

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

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

April 27, 2012

Abstract

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.

Key words and phrases: Brain Tissue; Diffusion Tensor; Eigenvalue; Fiber Tracts; Local Test; Quantitative Scalar.

Running title: Local Tests for Anisotropic Areas.

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 of \mathbf{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.

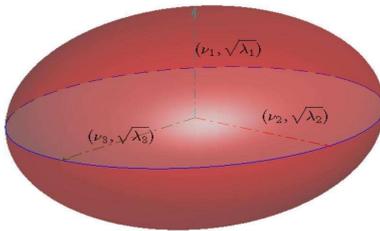


Figure 1: *Ellipsoid representation of a DT.*

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 methods 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 is widely adopted by neuroscientists in the past decade. Nonetheless, FA and RA are essentially defined as functions of the eigenvalues of 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 non-degenerate. 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 neighborhood 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 i th 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 *et al.* 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 i th 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 *et al.* (1994), in which the following multivariate linear and non-linear regression models are proposed. For a single-voxel,

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

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

where $i = 1, \dots, r$, $\mathbf{x}_i = (g_{i,1}^2, g_{i,2}^2, g_{i,3}^2, 2g_{i,1}g_{i,2}, 2g_{i,1}g_{i,3}, 2g_{i,2}g_{i,3})^T$, 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 *et al.* (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 *et al.* (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, i.e. 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. In contrast, for voxels in brain areas without fiber tracts, i.e. 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 non-degenerate, 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 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.

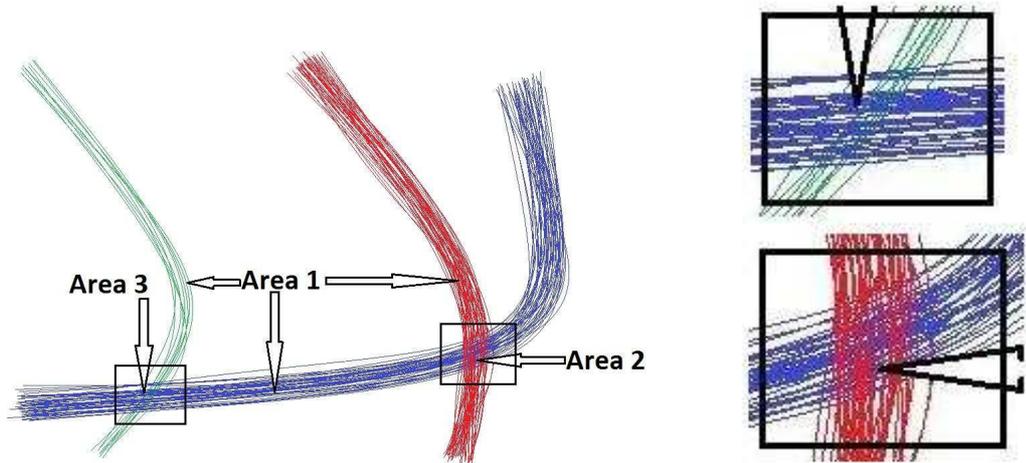


Figure 2: An illustrative example of fiber tracts.

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

$$\text{FA} = \sqrt{\frac{3}{2} \frac{\sum_{k=1}^3 \{\lambda_{(k)} - \bar{\lambda}_{(\cdot)}\}^2}{\sum_{k=1}^3 \lambda_{(k)}^2}}, \quad \text{RA} = \frac{3}{\sqrt{2}} \frac{\sqrt{\sum_{k=1}^3 \{\lambda_{(k)} - \bar{\lambda}_{(\cdot)}\}^2}}{\sum_{k=1}^3 \lambda_{(k)}},$$

where $\bar{\lambda}_{(\cdot)} = \{\lambda_{(1)} + \lambda_{(2)} + \lambda_{(3)}\}/3$, 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, $\text{RA} = 0$ and $\text{FA} = 0$ in isotropic voxels (i.e. $\lambda_{(1)} = \lambda_{(2)} = \lambda_{(3)}$), whereas $\text{RA} = \sqrt{3}$ and $\text{FA} = 1$ 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 non-vanishing 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 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 rectangular shown in the leftmost panel of Figure 3. The tracked fibers passing through the ROI using distinct FA thresholding criteria, namely $\text{FA} > c$ with $c = 0.3, 0.45$, and 0.6 , are displayed in the rest 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, is not well supported in theory. In other words, there is lack of criterion to choose the threshold for FA, and justify the goodness of the tracking results.

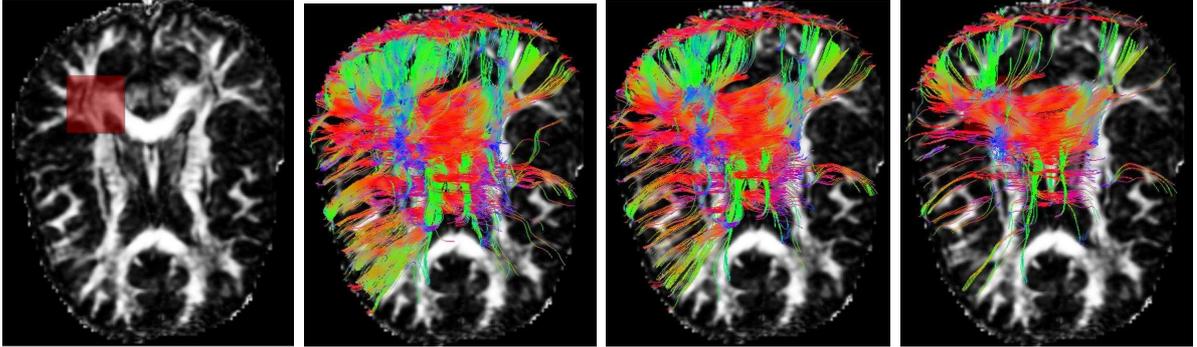


Figure 3: *Fiber tracking results by distinct FA thresholds. From left to right: FA map with a ROI highlighted; constructed fibers passing through the ROI by FA thresholds 0.3, 0.45, and 0.6 respectively.*

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, i.e. $r \rightarrow \infty$. 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 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 $v \in \{1, \dots, N\}$, 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} , i.e. $\lambda_k \in \{\lambda_{(k)} : k = 1, 2, 3\}$, 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 $E(\epsilon_k) = 0$ such that

$$\lambda_k = \lambda_{(k)}^* + \epsilon_k. \quad (3.1)$$

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 , i.e. $\lambda_{1,k} = \lambda_k$.

Similar in spirit to the random complete block design, we model $E(\lambda_{j,k})$, the expected eigenvalue of the j th 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 j th neighboring voxel, denoted by β_j . That is $E(\lambda_{j,k}) = \lambda_{(k)}^* + \beta_j$, which leads to the additive model,

$$\lambda_{j,k} = \lambda_{(k)}^* + \beta_j + \epsilon_{j,k}, \quad (3.2)$$

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; $\{\epsilon_{j,k} : k = 1, 2, 3\}_{j=1}^n$ 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, i.e. $\lambda_{(3)}^* = \lambda_{(2)}^* = \lambda_{(1)}^*$, are then rely on the fact that $\lambda_{j,(k)} = \lambda_{(1)}^* + \beta_j + \epsilon_{j,(k)}$.

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 non-degenerate $\lambda_{(3)}^* > \lambda_{(2)}^* > \lambda_{(1)}^*$, 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,

$$H_0 : \lambda_{(3)}^* = \lambda_{(2)}^* = \lambda_{(1)}^* \quad \text{versus} \quad H_1 : \lambda_{(3)}^* > \lambda_{(2)}^* \text{ or } \lambda_{(2)}^* > \lambda_{(1)}^*, \quad (4.1)$$

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

$$H_0 : A\boldsymbol{\lambda}^* = \mathbf{0} \quad \text{versus} \quad H_1 : A(l, \cdot)\boldsymbol{\lambda}^* > 0 \text{ for } l = 1, \text{ or } 2, \dots, \text{ or } \mu, \quad (4.2)$$

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

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \quad (4.3)$$

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

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

where $\mathbf{U}_n = (A\bar{\boldsymbol{\lambda}}/\sqrt{\text{MSE}}) \cdot \sqrt{\bar{S}^2}$; $\bar{\boldsymbol{\lambda}} = (\bar{\lambda}_{\cdot,(3)}, \bar{\lambda}_{\cdot,(2)}, \bar{\lambda}_{\cdot,(1)})^T$; $\bar{\lambda}_{\cdot,(k)} = \sum_{j=1}^n \lambda_{j,(k)}/n$, $k = 1, 2, 3$; \bar{S}^2 is the mean of S_j^2 over the neighboring voxels of v ; $S_j^2 = \sum_{k=1}^3 \{\lambda_{j,(k)} - \bar{\lambda}_{j,(\cdot)}\}^2/2$, $j = 1, \dots, n$; $\text{MSE} = \sum_{j=1}^n \sum_{k=1}^3 \{\lambda_{j,(k)} - \bar{\lambda}_{\cdot,(k)} - \bar{\lambda}_{j,(\cdot)} + \bar{\lambda}_{\cdot,(\cdot)}\}^2 / \{2(n-1)\}$; let $\hat{\mathcal{V}}_0$ be an estimate of \mathcal{V}_0 , the set of isotropic voxels over the entire brain; $\hat{\boldsymbol{\theta}}_{\mathbf{U}_n}$ denotes the sample median of \mathbf{U}_n over $\hat{\mathcal{V}}_0$; $\hat{\boldsymbol{\Sigma}}$ denotes the sample covariance of $\sqrt{n}\mathbf{U}_n$ over $\hat{\mathcal{V}}_0$. $\hat{\mathcal{V}}_0$ 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 $\chi_{\mu, 1-\alpha}^2$ 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 s th iteration ($s = 1, 2, \dots$), 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 $\mathbb{K} < \chi_{\mu;1-\alpha}^2$.
- v. Repeat steps iii and iv until $\widehat{\boldsymbol{\theta}}_{\mathcal{U}_n}$ converges.

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

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

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

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

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

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

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

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 $A\bar{\boldsymbol{\lambda}}/\sqrt{\text{MSE}}$ 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 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 $\sqrt{S^2}$ 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 non-fiber areas, the assumption that the collected eigenvalues in the selected neighboring voxels follow model (3.2) may not be appropriate. In this case, $\sqrt{\text{MSE}}$ tends to inflate, and consequently, the statistical power of the first part $A\bar{\lambda}/\sqrt{\text{MSE}}$ is limited. The second part then counteracts the effect of $\sqrt{\text{MSE}}$, and therefore is particularly useful when the resolution of the DTI data is limited.
- Furthermore, $\hat{\theta}_{\mathbf{U}_n}$, 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 non-degenerate, 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 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} (for example, choose the neighborhood as a fixed cube). However, for experiments with limited resolution, 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 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 , $l = 1, \dots, xyz$, as

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

where $d_p(v, v_l)$ denotes the physical distance between voxels v and v_l ; $d_{\mathbf{D}}(\mathbf{D}(v), \mathbf{D}(v_l)) = \sqrt{\text{trace}[\{\mathbf{D}(v) - \mathbf{D}(v_l)\}^2]}$ is a measurement of the diffusion similarity between tensors in voxels v and v_l (Alexander *et al.* 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.

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. 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 $n \rightarrow \infty$,*

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

We would like to point out that the construction of \mathbb{K} and the theoretical derivations of Theorem 1 are non-trivial and challenging. Following the discussion in Section 3, for each voxel v , we postulate that there exist unobservable one-to-one correspondences, i.e. $\{(\lambda_k, \lambda_{(k)}^*) : k = 1, 2, 3\}$, 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 $\lambda_{j,(k)} = \lambda_{(1)}^* + \beta_j + \epsilon_{j,(k)}$. As such, the bias components of eigenvalue estimates in the neighboring voxels of v are carried by $\epsilon_{j,(k)}$, whose effects in our test statistic are counteracted by $\widehat{\boldsymbol{\theta}}_{U_n}$ 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 $n \rightarrow \infty$,*

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

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

Theorem 2 shows that as long as $\mathbf{g}(\tilde{\mathbf{a}}(v)) \neq \mathbf{g}(\mathbf{b})$, $M > 0$ 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; $\tilde{\mathbf{a}}(v) = (E\{2S_1^2(v)\}, E\{\lambda_{1,(3)}(v)\}, E\{\lambda_{1,(2)}(v)\}, E\{\lambda_{1,(1)}(v)\})^T$; $\mathbf{b} = (E(2S_\epsilon^2), E\{\epsilon_{1,(3)}(1)\}, E\{\epsilon_{1,(2)}(1)\}, E\{\epsilon_{1,(1)}(1)\})^T$; $S_1^2(v)$ and S_ϵ^2 are respectively the sample variances of $\{\lambda_{1,(k)}(v), k = 1, 2, 3\}$ and $\{\epsilon_{1,(k)}(1), k = 1, 2, 3\}$. Thus, under 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 $\epsilon_{j,k}$ has a symmetric distribution about 0, i.e. $\epsilon_{j,k}$ has the same distribution as $-\epsilon_{j,k}$, 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 *et al.* 2004; Zhang *et al.* 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 *et al.* (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.

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 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 the results in Section 6.3: when $n = 25$, the sampling distribution of \mathbb{K} agrees reasonably well with the χ^2 distribution.

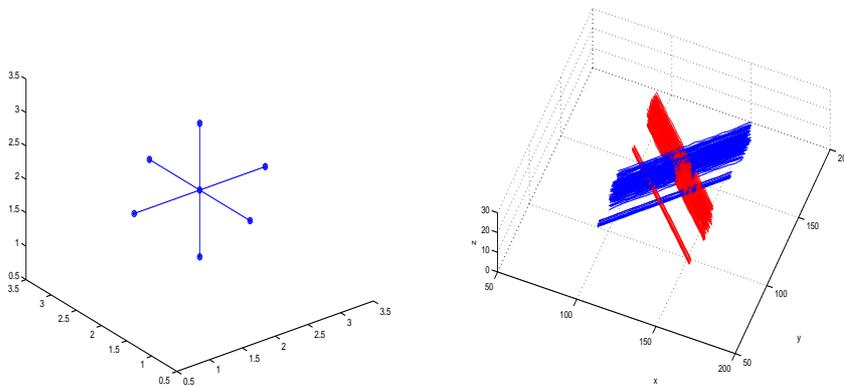


Figure 4: *Left panel - neighbors of a voxel used in the FDR_L procedure. Right panel - geometry of the simulated brain.*

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* (non-fiber areas), *prolate* (single blue and red bundles), *oblate* (intersected areas of blue bundles), and *non-degenerate* (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 ($i = 1, \dots, 12$) in each $v [= (v_x, v_y, v_z)]$ as

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

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 $\{\mathbf{g}_i : i = 1, \dots, 12\}$ 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 $\epsilon_{i,x}(v)$ and $\epsilon_{i,y}(v)$ are simulated as independent and normally distributed with variance σ^2 , which is varied to provide signal to noise ratios (SNRs), where $\text{SNR} = \phi_0^*(v)/\sigma$. We examine four distinct SNRs, $\{5, 10, 15, 20\}$, each corresponding to one set of the simulated DWI data.

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 non-fiber areas. The QQ plots of the (1st to 99th) percentiles of these \mathbb{K} 's against those of the χ_μ^2 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.

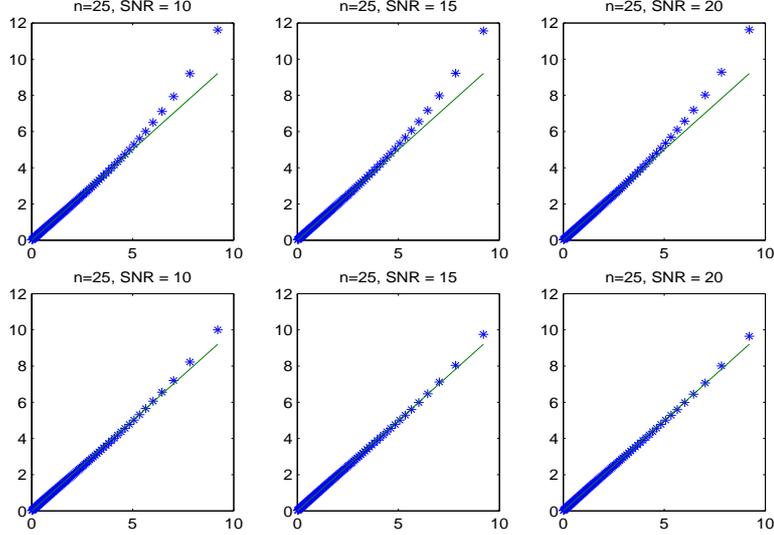


Figure 5: Empirical percentiles of \mathbb{K} (y -axis) versus percentiles of χ_μ^2 distribution (x -axis). Top panels - candidate cubic $5 \times 5 \times 3$, $n = 25$. Bottom panels - candidate cubic $11 \times 11 \times 3$, $n = 81$. From left to right panels: SNR = 10, 15, and 20. Solid line - the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

6.4 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 $\{T_1, \dots, T_N\}$ be the set of statistics for all voxels over the entire brain, where T_v for any voxel $v \in \{1, \dots, N\}$ 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 \tilde{p} -value if the \mathbb{K} -FDR_L approach is used, where \tilde{p} stands for the median smoothed p -value of \mathbb{K} (Zhang *et al.* 2011). For any given threshold t , if we classify a voxel v as anisotropic based on $T_v \in R(t)$, where $R(t) = \{T_v : T_v \geq t, v = 1, \dots, N\}$ when T_v is the FA or Smooth-FA value; $R(t) = \{T_v : T_v \leq t, v = 1, \dots, N\}$ when T_v is the p - or \tilde{p} -value,

then,

$$\begin{aligned} \text{sensitivity: } se(t) &\equiv \frac{\sum_{v=1}^N I\{T_v \in R(t), H_1 \text{ is true}\}}{|\mathcal{V}_0|} \\ \text{specificity: } sp(t) &\equiv \frac{\sum_{v=1}^N I\{T_v \notin R(t), H_0 \text{ is true}\}}{N - |\mathcal{V}_0|}, \end{aligned}$$

where $|\mathcal{V}_0|$ is the number of isotropic voxels over the entire brain. The ROC curve is then constructed as the 2D curve $(se(t), 1 - sp(t))$ 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 \tilde{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 \tilde{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 \tilde{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.

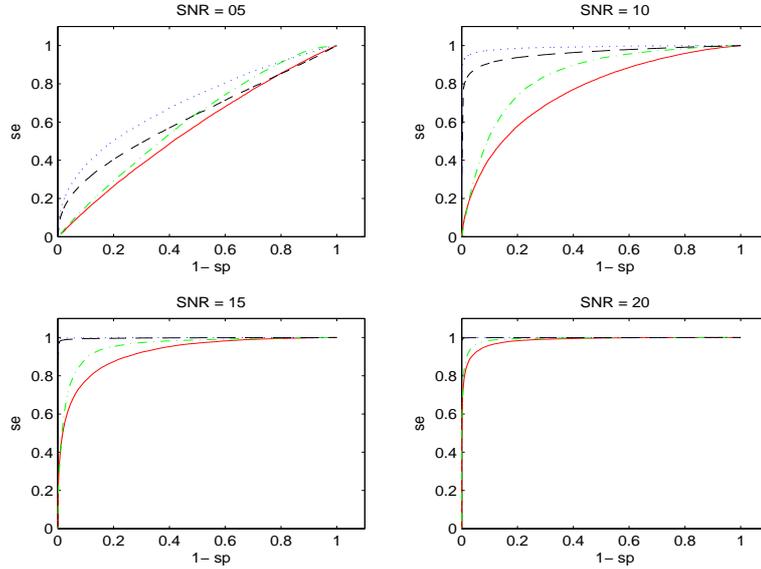


Figure 6: ROC curves for SNR = 5, 10, 15, and 20. Dashed black line - p -value of \mathbb{K} ; dotted blue line - \tilde{p} -value of \mathbb{K} ; solid red line - FA; dash-dotted green line - Smooth-FA.

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 $\text{SNR} = 10$. Those of $\text{SNR} = 5, 15$, 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 $\mathbb{K}\text{-FDR}_L$. The results of sensitivity and specificity by all four methods are displayed in Table 1. Clearly, $\mathbb{K}\text{-FDR}_L$ maximizes both the sensitivity (0.8845) and the specificity (0.9982) in this example. $\mathbb{K}\text{-FDR}$ achieves similar specificity (0.9957) but slightly smaller sensitivity (0.7522) compared to $\mathbb{K}\text{-FDR}_L$. However, in order to yield comparable sensitivity (0.8845) as $\mathbb{K}\text{-FDR}_L$, FA-threshold and Smooth-FA-threshold approaches produce much lower specificities (0.4012 for FA; 0.6242 for Smooth-FA).

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
$\mathbb{K}\text{-FDR}$	0.7522	0.9957
$\mathbb{K}\text{-FDR}_L$	0.8845	0.9982

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 left-most panel of Figure 8. Figure 7 displays the colormaps of the FA, Smooth-FA, $-\log(p)$, and $-\log(\tilde{p})$, with all $-\log(p)$ and $-\log(\tilde{p})$ 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 , $\mathbb{K}\text{-FDR}$, and $\mathbb{K}\text{-FDR}_L$ for $\text{SNR} = 10$.

As clearly evidenced in Figures 7 and 8, $\mathbb{K}\text{-FDR}$ and $\mathbb{K}\text{-FDR}_L$ 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 non-significant 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 tracking results. It has been seen in the second to right and rightmost panels of Figures 7 and 8 that the detected results by $\mathbb{K}\text{-FDR}$ and $\mathbb{K}\text{-FDR}_L$ well capture the primary features of the simulated anisotropic areas. Compared

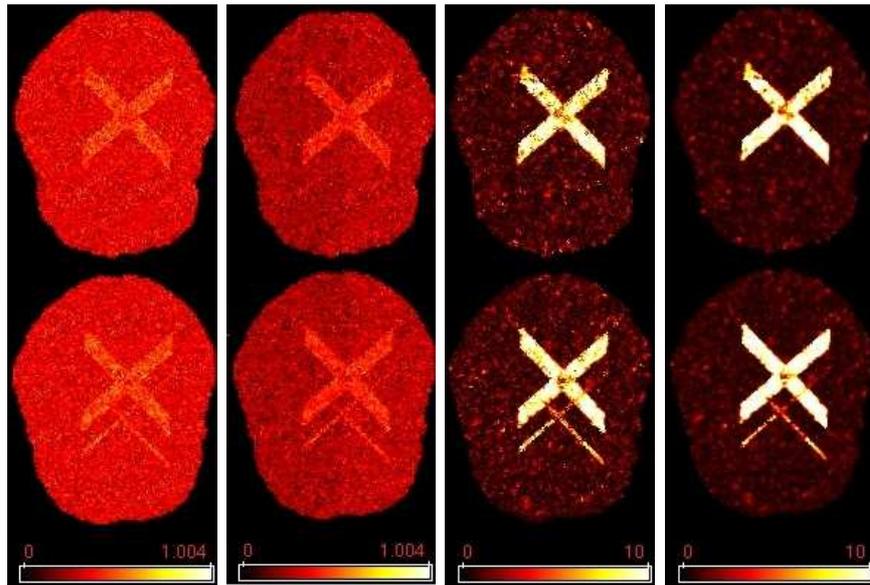


Figure 7: Comparison of color maps for simulated data set. From left to right: FA; Smooth-FA; $-\log(p)$ by \mathbb{K} ; $-\log(\tilde{p})$ by \mathbb{K} . SNR = 10.

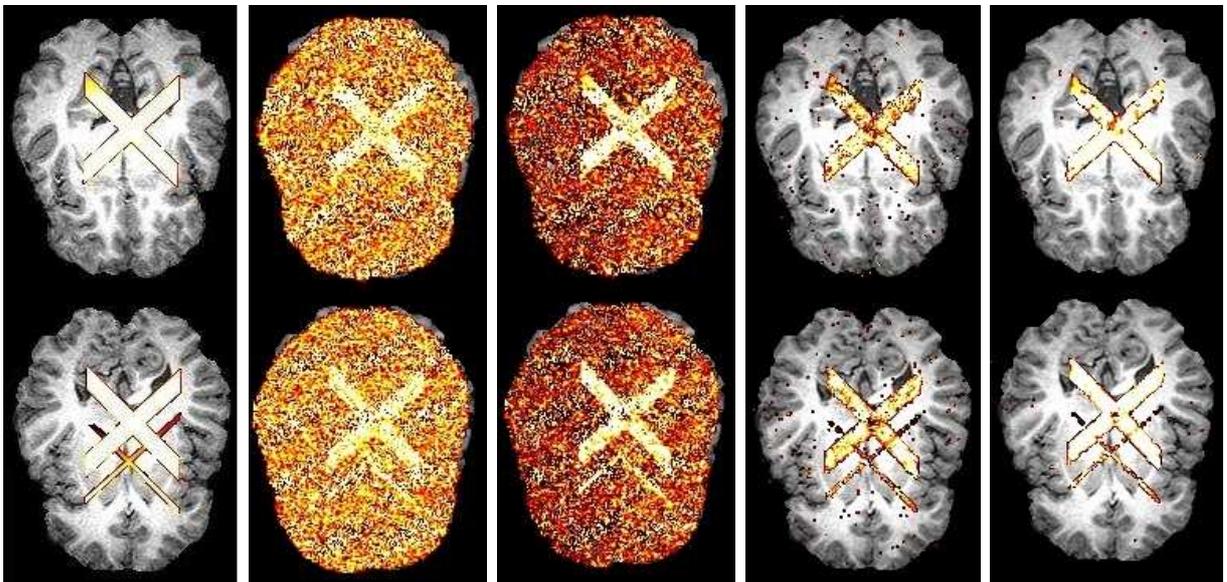


Figure 8: Comparison of brain anisotropic areas discovered for the simulated data set. From left to right: simulated brain anisotropic areas; FA > 0.3003; Smooth-FA > 0.2803; \mathbb{K} -FDR; \mathbb{K} -FDR_L. SNR = 10. The control level is 0.01.

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 $0.9375\text{mm} \times 0.9375\text{mm}$, field of view = 24cm^2 , matrix = 256×256 ; 30 axial slices are acquired along the z -axis, slice thickness = 3mm . A single reference image at $b = 0$ and 12 diffusion-attenuated images with non-collinear directions of diffusion gradients at $b = 1000\text{s/mm}^2$ 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 other four subjects display similar scenarios and are omitted.

The acquired data set contains $256 \times 256 \times 30 = 1966080$ voxels with 400309 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 colormaps of FA, Smooth-FA, $-\log(p)$, and $-\log(\tilde{p})$ 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), $\text{FA} > 0.35$ and $\text{Smooth-FA} > 0.35$ detect more scattered tiny areas than \mathbb{K} -FDR and \mathbb{K} -FDR_L. Those are highly likely to be faulty findings. Furthermore, overview the areas located close to the top of the highlighted areas, $\text{FA} > 0.35$ and $\text{Smooth-FA} > 0.35$ 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 non-fiber areas. In the meanwhile, as illustrated in Section 2.3, since the number of gradients r in each voxel is small, FA-threshold and Smooth-FA-threshold approaches can not clearly infer and control the error rate of the identification, and therefore lack the rigorous criterion in selecting the appropriate 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 colormaps of FA, Smooth-FA, $-\log(p)$, and $-\log(\tilde{p})$. The colormaps of both $-\log(p)$ and $-\log(\tilde{p})$ 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.

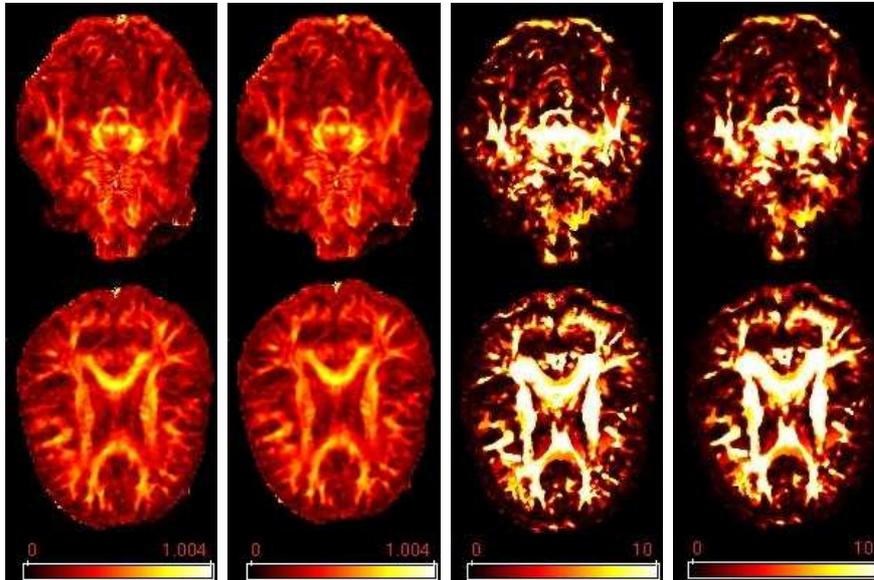


Figure 9: Comparison of color maps for the real brain data set. From left to right: FA; Smooth-FA; $-\log(p)$ by \mathbb{K} ; $-\log(\tilde{p})$ by \mathbb{K} .

To further illustrate the efficacy of our methods in reducing the scattered faulty findings, we summarize the frequency of identified isolated anisotropic voxels over each subject’s brain for all four methods in Table 2. In particular, denote by $\mathcal{N}(v) = \{v' = (v'_x, v'_y, v'_z) : |v'_i - v_i| \leq 1, \text{ for all } i = x, y, \text{ and } z\}$ 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, i.e. 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 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

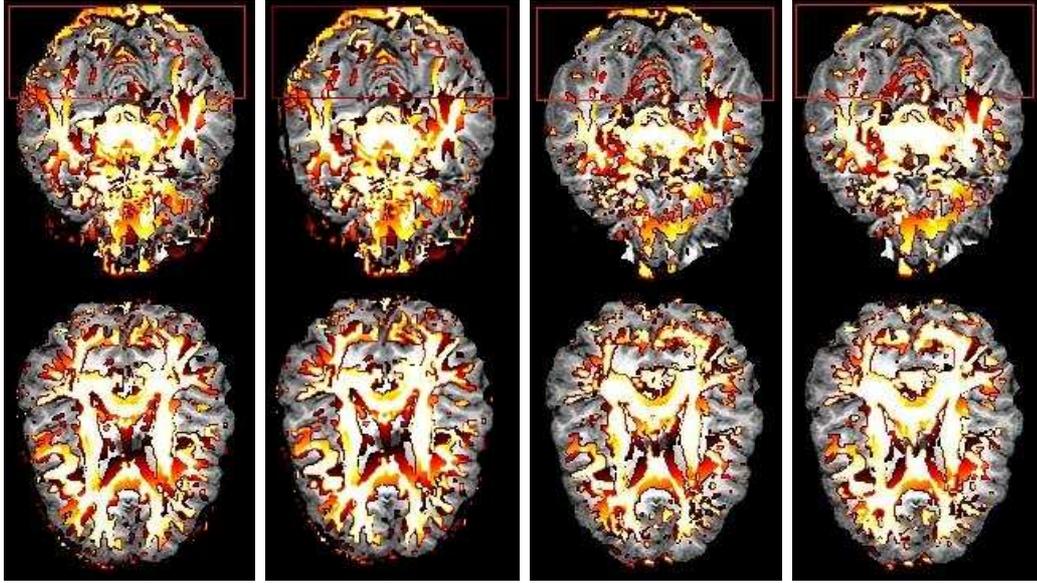


Figure 10: Comparison of brain anisotropic areas discovered for the real brain data set. From left to right: $FA > 0.35$; $Smooth-FA > 0.35$; $-\log(p)$ by \mathbb{K} ; $-\log(\hat{p})$ by \mathbb{K} . The control level is 0.01.



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

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

Methods	Subject					Subject				
	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

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 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 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, i.e. 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 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, 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 relies 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). 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 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 the 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.
- The 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.

Supplemental Document

Supplement: proofs for Theorems 1 and 2, a short summary of FDR_L procedure (Zhang *et al.* 2011), steps for constructing \mathbb{K} based on DWI data, and some more simulation results. The supplementary material would be provided at this location.

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

References

- [1] Alexander, D., Gee, J. and Bajcsy, R. (1999). "Similarity measures for matching diffusion tensor images," *Proceedings of the British Machine Vision Conference*, 1999, 93-102.
- [2] Anderson, A.W. (2001). "Theoretical analysis of the effects of noise on diffusion tensor imaging," *Magnetic Resonance in Medicine*, 46, 1174-1188.
- [3] Basser, P.J., Mattiello, J. and LeBihan, D. (1994). "Estimation of the effective self-diffusion tensor from NMR spin echo," *Journal of Magnetic Resonance, Series B*, 103, 247-254.
- [4] Basser, P.J. and Pierpaoli, C. (1996). "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI," *Journal of Magnetic Resonance, Series B*, 111, 209-219.
- [5] Benjamini, Y. and Hochberg, Y. (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B*, 57, 289-300.
- [6] Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S. and Woolrich, M.W. (2007). "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?" *NeuroImage*, 34, 144-155.
- [7] Chang, L.C., Jones, D.K., and Pierpaoli, C. (2005). "RESTORE: robust estimation of tensors by outlier rejection," *Magnetic Resonance in Medicine*, 53, 1088-1095.

- [8] Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H. and Raichle, M.E. (1999). "Tracking neuronal fiber pathways in the living human brain," *Proceedings of the National Academy of Sciences of the United States of America*, 96, 10422-10427.
- [9] 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," *Nature Neuroscience*, 8, 519-526.
- [10] Douek, P., Turner, R., Pekar, J., Patronas, N. and Le, Bihan D. (1991). "MR color mapping of myelin fiber orientation," *Journal of Computer Assisted Tomography*, 15, 923-929.
- [11] 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.
- [12] Grigis, A., Renard, F., Noblet, V., Heinrich, C., Heitz, F., Armspach, J. (2011). "A new high order tensor decomposition: Application to reorientation," *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, 258-261.
- [13] Hasan, K.M., Parker, D.L. and Alexander, A.L. (2001). "Comparison of gradient encoding schemes for diffusion-tensor MRI," *Journal of Magnetic Resonance Imaging*, 13,769-780.
- [14] Heim, S., Fahrmeir, L., Eilers, P.H.C. and Marx, B.D. (2007). "3d space-varying coefficient models with application to diffusion tensor imaging," *Computational Statistics and Data Analysis*, 51, 6212-6228.
- [15] Henkelman, R. M. (1985), "Measurement of Signal Intensities in the Presence of Noise in MR Images," *Medical Physics*, 12, 232-233.
- [16] Jones, D.K. (2003). "Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI," *Magnetic Resonance in Medicine*, 49, 7-12.
- [17] Johansen-Berg, H. and Behrens, T.E.J. (2009). *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy*, Elsevier, Inc.
- [18] Lazar, M. and Alexander, A.L. (2003). "An error analysis of white matter tractography methods: synthetic diffusion tensor field simulations," *NeuroImage*, 20, 1140-1153.
- [19] Li, Y.M., Zhu H.T., Shen D.G., Lin W.L., Gilmore J., and Ibrahim J.G. (2011). "Multiscale adaptive regression models for neuroimaging data," *Journal of the Royal Statistical Society, Series B*, 73, 559-578.
- [20] Mangin, J.F., Poupon, C., Clark C., Le Bihan, D. and Bloch, I. (2002). "Distortion correction and robust tensor estimation for MR diffusion imaging," *Medical Image Analysis*, 6, 191-198.

- [21] 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.
- [22] O'Donnell, L.J. and Westin, C.F. (2007). "Automatic tractography segmentation using a high-dimensional white matter atlas," *IEEE Transactions on Medical Imaging*, 26(11), 1562-1575.
- [23] Pepe, M.S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford: Oxford University Press.
- [24] Pierpaoli, C., and Basser, P.J. (1996). "Toward a quantitative assessment of diffusion anisotropy," *Magnetic Resonance in Medicine*, 36, 893-906.
- [25] 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.
- [26] Polzehl, J. and Tabelow K. (2009) "Structural Adaptive Smoothing in Diffusion Tensor Imaging: The R Package dti," *Journal of Statistical Software*, 31, issue 9.
- [27] Salvador, R., Pena, A., Menon, D.K., Carpenter, T.A., Pickard, J.D. and Bullmore, E.T. (2005). "Formal characterization and extension of the linearized diffusion tensor model," *Human Brain Mapping*, 24, 144-155.
- [28] 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.
- [29] Storey, J.D. (2002). "A direct approach to false discovery rates," *Journal of the Royal Statistical Society, Series B*, 64, 479-498.
- [30] 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," *Journal of the Royal Statistical Society, Series B*, 66, Part 1, 187-205.
- [31] Tabelow, K., Polzehl, J., Spokoiny, V. and Voss, H.U. (2008). "Diffusion tensor imaging: structural adaptive smoothing," *NeuroImage*, 39, 1763 - 1773.
- [32] 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," *Magnetic Resonance in Medicine*, 31, 154-163.
- [33] Xu, D., Mori, S., Solaiyappan, M., van Zijl, P. C. and Davatzikos, C. (2002). "A framework for callosal fiber distribution analysis," *NeuroImage*, 17, 1131-1143.
- [34] Yu, T. (2009). *Local Tests for Detecting Human Brain Isotropy-Anisotropy Areas on DT-MRI*, PhD thesis, University of Wisconsin-Madison, Wisconsin.
- [35] 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.

- [36] 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.
- [37] Zhu, H., Zhang H., Ibrahim, J. and Peterson, B. (2007). “Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data,” *Journal of the American Statistical Association*, 102, 1085-1102.