Savadjiev et al., 2008; Wang et al., 2010, 2012; Zhu et al., 2011) and so on. All of these latter tasks will benefit from the estimation of smooth diffusion tensors.

Estimating the diffusion tensors (DTs) from DWI is a challenging problem, since the DWI data are invariably affected by noise during its acquisition process (Poupon et al., 2008b; Tang et al., 2009; Tristan-Vega and Aja-Fernandez, 2010). Therefore, a robust DTI estimation method which is able to perform a feature preserving denoising is desired. For most of the existing methods, the DTs are estimated by using the raw diffusion weighted echo intensity image (DWI). At each voxel of the 3D image lattice, the diffusion signal intensity S is related with its diffusion tensor D ∈ SPD(3) via the Stejskal–Tanner equation (Stejskal and Tanner, 1965)

S = S₀ exp (−bgᵀDg).

(1)

where S₀ is the signal intensity without diffusion, b is the b-value and g is the direction of the diffusion sensitizing gradient.

There are various methods (Barmpouris et al., 2009a; Batchelor et al., 2005; Chang et al., 2005; Chefd’hotel et al., 2004; Fillard et al., 2007; Hamarneh and Hradsky, 2007; Mangin et al., 2002; Mishra et al., 2006; Niethammer et al., 2006; Pennc et al., 2006; Poupon et al., 2008b; Salvador et al., 2005; Tang et al., 2009; Tristan-Vega...
and Aja-Fernandez, 2010; Tschumperle and Deriche, 2003, 2005; Vemuri et al., 2001; Wang et al., 2003, 2004) in existing literature, to estimate $D$ from $S$. A very early one is a direct tensor estimation (Westin et al., 2002), which gives an explicit solution for $D$ and $S$. Though time efficient, it is sensitive to noise because only 7 gradient directions are used to estimate $D$ and $S$. Another method is the minimum recovery error (MRE) estimation or least squares fitting (Basser et al., 1994) which minimizes the error when recovering the DTs from the DWI. MRE is better than direct estimation because it uses more gradient directions, which increase its reliability. However, it does not smooth the DWI or the DTI, and thus it is subject to noise in the input data.

With this in mind, many denoising frameworks (Gilboa et al., 2004; Spira et al., 2007) have been proposed to improve the signal to noise ratio (SNR). Some methods perform denoising on the DWI and then estimate the DTI. Typical approaches to DWI denoising are designed according to the statistical properties of the noise. Most of these approaches assume that the noise follows the Rician distribution (Descoteaux et al., 2008; Koay and Basser, 2006; Landman et al., 2007; Piurica et al., 2003), and when denoising, they use the second order moment of the Rician noise (McGibney and Smith, 1993), maximum likelihood (ML) (Sijbers and den Dekker, 2004) which minimizes the error when recovering the DTs from the raw DWI and then perform denoising on the tensor field. The proposed model is given by the following equation:

$$
\min_{S_0, D \in \text{SPD}} E(S_0, D) = (1-\alpha-\beta) \int_\Omega \sum_{i=1}^n \left( S_i - S_0 \exp \left( -\frac{b g_i D g_i}{2} \right) \right)^2 dx + \alpha \int_{\Omega \times \Omega} w_1(x, y) (S_0(x) - S_0(y))^2 dy dx + \beta \int_{\Omega \times \Omega} w_2(x, y) \delta(D, D(y)) dy dx,
$$

where $\Omega$ is the domain of the image, $n$ is the number of diffusion gradients, $v(x)$ is the search window at voxel $x$, and $\delta(D, D(y))$ is the total Kullback-Leibler (tKL) divergence (Vemuri et al., 2011) between tensors $D$ and $D(y)$ which will be explained in detail later. The first term captures the non-linear data fitting error, the second and third terms are smoothness constraints on $S_0$ and $D$. $\alpha$ and $\beta$ are constants balancing the fitting error and the smoothness. $w_1(x, y)$ and $w_2(x, y)$ are the regularization weights for $S_0$ and $D$. Since $S_0$ and $S$ are linearly related, while $D$ and $S$ are “logarithmically” related, so we use different methods to calculate $w_1(x, y)$ and $w_2(x, y)$.

The discrete case of Eq. (2) is

$$
\min_{S_0, D \in \text{SPD}} E(S_0, D) = (1-\alpha-\beta) \sum_{x \in \Omega} \sum_{i=1}^n \left( S_i - S_0 \exp \left( -\frac{b g_i D g_i}{2} \right) \right)^2 + \alpha \sum_{x \in \Omega} \sum_{y \in \Omega} w_1(x, y) (S_0(x) - S_0(y))^2 + \beta \sum_{x \in \Omega} \sum_{y \in \Omega} w_2(x, y) \delta(D, D(y)).
$$

Since most of the time, DTI estimation problems are in the discrete case, we will focus on the discrete case in this work.

**Computation of the weights $w_1(x, y)$ and $w_2(x, y)$**

$w_1(x, y)$ and $w_2(x, y)$ are the regularization weights of the smoothness terms. If $w_1(x, y)$ is large, it requires $S_0$ and $S_0(y)$ to be similar. Similarly, if $w_2(x, y)$ is large, it requires $D$ and $D(y)$ to be similar. Usually, one requires $S_0$’s and $D$’s to be respectively similar only if the corresponding diffusion signals are similar. We will compute $w_1(x, y)$ and $w_2(x, y)$ according to the statistical properties of the diffusion weighted signals. It has been recognized that the diffusion signal $S$ follows the Rician distribution (Descoteaux et al., 2008; Koay and Basser, 2006; Piurica et al., 2003), i.e.,

$$
p(S, S^*, \sigma^2) = \frac{S}{\sigma^2} \exp \left( -\frac{S^2 + S^2}{2\sigma^2} \right) I_0 \left( \frac{SS^*}{\sigma^2} \right),
$$

where $S$ is the noisy image, the larger will $\alpha$ and $\beta$ be, and vice-versa. The noise in the images can be estimated by using any of the popular methods described in Aja-Fernandez et al. (2008) and Tristan-Vega and Aja-Fernandez (2010).
\[
\begin{align*}
\text{where } S & \text{ is the signal without noise, } \sigma \text{ is the variance of the Rician noise. Since } S_0 \text{ is linearly related with } S \text{ and } D \text{ is "logarithmically" related with } S, \text{ we set the regularization weights for } S_0 \text{ and } D \text{ to be}
\end{align*}
\]

\[
\begin{align*}
w_1(x, y) & = \frac{1}{Z_1(x)} \exp \left( -\frac{||S(N(x)) - S(N(y))||^2}{h \sigma^2} \right), \\
w_2(x, y) & = \frac{1}{Z_2(x)} \exp \left( -\frac{|| \log S(N(x)) - \log S(N(y)) ||^2}{h \sigma^2} \right),
\end{align*}
\]

where \( Z_1 \) and \( Z_2 \) are normalizers, \( h \) is the parameter filtering (Coupe et al., 2006), and \( \sigma \) is the standard deviation of the noise, which is estimated by using the first moment of the background (Aja-Fernandez et al., 2009; Tristan-Vega and Aja-Fernandez, 2010). \( N(x) \) and \( N(y) \) denote the neighborhoods of \( x \) and \( y \) respectively. The neighborhood of \( x \) can be viewed as the voxels around \( x \) or a square centered at \( x \) with a user defined radius. Furthermore,

\[
\begin{align*}
||S(N(x)) - S(N(y))||^2 & = \sum_j ||S(\mu_j) - S(\nu_j)||^2, \quad \text{and} \\
|| \log S(N(x)) - \log S(N(y)) ||^2 & = \sum_j || \log S(\mu_j) - \log S(\nu_j) ||^2.
\end{align*}
\]

where \( \mu_j \) and \( \nu_j \) are the \( j \)th voxels in the neighborhoods \( N(x) \) and \( N(y) \) respectively, and \( m \) is the number of voxels in each neighborhood.

From Eq. (5), we can see that if the signal intensities for two voxels are similar, \( w_1(x, y) \) and \( w_2(x, y) \) are large. Consequently according to Eq. (3), \( S_0(x) \) and \( S_0(y) \), \( D \) and \( D(y) \) should be similar respectively.

NLM is known for its high accuracy and high computational complexity. To address the computational load problem, we use two methods. One is to decrease the number of computations performed by selecting voxels in the search window, and the other is to make use of parallel computing. Concretely, we will prefilter the voxels in the search window which are not similar to the voxel under consideration if their diffusion weighted signal intensities are not similar. This is specified as

\[
\begin{align*}
w_1(x, y) & = \begin{cases} 
1 & \frac{1}{Z_1(x)} \exp \left( -\frac{||S(N(x)) - S(N(y))||^2}{h \sigma^2} \right), \\
0 & \text{if } ||S(N(x))||^2 \in [\tau_1, \tau_2], \\
0 & \text{otherwise}, \\
\end{cases} \\
w_2(x, y) & = \begin{cases} 
1 & \frac{1}{Z_2(x)} \exp \left( -\frac{|| \log S(N(x)) - \log S(N(y)) ||^2}{h \sigma^2} \right), \\
0 & \text{if } || \log S(N(x)) ||^2 \in [\tau_1, \tau_2], \\
0 & \text{otherwise}, \\
\end{cases}
\end{align*}
\]

\( \tau_1 \) and \( \tau_2 \) are the thresholds for prefiltering. We set \( \tau_1 = 0.1 \) and \( \tau_2 = 10 \) in our experiments.

In the context of parallel computing, we divide the computations into smaller parts and assign the computations to several processors. Since the smaller parts for NLM are not correlated, thus it can improve the efficiency a lot by using parallel computing. In our case, we divide the volumes into 8 subvolumes, and assign each subvolume to one processor, and a desktop with 8 processors is used. This multi-threading technique greatly enhances the efficiency.

**Computation of the tKL divergence**

\( \text{tKL} \) divergence is a special case of the recently proposed total Breman divergence (tBD) (Liu et al., 2010; Vemuri et al., 2011). This divergence measure is based on the orthogonal distance between the convex generating function of the divergence and its tangent approximation at the second argument of the divergence. The total Breman divergence \( \delta_t \) associated with a real valued strictly convex and differentiable function \( f \) defined on a convex set \( X \) between points \( x, y \in X \) is defined as,

\[
\delta_t(x, y) = \frac{f(x) - f(y) - \langle x - y, \nabla f(y) \rangle}{\sqrt{1 + \langle \nabla f(y) \rangle^2}},
\]

where \( \langle \cdot, \cdot \rangle \) is the inner product, and \( ||\nabla f(y)||^2 = \langle \nabla f(y), \nabla f(y) \rangle \) generally. tBD has been proven to have the property of being intrinsically robust to noise and outliers. Furthermore, it yields a closed form formula for computing the median (an \( l_1 \)-norm average) for a set of symmetric positive definite tensors. When \( f(x) = -\log x \) and \( X \) is the set of probability density functions (pdf), Eq. (7) becomes the tKL divergence, which is

\[
\delta_t(x, y) = \frac{\int x \log x - y (1 + \log y)^2}{\sqrt{1 + y (1 + \log y)^2}}.
\]

Motivated by an earlier use of the tKL divergence as a dissimilarity measure between DTs for DTI segmentation (Vemuri et al., 2011), we use tKL to measure the dissimilarity between tensors and apply it in the DTI regularization. It has been shown that the tKL divergence (Vemuri et al., 2011) based \( l_1 \)-norm average, termed by the t-center, is invariant to special linear group transformations (denoted by SL(n)). \(^{23}\) This is detailed in the following.

Since order-2 SPD tensors can be seen as covariance matrices of zero mean Gaussian pdf (Wang et al., 2004). Let \( P, Q \in SPD(l) \), then their corresponding pdf are

\[
p(t, P) = \frac{1}{(2\pi)^{ldetP}} \exp \left( -\frac{t^T P^{-1} t}{2} \right),
\]

\[
q(t, Q) = \frac{1}{(2\pi)^{ldetQ}} \exp \left( -\frac{t^T Q^{-1} t}{2} \right),
\]

and the tKL between them is explicitly given by,

\[
\delta_t(P, Q) = \frac{\int p \log q dt}{\sqrt{1 + \int (1 + \log q)^2 q dt}} = \log (\det(P^{-1} Q)) + tr (Q^{-1} P) - l
\]

where \( c_1 = \frac{1}{2} + \frac{\log 2 \pi}{ldetP} + \frac{\log 2 \pi l}{ldetQ} \) and \( c_2 = \frac{l (1 + \log 2 \pi)}{ldetQ} \).

Moreover, the partial minimization of the third term in Eq. (3)

\[
\min_{D} \sum_{y \in V(x)} \delta_t(D, D(y))
\]

leads to the t-center for the set of \( D(y) \). The t-center has been well studied in Vemuri et al. (2011). Given a set of tensors \( \{Q_i\} \), the t-center \( P^* \) minimizes the \( l_1 \)-norm divergence to all the tensors, i.e.,

\[
P^* = \arg \min_{P} \sum_t \delta_t(P, Q_i).
\]

\(^{23}\) An \( n \times n \) matrix \( A \in SL(n) \) implies \( \det(A) = 1 \).
and \( \mathbf{P}^* \) is explicitly expressed as

\[
\mathbf{P}^* = \left( \sum_{i} a_i \mathbf{Q}_i^{-1} \right)^{-1},
\]

\[
a_i = \left( 2 c_1 + \frac{(\log(\det \mathbf{Q}_i))}{4} - c_2 \log(\det \mathbf{Q}_i) \right)^{-1}.
\]

The \( t \)-center for a set of DTs is the weighted harmonic mean, which is in closed form. Moreover, the weight is invariant to \( \text{SL}(n) \) transformations, i.e., \( a_i(\mathbf{Q}) = a_i(\mathbf{A}^T \mathbf{Q} \mathbf{A}), \forall \mathbf{A} \in \text{SL}(n) \). The \( t \)-center after the transformation becomes

\[
\mathbf{P}^* = \left( \sum_{i} a_i (\mathbf{A}^T \mathbf{Q}_i \mathbf{A})^{-1} \right)^{-1} = \mathbf{A}^T \mathbf{P}^* \mathbf{A}.
\]

This means that if \( \{ \mathbf{Q}_i \}_{i=1}^n \) are transformed by some member of \( \text{SL}(n) \), the \( t \)-center will undergo the same transformation. It was also found that the \( t \)-center will be robust to noise in that the weight will be smaller if the tensor has more noise (Vemuri et al., 2011). These properties make it an appropriate tool for the DTI applications.

**The SPD constraint**

It is known that if a matrix \( \mathbf{D} \in \text{ SPD} \), there exists a unique lower diagonal matrix \( \mathbf{L} \) with its diagonal values all positive, and \( \mathbf{D} = \mathbf{L} \mathbf{L}^T \) (Golub and Loan, 1996). This is the well known Cholesky factorization theorem. Wang et al. (2003, 2004) were the first to use Cholesky factorization to enforce the positive definiteness condition on the estimated smooth diffusion tensors from the DWI data. They use the argument that testing for positive definiteness is equivalent to testing for positive semidefiniteness under finite precision arithmetic and hence their cost function minimization is set on the space of positive semidefinite matrices, which is a closed set that facilitates the existence of a solution within that space. Unlike (Wang et al. 2003, 2004), we use Cholesky decomposition and tKL divergence to regularize the smoothness of the tensor field, and this automatically ensures the diagonal values of \( \mathbf{L} \) to be positive. This argument is validated as follows.

Substituting \( \mathbf{D} = \mathbf{L} \mathbf{L}^T \) into Eq. (9), we get

\[
\begin{align*}
\delta (t) &= \sum_{i=1}^{n} (\log L_{ii}(\mathbf{y}) - \log L_{ii}) + \operatorname{tr}( \mathbf{L}^{-T}(\mathbf{y}) \mathbf{L}^{-1}(\mathbf{y}) \mathbf{L}^T ) - 1.5 \\
&= \sqrt{c_1 + \frac{\sum_{i=1}^{n} \log[L_{ii}(\mathbf{y})]}{4} - c_2 \sum_{i=1}^{n} \log[L_{ii}(\mathbf{y})]}. 
\end{align*}
\]

Because by using the “log” function in the computation, Eq. (12) automatically requires \( \mathbf{L}_{ii} \) to be positive, therefore we do not need to manually force the tensor to be SPD. The detailed explanation is given in Appendix A.

**Numerical solution**

In this section, we present the numerical solution to the variation principle (3). The partial derivative equations of Eq. (3) with respect to \( \mathbf{S}_0 \) and \( \mathbf{L} \) can be computed explicitly and are:

\[
\frac{\partial \mathbf{E}}{\partial \mathbf{S}_0} = -2(1-\alpha-\beta) \sum_{\mathbf{y} \in \Omega(\mathbf{x})} \mathbf{w}_i - 2\alpha \sum_{\mathbf{y} \in \Omega(\mathbf{x})} \mathbf{w}_i (\mathbf{S}_0 - \mathbf{S}_0(\mathbf{y})),
\]

\[
\frac{\partial \mathbf{E}}{\partial \mathbf{L}} = 4(1-\alpha-\beta) \sum_{\mathbf{y} \in \Omega(\mathbf{x})} \mathbf{b}_0 \mathbf{y}^2 \mathbf{g}_i \mathbf{g}_i^T \mathbf{L}^{-1} \mathbf{L}^{-1} (\mathbf{y}) \mathbf{L}^{-1} (\mathbf{y}),
\]

\[
-2\beta \sum_{\mathbf{y} \in \Omega(\mathbf{x})} \mathbf{w}_i (\mathbf{y}) (\mathbf{L}^{-1} - \mathbf{L}^{-T} (\mathbf{y}) \mathbf{L}^{-1} (\mathbf{y})) \mathbf{L}^{-1} (\mathbf{y}) \mathbf{L}^{-1} (\mathbf{y}) \mathbf{L}^{-T} (\mathbf{y}) \mathbf{L}^{-1} (\mathbf{y}) \mathbf{L}^{-1} (\mathbf{y}),
\]

\[
\sqrt{c_1 + \frac{\sum_{i=1}^{n} \log[L_{ii}(\mathbf{y})]}{4} - c_2 \sum_{i=1}^{n} \log[L_{ii}(\mathbf{y})]}.
\]

To solve Eq. (13), we use the limited memory quasi-Newton method described in Nocedal and Wright (2000). This method is useful for solving large problems with a lot of variables, as is in our case. This method maintains simple and compact approximations of Hessian matrices making them require, as the name suggests, modest...
storage, besides yielding a linear rate of convergence. Specifically, we use the linear Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method (Nocedal and Wright, 2000) to construct the Hessian approximation.

Experiments

We evaluated our method on synthetic datasets with various levels of noise, and real datasets, including rat spinal cord datasets and human brain datasets. Based on the estimated tensor fields by using the technique presented in this paper, we achieved DTI segmentation for the rat spinal cord datasets, and some preliminary fiber tracking on human brain datasets. However, since the main thrust of this paper is the estimation of smooth diffusion tensor fields, the segmentation and fiber tracking results are not presented here.

We compared our method with other state-of-the-art techniques including VF (Tschumperle and Deriche, 2003), LMMSE (Tristan-Vega and Aja-Fernandez, 2010) and NLM, we used an existing code for DWI denoising and used our own implementation for LMMSE, we used the implementation in 3DSlicer 3.6. For VF by ourselves since we did not use the linear Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method for every experiment, and chose the set of parameters yielding the best results. The visual and numerical results show that our method yields better results than the competing methods.

DTI estimation from synthetic datasets

There are two groups of synthetic datasets. The first one is a 16 × 16 DTI with two homogeneous regions as shown in Fig. 1(a). Each region is a repetition of a tensor, and the two tensors are D1 := [3.3, 1.8, 1.3, 3.0, 0.1, 1.2] and D2 := [3, 2.2, 3, −1, 0, 0]. To generate the DWI based on this DTI, we let S0 = 5, b = 1500s/mm², and g be 22 uniformly-spaced directions on the unit sphere starting from (1,0,0). Substituting the DTI, S0, g into the Stejskal–Tanner equation, we generate a 16 × 16 × 22 DWI S. One representative slice of S is shown in Fig. 1(b). Then following the method proposed in Kooi and Basser (2006), we add the Rician noise to S and get Š, by using the formula Š(x) = \sqrt{(S(x) + n(x))² + n²}, where n₁, n₂ ∼ N(0,σ²). By varying σ, we get different levels of noise and therefore a wide range of SNR (SNR = \frac{n}{max(S)}).

In our experiments, we set α = 0.1 and β = 0.4. The search window size is set to be 25 and the neighborhood size is 9. Fig. 1(c) shows the slice in Fig. 1(b) after adding noise (SNR = 60). The estimated DTI from using MRE, VF, NLM, LMMSE and the proposed method are shown in Fig. 1. The figure visually depicts that our method can estimate the tensor field more accurately.

To quantitatively assess the proposed variational unified model, we determine the accuracy of the computed principle eigenvectors of the tensors. Let c₀ be the average angle between the principle eigenvectors of the estimated tensor field and the original known tensor field. Besides we compare the difference, denoted as cS₀, between the estimated and ground truth S₀. The results are shown in Table 1, from which it is evident that our method outperforms others and the significance in performance is more evident at higher noise levels. Even though the accuracy of NLM and our proposed method is very similar at high SNR, however, our method is much more computationally efficient than NLM. The average CPU time taken to converge for our method on a desktop computer with an Intel Core 2.8 GHz, 24 GB of memory, GNU Linux and MATLAB (Version 2010a) is 3.51 s, whereas, NLM requires 5.26 s (note both methods are executed by using multi-core processors).

We also evaluated the importance of the two regularization terms separately. α = 0 means removing the regularization on S₀, while β = 0 means removing the regularization on D. For these two cases, we evaluate the c₀ and cS₀, and the results are shown in the last two columns of Table 1. The results show that removing either the regularization term will increase the DTI estimation error. This implies that both regularization terms are necessary in order to get accurate estimation results.

We also evaluated our method on the 64 × 64 fiber cup dataset (Fillard et al., 2011; Poupon et al., 2008a) with a voxel size of 3 × 3 × 3 mm³, a b-value of 1500s/mm² and 130 gradient directions. For the parameter settings in the proposed method, we chose α = 0.1 and β = 0.4. The search window size is 9 and the neighborhood size is 4. We showed the estimated S₀, D11, D12, D13, D22, D23, D33, FA and the visualization of the estimated DTI by using fanDTasias (Barmouptsis et al., 2009a) in Fig. 2. The results depict that the proposed method can give a well smoothed and feature preserved tensor field.

DTI estimation from real datasets

We also did DTI estimation on a 100 × 80 × 32 × 52 3D rat brain DWI. The data was acquired by using a PGSE technique with TR = 1.5 s, TE = 28.3 ms, bandwidth = 35Khz, 52 diffusion weighted images with a b-value of 1334 s/mm².

We compared with several other methods on the DTI estimation, however, to save space, we only show the results of MRE, LMMSE and our proposed method. We present D11, D22, D33, FA, and mean trace for each estimated result. The DTI estimation results of MRE, LMMSE and our proposed method are shown in Figs. 3, 4 and 5 respectively.

We used a human brain DWI dataset (256 × 256 × 72) provided by Alfred Anwander of the Max Planck Institute for Human Neuroscience (Makuuchi et al., 2009). The DWIs were acquired with a whole-body 3 T Magnetom TRIO operating at 3 T (Siemens Medical Solutions).
equipped with an 8-channel head array coil. The twice-refocused
spin-echo EPI sequence (TR = 12 s, TE = 100 ms) consists of 22 diffu-
sion gradients with a $b$-value of 800 s/mm$^2$.

Fig. 2. From left to right, top to bottom are the estimated $S_0$, $D_{11}$, $D_{12}$, $D_{13}$, $D_{22}$, $D_{23}$, $D_{33}$, FA and the visualization of the estimated DTI by using fanDTasia.

Fig. 3. From left to right, top to bottom are $D_{11}$, $D_{22}$, $D_{33}$, $S_0$, FA, and mean trace of the estimated tensor field by using MRE on the rat cord data set.

Fig. 4. From left to right, top to bottom are $D_{11}$, $D_{22}$, $D_{33}$, $S_0$, FA, and mean trace of the estimated tensor field by using LMMSE on the rat cord data set.
For the parameters of our model, we chose $\alpha = 0.15$ and $\beta = 0.45$. The search window size is 64 and the neighborhood size is 27. We compared with several other methods on the DTI estimation, however, to save space, we only show the results of MRE, LMMSE and our proposed method. We present $D_{11}$, $D_{22}$, $D_{33}$, $S_0$, FA, and mean trace for each estimated result. The DTI estimation results of MRE, LMMSE and our proposed method are shown in Figs. 6, 7 and 8 respectively. The comparisons indicate that the proposed DTI estimation method generates better results.

Conclusions

We proposed a robust variational non-local means based unified approach for simultaneous denoising and DTI estimation. The proposed method is a combination of a variational framework, non-local means and an intrinsically robust divergence measure to regularize the DTI estimation. In the variational principle, we used non-linear diffusion tensor fitting term, along with a combination of non-local means and the tKL based smoothness measure for denoising. To speed up the NLM method, we performed prefiltering on the voxels in the search window to reduce the number of computations and made use of parallel computing to distribute the computational load. This variational non-local approach was validated with both synthetic and real datasets and was shown to be more accurate than competing methods in the literature. The results show that our method depicts better noise removal while preserving the structure information even at high levels of noise.

For future work, we plan to develop a GPU-based implementation to further reduce the computation time. After getting a more comprehensive tensor estimation technique, we will utilize it as a preprocessing step in applications to fiber tracking and DTI segmentation.

Acknowledgments

This work was in part supported by the NIH grant NS066340 to Vemuri, the University of Florida Allumni Fellowship and the INRIA Internships program to Liu, the Association France Parkinson and the ANR NucleiPark project (French National Research Agency ANR, Neurodegenerative and Psychiatric Diseases program) to Deriche.

We thank Dr. Alfred Anwander of the Max Planck Institute for providing the human brain datasets and Dr. John Forder of the Department of Radiology, University of Florida, McKnight Brain Institute for providing the rat spinal cord datasets.

Appendix A. Automatically ensuring the positivity of $\text{diag}(L)$

Since

$$L = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix},$$

and

$$D = LL^T = \begin{pmatrix} L_{21} & L_{11}L_{21} & L_{11}L_{31} \\ L_{21} & L_{21} + L_{22} & L_{21}L_{31} + L_{21}L_{32} \\ L_{31} & L_{31}L_{22} & L_{31}L_{32} + L_{32}^2 + L_{33}^2 \end{pmatrix},$$

Fig. 5. From left to right, top to bottom are $D_{11}$, $D_{22}$, $D_{33}$, $S_0$, FA, and mean trace of the estimated tensor field by using our proposed method on the rat cord data set.

Fig. 6. From left to right, top to bottom are $D_{11}$, $D_{22}$, $D_{33}$, $S_0$, FA, and mean trace of the estimated tensor field by using MRE on the human brain dataset.

Appendix A. Automatically ensuring the positivity of $\text{diag}(L)$
Let \( L_0 = \exp(a_i), i = 1, 2, 3 \), then by solving \( a_i \) we can ensure the \( L_0 \) is positive. Therefore, the positiveness of the diagonal values of \( L \) is transferred to solve \( a_i \).

Now \( L \) is converted to \( \hat{L} \):

\[
\hat{L} = \left( \begin{array}{ccc}
\exp(a_1) & 0 & 0 \\
L_{21} & \exp(a_2) & 0 \\
L_{31} & L_{32} & \exp(a_3)
\end{array} \right).
\]

\[
\frac{\partial E}{\partial L} = 4(1-\alpha-\beta) \sum_{i=1}^{n} b S_i y_i L_i^T S_i g_i^T
\]

\[
-2\beta \sum_{y = (x, y)} w_2(x, y) \frac{(L_i^{-1} - L_i^{-1}(y)L_i^{-1}(y))}{\sqrt{c_1 + \sum_{i=1}^{3} \log L_{ii}(y)} + c_2 \sum_{i=1}^{3} \log L_{ii}(y)}
\]

\[
\frac{\partial E}{\partial \hat{L}} = \frac{\partial E}{\partial L} \cdot J
\]

(A.1)

where “\( \cdot \)” is element-wise product, and

\[
J = \left( \begin{array}{ccc}
\exp(a_1) & 0 & 0 \\
1 & \exp(a_2) & 0 \\
1 & 1 & \exp(a_3)
\end{array} \right).
\]

Therefore, the partial derivative is in closed form. This can be solved by using the L-BFGS method.

**References**


