The Annals of Applied Statistics
2013, Vol. 7, No. 1, 201–225
DOI: 10.1214/12-AOAS573

LOCAL TESTS FOR IDENTIFYING ANISOTROPIC DIFFUSION AREAS IN HUMAN BRAIN WITH DTI

BY TAO YU[1], CHUNMING ZHANG[2], ANDREW L. ALEXANDER[3] AND RICHARD J. DAVIDSON[3]

National University of Singapore, University of Wisconsin-Madison, University of Wisconsin-Madison and University of Wisconsin-Madison

Diffusion tensor imaging (DTI) plays a key role in analyzing the physical structures of biological tissues, particularly in reconstructing fiber tracts of the human brain in vivo. On the one hand, eigenvalues of diffusion tensors (DTs) estimated from diffusion weighted imaging (DWI) data usually contain systematic bias, which subsequently biases the diffusivity measurements popularly adopted in fiber tracking algorithms. On the other hand, correctly accounting for the spatial information is important in the construction of these diffusivity measurements since the fiber tracts are typically spatially structured. This paper aims to establish test-based approaches to identify anisotropic water diffusion areas in the human brain. These areas in turn indicate the areas passed by fiber tracts. Our proposed test statistic not only takes into account the bias components in eigenvalue estimates, but also incorporates the spatial information of neighboring voxels. Under mild regularity conditions, we demonstrate that the proposed test statistic asymptotically follows a χ^2 distribution under the null hypothesis. Simulation and real DTI data examples are provided to illustrate the efficacy of our proposed methods.

1. Introduction. Diffusion tensor imaging (DTI) has been widely used by neuroscientists to reconstruct the pathways of white matter fibers in human brain in vivo. DTI data are usually estimated from diffusion weighted imaging (DWI) data acquired in magnetic resonance experiments by some statistical model (see Section 2.1). A set of DTI data is typically composed of diffusion tensors (DTs), each contained in a corresponding voxel. Here, a voxel stands for a volume element in a 3D imaging space. Each DT, denoted by $\mathbf{D}$, can be represented by a 3×3 symmetric positive definite matrix, which together with its decomposed eigenvalue–eigenvector pairs $\{(\lambda_{(k)}, \mathbf{v}_{(k)}) : \lambda_{(3)} \geq \lambda_{(2)} \geq \lambda_{(1)}, k = 1, 2, 3\}$ geometrically characterizes the degree and orientation of the water diffusion in that particular voxel. More specifically, the eigenvectors and the square root of eigenvalues

Received September 2011; revised April 2012.

[1]Supported in part by NUS Grant R-155-000-100-133.

[2]Supported in part by NSF Grant DMS-11-06586; Wisconsin Alumni Research Foundation.

[3]Supported in part by National Institute of Mental Health Grants R01-MH43454 and P50-MH069315 to RJD and by Grant P30 HD003352 to the Waisman Center (PI: M. Seltzer).

Key words and phrases. Brain tissue, diffusion tensor, eigenvalue, fiber tracts, local test, quantitative scalar.

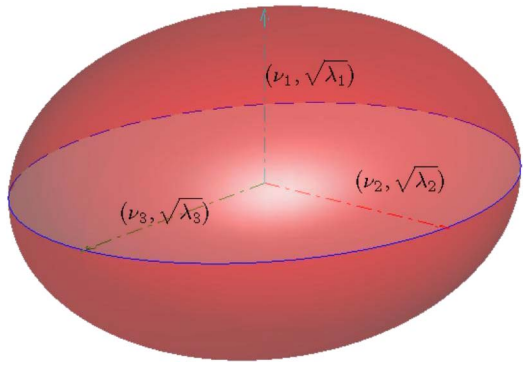

FIG. 1. *Ellipsoid representation of a DT.*

of **D**, respectively, correspond to the orientations and the lengths of axes in an ellipsoid representation (see Figure 1). The distance between the center and any point on the surface of the ellipsoid measures the rate of the water diffusion along that particular orientation in the voxel.

One of the main themes in DTI research is to identify the anisotropic water diffusion brain areas, which facilitates the downstream fiber tracking process. There are two general strategies aiming to address this problem. The first is to construct scalar measurements. The anisotropic water diffusion areas are then identified based on thresholding these scalar measurements. Reviews of this type of method can be found in Moseley et al. (1990), Douek et al. (1991), van Gelderen et al. (1994), Basser and Pierpaoli (1996), Johansen-Berg and Behrens (2009), among others. Thresholding the fractional anisotropy (FA) and the relative anisotropy (RA) [see Section 2.3; Basser and Pierpaoli (1996)] has gained popularity and has been widely adopted by neuroscientists in the past decade. Nonetheless, FA and RA are essentially defined as functions of the eigenvalues of the DT estimate in every brain voxel. These eigenvalues typically carry systematic bias [see Section 3; Pierpaoli and Basser (1996), Zhu et al. (2007), Jones (2003), Lazar and Alexander (2003)], the magnitudes of which are sensitive to the distribution of the noise carried in raw DWI data, and therefore significantly affect the effectiveness and validity of FA (RA) based methods in practice. The second is to classify the morphologies of DTs via test-based approaches [Zhu et al. (2006), Zhu et al. (2007)], which not only quantify the degree of water diffusivity in each voxel, but also lend theoretical supports to statistical inference. However, to the best of our knowledge, all existing approaches on this respect are single-voxel based. The validity and performance of these methods rely essentially on the technical requirement that the number of diffusion gradients in the DWI experiment is large and ultimately diverging to infinity. Moreover, the DTI data are typically spatially structured. Ignoring the spatial information may diminish the effectiveness of the methods.

In this paper, we develop new test-based approaches to identify the anisotropic water diffusion brain areas, which are usually associated with areas passed by fiber tracts. To this end, for each voxel, we examine the testing problem "*all three eigenvalues are equivalent*" against "*at least two eigenvalues are different*." The former corresponds to isotropically diffused DTs, whereas the latter includes anisotropically diffused DTs with morphologies of prolate, oblate and nondegenerate. Our proposed test statistic accommodates the spatial information of the imaging space by taking into account eigenvalues in neighboring voxels. Under mild regularity conditions, we demonstrate that our proposed test statistic asymptotically follows a χ^2 distribution. Therefore, the performance of our methods in the identification of fiber areas is not affected by the bias components carried in the eigenvalue estimates. In theory, one of the main technical requirements is the divergence of the number of neighboring voxels involved in the construction of the statistic. This differs from the divergence of the number of diffusion gradients typically assumed by test-based approaches. Therefore, our methods shed light on alternative ways of improving the identification accuracy of anisotropic water diffusion areas. Furthermore, an adaptive procedure to select varied neighborhoods is proposed to solidify the performance of our proposed approaches when the acquired imaging data have limited resolution. Simulation studies and real data examples are provided to illustrate the efficacy of our proposed methods.

The rest of the paper is organized as follows. Section 2 introduces the background related to our study. Section 3 establishes a statistical model based on eigenvalues in the selected neighborhood of a single voxel. Section 4 describes the procedure of constructing our proposed test statistic, and an adaptive method for selecting neighboring voxels. Section 5 explores the theoretical properties of our proposed test statistic. Section 6 presents simulation results. There, our methods are compared with FA-threshold and Smooth-FA-threshold approaches. Section 7 applies all approaches on real brain DTI data. Section 8 discusses our findings in this paper. Technical conditions and proofs are given in a supplemental document.

2. Background. We begin with a brief introduction of DWI data, DTI data and existing statistical models for estimating DTI data from DWI data. Then, we summarize the associations among fiber tracts, tissue types, water diffusivity and DT types. After that, we overview the quantitative scalars, FA and RA, popularly used in fiber tracking algorithms.

2.1. *From DWI to DTI.* In this section we first give a brief introduction of the structures of DWI and DTI data, where the former are acquired from the diffusion weighted magnetic resonance experiment, while the latter are estimated from the former based on some statistical model. Then, we summarize the existing statistical models for estimating DTI data from DWI data.

We assume that the DWI data over the brain of a given subject contain N voxels, each of which consists of diffusion-weighted measurements. Denote by

ϕ_0, b, $\{(\phi_i, \mathbf{g}_i)\}_{i=1}^r$, the acquired diffusion-weighted measurements at a given voxel over the brain in a DWI experiment. Here, the ith diffusion gradient $\mathbf{g}_i = (g_{i,1}, g_{i,2}, g_{i,3})^T$, with $\mathbf{g}_i^T \mathbf{g}_i = 1$, is chosen by the experimenter before the DWI experiment starts, and serves as a scanning direction in the experiment [Hasan, Parker and Alexander (2001)]; b is the b-factor, whose value is determined by a function of parameter settings in the DWI experiment [Stejskal and Tanner (1965), Moseley et al. (1990), Anderson (2001)]; both $\mathbf{g}_i$ and b usually adopt the same values over all voxels in a DWI experiment; ϕ_i denotes the diffusion attenuated signal, acquired on the ith diffusion gradient $\mathbf{g}_i$ at b; ϕ_0 is the reference signal obtained at $b = 0$; $\{\phi_i\}_{i=1}^r$ and ϕ_0 compose the responses of the DWI experiment for each voxel; r is the number of acquired attenuated signals for each voxel.

Accordingly, a single-voxel of the DTI data contains a 3×3 symmetric, positive definite DT matrix,

$$\mathbf{D} = \begin{bmatrix} D_{1,1} & D_{1,2} & D_{1,3} \\ D_{1,2} & D_{2,2} & D_{2,3} \\ D_{1,3} & D_{2,3} & D_{3,3} \end{bmatrix},$$

which carries the intrinsic information of water diffusion in that particular voxel. The elements of $\mathbf{D}$ can be reorganized as a 6×1 vector $\mathbf{d} = (D_{1,1}, D_{2,2}, D_{3,3}, D_{1,2}, D_{1,3}, D_{2,3})^T$.

The connections between DWI and DTI data are first investigated by the seminal work of Basser, Mattiello and LeBihan (1994), in which the following multivariate linear and nonlinear regression models are proposed. For a single-voxel,

$$\text{(2.1)} \qquad \text{multivariate linear model:} \quad \log(\phi_i) = \log(\phi_0) - b\mathbf{x}_i^T \mathbf{d} + \varepsilon_i,$$

$$\text{(2.2)} \qquad \text{multivariate nonlinear model:} \quad \phi_i = \phi_0 \exp\bigl(-b\mathbf{x}_i^T \mathbf{d}\bigr) + \eta_i,$$

where $i = 1, \ldots, r$, $\mathbf{x}_i = (g_{i,1}^2, g_{i,2}^2, g_{i,3}^2, 2g_{i,1}g_{i,2}, 2g_{i,1}g_{i,3}, 2g_{i,2}g_{i,3})^T$, ε_i and η_i are random errors. $\mathbf{d}$ is then estimated from either model by regression techniques.

A number of alternative statistical models as well as fitting procedures, besides Basser, Mattiello and LeBihan (1994), have been proposed to obtain sophisticated DT estimates from DWI data concerning various aspects, such as robustness, bias, non-Gaussian errors, spatial smoothness, model validity, etc. Examples include Mangin et al. (2002), Chang, Jones and Pierpaoli (2005), Salvador et al. (2005), Heim et al. (2007), Zhu et al. (2007), Tabelow et al. (2008) and many others.

2.2. *Fiber tracts, tissue types, water diffusivity and DT types.* The associations among fiber tracts, tissue types, water diffusivity and DT types are summarized as follows. For voxels located in fiber tracts, that is, white matter brain areas, water tends to present higher diffusivity along the dominant orientation of fibers than that in other orientations. DTs in these voxels are anisotropic, characterized by the heterogeneity in lengths of axes in the corresponding ellipsoid representations.

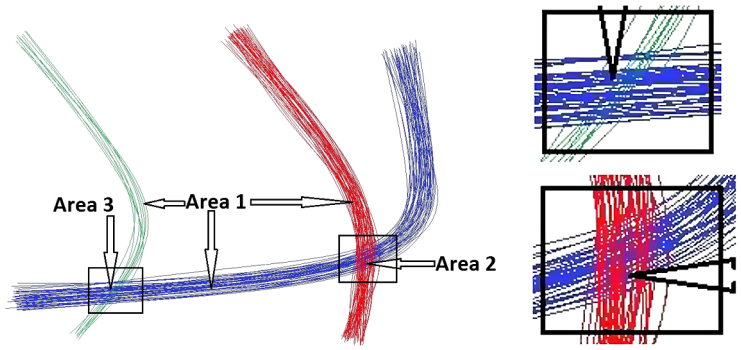

FIG. 2. *An illustrative example of fiber tracts.*

In contrast, for voxels in brain areas without fiber tracts, that is, grey matter areas, DTs are isotropic. In these areas, the diffusivity of water in all orientations is roughly the same. Therefore, the corresponding represented ellipsoids are spherically shaped. The morphology of an anisotropic DT can usually be classified into one of the three categories, namely, prolate, oblate and nondegenerate, respectively, corresponding to brain voxels located in uniquely orientated fiber tracts (*Area* 1 in Figure 2), crossed fiber tracts with similar intensities on two or more different orientations (*Area* 2 in Figure 2, characterized by similar fiber denseness for the red and blue bundles in their intersected parts) and crossed fiber tracts with distinct intensities on different orientations (*Area* 3 in Figure 2, characterized by the scenario that the denseness of blue tracts is higher than that of green tracts in their intersected parts).

A vast number of tractography algorithms have been proposed in order to reconstruct fiber tracts in the human brain based on DTI data. A list of examples of these algorithms can be found in Conturo et al. (1999), Gössl et al. (2002), Xu et al. (2002), Behrens et al. (2007), O'Donnell and Westin (2007) and many others. We observe that most of these approaches are founded on the derived voxel-wise scalar quantities, such as FA and RA, which summarize/extract microstructural information of water diffusion carried by DWI or DTI data.

2.3. *Quantitative scalars*: *FA and RA*. The fractional anisotropy (FA) and relative anisotropy (RA) [Basser and Pierpaoli (1996)] are quantitative scalars widely used by neuroscientists to measure the water diffusivity in brain tissues and construct algorithms for tracking fibers, since they are computationally simple and invariant in the choice of the laboratory coordinate system and diffusion gradients. For a given voxel, FA and RA are defined as

$$\mathrm{FA} = \sqrt{\frac{3}{2}\frac{\sum_{k=1}^{3}\{\lambda_{(k)} - \overline{\lambda}_{(\cdot)}\}^2}{\sum_{k=1}^{3}\lambda_{(k)}^2}}, \qquad \mathrm{RA} = \frac{3}{\sqrt{2}}\frac{\sqrt{\sum_{k=1}^{3}\{\lambda_{(k)} - \overline{\lambda}_{(\cdot)}\}^2}}{\sum_{k=1}^{3}\lambda_{(k)}},$$

where $\overline{\lambda}_{(\cdot)} = \{\lambda_{(1)} + \lambda_{(2)} + \lambda_{(3)}\}/3$, with $\lambda_{(3)} \geq \lambda_{(2)} \geq \lambda_{(1)}$ being the ordered eigenvalues of the DT estimate.

Under the ideal but unrealistic assumption that the estimated DT is noise free, RA $= 0$ and FA $= 0$ in isotropic voxels [i.e., $\lambda_{(1)} = \lambda_{(2)} = \lambda_{(3)}$], whereas RA $= \sqrt{3}$ and FA $= 1$ in purely anisotropic voxels [i.e., $\lambda_{(3)} \gg \lambda_{(2)} = \lambda_{(1)}$]. However, the noise carried by the DWI data contaminates the DT estimates, and subsequently introduces systematic bias into the derived eigenvalues [Pierpaoli and Basser (1996)]. Although these bias components have been investigated by numerical evaluations [Jones (2003), Lazar and Alexander (2003)] as well as in theory [Zhu et al. (2007)], the magnitudes are sensitive to the distribution of the noise in DWI experiments. Consequently, these bias components introduce uncertainty into the constructed FA and RA. In practice, brain areas with small but nonvanishing FA (RA) usually correspond to grey matter areas, whereas those with large FA (RA) are typically areas passed by fiber tracts. Some tractography algorithms are based on FA or RA with thresholds (e.g., the tractography algorithm integrated in `MedINRIA`, a publicly available software at http://www-sop.inria.fr/asclepios/software/MedINRIA). The thresholds, however, are usually manually chosen by investigators based on their historical knowledge of the DTI data and the structure of the human brain.

Figure 3 illustrates how the fiber tracking results are affected by distinct experiential thresholds of the FA, where the detailed data information is given in Section 7. We use a region of interest (ROI) with 30×30 voxels, which is highlighted by a red rectangle shown in the leftmost panel of Figure 3. The tracked fibers passing through the ROI using distinct FA thresholding criteria, namely, FA $> c$ with $c = 0.3, 0.45$ and 0.6, are displayed in the remaining three panels of Figure 3. All sub-figures in Figure 3 are constructed by `MedINRIA`, where different colors stand for different principal orientations of the corresponding DTs. We observe that the fiber tracking results are sensitive to the FA threshold, the choice of which, owing to the uncertainty of the bias magnitudes in eigenvalue estimates,

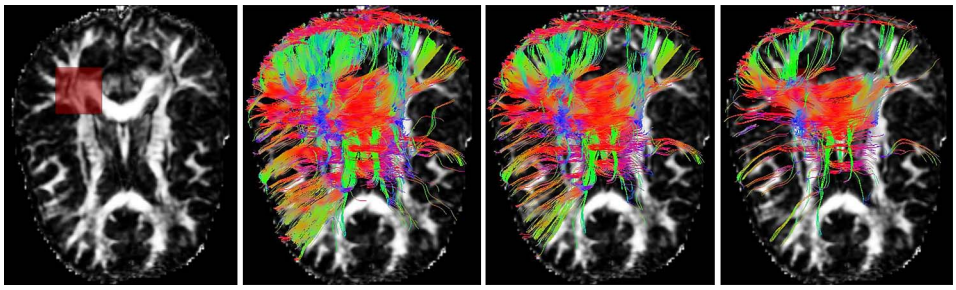

FIG. 3. *Fiber tracking results by distinct FA thresholds. From left to right*: *FA map with a ROI highlighted*; *constructed fibers passing through the ROI by FA thresholds* 0.3, 0.45 *and* 0.6, *respectively.*

is not well supported in theory. In other words, there is lack of a criterion to choose the threshold for FA, and justify the goodness of the tracking results.

A more sophisticated approach based on FA is due to Zhu et al. (2006), in which the asymptotic distribution of a test statistic established on FA is investigated. Therefore, it is capable of suggesting a threshold to be used. This approach, however, is not applicable in our study, since the theory needs the assumption that the number of gradients for every brain voxel is diverging to infinity, that is, $r \to \infty$. In contrast, in our study $r = 12$.

Another limitation of FA and RA with threshold approaches is that the DTI data are typically spatially structured. FA and RA, however, are defined as functions of the eigenvalues of the DT in every single voxel. These approaches identify the existence of fibers in each voxel while ignoring the spatial information, and therefore may diminish the effectiveness in the downstream fiber tracking algorithms.

We observe that there exist approaches which accommodate the spatial information in the stage of DT estimation, such as Heim et al. (2007) and Tabelow et al. (2008). Intuitively, by incorporating the spatial information, these methods improve the DT estimation, and therefore the accuracy of FA and RA. Hereafter, we refer to the FA based on DT estimated by the approach in Tabelow et al. (2008) and Polzehl and Tabelow (2009) (implement in `R` package `dti`) as *Smooth-FA*. In our numerical studies, Smooth-FA with threshold, namely, *Smooth-FA-threshold*, is employed as an approach to identify anisotropic brain voxels, and is compared with our proposed methods.

3. Statistical model on local eigenvalues. We consider the DTI data of a given subject with N voxels. In each voxel $v \in \{1, \ldots, N\}$, denote by $\mathbf{D}(v)$ (without confusion, we concisely denote $\mathbf{D}$) the DT estimated from a statistical model in Section 2.1. Let $\lambda_{(3)} \geq \lambda_{(2)} \geq \lambda_{(1)}$ be the ordered eigenvalues of $\mathbf{D}$. In practice, these eigenvalues are adopted to estimate the ordered true eigenvalues, denoted by $\lambda^*_{(3)} \geq \lambda^*_{(2)} \geq \lambda^*_{(1)}$.

It has been demonstrated both in numerical studies and in theory that $E\{\lambda_{(3)}\} > \lambda^*_{(3)}$ and $E\{\lambda_{(1)}\} < \lambda^*_{(1)}$. Such a bias is actually caused by the sorting procedure in the decomposition of $\mathbf{D}$. In other words, we can postulate that one of the eigenvalues of $\mathbf{D}$, that is, $\lambda_k \in \{\lambda_{(k)} : k = 1, 2, 3\}$, is associated with $\lambda^*_{(k)}$ (but we don't know which one it is). Here, "associated" means there exists a random error ε_k with $E(\varepsilon_k) = 0$ such that

$$\lambda_k = \lambda^*_{(k)} + \varepsilon_k. \tag{3.1}$$

However, $\lambda^*_{(k)}$ is estimated by $\lambda_{(k)}$ instead of λ_k. Therefore, $E\{\lambda_{(3)}\} > E(\lambda_3) = \lambda^*_{(3)}$. Similar arguments lead to $E\{\lambda_{(1)}\} < \lambda^*_{(1)}$. The magnitudes of the bias components are sensitive to the distribution of the noise in DWI experiments, introducing uncertainty into the quantitative scalars, such as FA and RA.

In our approach, instead of using the eigenvalues at a single voxel v only, we choose a set of n neighboring voxels located adjacent to v. Denote by

$\{\lambda_{j,(k)}(v) : k = 1, 2, 3\}_{j=1}^{n}$ [i.e., $\lambda_{j,(k)}$] the eigenvalue estimates in the selected neighboring voxels, whose permuted version associated with the set of true eigenvalues is denoted by $\{\lambda_{j,k}(v) : k = 1, 2, 3\}_{j=1}^{n}$ (i.e., $\lambda_{j,k}$). Here, $j = 1$ corresponds to voxel v, that is, $\lambda_{1,k} = \lambda_k$.

Similar in spirit to the random complete block design, we model $E(\lambda_{j,k})$, the expected eigenvalue of the jth neighboring voxel of v, as the addition of two components, namely, the true eigenvalue $\lambda^*_{(k)}$ for voxel v and the difference of eigenvalues between voxel v and its jth neighboring voxel, denoted by β_j. That is, $E(\lambda_{j,k}) = \lambda^*_{(k)} + \beta_j$, which leads to the additive model,

$$\lambda_{j,k} = \lambda^*_{(k)} + \beta_j + \varepsilon_{j,k}, \tag{3.2}$$

where the eigenvalues $\{\lambda^*_{(k)} : k = 1, 2, 3\}$ in voxel v are of primary interest and therefore treated as the treatment effects, whereas the differences of eigenvalues between voxel v and its neighboring voxels $\{\beta_j\}_{j=1}^{n}$ are the blocking effects and serve as nuisance parameters; $\{\varepsilon_{j,k} : k = 1, 2, 3\}_{j=1}^{n}$ are random errors. Clearly, $\beta_1 = 0$, such that (3.2) complies with (3.1) when $j = 1$.

Nonetheless, as discussed above, based on the DT estimates, we can only obtain $\{\lambda_{j,(k)}, k = 1, 2, 3\}$, the ordered version of $\{\lambda_{j,k}, k = 1, 2, 3\}$. In other words, the values of $\lambda_{j,k}$ are not available in practice. Therefore, the well-developed techniques for the additive model in a complete block design are not directly applicable in analyzing model (3.2). We borrow its basic idea and integrate the corresponding contrast test statistic as one of the main parts in our proposed test statistic. The technical properties of our proposed test statistic in isotropically diffused DT voxels, that is, $\lambda^*_{(3)} = \lambda^*_{(2)} = \lambda^*_{(1)}$, then rely on the fact that $\lambda_{j,(k)} = \lambda^*_{(1)} + \beta_j + \varepsilon_{j,(k)}$.

4. Local test based on neighboring eigenvalues.

4.1. *Test of hypotheses.* Classifying the tissue types for a particular brain area plays a key role in the downstream fiber tracking algorithms. In voxels without fibers, water tends to diffuse in an isotropic manner, characterized by the eigenvalue property $\lambda^*_{(3)} \approx \lambda^*_{(2)} \approx \lambda^*_{(1)}$. In contrast, water in voxels passed by fiber tracts usually presents anisotropic diffusion. DTs in these voxels typically have three possible morphologies, namely, prolate $\lambda^*_{(3)} > \lambda^*_{(2)} \approx \lambda^*_{(1)}$, oblate $\lambda^*_{(3)} \approx \lambda^*_{(2)} > \lambda^*_{(1)}$ and nondegenerate $\lambda^*_{(3)} > \lambda^*_{(2)} > \lambda^*_{(1)}$, respectively, corresponding to brain voxels located in uniquely oriented fiber areas, crossed fiber areas with roughly the same fiber intensities and crossed fiber areas with distinct fiber intensities. In summary, anisotropy is characterized by either $\lambda^*_{(3)} > \lambda^*_{(2)}$ or $\lambda^*_{(2)} > \lambda^*_{(1)}$. Therefore, we consider the hypothesis testing problem,

$$H_0 : \lambda^*_{(3)} = \lambda^*_{(2)} = \lambda^*_{(1)} \quad \text{vs.} \quad H_1 : \lambda^*_{(3)} > \lambda^*_{(2)} \quad \text{or} \quad \lambda^*_{(2)} > \lambda^*_{(1)} \tag{4.1}$$

for each voxel v. This testing problem can be reformulated as a form of contrast test,

$$H_0 : A\boldsymbol{\lambda}^* = \mathbf{0} \quad \text{vs.} \quad H_1 : A(l, \cdot)\boldsymbol{\lambda}^* > 0 \qquad \text{for } l = 1, \text{or}, \ldots, \text{or } \mu, \tag{4.2}$$

where $\boldsymbol{\lambda}^* = (\lambda^*_{(3)}, \lambda^*_{(2)}, \lambda^*_{(1)})^T$, A is a full row rank matrix with $\operatorname{rank}(A) = \mu$ and $\sum_{l_2=1}^{3} A(l_1, l_2) = 0$ for $l_1 = 1, \ldots, \mu$. There exists more than one possible choice of A, such that hypothesis testing problems (4.1) and (4.2) are equivalent. For example,

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \tag{4.3}$$

which is adopted in our numerical studies. Clearly, $\mu = 2$.

To test (4.2), for each voxel v, we propose the test statistic $\mathbb{K}$, represented by

$$\mathbb{K} = n(\mathbf{U}_n - \widehat{\boldsymbol{\theta}}_{\mathbf{U}_n})^T \widehat{\Sigma}^{-1} (\mathbf{U}_n - \widehat{\boldsymbol{\theta}}_{\mathbf{U}_n}), \tag{4.4}$$

where $\mathbf{U}_n = (A\overline{\boldsymbol{\lambda}}_\cdot / \sqrt{\mathrm{MSE}}) \cdot \sqrt{\overline{S^2_\cdot}}$; $\overline{\boldsymbol{\lambda}}_\cdot = (\overline{\lambda}_{\cdot,(3)}, \overline{\lambda}_{\cdot,(2)}, \overline{\lambda}_{\cdot,(1)})^T$; $\overline{\lambda}_{\cdot,(k)} = \sum_{j=1}^{n} \lambda_{j,(k)}/n$, $k = 1, 2, 3$; $\overline{S^2_\cdot}$ is the mean of S^2_j over the neighboring voxels of v; $S^2_j = \sum_{k=1}^{3} \{\lambda_{j,(k)} - \overline{\lambda}_{j,(\cdot)}\}^2/2$, $j = 1, \ldots, n$; $\mathrm{MSE} = \sum_{j=1}^{n} \sum_{k=1}^{3} \{\lambda_{j,(k)} - \overline{\lambda}_{\cdot,(k)} - \overline{\lambda}_{j,(\cdot)} + \overline{\lambda}_{\cdot,(\cdot)}\}^2/\{2(n-1)\}$; let $\widehat{\mathcal{V}}_0$ be an estimate of $\mathcal{V}_0$, the set of isotropic voxels over the entire brain; $\widehat{\boldsymbol{\theta}}_{\mathbf{U}_n}$ denotes the sample median of $\mathbf{U}_n$ over $\widehat{\mathcal{V}}_0$; $\widehat{\Sigma}$ denotes the sample covariance of $\sqrt{n}\mathbf{U}_n$ over $\widehat{\mathcal{V}}_0$. $\widehat{\mathcal{V}}_0$ is obtained from the following iteration steps:

(i) Evaluate $\mathbf{U}_n$'s over all voxels of interest. For some pre-given significant level α, let $\chi^2_{\mu;1-\alpha}$ be the $(1-\alpha)$th quantile of the chi-square distribution with μ degrees of freedom.

(ii) Let $\widehat{\mathcal{V}}_{0,0}$ include all voxels of interest.

(iii) In the sth iteration ($s = 1, 2, \ldots$), let $\widetilde{\mathbb{K}}$ denote the statistic computed from (4.4) based on $\widehat{\mathcal{V}}_{0,s-1}$, and let $\mathbb{K} = c\widetilde{\mathbb{K}}$.

(iv) Include voxel v in $\widehat{\mathcal{V}}_{0,s}$ if the corresponding statistic $\mathbb{K} < \chi^2_{\mu;1-\alpha}$.

(v) Repeat steps (iii) and (iv) until $\widehat{\boldsymbol{\theta}}_{\mathbf{U}_n}$ converges.

In step (iii) above, the constant c is to correct the bias in the evaluation of $\mathbb{K}$. This is because in the sth iteration, $\widehat{\boldsymbol{\theta}}_{\mathbf{U}_n}$ in $\mathbb{K}$ is constructed by only using voxels in $\widehat{\mathcal{V}}_{0,s-1}$, which includes voxels with $\mathbb{K}$ capped by $\chi^2_{\mu,1-\alpha}$ in the $(s-1)$th iteration. The value of c is derived as follows:

$$\begin{aligned} E\{\widetilde{\mathbb{K}}\mathbf{I}(c\widetilde{\mathbb{K}} < \chi^2_{\mu;1-\alpha})\} &\approx \frac{1}{|\widehat{\mathcal{V}}_{0,s-1}|} \sum_{v: c\widetilde{\mathbb{K}} < \chi^2_{\mu;1-\alpha}} n(\mathbf{U}_n - \widehat{\boldsymbol{\theta}}_{\mathbf{U}_n})^T \widehat{\Sigma}^{-1} (\mathbf{U}_n - \widehat{\boldsymbol{\theta}}_{\mathbf{U}_n}) \\ &\approx \frac{1}{|\widehat{\mathcal{V}}_{0,s-1}|} \sum_{v \in \widehat{\mathcal{V}}_{0,s-1}} \{\sqrt{n}(\mathbf{U}_n - \overline{\mathbf{U}}_\cdot)\}^T \widehat{\Sigma}^{-1} \{\sqrt{n}(\mathbf{U}_n - \overline{\mathbf{U}}_\cdot)\} \\ &= \frac{|\widehat{\mathcal{V}}_{0,s-1}| - 1}{|\widehat{\mathcal{V}}_{0,s-1}|} \mu \approx \mu, \end{aligned}$$

where "=" is followed by the fact that in step (iii), $\widehat{\Sigma}$ is the sample covariance of $\sqrt{n}\mathbf{U}_n$ over $\widehat{\mathcal{V}}_{0,s-1}$; $\mathrm{I}(\cdot)$ is the indicator function. Therefore,

$$E\{\mathbb{K}\mathrm{I}(\mathbb{K} < \chi^2_{\mu;1-\alpha})\} = E\{c\widetilde{\mathbb{K}}\mathrm{I}(c\widetilde{\mathbb{K}} < \chi^2_{\mu;1-\alpha})\} \approx c\mu. \tag{4.5}$$

On the other hand, according to Theorem 1 in Section 5, $\mathbb{K}$ is approximately χ^2_μ distributed,

$$E\{\mathbb{K}\mathrm{I}(\mathbb{K} < \chi^2_{\mu;1-\alpha})\} \approx \int_0^{\chi^2_{\mu;1-\alpha}} \frac{t^{\mu/2}}{2^{\mu/2}\Gamma(\mu/2)} e^{-t/2}\,dt. \tag{4.6}$$

Combining (4.5) and (4.6), we set $c = \frac{1}{\mu}\int_0^{\chi^2_{\mu;1-\alpha}} \frac{t^{\mu/2}}{2^{\mu/2}\Gamma(\mu/2)} e^{-t/2}\,dt$.

REMARK 1. We now make some remarks concerning the construction of the statistic in (4.4). The main term $\mathbf{U}_n$ in $\mathbb{K}$ consists of two parts:

- The first part $A\overline{\boldsymbol{\lambda}}_\cdot/\sqrt{\mathrm{MSE}}$ mimics the t statistic of the contrast test for the additive model in a complete block design. Referring to the proofs of Theorems 1 and 2, under certain regularity conditions, it approaches infinity with rate $\sqrt{n}$, when there are significant differences among $\{\lambda^*_{(k)} : k = 1, 2, 3\}$, whereas it converges to a fixed constant when $\lambda^*_{(3)} = \lambda^*_{(2)} = \lambda^*_{(1)}$. Therefore, it has good statistical power in identifying the differences among $\{\lambda^*_{(k)} : k = 1, 2, 3\}$, when model (3.2) is valid.
- The second part $\sqrt{\overline{S^2_\cdot}}$ is added to increase the power of the test statistic $\mathbb{K}$ on the boundary of fibers. For any voxel on the boundary of fibers, in the sense that its selected neighborhood contains voxels belonging to both fiber and nonfiber areas, the assumption that the collected eigenvalues in the selected neighboring voxels follow model (3.2) may not be appropriate. In this case, $\sqrt{\mathrm{MSE}}$ tends to inflate and, consequently, the statistical power of the first part $A\overline{\boldsymbol{\lambda}}_\cdot/\sqrt{\mathrm{MSE}}$ is limited. The second part then counteracts the effect of $\sqrt{\mathrm{MSE}}$, and therefore is particularly useful when the resolution of the DTI data is limited.
- Furthermore, $\widehat{\boldsymbol{\theta}}_{\mathbf{U}_n}$, the sample median instead of sample mean of $\mathbf{U}_n$, is adopted just to ensure the robustness of the approach.

REMARK 2. In this paper we establish testing procedures for identifying the brain areas with fiber tracts. However, for voxels located in fiber tracts, their DTs have three possible morphologies, namely, prolate, oblate, and nondegenerate, respectively, corresponding to eigenvalue properties $\lambda^*_{(3)} > \lambda^*_{(2)} \approx \lambda^*_{(1)}$, $\lambda^*_{(3)} \approx \lambda^*_{(2)} > \lambda^*_{(1)}$, and $\lambda^*_{(3)} > \lambda^*_{(2)} > \lambda^*_{(1)}$. One may also implement similar testing procedures as those in this paper to further classify these three possibilities. We leave the details out for presentational brevity.

4.2. *Adaptive selection of neighborhood.* Following Sections 4.1 and 5, the asymptotic theories for $\mathbb{K}$ need that the neighborhood size n is large and that model (3.2) is satisfied. From the experimental point of view, as long as the resolution in a DWI experiment is sufficiently good, such that the proportion of voxels located on the boundary of fibers over the entire brain shrinks, we can simply employ a fix-shaped neighborhood in the construction of $\mathbb{K}$ (e.g., choose the neighborhood as a fixed cube). However, for experiments with limited resolution, a fix-shaped neighborhood may not be a good choice, because in this case, the assumption that the eigenvalues in the neighboring voxels follow model (3.2) may not be well satisfied, in order to ensure n large required by Theorem 1. Such a problem is particularly severe for voxels located on the boundary of fibers. Therefore, development of a varied neighborhood is necessary.

We propose an adaptive method to select the neighboring voxels based on the philosophy below. First, the adjacent voxels of v should have a better chance to be selected as neighboring voxels than those far away. Second, to ensure the validity of model (3.2), if the tensor in v is isotropic, the selected neighborhood should mainly consist of voxels with isotropic tensors. Likewise, if the tensor in v is anisotropic, the selected neighborhood should be in favor of voxels with anisotropic tensors. Therefore, we incorporate the physical distances and similarity measures of DTs between voxel v and its nearby neighbors to establish the criteria for selecting neighboring voxels.

For each voxel v and fixed number n of neighboring voxels, we summarize our proposed adaptive neighborhood selection approach as follows:

(1) Fix a cube-shaped domain centered at v with reasonably large size $x \times y \times z$, whose voxels are candidates. Here x, y, z are integers and $xyz \geq n$.

(2) Define the similarity score function f between voxel v and its neighboring candidate v_l, $l = 1, \ldots, xyz$, as

$$f(v, v_l) = d_{\mathbf{D}}\big(\mathbf{D}(v), \mathbf{D}(v_l)\big) \exp\big\{C \cdot d_p(v, v_l)\big\},$$

where $d_p(v, v_l)$ denotes the physical distance between voxels v and v_l; $d_{\mathbf{D}}(\mathbf{D}(v), \mathbf{D}(v_l)) = \sqrt{\operatorname{trace}[\{\mathbf{D}(v) - \mathbf{D}(v_l)\}^2]}$ is a measurement of the diffusion similarity between tensors in voxels v and v_l [Alexander, Gee and Bajcsy (1999)]; $C \geq 0$ is added to balance the contribution of $d_p(v, v_l)$ and $d_{\mathbf{D}}(\mathbf{D}(v), \mathbf{D}(v_l))$.

(3) Select n voxels with the lowest f values as the neighboring voxels.

We observe that C in f is adopted to balance the contribution of $d_{\mathbf{D}}(\mathbf{D}(v), \mathbf{D}(v_l))$ and $d_p(v, v_l)$. When the resolution of the DTI data is high, $d_p(v, v_l)$ is close to 0. Consequently, for any fixed C, $\exp\{C \cdot d_p(v, v_l)\}$ approaches the constant 1. Therefore, for high resolution DTI data, the proposed approach is not sensitive to the choice of C. Throughout our numerical studies, we fix $C = 0.1$.

We would like to point out that the proposed approach above is similar in spirit to the adaptive approaches in Tabelow et al. (2008) and Li et al. (2011), where iterative testing procedures are used to adaptively control the contribution of neighboring voxels in their proposed algorithms. Compared with their approaches, our

approach is computationally more economic. The effectiveness of our proposed approach above has been demonstrated in Yu (2009) by simulation studies.

We summarize our proposed procedure of constructing $\mathbb{K}$ in the supplemental document [Yu et al. (2013)].

5. Theoretical properties. In this section we explore the theoretical properties of our proposed test statistic $\mathbb{K}$. The technical details are given in the supplemental document [Yu et al. (2013)]. Theorem 1 below establishes the asymptotic null distribution of $\mathbb{K}$, when the number n of neighboring voxels is large.

THEOREM 1. *Assume model* (3.2) *and Condition* A *in the supplemental document. Then for* $\mathbb{K}$ *defined in* (4.4), *under the null hypothesis in* (4.2), *as* $n \to \infty$,

$$\mathbb{K} \xrightarrow{\mathcal{L}} \chi^2_\mu.$$

We would like to point out that the construction of $\mathbb{K}$ and the theoretical derivations of Theorem 1 are nontrivial and challenging. Following the discussion in Section 3, for each voxel v, we postulate that there exist unobservable one-to-one correspondences, that is, $\{(\lambda_k, \lambda^*_{(k)}) : k = 1, 2, 3\}$, between the estimated and true eigenvalues. The ordered eigenvalue estimates $\lambda_{(3)} \geq \lambda_{(2)} \geq \lambda_{(1)}$, however, are neither unbiased estimates for $\lambda^*_{(3)} \geq \lambda^*_{(2)} \geq \lambda^*_{(1)}$, nor independent. We address these bias components in the construction of $\mathbb{K}$ and the corresponding proof of Theorem 1 based on the intuition as follows. Referring to model (3.2), for an isotropic voxel v, the collected neighboring voxels can be modeled as $\lambda_{j,(k)} = \lambda^*_{(1)} + \beta_j + \varepsilon_{j,(k)}$. As such, the bias components of eigenvalue estimates in the neighboring voxels of v are carried by $\varepsilon_{j,(k)}$, whose effects in our test statistic are counteracted by $\widehat{\boldsymbol{\theta}}_{\mathbf{U}_n}$ constructed based on spatial information of the entire brain.

To appreciate the discriminating power of $\mathbb{K}$ in the identification of anisotropic brain areas, the asymptotic power of $\mathbb{K}$ is established in Theorem 2 below.

THEOREM 2. *Assume model* (3.2) *and Condition* A *in the supplemental document. Then for voxel* v, *under the fixed alternative* H_1 *in* (4.2), *as* $n \to \infty$,

$$n^{-1}\mathbb{K} \xrightarrow{\mathcal{P}} M,$$

where M *is given by* (A.4.3) *in the supplemental document.*

Theorem 2 shows that as long as $\mathbf{g}(\widetilde{\mathbf{a}}(v)) \neq \mathbf{g}(\mathbf{b})$, $M > 0$ and $\mathbb{K} \xrightarrow{\mathcal{P}} +\infty$ at rate n, under the fixed alternative H_1. Here, $\mathbf{g}(\cdot)$ is defined by (A.3.2) in the supplemental document; $\widetilde{\mathbf{a}}(v) = (E\{2S_1^2(v)\}, E\{\lambda_{1,(3)}(v)\}, E\{\lambda_{1,(2)}(v)\}, E\{\lambda_{1,(1)}(v)\})^T$; $\mathbf{b} = (E(2S_\varepsilon^2), E\{\varepsilon_{1,(3)}(1)\}, E\{\varepsilon_{1,(2)}(1)\}, E\{\varepsilon_{1,(1)}(1)\})^T$; $S_1^2(v)$ and S_ε^2 are, respectively, the sample variances of $\{\lambda_{1,(k)}(v), k = 1, 2, 3\}$ and $\{\varepsilon_{1,(k)}(1), k = 1, 2, 3\}$.

Thus, under the fixed alternative, the power of our proposed test statistic $\mathbb{K}$ tends to 1 except in rare situations. Corollary 1 below gives one specific example of $M > 0$. The proof is straightforward and omitted.

COROLLARY 1. *Assume conditions in Theorem 2. Suppose that* $\varepsilon_{j,k}$ *has a symmetric distribution about* 0, *that is,* $\varepsilon_{j,k}$ *has the same distribution as* $-\varepsilon_{j,k}$, *and* $E\{\lambda_{1,(3)}\} - E\{\lambda_{1,(2)}\} \neq E\{\lambda_{1,(2)}\} - E\{\lambda_{1,(1)}\}$. *Then* $M > 0$.

6. Simulation study.

6.1. *Basic settings for numerical work.* Since in a real DTI data set, the number of voxels, each of which corresponds to a hypothesis test, is typically large, false discovery rate (FDR) techniques [Benjamini and Hochberg (1995), Storey (2002), Storey, Taylor and Siegmund (2004), Zhang, Fan and Yu (2011)] are incorporated in our numerical works to control the error rates. Two FDR procedures are employed in our study, namely, the conventional FDR procedure by Storey (2002) and the FDR_L procedure by Zhang, Fan and Yu (2011), which is capable of capturing the spatial information in imaging data. A short summary of the FDR_L procedure is provided in the supplemental document [Yu et al. (2013)].

Some settings of parameters throughout our numerical works are given as follows. For FDR and FDR_L procedures, set false discovery control level as 0.01 and tuning parameter $\lambda = 0.2$. The neighborhood for the FDR_L procedure is set as the nearest 7 voxels shown in the left panel of Figure 4. When evaluating $\mathbb{K}$ for each voxel, A is given by (4.3). Apply the adaptive neighborhood selection approach proposed in Section 4.2 to select neighboring voxels, where the domain for

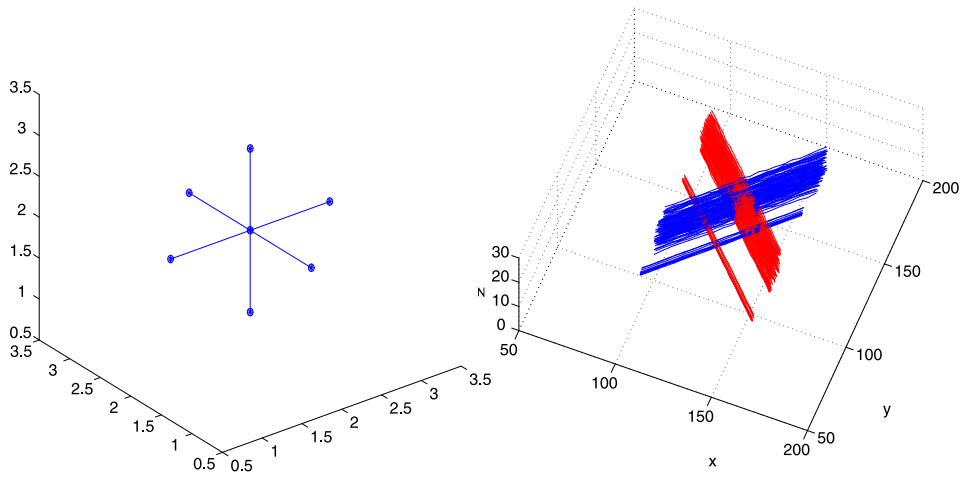

FIG. 4. *Left panel—neighbors of a voxel used in the* FDR_L *procedure. Right panel—gometry of the simulated brain.*

the candidate neighboring voxels is set as a $5 \times 5 \times 3$ cube centered at v. There, $n = 25$ voxels are selected based upon f with $C = 0.1$. The choice of n is referred to in the results in Section 6.3: when $n = 25$, the sampling distribution of $\mathbb{K}$ agrees reasonably well with the χ^2 distribution.

6.2. *Data simulation*. We simulate several sets of DWI data over the entire brain. For each set, a 3D imaging space with the same brain areas as the real DWI data in Section 7 is simulated, where fiber tracts are simulated to have the geometrical structure displayed in the right panel of Figure 4. DTs are simulated according to the locations of voxels in the imaging space. To this end, four distinct sets of eigenvalues, namely, $[0.7, 0.7, 0.7]$, $[1.0, 0.55, 0.55]$, $[0.8, 0.8, 0.5]$ and $[0.9, 0.7, 0.5]$ (units: $10^{-3}\text{mm}^2/\text{s}$), are adopted to, respectively, simulate DTs with morphologies of *isotropic* (nonfiber areas), *prolate* (single blue and red bundles), *oblate* (intersected areas of blue bundles) and *nondegenerate* (intersected areas between blue and red bundles), such that all simulated DTs share the same mean diffusivity $\bar{\lambda}^*_{(\cdot)} = 0.7 \times 10^{-3}\text{mm}^2/\text{s}$, a typical value in real human brains [Pierpaoli et al. (1996), Anderson (2001)].

Since the acquired attenuated signal intensity, $\phi_i(v)$, at each voxel v and gradient $\mathbf{g}_i$ in real DWI data is typically generated by the square-root of the sum of squares of two random numbers in the DWI experiment [Henkelman (1985), Salvador et al. (2005), Zhu et al. (2007)], we simulate a reference signal ($i = 0$) and $r = 12$ diffusion attenuated signals ($i = 1, \ldots, 12$) in each v $[= (v_x, v_y, v_z)]$ as

$$\phi_i(v) = \sqrt{\left[\phi_0^*(v) \exp\{-b\mathbf{g}_i^T \mathbf{D}^*(v)\mathbf{g}_i\} + \varepsilon_{i,x}(v)\right]^2 + \varepsilon_{i,y}^2(v)},$$

where $\phi_0^*(v) = 1200$ when $v_x \in (0, 128]$, $\phi_0^*(v) = 1800$ when $v_x \in (128, 256]$; the b factor $b = 1000$ when $i > 0$, $b = 0$ when $i = 0$; diffusion gradients $\{\mathbf{g}_i : i = 1, \ldots, 12\}$ are adopted from the real DWI data in Section 7; $\mathbf{D}^*(v) = Q(v)\Lambda^*(v)Q^T(v)$; $\Lambda^*(v) = \text{diag}(\lambda^*_{(3)}(v), \lambda^*_{(2)}(v), \lambda^*_{(1)}(v))$; $Q(v)$ is a 3×3 orthogonal matrix whose column vectors are composed of the eigenvectors of the simulated $\mathbf{D}^*(v)$,

$$Q(v) = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{or} \quad Q(v) = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Clearly, the former corresponds to the red and the corresponding parallel narrow blue bundles in the right panel of Figure 4, while the latter models the other two blue bundles. The random errors $\varepsilon_{i,x}(v)$ and $\varepsilon_{i,y}(v)$ are simulated as independent and normally distributed with variance σ^2, which is varied to provide signal to noise ratios (SNRs), where SNR $= \phi_0^*(v)/\sigma$. We examine four distinct SNRs, {5, 10, 15, 20}, each corresponding to one set of the simulated DWI data.

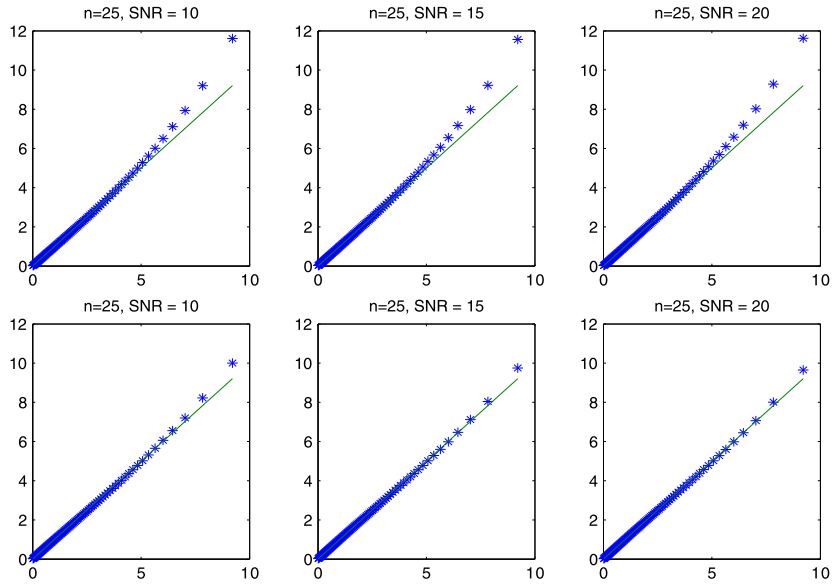

FIG. 5. *Empirical percentiles of* $\mathbb{K}$ (*y-axis*) *versus percentiles of* χ^2_μ *distribution* (*x-axis*). *Top panels—candidate cubic* $5 \times 5 \times 3$, $n = 25$. *Bottom panels—candidate cubic* $11 \times 11 \times 3$, $n = 81$. *From left to right panels*: SNR = 10, 15 *and* 20. *Solid line—the* 45 *degree reference line*.

6.3. *Agreement between* χ^2 *distribution and* $\mathbb{K}$. With the DWI data sets simulated in Section 6.2, we estimate the DT in each voxel by regression model (2.1) and the corresponding eigenvalues by Schur decomposition. For each DWI data set, two sets (I and II) of $\mathbb{K}$ are constructed according to different settings of the adaptive neighborhood selection approach in Section 4.2. In particular, for *Set* I, the size of the candidate neighborhood and the number of selected neighboring voxels are, respectively, chosen as $5 \times 5 \times 3$ and $n = 25$, while those for *Set* II as $11 \times 11 \times 3$ and $n = 81$. Other settings are given in Section 6.1.

For each simulated data set, we collect all $\mathbb{K}$'s whose corresponding voxels are located inside the simulated nonfiber areas. The QQ plots of the (1st to 99th) percentiles of these $\mathbb{K}$'s against those of the χ^2_μ distribution are displayed in Figure 5, with top panels based on *Set* I, bottom panels on *Set* II. The left, middle and right panels correspond to SNR $= 10$, 15 and 20, respectively. Results in Figure 5 demonstrate that the sampling distributions of $\mathbb{K}$, under both $n = 25$ and $n = 81$, agree reasonably well with the χ^2 distribution.

6.4. *Receiver operating characteristic curve*. We compare our methods ($\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$) with FA with threshold (i.e., *FA-threshold*) and Smooth-FA with threshold (i.e. *Smooth-FA-threshold*) approaches by the receiver operating characteristic (ROC) curve, a widely adopted statistical tool for evaluating the accuracy of continuous diagnostic tests [Pepe (2003)].

Let $\{T_1, \ldots, T_N\}$ be the set of statistics for all voxels over the entire brain, where T_v for any voxel $v \in \{1, \ldots, N\}$ is the FA or Smooth-FA value, if the FA-threshold or Smooth-FA-threshold approach is used; is the p-value based on $\mathbb{K}$, if the $\mathbb{K}$-FDR approach is used; is the $\widetilde{p}$-value if the $\mathbb{K}$-FDR$_L$ approach is used, where $\widetilde{p}$ stands for the median smoothed p-value of $\mathbb{K}$ [Zhang, Fan and Yu (2011)]. For any given threshold t, if we classify a voxel v as anisotropic based on $T_v \in R(t)$, where $R(t) = \{T_v : T_v \geq t, v = 1, \ldots, N\}$ when T_v is the FA or Smooth-FA value; $R(t) = \{T_v : T_v \leq t, v = 1, \ldots, N\}$ when T_v is the p- or $\widetilde{p}$-value, then,

$$\text{sensitivity:} \quad \text{se}(t) \equiv \frac{\sum_{v=1}^{N} I\{T_v \in R(t), H_1 \text{ is true}\}}{|\mathcal{V}_0|},$$

$$\text{specificity:} \quad \text{sp}(t) \equiv \frac{\sum_{v=1}^{N} I\{T_v \notin R(t), H_0 \text{ is true}\}}{N - |\mathcal{V}_0|},$$

where $|\mathcal{V}_0|$ is the number of isotropic voxels over the entire brain. The ROC curve is then constructed as the 2D curve $(\text{se}(t), 1 - \text{sp}(t))$ when t ranges between 0 and 1. The area under the ROC curve (AUC) is a popularly adopted measure of the accuracy of the test. More precisely, AUC is ranging between 0 and 1, and the larger the AUC, the better the method.

Following Sections 6.1 and 6.2, the ROC curves for FA, Smooth-FA, p and $\widetilde{p}$ are displayed in Figure 6 for data sets with SNR $= 5, 10, 15$ and 20, respectively. It has been seen from Figure 6 that the ROC curves of $\widetilde{p}$ (dotted blue lines) and p (dashed black lines) are consistently located above those of FA (solid red lines) and Smooth-FA (dash-dotted green lines) for all simulated data sets, indicating that $\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$ are capable of achieving better classification accuracy than FA-threshold and Smooth-FA-threshold approaches. Furthermore, the AUCs of $\widetilde{p}$ in all examined data sets are superior to those of p, suggesting that $\mathbb{K}$-FDR$_L$ performs the best among all four approaches.

6.5. *Test results*. In this section we present our test results in simulated data sets. Settings of our computations are given in Section 6.1. For the sake of clarity, we only present the results of SNR $= 10$. Those of SNR $= 5, 15$ and 20 display similar phenomenon and are omitted.

The FA threshold (>0.3003) and Smooth-FA threshold (>0.2803) are tentatively chosen such that their identified results share the same sensitivity as $\mathbb{K}$-FDR$_L$. The results of sensitivity and specificity by all four methods are displayed in Table 1. Clearly, $\mathbb{K}$-FDR$_L$ maximizes both the sensitivity (0.8845) and the specificity (0.9982) in this example. $\mathbb{K}$-FDR achieves similar specificity (0.9957) but slightly smaller sensitivity (0.7522) compared to $\mathbb{K}$-FDR$_L$. However, in order to yield comparable sensitivity (0.8845) as $\mathbb{K}$-FDR$_L$, FA-threshold and Smooth-FA-threshold approaches produce much lower specificities (0.4012 for FA; 0.6242 for Smooth-FA).

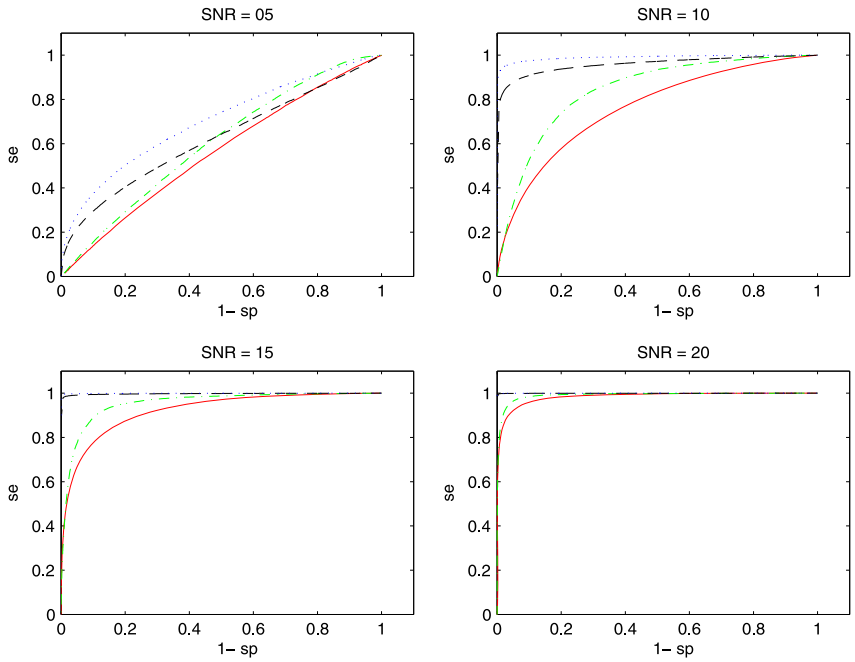

FIG. 6. *ROC curves for SNR* = 5, 10, 15 *and* 20. *Dashed black line—p-value of* $\mathbb{K}$; *dotted blue line—*$\widetilde{p}$*-value of* $\mathbb{K}$; *solid red line—FA*; *dash-dotted green line—Smooth-FA*.

The performances of the four methods are further compared on two selected axial slices. Throughout our simulation and real data examples, we apply the same registration transformations from the brain data to the T1 high-resolution image of the subject's brain. The two slices with the simulated brain anisotropic areas highlighted are given in the leftmost panel of Figure 8. Figure 7 displays the color maps of the FA, Smooth-FA, $-\log(p)$ and $-\log(\widetilde{p})$, with all $-\log(p)$ and $-\log(\widetilde{p})$ values greater than 10 set equal to 10 to improve the visualization. Figure 8 compares the detected anisotropic areas by FA > 0.3003, Smooth-FA > 0.2803, $\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$ for SNR $= 10$.

TABLE 1
Sensitivity and specificity, SNR $= 10$, *the FDR control level is* 0.01

	Sensitivity	**Specificity**
FA > 0.3003	0.8849	0.4012
Smooth-FA > 0.2803	0.8852	0.6242
$\mathbb{K}$-FDR	0.7522	0.9957
$\mathbb{K}$-FDR$_L$	0.8845	0.9982

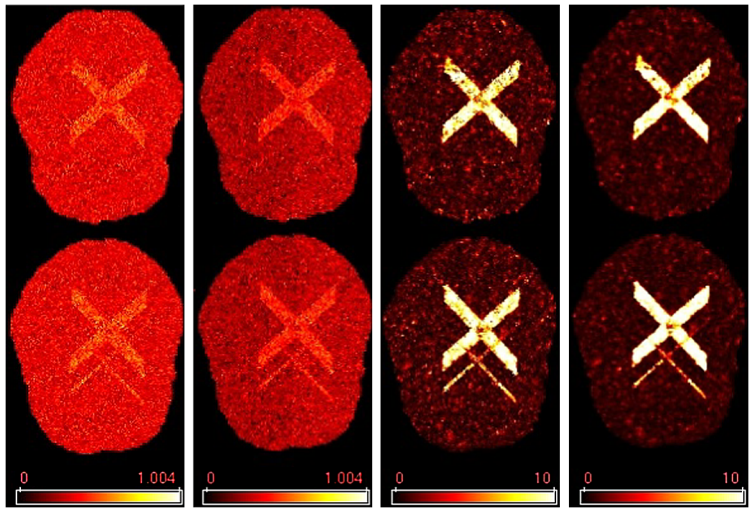

FIG. 7. *Comparison of color maps for simulated data set. From left to right*: *FA*; *Smooth-FA*; $-\log(p)$ *by* $\mathbb{K}$; $-\log(\tilde{p})$ *by* $\mathbb{K}$. SNR = 10.

As clearly evidenced in Figures 7 and 8, $\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$ not only provide detected results with better accuracy than FA-threshold and Smooth-FA-threshold approaches, but also yield better contrasts between the significant areas and the nonsignificant ones. In contrast, the results from FA ($>$0.3003) and Smooth-FA ($>$0.2803) not only fail to detect some truly significant voxels, but also present tiny scattered faulty findings, which expect to contaminate the downstream fiber

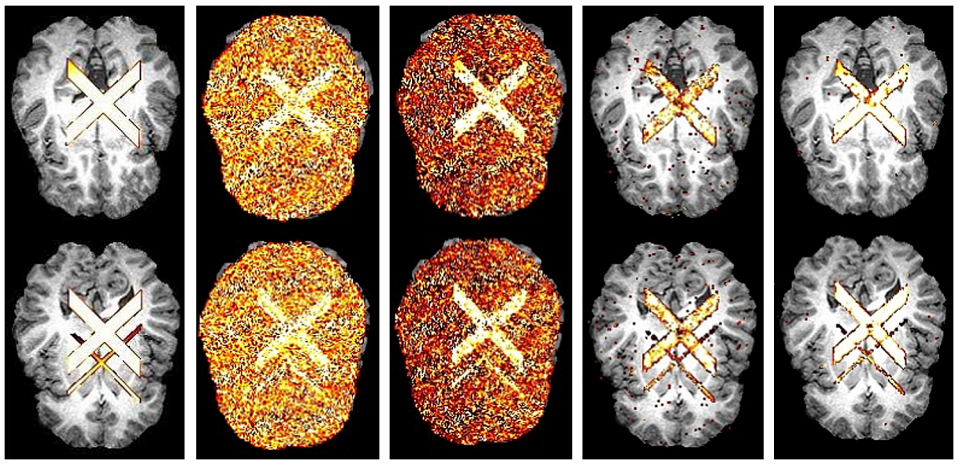

FIG. 8. *Comparison of brain anisotropic areas discovered for the simulated data set. From left to right*: *simulated brain anisotropic areas*; FA $>$ 0.3003; Smooth-FA $>$ 0.2803; $\mathbb{K}$-FDR; $\mathbb{K}$-FDR$_L$. SNR = 10. *The control level is* 0.01.

tracking results. It has been seen in the second to right and rightmost panels of Figures 7 and 8 that the detected results by $\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$ well capture the primary features of the simulated anisotropic areas. Compared with $\mathbb{K}$-FDR, the $\mathbb{K}$-FDR$_L$ approach offers slightly more accurate identifications in both isotropic and anisotropic water diffusion areas.

7. Real data example. We apply our proposed testing procedures on five subjects, whose DWIs were acquired by the magnetic resonance (MR) experiments described below.

The brain magnetic resonance images (MRIs, including DWI and fMRI) of each subject were acquired with a GE SIGNA 3-T scanner equipped with high-speed gradients and a whole-head transmit-receive quadrature birdcage headcoil (GE Medical Systems). The anatomical scan for each subject took approximately 20 minutes [Dalton et al. (2005)]. In the anatomical scanning, the size of each voxel in an xy-plane is 0.9375 mm $\times$ 0.9375 mm, field of view $= 24$ cm^2, matrix $= 256 \times 256$; 30 axial slices are acquired along the z-axis, slice thickness $= 3$ mm. A single reference image at $b = 0$ and 12 diffusion-attenuated images with noncollinear directions of diffusion gradients at $b = 1000$ s/mm^2 were obtained. Since we focus on the analysis of the anatomical structures of the human brain in this paper, the detailed information for the functional scans is omitted.

Using the DWI data of a single subject as the representative, we first present and compare the results by all four methods on two selected axial slices of the brain. The results for the other four subjects display similar scenarios and are omitted.

The acquired data set contains $256 \times 256 \times 30 = 1{,}966{,}080$ voxels with 400,309 voxels located inside the brain. In each voxel, the DT is estimated from regression model (2.1). After that, the corresponding eigenvalues are obtained by Schur decomposition.

All settings are given in Section 6.1. The color maps of FA, Smooth-FA, $-\log(p)$ and $-\log(\widetilde{p})$ are displayed in Figure 9 on two selected axial slices, whereas the corresponding detected anisotropic diffusion brain areas by all four methods are provided in Figure 10. As evidenced in Figure 10, compared with the identified anisotropic areas by $\mathbb{K}$-FDR or $\mathbb{K}$-FDR$_L$, FA-threshold and Smooth-FA-threshold approaches produce more noisy detections. For example, inspection of areas highlighted by red rectangles in the top panels of Figure 10 (enlarged in Figure 11), FA > 0.35 and Smooth-FA > 0.35 detect more scattered tiny areas than $\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$. Those are highly likely to be faulty findings. Furthermore, an overview of the areas located close to the top of the highlighted areas, FA > 0.35 and Smooth-FA > 0.35 present more findings than $\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$. However, those areas are located close to the boundary of the brain, and are most likely to be nonfiber areas. In the meantime, as illustrated in Section 2.3, since the number of gradients r in each voxel is small, FA-threshold and Smooth-FA-threshold approaches cannot clearly infer and control the error rate of the identification, and therefore lack the rigorous criterion in selecting the appropriate

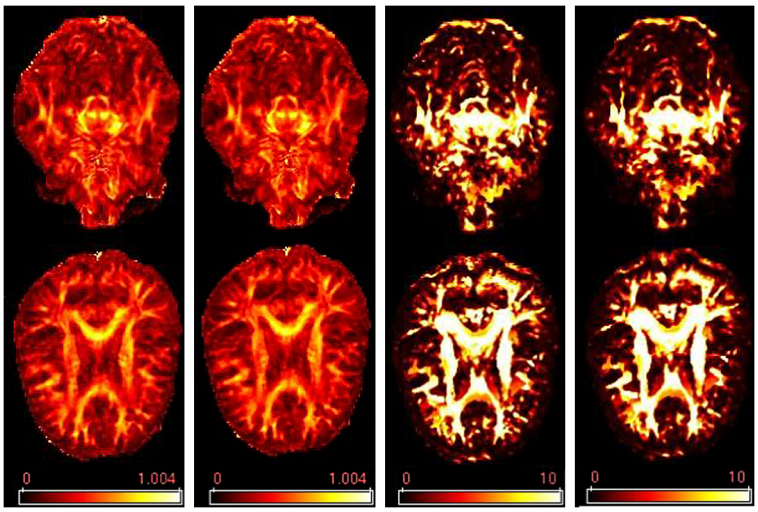

FIG. 9. *Comparison of color maps for the real brain data set. From left to right*: *FA*; *Smooth-FA*; $-\log(p)$ *by* $\mathbb{K}$; $-\log(\widetilde{p})$ *by* $\mathbb{K}$.

thresholds in practice. The superiority of our proposed methods to FA-threshold and Smooth-FA-threshold approaches can be further illustrated by Figure 9, the color maps of FA, Smooth-FA, $-\log(p)$ and $-\log(\widetilde{p})$. The color maps of both $-\log(p)$ and $-\log(\widetilde{p})$ show better contrasts between the anisotropic and isotropic areas than those of FA and Smooth-FA, indicating that $\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$ more effectively separate anisotropic diffusion areas from isotropic ones.

To further illustrate the efficacy of our methods in reducing the scattered faulty findings, we summarize the frequency of identified isolated anisotropic

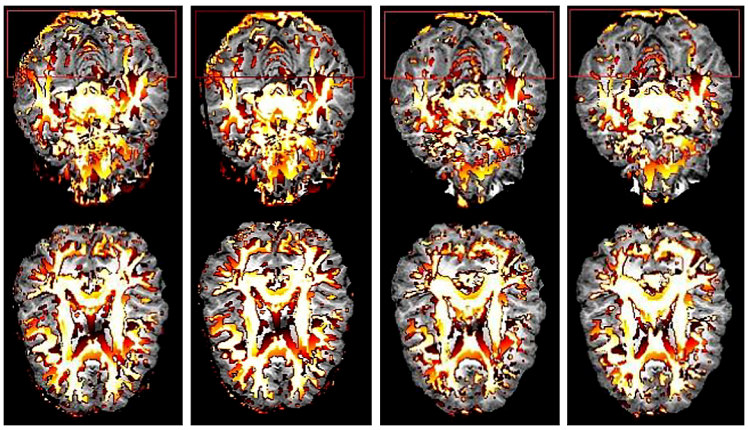

FIG. 10. *Comparison of brain anisotropic areas discovered for the real brain data set. From left to right*: FA > 0.35; Smooth- FA > 0.35; $-\log(p)$ *by* $\mathbb{K}$; $-\log(\widetilde{p})$ *by* $\mathbb{K}$. *The control level is* 0.01.

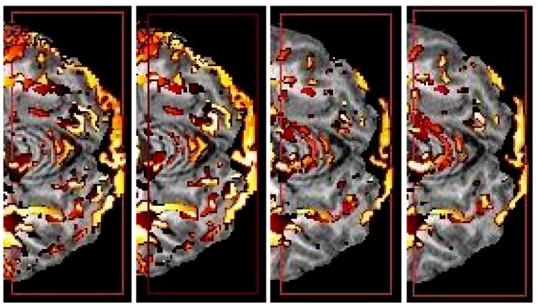

FIG. 11. *Enlarged areas in the rectangles of Figure 10. Rotated* 90°.

voxels over each subject's brain for all four methods in Table 2. In particular, denote by $\mathcal{N}(v) = \{v' = (v'_x, v'_y, v'_z) : |v'_i - v_i| \le 1, \text{ for all } i = x, y, \text{ and } z\}$ the nearest neighboring voxels of v. Table 2 displays the number of voxels carried in $\mathcal{S}_{1,m}$ (left) and $\mathcal{S}_{2,m}$ (right) over each subject's brain. Here $\mathcal{S}_{u,m} = \{v : \text{method } m \text{ identify } u \text{ voxels in } \mathcal{N}(v) \text{ as anisotropic}\}$. Clearly, $\mathcal{S}_{1,m}$ carries voxels where only the voxel v itself is identified by method m over voxels in $\mathcal{N}(v)$. Likewise, $\mathcal{S}_{2,m}$ contains voxels where only two voxels, that is, the voxel v itself and another voxel, are identified by method m over voxels in $\mathcal{N}(v)$. We observe that voxels in $\mathcal{S}_{1,m}$ and $\mathcal{S}_{2,m}$ are highly likely to be faulty findings by method m, since they are "isolated" from other identified anisotropic voxels; fiber tracts, in contrast, are typically spatially connected.

It has been seen from Table 2 that our methods continue to outperform FA-threshold and Smooth-FA-threshold approaches by producing much less isolated findings. Compared with $\mathbb{K}$-FDR, $\mathbb{K}$-FDR$_L$ produces even a smaller number of isolated identifications. Such a result is not surprising considering what has been observed in Figures 7–10. Combining all the numerical results above, we therefore recommend the identifications by $\mathbb{K}$-FDR$_L$ as the final results.

8. Discussion. In DTI studies, one of the important research topics is to refine the identification of the anisotropic water diffusion areas of human brain in vivo. There are two general strategies aiming to address this problem. The first is

TABLE 2
Number of voxels carried in $\mathcal{S}_{1,m}$ (left) and $\mathcal{S}_{2,m}$ (right) over each subject's brain

	Subject					Subject				
Methods	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
FA > 0.35	323	360	348	356	280	718	770	720	758	671
Smooth-FA > 0.35	279	288	327	298	227	775	793	657	721	660
$\mathbb{K}$-FDR	155	177	166	173	189	315	409	357	403	355
$\mathbb{K}$-FDR$_L$	63	77	100	93	103	129	158	149	212	138

to improve the DT estimation. A downstream procedure is then needed to identify the anisotropic water diffusion areas. The second is to refine the construction of the scalar measurements or establish more powerful test statistics for every brain voxel. The identification is then based on thresholding the measurements or a certain testing procedure. We observe that the second provides more intrinsic insight into the water diffusivity in each voxel, and therefore is more effective in allusion to the identification of anisotropic water diffusion areas.

From an experimental point of view, there are two ways to improve the acquisition schemes for the DWI data. One is to increase the number of diffusion gradients for every brain voxel, the other is to improve the resolution of the imaging space, that is, increase the number of brain voxels. To the best of our knowledge, existing methods for constructing scalar measurements or test statistics are all single voxel based. Therefore, the corresponding inferences improve only when the number of diffusion gradients in each voxel increases, while ignoring the possible improvement of the resolution of the imaging space. The methods proposed in this paper fill this gap by incorporating the eigenvalues in the neighboring voxels in the construction of the test statistic $\mathbb{K}$.

In this study we have established the asymptotic distribution of our proposed test statistic. One of the main assumptions required by our theoretical results is that the number of neighboring voxels for constructing $\mathbb{K}$ is large. This assumption can be well achieved when the resolution of the imaging data is high. As such, the bias components carried in the eigenvalue estimates no longer play a key role in the identification of anisotropic water diffusion brain areas. In both simulation and real data analysis, we have observed that our proposed $\mathbb{K}$-FDR and $\mathbb{K}$-FDR$_L$ approaches lead to different identification results from FA-threshold and Smooth-FA-threshold approaches, popularly adopted in the DTI community. In particular, the scattered findings by our methods are much less than those by FA-threshold and Smooth-FA-threshold approaches, indicating that by incorporating neighboring information, our methods are capable of screening out those isolated voxels which are highly likely to be faulty findings. Results based on simulated DWI data demonstrate that our proposed test statistic $\mathbb{K}$ agrees reasonably well with the χ^2 distribution when $n = 25$ (or larger), and our methods achieve better accuracy than FA-threshold and Smooth-FA-threshold approaches in the identification of anisotropic brain voxels. Furthermore, the Smooth-FA-threshold approach is capable of partially solving the bias problem in the eigenvalue estimates [Polzehl and Tabelow (2009)]. However, the performance of the approach heavily replies on the estimation of the heteroscedastic variances over the entire brain. These variances, in turn, are modeled by a linear model and estimated using the reference signals $\phi_0(v)$. We observe based on simulation studies that when the reference signals over the entire brain are not homogeneous or do not share comparable variances as attenuated signals $\phi_i(v)$, the performance of the approach varies [see more simulation results provided in the supplemental document: Yu et al. (2013)]. In contrast, under all these cases, our proposed $\mathbb{K}$-FDR$_L$ approach consistently offers descent results. We therefore conclude that over all four methods, our proposed $\mathbb{K}$-FDR$_L$

approach performs the best over all our simulation studies. Unlike the simulation examples, we are unable to show the true anisotropic (fiber) areas of the human brains in real DTI data.

We would like to point out that in DTI studies, identification of anisotropic water diffusion areas is just one step of the full analysis. Downstream analysis, such as fiber tracking, is usually needed to fully capture the physical structure of the human brain.

In this paper we have focused on establishing testing procedures to distinguish anisotropic DT voxels from isotropic ones based on second order DT models. Our proposed methods have been compared with the FA-threshold and Smooth-FA-threshold approaches. The results of this paper can serve as a benchmark in analyzing the anatomical structures of human brains based on DTI data. We observe that the following topics are highly related to this paper, and may absorb the interest of the community in future research:

- With the idea of incorporating spatial information, single-voxel-based approaches, including FA, can possibly be improved by appropriately accounting for the information in the neighboring voxels. Furthermore, in this paper, $\mathbb{K}$ is established on the DTs estimated from model (2.1). Similar approaches can be constructed based on more sophisticated DT estimates from other approaches.
- A similar strategy as that in this paper can be adopted to establish testing procedures for teasing apart the morphologies of anisotropic DTs. Furthermore, the basic ideas can be further extended to establish testing procedures for identifying the presence of signals for data with spatial structures, though we focus on DTI data in this paper.
- Another popular topic in DTI research is to consider higher-order tensor models [Grigis et al. (2011) and therein], which are powerful tools for investigating the fiber structures when there are several fiber bundles in a single voxel. The local test idea in this paper can be possibly extended to identify the number of intersected fiber bundles based on those higher-order tensor models.

Acknowledgments. We thank the Editor, the Associate Editor and four referees for invaluable suggestions, which greatly improved the quality of this paper.

SUPPLEMENTARY MATERIAL

Supplement to "Local tests for identifying anisotropic diffusion areas in human brain with DTI" (DOI: 10.1214/12-AOAS573SUPP; .pdf). This file provides proofs for Theorems 1 and 2, a short summary of FDR_L procedure [Zhang, Fan and Yu (2011)], steps for constructing $\mathbb{K}$ based on DWI data and some more simulation results.

REFERENCES

ALEXANDER, D., GEE, J. and BAJCSY, R. (1999). Similarity measures for matching diffusion tensor images. In *Proceedings of the 10th British Machine Vision Conference, 13–16 September 1999, University of Nottingham* 93–102.

ANDERSON, A. W. (2001). Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magn. Reson. Med.* **46** 1174–1188.

BASSER, P. J., MATTIELLO, J. and LEBIHAN, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. B* **103** 247–254.

BASSER, P. J. and PIERPAOLI, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B* **111** 209–219.

BEHRENS, T. E. J., BERG, H. J., JBABDI, S., RUSHWORTH, M. F. S. and WOOLRICH, M. W. (2007). Probabilistic dissusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* **34** 144–155.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

CHANG, L.-C., JONES, D. K. and PIERPAOLI, C. (2005). RESTORE: Robust estimation of tensors by outlier rejection. *Magn. Reson. Med.* **53** 1088–1095.

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. *Proc. Natl. Acad. Sci. USA* **96** 10422–10427.

DALTON, K. M., NACEWICZ, B. M., JOHNSTONE, T., SCHAEFER, H. S., GERNSBACHER, M. A., GOLDSMITH, H. H., ALEXANDER, A. L. and DAVIDSON, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nat. Neurosci.* **8** 519–526.

DOUEK, P., TURNER, R., PEKAR, J., PATRONAS, N. and BIHAN, D. L. (1991). MR color mapping of myelin fiber orientation. *J. Comput. Assist. Tomogr.* **15** 923–929.

GÖSSL, C., FAHRMEIR, L., PÜTZ, B., AUER, L. M. and AUER, D. P. (2002). Fiber tracking from DTI using linear state space models: Detectability of the pyramidal tract. *NeuroImage* **16** 378–388.

GRIGIS, A., RENARD, F., NOBLET, V., HEINRICH, C., HEITZ, F. and ARMSPACH, J. (2011). A new high order tensor decomposition: Application to reorientation. In *ISBI* 2011, *Chicago*, 8*th IEEE International Symposium on Biomedical Imaging* 258–261.

HASAN, K. M., PARKER, D. L. and ALEXANDER, A. L. (2001). Comparison of gradient encoding schemes for diffusion-tensor MRI. *Journal of Magnetic Resonce Imaging* **13** 769–780.

HEIM, S., FAHRMEIR, L., EILERS, P. H. C. and MARX, B. D. (2007). 3D space-varying coefficient models with application to diffusion tensor imaging. *Comput. Statist. Data Anal.* **51** 6212–6228. MR2407709

HENKELMAN, R. M. (1985). Measurement of signal intensities in the presence of noise in MR images. *Medical Physics* **12** 232–233.

JOHANSEN-BERG, H. and BEHRENS, T. E. J. (2009) *Diffusion MRI: From Quantitative Measurement to in-vivo Neuroanatomy*. Academic Press, London.

JONES, D. K. (2003). Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magn. Reson. Med.* **49** 7–12.

LAZAR, M. and ALEXANDER, A. L. (2003). An error analysis of white matter tractography methods: Synthetic diffusion tensor field simulations. *NeuroImage* **20** 1140–1153.

LI, Y., ZHU, H., SHEN, D., LIN, W., GILMORE, J. H. and IBRAHIM, J. G. (2011). Multiscale adaptive regression models for neuroimaging data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 559–578. MR2853730

MANGIN, J. F., POUPON, C., CLARK, C., LE BINHAN, D. and BLOCH, I. (2002). Distortion correction and robust tensor estimation for MR diffusion imaging. *Medical Image Analysis* **6** 191–198.

MOSELEY, M. E., COHEN, Y., KUCHARCZYK, J., MINTOROVITCH, J., ASGARI, H. S., WENDLAND, M. F., TSURUDA, J. and NORMAN, D. (1990). Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology* **176** 439–445.

O'DONNELL, L. J. and WESTIN, C.-F. (2007). Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. Med. Imaging* **26** 1562–1575.

PEPE, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. *Oxford Statistical Science Series* **28**. Oxford Univ. Press, Oxford. MR2260483

PIERPAOLI, C. and BASSER, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy. *Magn. Reson. Med.* **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.

POLZEHL, J. and TABELOW, K. (2009). Structural adaptive smoothing in diffusion tensor imaging: The R package dti. *Journal of Statistical Software* **31** 1–24.

SALVADOR, R., PEÑA, 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. *Hum. Brain Mapp.* **24** 144–155.

STEJSKAL, E. O. and TANNER, J. E. (1965). Spin diffusion measurements: Spin echoes in the presence of a time-depend field gradient. *Journal of Chemical Physics* **42** 288–292.

STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 479–498. MR1924302

STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 187–205. MR2035766

TABELOW, K., POLZEHL, J., SPOKOINY, V. and VOSS, H. U. (2008). Diffusion tensor imaging: Structural adaptive smoothing. *NeuroImage* **39** 1763–1773.

VAN GELDEREN, P., DE VLEESCHOUWER, M. H., DESPRES, D., PEKAR, J., VAN ZIJL, P. C. and MOONEN, C. T. (1994). Water diffusion and acute stroke. *Magn. Reson. Med.* **31** 154–163.

XU, D., MORI, S., SOLAIYAPPAN, M., VAN ZIJL, P. C. M. and DAVATZIKOS, C. (2002). A framework for callosal fiber distribution analysis. *NeuroImage* **17** 1131–1143.

YU, T. (2009). Local tests for detecting human brain isotropy-anisotropy areas on DT–MRI. Ph.D. thesis, Univ. Wisconsin-Madison. MR2714030

YU, T., ZHANG, C. M., ALEXANDER, A. L. and DAVIDSON, R. J. (2013). Supplement to "Local tests for identifying anisotropic diffusion areas in human brain with DTI." DOI:10.1214/12-AOAS573SUPP.

ZHANG, C., FAN, J. and YU, T. (2011). Multiple testing via FDR_L for large-scale imaging data. *Ann. Statist.* **39** 613–642. MR2797858

ZHU, H., XU, D., AMIR, R., HAO, X., ZHANG, H., ALAYAR, K., BANSAL, R. and PETERSON, B. (2006). A statistical framework for the classification of tensor morphology in diffusion tensor images. *Magnetic Resonance Imaging* **24** 569–582.

ZHU, H., ZHANG, H., IBRAHIM, J. G. and PETERSON, B. S. (2007). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data. *J. Amer. Statist. Assoc.* **102** 1085–1102. MR2412530

T. YU
DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117546
E-MAIL: stayt@nus.edu.sg

C. ZHANG
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706
USA
E-MAIL: cmzhang@stat.wisc.edu

A. L. ALEXANDER
R. J. DAVIDSON
WAISMAN CENTER
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53705
USA
E-MAIL: alalexander2@wisc.edu
rjdavids@wisc.edu

Supplement to "Local Tests for Identifying Anisotropic Diffusion Areas in Human Brain with DTI"

Tao Yu, Chunming Zhang, Alexander L. Alexander and Richard J. Davidson

May 28, 2012

1 Introduction

This is a supplementary document to the corresponding paper submitted to the *Annals of Applied Statistics*. It consists of the following four main parts. Sections 2 - 4 provide the regularity conditions and detailed proofs for Theorems 1 and 2 in Section 5 of the submitted paper. Section 5 shortly summarizes the FDR_L procedure proposed in Zhang *et al.* (2011). Section 6 briefly describes the computational procedure for constructing our proposed test statistic $\mathbb{K}$ for every brain voxel based on DWI data. Section 7 presents some more simulation results.

2 Regularity Conditions

We first impose some technical assumptions, which are not the weakest possible. For expository brevity, we assume the model errors $\boldsymbol{\epsilon}_j(v)$ are independent, which can be easily relaxed to block dependent situation.

Condition A:

A1. For each voxel v, in its neighborhood (containing n voxels), the computed eigenvalues satisfy model (3.2), but we only observe their ordered version, say, $\lambda_{j,(3)} \geq \lambda_{j,(2)} \geq \lambda_{j,(1)}$, $j = 1, \ldots, n$.

A2. For $v \in \mathcal{V}_0$, the voxel set where H_0 of (4.2) is true, $\boldsymbol{\epsilon}_j(v) = (\epsilon_{j,1}(v), \epsilon_{j,2}(v), \epsilon_{j,3}(v))^T$, are independent for $j = 1, \ldots, n$ and identically distributed for $v \in \mathcal{V}_0$, $j = 1, \ldots, n$; $E\{\boldsymbol{\epsilon}_j(v)\} = \mathbf{0}$, $\text{var}\{\boldsymbol{\epsilon}_j(v)\} = \Sigma_{\boldsymbol{\epsilon}}$, some positive definite matrix. Furthermore, we assume that there exists some $\alpha > 0$ such that $E\{\epsilon_{j,k}^{4+\alpha}(v)\} < \infty$ and $E\{\epsilon_{j,(k)}^{2+2\alpha}(v)\} < \infty$ for $v \in \mathcal{V}_0$, $j = 1, \ldots, n$ and $k = 1, 2, 3$.

A3. For different voxels, their computed eigenvalues are independent.

A4. $n \to \infty$ as $N \to \infty$.

A5. $n \log(n)/|\mathcal{V}_0| = o(1)$ as $N \to \infty$.

A6. $\widehat{\mathcal{V}}_0 = \mathcal{V}_0$.

A7. For any voxel in the isotropic regions, MSE is bounded away from 0, that is, there exists some $\delta > 0$, such that MSE $> \delta$ for any $v \in \mathcal{V}_0$. Furthermore, $d^2(\mathbf{b}) > \delta$, where $d(\mathbf{x})$ and $\mathbf{b}$ are defined in the proof of Lemma 1.

3 Lemmas 1 and 2

Lemma 1 *Assume Condition* **A**. *For each voxel* v, *define* $\boldsymbol{T}_n = \sqrt{n}(\boldsymbol{U}_n - \widehat{\boldsymbol{\theta}}_{\boldsymbol{U}_n})$. *Then for any* $v \in \mathcal{V}_0$, *as* $n \to \infty$, $\boldsymbol{T}_n \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Sigma)$ *for some covariance matrix* Σ.

Proof: Let $\mathbf{y}_j(v) = (2S_j^2(v), \lambda_{j,(3)}(v) - \beta_j(v), \lambda_{j,(2)}(v) - \beta_j(v), \lambda_{j,(1)}(v) - \beta_j(v))^T$, where $S_j^2(v) = \sum_{k=1}^3 \{\lambda_{j,(k)}(v) - \overline{\lambda}_{j,(\cdot)}(v)\}^2/2$. Note that under H_0, i.e. $v \in \mathcal{V}_0$, $\lambda^*_{(1)}(v) = \lambda^*_{(2)}(v) = \lambda^*_{(3)}(v)$, direct derivation shows

$$E\{S_j^2(v)\} = E\{S_{\boldsymbol{\epsilon}}^2\}, \quad \text{and} \quad \lambda_{j,(k)}(v) - \beta_j(v) = \lambda^*_{(1)}(v) + \epsilon_{j,(k)}(v),$$

where $S_{\boldsymbol{\epsilon}}^2$ is the sample variance of $\{\epsilon_{1,1}(1), \epsilon_{1,2}(1), \epsilon_{1,3}(1)\}$. Thus,

$$\begin{aligned} E\{\mathbf{y}_1(v)\} &= (E(2S_{\boldsymbol{\epsilon}}^2), \lambda^*_{(1)}(v) + E\{\epsilon_{1,(3)}(1)\}, \lambda^*_{(1)}(v) + E\{\epsilon_{1,(2)}(1)\}, \lambda^*_{(1)}(v) + E\{\epsilon_{1,(1)}(1)\})^T \\ &\equiv (a_1, a_2(v), a_3(v), a_4(v))^T \equiv \mathbf{a}(v), \\ \text{var}\{\mathbf{y}_1(v)\} &= \text{var}\{(2S_{\boldsymbol{\epsilon}}^2, \epsilon_{1,(3)}(1), \epsilon_{1,(2)}(1), \epsilon_{1,(1)}(1))^T\} \equiv \Omega. \end{aligned}$$

By the central limit theorem,

$$\sqrt{n}\{\overline{\mathbf{y}}_\cdot(v) - \mathbf{a}(v)\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Omega). \tag{A.3.1}$$

Let $\boldsymbol{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_\mu(\mathbf{x}))^T$ be $\mathbb{R}^4 \to \mathbb{R}^\mu$, where

$$\begin{aligned} g_l(\mathbf{x}) &= \frac{A(l,\cdot)(x_2, x_3, x_4)^T}{d(\mathbf{x})} \Big(\frac{x_1}{2}\Big)^{0.5}, \qquad l = 1, \ldots, \mu, \\ d(\mathbf{x}) &= \sqrt{\frac{1}{2}x_1 - \frac{1}{3}(x_2^2 + x_3^2 + x_4^2 - x_2x_3 - x_2x_4 - x_3x_4)}. \end{aligned} \tag{A.3.2}$$

Delta method shows,

$$\sqrt{n}\{\boldsymbol{g}(\overline{\mathbf{y}}_\cdot(v)) - \boldsymbol{g}(\mathbf{a}(v))\} \xrightarrow{\mathcal{L}} \nabla \boldsymbol{g}(\mathbf{a}(v))^T \cdot N(\mathbf{0}, \Omega). \tag{A.3.3}$$

Direct computations show that under H_0, $\boldsymbol{g}(\mathbf{a}(v)) = \boldsymbol{g}(\mathbf{b})$, $\nabla\boldsymbol{g}(\mathbf{a}(v)) = \nabla\boldsymbol{g}(\mathbf{b})$ and $d(\mathbf{a}(v)) = d(\mathbf{b})$, where $\mathbf{b} = (E(2S_{\boldsymbol{\epsilon}}^2), E\{\epsilon_{1,(3)}(1)\}, E\{\epsilon_{1,(2)}(1)\}, E\{\epsilon_{1,(1)}(1)\})^T \equiv (b_1, b_2, b_3, b_4)^T$ independent with v. Note the fact that, $d(\mathbf{x})^2 = x_1/2 - \{(x_2 - \widetilde{x})^2 + (x_3 - \widetilde{x})^2 + (x_4 - \widetilde{x})^2\}/2$, where $\widetilde{x} = (x_2 + x_3 + x_4)/3$. For each voxel v,

$$\begin{aligned} d^2(\overline{\mathbf{y}}_{\cdot}(v)) &= \frac{1}{2}\cdot 2\cdot \overline{S_{\cdot}^2}(v) - \frac{1}{2}\cdot\sum_{k=1}^{3}\{\overline{\lambda}_{\cdot,(k)}(v) - \overline{\lambda}_{\cdot,(\cdot)}(v)\}^2 \\ &= \frac{1}{n}\sum_{j=1}^{n}\Big[\frac{1}{2}\sum_{k=1}^{3}\{\lambda_{j,(k)}(v) - \overline{\lambda}_{j,(\cdot)}(v)\}^2\Big] - \frac{1}{2}\cdot\sum_{k=1}^{3}\{\overline{\lambda}_{\cdot,(k)}(v) - \overline{\lambda}_{\cdot,(\cdot)}(v)\}^2 \\ &= \frac{1}{2n}\sum_{j=1}^{n}\sum_{k=1}^{3}\{\lambda_{j,(k)}(v) - \overline{\lambda}_{\cdot,(k)}(v) - \overline{\lambda}_{j,(\cdot)}(v) + \overline{\lambda}_{\cdot,(\cdot)}(v)\}^2 \\ &= \frac{n-1}{n}\mathrm{MSE}, \end{aligned} \tag{A.3.4}$$

thus, $\boldsymbol{g}(\overline{\mathbf{y}}_{\cdot}(v)) = \boldsymbol{U}_n\sqrt{n}/\sqrt{n-1}$, which together with (A.3.3) shows that for each voxel $v \in \mathcal{V}_0$,

$$\sqrt{n}\{\boldsymbol{U}_n - \boldsymbol{g}(\mathbf{b})\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Sigma), \tag{A.3.5}$$

where $\Sigma = \nabla\boldsymbol{g}(\mathbf{b})^T\Omega\nabla\boldsymbol{g}(\mathbf{b})$. Denote $\widehat{\boldsymbol{\theta}}_{\boldsymbol{V}_n} = (\widehat{\boldsymbol{\theta}}_{V_1}, \ldots, \widehat{\boldsymbol{\theta}}_{V_\mu})^T$, where $\widehat{\boldsymbol{\theta}}_{V_l}$, $l = 1, \ldots, \mu$ is the median of $V_l \equiv \sqrt{n}\{U_l - g_l(\mathbf{b})\}$ over $v \in \mathcal{V}_0$, U_l, $l = 1, \ldots, \mu$ is the lth component of the random vector $\boldsymbol{U}_n$. Then we immediately have $\widehat{\boldsymbol{\theta}}_{\boldsymbol{V}_n} = \sqrt{n}\{\widehat{\boldsymbol{\theta}}_{\boldsymbol{U}_n} - \boldsymbol{g}(\mathbf{b})\}$, and thus,

$$\begin{aligned} \boldsymbol{T}_n &= \sqrt{n}(\boldsymbol{U}_n - \widehat{\boldsymbol{\theta}}_{\boldsymbol{U}_n}) \\ &= \sqrt{n}\{\boldsymbol{U}_n - \boldsymbol{g}(\mathbf{b})\} - \sqrt{n}\{\widehat{\boldsymbol{\theta}}_{\boldsymbol{U}_n} - \boldsymbol{g}(\mathbf{b})\} \\ &= \sqrt{n}\{\boldsymbol{U}_n - \boldsymbol{g}(\mathbf{b})\} - \widehat{\boldsymbol{\theta}}_{\boldsymbol{V}_n}. \end{aligned} \tag{A.3.6}$$

For any n, let F_n be the joint cumulative distribution function (c.d.f.) of random vector $\boldsymbol{V}_n$, where $\boldsymbol{V}_n = (V_1, \ldots, V_\mu)^T$. Then F_n's are continuous c.d.f.'s and $F_n \to \Phi_\Sigma$ the c.d.f of $N(\mathbf{0}, \Sigma)$, based upon our assumption and (A.3.5). Furthermore, apply Proposition 1.6 in Shao (2003), we have $F_n(\mathbf{x}) \to \Phi_\Sigma(\mathbf{x})$ uniformly in $\mathbf{x}$, which immediately leads to the convergence property of $\boldsymbol{\theta}_{\boldsymbol{V}_n}$, the theoretical median of $F_n(\mathbf{x})$'s,

$$\boldsymbol{\theta}_{\boldsymbol{V}_n} \to \mathbf{0}, \text{ as } n \to \infty. \tag{A.3.7}$$

Without loss of generality, we assume that $|\mathcal{V}_0|/n \equiv m_n$ is an integer. Note that for each voxel v, $\boldsymbol{V}_n$ only depends on its n neighboring voxels used for constructing $\boldsymbol{U}_n$. By blocking arguments, the set of $\boldsymbol{V}_n$ for $v \in \mathcal{V}_0$ can be partitioned into n disjoint subsets $\mathcal{G}_1, \ldots, \mathcal{G}_n$, such that all $\boldsymbol{V}_n$'s, in each subset, are independent and identically distributed (i.i.d.) random vectors. For presentation convenience, denote $\mathcal{G}_j = \{V_1^{(j)}, \ldots, V_{m_n}^{(j)}\}$, $j =$

$1, \ldots, n$. Let $\widehat{\boldsymbol{\theta}}^{(j)}_{\boldsymbol{V}_n}, j = 1, \ldots, n$ be the sample median of $\boldsymbol{V}_n$ over $\mathcal{G}_j$. Apply Theorem 5.9 of Shao (2003), for any $\epsilon > 0$,

$$\begin{aligned} & P(\max_{1\le j\le n} \|\widehat{\boldsymbol{\theta}}^{(j)}_{\boldsymbol{V}_n} - \boldsymbol{\theta}_{\boldsymbol{V}_n}\| > \epsilon) \le \sum_{j=1}^{n} P(\|\widehat{\boldsymbol{\theta}}^{(j)}_{\boldsymbol{V}_n} - \boldsymbol{\theta}_{\boldsymbol{V}_n}\| > \epsilon) \\ \le & \sum_{j=1}^{n}\sum_{l=1}^{\mu} P(|\widehat{\boldsymbol{\theta}}^{(j)}_{V_l} - \boldsymbol{\theta}_{V_l}| > \epsilon/\sqrt{\mu}) \le 2nC \sum_{l=1}^{\mu} e^{-2m_n\delta^2_{\epsilon,n,l}}, \end{aligned} \tag{A.3.8}$$

where

$$\delta_{\epsilon,n,l} = \min\{F_{n,l}(\boldsymbol{\theta}_{V_l} + \epsilon/\sqrt{\mu}) - 0.5, 0.5 - F_{n,l}(\boldsymbol{\theta}_{V_l} - \epsilon/\sqrt{\mu})\} \to \delta_{\epsilon,l} > 0, \tag{A.3.9}$$

as $n \to \infty$ by (A.3.7) and the uniform convergence of $F_n(\mathbf{x})$; $F_{n,l}$ is the lth marginal c.d.f. of F_n. (A.3.8) together with A5 in Condition **A** and (A.3.9), leads to

$$P(\max_{1\le j\le n} \|\widehat{\boldsymbol{\theta}}^{(j)}_{\boldsymbol{V}_n} - \boldsymbol{\theta}_{\boldsymbol{V}_n}\| > \epsilon) \to 0. \tag{A.3.10}$$

Note the fact that for each $l = 1, \ldots, \mu$, $\min_{1\le j\le n} \widehat{\boldsymbol{\theta}}^{(j)}_{V_l} \le \widehat{\boldsymbol{\theta}}_{V_l} \le \max_{1\le j\le n} \widehat{\boldsymbol{\theta}}^{(j)}_{V_l}$. Therefore, for any value of $\boldsymbol{\theta}_{\boldsymbol{V}_n}$, $\|\widehat{\boldsymbol{\theta}}_{\boldsymbol{V}_n} - \boldsymbol{\theta}_{\boldsymbol{V}_n}\| \le \max_{1\le j\le n} \|\widehat{\boldsymbol{\theta}}^{(j)}_{\boldsymbol{V}_n} - \boldsymbol{\theta}_{\boldsymbol{V}_n}\|$. Applying (A.3.10), we have

$$P(\|\widehat{\boldsymbol{\theta}}_{\boldsymbol{V}_n} - \boldsymbol{\theta}_{\boldsymbol{V}_n}\| > \epsilon) \le P(\max_{1\le j\le n} \|\widehat{\boldsymbol{\theta}}^{(j)}_{\boldsymbol{V}_n} - \boldsymbol{\theta}_{\boldsymbol{V}_n}\| > \epsilon) \to 0. \tag{A.3.11}$$

Combine (A.3.7) and (A.3.11), we have

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{V}_n} \xrightarrow{\mathcal{P}} \mathbf{0}. \tag{A.3.12}$$

Combine (A.3.5), (A.3.6) and (A.3.12), the desired result follows. ■

Lemma 2 *Assume Condition* **A**. *Then* $\widehat{\Sigma} \xrightarrow{\mathcal{P}} \Sigma$.

Proof: Consider the (l_1, l_2)th component in $\widehat{\Sigma}$,

$$\begin{aligned} \widehat{\Sigma}(l_1, l_2) &= \frac{n}{|\mathcal{V}_0| - 1} \sum_{v\in\mathcal{V}_0} (U_{l_1} - \overline{U}_{l_1})(U_{l_2} - \overline{U}_{l_2}) \\ &= \frac{1}{|\mathcal{V}_0| - 1} \sum_{v\in\mathcal{V}_0} (V_{l_1} - \overline{V}_{l_1})(V_{l_2} - \overline{V}_{l_2}) \\ &= \frac{1}{|\mathcal{V}_0| - 1} \sum_{v\in\mathcal{V}_0} V_{l_1}V_{l_2} - \frac{|\mathcal{V}_0|}{|\mathcal{V}_0| - 1} \overline{V}_{l_1}\overline{V}_{l_2} \\ &\equiv \frac{|\mathcal{V}_0|}{|\mathcal{V}_0| - 1}(I_1 + I_2), \end{aligned} \tag{A.3.13}$$

where $\overline{U}_{l_1}$ is the mean value of U_{l_1} over $\mathcal{V}_0$ and likewise, $\overline{V}_{l_1}$ is the mean value of V_{l_1} over $\mathcal{V}_0$. For any $\mathbf{x} \in \mathbb{R}^4$ such that $d^2(\mathbf{x}) > \delta$, $g_l(\mathbf{x})\,\mathrm{I}\{d^2(\mathbf{x}) > \delta\}$, for each $l = 1, \ldots, \mu$, is Lipschitz continuous. Therefore, refer to (A.3.4), for any voxel v, when MSE $> \delta$,

$$\begin{aligned} |V_l| &= \sqrt{n}|U_l - g_l(\mathbf{b})| \\ &= \sqrt{n}|U_l\,\mathrm{I}(\mathrm{MSE} > \delta) - g_l(\mathbf{b})\,\mathrm{I}\{d^2(\mathbf{b}) > \delta\}| \\ &= \sqrt{n}|g_l(\overline{\mathbf{y}}_{\cdot}(v))\,\mathrm{I}\{d^2(\overline{\mathbf{y}}_{\cdot}(v)) > \delta\} - g_l(\mathbf{a}(v))\,\mathrm{I}\{d^2(\mathbf{a}(v)) > \delta\}| \\ &\leq L\sqrt{n}\|\overline{\mathbf{y}}_{\cdot}(v) - \mathbf{a}(v)\|, \end{aligned} \tag{A.3.14}$$

for some constant $L \geq 0$. Consider the random variable sequence $n\|\overline{\mathbf{y}}_{\cdot}(v) - \mathbf{a}(v)\|^2$,

i. From (A.3.1), $n\|\overline{\mathbf{y}}_{\cdot}(v) - \mathbf{a}(v)\|^2 \xrightarrow{\mathcal{L}} Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2$, where $(Z_1, Z_2, Z_3, Z_4)^T \sim N(\mathbf{0}, \Omega)$, and $E(Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2) = 4\mathrm{var}(S_{\boldsymbol{\epsilon}}^2) + \mathrm{var}\{\epsilon_{1,(3)}(1)\} + \mathrm{var}\{\epsilon_{1,(2)}(1)\} + \mathrm{var}\{\epsilon_{1,(1)}(1)\}$.

ii. Under H_0,

$$\begin{aligned} & E(n\|\overline{\mathbf{y}}_{\cdot}(v) - \mathbf{a}(v)\|^2) \\ = \;& nE[\{2\overline{S_{\cdot}^2}(v) - a_1(v)\}^2 + \{\overline{\lambda}_{\cdot,(3)}(v) - \overline{\beta}_{\cdot}(v) - a_2(v)\}^2 \\ & + \{\overline{\lambda}_{\cdot,(2)}(v) - \overline{\beta}_{\cdot}(v) - a_3(v)\}^2 + \{\overline{\lambda}_{\cdot,(1)}(v) - \overline{\beta}_{\cdot}(v) - a_4(v)\}^2] \\ = \;& nE[\{2\overline{S_{\boldsymbol{\epsilon},\cdot}^2}(v) - b_1\}^2 + \{\overline{\epsilon}_{\cdot,(3)}(v) - b_2\}^2 + \{\overline{\epsilon}_{\cdot,(2)}(v) - b_3\}^2 + \{\overline{\epsilon}_{\cdot,(1)}(v) - b_4\}^2] \\ = \;& 4\mathrm{var}(S_{\boldsymbol{\epsilon}}^2) + \mathrm{var}\{\epsilon_{1,(3)}(1)\} + \mathrm{var}\{\epsilon_{1,(2)}(1)\} + \mathrm{var}\{\epsilon_{1,(1)}(1)\} \\ \to \;& E(Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2). \end{aligned}$$

Parts i and ii show that, for each $v \in \mathcal{V}_0$, $n\|\overline{\mathbf{y}}_{\cdot}(v) - \mathbf{a}(v)\|^2$ is *uniformly integrable* (*u.i.*) in n (van der Vaart 1998, Theorem 2.20). This together with (A.3.14) shows

$$V_{l_1}(v)V_{l_2}(v), \quad l_1, l_2 = 1, \ldots \mu \text{ is } u.i. \text{ in } n \text{ for each } v \in \mathcal{V}_0.$$

Recall the partition in Lemma 1,

$$I_1 = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m_n}\sum_{i=1}^{m_n} V_{i,l_1}^{(j)}V_{i,l_2}^{(j)} \equiv \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m_n}\sum_{i=1}^{m_n} X_{n,j,i}. \tag{A.3.15}$$

Then $X_{n,j,i}$ satisfies: (I) For $i = 1, \ldots m_n$ they are i.i.d. random variables. (II) For any fixed j and i, they are *u.i.* in n. (III) Applying (A.3.14), under H_0, for every n, j, i and its associated voxel v,

$$|X_{n,j,i}| \leq nL^2\|\overline{\mathbf{y}}_{\cdot}(v) - \mathbf{a}_1(v)\|^2. \tag{A.3.16}$$

iii. Now we show $\sum_{i=1}^{m_n} P(|X_{n,j,i}| > m_n) \to 0$.

$$\sum_{i=1}^{m_n} P(|X_{n,j,i}| > m_n) = m_n P(|X_{n,j,1}| > m_n) = m_n \int_{|X_{n,j,1}| > m_n} dP$$

$$\leq \int_{|X_{n,j,1}|>m_n} |X_{n,j,1}| dP = E|X_{n,j,1}\,\mathrm{I}(X_{n,j,1} > m_n)| \to 0,$$

since $X_{n,j,i}$ is *u.i.*.

iv. Next we show $\frac{1}{m_n^2}\sum_{i=1}^{m_n} E\{X_{n,j,i}^2\,\mathrm{I}(|X_{n,j,i}| \leq m_n)\} \to 0$. Setting $\beta = \alpha/4$, applying (A.3.16),

$$\begin{aligned} X_{n,j,1}^2\,\mathrm{I}(|X_{n,j,1}| \leq m_n) &= |X_{n,j,1}|^{1+\beta}\{|X_{n,j,1}|^{1-\beta}\,\mathrm{I}(|X_{n,j,1}| \leq m_n)\} \\ &\leq n^{1+\beta}L^{2+2\beta}\|\overline{\mathbf{y}}_{\cdot}(v) - \mathbf{a}(v)\|^{2+2\beta} m_n^{1-\beta}, \end{aligned} \tag{A.3.17}$$

where v is the voxel associated with $X_{n,j,1}$. Thus,

$$\begin{aligned} & \frac{1}{m_n^2}\sum_{i=1}^{m_n} E\{X_{n,j,i}^2\,\mathrm{I}(|X_{n,k,i}| \leq m_n)\} \\ = \;& \frac{1}{m_n} E\{X_{n,j,1}^2\,\mathrm{I}(|X_{n,j,1}| \leq m_n)\} \\ \leq \;& \frac{L^{2+2\beta}}{m_n^{\beta}} n^{1+\beta} E[\{\|\overline{\mathbf{y}}_{\cdot}(v) - \mathbf{a}(v)\|^2\}^{1+\beta}] \qquad \text{(A.3.18)} \\ = \;& \frac{L^{2+2\beta}}{m_n^{\beta}} n^{1+\beta} E[\{\|\overline{\mathbf{y}}_{\cdot}(1) - \mathbf{a}(1)\|^2\}^{1+\beta}] \\ = \;& \frac{L^{2+2\beta}}{m_n^{\beta}} n^{1+\beta} E\Big[[\{2\overline{S^2_{\boldsymbol{\epsilon},\cdot}}(1) - b_1\}^2 + \{\overline{\epsilon}_{\cdot,(3)}(1) - b_2\}^2 \\ & + \{\overline{\epsilon}_{\cdot,(2)}(1) - b_3\}^2 + \{\overline{\epsilon}_{\cdot,(1)}(1) - b_4\}^2]^{1+\beta}\Big] \qquad \text{(A.3.19)} \\ \leq \;& \frac{L^{2+2\beta}}{m_n^{\beta}} n^{1+\beta} \Big[[E[\{2\overline{S^2_{\boldsymbol{\epsilon},\cdot}}(1) - b_1\}^{2(1+\beta)}]]^{1/(1+\beta)} + [E[\{\overline{\epsilon}_{\cdot,(3)}(1) - b_2\}^{2(1+\beta)}]]^{1/(1+\beta)} \\ & + [E[\{\overline{\epsilon}_{\cdot,(2)}(1) - b_3\}^{2(1+\beta)}]]^{1/(1+\beta)} + [E[\{\overline{\epsilon}_{\cdot,(1)}(1) - b_4\}^{2(1+\beta)}]]^{1/(1+\beta)}\Big]^{1+\beta} \\ = \;& \frac{L^{2+2\beta}}{m_n^{\beta}} n^{1+\beta} \{J_1^{1/(1+\beta)} + J_2^{1/(1+\beta)} + J_3^{1/(1+\beta)} + J_4^{1/(1+\beta)}\}^{1+\beta} \qquad \text{(A.3.20)} \end{aligned}$$

where (A.3.17) is followed by (A.3.16); (A.3.18) is from the definitions of $\overline{\mathbf{y}}_{\cdot}(1)$ and $\mathbf{a}(1)$; (A.3.19) applies Minkowski's inequality (Shao 2003, expression (1.43)). Note that,

$$\begin{aligned} J_1 &= E[\{2\overline{S^2_{\boldsymbol{\epsilon},\cdot}}(1) - b_1\}^{2(1+\beta)}] \\ &= \frac{1}{n^{2+2\beta}} E\Big[\Big[\sum_{j=1}^{n}\{2S^2_{\boldsymbol{\epsilon},j}(1) - b_1\}\Big]^{2(1+\beta)}\Big] \\ &\leq \frac{1}{n^{2+2\beta}} \frac{C}{n^{1-2(1+\beta)/2}} \sum_{j=1}^{n} E[\{2S^2_{\boldsymbol{\epsilon},j}(1) - b_1\}^{2(1+\beta)}] \qquad \text{(A.3.21)} \\ &= \frac{C}{n^{1+\beta}} E\{(2S^2_{\boldsymbol{\epsilon}} - b_1)^{2(1+\beta)}\}, \end{aligned}$$

where (A.3.21) applies an inequality due to Marcinkiewicz and Zygmund (Shao 2003, expression (1.45)); C is a constant depending only on β. Thus, there exist constant $C > 0$ depending only on β, such that

$$J_1 \leq \frac{C}{n^{1+\beta}} E\{\epsilon_{1,1}^{4+4\beta}(1)\} \leq \frac{C}{n^{1+\beta}} C' \tag{A.3.22}$$

for some $C' > 0$, based upon A2 in Condition **A**. Exactly the same procedures show that there exist constants $C > 0$ depending only on β and $C' > 0$ such that

$$J_2 \leq \frac{C}{n^{1+\beta}} C', \quad J_3 \leq \frac{C}{n^{1+\beta}} C', \quad J_4 \leq \frac{C}{n^{1+\beta}} C', \tag{A.3.23}$$

Combine (A.3.20), (A.3.22) and (A.3.23),

$$\frac{1}{m_n^2} \sum_{i=1}^{m_n} E\{X_{n,j,i}^2 \, \mathrm{I}_{(|X_{n,j,i}| \leq m_n)}\} \leq \frac{L^{2+2\beta}}{m_n^\beta} n^{1+\beta} \cdot \frac{C 4^{1+\beta}}{n^{1+\beta}} C' = \frac{C C' 4^{1+\beta} L^{2+2\beta}}{m_n^\beta} \to 0,$$

which completes our proof of part iv.

Based on parts iii and iv, and "Weak Law for Triangular Arrays" (Durrett, 1991, Theorem 5.8, Chapter 1), we have $\sum_{i=1}^{m_n} [X_{n,j,i} - E\{X_{n,j,i} \, \mathrm{I}(|X_{n,j,i}| \leq m_n)\}]/m_n \xrightarrow{\mathcal{P}} 0$. That is,

$$\frac{1}{m_n} \sum_{i=1}^{m_n} X_{n,j,i} - E\{X_{n,j,1} \, \mathrm{I}(|X_{n,j,1}| \leq m_n)\} \xrightarrow{\mathcal{P}} 0.$$

Note that $X_{n,j,i}$ is *u.i.*, so does $X_{n,j,1} \, \mathrm{I}(|X_{n,j,1}| \leq m_n)$, which, together with the fact that $X_{n,j,1} \, \mathrm{I}(|X_{n,j,1}| \leq m_n) \xrightarrow{\mathcal{L}} W_{l_1} W_{l_2}$, where $W = (W_1, \ldots, W_\mu)^T \sim N(\mathbf{0}, \Sigma)$, shows

$$\frac{1}{m_n} \sum_{i=1}^{m_n} X_{n,j,i} \xrightarrow{\mathcal{P}} \Sigma(l_1, l_2). \tag{A.3.24}$$

On the other hand,

$$E\Big(\frac{1}{m_n} \sum_{i=1}^{m_n} X_{n,j,i}\Big) = E(X_{n,j,1}) \to \Sigma(l_1, l_2), \tag{A.3.25}$$

since $X_{n,j,i}$ is *u.i.* and $X_{n,j,1} \xrightarrow{\mathcal{L}} W_{l_1} W_{l_2}$. (A.3.24) and (A.3.25) show that $\frac{1}{m_n} \sum_{i=1}^{m_n} X_{n,j,i}$ is *u.i.*. Recall (A.3.15). Consider

$$\begin{aligned} E\{|I_1 - \Sigma(l_1, l_2)|\} &\leq \frac{1}{n} \sum_{j=1}^{n} E\Big\{\Big|\frac{1}{m_n} \sum_{i=1}^{m_n} X_{n,j,i} - \Sigma(l_1, l_2)\Big|\Big\} \\ &= E\Big\{\Big|\frac{1}{m_n} \sum_{i=1}^{m_n} X_{n,1,i} - \Sigma(l_1, l_2)\Big|\Big\} \to 0, \end{aligned}$$

which shows $I_1 \xrightarrow{\mathcal{L}} \Sigma(l_1, l_2)$. Thus,

$$I_1 \xrightarrow{\mathcal{P}} \Sigma(l_1, l_2). \tag{A.3.26}$$

Similarly, $I_2 \xrightarrow{\mathcal{P}} 0$, which together with (A.3.13) and (A.3.26) entails the result. ■

4 Proofs of Theorems

Proof of Theorem 1

Proof: Combining Lemma 1–2, the result follows. ■

Proof of Theorem 2

Proof: We adopt all the notations in Lemmas 1 and 2. Follow the same steps as Lemma 1, we have

$$\sqrt{n}\{\boldsymbol{U}_n - \boldsymbol{g}(\widetilde{\mathbf{a}}(v))\} \xrightarrow{\mathcal{L}} \nabla \boldsymbol{g}(\widetilde{\mathbf{a}}(v))^T \cdot N(\mathbf{0}, \Omega(v)), \tag{A.4.1}$$

where $\widetilde{\mathbf{a}}(v) = (E\{2S_1^2(v)\}, E\{\lambda_{1,(3)}(v)\}, E\{\lambda_{1,(2)}(v)\}, E\{\lambda_{1,(1)}(v)\})^T$, $\Omega(v) = \text{var}\{\mathbf{y}_1(v)\} = \text{var}\{2S_1^2(v), \lambda_{1,(3)}(v), \lambda_{1,(2)}(v), \lambda_{1,(1)}(v)\}$. Recall (A.3.6),

$$\begin{aligned} \boldsymbol{T}_n &= \sqrt{n}\{\boldsymbol{U}_n - \boldsymbol{g}(\mathbf{b})\} - \widehat{\boldsymbol{\theta}}_{\boldsymbol{V}_n} \\ &= \sqrt{n}\{\boldsymbol{U}_n - \boldsymbol{g}(\widetilde{\mathbf{a}}(v))\} + \sqrt{n}\{\boldsymbol{g}(\widetilde{\mathbf{a}}(v)) - \boldsymbol{g}(\mathbf{b})\} - \widehat{\boldsymbol{\theta}}_{\boldsymbol{V}_n}. \end{aligned} \tag{A.4.2}$$

Combine (A.3.12), (A.4.1) and (A.4.2), we have $\boldsymbol{T}_n/\sqrt{n} \xrightarrow{\mathcal{P}} \boldsymbol{g}(\widetilde{\mathbf{a}}(v)) - \boldsymbol{g}(\mathbf{b})$. Thus, $n^{-1}\mathbb{K} = (n^{-1/2}\boldsymbol{T}_n)^T \widehat{\Sigma}^{-1} (n^{-1/2}\boldsymbol{T}_n) \xrightarrow{\mathcal{P}} M$, where

$$M = \{\boldsymbol{g}(\widetilde{\mathbf{a}}(v)) - \boldsymbol{g}(\mathbf{b})\}^T \Sigma^{-1} \{\boldsymbol{g}(\widetilde{\mathbf{a}}(v)) - \boldsymbol{g}(\mathbf{b})\}. \quad ■ \tag{A.4.3}$$

5 Summary of FDR_L Approach

The FDR_L approach proposed by Zhang *et al.* (2011) is a powerful tool in dealing with multiple test problems for spatially located imaging data. In this paper, we apply the FDR_L approach to the following multiple testing problem. For each voxel v, we are interested in testing

$$H_0(v) : \text{DT in } v \text{ is isotropic} \quad \text{versus} \quad H_1(v) : \text{DT in } v \text{ is anisotropic.}$$

Denote by $\{p(v), v \in \mathcal{V}\}$ the collection of p-values, based upon the constructed test statistic $\mathbb{K}$ by our approach, over all voxels of the entire brain. In FDR_L approach, instead of using $\{p(v), v \in \mathcal{V}\}$ directly, a set of locally aggregated p-values, $\{\widetilde{p}(v) : v \in \mathcal{V}\}$, is acquired by the following procedure.

Step 1. Choose a local neighborhood with size l.

Step 2. At each voxel v, find the set N_v of its neighborhood voxels, and the set $\{p(v') : v' \in N_v\}$ of the corresponding p-values.

Step 3. At each voxel v, apply a transformation h: $[0,1]^l \mapsto [0,1]$ to the set of p-values in Step 2, leading to a "locally aggregated" quantity, $\widetilde{p}(v) = h(\{p(v') : v' \in N_v\})$.

There are various possible choices of h. In this paper, we set h the same as that adopted in Zhang *et al.* (2011), i.e. the median filter. The neighborhood N_v for each voxel is set as the nearest 7 voxels shown in the left panel of Figure 3 in our paper. Utilizing $\{\widetilde{p}(v) : v \in \mathcal{V}\}$, the test outcomes based upon a significant rule can be summarized in Table 1.

Table 1: *Using* $\{\widetilde{p}(v) : v \in \mathcal{V}\}$*, outcomes of tests based upon a significant rule*

	$H_0(v)$ retained	$H_0(v)$ rejected	Total
$H_0(v)$ true	$\widetilde{U}$	$\widetilde{V}$	N_0
$H_0(v)$ false	$\widetilde{T}$	$\widetilde{S}$	N_1
Total	$\widetilde{W}$	$\widetilde{R}$	N

The definition for the false discovery rate (FDR) based upon the "locally aggregated" p-values becomes

$$\mathrm{FDR}_L(t) = E\Big\{\frac{\widetilde{V}(t)}{\widetilde{R}(t) \vee 1}\Big\}, \tag{A.5.1}$$

where $\widetilde{V}(t)$ and $\widetilde{R}(t)$ are those corresponding quantities in Table 1 when "reject $H_0(v)$ if $\widetilde{p}(v) \le t$" is used as the significant rule. $\mathrm{FDR}_L(t)$ is then estimated by

$$\widehat{\mathrm{FDR}}_L(t) = \frac{\widetilde{W}(\lambda)\widehat{G}(t)}{\{\widetilde{R}(t) \vee 1\}\{1 - \widehat{G}(\lambda)\}},$$

where $\widehat{G}(\cdot)$, a non-parametric estimate of the sample distribution of $\{\widetilde{p}(v) : v \in \mathcal{V}_0\}$, is given by

$$\widehat{G}(t) = \begin{cases} \frac{\sum_{v\in\mathcal{V}} \mathrm{I}\{\widetilde{p}(v)\ge(1-t)\}}{2\sum_{v\in\mathcal{V}} \mathrm{I}\{\widetilde{p}(v)>0.5)+\sum_{v\in\mathcal{V}} \mathrm{I}\{\widetilde{p}(v)=0.5)}, & \text{if } 0 \le t \le 0.5, \\ 1 - \frac{\sum_{v\in\mathcal{V}} \mathrm{I}\{\widetilde{p}(v)>t)}{2\sum_{v\in\mathcal{V}} \mathrm{I}\{\widetilde{p}(v)>0.5)+\sum_{v\in\mathcal{V}} \mathrm{I}\{\widetilde{p}(v)=0.5)}, & \text{if } 0.5 < t \le 1. \end{cases}$$

For a given FDR control level α, the $H_0(v)$ is rejected if its associated $\widetilde{p}(v)$ is smaller than or equal to

$$t_\alpha(\widehat{\mathrm{FDR}}_L) \equiv \sup\{0 \le t \le 1 : \widehat{\mathrm{FDR}}_L(t) \le \alpha\}.$$

6 Summary of Constructing $\mathbb{K}$

We summarize the computational procedures for constructing $\mathbb{K}$ for each brain voxel based on DWI data as follows.

Step 1. Use the DWI data collected from the experiment to estimate the DT for every brain voxel based on model (2.1) [or (2.2)] in the paper.

Step 2. Decompose the estimated DT above in every brain voxel by Schur decomposition to obtain the ordered eigenvalue estimates.

Step 3. For each voxel,

- use the estimated DTI data in Step 1 and apply the adaptive neighborhood selection procedure in Section 4.2 to choose n (=25) neighboring voxels;
- then collect the eigenvalues (estimated in Step 2) in these selected neighboring voxels to compute $\boldsymbol{U}_n$ as defined in (4.4).

Step 4. Use the five subsequent steps described after equation (4.4) to construct $\widehat{\boldsymbol{\theta}}_{\boldsymbol{U}_n}$ and $\mathbb{K}$.

7 More Simulation Results

In this section, we provide some more simulation results to support the findings in our paper submitted to the *Annals of Applied Statistics*. When simulating the DWI data, all the settings are adopted the same as those in Section 6.2 in our paper, unless otherwise stated. We adopt SNR=10 throughout this section.

- Case 1: let $\phi_0^*(v) = 1500$. We simulate a reference signal and $r = 12$ diffusion attenuated signals in each v as follows.

$$\phi_i(v) = \sqrt{[\phi_0^*(v)\exp\{-b\boldsymbol{g}_i^T\mathbf{D}^*(v)\boldsymbol{g}_i\} + \epsilon_{i,x}(v)]^2 + \epsilon_{i,y}^2(v)},$$

  where $i = 0, \ldots, 12$, $\epsilon_{i,x}(v)$ and $\epsilon_{i,y}(v)$ are simulated as `iid` $N(0, \{\phi_0^*(v)/\text{SNR}\}^2)$ random variables.

  In this case, the simulated $\phi_0(v)$ are homogeneous over the entire brain, sharing the same variance as all attenuated signals. We observe from Figures 1 and 2 that under such an optimal setting, Tabelow-FA-threshold approach works as well as our proposed $\mathbb{K}$-FDR$_L$ approach. The performances of all four methods are ordered as FA-threshold$<$ $\mathbb{K}$-FDR$<$ Tabelow-FA-threshold $\approx$ $\mathbb{K}$-FDR$_L$.

- Case 2: let $\phi_0^*(v) = 1500$. We simulate $r = 12$ diffusion attenuated signals in each v as follows.

$$\phi_i(v) = \sqrt{[\phi_0^*(v)\exp\{-b\boldsymbol{g}_i^T\mathbf{D}^*(v)\boldsymbol{g}_i\} + \epsilon_{i,x}(v)]^2 + \epsilon_{i,y}^2(v)},$$

  where $i = 1, \ldots, 12$, $\epsilon_{i,x}(v)$ and $\epsilon_{i,y}(v)$ are simulated as `iid` $N(0, \{\phi_0^*(v)/\text{SNR}\}^2)$ random variables. The reference signals are simulated as

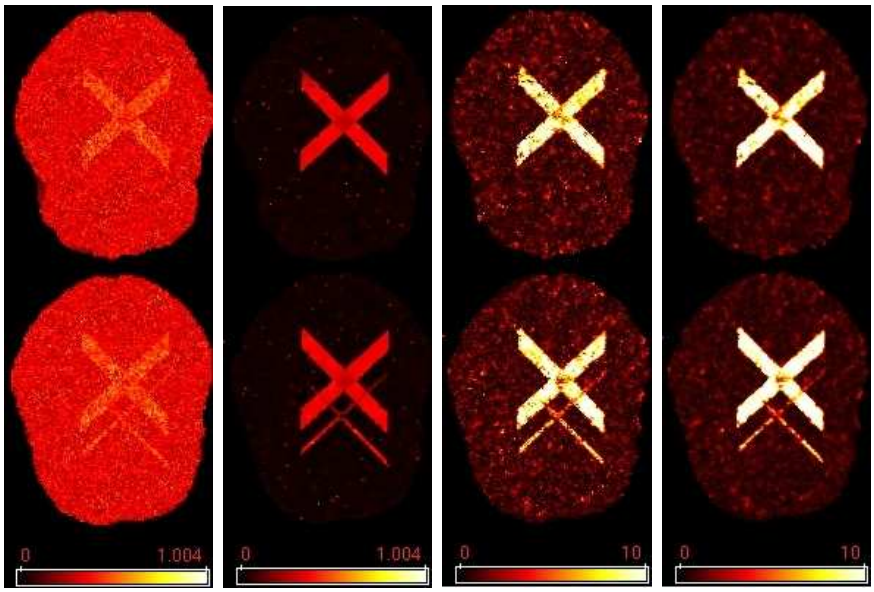

Figure 1: *Case 1. The caption is the same as that of Figure 7 in the paper.*

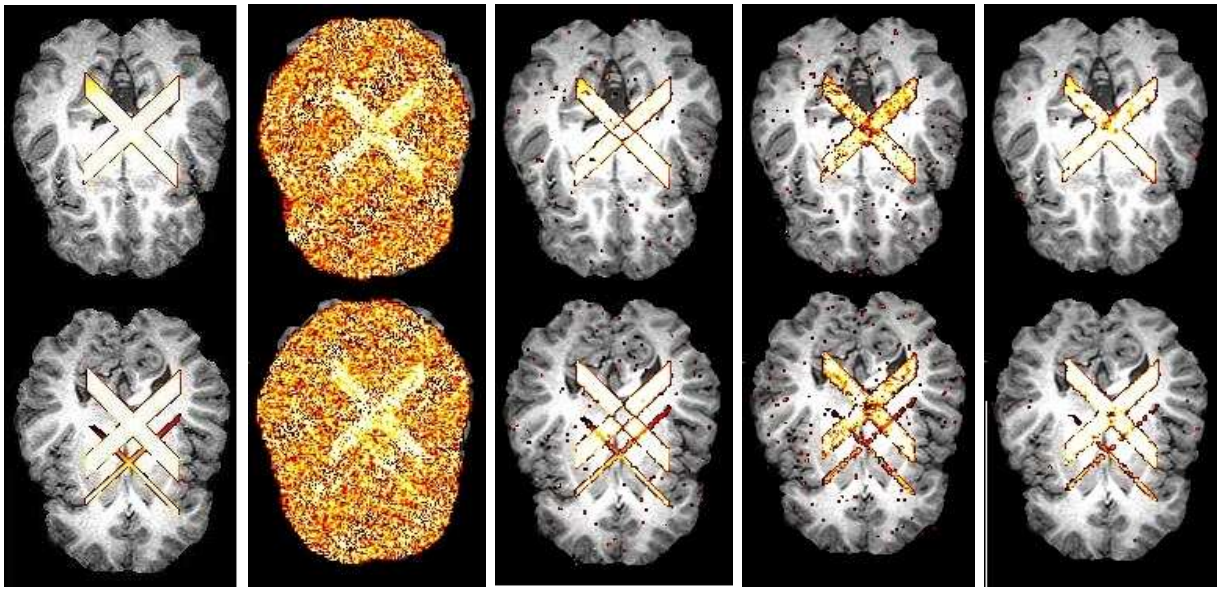

Figure 2: *Case 1. The caption is the same as that of Figure 8 in the paper.*

$$\phi_0(v) = \sqrt{\{\phi_0^*(v) + \epsilon_{0,x}(v)\}^2 + \epsilon_{0,y}^2(v)},$$

where $\epsilon_{0,x}(v)$ and $\epsilon_{0,y}(v)$ are simulated as `iid` $N(0, \{\phi_0^*(v)/20\}^2)$ random variables. In other words, the signal to noise ratio for the reference signals is simulated as 20, which is twice as large as that of attenuated signals over the entire brain.

Overview Figures 3 and 4, we can clearly see that the performances of all four methods are now ordered as FA-threshold< Tabelow-FA-threshold < $\mathbb{K}$-FDR < $\mathbb{K}$-FDR$_L$.

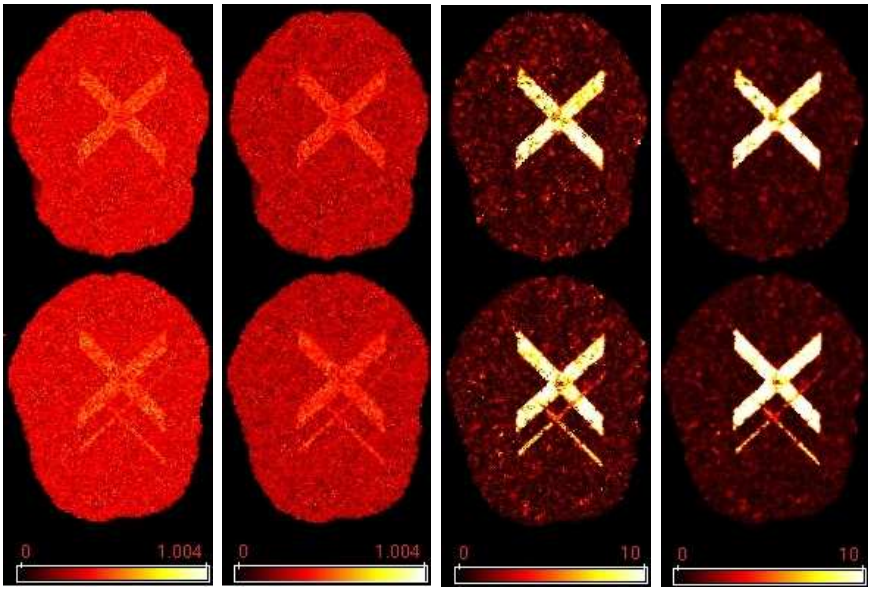

Figure 3: *Case 2. The caption is the same as that of Figure 7 in the paper.*

- Case 3. All the settings are adopted the same as those for Case 2, except that $\epsilon_{0,x}(v)$ and $\epsilon_{0,y}(v)$ are simulated as `iid` $N(0, \{\phi_0^*(v)/5\}^2)$ random variables, i.e. the signal to noise ratio of the reference signals is half as large as that of attenuated signals over the entire brain.

  Inspection of Figure 6, Tabelow-FA-threshold method "over-corrected" the bias components in the eigenvalue estimates, indicated by the loss of identification in the crossed areas of fiber tracts. As such, we would suggest FA-threshold< $\mathbb{K}$-FDR < Tabelow-FA-threshold < $\mathbb{K}$-FDR$_L$.

References

[1] Durrett, R. (1991). *Probability: Theory and Examples*, Duxbury Press.

[2] Shao, J. (2003). *Mathematical Statistics* (2nd ed.), Springer.

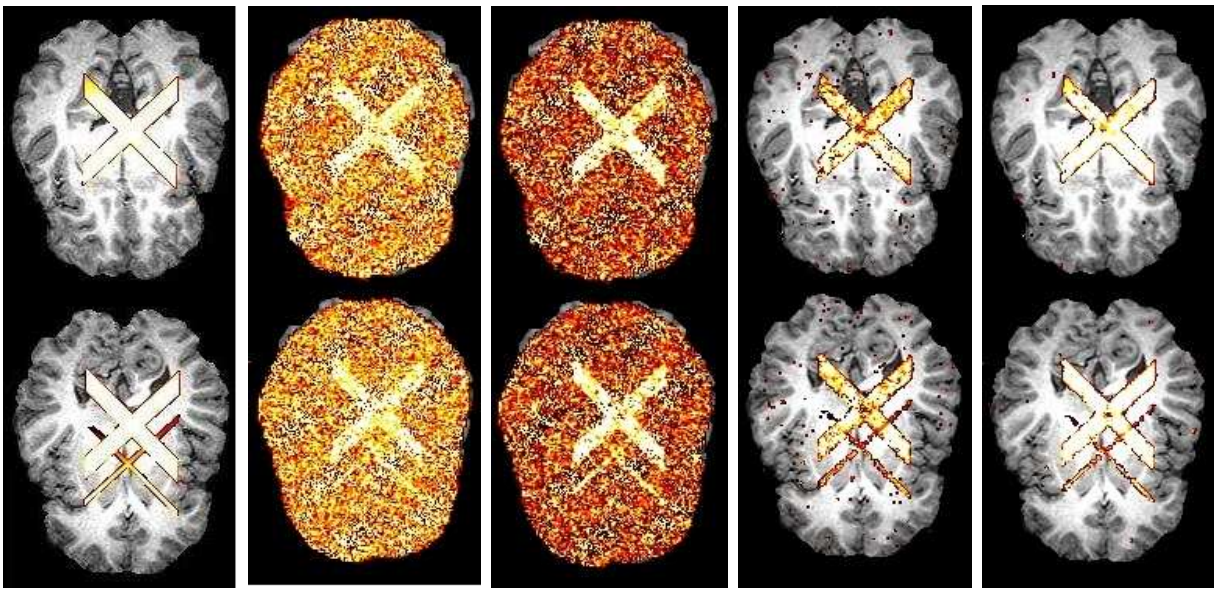

Figure 4: *Case 2. The caption is the same as that of Figure 8 in the paper.*

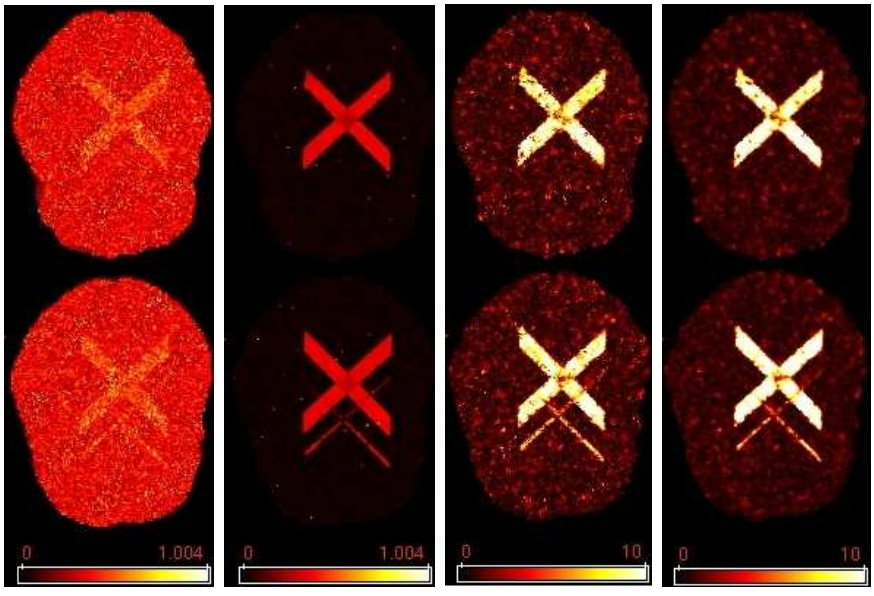

Figure 5: *Case 3. The caption is the same as that of Figure 7 in the paper.*

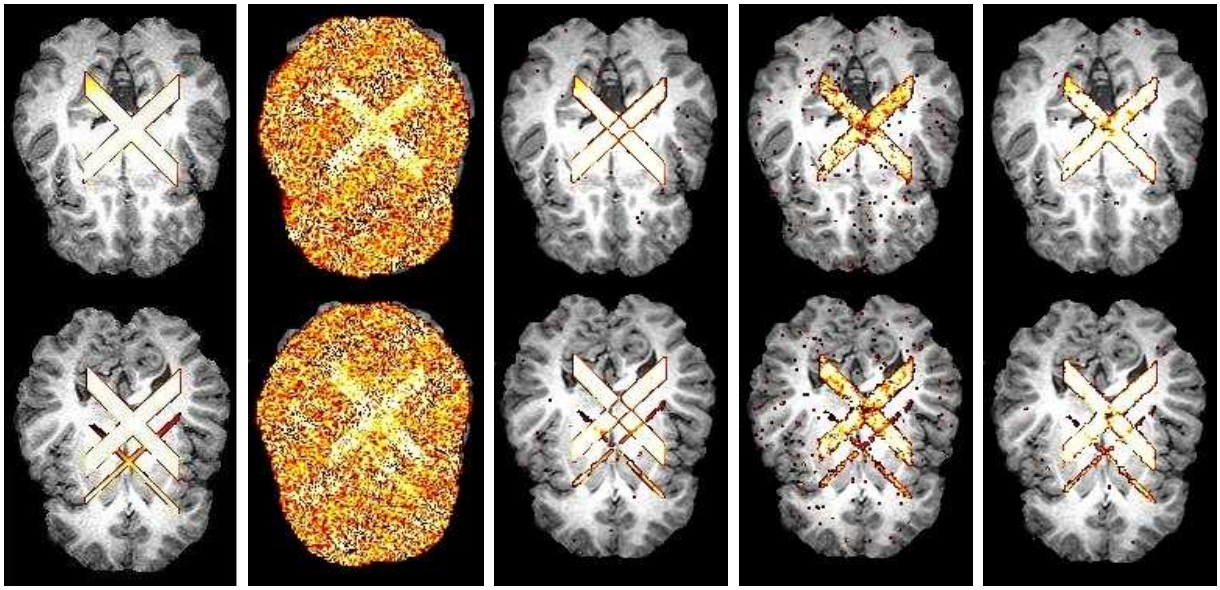

Figure 6: *Case 3. The caption is the same as that of Figure 8 in the paper.*

[3] van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press.

[4] Zhang, C.M., Fan, J. and Yu, T. (2011). "Multiple testing via FDR_L for large scale imaging data," *Annals of Statistics*, 39, 613-642.