DEPARTMENT OF STATISTICS
University of Wisconsin
1210 West Dayton St.
Madison, WI 53706

TECHNICAL REPORT NO. 1081

November 10, 2003

Probabilistic Connectivity Measure in Diffusion Tensor Imaging via Anisotropic Kernel Smoothing

[1,2,3]Moo K. Chung[1], [3,4]Mariana Lazar, [3,4]Andrew L. Alexender,
[1,3] Yuefeng Lu, [3]Richard Davidson

[1]Department of Statistics
[2]Department of Biostatistics and Medical Informatics
[3]Keck Laboratory for Functional Brain Imaging and Behavior
[4]Department of Medical Physics
University of Wisconsin-Madison
Madison, WI 53706

[1]Corresponding Author. email:mchung@stat.wisc.edu, http://www.stat.wisc.edu/~mchung

Probabilistic Connectivity Measure in Diffusion Tensor Imaging via Anisotropic Kernel Smoothing

[1,2,3]Moo K. Chung[†], [3]Mariana Lazar, [3,4]Andrew L. Alexender,
[1,3] Yuefeng Lu, [3]Richard Davidson
[1]Department of Statistics
[2]Department of Biostatistics
[3]Keck Laboratory for Functional Brain Imaging and Behavior
[4]Department of Medical Physics
University of Wisconsin-Madison
Madison, WI 53706

November 10, 2003

Abstract

We present a novel probabilistic approach of representing the connectivity of the brain white fiber in diffusion tensor imaging via anisotropic Gaussian kernel smoothing. Our approach is simpler than solving a diffusion equation, which has been used in probabilistic representation of white matter connectivity by other researchers. Also the connectivity metric is deterministic in a sense that it avoids using Monte-Carlo random walk simulation in constructing the transition probability so the resulting connectivity maps do not change from one computational run to another. As a further usefulness of this new method, the same computational framework can also be used in smoothing functional and structural signals along the white fiber tracks.

1 Introduction

Diffusion tensor imaging (DTI) is a new technique that provides the directional information of water diffusion in the white matter of the brain. The directional information is usually represented as a symmetric positive definite 3×3 matrix which is usually termed as the *diffusion tensor* or

[†]Corresponding Author. email://mchung@stat.wisc.edu, http://www.stat.wisc.edu/∼mchung

diffusion coefficients. The diffusion tensor can be used to estimate the patterns of white matter connectivity. In white matter tractography, a continuous path of connection between two brain regions is estimated mainly from the eigenvectors of the diffusion tensor. Most of current white matter tractography is based on streamlines (Conturo *et al.*, 1999; Mori *et al.*, 1999; Basser *et al.*, 2000) or the variations on the streamlines such as tensor deflection method (Weinstein *et al.*, 1999; Lazar *et al.*, 2003). Also Tench *et al.* (2002) introduced a hybrid streamline-based tractography where the direction of principal eigenvector is modelled stochastically to overcome the shortcomings of DTI. The white fiber tracking is prone to cumulative acquisition noise and partial volume effect so the estimated white fiber tracks might possibly be erroneous in some cases (Alexander *et al.*, 2001; Basser *et al.*, 2000; Tench *et al.*, 2000). So it is crucial to develop a connectivity metric that is robust under the effect of acquisition noise and partial voluming. Such a robust metric can be possibly used in voxel-based morphometry (VBM) (Ashburner *et al.*, 2000) in detecting the region of the connectivity difference between two clinical groups of brain images. In the classical VBM, the gray and white matter concentration probability metrics are computed and used for statistical inference on gray and white matter concentration at each voxel. In DTI, instead of white and gray matter concentration probability, we can introduce the concept of connection probability metric which measures the strength of how two regions of the brain are connected via the white fiber tracks.

Before we go to the detailed discussion of our new method, let us review the relevant previous works on probabilistic connectivity measures in DTI that have been used in brain imaging. The probabilistic approach to the white fiber tracking is somewhat new. Koch *et al.* (2002) introduced a Monte-Carlo random walk simulation that uses a different transition probability than our own. Their algorithm has a certain restrictions built in the random walk so that it was only allowed to jump in a direction within 90 degree from the previous jump direction, which restricts the jump to a very small number of voxels in the neighborhood. Furthermore they only considered the voxels with the fractional anisotropy (FA) index (Basser and Pierpaoli, 1996) and sum of the eigenvectors bigger than certain thresholds. Then based on the Monte-Carlo simulation of 4000 random walks, they computed the probabilistic connectivity measure. Our kernel-smoothing based result would be somewhat compatible to the Monte-Carlo approach of Kosh *et al.* (2002) with unrestricted random walk and different local weights. However, our approach would be easier to implement numerically compared to the Monte-Carlo simulation of Kosh *et al.* (2002).

Hagmann *et al.* (2000) used a hybrid approach combining Monte-Carlo random walk simulation with information about the white fiber track curvature function in the corpus callosum. Then assuming bivariate normal distribution of the random walk hitting a vertical plane at some distance apart, they estimated the covariance matrix and performed a statistical hypothesis testing of the homogeneity of covariance matrix in the different regions of the corpus callosum.

Batchelor *et al.* (2002) solved an anisotropic heat equation where the diffusion coefficients of the heat equation are the diffusion coefficients of DTI. To get the probabilistic measure of the connectivity, the diffusion equation is solved with the initial condition where every vertex is zero except a seed region where it is given the value one. Then the value 1 is diffused though the brain and the numerical value that should be between 0 and 1 is taken as a probability of connection. Mathematically it is equivalent as the Monte-Carlo random walk simulation (Kosh *et al.*, 2002)

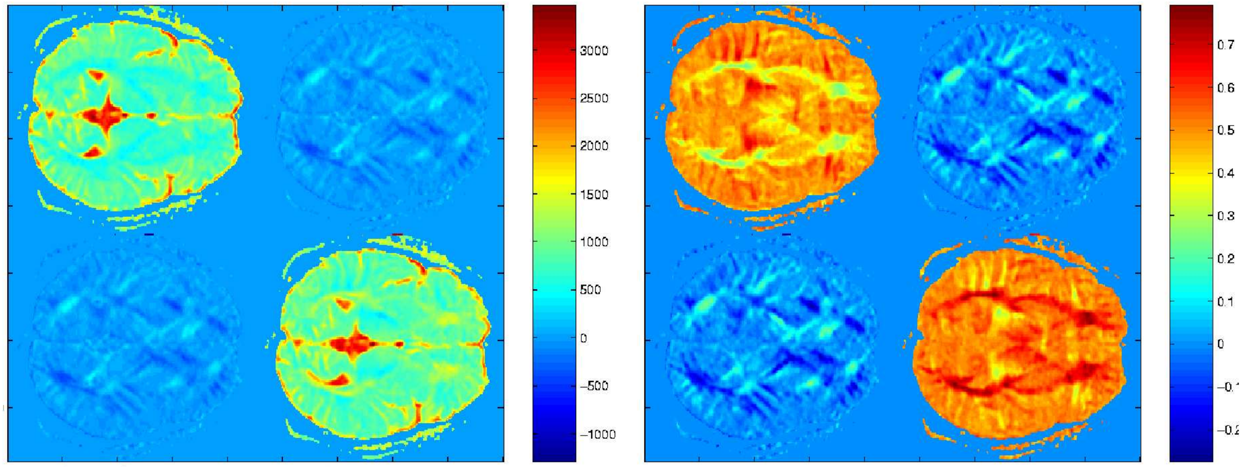

Figure 1: Left: the original diffusion coefficients D_0 for 2D slice, Right: normalized diffusion coefficients $D = D_0/\mathrm{tr}D_0$ for 2D slice. D gives a natural Riemannian metric tensor $G = D$ to DTI. For display purpose, 2 dimensional version of anisotropic Gaussian kernel smoothing are presented.

with no restriction.

Our kernel smoothing approach presented here is different from Kosh *et al.* (2002) and Batchelor *et al.* (2002) but it is compatible to these two methods. In our kernel-based approach, the connectivity metric is taken as the transitional probability density of a continues diffusion process and it is estimated via spatially adaptive anisotropic Gaussian kernel smoothing. This is simple to implement numerically compared to solving an anisotropic diffusion equation numerically. Further, unlike the Monte-Carlo random walk simulation approach, it always give the deterministic metric that does not change from one run to another so it would be better suited in VBM-like morphometry. As a further usefulness of this nobel technique, this smoothing technique can be used to smooth out data to increase the signal-to-noise ratio (SNR) while preserving the directional information of DTI. This kind of directional smoothing would be needed for DTI template construction and DTI registration. In this context, Gössl *et al.* (2002) applied an isotropic Gaussian kernel smoothing with FWHM of 0.75 voxels to get a sufficiently smooth, continuous representation of the DTI data in white fiber tracking using the linear state space model. However, the isotropic smoothing tends to smooth out the directional characteristic of the DTI data. Even though Gössl *et al.* (2002) reported that the use of isotropic kernel smoothing performs sufficiently well for their computation, but they indicated that a more sophisticated smoothing algorithm should be investigated. Our anisotropic kernel smoothing approach might shed a light on this respect.

2 Anisotropic Gaussian kernel Smoothing

Let $\mathbf{x} = (x_1, \cdots, x_n)' \in \mathbb{R}^n$. The n-dimensional isotropic Gaussian kernel is given by

$$K(\mathbf{x}) = \exp(-\mathbf{x}'\mathbf{x}/2)/(2\pi)^{n/2},$$

which is the joint multivariate normal probability density function of n independent standard normal random variables. Note that $\int_{\mathbb{R}^n} K(\mathbf{x})\ d\mathbf{x} = 1$. The *full width at half maximum* (FWHM) which is the usual unit for filter size among brain imaging researchers is given by

$$\text{FWHM} = 2(\ln 4)^{1/2}.$$

It is the length of the full width of the kernel at half maximum. The *anisotropic Gaussian kernel* is defined as the generalization of isotropic kernel $K_H(x) = K(H^{-1}\mathbf{x})/\det(H)$, where H is a $n \times n$ constant *bandwidth matrix*. Note that $K_H(x)$ is a multivariate normal density with the mean zero and the covariance matrix HH'. Since levelset of the anisotropic Gaussian kernel is ellipsoidal, we generalize FWHM of isotropic kernel. Suppose λ_i are eigenvalues of HH', it can be shown that $\sqrt{\lambda_i}\ln 4$ are FWHM along the principal axis of ellipsoid $\mathbf{x}'(HH')^{-1}\mathbf{x} = \ln 4$. Taking the average generalized-FWHM to be locally

$$\text{FWHM} = \frac{2(\ln 4)^{1/2}}{n}\sum_{i=1}^{n}\sqrt{\lambda_i} = \frac{2(\ln 4)^{1/2}}{n}\ \text{tr}\ (HH')^{1/2}.$$

Anisotropic Gaussian kernel smoothing is defined as the convolution of signal f and kernel K_H:

$$F(\mathbf{x}) = K_H * f(\mathbf{x}) = \int_{\mathbb{R}^n} K_H(\mathbf{x}-\mathbf{y})f(\mathbf{y})\ dy. \tag{1}$$

Note that $K_H * f(\mathbf{x}) = f * K_H(\mathbf{x})$. Because of the computational difficulty of integrating from $-\infty$ to ∞ in (1), the integral is usually truncated within a close and bounded domain and corrected for the truncation. The kernel $K_H(\mathbf{x})$ decreases exponentially as $|\mathbf{x}|$ increases so the most of weight should be concentrated near the origin. For example, when we integrate the kernel K in the closed cube $[-2.58, 2.58]^n$, we get 0.99^n which is sufficiently close enough to the total probability 1. Let us denote the truncated and normalized kernel as

$$\tilde{K}_H(\mathbf{x}) = \frac{K_H(\mathbf{x})\mathbf{1}_B(\mathbf{x})}{\int_B K_H(\mathbf{y})\ d\mathbf{y}},$$

where $\mathbf{1}_B$ is an index function that gives zero everywhere except B where it gives the value one. B can be taken as either a closed ball or cube centered around the origin. Then we approximate the convolution (1) as

$$F(\mathbf{x}) = \tilde{K}_H * f(\mathbf{x}) = \frac{\int_{B_\mathbf{x}} K_H(\mathbf{x}-\mathbf{y})f(\mathbf{y})\ d\mathbf{y}}{\int_B K_H(\mathbf{y})\ d\mathbf{y}}, \tag{2}$$

where $B_\mathbf{x}$ is the translation of B to $\mathbf{x}$, i.e. $B_\mathbf{x} = \mathbf{x} + B$. The two important properties of the kernel smoothing are

$$|K_H * f(\mathbf{x})| \leq |f(\mathbf{x})|, \tag{3}$$

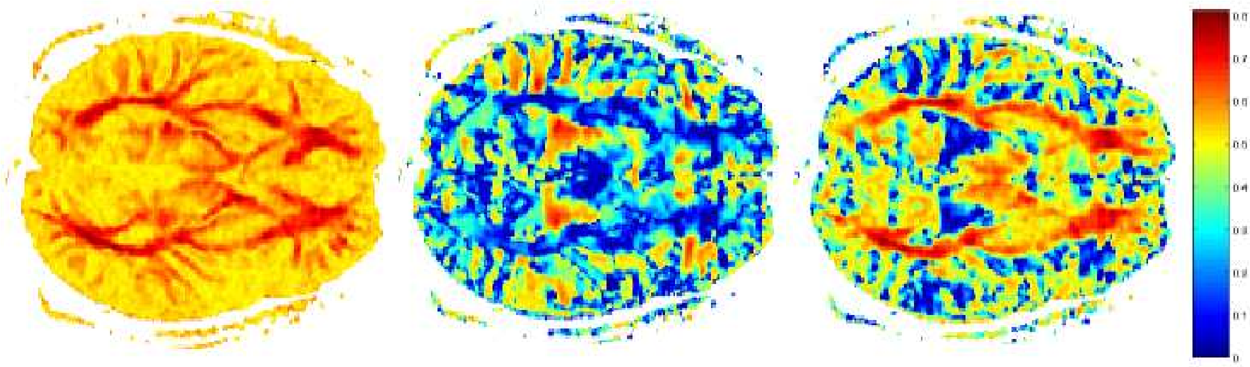

Figure 2: Left: the principal eigenvalues of D. Middle: x-component of the principal eigenvectors. Right: y-component of the principal eigenvectors. Note that the x coordinate direction is along the vertical axis while y direction is along the horizontal axis following MATLAB convention.

$$\int_{\mathbb{R}^n} K_H * f(\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^n} f(\mathbf{x})\, d\mathbf{x}. \tag{4}$$

So if f is a probability density function, then integral (4) is one showing the conservation of the total probability under kernel smoothing. This is also true for $\tilde{K}_H$ and this fact will be used in the later section. For numerical implementation, we define $B_{\mathbf{x}}$ to be a collection of voxels that forms a cube around $\mathbf{x}$. The discrete version of the truncated kernel is then

$$\tilde{K}_H(\mathbf{x}) = K_H(\mathbf{x}) / \sum_{x_j \in B_{\mathbf{x}}} K_H(\mathbf{x}_j)$$

and the discrete version of the kernel smoothing

$$F(\mathbf{x}) = \tilde{K}_H * f(\mathbf{x}) = \sum_{x_j \in B_{\mathbf{x}}} \tilde{K}_H(\mathbf{x} - \mathbf{x}_j) f(\mathbf{x}_j), \tag{5}$$

Note that $\sum_{x_j \in B_{\mathbf{x}}} \tilde{K}_{\Delta t}(\mathbf{x}) = 1$ for all $\mathbf{x} \in \mathbb{R}^n$. So we can view the discrete version of anisotropic Gaussian kernel as a weighted local averaging.

3 Riemannian Metric Tensors

Let D_0 be raw diffusion coefficients. D_0 is represented as $n \times n$ matrix in $\mathbb{R}^n$. We normalize it by $D = D_0/\mathrm{tr}D_0$. This normalization guarantees that the sum of eigenvalues of D to be 1. Consider a vector field $\mathbf{V} = (V_1, \cdots, V_n)'$ which is the principal eigenvector of D with $\|\mathbf{V}\| = 1$ with the corresponding principal eigenvalue λ. Now suppose that we would like to smooth signals along the vector fields $\lambda\mathbf{V}$ such that we smooth more along the larger vector fields. Suppose that the stream line or flow $\mathbf{x} = \psi(t)$ corresponding to the vector field is given by

$$\frac{d\psi}{dt} = (\lambda \mathbf{V}) \circ \psi(t).$$

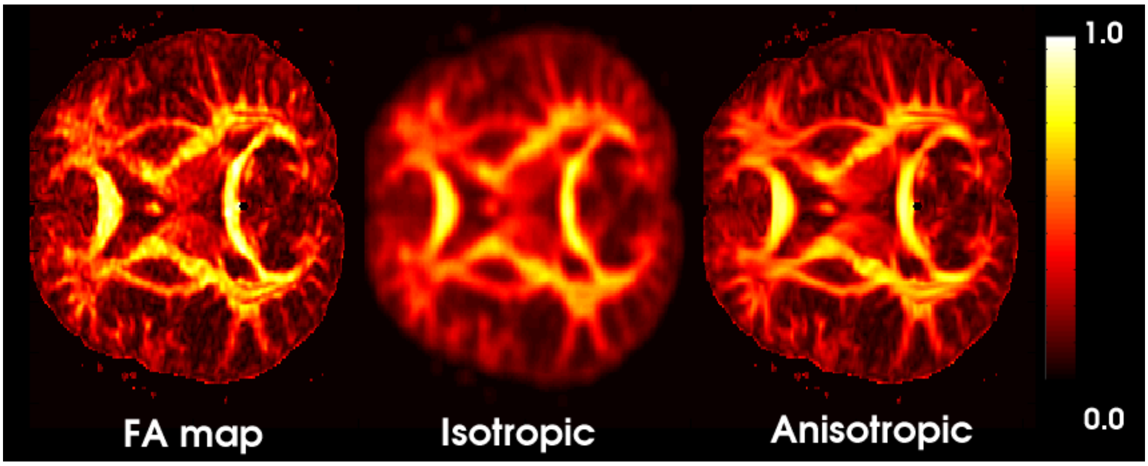

Figure 3: Left: normalized FA-map. Middle: isotropic Gaussian kernel smoothing with 12mm FWHM. Right: anisotropic Gaussian kernel smoothing with generalized 12mm FWHM.

This ordinary differential equation gives a family of integral curves whose tangent vector is $\lambda\mathbf{V}$ (Betounes, 1998). The line element is

$$d\psi^2 = \lambda^2(V_1^2\, dx_1^2 + \cdots + V_n^2 dx_n^2).$$

So $g_{ij} = \lambda^2 V_i^2 \delta_{ij}$. We want to smooth more along the larger metric distance so we let $HH' = 2tG, G = (g_{ij})$, i.e.

$$HH' = 2\lambda^2 t \begin{pmatrix} V_1^2 & 0 & \cdots & 0 \\ 0 & V_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & V_n^3 \end{pmatrix}.$$

By introducing the scaling parameter t, we left a room for adjusting the amount of smoothing. For the above choice of the covariance matrix, anisotropic kernel is given by

$$K_t(\mathbf{x}) = (4\pi\lambda^2 t)^{-n/2} \prod_{j=1}^{n} \frac{1}{|V_j|} \exp\Big(- \frac{x_j^2}{4\lambda^2 t V_j^2}\Big). \tag{6}$$

For this type of kernel, our generalized FWHM is locally

$$\text{FWHM} = 4\lambda\sqrt{t}\Big(\frac{|V_1 \cdots V_n|}{2}\Big)^{1/2} \frac{1}{n}\sum_{i=1}^{n} V_i.$$

In the case $V_j = 0$ while $|\mathbf{V}| \neq 0$ in (6), the kernel diverges. To avoid the divergence in numerical implementation, we can introduce very small ϵ and use $\mathbf{V} + \epsilon$ instead of $\mathbf{V}$ in computing kernel. In the numerical computation, we used $\epsilon = 10^{-10}$. The advantage of using only the principal eigenvectors would be the simplicity of the implementation while the drawback is that the computation of the principal eigenvectors takes a fair amount of time in MATLAB as well as the fact that the method does not completely utilize other eigenvectors. Further the above kernel has a problem in the case of constant field $\mathbf{V} = (1, \cdots, 1)/\sqrt{n}$. In this particular case the kernel is isotropic. A way to avoid this is to introduce the cross-product terms into the matric tensor. Metric like $g_{ij} = V_i V_j$ would give such structure. However such metric tensor is not positive definite.

To address this problem, we formulate the estimation of HH' as the least-squares estimation problem. We wish to find the covariance matrix HH' of the kernel that gives $\mathbf{V}$ as a eigenvector. Without the full set of eigenvalues and eigenvectors, the exact determination of the covariance matrix is not possible. So we estimate the components of HH' by minimizing the sum of the errors, i.e. the least-squares estimation. Solving $HH'\mathbf{V} = \lambda\mathbf{V}$, we have an estimate $\widehat{HH'} = \lambda\mathbf{V}\mathbf{V}^-$, where $\mathbf{V}^-$ is the Moore-Penrose generalized inverse of $\mathbf{V}$. It is trivial to show that for unit column vector $\mathbf{V}^- = \mathbf{V}'$. However $\mathbf{V}\mathbf{V}^-$ singular. To get the positive definite symmetric estimate $\widehat{HH'}$, we add small epsilon in the diagonal term:

$$(HH' - \epsilon I)\mathbf{V} = (\lambda - \epsilon)\mathbf{V}$$

for some $0 < \epsilon < \lambda$. Then taking the Moore-Penrose inverse, we get

$$\widehat{HH'} = \epsilon I + (\lambda - \epsilon)\mathbf{V}\mathbf{V}^-.$$

When ϵ is relatively small compared to λ, i.e. $\lambda - \epsilon \approx \lambda$ so

$$\widehat{HH'} = \epsilon I + \lambda\mathbf{V}\mathbf{V}'. \tag{7}$$

Note that this basically corresponds to matching HH' to the Riemannian metric tensor of the form

$$d\psi^2 = (\epsilon\delta_{ij} + V_i V_j)dx_i dx_j.$$

From Corollary 18.2.11 of Harville (1997), the inverse of $\widehat{HH'}$ is given by

$$\widehat{HH'}^{-1} = \frac{1}{\epsilon}I + \frac{\epsilon - \lambda}{\lambda\epsilon}\mathbf{V}\mathbf{V}' \approx \frac{1}{\epsilon}I - \frac{1}{\epsilon}\mathbf{V}\mathbf{V}'.$$

Also from Corollary 13.7.4 of Harville (1997), the determinant of $\widehat{HH'}$ is approximated as

$$\det(\widehat{HH'}) = \epsilon^n + \frac{\epsilon^{n-1}}{\lambda - \epsilon} \approx \frac{\epsilon^{n-1}}{\lambda}$$

In this case the anisotropic kernel is given by

$$K_\epsilon(\mathbf{x}) = \alpha\frac{(\epsilon\lambda)^{1/2}}{(2\pi\epsilon)^{n/2}}\exp\Big(\frac{\mathbf{x}'(I - \mathbf{v}\mathbf{v}')\mathbf{x}}{2\epsilon}\Big),$$

where α is a normalizing constant to make $K_\epsilon(\mathbf{x})$ a probability density. It can be computed numerically at each voxel. ϵ can be used to control the amount of smoothing.

The above approach is based on constructing the Riemannian metric tensors based on the principal eigenvectors and eigenvalues of the diffusion tensor D. If one want to utilize the full diffusion tensor information that is available in DTI, we note the fact that the inverse of the diffusion tensor

D gives a natural Riemannian metric tensor along the white matter fibers in the brain, i.e. $G = D$. For the rigorous mathematical justification, see De Lara (1995). A similar Riemannian metric tensor approach in connection with DTI can be found in Chefd'hotel *et al.* (2002) and O'Donnell *et al.* (2002). In particular O'Donnell *et al.* (2002) matched the diffusion tensor D to the inverse of the metric tensor G and applied it in computing the tensor-warped distances in the white fibers. We may simply let $HH' = 2tD$ but any reasonable functional relationship between HH' and D such as $HH' = D^2$ or $HH' = (I + D)^2$ can be used for the covariance matrix depending on how one want to smooth data along the diffusion tensor fields. Based on the natural Riemannian metric tensor D of the diffusion, our anisotropic kernel is given by

$$K_t(\mathbf{x}) = \frac{\exp(-\mathbf{x}'D^{-1}\mathbf{x}/4t)}{(4\pi t)^{n/2}(\det D)^{1/2}}.$$

Assuming D is constant everywhere, it can be shown that $K_t * f(\mathbf{x})$ is a solution to an anisotropic diffusion equation

$$\frac{\partial g}{\partial t} = \nabla \cdot (D\nabla g) \tag{8}$$

with the initial condition $g(\mathbf{x}, 0) = f(\mathbf{x})$ after time t. If D is not constant, $K_t * f(\mathbf{x})$ is an approximate solution to (8) in the small neighborhood of $\mathbf{x}$ where D can be considered as constant. The exact solution to the equation (8) with the initial condition $g(\mathbf{x}, 0) = \delta(\mathbf{x})$ has been used as the probabilistic representation of white fiber track connectivity by Batchelor *et al.* (2002). Note the conservation of total probability $\int_{\mathbb{R}^n} g(\mathbf{x}, t)\ d\mathbf{x} = 1$ for all t. So in the diffusion equation approach, at each iteration step, the connectivity probability will always sum up to one and this will be also true for our kernel approach. When D is not constant, we have an adaptive filter where the filter size is given by

$$\text{FWHM} = \frac{2(\ln 4)^{1/2}}{n} tr(2tD)^{1/2} = \frac{(2\ln 4)^{1/2}}{\sqrt{tn}} \sum_{i=1}^{n} \sqrt{\lambda_i},$$

where λ_i are the eigenvalues of D.

4 Transition Probability

Let $P_t(\mathbf{p}, \mathbf{q})$ be the *transition probability density* of a particle going from $\mathbf{p}$ to $\mathbf{q}$ under diffusion process. This is the probability density of the particle hitting $\mathbf{q}$ at time t when the particle is at $\mathbf{p}$ at time 0. Then the *transition probability* of going from point $\mathbf{p}$ to another region $\mathbf{Q}$ is given by

$$P_t(\mathbf{p}, \mathbf{Q}) = \int_{\mathbf{Q}} P_t(\mathbf{p}, \mathbf{x})\ d\mathbf{x}.$$

Note that $\int_{\mathbb{R}^n} P_t(\mathbf{p}, \mathbf{x})\ d\mathbf{x} = 1$. For brain imaging, region $\mathbf{Q}$ would be a collection of voxels and it might possibly consist of a single voxel $\mathbf{p}$. So we will interchangeably use $P_t(\mathbf{p}, \mathbf{q})$ as either transition probability density or transition probability if there is no ambiguity. The transition

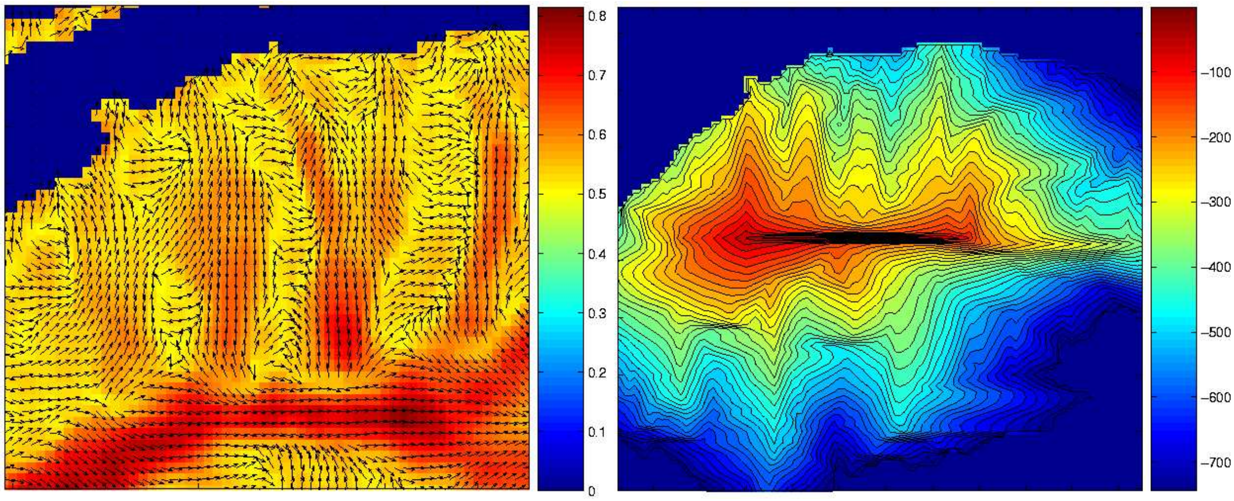

Figure 4: Left: The principal eigenvalues and eigenvectors of the diffusion tensor. The arrows are the principal eigenvectors. The anisotropic kernel technique smooths along the direction of the principal eigenvectors and the amount of the smoothing is related to the principal eigenvalues. Right: The connection probability computed via the kernel smoothing and displayed in logscale. Comparing the left and right figures, we see the log transition probability has higher probability along the white fiber path ways.

probability is the most natural probabilistic measure associated with diffusion process and we will develop our connectivity measure based on the transition probability.

If the diffusion coefficient D is constant in $\mathbb{R}^n$, it can be shown that $P_t(\mathbf{p}, \mathbf{q}) = K_t(\mathbf{q} - \mathbf{p})$. (Stevens, 1995). Since D is varying over the brain regions, it will be only valid when $\mathbf{p}$ and $\mathbf{q}$ are short distance apart and we may take $D(\mathbf{x})$ to be constant in the neighborhood of $\mathbf{x}$.

The transition probability of a particle going from $\mathbf{p}$ to any arbitrary $\mathbf{q}$ is the total sum of the probabilities of going from $\mathbf{p}$ to $\mathbf{q}$ through all possible intermediate point $\mathbf{x} \in \mathbb{R}^n$. Therefore,

$$P_t(\mathbf{p}, \mathbf{q}) = \int_{\mathbb{R}^n} P_s(\mathbf{p}, \mathbf{x}) P_{t-s}(\mathbf{x}, \mathbf{q}) \, d\mathbf{x} \tag{9}$$

for any $0 < s < t$. It is traditionally called the Chapman-Kolmogorov equation (Paul and Baschnagel, 1999). The equation still hold in the case when s is either 0 or t, since it that case one of the probability in the integral will become the Dirac-delta function and in turn the integral collapse to the probability on the left side.

Note that the probability $P(\mathbf{p}, \mathbf{x})$ will decrease exponentially as the distance between $\mathbf{p}$ and $\mathbf{x}$ increase so following a similar argument as in the case of anisotropic kernel smoothing, we approximate the above integral on a small region $B_{\mathbf{p}}$ centered around $\mathbf{p}$. For any point $\mathbf{x} \in B_{\mathbf{p}}$, $P_s(\mathbf{p}, \mathbf{x}) \doteq K_s(\mathbf{p} - \mathbf{x})$. Then for any arbitary two points $\mathbf{p}$ and $\mathbf{q}$,

$$P_t(\mathbf{p}, \mathbf{q}) \doteq \frac{\int_{B_{\mathbf{p}}} K_s(\mathbf{p} - \mathbf{x}) P_{t-s}(\mathbf{x}, \mathbf{q}) \, d\mathbf{x}}{\int_{B_{\mathbf{p}}} K_s(\mathbf{p} - \mathbf{x}) \, d\mathbf{x}}.$$

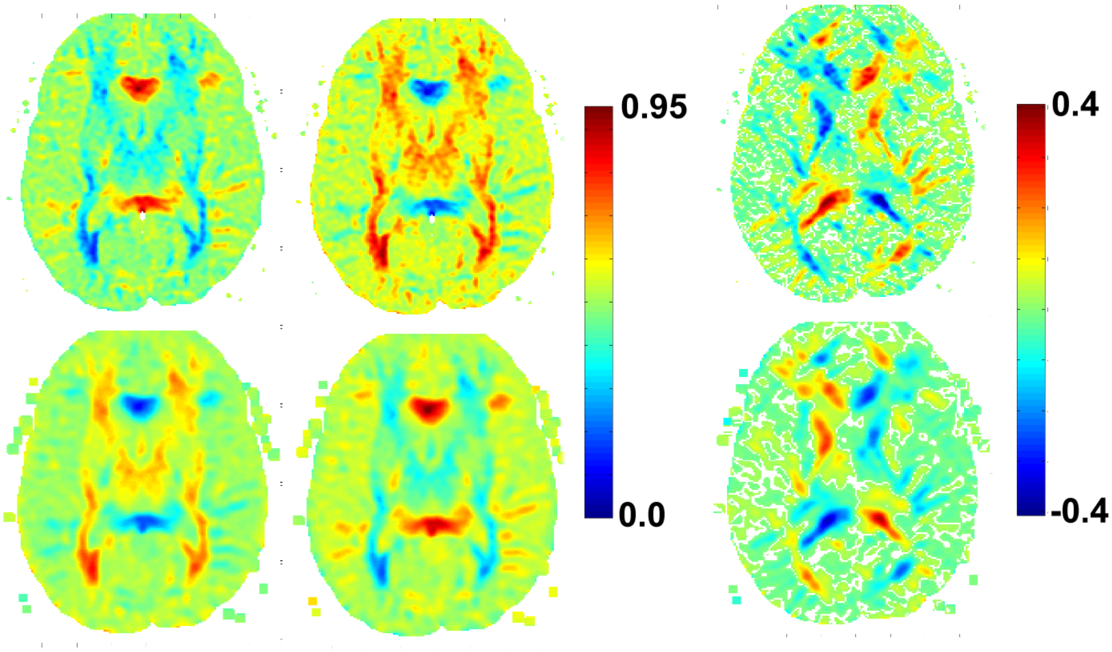

Figure 5: Top: the original diffusion tensors d_{11}, d_{22} and d_{12}. Bottom: the inverse of the diffusion tensors, d^{11}, d^{22} and d^{12}, which are need for computing the anisotropic kernel. Note that d^{ij} have been smoothed via the Cholesky factorization and normalized by the trace of the inverse.

When $s \to 0$, the approximation becomes exact since all the weights of the kernel will be in $B_{\mathbf{p}}$. Again the denominator is a correction term for compensating the underestimation of the numerator. Note that this is the integral version of Gaussian kernel smoothing of data $P_{t-s}(\mathbf{x}, \mathbf{q})$ for given $\mathbf{q}$. Comparing with the formulation of the Gaussian kernel smoothing, we immediately see that for given $\mathbf{p}$,

$$P_t(\mathbf{p}, \mathbf{q}) \doteq \tilde{K}_s * P_{t-s}(\mathbf{p}, \mathbf{q}),$$

where the convolution is with respect to the first argument $\mathbf{p}$. Again note that when $s \to 0$, the equation becomes exact. This formulation is somewhat simpler to solve numerically by iteration than solving the Chapman-Kolmogorov equation. Suppose that $t = N\Delta t$ and $s = \Delta t$. Then we have iteration

$$F_j(\mathbf{q}) = \tilde{K}_{\Delta t} * F_{j-1}(\mathbf{q}), \tag{10}$$

where $F_j(\mathbf{q}) = P_{j\Delta t}(\mathbf{p}, \mathbf{q})$ for a given $\mathbf{p}$ and the initial condition $F_0(\mathbf{q}) = P_0(\mathbf{p}, \mathbf{q}) = \delta(\mathbf{p} - \mathbf{q})$. The reason we get the Dirac-delta function is that the transition probability of a particle at $\mathbf{p}$ hitting any point $\mathbf{a}$ instanenously is zero except when $\mathbf{q} = \mathbf{p}$. The important properties of our iterative procedure is the conservation of the total probability at each iteration. From (10),

$$\int_{\mathbb{R}^n} F_{j+1}(\mathbf{x})\, d\mathbf{x} = \int_{B_{\mathbf{x}}} \tilde{K}_{\Delta t}(\mathbf{x} - \mathbf{y})\, d\mathbf{y} \int_{\mathbb{R}^n} F_j(\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^n} F_j(\mathbf{x})\, d\mathbf{x}.$$

Since $F_1(\mathbf{q}) = \tilde{K}_{\Delta t} * \delta(\mathbf{q}) = \tilde{K}_{\Delta t}(\mathbf{q})$, F_1 is a probability function and it will integrate to one so $\int_{\mathbb{R}^n} F_j(\mathbf{x})\, d\mathbf{x} = 1$. Hence F_j is also a probability function at each iteration. As we run the iteration,

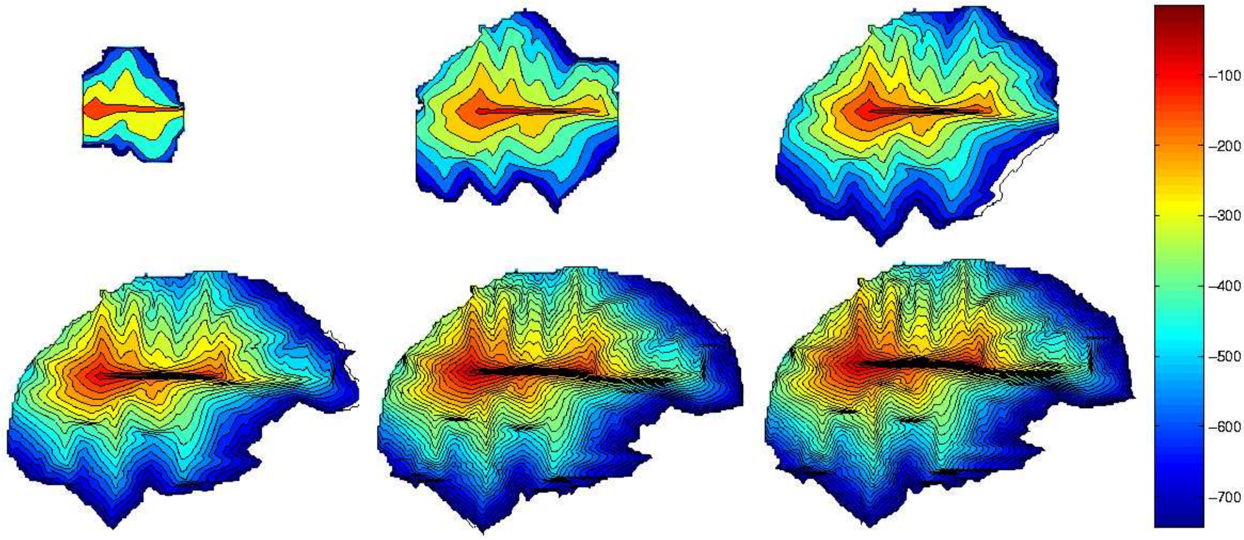

Figure 6: Log-transition probability at 20, 40, 60, 100,180 and 200 iterations respectively showing the anisotropic propagation of the total probability 1. $t = 0.1$ has been used for smoothing with the kernel (6).

the total probability will be dispersed over all region of brain from the seed (Figure 6). If there are one million voxels within the brain, in average, each voxel will have the connection probability of one over a million, which is extremely small. So eventhough the connectivity measure based on the transition probability is a mathematically sound one, it may not be a good one for visualization. So what need is the log-scale of the transition probability, i.e. $\rho = \log P_t(\mathbf{p}, \mathbf{q})$ and we propose this as a probabilistic metric for measuring the strength of the anatomical connecitivy. We will refer this metric as the *log-transition probability*. For simplicity we may let $\mathbf{p} = \mathbf{0}$ and let $\rho(\mathbf{q}) = \log P_t(\mathbf{0}, \mathbf{q})$ for fixed t. If the diffusion coefficient is constant, the log-transition probility can be represented in a simple formula $\rho(\mathbf{x}) = -\mathbf{x}' D^{-1}\mathbf{x} - \sum_{i=1}^{n} \log \lambda_i - \frac{n}{2}\log(4\pi t)$, where λ_i are the eigenvalues of D. When $D = I$, $\rho(\mathbf{x}) = -\mathbf{x}'\mathbf{x} - \frac{n}{2}\log(4\pi t)$. For a region of intest $\mathbf{Q}$, the log-transition probability of reaching $\mathbf{Q}$ would be $\rho(\mathbf{Q}) = \log \int_{\mathbf{Q}} P_t(\mathbf{0}, \mathbf{x})\, d\mathbf{x}$.

5 Results

DTI of normal subjects were obtained using 1.5 tesla SIGNA scanners. A conventional single-shot spin echo EPI pulse sequence was modified to obtain diffusion-weighted (DW) images from any arbitrary set of specified diffusion-weighting directions (Lazar *et al.*, 2003).

Because of noise involved in DTI, smoothing DTI images $D = (d_{ij})$ are necessary but we can not smooth each component d_{ij} of the diffusion tensor D separately. Doing so will violate the positive definiteness. So we perform the Cholesky factorization to D and smooth the elements of the Cholesky factor (Figure 4, 6 and 7). Let $R = (r_{ij})$ be the upper triangular Cholesky factor such that $D = R'R$. Then we smooth each r_{ij} with isotropic kernel with small filter size and reconstruct

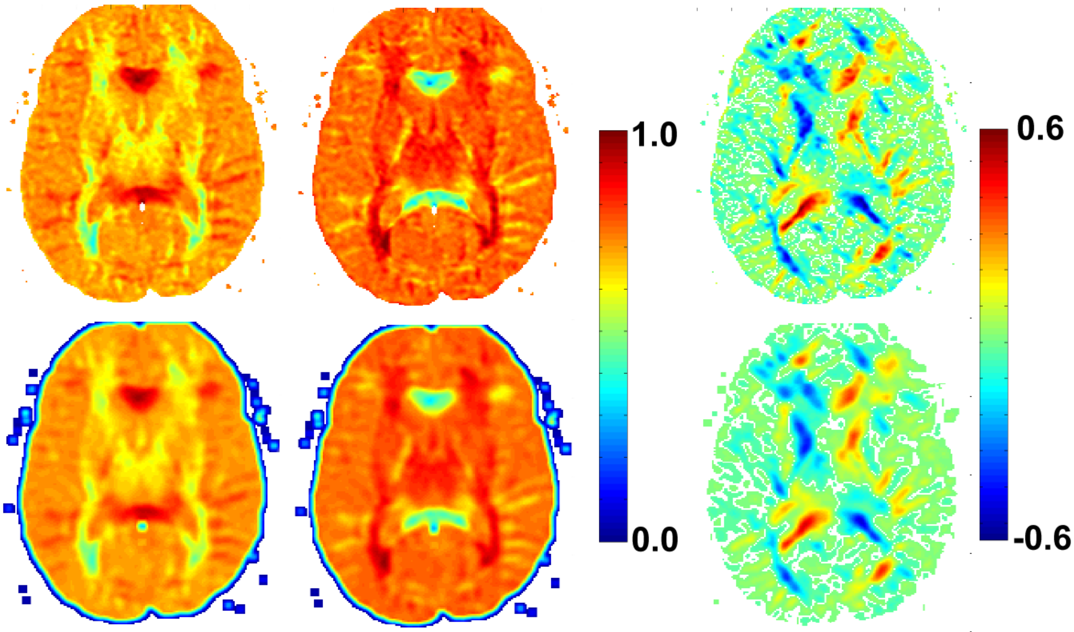

Figure 7: Top: Cholesky factors r_{11}, r_{22}, r_{12} respectively. Bottom: smoothed cholesky factors with 8mm FWHM isotropic Gaussian kernel.

back.

In discretizing the integral version of kernel smoothing (2), we assume that the center of voxels in DTI to be on the integer lattice $\mathbb{Z}^n$. The affine transformation that maps DTI to the lattice $\mathbb{Z}^n$ without changing the maximal connectivity measure is used to simplify the computation. Take $B_{\mathbf{x}}$ to be the collection of voxels incident to voxel $\mathbf{x}$ and $\mathbf{x}$ itself. There are 3^n voxels in $B_{\mathbf{x}}$. The smaller the bandwidth parameter Δt, the more concentrated the weighting is to these incident voxels. The smaller support $B_{\mathbf{x}}$ is used to speed up the computational time. Let's illustrate how the method works in the case of isotropic kernel. Anisotropic kernel case is similar except that the kernel changes from voxel to voxel. The 2D isotropic kernel when $H = I$ for $\Delta t = 0.1$ is given by

$$\begin{pmatrix} 0.0054 & 0.0653 & 0.0054 \\ 0.0653 & 0.7958 & 0.0653 \\ 0.0054 & 0.0653 & 0.0054 \end{pmatrix}.$$

Contributions outside 9 voxels are negligible (less than 0.0001) so we use $\Delta t \leq 0.1$ and 9 voxel neighborhood for our computation. If a more accurate connection probability is desired, one may tempted to use more neighborhood; however, that is not necessary if Δt is very small. Note that $\sum_{\mathbf{x}_j \in B_{\mathbf{x}}} K_{\Delta t}(\mathbf{x}_j) = 1.0785$ while $\sum_{\mathbf{x}_j \in \mathbb{Z}^2} K_{\Delta t}(\mathbf{x}_j) = 1.0787$, about a 0.0002% difference. So even after 500 iterations using the above truncated kernel, there will be only 0.1% difference to true value in the discrete lattice. Since the dimension of DTI is less than 500^3, the probability of

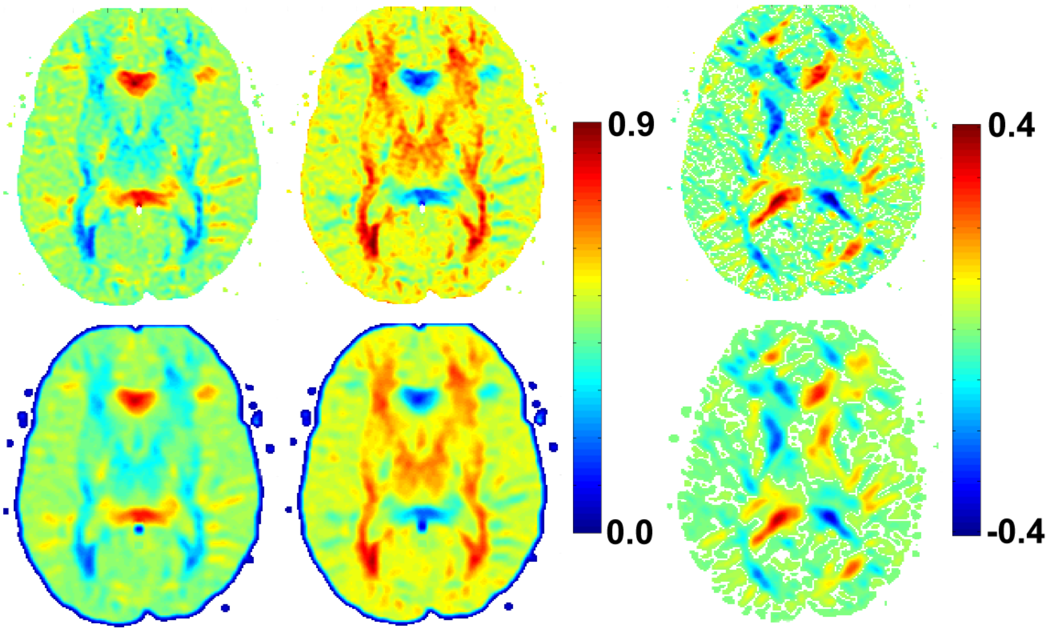

Figure 8: Top: normalized diffusion tensors d_{11}, d_{22}, d_{12} respectively. Bottom: smoothed diffusion tensors. Note that we are smoothing the Cholesky factors r_{ij} and reconstructing back into d_{ij} via $D = R'R$.

connection can be computed in less than 500 iterations. Normalizing the above kernel, we get

$$\tilde{K}_{\Delta t} = \begin{pmatrix} 0.0050 & 0.0606 & 0.0050 \\ 0.0606 & 0.7378 & 0.0606 \\ 0.0050 & 0.0606 & 0.0050 \end{pmatrix}.$$

The discrete version of the Dirac-delta function is the Kronecker's delta. So we let $F_0(\mathbf{q}) = 0$ everywhere except $F_0(\mathbf{p}) = 1$. Then using the discrete convolution, we computed

$$F_{j+1}(\mathbf{q}) = \tilde{K}_{\Delta t} * F_j(\mathbf{q}).$$

At each stage of iteration the total probability is conserved, i.e. $\sum_{\mathbf{q} \in \mathbb{Z}^n} F_j(\mathbf{q}) = 1$.

Discussions

We introduced a novel approach of spatially adaptive anisotropic Gaussian kernel smoothing in representing the white fiber track connectivity via the concept of the transition probability of a diffusion process. Compared to the previous approaches of Monte-Carlo simulation or diffusion equation, our kernel method is simpler to implement. Further our anisotropic kernel method can be used to smooth data along the white fiber tracks to get the continuous and smooth representation of data while preserving the directional characteristic of DTI. So it is hoped that the anisotropic kernel method can be further investigated in relation to the registration, segmentation and other

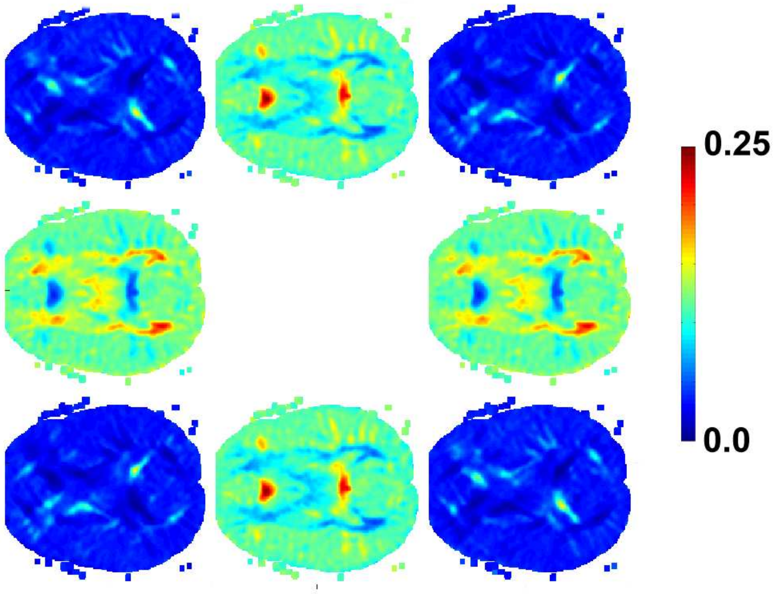

Figure 9: The transition probabilities from the center voxel to neighboring voxels. Using these probabilities as the weights, kernel smoothing is applied as the weighted averaging.

image processing in DTI. Our spatially adaptive anisotropic kernel smoothing is compatible to solving an anisotropic diffusion equation in computational speed since both approaches can be viewed as an iterative adaptive local weighted averaging. To speed up the kernel smoothing, we may use decomposition scheme of Geusebroek in future (Geusebroek *et al.*, 2002).

For simulating DTI for nonparametric inference, one may generate isotropic Gaussian random field by convolving white noise with isotropic Gaussian kernel. Let $w(\mathbf{x})$ be the white noise whose covariance function is of the form $\sigma_w^2\delta(\mathbf{x}-\mathbf{y})$ for some constant σ_w^2 (Dougherty, 1999). Let $K_t(\mathbf{x}) = \exp(-\mathbf{x}'\mathbf{x}/4t)/(4\pi t)^{n/2}$ be the isotropic kernel. Then we generate a Gaussian field via $e(\mathbf{x}) = K_t * w(\mathbf{x})$. The covariance function of e is

$$R_e(\mathbf{x},\mathbf{y}) = \int_{\mathbb{R}^2} K_t(\mathbf{x}-\mathbf{x}')K_t(\mathbf{y}-\mathbf{y}')R_w(\mathbf{x}',\mathbf{y}')\, d\mathbf{x}'\, d\mathbf{y}'.$$

From the property of the Dirac-delta function, the variance of e can be shown to be

$$\sigma_e^2 = \sigma_w^2 \int_{\mathbb{R}} K_t^2(\mathbf{x}-\mathbf{x}')\, d\mathbf{x}'.$$

Since $K_t^2(\mathbf{x}) = K_{t/2}(\mathbf{x})/(8\pi t)^{n/2}$, by integrating the above integral we have $\sigma_e^2 = \sigma_w^2/(8\pi t)^{n/2}$. Based on this relationship, we simulated an isotropic Gaussian field with the specific variance σ_e^2 from the discrete version of white noise with variance σ_w^2. σ_e^2 is estimated from DTI. In our data, all voxels where the diffusion tensor is positive definite, the sample variance for each e_{ij} ranged from 0.006 to 0.009. For simulation we let $\sigma_e^2 = 0.01$ to give a little bit more noise than actual DTI to show the robustness of our connectivity metric.

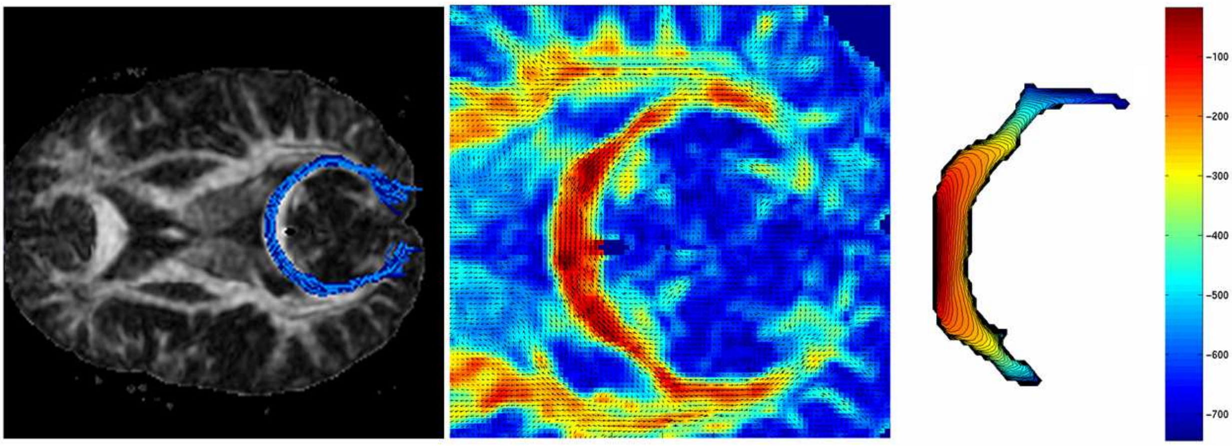

Figure 10: Left: White matter tracking based on the tensor deflection algorithm (Lazar *et al.*, 2003) Middle: Arrows represent the principal eigenvectors. Right: The log transitional probability of connectivity from the seed point taken at the splenium of the corpus callosum. The smoothing was performed in the region FA ≥ 0.2.

If one is interested in inference on the transition probability metric, one would test the hypothesis:

$$H_0 : \rho(\mathbf{x}) = \rho_0 \text{ vs. } H_1 : \rho(\mathbf{x}) > \rho_0$$

for fixed $\mathbf{x}$. An appropriate test statistic would be

$$Z = \frac{\bar{\rho}(\mathbf{x}) - \rho_0}{S(\rho(\mathbf{x}))},$$

where $\bar{\rho}$ and $S(\rho)$ are the sample mean and standard deviation of a sample log-transition probability image $\rho_1, \cdots, \rho_m$ respectively. Unfortunately, our statistic Z will not be the T random fields as defined in Worsley *et al.* (1996) due to the nonlinearity of the kernel smoothing used. For the inference, we need to know the probability distribution of ρ from the distribution of D; however, it is hard to compute the exact distribution analytically. So we may estimate the distribution of ρ and Z via bootstrapping the threshold would be based on computing the quantiles of the estimated distribution.

Acknowledgement

Authors wish to thank Keith Worsley of McGill University and John Ashburner of Functional Imaging Laboratory, London for useful discussions about simulating diffusion tensors. Also thanks is to Sigurd Angenent of the department of mathematics for the discussion of the Riemannian metric tensor that has the cross-product terms.

References

1. Alexander, A.L., Hasan, H.M., Lazar, M., Tsuruda, J.S., Parker, D.L. 2001. Analysis of Partial Volume Effects in Diffusion-tensor MRI. *Magn. Res. Med* **45**:770-780.

2. Anderson, T.W. 1984. *An introduction to multivariate statistical analysis*, Wiley, New York.

3. Ashburner, J. and Friston, K.J. 2000. Voxel-Based Morphometry: The Methods. *NeuroImage* **11**:805-821.

4. Basser, P.J., and Pierpaoli, C. 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B* **111:**209-219.

5. Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A. 2000. In Vivo Tractography using DT-MRI data. *Magn. Res. Med.* **44**:625-632.

6. Batchelor, P.G., Hill, D.L.G., Atkinson, D., Calamanten, F., Atkinson, D. 2001. Study of Connectivity in the Brain Using the Full Diffusion Tensor from MRI, IPMI 2001, *Lecture Notes in Computer Science* **2082**:121-133.

7. Betounes, D. 1998. *Partial Differential Equations for computational Science.* Springer-Verlag New York, Inc.

8. Chefd'hotel, C., Tschumperlé, D., Deriche, R., Faugeras, O. 2002. Constrained flows of matrix-valued functions: Application to diffusion tensor regularization. *Lecture Notes in Computer Science* **2350**:251-265, Springer-Verlag, Berlin.

9. Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., Raichle, M.E. 1999. Tracking Neuronal Fiber Pathways in the Living Human Brain, *Proc. Natl. Acad. Sci.* **96**:10422-10427.

10. De Lara, M.C. 1995. Geometric and Symmetric Properties of a Nondegenerate Diffusion Process, *Annals of Probability*. **23**:1557-1604.

11. Dougherty, E.R. 1998. *Random processes for image and signal processing*. IEEE Press.

12. Hagmann, P., Thiran, J.-P., Vandergheynst, P., Clarke, S., Meuli, R. 2000. Statistical fiber tracking on DT-MRI data as a potential tool for morphological brain studies. *ISMRM Workshop on Diffusion MRI: Biophysical Issues*.

13. Harville, D.A. 1997. *Matrix algebra from a statistician's perspective*, Springer-Verlag New York.

14. Gössl, C., Fahrmeir, L., Pütz, B., Auer, L.M., and Auer, D.P. 2002. Fiber tracking from DTI using linear state space models: delectability of the pyramidal tract. *NeuroImage* **16:**378-388.

15. Geusebroek, J.-M., Smeulders, A. W. M., van de Weijer, J. 2000. Fast Anisotropic Gauss Filtering, *Lecture Notes in Computer Science* **2350**:99-112, Springer-Verlag, Berlin.

16. Koch, M.A., Norris, D.G., and Hund-Georgiadis, M. 2002. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage* **16**:241-250.

17. Lazar, M., Weinstein, D.M., Tsuruda, J.S., Hasan, K.M., Arfanakis, K., Meyerand, M.E., Badie, B., Rowley, H.A., Haughton, V., Field, A., Alexander. 2003. L. White Matter Tractography using Diffusion Tensor Deflection. *Human Brain Mapping* **18**:306-321.

18. O'Donnell, L., Haker, S., Westin, C.-F., New Approaches to Estimation of White Matter Connectivity in Diffusion Tensor MRI: Elliptic PDEs and Geodesics in a Tensor-Warped Space. 2002. Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 02.

19. Paul, W., Baschnagel, J., *Stochastic Processes from Physics to Finance*. Springer-Verlag Berlin.

20. Mori, S., Crain, B., Chacko, V.P., van Zijl, P.C.M. 1999. Three Dimensional Tracking of Axonal Projections in the Brain by Magnetic Resonance Imaging. *Ann. Neurol.* **45**:265-269.

21. Stevens, C.F. 1995. *The six core theories of modern physics*, The MIT Press.

22. Tench, C.R., Morgan, P.S., Blumhardt, L.D., Constantinescu, C. 2002. Improved White Matter Fiber Tracking Using Stochastic Labelling, *Magn. Res. Med.* **48**:677-683.

23. Weinstein, D.M., Kindlmann, G.,Lundberg, E. 1999. Tensorlines: Advection-Diffusion Based Propagation Through Diffusion Tensor Fields. *IEEE Visualization* 249-253. http://www.sci.utah.edu/publications/vis99_tensorlines-paper.pdf

24. Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J. and Evans, A.C. 1996. A unified statistical appraoch for determining significant signals in images of cerebral activation. *Human Brain Mapping* **4**:58-73.