# Error Propagation Framework for Diffusion Tensor Imaging via Diffusion Tensor Representations

Cheng Guan Koay*, Lin-Ching Chang, Carlo Pierpaoli, and Peter J. Basser

***Abstract*—An analytical framework of error propagation for diffusion tensor imaging (DTI) is presented. Using this framework, any uncertainty of interest related to the diffusion tensor elements or to the tensor-derived quantities such as eigenvalues, eigenvectors, trace, fractional anisotropy (FA), and relative anisotropy (RA) can be analytically expressed and derived from the noisy diffusion-weighted signals. The proposed framework elucidates the underlying geometric relationship between the variability of a tensor-derived quantity and the variability of the diffusion weighted signals through the nonlinear least squares objective function of DTI. Monte Carlo simulations are carried out to validate and investigate the basic statistical properties of the proposed framework.**

***Index Terms*—Cone of uncertainty, covariance structures, diffusion tensor imaging, diffusion tensor representations, error propagation, invariant Hessian.**

## I. Introduction

DIFFUSION tensor imaging (DTI) is a unique noninvasive magnetic resonance imaging technique capable of probing tissue microstructure in the brain [1]–[7]. DTI is a well-established diagnostic technique and has provided fresh impetus in monitoring human brain morphology and development [6]–[12]. Therefore, an accurate quantification of uncertainties in tensor elements as well as in tensor derived quantities, such as the eigenvalues, eigenvectors, trace, fractional anisotropy (FA), and relative anisotropy (RA), is needed so that statistical inferences can inform clinical decision making.

Accurate characterization of variability in tensor-derived quantities is of great relevance in various stages of DTI data analysis—from exploratory and diagnostic testing to hypothesis testing, experimental design and tensor classification. To date, many studies have been conducted on optimal design and the effects of noise in DTI [13]–[24]. In the context of variability studies on tensor-derived quantities in DTI, there are currently two different methods—perturbation and error propagation—which have been studied in the work of Anderson [17], Chang *et al.* [21], and Poonawalla [19]. However, these studies were based on the linear objective function of DTI, which may not be appropriate for diffusion that is anisotropic [25]. Further, Poonawalla [19] focused only on anisotropy (or scalar) calculations.

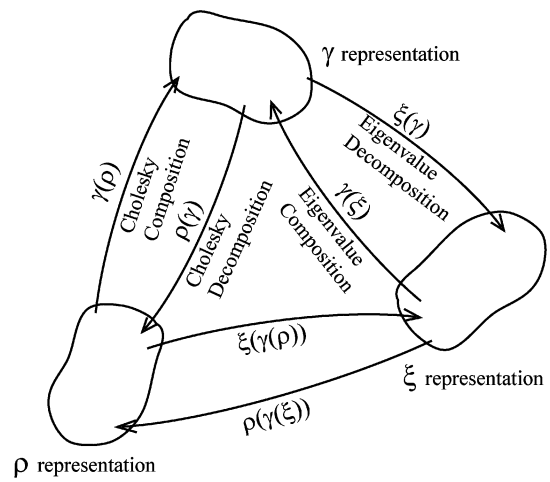



Fig. 1. Different representations and coordinate transformations of the diffusion tensor. As defined in the text, $\gamma$ is the ordinary representation of the diffusion tensor together with the logarithm of the reference signal. Similarly, $\rho$ and $\xi$ are representations derived from the Cholesky composition and from the Eigenvalue composition, respectively. Note that decompositions, Cholesky or Eigenvalue, are more numerical in character whereas their compositions are more analytical or rather, analytically more tractable.

In this paper, our goal is to present a general analytical error propagation framework for DTI based on the nonlinear objective functions of DTI and to show the relevance of various diffusion tensor representations to DTI error propagation. Fig. 1 shows three basic diffusion tensor representations and their mappings. The proposed theoretical framework allows the uncertainty to be calculated for any tensor-derived quantity including the eigenvector—the main geometric object of DTI tractography. Within this framework, the cone of uncertainty [26]–[30] can be quantitatively estimated; this framework coupled with the observation of Jeong *et al.* [30] and Lazar *et al.* [29] provides converging evidence that the cone of uncertainty is generally elliptical. A fresh approach is taken to show both the geometric and analytical aspects of the proposed framework without heavy machinery from differential geometry and tensor calculus [31]–[40].

Monte Carlo simulations are carried out to investigate the basic statistical properties of the proposed framework. Some material here was previously presented in abstract form [23].

Manuscript received December 21, 2006; revised February 27, 2007. This work was supported by the Intramural Research Program of the National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD. *Asterisk indicates corresponding author.*

*C. G. Koay is with the National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892 USA (e-mail: guankoac@mail.nih.gov).

L.-C. Chang, C. Pierpaoli, and P. J. Basser are with the National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892 USA.

Digital Object Identifier 10.1109/TMI.2007.897415

## II. Methods

### A. Nonlinear DTI Estimation in Different Diffusion Tensor Representations

In a typical DTI experiment, the measured signal in a single voxel has the following form [1], [4], [41]:

$$s = S_0 \exp(-b\mathbf{g}^T\mathbf{D}\mathbf{g}) \tag{1}$$

where the measured signal $s$ depends on the diffusion encoding gradient vector $\mathbf{g}$ of unit length, the diffusion weight $b$, the reference signal $S_0$, and the diffusion tensor $\mathbf{D}$. The symbol "$T$" denotes the matrix or vector transpose. Given $n \geq 7$ sampled signals based on at least six noncollinear gradient directions and at least one sampled reference signal, the diffusion tensor estimate can be found by minimizing various objective functions with respect to different representations of the diffusion tensor in (1). Different representations of the diffusion tensor provide different insights and information about the diffusion tensor itself. We will use three distinct diffusion tensor representations that have applications to DTI and show how they can be used in DTI error propagation.

In general, the objective functions for the nonlinear least squares problem in different diffusion tensor representations can be expressed as follows:

$$f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{n}\left(\underbrace{s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right]}_{r_i(\boldsymbol{\gamma})}\right)^2 \tag{2}$$

$$f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho})) = \frac{1}{2}\sum_{i=1}^{n}\left(\underbrace{s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j(\boldsymbol{\rho})\right]}_{r_i(\boldsymbol{\gamma}(\boldsymbol{\rho}))}\right)^2 \tag{3}$$

$$f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\boldsymbol{\xi})) = \frac{1}{2}\sum_{i=1}^{n}\left(\underbrace{s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j(\boldsymbol{\xi})\right]}_{r_i(\boldsymbol{\gamma}(\boldsymbol{\xi}))}\right)^2 \tag{4}$$

where $s_i$ is the measured diffusion weighted signal with noise; $\hat{s}_i(\boldsymbol{\gamma}) = \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right]$ is the diffusion weighted function evaluated at $\boldsymbol{\gamma}$; $\hat{s}_i(\boldsymbol{\gamma}(\boldsymbol{\rho})) = \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j(\boldsymbol{\rho})\right]$ is the diffusion weighted function evaluated at $\boldsymbol{\gamma}(\boldsymbol{\rho})$; $\hat{s}_i(\boldsymbol{\gamma}(\boldsymbol{\xi})) = \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j(\boldsymbol{\xi})\right]$ is the diffusion weighted function evaluated at $\boldsymbol{\gamma}(\boldsymbol{\xi})$; and see the equations at the bottom of the page.

The three representations of the diffusion tensor are

$$\begin{aligned}\boldsymbol{\gamma} &= \begin{bmatrix}\gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 & \gamma_6 & \gamma_7\end{bmatrix}^T \\ &= [\ln(\alpha) \quad D_{xx} \quad D_{yy} \quad D_{zz} \quad D_{xy} \quad D_{yz} \quad D_{xz}]^T \end{aligned} \tag{5}$$

$$\boldsymbol{\rho} = [\rho_1 \quad \rho_2 \quad \rho_3 \quad \rho_4 \quad \rho_5 \quad \rho_6 \quad \rho_7]^T \tag{6}$$

$$\begin{aligned}\boldsymbol{\xi} &= [\xi_1 \quad \xi_2 \quad \xi_3 \quad \xi_4 \quad \xi_5 \quad \xi_6 \quad \xi_7]^T \\ &= [\ln(\alpha) \quad \lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \theta \quad \phi \quad \psi]^T \end{aligned} \tag{7}$$

where $\alpha$ is the parameter for the nondiffusion weighted signal.

We shall denote $\boldsymbol{\gamma}$, $\boldsymbol{\rho}$, and $\boldsymbol{\xi}$ as the ordinary, the Cholesky, and the Euler representations, respectively. The meaning of each term mentioned here will be obvious in the following discussion. Fig. 1 shows the mappings between different spaces or representations.

To construct the mappings $\boldsymbol{\gamma}(\boldsymbol{\rho})$ and $\boldsymbol{\gamma}(\boldsymbol{\xi})$, we use the main ideas from the Cholesky decomposition of a symmetric positive definite matrix and the eigenvalue decomposition of a symmetric matrix [42] in reverse. The connections between (5) and (6) and between (5) and (7) can then be established based on the following equations:

$$\mathbf{D}(\boldsymbol{\rho}) = \mathbf{U}^T\mathbf{U} \tag{8}$$

and

$$\mathbf{D}(\boldsymbol{\xi}) = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T = \sum_{i=1}^{3}\lambda_i\mathbf{q}_i\mathbf{q}_i^T \tag{9}$$

where $\mathbf{U}$ is an upper triangular matrix with nonzero diagonal elements and $\mathbf{q}_i$ are the column vectors of $\mathbf{Q}$ which depend on the Euler angles. Without loss of generality, we shall assume the eigenvalues are arranged in decreasing order, i.e., $\lambda_1 \geq \lambda_2 \geq$

$$\begin{aligned} r(\boldsymbol{\gamma}) &= [r_1(\boldsymbol{\gamma}) \quad \ldots \quad r_n(\boldsymbol{\gamma})]^T \\ r(\boldsymbol{\gamma}(\boldsymbol{\rho})) &= [r_1(\boldsymbol{\gamma}(\boldsymbol{\rho})) \quad \ldots \quad r_n(\boldsymbol{\gamma}(\boldsymbol{\rho}))]^T \\ r(\boldsymbol{\gamma}(\boldsymbol{\xi})) &= [r_1(\boldsymbol{\gamma}(\boldsymbol{\xi})) \quad \ldots \quad r_n(\boldsymbol{\gamma}(\boldsymbol{\xi}))]^T \end{aligned}$$

and

$$W = \begin{pmatrix} 1 & -b_1g_{1x}^2 & -b_1g_{1y}^2 & -b_1g_{1z}^2 & -2b_1g_{1x}g_{1y} & -2b_1g_{1y}g_{1z} & -2b_1g_{1x}g_{1z} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -b_ng_{nx}^2 & -b_ng_{ny}^2 & -b_ng_{nz}^2 & -2b_ng_{nx}g_{ny} & -2b_ng_{ny}g_{nz} & -2b_ng_{nx}g_{nz} \end{pmatrix}$$

$\lambda_3$. Each column vector of $\mathbf{Q}$ is also an eigenvector of $D(\boldsymbol{\xi})$. Particularly, we have

$$\mathbf{U}(\boldsymbol{\rho}) = \begin{pmatrix} \rho_2 & \rho_5 & \rho_7 \\ 0 & \rho_3 & \rho_6 \\ 0 & 0 & \rho_4 \end{pmatrix}$$

$$\mathbf{Q} = \mathbf{R}_z(\phi)\mathbf{R}_y(\theta)\mathbf{R}_z(\psi) = \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{pmatrix} \tag{10}$$

$$\mathbf{q}_1 = \begin{pmatrix} Q_{11} \\ Q_{21} \\ Q_{31} \end{pmatrix} = \begin{pmatrix} \cos(\theta)\cos(\phi)\cos(\psi) - \sin(\phi)\sin(\psi) \\ \cos(\theta)\cos(\psi)\sin(\phi) + \cos(\phi)\sin(\psi) \\ -\cos(\psi)\sin(\theta) \end{pmatrix} \tag{11}$$

$$\mathbf{q}_2 = \begin{pmatrix} Q_{12} \\ Q_{22} \\ Q_{32} \end{pmatrix} = \begin{pmatrix} -\cos(\psi)\sin(\phi) - \cos(\theta)\cos(\phi)\sin(\psi) \\ \cos(\phi)\cos(\psi) - \cos(\theta)\sin(\phi)\sin(\psi) \\ \sin(\theta)\sin(\psi) \end{pmatrix} \tag{12}$$

$$\mathbf{q}_3 = \begin{pmatrix} Q_{13} \\ Q_{23} \\ Q_{33} \end{pmatrix} = \begin{pmatrix} \cos(\phi)\sin(\theta) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{pmatrix} \tag{13}$$

and

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \tag{14}$$

where $\lambda_1$, $\lambda_2$, and $\lambda_3$ are the eigenvalues of $\mathbf{D}(\boldsymbol{\xi})$. If $\mathbf{D}(\boldsymbol{\xi})$ is positive definite then its eigenvalues are positive. Finally, the rotation matrices, $\mathbf{R}_x(\Omega)$, $\mathbf{R}_y(\Omega)$, and $\mathbf{R}_z(\Omega)$ represent rotations through angle $\Omega$ around the $x$, $y$, and $z$ axes, respectively, and are defined in Appendix I.

Given (8) and (9), the mappings, $\boldsymbol{\gamma}(\boldsymbol{\rho})$ and $\boldsymbol{\gamma}(\boldsymbol{\xi})$, can be expressed as

$$\boldsymbol{\gamma}(\boldsymbol{\rho}) = \begin{pmatrix} \gamma_1(\boldsymbol{\rho}) \\ \gamma_2(\boldsymbol{\rho}) \\ \gamma_3(\boldsymbol{\rho}) \\ \gamma_4(\boldsymbol{\rho}) \\ \gamma_5(\boldsymbol{\rho}) \\ \gamma_6(\boldsymbol{\rho}) \\ \gamma_7(\boldsymbol{\rho}) \end{pmatrix} = \begin{pmatrix} \gamma_1(\boldsymbol{\rho}) \\ D_{xx}(\boldsymbol{\rho}) \\ D_{yy}(\boldsymbol{\rho}) \\ D_{zz}(\boldsymbol{\rho}) \\ D_{xy}(\boldsymbol{\rho}) \\ D_{yz}(\boldsymbol{\rho}) \\ D_{xz}(\boldsymbol{\rho}) \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2^2 \\ \rho_3^2 + \rho_5^2 \\ \rho_4^2 + \rho_6^2 + \rho_7^2 \\ \rho_2\rho_5 \\ \rho_3\rho_6 + \rho_5\rho_7 \\ \rho_2\rho_7 \end{pmatrix} \tag{15}$$

$$\boldsymbol{\gamma}(\boldsymbol{\xi}) = \begin{pmatrix} \gamma_1(\boldsymbol{\xi}) \\ \gamma_2(\boldsymbol{\xi}) \\ \gamma_3(\boldsymbol{\xi}) \\ \gamma_4(\boldsymbol{\xi}) \\ \gamma_5(\boldsymbol{\xi}) \\ \gamma_6(\boldsymbol{\xi}) \\ \gamma_7(\boldsymbol{\xi}) \end{pmatrix} = \begin{pmatrix} \gamma_1(\boldsymbol{\xi}) \\ D_{xx}(\boldsymbol{\xi}) \\ D_{yy}(\boldsymbol{\xi}) \\ D_{zz}(\boldsymbol{\xi}) \\ D_{xy}(\boldsymbol{\xi}) \\ D_{yz}(\boldsymbol{\xi}) \\ D_{xz}(\boldsymbol{\xi}) \end{pmatrix}. \tag{16}$$

Since the expressions for (16) are lengthy but easy to compute, we have collected them in Appendix II.

It is important to note that the inverse mapping of $\boldsymbol{\gamma}(\boldsymbol{\rho})$, $\boldsymbol{\rho}(\boldsymbol{\gamma})$, which can be constructed analytically, is well defined only when the diffusion tensor contained within $\boldsymbol{\gamma}$ is positive definite, otherwise the modified Cholesky decomposition is needed to force the diffusion tensor to be sufficiently positive definite [43]. However, the solution obtained from the modified Cholesky decomposition is generally not a minimizer of $f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho}))$. The solution is, nevertheless, useful as an initial guess of the minimization of $f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho}))$. A specific algorithm of this type of minimization, where the resultant diffusion tensor estimate is both positive definite and a minimizer of $f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho}))$, can be found in [25]. Finally, the analytical expression of $\boldsymbol{\rho}(\boldsymbol{\gamma}))$ based on the Cholesky decomposition is shown in Appendix II.

Another mapping of interest is $\boldsymbol{\xi}(\boldsymbol{\gamma})$. The construction of $\boldsymbol{\xi}(\boldsymbol{\gamma})$, which requires the eigenvalue decomposition of a symmetric matrix, e.g., using the Jacobi method (34), has two main advantages. First, it is numerically more stable and more accurate than the analytical approach of [44]. Second, it can be used even when the diffusion tensor within $\boldsymbol{\gamma}$ is not positive definite—an additional advantage over the analytical approach of [44]. Once the orthogonal matrix $\mathbf{Q}$ is obtained by diagonalization, we still need to solve for the Euler angles, $\theta$, $\phi$, and $\psi$. The solution to this problem is simple but, for completeness, we have collected these results in Appendix I. The Euler representation is more useful than the representation proposed by Hext [45], a special case of which was used by Anderson [17] and Chang *et al.* [21] in computing the covariance between eigenvalues, because the covariance matrix of the major eigenvector of the diffusion tensor can be constructed in the Euler representation. Appendix III contains further comments on the representation by Hext.

The first two objective functions, (2) and (3), have been used in many studies [25], [46]–[50], and the theoretical and algorithmic framework for these objective functions was investigated by Koay *et al.* [25]. To date, the third objective function, (4), has not been used for DTI estimation because the direct estimation of the eigenvalues and eigenvectors by (4) is impractical due to the cost of computation, particularly for the initial solution and for those trigonometric functions occurring in the rotation matrix. Nonetheless, (4) as expressed in the Euler representation does provide a foundation for DTI error propagation that is conceptually elegant and algorithmically practical.

We introduce the proposed framework with respect first to the ordinary representation for a scalar function in Section II-B and then for a vector function in Section II-C. In Section II-D, we discuss commonly used scalar and vector tensor-derived quantities and their corresponding gradient vectors, while Section II-E covers the diffusion tensor representation and analytical formulas for the invariant Hessian structures, a new concept to be defined later, with respect to different diffusion tensor representations. Section II-F discusses selected applications of the proposed framework. Fig. 7 shows the schematic diagram of the necessary steps needed to obtain appropriate covariance structures. The segment above the dotted line in Fig. 7 deals with diffusion tensor estimations; these techniques can be found in [25], while the segment below the dotted line pertains to the proposed framework.

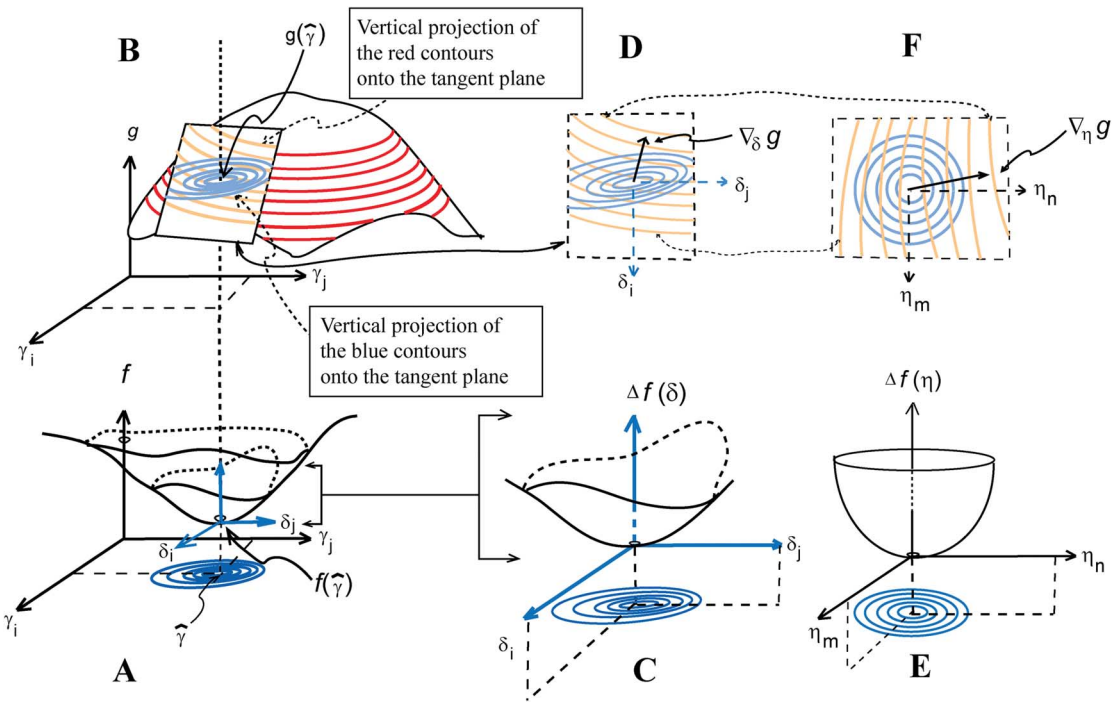



Fig. 2. (A) Hyper-surface of the nonlinear objective function, $f_{\mathrm{NLS}}$, with respect to the $\boldsymbol{\gamma}$ coordinate system with a minimum value of $f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})$ at $\hat{\boldsymbol{\gamma}}$. A new coordinate system centered at $f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})$ is also shown here and will be denoted as the $\boldsymbol{\delta}$-coordinate system. (B) Typical hyper-surface of a tensor-derived quantity with respect to a $\boldsymbol{\gamma}$-coordinate system. The contours of $f$ are projected vertically onto the tangent plane of $g$. This tangent plane of $g$ at $g(\hat{\boldsymbol{\gamma}})$ shows the intersection between the contours of $f_{\mathrm{NLS}}$ and those of $g$. (C) The magnified image of the region centered at $f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})$ with respect to the $\boldsymbol{\delta}$-coordinate system. (D) The magnified image of the tangent plane of $g$ at $g(\hat{\boldsymbol{\gamma}})$. The gradient vector of $g(\hat{\boldsymbol{\gamma}})$ shows the direction of greatest ascent with respect to the landscape of $g$ around $g(\hat{\boldsymbol{\gamma}})$. (E) New look of the hyper-surface of $f_{\mathrm{NLS}}$ with respect to the transformed $\boldsymbol{\eta}$-coordinate system as defined in both (20) and (21) where the change in $f_{\mathrm{NLS}}$ looks uniform in all directions of any unit vector. (F) Tangent plane of $g(\hat{\boldsymbol{\gamma}})$ with respect to the $\boldsymbol{\eta}$-coordinate system.

### *B. Error Propagation Framework For Scalar Functions*

Let $f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = (1/2)\sum_{i=1}^{n}\left(s_i - \exp\left[\sum_{j=1}^{7} W_{ij}\gamma_j\right]\right)^2$ be the NLS objective function in the ordinary representation. Let $g$ be any smooth function (tensor-derived quantity) of $\boldsymbol{\gamma}$ and let $\hat{\boldsymbol{\gamma}}$ be the NLS estimate, i.e., $\hat{\boldsymbol{\gamma}}$ is a minimizer of $f_{\mathrm{NLS}}$. The connection between the uncertainty of $g$ and of $f_{\mathrm{NLS}}$ can be represented geometrically.

To examine the effect of the variability of $f_{\mathrm{NLS}}$ on the variability of $g$, we first focus on the region around $f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})$ (the blue contour) and its relation to the function $g$ by projecting the contour around $f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})$ to the tangent plane of $g$ at $g(\hat{\boldsymbol{\gamma}})$ [Fig. 2(A) and (B)].

By second-order Taylor expansion, the change in $f_{\mathrm{NLS}}$ is

$$\Delta f_{\mathrm{NLS}}(\boldsymbol{\delta}) = f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}} + \boldsymbol{\delta}) - f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}}) \approx \frac{1}{2}\boldsymbol{\delta}^T \nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})\boldsymbol{\delta} \tag{17}$$

where $\boldsymbol{\delta}(\boldsymbol{\gamma}) \equiv \boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}$ and $\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})$ is the Hessian matrix of $f_{\mathrm{NLS}}$. Here, we can safely assume that $\nabla f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}}) = \mathbf{0}$ because $\hat{\boldsymbol{\gamma}}$ minimizes $f_{\mathrm{NLS}}$ [Fig. 2(C)]. In the same vein, the first-order change in $g$ is

$$\Delta g(\boldsymbol{\eta}) = g(\hat{\boldsymbol{\gamma}} + \boldsymbol{\eta}) - g(\hat{\boldsymbol{\gamma}}) \approx \nabla_{\boldsymbol{\eta}}^T g \boldsymbol{\eta} \tag{18}$$

where $\boldsymbol{\eta}$ is defined later. If $\hat{\boldsymbol{\gamma}}$ minimizes $f_{\mathrm{NLS}}$ then the Hessian matrix $\nabla^2 f_{\mathrm{NLS}}$ is positive definite at $\hat{\boldsymbol{\gamma}}$ and can be written as

$$\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T = \mathbf{Q}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\mathbf{Q}^T = \mathbf{Q}\boldsymbol{\Lambda}^{1/2}(Q\boldsymbol{\Lambda}^{1/2})^T \tag{19}$$

where $\mathbf{Q}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix with positive elements. Therefore, we can express the change in $f_{\mathrm{NLS}}$ as

$$\Delta f_{\mathrm{NLS}}(\boldsymbol{\eta}) = \frac{1}{2}\boldsymbol{\eta}^T\boldsymbol{\eta} \tag{20}$$

such that

$$\boldsymbol{\eta} \equiv (\mathbf{Q}\boldsymbol{\Lambda}^{1/2})^T\boldsymbol{\delta}. \tag{21}$$

In the $\boldsymbol{\eta}$-coordinate system, the change in $f_{\mathrm{NLS}}$ looks uniform in all directions of $\boldsymbol{\eta}$ since (20) is the equation of a hyper-sphere [Fig. 2(E)]. To measure or capture the change in $g$ in a consistent manner, $\boldsymbol{\eta}$ has to satisfy (20) and be parallel to $\nabla_{\boldsymbol{\eta}} g$; the theoretical reason behind the latter condition is related to another condition, which we shall refer to as the *consistency* condition. This condition is best explained using a geometric figure and is discussed in the caption of Fig. 3. These conditions then lead naturally to the following formula:

$$\boldsymbol{\eta} = \frac{\sqrt{2\Delta f_{\mathrm{NLS}}}\nabla_{\boldsymbol{\eta}} g}{||\nabla_{\boldsymbol{\eta}} g||} \tag{22}$$

where $\nabla_{\boldsymbol{\eta}} g/||\nabla_{\boldsymbol{\eta}} g||$ is a unit vector along $\nabla_{\boldsymbol{\eta}} g$. Therefore, (18) can be written as

$$\begin{aligned}\Delta g(\boldsymbol{\eta}) &\approx \nabla_{\boldsymbol{\eta}}^T g \boldsymbol{\eta} = \nabla_{\boldsymbol{\eta}}^T g\left(\frac{\sqrt{2\Delta f_{\mathrm{NLS}}}\nabla_{\boldsymbol{\eta}} g}{||\nabla_{\boldsymbol{\eta}} g||}\right) \\ &= \sqrt{2\Delta f_{\mathrm{NLS}}}||\nabla_{\boldsymbol{\eta}} g||.\end{aligned} \tag{23}$$

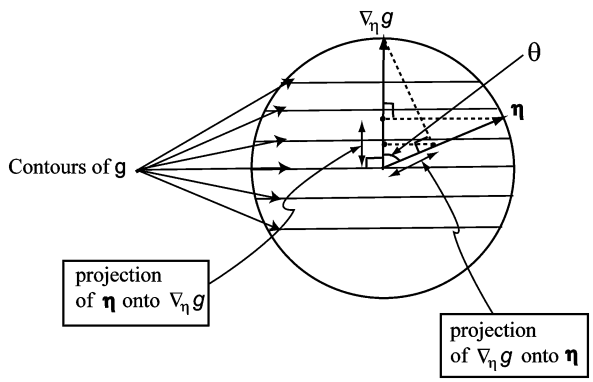



Fig. 3. The *consistency* condition. As in Fig. 2(F), suppose the contours of $g$ are projected onto the tangent plane of $g$, depicted here as a circle. The contours of $g$ on the tangent plane provide a means of measuring change or variation but this type of change is 1-D, that is perpendicular to the contours, i.e., parallel to $\nabla_{\boldsymbol{\eta}} g$. Without loss of generality, we will assume that both $\boldsymbol{\eta}$ and $\nabla_{\boldsymbol{\eta}} g$ are normalized to unit length and suppose that $\boldsymbol{\eta}$ is not parallel to the gradient of $g$. This implies the projections of $\boldsymbol{\eta}$ onto $\nabla_{\boldsymbol{\eta}} g$ and of $\nabla_{\boldsymbol{\eta}} g$ onto $\boldsymbol{\eta}$ no longer fall onto the same contours. Therefore, the change in $g$ cannot be measured *consistently* if $\boldsymbol{\eta}$ is not parallel to $\nabla_{\boldsymbol{\eta}} g$. If $\boldsymbol{\eta}$ is perpendicular to $\nabla_{\boldsymbol{\eta}} g$ then the change in $g$ is always zero by (18). Therefore, $\boldsymbol{\eta}$ must be parallel to $\nabla_{\boldsymbol{\eta}} g$.

[Fig. 2(F)]. By changing the variables from the $\boldsymbol{\eta}$-coordinate system back to the $\boldsymbol{\delta}(\boldsymbol{\gamma})$-coordinate system and squaring (23) [Fig. 2(D)–(F)], we arrive at the error propagation equation for DTI [51]

$$\Delta g(\boldsymbol{\delta})^2 \approx 2\Delta f(\boldsymbol{\delta}) \nabla_{\boldsymbol{\gamma}}^T g(\hat{\boldsymbol{\gamma}}) [\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1} \nabla_{\boldsymbol{\gamma}} g(\hat{\boldsymbol{\gamma}}). \quad (24)$$

The derivation of (24) is provided in Appendix IV.

Note that there is freedom in setting the magnitude of the change in $f_{\mathrm{NLS}}$, $\Delta f_{\mathrm{NLS}}(\boldsymbol{\delta})$. However, it is more meaningful to use the following definition $\Delta f_{\mathrm{NLS}}(\boldsymbol{\delta}) = f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})/(n-p)$ where $n-p$ is the number of degrees-of-freedom. Here, $p = 7$ for DTI, i.e., the number of tensor elements and one reference signal. This definition is meaningful because $2f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})/(n-7)$ is an unbiased estimate of the variance of the diffusion weighted (DW) signals [25], so that $\Delta g(\boldsymbol{\delta})^2$ can serve as an estimate of the variance of $g$. More importantly, if the change in $f_{\mathrm{NLS}}$ were to be taken as some multiple of the DW signal variance instead of one unit of the DW signal variance, then $\Delta g(\boldsymbol{\delta})^2$ would no longer be in agreement with the familiar notion of variance in statistics.

In subsequent discussion, we will denote $\sigma_{\mathrm{DW}}^2$ for $2f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})/(n-7)$. As an example, the variance of $\gamma_1 = \ln(\alpha)$ can be calculated by setting $g = \ln(\alpha)$ in (24), which yields $\sigma_{\ln(\alpha)}^2 \equiv \Delta \ln(\alpha)^2 \approx \sigma_{\mathrm{DW}}^2 \left[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})^{-1}\right]_{11}$. Similarly, for $\gamma_2 = D_{xx}$, $\sigma_{D_{xx}}^2 \equiv \Delta D_{xx}^2 \approx \sigma_{\mathrm{DW}}^2 \left[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})^{-1}\right]_{22}$.

We can also work with the objective functions $f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho}))$ and $f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\boldsymbol{\xi}))$ instead of $f_{\mathrm{NLS}}(\boldsymbol{\gamma})$ so that the variances of interest with respect to a particular representation can be computed without elaborate computation. However, it is important to realize that the Hessian matrix of the ordinary representation is fundamentally different from the Hessian matrices of the Euler and the Cholesky representations because the latter matrices do not transform like a tensor. Although a detailed discussion on tensor transformation laws is beyond the scope of this paper, we shall pursue along a different line by constructing covariance matrices of the Euler and the Cholesky representations based on the technique explicated in Section II-C. We will show that fundamental geometric objects in error propagation from which the covariance matrices are derived are the invariant Hessian matrices, and not the Hessian matrices. Briefly, an invariant Hessian matrix is defined to be the term in the Hessian matrix that is invariant with respect to coordinate transformations.

One of the goals in this paper is to show that with one condition—which is that the tensor estimate has to be positive definite [25], [47], [50]—separate minimizations of each objective function are unnecessary and the variances computed from one representation can also be obtained rather easily from another representation by a continuous coordinate transformation between the representations. Before we discuss the technique of coordinate transformation between representations, we will work on error propagation for vector functions and on practical variance or covariance computations of commonly used tensor-derived quantities in Sections II-C and II-D.

### C. Error Propagation Framework for Vector Functions

The discussion thus far has focused mainly on the proposed framework for any scalar function of $\boldsymbol{\gamma}$. Here, we will extend the framework to include vector functions so that quantities of interest such as the variance-covariance matrix of $\boldsymbol{\gamma}$ or of the major eigenvector of the diffusion tensor can be obtained. Without loss of generality, we will assume the vector function $\mathbf{g} = [g_1, g_2, g_3]^T$ consists of three scalar functions. By the first-order Taylor expansion of $g$ at $\hat{\boldsymbol{\gamma}}$, we have

$$\begin{aligned} \Delta g(\boldsymbol{\eta}) &= g(\hat{\boldsymbol{\gamma}} + \boldsymbol{\eta}) - g(\hat{\boldsymbol{\gamma}}) \\ &= \begin{pmatrix} g_1(\hat{\boldsymbol{\gamma}} + \boldsymbol{\eta}) - g_1(\hat{\boldsymbol{\gamma}}) \\ g_2(\hat{\boldsymbol{\gamma}} + \boldsymbol{\eta}) - g_2(\hat{\boldsymbol{\gamma}}) \\ g_3(\hat{\boldsymbol{\gamma}} + \boldsymbol{\eta}) - g_3(\hat{\boldsymbol{\gamma}}) \end{pmatrix} \\ &\approx \begin{pmatrix} \frac{\partial g_1}{\partial \eta_1} & \cdots & \frac{\partial g_1}{\partial \eta_p} \\ \frac{\partial g_2}{\partial \eta_1} & \cdots & \frac{\partial g_2}{\partial \eta_p} \\ \frac{\partial g_3}{\partial \eta_1} & \cdots & \frac{\partial g_3}{\partial \eta_p} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_p \end{pmatrix}. \end{aligned} \quad (25)$$

The variance-covariance matrix of $\mathbf{g}$ can be defined as (26)–(28), shown at the bottom of the page.

Under the consistency condition, there are three possibilities in choosing $\boldsymbol{\eta}$: $\boldsymbol{\eta} = \sigma_{\mathrm{DW}} \nabla_{\boldsymbol{\eta}} g_1 / \|\nabla_{\boldsymbol{\eta}} g_1\|$, $\boldsymbol{\eta} = \sigma_{\mathrm{DW}} \nabla_{\boldsymbol{\eta}} g_2 / \|\nabla_{\boldsymbol{\eta}} g_2\|$, and $\boldsymbol{\eta} = \sigma_{\mathrm{DW}} \nabla_{\boldsymbol{\eta}} g_3 / \|\nabla_{\boldsymbol{\eta}} g_3\|$. But the correct choice of $\boldsymbol{\eta}$ in each element of the matrix in (28) is again determined by the same *consistency* condition. This condition also ensures that the matrix in (28) is symmetric. As an example, let us consider the case of two distinct tangent planes, say of $g_1$ and of $g_2$. In this case, $\boldsymbol{\eta}$, which appears in the off-diagonal term, $\left(\nabla_{\boldsymbol{\eta}}^T g_1 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_2\right)$, of (28), can be taken to be either $\sigma_{\mathrm{DW}} \nabla_{\boldsymbol{\eta}} g_1 / \|\nabla_{\boldsymbol{\eta}} g_1\|$ or $\sigma_{\mathrm{DW}} \nabla_{\boldsymbol{\eta}} g_2 / \|\nabla_{\boldsymbol{\eta}} g_2\|$. It is important to note here that either one will yield the same result, which is $\nabla_{\boldsymbol{\eta}}^T g_1 \cdot \nabla_{\boldsymbol{\eta}} g_2$. In other words, the projection of $\nabla_{\boldsymbol{\eta}}^T g_1$ onto the tangent plane of $g_2$ or the projection of $\nabla_{\boldsymbol{\eta}}^T g_2$ onto the tangent plane of $g_1$ will yield the same covariation. Taking the *consistency* condition into account, the final expression of (28) can be written as

$$\boldsymbol{\Sigma}_g(\boldsymbol{\eta}) = \sigma_{\mathrm{DW}}^2 \begin{pmatrix} \nabla_{\boldsymbol{\eta}}^T g_1 \cdot \nabla_{\boldsymbol{\eta}} g_1 & \nabla_{\boldsymbol{\eta}}^T g_1 \cdot \nabla_{\boldsymbol{\eta}} g_2 & \nabla_{\boldsymbol{\eta}}^T g_1 \cdot \nabla_{\boldsymbol{\eta}} g_3 \\ \nabla_{\boldsymbol{\eta}}^T g_2 \cdot \nabla_{\boldsymbol{\eta}} g_1 & \nabla_{\boldsymbol{\eta}}^T g_2 \cdot \nabla_{\boldsymbol{\eta}} g_2 & \nabla_{\boldsymbol{\eta}}^T g_2 \cdot \nabla_{\boldsymbol{\eta}} g_3 \\ \nabla_{\boldsymbol{\eta}}^T g_3 \cdot \nabla_{\boldsymbol{\eta}} g_1 & \nabla_{\boldsymbol{\eta}}^T g_3 \cdot \nabla_{\boldsymbol{\eta}} g_2 & \nabla_{\boldsymbol{\eta}}^T g_3 \cdot \nabla_{\boldsymbol{\eta}} g_3 \end{pmatrix}. \quad (29)$$

or, equivalently

$$\lfloor \mathbf{\Sigma}_g(\boldsymbol{\eta}) \rfloor_{ij} \equiv \sigma_{ij}(\boldsymbol{\eta}) = \sigma_i(\boldsymbol{\eta})\sigma_j(\boldsymbol{\eta})\rho_{ij}(\boldsymbol{\eta}) \tag{30}$$

where $\sigma_i(\boldsymbol{\eta}) = \sigma_{\mathrm{DW}}\|\nabla_{\boldsymbol{\eta}} g_i\|$ and $\rho_{ij}(\boldsymbol{\eta}) = \left(\nabla_{\boldsymbol{\eta}}^T g_i / \|\nabla_{\boldsymbol{\eta}} g_i\|\right) \cdot \left(\nabla_{\boldsymbol{\eta}} g_j / \|\nabla_{\boldsymbol{\eta}} g_j\|\right)$ is the correlation coefficient. Note that $\left(\nabla_{\boldsymbol{\eta}}^T g_i / \|\nabla_{\boldsymbol{\eta}} g_i\|\right)$ is a unit vector parallel to $\nabla_{\boldsymbol{\eta}} g_i$. Finally, the variance-covariance matrix in the $\boldsymbol{\delta}(\boldsymbol{\gamma})$ coordinate system has the following expression:

$$\left[\mathbf{\Sigma}_{g(\hat{\boldsymbol{\gamma}})}\right]_{ij} = \sigma_{\mathrm{DW}}^2 \nabla_{\boldsymbol{\gamma}}^T g_i(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\gamma}} g_j(\hat{\boldsymbol{\gamma}}) \tag{31}$$

or

$$\lfloor \mathbf{\Sigma}_{g(\hat{\boldsymbol{\gamma}})} \rfloor_{ij} \equiv \sigma_{ij} = \sigma_i \sigma_j \rho_{ij} \tag{32}$$

where

$$\sigma_i = \sigma_{\mathrm{DW}} \sqrt{\nabla_{\boldsymbol{\gamma}}^T g_i(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\gamma}} g_i(\hat{\boldsymbol{\gamma}})} \tag{33}$$

and (34), shown at the bottom of the page, is the correlation coefficient. As an example, it can be shown that the variance-covariance matrix of $\boldsymbol{\gamma}$ can be expressed as $\Sigma_{\boldsymbol{\gamma}} = \sigma_{\mathrm{DW}}^2[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}$; see Appendix V for various Hessian structures and Appendix VI for the derivation of $\Sigma_{\boldsymbol{\gamma}} = \sigma_{\mathrm{DW}}^2[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}$.

The reader should be cautious not to be misled into thinking that the covariance matrix of the Euler representation, $\mathbf{\Sigma}_{\boldsymbol{\xi}}$, is simply $\sigma_{\mathrm{DW}}^2[\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))]^{-1}$. These two quantities are closely related but they are not equivalent. As mentioned earlier, the Hessian matrix of the Euler representation is not a tensor. This means that its inverse will not be invariant with respect to coordinate transformations. In Appendix VI, it is shown that the covariance matrix of the Euler representation, $\mathbf{\Sigma}_{\boldsymbol{\xi}}$, is equal to $\sigma_{\mathrm{DW}}^2[\mathrm{inv}(\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})))]^{-1}$, where $\mathrm{inv}(\nabla^2 f)$ denotes the invariant Hessian matrix of $f$, which is the part of the Hessian matrix that is invariant with respect to coordinate transformations. It is noteworthy that we can discover these invariant Hessian structures within the proposed framework using the technique discussed in this section, see Appendix VI.

### D. Scalar and Vector Functions of the Diffusion Tensor

As mentioned earlier, variance computation for certain tensor-derived quantities can be greatly simplified by using the appropriate diffusion tensor representation. The most commonly used tensor-derived quantities are listed below [2], [5]

1) Trace:

$$\mathrm{Trace} = \mathrm{Tr}(\mathbf{D}) = D_{xx} + D_{yy} + D_{zz}. \tag{35}$$

2) Fractional Anisotropy:

$$\mathrm{FA} = \sqrt{\frac{3}{2}\left[1 - \frac{1}{3}\frac{\mathrm{Tr}(D)^2}{\mathrm{Tr}(D^2)}\right]}$$

or

$$\mathrm{FA} = \frac{1}{\sqrt{2}}\left(\frac{(\lambda_1-\lambda_2)^2 + (\lambda_2-\lambda_3)^2 + (\lambda_3-\lambda_1)^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}\right)^{1/2}. \tag{36}$$

3) Relative Anisotropy:

$$\mathrm{RA} = \sqrt{\frac{3}{2}\left[\frac{\mathrm{Tr}(D^2)}{\mathrm{Tr}(D)^2} - \frac{1}{3}\right]}$$

or

$$\mathrm{RA} = \frac{1}{\sqrt{2}}\frac{\left((\lambda_1-\lambda_2)^2 + (\lambda_2-\lambda_3)^2 + (\lambda_3-\lambda_1)^2\right)^{1/2}}{\lambda_1+\lambda_2+\lambda_3}. \tag{37}$$

4) Eigenvalues of $\mathbf{D}$:

$$\lambda_1, \lambda_2, \text{ and } \lambda_3. \tag{38}$$

as defined in (9) and (14).

5) Eigenvalues of $\mathbf{D}$:
$\mathbf{q}_1, \mathbf{q}_2,$ and $\mathbf{q}_3$ as defined in (9), and (11) -(13). (39)

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{g}}(\boldsymbol{\eta}) &\equiv \Delta\mathbf{g}(\boldsymbol{\eta}) \cdot \Delta\mathbf{g}(\boldsymbol{\eta})^T && (26)\\
&= \begin{pmatrix} \frac{\partial g_1}{\partial \eta_1} & \cdots & \frac{\partial g_1}{\partial \eta_p} \\ \frac{\partial g_2}{\partial \eta_1} & \cdots & \frac{\partial g_2}{\partial \eta_p} \\ \frac{\partial g_3}{\partial \eta_1} & \cdots & \frac{\partial g_3}{\partial \eta_p} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_p \end{pmatrix} \begin{pmatrix} \eta_1 & \cdots & \eta_p \end{pmatrix} \begin{pmatrix} \frac{\partial g_1}{\partial \eta_1} & \frac{\partial g_2}{\partial \eta_1} & \frac{\partial g_3}{\partial \eta_1} \\ \vdots & \vdots & \vdots \\ \frac{\partial g_1}{\partial \eta_p} & \frac{\partial g_2}{\partial \eta_p} & \frac{\partial g_3}{\partial \eta_p} \end{pmatrix} && (27)\\
&= \begin{pmatrix} \left(\nabla_{\boldsymbol{\eta}}^T g_1 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_1\right) & \left(\nabla_{\boldsymbol{\eta}}^T g_1 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_2\right) & \left(\nabla_{\boldsymbol{\eta}}^T g_1 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_3\right) \\ \left(\nabla_{\boldsymbol{\eta}}^T g_2 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_1\right) & \left(\nabla_{\boldsymbol{\eta}}^T g_2 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_2\right) & \left(\nabla_{\boldsymbol{\eta}}^T g_2 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_3\right) \\ \left(\nabla_{\boldsymbol{\eta}}^T g_3 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_1\right) & \left(\nabla_{\boldsymbol{\eta}}^T g_3 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_2\right) & \left(\nabla_{\boldsymbol{\eta}}^T g_3 \cdot \boldsymbol{\eta}\right)\left(\boldsymbol{\eta}^T \cdot \nabla_{\boldsymbol{\eta}} g_3\right) \end{pmatrix}. && (28)
\end{aligned}$$

$$\rho_{ij} = \frac{\nabla_{\boldsymbol{\gamma}}^T g_i(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\gamma}} g_j(\hat{\boldsymbol{\gamma}})}{\sqrt{\nabla_{\boldsymbol{\gamma}}^T g_i(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\gamma}} g_i(\hat{\boldsymbol{\gamma}})}\sqrt{\nabla_{\boldsymbol{\gamma}}^T g_j(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\gamma}} g_j(\hat{\boldsymbol{\gamma}})}} \tag{34}$$

It should be noted here that the first component in each representation, $\boldsymbol{\gamma}$, $\boldsymbol{\rho}$, and $\boldsymbol{\xi}$, is $\ln(\alpha)$ and, therefore, the partial derivative of any tensor-derived quantity with respect to $\ln(\alpha)$ is 0.

Since $\mathrm{Tr}(\mathbf{D}) = \mathrm{Tr}(\boldsymbol{\Lambda})$ and $\mathrm{Tr}(\mathbf{D}^2) = \mathrm{Tr}(\boldsymbol{\Lambda}^2)$, it is clear that the variance computation for (35)–(37) is equally tractable in both the ordinary and the Euler representations. However, the variance-covariance computation of (38) and (39) is most convenient in the Euler representation.

The formulas listed below are the gradients of the most commonly used tensor-derived quantities; the first three formulas are expressed with respect to both the ordinary representation and the Euler representation, while the last two are expressed with respect to the Euler representation only.

1) Gradient of Trace:

$$\nabla_{\boldsymbol{\gamma}}\mathrm{Tr} = [0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0]^T$$

or

$$\nabla_{\boldsymbol{\xi}}\mathrm{Tr} = [0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0]^T. \tag{40}$$

2) Gradient FA: Let $a = (\lambda_1-\lambda_2)^2+(\lambda_2-\lambda_3)^2+(\lambda_3-\lambda_1)^2$ and $b = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$ then $\mathrm{FA} = (1/\sqrt{2})\,(a/b)^{1/2}$. The gradient of FA with respect to the Euler representation is $\nabla_{\boldsymbol{\xi}}\mathrm{FA}= [0 \quad \partial\mathrm{FA}/\partial\lambda_1 \quad \partial\mathrm{FA}/\partial\lambda_2 \quad \partial\mathrm{FA}/\partial\lambda_3 \quad 0 \quad 0 \quad 0]^T$, where

$$\frac{\partial \mathrm{FA}}{\partial \lambda_1} = -\frac{\lambda_1}{b}\mathrm{FA} + \frac{(\lambda_1-\lambda_2)+(\lambda_1-\lambda_3)}{\sqrt{2ab}}$$
$$\frac{\partial \mathrm{FA}}{\partial \lambda_2} = -\frac{\lambda_2}{b}\mathrm{FA} + \frac{(\lambda_2-\lambda_1)+(\lambda_2-\lambda_3)}{\sqrt{2ab}}$$

and

$$\frac{\partial \mathrm{FA}}{\partial \lambda_3} = -\frac{\lambda_3}{b}\mathrm{FA} + \frac{(\lambda_3-\lambda_2)+(\lambda_3-\lambda_1)}{\sqrt{2ab}}.$$

The gradient of FA with respect to the ordinary representation has the following components:

$$\frac{\partial \mathrm{FA}}{\partial D_{ij}} = \frac{1}{\mathrm{FA}}\frac{\mathrm{Tr}(D)}{\mathrm{Tr}(D^2)}\left(\left[\frac{1}{1+\delta_{ij}}\frac{\mathrm{Tr}(D)}{\mathrm{Tr}(D^2)}D_{ij}\right] - \frac{\delta_{ij}}{2}\right). \tag{41}$$

3) Gradient of RA: Let $\mathrm{Tr} = \lambda_1 + \lambda_2 + \lambda_3$, the gradient of RA with respect to the Euler representation is $\nabla_{\boldsymbol{\xi}}RA= [0 \quad \partial RA/\partial\lambda_1 \quad \partial RA/\partial\lambda_2 \quad \partial RA/\partial\lambda_3 \quad 0 \quad 0 \quad 0]^T$, where

$$\frac{\partial RA}{\partial \lambda_1} = -\frac{RA}{\mathrm{Tr}} + \frac{(\lambda_1-\lambda_2)+(\lambda_1-\lambda_3)}{2RATr^2}$$
$$\frac{\partial RA}{\partial \lambda_2} = -\frac{RA}{\mathrm{Tr}} + \frac{(\lambda_2-\lambda_1)+(\lambda_2-\lambda_3)}{2RATr^2}$$

and

$$\frac{\partial RA}{\partial \lambda_3} = -\frac{RA}{\mathrm{Tr}} + \frac{(\lambda_3-\lambda_1)+(\lambda_3-\lambda_2)}{2RATr^2}.$$

The gradient of RA with respect to the ordinary representation has the following components:

$$\frac{\partial RA}{\partial D_{ij}} = \frac{3}{RA}\frac{\mathrm{Tr}(\mathbf{D}^2)}{\mathrm{Tr}(\mathbf{D})^3}\left(\left[\frac{1}{1+\delta_{ij}}\frac{\mathrm{Tr}(\mathbf{D})}{\mathrm{Tr}(\mathbf{D}^2)}D_{ij}\right] - \frac{\delta_{ij}}{2}\right). \tag{42}$$

4) Gradients of the eigenvalues are $[\nabla_{\boldsymbol{\xi}}\lambda_1]_i = \delta_{2i}$

$$[\nabla_{\boldsymbol{\xi}}\lambda_2]_i = \delta_{3i} \text{ and } [\nabla_{\boldsymbol{\xi}}\lambda_3]_i = \delta_{4i}. \tag{43}$$

5) Gradients of a component of an eigenvector: i.e., $\mathbf{q}_1 = [Q_{11} \quad Q_{21} \quad Q_{31}]^T$. Since the expressions are more involved, we have collected these formulas in Appendix VII.

Some of the preliminary formulas used to derive (40)–(42) are collected in Appendix VIII.

### *E. Coordinate Transformation Between Different Tensor Representations*

As discussed in Section II-C and Appendix VI, invariant Hessian structures are very important to variance-covariance computation. Particularly, we have used the technique explicated in Section II-C to derive the invariant Hessian structures of the ordinary and Euler representations in Appendix VI. For convenience, these structures are explicitly given here

$$\mathrm{inv}(\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma})) = \nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W} \tag{44}$$

$$\mathrm{inv}(\nabla^2 f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho}))) = \mathbf{J}_{\boldsymbol{\rho}}^T(\boldsymbol{\gamma})\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}) \tag{45}$$

$$\mathrm{inv}(\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\boldsymbol{\xi}))) = \mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma})\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}) \tag{46}$$

where the invariant Hessian matrix of $f$ is denoted by $\mathrm{inv}(\nabla^2 f)$. Please refer to Appendix V for the terms defined above.

We have previously mentioned why the expression, $\sigma_{\mathrm{DW}}^2[\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))]^{-1}$, should not be taken as the definition of the covariance matrix of the Euler representation. On intuitive ground, variance or covariance of a quantity should be invariant with respect to coordinate transformations, or equivalently, it can be said that variance or covariance of a quantity should transform like a tensor. Here, we will show how the invariance property of the covariance matrix is violated if one insists on using $\sigma_{\mathrm{DW}}^2[\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))]^{-1}$ as the covariance matrix of the Euler representation. According to (31), we need to construct $\nabla_{\boldsymbol{\xi}}\gamma_j(\boldsymbol{\xi})$ so that $\left[\boldsymbol{\Sigma}_{\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})}\right]_{ij} = \nabla_{\boldsymbol{\xi}}^T\gamma_i(\hat{\boldsymbol{\xi}})\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\nabla_{\boldsymbol{\xi}}\gamma_j(\hat{\boldsymbol{\xi}})$ or

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})} &= \left[\nabla_{\boldsymbol{\xi}}\gamma_1(\hat{\boldsymbol{\xi}}) \quad \ldots \quad \nabla_{\boldsymbol{\xi}}\gamma_j(\hat{\boldsymbol{\xi}})\right]^T \\ &\quad \times \boldsymbol{\Sigma}_{\boldsymbol{\xi}}\left[\nabla_{\boldsymbol{\xi}}\gamma_1(\hat{\boldsymbol{\xi}}) \quad \ldots \quad \nabla_{\boldsymbol{\xi}}\gamma_j(\hat{\boldsymbol{\xi}})\right] \\ &= \mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})).\end{aligned} \tag{47}$$

Since the covariance structure should be invariant with respect to coordinate transformation, one expects $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})}$ in (47) to be equal to $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \sigma_{\mathrm{DW}}^2[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}$. However, the invariance property is violated if one substitutes the following expression:

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\xi}} &= \sigma_{\mathrm{DW}}^2[\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))]^{-1} \\ &= \sigma_{\mathrm{DW}}^2\Bigg(\mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}}) \\ &\qquad \times \mathbf{W}\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) + \sum_{i=1}^{n} r_i\hat{s}_i\mathbf{T}_i\Bigg)^{-1}\end{aligned}$$

into (47)

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})} &= \sigma_{\mathrm{DW}}^2 \mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) \\ &\quad \times \left( \mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}}) \right. \\ &\qquad \left. \times \mathbf{W}\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) + \sum_{i=1}^{n} r_i \hat{s}_i \mathbf{T}_i \right)^{-1} \mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) \\ &= \sigma_{\mathrm{DW}}^2 \left[ \left( \mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) \right)^{-1} \left( \mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}}) \right. \right. \\ &\qquad \left. \times \mathbf{W}\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) + \sum_{i=1}^{n} r_i \hat{s}_i \mathbf{T}_i \right) \\ &\qquad \left. \times \left( \mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) \right)^{-1} \right]^{-1} \\ &= \sigma_{\mathrm{DW}}^2 \left[ \left( \nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) + \sum_{i=1}^{n} r_i \hat{s}_i \left[ \left( \mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) \right)^{-1} \right. \right. \right. \\ &\qquad \left. \left. \left. \times \mathbf{T}_i \left( \mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) \right)^{-1} \right] \right) \right]^{-1}. \end{aligned} \tag{48}$$

Comparing (48) and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \sigma_{\mathrm{DW}}^2 [\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}$, we see that the additional error introduced in the estimation of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})}$ violates the invariance property of the covariance matrix.

In brief, the covariance matrices in various tensor representations are derived from the invariant Hessian structures and their expressions are given below

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} &= \sigma_{\mathrm{DW}}^2 [\mathrm{inv}(\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}))]^{-1} \\ &= \sigma_{\mathrm{DW}}^2 [\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}]^{-1} \end{aligned} \tag{49}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\rho}} &= \sigma_{\mathrm{DW}}^2 [\mathrm{inv}(\nabla^2 f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho})))]^{-1} \\ &= \sigma_{\mathrm{DW}}^2 [\mathbf{J}_{\boldsymbol{\rho}}^T(\boldsymbol{\gamma})\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma})]^{-1} \end{aligned} \tag{50}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\xi}} &= \sigma_{\mathrm{DW}}^2 [\mathrm{inv}(\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})))]^{-1} \\ &= \sigma_{\mathrm{DW}}^2 [\mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))]^{-1}. \end{aligned} \tag{51}$$

### *F. Applications*

*1) Average Variance-Covariance Matrix:* The average variance-covariance matrix for a given diffusion tensor is a very useful quantity to compute in a simulation study; it is directly related to average DW signals where the estimated signals are assumed to be fitted perfectly to the observed signals, i.e., the residuals are zero. One can see then that the average variance-covariance matrices can be easily derived from (49)–(51), and are given by

$$\langle \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \rangle = \sigma_{\mathrm{DW}}^2 [\mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W}]^{-1} \tag{52}$$

$$\langle \boldsymbol{\Sigma}_{\boldsymbol{\rho}} \rangle = \sigma_{\mathrm{DW}}^2 [\mathbf{J}_{\boldsymbol{\rho}}^T(\boldsymbol{\gamma})\mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W}\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma})]^{-1} \tag{53}$$

$$\langle \boldsymbol{\Sigma}_{\boldsymbol{\xi}} \rangle = \sigma_{\mathrm{DW}}^2 [\mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma})\mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W}\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma})]^{-1}. \tag{54}$$

The symbol, $\langle \boldsymbol{\Sigma} \rangle$, represents the average quantity of $\boldsymbol{\Sigma}$. The method of averaging and the derivations of (52)–(54) are discussed in Appendix IX.

It should be noted here that $\sigma_{\mathrm{DW}}^2$ has to be defined differently from the previous definition, which was based on the estimated DW signal variance, because the residual sum of squares is now assumed to be zero. Therefore, $\sigma_{\mathrm{DW}}^2$ has to be taken from a known variance with respect to the Rician-distributed DW signals. The technique on transforming the variance with respect to the Gaussian-distributed complex signals to the variance with respect to Rician-distributed magnitude signals at a prescribed level of signal-to-noise ratio (SNR) can be found in Koay *et al.* [25], [52]. For $\mathrm{SNR} > 5$, $\sigma_{\mathrm{DW}}^2$ is an acceptable approximation to $\sigma_{\mathrm{Gaussian}}^2$; $\sigma_{\mathrm{Gaussian}}$ represents the standard deviation of the Gaussian-distributed complex signals.

Once the average covariance matrices with respect to various tensor representations are known, the mean variance of any tensor-derived quantity or the mean variance-covariance between any two tensor-derived quantities can then be computed based on the techniques explained in the preceding sections. As an example, the mean variance of $D_{xx}$ can be expressed as

$$\sigma_{D_{xx}}^2 \equiv \Delta D_{xx}^2 \approx \sigma_{\mathrm{DW}}^2 \left[ \left( \mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W} \right)^{-1} \right]_{22}.$$

Here, we use this approach to show the rotational variance of Tr of a prolate tensor based on the variance-covariance matrix in (52), Fig. 4. This framework will also be useful in analyzing the effect of gradient sampling schemes on tensor-derived quantities without the need for a computationally intensive bootstrap to quantify uncertainty, see Jones [53]. It is clear that the variance of trace exhibits rotational asymmetry, Fig. 4. Increasing the number of gradient directions will not reduce the systematic variation, Fig. 4. The theoretical reason for this phenomenon is that the experimental design for DTI is not rotationally invariant [54].

*2) Elliptical Cones of Uncertainty of the Principal Eigenvectors:* Based on the technique expounded in Sections II-C and D, the variance-covariance matrix of the components of an eigenvector can be computed quite easily. This particular variance-covariance matrix is useful in constructing the elliptical cone of uncertainty about that eigenvector.

Without loss of generality, we shall take the major vector of a diffusion tensor to illustrate the method in this section. By (11), $\boldsymbol{q}_1 = [\, Q_{11} \quad Q_{21} \quad Q_{31} \,]^T$, and (31), we have

$$\left[ \boldsymbol{\Sigma}_{\boldsymbol{q}_1} \right]_{ij} = \nabla_{\boldsymbol{\xi}}^T Q_{i1}(\hat{\boldsymbol{\gamma}}) \boldsymbol{\Sigma}_{\boldsymbol{\xi}} \nabla_{\boldsymbol{\xi}} Q_{j1}(\hat{\boldsymbol{\gamma}}). \tag{55}$$

According to the perturbation method proposed by Hext [45], $\mathbf{q}_1$ is normal to the plane of the elliptical cone of uncertainty. In other words, the eigenvector that is associated with the smallest eigenvalue of $\boldsymbol{\Sigma}_{\boldsymbol{q}_1}$ is parallel to $\mathbf{q}_1$, therefore, the other two eigenvectors are perpendicular to $\mathbf{q}_1$. The same observation can be made within the proposed framework. That is, the equation of a sphere $Q_{11}^2 + Q_{21}^2 + Q_{31}^2 = 1$, will force $\boldsymbol{\Sigma}_{\boldsymbol{q}_1}$ to be a matrix of rank 2, therefore, the smallest eigenvalue of $\boldsymbol{\Sigma}_{\boldsymbol{q}_1}$ is essentially zero. Another argument for this observation is based on the dyadics formulation; it is presented in Appendix X. We shall outline the basic idea with an example. If we have $\boldsymbol{\gamma}$, as shown

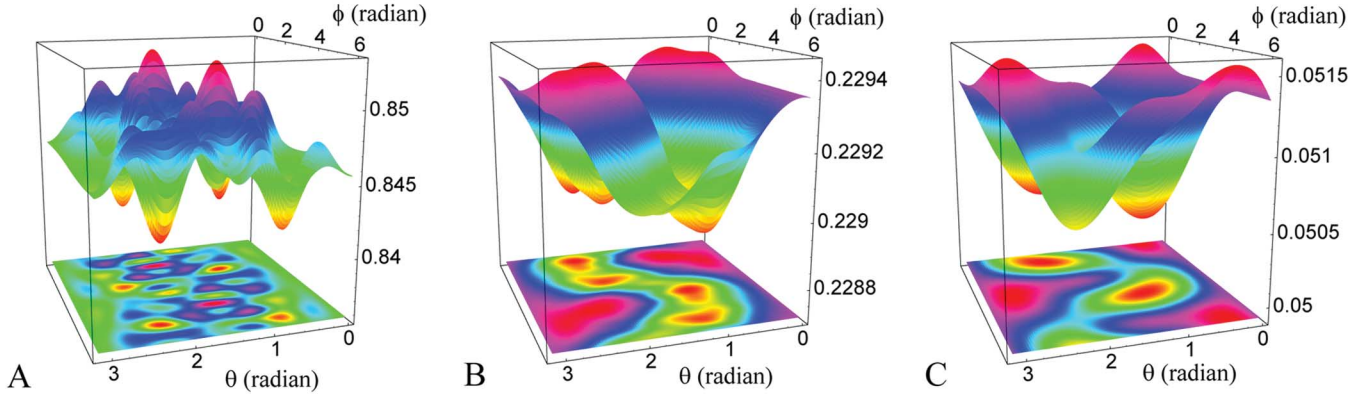



Fig. 4. Rotational Asymmetry in the variance of Trace for a prolate tensor. Generally, the rotation of a typical tensor requires three parameters, i.e., Euler angles. But, analysis of rotational asymmetry of any tensor-derived quantity can be studied using a prolate tensor where only two parameters are sufficient, i.e., the major eigenvector of the prolate tensor can be parametrized by $[\sin(\theta)\cos(\varphi) \quad \sin(\theta)\sin(\varphi) \quad \cos(\theta)]^T$ with $0 \leq \theta < \pi$ and $0 \leq \varphi < 2\pi$. The plots above are computed with a prolate tensor having FA of 0.586 and eigenvalues of $\lambda_1 = 1.24 \times 10^{-3}\ \mathrm{mm}^2/\mathrm{s}$ and $\lambda_2 = \lambda_3 = 4.30 \times 10^{-4}\ \mathrm{mm}^2/\mathrm{s}$ at $\mathrm{SNR} = 25$. Images A–C were computed with different numbers of gradient directions: 23, 85, and 382, respectively. In each plot, the final design matrix $\mathbf{W}$ was constructed from four spherical shells having b values of 0, 500, 1000, and 1500 $\mathrm{s/mm}^2$. The color-coded variation is specific to each plot but the numerical scale, which has been normalized to the unit interval, $[0, 1]$, from $\lfloor 0, 2.0 \times 10^{-9}\ \mathrm{mm}^4/\mathrm{s}^2 \rfloor$, is common to all.

in the equation at the bottom of the page, then the major eigenvector is $\mathbf{q}_1 = [0.9027 \quad 0.3139 \quad -0.2940]^T$ and the major eigenvalue is $0.00114\ \mathrm{mm}^2/\mathrm{s}$.

We shall denote the lower right $3 \times 3$ submatrix of $\langle \mathbf{\Sigma}_{\boldsymbol{\xi}} \rangle$ as $\lfloor \langle \mathbf{\Sigma}_{\boldsymbol{\xi}} \rangle \rfloor_{5:7,5:7}$ and

$$[\langle \mathbf{\Sigma}_{\boldsymbol{\xi}} \rangle]_{5:7,5:7} = \begin{pmatrix} 10.271 & -14.098 & 13.473 \\ -14.098 & 604.232 & -575.827 \\ 13.473 & -575.827 & 579.015 \end{pmatrix} \times 10^{-5}$$

based on the SNR level of 50 and on a design matrix, $\mathbf{W}$, that was constructed from a 35 gradient direction set with four spherical shells having $b$ values of 0, 500, 1000, and 1500 $\mathrm{s/mm}^2$.

Similarly, we shall denote the lower $3 \times 3$ submatrix of $\nabla_{\boldsymbol{\xi}} \mathbf{q}_1 \equiv [\nabla_{\boldsymbol{\xi}} Q_{11}(\hat{\boldsymbol{\gamma}}), \nabla_{\boldsymbol{\xi}} Q_{21}(\hat{\boldsymbol{\gamma}}), \nabla_{\boldsymbol{\xi}} Q_{31}(\hat{\boldsymbol{\gamma}})]$ as $[\nabla_{\boldsymbol{\xi}} \mathbf{q}_1]_{5:7,1:3}$, particularly, we have

$$[\nabla_{\boldsymbol{\xi}} \mathbf{q}_1]_{5:7,1:3} = \begin{pmatrix} -0.2863 & -0.0670 & -0.9505 \\ -0.3139 & 0.9027 & 0.0 \\ -0.3197 & 0.9470 & 0.0295 \end{pmatrix}$$

for our example.

The variance-covariance matrix can then be computed as follows:

$$\mathbf{\Sigma}_{\mathbf{q}_1} = \left( [\nabla_{\boldsymbol{\xi}} \mathbf{q}_1]_{5:7,1:3} \right)^T [\langle \mathbf{\Sigma}_{\boldsymbol{\xi}} \rangle]_{5:7,5:7} \left( [\nabla_{\boldsymbol{\xi}} \mathbf{q}_1]_{5:7,1:3} \right)$$

which has the following numerical values

$$\begin{pmatrix} 3.9182 & -8.9822 & 2.4405 \\ -8.9822 & 27.179 & 1.4387 \\ 2.4405 & 1.4387 & 9.0289 \end{pmatrix} \times 10^{-5}.$$

The eigenvalue–eigenvector pairs of this matrix are

$$\left\{ 3.0259 \times 10^{-4}, [0.32035, \quad -0.9469, \quad -0.0273]^T \right\}$$
$$\left\{ 9.8667 \times 10^{-5}, [0.2870 \quad 0.0695 \quad 0.9554]^T \right\}$$

and

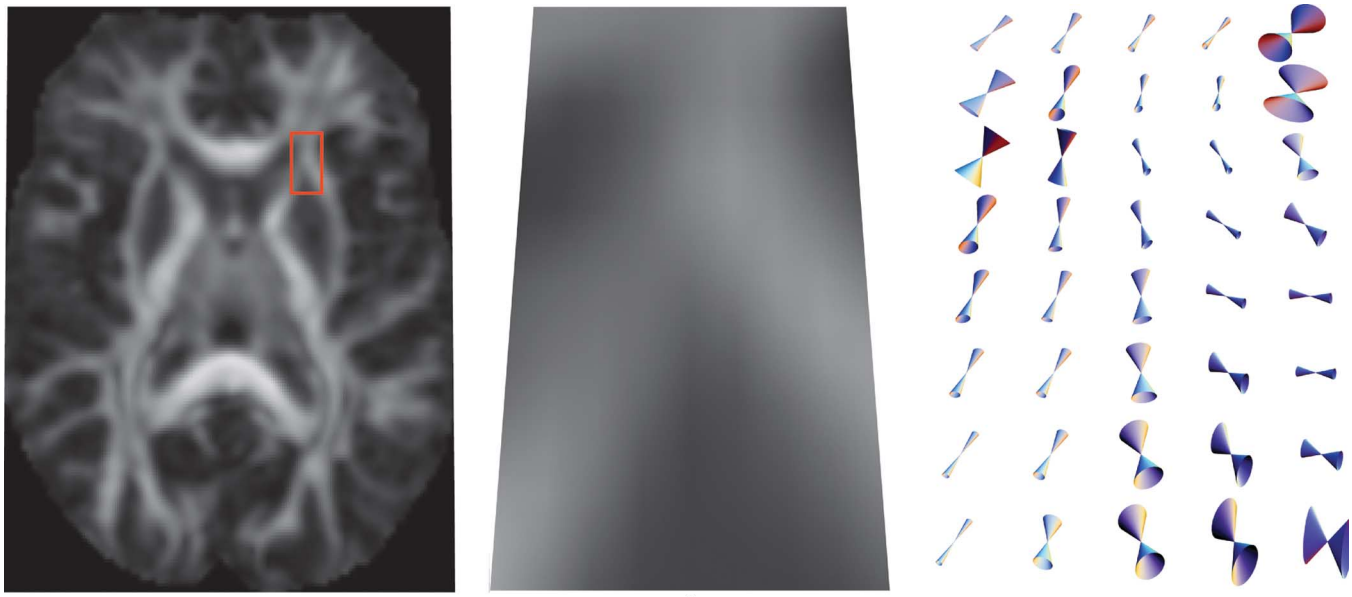
$$\left\{ 0 \cong 2.4915 \times 10^{-20}, [0.9027 \quad 0.3139 \quad -0.2940]^T \right\}.$$

It is quite clear then that $\mathbf{q}_1$ is parallel to the minor eigenvector of $\mathbf{\Sigma}_{\mathbf{q}_1}$. Note that the other two eigenvectors of $\mathbf{\Sigma}_{\mathbf{q}_1}$ are not generally equal to the medium and minor eigenvectors of the diffusion tensor. Once the eigenvalue–eigenvector pairs of $\mathbf{\Sigma}_{\mathbf{q}_1}$ and $\mathbf{q}_1$ are computed, the $100(1-\beta)\%$ elliptical confidence cone can be constructed quite easily. We shall mention here a simple but important method for visualizing the confidence cone. We prefer to use the approach proposed by Hext [45] in which the confidence cone is projected onto the unit sphere, thus avoiding an important visual ambiguity: if the height of a confidence cone were to be scaled proportional to some function of the major eigenvalue, then the spread of the cone would be a function not only of the two nonzero eigenvalues of $\mathbf{\Sigma}_{\mathbf{q}_1}$ but also of the major eigenvalue of the diffusion tensor. It would then be harder to compare two neighboring confidence cones visually. Fig. 5 shows an example of the elliptical confidence cones constructed from the human brain data.

## III. Results

A variance-covariance estimate can be obtained from a set of DW measurements. Therefore, repeated DW measurements can

$$\boldsymbol{\gamma} = [\ln(1000.0) \times 10^{+4}\ \tfrac{\mathrm{s}}{\mathrm{mm}^2} \quad 10.208 \quad 6.7889 \quad 4.0029 \quad 1.3871 \quad -0.66383 \quad 2.1784]^T \times 10^{-4}\ \frac{\mathrm{mm}^2}{\mathrm{s}}$$



(a) (b) (c)

Fig. 5. Elliptical confidence cones. (A) FA map. (B) Magnified image of the region bounded by a red square on the FA map. (C) Corresponding elliptical 95% confidence cones on that region at SNR level of 15.

be carried out to measure the uncertainty of the variance-covariance estimate by a graphical method based on histogram analysis. This approach will provide a reasonable measure of the distributional properties of these estimates. Further, the classical sample variance-covariance formulas can be employed to compare with the analytically derived value of these estimates. Monte Carlo simulations similar to those of Pierpaoli and Basser [5] were carried out to validate the proposed method.

For simplicity, we shall use the simulation condition (including the parameter vector, $\boldsymbol{\gamma}$) similar to that of Section II-F2 except at a single SNR level of 15. Further, 50 000 repeated measurements were generated to facilitate statistical comparison. Briefly, this parameter vector has Tr of $0.0021\ \mathrm{mm}^2/\mathrm{s}$ and FA of 0.5278. Further, its major eigenvector is $[0.9027 \quad 0.3139 \quad -0.2940]^T$.

The sample statistics, and the results from the proposed framework with respect to two different covariance matrices, $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ are listed in Table I. The sample statistics, listed as (I) in Table I, are computed based on classical statistical expressions for sample mean and sample variance. Similarly, the sample covariance matrix of the major eigenvector is based on classical statistics but the sample eigenvectors, $\{\boldsymbol{q}_{i1}, \ldots, \boldsymbol{q}_{iN}\}$ for $i = 1, 2, 3$, have to be properly oriented so that their directions are on the same hemisphere as the estimated mean major eigenvector. The estimated mean major eigenvector is computed based on the dyadic product formulation [27] where the major eigenvector of $\langle \boldsymbol{q}_1 \boldsymbol{q}_1^T \rangle = 1/N \sum_{j=1}^{N} \boldsymbol{q}_{1j} \boldsymbol{q}_{1j}^T$ corresponds to the mean major eigenvector. An important observation about this dyadic product is that the medium and the minor eigenvectors of $\langle \boldsymbol{q}_1 \boldsymbol{q}_1^T \rangle$ can be used to construct the covariance matrix of the major eigenvector. The argument for this observation is presented in Appendix X. The results on (II) and (III) are obtained from the average covariance matrix discussed in Section II-F1. The results on (IV) and (V) are obtained by averaging the 50 000 variance estimates of Tr and FA; these variance estimates, (IV) and (V), are obtained from the proposed framework with respect to the ordinary and the Euler representations, respectively. Further, the DW signal variances were estimated from each nonlinear fit using the modified full Newton method described in [25]. To complement the results in Table I, we also show the distributional property of these variance estimates in Fig. 6.

TABLE I
SIMULATION RESULTS BASED ON VARIOUS METHODS DISCUSSED IN THIS PAPER

| | (I) Sample statistics | (II) Expected value based on $<\Sigma_\gamma>$ | (III) Expected value based on $<\Sigma_\xi>$ | (IV) Sample statistics based on $\Sigma_\gamma$ | (V) Sample statistics based on $\Sigma_\xi$ |
|---|---|---|---|---|---|
| Variance of Trace | $2.9331\times10^{-9}$ | $3.0331\times10^{-9}$ | $3.0331\times10^{-9}$ | $2.9529\times10^{-9}$ | $2.9529\times10^{-9}$ |
| Variance of FA | $8.0014\times10^{-4}$ | $8.1162\times10^{-4}$ | $8.1162\times10^{-4}$ | $8.1402\times10^{-4}$ | $8.1402\times10^{-4}$ |

Before presenting the results on the covariance matrix of the major eigenvector, we will show some results on the dyadics formalism for later comparison. The average dyadics from the 50 000 samples of eigenvectors turns out to be

$$\begin{pmatrix} 0.8116 & 0.2813 & -0.2639 \\ 0.2813 & 0.1014 & -0.0918 \\ -0.2639 & -0.0918 & 0.0871 \end{pmatrix}$$

and the corresponding eigenvalue–eigenvector pairs are

$$\{\lambda_1, \boldsymbol{\Psi}_1\} = \{9.954 \times 10^{-1}, [-0.9027 \quad -0.3142 \quad 0.2941]^T\}$$
$$\{\lambda_2, \boldsymbol{\Psi}_2\} = \{3.452 \times 10^{-3}, [-0.3210 \quad 0.9467 \quad 0.0260]^T\}$$

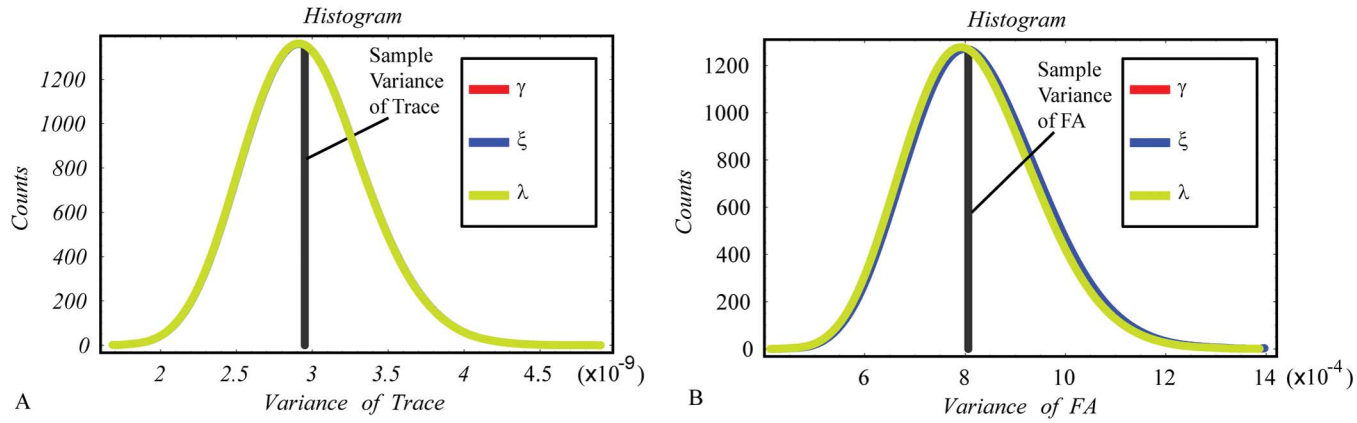



Fig. 6. Histograms of the variance estimates of (A) trace and of (B) FA based on three different covariance matrices: $\mathbf{\Sigma_\gamma}$ (red), $\mathbf{\Sigma_\xi}$ (blue), and $\mathbf{\Sigma_\lambda}$ (green). The construction of $\mathbf{\Sigma_\lambda}$ is discussed in Appendix III and it is related to the Hext representation. Note that on Fig. 6(A) and (B), the lines are superimposed. Sample variance of trace (FA), which is computed from the 50 000 trace (FA) estimates, is shown in Fig. 6(A) and (B) as a vertical line.

and

$$\{\lambda_3, \mathbf{\Psi}_3\} = \{1.123 \times 10^{-3}, \quad [0.2865 \quad 0.0709 \quad 0.9555]^T\}$$

The average vector before and after normalization is $\mathbf{t} = 1/N \sum_{i=1}^{N} \mathbf{q}_{1i} = \{0.9006, 0.3135, -0.2933\}$, and $\hat{\mathbf{t}} = \{0.9027, 0.3142, -0.2940\}$, respectively, with a vector norm of $\|\mathbf{t}\| = 0.9977$ and $(\lambda_1 + 1 - 2\|\mathbf{t}\|) \approx 1.2038 \times 10^{-5}$; $(\lambda_1 + 1 - 2\|\mathbf{t}\|)$ is an approximation to the minor eigenvalue of the covariance matrix of the major eigenvector of the diffusion tensor, see Appendix X.

Here, we present the results on the covariance matrices of the major eigenvector

$$\begin{pmatrix} 4.583 \times 10^{-4} & -1.024 \times 10^{-3} & 2.751 \times 10^{-4} \\ -1.024 \times 10^{-3} & 3.100 \times 10^{-3} & 1.602 \times 10^{-4} \\ 2.751 \times 10^{-4} & 1.602 \times 10^{-4} & 1.029 \times 10^{-3} \end{pmatrix}$$
$$\begin{pmatrix} 4.844 \times 10^{-4} & -1.109 \times 10^{-3} & 3.035 \times 10^{-4} \\ -1.109 \times 10^{-3} & 3.256 \times 10^{-3} & 7.219 \times 10^{-5} \\ 3.035 \times 10^{-4} & 7.219 \times 10^{-5} & 1.009 \times 10^{-3} \end{pmatrix}$$

and

$$\begin{pmatrix} 5.218 \times 10^{-4} & -1.169 \times 10^{-3} & 3.066 \times 10^{-4} \\ -1.169 \times 10^{-3} & 3.460 \times 10^{-3} & 8.833 \times 10^{-5} \\ 3.066 \times 10^{-4} & 8.833 \times 10^{-5} & 1.047 \times 10^{-3} \end{pmatrix}$$

which are obtained, respectively, by methods, (I), (III), and (V) listed in Table I. Their corresponding eigenvalue–eigenvector pairs are shown in the equation at the bottom of the page

Clearly, $(\lambda_1 + 1 - 2\|\mathbf{t}\|) \approx 1.2038 \times 10^{-5}$, a result from the average dyadics, is a good approximation to the minor eigenvalue of the sample covariance matrix of the major eigenvector, $\lambda_3^{(I)}$. Further, the medium and minor eigenvalue–eigenvector pairs from the average dyadics respectively are very close to the largest and medium eigenvalue–eigenvector pairs of the sample covariance matrix of the major eigenvector. This result validates the analysis represented in Appendix X.

## IV. Discussion

In this work, our main objective is to present as simply as possible both the geometric and analytical ideas that underlie the proposed framework of error propagation so that the translation of this work into practice is clear to interested readers.

Here, we outline the main findings of this work. As a technique of error propagation, the proposed framework has several desirable features—namely, that the uncertainty of any tensor-derived quantity, scalar or vector, can be estimated by using the appropriate diffusion tensor representation; that the covariance matrices with respect to different diffusion tensor representations can be analytically expressed; and that covariance estimation is very accurate and is a natural by-product of the modified full Newton method of tensor estimation, a description of which can be found in [25]. Fig. 7 shows schematically the necessary steps needed to obtain the covariance matrices of interest. The sample statistics and the simulation results obtained from the proposed framework agreed reasonably well, see Fig. 6.

The concept of the average covariance matrix is introduced and applied to the issue of rotational asymmetry of the variance of the trace. This particular approach circumvents the need for bootstrap methods [18], [53] in this type of investigation. It is not hard to see that a covariance matrix with respect to a diffusion tensor representation corresponding to a particular tensor can be generated with great ease and efficiency. This technique of generating covariance matrices will be very useful in simulation studies but we should emphasize here that it is based on the limiting case of zero-residual. Therefore, readers who are interested in analyzing experimental DTI data should use the covariance matrices in (49)–(51) of Section II-E rather than the average covariance matrices discussed in Section II-F1. A similar idea related to the average covariance matrix is that of the average Hessian matrix of the ordinary representation, which is also known as the precision matrix [55]. The precision matrix is very useful in DTI experimental design [54], [55], and it can also be used in constructing the Hotelling's $T^2$-statistic for testing group differences or the Mahalanobis distance for tensor classification. However, we expect the invariant Hessian matrix of the Euler representation to be more useful than its regular counterpart, Hessian matrix, for tensor classification. These are

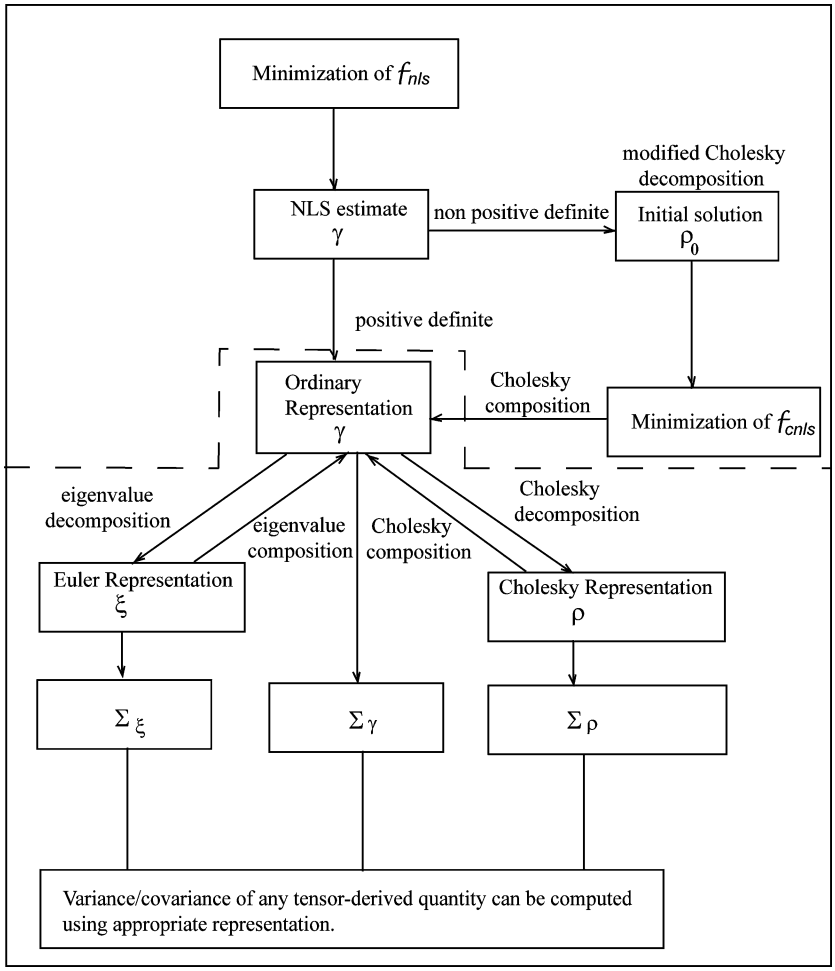



Fig. 7. Overview of the proposed error propagation framework for diffusion tensor imaging. The segment above the dotted line deals with tensor estimations; (these techniques can be found in [25]); while the segment below the dotted line pertains to the proposed framework.

areas of our current interest and we shall present them in future work.

The confidence cone, or the cone of uncertainty, of the major eigenvector in DTI—a concept introduced by Basser [26] and expounded upon by Basser and Pajevic [27], was brought to bear in fiber tract visualization by Jones [28]. But, the shape of the confidence cone discussed in these work has always been simplified or reduced to being circular. The observation of Jeong *et al.* [30] and Lazar *et al.* [29] provided clear evidence that the cone of uncertainty is generally elliptical in cross section. In this work, we have presented several analytical tools, based on the proposed framework, the perturbation method, and dyadic formalism, for constructing the elliptical cone of uncertainty. According to the result derived in Appendix X, it is noteworthy that the length and direction of the major and minor axes of the ellipse of the confidence cone are just the medium and minor eigenvalue–eigenvector pairs of the average dyadics of the particular eigenvector—a fact that had escaped notice for sometime.

The proposed framework can also be used to analyze DTI data retrospectively to investigate the reproducibility of a DTI parameter of interest or of the fiber orientation. For example, if there is an insufficient number of diffusion-weighted images to perform a bootstrap analysis, at least the uncertainty in the tensor elements and tensor-derived quantities can still be estimated within the proposed framework.

Although we have presented some cogent reasons—the unifying principles of diffusion tensor representations, of Taylor approximations of scalar and vector functions and, more importantly, of invariant Hessian and covariance structures of the nonlinear least squares objective function of DTI—for preferring the proposed framework to the perturbation method, the perturbation method is nevertheless a useful technique [17]. The diffusion tensor representations studied here are logically equivalent but they are not equally useful or significant. It is the variety of applications that made one diffusion tensor representation to be preferable to another.

We have shown that invariant Hessian matrices are more important than the Hessian matrices in DTI error propagation because covariance matrices are directly linked to them. Further, we also showed how these invariant Hessian matrices can be obtained from the proposed framework without employing the technique of covariant derivatives in tensor calculus and differential geometry.

## V. Conclusion

We have developed an analytical and geometrically intuitive error propagation framework for diffusion tensor imaging.

We have presented the nuts and bolts of various aspects of diffusion representations for understanding variability in any tensor derived quantity, vector, or scalar. This framework provides an analytical and efficient method for understanding the dependence of variance of a tensor-derived quantity on orientation or gradient schemes. Furthermore, it provides an approach for computing the necessary parameters in order to construct the elliptical confidence cone of an eigenvector. This particular technique will be very useful in fiber tractography, group analysis of diffusion tensor data and tensor classification. It is also clear that the proposed framework can be adapted to other nonlinear least squares problems.

## Appendix I
## Rotation Matrices and a Method for Finding Euler Angles

The rotation matrices, $\mathbf{R_x}(\Omega)$, $\mathbf{R_y}(\Omega)$, and $\mathbf{R_z}(\Omega)$ represent rotations through angle $\Omega$ around the $x$, $y$, and $z$ axes, respectively, and are defined as follows:

$$\mathbf{R_x}(\Omega) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\Omega) & -\sin(\Omega) \\ 0 & \sin(\Omega) & \cos(\Omega) \end{pmatrix}$$

$$\mathbf{R_y}(\Omega) = \begin{pmatrix} \cos(\Omega) & 0 & \sin(\Omega) \\ 0 & 1 & 0 \\ -\sin(\Omega) & 0 & \cos(\Omega) \end{pmatrix}$$

and

$$\mathbf{R_z}(\Omega) = \begin{pmatrix} \cos(\Omega) & -\sin(\Omega) & 0 \\ \sin(\Omega) & \cos(\Omega) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The following discussion is on obtaining the Euler angles from the proper rotation matrix, $\mathbf{Q}$, which can be expressed columnwise as

$$\begin{pmatrix} Q_{11} \\ Q_{21} \\ Q_{31} \end{pmatrix} = \begin{pmatrix} \cos(\theta)\cos(\phi)\cos(\psi) - \sin(\phi)\sin(\psi) \\ \cos(\theta)\cos(\psi)\sin(\phi) + \cos(\phi)\sin(\psi) \\ -\cos(\psi)\sin(\theta) \end{pmatrix}$$

$$\begin{pmatrix} Q_{12} \\ Q_{22} \\ Q_{32} \end{pmatrix} = \begin{pmatrix} -\cos(\psi)\sin(\phi) - \cos(\theta)\cos(\phi)\sin(\psi) \\ \cos(\phi)\cos(\psi) - \cos(\theta)\sin(\phi)\sin(\psi) \\ \sin(\theta)\sin(\psi) \end{pmatrix}$$

and

$$\begin{pmatrix} Q_{13} \\ Q_{23} \\ Q_{33} \end{pmatrix} = \begin{pmatrix} \cos(\phi)\sin(\theta) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{pmatrix}.$$

By proper rotation, we mean that the determinant of $\mathbf{Q}$ should be positive one. If negative one is encountered, we can always change $\mathbf{Q}$ to its additive inverse, $-\mathbf{Q}$. Once this step is checked, the Euler angles can then be found as follows:

1) $\theta = \theta_1 = \cos^{-1}(Q_{33})$
2) If $\theta \neq 0$, then
   a) $\phi = \mathrm{atan2}(Q_{23}, Q_{13})$
   b) $\psi = \mathrm{atan2}(Q_{32}, -Q_{31})$.

The function atan2 is defined in many programming languages such as C and Java.

In the case where $\theta = 0$, the rotation matrix $\mathbf{Q}$ can be shown to reduce to

$$\begin{pmatrix} \cos(\phi+\psi) & -\sin(\phi+\psi) & 0 \\ \sin(\phi+\psi) & \cos(\phi+\psi) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is clear that $\phi$ and $\psi$ can not be uniquely determined and we can set one of them to zero. Let $\psi = 0$, then $\phi = \mathrm{atan2}(-Q_{12}, Q_{22})$.

## Appendix II
## Mappings Between Various Representations

The components of $\gamma(\xi)$ are defined below

$$\begin{aligned}
\gamma_1(\boldsymbol{\xi}) &= \xi_1 = \ln(\alpha) \\
\gamma_2(\boldsymbol{\xi}) &= D_{xx}(\boldsymbol{\xi}) = [Q_{11}]^2\,\xi_2 + [Q_{12}]^2\,\xi_3 + [Q_{13}]^2\,\xi_4 \\
\gamma_3(\boldsymbol{\xi}) &= D_{yy}(\boldsymbol{\xi}) = [Q_{21}]^2\,\xi_2 + [Q_{22}]^2\,\xi_3 + [Q_{23}]^2\,\xi_4 \\
\gamma_4(\boldsymbol{\xi}) &= D_{zz}(\boldsymbol{\xi}) = [Q_{31}]^2\,\xi_2 + [Q_{32}]^2\,\xi_3 + [Q_{33}]^2\,\xi_4 \\
\gamma_5(\boldsymbol{\xi}) &= D_{xy}(\boldsymbol{\xi}) \\
&= [Q_{11}]\,[Q_{21}]\,\xi_2 + [Q_{12}]\,[Q_{22}]\,\xi_3 + [Q_{13}]\,[Q_{23}]\,\xi_4 \\
\gamma_6(\boldsymbol{\xi}) &= D_{yz}(\boldsymbol{\xi}) \\
&= [Q_{21}]\,[Q_{31}]\,\xi_2 + [Q_{22}]\,[Q_{32}]\,\xi_3 + [Q_{23}]\,[Q_{33}]\,\xi_4 \\
\gamma_7(\boldsymbol{\xi}) &= D_{xz}(\boldsymbol{\xi}) \\
&= [Q_{11}]\,[Q_{31}]\,\xi_2 + [Q_{12}]\,[Q_{32}]\,\xi_3 + [Q_{13}]\,[Q_{33}]\,\xi_4
\end{aligned}$$

where the components of $\mathbf{Q}$ are functions of $\xi_5$, $\xi_6$, and $\xi_7$.

---

$$\begin{aligned}
\{\lambda_1^{(I)}, \boldsymbol{\Psi}_1^{(I)}\} &= \{3.452 \times 10^{-3}, [-0.3213 \quad 0.9466 \quad 0.0261]^T\} \\
\{\lambda_2^{(I)}, \boldsymbol{\Psi}_2^{(I)}\} &= \{1.123 \times 10^{-3}, [0.2863 \quad 0.0708 \quad 0.9555]^T\} \text{ and} \\
\{\lambda_3^{(I)}, \boldsymbol{\Psi}_3^{(I)}\} &= \{1.204 \times 10^{-5}, [-0.9027 \quad -0.3145 \quad 0.2938]^T\} \text{ for method (I)} \\
\{\lambda_1^{(III)}, \boldsymbol{\Psi}_1^{(III)}\} &= \{3.646 \times 10^{-3}, [0.3320 \quad -0.9432 \quad -0.0124]^T\} \\
\{\lambda_2^{(III)}, \boldsymbol{\Psi}_2^{(III)}\} &= \{1.104 \times 10^{-3}, [0.2735 \quad 0.1088 \quad 0.9557]^T\} \text{ and} \\
\{\lambda_3^{(III)}, \boldsymbol{\Psi}_3^{(III)}\} &= \{3.160 \times 10^{-19} \approx 0, [-0.9028 \quad -0.3139 \quad 0.2940]^T\} \text{ for method (III) and} \\
\{\lambda_1^{(V)}, \boldsymbol{\Psi}_1^{(V)}\} &= \{3.868 \times 10^{-3}, [0.3302 \quad -0.9439 \quad 0.0063]^T\} \\
\{\lambda_2^{(V)}, \boldsymbol{\Psi}_2^{(V)}\} &= \{1.145 \times 10^{-3}, [0.2766 \quad 0.1032 \quad 0.9554]^T\} \text{ and} \\
\{\lambda_3^{(V)}, \boldsymbol{\Psi}_3^{(V)}\} &= \{1.515 \times 10^{-5}, [0.9025 \quad 0.3137 \quad -0.2951]^T\} \text{ and method (V)}
\end{aligned}$$

For completeness, we will show analytical formulas for each component of $\boldsymbol{\rho}(\boldsymbol{\gamma})$ by the Cholesky decomposition with the assumption that the diffusion tensor within $\boldsymbol{\gamma}$ is positive definite otherwise, as mentioned in the text, the modified Cholesky decomposition is to be used for constructing $\boldsymbol{\rho}(\boldsymbol{\gamma})$[25], [50]. Before presenting the formulas, we shall define the following terms to simplify the expression of $\boldsymbol{\rho}(\boldsymbol{\gamma})$:

$\mathbf{D}_1 = D_{xx}$, $\mathbf{D}_2 = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix}$, $\mathbf{D}_3 = \mathbf{D}$, and $\det(\cdot)$ is the matrix determinant.

$\boldsymbol{\rho}(\boldsymbol{\gamma})$ can be expressed as follows:

$$\begin{aligned}
\rho_1(\boldsymbol{\gamma}) &= \gamma_1 = \ln(\alpha) \\
\rho_2(\boldsymbol{\gamma}) &= \det(\mathbf{D}_1)^{1/2} = \det(D_{xx})^{1/2} = \det(\gamma_2)^{1/2} \\
\rho_3(\boldsymbol{\gamma}) &= \left(\frac{\det(\mathbf{D}_2)}{\det(\mathbf{D}_1)}\right)^{1/2} \\
\rho_4(\boldsymbol{\gamma}) &= \left(\frac{\det(\mathbf{D}_3)}{\det(\mathbf{D}_2)}\right)^{1/2} \\
\rho_5(\boldsymbol{\gamma}) &= \frac{D_{xy}}{\det(\mathbf{D}_1)^{1/2}} \\
\rho_6(\boldsymbol{\gamma}) &= \frac{D_{xx}D_{yz} - D_{xy}D_{xz}}{\det(\mathbf{D}_1)^{1/2}\det(\mathbf{D}_2)^{1/2}} \\
\rho_7(\boldsymbol{\gamma}) &= \frac{D_{xz}}{\det(\mathbf{D}_1)^{1/2}}
\end{aligned}$$

## Appendix III
## Representation by Hext

The representation proposed by Hext [45] is the mapping relating the components of $\boldsymbol{\Lambda}$ to those of $\mathbf{D}$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & \lambda_4 & \lambda_6 \\ \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_6 & \lambda_5 & \lambda_3 \end{pmatrix} = \mathbf{Q}^T\mathbf{D}\mathbf{Q} \tag{C1}$$

where $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$, but the off-diagonal elements of $\boldsymbol{\Lambda}$ are not necessarily zero. A special case of (C1) with $\boldsymbol{\Lambda}$ being a diagonal matrix was used by Anderson [17] to compute the covariance between two eigenvalues.

Adapting (C1) to the convention used in this paper, we can show that the linear relation in vector form can be expressed as shown in the equation at the bottom of the page.

We shall denote the above equation as $\boldsymbol{\lambda} = \mathbf{P}\boldsymbol{\gamma}$, the first-order differential can be written as $d\boldsymbol{\lambda} = \mathbf{P}d\boldsymbol{\gamma}$ so that we can identify the elements of $\mathbf{P}$ as $[\mathbf{P}]_{ij} = \partial\lambda_i/\partial\gamma_j$ or $\mathbf{P} = \mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\lambda})$. If the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ is given then it can be shown that $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} = \mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\lambda})\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\mathbf{J}_{\boldsymbol{\gamma}}^T(\boldsymbol{\lambda}) = \mathbf{P}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\mathbf{P}^T$. See Section II-E and Appendix VI for the technique for transforming covariance matrices from one representation to another.

It is evident that this representation $\boldsymbol{\lambda}$ has a simpler expression than that of the proposed Euler representation, $\xi$. However, this representation cannot answer questions regarding the uncertainties in the eigenvectors, i.e., the elliptical cone of uncertainty of the major eigenvector, without resorting to the perturbation method.

## Appendix IV
## Derivation of a Key Equation on Error Propagation

As defined in the main text, we have $\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}}) = \mathbf{Q}\boldsymbol{\Lambda}^{1/2}(\mathbf{Q}\boldsymbol{\Lambda}^{1/2})^T$ and $\boldsymbol{\eta} \equiv (\mathbf{Q}\boldsymbol{\Lambda}^{1/2})^T\boldsymbol{\delta} = \boldsymbol{\Lambda}^{1/2}\mathbf{Q}^T\boldsymbol{\delta}$ where $\mathbf{Q}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix with positive elements. Therefore, we can write $\boldsymbol{\delta} = \mathbf{Q}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\eta}$. This is equivalent to the following expressions in component form:

$$\eta_i = \sum_j \frac{\partial\eta_i}{\partial\delta_j}\delta_j \tag{D1}$$

and

$$\delta_i = \sum_j \frac{\partial\delta_i}{\partial\eta_j}\eta_j \tag{D2}$$

where $\dfrac{\partial\eta_i}{\partial\delta_j} = [\boldsymbol{\Lambda}^{1/2}\mathbf{Q}^T]_{ij}$ and $\dfrac{\partial\delta_i}{\partial\eta_j} = [\mathbf{Q}\boldsymbol{\Lambda}^{-1/2}]_{ij}$

$$\begin{aligned}
\Delta g(\boldsymbol{\delta})^2 &\approx 2\Delta f_{\mathrm{NLS}}(\boldsymbol{\delta})\|\nabla_{\boldsymbol{\eta}} g\|^2 \\
&= 2\Delta f(\boldsymbol{\delta})\sum_i \frac{\partial g(\boldsymbol{\eta})}{\partial\eta_i}\frac{\partial g(\boldsymbol{\eta})}{\partial\eta_i} \\
&= 2\Delta f(\boldsymbol{\delta})\sum_i \left[\sum_j \frac{\partial\delta_j}{\partial\eta_i}\frac{\partial g(\boldsymbol{\delta})}{\partial\delta_j}\right] \\
&\quad \times \left[\sum_k \frac{\partial\delta_k}{\partial\eta_i}\frac{\partial g(\boldsymbol{\delta})}{\partial\delta_k}\right] \\
&= 2\Delta f(\boldsymbol{\delta})\sum_j\sum_k \left(\frac{\partial g(\boldsymbol{\delta})}{\partial\delta_j}\right) \\
&\quad \times \left[\sum_i \frac{\partial\delta_j}{\partial\eta_i}\frac{\partial\delta_k}{\partial\eta_i}\right]\left(\frac{\partial g(\boldsymbol{\delta})}{\partial\delta_k}\right) \\
&= 2\Delta f(\boldsymbol{\delta})\sum_j\sum_k \left(\frac{\partial g(\boldsymbol{\delta})}{\partial\delta_j}\right)
\end{aligned}$$

$$\begin{pmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & Q_{11}^2 & Q_{21}^2 & Q_{31}^2 & 2Q_{11}Q_{21} & 2Q_{21}Q_{31} & 2Q_{11}Q_{31} \\
0 & Q_{12}^2 & Q_{22}^2 & Q_{32}^2 & 2Q_{12}Q_{22} & 2Q_{22}Q_{32} & 2Q_{12}Q_{32} \\
0 & Q_{13}^2 & Q_{23}^2 & Q_{33}^2 & 2Q_{13}Q_{23} & 2Q_{23}Q_{33} & 2Q_{13}Q_{33} \\
0 & Q_{11}Q_{12} & Q_{21}Q_{22} & Q_{31}Q_{32} & Q_{11}Q_{22}+Q_{21}Q_{12} & Q_{21}Q_{32}+Q_{31}Q_{22} & Q_{11}Q_{32}+Q_{31}Q_{12} \\
0 & Q_{12}Q_{13} & Q_{22}Q_{23} & Q_{32}Q_{33} & Q_{12}Q_{23}+Q_{22}Q_{13} & Q_{22}Q_{33}+Q_{32}Q_{23} & Q_{12}Q_{33}+Q_{32}Q_{13} \\
0 & Q_{11}Q_{13} & Q_{21}Q_{23} & Q_{31}Q_{33} & Q_{11}Q_{23}+Q_{21}Q_{13} & Q_{21}Q_{33}+Q_{31}Q_{23} & Q_{11}Q_{33}+Q_{31}Q_{13}
\end{pmatrix} \begin{pmatrix} \ln S_0 \\ D_{xx} \\ D_{yy} \\ D_{zz} \\ D_{xy} \\ D_{yz} \\ D_{xz} \end{pmatrix}$$

$$\begin{aligned}
&\times \left[\sum_i [\mathbf{Q}\mathbf{\Lambda}^{-1/2}]_{ji}[\mathbf{Q}\mathbf{\Lambda}^{-1/2}]_{ki}\right]\left(\frac{\partial g(\boldsymbol{\delta})}{\partial \delta_k}\right)\\
&= 2\Delta f(\boldsymbol{\delta})\sum_j\sum_k\left(\frac{\partial g(\boldsymbol{\delta})}{\partial \delta_j}\right)\\
&\quad\times \left[\sum_i [\mathbf{Q}\mathbf{\Lambda}^{-1/2}]_{ji}[(\mathbf{Q}\mathbf{\Lambda}^{-1/2})^T]_{ik}\right]\left(\frac{\partial g(\boldsymbol{\delta})}{\partial \delta_k}\right)\\
&= 2\Delta f(\boldsymbol{\delta})\sum_j\sum_k\left(\frac{\partial g(\boldsymbol{\delta})}{\partial \delta_j}\right)\\
&\quad\times \left([\mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^T]_{jk}\right)\left(\frac{\partial g(\boldsymbol{\delta})}{\partial \delta_k}\right)\\
&= 2\Delta f(\boldsymbol{\delta})\nabla_{\boldsymbol{\delta}}^T g(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\gamma}} g(\hat{\boldsymbol{\gamma}}). \quad \text{(D3)}
\end{aligned}$$

From the above derivation, we also see that $||\nabla_{\boldsymbol{\eta}} g||^2 = \nabla_{\boldsymbol{\gamma}}^T g(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\eta}} g(\hat{\boldsymbol{\gamma}})$ and $\nabla_{\boldsymbol{\eta}} g = \mathbf{\Lambda}^{-1/2}\mathbf{Q}^T\nabla_{\boldsymbol{\delta}} g$.

## Appendix V
## Hessian Structures in Different Representations

Here, we provide explicit Hessian expressions with respect to various representations studied in this paper

$$\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) = \mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W} \quad \text{(E1)}$$

$$\nabla^2 f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho})) = \mathbf{J}_{\boldsymbol{\rho}}^T(\boldsymbol{\gamma})\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma}) + \sum_{i=1}^n r_i\hat{s}_i\mathbf{P}_i \quad \text{(E2)}$$

$$\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\boldsymbol{\xi})) = \mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma})\mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}) + \sum_{i=1}^n r_i\hat{s}_i\mathbf{T}_i \quad \text{(E3)}$$

where $S$ and $\hat{S}$ are diagonal matrices whose diagonal elements are the observed and the estimated diffusion weighted signals, respectively, i.e.,

$$\mathbf{S} = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{pmatrix}, \text{ and } \hat{\mathbf{S}} = \begin{pmatrix} \hat{s}_1 & & \\ & \ddots & \\ & & \hat{s}_n \end{pmatrix}.$$

Further, we have $\mathbf{R} = \mathbf{S} - \hat{\mathbf{S}}$, $[\mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma})]_{ij} \equiv \partial\gamma_i/\partial\rho_j$, $[\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma})]_{ij} \equiv \partial\gamma_i/\partial\xi_j$, $[\mathbf{P}_q]_{kl} = \sum_{i=1}^7(-W_{qi})\partial^2\gamma_i/\partial\rho_k\partial\rho_l$, and $[\mathbf{T}_q]_{kl} = \sum_{i=1}^7(-W_{qi})\partial^2\gamma_i/\partial\xi_k\partial\xi_l$. Equations (E1) and (E2) have been previously derived and studied by Koay *et al.* [25].

## Appendix VI
## Covariance Structures in Different Representations

In (31), we have the following equation:

$$\left[\mathbf{\Sigma}_{g(\hat{\boldsymbol{\gamma}})}\right]_{ij} = \sigma_{\mathrm{DW}}^2\nabla_{\boldsymbol{\gamma}}^T g_i(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\gamma}} g_j(\hat{\boldsymbol{\gamma}}).$$

To construct the covariance matrix with respect to the ordinary representation, we write

$$\begin{aligned}
[\mathbf{\Sigma}_{\boldsymbol{\gamma}}]_{ij} \equiv \left[\mathbf{\Sigma}_{\boldsymbol{\gamma}(\hat{\boldsymbol{\gamma}})}\right]_{ij} &= \sigma_{\mathrm{DW}}^2\nabla_{\boldsymbol{\gamma}}^T\boldsymbol{\gamma}_i(\hat{\boldsymbol{\gamma}})\left[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})\right]^{-1}\\
&\quad\times \nabla_{\boldsymbol{\gamma}}\boldsymbol{\gamma}_j(\hat{\boldsymbol{\gamma}})
\end{aligned}$$

or

$$\begin{aligned}
\mathbf{\Sigma}_{\boldsymbol{\gamma}} \equiv \mathbf{\Sigma}_{\boldsymbol{\gamma}(\hat{\boldsymbol{\gamma}})} &= \sigma_{\mathrm{DW}}^2\left[\nabla_{\boldsymbol{\gamma}}\gamma_1(\hat{\boldsymbol{\gamma}}) \quad \ldots \quad \nabla_{\boldsymbol{\gamma}}\boldsymbol{\gamma}_j(\hat{\boldsymbol{\gamma}})\right]^T\\
&\quad\times [\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\left[\nabla_{\boldsymbol{\gamma}}\boldsymbol{\gamma}_1(\hat{\boldsymbol{\gamma}}) \quad \ldots \quad \nabla_{\boldsymbol{\gamma}}\boldsymbol{\gamma}_j(\hat{\boldsymbol{\gamma}})\right]\\
&= \sigma_{\mathrm{DW}}^2\mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\gamma}}))[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\mathbf{J}_{\boldsymbol{\gamma}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\gamma}}))\\
&= \sigma_{\mathrm{DW}}^2\mathbf{I}\left[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})\right]^{-1}\mathbf{I}^T\\
&= \sigma_{\mathrm{DW}}^2[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}
\end{aligned}$$

where $\mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\gamma}})) = \mathbf{I}$ and $\mathbf{I}$ denotes the identity matrix.

To construct the covariance matrix with respect to the Euler representation, we write

$$[\mathbf{\Sigma}_{\boldsymbol{\xi}}]_{ij} \equiv \left[\mathbf{\Sigma}_{\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}})}\right]_{ij} = \sigma_{\mathrm{DW}}^2\nabla_{\boldsymbol{\gamma}}^T\boldsymbol{\xi}_i(\hat{\boldsymbol{\gamma}})[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\nabla_{\boldsymbol{\gamma}}\boldsymbol{\xi}_j(\hat{\boldsymbol{\gamma}})$$

or

$$\begin{aligned}
\mathbf{\Sigma}_{\boldsymbol{\xi}} \equiv \mathbf{\Sigma}_{\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}})} &= \sigma_{\mathrm{DW}}^2\left[\nabla_{\boldsymbol{\gamma}}\boldsymbol{\xi}_1(\hat{\boldsymbol{\gamma}}) \quad \ldots \quad \nabla_{\boldsymbol{\gamma}}\boldsymbol{\xi}_j(\hat{\boldsymbol{\gamma}})\right]^T\\
&\quad\times [\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\left[\nabla_{\boldsymbol{\gamma}}\boldsymbol{\xi}_1(\hat{\boldsymbol{\gamma}}) \quad \ldots \quad \nabla_{\boldsymbol{\gamma}}\boldsymbol{\xi}_j(\hat{\boldsymbol{\gamma}})\right]\\
&= \sigma_{\mathrm{DW}}^2\mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}}))[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\mathbf{J}_{\boldsymbol{\gamma}}^T(\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}}))\\
&= \sigma_{\mathrm{DW}}^2[\mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}}))]^{(-1)(-1)}[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]^{-1}\\
&\quad\times [\mathbf{J}_{\boldsymbol{\gamma}}^T(\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}}))]^{(-1)(-1)}\\
&= \sigma_{\mathrm{DW}}^2\Big[\left[\mathbf{J}_{\boldsymbol{\gamma}}^T(\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}}))\right]^{(-1)}\\
&\qquad\times [\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})][\mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}}))]^{(-1)}\Big]^{-1}\\
&= \sigma_{\mathrm{DW}}^2[\mathbf{J}_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))[\nabla^2 f_{\mathrm{NLS}}(\hat{\boldsymbol{\gamma}})]\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))]^{-1}. \quad \text{(F1)}
\end{aligned}$$

Two identities: $\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}}))\cdot\mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}})) = \mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\xi}(\hat{\boldsymbol{\xi}})) = \mathbf{I}$ and $\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}(\hat{\boldsymbol{\xi}})) = [\mathbf{J}_{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\hat{\boldsymbol{\gamma}}))]^{-1}$ were used in the derivation of (F1).

Equation (F1) is very important because we have discovered the part of the Hessian matrix that is invariant with respect to transformation without using the concept of covariant derivative in tensor calculus. Interestingly, the invariant Hessian matrix of the Euler representation is exactly the first term of the Hessian matrix in (E2).

## Appendix VII
## Gradient Computation: Eigenvectors

The gradient of the first, second, and third components of $\mathbf{q}_1$ can be written as

$$\nabla_{\boldsymbol{\xi}} Q_{11} = \begin{pmatrix} 0\\ 0\\ 0\\ 0\\ -\cos(\phi)\cos(\psi)\sin(\theta)\\ -\cos(\theta)\cos(\psi)\sin(\phi) - \cos(\phi)\sin(\psi)\\ -\cos(\psi)\sin(\phi) - \cos(\theta)\cos(\phi)\sin(\psi) \end{pmatrix}$$

$$\nabla_{\boldsymbol{\xi}} Q_{21} = \begin{pmatrix} 0\\ 0\\ 0\\ 0\\ -\cos(\psi)\sin(\theta)\sin(\phi)\\ \cos(\theta)\cos(\phi)\cos(\psi) - \sin(\phi)\sin(\psi)\\ \cos(\phi)\cos(\psi) - \cos(\theta)\sin(\phi)\sin(\psi) \end{pmatrix}$$

and

$$\nabla_{\boldsymbol{\xi}} Q_{31} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\cos(\theta)\cos(\psi) \\ 0 \\ \sin(\theta)\sin(\psi) \end{pmatrix}.$$

Similarly, the gradient of the components of $\mathbf{q}_2$ and $\mathbf{q}_3$ can be computed quite easily.

## Appendix VIII
## Gradient Computation: Tensor-Derived Quantities

A few notations and conventions are introduced here to keep the formulas shown in (38)–(40) in a compact form.

1) The indices $i = 1, 2, 3$ denote $x, y, z$, respectively.
2) $\delta_{ij}$ is the Kronecker delta function, i.e., $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$.
3) The formulas for the partial derivatives with respect to the off-diagonal elements of $\mathbf{D}$ is symmetrized, i.e., $\partial \mathrm{FA}/\partial D_{ij} = \partial \mathrm{FA}/\partial D_{ji}$ for $i \neq j$.

For convenience, the formulas that are frequently used are listed here

$$\frac{\partial \mathrm{Tr}(\mathbf{D})}{\partial D_{ij}} = \delta_{ij} \tag{H1}$$

$$\frac{\partial \mathrm{Tr}(\mathbf{D}^2)}{\partial D_{ij}} = \frac{4}{(1+\delta_{ij})} D_{ij} \tag{H2}$$

$$\frac{\partial}{\partial_{ij}} \frac{\mathrm{Tr}(\mathbf{D})^m}{\mathrm{Tr}(\mathbf{D}^2)^n} = m \frac{\mathrm{Tr}(\mathbf{D})^{m-1}}{\mathrm{Tr}(\mathbf{D}^2)^n} \delta_{ij} - n \frac{\mathrm{Tr}(\mathbf{D})^m}{\mathrm{Tr}(\mathbf{D}^2)^{n+1}} \frac{4}{(1+\delta_{ij})} D_{ij} \tag{H3}$$

and

$$\frac{\partial}{\partial D_{ij}} \frac{\mathrm{Tr}(\mathbf{D}^2)^m}{\mathrm{Tr}(\mathbf{D})^n} = m \frac{\mathrm{Tr}(\mathbf{D}^2)^{m-1}}{\mathrm{Tr}(\mathbf{D})^n} \frac{4}{(1+\delta_{ij})} D_{ij} - n \frac{\mathrm{Tr}(\mathbf{D}^2)^m}{\mathrm{Tr}(\mathbf{D})^{n+1}} \delta_{ij} \tag{H4}$$

## Appendix IX
## Average Covariance Matrix

In the zero-residual case, which is very useful in simulation studies where the ground truth is known, the invariant Hessian expressions in (49)–(51) reduce to

$$\mathrm{inv}(\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma})) = \mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W} \tag{I1}$$

$$\mathrm{inv}(\nabla^2 f_{\mathrm{CNLS}}(\boldsymbol{\gamma}(\boldsymbol{\rho}))) = \mathbf{J}_{\boldsymbol{\rho}}^T(\boldsymbol{\gamma}) \mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W} \mathbf{J}_{\boldsymbol{\rho}}(\boldsymbol{\gamma}) \tag{I2}$$

and

$$\mathrm{inv}(\nabla^2 f_{\mathrm{ENLS}}(\boldsymbol{\gamma}(\boldsymbol{\xi}))) = J_{\boldsymbol{\xi}}^T(\boldsymbol{\gamma}) \mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W} \mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\gamma}) \tag{I3}$$

Further, we have $\left\langle \hat{\mathbf{S}}^2 \right\rangle \equiv \left(1/N \sum_{i=1} \hat{\mathbf{S}}_i^2\right) = \hat{\mathbf{S}}^2$, where $\hat{\mathbf{S}}$ is known. As an example, the average invariant Hessian matrix of (I1) can be expressed as follows:

$$\begin{aligned} \left\langle \mathrm{inv}(\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma})) \right\rangle &= \left\langle \nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}) \right\rangle = \frac{1}{N} \sum_{i=1} \mathbf{W}^T \hat{\mathbf{S}}_i^2 \mathbf{W} \\ &= \mathbf{W}^T \left( \frac{1}{N} \sum_{i=1} \hat{\mathbf{S}}_i^2 \right) \mathbf{W} \\ &= \mathbf{W}^T \left\langle \hat{\mathbf{S}}^2 \right\rangle \mathbf{W} = \mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W}. \end{aligned} \tag{I4}$$

Therefore, the average covariance matrix is

$$\langle \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \rangle \approx \sigma_{\mathrm{DW}}^2 \left\langle [\mathrm{inv}(\nabla^2 f_{\mathrm{NLS}}(\boldsymbol{\gamma}))]^{-1} \right\rangle = \sigma_{\mathrm{DW}}^2 [\mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W}]^{-1}. \tag{I5}$$

In other words, we expect the arithmetic mean of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ to approach $\sigma_{\mathrm{DW}}^2 [\mathbf{W}^T \hat{\mathbf{S}}^2 \mathbf{W}]^{-1}$ as the number of samples of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ increases. Note that the arithmetic mean and the method of averaging used in obtaining (I5) are different but we expect the difference between these two quantities to be negligible for a large sample.

## Appendix X
## Connection Between the Elliptical Cone of Uncertainty and the Average Dyadics

Let $\{\mathbf{q}_{11}, \ldots, \mathbf{q}_{1N}\}$ be the collection of properly oriented major eigenvectors with respect to the mean major eigenvector and let $\left\langle \mathbf{q}_1 \mathbf{q}_1^T \right\rangle = 1/N \sum_{j=1}^N \mathbf{q}_{1j} \mathbf{q}_{1j}^T$ be the average dyadics [27]. Further, let the eigenvalue decomposition of the average dyadics be $\sum_{i=1}^3 \lambda_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T$ where $\lambda_1 \geq \lambda_2 \geq \lambda_3$.

According to [56], the maximum likelihood estimate of the mean of $\{\mathbf{q}_{11}, \ldots, \mathbf{q}_{1N}\}$ is $\boldsymbol{\psi}_1$. We shall now show that these eigenvalue–eigenvector pairs, $\{\lambda_2, \boldsymbol{\psi}_2\}$ and $\{\lambda_3, \boldsymbol{\psi}_3\}$, are related to the length and direction of the major and the minor axes of the confidence cone of the major eigenvector, $\mathbf{q}_1$. In other words, these two eigenvalue–eigenvector pairs are related to the covariance matrix of $\{\mathbf{q}_{11}, \ldots, \mathbf{q}_{1N}\}$. The argument goes as follows. Let the sample covariance of $\{\mathbf{q}_{11}, \ldots, \mathbf{q}_{1N}\}$ be defined as

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{q}_1} &= \frac{1}{N-1} \sum_{i=1}^N (\mathbf{q}_{1i} - \boldsymbol{\psi}_1)(\mathbf{q}_{1i} - \boldsymbol{\psi}_1)^T \\ &= \frac{1}{N-1} \sum_{i=1}^N (\mathbf{q}_{1i}\mathbf{q}_{1i}^T - \boldsymbol{\psi}_1 \mathbf{q}_{1i}^T - \mathbf{q}_{1i} \boldsymbol{\psi}_1^T + \boldsymbol{\psi}_1 \boldsymbol{\psi}_1^T) \\ &= \frac{N}{N-1} \left( \frac{1}{N} \sum_{i=1}^N \mathbf{q}_{1i}\mathbf{q}_{1i}^T - \boldsymbol{\psi}_1 \left( \frac{1}{N} \sum_{i=1}^N \mathbf{q}_{1i}^T \right) - \left( \frac{1}{N} \sum_{i=1}^N \mathbf{q}_{1i} \right) \boldsymbol{\psi}_1^T + \boldsymbol{\psi}_1 \boldsymbol{\psi}_1^T \right). \end{aligned} \tag{J1}$$

Let $\mathbf{t} = 1/N \sum_{i=1}^{N} \mathbf{q}_{1i}$, then $\hat{\mathbf{t}} = \mathbf{t}/||\mathbf{t}||$. If we assume that $\hat{\mathbf{t}} \approx \boldsymbol{\psi}_1$, which is not unreasonable because $\boldsymbol{\psi}_1$ is an estimate of the mean major eigenvector, then we have

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{q}_1} &\cong \frac{N}{N-1} \\ &\quad \times \left( \sum_{i=1}^{3} \lambda_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T - ||\mathbf{t}|| \, \boldsymbol{\psi}_1 \boldsymbol{\psi}_1^T - ||\mathbf{t}|| \, \boldsymbol{\psi}_1 \boldsymbol{\psi}_1^T + \boldsymbol{\psi}_1 \boldsymbol{\psi}_1^T \right) \\ \boldsymbol{\Sigma}_{\mathbf{q}_1} &\cong \frac{N}{N-1} \\ &\quad \times \left( \lambda_2 \boldsymbol{\psi}_2 \boldsymbol{\psi}_2^T + \lambda_3 \boldsymbol{\psi}_3 \boldsymbol{\psi}_3^T + (\lambda_1 + 1 - 2\,||\mathbf{t}||) \boldsymbol{\psi}_1 \boldsymbol{\psi}_1^T \right).\end{aligned}$$

When $N$ is large we have $\lambda_1 \approx 1$ and $||\mathbf{t}|| \approx 1$, so that $(\lambda_1 + 1 - 2\,||\mathbf{t}||) \approx 0$. The sample covariance is then reduced to

$$\boldsymbol{\Sigma}_{\mathbf{q}_1} \cong \frac{N}{N-1} \left( \lambda_2 \boldsymbol{\psi}_2 \boldsymbol{\psi}_2^T + \lambda_3 \boldsymbol{\psi}_3 \boldsymbol{\psi}_3^T \right). \tag{J2}$$

Essentially, the dyadic product formulation suggested in [27] is sufficient to construct the elliptical confidence cone without having to use (J1). In retrospect, the construction of the confidence cone using (J2) bypasses the need to reorient the sample eigenvectors such that they are pointing on the same hemisphere as the mean major eigenvector.

## ACKNOWLEDGMENT

C. G. Koay dedicates this work to the memory of Dr. M. C. Toh. The authors would like to thank L. Salak for editing this paper and thank the anonymous reviewers for their helpful comments.

## REFERENCES

[1] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophys. J.*, vol. 66, pp. 259–267, 1994.

[2] P. J. Basser, "Inferring microstructural features and the physiological state of tissues from diffusion-weighted images," *NMR Biomed.*, vol. 8, pp. 333–344, 1995.

[3] P. J. Basser and C. Pierpaoli, "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI," *J. Magn. Reson. B*, vol. 111, pp. 209–219, 1996.

[4] P. J. Basser, J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *J. Magn. Reson. B*, vol. 103, pp. 247–254, 1994.

[5] C. Pierpaoli and P. J. Basser, "Toward a quantitative assessment of diffusion anisotropy," *Magn. Reson. Med.*, vol. 36, pp. 893–906, 1996.

[6] C. Pierpaoli, P. Jezzard, P. J. Basser, A. Barnett, and G. D. Chiro, "Diffusion tensor MR imaging of the human brain," *Radiology*, vol. 201, pp. 637–648, 1996.

[7] J. J. Neil, S. I. Shiran, R. C. McKinstry, G. L. Schefft, A. Z. Snyder, C. R. Almli, E. Akbudak, J. A. Aronovitz, J. P. Miller, B. C. Lee, and T. E. Conturo, "Normal brain in human newborns: Apparent diffusion coefficient and diffusion anisotropy measured by using diffusion tensor MR imaging," *Radiology*, vol. 209, pp. 57–66, 1998.

[8] A. M. Ulug, N. Beauchamp, Jr, R. N. Bryan, and P. C. van Zijl, "Absolute quantitation of diffusion constants in human stroke," *Stroke*, vol. 28, pp. 483–490, 1997.

[9] J. S. Shimony, R. C. McKinstry, E. Akbudak, J. A. Aronovitz, A. Z. Snyder, N. F. Lori, T. S. Cull, and T. E. Conturo, "Quantitative diffusion-tensor anisotropy brain MR imaging: Normative human data and anatomic analysis," *Radiology*, vol. 212, pp. 770–784, 1999.

[10] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi, "In vivo fiber tractography using DT-MRI data," *Magn. Reson. Med.*, vol. 44, pp. 625–632, 2000.

[11] M. Lazar, D. M. Weinstein, J. S. Tsuruda, K. M. Hasan, K. Arfanakis, M. E. Meyerand, B. Badie, H. A. Rowley, V. Haughton, A. Field, and A. L. Alexander, "White matter tractography using diffusion tensor deflection," *Hum. Brain Mapp.*, vol. 18, pp. 306–321, 2003.

[12] D. K. Jones, A. R. Travis, G. Eden, C. Pierpaoli, and P. J. Basser, "PASTA: Pointwise assessment of streamline tractography attributes," *Magn. Reson. Med.*, vol. 53, pp. 1462–1467, 2005.

[13] D. K. Jones, M. A. Horsfield, and A. Simmons, "Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging," *Magn. Reson. Med.*, vol. 42, pp. 515–525, 1999.

[14] K. M. Hasan, D. L. Parker, and A. L. Alexander, "Comparison of gradient encoding schemes for diffusion-tensor MRI," *J Magn. Reson. Imag.*, vol. 13, pp. 769–780, 2001.

[15] P. G. Batchelor, D. Atkinson, D. L. Hill, F. Calamante, and A. Connelly, "Anisotropic noise propagation in diffusion tensor MRI sampling schemes," *Magn. Reson. Med.*, vol. 49, pp. 1143–1151, 2003.

[16] S. Skare, T. Li, B. Nordell, and M. Ingvar, "Noise considerations in the determination of diffusion tensor anisotropy," *Magn. Reson. Imag.*, vol. 18, pp. 659–669, 2000.

[17] A. W. Anderson, "Theoretical analysis of the effects of noise on diffusion tensor imaging," *Magn. Reson. Med.*, vol. 46, pp. 1174–1188, 2001.

[18] S. Pajevic and P. J. Basser, "Parametric and non-parametric statistical analysis of DT-MRI data," *J. Magn. Reson.*, vol. 161, pp. 1–14, 2003.

[19] A. H. Poonawalla and X. J. Zhou, "Analytical error propagation in diffusion anisotropy calculations," *J. Magn. Reson. Imag.*, vol. 19, pp. 489–498, 2004.

[20] Y. C. Wu, A. S. Field, M. K. Chung, B. Badie, and A. L. Alexander, "Quantitative analysis of diffusion tensor orientation: Theoretical framework," *Magn. Reson. Med.*, vol. 52, pp. 1146–1155, 2004.

[21] L. C. Chang, C. G. Koay, C. Pierpaoli, and P. J. Basser, "Variance of estimated DTI-derived parameters via first-order perturbation methods," *Magn. Reson. Med.*, vol. 57, pp. 141–149, 2007.

[22] J. D. Carew, C. G. Koay, G. Wahba, A. L. Alexander, P. J. Basser, and M. E. Meyerand, "The asymptotic distribution of diffusion tensor and fractional anisotropy estimates," *Proc. Int. Soc. Magn. Reson. Med.*, 2006.

[23] C. G. Koay, L. C. Chang, C. Pierpaoli, and P. J. Basser, "Error propagation framework for diffusion tensor imaging," *Proc. Int. Soc. Magn. Reson. Med.*, 2006.

[24] Y. Shen, I. Pu, and C. Clark, "Analytical expressions for noise propagation in diffusion tensor imaging," *Proc. Int. So. Magn. Reson. Med.*, 2006.

[25] C. G. Koay, L. C. Chang, J. D. Carew, C. Pierpaoli, and P. J. Basser, "A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging," *J. Magn. Reson.*, vol. 182, pp. 115–125, 2006.

[26] P. J. Basser, "Quantifying errors in fiber-tract direction and diffusion tensor field maps resulting from MR noise," *Proc. Int. Soc. Magn. Reson. Med.*, 1997.

[27] P. J. Basser and S. Pajevic, "Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise," *Magn. Reson. Med.*, vol. 44, pp. 41–50, 2000.

[28] D. K. Jones, "Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI," *Magn. Reson. Med.*, vol. 49, pp. 7–12, 2003.

[29] M. Lazar and A. L. Alexander, "Bootstrap white matter tractography (BOOT-TRAC)," *NeuroImage*, vol. 24, pp. 524–532, 2005.

[30] H. K. Jeong, Y. Lu, Z. Ding, and A. W. Anderson, "Characterizing cone of uncertainty in diffusion tensor MRI," *Proc. Int. Soc. Magn. Reson. Med.*, 2005.

[31] S. I. Amari, *Differential Geometrical Methods in Statistics*. New York: Springer-Verlag, 1985, vol. 28.

[32] S. I. Amari and H. Nagaoka, *Methods of Information Geometry*. Providence, RI: Amer. Math. Soc., 2000, vol. 191.

[33] M. P. d. Carmo, *Riemannian Geometry*. Boston, MA: Birkhäuser, 1992.

[34] E. Cartan, *Leçons sur la Géométrie des Espaces de Riemann*. Paris, France: Gauthier-Villars, 1946.

[35] A. S. Eddington, *The Mathematical Theory of Relativity*, 2nd ed. New York: Cambridge Univ. Press, 1924.

[36] L. P. Eisenhart, *Riemannian Geometry*. Princeton, NJ: Princeton Univ. Press, 1950.

[37] D. Laugwitz, *Differential Geometry and Riemannian Geometry*. New York: Academic, 1965.

[38] M. K. Murray and J. W. Rice, *Differential Geometry and Statistics*. Roca Raton: CRC Press, 1993, vol. 48.

[39] E. Schrödinger, *Space-Time Structure*. New York: Cambridge Univ. Press, 1950.
[40] P. Marriott and M. Salmon, "An introduction to differential geometry in econometrics," in *Applications of Differential Geometry to Econometrics*, P. Marriott and M. Salmon, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2000, pp. 7–63.
[41] E. O. Stejskal and J. E. Tanner, "Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient," *J. Chem. Phys.*, vol. 42, pp. 288–292, 1965.
[42] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[43] P. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. New York: Academic, 1981.
[44] K. M. Hasan, P. J. Basser, D. L. Parker, and A. L. Alexander, "Analytical computation of the eigenvalues and eigenvectors in DT-MRI," *J. Magn. Reson.*, vol. 152, pp. 41–47, 2001.
[45] G. R. Hext, "The estimation of second-order tensors, with related tests and designs," *Biometrika*, vol. 50, pp. 353–373, 1963.
[46] L. C. Chang, D. K. Jones, and C. Pierpaoli, "RESTORE: Robust estimation of tensors by outlier rejection," *Magn. Reson. Med.*, vol. 53, pp. 1088–1095, 2005.
[47] 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 complex DWI," *IEEE Trans. Med. Imag.*, vol. 23, no. 8, pp. 930–939, Aug. 2004.
[48] J. F. Mangin, C. Poupon, C. Clark, D. LeBihan, and I. Bloch, "Distortion correction and robust tensor estimation for MR diffusion imaging," *Med. Image Anal.*, vol. 6, pp. 191–198, 2002.
[49] D. K. Jones and P. J. Basser, "Squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted MR data," *Magn. Reson. Med.*, vol. 52, pp. 979–993, 2004.
[50] C. G. Koay, J. D. Carew, A. L. Alexander, P. J. Basser, and M. E. Meyerand, "Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging," *Magn. Reson. Med.*, vol. 55, pp. 930–936, 2006.
[51] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 2nd ed. New York: McGraw-Hill, 1992.
[52] C. G. Koay and P. J. Basser, "Analytically exact correction scheme for signal extraction from noisy magnitude MR signals," *J. Magn. Reson.*, vol. 179, pp. 317–322, 2006.
[53] D. K. Jones, "The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: A Monte Carlo study," *Mag. Reson. Med.*, vol. 51, pp. 807–815, 2004.
[54] D. K. Jones, C. G. Koay, and P. J. Basser, "It is not possible to design a rotationally invariant sampling scheme for DT-MRI," *Proc. Int. Soc. Magn. Reson. Med.*, 2007.
[55] P. J. Basser and S. Pajevic, "A normal distribution for tensor-valued random variables: Applications to diffusion tensor MRI," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 785–794, Jul. 2003.
[56] C. Bingham, "An antipodally symmetric distribution on the sphere," *Ann. Stat.*, vol. 2, pp. 1201–1225, 1974.