Harvard University

Harvard University Biostatistics Working Paper Series

Year 2009 *Paper* 96

Group Comparison of Eigenvalues and Eigenvectors of Diffusion Tensors

Armin Schwartzman* Robert F. Dougherty†

Jonathan E. Taylor‡

*Harvard School of Public Health and Dana Farber Cancer Institute, armins@hsph.harvard.edu

†Stanford University, bobd@stanford.edu

‡Stanford University, jonathan.taylor@stanford.edu

Group Comparison of Eigenvalues and Eigenvectors of Diffusion Tensors

Armin Schwartzman, Robert F. Dougherty, Jonathan E. Taylor

February 13, 2009

Authors' footnote

Armin Schwartzman is Assistant Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115 (E-mail: armins@hsph.harvard.edu). Robert F. Dougherty is Senior Research Scientist, Department of Psychology, Stanford University, Stanford, CA 94305 (E-mail: bobd@stanford.edu). Jonathan E. Taylor is Associate Professor, Department of Statistics, Stanford University, Stanford, CA 94305 (E-mail: jonathan.taylor@stanford.edu).
This work was partially supported by NIH grant EY015000 and NSF grant DMS-0405970.

Abstract

Diffusion tensor images (DTI) differ from most medical images in that values at each voxel are not scalars, but 3×3 symmetric positive definite matrices called diffusion tensors (DTs). The anatomical characteristics of the tissue at each voxel are reflected by the DT eigenvalues and eigenvectors. In this article we consider the problem of testing whether two groups of DTIs are equal at each voxel in terms of the DT's eigenvalues, eigenvectors, or both. Because eigen-decompositions are highly nonlinear, existing likelihood ratio statistics (LRTs) for testing differences in the set of eigenvalues or the frame of eigenvectors assume an orthogonally invariant covariance structure between the DT entries. While retaining the form of the LRTs, we derive new approximations to their true distributions when the covariance between the DT entries is arbitrary and possibly different between the two groups. The approximate distributions are those of other similar LRT statistics computed at the tangent space to the parameter manifold at the true value of the parameter, but plugging in an estimate for the point of application of the tangent space. The resulting distributions, which are weighted sums of χ^2s, are further approximated by scaled χ^2 distributions by matching the first two moments. For application to DTI data, a log transformation that converts positive definite matrices into real symmetric matrices is appropriate but not necessary. Voxelwise application of the test statistics leads to a multiple testing problem, which is solved by false discovery rate inference. The above methods are illustrated in a DTI group comparison of boys vs. girls.

Keywords: diffusion tensor imaging, random matrix, likelihood ratio test, manifold-valued data, Satterthwaite approximation, multiple testing

1 Introduction

Diffusion Tensor Imaging (DTI) is a novel modality of magnetic resonance imaging that allows visualization in-vivo of the internal anatomical structure of the brain's white matter (Basser and Pierpaoli, 1996; LeBihan et al., 2001). DTI images are 3D rectangular arrays that contain at every voxel (volume pixel) not a scalar but a 3×3 symmetric positive definite matrix, also called diffusion tensor (DT). The DT describes the local pattern of water diffusion and can be thought of as the covariance matrix of a 3D Gaussian distribution that models the Brownian motion of the water molecules in the voxel. However, the displacement of the molecules is not observable. Instead, the DT is reconstructed from measurements of the diffusion coefficient in at least 6 directions in space. The DT serves as a proxy for local anatomical structure. The DT's eigenvalues measure diffusivity and are indicative of the type of tissue and its health, while the eigenvectors relate to the spatial orientation of the underlying neural fibers.

A common statistical problem in DTI group studies is to find regions of the brain whose anatomical characteristics differ between two groups of subjects. The analysis typically consists of normalizing the images to a common template so that each voxel corresponds to the same anatomical structure in all the images, and then applying two-sample tests at each voxel. Because of familiarity with univariate statistics, analyses are often restricted to scalar quantities derived from the DT such as trace and fractional anisotropy (FA), both functions of the DT's eigenvalues and related respectively to the total diffusivity and the degree of anisotropy within a voxel. Current multivariate approaches provide inference for the DT as a single multivariate unit, but not for its eigenvalues and eigenvectors (Basser and Pajevic, 2003; Schwartzman, 2006; Whitcher et al., 2007), or focus on the principal diffusion direction (PDD), the eigenvector corresponding to the largest eigenvalue, failing to take into account the role of

the eigenvalues in the estimation (Jones et al., 2002; Wu et al., 2004; Schwartzman et al., 2005, 2008a).

In this paper, we introduce a new practical multivariate approach to making inferences about the DT's eigenstructure in group DTI studies. Specifically, given two groups of observed DTs, we consider a test of whether the two group means have the same set of eigenvalues while treating the eigenvectors as nuisance parameters, and a test of whether they have the same frame of eigenvectors while treating the eigenvalues as nuisance parameters. These can be thought of as multivariate extensions of trace/FA tests and principal diffusion direction tests, respectively. These tests are non-trivial as they involve parameter sets that are submanifolds of the set of symmetric matrices times itself.

Because eigenvalues and eigenvectors are highly nonlinear functions of the data, likelihood ratio test (LRT) statistics for these cases have been derived elsewhere assuming an orthogonally invariant (OI) covariance structure between the matrix entries and assuming that the covariances in the two groups are the same (Schwartzman et al., 2008b). The eigenvector test assuming an OI covariance has been useful for identifying white matter tracts affected by necrosis in a brain cancer survivor (Rauschecker et al., 2009). However, to be used more generally in practical DTI analysis, an arbitrary covariance structure between the tensor elements must be allowed. In this paper we retain the form of the LRT statistics derived under the OI assumption but derive new approximations to their distributions when the covariance between the DT entries is arbitrary and possibly different in the two groups. The approximate distributions are those of other LRT statistics obtained when the parameter manifolds are replaced by their tangent spaces at the estimated value of the parameter. These distributions are weighted mixtures of χ^2's, which are then further approximated by scaled χ^2 distributions using the method of moments. The approximations are asymptotically valid as the number of observations per voxel increases.

In the DTI literature, it has been proposed to analyze DTI data after a matrix log transformation (Arsigny et al., 2005; Fletcher and Joshi, 2007; Schwartzman, 2006). The matrix log, computed by taking the log of the eigenvalues and keeping the eigenvectors intact, maps the observed positive definite matrices to the set of symmetric matrices. Technically, this is required for Gaussian modeling as the symmetric matrices form a vector space while the positive definite matrices do not. In the statistics literature, this idea was proposed earlier for modeling covariance matrices (Leonard and Hsu, 1992; Chiu et al., 1996). Whether the log transform should be applied in practice to DTI data is a subject of current debate (Whitcher et al., 2007). The methods developed in this paper are applicable in either case because the matrix log affects only the eigenvalues in a one-to-one fashion, so the various hypotheses about eigenvalues and eigenvectors can be equivalently stated in both domains. In addition, because the inference is based on large-sample asymptotics, the positive definite restrictions are of no practical importance. For simplicity, we assume that the data lives in the set of 3×3 symmetric matrices, regardless of whether the eigenvalues have been log-transformed prior to the analysis.

Application of the test statistics at each voxel leads to a massive multiple testing problem involving hundres of thousands of tests. To address this problem we use false discovery rate (FDR) inference, now commonly used in the analysis of neuroimaging data (Genovese et al., 2002; Logan and Rowe, 2004; Schwartzman et al., 2009).

We demonstrate the above methods in data from an observational study of brain anatomy in children (Dougherty et al., 2007). Here we focus on a cross-sectional analysis whose goal is to find brain regions that differ significantly between boys and girls at age 10. The most important test for this data analysis is the two-sample eigenvector test, which reveals differences in neural fiber orientation between boys and girls mostly in the posterior left hemisphere.

2 Theory

2.1 A Gaussian signal-plus-noise model

Let $\mathcal{S}_p$ denote the set of $p \times p$ symmetric matrices ($p \geq 2$). For observed $Y \in \mathcal{S}_p$, let

$$Y = M + Z, \tag{1}$$

where $M \in \mathcal{S}_p$ (capital μ) is a mean parameter and $Z \in \mathcal{S}_p$ has zero mean. In DTI, $p = 3$, Y is the observed diffusion tensor (or its matrix logarithm) at a particular voxel, M represents the population mean at that voxel, and Z incorporates both inter-subject variability and measurement noise at that voxel. The effect of measurement noise (Zhu et al., 2007) is not given separate treatment here as it may be assumed to be negligible in comparison to the anatomical variability between subjects.

An arbitrary covariance between the entries of Y may be specified via the operator

$$\text{vecd}(Y) = (\text{diag}(Y)', \sqrt{2}\text{offdiag}(Y)')',$$

defined as a column vector of length $q = p(p+1)/2$ where $\text{diag}(Y)$ is a $p \times 1$ vector containing the diagonal entries of Y and $\text{offdiag}(Y)$ is a $(q-p) \times 1$ vector containing the off-diagonal entries of Y copied from below the diagonal columnwise (or above the diagonal rowwise). In DTI, $\text{vecd}(Y) = (Y_{11}, Y_{22}, Y_{33}, \sqrt{2}Y_{12}, \sqrt{2}Y_{13}, \sqrt{2}Y_{23})'$ and $q = 6$. The operator $\text{vecd}(\cdot)$ was chosen for the convenient property that it converts the Frobenius norm of symmetric matrices into the Euclidean norm for vectors, that is

$$\|Y\|^2 = \text{tr}(Y^2) = \text{vecd}(Y)'\text{vecd}(Y) = \|\text{vecd}(Y)\|^2. \tag{2}$$

If Z is non-degenerate Gaussian, the distribution of Y, denoted $Y \sim N_{pp}(M, \Sigma)$, may be called symmetric-matrix-variate normal and written as the multivariate normal density

$$Y \sim N_{pp}(M, \Sigma) \quad \Leftrightarrow \quad \text{vecd}(Y) \sim N_q(\text{vecd}(M), \Sigma) \tag{3}$$

where Σ is a $q \times q$ positive definite matrix.

Schwartzman et al. (2008b) derive various LRTs for the mean parameter M when M is restricted to subsets of $\mathcal{S}_p$ defined in terms of eigenvalues and eigenvectors of M. The LRT statistics derived there assume that the covariance Σ in (3) is OI. Briefly, Σ is called OI if the distribution of $Z = Y - M$ is the same as that of QZQ' for any $Q \in \mathcal{O}(p)$, the set of $p \times p$ orthogonal matrices ($Q'Q = QQ' = I_p$). An OI Σ can be parametrized by two scalar parameters, one that controls the dependence between the diagonal entries and a global variance parameter (Mallows, 1961). A special case for independent entries is the spherical covariance, corresponding to the distribution known in random matrix theory as Gaussian orthogonal ensemble (GOE) (Mehta, 1991).

While restrictive for data, the OI assumption has the advantage that it avoids the need to vectorize the data matrices and therefore allows obtaining closed-form solutions to the maximum likelihood estimates (MLEs) and LRTs for many hypotheses defined in terms of eigenvalues and eigenvectors of M. Moreover, Schwartzman et al. (2008b) show that in many cases this covariance structure is the only one that allows derivation of closed-form expressions that do not depend on estimates of the covariance parameters. In this article we are interested in applying the LRT statistics derived by Schwartzman et al. (2008b) when the true covariance Σ in the data is not necessarily OI.

2.2 Two-sample tests

Let $Y_1, \ldots, Y_{n_1}$ and $Y_{n_1+1}, \ldots, Y_n$, $n = n_1 + n_2$, be two independent i.i.d. samples from $N_{pp}(M_1, \Sigma_1)$ and $N_{pp}(M_2, \Sigma_2)$, respectively. We consider three tests.

The first test, referred to hereafter as the eigenvalue test, is a test of whether M_1 and M_2 have the same eigenvalues, while the eigenvectors are unrestricted and treated as nuisance parameters. The test is $H_0 : (M_1, M_2) \in \mathcal{M}_{2,D}$ vs. $H_A : (M_1, M_2) \notin \mathcal{M}_{2,D}$, where

$$\mathcal{M}_{2,D} = \{(M_1, M_2) : M_1 = U_1 D U_1', M_2 = U_2 D U_2'\} \tag{4}$$

for unspecified $D \in \mathcal{D}_p$, the set of $p \times p$ diagonal matrices, and $U_1, U_2 \in \mathcal{O}_p$. For simplicity, D is assumed to have p distinct eigenvalues.

The second test, referred to hereafter as the eigenvector test, is a test of whether M_1 and M_2 have the same eigenvectors, when the eigenvalues are treated as nuisance parameters and assumed equal between the two populations. The test is $H_0 : M_1 = M_2$ vs. $H_A : (M_1, M_2) \in \mathcal{M}_{2,D}$, where $\mathcal{M}_{2,D}$ is given by (4).

For comparison, we also consider the test $H_0 : M_1 = M_2$ vs. $H_A : M_1 \neq M_2$, referred to hereafter as the full matrix test, where no particular attention is paid to the eigenstructure. In this case, the Y_i's can be seen as multivariate samples of dimension q from model (3). Since Σ_1 and Σ_2 are not assumed equal, this is a multivariate Behrens-Fisher problem. Defining the group averages $\bar{Y}_1 = (1/n_1)\sum_{i=1}^{n_1} Y_i$ and $\bar{Y}_2 = (1/n_2)\sum_{i=n_1+1}^{n} Y_i$, Hotelling's T^2 statistic is given by

$$T^2 = \mathbf{d}' S^{-1} \mathbf{d}, \qquad \mathbf{d} = \mathrm{vecd}(\bar{Y}_1 - \bar{Y}_2), \qquad S = \frac{S_1}{n_1} + \frac{S_2}{n_2} \tag{5}$$

where

$$\begin{aligned} S_1 &= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \mathrm{vecd}(Y_i - \bar{Y}_1)\mathrm{vecd}(Y_i - \bar{Y}_1)' \\ S_2 &= \frac{1}{n_2 - 1} \sum_{i=n_1+1}^{n} \mathrm{vecd}(Y_i - \bar{Y}_2)\mathrm{vecd}(Y_i - \bar{Y}_2)' \end{aligned} \tag{6}$$

are the group sample covariances. The null distribution of (5) may be approximated by redefining

$$T_F = \frac{f - q + 1}{qf} T^2 \underset{H_0}{\dot{\sim}} F(q, f - q + 1) \tag{7}$$

where f is the approximate number of degrees of freedom of Yao (1965) given by

$$\frac{1}{f} = \frac{1}{n_1 - 1}\left(\frac{\mathbf{d}' S^{-1} S_1 S^{-1} \mathbf{d}}{n_1 \mathbf{d}' S^{-1} \mathbf{d}}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{\mathbf{d}' S^{-1} S_2 S^{-1} \mathbf{d}}{n_2 \mathbf{d}' S^{-1} \mathbf{d}}\right)^2.$$

2.3 LRT statistics

Computation of the LRT statistic in both the eigenvalue and eigenvector tests involves estimation of the pair $\boldsymbol{M} = (M_1, M_2)$ under the hypothesis $\boldsymbol{M} \in \mathcal{M}_{2,D}$. The set $\mathcal{M}_{2,D}$

(4) is a curved submanifold of $\mathcal{S}_p \times \mathcal{S}_p$. Assuming $\Sigma_1 = \Sigma_2 = \Sigma$, Schwartzman et al. (2008b) showed that estimation of $\boldsymbol{M} \in \mathcal{M}_{2,D}$ is analytically tractable and separable from the estimation of Σ if and only if Σ is OI. In such case, the MLE of $\boldsymbol{M}$ is the same that would be obtained if Σ were assumed spherical with variance one, i.e. $\Sigma = I_q$. In other words, the MLE is the same as the least squares solution and is found by orthogonal projection.

Specifically, let $M_1 = U_1 D_1 U_1'$ and $M_2 = U_2 D_2 U_2'$, $\bar{Y}_1 = V_1 \Lambda_1 V_1'$, $\bar{Y}_2 = V_2 \Lambda_2 V_2'$ and $\bar{Y} = V \Lambda V'$ be eigendecompositions, all with eigenvalues in decreasing order. Assuming spherical covariances $\Sigma_1 = \Sigma_2 = I_q$, the MLE of $\boldsymbol{M} = (M_1, M_2)$ when $\boldsymbol{M} \in \mathcal{M}_{2,D}$ is given by

$$\hat{\boldsymbol{M}} = (\hat{M}_1, \hat{M}_2) = (\hat{U}_1 \hat{D} \hat{U}_1', \hat{U}_2 \hat{D} \hat{U}_2') \ \in \ \mathcal{M}_{2,D}, \tag{8}$$

where $\hat{D} = \bar{\Lambda} = (n_1 \Lambda_1 + n_2 \Lambda_2)/n$ and $\hat{U}_1$ and $\hat{U}_2$ are any matrices of the form $\hat{U}_1 = V_1 Q_1$ and $\hat{U}_2 = V_2 Q_2$, where Q_1 and Q_2 are diagonal matrices with diagonal entries equal to ± 1 (Schwartzman et al., 2008b, Theorem 5.1).

LRT statistics are constructed as follows. More generally, suppose $\mathcal{M}_0 \subset \mathcal{M}_A \subset \mathcal{S}_p \times \mathcal{S}_p$ are nested hypotheses about the pair $\boldsymbol{M} = (M_1, M_2)$. Assuming $\Sigma_1 = \Sigma_2 = I_q$, the LRT statistic, defined as minus twice the log ratio of the maximized likelihoods under $\mathcal{M}_A$ and $\mathcal{M}_0$ respectively, is given by

$$T = \|\bar{\boldsymbol{Y}} - \hat{\boldsymbol{M}}_0\|^2 - \|\bar{\boldsymbol{Y}} - \hat{\boldsymbol{M}}_A\|^2 \tag{9}$$

where $\bar{\boldsymbol{Y}} = (\bar{Y}_1, \bar{Y}_2)$, $\hat{\boldsymbol{M}}_0$ and $\hat{\boldsymbol{M}}_A$ are the MLEs of $\boldsymbol{M}$ under $\mathcal{M}_0$ and $\mathcal{M}_A$, respectively, and the norm $\|\cdot\|_2^2$ in $\mathcal{S}_p \times \mathcal{S}_p$ is defined in terms of the Frobenius norm (2) by

$$\|\bar{\boldsymbol{Y}} - \hat{\boldsymbol{M}}\|^2 = n_1 \|\bar{Y}_1 - \hat{M}_1\|^2 + n_2 \|\bar{Y}_2 - \hat{M}_2\|^2 \tag{10}$$

For the eigenvalue test, the MLE $\hat{\boldsymbol{M}}_0$ under $H_0 : (M_1, M_2) \in \mathcal{M}_{2,D}$ is equal to (8). Under $H_A : (M_1, M_2) \notin \mathcal{M}_{2,D}$, the MLE is $\hat{\boldsymbol{M}}_A = (\bar{Y}_1, \bar{Y}_2)$. Replacing in (9) gives that,

when $\Sigma_1 = \Sigma_2 = I_q$, the LRT statistic is

$$T_D = \frac{n_1 n_2}{n} \|\Lambda_1 - \Lambda_2\|^2 \tag{11}$$

and is asymptotically $\chi^2(p)$ under H_0 as $n_1, n_2 \to \infty$ (Schwartzman et al., 2008b, Corollary 5.1). Not surprisingly, the test statistic measures the distance between the eigenvalue matrices of the two group averages. The number of degrees of freedom corresponds to the fact that p eigenvalues are being tested.

Similarly, for the eigenvector test, the MLE under $H_0 : M_1 = M_2$ is $\hat{\boldsymbol{M}}_0 = (\bar{Y}, \bar{Y})$, where $\bar{Y} = (n_1 \bar{Y}_1 + n_2 \bar{Y}_2)/n$. Under $H_A : (M_1, M_2) \in \mathcal{M}_{2,D}$, the MLE $\hat{\boldsymbol{M}}_A$ is equal to (8). Replacing in (9) gives that, when $\Sigma_1 = \Sigma_2 = I_q$, the LRT statistic is

$$T_U = \frac{2 n_1 n_2}{n} \big[\mathrm{tr}(\Lambda_1 \Lambda_2) - \mathrm{tr}(\bar{Y}_1 \bar{Y}_2) \big] \tag{12}$$

and is asymptotically $\chi^2(q-p)$ under H_0 as $n_1, n_2 \to \infty$ (Schwartzman et al., 2008b, Corollary 5.2). The number of degrees of freedom in this case corresponds to the fact that the set of eigenvectors being tested has dimension $q-p$. The functional form of the test statistic (12) is interesting. When the eigenvectors of $\bar{Y}_1$ and $\bar{Y}_2$ are equal, the test statistic is equal to zero. As the angles between the eigenvectors of $\bar{Y}_1$ and $\bar{Y}_2$ increase, with the eigenvalues remaining constant, the inner product $\mathrm{tr}(\bar{Y}_1 \bar{Y}_2)$ decreases, thus increasing the value of the test statistic.

2.4 Distributions under arbitrary covariance

The null distributions given above of the eigenvalue and eigenvector LRT statistics (11) and (12) are asymptotic and valid only under the assumption that $\Sigma_1 = \Sigma_2 = I_q$. The goal of this section is to derive approximate asymptotic null distributions for those LRT statistics when the true Σ_1 and Σ_2 are arbitrary. This task presents two difficulties. First, the LRT statistics are highly nonlinear functions of the data, a consequence of the curvature of the parameter sets involved. Second, the eigenvectors of the true covariance matrices may be oblique to the parameter sets.

Our general approach is as follows. For large n, $\bar{\boldsymbol{Y}} = (\bar{Y}_1, \bar{Y}_2)$ is close to $\boldsymbol{M} = (M_1, M_2)$ with high probability and the effect of the curvature of the parameter manifolds $\mathcal{M}_A$ and $\mathcal{M}_0$ near $\boldsymbol{M}$ becomes negligible. Therefore the distribution of the LRT statistic (9) is close to the distribution of another LRT statistic $T^*(\boldsymbol{M})$ computed on the tangent spaces to $\mathcal{M}_A$ and $\mathcal{M}_0$ at $\boldsymbol{M}$. Denote these tangent spaces by $\mathcal{T}_0(\boldsymbol{M})$ and $\mathcal{T}_A(\boldsymbol{M})$, respectively, and notice that $\mathcal{T}_0(\boldsymbol{M}) \subset \mathcal{T}_A(\boldsymbol{M})$. Similar to the derivation of (9), the tangent LRT statistic for testing $\mathcal{T}_0(\boldsymbol{M})$ vs. $\mathcal{T}_A(\boldsymbol{M})$ under the assumption of spherical covariance is

$$T^*(\boldsymbol{M}) = \|\bar{\boldsymbol{Y}} - \hat{\boldsymbol{M}}_0^*\|^2 - \|\bar{\boldsymbol{Y}} - \hat{\boldsymbol{M}}_A^*\|^2 = \|\hat{\boldsymbol{M}}_A^* - \hat{\boldsymbol{M}}_0^*\|^2 \tag{13}$$

where $\hat{\boldsymbol{M}}_0^*$ and $\hat{\boldsymbol{M}}_A^*$ are the MLEs of $\bar{\boldsymbol{Y}}$ on $\mathcal{T}_0(\boldsymbol{M})$ and $\mathcal{T}_A(\boldsymbol{M})$ respectively, obtained by orthogonal projection (see Figure 1). Define $\boldsymbol{Z} = \bar{\boldsymbol{Y}} - \boldsymbol{M}$. Since the segment $\hat{\boldsymbol{M}}_A^* - \hat{\boldsymbol{M}}_0^* \subset \mathcal{T}_A(\boldsymbol{M})$ is orthogonal to the segment $\hat{\boldsymbol{M}}_0^* - \boldsymbol{M} \subset \mathcal{T}_0(\boldsymbol{M}) \subset \mathcal{T}_A(\boldsymbol{M})$, the tangent LRT statistic (13) may also be written as

$$T^*(\boldsymbol{M}) = \left\| \mathrm{Proj}_{\mathcal{T}_0^\perp(\boldsymbol{M})} \left(\mathrm{Proj}_{\mathcal{T}_A(\boldsymbol{M})} \boldsymbol{Z} \right) \right\|^2 \tag{14}$$

where $\mathcal{T}_0^\perp(\boldsymbol{M})$ is the subspace orthogonal to $\mathcal{T}_0(\boldsymbol{M})$ and $\mathrm{Proj}(\cdot)$ denotes orthogonal projection.

Given that $\boldsymbol{Z} = (Z_1, Z_2) = (\bar{Y}_1 - M_1, \bar{Y}_2 - M_2)$, define the $2q \times 1$ multivariate normal vector

$$\mathrm{vecd}(\boldsymbol{Z}) = \begin{pmatrix} \mathrm{vecd}(Z_1) \\ \mathrm{vecd}(Z_2) \end{pmatrix} \sim N_{2q} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_1/n_1 & 0 \\ 0 & \Sigma_2/n_2 \end{pmatrix} \right] \tag{15}$$

Theorems 1 and 2 below show that $T^*(\boldsymbol{M})$ is a quadratic form of $\mathrm{vecd}(\boldsymbol{Z})$ and give expressions for the specific two-sample tests of interest. For simplified notation, define the elementary matrices $E_{ij} = (\mathbf{e}_i \mathbf{e}_j' + \mathbf{e}_j \mathbf{e}_i')/2$, where $\mathbf{e}_i$ and $\mathbf{e}_j$ denote column vectors with a single 1 in positions i and j, respectively. Specifically, if $i = j$, E_{ii} contains a single 1 in position (i, i), and if $i \neq j$, $E_{ij} = E_{ji}$ contains $1/2$ in positions (i, j) and (j, i), and zeros elsewhere.

Theorem 1. *Consider the test* $H_0 : (M_1, M_2) \in \mathcal{M}_{2,D}$ *vs.* $H_A : (M_1, M_2) \notin \mathcal{M}_{2,D}$, *where* $\mathcal{M}_{2,D}$ *is given by* (4). *The tangent LRT statistic* (14) *is*

$$T^*(\boldsymbol{M}) = \text{vecd}(\boldsymbol{Z})'\boldsymbol{\Omega}(\boldsymbol{M})\text{vecd}(\boldsymbol{Z}), \qquad \boldsymbol{\Omega}(\boldsymbol{M}) = \frac{n_1 n_2}{n}\sum_{i=1}^{p}\omega_i\omega_i' \tag{16}$$

where

$$\omega_i = \begin{pmatrix} \text{vecd}(U_1 E_{ii} U_1') \\ -\text{vecd}(U_2 E_{ii} U_2') \end{pmatrix}$$

Theorem 2. *Consider the test* $H_0 : M_1 = M_2$ *vs.* $H_A : (M_1, M_2) \in \mathcal{M}_{2,D}$, *where* $\mathcal{M}_{2,D}$ *is given by* (4). *The tangent LRT statistic* (14) *is*

$$T^*(\boldsymbol{M}) = \text{vecd}(\boldsymbol{Z})'\boldsymbol{\Omega}(\boldsymbol{M})\text{vecd}(\boldsymbol{Z}), \qquad \boldsymbol{\Omega}(\boldsymbol{M}) = \frac{n_1 n_2}{n}\sum_{i=1}^{p}\sum_{j=1}^{p}\omega_{ij}\omega_{ij}' \tag{17}$$

where

$$\omega_{ij} = \begin{pmatrix} \text{vecd}(E_{ij}) - J(U_1)h_{ij} \\ -\text{vecd}(E_{ij}) + J(U_2)h_{ij} \end{pmatrix}$$

$$J(U) = \begin{pmatrix} \text{vecd}(UE_{11}U') & \text{vecd}(UE_{22}U') & \cdots & \text{vecd}(UE_{pp}U') \end{pmatrix}$$

$$h_{ij} = \text{diag}(n_1 U_2' E_{ij} U_2 + n_2 U_1' E_{ij} U_1)/n$$

In both (16) and (17), $T^*(\boldsymbol{M})$ has the same form and its distribution is the same as that of a weighted sum of χ^2 variables, given by Proposition 1 below. Notice that since the distribution of $\text{vecd}(\boldsymbol{Z})$ does not depend on $\boldsymbol{M}$, the dependence on $\boldsymbol{M}$ is only through $\boldsymbol{\Omega}(\boldsymbol{M})$. Further, $\boldsymbol{\Omega}(\boldsymbol{M})$ is a function of the eigenvectors of M_1 and M_2 only, not their eigenvalues.

Proposition 1. *Let* $\boldsymbol{\Sigma}$ *be the covariance matrix of* $\text{vecd}(\boldsymbol{Z})$ *given by* (15) *and let* k *be the rank of* $\boldsymbol{\Omega}(\boldsymbol{M})$. *Then the distribution of* $T^*(\boldsymbol{M}) = \text{vecd}(\boldsymbol{Z})'\boldsymbol{\Omega}(\boldsymbol{M})\text{vecd}(\boldsymbol{Z})$ *is the same as that of* $\mathbf{z}'\boldsymbol{\Lambda}\mathbf{z} = \sum_{i=1}^{k}\lambda_i z_i^2$, *where* $\mathbf{z} \sim N_{2q}(0, I_{2q})$ *and* $\boldsymbol{\Lambda}$ *is a diagonal matrix containing the* k *nonzero eigenvalues of* $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Omega}(\boldsymbol{M})\boldsymbol{\Sigma}^{1/2}$.

The distribution of $T^*(\boldsymbol{M})$ can be further approximated as the distribution of a scaled χ^2 variable $a\chi^2_\nu$ matching their first two moments and then solving for $a, \nu > 0$,

an idea generally known as Welch-Satterthwaite approximation (Stuart and Ord, 1994; Kuonen, 1999; Casella and Berger, 2002). Applying this to the equivalent quadratic form $\mathbf{z}'\Lambda\mathbf{z}$ in Proposition 1 gives that the first two moments of $T^*(\boldsymbol{M})$ are the same as those of a $a\chi^2_\nu$ variable with

$$a = \frac{\mathrm{tr}(\boldsymbol{\Lambda}^2)}{\mathrm{tr}(\boldsymbol{\Lambda})} = \frac{\mathrm{tr}(\boldsymbol{\Sigma\Omega\Sigma\Omega})}{\mathrm{tr}(\boldsymbol{\Sigma\Omega})}, \qquad \nu = \frac{(\mathrm{tr}\boldsymbol{\Lambda})^2}{\mathrm{tr}(\boldsymbol{\Lambda}^2)} = \frac{(\mathrm{tr}\boldsymbol{\Sigma\Omega})^2}{\mathrm{tr}(\boldsymbol{\Sigma\Omega\Sigma\Omega})}.$$

In practice, both $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}(\boldsymbol{M})$ are unknown. We therefore use a plug-in estimate of a and ν where the required covariances Σ_1 and Σ_2 in (15) are replaced by the sample covariances (6) and $\boldsymbol{\Omega}(\boldsymbol{M})$ is replaced by $\boldsymbol{\Omega}(\hat{\boldsymbol{M}})$, computed according to (16) and (17) but using the empirical eigenvectors V_1 and V_2 of $\bar{Y}_1$ and $\bar{Y}_2$ instead of the true eigenvectors U_1 and U_2 of M_1 and M_2. Given T from (16) or (17), appropriate $\hat{a}$ and $\hat{\nu}$ are found as explained above and the p-value is computed as $1 - \hat{F}(T)$, where $\hat{F}$ is the cumulative distribution function of $\hat{a}\chi^2(\hat{\nu})$.

3 Numerical Studies

3.1 Distribution

In order to evaluate the accuracy of the approximate distributions derived above, the following simulations were conducted. Taking $p = 3$, two groups of $n_1 = n_2 = 50$ i.i.d. samples were generated according to model (3) with $M_0 = \mathrm{Diag}(1, 2, 4)$, i.e. diagonal with diagonal entries 1, 2, and 4, and $\Sigma_1 = \Sigma_2$ chosen at random as Wishart$(I_6, 6)$. The eigenvalue and eigenvector test statistics (11) and (12) were computed as well as their approximate distributions as described at the end of Section 2.4. The above procedure was repeated 10000 times, resulting in 10000 test statistic values and 10000 estimated values of a and ν in both cases.

Figure 2 shows the distribution of the p-values for both test statistics. The proximity of the distribution to the 45° line is an indicator of the goodness of the approxi-

mation. Similar results were obtained for various instances of the randomly generated Σ.

3.2 Power

The goal of the following simulations is to assess the power of the proposed test statistics against various alternative hypotheses. Two groups of $n_1 = n_2 = 50$ i.i.d. samples for $p = 3$ were generated, under the null hypothesis with $M_0 = \text{Diag}(1, 2, 4)$ and $\Sigma_1 = \Sigma_2$ chosen at random as $0.5\text{Wishart}(I_6, 6)$, and under each alternative as described below. Eigenvalue and eigenvector test statistics were computed as well as their approximate tangent space distributions. The above procedure was repeated 10000 times, resulting in 10000 test statistic values and 10000 estimated a and ν for each case. P-values and statistical power were computed.

The first set of alternatives is defined by changes in the eigenvalues of M_0 only. These are of the form $M = M_0 + \Delta M$, where ΔM is diagonal. Figure 3 summarizes the results for various alternatives in increasing order of difficulty. Alternatives are more difficult to detect when ΔM is smaller, but also when the eigenvalues of M are closer to one another. Interestingly, the eigenvalue test is more powerful than the full matrix test in the harder cases but not in the easiest case. By design, the eigenvector test has very little power here, but is still a little sensitive to changes in eigenvalues which, with noise, can be mistaken for rotations. It should be noted that the results are strongly dependent on the noise covariance Σ and how aligned are its eigenvectors with the direction of change of ΔM. Further simulations show that depending on Σ, the full matrix test may be more powerful than the eigenvalue test in all four cases above.

The second set of alternatives is defined by changes in the eigenvectors. These are

of the form $M = QM_0Q'$, where $Q \in \mathcal{O}_3$ is given by Rodrigues' rotation formula

$$Q = \exp\left[\theta \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}\right]$$

Here, $\mathbf{a} = (a_1, a_2, a_3)$ is a unit vector indicating the axis of rotation and θ is the rotation angle around that axis. Figure 4 summarizes the results for various alternatives in increasing order of difficulty. In all panels, the eigenvector test is more powerful than the full matrix test. In panel (a), the axis of rotation is oblique to all the eigenvectors of M. In panels (b) and (c), rotations are around the axis corresponding to the smallest eigenvalue and corresponding to the largest eigenvalue. Panel (d) is the same as panel (a) but with noise amplified by a factor of 4. By design, the eigenvalue test has no power here. In fact, it is sometimes counterproductive depending on the particular orientation of the eigenvectors of Σ. In general, the results depend on Σ and further simulations show that the full matrix test can be sometimes more powerful than the eigenvalue test.

4 Data Example

4.1 Data description

Our dataset concerns an observational study of brain anatomy in children (Dougherty et al., 2007). A total of 55 children were recruited for the study and brain scans were taken about once a year. The children were between the ages of 8 and 12 years old at the time of the first measurment. For the purposes of this paper, we extracted a subset of the data consisting of DTI images from only the ten-year-old children. This subset contains 34 brain images (one per child) and included 12 boys and 22 girls. The goal of the analisys is to find regions of anatomical difference between the boys and the girls.

Voxelwise analysis requires that each voxel corresponds to the same brain structure across all subjects. To this end, the DTI images were normalized to a common coordinate system based on a custom pediatric template (Dougherty et al., 2005). Normalization resulted in each image being a voxel array of size $81 \times 106 \times 76$ in a rectangular grid with $2 \times 2 \times 2$ mm regular spacings. The coordinate axes are defined so that the x, y and z axis point respectively to the right, front and top of the brain.

Neuroimaging investigators often restrict their analyses to a relevant subset of the brain, called search region or mask. Here we take a liberal approach and do not restrict the search region to any particular region inside the brain. A whole-brain mask that defines all voxels inside the brain was computed by segmenting each subject's spatially normalized brain and taking the intersection of all such segmentations. The final mask contains $m = 105822$ voxels.

4.2 Data analysis

In order to eliminate the positive definite constraints before analysis, the eigenvalues of the DT data were log-transformed voxelwise. Next, at each voxel, the two-sample test statistics (7), (11) and (12) were computed, as well as their approximate null distributions and corresponding p-values, as described by (7) and Section 2.4. For our data, $p = 3$, $q = p(p+1)/2 = 6$, $n_1 = 12$, and $n_2 = 22$.

Figure 5a shows the empirical distribution of the p-values for the three test types. The eigenvector test produces the highest number of low p-values, as indicated by the slope of the distribution function at zero. This suggests that the eigenvector test may be more powerful for finding differences between the two groups in this dataset.

Figure 6 shows maps of the p-values at a transverse slice 36 mm above the anterior commisure, an anatomical landmark commonly used for spatial normalization (Talairach convention). At this slice some of the most prominent differences between the two groups appear. Some regions are highlighted by the threee tests to different

extents, but the eigenvector test shows the strongest signal.

Determining significance of the above p-values is a multiple testing problem involving $N = 105822$ tests. For simplicity, a single threshold was obtained using the classical BH FDR procedure (Benjamini and Hochberg, 1995) at FDR level 0.02. Table 1 summarizes the results for the eigenvalue and eigenvector tests. Reported are the p-value thresholds and the fraction of discoveries R/m, where R denotes the number of rejected null hypotheses at that threshold. Here we see again that the eigenvector test produces many more discoveries.

Figure 7a shows the interesting voxels at this FDR level obtained from the eigenvector test at the same slice as Figure 6. The interesting regions extend to other slices beyond the one shown. For reference, Figure 7b shows a map of the total variance at each voxel, i.e. the trace of the covariance matrix S given by (5). In comparison with the underlying anatomy in panel a, it is apparent that the lowest variance is observed in the highly coherent neural fibers of the white matter.

Taking advantage of the spatial structure of the data to increase power, the entire analysis above was repeated after spatially convolving the log-transformed DT data entry-by-entry with a uniform box kernel of size $b \times b \times b$ for $b = 5$. Voxels with neighbors outside the mask were eliminated, resulting in a smaller mask with $N = 73462$ voxels. Test statistics, approximate distributions and p-values were recomputed at each voxel. The new distribution of p-values is shown in Figure 5b and the new FDR thresholds are summarized in the last two columns of Table 1. Both results suggest an increase in power for the eigenvector test, but not necessarily for the eigenvalue test. The p-value map and corresponding interesting regions are shown in Figure 8 for the same slice as above. Notice the difference in texture with Figures 6c and 7a.

In terms of model checking, a test of was applied to the original data in order to check whether the covariance matrices Σ_1 and Σ_2 could have been modeled as OI (Schwartzman et al., 2008b, Proposition 3.1). The null hypothesis of orthogonal invari-

ance was rejected at the 0.05 level for 83.8% of the voxels, thus warranting the need for the approximations developed in Section 2.4. In addition, a Wilks' Lambda test was applied at each voxel in order to test the null hypothesis that the group covariance matrices Σ_1 and Σ_2 are equal. The null hypothesis was rejected at the 0.05 level for 88.6% of voxels, thus warranting the need to assume different covariances between the groups.

5 Discussion

This article has presented tests for detecting differences in eigenvalues and/or eigenvectors of symmetric matrices between the means of two groups and has demonstrated their application to comparison of groups of DTs with arbitrary covariance between the DT entries. Of the three tests considered, the eigenvector test was the most powerful for detecting differences between the two groups in the analyzed dataset. It is remarkable that the eigenvector test was able to find many regions of difference despite the large multiple testing problem over 105822 voxels in the entire brain.

The tests in this paper were applied after a matrix log transformation that maps positive definite matrices to real symmetric matrices. Since the matrix log transformation is bijective and only affects the eigenvalues, the conclusions of the tests are directly interpretable in the original domain. The matrix log transformation is required for validity of the signal-plus-noise model. In practice, however, it may be argued that as n increases, the distributions of the group DT averages become more concentrated around their means and the log transformation has less of an effect. This depends on the Mahalanobis distance from the DT means to the boundaries of the positive definite cone, which is determined by the DT's smallest eigenvalue. Since the trace of the DT is more or less constant over the brain, the matrix log transformation may be more important for highly anisotropoic voxels.

The χ^2 null distributions used in this paper were derived assuming a Gaussian noise model. A Lilliefors test at each voxel revealed that the distribution across subjects of both the DT entries and the log-DT entries at each voxel is never normal. However, it is easy to see from the derivations that in reality not the data but only the group averages are required to be normally distributed. Asymptotically, this is guaranteed by the central limit theorem, and may not be an unreasonable assumption for the moderate sample size in this dataset. Another option not explored here is to estimate the null distribution from the data using an empirical null (Efron, 2007b; Schwartzman, 2008).

An advantage of the methodology presented in this paper is that test statistics have a closed form and are easy to compute for hundreds of thousands of voxels in the brain. A disadvantage has been the need to find approximations to their null distribution, which we have done with both theoretical and empirical tools. An alternative approach is to directly compute the LRT statistics numerically at each voxel by numerically maximizing the full likelihood. While much more computationally costly, it might provide better statistical power and it is worth exploring in future work.

For simultaneous inference across the brain, FDR was chosen for its interpretability and simplicity. As usual in FDR inference, our analysis was based on the marginal distribution of the test statistics. Other approaches such as random field corrections, common in fMRI studies, make explicit use of the spatial correlation in the data (Worsley et al., 2004). Unlike fMRI data, which has implicit physiological blurring, the tensor measurements are mostly independent by design with minimal spatial blurring introduced during image acquisition and preprocessing. The distinct patterns of the underlying anatomy, such as fiber tracts, are features of the brain, not artifacts of the measurement. Other recent statistical analyses of DTI data model voxels independently (Zhu et al., 2007).

While some dependency may exist between neighboring voxels, it is of a local nature

and qualifies as weak dependence so FDR control is still guaranteed (Storey et al., 2004). Here, the spatial structure was used by reanalyzing the data after local spatial smoothing of the log-DT data. By borrowing information from neighboring voxels, smoothing helped increase the power of the eigenvector test for some regions, an effect seen before in DTI data by (Schwartzman et al., 2008a). However, in other regions, smoothing reduces the power by conflating the different eigenvector frames of distinct neighboring anatomical structures.

For an anatomical interpretation of the data analysis results, Figures 9 and 10 show respectively the first and second eigenvectors of the average log-DTs for both groups at the same slice as the previous figures. Since the eigenvector test tests for differences in the full frame of eigenvectors, detected differences may be in the first or second eigenvector (the third is fixed once the first and second are fixed). Many of the regions identified in Figures 7a and 8b correspond to boundaries between anatomical structures. In the PDD map (first eigenvector), some differences in the crossing paterns can be seen in the left hemisphere between the superior longitudinal fasciculus (SLF) (green) and the corona radiata (blue/purple). Some differences in the crossing paterns can be seen also in this region in the second eigenvector map, which indicate the influence of neural fibers that are not parallel to the PDD. Previous DTI studies of FA have found asymmetries between the two brain hemispheres, particularly in the SLF (Büchel et al., 2004). Our findings encourage further study of the neural connectivity involving this region, possibly by means of DTI tractography.

Although the methods in this paper have been illustrated in a voxelwise comparison of two groups of subjects, other potential applications include analysis of symmetry comparing contralateral brain hemispheres, and assessment of error in fiber tracing and spatial registration/normalization.

A Proofs

In this section, $\text{Diag}(A)$ denotes a diagonal matrix with the same diagonal entries as A, and $\text{Offdiag}(A) = A - \text{Diag}(A)$ denotes a matrix with the same off-diagonal entries A and zero diagonal.

Lemma 1.

Denote by $\mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,D})$ the tangent space to the manifold $\mathcal{M}_{2,D}$ given by (4) at an arbitrary point $\boldsymbol{M} = (U_1 D U_1', U_2 D U_2') \in \mathcal{M}_{2,D}$. Then $\mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,D})$ coincides with the set of all pairs $\boldsymbol{X} = (X_1, X_2) \in \mathcal{S}_p \times \mathcal{S}_p$ such that

$$\boldsymbol{X} = (U_1 B_1 U_1' + U_1 C U_1',\ U_2 B_2 U_2' + U_2 C U_2')$$

where $B_1, B_2 \in \mathcal{S}_p$ have $\text{diag}(B_1) = \text{diag}(B_2) = 0$ and $C \in \mathcal{D}_p$.

Proof of Lemma 1.

(i) Let

$$\boldsymbol{M}(t) = \big(U_1 e^{A_1 t}(D + Ct)e^{-A_1 t} U_1',\ U_2 e^{A_2 t}(D + Ct)e^{-A_2 t} U_1'\big)$$

be an arbitrary curve in $\mathcal{M}_{2,D}$ passing through $\boldsymbol{M}$ at $t = 0$, where $C \in \mathcal{D}_p$ and $A_1, A_2 \in \mathcal{A}_p$, the set of $p \times p$ antisymmetric matrices. This parametrization relies on the fact that e^{At} parametrizes $\mathcal{O}(p)$ (Edelman et al., 1998; Lang, 1999; Moakher, 2002). Taking the derivative with respect to t and evaluating at $t = 0$ gives that any tangent vector $\boldsymbol{X} \in \mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,D})$ has the form

$$\boldsymbol{X} = \big(U_1(A_1 D - D A_1)U_1' + U_1 C U_1',\ U_2(A_2 D - D A_2)U_2' + U_2 C U_2'\big).$$

Letting $B_1 = A_1 D - D A_1$ and $B_2 = A_2 D - D A_2$, it is easy to see that both B_1, B_2 are symmetric matrices with zero diagonal.

Conversely, the $\boldsymbol{X}$'s span the tangent space. This is because for every B symmetric with zero diagonal, there exists an antisymmetric A such that $B = AD - DA$. To see

this, observe that the entry (i,j), $i \neq j$, of $AD - DA$ is

$$(AD - DA)_{ij} = \mathbf{e}_i'(AD - DA)\mathbf{e}_j = \mathbf{e}_i' A d_{jj} \mathbf{e}_j - \mathbf{e}_i' d_{ii} A \mathbf{e}_j = (d_{jj} - d_{ii}) a_{ij}$$

where $\mathbf{e}_i$ denotes a column vector with a single 1 in position i, d_{ii} are the diagonal entries of D, and a_{ij} are the entries of A. Therefore, for every B with entries b_{ij}, the corresponding A has entries $a_{ij} = b_{ij}/(d_{jj} - d_{ii})$, $i \neq j$. □

Proof of Theorem 1.

For this test, $\mathcal{T}_0(\boldsymbol{M}) = \mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,D})$ and $\mathcal{T}_A(\boldsymbol{M}) = \mathcal{S}_p \times \mathcal{S}_p$, so (14) reduces to $T^*(\boldsymbol{M}) = \|\text{Proj}_{\mathcal{T}_0^\perp(\boldsymbol{M})} \boldsymbol{Z}\|^2$. By Lemma 1, the projection of $\boldsymbol{Z}$ onto $\mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,D})$ is the minimizer of the square distance

$$\begin{aligned} \|\boldsymbol{Z} - \boldsymbol{X}\|^2 &= \|(Z_1, Z_2) - (U_1 B_1 U_1' + U_1 C U_1',\ U_2 B_2 U_2' + U_2 C U_2')\|^2 \\ &= n_1 \|U_1' Z_1 U_1 - B_1 - C\|^2 + n_2 \|U_2' Z_2 U_2 - B_2 - C\|^2 \end{aligned}$$

For simplicity, let $W_1 = U_1' Z_1 U_1$ and $W_2 = U_2' Z_2 U_2$. We can write

$$\begin{aligned} \|\boldsymbol{Z} - \boldsymbol{X}\|^2 &= n_1 \|\text{Diag}(W_1 - C)\|^2 + 2n_1 \|\text{Offdiag}(W_1 - B_1)\|^2 \\ &\quad + n_2 \|\text{Diag}(W_2 - C)\|^2 + 2n_1 \|\text{Offdiag}(W_2 - B_2)\|^2 \end{aligned}$$

This is minimized over B_1, B_2 and C for

$$\tilde{B}_1 = \text{Offdiag}(W_1), \quad \tilde{B}_2 = \text{Offdiag}(W_2), \quad \tilde{C} = \left[n_1 \text{Diag}(W_1) + n_2 \text{Diag}(W_2)\right]/n \quad (18)$$

The norm $T^*(\boldsymbol{M})$ of the projection of $\boldsymbol{Z}$ onto the space orthogonal to $\mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,D})$ is the same as the minimal distance of $\boldsymbol{Z}$ to $\mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,D})$. Replacing the optimal $\tilde{\boldsymbol{X}}$ with $\tilde{B}_1$, $\tilde{B}_2$ and $\tilde{C}$ in the expression of the distance gives that $T^*(\boldsymbol{M}) = \|\boldsymbol{Z} - \tilde{\boldsymbol{X}}\|^2$ is equal

to

$$\begin{aligned} T^*(\boldsymbol{M}) = \|\boldsymbol{Z} - \tilde{\boldsymbol{X}}\|^2 &= n_1\|\mathrm{Diag}(W_1 - \tilde{C})\|^2 + n_2\|\mathrm{Diag}(W_2 - \tilde{C})\|^2 \\ &= \frac{n_1 n_2}{n}\|\mathrm{Diag}(W_1 - W_2)\|^2 \\ &= \frac{n_1 n_2}{n}\sum_{i=1}^{p}\left\{\mathrm{tr}\left[\mathbf{e}_i'(W_1 - W_2)\mathbf{e}_i\right]\right\}^2 \\ &= \frac{n_1 n_2}{n}\sum_{i=1}^{p}\left[\mathrm{tr}(Z_1 U_1 \mathbf{e}_i\mathbf{e}_i' U_1') - \mathrm{tr}(Z_2 U_2 \mathbf{e}_i\mathbf{e}_i' U_2')\right]^2 \\ &= \frac{n_1 n_2}{n}\sum_{i=1}^{p}\left[\mathrm{vecd}(Z_1)'\mathrm{vecd}(U_1\mathbf{e}_i\mathbf{e}_i' U_1') - \mathrm{vecd}(Z_2)'\mathrm{vecd}(U_2\mathbf{e}_i\mathbf{e}_i' U_2')\right]^2 \end{aligned}$$

which yields the result. □

Proof of Theorem 2.

In the computation of (14) for this test, $\mathcal{T}_0(\boldsymbol{M})$ is the tangent space to the set $\mathcal{M}_{2,M} = \{\boldsymbol{M} = (M, M) : M \in \mathcal{S}_p\}$, which can be identified with $\mathcal{M}_{2,M}$ itself, while $\mathcal{T}_A(\boldsymbol{M}) = \mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,D})$. From the proof of Theorem 1 we have that the projection of $\boldsymbol{Z}$ on $\mathcal{T}_A(\boldsymbol{M})$ is

$$\mathrm{Proj}_{\mathcal{T}_A(\boldsymbol{M})}\boldsymbol{Z} = (\tilde{X}_1, \tilde{X}_2) = (U_1\tilde{B}_1U_1' + U_1\tilde{C}U_1',\ U_2\tilde{B}_2U_2' + U_2\tilde{C}U_2')$$

where $\tilde{B}_1$, $\tilde{B}_2$ and $\tilde{C}$ are given by (18). By Lemma 1, the projection of $\mathrm{Proj}_{\mathcal{T}_A(\boldsymbol{M})}\boldsymbol{Z}$ onto $\mathcal{T}_A(\boldsymbol{M}) = \mathcal{T}_{\boldsymbol{M}}(\mathcal{M}_{2,M})$ is the minimizer of the square distance

$$\|(\tilde{X}_1, \tilde{X}_2) - (M, M)\|^2 = n_1\|\tilde{X}_1 - M\|^2 + n_2\|\tilde{X}_2 - M\|^2$$

This is minimized over $M \in \mathcal{S}_p$ for $\tilde{M} = (n_1\tilde{X}_1 + n_2\tilde{X}_2)/n$. The norm $T^*(\boldsymbol{M})$ of the projection of $\mathrm{Proj}_{\mathcal{T}_A(\boldsymbol{M})}\boldsymbol{Z}$ onto the space orthogonal to $\mathcal{T}_0(\boldsymbol{M})$ is the same as the minimal distance of $\mathrm{Proj}_{\mathcal{T}_A(\boldsymbol{M})}\boldsymbol{Z}$ to $\mathcal{T}_0(\boldsymbol{M})$. Replacing the optimal $\tilde{M}$ in the expression of the distance gives

$$T^*(\boldsymbol{M}) = \|(\tilde{X}_1, \tilde{X}_2) - (\tilde{M}, \tilde{M})\|^2 = n_1\|\tilde{X}_1 - \tilde{M}\|^2 + n_2\|\tilde{X}_2 - \tilde{M}\|^2 = \frac{n_1 n_2}{n}\|\tilde{X}_1 - \tilde{X}_2\|^2$$

Replacing $\tilde{X}_1$ and $\tilde{X}_2$ using (18), and adding and subtracting $U_1\text{Diag}(W_1)U_1'$ and $U_2\text{Diag}(W_2)U_2'$, we get that $T^*(\boldsymbol{M})$ can be written as

$$T^*(\boldsymbol{M}) = \frac{n_1 n_2}{n}\|Z_1 - H(W_1) - Z_2 + H(W_2)\|^2 \tag{19}$$

where

$$\begin{aligned} H(W_1) &= (n_2 U_1\text{Diag}(W_1)U_1' + n_1 U_2\text{Diag}(W_1)U_2')/n \\ H(W_2) &= (n_2 U_1\text{Diag}(W_2)U_1' + n_1 U_2\text{Diag}(W_2)U_2')/n \end{aligned}$$

The entries of $H(W_1)$ are

$$\text{tr}(\mathbf{e}_i' H(W_1)\mathbf{e}_j) = \text{tr}\left[\text{Diag}(W_1)(n_2 U_1'\mathbf{e}_j\mathbf{e}_i'U_1 + n_1 U_2'\mathbf{e}_j\mathbf{e}_i'U_2)/n\right] = \text{diag}(W_1)'\mathbf{h}_{ij}$$

where $\mathbf{h}_{ij}$ is the $p \times 1$ vector

$$\mathbf{h}_{ij} = \text{diag}(n_1 U_2' E_{ij} U_2 + n_2 U_1' E_{ij} U_1)/n$$

Furthermore, we can write $\text{diag}(W_1) = J(U_1)'\text{vecd}(Z_1)$, where

$$J(U) = \begin{pmatrix} \text{vecd}(UE_{11}U') & \text{vecd}(UE_{22}U') & \cdots & \text{vecd}(UE_{pp}U') \end{pmatrix}$$

This can be verified by checking that the entries of the $p \times 1$ vector $\text{diag}(W_1)$ are

$$\mathbf{e}_i'\text{Diag}(W_1)\mathbf{e}_i = \text{tr}(\mathbf{e}_i'U_1 Z_1 U_1'\mathbf{e}_i) = \text{tr}(U_1'\mathbf{e}_i\mathbf{e}_i'U_1 Z_1) = \text{vecd}(U_1'E_{ii}U_1)'\text{vecd}(Z_1)$$

Similar expressions are btained for $H(W_2)$. Going back to (19) we have that

$$\begin{aligned} T^*(\boldsymbol{M}) &= \frac{n_1 n_2}{n}\sum_{i=1}^{p}\sum_{i=1}^{p}\left[\text{tr}\big(\mathbf{e}_i'Z_1\mathbf{e}_j - \mathbf{e}_i'H(W_2)\mathbf{e}_j - \mathbf{e}_i'Z_2\mathbf{e}_j + \mathbf{e}_i'H(W_2)\mathbf{e}_j\big)\right]^2 \\ &= \frac{n_1 n_2}{n}\sum_{i=1}^{p}\sum_{i=1}^{p}\big[\text{vecd}(Z_1)'\text{vecd}(\mathbf{e}_j\mathbf{e}_i') - \text{vecd}(Z_1)'J(U_1)h_{ij} \\ &\qquad - \text{vecd}(Z_2)'\text{vecd}(\mathbf{e}_j\mathbf{e}_i') - \text{vecd}(Z_2)'J(U_2)h_{ij}\big]^2 \end{aligned}$$

which gives the result. □

Proof of Proposition 1.

This can be seen as a special case of Theorem 3.4.4a of Mardia et al. (1979), but can also be proven directly as follows. Let $\mathbf{z} = \mathbf{\Sigma}^{-1/2}\text{vecd}(\boldsymbol{Z}) \sim N_{2q}(0, I_{2q})$. We have

$$T^* = \text{vecd}(\boldsymbol{Z})'\mathbf{\Omega}\text{vecd}(\boldsymbol{Z}) = \mathbf{z}'\mathbf{\Sigma}^{1/2}\mathbf{\Omega}\mathbf{\Sigma}^{1/2}\mathbf{z} = (Q\mathbf{z})'\mathbf{\Lambda}(Q\mathbf{z})$$

where $\mathbf{\Lambda}$ and C are respectively the eigenvalue and eigenvector matrices of $\mathbf{\Sigma}^{1/2}\mathbf{\Omega}\mathbf{\Sigma}^{1/2}$. The result follows because $Q\mathbf{z} \sim N_{2q}(0, I_{2q})$. □

References

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Fast and simple calculus on tensors in the log-Euclidean framework. In *MICCAI 2005, LNCS 3749*, pages 115–122, 2005.

Peter J. Basser and Sinisa Pajevic. A normal distribution for tensor-valued random variables: applications to diffusion tensor MRI. *IEEE Trans Med Imaging*, 22(7): 785–794, 2003.

Peter J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated b y quantitative-diffusion-tensor MRI. *J Magn Reson B*, 111:209–219, 1996.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*, 57(1):289–300, 1995.

C. Büchel, T. Raedler, M. Sommer, M. Sach, C. Weiller, and M. A. Koch. White matter asymmetry in the human brain: a diffusion tensor MRI study. *Cereb Cortex*, 14(9):945–951, 2004.

Georege Casella and Roger L. Berger. *Statistical Inference*. Duxbury, Pacific Grove, California, 2nd edition, 2002.

Tom Y. M. Chiu, Tom Leonard, and Kam Wah Tsui. The matrix-logarithmic covariance model. *J Amer Statist Assoc*, 91(433):198–210, 1996.

Robert F. Dougherty, Michal Ben-Shachar, Gayle K. Deutsch, Polina Potanina, and Brian A. Wandell. Occipital-callosal pathways in children: Validation and atlas development. *Ann N Y Acad Sci*, 1064:98–112, 2005.

Robert F. Dougherty, Michal Ben-Shachar, Gayle K. Deutsch, Arvel Hernandez, Glenn R. Fox, and Brian A. Wandell. Temporal-callosal pathway diffusivity predicts phonological skills in children. *PNAS*, 104(20):8556–8561, 2007.

Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J Matrix Anal Appl*, 20(2):303–353, 1998.

Bradley Efron. Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *J Amer Statist Assoc*, 99(465):96–104, 2004.

Bradley Efron. Correlation and large-scale simultaneous hypothesis testing. *J Amer Statist Assoc*, 102(477):93–103, 2007a.

Bradley Efron. Size, power and false discovery rates. *Ann Statist*, 35(4):1351–1377, 2007b.

P. Thomas Fletcher and Sarang Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007.

Christopher R. Genovese, Nicole A. Lazar, and Thomas E. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15:870–878, 2002.

Derek K. Jones, Lewis D. Griffin, Daniel C. Alexander, Marco Catani, Mark Horsfield, Robert Howard, and Steve C. R. Williams. Spatial normalization and averaging of diffusion tensor MRI data sets. *Neuroimage*, 17:592–617, 2002.

D. Kuonen. Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*, 86(4):929–935, 1999.

Serge Lang. *Fundamentals of Differential Geometry*. Springer-Verlag, New York, 1999.

Denis LeBihan, Jean-François Mangin, Cyril Poupon, Chris A. Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. Diffusion tensor imaging: Concepts and applications. *J Magn Reson Imaging*, 13:534–546, 2001.

Tom Leonard and John S. J. Hsu. Bayesian inference for a covariance matrix. *Ann Statist*, 20(4):1669–1696, 1992.

Brent R. Logan and Daniel B. Rowe. An evaluation of thresholding techniques in fMRI analysis. *Neuroimage*, 22:95–108, 2004.

C. L. Mallows. Latent vectors of random symmetric matrices. *Biometrika*, 48(1): 133–149, 1961.

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, San Diego, California, 1979.

Madan Lal Mehta. *Random Matrices*. Academic Press, San Diego, California, 2nd edition, 1991.

Maher Moakher. Means and averaging in the group of rotations. *SIAM Matrix Anal Appl*, 24(1):1–16, 2002.

Andreas M. Rauschecker, Gayle K. Deutsch, Michal Ben-Shachar, Armin Schwartzman, Lee M. Perry, and Robert F. Dougherty. Reading impairment in a patient with missing arcuate fasciculus. *Neuropsychologia*, 47(1):180–194, 2009.

Armin Schwartzman. Empirical null and false discovery rate inference for exponential families. *Ann Appl Statist*, 2(4):1332–1359, 2008.

Armin Schwartzman. *Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data*. PhD thesis, Stanford University, June 2006.

Armin Schwartzman, Robert F. Dougherty, and Jonathan E. Taylor. Cross-subject comparison of principal diffusion direction maps. *Magn Reson Med*, 53:1423–1431, 2005.

Armin Schwartzman, Robert F. Dougherty, and Jonathan E. Taylor. False discovery rate analysis of brain diffusion direction maps. *Ann Appl Statist*, 2(1):153–175, 2008a.

Armin Schwartzman, Walter Mascarenhas, and Jonathan E. Taylor. Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *Ann Statist*, 36(6): 2886–2919, 2008b.

Armin Schwartzman, Robert F. Dougherty, Jongho Lee, Dara Ghahremani, and Jonathan E. Taylor. Empirical null and false discovery rate analysis in neuroimaging. *Neuroimage*, 44(1):71–82, 2009.

John D. Storey, Jonathan E. Taylor, and David Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J R Statist Soc B*, 66(1):187–205, 2004.

Alan Stuart and J. Keith Ord. *Kendall's Advanced Theory of Statistics*. Edward Arnold, London, 6th edition, 1994.

Brandon Whitcher, Jonathan J. Wisco, Nouchine Hadjikhani, and David S. Tuch. Statistical group comparison of diffusion tensors via multivariate hypothesis testing. *Magn Reson Med*, 57:1065–1074, 2007.

Keith J. Worsley, Jonathan E. Taylor, F. Tomaiuolo, and J. Lerch. Unified univariate and multivariate random field theory. *Neuroimage*, 23:S189–195, 2004.

Yu-Chien Wu, Aaron S. Field, Moo K. Chung, Benham Badie, and Andrew L. Alexander. Quantitative analysis of diffusion tensor orientation: theoretical framework. *Magn Reson Med*, 52:1146–1155, 2004.

Ying Yao. An approximate degrees of freedom solution to the multivariate behrens fisher problem. *Biometrika*, 52(1/2):139–147, 1965.

Hongtu Zhu, Heping Zhang, Joseph G. Ibrahim, and Bradley S. Peterson. Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging. *J Amer Statist Assoc*, 102(480):1085–1102, 2007.

Table 1: Thresholds and fraction of rejected hypotheses R/m at FDR level 0.05.

	unsmoothed data		smoothed data, $b = 5$	
test type	threshold	R/m	threshold	R/m
eigenvalue	7.1×10^{-5}	0.0014	4.5×10^{-5}	0.0009
eigenvector	1.1×10^{-3}	0.0225	4.5×10^{-3}	0.0895

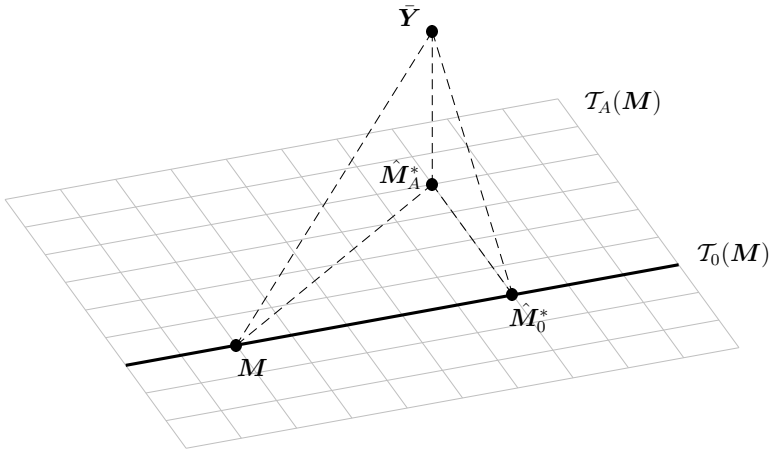

Figure 1: The tangent spaces $\mathcal{T}_0(\boldsymbol{M}) \subset \mathcal{T}_A(\boldsymbol{M})$ to $\mathcal{M}_0$ and $\mathcal{M}_A$ at M and the corresponding MLEs. The nested underlying manifolds $\mathcal{M}_0 \subset \mathcal{M}_A$ are not shown.

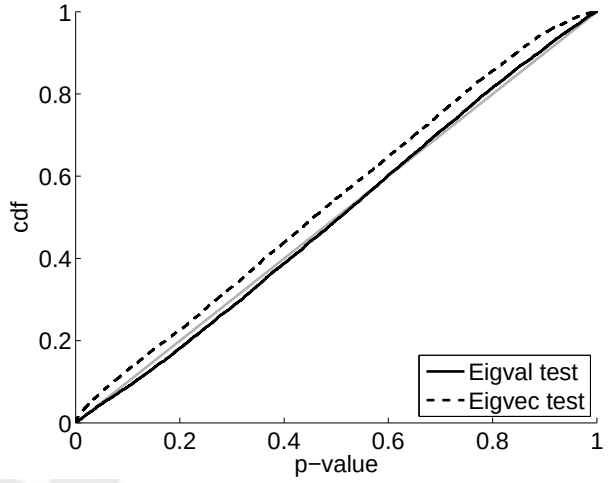

Figure 2: Null distribution of p-values for the two-sample eigenvalue and eigenvector test statistics, for $n_1 = 50$, $n_2 = 50$.

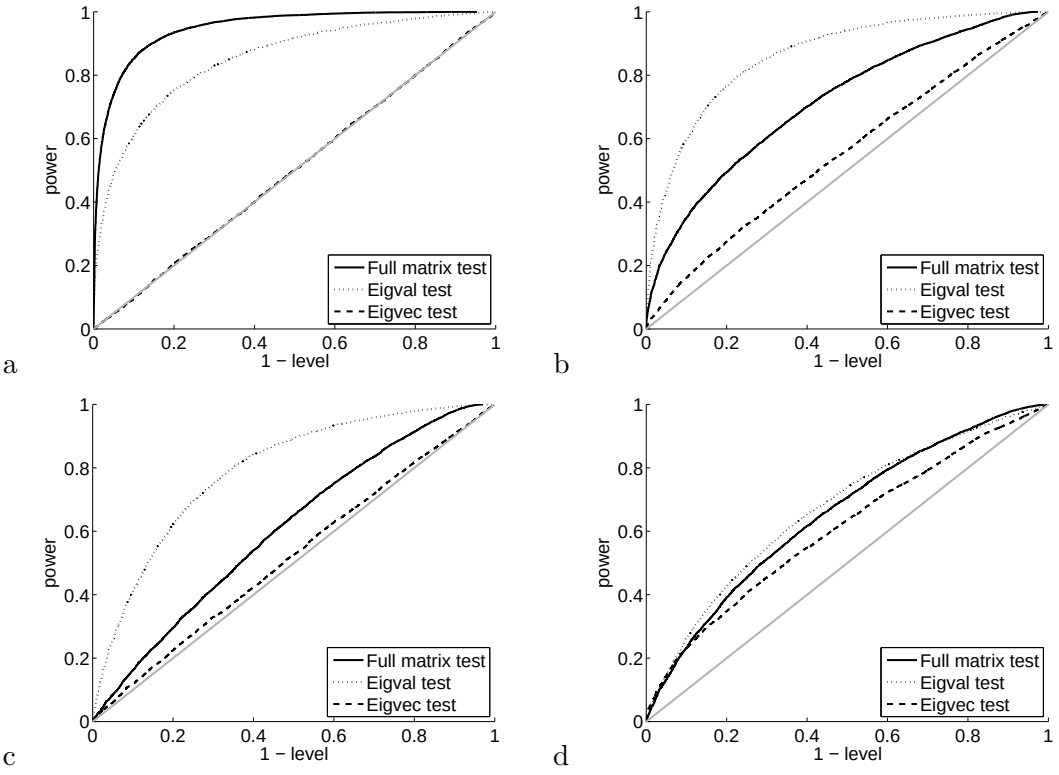

Figure 3: ROC curves for fixed changes in eigenvalues, $n_1 = 50$, $n_2 = 50$, $\Sigma = 0.5\text{Wishart}(I_6, 6)$. (a) $\Delta M = \text{Diag}(0.2, 0.4, 0.8)$. (b) $\Delta M = \text{Diag}(0.2, 0.4, -0.8)$. (c) $\Delta M = \text{Diag}(0.4, 0.4, -0.4)$. (d) $\Delta M = \text{Diag}(0.4, -0.4, -0.4)$.

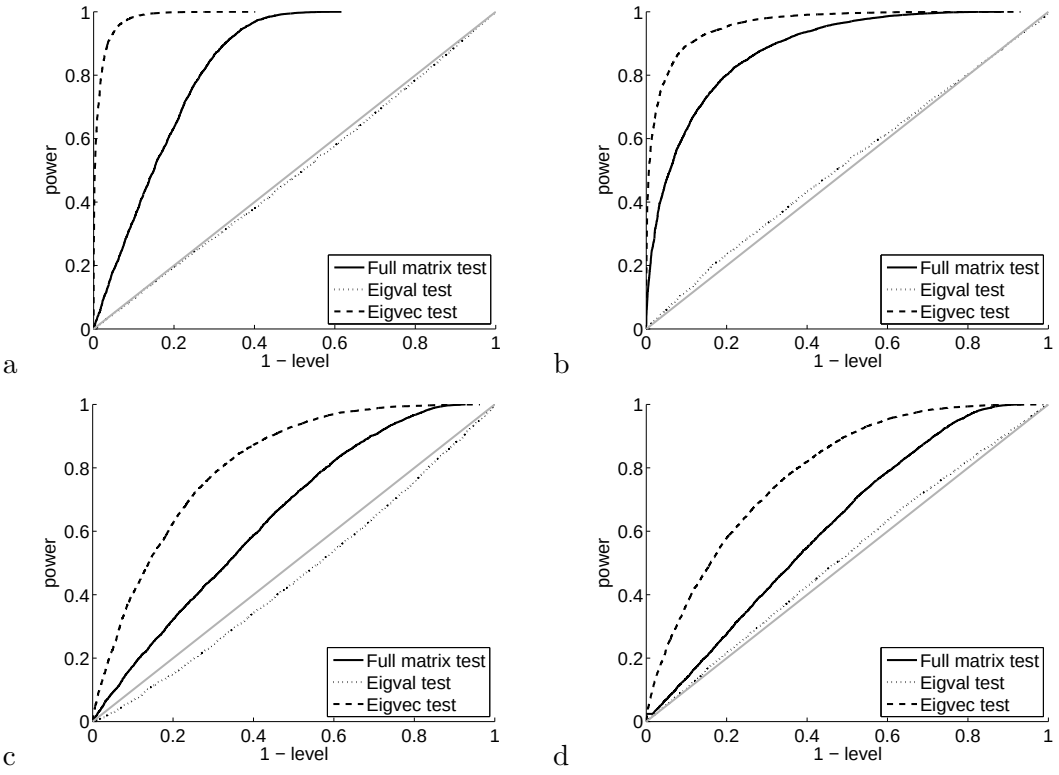

Figure 4: ROC curves for fixed changes in eigenvectors, $n_1 = 50$, $n_2 = 50$, $\Sigma = 0.5\text{Wishart}(I_6, 6)$. (a) $\boldsymbol{a} = (1,1,1)/\sqrt{3}$, $\theta = 0.5$. (b) $\boldsymbol{a} = (1,1,1)/\sqrt{3}$, $\theta = 0.5$. (c) $\boldsymbol{a} = (1,0,0)$, $\theta = 0.5$. (d) $\boldsymbol{a} = (0,0,1)$, $\theta = 0.5$.

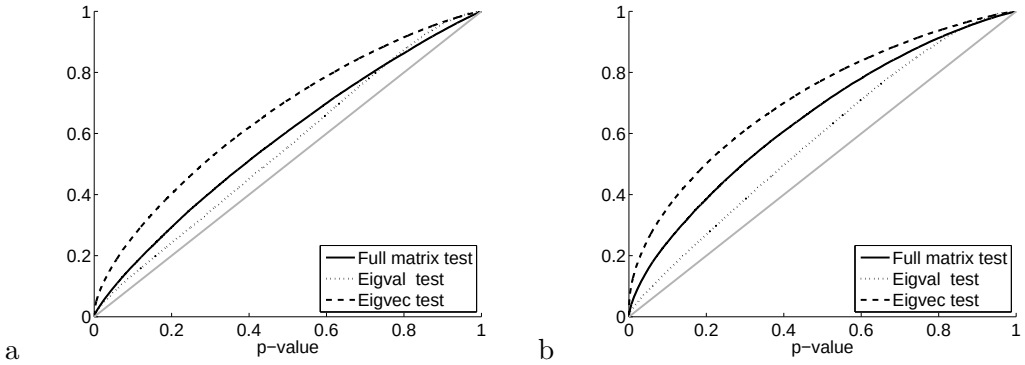

Figure 5: Empirical distribution of p-values for the three test types: (a) unsmoothed data, $N = 105822$ voxels; (b) smoothed data with $b = 5$, $N = 73462$ voxels.

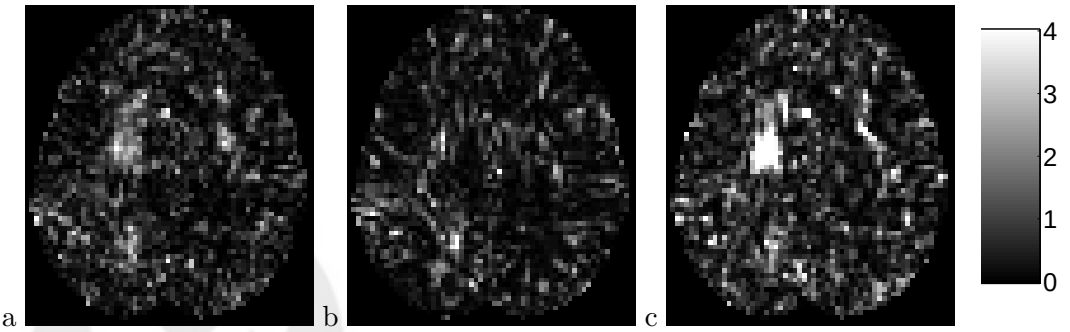

Figure 6: P-value maps in scale $-log_{10}(p)$ at a transverse slice 36 mm above the anterior commisure: (a) full matrix test; (b) eigenvalue test; (c) eigenvector test.

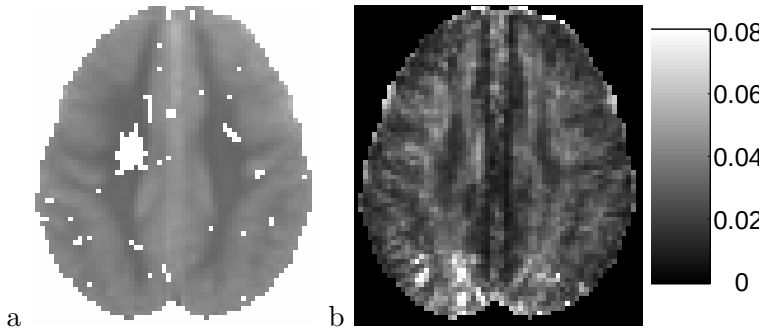

Figure 7: (a) Transverse slice 36 mm above the anterior commisure: Interesting voxels according to eigenvector test at FDR level 0.05. (b) Trace of the empirical covariance $S = S_1/n_1 + S_2/n_2$.

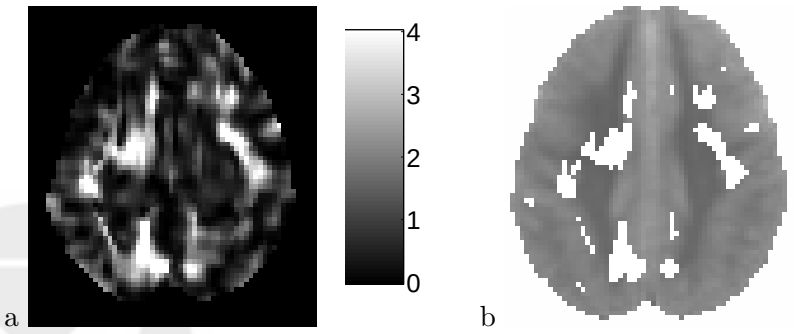

Figure 8: Spatial smoothing of the data with $b = 5$, transverse slice 36 mm above the anterior commisure: (a) P-value map in scale $-log_{10}(p)$. (c) Interesting voxels at FDR level 0.05.

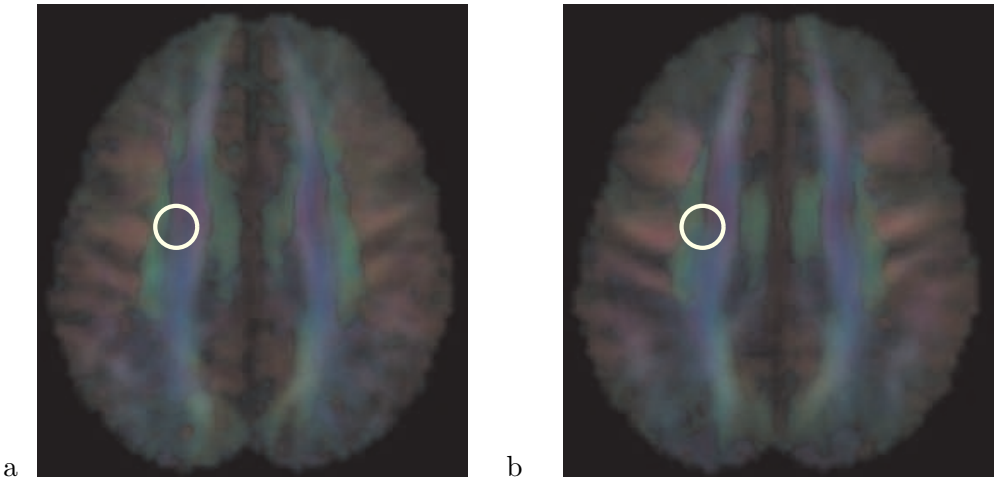

Figure 9: First eigenvector of DT mean for boys (a) and girls (b) at a transverse slice 36 mm above the anterior commisure. Colors indicate coordinate directions: right-left (red) and anterior-posterior (green) and superior-inferior (blue). Mixed colors represent directions that are oblique to the coordinate axes. The intensity is weighted by FA, so brighter regions have higher anisotropy.

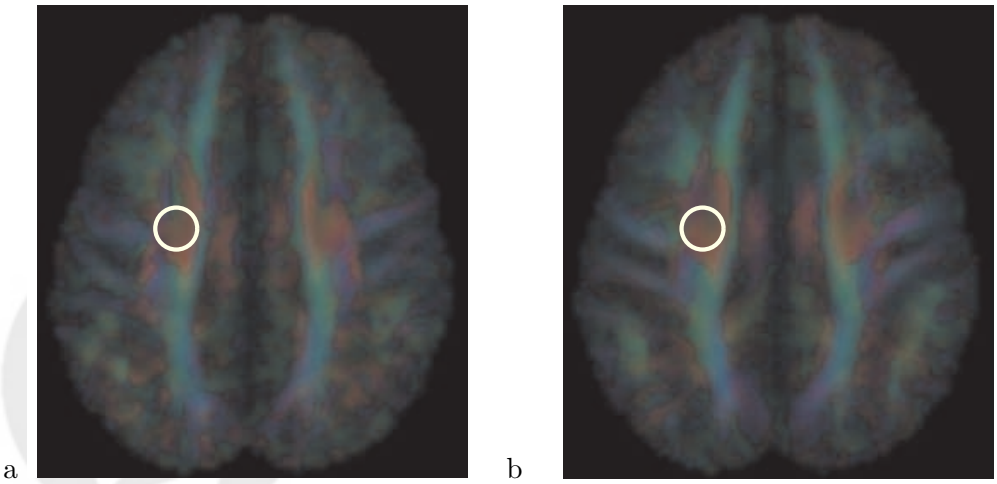

Figure 10: Second eigenvector of DT mean for boys (a) and girls (b) at the same slice and using the same color scheme as Figure 9.