A Constrained Variational Principle for Direct Estimation and Smoothing of the Diffusion Tensor Field From Complex DWI

Zhizhou Wang, *Student Member, IEEE*, Baba C. Vemuri*, *Fellow, IEEE*, Yunmei Chen, and Thomas H. Mareci

***Abstract*—In this paper, we present a novel constrained variational principle for simultaneous smoothing and estimation of the diffusion tensor field from complex valued diffusion-weighted images (DWI). The constrained variational principle involves the minimization of a regularization term of L^p norms, subject to a nonlinear inequality constraint on the data. The data term we employ is the original Stejskal-Tanner equation instead of the linearized version usually employed in literature. The complex valued nonlinear form leads to a more accurate (when compared to the linearized version) estimate of the tensor field. The inequality constraint requires that the nonlinear least squares data term be bounded from above by a known tolerance factor. Finally, in order to accommodate the positive definite constraint on the diffusion tensor, it is expressed in terms of Cholesky factors and estimated. The constrained variational principle is solved using the augmented Lagrangian technique in conjunction with the limited memory quasi-Newton method. Experiments with complex-valued synthetic and real data are shown to depict the performance of our tensor field estimation and smoothing algorithm.**

***Index Terms*—Constrained variational principle, diffusion tensor MRI, image smoothing.**

I. Introduction

DIFFUSION is a process of intermingling molecules as a result of random thermal agitation and in our context, refers specifically to the random translational motion of water molecules in the part of the anatomy being imaged with MR. In three-dimensional (3-D) space, diffusivity can be described by a 3×3 matrix $\mathbf{D}$ called diffusion tensor which is intimately related to the geometry and organization of the microscopic environment.

In their seminal work [2], Basser *et al.*, introduced diffusion tensor MRI (DT-MRI) as a new MR image modality from which anisotropic water diffusion can be inferred quantitatively. Since then, DT-MRI became a powerful method to study the tissue microstructure e.g., white matter connectivity in the brain *in vivo*. Diffusion-weighted echo intensity image (DWI) S_l and the diffusion tensor $\mathbf{D}$ are related through the Stejskal-Tanner equation as given by

$$S_l = S_0 e^{-\mathbf{b}_l:\mathbf{D}} = S_0 e^{-\sum_{i=1}^{3}\sum_{j=1}^{3} b_{l,ij} D_{ij}} \tag{1}$$

where $\mathbf{b}_l$ is the diffusion weighting matrix of the l-th magnetic gradient, ":" denotes the generalized inner product for matrices. The popular phase-encoding method for acquiring DT-MRI images will yield complex measurements, thus, S_l and S_0 will all be complex variables and (1) still holds in such cases. In the following, we assume $S_l = R_l + iI_l$ and $S_0 = R_0 + iI_0$ are complex variables, where $R_l = \mathrm{real}(S_l), I_l = \mathrm{imaginary}(S_l), R_0 = \mathrm{real}(S_0)$, and $I_0 = \mathrm{imaginary}(S_0)$. Take the magnitude of both sides in (1), we have

$$||S_l|| = ||S_0|| e^{-\mathbf{b}_l:\mathbf{D}} = ||S_0|| e^{-\sum_{i=1}^{3}\sum_{j=1}^{3} b_{l,ij} D_{ij}}. \tag{2}$$

Taking log on both sides of (2) yields the following linear Stejskal-Tanner equation:

$$\log(||S_l||) = \log(||S_0||) - \sum_{i=1}^{3}\sum_{j=1}^{3} b_{l,ij} D_{ij}. \tag{3}$$

Note that in the past, what have been considered are only the magnitude of the complex measurements and in particular the linearized equation (3), [21], [27].

Given several (at least seven) noncollinear diffusion weighted intensity measurements, $\mathbf{D}$ can be estimated via multivariate regression models from any of the above three equations. Diffusion anisotropy can then be computed to show microstructural and physiological features of tissues [3]. Especially in highly organized nerve tissue, like white matter, the diffusion tensor provides a complete characterization of the restricted motion of water through the tissue that can be used to infer fiber tracts. The development of diffusion tensor acquisition, processing, and analysis methods provides the framework for creating fiber tract maps based on this complete diffusion tensor analysis [8], [11].

For automatic fiber tract mapping, the diffusion tensor field (DTI) must be smoothed without losing relevant features. Currently there are two popular approaches, one involves smoothing the raw data $||S_l||$ while preserving relevant detail and then estimating the diffusion tensor $\mathbf{D}$ from the smoothed raw data (see [15] and [22]). The raw data in this context consists of several diffusion weighted images (DWIs) acquired for varying magnetic gradient strengths and directions. Note that at least seven values at each 3-D grid point in the data domain are required

Manuscript received September 27, 2003; revised April 29, 2004. This work was supported in part by the National Institutes of Health (NIH) under Grant NIH RO1 NS42075 and Grant NIH P41 RR16105. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was S. Pizer. *Asterisk indicates corresponding author.*

Z. Wang is with the Department of Computer Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA.

*B. C. Vemuri is with the Department of Computer Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: vemuri@cise.ufl.edu).

Y. Chen is with the Department of Mathematics, University of Florida, Gainesville, FL 32611 USA.

T. H. Mareci is with the Department of Biochemistry and Department of Physics, University of Florida, Gainesville, FL 32611 USA.

Digital Object Identifier 10.1109/TMI.2004.831218

to estimate the six unknowns in the 3×3 symmetric tensor $\mathbf{D}$ and one scale parameter $||S_0||$. The raw data smoothing or de-noising can be formulated using variational principles which in turn requires solution to partial differential equations (PDEs) ([1] and others in [5]) or at times directly using PDEs which are not necessarily arrived at from variational principles ([26] and others in [5]).

Another approach to restore the DTI is to smooth the principal diffusion direction after the diffusion tensor has been estimated from the raw noisy measurements $||S_l||$. In Poupon *et al.* [16], an energy function based on a Markovian model was used to regularize the noisy dominant eigenvector (DEV) field computed directly from the noisy estimates of $\mathbf{D}$ obtained from the measurements $||S_l||$ using the linearized Stejskal-Tanner equation (3). Coulon *et al.*, [9] proposed an iterative restoration scheme for principal diffusion direction based on direction map restoration work reported in [19]. Other sophisticated vector field restoration methods [12], [20] can potentially be applied to the problem of restoring the DEV fields computed from the noisy estimates of $\mathbf{D}$. Recently, Chefd'Hotel *et al.*, [7] presented an elegant geometric solution to the problem of smoothing a noisy $\mathbf{D}$ that was computed from $||S_l||$ using the log-linearized model (3) described above. They assume that the given (computed) tensor field $\mathbf{D}$ from $||S_l||$ is positive definite and develop a clever approach based on differential geometry of manifolds to achieve constrained smoothing where the smoothed tensor field is constrained to be positive semi-definite. Interesting results of mapped fibers are shown for human brain MRI.

The idea of simultaneous estimation and smoothing of the diffusion tensors from DWI was pioneered by Wang *et al.*, [25] and improved upon by Wang *et al.*, in [24]. This improvement involved methods to overcome the problem of manual choice of regularization control parameters. In both these works [25], [24], the estimated smooth tensor field was guaranteed to be positive semi definite. Moreover, these works were a report of the first use of the nonlinear Stejskal-Tanner equation in the estimation of the diffusion tensors. Recently, in Tschumperlé *et al.*, [21], a robust version of the linearized Stejskal-Tanner equation is used as the data term along with a robust regularization term in a unified variational principle to estimate a smooth $\mathbf{D}$ from the noisy signal measurements. Note that the data term uses a linearized version of the Stejskal-Tanner equation as in earlier works [9], [7], [16], [22]. In this paper, we further extend our model in [24] to the restoration of the DTI from complex valued DWIs. Specifically, we propose a novel formulation of the DTI estimation and smoothing as a constrained optimization problem. The specific approach we use is called the augmented Lagrangian technique which allows one to deal with inequality constraints. *The novelty of our formulation lies in the ability to directly, in a single step, estimate a smooth* $\mathbf{D}$ *from the noisy complex measurements* S_l *with the preservation of its positiveness. The formulation does not require any adhoc methods of setting tuning parameters to achieve the solution. These are the key features distinguishing our solution method from methods reported in literature to date*.

In contrast to our solution (to be described subsequently in detail), most of the earlier approaches used a two step method involving 1) computation of a $\mathbf{D}$ from $||S_l||$ using a linear least-squares approach and then 2) computing a smoothed $\mathbf{D}$ via either smoothing of the eigenvalues and eigenvectors of $\mathbf{D}$ or using the matrix flows approach in [7]. The problem with the two step approach to computing $\mathbf{D}$ is that the estimated $\mathbf{D}$ in the first step using the log-linearized model need not be positive definite or even semi-definite. Moreover, it is hard to trust the fidelity of the eigenvalues and vectors computed from such matrices even if they are to be smoothed subsequently prior to mapping out the nerve fiber tracts.

Briefly, our model seeks to minimize a cost function involving, the sum of an L^p norm based gradient of the Cholesky factor $\mathbf{L}$ which ensure the positiveness of $\mathbf{D}$ by the Cholesky factorization LL^T—and an L^p norm based gradient of S_0, subject to a nonlinear data constraint based on the complex (not linearized) Stejskal-Tanner equation (1). The model is posed as a constrained variational principle which can be minimized by either discretizing the variational principle itself or the associated Euler-Lagrange equation. We choose the former and use the augmented Lagrangian method together with the limited memory quasi-Newton method to achieve the solution.

Rest of the paper is organized as follows: in Section II, the detailed variational formulation is described along with the nonlinear data constraints, the positive definite constraint and the augmented Lagrangian solution. Section III contains the detailed description of the discretization as well as the algorithmic description of the augmented Lagrangian framework. In Section IV, we present experiments on application of our model to synthetic as well as real data. Synthetic data experiments are conducted to present comparison of DTI restoration results with a recently presented work of Coulon *et al.*, [9]. Moreover, results of comparison between the use of the linearized Stejskal-Tanner model and the nonlinear form of the same are presented as well.

II. Constrained Variational Principle Formulation

Our solution to the recovery of a piecewise smooth DTI from the complex measurements S_l is posed as a constrained variational principle. We seek to minimize a measure of lack of smoothness in S_0 and the diffusion tensor $\mathbf{D}$ being estimated using an L^p norm of the gradient in S_0 and an L^p norm of the gradient in the Cholesky factor $\mathbf{L}$. This measure is then constrained by a nonlinear data fidelity term related to the complex Stejskal-Tanner equation (1). The nonlinear data term is constrained by an inequality which requires that it be bounded from above by a possibly known tolerance factor. The positiveness constraint on the diffusion tensor being estimated is achieved via the use of the Cholesky factorization theorem from computational linear algebra. The constrained variational principle is discretized and posed using the augmented Lagrangian technique [14]. Each subproblem in the augmented Lagrangian framework is then solved using the limited memory quasi-Newton scheme [14]. The novelty of our formulation lies in the unified framework for recovering and smoothing of the DTI from the raw data S_l. In addition, to our knowledge, this is the first formulation which allows for simultaneous estimation and smoothing of $\mathbf{D}$ as well as one in which the regularization parameter is not set in an adhoc way.

Let $S_0(\mathbf{x}) = R_0(\mathbf{x}) + iI_0(\mathbf{x})$ be the complex DWI when no diffusion-encoding magnetic gradient is present, $\mathbf{D}(\mathbf{x})$ be the unknown symmetric positive definite diffusion tensor, $\mathbf{L}\mathbf{L}^T(\mathbf{x})$ be the Cholesky factorization of the diffusion tensor with $\mathbf{L}$ being a lower triangular matrix, $S_l(\mathbf{x}) = R_l(\mathbf{x}) + iI_l(\mathbf{x}), l = 1, \ldots, N$ are the complex DWIs measured after application of a diffusion-encoded magnetic gradient of known strength and direction and N is the total number of measured DWIs The constrained variational principle is

$$\min_{S_0,\mathbf{L}} \mathcal{E}(S_0, \mathbf{L}) = \int_\Omega [|\nabla R_0|^{p_1} + |\nabla I_0|^{p_1} + |\nabla \mathbf{L}|^{p_2}]\, d\mathbf{x}$$
$$\text{s.t. } \mathcal{C}(S_0, \mathbf{L}) = \alpha\sigma^2 - \int_\Omega \sum_{l=1}^{N} \left(FR_l^2 + FI_l^2\right) d\mathbf{x} \geq 0 \quad (4)$$

where Ω is the image domain, the first and the second terms in the variational principle are L^p smoothness constraints on the real and imaginary part of S_0, the third term is an L^p smoothness constraint on $\mathbf{L}$, where $p_1 > 6/5$ for R_0 and I_0 and $p_2 \geq 1$ for $\mathbf{L}$. $|\nabla \mathbf{L}|^{p_2} = \sum_d |\nabla L_d|^{p_2}$, where $d \in \mathcal{D} = \{xx, yy, zz, xy, yz, xz\}$ are indices to the six nonzero components of $\mathbf{L}$. The lower bounds on the value of p_1 and p_2 are chosen so as to make the proof of existence of a solution for this minimization (see Section II-D and the Appendix) mathematically tractable. α is a constant scale factor and σ^2 is the standard deviation of the noise in the measurements S_l. FR_l and FI_l are defined as

$$FR_l = R_l - R_0 e^{-\mathbf{b_l}:\mathbf{L}\mathbf{L}^\mathbf{T}}, \quad FI_l = I_l - I_0 e^{-\mathbf{b_l}:\mathbf{L}\mathbf{L}^\mathbf{T}}$$

A. The Complex Nonlinear Data Constraint

The Stejskal-Tanner equation (1) shows the relation between the complex DWI S_l and the diffusion tensor $\mathbf{D}$. However, multivariate linear regression based on (3) has been used to estimate the diffusion tensor $\mathbf{D}$ [2]. It was pointed out in [2] that these results agree with nonlinear regression based on the magnitude Stejskal-Tanner equation (2). However, if the signal-to-noise ratio (SNR) is low and the number of S_l is not very large (unlike in [2] where $N = 315$ or $N = 294$), the result from multivariate linear regression will differ from the nonlinear regression significantly. A robust estimator belonging to the M-estimator family was used by Poupon *et al.*, [16], however, its performance is not discussed in detail. In Westin *et al.*, [27]), an analytical solution is derived from (3) by using a dual tensor basis, however, it should be noted that this can only be used for computing the diffusion tensor $\mathbf{D}$ when there is no noise in the measurements S_l or the SNR is extremely high.

Our aim is to provide an accurate estimation of the diffusion tensor $\mathbf{D}$ for practical use, where the SNR may not be high and the total number of DWIs N is restricted to a moderate number. The nonlinear data fidelity term based on the complex Stejskal-Tanner equation (1) is fully justified for use in such situations. This nonlinear data term is part of an inequality constraint that imposes an upper bound on the closeness of the measurements S_l to the mathematical model $S_0 e^{-\mathbf{b_l}:\mathbf{L}\mathbf{L}^\mathbf{T}}$. The bound $\alpha\sigma^2$ may be estimated automatically [13], [17].

B. The L^p Smoothness Constraint

In Blomgren *et al.*, [4], it was shown that L^p smoothness constraint does not admit discontinuous solutions as the TV-norm does when $p > 1$. However, when p is chosen close to 1, its behavior is close to the TV-norm for restoring edges. In our constrained model, we need $p_1 > 6/5$ for regularizing S_0 and $p_2 \geq 1$ for $\mathbf{L}$ to ensure existence of the solution described in Section II-D. Note that what is of importance here is the estimation of the diffusion tensor $\mathbf{D}$ and therefore, the edge-preserving property in the estimation process is more relevant for the case of $\mathbf{D}$ than for S_0. In our experiment, we choose $p_1 = 1.205$ for S_0 and $p_2 = 1.00$ for $\mathbf{L}$.

C. The Positive Definite Constraint

In general, a matrix $\mathbf{A} \in \Re^{n\times n}$ is said to be positive definite if $\mathbf{x}^\mathbf{T}\mathbf{A}\mathbf{x} > 0$, for all $\mathbf{x} \neq \mathbf{0}$ in $\Re^n$. The diffusion tensor $\mathbf{D}$ happens to be a symmetric positive definite matrix but due to the noise in the data S_l, it is hard to recover a $\mathbf{D}$ that retains this property unless one includes it explicitly as a constraint. One way to impose this constraint is using the Cholesky factorization theorem, which states that: *If* $\mathbf{A}$ *is a symmetric positive definite matrix then, there exists a unique factorization* $\mathbf{A} = \mathbf{L}\mathbf{L}^\mathbf{T}$ *where,* $\mathbf{L}$ *is a lower triangular matrix with positive diagonal elements*. After doing the Cholesky factorization, we have transfered the positive definiteness constraint on the matrix $\mathbf{D}$ to an inequality constraint on the diagonal elements of $\mathbf{L}$. This is, however, still hard to satisfy theoretically because, the set on which the minimization takes place is an open set. However, in practise, with finite precision arithmetic, testing for a positive definiteness constraint is equivalent to testing for positive semi-definiteness. This is because for any symmetric positive definite matrix $\mathbf{D}$, its machine representation $\tilde{\mathbf{D}} = \mathbf{D} + \mathbf{E}$ with $\|\mathbf{E}\| \leq \epsilon\|\mathbf{D}\|$, where ϵ is a small multiple of the machine precision. When $\mathbf{D}$ is a small symmetric positive definite matrix, $\tilde{D}$ can become a semi-definite matrix, it follows that in finite precision arithmetic, testing for definiteness is equivalent to testing for semi-definiteness. Thus, we repose the positive definiteness constraint on the diffusion tensor matrix as, $\mathbf{x}^\mathbf{T}\tilde{\mathbf{D}}\mathbf{x} \geq 0$ which is satisfied when $\tilde{\mathbf{D}} = \mathbf{L}\mathbf{L}^\mathbf{T}$.

D. Existence of a Solution

Justification for using the augmented Lagrangian method for constrained problems is given in [14], thus, we only need to prove there is a solution for the following subproblem:

$$\min_{(S_0,\mathbf{L})\in\mathcal{A}} \mathcal{L}(S_0, \mathbf{L}; \lambda, \mu) = \begin{cases} \mathcal{E}(S_0, \mathbf{L}) - \lambda\mathcal{C}(S_0, \mathbf{L}) + \frac{\mathcal{C}^2(S_0,\mathbf{L})}{2\mu}, & \text{if } \mathcal{C}(S_0, \mathbf{L}) \leq \mu\lambda \\ \mathcal{E}(S_0, \mathbf{L}) - \frac{\mu}{2}\lambda^2, & \text{otherwise} \end{cases} \quad (5)$$

where $\lambda \geq 0$ is an estimate of the Lagrange multiplier, $\mu > 0$ is a penalty parameter and $\mathcal{A} = \{(S_0, \mathbf{L}) \,|\, \mathbf{L} \in L^p(\Omega) \text{ where } p \geq 1 \text{ and } R_0, I_0 \in W^{1,p}(\Omega) \text{ where } p > 6/5\}$. Here $\Omega \subset \Re^3, L^p(\Omega)$ denotes the space of functions with bounded L^p norms, $L^2(\Omega)$ is the space of square integrable functions on Ω and $W^{1,p}(\Omega)$ denotes the Sobolev space of order p on Ω [10].

Theorem 1: Suppose $R_l, I_l \in L^2(\Omega)$, then the augmented Lagrangian formulation (5) has a solution $(S_0, \mathbf{L}) \in \mathcal{A}$.

Proof of this theorem is given in the Appendix.

Finding a solution of the constrained variation principle (4) involves solving a sequence of equations of the form in (5) with fixed λ and μ at each stage. It is much more difficult than when dealing with the problems of recovering and smoothing separately. However, there are benefits of posing the problem in this constrained unified framework, namely, one does not accumulate the errors from a two stage process. Moreover, this framework incorporates the nonlinear data term which is more appropriate for low SNR values prevalent when the magnitude of the diffusion-encoded magnetic gradient is high. Also, the noise model is correct for the nonlinear complex data model unlike the log-linearized case. Lastly, in the constrained formulation, it is now possible to pose mathematical questions of existence and uniqueness of the solution—which was not possible in earlier formulations reported in literature.

III. Numerical Methods

The minimization problem given by (4) can only be solved numerically. Here, we discretize the constrained variational principle (4), transform it into a sequence of unconstrained problems by using the augmented Lagrangian method and then employ the limited quasi-Newton technique [14] to solve them. Note that this framework allows us to solve the minimization without resorting to adhoc methods of choosing the "tuning" parameters. Also limited memory quasi-Newton is the method of choice here due to the advantages it affords in the context of memory/storage savings.

A. Discretized Constrained Variational Principle

We use the standard finite difference method to discretize the problem. Let

$$\begin{aligned}
FR_{l,ijk} &= R_{l,ijk} - R_{0,ijk} e^{-\mathbf{b}_l : \mathbf{L}_{ijk} \mathbf{L}_{ijk}^T} \\
FI_{l,ijk} &= I_{l,ijk} - I_{0,ijk} e^{-\mathbf{b}_l : \mathbf{L}_{ijk} \mathbf{L}_{ijk}^T} \\
F_{l,ijk} &= FR_{l,ijk} + iFI_{l,ijk} \\
|\nabla R_0|_{ijk} &= [\sqrt{(\Delta_x^+ R_0)^2 + (\Delta_y^+ R_0)^2 + (\Delta_z^+ R_0)^2 + \epsilon}]_{ijk} \\
|\nabla I_0|_{ijk} &= [\sqrt{(\Delta_x^+ I_0)^2 + (\Delta_y^+ I_0)^2 + (\Delta_z^+ I_0)^2 + \epsilon}]_{ijk} \\
|\nabla L_d|_{ijk} &= [\sqrt{(\Delta_x^+ L_d)^2 + (\Delta_y^+ L_d)^2 + (\Delta_z^+ L_d)^2 + \epsilon}]_{ijk} \\
|\nabla \mathbf{L}|_{ijk}^p &= \sum_{d \in \mathcal{D}} |\nabla L_d|_{ijk}^p
\end{aligned}$$

where Δ_x^+, Δ_y^+, and Δ_z^+ are forward difference operators, ϵ is a small positive number used to avoid singularities of the L^p norm when $p < 2$. Now the discretized constrained variational principle can be written as

$$\begin{aligned}
&\min_{S_0, \mathbf{L}} \mathcal{E}(S_0, \mathbf{L}) = \sum_{i,j,k} \left(|\nabla R_0|_{ijk}^{p_1} + |\nabla I_0|_{ijk}^{p_1} + |\nabla \mathbf{L}|_{ijk}^{p_2} \right) \\
&\text{s.t. } \mathcal{C}(S_0, \mathbf{L}) = \alpha \sigma^2 - \sum_{i,j,k} \sum_{l=1}^{N} ||F_{l,ijk}||^2 \geq 0.
\end{aligned} \tag{6}$$

B. Augmented Lagrangian Method

The above problem is now posed using the augmented Lagrangian method, where a sequence of related unconstrained subproblems are solved, and the limit of these solutions is the solution to (6). Following the description in [14], the kth subproblem of (6) is given by

$$\begin{aligned}
&\min \mathcal{L}(S_0, \mathbf{L}, s; \lambda_k, \mu_k) \\
&\quad = \mathcal{E}(S_0, \mathbf{L}) - \lambda_k [\mathcal{C}(S_0, \mathbf{L}) - s] \\
&\qquad + \frac{1}{2\mu_k} [\mathcal{C}(S_0, \mathbf{L}) - s]^2
\end{aligned} \tag{7}$$

where $s \geq 0$ is a slack variable, μ_k, λ_k are the barrier parameter and the Lagrange multiplier estimate for the k-th subproblem, respectively.

One can explicitly compute the slack variable s at the minimum as $s = \max(\mathcal{C}(S_0, \mathbf{L}) - \mu_k \lambda_k, 0)$ and substitute it in (7) to get an equivalent subproblem in $(S_0, \mathbf{L})$ given by

$$\begin{aligned}
&\min \mathcal{L}(S_0, \mathbf{L}; \lambda_k, \mu_k) \\
&= \begin{cases} \mathcal{E}(S_0, \mathbf{L}) - \lambda_k \mathcal{C}(S_0, \mathbf{L}) + \frac{\mathcal{C}^2(S_0, \mathbf{L})}{2\mu_k}, & \text{if } \mathcal{C}(S_0, \mathbf{L}) \leq \mu_k \lambda_k \\ \mathcal{E}(S_0, \mathbf{L}) - \frac{\mu_k}{2} \lambda_k^2, & \text{otherwise} \end{cases}
\end{aligned} \tag{8}$$

The following algorithm summarizes the procedure to find the solution for (6).

Algorithm 1 Augmented Lagrangian Algorithm
1) Initialize $S_0(0), \mathbf{L}(0)$ using the nonlinear regression, choose an initial μ_0 and λ_0.
2) **for** $k = 1, 2, \ldots$
3) Find an approximate minimizer $S_0(k), \mathbf{L}(k)$ of $\mathcal{L}(\cdot, \cdot; \lambda_k, \mu_k)$ as in (8) starting with $S_0(k-1), \mathbf{L}(k-1)$;
4) If final convergence test is satisfied
5) **STOP** with an approximate solution $S_0(k), \mathbf{L}(k)$;
6) Update the Lagrange multiplier using $\lambda_{k+1} = \max(\lambda_k - \mathcal{C}(S_0, \mathbf{L})/\mu_k, 0)$;
7) Choose a new penalty parameter $\mu_{k+1} = \mu_k/2$;
8) Set the new starting point for the next iteration to $S_0(k)$, $\mathbf{L}(k)$;
9) **end(for)**

C. Limited Memory Quasi-Newton Method

Due to the large number of unknown variables in the minimization, we solve the subproblem using limited memory quasi-Newton technique. Quasi-Newton-like methods compute the approximate Hessian matrix at each iteration of the optimization by using only the first derivative information. In the Limited-Memory Broyden–Fletcher–Goldfarb–Shano (BFGS) method, search direction is computed without storing the approximated Hessian matrix which can be a very large matrix in general ($O(N^6)$ size for $O(N^3)$ unknowns).

Let $\mathbf{x} = (S_0, \mathbf{L})$ be the vector of variables, and $f(\mathbf{x}) = \mathcal{L}(S_0, \mathbf{L}; \lambda, \mu)$ denote the augmented Lagrangian function (8) to be minimized. For simplicity, we write $f(\mathbf{x}) = \mathcal{L}(S_0, \mathbf{L})$ by omitting the fixed parameter λ and μ in the following description. At kth iteration, let $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ be the update of the variable vector $\mathbf{x}$, $\mathbf{y}_k = \nabla f_{k+1} - \nabla f_k$ the update of the gradient

and $\mathbf{H}_\mathbf{k}^{-1}$ the approximation of the inverse of the Hessian. The inverse of the approximate Hessian can be approximated using the BFGS update formula

$$\mathbf{H}_{k+1}^{-1} = \mathbf{V}_k \mathbf{H}_k^{-1} \mathbf{V}_k^T + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \tag{9}$$

where $\mathbf{V_k} = I - (\mathbf{y_k s_k^T})/(\mathbf{y_k^T s_k})$.

Then we can use the following L-BFGS two-loop recursion iterative procedure, which computes the search direction $\mathbf{H}_k^{-1}\nabla f_k$ efficiently by using last m pairs of $(\mathbf{s}_k, \mathbf{y}_k)$ [14].

Algorithm 2 Search Direction Update Algorithm

$\mathbf{q} \leftarrow \nabla f_k$;
2) **for** $i = k-1, k-2, \ldots, k-m$
$\alpha_i \leftarrow \rho_i \mathbf{s}_i^T \mathbf{q}$;
4) $\mathbf{q} \leftarrow \mathbf{q} - \alpha_i \mathbf{y}_i$;
end(for)
6) $\mathbf{r} \leftarrow (\mathbf{H}_k^0)^{-1}\mathbf{q}$;
for $i = k-m, k-m-1, \ldots, k-1$
8) $\beta \leftarrow \rho_i \mathbf{y}_i^T \mathbf{r}$;
$\mathbf{r} \leftarrow \mathbf{r} + \mathbf{s}_i(\alpha_i - \beta)$
10) **end(for)**
stop with result $\mathbf{H_k^{-1}}\nabla f_k = \mathbf{r}$.

where $\rho_k = (1)/(\mathbf{y}_k^T \mathbf{s}_k)$ and $(\mathbf{H}_k^0)^{-1}$ is the initial approximation of the inverse of the Hessian, we set $(\mathbf{H}_k^0)^{-1} = \gamma_k \mathbf{I}$ where $\gamma_k = (\mathbf{s}_{k-1}^T \mathbf{y}_{k-1})/(\mathbf{y}_{k-1}^T \mathbf{y}_{k-1})$.

The gradient of our energy function is

$$\begin{aligned}\nabla f(x) = \Bigg(&\frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial R_0}, \frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial I_0}, \frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial L_{xx}}, \\ &\frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial L_{yy}}, \frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial L_{zz}}, \frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial L_{xy}}, \\ &\frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial L_{yz}}, \frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial L_{xz}}\Bigg)\end{aligned} \tag{10}$$

Each term in the above equation can be easily computed analytically, for example

$$\frac{\partial \mathcal{L}(S_0, \mathbf{L})}{\partial L_{d,ijk}} = \begin{cases} \sum_{i'j'k'} \frac{\partial |\nabla L_d|_{i'j'k'}^p}{\partial L_{d,ijk}}, & \text{if } \mathcal{C}(S_0, \mathbf{L}) > \mu_k \lambda_k \\ \sum_{i'j'k'} \frac{\partial |\nabla L_d|_{i'j'k'}^p}{\partial L_{d,ijk}}, & \text{otherwise} \\ \quad +2\left(\lambda - \frac{\mathcal{C}(S_0, \mathbf{L})}{\mu}\right) & \\ \quad \times \sum_{l=1}^N \left(FR_{l,ijk} \frac{\partial FR_{l,ijk}}{\partial L_{d,ijk}} + FI_{l,ijk} \frac{\partial FI_{l,ijk}}{\partial L_{d,ijk}}\right) & \end{cases} \tag{11}$$

here $\sum_{i'j'k'}$ is computed over a neighborhood of the voxel (i, j, k) where the forward differences involves the variables $L_{d,ijk}$ and

$$\begin{aligned}\frac{\partial FR_{l,ijk}}{\partial L_{xx,ijk}} &= R_{0,ijk} e^{-\mathbf{b}_l : \mathbf{L}_{ijk}\mathbf{L}_{ijk}^T} \times \frac{\partial \mathbf{b}_l : \mathbf{L}_{ijk}\mathbf{L}_{ijk}^T}{\partial L_{xx,ijk}} \\ \frac{\partial FI_{l,ijk}}{\partial L_{xx,ijk}} &= I_{0,ijk} e^{-\mathbf{b}_l : \mathbf{L_{ijk}}\mathbf{L_{ijk}^T}} \times \frac{\partial \mathbf{b}_l : \mathbf{L}_{ijk}\mathbf{L}_{ijk}^T}{\partial L_{xx,ijk}} \\ \frac{\partial \mathbf{b}_l : \mathbf{L}_{ijk}\mathbf{L}_{ijk}^T}{\partial L_{xx,ijk}} &= (2b_{l,xx}L_{xx,ijk} \\ &\quad + 2b_{l,xy}L_{xy,ijk} + 2b_{l,xz}L_{xz,ijk}).\end{aligned}$$

D. Implementation Issues

The constraint in (6) is directly related to the standard deviation σ of the noise which can be computed as in [17]. Since there are N complex measurements and 8 unknown parameters at each voxel (Note: S_0 is complex-valued, so it is treated as two unknowns), we have

$$\sigma = \frac{\sum_{l=1}^N \left\| S_l(\mathbf{x}) - S_0(\mathbf{x}) e^{-\mathbf{b}_l : \mathbf{L}(\mathbf{x})\mathbf{L}(\mathbf{x})^T} \right\|^2}{2N - 8}.$$

Initialization is very crucial for nonlinear optimization. In our case, we use the following nonlinear regression with positive definiteness constraint as the initial guess:

$$\min \mathcal{E}(S_0(\mathbf{x}), \mathbf{L}(\mathbf{x})) = \sum_{l=1}^N \left\| S_l(\mathbf{x}) - S_0(\mathbf{x}) e^{-\mathbf{b}_l : \mathbf{L}(\mathbf{x})\mathbf{L}(\mathbf{x})^T} \right\|^2. \tag{12}$$

The above minimization is a simple nonlinear least square problem and can be efficiently solved by the Levenberg-Marquardt method [14] using the results of the corresponding linearized least square problem as an initial guess.

There are a few practical issues in implementing the augmented Lagrangian method and the quasi-Newton method, these are settled by using the suggestions in [14] or empirically. For example, in the augmented Lagrangian method (see algorithm 1), we start with a penalty control parameter $\mu = 5.0$, decrease it by a factor of 2 in each step untill it is less than 0.01. We also choose $\lambda_0 = 1.0$. Note that the augmented Lagrangian method is quite robust with respect to the choice of μ_0 and λ_0 since μ_0 will eventually decrease to 0 and λ_0 approaches the Lagrange multiplier. The final convergence test has two criteria: The subproblem converges and $\mu < 0.01$. As the subproblem is just a standard unconstrained minimization problem, the criteria to check whether it converges or not is achieved using any of the standard criteria in iterative optimization schemes [14] and for the line search, we employ cubic interpolation and Wolfe convergence criterion, see [14] for more details. For the limited memory quasi-Newton method, we use the last five update pairs to update the search direction.

IV. Experimental Results

In this section, we present three sets of experiments on the application of our direct tensor smoothing and estimation model. One is on complex valued synthetic data sets and the other two are on a complex valued DWI data acquired from a normal rat brain. For the synthetic data example, we compare the results obtained from our estimation procedure with competing methods published in literature.

A. Complex Valued Synthetic Data

We synthesized an anisotropic tensor field on a 3-D lattice of size $32 \times 32 \times 8$. The volume consists of two homogeneous regions with the following values for S_0 and $\mathbf{D}$:

$$\begin{aligned}\text{Region1}: S_0 &= 10.0e^{i\theta} \\ \mathbf{D} &= 0.001 \\ &\quad \times [0.970 \quad 1.751 \quad 0.842 \quad 0.0 \quad 0.0 \quad 0.0] \\ \text{Region2}: S_0 &= 8.0e^{i\theta}\end{aligned}$$

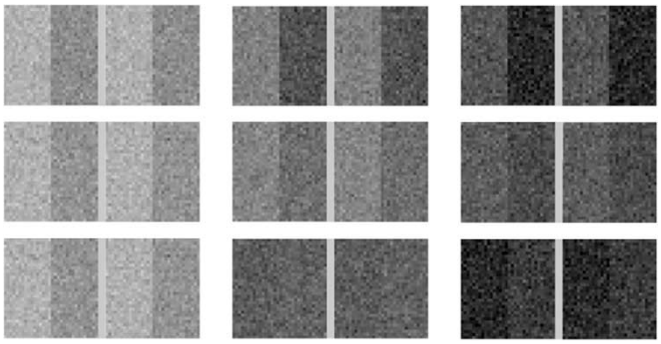

Fig. 1. A slice of several volumetric complex DWIs generated with $\sigma_N = 0.5$. Left to right: real and imaginary pairs of complex DWIs with varying magnitude of diffusion encoded magnetic gradient. Top to bottom: complex DWIs for varying directions of diffusion encoded magnetic gradient.

$$\mathbf{D} = 0.001 \times [1.556 \quad 1.165 \quad 0.842 \quad 0.338 \quad 0.0 \quad 0.0]$$

where the tensor $\mathbf{D}$ is depicted as

$$\mathbf{D} = [d_{xx}, d_{yy}, d_{zz}, d_{xy}, d_{xz}, d_{yz}]$$

the DEV of the first region is along the y axis, while that of the second region is in the xy plane and inclined at 60° to the y axis. θ is chosen to be 45° to yield an even distribution of the real and the imaginary part.

The complex DWIs S_l are generated using the Stejskal-Tanner equation at each voxel $\mathbf{x}$ given by

$$S_l(\mathbf{x}) = S_0(\mathbf{x})e^{-\mathbf{b}_l:\mathbf{D}(\mathbf{x})} + n(\mathbf{x}) \tag{13}$$

where $n(\mathbf{x}) \sim N(0, \sigma_N) + iN(0, \sigma_N)$, $N(0, \sigma_N)$ is a zero mean Gaussian noise with standard deviation σ_N. As the signal measured before the Fourier transform in MRI formation is complex, it is reasonable to assume the noise is an additive complex Gaussian noise. The noise in the DWIs remains to be complex valued after the Fourier Transform. Thus, our simulated data reflects the correct physics of MRI imaging. Note that the noise in the magnitude of the complex DWIs have a Rician distribution and is approximated by a Gaussian distribution when the SNR is high [18]. We choose the 7 commonly used configurations x, y, z, xy, xz, yz, and xyz for the directions of the diffusion-encoded magnetic gradient as in [2] and use 3 different field strengths in each direction (100, 500, and 1000 s/mm^2). Thus, we have a total of 21 different $\mathbf{b}$ measurements. A slice of the generated data set is shown in Fig. 1, note that the DWIs are different when either the directions or the magnitudes of diffusion-encoded magnetic gradient are different.

For better illustration of the superior performance of our model, we compare performance with the following methods in our experiments: 1) **Linear**—linear regression on (3) as used in [2]; 2) **Nonlinear**—nonlinear regression applied to (1); 3) $\mathbf{Linear + EVS}$ (eigenvector smoothing)—linear regression followed by the DEV smoothing method described in Coulon *et al.*, [9]; 4) $\mathbf{Nonlinear + EVS}$—nonlinear regression plus the smoothing as in 3), and 5) **Ours**—Our method. Note that the EVS method in [9] is a direction field restoration scheme that preserves discontinuities based on the work of Chan and Shen [6].

Fig. 2 shows an ellipsoid visualization of the restored DTI for the synthetic data set with $\sigma_N = 0.5$. It is evident that our method restored the noisy tensor field quite well in comparison to the nonlinear regression method which did not perform well and the linear regression technique which performed the worst.

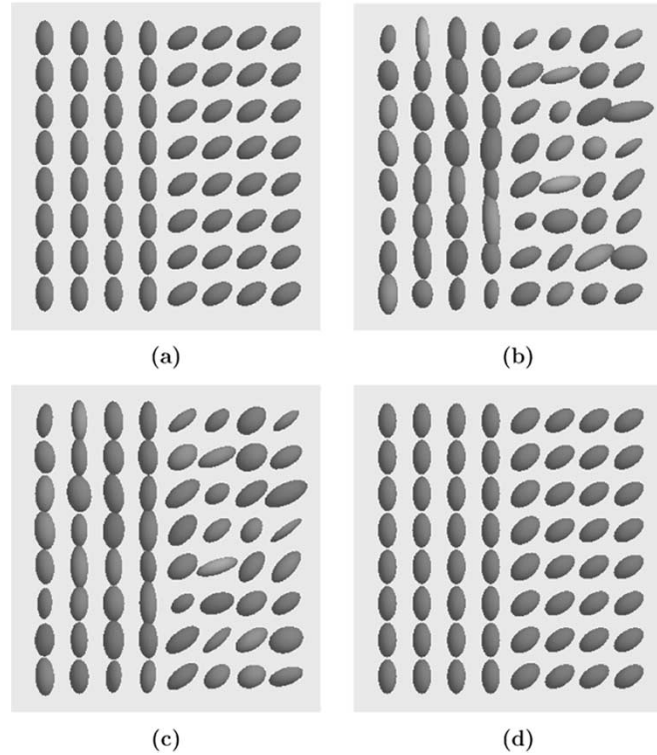

Fig. 2. A slice of the original (ground truth) and the estimated DTIs, respectively, for the noisy synthetic data with $\sigma_N = 0.5$. (a) Original, (b) linear, (c) nonlinear, and (d) our model.

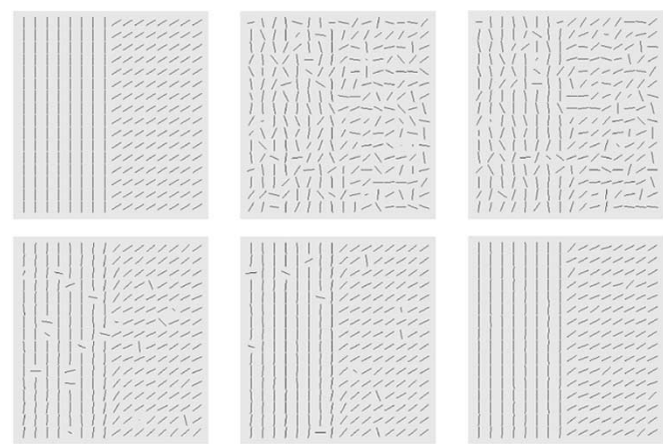

Fig. 3. A slice of the computed DEV field for the noisy synthetic data with $\sigma_N = 1.5$. Top left image is the DEV field computed from the original tensor field, and the other images arranged from left to right, top to bottom are the DEV field computed from estimated tensor field using the following methods: linear, nonlinear, $\mathrm{linear + EVS}$, $\mathrm{nonlinear + EVS}$ and our model.

For further comparison, Fig. 3 shows the DEV computed from the original and the restored DTI using all five methods as mentioned before. This figure clearly shows that our model yielded the best estimation of the original DEV field. The method in [9], however, did not work well at voxels where the estimated DEVs are almost orthogonal to those in their neighborhoods. In such cases, the method of Coulon *et al.* will treat them as discontinuities and does not smooth them.

TABLE I
COMPARISON OF THE ACCURACY OF THE ESTIMATED DEVS USING DIFFERENT METHODS FOR DIFFERENT NOISE LEVELS

$\sigma_n = 0.5$

	Linear	Nonlinear	Linear+EVS	Nonlinear+EVS	Ours
μ_θ	9.57	7.44	1.63	1.19	0.80
σ_θ	6.93	5.08	1.57	0.84	0.96

$\sigma_n = 1.0$

	Linear	Nonlinear	Linear+EVS	Nonlinear+EVS	Ours
μ_θ	22.28	16.94	6.67	3.78	1.99
σ_θ	17.46	13.86	13.06	8.58	2.74

$\sigma_n = 1.5$

	Linear	Nonlinear	Linear+EVS	Nonlinear+EVS	Ours
μ_θ	33.14	26.40	14.55	9.09	4.08
σ_θ	22.60	20.19	22.39	17.46	4.70

Though it is possible to treat these locations as outliers, it is difficult to set a reasonable criteria to achieve the same. It is interesting to notice that the $\mathrm{Nonlinear} + \mathrm{EVS}$ method which serves as an improvement of the existing $\mathrm{Linear} + \mathrm{EVS}$ method can diminish this problem. Additional quantitative measures, described below, confirm the visual comparison results.

To quantitatively assess the proposed model, we compare the accuracy of the DEV computed using the previously mentioned methods. Let θ be the angle (in degrees) between the estimated DEV and the original DEV, Table I shows the mean μ_θ and standard deviation σ_θ of θ using different methods for the synthetic data with different levels of additive Gaussian noises. A better method is one that yields smaller values. From this table, we can see that our model yields lower error values than all other methods under various noise levels. It is also clear from this table that the methods using the original nonlinear complex Stejskal-Tanner equation (1) are more accurate than those using the linearized one (3). The advantage of our method and the nonlinear approaches are more apparent when the noise level is higher, which supports our discussion in Section II-A.

B. Complex DWI of a Normal Rat Brain

The normal rat brain data is imaged using a 17.6-T (750-MHz) Bruker Avance Imaging Spectrometer system with the following settings: $\mathrm{TR} = 3058$ ms, $\mathrm{TE} = 28.8$ ms, $\Delta = 17.8$ ms, $\delta = 2.4$ ms, $\mathrm{diffusion\ time} = 17.0$ ms, and $\mathrm{bandwidth} = 40$ kHz. The field of view is $15 \times 15 \times 21$ mm^3 with a resolution of $117 \times 117 \times 270$ um^3. The same set of seven diffusion-encoded magnetic directions as the synthetic data are used with two different magnitudes (100, 500 s/mm^2). With a number of averages equal to 8 for each signal measurement S_l, the raw data is a set of 14 complex DWI volume data, each with a size of $128 \times 128 \times 78$.

We extract a $128 \times 128 \times 10$ volume in the region of the corpus callosum for our first experiment. Fig. 6 depicts the restored images of the six independent components of the estimated diffusion tensor. As a comparison, Figs. 4 and 5 show the same component images computed from th eraw data using linear regression applied to (3) and the nonlinear regression applied to (1). For display purposes, we use the same brightness and contrast enhancement for displaying the corresponding images in all the three figures. We also present the computed DEV of the estimated diffusion tensor in Fig. 7. We did not compare with the EVS methods because the sorting problem of the eigenvectors is very severe in free water or other isotropic regions, thus, it is necessary to exclude those regions to make effective usage of EVS methods. This involves a segmentation issue which is nontrivial.

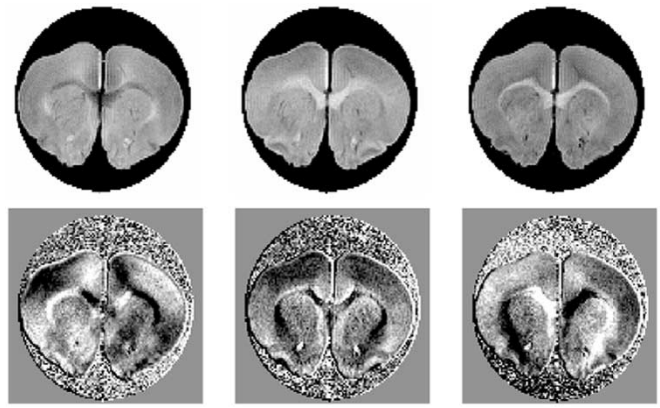

Fig. 4. A slice of the normal rat brain DTI estimated using multivariate linear regression without smoothing, viewed channel by channel. Top row, left to right: D_{xx}, D_{yy}, and D_{zz}. Bottom row, left to right: D_{xy}, D_{xz}, and D_{yz}.

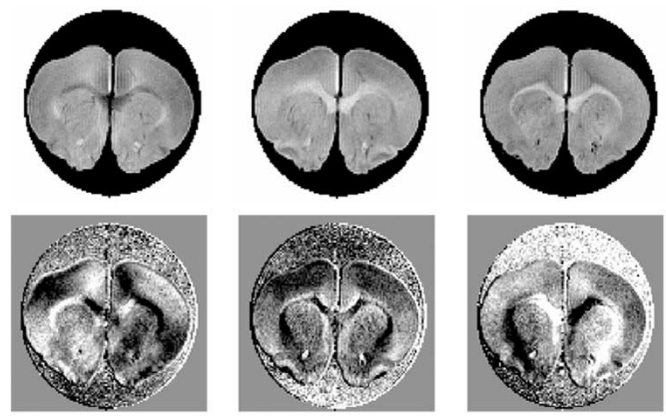

Fig. 5. A slice of the normal rat brain DTI estimated using multivariate nonlinear regression without smoothing, viewed channel by channel. Arrangement of the figures are the same as in Fig. 4.

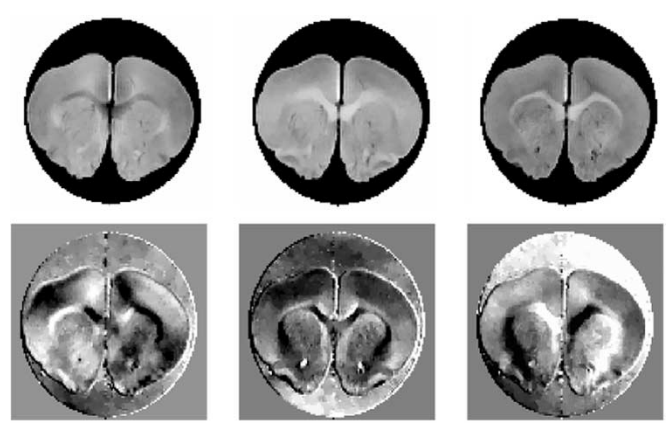

Fig. 6. A slice of the normal rat brain DTI estimated using using our proposed method, viewed channel by channel. Arrangement of the figures are the same as in Fig. 4.

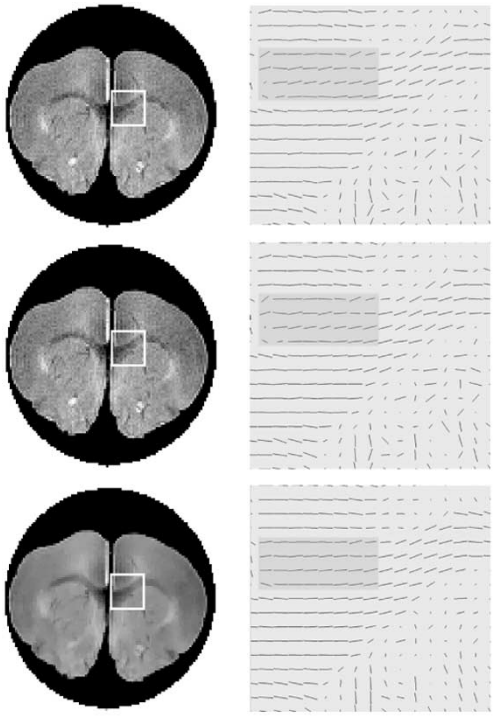

Fig. 7. A slice of the computed DEV field from a normal rat brain. Top to bottom: Linear, nonlinear regression and our model. Left column: Region of interest for depicting the DEV indicated by the white box superposed on the D_{xx} image. Right column: The computed DEV field inside the white rectangle on the left and the smoothing effect of our model is clearly visible in the shaded box.

We then extracted a $10\times 127 \times 78$ volume in the region of the cerebellum and show the sagittal view of the results in Fig. 8. The brightness and contrast enhancement of the figures are the same as in the previous experiment. In Figs. 6 and 8, the edge preserving smoothing is evident especially in the off diagonal terms of the diffusion tensor $\mathbf{D}$ which are essential in evaluating the structural anisotropy. We also notice that there are some differences in the region of free water between Figs. 4 and 5 visible in the off-diagonal terms while there is no difference visible inside the corpus callosum between these two figures. However, Fig. 7 gives more insight into this via the depiction of the computed DEVs. Note that smoothing effect on DEV is evident in the shadowed box in Fig. 7. Our model for estimating a smooth DTI from the noisy DWIs may be used to achieve better accuracy in fiber tractography. Our future efforts will focus on applying the model described in this paper to achieve better fiber tract maps.

V. Summary and Discussion

In this paper, we presented a novel constrained variational principle formulation for simultaneous smoothing and estimation of the symmetric positive definite DTI from complex DWIs. To our knowledge, this is the first attempt at simultaneous smoothing and estimation of the symmetric positive definite DTI from the complex DWI data. We used the Cholesky decomposition to incorporate the symmetric positive definiteness constraint on the diffusion tensor to be estimated. The constrained variational principle formulation is transformed into a sequence of unconstrained problems using the augmented Lagrangian technique and solved numerically. Proof of the existence of a solution for the minimization problem posed in the augmented Lagrangian framework is presented.

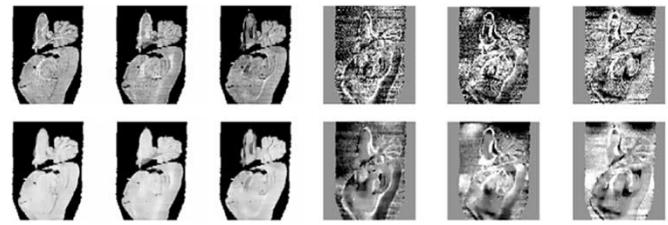

Fig. 8. A slice of the normal rat brain DTIs in the region of the cerebellum viewed channel by channel. The DTIs are estimated using multivariate nonlinear regression without smoothing (top row) and our proposed method (bottom row). Both the top and the bottom rows are sagittal views and are arranged left to right as: $D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{xz}$, and D_{yz}.

Results of comparison between our method and a representative [9] from the competing schemes are shown for synthetic data under a variety of situations involving the use of linearized and nonlinear data acquisition models depicting the influence of the choice of the data acquisition model on the estimation. It was concluded that using the complex nonlinear data model yields better accuracy in comparison to the log-linearized model. Also, superior performance of our method in estimating the tensor field over the chosen competing method was demonstrated for the synthetic data experiment. The estimated diffusion tensors are quite smooth without loss of essential features when inspected visually via the use of ellipsoid visualization as well as DEV field visualization. The superior performance is borne out via a quantitative statistical comparison of the angle between estimated DEV and ground truth DEV. Additional quantitative validation may be performed by comparing the estimated fiber tracts from the smooth tensor field obtained here to those obtained from histology as was done in our earlier work [23] and will be the focus of our future efforts.

Though the presented work focuses on diffusion tensor imaging, the constrained variational principle and the applied numerical methods can be easily tailored for processing the *high angular resolution diffusion imaging* (HARDI) data and other image data sets as well.

Appendix

Consider the augmented Lagrangian formulation (5) which serves as a subproblem of (4), the existence theorem will be stated and proved after the following two lemmas. If not stated, the definitions and theorems employed in the proof can be found in [10].

Lemma 1: Let $\mathcal{A}_1 = \{(S_0, \mathbf{L}) \,|\, (S_0, \mathbf{L}) \in \mathcal{A} \text{ and } \mathcal{C}(S_0, \mathbf{L}) \leq \mu\lambda\}$, and suppose $R_l, I_l \in L^2(\Omega)$, then the following minimization problem (14) has a solution $(S_0^*, \mathbf{L}^*) \in \mathcal{A}_1$ if $\mathcal{A}_1 \neq \emptyset$:

$$\min_{(S_0,\mathbf{L})\in\mathcal{A}_1} \mathcal{L}(S_0, \mathbf{L}; \lambda, \mu) = \mathcal{E}(S_0, \mathbf{L}) - \lambda\mathcal{C}(S_0, \mathbf{L}) + \frac{1}{2\mu}\mathcal{C}^2(S_0, \mathbf{L}). \quad (14)$$

Proof: We will verify the following three statements, one by one, and then prove this lemma.

- The first term $\mathcal{E}(S_0, \mathbf{L})$ in (14) is lower semi-continous with respect to $\mathbf{L}$ in $L^p(\Omega), p \geq 1$ and weakly lower semi-continous with respect to S_0 in $W^{1,p}(\Omega), p \geq q1$.
- The second term $\mathcal{C}(S_0, \mathbf{L})$ in (14) is continous with respect to S_0 in $W^{1,p}(\Omega)$ when $p > 6/5$, and is continous with respect to $\mathbf{L}$ in $L^p(\Omega)$ when $p \geq 1$.
- The third term $\mathcal{C}^2(S_0, \mathbf{L})$ in (14) has the same continuity property as the second term.

As $\mathcal{L}$ in (14) is lower bounded, there exists a minimizing sequence $(S_0^n, \mathbf{L}^n)$ for it, where $L_d^n \in L^p(\Omega), d \in \mathcal{D}, p \geq 1, R_0^n$, and $I_0^n \in W^{1,p}(\Omega), p > 6/5$ and $\mathcal{C}(S_0^n, \mathbf{L}^n) \leq \mu\lambda$. Then

$\{L_d^n\}_{n=1}^{\infty}, d \in \mathcal{D}$ is bounded in $L^p(\Omega), p \geq 1$.
$\{R_0^n\}_{n=1}^{\infty}$ and $\{I_0^n\}_{n=1}^{\infty}$ are bounded in $W^{1,p}(\Omega)$.

Therefore, there is a subsequence $\{(S_0^{n_k}, \mathbf{L}^{n_k})\}, L_d^* \in L^p(\Omega), p \geq 1, d \in \mathcal{D}$ and $R_0^*, I_0^* \in W^{1,p}(\Omega), p > 6/5$ such that when $n_k \to \infty$

- $\{L_d^{n_k}\} \to L_d^*, d \in \mathcal{D}$ in $L^1(\Omega)$ and a.e. on Ω. (From the compactness property of $L^p(\Omega), p \geq 1$).
- $\{R_0^{n_k}\} \rightharpoonup R_0^*$ and $\{I_0^{n_k}\} \rightharpoonup I_0^*$ in $W^{1,p}(\Omega)$. (From the weak compactness of $W^{1,p}$)
- $\{R_0^{n_k}\} \to R_0^*$ and $\{I_0^{n_k}\} \to I_0^*$ in $L^2(\Omega)$ and a.e on Ω. (From Rellich-Kondrachov Compactness Theorem when $p > 6/5$ and $n = 3$, **this is why we need** $p > 6/5$!).

1) By the lower semi-continuity of L^p norm in $L^p(\Omega), p \geq 1$ and the lower semi-continuity of L^p norm in $W^{1,p}(\Omega), p > 6/5$, we have

$$\mathcal{E}(S_0^*, \mathbf{L}^*) \leq \liminf_{n_k \to \infty} \int_\Omega \left[|\nabla R_0^{n_k}(\mathbf{x})|^{p_1} + |\nabla I_0^{n_k}(\mathbf{x})|^{p_1} + |\nabla \mathbf{L}^*(\mathbf{x})|^{p_2} \right] d\mathbf{x}.$$

2) Next, we claim

$$\mathcal{C}(S_0^*, \mathbf{L}^*) = \lim_{n_k \to \infty} \mathcal{C}\left(S_0^{n_k}, \mathbf{L}^{n_k}\right). \tag{15}$$

Since $\mathcal{C}(S_0^*, \mathbf{L}^*) = \alpha\sigma^2 - \int_\Omega \sum_{l=1}^N \|S_l - S_0^* e^{-\mathbf{b}_l : \mathbf{L}^* \mathbf{L}^{n_k}}\|^2 d\mathbf{x}$, we only need to show

$$\int_\Omega \sum_{l=1}^N \left\| S_l - S_0^* e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \right\|^2 d\mathbf{x} = \lim_{n_k \to \infty} \int_\Omega \sum_{l=1}^N \left\| S_l - S_0^{n_k} e^{-\mathbf{b_l} : \mathbf{L^{n_k}} (\mathbf{L^{n_k}})^\mathbf{T}} \right\|^2 d\mathbf{x}. \tag{16}$$

This will be proved in several stages

a) Let $FR_l^* = R_l - R_0^* e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}}$ and $FR_{l,n_k} = R_l - R_0^{n_k} e^{-\mathbf{b_l} : \mathbf{L^{n_k}} (\mathbf{L^{n_k}})^\mathbf{T}}$, then

$$\begin{aligned}
&\int_\Omega [FR_{l,n_k} - FR_l^*]^2 d\mathbf{x} \\
&= \int_\Omega \left[R_0^{n_k} e^{-\mathbf{b_l} : \mathbf{L^{n_k}} (\mathbf{L^{n_k}})^\mathbf{T}} - R_0^* e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \right]^2 d\mathbf{x} \\
&\leq 2 \int_\Omega \left[R_0^{n_k} \left(e^{-\mathbf{b_l} : \mathbf{L^{n_k}} (\mathbf{L^{n_k}})^\mathbf{T}} - e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \right) \right]^2 d\mathbf{x} \\
&\quad + 2 \int_\Omega \left[\left(R_0^{n_k} - R_0^* \right) e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \right]^2 d\mathbf{x} \\
&\triangleq 2A + 2B.
\end{aligned}$$

Since $\|R_0^{n_k}\|_{W^{1,p}(\Omega)}$ is uniformly bounded in n_k, by the Sobolev embedding theorem, for $\mathcal{R}^3$, we have

$$\|R_0^{n_k}\|_{L^2(\Omega)} \leq C \|R_0^{n_k}\|_{W^{1,p}(\Omega)} \leq C.$$

Noting the fact that $\mathbf{L^{n_k}}$ has a strong convergence toward $\mathbf{L}^*$ in $L^1(\Omega)$, we have from the Dominant Convergence Theorem [10]

$$\lim_{n_k \to \infty} \left\| e^{-\mathbf{b_l} : \mathbf{L^{n_k}} (\mathbf{L^{n_k}})^\mathbf{T}} - e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \right\|_{L^2(\Omega)} = 0.$$

Thus, we have

$$\begin{aligned}
A &\triangleq \int_\Omega \left[R_0^{n_k} \left(e^{-\mathbf{b_l} : \mathbf{L^{n_k}} (\mathbf{L^{n_k}})^\mathbf{T}} - e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \right) \right]^2 d\mathbf{x} \\
&\leq \|R_0^{n_k}\|_{L^2(\Omega)}^2 \left\| e^{-\mathbf{b_l} : \mathbf{L^{n_k}} (\mathbf{L^{n_k}})^\mathbf{T}} - e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \right\|_{L^2(\Omega)}^2 \\
&\xrightarrow{n_k \to \infty} 0
\end{aligned}$$

From the strong convergence of $R_0^{n_k}$ to R_0^* in $L^2(\Omega)$ and $e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \leq 1$, we have

$$|B| \leq \int_\Omega |R_0^{n_k} - R_0^*|^2 d\mathbf{x} \triangleq \|R_0^{n_k} - R_0^*\|_{L^2(\Omega)} \xrightarrow{n_k \to \infty} 0$$

Now as $A \xrightarrow{n_k \to \infty} 0$ and $B \xrightarrow{n_k \to \infty} 0$, we have

$$\lim_{n_k \to \infty} \int_\Omega [FR_l^* - FR_{l,n_k}]^2 d\mathbf{x} = 0$$

$$\Longrightarrow \int_\Omega FR_l^{*2} d\mathbf{x} = \lim_{n_k \to \infty} \int_\Omega FR_{l,n_k}^2 d\mathbf{x}. \tag{17}$$

b) Similarly as previous step, we can prove

$$\int_\Omega \left[I_l - I_0^* e^{-\mathbf{b_l} : \mathbf{L}^* \mathbf{L}^{*T}} \right]^2 d\mathbf{x} = \lim_{n_k \to \infty} \int_\Omega \left[I_l - I_0^{n_k} e^{-\mathbf{b_l} : \mathbf{L^{n_k}} (\mathbf{L^{n_k}})^\mathbf{T}} \right]^2 d\mathbf{x}. \tag{18}$$

Combining with (a), it is easy to verify (16).

3) Now we will show that

$$\mathcal{C}(S_0^*, \mathbf{L}^*)^2 = \lim_{n_k \to \infty} \mathcal{C}\left(S_0^{n_k}, \mathbf{L}^{n_k}\right)^2. \tag{19}$$

The above can be easily verified since $\mathcal{C}(S_0^*, \mathbf{L}^*), \mathcal{C}(S_0^{n_k}, \mathbf{L}^{n_k})$ are bounded and

$$\mathcal{C}(S_0^*, \mathbf{L}^*) = \lim_{n_k \to \infty} \mathcal{C}\left(S_0^{n_k}, \mathbf{L}^{n_k}\right). \tag{20}$$

Finally, we have from 1), 2), and 3)

$$\begin{aligned}
\mathcal{L}(S_0^*, \mathbf{L}^*; \lambda, \mu) &\leq \liminf_{n_k \to \infty} \mathcal{L}\left(S_0^{n_k}, \mathbf{L}^{n_k}; \lambda, \mu\right) \\
&= \inf \mathcal{L}(S_0, \mathbf{L}; \lambda, \mu)
\end{aligned} \tag{21}$$

Therefore, $(S_0^*, \mathbf{L}^*)$ is a minimizer of $\mathcal{L}(S_0, \mathbf{L}; \lambda, \mu)$ as defined in (14). ■

Lemma 2: Let $\mathcal{A}_2 = \{(S_0, \mathbf{L}) \,|\, (S_0, \mathbf{L}) \in \mathcal{A} \text{ and } \mathcal{C}(S_0, \mathbf{L}) > \mu\lambda\}$, and suppose $R_l, I_l \in L^2(\Omega)$, then the following minimization problem (22) has a solution $(S_0^*, \mathbf{L}^*) \in \mathcal{A}_2$ if $\mathcal{A}_2 \neq \emptyset$

$$\min_{(S_0, \mathbf{L}) \in \mathcal{A}_2} \mathcal{L}(S_0, \mathbf{L}; \lambda, \mu) = \mathcal{E}(S_0, \mathbf{L}) - \frac{\mu}{2}\lambda^2 \tag{22}$$

The proof is similar as in the Lemma 1.

Theorem 1: Suppose $R_l, I_l \in L^2(\Omega)$, then the augmented Lagrangian formulation (5) has a solution $(S_0^*, \mathbf{L}^*) \in \mathcal{A}$.

Proof: It is easy to see $\mathcal{A} \neq \emptyset$, as a matter of fact, constant functions will be members of $\mathcal{A}$. Thus, there will be three cases: 1) $\mathcal{A}_1 \neq \emptyset$ and $\mathcal{A}_2 = \emptyset$; 2) $\mathcal{A}_2 \neq \emptyset$ and $\mathcal{A}_1 = \emptyset$; and 3) $\mathcal{A}_1 \neq \emptyset$ and $\mathcal{A}_2 \neq \emptyset$. Here we provide a proof for case 3, case 1 and 2 are trival to prove. Let $(S_0^1, \mathbf{L}^1)$ be the solution for (14) and

$(S_0^2, \mathbf{L}^2)$ be the solution for (22), it is not hard to see that the solution of (5) is

$$(S_0^*, \mathbf{L}^*) = \begin{cases} \left(S_0^1, \mathbf{L}^1\right), & \text{if } \mathcal{L}\left(S_0^1, \mathbf{L}^1; \lambda, \mu\right) \leq \mathcal{L}\left(S_0^2, \mathbf{L}^2; \lambda, \mu\right) \\ \left(S_0^2, \mathbf{L}^2\right), & \text{otherwise} \end{cases}.$$

■

ACKNOWLEDGMENT

MR images were obtained at the AMRIS facility in the McKnight Brain Institute of the University of Florida. The authors would also like to thank E. Özarslan for discussions on the physics of imaging.

REFERENCES

[1] L. Alvarez, P.-L. Lions, and J.-M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion. II," in *SIAM J. Numer. Anal.*, vol. 29, June 1992, pp. 845–866.

[2] P. J. Basser, J. Mattiello, and D. Lebihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *J. Magn. Reson.*, vol. 103, no. 3, pp. 247–254, 1994.

[3] P. J. Basser and C. Pierpaoli, "Microstructural and physiological features of tissue elucidated by quantitative diffusion-tensor MRI," *J. Magn. Reson.*, vol. 111, no. 3, pp. 209–219, 1996.

[4] P. Blomgren, T. F. Chan, and P. Mulet, "Extensions to Total Variation Denoising," UCLA, Los Angeles, CA, Tech. Rep. Rep. 97-42, Sept. 1997.

[5] V. Caselles, J. M. Morel, G. Sapiro, and A. Tannenbaum, *IEEE Trans. Image Processing (Special Issue on PDEs and Geometry-Driven Diffusion in Image Processing and Analysis)*, vol. 7, Mar. 1998.

[6] T. F. Chan and J. Shen, "Variational restoration of nonflat image features: Model and algorithm," in *SIAM J. Appl. Math.*, vol. 61, 2000, pp. 1338–1361.

[7] C. Chefd'hotel, O. Faugeras, D. Tschumperlé, and R. Deriche, "Constrained flows of matrix-valued functions: Application to diffusion tensor regularization," in *Lecture Notes in Computer Science*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds. Berlin, Germany: Springer-Verlag, 2002, vol. 2351, Proc. 7th Eur. Conf. Comput. Vision, pp. 251–265.

[8] T. E. Conturo, N. F. Lori, T. Cull, E. Akbudak, A. Z. Snyder, J. S. Shimony, R. C. McKinstry, H. Burton, and M. E. Raichle, "Tracking neuronal fiber pathways in the living human brain," *Proc. Nat. Acad. Sci.*, pp. 10 422–10 427, 1999.

[9] O. Coulon, D. C. Alexander, and S. R. Arridge, "A regularization scheme for diffusion tensor magnetic resonance images," in *Lecture Notes in Computer Science*, M. F. Insana and R. M. Leahy, Eds. Berlin, Germany: Springer-Verlag, 2001, vol. 2082, Proc. Information Processing in Medical Imaging, pp. 92–105.

[10] L. C. Evans, *Partial Differential Equations*: American Mathematical Society, 1997.

[11] D. K. Jones, A. Simmons, S. C. R. Williams, and M. A. Horsfield, "Noninvasive assessment of axonal fiber connectivity in the human brain via diffusion tensor MRI," *Magn. Reson. Med.*, vol. 42, pp. 37–41, 1999.

[12] R. Kimmel, R. Malladi, and N. A. Sochen, "Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images," *Intl. J. Comput. Vis.*, vol. 39, no. 2, pp. 111–129, 2000.

[13] B. W. Lindgren, *Statistical Theory*. London, U.K.: Chapman Hall/CRC, 1939.

[14] J. Nocedal and S. J. Wright, *Numerical Optimization*: Springer, 2000.

[15] G. J. M. Parker, J. A. Schnabel, M. R. Symms, D. J. Werring, and G. J. Baker, "Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging," *J. Magn. Reson. Imag.*, vol. 11, no. 6, pp. 702–710, 2000.

[16] C. Poupon, J. F. Mangin, C. A. Clark, V. Frouin, J. Regis, D. Bihan, and I. Block, "Toward inference of human brain connectivity from MR diffusion tensor data," *Med. Image Anal.*, vol. 5, no. 1, pp. 1–15, Mar. 2001.

[17] G. A. F. Seber, *Linear Regression Analysis*. New York: Wiley, 1977.

[18] J. Sijbers, "Signal and noise estimation from magnetic resonance images," Ph.D. dissertation, Dept. Natuurkunde, Universitaire Instelling Antwerpen, May 1998.

[19] B. Tang, G. Sapiro, and V. Caselles, "Diffusion of general data on nonflat manifolds via harmonic maps theory: The direction diffusion case," *Int. J. Comput. Vis.*, vol. 36, no. 2, pp. 149–161, 2000.

[20] D. Tschumperlé and R. Deriche, "Orthonormal vector sets regularization with PDE's and applications," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 237–252, Dec. 2002.

[21] ——, "Variational frameworks for DT-MRI estimation, regularization, and visualization," in *Proc. 9th Intl. Conf. Comput. Vision*, vol. 2, 2003, pp. 116–121.

[22] B. C. Vemuri, Y. Chen, M. Rao, T. McGraw, Z. Wang, and T. H. Mareci, "Fiber tract mapping from diffusion tensor MRI," in *Proc. IEEE Workshop Variational and Level Set Methods in Computer Vision*, 2001, pp. 81–88.

[23] B. C. Vemuri, Y. Chen, M. Rao, Z. Wang, T. McGraw, T. H. Mareci, S. J. Blackband, and P. Reier, "Automatic fiber tractography from DTI and ITS validation," in *Proc. IEEE Int. Symp. Biomedical Imaging*, 2002, pp. 501–504.

[24] Z. Wang, B. C. Vemuri, Y. Chen, and T. H. Mareci, "A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from DWI," in *Lecture Notes in Computer Science*, vol. 2732, Proc. Information Processing in Medical Imaging, C. J. Taylor and J. A. Noble, Eds. Berlin, Germany, 2003, pp. 660–671.

[25] ——, "Simultaneous smoothing and estimation of the tensor field from diffusion tensor MRI," in *Proc. IEEE Computer Society Conf. Comput. Vision and Pattern Recognition*, vol. I, 2003, pp. 461–466.

[26] J. Weickert, "A review of nonlinear diffusion filtering," in *Lecture Notes in Computer Science*, vol. 1252, Proc. 1st Int. Conf. Scale-Space Theory in Computer Vision, B. M. terHaarRomeny, L. Florack, J. J. Koenderink, and M. A. Viergever, Eds. Berlin, Germany, 1997, pp. 3–28.

[27] C. F. Westin, S. E. Maier, H. Mamata, A. Nabavi, F. A. Jolesz, and R. Kikinis, "Processing and visualization for diffusion tensor MRI," *Med. Image Anal.*, vol. 6, no. 2, pp. 93–108, 2002.