Statistical Analysis of Diffusion Tensors in Diffusion-Weighted Magnetic Resonance Imaging Data [With Omments, Rejoinder]

Author(s): Hongtu Zhu, Heping Zhang, Joseph G. Ibrahim, Bradley S. Peterson, Armin Schwartzmann and Nicole A. Lazar

Source: *Journal of the American Statistical Association*, Dec., 2007, Vol. 102, No. 480 (Dec., 2007), pp. 1085-1113

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/27639963

Statistical Analysis of Diffusion Tensors in Diffusion-Weighted Magnetic Resonance Imaging Data

Hongtu ZHU, Heping ZHANG, Joseph G. IBRAHIM, and Bradley S. PETERSON

Diffusion tensor imaging has been widely used to reconstruct the structure and orientation of fibers in biological tissues, particularly in the white matter of the brain, because it can track the effective diffusion of water along those fibers. The raw diffusion-weighted images from which diffusion tensors are estimated, however, inherently contain noise. Noise in the images produces uncertainty in the estimation of the tensors (which are 3×3 positive-definite matrices) and of their derived quantities, including eigenvalues, eigenvectors, and the fiber pathways that are reconstructed based on those tensor elements. The aim of this article is to provide a comprehensive theoretical framework of statistical inference for quantifying the effects of noise on diffusion tensors, on their eigenvalues and eigenvectors, and on their morphological classification. We propose a semiparametric model to account for noise in diffusion-weighted images. We then develop a one-step, weighted least squares estimate of the tensors and justify use of the one-step estimates based on our theoretical framework and computational results. We also quantify the effects of noise on the eigenvalues and eigenvectors of the estimated tensors by establishing their limiting distributions. We construct pseudo-likelihood ratio statistics to classify tensor morphologies. Simulation studies show that our theoretical results can accurately predict the stochastic behavior of the estimated eigenvalues and eigenvectors, as well as the bias that is introduced by sorting the eigenvalues by their magnitudes. Implementation of these methods is illustrated in a diffusion-weighted dataset from seven healthy human subjects.

KEY WORDS: Diffusion tensor; Eigenvalues; Eigenvectors; Principal direction; Random matrices; Weighted least squares.

1. INTRODUCTION

Diffusion Tensor Imaging (DTI) tracks the effective diffusion of water in the human brain in vivo. Because water tends to diffuse along the pathways of white-matter fibers, tracking its diffusion with DTI allows investigators to map the microstructure and organization of those pathways (Basser and Jones 2002; Le Bihan 2003; Kingsley 2006a–c). DTI geometrically characterizes diffusion within each voxel of an imaging space as a 3×3 diffusion tensor $\mathbf{D}$, with three eigenvalue–eigenvector pairs $\{(\lambda_i, \mathbf{v}_i) : i = 1, 2, 3\}$ quantifying the direction and degree of diffusivity, respectively, where $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Many tractography algorithms attempt to reconstruct fiber tracts by consecutively connecting the principal directions ($\mathbf{v}_1$) of the diffusion tensors (DTs) in adjacent voxels (Conturo et al. 1999; Xu, Mori, Solaiyappan, van Zijl, and Davatzikos 2002). Statistical analysis of estimated DTI measures (e.g., eigenvalues and eigenvectors) and fiber tracts can provide a quantitative assessment for the integrity of anatomical connectivity in white matter. In turn, the results from these statistical analyses can be used to understand better the development and disturbances of white matter in the central nervous system. DTI has been used to study a wide array of neurological and neuropsychiatric illnesses (Lim and Helpern 2002; Brain Development Cooperative Group and Evans 2006).

DTs are estimated from the raw data contained in diffusion-weighted (DW) images. The process of transforming DW images into estimated DTs that can be used for accurate tracking of fiber pathways entails a number of steps. First, DW images inherently contain varying amounts of noise that must be modeled appropriately if DTs are to be estimated accurately; failure to do so may lead to a biased estimate of DTs and to an incorrect estimate of their covariance matrices. After appropriately modeling the noise in DW images and estimating the tensors, we must then quantify the effects of noise on the estimated eigenspace components. Because many algorithms for fiber tracking reconstruct the directions of fiber pathways based on the principal directions of diffusion, quantifying the effects of noise on these eigenspace components in particular is crucially important for the accurate tracking of fibers. However, because the noise-induced stochastic behavior of the principal direction of a tensor is primarily determined by the overall morphology of the tensor, we must first classify that morphology, which is typically designated as nondegenerate (in which all eigenvalues differ), oblate ($\lambda_1 = \lambda_2 > \lambda_3$), prolate ($\lambda_1 > \lambda_2 = \lambda_3$), or isotropic ($\lambda_1 = \lambda_2 = \lambda_3$). When the effects of noise on the eigenspace components have been assessed and the morphologies of the DTs have been classified, fiber tracking can begin.

Three statistical questions emerge from this process of transforming diffusion-weighted images into estimated DTs and eigenspace components: (1) How can we obtain an accurate estimate of the diffusion tensor and its covariance matrix when the diffusion-weighted magnetic resonance (MR) images contain various noise components, including random and structured noise (such as noise from bulk motion or cardiac pulsation)? (2) How can we quantify the effects of noise on the DTs, including their eigenvalues and eigenvectors? (3) Does the presence of the noise that is inherent in DW images affect in any way our morphological classifications of DTs and, if so, how? In this article, we will address these three questions systematically and rigorously within a statistical theoretical framework.

Hongtu Zhu is Associate Professor of Biostatistics, Department of Biostatistics and Biomedical Research Imaging Center, University of North Carolina, Chapel Hill, NC 27599-7420 (E-mail: *hzhu@bios.unc.edu*). Heping Zhang is Professor of Biostatistics, Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520-8034 (E-mail: *heping.zhang@yale.edu*). Joseph G. Ibrahim is Alumni Distinguished Professor of Biostatistics, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420 (E-mail: *ibrahim@bios.unc.edu*). Bradley S. Peterson is Professor of Psychiatry, Department of Psychiatry, Columbia University Medical Center and the New York State Psychiatric Institute, 1051 Riverside Drive, Unit 74, New York, NY 10032 (E-mail: *petersob@childpsych.columbia.edu*). We thank the editor, the associate editor, and two anonymous referees for valuable suggestions, which greatly helped to improve our presentation. Thanks to Dr. Jason Royal for his invaluable editorial assistance. This work was supported in part by NSF grant SES-06-43663 to Dr. Zhu, NIDA grants DA016750 and DA017713 to Dr. Zhang, NIDA grant DA017820 and NIMH grants MH068318 and K02-74677 to Dr. Peterson, NIH grants GM 70335 and CA 74015 to Dr. Ibrahim, as well as by the Suzanne Crosby Murphy Endowment at Columbia University Medical Center and by the Thomas D. Klingenstein and Nancy D. Perlman Family Fund.

Journal of the American Statistical Association
December 2007, Vol. 102, No. 480, Applications and Case Studies
DOI 10.1198/016214507000000581

Much effort has been devoted to modeling appropriately the noise components of DW images so as to improve the accuracy of estimating at each voxel a diffusion tensor and its derived quantities, such as its principal direction. In the presence of random noise only, the signal intensity in DW images follows a Rician distribution (Henkelman 1985; Gudbjartsson and Patz 1995). In the presence of only a small amount of random noise within DW images, the log-transformed signal intensity can be approximated by a weighted Gaussian distribution (Basser, Mattiello, and Le Bihan 1994; Anderson 2001; Salvador et al. 2005). However, in addition to random noise, DW images always contain varying amounts of noise from other sources (e.g., susceptibility artifacts and rigid-body motion). Although some postprocessing techniques, including image coregistration, may be applied to correct for the presence of the noise from other sources, these techniques can significantly alter the properties of the noise in DW images, including its distribution and variance (Rohde, Barnett, Basser, and Pierpaoli 2005). Thus, the distribution of noise in DW images will likely deviate from both the Gaussian and Rician distributions (Rohde, Barnett, Basser, Marenco, and Pierpaoli 2004), and any strategy for modeling of noise in the postprocessed DW images must extend beyond the sole application of Gaussian and Rician distributions.

Given that noise in DW images also introduces uncertainty into the eigenvalues and eigenvectors of the DTs (Jones 2003; Lazar and Alexander 2003), numerical simulations have been used increasingly to quantify uncertainty in the three eigenvalue–eigenvector pairs of the estimated tensors, as well as to assess how these estimated eigenspace components ultimately influence the performance of tractography algorithms. These simulations have shown, for example, that estimates of the largest eigenvalue in a tensor usually overestimate the true value of λ_1 and that estimates of the smallest eigenvalue usually underestimate λ_3 (Pierpaoli and Basser 1996). These differences between the estimated and true eigenvalues, referred to as "sorting bias," subsequently bias the estimation of invariant measures that are calculated from the values of these estimated eigenvalues (Pierpaoli and Basser 1996; Basser and Pajevic 2000). Although previous investigators have derived first- and second-order expansions of the estimated eigenvalues and eigenvectors for nondegenerate tensors (Anderson 2001), their results cannot predict the bias observed in degenerate tensors and their derived quantities, such as eigenvalues (Basser and Pajevic 2003). A nonparametric bootstrapping method (Efron 1979; Efron and Tibshirani 1993; Basser and Jones 2002, p. 465; Pajevic and Basser 2003) has also been used to quantify numerically the effects of noise on the eigenvalues and eigenvectors, and fiber tracts. However, because bootstrapping methods do rely on asymptotic results (Shao and Tu 1995), one can question whether approximating the uncertainty of eigenvalues and eigenvectors, and particularly the trajectories of fiber tracts using the bootstrapping methods, is ultimately valid. Therefore, mathematically quantifying the effects of noise on the eigenvalues and eigenvectors of the diffusion tensors and fiber tracts is of paramount importance.

Numerous invariant measures of anisotropy have been developed for the classification of tensor morphologies (Basser 1997; Skare, Li, Nordell, and Ingvar 2000; Hasan, Basser, Parker, and Alexander 2001). Examples include Fractional Anisotropy (FA; Basser 1997), for which small values indicate that the diffusion tensor is nearly isotropic. Comparing a specific invariant measure with a predefined fixed value, or "threshold," is often used to determine whether a tensor is degenerate within a particular voxel and, therefore, whether a tractography algorithm should terminate, signaling the end of that particular fiber pathway (Mori and van Zijl 2002; Lazar and Alexander 2005). Thresholds are often selected arbitrarily (a common FA threshold, for example, is .20; Jones 2003), producing either large Type I or Type II errors in classifying tensor morphologies (Zhu et al. 2006). Therefore, developing sensitive measures of tensor morphology, as well as a rigorous and rational strategy for determining thresholds of these measures that capture within a single scalar index one of several of the most salient features of that morphology, is critially important for the correct morphological classification of diffusion tensors and, ultimately, for the valid reconstruction of fiber tracts.

We propose herein a set of three solutions for modeling noise in DW images. First, we propose use of a semiparametric model, which allows for a large class of distributions for the noise component, to fit the log-transformed signal intensities in diffusion-weighted MR data. Second, we propose a one-step Weighted Least Squares (WLS) estimate of the diffusion tensors in this semiparametric model (Carroll, Wu, and Ruppert 1988). Calculating the one-step WLS estimate of the tensors across all voxels in an imaging volume is computationally highly efficient, which is valuable when employing computationally intensive statistical methods such as nonparametric bootstrapping. Third, under the semiparametric model, we quantify the effects of noise on the tensor estimation by establishing a strong convergence rate and by obtaining the covariance matrix of the one-step WLS estimate of the tensor.

We statistically quantify the effects of noise on the eigenvalues and eigenvectors of the estimated tensors. Noise can introduce error into estimation of these components and into the classification of tensor morphology; therefore, even if a tensor in reality has equal eigenvalues (i.e., even if it is "degenerate"), noise makes those estimated tensors distinct in their estimated values. Thus, degenerate tensors can always be estimated and classified as nondegenerate, yielding erroneous principal directions of diffusion. Fiber tracking based on these erroneous principal directions will, in turn, produce fiber pathways that are incorrectly reconstructed. However, because the distinctness of the three estimated eigenvalues is insufficient for quantifying the effects of noise on all eigenspace components of the tensors, we must derive the asymptotic expansions and limiting distributions of the eigenvalues and eigenvectors of both the degenerate and nondegenerate diffusion tensors.

We reformulate the morphological classification problem within a hypothesis testing framework so as to provide a means of estimating confidence when classifying the morphology of any given tensor. We develop three sensitive measures of tensor morphology using pseudo-likelihood ratio statistics, and then determine rigorous thresholds of those statistics based on their limiting distributions under the null hypothesis.

Section 2 presents solutions to the statistical issues we have just outlined. In Section 3 we conduct simulation studies to evaluate the effects of noise on estimation of eigenvalues and

eigenvectors, and we assess the finite performance of the one-step WLS estimate of the tensors and the pseudo-likelihood ratio statistics. Section 4 illustrates an application of the proposed methods in a real dataset. We present concluding remarks in Section 5.

2. THEORY

2.1 Heteroscedastic Linear Model

We usually acquire n diffusion-weighted images for each subject, with each image containing N voxels, and each of those voxels consisting of n diffusion-weighted measurements. Let $\{(S_i, \mathbf{r}_i, b_i) : i = 1, \ldots, n\}$ be n diffusion-weighted measurements at a single voxel in the human brain, where S_i denotes the signal intensity of the MR image, $\mathbf{r}_i = (r_{i,1}, r_{i,2}, r_{i,3})^T$ is the ith direction of the diffusion gradient such that $\mathbf{r}_i^T \mathbf{r}_i = 1$, and b_i is the corresponding b factor of each ith diffusion-weighted MR image. The b factor denotes the magnitude of the diffusion gradients (Stejskal and Tanner 1965; Anderson 2001; Kingsley 2006b).

In magnetic resonance imaging, we often need to characterize random noise in the magnitude of the observed signal intensity. The magnitude is generated by the square root of the sum of two squared numbers. If these two numbers are independent normal random variables, then their magnitude follows a Rician distribution (Henkelman 1985; Gudbjartsson and Patz 1995; Rowe 2005). Specifically, $S_i = \sqrt{R_i^2 + I_i^2}$ is the magnitude of the complex-data (R_i, I_i) in a given voxel at the ith acquisition for $i = 1, \ldots, n$. Let ϕ_i be the phase data in a given voxel such that $R_i = S_i \sin(\phi_i)$ and $I_i = S_i \cos(\phi_i)$. If the signal intensities contain only random noise, then R_i and I_i are independent and follow Gaussian distributions with the same variance σ^2 and with means $\mu_{R,i}$ and $\mu_{I,i}$, respectively. Thus, using the Jacobian transformation, the joint density of (S_i, ϕ_i) can be written as

$$p(S_i, \phi_i) = \frac{S_i}{2\pi\sigma^2} \exp\{-.5\sigma^{-2}(S_i \sin(\phi_i) - \mu_{R,i})^2 - .5\sigma^{-2}(S_i \cos(\phi_i) - \mu_{I,i})^2\}.$$

Integrating out ϕ_i, we obtain a Rician distribution with parameters μ_i and σ^2, which is given by

$$p(S_i|\mu_i, \sigma^2) = \frac{S_i}{\sigma^2} \exp\{-.5\sigma^{-2}(S_i^2 + \mu_i^2)\} B_0\left(\frac{\mu_i S_i}{\sigma^2}\right) \times \mathbf{1}(S_i \geq 0), \quad (1)$$

where $\mu_i = \sqrt{\mu_{R,i}^2 + \mu_{I,i}^2}$, $\mathbf{1}(\cdot)$ is an indicator function, and $B_0(z)$ denotes the zeroth-order modified Bessel function of the first kind. For diffusion-weighted images, a simple model of diffusion assumes $\mu_i = S_0 \exp(-b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i)$, where $\mathbf{D}$ is a 3×3 diffusion tensor and S_0 is the signal intensity in the absence of any diffusion-weighted gradient. The distribution of $\log(S_i)$ has been shown to be well approximated by a Gaussian distribution with mean $\log \mu_i$ and variance σ^2/μ_i^2 (Salvador et al. 2005) when the value of μ_i/σ is moderate and relatively large (e.g., greater than 5), which is the case in most current imaging studies.

We consider a heteroscedastic linear model to fit the log-transformed signal intensities $\log S_i$ as follows:

$$\log S_i = \log S_0 - b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i + \eta_i = \mathbf{z}_i^T \theta + \exp(-\mathbf{z}_i^T \theta)\sigma\epsilon_i \quad (2)$$

for $i = 1, \ldots, n$, where $\theta^T = (\log S_0, \beta^T)$, $\eta_i = \exp(-\mathbf{z}_i^T \theta)\sigma\epsilon_i$, and the errors ϵ_i are independent random variables that have zero means and finite variances. We define $\beta^T = (D_{11}, D_{12}, D_{13}, D_{22}, D_{23}, D_{33})$ and $\mathbf{z}_i^T = (1, -b_i(r_{i,1}^2, 2r_{i,1}r_{i,2}, 2r_{i,1}r_{i,3}, r_{i,2}^2, 2r_{i,2}r_{i,3}, r_{i,3}^2)^T)^T$. We set $\mathrm{Var}(\epsilon_1)$ at 1 for identifiability purposes. Model (2) allows a large class of distributions for ϵ_i, including the Gaussian distribution, and different distributions and variances for differing ϵ_i. If all the ϵ_i are standard Gaussian random variables, then model (2) reduces to a Gaussian model (Anderson 2001; Salvador et al. 2005).

The WLS algorithm for model (2) can be summarized as follows:

- In step 1, set $k = 0$ and select an initial estimate $\widehat{\theta}^{(k)}$, such as the ordinary least squares estimate $\widehat{\theta}_{\mathrm{LS}} = (\sum_{i=1}^n \mathbf{z}_i \times \mathbf{z}_i^T)^{-1} \sum_{i=1}^n \mathbf{z}_i \log S_i$.
- In step 2, calculate $\omega_i^{(k)} = \exp(2\mathbf{z}_i^T \widehat{\theta}^{(k)})$ for $i = 1, \ldots, n$.
- In step 3, update $\widehat{\theta}^{(k)}$ to $\widehat{\theta}^{(k+1)}$ by using

$$\widehat{\theta}^{(k+1)} = \left(\sum_{i=1}^n \omega_i^{(k)} \mathbf{z}_i \mathbf{z}_i^T\right)^{-1} \sum_{i=1}^n \omega_i^{(k)} \mathbf{z}_i \log S_i. \quad (3)$$

- In step 4, repeat steps 2 and 3 for k_0 iterations and obtain $\widehat{\theta}^{(k_0)}$.
- In step 5, estimate $\widehat{\sigma}^2 = \sum_{i=1}^n (\log S_i - \mathbf{z}_i^T \widehat{\theta}^{(k_0)})^2 \omega_i^{(k_0)} / (n-7)$.

The WLS estimates are computationally simple in that they require simple algebraic manipulations; they also have some good statistical properties, such as robustness against small misspecifications in the variances of the errors (Carroll and Ruppert 1982a,b; Carroll et al. 1988). Moreover, for any reasonable $\widehat{\theta}^{(0)}$, the number of iterations k_0 in the algorithm can be as small as $k_0 = 1$, because all WLS estimates $\widehat{\theta}^{(k_0)}$ for any $k_0 \geq 1$ are asymptotically equivalent (Carroll and Ruppert 1982a). The second-order asymptotic expansion of $\widehat{\theta}^{(k_0)}$ reveals that only one iteration starting from $\widehat{\theta}_{\mathrm{LS}}$ is needed to obtain an efficient estimate of θ under certain conditions (Carroll et al. 1988). Numerically, when $n = 30$, results from a simulation study in Section 3.1 reveal that the WLS estimates $\widehat{\theta}^{(1)}$ and $\widehat{\theta}^{(5)}$ are very close. Thus, we can use $\widehat{\theta}_{\mathrm{LS}}$ as an initial estimate and take $\widehat{\theta}^{(1)}$ as the final WLS estimate of θ.

We introduce some notation to characterize the properties of the WLS estimates of diffusion tensors. Let $\theta_*^T = (\log S_{0*}, \beta_*^T)$ be the true value of θ; let D_* denote the diffusion tensor corresponding to β_*; and let $\|\cdot\|$ denote the Euclidean norm of a vector or a matrix. We also define $A_n = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T$, $B_n(\theta) = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T \exp(2\mathbf{z}_i^T \theta)$, $G_n(\theta) = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T \exp(4\mathbf{z}_i^T \theta) E(\eta_i^2)$, and $F_n(\theta) = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T \exp(4\mathbf{z}_i^T \theta) \mathbf{e}_i(\theta)^2$, where $\mathbf{e}_i(\theta) = (\log S_i - \mathbf{z}_i^T \theta)^2$.

We quantify the effects of noise on a diffusion tensor by establishing its strong consistency rate and asymptotic normality. We obtain the following theorems, whose detailed proofs can be found in a supplementary technical report available at *http://www.bios.unc.edu/~hzhu/DTIreport.pdf.*

Theorem 1. (a) If assumptions (C1)–(C3) in the Appendix are satisfied and $\|\widehat{\theta}^{(0)} - \theta_*\| \le \delta'$ for any fixed $\delta' > 0$, then

$$\widehat{\theta}^{(k)} - \theta_* = o\left(\left\{[\log \lambda_{\min}(A_n)]^{1+\delta}/\lambda_{\min}(A_n)\right\}^{1/2}\right) \quad \text{a.s.} \quad (4)$$

for any $\delta > 0$ and $k \ge 1$, where $\lambda_{\min}(A_n)$ is the minimum eigenvalue of A_n.

(b) Under assumptions (C1)–(C5), for any $k \ge 2$, we have

$$\left[G_n\left(\widehat{\theta}^{(k)}\right)\right]^{-1/2} B_n\left(\widehat{\theta}^{(k)}\right)\left(\widehat{\theta}^{(k)} - \theta_*\right) \to^L \mathrm{N}(0, \mathbf{I}_7) \quad (5)$$

as $n \to \infty$, where $\mathbf{I}_7$ is the 7×7 identity matrix and $\to^L$ denotes convergence in distribution. If $\widehat{\theta}^{(0)} = \widehat{\theta}_{\mathrm{LS}}$, then (5) holds for any $k \ge 1$.

(c) Under assumptions (C1)–(C6), for any $k \ge 2$, we have

$$G_n(\theta_*)^{-1/2} F_n\left(\widehat{\theta}^{(k)}\right) G_n(\theta_*)^{-1/2} - \mathbf{I}_7 \to \mathbf{0}_7 \quad \text{in probability (elementwise)}, \quad (6)$$

where $\mathbf{0}_7$ is a 7×7 matrix with all zero entries.

Theorem 1 explicitly gives the strong convergence rate and the covariance matrix of $\widehat{\theta}^{(k)}$. If $\lambda_{\min}(A_n)$ is $O(n)$, then $\hat{\theta}^{(k)}$ converges almost surely to θ_* at the rate of $o(n^{-1/2}(\log n)^{1/2+\delta})$ for any $\delta > 0$. The covariance matrix of $\widehat{\theta}^{(k)}$ under model (2) differs from that of the WLS estimate under the Gaussian model with homogeneous variance (Anderson 2001; Kingsley 2006c). However, according to Theorem 1(b) and (c), the covariance matrix of $\widehat{\theta}^{(k)}$ under model (2) can be consistently estimated by $[B_n(\widehat{\theta}^{(k)})]^{-1} F_n(\widehat{\theta}^{(k)})[B_n(\widehat{\theta}^{(k)})]^{-1}$. Furthermore, we propose an empirically better estimator of $\mathrm{Cov}[\widehat{\theta}^{(k)}]$, denoted by $\widehat{\Sigma}^{(k)}$, as follows:

$$\left[B_n\left(\widehat{\theta}^{(k)}\right)\right]^{-1}\left[\sum_{i=1}^{n} \mathbf{z}_i^T \mathbf{z}_i \exp\left(4\mathbf{z}_i^T \widehat{\theta}^{(k)}\right)\left(\log S_i - \mathbf{z}_i^T \widehat{\theta}^{(k)}\right)^2 \times \left(1 - t_i^{(k)}\right)^{-1}\right]\left[B_n\left(\widehat{\theta}^{(k)}\right)\right]^{-1}, \quad (7)$$

where $t_i^{(k)} = \omega_i^{(k)} \mathbf{z}_i^T (\sum_{j=1}^n \omega_j^{(k)} \mathbf{z}_j \mathbf{z}_j^T)^{-1} \mathbf{z}_i$. Compared with $[B_n(\widehat{\theta}^{(k)})]^{-1} F_n(\widehat{\theta}^{(k)})[B_n(\widehat{\theta}^{(k)})]^{-1}$, the estimate in (7) is better because we have explicitly accounted for the variability in the estimated residuals $\mathbf{e}_i(\widehat{\theta}^{(k)})$ (MacKinnon and White 1985).

2.2 Effects of Noise on Eigenvalues and Eigenvectors

We consider a decomposition of $\mathbf{D}$ as $\mathbf{D} = \Gamma \Lambda \Gamma^T$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$ and $\Gamma = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is an orthogonal matrix. Geometrically, diffusion tensors can be represented as an ellipsoid describing three eigenvectors $\{\mathbf{v}_i, i = 1, 2, 3\}$ scaled with the square root of their corresponding eigenvalues $\{\lambda_i, i = 1, 2, 3\}$. An elongated ellipsoid represents high diffusivity in the principal direction $\mathbf{v}_1$ associated with λ_1, which may be interpreted as the dominant orientation of fibers passing through that particular voxel. However, in isotropic tensors, the principal diffusion could be any direction in three-dimensional space; in oblate tensors, any direction on the plane orthogonal to $\mathbf{v}_3$ could be the principal direction. Currently, those oblate and isotropic tensors pose a significant challenge for existing algorithms for fiber tracking (Mori and van Zijl 2002; Parker, Haroon, and Wheeler-Kingshott 2003).

In practice, we can only obtain $\widehat{\mathbf{D}}$ and its three eigenvalue–eigenvector pairs denoted by $\{(m_i, \mathbf{e}_i) : i = 1, 2, 3\}$, such that $m_1 \ge m_2 \ge m_3$. Thus, $\widehat{\mathbf{D}} = \mathbf{E}\mathbf{M}\mathbf{E}^T$, where $\mathbf{M} = \mathrm{diag}(m_1, m_2, m_3)$ and $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is an orthogonal matrix. Because of the presence of noise that is inherent in diffusion-weighted MR images, $\{(m_i, \mathbf{e}_i) : i = 1, 2, 3\}$ are generally different from the true eigenvalue–eigenvector pairs $\{(\lambda_i, \mathbf{v}_i) : i = 1, 2, 3\}$. For instance, previous simulation studies have shown that the estimated eigenvalues $\{m_i : i = 1, 2, 3\}$ are always distinct regardless of the presence of degenerate and nondegenerate tensors (Pierpaoli and Basser 1996; Basser and Pajevic 2000). Falsely attributing distinct directionality to the principal directions of the tensors that are in reality degenerate will wreak havoc for the current algorithms for fiber tracking. The distinctness of $\{m_i : i = 1, 2, 3\}$ has not yet been investigated theoretically.

In the following, we establish the distinctness of the three eigenvalues for $\widehat{\mathbf{D}}$, which are determined by $\widehat{\theta}^{(1)}$ starting from $\widehat{\theta}_{\mathrm{LS}}$.

Theorem 2. (a) If assumption (C7) in the Appendix is satisfied, then the three eigenvalues of $\widehat{\mathbf{D}}$ based on $\widehat{\theta}_{\mathrm{LS}}$ are distinct with probability 1 when $n \ge 7$.

(b) If assumptions (C1)–(C3) and (C7) in the Appendix are satisfied and $\widehat{\theta}^{(0)} = \widehat{\theta}_{\mathrm{LS}}$ satisfies $\|\widehat{\theta}_{\mathrm{LS}} - \theta_*\| \le \delta'$ for a given $\delta' > 0$, then the eigenvalues of $\widehat{\mathbf{D}}$ based on $\widehat{\theta}^{(1)}$ are distinct with probability 1 when $n \ge 7$.

Theorem 2 reveals that the distinctness of the estimated eigenvalues persists in all regions of an image in vivo, confirming the sorting bias (Pierpaoli and Basser 1996; Basser and Pajevic 2000). Therefore, we always conclude that $m_1 > m_2 > m_3$, and we obtain incorrect principal directions of diffusion within the regions that contain isotropic and oblate tensors.

Because the distinctness of the estimated eigenvalues is not adequate for understanding the stochastic behavior of $\{(m_i, \mathbf{e}_i) : i = 1, 2, 3\}$, we derive the limiting distributions of the eigenvalues and eigenvectors of $\widehat{\mathbf{D}}$ for both degenerate and nondegenerate tensors.

We introduce the following notation. Recall that $\mathbf{D} = \Gamma \Lambda \Gamma^T$, $\Gamma^T \mathbf{D} \Gamma = \Lambda$, and $\widehat{\mathbf{D}} = \mathbf{E}\mathbf{M}\mathbf{E}^T$. We use $\mathrm{Vecs}(\mathbf{U})$ to represent $(U_{11}, U_{12}, U_{13}, U_{22}, U_{23}, U_{33})^T$ for any 3×3 symmetric matrix $\mathbf{U} = (U_{ij})$. Thus, using Theorem 1(b), we have

$$\mathbf{U}_n = \sqrt{n}(\mathbf{T}_n - \Lambda) = \sqrt{n}(\Gamma^T \widehat{\mathbf{D}} \Gamma - \Lambda) \to^L \mathbf{U}, \quad (8)$$

where $\mathrm{Vec}(\mathbf{U})$ is a multivariate normal random vector with mean $\mathbf{0}$ and covariance matrix $\Sigma_{\mathbf{U}}$. Furthermore, let $\mathbf{C}_n^T = \Gamma^T \mathbf{E}$. Then $\mathbf{T}_n$ can be written as $\mathbf{T}_n = \Gamma^T \widehat{\mathbf{D}} \Gamma = \mathbf{C}_n^T \mathbf{M} \mathbf{C}_n$ and $\mathbf{C}_n^T \mathbf{C}_n = \mathbf{I}_3$.

Theorem 3. If assumptions (C8) and (C9) in the Appendix are satisfied and if $\mathbf{D}$ is an isotropic tensor, then the density of the limiting distribution of $\mathbf{H}_n = \mathrm{diag}(h_{n1}, h_{n2}, h_{n3}) = \sqrt{n}\,\mathrm{diag}(\mathbf{M} - \lambda \mathbf{I}_3)$ and $\mathbf{E}$, denoted by $p(\mathbf{h}, \mathbf{C})$, is proportional to

$$(h_1 - h_2)(h_2 - h_3)(h_1 - h_3) \times \exp\left\{-\frac{1}{2}\,\mathrm{Vecs}(\mathbf{C}^T \mathbf{H} \mathbf{C})^T \Sigma_{\mathbf{U}}^{-1}\,\mathrm{Vecs}(\mathbf{C}^T \mathbf{H} \mathbf{C})\right\}, \quad (9)$$

where $\mathbf{C} = (c_{ij})$, a 3×3 matrix, satisfies $\mathbf{C}^T \mathbf{C} = \mathbf{I}_3$ and $c_{ii} > 0$ for $i = 1, 2, 3$ and $h_1 > h_2 > h_3$, where $\mathbf{H} = \mathrm{diag}(\mathbf{h})$ and $\mathbf{h} = (h_1, h_2, h_3)^T$. In addition, $E(h_1) > E(h_2) = 0 > E(h_3)$ and $E(h_1 + h_2 + h_3) = 0$, where E denotes the expectation with respect to $p(\mathbf{h}, \mathbf{C})$ given in (9).

For the oblate tensor, we must introduce additional notation, as follows:

$$\Lambda = \begin{pmatrix} \lambda_1 \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0}^T & \lambda_3 \end{pmatrix}, \qquad \mathbf{U}_n = \begin{pmatrix} \mathbf{U}_{n,11} & \mathbf{U}_{n,12} \\ \mathbf{U}_{n,21} & \mathbf{U}_{n,22} \end{pmatrix},$$

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{0} \\ \mathbf{0}^T & m_3 \end{pmatrix}, \qquad \mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix},$$

$$\mathbf{H}_n = \begin{pmatrix} \mathbf{H}_{n,1} & \mathbf{0} \\ \mathbf{0}^T & h_{n,3} \end{pmatrix}, \qquad \mathbf{C}_n = \begin{pmatrix} \mathbf{C}_{n,11} & \mathbf{C}_{n,12} \\ \mathbf{C}_{n,21} & \mathbf{C}_{n,22} \end{pmatrix},$$

where $\mathbf{M}_1 = \mathrm{diag}(m_1, m_2)$, $\mathbf{H}_{n,1} = \sqrt{n}(\mathbf{M}_1 - \lambda_1 \mathbf{I}_2)$, and $h_{n,3} = \sqrt{n}(m_3 - \lambda_3)$. In addition, we assume $\mathrm{Cov}[\mathrm{Vecs}(\mathbf{U}_{11})] = \Sigma_{\mathbf{U}_{11}}$, $\mathrm{Var}(\mathbf{U}_{22}) = \Sigma_{\mathbf{U}_{22}}$, and $\mathrm{Cov}(\mathbf{U}_{12}) = \Sigma_{\mathbf{U}_{12}}$.

Theorem 4. If assumptions (C8) and (C9) in the Appendix are satisfied and if $\mathbf{D}$ is an oblate tensor, then we can conclude that:

(a) The density of the limiting distribution of $\mathrm{diag}(\mathbf{H}_{n,1})$ and $\mathbf{C}_{n,11}$ is proportional to

$$(h_1 - h_2) \exp\left\{ -\frac{1}{2} \mathrm{Vecs}(\mathbf{C}_{11}^T \mathbf{H}_1 \mathbf{C}_{11})^T \Sigma_{\mathbf{U}_{11}}^{-1} \mathrm{Vecs}(\mathbf{C}_{11}^T \mathbf{H}_1 \mathbf{C}_{11}) \right\}, \tag{10}$$

where $\mathbf{C}_{11} = (c_{11}, c_{12}; c_{21}, c_{22})$ is a 2×2 matrix satisfying $\mathbf{C}_{11}^T \mathbf{C}_{11} = \mathbf{I}_2$, $c_{11} > 0$, and $c_{22} > 0$, and $\mathbf{H}_1 = \mathrm{diag}(h_1, h_2)$ such that $h_1 > h_2$. In addition, $E(h_1 + h_2) = 0$ and $E(h_2) < 0 < E(h_1)$. In general,

$$\mathbf{C}_{n,11}^T \mathbf{H}_{n,1} \mathbf{C}_{n,11} = \mathbf{U}_{n,11} + n^{-1/2}(\lambda_1 - \lambda_3)^{-1} \mathbf{U}_{n,12} \mathbf{U}_{n,12}^T + o_p\left(n^{-1/2}\right). \tag{11}$$

(b) As $n \to \infty$, $h_{n,3} \to^L \mathbf{U}_{22}$ and

$$h_{n,3} = \mathbf{U}_{n,22} - n^{-1/2}(\lambda_1 - \lambda_3)^{-1} \mathbf{U}_{n,12}^T \mathbf{U}_{n,12} + o_p\left(n^{-1/2}\right). \tag{12}$$

(c) $\sqrt{n}\mathbf{C}_{n,11}^T \mathbf{C}_{n,12} = -\sqrt{n}\mathbf{C}_{n,21}^T + O_p(n^{-1})$ and

$$\mathbf{C}_{n,22} = 1 - n^{-1}(\lambda_1 - \lambda_3)^{-2} \mathbf{U}_{n,21} \mathbf{C}_{n,11}^T \mathbf{C}_{n,11} \mathbf{U}_{n,12} + o_p(n^{-1}).$$

Furthermore, $-\sqrt{n}\mathbf{C}_{n,21}^T$ can be written as

$$(\lambda_1 - \lambda_3)^{-1}\big[\mathbf{I}_2 - n^{-1/2}(\lambda_1 - \lambda_3)^{-1} \times (\mathbf{C}_{n,11}^T \mathbf{H}_{n,1} \mathbf{C}_{n,11} - \mathbf{U}_{n,22}\mathbf{I}_2)\big]\mathbf{U}_{n,12} + o_p\left(n^{-1/2}\right). \tag{13}$$

Thus, both $-\sqrt{n}\mathbf{C}_{n,12}$ and $\sqrt{n}\mathbf{C}_{n,11}\mathbf{C}_{n,21}^T$ converge to $\mathbf{U}_{12}/(\lambda_1 - \lambda_3)$ in distribution as $n \to \infty$.

(d) The eigenvectors $\{\mathbf{e}_i : i = 1, 2, 3\}$ satisfy

$$\begin{aligned} (\mathbf{e}_1, \mathbf{e}_2) &= (\mathbf{v}_1, \mathbf{v}_2)\mathbf{C}_{n,11}^T + \frac{n^{-1/2}\mathbf{v}_3 \mathbf{U}_{n,12}^T \mathbf{C}_{n,11}^T}{\lambda_1 - \lambda_3} + o_p\left(n^{-1/2}\right), \\ \sqrt{n}(\mathbf{e}_3 - \mathbf{v}_3) &= -(\lambda_1 - \lambda_3)^{-1}(\mathbf{v}_1, \mathbf{v}_2)\mathbf{U}_{n,12} + o_p(1) \\ &= (\mathbf{v}_1, \mathbf{v}_2) n^{1/2} \mathbf{C}_{n,21}^T - \frac{.5 n^{-1/2} \mathbf{v}_3 \mathbf{U}_{n,21} \mathbf{C}_{n,11}^T \mathbf{C}_{n,11} \mathbf{U}_{n,12}{}^2}{\lambda_1 - \lambda_3} + o_p\left(n^{-1/2}\right). \end{aligned} \tag{14}$$

For the prolate tensor, we need to modify the corresponding six matrices introduced for the oblate tensor. In particular, we modify Λ, $\mathbf{M}$, and $\mathbf{H}_n$ as follows:

$$\Lambda = \begin{pmatrix} \lambda_1 & \mathbf{0}^T \\ \mathbf{0} & \lambda_3 \mathbf{I}_2 \end{pmatrix}, \qquad \mathbf{M} = \begin{pmatrix} m_1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{M}_2 \end{pmatrix},$$

$$\mathbf{H}_n = \begin{pmatrix} h_{n,1} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{H}_{n,2} \end{pmatrix},$$

where $\mathbf{M}_2 = \mathrm{diag}(m_2, m_3)$, $\mathbf{h}_{n,1} = \sqrt{n}(m_1 - \lambda_1)$, and $\mathbf{H}_{n,2} = \sqrt{n}(\mathbf{M}_2 - \lambda_3 \mathbf{I}_2)$. We use the same notation for $\mathbf{U}_n$, $\mathbf{U}$, and $\mathbf{C}_n$, although we have made several necessary modifications.

Corollary 1. If assumptions (C8) and (C9) are satisfied and if $\mathbf{D}$ is a prolate tensor, then we have the following results:

(a) The density of the limiting distribution of $\mathrm{diag}(\mathbf{H}_{n,2})$ and $\mathbf{C}_{n,22}$ is proportional to

$$(h_2 - h_3) \exp\left\{ -\frac{1}{2} \mathrm{Vecs}(\mathbf{C}_{22}^T \mathbf{H}_2 \mathbf{C}_{22})^T \Sigma_{\mathbf{U}_{22}}^{-1} \mathrm{Vecs}(\mathbf{C}_{22}^T \mathbf{H}_2 \mathbf{C}_{22}) \right\}, \tag{15}$$

where $\mathbf{C}_{22} = (c_{22}, c_{23}; c_{32}, c_{33})$ is a 2×2 matrix satisfying $\mathbf{C}_{22}^T \mathbf{C}_{22} = \mathbf{I}_2$, $c_{22} > 0$, $c_{33} > 0$, and $\mathbf{H}_2 = \mathrm{diag}(h_2, h_3)$ such that $h_2 > h_3$. In addition, $E(h_2 + h_3) = 0$ and $E(h_3) < 0 < E(h_2)$. In general,

$$\mathbf{C}_{n,22} \mathbf{H}_{n,2} \mathbf{C}_{n,22}^T = \mathbf{U}_{n,22} - n^{-1/2}(\lambda_1 - \lambda_2)^{-1} \mathbf{U}_{n,21} \mathbf{U}_{n,21}^T + o_p\left(n^{-1/2}\right). \tag{16}$$

(b) As $n \to \infty$, $h_{n,1} \to^L \mathbf{U}_{11}$ and

$$h_{n,1} = \mathbf{U}_{n,11} + n^{-1/2}(\lambda_1 - \lambda_3)^{-1} \mathbf{U}_{n,12} \mathbf{U}_{n,12}^T + o_p\left(n^{-1/2}\right). \tag{17}$$

(c) $\sqrt{n}\mathbf{C}_{n,21}^T \mathbf{C}_{n,22} = -\sqrt{n}\mathbf{C}_{n,12} + O_p(n^{-1})$ and

$$\mathbf{C}_{n,11} = 1 - n^{-1}(\lambda_1 - \lambda_3)^{-2} \mathbf{U}_{n,12} \mathbf{C}_{n,22}^T \mathbf{C}_{n,22} \mathbf{U}_{n,21} + o_p\left(n^{-1}\right).$$

Furthermore, $\sqrt{n}\mathbf{C}_{n,12}$ can be written as

$$(\lambda_1 - \lambda_3)^{-1} \mathbf{U}_{n,12}\big[\mathbf{I}_2 + n^{-1/2}(\lambda_1 - \lambda_3)^{-1} \times (\mathbf{C}_{n,22}^T \mathbf{H}_{n,2} \mathbf{C}_{n,22} - \mathbf{U}_{n,11}\mathbf{I}_2)\big] + o_p\left(n^{-1/2}\right). \tag{18}$$

Thus, both $\sqrt{n}\mathbf{C}_{n,12}$ and $-\sqrt{n}\mathbf{C}_{n,21}^T \mathbf{C}_{n,22}$ converge to $\mathbf{U}_{12}/(\lambda_1 - \lambda_3)$ in distribution as $n \to \infty$.

(d) The eigenvectors $\{\mathbf{e}_i : i = 1, 2, 3\}$ satisfy

$$\begin{aligned} \sqrt{n}(\mathbf{e}_1 - \mathbf{v}_1) &= (\lambda_1 - \lambda_3)^{-1}(\mathbf{v}_2, \mathbf{v}_3)\mathbf{U}_{n,12}^T + o_p(1) \\ &= (\mathbf{v}_2, \mathbf{v}_3) n^{1/2} \mathbf{C}_{n,12}^T - \frac{.5 n^{-1/2} \mathbf{v}_1 \mathbf{U}_{n,12} \mathbf{C}_{n,22}^T \mathbf{C}_{n,22} \mathbf{U}_{n,21}{}^2}{\lambda_1 - \lambda_3} + o_p\left(n^{-1/2}\right), \\ (\mathbf{e}_2, \mathbf{e}_3) &= (\mathbf{v}_2, \mathbf{v}_3)\mathbf{C}_{n,22}^T - \frac{n^{-1/2}\mathbf{v}_1 \mathbf{U}_{n,12}^T \mathbf{C}_{n,22}^T}{\lambda_1 - \lambda_3} + o_p\left(n^{-1/2}\right). \end{aligned} \tag{19}$$

For the nondegenerate tensor, we need to modify the six matrices as follows:

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3), \qquad \mathbf{M} = \mathrm{diag}(m_1, m_2, m_3),$$
$$\mathbf{H}_n = \mathrm{diag}(h_{n,1}, h_{n,2}, h_{n,3}), \qquad \mathbf{U} = (U_{ij}),$$
$$\mathbf{U}_n = (U_{n,ij}), \qquad \mathbf{C}_n = (c_{n,ij}).$$

In addition, we define $c_{n,ij} = n^{-1/2} f_{n,ij}$ for $i \neq j$. Let $\lambda_{i,j}$ be $\lambda_i - \lambda_j$ for all $i, j = 1, 2, 3$.

Corollary 2. If assumptions (C8) and (C9) are satisfied and if $\mathbf{D}$ is nondegenerate, then we have the following results:

(a) Let $\sigma_{ii} = \mathrm{Var}(U_{ii})$. Then $h_{n,i} = U_{n,ii} + o_p(1) \to^L \mathrm{N}(0, \sigma_{ii})$ for $i = 1, 2, 3$ and

$$h_{n,i} = U_{n,ii} + n^{-1/2} \sum_{j \neq i} \lambda_{i,j}^{-1} U_{n,ij}^2 + o_p\big(n^{-1/2}\big). \tag{20}$$

(b) $c_{n,ii} = 1 - n^{-1} \sum_{j \neq i} f_{n,ji}^2 + o_p(n^{-1})$ and $f_{n,ij} + f_{n,ji} + n^{-1/2} f_{n,ki} f_{n,kj} \mathbf{1}(k \neq i, k \neq j) + o_p(n^{-1/2}) = 0$ for $i < j$ and $i, j, k = 1, 2, 3$. Moreover, for $i < j$, $k \neq i$, and $k \neq j$,

$$\begin{aligned} f_{n,ij} \lambda_{i,j} &= U_{n,ij} + n^{-1/2} U_{n,ki} U_{n,kj} / \lambda_{i,k} \\ &\quad - n^{-1/2} (U_{n,ii} - U_{n,jj}) U_{n,ij} / \lambda_{i,j} \\ &\quad + o_p\big(n^{-1/2}\big). \end{aligned} \tag{21}$$

(c) The eigenvectors $\{\mathbf{e}_i : i = 1, 2, 3\}$ satisfy

$$\begin{aligned} \sqrt{n}(\mathbf{e}_i - \mathbf{v}_i) &= \sum_{j \neq i} \lambda_{i,j}^{-1} U_{n,ij} \mathbf{v}_j + o_p(1) \\ &= \sum_{j \neq i} \mathbf{v}_j f_{n,ij} - 0.5 \mathbf{v}_i n^{-1/2} \sum_{j \neq i} f_{n,ij}^2 \\ &\quad + o_p\big(n^{-1/2}\big). \end{aligned} \tag{22}$$

Theorems 3 and 4 have several important implications for the analysis of diffusion tensor images. For instance, Theorem 3 gives the explicit form of the joint limiting distribution of the estimated eigenvalues and eigenvectors for an isotropic tensor. Therefore, we can directly sample from (9) to approximate the stochastic behavior of $\{(m_i, \mathbf{e}_i) : i = 1, 2, 3\}$ for an isotropic tensor. Theorem 3 also confirms that m_1 overestimates λ, and m_3 underestimates λ for the isotropic tensor (Pierpaoli and Basser 1996; Basser and Pajevic 2000). Explicitly, because $m_i = \lambda + \sqrt{n}(m_i - \lambda) n^{-1/2}$ can be approximated by $\lambda + h_i n^{-1/2}$, $E(m_i)$ may be close to $\lambda + E(h_i) n^{-1/2}$ for $i = 1, 2, 3$. Thus, we have $E(m_1) > \lambda$, $E(m_2) \approx \lambda$, and $E(m_3) < \lambda$ using Theorem 3. Therefore, compared with m_1 and m_3, m_2 and $\mathrm{tr}(\widehat{\mathbf{D}})/3$ are better estimates of λ with smaller bias, because $\mathrm{tr}(\widehat{\mathbf{D}})/3 \approx \lambda + n^{-1/2}(h_1 + h_2 + h_3)/3$, $E(h_2) = 0$, and $E(\sum_{i=1}^3 h_i) = 0$. We can also construct confidence intervals for the eigenvalues $\{\lambda_i : i = 1, 2, 3\}$. For instance, for a nondegenerate tensor, a $1 - \alpha$ confidence interval of λ_i is given by $[m_i - z_{\alpha/2}\sigma_{ii}, m_i + z_{\alpha/2}\sigma_{ii}]$, where $z_{\alpha/2}$ is an upper $\alpha/2$ percentile of a standard normal distribution. Moreover, we can use (19) and (22) to quantify the variability of the true principal directions for the nondegenerate and prolate tensors.

2.3 Classification of Tensor Morphologies

Following the reasoning described in Zhu et al. (2006), we statistically test three hypotheses to determine the morphology of a tensor. We specify these hypotheses as follows:

$$\begin{aligned} H_0^{(1)} &: \lambda_1 = \lambda_3 \quad \text{vs.} \quad H_1^{(1)} : \lambda_1 \neq \lambda_3, \\ H_0^{(2)} &: \lambda_1 = \lambda_2 \quad \text{vs.} \quad H_1^{(2)} : \lambda_1 \neq \lambda_2, \\ H_0^{(3)} &: \lambda_2 = \lambda_3 \quad \text{vs.} \quad H_1^{(3)} : \lambda_2 \neq \lambda_3. \end{aligned} \tag{23}$$

For a given significance level α, we can test these three hypotheses at every voxel of the image. If we do not reject $H_0^{(1)}$, then we classify the diffusion tensor as isotropic; otherwise, we then test the second and third hypotheses. If $H_0^{(2)}$ is not rejected, but both $H_0^{(1)}$ and $H_0^{(3)}$ are rejected, then we classify the diffusion tensor as oblate because of the lack of evidence that this diffusion tensor is not oblate. If both $H_0^{(1)}$ and $H_0^{(2)}$ are rejected, but $H_0^{(3)}$ is not rejected, then the diffusion tensor is classified as prolate. If all $H_0^{(i)}$ $(i = 1, 2, 3)$ are rejected, then the diffusion tensor is classified as nondegenerate.

For each of the three hypotheses, we develop a pseudo-likelihood ratio test statistic based on a pseudo-log-likelihood function defined by

$$\ell_n(\theta | \widehat{\theta}_{\mathrm{LS}}) = -\sum_{i=1}^n (\log S_i - \mathbf{z}_i^T \theta)^2 \exp(2\mathbf{z}_i^T \widehat{\theta}_{\mathrm{LS}}). \tag{24}$$

The parameter spaces for the three null hypotheses $H_0^{(i)}$ $(i = 1, 2, 3)$ can be written as follows: $\Theta(1) = \{\theta : \lambda_1 = \lambda_3\} \cap \Theta$, $\Theta(2) = \{\theta : \lambda_1 = \lambda_2\} \cap \Theta$, and $\Theta(3) = \{\theta : \lambda_2 = \lambda_3\} \cap \Theta$, where $\Theta = \{\theta : \log S_0 > -\infty, \mathbf{D} \geq \mathbf{0}\}$. Let $\widehat{\theta}(i)$ be the maximizer of $\ell_n(\theta | \widehat{\theta}_{\mathrm{LS}})$ as θ varies in $\Theta(i)$. For each i, the pseudo-likelihood ratio statistic for testing $H_0^{(i)}$ against $H_1^{(i)}$ is defined as

$$\mathrm{PLRT}(i) = 2\big[\ell_n\big(\widehat{\theta}^{(1)} | \widehat{\theta}_{\mathrm{LS}}\big) - \ell_n(\widehat{\theta}(i) | \widehat{\theta}_{\mathrm{LS}})\big]. \tag{25}$$

In the following, we derive the limiting distributions of $\mathrm{PLRT}(i)$ for $i = 1, 2, 3$. Let $X(i)$ $(i = 1, 2, 3)$ be three weighted chi-squared random variables (a weighted chi-squared random variable is a linear combination of independent χ_1^2 random variables; Schott 2003).

Theorem 5. Under assumptions (C1)–(C6) and (C10), the following results hold as $n \to \infty$.

(a) If $H_0^{(1)}$ is true, then $\mathrm{PLRT}(1) \to^L X(1)$.

(b) For $i = 2$ and 3, $\mathrm{PLRT}(i)$ converges, in distribution, either to $X(i)$ for an anisotropic tensor $\mathbf{D}$ or to the maximum of a weighted χ^2 process for an isotropic tensor $\mathbf{D}$ when $H_0^{(i)}$ is true.

(c) If $\mathrm{Var}(\epsilon_i) = \sigma^2$ for $i = 1, \ldots, n$, then $X(1)$ is a $\sigma^2 \chi_5^2$ random variable and $X(2)$ and $X(3)$ follow $\sigma^2 \chi_2^2$ distributions.

Theorem 5 characterizes the limiting distributions of $\mathrm{PLRT}(i)$ under the null hypotheses. In particular, if the variances of the ϵ_i are homogeneous, then we can estimate σ^2 and use χ_5^2 and χ_2^2 as null distributions to test the three hypotheses in (23). However, if the homogeneous variance assumption on the ϵ_i is invalid, then we need to approximate the weighted χ^2 random variables $X(i)$ for $i = 1, 2, 3$. The procedures for approximating $X(i)$ can be found in the supplementary technical report.

3. SIMULATIONS

We conducted three Monte Carlo simulations to illustrate and examine the accuracy of using the asymptotic results under differing signal-to-noise ratios (SNRs; see Secs. 3.1–3.3). First, we compared the stochastic behavior (e.g., bias) of the WLS estimate $\widehat{\theta}^{(k_0)}$ starting from $\widehat{\theta}^{(0)} = \widehat{\theta}_{\mathrm{LS}}$ and evaluated the accuracy of using $\widehat{\Sigma}^{(k_0)}$ in (7) as an estimate of $\mathrm{Cov}[\widehat{\theta}^{(k_0)}]$ for $k_0 = 1$ and $k_0 = 5$. Second, we used the results of Theorems 3 and 4 to predict the stochastic behavior of the estimated eigenvalues and eigenvectors for both degenerate and nondegenerate tensors at low to moderate SNRs (e.g., SNR ≥ 5). Finally, we evaluated the Type I and Type II errors of PLRT(i) ($i = 1, 2, 3$) when used as test statistics for the classification of tensor morphologies.

We generated the simulated diffusion-weighted images as follows. The value of S_0 was fixed at 1,500, but the values of σ_0 were varied to provide differing SNRs (SNR $= S_0/\sigma_0$), such as 5 and 10. Six differing SNRs $\{5, 10, 15, 20, 25, 30\}$ were selected for all Monte Carlo simulations. We used an imaging acquisition scheme $\{(b_i, \mathbf{r}_i) : i = 1, \ldots, 30\}$ that consists of $m = 5$ baseline images with $b = 0$ s/mm^2 and $n - m = 25$ directions of diffusion gradients arranged uniformly in three-dimensional space at $b = 1{,}000$ s/mm^2 (Hardin, Sloane, and Smith 1994). For a given diffusion tensor $\mathbf{D}$, x_i and y_i were generated from a Gaussian random generator with mean 0 and standard deviation σ_0. Finally, we calculated $S_i = \sqrt{(S_0 \exp(-b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i) + x_i)^2 + y_i^2}$ as the resulting diffusion-weighted data at the ith acquisition.

In all simulation studies, we used four diagonal diffusion tensors $\mathbf{D}$. These four diagonal tensors $\mathbf{D}_i$ ($i = 1, 2, 3, 4$), whose three diagonal elements were, respectively, [.7, .7, .7], [.8, .8, .5], [1.0, .55, .55], and [.9, .7, .5] (units: 10^{-3} mm^2/s), were selected to simulate diffusion-weighted data. The four diffusion tensors $\mathbf{D}_i$ ($i = 1, 2, 3, 4$) were, respectively, isotropic, oblate, prolate, and nondegenerate in shape. For each i, the mean diffusivity $\bar{\lambda} = \mathrm{tr}(\mathbf{D}_i)/3$ was set equal to $.7 \times 10^{-3}$ mm^2/s, a value typical in the human brain (Pierpaoli, Jezzard, Basser, Barnett, and Chiro 1996; Anderson 2001).

3.1 Weighted Least Squares Estimates

For each diffusion tensor at each SNR, 10,000 diffusion-weighted datasets were generated. Then, for each simulated diffusion-weighted dataset, we calculated the WLS estimates $\widehat{\theta}^{(k_0)}$ and their corresponding estimates of variance $\mathrm{diag}(\widehat{\Sigma}^{(k_0)})$, when $k_0 = 1$ and 5. We finally calculated the bias, the mean of the standard deviation estimates, and the root mean squared error obtained from the 10,000 estimates based on 10,000 simulated diffusion-weighted datasets.

The one-step WLS estimate $\hat{\theta}^{(1)}$ is numerically close to the five-step WLS estimate $\hat{\theta}^{(5)}$ (see Table 1). Compared with $k_0 = 1$, the larger $k_0 = 5$ leads to less bias in the estimates when SNR $= 5$, but comparable bias in the estimates when SNR ≥ 10. Compared with $\hat{\theta}^{(1)}$, $\hat{\theta}^{(5)}$ has larger root mean squared errors for all six SNRs. All relative efficiencies (the ratio of the mean of the standard deviation estimates to the root mean squared error; RMSE) are close to 1.0, indicating that $\mathrm{diag}(\widehat{\Sigma}^{(k_0)})$ in (7) is an accurate estimate of $\mathrm{diag}(\mathrm{Cov}[\widehat{\theta}^{(k_0)}])$. As expected, the root mean squared error decreases as the value of SNR increases.

3.2 Stochastic Behavior of Eigenvalues and Eigenvectors

We further evaluated the accuracy of the asymptotic results obtained for the estimated eigenvalues and eigenvectors. For each diffusion tensor at each SNR, 10,000 diffusion-weighted datasets were simulated, and then we calculated the WLS estimates $\widehat{\theta}^{(1)}$ and their eigenvalue–eigenvector pairs $\{(m_j, \mathbf{e}_j) : j = 1, 2, 3\}$. Finally, we estimated the means and standard deviations of the eigenvalues and the bias $E(m_i) - \lambda_i$. For each diffusion tensor at each SNR, we also generated eigenvalues and eigenvectors from their asymptotic expansions in Theorems 3 and 4 (see the following paragraphs for a detailed description of the methods). Finally, we compared the results based on the asymptotic results in Theorems 3 and 4 to the empirical results based on the 10,000 simulated datasets for each diffusion tensor at each SNR.

For the isotropic tensor $\mathbf{D}_1$, we used the following procedure to generate eigenvalues and eigenvectors from the density (9). We first used (7) and (8) to calculate the covariance matrix $\Sigma_{\mathbf{U}}$, where the true diffusion tensor $\mathbf{D}_1$ was used. Then, we generated 10,000 3×3 symmetric matrices $\mathbf{U}_{(j)}$ from a Gaussian random generator, where $\mathrm{Vecs}(\mathbf{U}_{(j)})$ followed a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma_{\mathbf{U}}$. We then calculated a decomposition of $\mathbf{U}_{(j)}$ as $\mathbf{C}_{(j)}^T \mathbf{H}_{(j)} \mathbf{C}_{(j)}$ for each j, where $\mathbf{C}_{(j)}$ and $\mathbf{H}_{(j)}$ satisfied the conditions specified in Theorem 3. Thus, we obtained $\{h_{(j),i} : i = 1, 2, 3\}$, the diagonal elements of $\mathbf{H}_{(j)}$, and the three eigenvectors $\{\mathbf{e}_{(j),i} : i = 1, 2, 3\}$ associated with each column of $\mathbf{C}_{(j)}$. Finally, we obtained $\{m_{(j),i} = .7 + n^{-1/2} h_{(j),i} : i = 1, 2, 3\}$ and $\{\mathbf{e}_{(j),i} : i = 1, 2, 3\}$ for all $j = 1, \ldots, J_0$, where $J_0 = 10{,}000$; moreover, for each i, we approximated $E(m_i)$ by the mean value of all $m_{(j),i}$.

For each of the diffusion tensors $\mathbf{D}_i$ ($i = 2, 3, 4$), we used the following procedure to generate eigenvalues and eigenvectors according to the asymptotic results in Theorem 4. For simplicity, we only give detailed information for the oblate tensor $\mathbf{D}_2$ as follows. We calculated $\Sigma_{\mathbf{U}}$ for $\mathbf{D}_2$ using (7) and (8), and then we generated Gaussian random matrices $\{\mathbf{U}_{(j)} : j = 1, \ldots, J_0\}$. Using (11), we obtained both the first-order and the second-order approximations of $\mathbf{C}_{n,11}$ and $\mathbf{H}_{n,1}$. For the first-order approximation, we decomposed $\mathbf{U}_{(j),11}$ directly into $\mathbf{C}_{(j),11}^T \mathbf{H}_{(j),1} \mathbf{C}_{(j),11}$, whereas, for the second-order approximation, we decomposed $\mathbf{U}_{(j),11} + n^{-1/2}(\lambda_1 - \lambda_3)^{-1} \mathbf{U}_{(j),12} \mathbf{U}_{(j),12}^T$ into $\mathbf{C}_{(j),11}^T \mathbf{H}_{(j),1} \mathbf{C}_{(j),11}$, where $\mathbf{U}_{(j),kl}$ ($k, l = 1, 2$) are submatrices of $\mathbf{U}_{(j)}$. Subsequently, we obtained first- and second-order approximations of $\{m_{(j),i} = \lambda_1 + n^{-1/2} h_{(j),i} : i = 1, 2\}$ for $j = 1, \ldots, J_0$, and we calculated the mean values of $m_{(j),i}$ for $i = 1, 2$. We further substituted $\mathbf{U}_{(j)}$ into (12) to obtain $h_{(j),3} = \mathbf{U}_{(j),22} - n^{-1/2}(\lambda_1 - \lambda_3)^{-1} \mathbf{U}_{(j),12}^T \mathbf{U}_{(j),12}$ and calculated $m_{(j),3} = .5 + n^{-1/2} h_{(j),3}$ for all j. We substituted $\mathbf{U}_{(j),12}$ and the second-order approximation of $\mathbf{C}_{(j),11}$ into (14) to obtain the first- and second-order approximations of $\mathbf{e}_{(j),1}$ as follows: For the first-order approximation, $\mathbf{e}_{(j),1}$ was approximated by the normalized vector of $(\mathbf{v}_1, \mathbf{v}_2) \mathbf{C}_{(j),11}^T \mathbf{u}$, whereas, for the second-order approximation, $\mathbf{e}_{(j),1}$ was approximated by the normalized vector of $(\mathbf{v}_1, \mathbf{v}_2) \mathbf{C}_{(j),11}^T \mathbf{u} + n^{-1/2} \mathbf{v}_3 \mathbf{U}_{(j),12}^T \mathbf{C}_{(j),11}^T \mathbf{u}$, where $\mathbf{u}^T = (1, 0)$.

Figure 1 summarizes the results for the isotropic tensor $\mathbf{D}_1$. Based on the simulated DW data, the mean value of m_1 was

Table 1. Bias ($\times 10^{-3}$), RMSE ($\times 10^{-2}$), and SD ($\times 10^{-2}$) of two selected components of $\widehat{\theta}^{(k_0)}$ starting from $\widehat{\theta}^{(0)} = \widehat{\theta}_{\mathrm{LS}}$ for $k_0 = 1$ and 5

	$k_0 = 1$			$k_0 = 5$			$k_0 = 1$			$k_0 = 5$		
SNR	Bias	RMSE	SD	Bias	RMSE	SD	Bias	RMSE	SD	Bias	RMSE	SD
	$\mathbf{D}$: $[D_{11}, D_{12}, D_{13}, D_{22}, D_{23}, D_{33}] = [.7, 0, 0, .7, 0, .7]$ (units: $\times 10^{-3}$ mm^2/s)											
	$D_{11} = .7$						$D_{13} = 0$					
5	−13.37	21.51	20.58	−8.74	22.76	20.66	−.50	15.25	14.69	−.52	16.54	14.77
10	−1.06	10.86	10.60	−.79	11.00	10.61	−.25	7.91	7.64	−.26	8.06	7.65
15	−.16	7.14	7.05	−.10	7.18	7.05	−1.32	5.21	5.08	−1.33	5.30	5.08
20	−.14	5.41	5.27	−.13	5.43	5.27	.43	3.91	3.80	.43	3.93	3.80
25	−.05	4.34	4.22	−.05	4.35	4.22	−.38	3.14	3.06	−.38	3.15	3.06
30	.04	3.62	3.52	.04	3.62	3.52	.21	2.50	2.55	.21	2.61	2.55
	$\mathbf{D}$: $[D_{11}, D_{12}, D_{13}, D_{22}, D_{23}, D_{33}] = [.8, 0, 0, .8, 0, .5]$ (units: $\times 10^{-3}$ mm^2/s)											
	$D_{11} = .8$						$D_{13} = 0$					
5	−19.67	21.97	21.40	−7.39	23.59	21.60	.03	14.78	14.60	.26	15.91	14.32
10	−2.11	11.37	11.06	.06	11.55	11.08	.30	7.59	7.36	.36	7.71	7.36
15	−1.17	7.57	7.37	−.30	7.68	7.38	−2.18	5.02	4.91	−.39	5.05	4.91
20	−.94	5.65	5.55	−.48	5.67	5.55	.36	3.76	3.67	.33	3.77	3.67
25	−.54	4.49	4.43	−.25	4.50	4.43	.33	2.99	2.95	.33	3.00	2.95
30	.39	3.78	3.68	.59	3.78	3.68	.06	2.53	2.46	.06	2.53	2.46
	$\mathbf{D}$: $[D_{11}, D_{12}, D_{13}, D_{22}, D_{23}, D_{33}] = [1.0, 0, 0, .55, 0, .55]$ (units: $\times 10^{-3}$ mm^2/s)											
	$D_{11} = 1.0$						$D_{13} = 0$					
5	−47.43	23.64	22.86	−18.80	25.95	23.29	.10	15.52	15.08	.16	16.91	15.20
10	−5.94	12.41	12.17	1.18	12.73	12.24	−.01	8.08	7.89	−.02	8.25	7.90
15	−3.99	8.30	8.09	−1.05	8.39	8.11	−.55	5.34	5.22	−.54	5.39	5.22
20	−2.05	6.25	6.08	−.43	6.28	6.09	.02	4.02	3.94	.03	4.04	3.94
25	−1.45	4.97	4.86	−.44	4.98	4.87	1.15	3.27	3.14	1.16	3.28	3.14
30	−1.28	4.12	4.05	−.57	4.13	4.05	−.33	2.66	2.62	−.33	2.67	2.62
	$\mathbf{D}$: $[D_{11}, D_{12}, D_{13}, D_{22}, D_{23}, D_{33}] = [.9, 0, 0, .7, 0, .5]$ (units: $\times 10^{-3}$ mm^2/s)											
	$D_{11} = .9$						$D_{13} = 0$					
5	−33.75	23.00	22.24	−13.62	24.93	22.55	−2.58	15.20	14.68	−2.79	16.49	14.75
10	−4.29	11.84	11.57	.15	12.09	11.61	.31	7.79	7.53	.34	7.93	7.54
15	−2.11	7.90	7.73	−.31	7.96	7.74	−.48	5.13	5.04	−.47	5.17	5.04
20	−1.84	5.93	5.80	−.83	5.96	5.80	−.19	3.90	3.77	−.18	3.92	3.77
25	−.27	4.68	4.64	.35	4.69	4.64	−.19	3.10	3.01	−.18	3.11	3.01
30	−.58	4.03	3.87	−.14	4.04	3.87	.43	2.56	2.51	.43	2.57	2.51

NOTE: Bias denotes the bias of the mean of the WLS estimates; RMSE denotes the root mean-squared error; SD denotes the mean of the standard deviation estimates. Six different SNRs {5, 10, 15, 20, 25, 30} and 10,000 simulated datasets were used for each case. Only diagonal diffusion tensors were considered.

greater than .7, that of m_2 was close to 0.7, and that of m_3 was smaller than .7 [Fig. 1(a)]. As expected, we observed that the bias of m_i (the mean value of $m_i - \lambda$) decreased as the SNR increased. In contrast, based on the generated eigenvalues using the asymptotic results in Theorem 3, the bias of the estimated eigenvalues at all SNRs can be predicted accurately [Fig. 1(a)]. Besides the bias, inspecting the Q–Q plot of m_i against $m_{(j),i}$ [Fig. 1(b)] revealed that the limiting density (9) can accurately predict the stochastic behavior of m_1 at SNR = 20. In terms of eigenvectors, at SNR = 20, the distribution of $\mathbf{e}_1$ estimated from the simulated DW data was a uniform distribution on the unit sphere [Figs. 1(c) and 1(d)] and was close to the distribution of the generated $\mathbf{e}_1$ based on Theorem 3 [Figs. 1(e) and 1(f)].

Figure 2 summarizes the results for $\mathbf{D}_i$ $(i = 2, 3, 4)$ as follows. First, for the oblate tensor $\mathbf{D}_2$, the second-order approximation of m_i $(i = 1, 2)$ performed better than the first-order approximation of m_i $(i = 1, 2)$ when SNR ≤ 10, whereas both the first- and second-order approximations of m_i $(i = 1, 2)$ were accurate when SNR > 10 [Fig. 2(a)]. The prediction of bias for m_3 based on (12) was highly accurate at all SNRs [Fig. 2(a)]. Moreover, in terms of $\mathbf{e}_1$, the second-order approximation of $\mathbf{e}_1$ led to a better prediction of the simulated distribution of $\mathbf{e}_1$ at SNR = 20 than did the first-order approximation of $\mathbf{e}_1$, because the second-order approximation accounted for additional variation along $\mathbf{v}_3$, while the first-order approximation did not [Figs. 2(b) and 2(c)]. Second, for the prolate tensor $\mathbf{D}_3$, the second-order approximations of m_i were accurate at all SNRs [Fig. 2(d)]. The second-order approximation of $\mathbf{e}_1$ led to accurate predictions even at small SNRs and the discrepancies between the first-order and second-order approximations were negligible when SNR > 10 [Fig. 2(e)]. Moreover, the first-order approximation of $\mathbf{e}_1$ provided a good prediction of the estimated

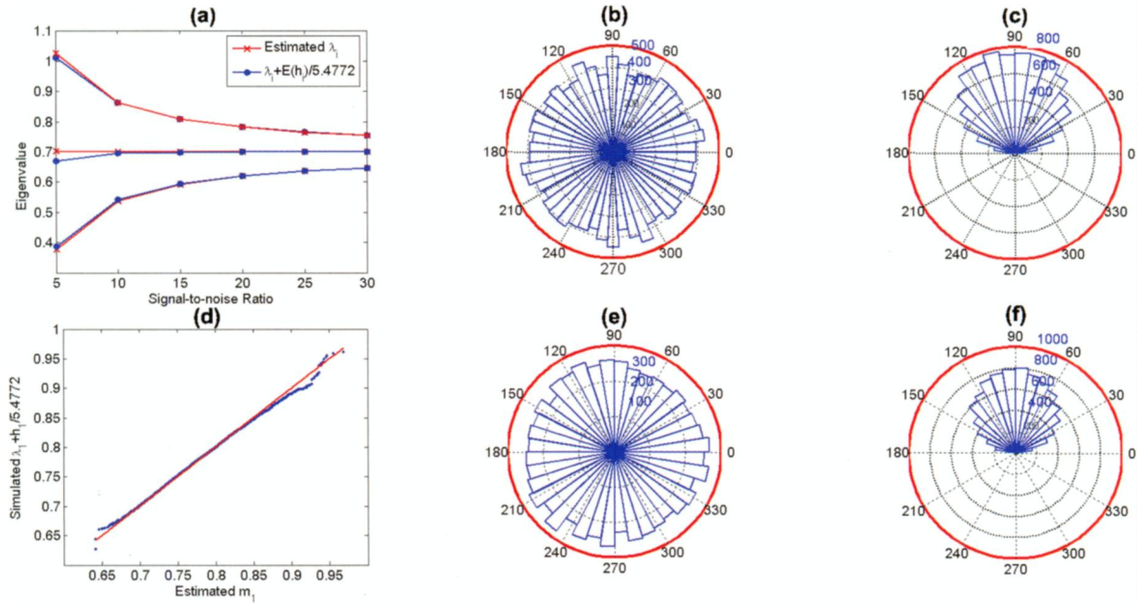

Figure 1. Results from a simulation study of the tensor $\mathbf{D}_1 = .7\mathbf{I}_3$ (units: 10^{-3} mm^2/s). (a) shows $E(m_i) = \lambda + n^{-1/2}E(h_i)$ $(i = 1, 2, 3)$ and the mean value of the estimated eigenvalues m_i $(i = 1, 2, 3)$ as a function of SNR from 5 to 30 based on 10,000 simulated DW datasets. (d) shows the Q–Q plot of the estimated eigenvalues m_1 based on 10,000 simulated DW datasets against eigenvalues $\{m_{(j),1} = .7 + n^{-1/2}h_{(j),1} : j = 1, \ldots, 10{,}000\}$ at SNR $= 20$, where the $h_{(j),1}$ are simulated from the limiting density (9). (b) and (c) show the angle histogram plots of θ and ϕ based on 10,000 simulated DW datasets at SNR $= 20$, respectively, where $\theta \in [0, 2\pi]$ and $\phi \in [0, \pi]$ are subcomponents of $(1, \theta, \phi)$, the spherical coordinate of $\mathbf{e}_1$. (e) and (f) show the angle histogram plots of θ and ϕ based on 10,000$\mathbf{e}_1$ that are simulated from the limiting density (9).

principal directions at SNR $= 20$ [Fig. 2(f)]. Finally, for the nondegenerate tensor $\mathbf{D}_4$, the second-order approximations of m_i $(i = 1, 2, 3)$ were accurate at all SNRs [Fig. 2(g)]. In addition, the first-order approximation of $\mathbf{e}_1$ was relatively accurate for SNRs ≥ 15 [Figs. 2(h) and 2(i)].

3.3 Type I and II Error Rates of PLRT(i)

We evaluated the performance of each of the PLRT(i) when used as the test statistics for the three hypotheses pertaining to the classifications of tensor morphologies (e.g., isotropic or not). Different diagonal tensors $\mathbf{D}_i$ $(i = 1, 2, 3, 4)$, whose three diagonal elements were, respectively, [.7, .7, .7], [.8, .8, .5], [1.0, .55, .55], and [.9, .7, .5] (units: 10^{-3} mm^2/s), were chosen for the`various test statistics, because each PLRT(i) was developed for diffusion tensors with different morphologies under the null hypothesis. To evaluate PLRT(1), we chose three diagonal diffusion tensors whose three diagonal elements were, respectively, $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_4$ (units: 10^{-3} mm^2/s). To evaluate PLRT(2), we chose three other diagonal tensors whose three diagonal elements were, respectively, given by $\mathbf{D}_2$, $\mathbf{D}_4$, and $\mathbf{D}_3$ (units: 10^{-3} mm^2/s). To evaluate PLRT(3), we chose three other tensors whose three diagonal elements were $\mathbf{D}_3$, $\mathbf{D}_2$, and $\mathbf{D}_4$ (units: 10^{-3} mm^2/s), respectively. For each simulation, two significance levels, 5% and 1%, were considered and 10,000 replications were used to estimate the nominal significance levels (or rejection rates). For a fixed α, if the Type I rejection rate is smaller than α, then the test is conservative, whereas if the Type I rejection rate is greater than α, then the test is anticonservative, or liberal.

Under the null hypothesis, the estimated significance levels of the PLRT(i) were reasonably close to the nominal significance levels for this small sample of 30 DW images (Table 2). Overall, although the Type I errors for the three test statistics were not excessive, these results indicate that the use of the scaled chi-squared distribution as a simple and reasonable approximation to the distribution of PLRT(i) under the null hypothesis requires further improvement for small sample sizes. Finding a better approximation to the distribution PLRT(i) under the null hypothesis warrants further research. Moreover, as expected, statistical power increased with the degree of anisotropy and the SNR values.

4. APPLICATION

We acquired diffusion-weighted MR images of the brains of seven healthy adult volunteers (four men and three women; all right handed; mean age 28 ± 4.2 years) on a GE 3.0-T whole-body magnetic resonance imaging (MRI) scanner (Milwaukee, WI). The imaging acquisition scheme $\{(b_i, \mathbf{r}_i) : i = 1, \ldots, 30\}$ consisted of $m = 5$ baseline images with $b = 0$ s/mm^2 and $n - m = 25$ directions of diffusion gradients that were arranged uniformly in three-dimensional space at $b = 1{,}000$ s/mm^2 (Hardin et al. 1994). Each diffusion-weighted image contained $256 \times 256 \times 34$ voxels.

For each subject, we used a heteroscedastic linear model (2) to construct the diffusion tensors. We subsequently calculated at each voxel the WLS estimate $\widehat{\theta}^{(1)}$, the SNR (S_0/σ), three eigenvalue–eigenvector pairs $\{(m_i, \mathbf{e}_i) : i = 1, 2, 3\}$, the invariant measures, including CL $= (m_1 - m_2)/I_1$, CP $= 2(m_2 - m_3)/I_1$, RA $= \sqrt{1 - 3I_2 I_1^{-2}}$, and FA $= \sqrt{1 - I_2(I_1^2 - 2I_2)^{-1}}$, and our three test statistics PLRT(i) and their associated p values, where $m_1 \geq m_2 \geq m_3$, $I_1 = \mathrm{tr}(\hat{\mathbf{D}})$, $I_2 = m_1 m_2 + m_1 m_3 +$

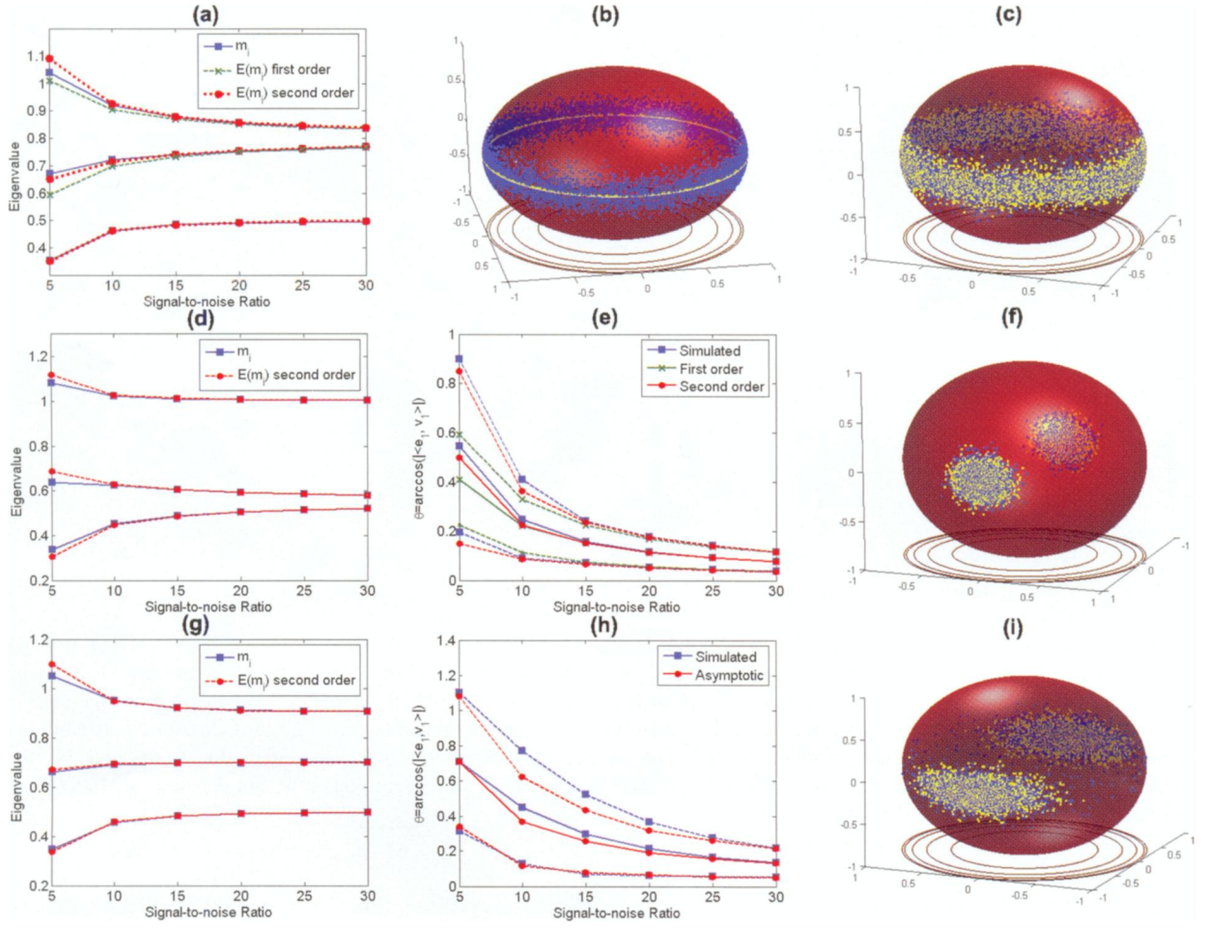

Figure 2. Results from a simulation study of three diagonal tensors $\mathbf{D}_i$ $(i = 2, 3, 4)$. (a)–(c) summarize results for $\mathbf{D}_2$: [.8, .8, .5] (units: 10^{-3} mm^2/s); (d)–(f) are for $\mathbf{D}_3$: [1.0, .55, .55] (units: 10^{-3} mm^2/s); and (g)–(i) are for $\mathbf{D}_4$: [.9, .7, .5] (units: 10^{-3} mm^2/s). Moreover, 10,000 simulated datasets were used for all cases. (a) shows the first- and second-order approximations of $E(m_i)$ $(i = 1, 2)$, the second-order approximation of $E(m_3)$, and the mean value of m_i $(i = 1, 2, 3)$ as a function of the SNRs from 5 to 30. At SNR = 20, (b) shows the scatterplots of $\mathbf{e}_1$ (blue points) and $\mathbf{e}_{(j),1}$ (yellow points) simulated from the first-order approximation, whereas (c) shows the scatterplots of $\mathbf{e}_1$ (blue points) and $\mathbf{e}_{(j),1}$ (yellow points) simulated from the second-order approximation. (d) and (g) show the second-order approximations of $E(m_i)$ $(i = 1, 3)$ and the mean value of m_i $(i = 1, 2, 3)$ as a function of the SNRs from 5 to 30 for $\mathbf{D}_2$ and $\mathbf{D}_3$, respectively. (e) and (h) show the mean value and the standard deviation of $\arccos(|\langle \mathbf{e}_1, \mathbf{v}_1 \rangle|)$ for $\mathbf{D}_2$ and $\mathbf{D}_3$, respectively: For $\mathbf{D}_2$, $\mathbf{e}_1$ is based on simulated datasets (blue squares), the first-order approximation (green triangles), or the second-order approximation (red circles); for $\mathbf{D}_3$, $\mathbf{e}_1$ is based on either simulated datasets (blue squares) or the first-order approximation (red circles). (f) and (i) show the scatterplots of the estimated $\mathbf{e}_1$ (blue points) and eigenvectors $\mathbf{e}_{(j),1}$ (yellow points) simulated from the first-order approximation for $\mathbf{D}_2$ and $\mathbf{D}_3$, respectively, at SNR = 20.

$m_2 m_3$, and $I_3 = m_1 m_2 m_3$. We further set the significance level at 1% and used the p values of PLRT(i) $(i = 1, 2, 3)$ to classify the morphology of the DT at each voxel. Furthermore, based on the tensor morphology at each voxel, we constructed the confidence intervals of the three eigenvalues and the confidence cones of $\mathbf{e}_1$ using the asymptotic results in Theorems 3 and 4. We tracked fibers in a selected region of interest (ROI) using a commonly available software package (DTI Track 2005; Fillard 2005).

Using a single representative subject, we presented the maps of CL, CP, and FA, the $-\log_{10}(p)$ maps of PLRT(i) $(i = 1, 2, 3)$, the map of morphological types, and the map of principal directions at a selective slice in Figures 3(a)–3(h). In the $-\log_{10}(p)$ value maps of PLRT(i) $(i = 1, 2, 3)$, a voxel having a p value less than .01, which corresponds to a $-\log_{10}(p)$ value of greater than 2, was regarded as significant, and all $(-\log_{10}(p))$ values greater than 8 were set equal to 8 to improve the visualization of the $-\log_{10}(p)$ values [Figs. 3(b), 3(e), and 3(h)]. In the map of the tensor morphologies, a four-color scheme was used to represent the four differing morphologies: blue for isotropic tensors, red for oblate tensors, yellow for prolate tensors, and white for nondegenerate tensors [Fig. 3(c)]. We also superimposed the oblate voxels (in yellow) on a three-color map of principal directions (green, inferior–superior; red, left–right; blue, anterior–posterior) [Fig. 3(f)].

The $-\log_{10}(p)$ values of PLRT(i) $(i = 1, 2, 3)$ were more sensitive and specific in detecting degenerate and nondegenerate tensors [Figs. 3(b), 3(e), and 3(h)]. For instance, in the map of linear anisotropy measures [Fig. 3(d)], although red and white voxels had relatively large differences between λ_1 and λ_2, whether the diffusion tensors represented in blue are truly oblate, however, is unclear. The $-\log_{10}(p)$ maps of PLRT(2)

Table 2. Comparisons of the rejection rates for the test statistics PLRT(i) ($i = 1, 2, 3$) under the single-tensor models

	Statistic: PLRT(1); $\mathbf{D}$: $[D_{11}, D_{22}, D_{33}]$ (units: 10^{-3} mm^2/s)					
	$H_0^{(1)}$: [.7, .7, .7]		$H_1^{(1)}$: [.8, .8, .5]		$H_1^{(1)}$: [.9, .7, .5]	
SNR	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$
5	.028	.084	.072	.177	.077	.189
10	.027	.083	.238	.428	.286	.493
15	.026	.082	.565	.753	.678	.848
20	.025	.079	.867	.951	.933	.979
25	.022	.078	.982	.997	.996	.999
30	.023	.077	.998	1.000	.999	1.000
	Statistic: PLRT(2); $\mathbf{D}$: $[D_{11}, D_{22}, D_{33}]$ (units: 10^{-3} mm^2/s)					
	$H_0^{(2)}$: [.8, .8, .5]		$H_1^{(2)}$: [.9, .7, .5]		$H_1^{(2)}$: [1.0, .55, .55]	
SNR	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$
5	.019	.063	.017	.060	.033	.106
10	.017	.062	.055	.151	.274	.495
15	.014	.057	.166	.344	.754	.909
20	.015	.061	.348	.562	.975	.996
25	.013	.056	.565	.771	.999	1.000
30	.014	.057	.761	.905	1.000	1.000
	Statistic: PLRT(3); $\mathbf{D}$: $[D_{11}, D_{22}, D_{33}]$ (units: 10^{-3} mm^2/s)					
	$H_0^{(3)}$: [1.0, .55, .55]		$H_1^{(3)}$: [.8, .8, .5]		$H_1^{(3)}$: [.9, .7, .5]	
SNR	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$
5	.021	.069	.016	.062	.015	.060
10	.019	.069	.095	.231	.072	.185
15	.017	.065	.340	.574	.212	.405
20	.018	.070	.699	.873	.442	.662
25	.016	.065	.931	.984	.687	.854
30	.017	.064	.992	.999	.859	.954

NOTE: Six different SNRs {5, 10, 15, 20, 25, 30} and 10,000 simulated datasets were used for each case. Two significance levels, 5% and 1%, and only diagonal diffusion tensors were considered.

for the oblate tensors [Fig. 3(e)], in contrast, identified many voxels that had large $-\log_{10}(p)$ values and had relatively small values of the linear anisotropy measure [Fig. 3(d)]. In the map of principal directions [Fig. 3(f)], oblate tensors occurred primarily in voxels where fiber tracts cross, as well as along the boundaries of different tissue types.

Tensor morphologies in a region of interest were examined to illustrate the proposed methods for tensor classification. The ROI [Figs. 4(a) and 4(b)] contained 900 (30×30) voxels representing diffusion tensors with differing morphologies. The percentage of the total that falls into each tensor group is as follows: 68.78% were nondegenerate, 13.78% were oblate, 11.67% were prolate, and 5.77% were isotropic. We observed from a three-color map of principal directions [Fig. 4(c)] that three fibers oriented from left to right (red), two fibers oriented from inferior to superior (green), and oblate tensors (yellow points) were located primarily in voxels where fiber tracts cross. The tracking algorithm confirmed those three red fibers, two green fibers, and one blue fiber [Fig. 4(e)]. We further applied asymptotic results from Theorems 3 and 4 to a diffusion tensor in a selected voxel of the ROI [the fourth ellipsoid from the right in the last row of Fig. 4(d)]. The three eigenvalues of the estimated diffusion tensor were calculated as .9631, .6722, and .5619 (units: 10^{-3} mm^2/s), respectively. The p value of PLRT(1) is smaller than 10^{-8}, the p value of PLRT(2) is $10^{-7.27}$, and the p value of PLRT(3) is $10^{-1.349}$. Thus, at the 1% significance level, this tensor was classified as prolate in shape. Furthermore, the principal direction $\mathbf{e}_1$ of the estimated diffusion tensor was calculated as either (.926, .229, .300) or $-$(.926, .229, .300). We also used (19) to construct a cone that approximated the distribution of the principal direction $\mathbf{e}_1$ of this tensor [Fig. 4(f)], and we used (17) to construct a $1-\alpha$ confidence interval of λ_1 as $[.9631 - .041 \times z_{\alpha/2}, .9631 + .041 \times z_{\alpha/2}]$.

After classifying the morphology of the DT at each voxel, we examined the histogram of m_1, the plots of m_1 versus m_2 and m_2 versus m_3, and the histogram of FA for each morphological class of tensors [Figs. 5(a)–5(p)]. For isotropic tensors, the histogram of m_1 was skewed to the right and m_1 was widely spread from .5 to 4.0 (units: 10^{-3} mm^2/s) [Fig. 5(a)], whereas, for the other three classes of tensors, the histogram of m_1 was bell shaped and m_1 was mainly distributed from .5 to 2.0 (units: 10^{-3} mm^2/s) [Figs. 5(e), 5(i), and 5(m); Pierpaoli et al. 1996]. As expected, for degenerate tensors, the difference between two

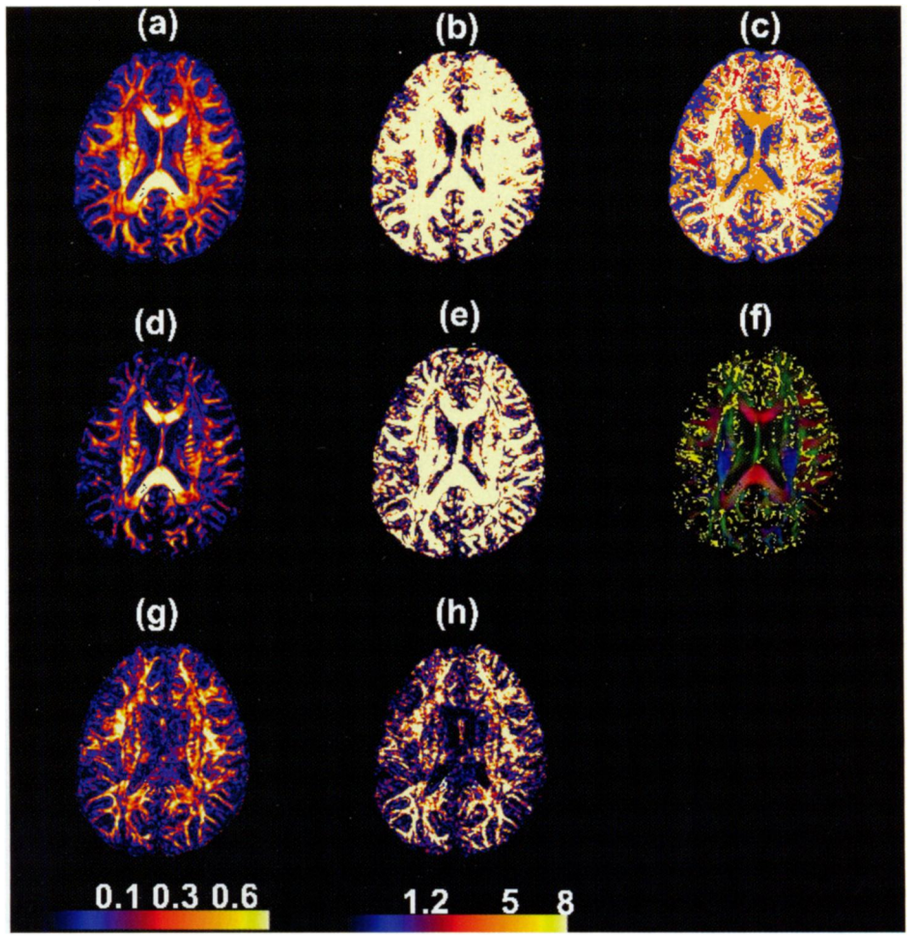

Figure 3. Maps of invariant measures: (a) FA, (d) CL, and (g) CP; the $-\log_{10}(p)$ value maps: (b) PLRT(1), (e) PLRT(2), and (h) PLRT(3); (c) map of tensor morphologies; and (f) map of principal directions at a selective slice from a single subject. Tensor morphologies in panel (c): white, nondegenerate; red, oblate; yellow, prolate; and blue, isotropic. Principal direction maps in panel (f): yellow, overlay indicates tensors having an oblate shape. The color scale in (b), (e), and (h) reflects the size of the values of $-\log_{10}(p)$ with black to blue representing smaller values (0–1) and red to white representing larger values (1.88–8).

consecutive eigenvalues was close to 0, even in the presence of the sorting bias [Figs. 5(b), 5(c), 5(f), and 5(g)]. For instance, the values of $m_1 - m_2$ were small for oblate tensors [Fig. 5(f)]. For all classes of tensors, the differences between m_2 and m_3 were relatively small, because all points (m_2, m_3) were positioned near the red line $m_2 = m_3$ [Figs. 5(c), 5(g), 5(k), and 5(o)]. For many prolate and nondegenerate tensors, the values of m_1 were much larger than those of m_2 [Figs. 5(j) and 5(n)]. The histograms of the FA for all four classes of tensors were skewed to the right, and the median of the FA values increased with the degree of anisotropy (nondegenerate > prolate > oblate > isotropic) [Figs. 5(d), 5(h), 5(l), and 5(p)].

We also constructed 95% confidence intervals of the three eigenvalues, the true FA for nonisotropic tensors, and the true CL for prolate and nondegenerate tensors, and presented them at a selective slice of a representative subject in Figures 6(a)–6(o). The three eigenvalues in each of the voxels containing cerebrospinal fluid were greater than 2.5 (10^{-3} mm^2/s). Except for the voxels containing cerebrospinal fluid, m_1 in most voxels ranged from .5 (10^{-3} mm^2/s) to 1.5 (10^{-3} mm^2/s) across the slice, whereas the second eigenvalue and the smallest eigenvalue in most of the voxels were smaller than 1.2 (10^{-3} mm^2/s) [Figs. 6(a)–6(i)]. Because of the inherent sorting bias, the estimated CL value was always larger than 0 for the oblate and isotropic tensors in probability (Thm. 2). Thus, we could not use the estimated CL value as a statistic to construct 95% confidence intervals for the true CL (= 0). However, in prolate and isotropic tensors, we applied the asymptotic results in Theorem 4 to construct the 95% confidence intervals for the true CL. For FA in nonisotropic tensors, the delta method was used to construct the 95% confidence intervals for true FA based on the asymptotic normality in Theorem 1(b) and a Taylor's series expansion (van der Vaart 1998; Zhu et al. 2006).

We studied the prevalence of the four standard morphological classes of tensors (isotropic, oblate, prolate, and nondegenerate) in vivo in the seven adult brains using our statistical framework for the classification of tensor morphology. We determined the standard types of DTs using the test statistics PLRT(i) and their associated p values at the 1% significance level for each of the three hypothesis tests. These percentages, shown in Table 3, were close to those obtained previously in Zhu et al. (2006).

We calculated the means and standard deviations of eight quantities, including m_i $(i = 1, 2, 3)$, tr($\hat{\mathbf{D}}$), CL, CP, RA, and

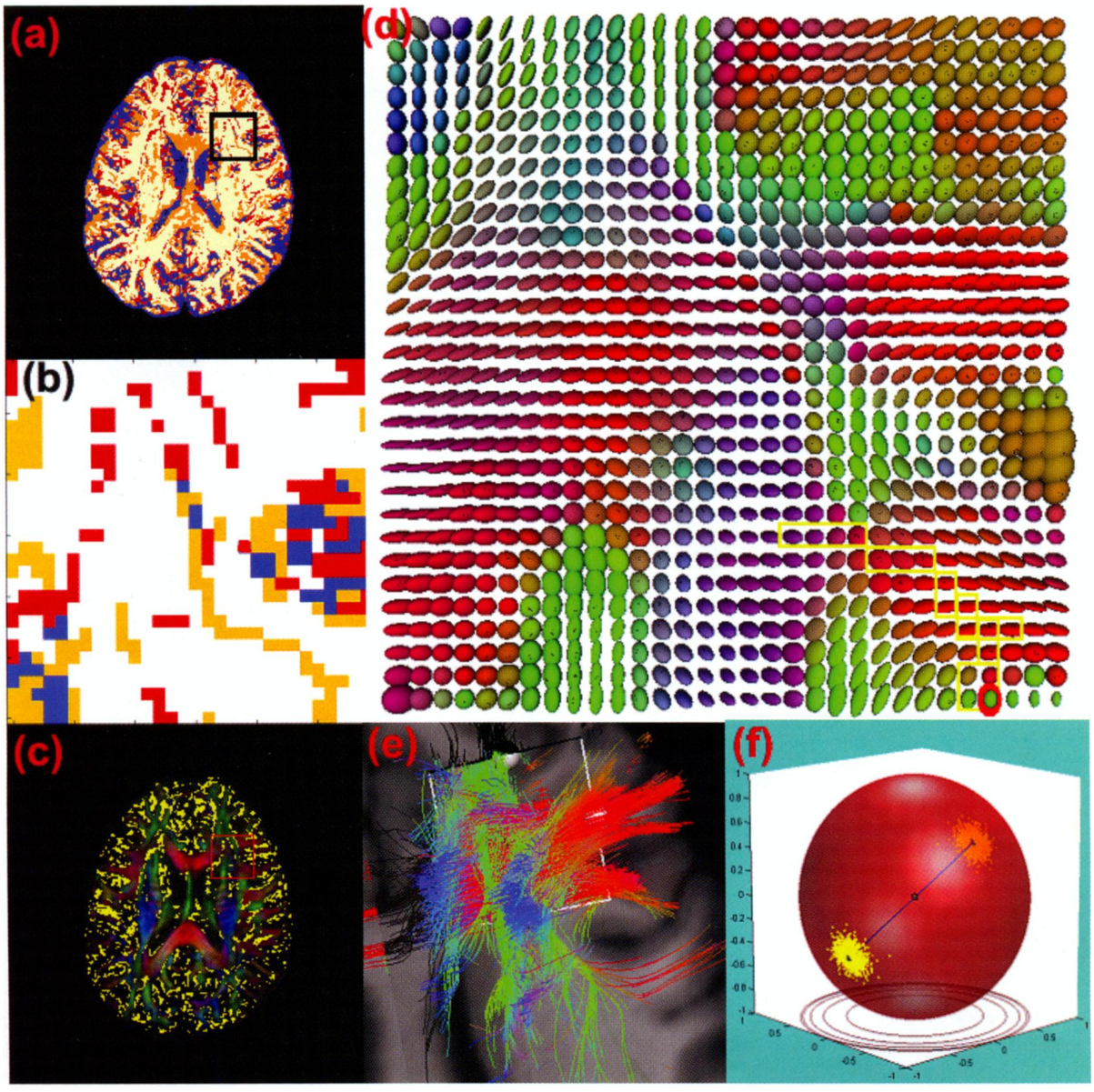

Figure 4. Maps of classified DT morphologies. (a) Within the ROI highlighted inside a black square; (b) the morphological map in the ROI; (c) the principal direction map within the ROI; (d) the ellipsoid map in the ROI; (e) the fiber tracts passing through the ROI; (f) the scatterplot of principal directions (yellow points) simulated from the first-order asymptotic expansion of $\mathbf{e}_1$ and the estimated principal directions $\mathbf{e}_1 = \pm(.926, .229, .300)$ (blue lines) for a selective tensor, whose ellipsoid is the fourth one from the right in the last row of (d).

FA, within each morphological class for each subject and their means and standard deviations across the seven subjects. The cross-subject variations of these eight quantities were relatively small. The m_i $(i = 1, 2, 3)$ and $\mathrm{tr}(\hat{\mathbf{D}})$ in isotropic tensors were much larger than those in nonisotropic tensors. Because RA and FA were invariant measures for quantifying the difference among m_i $(i = 1, 2, 3)$, the means of RA and FA increased with the degree of anisotropy, as expected. However, because CL only measured the difference between m_1 and m_2, the means of CL in prolate and nondegenerate tensors were much larger than those in isotropic and oblate tensors. Similarly, because CP only measured the difference between m_2 and m_3, the means of CP in oblate and nondegenerate tensors were much larger than those in isotropic and prolate tensors.

5. DISCUSSION

We have presented a set of answers for three interrelated questions that are central to the statistical analysis of DTI data. First, we have proposed a heteroscedastic linear model to analyze noise-laden diffusion-weighted MR images. To estimate the unknown parameter θ, we have used both theoretical results and numerical simulations to justify and support the use of the WLS estimate $\hat{\theta}^{(1)}$ starting from $\hat{\theta}_{\mathrm{LS}}$. We have also derived an explicit form for estimating $\mathrm{Cov}(\hat{\theta}^{(1)})$. For quantifying the effects of noise on the eigenspace components of the DTs, we have established the asymptotic expansions and limiting distributions of the estimated eigenvalues and eigenvectors for both degenerate and nondegenerate tensors. Our asymp-

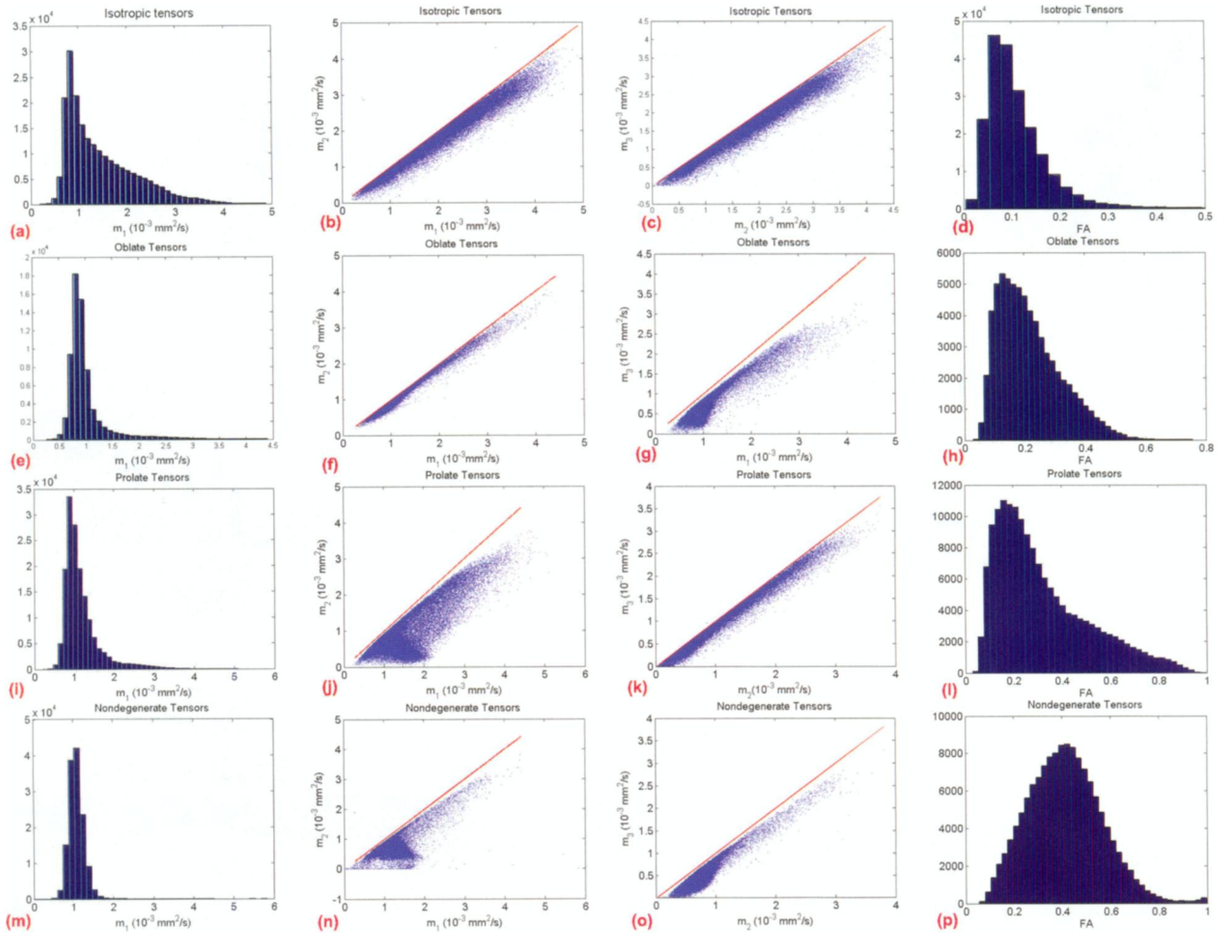

Figure 5. Maps of the histogram of m_1, the plot of m_1 versus m_2, the plot of m_2 versus m_3, and the histogram of FA for four morphological types from a single subject. Columns from left to right show the histogram of m_1, the plot of m_1 versus m_2, the plot of m_2 versus m_3, and the histogram of FA. Rows from top to bottom show isotropic [(a)–(d)], oblate [(e)–(h)], prolate [(i)–(l)], and nondegenerate tensors [(m)–(o)].

totic results for estimated eigenvalues and eigenvectors agree with the results obtained using various Monte Carlo simulations at relatively low to high SNRs. Finally, we have developed PLRT(i) to classify the morphology of DTs at each voxel as one of four standard types—nondegenerate, oblate, prolate, or isotropic. The null limiting distributions of PLRT(i) were used to determine rigorous statistical thresholds for the classification of tensor morphologies. In addition, we have demonstrated the effectiveness of our theoretical procedure by applying it to a real dataset to characterize the degree of uncertainty in the estimated eigenvalues and eigenvectors at each voxel of the human brain in vivo.

Our results differ substantially from those using a previous method (Behrens et al. 2003) and in several aspects. First, the previous method is a fully parametric approach that assumes a Gaussian distribution with homogeneous variance for the error components [see eq. (9) in Behrens et al. 2003], whereas ours is a semiparametric approach that allows a large class of distributions for the error components. The previous method is Bayesian and conducts statistical inference based on the posterior distribution of parameters of interest, such as the largest eigenvalue, whereas ours is a frequentist approach that conducts statistical inference based on the asymptotic results (e.g., the asymptotic distribution) of the estimate and test statistic. Finally, the previous method estimates the probability of the existence of fiber tracts between any two points, whereas ours quantifies the effects of noise on the estimation of diffusion tensors, their eigenvalues and eigenvectors, and classification of tensor morphologies.

Our methods are useful for addressing other important issues in the field of diffusion tensor imaging. We discuss several of those here.

Invariant Measures. Our results can be used to study the statistical properties (e.g., small-sample properties and limiting distributions) of invariant measures derived from estimated eigenvalues and eigenvectors, including fractional anisotropy (Skare et al. 2000; Mori and van Zijl 2002). For instance, we can apply Theorems 3 and 4 to the derivation of the limiting distribution of linear and planar anisotropy measures for both degenerate and nondegenerate tensors (Hasan et al. 2001). These statistical properties are useful for undertaking further statistical inference on the quantities derived from DTs, such as the calculation of their means, standard errors, and confidence intervals, as well as for determining rational and nonarbitrary thresholds for classifying the presence of anisotropy, which are required in tractography algorithms (Mori and van Zijl 2002;

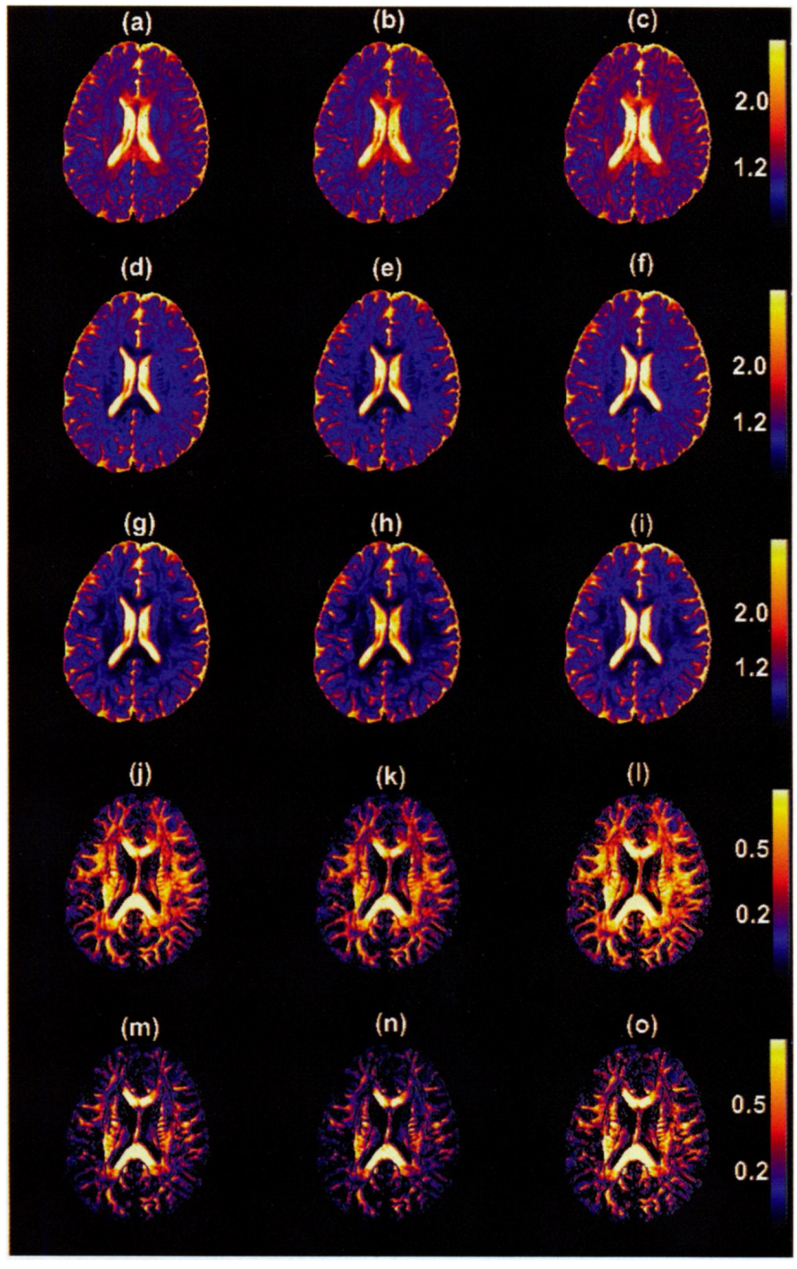

Figure 6. Maps of (a)–(c) m_1 and the 95% lower and upper confidence bounds of confidence intervals of λ_1; (d)–(f) m_2 and the 95% lower and upper confidence bounds of confidence intervals of λ_2; (g)–(i) m_3 and the 95% lower and upper confidence bounds of confidence intervals of λ_3; (j)–(l) FA and the 95% lower and upper confidence bounds of confidence intervals of true FA for nonisotropic tensors; and (m)–(o) CL and the 95% lower and upper confidence bounds of confidence intervals of true CL for prolate and nondegenerate tensors at a selective slice from a single subject. The color scale in the first three rows reflects the size of the values of m_i $(i = 1, 2, 3)$ with black to blue representing smaller values (0–1.2) (units: 10^{-3} mm^2/s) and red to white representing larger values (1.8–4) (units: 10^{-3} mm^2/s), whereas the color scale in the last two rows reflects the size of the values of CL and FA with black to blue representing smaller values (0–.2) and red to white representing larger values (.4–1).

Jones 2003). These statistical properties are also useful for determining the minimum signal-to-noise ratio and number of acquisitions to discriminate accurately differences in invariant measures, such as FA, across fibers (e.g., in the corpus callosum and internal capsule).

Acquisition Schemes. Our results can be used to study and select an optimal acquisition scheme, which minimizes certain design criteria (Jones, Horsfield, and Simmons 1999). For instance, to accurately estimate **D** accurately, we can use $\mathrm{tr}\{\mathrm{Cov}[\hat{\theta}^{(1)}]\}$, the trace of the covariance matrix of $\hat{\theta}^{(1)}$, to construct a design criterion, and then we can numerically minimize $\mathrm{tr}\{\mathrm{Cov}[\hat{\theta}^{(1)}]\}$ by varying the number of acquisitions, b factors, and diffusion gradients. Furthermore, to estimate the principal direction of prolate and nondegenerate tensors accurately, we can apply the results of (19) and (22) in Theorem 4 to construct a design criterion, such as the trace of the covari-

Table 3. The proportion, mean of eigenvalues, and mean of invariant measures of the DTs classified into four morphologies in seven adults subjects

Tensor	Per (%)	Statistics	m_1	m_2	m_3	tr(**D**)	RA	FA	CL	CP
			units: 10^{-3} mm^2/s							
ISO	34.74	M.M.	1.424	1.287	1.159	3.859	.067	.115	.038	.073
		s.d.	.100	.098	.095	.285	.005	.008	.003	.005
	3.45	M.sd.	.716	.658	.614	1.960	.043	.070	.030	.056
		s.d.	.051	.049	.047	.140	.004	.006	.003	.004
OB	11.22	M.M.	.993	.902	.662	2.557	.128	.215	.039	.206
		s.d.	.048	.049	.053	.145	.010	.015	.003	.015
	1.21	M.sd.	.383	.367	.356	1.086	.064	.102	.024	.111
		s.d.	.065	.065	.062	.193	.004	.005	.002	.006
PRO	27.88	M.M.	1.164	.750	.650	2.562	.201	.315	.174	.084
		s.d.	.034	.029	.030	.088	.009	.014	.007	.006
	.48	M.sd.	.476	.384	.373	1.139	.145	.193	.140	.051
		s.d.	.032	.046	.045	.123	.001	.002	.001	.004
ND	26.15	M.M.	1.074	.685	.439	2.198	.265	.415	.182	.232
		s.d.	.013	.018	.022	.046	.014	.019	.011	.011
	2.93	M.sd.	.234	.182	.188	.490	.118	.154	.113	.099
		s.d.	.012	.021	.022	.068	.006	.004	.007	.005

NOTE: ISO, isotropic; OB, oblate; PRO, prolate; ND, nondegenerate; M.M., mean of means; s.d., standard deviation; M.sd. mean of standard deviations; RA, rational anisotropy; FA, fractional anisotropy; CL, linear shape; CP, planar shape.

ance matrix of $\mathbf{e}_1$, and then we can optimize the acquisition scheme.

Nonparametric Bootstrapping. Although nonparametric bootstrapping methods have been proposed for the analysis of eigenvalues, eigenvectors, and their associated invariant scalar indices, as well as for use even in tractography algorithms, no asymptotic results until now have been provided to support the appropriate statistical use of bootstrapping methods in this context (Jones 2003; Pajevic and Basser 2003; Jones and Pierpaoli 2005; Lazar and Alexander 2005). Our results in Theorems 1–4 may help in establishing the validity of nonparametric bootstrapping methods in the analysis of diffusion tensor images (Shao and Tu 1995; sec. 1.6). We will present the asymptotic properties of nonparametric bootstrapping methods used for DTI in a separate article.

Relevance for Fiber-Tracking Algorithms. The uncertainty of the principal direction within each voxel has been developed into a general method for making probabilistically based maps of fiber tracts (Parker et al. 2003), even though a valid method for approximating the uncertainty of the principal direction has not been proposed in such a framework. We have, therefore, developed a method to statistically quantify the degree of uncertainty in estimating the principal direction $\mathbf{e}_1$ within each voxel when the diffusion tensor is either degenerate or nondegenerate. Therefore, we can produce more meaningful probabilistic maps for fiber tracts by combining the uncertainty of the principal direction $\mathbf{e}_1$ with the proposed method for constructing probabilistic maps of fiber tracts (Parker et al. 2003).

Spatial Normalization. Our results are also useful for coregistering DTI datasets across individuals. Methods for spatial normalization of diffusion tensor fields have been proposed based on the distribution of the principal direction within each voxel (Xu, Mori, Shen, van Zijl, and Davatzikos 2003). These methods, however, use the principal directions from neighboring voxels to approximate the distribution of the principal direction in a given voxel. Our results show that the distribution of the principal directions can be approximated using only diffusion-weighted data within the voxel itself.

Multiple-Tensor Models. Because our findings are limited to a model in which only a single tensor is present within each voxel, future investigations should consider developing models that account for the presence of multiple tensors within a single voxel (Alexander, Barker, and Arridge 2002; Frank 2002; Tuch et al. 2002; Wedeen, Hagmann, Tseng, Reese, and Weisskoff 2005). The limited spatial resolution of DW images will always include multiple tensors within the same voxel, and this reality will, therefore, always be a challenge for developing statistical models for tensor estimation and fiber tracking in DTI datasets. How to appropriately estimate the number of tensors within each voxel, and how to quantify the effects of noise on those multiple tensors and their associated eigenvalues and eigenvectors, remain daunting problems.

We hope that statistical methods will play an important role in addressing these and other challenges in the field of diffusion tensor imaging.

APPENDIX: ASSUMPTIONS

The following assumptions are needed to facilitate development of our methods, although they are not the weakest possible conditions.

(C1) The errors η_i are independent and $\sup_i E\eta_i^2 < \infty$.

(C2) $\lambda_{\min}(A_n) \to \infty$.

(C3) θ_* is an interior point of Θ and $\sup_i b_i < \infty$.

(C4) $\lim_{C\to\infty} \sup_i E[\eta_i^2 \mathbf{1}\{|\eta_i| > C\}] = 0$ and $\inf_i E[\eta_i^2] > 0$, where $\mathbf{1}(\cdot)$ denotes the indicator function.

(C5) $\max_{1\le i\le n} \mathbf{z}_i^T (A_n)^{-1} \mathbf{z}_i \to 0$ as $n \to \infty$.

(C6) $\sup_i E[\eta_i^4] < \infty$.

(C7) $\sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^T$ is always positive definite for $n \geq 7$, and the distribution of $(\log S_1, \ldots, \log S_n)$ is absolutely continuous with respect to n-dimensional Lebesgue measure.

(C8) The three eigenvalues of $\widehat{\mathbf{D}}$ are distinct with probability 1.

(C9) $\sqrt{n}\,\mathrm{Vec}(\widehat{\mathbf{D}} - \mathbf{D})$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ_D.

(C10) $\mathbf{Q}_n$ converges to a matrix $\mathbf{Q}$, which satisfies $0 < \lambda_{\min}(\mathbf{Q}) \leq \lambda_{\max}(\mathbf{Q}) < \infty$, where $\mathbf{Q}_n = G_{n,*}^{1/2} B_{n,*}^{-1} G_{n,*}^{1/2}$ and $\lambda_{\max}(\mathbf{Q})$ denotes the maximum eigenvalue of $\mathbf{Q}$.

Comments. Conditions (C1) and (C2) are sufficient and necessary conditions for $\hat{\theta}_{\mathrm{LS}}$ to be strongly consistent (Lai, Robbins, and Wei 1979; Chen, Hu, and Ying 1999). Condition (C3) is a natural condition to assume for diffusion tensor imaging, because the diffusion tensor is associated with the covariance matrix of a diffusion process and b_i, the b factor, usually ranges from 0 to 3,000 s/mm^2 (Kingsley 2006a–c). Conditions (C4)–(C6) are standard conditions to establish the asymptotic normality of $\hat{\theta}_{\mathrm{LS}}$ for a linear heteroscedastic model (Eicker 1963; White 1980). Condition (C7) is similar to the condition that was used for the sample covariance matrix in Okamoto (1973; Anderson 2003). Conditions (C1)–(C7) are sufficient conditions for conditions (C8) and (C9). Condition (C10) is required to ensure the existence of the asymptotic distributions of PLRT(i).

[Received June 2006. Revised January 2007.]

REFERENCES

Alexander, D. C., Barker, G. J., and Arridge, S. R. (2002), "Detection and Modelling of Non-Gaussian Apparent Diffusion Coefficient Profiles in Human Brain Data," *Magnetic Resonance in Medicine*, 48, 331–340.

Anderson, A. W. (2001), "Theoretical Analysis of the Effects of Noise on Diffusion Tensor Imaging," *Magnetic Resonance in Medicine*, 46, 1174–1188.

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis* (3rd ed.), New York: Wiley.

Basser, P. J. (1997), "New Histological and Physiological Stains Derived From Diffusion-Tensor MR Images," *Annals of the New York Academy of Sciences*, 820, 123–138.

Basser, P. J., and Jones, D. K. (2002), "Diffusion-Tensor MRI: Theory, Experimental Design and Data Analysis—A Technical Review," *NMR in Biomedicine*, 15, 456–467.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994), "Estimation of the Effective Self-Diffusion Tensor From the NMR Spin Echo," *Journal of Magnetic Resonance*, Ser. B, 103, 247–254.

Basser, P. J., and Pajevic, S. (2000), "Statistical Artifacts in Diffusion Tensor MRI (DT–MRI) Caused by Background Noise," *Magnetic Resonance in Medicine*, 44, 41–50.

——— (2003), "Dealing With Uncertainty in Diffusion Tensor MR Data," *Israel Journal of Chemistry*, 43, 129–144.

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., and Smith, S. M. (2003), "Characterization and Propagation of Uncertainty in Diffusion-Weighted Imaging," *Magnetic Resonance in Medicine*, 50, 1077–1088.

Brain Development Cooperative Group and Evans, A. C. (2006), "The NIH MRI Study of Normal Brain Development," *NeuroImage*, 30, 184–202.

Carroll, R. J., and Ruppert, D. (1982a), "Robust Estimation in Heteroscedastic Linear Models," *The Annals of Statistics*, 10, 429–441.

——— (1982b), "A Comparison Between Maximum Likelihood and Generalized Least Squares in a Heteroscedastic Linear Model," *Journal of the American Statistical Association*, 77, 878–882.

Carroll, R. J., Wu, J. F., and Ruppert, D. (1988), "The Effect of Estimating Weights in Weighted Least Squares," *Journal of the American Statistical Association*, 83, 1045–1054.

Chen, K., Hu, I., and Ying, Z. (1999), "Strong Consistency of Maximum Quasi-Likelihood Estimators in Generalized Linear Models With Fixed and Adaptive Designs," *The Annals of Statistics*, 27, 1155–1163.

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., Burton, H., and Raichle, M. E. (1999), "Tracking Neuronal Fiber Pathways in the Living Human Brain," *Proceedings of the National Academy of Sciences of the United States of America*, 96, 10422–10427.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, London: Chapman & Hall.

Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *Annals of Mathematical Statistics*, 34, 447–456.

Fillard, P. (2005), "DTI Track 2005," available at *http://www-sop.inria.fr/epidaure/personnel/Pierre.Fillard/softwares/*.

Frank, L. R. (2002), "Characterization of Anisotropy in High Angular Resolution Diffusion-Weighted MRI," *Magnetic Resonance in Medicine*, 47, 1083–1099.

Gudbjartsson, H., and Patz, S. (1995), "The Rician Distribution of Noisy MRI Data," *Magnetic Resonance in Medicine*, 34, 910–914.

Hardin, R. H., Sloane, N. J. A., and Smith, W. D. (1994), "Minimal Energy Arrangements of Points on a Sphere With Minimal $1/r$ Potential," available at *http://www.research.att.com/njas/electrons/*.

Hasan, K. M., Basser, P. J., Parker, D. L., and Alexander, A. L. (2001), "Analytical Computation of the Eigenvalues and Eigenvectors in DT–MRI," *Journal of Magnetic Resonance*, 152, 41–47.

Henkelman, R. M. (1985), "Measurement of Signal Intensities in the Presence of Noise in MR Images," *Medical Physics*, 12, 232–233.

Jones, D. K. (2003), "Determining and Visualizing Uncertainty in Estimates of Fiber Orientation From Diffusion Tensor MRI," *Magnetic Resonance in Medicine*, 49, 7–12.

Jones, D. K., and Pierpaoli, C. (2005), "Confidence Mapping in Diffusion Tensor Magnetic Resonance Imaging Tractography Using a Bootstrap Approach," *Magnetic Resonance in Medicine*, 53, 1143–1149.

Jones, D. K., Horsfield, M. A., and Simmons, A. (1999), "Optimal Strategies for Measuring Diffusion in Anisotropic Systems by Magnetic Resonance Imaging," *Magnetic Resonance in Medicine*, 42, 515–525.

Kingsley, P. B. (2006a), "Introduction to Diffusion Tensor Imaging Mathematics—I: Tensors, Rotations, and Eigenvectors," *Concepts in Magnetic Resonance, A*, 28, 101–122.

——— (2006b), "Introduction to Diffusion Tensor Imaging Mathematics—II: Anisotropy, Diffusion-Weighting Factors, and Gradient Encoding Schemes," *Concepts in Magnetic Resonance, A*, 28, 123–154.

——— (2006c), "Introduction to Diffusion Tensor Imaging Mathematics—III: Tensor Calculation, Noise, Simulations, and Optimization," *Concepts in Magnetic Resonance, A*, 28, 155–179.

Lai, T. L., Robbins, H., and Wei, C. Z. (1979), "Strong Consistency of Least Squares Estimates in Multiple Regression II," *Journal of Multivariate Analysis*, 9, 343–361.

Lazar, M., and Alexander, A. L. (2003), "An Error Analysis of White Matter Tractography Methods: Synthetic Diffusion Tensor Field Simulations," *NeuroImage*, 20, 1140–1153.

——— (2005), "Bootstrap White Matter tractography (BOOT–TRAC)," *NeuroImage*, 24, 524–532.

Le Bihan, D. (2003), "Looking Into the Functional Architecture of the Brain With Diffusion MRI," *Nature Reviews Neuroscience*, 4, 469–480.

Lim, K. O., and Helpern, J. A. (2002), "Neuropsychiatric Applications of DTI—A Review," *NMR in Biomedicine*, 15, 587–593.

MacKinnon, J. G., and White, H. (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators With Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305–325.

Mori, S., and van Zijl, P. C. M. (2002), "Fiber Tracking: Principles and Strategies—A Technical Review," *NMR in Biomedicine*, 15, 468–480.

Okamoto, M. (1973), "Distinctness of the Eigenvalues of a Quadratic Form in a Multivariate Sample," *The Annals of Statistics*, 1, 763–765.

Pajevic, S., and Basser, P. J. (2003), "Parametric and Non-Parametric Statistical Analysis of DT–MRI Data," *Journal of Magnetic Resonance*, 161, 1–14.

Parker, G. J. M., Haroon, H. A., and Wheeler-Kingshott, C. A. M. (2003), "A Framework for a Streamline-Based Probabilistic Index of Connectivity (PICo) Using a Structural Interpretation of MRI Diffusion Measurements," *Journal of Magnetic Resonance Imaging*, 21, 505–512.

Pierpaoli, C., and Basser, P. J. (1996), "Toward a Quantitative Assessment of Diffusion Anisotropy," *Magnetic Resonance in Medicine*, 36, 893–906.

Pierpaoli, C., Jezzard, P., Basser, P. J., Barnett, A., and Chiro, G. D. (1996), "Diffusion Tensor MR Imaging of the Human Brain," *Radiology*, 201, 637–648.

Rohde, G. K., Barnett, A. S., Basser, P. J., Marenco, S., and Pierpaoli, C. (2004), "Comprehensive Approach for Correction of Motion and Distortion in Diffusion-Weighted MRI," *Magnetic Resonance in Medicine*, 51, 103–114.

Rohde, G. K., Barnett, A. S., Basser, P. J., and Pierpaoli, C. (2005), " Estimating Intensity Variance Due to Noise in Registered Images: Application to Diffusion Tensor MRI," *NeuroImage*, 26, 673–684.

Rowe, D. R. (2005), "Parameter Estimation in the Magnitude-Only and Complex-Valued fMRI Data Models," *NeuroImage*, 25, 1124–1132.

Salvador, R., Pena, A., Menon, D. K., Carpenter, T. A., Pickard, J. D., and Bullmore, E. T. (2005), "Formal Characterization and Extension of the Linearized Diffusion Tensor Model," *Human Brain Mapping*, 24, 144–155.

Schott, J. R. (2003), "Weighted Chi-Squared Tests for Partial Common Principal Component Subspaces," *Biometrika*, 90, 411–421.
Shao, J., and Tu, D. (1995), *The Jackknife and Bootstrap*, New York: Springer-Verlag.
Skare, S., Li, T. Q., Nordell, B., and Ingvar, M. (2000), "Noise Considerations in the Determination of Diffusion Tensor Anisotropy," *Magnetic Resonance Imaging*, 18, 659–669.
Stejskal, E. O., and Tanner, J. E. (1965), "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient," *Journal of Chemical Physics*, 42, 288–292.
Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., and Wedeen, V. J. (2002), "High Angular Resolution Diffusion Imaging Reveals Intravoxel White Matter Fiber Heterogeneity," *Magnetic Resonance in Medicine*, 48, 577–582.
van der Vaart, A. W. (1998), *Asymptotic Statistics*, London: Cambridge University Press.
Wedeen, V. J., Hagmann, P., Tseng, W. I., Reese, T. G., and Weisskoff, R. M. (2005), "Mapping Complex Tissue Architecture With Diffusion Spectrum Magnetic Resonance Imaging," *Magnetic Resonance in Medicine*, 54, 1377–1386.
White, H. L. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.
Xu, D., Mori, S., Shen, D., van Zijl, P., and Davatzikos, C. (2003), "Spatial Normalization of Diffusion Tensor Fields," *Magnetic Resonance in Medicine*, 50, 175–182.
Xu, D., Mori, S., Solaiyappan, M., van Zijl, P. C., and Davatzikos, C. (2002), "A Framework for Callosal Fiber Distribution Analysis," *NeuroImage*, 17, 1131–1143.
Zhu, H., Xu, D., Amir, R., Hao, X., Zhang, H., Alayar, K., Ravi, B., and Peterson, B. (2006), "A Statistical Framework for the Classification of Tensor Morphologies in Diffusion Tensor Images," *Magnetic Resonance Imaging*, 24, 569–582.

Comment

Armin SCHWARTZMAN

1. INITIAL REMARKS

This interesting and timely article attempts an important goal: to formalize the inference about diffusion tensors from diffusion weighted images in a single subject in the presence of measurement and artifact noise. The article's main contributions are:

- A heteroscedastic linear model to account for noise in diffusion-weighted MRI data along with theoretical support for the use of a (one-step) weighted least squares algorithm to solve it.
- Asymptotic distributions of the estimated eigenstructure of the diffusion tensor under degenerate and nondegenerate cases, in addition to pseudolikelihood ratio tests for classifying each tensor into one of those cases.

As the authors explain, inferences about the diffusion tensor in a single subject are usually based on quantities derived from the tensor, the most common being scalar functions of the eigenvalues such as fractional anysotropy (FA) and trace, and the principal diffusion direction (PDD), the eigenvector corresponding to the largest eigenvalue. Whereas standard statistics are often used to analyze the scalar quantities, formal modeling of the PDD is not usually seen. Perhaps this is because statistical methods for unit vectors in three-dimensional space are not as widely known in the general scientific community, even though they have been studied extensively in the field of directional statistics (Mardia and Jupp 2005).

Especially because tractography algorithms are based on the PDD, it is important to have a characterization of the uncertainty in that vector as a result of noise. The authors provide this in an asymptotic sense as the number of measurements gets large. When the true tensor is oblate (i.e., the two largest eigenvalues are equal) or isotropic (i.e., all three eigenvalues are equal), the PDD is not defined, making the uncertainty infinite. Algorithms often deal with this problem by thresholding a function of the eigenvalues such as FA or CL (Westin et al. 2002), the idea being that if the tensor is not sufficiently anisotropic, then the PDD is not to be trusted. Instead of numerically trying appropriate thresholds, the authors mathematically derive the uncertainty and provide formal tests to classify whether the tensor is isotropic, oblate, prolate, or fully anisotropic.

A difficulty with the article is that the mathematical results obscure how the obtained quantities actually can be computed and used in practice. More specifically, theorem 1 proves the asymptotic normality of $\hat{\theta}$ and provides a way to estimate the asymptotic covariance, whereas the results of section 2.2 give the asymptotic distribution of the eigenstructure of the tensor as a function of the asymptotic covariances Σ_U, $\Sigma_{\mathbf{U}_{11}}$, and $\Sigma_{\mathbf{U}_{22}}$. However, the results do not explain how these parameters depend on the underlying true diffusion and noise parameters. Clearly, the asymptotic covariances depend on the acquisition scheme $\mathbf{z}_1, \ldots, \mathbf{z}_n$. The authors have preferred to give general results and to not commit themselves to particular acquisitions.

Admittedly, general analytical expressions are difficult to obtain, so a solution offered by the authors is to compute the asymptotic covariances by simulation, which is what they implement in section 3. The particular acquisition used by the authors in their simulation consists of $m = 5$ baseline images with $b = 0$ and $n - m = 25$ directions of diffusion gradients arranged uniformly in three-dimensional space with $b \neq 0$. In this comment, I hope to provide some insight into the asymptotic covariance parameters, based on an asymptotic version of this acquisition scheme. Specifically, I derive analytic forms of the asymptotic covariances of $\hat{\theta}$ and $\mathbf{U}$, $\mathbf{U}_{11}$, and $\mathbf{U}_{22}$. It turns out that these provide a surprising answer to the question of what is the optimal ratio between the number of measurements at $b = 0$ versus the number of measurements at $b \neq 0$.

Armin Schwartzman is Assistant Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115 (E-mail: *armins@hsph.harvard.edu*).

Journal of the American Statistical Association
December 2007, Vol. 102, No. 480, Applications and Case Studies
DOI 10.1198/016214507000001193

2. ACQUISITION

As the authors mention in section 5, their results can be used to design acquisition schemes. The key conditions for theorem 1 that are affected by the acquisition scheme are (C2) and (C5), so it makes sense to ask when these hold. First, note that if $b_i = b$ constant for all $i = 1, \ldots, n$, then A_n is not invertible for any n and has an eigenvalue 0 with eigenvector $(b, 1, 0, 0, 1, 0, 1)^T/\sqrt{3+b^2}$. For this reason, it is common in DWI to make some measurements, say m of them, at $b = 0$, and the remaining $n - m$ measurements at a constant value $b > 0$. In their simulation, the authors use $m = 5$ and $n - m = 25$. A question that may be asked is what is a good proportion between the two.

Let $\mathbf{r}$ be a random unit vector on the sphere and define $\mathbf{x} = (r_1^2, 2r_1r_2, 2r_1r_3, r_2^2, 2r_2r_3, r_3^2)^T$. Let r_i be iid samples from the distribution of $\mathbf{r}$ and, without loss of generality, assume that $b_i = 0$ for $i = 1, \ldots, m$ and $b_i = b$ (constant) for $i = m+1, \ldots, n$. Suppose both $m, n \to \infty$ but $m/n \to \gamma$, where, by definition, $0 \le \gamma \le 1$. The strong law of large numbers applies because the sphere is bounded. Thus $A_n/n \to A_\gamma = E(\mathbf{z}\mathbf{z}^T)$, where $\mathbf{z}$ is a random vector that takes the value $(1, 0, 0, 0, 0, 0, 0)^T$ with probability γ and $(1, -b\mathbf{x}^T)^T$ with probability $1 - \gamma$. Here A_γ can be written as

$$\frac{1}{n}A_n \to A_\gamma = \begin{pmatrix} 1 & -(1-\gamma)bE(\mathbf{x}^T) \\ -b(1-\gamma)E(\mathbf{x}) & (1-\gamma)b^2E(\mathbf{x}\mathbf{x}^T) \end{pmatrix}. \quad (1)$$

When $\gamma = 0$, A_γ has a zero eigenvalue corresponding to the same eigenvector $(b, 1, 0, 0, 1, 0, 1)^T/\sqrt{3+b^2}$ as A_n. When $\gamma = 1$, A_γ is obviously noninvertible. Therefore, a necessary condition for A_γ to be invertible is that $0 < \gamma < 1$. That this is also a sufficient condition is proved next when $\mathbf{r}$ is uniform on the sphere. The principle is somehow clearer when proved for general dimension p (in DWI, $p = 3$). In addition, presentation is easier if the elements of $\mathbf{x}$ are reordered as was done by Salvador et al. (2005), as follows.

Proposition 1. Let $\mathbf{r} = (r_1, \ldots, r_p)^T$ be a random unit vector uniformly distributed on the $(p-1)$-dimensional sphere, and let $\mathbf{x}$ be the $p(p+1)/2$-dimensional vector $\mathbf{x} = (r_1^2, \ldots, r_p^2, 2r_1r_2, \ldots, 2r_{p-1}r_p)^T$ formed by taking the p squared diagonals of $\mathbf{r}$ and then twice all $p(p-1)/2$ pairs of off-diagonals of $\mathbf{r}$. Then the matrix A_γ defined by the right side of (1) is positive definite if and only if $0 < \gamma < 1$.

Proof. By symmetry, all of the odd moments of $\mathbf{r}$ are 0. The relevant even moments are $E(r_i^2) = 1/p$, $E(r_i^4) = 3/(p(p+2))$, and $E(r_i^2r_j^2) = 1/(p(p+2))$ for $i \neq j$ (Mardia and Jupp 2005, p. 186). Therefore, $E(\mathbf{x}^T) = (\mathbf{1}_p^T/p, \mathbf{0}^T)$, where $\mathbf{1}_p$ denotes a vector of 1's of length p and $\mathbf{0}^T$ fills in up to length $p(p+1)/2$. In addition,

$$E(\mathbf{x}\mathbf{x}^T) = \frac{1}{p(p+2)}\begin{pmatrix} 2\mathbf{I}_p + \mathbf{1}_p\mathbf{1}_p^T & \mathbf{0} \\ \mathbf{0} & 4\mathbf{I}_{p(p-1)/2} \end{pmatrix}, \quad (2)$$

with inverse

$$(E(\mathbf{x}\mathbf{x}^T))^{-1} = \frac{p(p+2)}{4} \times \begin{pmatrix} 2\mathbf{I}_p - 2\mathbf{1}_p\mathbf{1}_p^T/(p+2) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p(p-1)/2} \end{pmatrix}, \quad (3)$$

and, interestingly, $E(\mathbf{x}^T)(E(\mathbf{x}\mathbf{x}^T))^{-1} = (\mathbf{1}_p^T, \mathbf{0}^T)$, so that $E(\mathbf{x}^T)(E(\mathbf{x}\mathbf{x}^T))^{-1}E(\mathbf{x}) = 1$. Replacing in (1) and using the formula for the determinant of a partitioned matrix (Schott 2005, p. 250), we get $|A_\gamma| = \gamma(1-\gamma)b^2|E(\mathbf{x}\mathbf{x}^T)|$, which is positive if and only if $0 < \gamma < 1$.

Once $0 < \gamma < 1$ has been established, an asymptotically equivalent acquisition scheme is to sample the iid $\mathbf{z}_i$'s as

$$\mathbf{z} = \begin{cases} (1, \mathbf{0}^T)^T & \text{with probability } \gamma \\ (1, -b\mathbf{x}_i^T)^T & \text{with probability } 1-\gamma, \end{cases} \quad (4)$$

where $\mathbf{0}^T$ fills in up to length $p(p+1)/2$ and the $\mathbf{x}_i$'s are iid samples of $\mathbf{x}$ according to the distribution of $\mathbf{r}$.

From the authors' theorem 1a, the rate of convergence of the estimate $\hat{\theta}^{(k)}$ depends on the smallest eigenvalue $\lambda_{\min}(A_n)$. The larger this eigenvalue, the faster the convergence. Because A_γ has finite positive eigenvalues, the foregoing acquisition scheme with uniform $\mathbf{r}$ is one in which $\lambda_{\min}(A_n)$ grows as $O(n)$, as the authors suggest at the end of section 2.1. This is simulated in Figure 1 using $\gamma = 1/6$, the proportion used by the authors in their simulation.

The proportion γ can be chosen to maximize the asymptotic smallest eigenvalue $\lambda_{\min}(A_\gamma)$. This is simulated in Figure 2. The figure shows that $\lambda_{\min}(A_\gamma)$ has a sharp maximum at $\gamma = 4/9$, substantially higher than $1/6$.

3. ASYMPTOTIC DISTRIBUTION OF EIGENVALUES

To better understand the results of the authors' section 2.2, explicit expressions for Σ_θ and Σ_U can be obtained as follows. Assume that the acquisition scheme where the $\mathbf{z}_i$'s are iid as in (4). For simplicity, suppose the measurement errors are homoscedastic. Appealing to the strong law of large numbers as before, we have

$$\frac{1}{n}B_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^T\exp(2\mathbf{z}_i^T\theta) \to B_\gamma = E(\mathbf{z}\mathbf{z}^T\exp(2\mathbf{z}^T\theta)), \quad (5)$$

and similarly, as stated in theorem 1c, $(1/n)G_n$ and $(1/n)F_n$ converge to the same limit $F_\gamma = E(\mathbf{z}\mathbf{z}^T\exp(4\mathbf{z}^T\theta) \times \exp(-2\mathbf{z}^T\theta)\sigma^2\epsilon^2) = \sigma^2B_\gamma$. Using Slutsky's theorem, the authors' equation (5) becomes $\sqrt{n}(\hat{\theta}^{(k)} - \theta_*) \to^L \mathrm{N}(0, \Sigma_\theta)$, where $\Sigma_\theta = B_\gamma^{-1}F_\gamma B_\gamma^{-1} = \sigma^2B_\gamma^{-1}$. Because $\theta = (S_0, \beta^T)^T$ and $\beta = \mathrm{vecs}(\mathbf{D})$, assumption (C9) holds under the conditions of theorem 1, and Σ_D is equal to the 6×6 lower diagonal block of Σ_θ.

When $\mathbf{D}$ is isotropic, two simplifications occur. First, $\Gamma = \mathbf{I}_3$, $\Lambda = \lambda\mathbf{I}_3$, and the authors' equation (8) reads $\mathbf{U}_n = \sqrt{n}(\hat{\mathbf{D}} - \mathbf{D}) \to \mathbf{U}$. Thus $\Sigma_U = \Sigma_D$. Second, according to the acquisition scheme (4), $\mathbf{z}^T\theta = \log S_0$ with probability γ and $\mathbf{z}^T\theta = \log S_0 - b\mathbf{r}^T(\lambda\mathbf{I}_3)\mathbf{r} = \log S_0 - b\lambda$ with probability $1-\gamma$; thus

$$B_\gamma = S_0^2\begin{pmatrix} \gamma + (1-\gamma)e^{-2b\lambda} & -(1-\gamma)e^{-2b\lambda}bE(\mathbf{x}^T) \\ -(1-\gamma)e^{-2b\lambda}bE(\mathbf{x}) & (1-\gamma)e^{-2b\lambda}b^2E(\mathbf{x}\mathbf{x}^T) \end{pmatrix}. \quad (6)$$

Note that $B_\gamma = S_0^2A_\gamma$ when $\lambda = 0$, where A_γ is as given by (1).

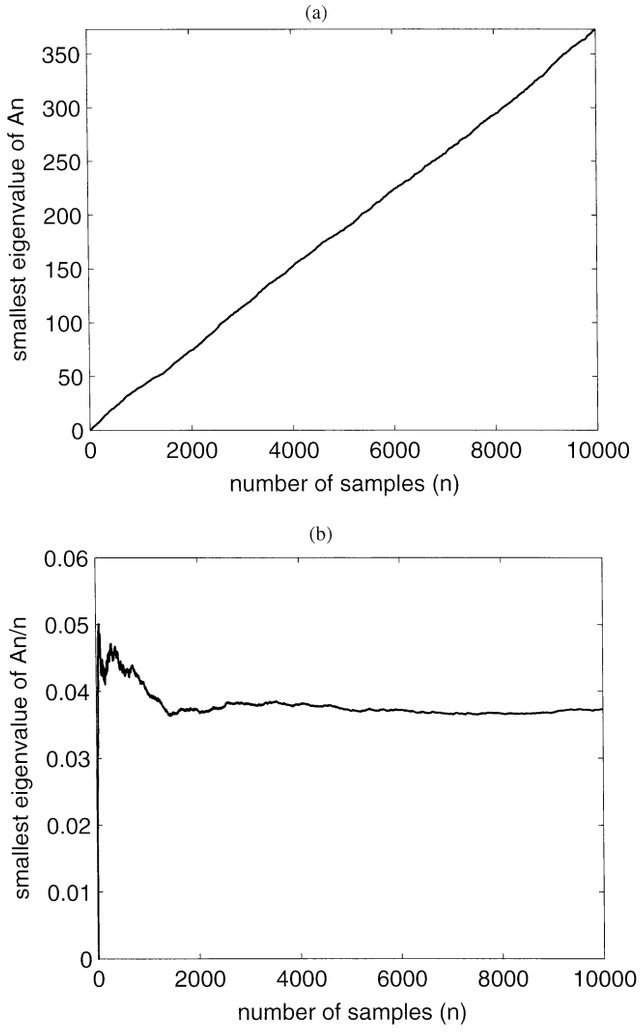

Figure 1. (a) Smallest eigenvalue $\lambda_{\min}(A_n)$ as a function of the number of samples n for spatially uniform acquisition using $\gamma = 1/6$. (b) Smallest eigenvalue $\lambda_{\min}(A_n)/n$. The convergence is to $\lambda_{\min}(A_\gamma) = .0373$.

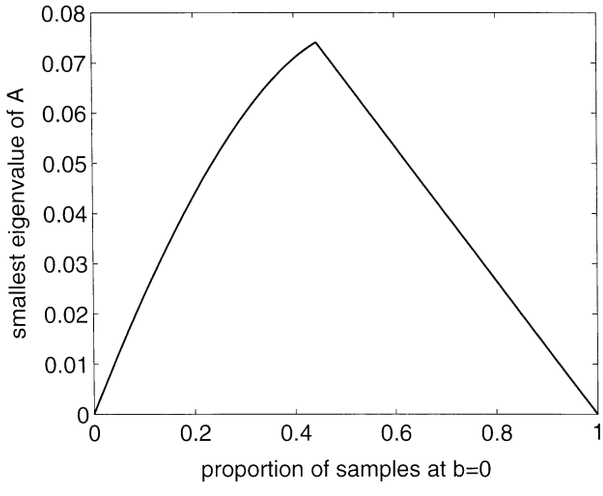

Figure 2. Smallest eigenvalue $\lambda_{\min}(A_\gamma)$ as a function of the proportion γ of samples with $b = 0$ for spatially uniform acquisition. The maximum occurs at $\gamma = 4/9$.

Recall that Σ_U is the lower 6×6 diagonal block of $\Sigma_\theta = \sigma^2 B_\gamma^{-1}$. Applying the Schur inversion formula for block matrices (Schott 2005, p. 247) to (6) and deleting the first row and first column gives

$$\Sigma_U = \frac{\sigma^2}{S_0^2 b^2}\left(\frac{(E(\mathbf{x}\mathbf{x}^T))^{-1}}{(1-\gamma)e^{-2b\lambda}} + \frac{(E(\mathbf{x}\mathbf{x}^T))^{-1}E(\mathbf{x})E(\mathbf{x}^T)(E(\mathbf{x}\mathbf{x}^T))^{-1}}{\gamma + (1-\gamma)e^{-2b\lambda}((1 - E(\mathbf{x}^T)(E(\mathbf{x}\mathbf{x}^T))^{-1}E(\mathbf{x}))}\right). \tag{7}$$

Expression (7) is valid for any distribution of $\mathbf{r}$ on the sphere and demonstrates the dependence on the signal-to-noise ratio and the advantage of using high values of b in the acquisition. When $\mathbf{r}$ is uniform, expressions (2) and (3) reduce (7) explicitly to

$$\Sigma_U = \frac{p(p+2)\sigma^2 e^{2b\lambda}}{4S_0^2 b^2(1-\gamma)}\begin{pmatrix} 2\mathbf{I}_p - C_p(\gamma)\mathbf{1}_p\mathbf{1}_p^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p(p-1)/2} \end{pmatrix}$$

with $C_p(\gamma) = 2/(p+2) - e^{-2b\lambda}(1-\gamma)/\gamma$. In particular for DWI ($p = 3$), we get

$$\Sigma_U = \frac{15\sigma^2 e^{2b\lambda}}{4S_0^2 b^2(1-\gamma)}\begin{pmatrix} 2\mathbf{I}_3 - C_3(\gamma)\mathbf{1}_3\mathbf{1}_3^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \end{pmatrix}, \tag{8}$$

with $C_3(\gamma) = 2/5 - e^{-2b\lambda}(1-\gamma)/\gamma$.

From this calculation, we learn that when $\mathbf{D}$ is isotropic, the diagonal entries of the estimation residuals $\mathbf{U}_n = \sqrt{n}(\hat{\mathbf{D}} - \lambda\mathbf{I}_3) \to \mathbf{U}$ [the covariance of which is indicated by the upper 3×3 diagonal block in (8)] are asymptotically uncorrelated with the off-diagonal entries (the covariance of which is indicated by the lower 3×3 diagonal block). The correlation between the diagonal entries of $\mathbf{U}$ themselves is symmetric with respect to those entries, as expected from the isotropy. In particular, the diagonal entries of $\mathbf{U}$ become uncorrelated $[C_3(\gamma) = 0]$ when the fraction γ is chosen to be equal to

$$\gamma_0 = \frac{1}{1 + 2e^{2b\lambda}/(p+2)},$$

are positively correlated when $\gamma < \gamma_0$ and negatively correlated if $\gamma > \gamma_0$. Moreover, when $\gamma = \gamma_0$, the diagonal entries of $\mathbf{U}$ have exactly twice the variance as the off-diagonal entries, and $\mathbf{U}$ has the distribution known in random matrix theory as the Gaussian orthogonal ensemble (Mehta 1991). This distribution has been proposed for modeling variability of diffusion tensors (Schwartzman 2006). Using $b\lambda = .7$ as done by the authors gives $\gamma_0 = .381$.

Because the interest is usually in estimating $\mathbf{D}$ rather than S_0, the fraction γ could be chosen to minimize the variance in the estimation of λ as opposed to that of the vector θ. Using $\hat{\lambda} = \mathrm{tr}(\mathbf{D})/3$ as the estimate of λ (as the authors suggest at the end of sec. 2.2), we get from (8) that the asymptotic variance of $\sqrt{n}(\hat{\lambda} - \lambda)$ is proportional to

$$f_p(\gamma) = \frac{2 - C_p(\gamma)}{1-\gamma} = \frac{2(p+1)}{(p+2)(1-\gamma)} + \frac{e^{-2b\lambda}}{\gamma}, \tag{9}$$

which is minimized at

$$\gamma^* = \frac{1}{1 + \sqrt{2e^{2b\lambda}(p+1)/(p+2)}}.$$

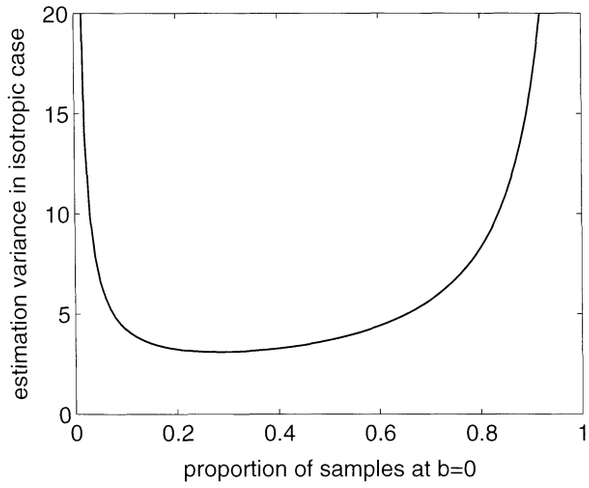

Figure 3. Graph proportional to the asymptotic variance in the estimation of the eigenvalues of $\mathbf{D}$. The minimum occurs at $\gamma = .28$.

Again, using $b\lambda = .7$ as done by the authors gives $\gamma^* = .282$ (Fig. 3). This value is smaller than the optimal γ obtained in the previous section, but the minimum is not as sharp.

When $\mathbf{D}$ is not fully isotropic, the calculations are complicated by the fact that they depend on the particular orientation in space of the anisotropic axes. However, it can be argued that if $\mathbf{r}$ is sampled uniformly, then the estimation variance should not depend on those axes. Then, without loss of generality, one may assume that $\Gamma = \mathbf{I}_3$. Then again, we have that $\Sigma_U = \Sigma_D$, and the required submatrices $\Sigma_{\mathbf{U}_{11}}$ and $\Sigma_{\mathbf{U}_{22}}$ can be taken as the appropriate submatrices of (8). Note that because of the symmetry, we get that in fact, $\Sigma_{\mathbf{U}_{11}} = \Sigma_{\mathbf{U}_{22}}$. Similarly, the variance σ_{ii} in the authors' corollary 2 is equal to any one of the upper three diagonal entries of (8) and thus is also proportional to (9). Interestingly, the minimization problem in terms of γ turns out to be the same as in the isotropic case.

The distribution of the eigenvector matrices $\mathbf{C}$, $\mathbf{C}_{11}$, and $\mathbf{C}_{22}$ in theorems 3 and 4 and corollaries 1 and 2 are more difficult to assess directly from the written densities. However, intuition dictates that $\mathbf{C}$ should be uniformly distributed on the orthogonal group $O(3)$, whereas $\mathbf{C}_{11}$ and $\mathbf{C}_{22}$ should be uniformly distributed on the orthogonal group $O(2)$ in the oblate and prolate cases. The authors did not write this observation explicitly, but their figures 1 and 2 support this conjecture. Figures 2(b) and 2(f) make this clear for the oblate and prolate case. Similarly, the angle histogram plots in figure 1 indicate a uniform distribution on the sphere, because the uniform density in spherical coordinates is $\sin(\phi)/(4\pi)$.

4. SUMMARY AND FINAL REMARKS

Using the authors' results, I have suggested a stochastic acquisition scheme [eq. (4)] and derived from it analytical forms for the covariances Σ_θ and Σ_U. These formulas provide insight into the covariance between the entries of the estimated diffusion tensor $\mathbf{D}$ and validate a model of variability for diffusion tensors proposed by Schwartzman (2006).

In addition, using these formulas, I have calculated the optimal fraction of measurements γ^* at $b = 0$ in two different ways. When the objective is to optimize the rate of convergence of the estimate $\hat{\theta}(k)$, the optimal fraction is $4/9$. When the objective is to minimize the asymptotic variance of the estimate of the eigenvalues of $\mathbf{D}$, the optimal fraction is .28, although this criterion is less sensitive to departures of γ from the optimum. Both values are higher than one would intuitively guess, because measurements at $b = 0$ do not provide information about $\mathbf{D}$. This might be necessary to compensate for the singularity of the design when those measurements are not present.

As a final remark, one issue that the authors overlook is the multiple-testing problem. This issue presents itself in two ways. First, it is present in the hierarchical testing scheme proposed in section 2.3. Classification of a diffusion tensor requires evaluation of one, two, or three tests in a certain order, so p values need to be adjusted accordingly. Second, a large-scale multiple testing problem is present when the classification scheme is applied to the tens of thousands of voxels in the brain. Because of this, the marginal p values reported in Section 4 and Figure 3 are much lower than they would be were multiple testing corrections applied.

ADDITIONAL REFERENCES

Mardia, K. V., and Jupp, P. E. (2005), *Directional Statistics*, West Sussex, U.K.: Wiley.

Mehta, M. L. (1991), *Random Matrices* (2nd ed.), San Diego: Academic Press.

Schott, J. R. (2005), *Matrix Analysis for Statistics*, Hoboken, NJ: Wiley.

Schwartzman, A. (2006), "Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data," unpublished doctoral dissertation, Stanford University, Dept. of Statistics.

Westin, C.-F., Maier, S. E., Mamata, H., Nabavi, A., Jolesz, F. A., and Kikinis, R. (2002), "Processing and Visualization for Diffusion Tensor MRI," *Medical Image Analysis*, 6, 93–108.

Comment

Nicole A. LAZAR

1. INTRODUCTION

Zhu, Zhang, Ibrahim, and Peterson (henceforth ZZIP) set an ambitious agenda of providing a statistical basis for understanding the effect of noise on the diffusion tensors, obtained from diffusion tensor imaging (DTI), which can be used to investigate the structure of the brain. The problem arises because exploration of the neural pathways is based on classification of the tensors into one of four types: nondegenerate, oblate, prolate,

Nicole A. Lazar is Associate Professor, Department of Statistics, University of Georgia, Athens, GA 30602 (E-mail: *nlazar@stat.uga.edu*).

Journal of the American Statistical Association
December 2007, Vol. 102, No. 480, Applications and Case Studies
DOI 10.1198/016214507000001184

and isotropic. This categorization depends on the equality and inequality of the eigenvalues that define the tensor at each pixel. However in practice, as the authors show, the eigenvalues are always different. Furthermore, there is a "sorting bias" in that standard methods in the DTI literature tend to overestimate the largest eigenvalue and underestimate the smallest eigenvalue. Reliable inference regarding the diffusion patterns in the brain depends on the ability to detect when the eigenvalues truly are distinct, and, when they are, which ones differ.

Deducing the structure of the brain, in the sense of identifying the neural pathways followed when a particular cognitive task is performed, goes hand in hand with the other central goal of neuroimaging—deducing brain function. Usually, this is framed as the rather crude question: "Where does activation take place?," although, more productively, we might ask the related question: "And how are regions of activation connected to each other?" This latter is a question of *connectivity*, and it has been suggested (see, e.g., Friston 1998) that in fact this is where the focus of functional neuroimaging should lie.

The statistical challenge of the connectivity question is clear, because it involves, at some level, a notion of *causality*. Neuroscientists often would like to conclude more than simply that "region A and region B coactivate." Although such a finding might be of interest, the statement could be refined to declare that "region A and region B are in a connected network, so that the activation of one precedes the activation of the other, with the activation of the former potentially causing that of the latter." The problem with attempting statistical models that would help quantify such statements is that the temporal and/or spatial resolution of the data (depending on the particular functional imaging modality used) generally are not sufficiently high to support them. For instance, the temporal resolution of functional magnetic resonance imaging (fMRI) is orders of magnitude too slow to model the flow of activation from one region to another, let alone to infer causality. At best, then, with present technologies, we can hope to understand which regions are connected, but evidently not the mechanisms behind those connections.

In this discussion, I focus on the following issues: (1) combinations of modalities to help elucidate connectivity of brain networks; (2) visualization of complex, high-dimensional imaging data; and (3) distribution of the ordered eigenvalues.

2. COMBINING MULTIPLE MODALITIES: DTI + fMRI

The importance of understanding brain connectivity, coupled with the limitations of any single imaging modality, is one motivation for researchers to look at the possibilities inherent in combining multiple imaging modalities. Namely, by taking advantage of the strengths available from each, one can attempt to build a more complete picture of the neurologic processes at the requisite level of spatial and temporal detail. Often scientists will consider multiple functional imaging techniques in tandem, for instance, electroencephalography (EEG) and fMRI or magnetoencephalography (MEG) and fMRI. However, I wonder at the potential for combining instead a structural modality, such as diffusion tensor MRI, with a functional modality, such as fMRI. Here both techniques are based on magnetic resonance, and it is an intriguing possibility to ponder the power of putting them together.

In general, if model-dti is the diffusion tensor model, such as that proposed by ZZIP (with parameter vector $\boldsymbol{\theta}_{dti}$), and model-fmri is the functional MRI model, which could be one of the many proposed in the literature (with parameter vector $\boldsymbol{\theta}_{fmri}$), then the notion is to consider such formulations as $response = f(\boldsymbol{\theta}_{dti}, \boldsymbol{\theta}_{fmri})$. A number of challenges with this type of formulation present themselves:

1. How to define *response*. DTI and fMRI measure different signals. These could perhaps be combined into a single response at each voxel; alternatively, *response* itself can be vector-valued, shifting to a multivariate setting. If the former approach is taken, one still has to decide how to combine the measured signals from the two modalities.
2. The form of $f(\cdot, \cdot)$. Clearly, a wide range of models are available at this level; different models would place more or less emphasis on the structural side of the image or the functional. Modern nonparametric methods, such as generalized additive models, might be helpful here, although the size of the data sets would no doubt present difficulties in computation and interpretation. Other possibilities would be to embed both the model-dti and the model-fmri in a hierarchical (Bayesian) framework, and to incorporate functional information directly into ZZIP's semiparametric model for DTI.
3. Models for the correlation across voxels, necessary for a more complete understanding of the brain networks of interest, are perforce more complicated now. Presumably, structure helps to dictate function at least to some extent, although I suspect that these types of correlations are largely unknown. How to examine the correlation structure both across and within modalities (BOLD fMRI and DTI) is a nontrivial problem; a detailed study of the correlation structure of fMRI data alone is not easy.
4. Visualization of results. ZZIP have some interesting graphics describing the distribution of tensor types in the brain, as well as the pathways deduced from their inferential approach. Building in inference regarding function adds a dynamic aspect to the problem—function evolving and changing over time—and good graphical representations become even more important. I am an advocate of "statistical movies" in which the model(s) and its (their) estimates are plotted along some index (e.g., time, Markov chain Monte Carlo run, iteration of EM algorithm, as appropriate). Such visual inferential tools are underdeveloped in the neuroimaging community; a model which combines analyses of structure and dynamic function seems like an ideal testing ground for these ideas.

Despite the difficulties, it is evident that a model that uses both structural information on neural pathways and functional information on neuronal activity in response to particular stimuli should yield valuable insight into the crucial question of connectivity. A dynamic component seems to be needed here, so I imagine that one would need to shy away from the standard analyses of fMRI data, which summarize an experiment in a static statistical parametric map.

3. VISUALIZATION OF IMAGING DATA

Expanding on the last of the points listed in the preceding section, good, effective graphics are an essential part of understanding a sophisticated analysis, such as that suggested by ZZIP. I particularly like the authors' figure 4(d) of the ellipsoid distribution in the region of interest, which presents a simple, yet rather elegant representation of the results. Due to its simplicity, one could even "layer" other pieces of information on top of it or within it—some notion of standard error, for example. Different shades of colors or different-sized ellipsoids or side-by-side plots of ellipsoid estimates with standard errors, could be used for this purpose. This would be a variant of the authors' figure 6, which gives lower and upper "confidence brains" for some of the measures that ZZIP explore.

In terms of dynamic visualization, I can think of several potentially useful possibilities. For instance, one could plot the pathways determined by the DTI model, as in the authors' figure 4, and overlay on these the measured fMRI signal or some model of the fMRI signal at each time point. Alternatively, one could plot the signal only for that group of voxels detected as active by some measure, again over time or even statically. If we embed model-dti and model-fmri in a hierarchical framework, as conceptualized in the previous section, then the analysis path will dictate appropriate graphical representations; for example, a Bayesian analysis will admit plotting posterior probabilities (jointly or marginally) of activation and the various tensor morphologies.

Plotting structure and function together raises a multitude of other questions, for instance:

- Do the voxels that are declared active by the fMRI model line up at all with the pathways detected by the DTI model?
- Are clusters of active voxels "linked" by the diffusion tensors?
- Do these links change over time, and if so, how?
- Is one type of tensor morphology more related to functional activation than others?
- What do these connections tell us about the functional connectivity in the brain?

Good visualization tools also are critical for comparing and contrasting the results from multiple slices of the brain and multiple subjects, particularly the latter. It is quite common when reporting on the results of a functional neuroimaging analysis to show a "representative" slice or subject, yet how much more convincing would it be to show summary graphics for multiple subjects, for example. This opens up a whole new array of questions, such as:

- How much variability is there in the pathways detected by model-dti?
- Do different subjects show the same patterns of tensor morphologies?
- Are some paths present for all, or nearly all, of the subjects?
- Do these correspond to stronger or more consistent activation across subjects from model-fmri?
- How consistent are the connected networks from subject to subject?

Cognitive neuroscientists are obviously interested in making general statements about groups of subjects; effective visualization can be one tool for making these statements. I have done some work on combining the results from individual subject maps into group maps (Lazar, Luna, Sweeney, and Eddy 2002), but there remain many important issues to explore and develop.

4. DISTRIBUTION OF ORDERED EIGENVALUES

A basic element of ZZIP's approach is to use the (ordered) eigenvalues in the decision on how to classify the diffusion tensor observed at each voxel. But, as they observe, there are three crucial difficulties with this approach. First, the eigenvalues will always be different from one another in any of the real DTI data sets in question; because several of the tensor classes are defined through equality of particular eigenvalues, this is problematic and leads to a question of statistical inference, namely: "How different is different?" The sorting bias mentioned earlier makes this determination harder. These problems lead to the final difficulty—the distribution of the ordered eigenvalues is hard to derive, and thus the inferential question is difficult to address.

ZZIP tackle the problem by building a model within which the distribution of the eigenvalues can be approximated; this also allows them to construct a sequence of hypothesis tests that yields the required classification at each pixel. They mention early that the bootstrap does not yield satisfactory results for this problem, but do not give much detail. Efron and Tibshirani (1993) provided a general discussion of bootstrap for the evaluation of eigenvalues; however, they did not explicitly mention the sorting bias.

To explore the performance of the bootstrap and the importance of the sorting bias more closely, I looked at a simple isotropic case. Namely, I generated three independent Normal samples of size $n = 50$, each with mean 0 and variance 4. This can be considered a random sample from $\mathrm{N}_3(0, 4 * I)$, a spherical distribution.

Next, I considered two naive nonparametric bootstrap schemes, what I call "cases" and "units." In the "cases" scheme, the three samples were treated as 3 measurements on 50 individuals, and indices were sampled with replacement from integers 1–50. If, for example, the index "23" was chosen, then the 23rd observation from each of the 3 samples was included in the bootstrap resample. In the "units" scheme, each of the 150 data points (50 in each of the 3 original samples) was considered a separate unit, and these were the candidates for resampling; thus indices were sampled with replacement from the integers 1–150. The resampled units were then arranged in an array that reproduced the structure of the original data matrix. Evidently the cases scheme is the more realistic one for our purposes, because it mimics the situation in which multiple measurements are available for each observation. Finally, I obtained the eigenvalues of the covariance and correlation matrices of each bootstrap resample under each of the two schemes. I examined the isotropic case, because this was the best test bed for assessing the severity of the sorting bias (with the smallest eigenvalue underestimated and the largest overestimated). Any scenario in which some of the eigenvalues are postulated a priori to be distinct should be easier to handle.

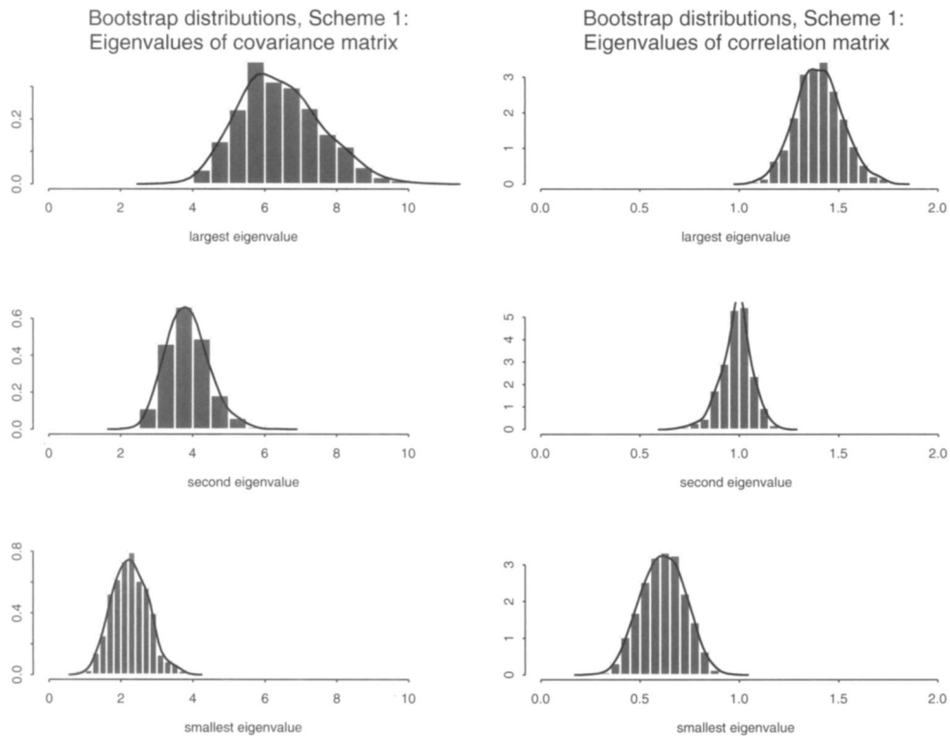

Figure 1. Bootstrap distributions for each ordered eigenvalue, case-based resampling scheme. The left column shows the eigenvalues for the covariance matrix of the samples; the right column shows those for the correlation matrix.

I now present some results graphically. Figure 1 shows the distribution for each of the eigenvalues separately, using the case-based resampling scheme; Figures 2 and 3 show the bootstrap distributions of all three eigenvalues, along with the true values (denoted by "e") and the bootstrap means (denoted by "b"), for the covariance and correlation matrices. Figures 4, 5, and 6 show the same for the unit-based resampling. As can be seen, the sorting bias is a particularly serious problem for the case-based resampling: the means of the sorted eigenvalues are more extreme than the true values for both the covariance and correlation matrices, and these are themselves already subject to a sorting bias. In other words, the case-based bootstrap exacerbates the bias. On the other hand, unit-based resampling appears to have a shrinkage effect. Indeed, for both sets of eigenvalues (covariance matrix and correlation matrix), the means of the sorted eigenvalues are less extreme than the true values. There is considerably more overlap of the bootstrap distributions for the unit-based bootstrap, another indication that the estimation bias is not as severe in this scheme. Table 1 gives the actual eigenvalues for the two matrices, along

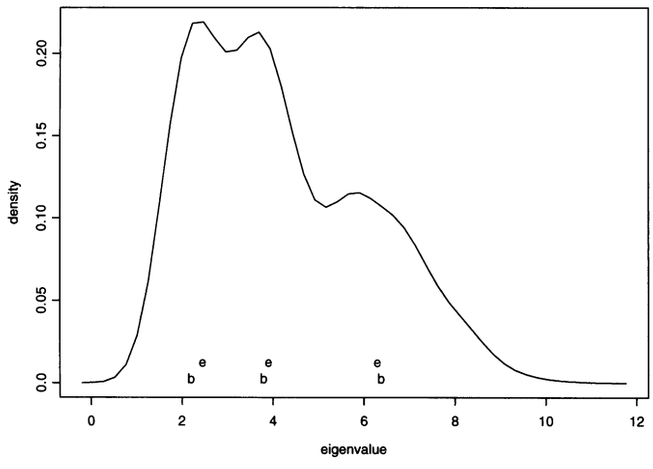

Figure 2. Bootstrap distribution (Scheme 1) for the eigenvalues of the covariance matrix, case-based resampling scheme. Here "b" indicates the bootstrap mean, and "e" is the eigenvalue in the original sample.

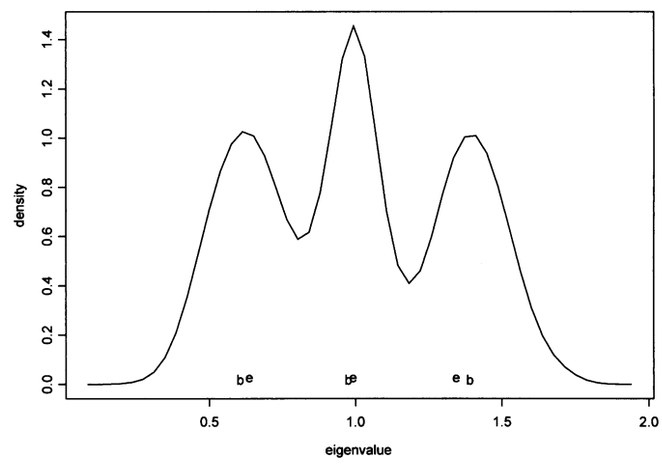

Figure 3. Bootstrap distribution (Scheme 1) for the eigenvalues of the correlation matrix, case-based resampling scheme. Here "b" indicates the bootstrap mean, and "e" is the eigenvalue in the original sample.

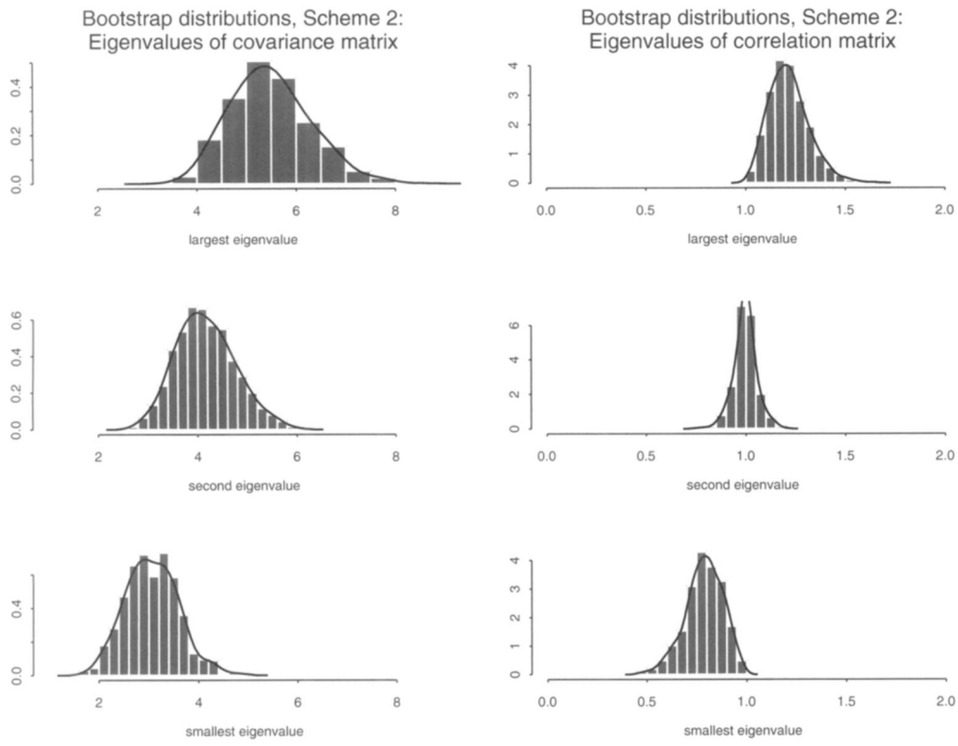

Figure 4. Bootstrap distributions for each ordered eigenvalue, unit-based resampling scheme. The left column shows the eigenvalues for the covariance matrix of the samples; the right column shows those for the correlation matrix.

with the bootstrap estimates from the case-based and unit-based analyses.

5. CONCLUSION

Mathematicians, statisticians, computer scientists, and engineers have had much to say about brain science, particularly over the last two decades as imaging techniques have flourished and advanced. However, the need continues for sophisticated, realistic models that effectively use the information available from the data. Although progress has been made, there remains room for improvement. The authors have made a valuable contribution to the important problems of discovering and understanding the neural pathways of the brain. Knowledge of these systems is critical for modeling connectivity between functioning brain regions. As in any rich field of research, one question leads to another; even a partial solution brings us that much closer to some of the answers that have been eluding scientists for decades, while at the same time stimulating more questions for future investigators to ponder. I congratulate the authors on

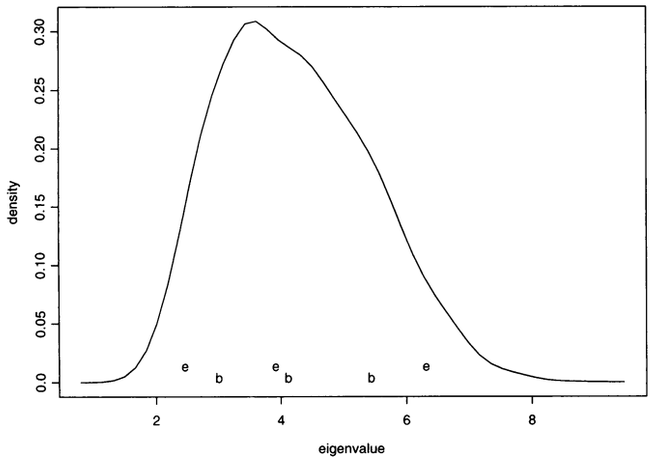

Figure 5. Bootstrap distribution (Scheme 2) for the eigenvalues of the covariance matrix, unit-based resampling scheme. Here "b" indicates the bootstrap mean, and "e" is the eigenvalue in the original sample.

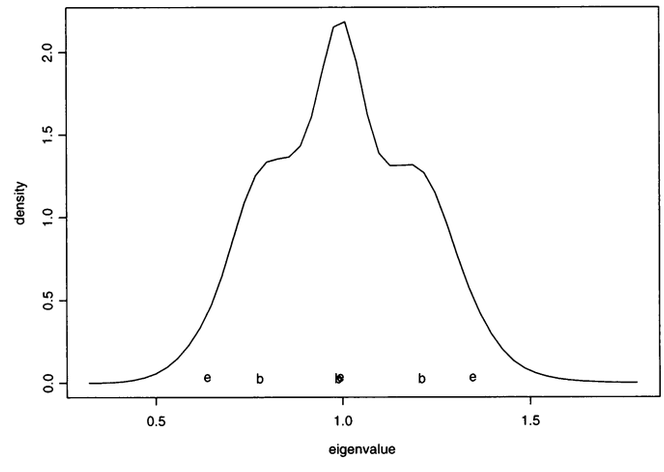

Figure 6. Bootstrap distribution (Scheme 2) for the eigenvalues of the correlation matrix, unit-based resampling scheme. Here "b" indicates the bootstrap mean, and "e" is the eigenvalue in the original sample.

Table 1. True eigenvalues and bootstrap estimates based on resampling cases or units

	Covariance			Correlation		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
True	6.357	3.961	2.509	1.353	1.001	.645
Case mean	6.438	3.853	2.257	1.401	.985	.615
Unit mean	5.484	4.163	3.059	1.218	.995	.786

NOTE: The case-based scheme results in eigenvalue estimates that are even more extreme than the true values, an exacerbated sorting bias. The unit-based resampling ameliorates the bias problem to a limited extent: the first eigenvalue is still overestimated and the third is still underestimated, but the bias is not as large.

an article that presents both possible solutions and stimulus for research in new directions.

ADDITIONAL REFERENCES

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Boca Raton, FL: Chapman & Hall/ CRC.

Friston, K. J. (1998), "Imaging Neuroscience: Principles or Maps?" *Proceedings of the National Academy of Sciences USA*, 95, 796–802.

Lazar, N. A., Luna, B., Sweeney, J. A., and Eddy, W. F. (2002), "Combining Brains: A Survey of Methods for Statistical Pooling of Information," *NeuroImage*, 16, 538–550.

Rejoinder

Hongtu ZHU, Heping ZHANG, Joseph G. IBRAHIM, and Bradley S. PETERSON

1. RESPONSE TO DR. SCHWARTZMAN

Dr. Schwartzman presented several interesting findings on the acquisition and asymptotic distributions of eigenvalues and eigenvectors. Specifically, he developed a stochastic acquisition scheme and derived the analytical forms of the covariance matrices Σ_θ and Σ_U. Based on these theoretical results, he obtained an optimal ratio for γ and showed that $\gamma > 0$ is a necessary condition to ensure the validity of our theorems 1–5. He also noted that for an isotropic tensor, $\mathbf{C}$ should be distributed uniformly on the orthogonal group $O(3)$ [figs. 1(c) and 1(d)], whereas similar observations should be valid for $\mathbf{C}_{11}$ for the oblate tensor and for $\mathbf{C}_{22}$ for the prolate tensor [figs. 2(b) and 2(c)].

Our results in theorem 1 can be used not only to choose the optimal ratio γ, but also to select an optimal acquisition scheme, including gradient directions and strengths of the diffusion gradients (Jones, Horsfield, and Simmons 1999). To accurately estimate $\mathbf{D}$ and its associated measures (e.g., eigenvalues and eigenvectors), we can use $\mathrm{tr}\{\mathrm{cov}[\hat{\theta}]\} = \mathrm{tr}\{[B_n(\theta)]^{-1}F_n(\theta)[B_n(\theta)]^{-1}\}$ to construct a design criterion. But, because $F_n(\theta)$ involves the variance of background noise and this is unknown at the design stage, using $\mathrm{tr}\{\mathrm{cov}[\hat{\theta}]\}$ as a design criterion is not tractable. Instead, we suggest minimizing $\mathrm{tr}\{[B_n(\theta)]^{-1}\}$ directly. Letting $\mathbf{z}_i^T = (1, \mathbf{x}_i^T)$, $B_n(\theta)$ can be written as $B_n(\theta) = S_0^2 \sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i^T \exp(2\mathbf{x}_i^T\beta)$. Because S_0 is the signal intensity at baseline and because our interest is primarily in the diffusion tensor, S_0 can be dropped. Therefore, for n acquisitions, and for a given diffusion tensor $\mathbf{D}$ or β, we minimize

$$C(\mathbf{b}, \mathbf{r}|\beta) = \mathrm{tr}\left\{\left[\sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i^T \exp(2\mathbf{x}_i^T\beta)\right]^{-1}\right\}, \tag{1}$$

where $\mathbf{b} = (b_1, \ldots, b_n)$ and $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)$. If we have a set of diffusion tensors $\{\beta_k : k = 1, \ldots, K_0\}$ from several regions of interest, then we can calculate the optimal $(\mathbf{b}, \mathbf{r})$ by solving

$$(\mathbf{b}_{opt}, \mathbf{r}_{opt}) = \underset{(\mathbf{b},\mathbf{r})}{\mathrm{argmin}} \sum_{k=1}^{K_0} C(\mathbf{b}, \mathbf{r}|\beta_k). \tag{2}$$

Moreover, because the gradient strengths b_i are determined by the parameters of the pulse sequence, such as the pulse separation time Δ, echo time (TE), and pulse duration δ (Alexander and Barker 2005), we can incorporate these parameters into (2), and then optimize the pulse sequence parameters and the gradient directions simultaneously. Thus we have a general method for determining the optimal acquisition scheme.

Two important issues pertaining to (2) are worth noting. The first issue is the importance of selecting the set of diffusion tensors in (2), which may be extracted from either the existing diffusion tensor data or specific ROIs. The second issue is to numerically find the optimal solution of (2), for which we can use optimization methods, such as simulated annealing, conjugate gradient methods, quasi-Newton methods, and genetic algorithms (Kirkpatrick, Gelatt, and Vecchi 1983; Mitchell 1996). We omit here the details of how to solve these two issues, deferring them for presentation elsewhere.

We conducted Monte Carlo simulations to compare the estimated DTI measures under two different imaging acquisition schemes, which we term traditional and optimal. Each of these acquisition schemes consisted of $m = 3$ baseline images, in which $b = 0$ s/mm^2, and $n - m = 12$ images, in which

Hongtu Zhu is Associate Professor of Biostatistics, Department of Biostatistics and Biomedical Research Imaging Center, University of North Carolina, Chapel Hill, NC 27599-7420 (E-mail: *hzhu@bios.unc.edu*). Heping Zhang is Professor of Biostatistics, Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520-8034 (E-mail: *heping.zhang@yale.edu*). Joseph G. Ibrahim is Alumni Distinguished Professor of Biostatistics, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420 (E-mail: *ibrahim@bios.unc.edu*). Bradley S. Peterson is Professor of Psychiatry, Department of Psychiatry, Columbia University Medical Center and the New York State Psychiatric Institute, 1051 Riverside Drive, Unit 74, New York, NY 10032 (E-mail: *petersob@childpsych.columbia.edu*). We thank the editor for organizing the discussion of our article and the discussants for their stimulating comments and insightful contributions to this work.

Journal of the American Statistical Association
December 2007, Vol. 102, No. 480, Applications and Case Studies
DOI 10.1198/016214507000001201

Table 2. Comparisons of the acquisition schemes

	λ_1	λ_2	λ_3	RA	FA	$\arccos\lvert\langle \mathbf{e}_1, \mathbf{v}_1\rangle\rvert$	
True	1.7	.2	.2	.7143	.8704	.0	SNR
Traditional acquisition: MM (SD)							
	$1.6469_{(.429)}$	$.3778_{(.223)}$	$-.0486_{(.198)}$	$.8187_{(.307)}$	$.8928_{(.124)}$	$.2560_{(.149)}$	5
	$1.7294_{(.277)}$	$.2973_{(.112)}$	$.0706_{(.110)}$	$.7473_{(.123)}$	$.8809_{(.069)}$	$.1298_{(.070)}$	10
	$1.7129_{(.180)}$	$.2685_{(.075)}$	$.1166_{(.073)}$	$.7292_{(.081)}$	$.8753_{(.047)}$	$.0875_{(.049)}$	15
	$1.7071_{(.130)}$	$.2522_{(.056)}$	$.1377_{(.055)}$	$.7237_{(.060)}$	$.8738_{(.035)}$	$.0651_{(.036)}$	20
	$1.7050_{(.103)}$	$.2417_{(.044)}$	$.1519_{(.043)}$	$.7203_{(.048)}$	$.8726_{(.028)}$	$.0520_{(.029)}$	25
	$1.7035_{(.086)}$	$.2366_{(.037)}$	$.1602_{(.036)}$	$.7178_{(.039)}$	$.8715_{(.024)}$	$.0434_{(.024)}$	30
Optimal acquisition: Mean (SD)							
	$1.7477_{(.529)}$	$.3691_{(.226)}$	$-.0478_{(.193)}$	$.8261_{(.299)}$	$.8985_{(.122)}$	$.2210_{(.142)}$	5
	$1.7217_{(.234)}$	$.2959_{(.108)}$	$.0843_{(.099)}$	$.7401_{(.111)}$	$.8781_{(.064)}$	$.1009_{(.053)}$	10
	$1.7081_{(.150)}$	$.2655_{(.071)}$	$.1253_{(.067)}$	$.7256_{(.073)}$	$.8738_{(.043)}$	$.0666_{(.035)}$	15
	$1.7031_{(.112)}$	$.2494_{(.053)}$	$.1445_{(.050)}$	$.7210_{(.055)}$	$.8725_{(.033)}$	$.0496_{(.026)}$	20
	$1.7032_{(.089)}$	$.2403_{(.042)}$	$.1555_{(.041)}$	$.7189_{(.043)}$	$.8720_{(.026)}$	$.0400_{(.021)}$	25
	$1.7019_{(.075)}$	$.2344_{(.035)}$	$.1640_{(.034)}$	$.7167_{(.036)}$	$.8710_{(.022)}$	$.0331_{(.017)}$	30

NOTE: Six different SNRs $\{5, 10, 15, 20, 25, 30\}$ and 10,000 simulated data sets were used for each case. Estimated eigenvalues, RA, FA, and $\arccos\lvert\langle \mathbf{e}_1, \mathbf{v}_1\rangle\rvert$ were calculated. MM denotes the mean of means, and SD denotes the standard deviation.

$b = 1{,}000$ s/mm^2. At $b = 1{,}000$ s/mm^2, the traditional acquisition involved 12 directions of diffusion gradients arranged uniformly in three-dimensional space, whereas the optimal acquisition consisted of 12 gradient directions that were the solution of (2) when $\mathbf{D} = \mathrm{diag}(1.7, .2, .2)$. Six different Signal to Noise Ratios (SNRs) $\{5, 10, 15, 20, 25, 30\}$ were selected for all Monte Carlo simulations. For the given diffusion-tensor at each SNR and acquisition scheme, 10,000 diffusion-weighted data sets were generated. The weighted least squares estimates $\widehat{\theta}^{(1)}$, their eigenvalue-eigenvector pairs, RA and FA values, and $\arccos\lvert\langle \mathbf{e}_1, \mathbf{v}_1\rangle\rvert$, where $\mathbf{v}_1$ is the true principal direction $\mathbf{v}_1 = (1, 0, 0)^T$, were calculated. Compared with the traditional acquisition, the optimal acquisition leads to smaller biases and smaller standard deviations for most of the calculated DTI measures, at all SNRs (Table 1). In particular, the principal directions estimated from the optimal acquisition are much closer to the true principal direction at all SNRs than those estimated from the traditional acquisition.

We agree in part with Dr. Schwartzman concerning the issue of multiple comparisons. Because three tests are performed on individual voxels, we could correct for multiple comparisons at each voxel by choosing a relatively small significance level. However, the problem of multiple comparisons over the entire cerebrum is not a central one for us, because we are not comparing voxels across the human brain simultaneously. Rather, our classification of tensor morphologies is a clustering method. We are primarily interested in determining the shape of the diffusion tensor at each voxel of the image to enhance the tracking of fiber pathways. The question here is "Is this voxel degenerate?," not "Are there any degenerate voxels in the image?"

2. RESPONSE TO DR. LAZAR

Dr. Lazar presented three very important issues in the analysis of DTI data sets. The first of these is the combination of DTI data with other imaging modalities, including functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), to identify and understand the neural networks of interacting brain regions. In various neuroimaging studies, fMRI has been widely used to investigate how various brain regions functionally couple together and how they change their level of activation in response to specific stimuli and behavioral tasks. Recent fMRI studies have established that low-frequency (<.08 Hz) physiological fluctuations are temporally correlated between functionally related areas, such as motor, auditory, and visual cortices, likely reflecting the connectivity of those regions (Lowe, Mock, and Sorenson 1998). Although various analytical methods, such as independent component analysis, are improving rapidly (Rogers, Morgan, Newton, and Gore 2007), measuring the low-frequency correlations in fMRI data remains a challenging task (Lowe et al. 1998). Nevertheless, using a map of regions that are functionally connected together with DTI data may provide a direct measurement of the degree to which fiber tracts anatomically connecting those regions that are presumably "functionally connected." We agree that eventually combing DTI data with fMRI data may improve our understanding of the relationship between brain function and its structure (Kim and Kim 2005; Schonberg, Pianka, Hendler, Pasternak, and Assaf 2006).

Dr. Lazar suggested a generic model for analytically combining DTI and fMRI data as follows:

$$response = f(\theta_{dti}, \theta_{fmri}), \tag{3}$$

where θ_{dti} is a map from the high-dimensional diffusion-weighted images to summary data, such as fiber tracts; θ_{fmri} is a map from the massive functional images to summary data; and $f(\cdot, \cdot)$ is a map for combining θ_{dti} and θ_{fmri}. Thus we have a hierarchical model from the raw DW and functional images at the bottom most level to the response variable at the top most level. We believe that this "*response*" should be a summary scalar or vector related to a specific scientific question, which should be formulated as a test of the equality/inequality of specific parameters/functions.

This hierarchical model (3) may be applied at both the subject and population levels. For instance, suppose that the question of interest is to investigate how a neural network, one that

permits interaction across several particular brain regions, differs across subjects or populations. Thus, for any subject, "*response*" may be defined as the degree of connectedness among each of the various regions within the neural network, θ_{dti} may be defined as the probability of the fiber tracts connecting those brain regions, θ_{fmri} may be defined as the path diagrams established by using some statistical methods such as Granger causality (Goebel, Roebroeck, Kim, and Formisano 2003; Roebroeck, Formisano, and Goebel 2005), and $f(\cdot,\cdot)$ may be a parametric/nonparametric function. Furthermore, we may incorporate some covariates of interest (e.g., age, gender) into (3) and develop a general model for the relationship of structural and functional connections as

$$response = f(\theta_{dti}, \theta_{fmri}, \mathbf{x}), \tag{4}$$

where $\mathbf{x}$ denotes a vector of covariates, such as age, gender, or diagnostic status. Furthermore, we need to develop computational algorithms for estimating the unknown parameters in (4) and to develop rigorous procedures for statistical inference, such as test statistics, to address specific hypotheses about the relationship of structural with functional connectivity.

We agree wholeheartedly with Dr. Lazar's comments on the importance of developing effective techniques of visualization for understanding the effects of any imaging analytic procedures. Tools for visualizing neuroimaging data can be implemented in at least two ways. The first of these is to plot a biological measurement of interest, which is typically represented by an intensity value, a vector, or a matrix, such as a diffusion tensor, at each voxel of the image. Representing a long vector or a large matrix in a single voxel can be extremely difficult. Furthermore, we need to represent these biological measures of interest at multiple voxels of each brain (often at hundreds of thousands of voxels). Although plotting univariate measurements in all voxels for each slice of an image is relatively simple, displaying multivariate imaging measurements together across multiple voxels [fig. 4(d)] is largely underdeveloped. The second way is to display imaging data from multiple biological measures from multiple individuals across multiple modalities. This display can be used to, for example, map brain variability across human populations, growth rates of brain tumors, and loss of brain tissue in patients with various neuropsychiatric diseases (Gogtay et al. 2004; Jones 2004). For any formal data analysis, effective graphics are critical for imaging scientists to understand the imaging data at hand, to model inherent noise in the data, and to develop sophisticated models and methods of statistical inference. Analyzing the high-dimensional and complex imaging data sets requires methodological development across multiple disciplines, including mathematics, statistics, physics, computer science, and neuroscience.

Dr. Lazar emphasized the importance of classifying tensor morphologies, which we partially addressed in our article, along with the difficulty of addressing inferential questions, such as the distributions of eigenvalues and eigenvectors, because of the presence of "sorting" bias. She also explored the performance of two nonparametric bootstrap methods (cases and units) under the presence of sorting bias. Her results clearly showed that the sorting bias truly influenced both bootstrap methods, although the unit-based bootstrap method seemed to perform better than the case-based one. She also pointed out our claims regarding the unsatisfactory results using bootstrap methods.

We would like to clarify why some bootstrap methods cannot avoid the sorting bias problem when characterizing the distributions of DT-related quantities (e.g., eigenvalues and eigenvectors). Bootstrapping methods, including repetition bootstrap and wild bootstrap, have recently received much attention in the DTI literature, because they may quantify numerically the effects of noise on the eigenvalues and eigenvectors as well as on the fidelity of fiber tracking (Pajevic and Basser 2003; Basser and Jones 2002; Whitcher, Tuch, Wisco, Sorenson, and Wang 2007; Chung, Lu, and Henry 2006; Lazar and Alexander 2005). Our results, however, showed that some bootstrapping methods, including the wild bootstrap, are invalid for degenerate tensors.

We show here only that the wild bootstrap fails to characterize the distribution of eigenvalues for isotropic tensors, but we will present the detailed derivations for all degenerate tensors elsewhere. For simplicity, we consider only the least squares estimator $\hat{\theta}_{LS}$ and its associated DT, $\hat{\mathbf{D}}_{LS}$. According to theorems 1 and 3, we can show that

$$\mathbf{U}_n = \sqrt{n}(\hat{\mathbf{D}}_{LS} - \lambda \mathbf{I}_3) \to^L \mathbf{U}$$

and

$$\mathbf{H}_n \to^L \mathbf{H}_{LS},$$

where $\mathrm{vecs}(\mathbf{U}) \sim \mathrm{N}(\mathbf{0}, \Sigma_{LS})$ and $\mathbf{H}_{LS}$ is a random variable, and its density $p(\mathbf{h})$ is given by

$$\int (h_1 - h_2)(h_2 - h_3)(h_1 - h_3) \times \exp\left\{-\frac{1}{2}\,\mathrm{vecs}(\mathbf{C}^T\mathbf{H}_{LS}\mathbf{C})^T \Sigma_{LS}^{-1}\,\mathrm{vecs}(\mathbf{C}^T\mathbf{H}_{LS}\mathbf{C})\right\} d\mathbf{C}.$$

We generate a wild bootstrap sample $\{(S_i^*, \mathbf{z}_i) : i = 1, \ldots, n\}$ as follows (Zhu et al. 2007):

$$\log(S_i^*) = \mathbf{z}_i^T \hat{\theta}_{LS} + a_i \tilde{\varepsilon}_i \varepsilon_i^*, \tag{5}$$

where $a_i = 1/[1 - \mathbf{z}_i^T (\sum_{j=1}^n \mathbf{z}_j \mathbf{z}_j^T)^{-1} \mathbf{z}_i]$, $\tilde{\varepsilon}_i = \log S_i - \mathbf{z}_i^T \hat{\theta}_{LS}$ and ε_i^* are independently and identically distributed as ± 1 with equal probability. After some calculations, we can show that

$$\begin{aligned}\hat{\theta}_{LS}^* &= \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T\right)^{-1} \sum_{i=1}^n \mathbf{z}_i \log(S_i^*) \\ &= \hat{\theta}_{LS} + \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T\right)^{-1} \sum_{i=1}^n \mathbf{z}_i a_i \tilde{\varepsilon}_i \varepsilon_i^*.\end{aligned}$$

Thus, using the large-sample theorem of van der Vaart and Wellner (1996; secs. 2.9 and 3.6), we can prove that conditional on the data $\{(S_i, \mathbf{z}_i) : i = 1, \ldots, n\}$,

$$\mathbf{U}_n^* = \sqrt{n}(\hat{\mathbf{D}}_{LS}^* - \hat{\mathbf{D}}_{LS}) \to^L \mathbf{U},$$

where $\hat{\mathbf{D}}_{LS}^*$ is the DT of $\hat{\theta}_{LS}^*$. Let $\tilde{\mathbf{U}}_n^* = P_n \mathbf{U}_n^* P_n^T$ and $\hat{\mathbf{D}}_{LS} = P_n^T \Delta_n P_n$ denotes the spectral decomposition of $\hat{\mathbf{D}}_{LS}$ with $\Delta_n = \mathrm{diag}(\psi(\hat{\mathbf{D}}_{LS}))$, where $\psi(\mathbf{M})$ denotes the vector of ordered eigenvalues of a matrix $\mathbf{M}$. Finally, using theorems 4.1 and 4.2 of Eaton and Tyler (1991), we can show that

$$\psi(\mathbf{U}_n^*) - \{\psi(\tilde{\mathbf{U}}_n^* + \mathbf{A}_n) - \psi(\mathbf{A}_n)\} \to^L \mathbf{0}, \tag{6}$$

where $\mathbf{A}_n = \sqrt{n}(\Delta_n - \lambda \mathbf{I}_3)$. Then, following the arguments of Eaton and Tyler (1991), we establish the proof of inconsistency of the wild bootstrap for isotropic tensors.

ADDITIONAL REFERENCES

Alexander, D. C., and Barker, G. J. (2005), "Optimal Imaging Parameters for Fibre-Orientation Estimation in Diffusion MRI," *NeuroImage*, 27, 357–367.

Chung, S., Liu, Y., and Henry, R. G. (2006), "Comparison of Bootstrap Approaches for Estimation of Uncertainties of DTI Parameters," *NeuroImage*, 33, 531–541.

Eaton, M. L., and Tyler, D. E. (1991), "On Wielandt's Inequality and Its Application to the Asymptotic Distribution of the Eigenvalues of a Random Symmetric Matrix," *The Annals of Statistics*, 19, 260–271.

Goebel, R., Roebroeck, A., Kim, D. S., and Formisano, E. (2003), "Investigating Directed Cortical Interactions in Time-Resolved FMRI Data Using Vector Autoregressive Modeling and Granger Causality Mapping," *Magnetic Resonance Imaging*, 21, 1251–1261.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Herman, D. H., Nugent, T. F., Clasen, L., Toga, A. W., Rapoport, J. L., and Thompson, P. M. (2004), "Dynamic Mapping of Human Cortical Development During Childhood Through Early Adulthood," *Proceedings of the National Academy of Sciences*, USA 101, 8174–8179.

Jones, D. K. (2004), "The Effect of Gradient Sampling Schemes on Measures Derived From Diffusion Tensor MRI: A Monte Carlo Study," *Magnetic Resonance in Medicine*, 51, 807–815.

Kim, D. S., and Kim, M. (2005), "Combining Functional and Diffusion Tensor MRI," *Annals of the New York Academy of Sciences*, 1064, 1–15.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983), "Optimization by Simulated Annealing," *Science*, 220, 671–680.

Lowe, M. J., Mock, B. J., and Sorenson, J. A. (1998), "Functional Connectivity in Single and Multislice Echo-Planar Imaging Using Resting-State Fluctuations," *NeuroImage*, 7, 119–132.

Mitchell, M. (1996), *An Introduction to Genetic Algorithms*, Cambridge, MA: MIT Press.

Roebroeck, A., Formisano, E., and Goebel, R. (2005), "Mapping Directed Influence Over the Brain Using Granger Causality and FMRI," *NeuroImage*, 25, 230–242.

Rogers, B. P., Morgan, V. L., Newton, A. T., and Gore, J. C. (2007), "Assessing Functional Connectivity in the Human Brain by FMRI," *Magnetic Resonance Imaging*, to appear.

Schonberg, T., Pianka, P., Hendler, T., Pasternak, O., and Assaf, Y. (2006), "Characterization of Displaced White Matter by Brain Tumors Using Combined DTI and FMRI," *NeuroImage*, 30, 1100–1111.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Process*, New York: Springer.

Whitcher, B., Tuch, D. S., Wisco, J. J., Sorenson, A. G., and Wang, L. (2007), "Using the Wild Bootstrap to Quantify Uncertainty in DTI," *Human Brain Mapping*, to appear.

Zhu, H. T., Ibrahim, J. G., Tang, N. S., Rowe, D. B., Hao, X. J., Bansal, R., and Peterson, B. (2007), "A Statistical Analysis of Brain Morphology Using Wild Bootstrapping," *IEEE Transactions on Medical Imaging*, 26, 954–966.