Medical Image Analysis 16 (2012) 876–888

Contents lists available at SciVerse ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

# Nonparametric Bayesian inference of the fiber orientation distribution from diffusion-weighted MR images

Enrico Kaden [a], Frithjof Kruggel [b,*]

[a] Department of Computer Science, University of Leipzig, Johannisgasse 26, 04103 Leipzig, Germany
[b] Department of Biomedical Engineering, University of California, Irvine, 204 Rockwell Engineering Center, Irvine, CA 92697-2715, USA

ARTICLE INFO

*Article history:*
Received 24 July 2011
Received in revised form 10 January 2012
Accepted 17 January 2012
Available online 1 February 2012

*Keywords:*
Spherical deconvolution
Lebesgue decomposition
Dirichlet process mixture
Reversible jump MCMC
Diffusion MR imaging

ABSTRACT

Diffusion MR imaging provides a unique tool to probe the microgeometry of nervous tissue and to explore the wiring diagram of the neural connections noninvasively. Generally, a forward model is established to map the intra-voxel fiber architecture onto the observable diffusion signals, which is reformulated in this article by adopting a measure-theoretic approach. However, the inverse problem, i.e., the spherical deconvolution of the fiber orientation density from noisy MR measurements, is ill-posed. We propose a nonparametric representation of the tangential distribution of the nerve fibers in terms of a Dirichlet process mixture. Given a second-order approximation of the impulse response of a fiber segment, the specified problem is solved by Bayesian statistics under a Rician noise model, using an adaptive reversible jump Markov chain Monte Carlo sampler. The density estimation framework is demonstrated by various experiments with a diffusion MR dataset featuring high angular resolution, uncovering the fiber orientation field in the cerebral white matter of the living human brain.



## 1. Introduction

The human brain forms a complex neural network with a connectional architecture that is still far from being known in full detail, even at the systems-level scale. Diffusion-weighted magnetic resonance (MR) imaging has enabled us to probe the microgeometry of nervous tissue noninvasively and to explore the wiring scheme of the neural connections in the individual living subject (Le Bihan et al., 1986). This technique measures the diffusion process of the spin-bearing water molecules in the underlying sample material. Depending on the direction along which the Brownian motion is observed, the microscopic tissue structure, especially the orientation of the axonal membranes, may hinder the diffusing molecules differently. As a result, the MR measurement exhibits anisotropic diffusion patterns if the nerve fibers are coherently oriented and the examination time is chosen sufficiently long. Moreover, a distinctive feature of brain organization is that the white matter fiber pathways are neither regularly ordered nor entirely arbitrary, but are grouped into coherent bundles. Therefore, it is interesting to know how the tangents at the fiber streamlines are oriented within a small environment. This knowledge about the intra-voxel tissue architecture allows to reconstruct the course of the nerve fibers. Subsequently, we can study the neural circuitry by means of interregional connections, which integrate the diversity of cortical areas and subcortical nuclei with their specialized functional roles.

For this purpose a biophysical model is established that maps the tissue geometry of white matter onto the diffusion process of water molecules and further onto the observable MR signals. Let us consider a population of nerve fibers located within a voxel. Assuming that the impulse response of a fiber segment is known up to its orientation, the spherical convolution with the tangential distribution of the fiber pathways yields the diffusion signal (von dem Hagen and Henkelman, 2002). The objective is then to estimate the fiber orientation distribution from a finite sample of noisy MR observations via spherical deconvolution (Tournier et al., 2004; Anderson, 2005), which is an ill-posed problem. Henceforth, we shall focus on recovery techniques that preserve the characteristic properties of a density function, which is nonnegative and integrates to one. For instance, Alexander (2005) proposed a maximum entropy method to solve the inverse problem. The fiber orientation density may be also represented by a finite mixture of Bingham distributions (Kaden et al., 2007), which provides an explanatory parameterization of the intra-voxel fiber architecture (e.g., the spread of the fiber direction within a bundle). Alternatively, the solution to this variational problem is found in a reproducing kernel Hilbert space of square-integrable functions on the sphere whose harmonic coefficients decay sufficiently fast (Kaden et al., 2008). This nonparametric function estimation generalizes previous linear methods truncating the spherical harmonic expansion

* Corresponding author.
*E-mail addresses:* enrico.kaden@gmail.com (E. Kaden), fkruggel@uci.edu (F. Kruggel).

1361-8415/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2012.01.004

and preserves the defining properties of the density function (i.e., its nonnegativity and normalization). Most recently, Patel et al. (2010) developed a mesh-based deconvolution framework. We refer the reader to Ramirez-Manzanares et al. (2007), Jian and Vemuri (2007), Tournier et al. (2007), Leow et al. (2009) and Yeh et al. (2011) for related reconstruction techniques, which may not guarantee all characteristic properties and/or estimate spherical functions only loosely connected to the fiber orientation distribution. See also Assemlal et al. (2011) for a comprehensive review.

The present article revisits the forward model by adopting a measure-theoretic perspective. In particular, we shall ask for an appropriate mathematical structure in order to describe the tangential distribution of the fiber streamlines. Indeed, measure theory (Bogachev, 2007) provides a rigorous definition of the fiber orientation distribution as a measure normalized with one, showing that previous approaches constitute special cases which are generalized in this work. The Lebesgue decomposition then allows us to categorize the measure into discrete, absolutely continuous, and singular continuous components, which each are illustrated by examples found in the existing literature. Additionally, we attempt to clarify the meaning of the (fiber orientation) density function and that the tangential distribution of the nerve fibers is not a probability measure, although this misconception about the fiber orientation distribution has gained some acceptance. Another objective is to define a dictionary in terms of which the measure is represented, since the general measure space is by far too large in order to be computationally manageable. In this paper we propose a nonparametric description of the fiber orientation distribution via a Dirichlet process mixture (Ferguson, 1973; Lo, 1984), which models a probability space on the measure space. Its constructive representation shows that these random distributions are absolutely continuous and fulfill the antipodal symmetry property. Thus, we will be able to generate synthetic fiber orientation distributions that closely resemble the observed microgeometry of white matter. More importantly, the Dirichlet process mixture encodes our prior knowledge about the tangential distribution of the nerve fibers.

After the exact definition of the fiber orientation distribution, a novel computational method is developed for the nonparametric estimation of the unknown spherical density. Given a second-order approximation of the impulse response of a fiber segment with specified diffusivity parameters, we formulate a statistical model that maps the intra-voxel tissue geometry onto the observable diffusion signals. The inverse problem aims to estimate the fiber orientation density, that is, an absolutely continuous measure described by a Dirichlet process mixture, from a finite set of noisy MR measurements, which is performed by Bayesian inference under a Rician noise model. The spin echo signal without diffusion weighting (i.e., the $T_2$-contrast) is integrated into the estimation framework. We also pay attention to the noise alterations due to the linear interpolation that is required for the correction of subject motion and the co-registration with an anatomical coordinate system. Subsequently, our paper introduces an original approach to sampling the posterior of this Bayesian model. The proposed technique consists of a finite approximation of the Dirichlet process through a discrete multivariate distribution (Ewens, 1972) and an adaptive version of the reversible jump Markov chain Monte Carlo (MCMC) sampler (Green, 1995), which might be interesting beyond the statistical analysis of diffusion MR images. Note that all characteristic properties of the fiber orientation distribution, namely its antipodal symmetry, nonnegativity, and normalization, will be ensured.

This article is organized as follows. Section 2 reformulates a recently proposed spherical convolution model from a measure-theoretic viewpoint. A major contribution presents the Lebesgue decomposition of the fiber orientation distribution. Section 3 focuses on the absolutely continuous component of the measure, which is here modeled by a Dirichlet process mixture. More technically, the stochastic process is represented through a stick-breaking model with bipolar Watson densities. Additionally, we sample from the Dirichlet process mixture and illustrate the simulated fiber orientation densities, which are very similar to the observable tissue geometry. Section 4 introduces the nonparametric density estimation via Bayesian inference. Weakly-informative priors are chosen that make as few assumptions about the tangential distribution of the nerve fibers as possible, in an attempt to let the data speak for themselves. An appendix addresses the computational issues of Bayesian nonparametrics. The Results section demonstrates the reliable estimation of the fiber orientation field even for three crossing fiber bundles, namely the corona radiata, the callosal fibers, and the superior longitudinal fasciculus, using a diffusion MR dataset featuring high angular resolution. We conclude with a discussion of the proposed models and computational methods, including a thorough comparison with parametric Bayesian inference and Gaussian process models.

## 2. Spherical convolution

Diffusion MR imaging aims to reveal quantitative information about the directional structure of nervous tissue. The task may be divided into two parts, that is, data synthesis and data analysis. In this section we establish a forward-generative model that maps the intra-voxel tissue geometry onto the observable diffusion signal.

### 2.1. Biophysical model

Human brain white matter is composed of (myelinated) axons, which are organized in bundles called fascicles, and an extraaxonal compartment including glial cells such as oligodendrocytes. Consider the infinitesimal neighborhood of a nerve fiber at the point $x \in \mathbb{R}^3$ that is oriented by the tangent vector $\omega(x) \in S^2$, where $S^2 = \{\omega \in \mathbb{R}^3 : \|\omega\| = 1\}$ denotes the two-dimensional unit sphere. If the fiber pathways are sufficiently smooth, a small fiber section (in the range of micrometers) resembles a portion of a cylindrical tube which is rotationally symmetric. Under the Gaussian phase assumption (Neuman, 1974), the diffusion signal of this fiber segment—including the surrounding volume typical of an axon, which is always present, even when the fibers abut—may be approximated by

$$h_b(g,\omega) = \exp(-b[(\lambda_\parallel - \lambda_\perp)\langle g,\omega\rangle^2 + \lambda_\perp]) \tag{1}$$

up to the second order. $\lambda_\parallel$ and $\lambda_\perp$ denote the effective longitudinal and transverse diffusion coefficients, respectively, with the constraint $\lambda_\parallel \geqslant \lambda_\perp \geqslant 0$, as the axonal membranes perpendicular to the fiber orientation $\omega \in S^2$ form the major structural barriers that confine the diffusing water molecules which explore the biological tissue by random walk. The MR experiment, whose sequence timing (i.e., the temporal profile of the diffusion sensitizing gradients) is kept fixed, is controlled by the diffusion weighting factor $b \geqslant 0$ and the normalized gradient direction $g \in S^2$. The impulse response function $h_b$ takes its values in the interval $[0,1]$ and is antipodally symmetric, that is, $h_b(g,\omega) = h_b(-g,\omega)$ for all $g, \omega \in S^2$. The apparent diffusivity parameters $\lambda_\parallel$ and $\lambda_\perp$ of a single fiber should not vary too much throughout the cerebral white matter. Compare also Kaden et al. (2008) for a generalization of this constant diffusivity assumption.

Henceforth, we assume that the impulse response $h_b$ of a small fiber segment is known up to its orientation $\omega \in S^2$. An important remark is that this function does not depend on the particular location of the fiber within a voxel. Let the two-dimensional unit

sphere $S^2$ be equipped with the Borel $\sigma$-algebra $\mathcal{B}$. The fiber orientation distribution $\mu : \mathcal{B} \to [0, \infty]$ defines a measure with $\mu(S^2) = 1$ (Bogachev, 2007) which quantifies the relative frequency of specific fiber orientations over a given Borel set. Note that $\mu$ resembles a probability distribution from a mathematical point of view, but has a different, namely biophysical meaning. Consider a population of nerve fibers within a voxel. Since the diffusion process between different compartments, whose $T_2$-relaxation should be similar, is supposed to be in slow exchange, the spherical convolution of the response function $h_b$ with the fiber orientation distribution $\mu$,

$$\frac{E_b(g)}{E_0} = P_{\text{iso}} \exp(-b\lambda_{\text{iso}}) + (1 - P_{\text{iso}}) \int_{S^2} h_b(g, \omega) d\mu(\omega), \tag{2}$$

yields the observable MR signal for this voxel (cf. von dem Hagen and Henkelman, 2002). $E_b(g)$ is the spin echo signal with the diffusion weighting $b \geqslant 0$ and $g \in S^2$, $E_0$ denotes the signal in the absence of diffusion encoding, which is required for the normalization of the $T_2$-contrast. The diffusion signal $e_b(g) = E_b(g)/E_0$ lies in the interval $[0,1]$ and is antipodally symmetric, i.e., $e_b(g) = e_b(-g)$ for all $g \in S^2$. The first term on the right-hand side approximates the isotropic signal for the space between the nerve fibers (excluding the extracellular space in the immediate neighborhood of an axon), which is filled with glial cells, interstitial fluid containing the extracellular matrix, etc. $P_{\text{iso}} \in [0,1]$ denotes its volume fraction and $\lambda_{\text{iso}}$ is the apparent diffusion coefficient. Therefore, anisotropic patterns observable in the MR signal are a consequence of the coherent orientation of the nerve fibers within a voxel, the dense packing of the (myelinated) axons in the white matter, and that the transverse diffusion coefficient of a fiber segment is smaller than the longitudinal diffusivity parameter due to the barriers formed by the axonal membranes.

### 2.2. Lebesgue decomposition

Next, we decompose the fiber orientation distribution into three different types of measures. If a distribution does not consists of any two of these principal components, it can be shown that the tangential distribution of the nerve fibers has to be of the remaining type. Let $\lambda$ denote the Haar measure $\lambda : \mathcal{B} \to [0, \infty]$ normalized with $\lambda(S^2) = 4\pi$, which is invariant with respect to the orientation-preserving rotations. According to Lebesgue's decomposition theorem (Bogachev, 2007), the fiber orientation distribution may be uniquely decomposed as $\mu = \mu_{\text{d}} + \mu_{\text{ac}} + \mu_{\text{sc}}$ with respect to $\lambda$. These components are defined as follows:

(i) The discrete measure $\mu_{\text{d}}$ forms the pure point component with a countable collection

$$\mu_{\text{d}}(S^2) = \sum_{\omega \in S^2} \mu_{\text{d}}(\{\omega\}),$$

which means that $\mu_{\text{d}}$ takes all its mass at individual points. For example, a Dirac measure $\delta_\omega$ describes the situation where the fibers in a voxel run parallel to each other, thus have the same tangent vector $\omega(x)$ at each point $x$. See Section 3.1 for the Dirichlet process model, but compare also Ramirez-Manzanares et al. (2007) who proposed a related approach.

(ii) $\mu_{\text{ac}}$ is absolutely continuous with respect to $\lambda$ if $\lambda(A) = 0$ implies $\mu_{\text{ac}}(A) = 0$ for every Borel set $A \in \mathcal{B}$. According to the Radon–Nikodým theorem, there exists a nonnegative Borel-measurable function $p$ such that

$$\mu_{\text{ac}}(A) = \int_A p(\omega) d\omega$$

for all $A \in \mathcal{B}$. $p \in L^1(S^2)$ is also called (fiber orientation) density. Whereas $\mu_{\text{ac}}$ is unique, the density function, which fulfills $p(\omega) \geqslant 0$, $\omega \in S^2$ and $\int_{S^2} p(\omega) d\omega = 1$, is only unique up to a set of Haar measure 0. Making further assumptions on the shape of this function, the literature suggests different approaches to recovering the fiber orientation density (e.g., Alexander, 2005; Kaden et al., 2007, 2008). See Section 3.2 for the Dirichlet process mixture.

(iii) $\mu_{\text{sc}}$ is the singular continuous measure with $\mu_{\text{sc}}(\{\omega\}) = 0$ for all $\omega \in S^2$, but there exists a set $A \subset S^2$ with Haar measure $\lambda(A) = 0$ and $\mu_{\text{sc}}(A^c) = 0$. $A^c$ denotes the complement of $A$ in $S^2$. An example for the singular continuous part is a bundle of parallel running fibers that is smoothly arched in a specific direction. Savadjiev et al. (2006) proposed to model the fiber pathways using a helicoidal model, which results into a fiber orientation distribution that is singular continuous.

Since the sign of the orientation of a fiber segment is arbitrary, the two opposite points $\omega$ and $-\omega$ on $S^2$ may be identified. The quotient manifold $\mathbb{R}P^2 \cong S^2/-\text{Id}$ of the two-dimensional unit sphere under the antipodal involution $\omega \mapsto -\omega$ is the real projective plane. Without loss of generality, $\mu$ is set to be antipodally symmetric, which means that a fiber segment is described as a line without any marked direction. Obviously, the fiber orientation density is also antipodally symmetric, i.e., $p(\omega) = p(-\omega)$ for all $\omega \in S^2$.

## 3. Dirichlet process mixture

The space of all measures is by far too large in order to be computationally manageable, even though the previous categorization of the fiber orientation distribution gives us important insight into its principle types. This section proposes two different dictionaries in terms of which the measure may be represented, choosing their support as large as possible.

### 3.1. Stick-breaking representation

The objective is to define a probability measure on the space of all fiber orientation distributions. In this work we propose that the tangential distribution of the nerve fibers

$$\mu \sim \mathcal{DP}(\alpha, H)$$

follows a Dirichlet process on the two-dimensional unit sphere $S^2$ equipped with the Borel $\sigma$-algebra $\mathcal{B}$. A stochastic process is called Dirichlet (Ferguson, 1973) if for any finite partition $A_1, \ldots, A_n \in \mathcal{B}$ of the sphere (i.e., the $A_i$ are measurable, disjoint, and $\bigcup_{i=1}^n A_i = S^2$) the random vector

$$(\mu(A_1), \ldots, \mu(A_n)) \sim \mathcal{D}(\alpha H(A_1), \ldots, \alpha H(A_n))$$

has an $n$-dimensional Dirichlet distribution. $\alpha > 0$ denotes a concentration parameter and $H$ the spherical base measure, which is assumed to be nonatomic (i.e., $H(\{\omega\}) = 0$ for all $\omega \in S^2$), for example, the uniform distribution $\mathcal{U}_{S^2}$. For every Borel set $A \in \mathcal{B}$ the expectation is $\mathbb{E}[\mu(A)] = H(A)$ and the variance writes $\text{var}[\mu(A)] = H(A)(1 - H(A))/(\alpha + 1)$. Note that $\mu$ is a discrete measure with probability one and thus describes the pure point part of the fiber orientation distribution. However, the realizations of the Dirichlet process are generally not antipodally symmetric, which is a characteristic property of the tissue microgeometry. Let $G = \{e, g\}$ be a finite group of measurable transformations with the identity element $e(\omega) = \omega$ and the antipodal involution $g(\omega) = -\omega$ for each $\omega \in S^2$. Following Dalal (1979),

$$\mu \sim \mathcal{DP}_G(\alpha, H)$$

is governed by a $G$-invariant Dirichlet process if the base measure $H$ is invariant with respect to the action of $G$ and for any finite $G$-invariant partition the resulting random vector has a Dirichlet

distribution as defined above. Hence, the random measure $\mu$ is antipodally symmetric.

An equivalent description of the Dirichlet process is the stick-breaking representation (Sethuraman, 1994), which takes the form

$$\mu = \sum_{i=1}^{\infty} \pi_i \delta_{\omega_i}$$

with the random weights $\pi_i = X_i \prod_{j=1}^{i-1}(1 - X_j)$ and $X_i \sim \mathcal{B}e(1, \alpha)$. The directions $\omega_i \sim H$ are distributed according to the base measure independent of $\pi_i$, where $\mathcal{B}e$ denotes a Beta distribution and $\delta_{\omega_i}$ the Dirac measure at $\omega_i \in S^2$. A $G$-invariant version writes

$$\mu = \sum_{i=1}^{\infty} \pi_i \frac{\delta_{\omega_i} + \delta_{-\omega_i}}{2}, \tag{3}$$

which pays attention to the fact that the tangential distribution of the nerve fibers is antipodally symmetric. The two weighted sums of point masses highlight the almost sure discreteness of the Dirichlet process. Further, it is easy to see that the characteristic properties of the fiber orientation distribution, namely antipodal symmetry (for the $G$-invariant version only), nonnegativity, and normalization, are fulfilled. Setting this model of the tangential distribution into the spherical convolution model (2), we obtain the forward model

$$\frac{E_b(g)}{E_0} = P_{\text{iso}} \exp(-b\lambda_{\text{iso}}) + (1 - P_{\text{iso}}) \sum_{i=1}^{\infty} \pi_i \exp(-b[(\lambda_{\parallel} - \lambda_{\perp}) \times \langle g, \omega_i \rangle^2 + \lambda_{\perp}]) \tag{4}$$

with the random weights $\pi_i \in [0,1]$ and $\sum_{i=1}^{\infty} \pi_i = 1$, which yields the observable MR signals. Note that both versions of the stick-breaking representation give the same diffusion signal, since the impulse response of a fiber segment is antipodally symmetric.

### 3.2. Infinite mixture model

Another approach is to model the fiber orientation distribution as an absolutely continuous measure (with respect to Haar measure). Consequently, there exists a fiber orientation density $p \in L^1(S^2)$ with the following properties: antipodal symmetry $p(\omega) = p(-\omega)$, $\omega \in S^2$, nonnegativity $p(\omega) \geqslant 0$, $\omega \in S^2$, and normalization $\int_{S^2} p(\omega) d\omega = 1$. Before proceeding, we introduce the Watson density which is defined as

$$f_{\mathcal{W}}(\omega; v, \kappa) = \frac{1}{c_{\mathcal{W}}(\kappa)} \exp(\kappa \langle \omega, v \rangle^2)$$

on the sphere $\omega \in S^2$ with the mean orientation $v \in S^2$ and the scaling parameter $\kappa \in \mathbb{R}$ (Watson, 1965). The normalization constant writes $c_{\mathcal{W}}(\kappa) = 4\pi {}_1\mathrm{F}_1(1/2; 3/2; \kappa)$, where ${}_1\mathrm{F}_1$ denotes the confluent hypergeometric function of a scalar argument. The Watson density $f_{\mathcal{W}}(\cdot; v, \kappa)$ can be regarded as a trivariate Gaussian distribution with a zero mean and the covariance matrix $-1/2(\kappa v v^t)^{-1}$ conditional upon $S^2$. Obviously, this density function is antipodally symmetric, nonnegative, and integrates to one over the sphere. We may differentiate the following cases with respect to the scaling parameter $\kappa$. If $\kappa$ is greater than zero, the distribution is bipolar. In the case of $\kappa < 0$, $f_{\mathcal{W}}$ denotes an equatorial density around the axis $v$. If $\kappa = 0$, then we obtain the uniform distribution on $S^2$. In this work we constrain the scaling parameter $\kappa \geqslant 0$ and consider only bipolar Watson densities. To define a fiber orientation distribution that has a density function, we propose a Dirichlet process mixture (Lo, 1984)

$$p(\omega) = \int_{S^2 \times \mathbb{R}^+} f_{\mathcal{W}}(\omega; v, \kappa) d\rho(v, \kappa), \tag{5}$$

where $\rho \sim \mathcal{DP}(\alpha, H)$ follows a Dirichlet process with the base measure $H$, here the product measure formed by the uniform distribution $\mathcal{U}_{S^2}$ for the parameter $v$ and the inverse Gamma distribution $\mathcal{IG}(\alpha_\kappa, \beta_\kappa)$ for the parameter $\kappa$. Although $\rho$ is a discrete random measure, the integration with a continuous function gives rise to a prior on the space of all fiber orientation densities. Note that both versions of the Dirichlet process, that is, the unconstrained and the $G$-invariant realization, yield the same tangential distribution due to the antipodal symmetry of the Watson density.

Since the Dirichlet process has a stick-breaking representation, the fiber orientation density model (5) can be rewritten as

$$p(\omega) = \sum_{i=1}^{\infty} \pi_i f_{\mathcal{W}}(\omega; v_i, \kappa_i), \tag{6}$$

which is a countably infinite mixture of Watson densities. The $\pi_i$ are given in the previous section, while $v_i \sim \mathcal{U}_{S^2}$ and $\kappa_i \sim \mathcal{IG}(\alpha_\kappa, \beta_\kappa)$ are distributed according to the base measure. Under the topology of weak convergence, the proposed mixture model includes all tangential distributions of the nerve fibers in its closure (Bogachev, 2007), which particularly means that any antipodally symmetric density function can be approximated by the Dirichlet process mixture as closely as desired. Therefore, the mixture model (5) with its stick-breaking representation forms a probability measure on the space of all fiber orientation distributions whose support is as large as possible with respect to the weak topology. It is easy to see that the characteristic properties of the fiber orientation density (i.e., antipodal symmetry, nonnegativity, and normalization) are fulfilled. Setting this model of a density function into the spherical convolution model (2), we obtain the forward model[1]

$$\frac{E_b(g)}{E_0} = P_{\text{iso}} \exp(-b\lambda_{\text{iso}}) + (1 - P_{\text{iso}}) \exp(-b\lambda_{\perp}) \sum_{i=1}^{\infty} \pi_i \times \frac{{}_1\mathrm{F}_1(1/2; 3/2; \kappa_i v_i v_i^t - b(\lambda_{\parallel} - \lambda_{\perp}) g g^t)}{{}_1\mathrm{F}_1(1/2; 3/2; \kappa_i)}, \tag{7}$$

which generates the observable MR signals. ${}_1\mathrm{F}_1$ is the confluent hypergeometric function of a matrix argument, here the real 3-by-3 symmetric matrix $\kappa_i v_i v_i^t - b(\lambda_{\parallel} - \lambda_{\perp}) g g^t$. See Appendix A.1 for more information about the special function. Note that the nonparametric description of the fiber orientation density yields a closed-form expression for the diffusion signal. In the case of $\kappa_i \to \infty$, the infinite mixture model (6) reduces to the $G$-invariant Dirichlet process (3) and the diffusion signal (7) simplifies to Eq. (4). Fig. 1 exemplifies three samples of the fiber orientation density drawn from the Dirichlet process mixture with the parameters $\alpha = 1.5$, $\alpha_\kappa = 17.5$, and $\beta_\kappa = 184.8$ (with maximum density at $\kappa_{\text{mode}} = 10$ and $\mathbb{P}[\kappa \leqslant 20 | \alpha_\kappa, \beta_\kappa] = 0.99$). The lower row shows the simulated diffusion signal with a $b$-value of 1500 s/mm$^2$. As a result, the proposed mixture model presents an appropriate method to simulate fiber orientation distributions that closely resemble the observed tissue geometry of white matter.

## 4. Nonparametric Bayesian inference

Henceforth, we assume that the fiber orientation distribution is an absolutely continuous measure that has a density function. This is a reasonable assumption because histological studies with confocal laser microscopy (Axer et al., 2001) suggest that the fiber architecture of white matter is rather heterogeneous, even in the body of the human corpus callosum. In the following we estimate the tangential distribution of the nerve fibers via Bayesian data

[1] Compare also Kaden et al. (2008) who proposed to model the fiber bundles through a finite mixture of Bingham distributions (i.e., a generalization of the bipolar Watson densities). Here, the mixture components are merely base elements for the close approximation of a generic density function, which is discussed further in Section 6.

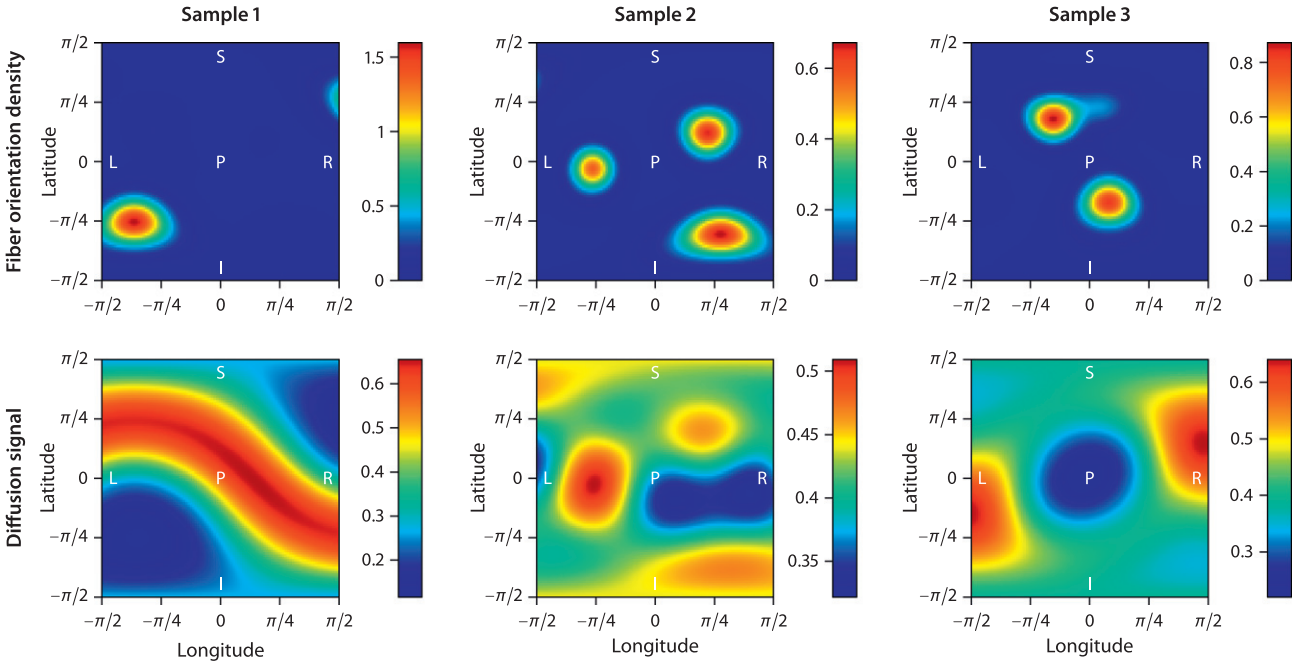


**Fig. 1.** Simulation of the fiber orientation density using a Dirichlet process mixture, which is exemplified for three samples. The bottom row shows the corresponding diffusion signal with a $b$-value of 1500 s/mm$^2$. The nonfiber compartment is neglected ($P_{\text{iso}} = 0$). The water diffusivity parameters of a fiber segment are set to $\lambda_{\parallel} = 0.0018$ and $\lambda_{\perp} = 0.0002$ mm$^2$/s. Abbreviations: left (L), right (R), inferior (I), superior (S), anterior (A), posterior (P).

analysis, using a finite approximation through a discrete multivariate distribution and an adaptive reversible jump MCMC sampler.

### 4.1. Ewens distribution

Bayesian statistics often give rise to a number of computational issues. In this work we have to deal with the infinite-dimensional character of Dirichlet process mixtures. Consider the infinite random vector

$$\pi = (\pi_1, \pi_2, \ldots)$$

with $0 \leqslant \pi_i \leqslant 1$ and $\sum_{i=1}^{\infty} \pi_i = 1$, which represents the partition structure of the interval $[0,1]$. Note that $\pi$ is independent of the data sequence $\theta_i = (v_i, \kappa_i)$ for $i = 1, 2, \ldots$, which is drawn from the base measure. The objective now is to find a finite-dimensional approximation to $\pi$ that enables Bayesian density estimation in a numerically convenient manner. Let $\theta = (\theta_1, \ldots, \theta_n)$ be a finite sample of a random distribution that is drawn from a Dirichlet process $\mathcal{DP}(\alpha, H)$ with the concentration parameter $\alpha > 0$. Even though the base measure $H$ is assumed to be absolutely continuous, the probability of the event $\{\theta_i = \theta_j\}$ for two samples with indices $i \neq j$ is greater than zero, which can be easily seen from the Pólya urn scheme (Blackwell and MacQueen, 1973). So we may construct a finite discrete-valued version of $\pi$ as follows: Suppose that there are exactly $k \leqslant n$ distinct values in the sample $\theta$. The finite random vector

$$\pi^{[n]} = \left(\pi_1^{[n]}, \ldots, \pi_k^{[n]}\right)$$

counts the number $n\pi_i^{[n]}$ (for $i = 1, \ldots, k$) of how many times the $i$th value appears in $\theta$. This quantization with $\pi_i^{[n]} \in \{q/n \in \mathbb{Q} : q = 1, \ldots, n\}$ satisfies $\sum_{i=1}^{k} \pi_i^{[n]} = 1$, where $1/n$ is the smallest fraction that can be represented. The finite discrete-valued random vector $\pi^{[n]}$ converges to $\pi$ with probability one if the sample size $n$ approaches infinity, which means that for any desired accuracy we can choose a number $n$ such that $\pi^{[n]}$ approximates $\pi$ within a certain threshold.

Alternatively, the sample $\theta$ of size $n$ may be described by the finite-dimensional vector $\mathfrak{C}(m_1, \ldots, m_n)$ of nonnegative integers which quantify that $m_1$ distinct values appear only once in $\theta$, $m_2$ distinct values occur exactly twice, and so on. Then $n = \sum_{i=1}^{n} i m_i$ provides the number of samples and $k = \sum_{i=1}^{n} m_i$ the partition size. In the following we are interested in the probability that a sample $\theta$ of size $n$ and thus the random vector $\pi^{[n]}$ belong to the class $\mathfrak{C}(m_1, \ldots, m_n)$, which is invariant with respect to the permutation of the indices in $\theta$ and $\pi^{[n]}$. The event $\{\pi^{[n]} \in \mathfrak{C}(m_1, \ldots, m_n)\}$ is governed by a discrete multivariate distribution (Ewens, 1972; Antoniak, 1974), also known as Ewens measure, which is defined by

$$f_{\mathcal{E}}(\mathfrak{C}(m_1, \ldots, m_n); \alpha, n) = \frac{\alpha^k}{\alpha^{(n)}} \prod_{i=1}^{n} \frac{i}{i^{m_i} m_i!},$$

where $\alpha > 0$ denotes the concentration parameter, $n$ is the number of samples, and $\alpha^{(n)} = \alpha(\alpha + 1) \cdot \ldots \cdot (\alpha + n - 1)$ defines the Pochhammer symbol. The probability that we observe exactly $k$ distinct values in a sample of size $n$ is given by

$$\mathbb{P}[k|\alpha, n] = \frac{\alpha^k}{\alpha^{(n)}} |S_n^{(k)}|.$$

$|S_n^{(k)}|$ is the absolute value of a Stirling number of the first kind. The expectation writes $\mathbb{E}[k|\alpha, n] = \alpha(\Psi(\alpha + n) - \Psi(\alpha))$, where $\Psi$ denotes the digamma function. It can be shown that $\lim_{n\to\infty} \alpha \ln(1 + n/\alpha)/\mathbb{E}[k|\alpha, n] = 1$ holds, which means that the expected partition size $k$, and hence the required storage space for $\pi^{[n]}$, grows logarithmically with the sample size $n$. As a consequence, we arrive at the discretized version of the Dirichlet process mixture (6), that is,

$$p(\omega) = \sum_{i=1}^{k} \pi_i^{[n]} f_{\mathcal{W}}(\omega; v_i, \kappa_i), \tag{8}$$

where $v_i \sim \mathcal{U}_{\mathcal{S}^2}$ and $\kappa_i \sim \mathcal{IG}(\alpha_\kappa, \beta_\kappa)$ are drawn from the base measure. The other stochastic processes and the corresponding diffusion signals are approximated in a similar way.

### 4.2. Statistical model

The Bayesian paradigm provides an inferential framework for statistical modeling and analysis, enabling the differentiation between relevant features of the tissue geometry and unwanted artifacts due to the measurement noise. Our objective is to estimate the packing density $1 - P_{\text{iso}} \in [0,1]$ of the fiber pathways and their tangential distribution $p \in L^1(S^2)$. A pulsed gradient spin echo experiment measures the diffusion-sensitized MR signals at a finite number of gradient directions $g \in S^2$ with a diffusion weighting factor $b > 0$. In the present work we assume that the noise in the magnitude signal follows a Rician distribution, which is defined by

$$f_{\mathcal{R}}(S; E, \varsigma^2) = \frac{S}{\varsigma^2} \exp\left(-\frac{S^2 + E^2}{2\varsigma^2}\right) I_0\left(\frac{SE}{\varsigma^2}\right).$$

$S$ denotes the observed signal, $E$ the true magnitude signal, $\varsigma^2$ is a parameter characterizing the noise level, and $I_0$ the zeroth-order modified Bessel function of the first kind. Note that the noise distribution is altered by the correction for subject motion and the co-registration with an anatomical coordinate system due to linear interpolation (Rohde et al., 2005). While under a Gaussian noise model these variance modifications $w$ can be easily taken into account (Kaden and Kruggel, 2011), the situation is much more complicated for Rician noise. Here the noise alterations are considered in terms of $S \sim \mathcal{R}(E, w\varsigma^2)$, which is a reasonable approximation when the signal-to-noise ratio is sufficiently high.

We choose weakly-informative priors that make as few assumptions about the hidden quantities as possible. The prior knowledge on the fiber orientation density is encoded by a discrete-valued version of the Dirichlet process mixture. The finite random vector $\pi^{[n]} \sim \mathcal{E}(\alpha, n)$ follows an Ewens distribution as described above, where the concentration parameter $\alpha \sim \mathcal{G}a$ is governed by a Gamma density and the sample size is set to $n = 128$ in our experiments. The base measure is formed by a uniform distribution $v_i \sim \mathcal{U}_{S^2}$ on the sphere and we choose the scaling parameters with $\kappa_i = 10$ for all indices $i$ (compare Section 3.2 for the rationale behind our prior belief). The volume fraction of the non-fiber compartment $P_{\text{iso}} \sim \mathcal{U}_{[0,1]}$ follows a uniform distribution on the interval $[0,1]$. We set the prior distribution of the noise level to $\varsigma^2 \sim \mathcal{IG}$, which is an inverse Gamma density with appropriately chosen hyperparameters. The MR signal $E_0$ without diffusion weighting, which normalizes the $T_2$-contrast, is estimated by a maximum likelihood approach using a Rician noise model (Sijbers and den Dekker, 2004). The likelihood function with the discrete version of the forward model (7) then updates our prior knowledge by extracting the information on the fiber orientation density from the noisy MR observations. The tangential density in adjacent voxels is implicitly supposed to be independent. According to Bayes' theorem, we obtain the posterior distribution that reflects the uncertainty about (not a possible variability in) the fixed but unknown density function after we have seen the diffusion-weighted signals. This distribution is approximately sampled by an adaptive version of the reversible jump Metropolis–Hastings algorithm. See Appendix A.2 for more details about the Monte Carlo sampler.

## 5. Results

### 5.1. Data acquisition and preprocessing

The diffusion-weighted MR dataset, which was kindly provided by the 2009 Pittsburgh Brain Connectivity Competition (Schneider, 2009), available online at http://www.braincompetition.org, was acquired by a 3 T Magnetom Trio scanner (Siemens, Erlangen) equipped with a 32-channel phased-array head coil. A spin echo EPI sequence measured 256 diffusion gradient directions with a $b$-value of 1500 s/mm$^2$. These directions and their antipodal points are uniformly distributed on the sphere. Further, 30 images without diffusion weighting were acquired. The sequence timing of the MR experiment was fixed with an echo time $t_{\text{TE}} = 108$ ms and the repetition time $t_{\text{TR}} = 11.6$ s. The diffusion images were reconstructed by a 6/8 partial Fourier encoding scheme. The measurement of 68 slices with 2 mm thickness (no slice gap) and a $128 \times 128$ image matrix (with a field of view $256 \times 256$ mm$^2$) covered the whole brain, resulting in an acquisition time of about 55 min. A male volunteer (aged 27 years) participated in this study. The dataset is corrected for subject motion with respect to the images having a $b$-value of 0 s/mm$^2$ and co-registered with a $T_1$-weighted MR volume using rigid-body transformations (Jenkinson et al., 2002), as implemented in FSL (2008). For this purpose the diffusion-weighted images, which have an isotropic voxel resolution of 2 mm, are resampled by linear interpolation. As a consequence, the noise variance is modified by the factor

$$w(x,y,z) = \sum_{i,j,k \in \mathbb{Z}} \Lambda(x/T - i)^2 \Lambda(y/T - j)^2 \Lambda(z/T - k)^2$$

at the location $(x,y,z) \in \mathbb{R}^3$, assuming that the original dataset is spatially uncorrelated (Rohde et al., 2005). $T$ is the step size of the uniform sampling, $\Lambda(s)$ denotes the hat function with $\Lambda(s) = 1 - |s|$ for $|s| \leqslant 1$ and $\Lambda(s) = 0$ otherwise. The $T_1$-weighted anatomical dataset is aligned with the stereotactic coordinate system (Talairach and Tournoux, 1988) without spatial normalization.

### 5.2. Simulations

In the following we demonstrate the reliability of the Dirichlet process mixture for the nonparametric estimation of the fiber orientation density from noisy MR measurements. Henceforth, the water diffusivity parameter of the nonfiber compartment is fixed at $\lambda_{\text{iso}} = 0.0012$ mm$^2$/s, and the effective diffusion coefficients of a fiber segment are set to $\lambda_{\parallel} = 0.0018$ and $\lambda_{\perp} = 0.0002$ mm$^2$/s (cf. Pierpaoli et al., 1996). Additionally, we consider different subsets of the acquired 256 diffusion encoding gradients. Various results are reported for 32, 64, and 128 gradient directions, which together with their antipodal points are distributed as uniformly as possible on the sphere. Using the acquisition protocol presented in the previous section, we simulate the diffusion-weighted MR signals for 64 gradient directions from the density function shown in Sample 3 of Fig. 1. After a period of 50,000 burn-in transitions, the reversible jump MCMC algorithm draws 50,000 samples from which every fifth is subsampled. Fig. 2 exemplifies the mixing efficiency of the Monte Carlo sampler, which approximately computes the posterior distribution conditional upon the synthetic data that are disturbed by Rician noise. The top row of this figure shows a trace plot of the model indicator $k$ for 10,000 samples, indicating that the transdimensional MCMC sampler is capable of moving between different parameter spaces. The bottom panel depicts the histogram of the partition size (left) and a box-and-whisker plot of the concentration parameter $\alpha > 0$ with respect to the number of mixture components. The latter diagram illustrates that if $\alpha$ is increased, the Ewens distributed random vector $\pi^{[n]}$ has a finer partition structure. In Bayesian nonparametric techniques, however, the model indicator $k$ is not of much interest. The mixture components should be interpreted as base elements required to represent the unknown density function. In particular, the partition size $k$ does not quantify the number of fiber bundles. The drawn sample may then be used to compute the probability of certain events or to estimate the expectation of a function with respect to the posterior distribution.

To assess the accuracy of the Bayesian density estimator, we conduct a simulation study. The biophysical model (7) generates

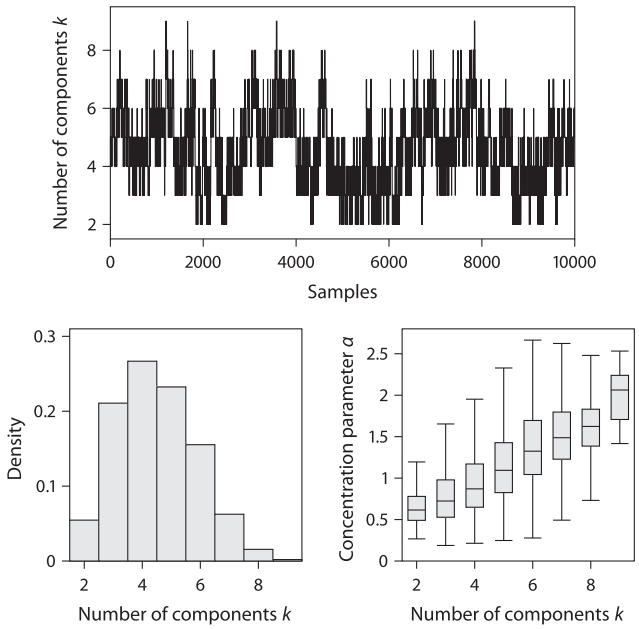


**Fig. 2.** Mixing efficiency of the reversible jump MCMC sampler. The top panel exemplifies a trace plot of the model indicator $k$ for 10,000 samples. The bottom row shows the histogram of the partition size (left) and a box-and-whisker plot of the concentration parameter $\alpha$ with respect to the number of mixture components (with 1.5 times the interquartile range).

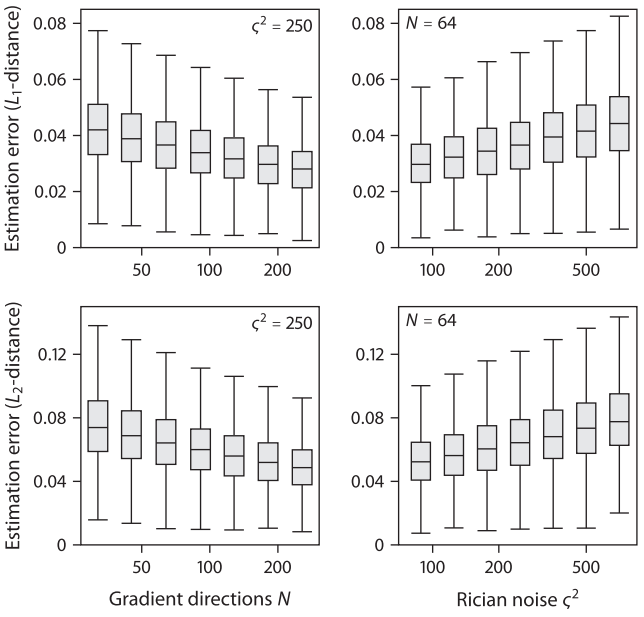


**Fig. 3.** Estimation error of the posterior mean with respect to the true density function. The top section shows the error based on the $L_1$-metric, while the bottom row depicts the root mean square error. From left to right, the reconstruction error is displayed for different numbers of gradient directions $N$ and various levels of Rician noise $\varsigma^2$. See text for further details on this simulation study.

the diffusion-weighted MR signals for a $b$-value of 1500 s/mm$^2$ using a Dirichlet process mixture. See also Section 3.2 for more information about the hyperparameters of this statistical model. We fix the spin echo signal $E_0 = 200$ (in the absence of diffusion weighting), which is in the range of the observed $T_2$-weighted MR signals for white matter, and the volume fraction of the nonfiber compartment is set to $P_{\text{iso}} = 0.1$. The factors $w^{(i)}$, which quantify the noise variance modifications due to linear interpolation, are picked at random from the diffusion dataset. Fig. 3 shows the estimation error of the posterior mean with respect to the original fiber orientation density. For this purpose we define the distance between two density functions $f, g \in L^p(S^2)$ as

$$d_p(f,g) = \left( \frac{1}{4\pi} \int_{S^2} |f(\omega) - g(\omega)|^p \, d\omega \right)^{1/p}$$

for $p \in \{1,2\}$. The top row of this figure depicts the error based on the $L_1$-metric, while the bottom section displays the root mean square error. We run 2500 trials to study the reconstruction error of the fiber orientation density under various scenarios. From left to right, the box-and-whisker plots (with 1.5 times the interquartile range) show the estimation error in dependence on the number of diffusion gradient directions $N$ and the level of Rician noise $\varsigma^2$. The fixed parameter is indicated in an upper corner of the diagrams. Note that this simulation study is an average-case analysis over the density functions drawn from the Dirichlet process mixture. Not surprisingly, the reconstruction error decreases when more gradient directions $N$ are measured and/or the Rician noise level $\varsigma^2$ is reduced.

### 5.3. Data analysis

Next, we showcase a statistical analysis of the acquired MR dataset using a subset of the gradient directions, here $N = 64$. Fig. 4 demonstrates the reconstruction of the diffusion signal $e_b(g)$, $g \in S^2$ and the fiber orientation density $p(\omega)$, $\omega \in S^2$ in three voxels taken from different brain regions. The posterior mean of the unknown density function is approximated by the sample average

$$\hat{p}(\omega) = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{k^{(t)}} \pi_i^{[n](t)} f_{\mathcal{W}}\left(\omega; v_i^{(t)}, \kappa_i\right)$$

of $T$ reversible jump MCMC iterations, where $\left\{k^{(t)}, \pi_i^{[n](t)}, v_i^{(t)}\right\}$ denotes the $t$-th output of the simulated chain. Note that this posterior expectation is defined over all models $k \in \{1,\ldots,n\}$. The posterior mean of the diffusion signal, which is depicted in the left column, is also computed by model averaging. The directional functions are shown in polar coordinates and, because these functions are antipodally symmetric, it is sufficient to display them in one hemisphere. The first row of Fig. 4 shows the tangential distribution of the nerve fibers in the splenium of the corpus callosum (CC), which consists of a single fiber bundle. The second and the third row provide evidence that the fiber population is composed of two and three fiber bundles, i.e., the crossing of the callosal fibers with the corona radiata (CR), and the crossing of the radiation of the corpus callosum, the corona radiata, and the superior longitudinal fasciculus (SLF), respectively. The fractional anisotropy of the latter voxel is estimated at 0.235, which indicates an almost isotropic diffusion process according to the classical diffusion tensor model (Basser et al., 1994). The middle column of this figure depicts the posterior mean including the 95% confidence band for a great circle on the sphere, which is shown as dotted line in the right column. The Bayesian credible regions are bounded by the 0.025- and the 0.975-quantile of the sample drawn from the reversible jump MCMC algorithm and thus characterize the pointwise uncertainty about the estimated fiber orientation density after we have observed the diffusion-weighted MR signals for 64 and 256 gradient directions, respectively.

Fig. 5 shows the Bayesian inference of the fiber orientation field in the centrum semiovale for different numbers of gradient directions using the effective diffusion coefficients given in the previous section. The voxelwise posterior mean of the fiber orientation

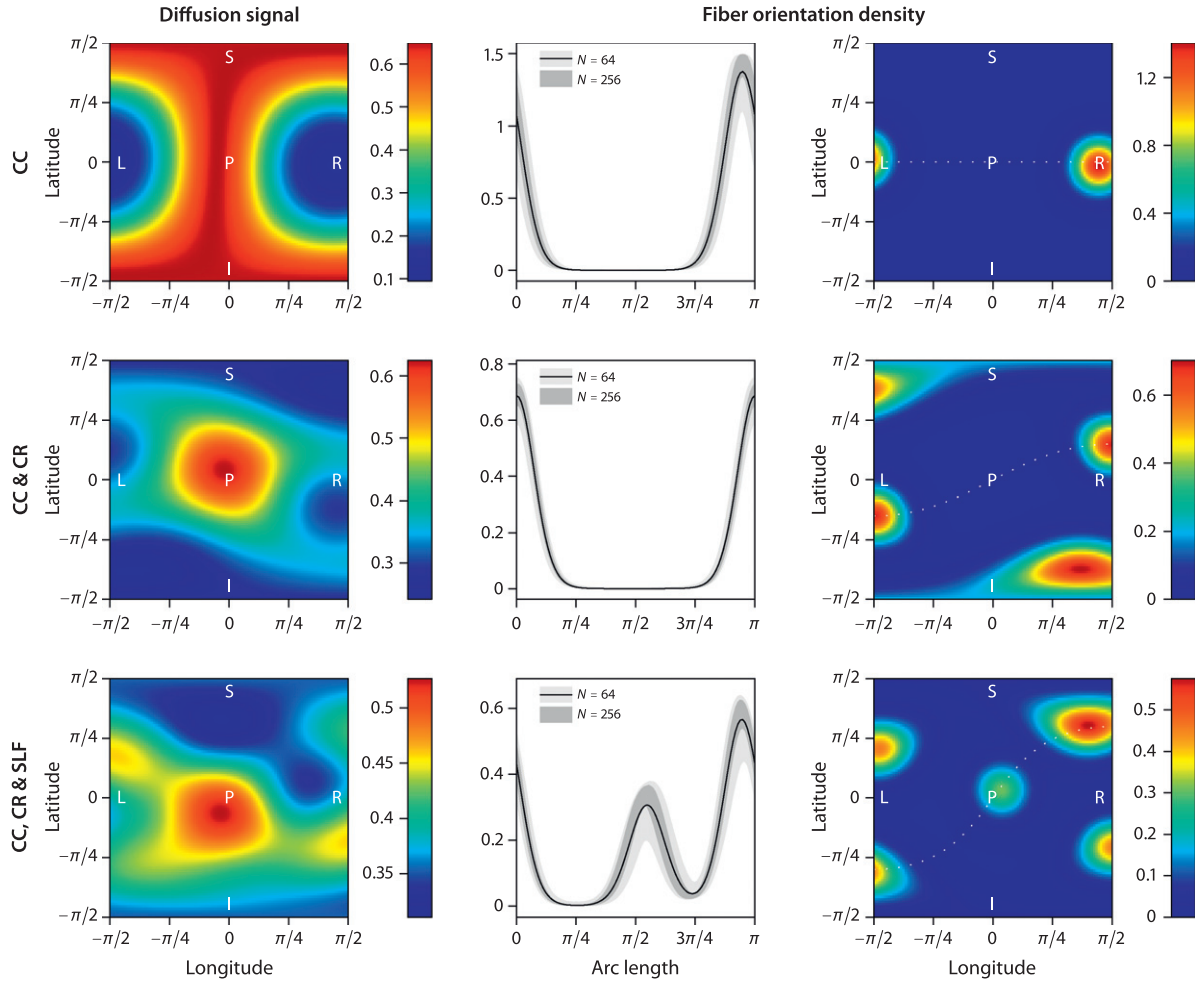


**Fig. 4.** Nonparametric Bayesian estimation of the diffusion signal and the fiber orientation density using 64 gradient directions, exemplified (from top to bottom) for the three voxels $(0,-35,15)$, $(20,-16,34)$, and $(-26,-16,38)$ in the stereotactic coordinate system (Talairach and Tournoux, 1988). Their fractional anisotropy is estimated at 0.887, 0.409, and 0.235, respectively. The second column depicts the posterior mean for a great circle on the sphere, which is shown as dotted line in the third column. The shaded areas display the 95% Bayesian confidence intervals for 64 and 256 diffusion encoding gradients.

density $\hat{p}$ is visualized by the quasi-spherical surface $S^2 \ni \omega \mapsto \hat{p}(\omega)\omega \in \mathbb{R}^3$. Red[2] indicates a left–right orientation, green an anterior–posterior direction, and blue a superior–inferior orientation. The underlying map shows the fractional anisotropy. Recall that the callosal fibers are commissural fibers which interconnect the two hemispheres, whereas the corona radiata includes, for instance, the pyramidal tract which is a projection fiber bundle primarily linking the motor cortex with the spinal cord. This figure exposes the intermingling of these two nerve fiber systems in the coronal plane. In particular, it is apparent that the commissural fibers passing the central part of the corpus callosum do not only project into the superior and medial areas, but also radiate towards the lateral regions of the cerebral cortex (cf. Pandya and Seltzer, 1986). The narrow band marked with a star consists of a set of white matter voxels that have small fractional anisotropy values. The bottom right panel displays a close-up of those voxels, which are composed of three distinct fiber bundles, namely the callosal fibers (red), the corona radiata (blue), and the superior longitudinal fasciculus (green). The latter pathway, which points out of the image plane, is an association fiber bundle that interconnects distant cortical areas within the same hemisphere. Fig. 5 also illustrates that a subset of the acquired 256 gradient directions is sufficient to disentangle the major fiber bundles in the cerebral white matter. Here we show results with 32, 64, and 128 gradient directions. The proposed estimation framework enables the analysis of the connectional neuroanatomy even in patients, since the acquisition time for 64 diffusion encoding gradients is not longer than 15 min on a clinical MR scanner system. Of course, a larger number of gradient directions improves the reconstruction quality of the fiber orientation field.

Fig. 6 maps the posterior mean of the number of mixture components $k$, the Rician noise level $\varsigma^2$, the volume fraction of the fiber compartment $1 - P_{\text{iso}}$, and the fractional anisotropy of the diffusion tensor model using 64 gradient directions (from left to right, top to bottom). The number in the upper right corner of these plots indicates the axial plane in the anatomical coordinate system (Talairach and Tournoux, 1988). The partition size in the discrete approximation of the Dirichlet process mixture is significantly lower in the corpus callosum and the internal capsule, since in these brain regions the diffusion signal is highly anisotropic. If the fiber population is more complex, the number of mixture components $k$ exhibits higher values. The volume fraction of the fiber compartment $1 - P_{\text{iso}}$, or equivalently the packing density of the nerve fibers, seems to be rather heterogeneous in the human brain. This parameter provides a good indicator for the cerebral white matter, also in comparison with alternative approaches based on the fractional anisotropy. Higher noise estimates observable in some regions may originate in tissue compartments with diffusion

[2] For interpretation of color in Figs. 1 and 4–6, the reader is referred to the web version of this article.

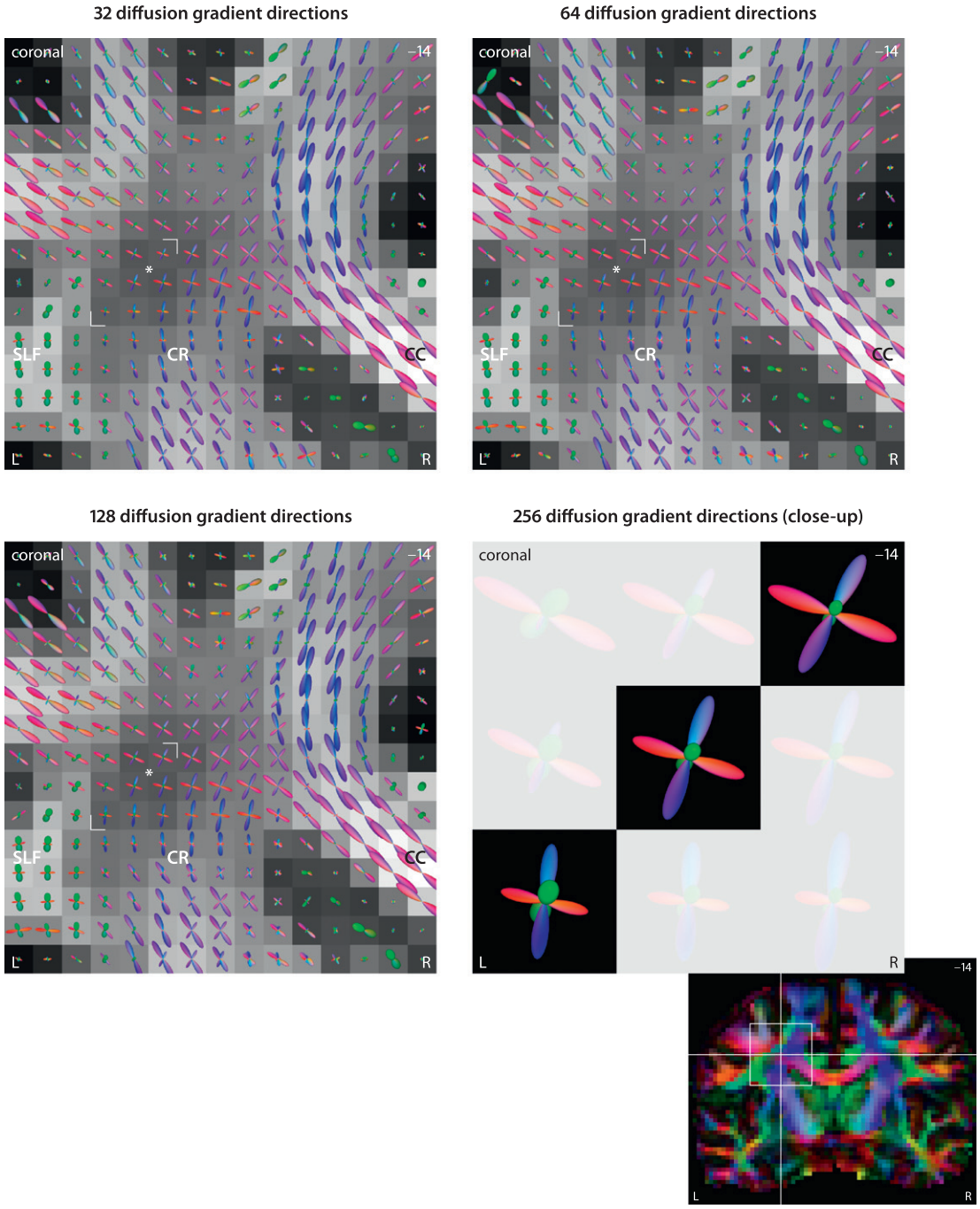


**Fig. 5.** The fiber orientation field uncovers the radiation of the corpus callosum (CC), the corona radiata (CR), and their crossing for different numbers of gradient directions. The bottom right panel shows a close-up of the narrow band marked with a star. The underlying map depicts the fractional anisotropy. The number in the upper right corner indicates the coronal plane in the anatomical coordinate system (Talairach and Tournoux, 1988). Abbreviation: superior longitudinal fasciculus (SLF).

coefficients different from $\{\lambda_{\text{iso}}, \lambda_{\parallel}, \lambda_{\perp}\}$, for example, cerebrospinal fluid and the image background, or are caused by artifacts due to magnetic susceptibility differences on material interfaces. The bottom row exemplifies an MR image without diffusion weighting and the diffusion signal for one of the gradient directions with a *b*-value of 1500 s/mm$^2$. Note that this dataset has lower signal intensities in the posterior portion of the brain, which is also visible in the Rician noise level. The acceptance rate of the transdimensional Monte Carlo algorithm is shown in Fig. 7, which summarizes the results of the computations for the axonal slice in the previous figure. Acceptance probabilities in the range of 0.1 and 0.6 are widely believed to be efficient (cf. Roberts et al., 1997).

## 6. Discussion

In the following we discuss the simplifying assumptions made to study complex biological systems such as the cerebral white matter using diffusion MR imaging. The impulse response of a small fiber segment may be modeled by a second-order approximation based on the Gaussian phase assumption (Neuman, 1974), which is a sensible choice especially under the low *b*-value regime. The water diffusivity parameters of a fiber segment are assumed to be invariant throughout the brain, which is obviously a simplification. A broad range of axonal diameters and multiple degrees of myelination are characteristic features of white matter

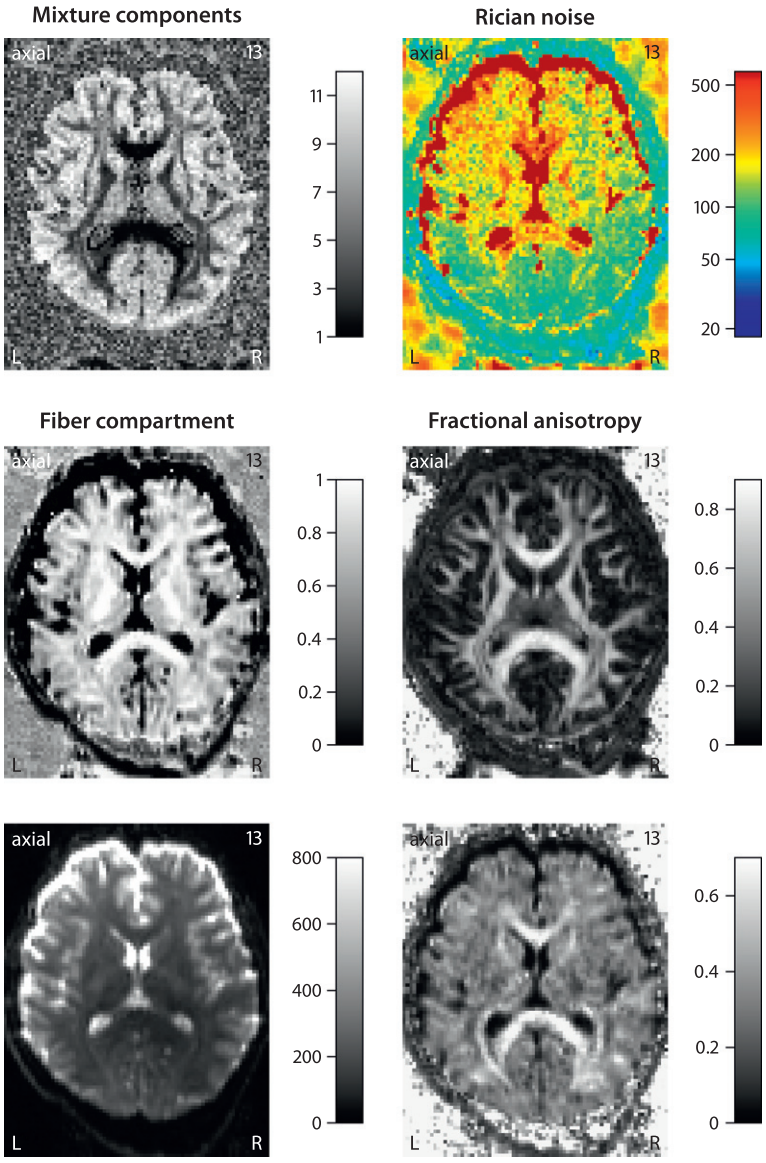


**Fig. 6.** The upper two rows show the estimation of the number of mixture components $k$, the Rician noise level $\varsigma^2$, the volume fraction of the fiber compartment $1 - P_{\text{iso}}$, and the fractional anisotropy of the diffusion tensor model. The bottom row exemplifies an MR image without diffusion weighting (left) and the diffusion signal for one of the gradient directions with a $b$-value of 1500 s/mm$^2$.

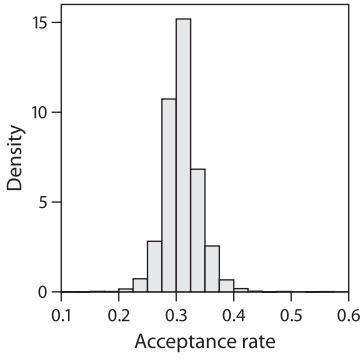


**Fig. 7.** Histogram of the acceptance rate for the reversible jump MCMC sampler. Note that the target acceptance probability is 0.3 in our experiments.

tissue, rendering the constant diffusivity assumption inappropriate to a certain extent. Hence, $\{\lambda_{\text{iso}}, \lambda_{\parallel}, \lambda_{\perp}\}$ should be interpreted as effective diffusion coefficients averaged over the brain. The proposed estimation framework may be extended to alternative response functions that describe the diffusion signal of a small fiber section more adequately. To construct a prior on the space of all fiber orientation densities, we suggest a Dirichlet process mixture with bipolar Watson densities. This stochastic process has various interpretations, for instance, the stick-breaking representation and the Pólya urn scheme, which both are used in the present work. The concentration parameter $\alpha$ of the Dirichlet process $\mathcal{DP}(\alpha, H)$ with the base measure $H$ has a somewhat complicated meaning. On the one hand, if the variability of $\mu \sim \mathcal{DP}(\alpha, H)$ should be maximized, we would let $\alpha$ go to zero because the variance writes $\text{var}[\mu(A)] = H(A)(1 - H(A))/(\alpha + 1)$ for a Borel set $A \in \mathcal{B}$ (Ferguson, 1973). On the other hand, in case of $\alpha \to 0$, $\mu$ converges in distribution to a Dirac measure. Furthermore, the Dirichlet process used to model the tangential distribution of the nerve fibers may be generalized to a greater class of stick-breaking processes. An example is the two-parameter Poisson–Dirichlet process (Pitman and Yor, 1997), whose sampling formula is a

generalization of the Ewens distribution. Nonparametric Bayesian inference then estimates the posterior distribution of the fiber orientation density conditional on the noisy MR observations of an integral transform. The uncertainty over the unknown density function is expressed by Bayesian confidence bands, dissociating between the neurobiological quantity, here the fiber orientation density, and the probability thereof.

A noteworthy outcome is that the tangential distribution of the nerve fibers is sparse within a small environment of white matter, which suggests to cluster the intra-voxel fiber population. Kaden et al. (2007) proposed to parameterize the fiber orientation density through a finite mixture of Bingham distributions. The underlying idea is to decompose the fiber population into a finite number of subpopulations (also called fiber bundles), using the criterion that the fibers within a bundle should be coherently oriented. This parametric model gives an explanatory description of the fiber architecture, for example, the number of fiber bundles, the spread of the fiber orientation within a bundle, and the volume fraction of the fiber subpopulations, solving the ill-posed spherical deconvolution in the finite-dimensional world. In contrast, a Dirichlet process mixture (with a bipolar Watson density) represents the tangential distribution of the fiber pathways in a nonparametric way. This infinite-dimensional approach makes as few assumptions on the shape of the unknown density function as possible, in an attempt to maximize the solution space of the inverse problem. Whereas each mixture component of the parametric model describes one fiber bundle with its structural properties, the components of a Dirichlet process mixture are merely base elements for the close approximation of a generic density function without any particular meaning. Moreover, the modes of the fiber orientation density generally do not correspond to the respective directions of the fiber bundles (Kaden et al., 2007). The number of (local) maxima cannot exceed the number of mixture components and may be lower, which means that the number of modes in the density function is a lower bound for the number of subpopulations. Consequently, fiber tracking along the modes of the fiber orientation density might introduce a systematic error.

Lastly, we compare the Dirichlet process mixture with a Gaussian process model, which both may act as a prior of the fiber orientation density before any diffusion MR signals have been observed. These two methodologies are examples of Bayesian nonparametrics, which do not assume that the tangential distribution of the nerve fibers lies within a specific parametric family. Kaden et al. (2008) suggested to solve the Fredholm integral equation of the first kind in a generic function space, making the vague assumption that the fiber orientation density is a suitably smooth function. More specifically, the solution to this Tikhonov regularization problem is found in an infinite-dimensional Hilbert space with a reproducing kernel that is derived from the spherical Laplace–Beltrami operator. This function estimation framework imposes a discretized nonnegativity and an exact normalization constraint in order to preserve the characteristic properties of a density function. The resulting convex optimization problem is solved by quadratic programming, which requires less computational resources (i.e., we find the global optimum within a short period of time) compared to the other two techniques. To quantify the uncertainty in the estimation of the fiber orientation density, the close relationship between reproducing kernel Hilbert spaces and Gaussian process models is exploited. Recall that a stochastic process is Gaussian if the marginal distribution for any finite subset of random variables is normally distributed. Thus, a Gaussian process defines a probability distribution on the space of generic functions, where the random functions drawn from this stochastic process are almost surely not normalized and may be negative. The Dirichlet process mixture, by contrast, is a probability measure on the space of all fiber orientation densities. In particular, the drawn random densities satisfy all intrinsic properties of a density function, allowing for the specification of Bayesian confidence intervals which respect that the fiber orientation density is nonnegative, as shown in Fig. 4. In summary, none of the three approaches is more favorable *per se*. Rather, it depends on the question to be answered which method is preferable for studying the intra-voxel fiber architecture in the individual living human brain.

## Acknowledgment

The diffusion-weighted MR dataset analyzed in the present study was kindly provided by the 2009 Pittsburgh Brain Connectivity Competition (Schneider, 2009), available online at http://www.braincompetition.org.

## Appendix A

### A.1. Confluent hypergeometric function

The confluent hypergeometric function ${}_1\mathrm{F}_1$ of a matrix argument (Herz, 1955) depends only on its eigenvalues. The real-valued eigenvalues of the 3-by-3 symmetric matrix $B = \kappa v v^t - b(\lambda_\parallel - \lambda_\perp) g g^t$ (see Eq. (7), neglecting the index $i$) write

$$\zeta_{1,2} = \frac{1}{2}\left(\kappa - b(\lambda_\parallel - \lambda_\perp) \pm \sqrt{(\kappa + b(\lambda_\parallel - \lambda_\perp))^2 - 4b(\lambda_\parallel - \lambda_\perp)\kappa \langle g, v\rangle^2}\right)$$

and $\zeta_3 = 0$. Consequently, we arrive at

$$\begin{aligned} {}_1\mathrm{F}_1(1/2; 3/2; B) &= {}_1\mathrm{F}_1(1/2; 3/2; \mathrm{diag}(\zeta_1, \zeta_2, \zeta_3)) \\ &= \frac{1}{2\pi} \sum_{i,j,k=0}^{\infty} \frac{\Gamma(i+1/2)\Gamma(j+1/2)\Gamma(k+1/2)}{\Gamma(i+j+k+3/2)} \frac{\zeta_1^i \zeta_2^j \zeta_3^k}{i!j!k!}, \end{aligned}$$

where $\Gamma$ denotes the Gamma function. Assuming $\zeta_1 \geqslant \zeta_2 \geqslant \zeta_3$ without loss of generality, Strelitz (1989) showed that this sum can be simplified to

$${}_1\mathrm{F}_1(1/2; 3/2; B) = \exp(\zeta_3) \sum_{i=0}^{\infty} \frac{(\zeta_1 - \zeta_3)^i}{(2i+1)i!} \sum_{j=0}^{i} b_{i,j} \left(\frac{\zeta_2 - \zeta_3}{\zeta_1 - \zeta_3}\right)^j$$

with the recursive construction

$$b_{i,0} = b_{i,i} = 1 \quad \text{and} \quad b_{i,j+1} = \frac{2j+1}{j+1} \frac{i-j}{2(i-j)-1} b_{i,j}.$$

For numerical computation the infinite sum is truncated in dependence on the desired accuracy. Alternatively, the confluent hypergeometric function may be computed by a saddlepoint approximation (Kume and Wood, 2005).

### A.2. Reversible jump MCMC

The posterior of the fiber orientation density is approximately sampled by Monte Carlo methods based on Markov chains. Note that this probability distribution is only known up to a multiplicative normalizing constant, which most likely does not exist in closed form. An in-depth treatment of Markov chains may be found in the seminal book by Meyn and Tweedie (1993). They showed, for instance, that under certain conditions an MCMC sampler converges to the target distribution (with respect to the total variation norm) irrespective of its initialization. In this work the algorithm is started with a fiber orientation density that is roughly uniform. Hence, after setting the sample size $n$, we choose the number of components $k = 24$ and fill the entries of the $k$-dimensional random vector $\pi_i^{[n]} \in \{q/n \in \mathbb{Q} : q = 1, \ldots, n\}$ with $\sum_{i=1}^{k} \pi_i^{[n]} = 1$ as evenly as possible. The other model parameters are either drawn at random from the prior distribution or are set to the expected value of what

is known prior to observing any diffusion MR signals. As the chain is frequently started in a region of low probability, we skip over its initial transient phase. Only the samples drawn after a sufficiently long burn-in are used to approximate the probability distribution conditional upon the measured data.

For the construction of a Markov chain we need to define an instrumental distribution which explores the posterior distribution via random walk and is easy to simulate. An urn model is employed to describe a proposal distribution $\mathcal{Q}$ on the $k$-dimensional random vector $n\pi^{[n]} \in \{1,\dots,n\}$. Let us assume that for $i = 1,\dots,k$ the nonempty urn $i$ contains $n\pi_i^{[n]}$ balls, where the number of balls in all $k$ urns sums up to $\sum_{i=1}^{k} n\pi_i^{[n]} = n$. A simple random step removes a ball from one of the urns with equal probability. If this urn is not empty after the removal, we add another urn to the model. Subsequently, the ball is randomly put back in one of these urns. After removing possibly empty urns, we obtain a new discrete-valued random vector $\rho^{[n]}$ with the sample size $n$ that has a dimension $\dim \rho^{[n]}$ of either $k-1$, $k$, or $k+1$. In the following we outline the transdimensional Monte Carlo sampler:

*initialization*;
**while** *the chain is not converged* **do**
  *simulate the proposal* $\rho^{[n]} \sim \mathcal{Q}(\pi^{[n]})$;
  **if** $\dim \rho^{[n]} < \dim \pi^{[n]}$ **then**
    *reversible jump MCMC sampler: death step*;
  **else if** $\dim \rho^{[n]} = \dim \pi^{[n]}$ **then**
    *fixed dimension Metropolis–Hastings sampling*;
  **else** (i.e., $\dim \rho^{[n]} > \dim \pi^{[n]}$)
    *reversible jump MCMC sampler: birth step*;
  **endif**
**done**.

If the dimension of the finite random vector has been changed, we proceed with reversible jump MCMC for across-model sampling. In the case where the number of mixture components is unaltered, a regular Metropolis–Hastings algorithm is used for within-model sampling.

The proposed discretization (8) of the Dirichlet process mixture gives rise to a variable dimension model, where the model indicator $k \in \{1,\dots,n\}$ counts the number of partitions in a sample of size $n$ and the parameter vector $\theta_k \in \Theta_k$ takes its values in a given space that has varying dimensions for different indices $k$. The posterior distribution is defined on the union of disjoint spaces $\Theta = \bigcup_{k=1}^{n} \{k\} \times \Theta_k$, which is also the general state space of the simulated Markov chain. Thus, we need to construct a sampler that is capable of moving between different parameter spaces associated with a finite collection of models. Without loss of generality, consider two models $k_1$ and $k_2$ with the dimensional constraint $k_1 < k_2$ and the state vectors $(k_1, \theta_{k_1})$ and $(k_2, \theta_{k_2})$, respectively. The reversible jump MCMC method (Green, 1995) augments the vector $(\theta_{k_1})$ with an auxiliary random variable $u_{k_1}$ so that both parameters $(\theta_{k_1}, u_{k_1})$ and $(\theta_{k_2})$ have the same dimension. A diffeomorphism $(\theta_{k_2}) = \Psi_{k_1 \to k_2}(\theta_{k_1}, u_{k_1})$, i.e., a differentiable bijective map whose inverse is also differentiable, then allows us to move between the two models $k_1$ and $k_2$ seamlessly. Note that there is a wide variability in the definition of the variable $u_{k_1}$ and the transform $\Psi_{k_1 \to k_2}$, indicating the universality of this approach. Here we suggest a birth-and-death sampler. If a new random vector $\rho^{[n]}$ is proposed that has a greater dimension than the current state, the partition size $k$ is increased accordingly and the model parameter $v_i$ of the additional mixture component is drawn from the prior distribution as proposal for the birth jump. Due to reversibility, the death step is simply the inverse transform: The number of mixture components $k$ is decreased by removing the associated model parameter $v_i$ if the proposed random vector $\rho^{[n]}$ has a lower dimension than $\pi^{[n]}$. The birth respective death jump is taken with a certain acceptance probability which satisfies the detailed balance condition and also involves the Jacobian of the transform $\Psi_{k_1 \to k_2}$. Otherwise the proposal is rejected and we use the current state. For further details, the reader is referred to Green (1995).

In addition, we employ a regular Metropolis–Hastings scheme for each variable and fixed dimension move that explores the other parameters (apart from the model dimension $k$ and the random vector $\pi^{[n]}$) by random walk. These local jumps may be described by a Watson distribution (for the parameters $v_i \in S^2$), a beta density (for $P_{\text{iso}} \in [0,1]$), and a normal distribution (for $\log\alpha$ and $\log\varsigma^2$), conditional upon the current state. The scaling of the random walk determines the speed at which the generated Markov chain reaches its stationary regime. When the proposed moves are small, the acceptance probability is high and the random walk explores the posterior distribution only slowly. If the step size is large, the acceptance rate is low (i.e., the proposals are often rejected) and it takes a long time for the sampler to converge to the target distribution. Roberts et al. (1997) showed that if the objective distribution is a normal density, the optimal scaling gives rise to an acceptance rate of 0.44 for one-dimensional models and 0.234 for models with infinite dimensions. Clearly, the obtained results may hold for other distributions only approximately. In the present work we implement an adaptation scheme that automatically optimizes the scaling of the random walk while exploring the posterior distribution (Roberts and Rosenthal, 2007). The step size is updated based on the history of the simulated sequence with the aim of an equilibrium acceptance rate of 0.3. Note that the adaptation amount at the time step $t$ vanishes as the chain proceeds. For example, the scaling parameter may be modified by sufficiently fast decreasing update values on the order of $\min\{0.05, 1/\sqrt{t}\}$. The scaling of the random walk is adapted only during the finite period of burn-in, afterwards the adaptation process is stopped and the step size is kept fixed. This finite adaptation scheme, where each individual Markov kernel with a fixed scaling is ergodic, ensures the asymptotic convergence to the posterior distribution. For more details, we refer the reader to Roberts and Rosenthal (2007). The adaptive MCMC algorithm facilitates a significant speed-up of the convergence rate, particularly if the scaling of the model parameters is rather heterogeneous.

## References


Alexander, D.C., 2005. Maximum entropy spherical deconvolution for diffusion MRI. In: Proceedings of the 19th International Conference on Information Processing in Medical Imaging, pp. 76–87.

Anderson, A.W., 2005. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. Magnetic Resonance in Medicine 54, 1194–1206.

Antoniak, C.E., 1974. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics 2, 1152–1174.

Assemlal, H.-E., Tschumperlé, D., Brun, L., Siddiqi, K., 2011. Recent advances in diffusion MRI modeling: angular and radial reconstruction. Medical Image Analysis 15, 369–396.

Axer, H., Axer, M., Krings, T., von Keyserlingk, D.G., 2001. Quantitative estimation of 3-D fiber course in gross histological sections of the human brain using polarized light. Journal of Neuroscience Methods 105, 121–131.

Basser, P.J., Mattiello, J., Le Bihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. Journal of Magnetic Resonance, Series B 103, 247–254.

Blackwell, D., MacQueen, J.B., 1973. Ferguson distributions via Pólya urn schemes. The Annals of Statistics 2, 353–355.

Bogachev, V.I., 2007. Measure Theory. Springer.

Dalal, S.R., 1979. Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. Stochastic Processes and their Applications 9, 99–107.

Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. Theoretical Population Biology 3, 87–112.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. The Annals of Statistics 1, 209–230.

FSL, 2008. FMRIB Software Library 4.1, University of Oxford. <http://www.fmrib.ox.ac.uk/fsl>.
Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.
Herz, C.S., 1955. Bessel functions of matrix arguments. The Annals of Mathematics 61, 474–523.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841.
Jian, B., Vemuri, B.C., 2007. A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE Transactions on Medical Imaging 26, 1464–1471.
Kaden, E., Kruggel, F., 2011. A reproducing kernel Hilbert space approach for Q-ball imaging. IEEE Transactions on Medical Imaging 30, 1877–1886.
Kaden, E., Knösche, T.R., Anwander, A., 2007. Parametric spherical deconvolution: inferring anatomical connectivity using diffusion MR imaging. NeuroImage 37, 474–488.
Kaden, E., Anwander, A., Knösche, T.R., 2008. Variational inference of the fiber orientation density using diffusion MR imaging. NeuroImage 42, 1366–1380.
Kume, A., Wood, A.T.A., 2005. Saddlepoint approximations for the Bingham and Fisher–Bingham normalising constants. Biometrika 92, 465–476.
Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M., 1986. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology 161, 401–407.
Leow, A.D., Zhu, S., Zhan, L., McMahon, K., de Zubicaray, G.I., Meredith, M., Wright, M.J., Toga, A.W., Thompson, P.M., 2009. The tensor distribution function. Magnetic Resonance in Medicine 61, 205–214.
Lo, A.Y., 1984. On a class of Bayesian nonparametric estimates: I. Density estimates. The Annals of Statistics 12, 351–357.
Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer.
Neuman, C.H., 1974. Spin echo of spins diffusing in a bounded medium. The Journal of Chemical Physics 60, 4508–4511.
Pandya, D.N., Seltzer, B. 1986. The topography of commissural fibers. In: Two Hemispheres – One Brain: Functions of the Corpus Callosum. Neurology and Neurobiology, vol. 17. Sixth International Symposium of the Center for Research in Neurological Sciences, Montreal, Canada, 1984, pp. 47–74.
Patel, V., Shi, Y., Thompson, P.M., Toga, A.W., 2010. Mesh-based spherical deconvolution: a flexible approach to reconstruction of non-negative fiber orientation distributions. NeuroImage 51, 1071–1081.
Pierpaoli, C., Jezzard, P., Basser, P.J., Barnett, A., Di Chiro, G., 1996. Diffusion tensor MR imaging of the human brain. Radiology 201, 637–648.
Pitman, J., Yor, M., 1997. The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. The Annals of Probability 25, 855–900.
Ramirez-Manzanares, A., Rivera, M., Vemuri, B.C., Carney, P., Mareci, T., 2007. Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry. IEEE Transactions on Medical Imaging 26, 1091–1102.
Roberts, G.O., Rosenthal, J.S., 2007. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. Journal of Applied Probability 44, 458–475.
Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. The Annals of Applied Probability 7, 110–120.
Rohde, G.K., Barnett, A.S., Basser, P.J., Pierpaoli, C., 2005. Estimating intensity variance due to noise in registered images: applications to diffusion tensor MRI. NeuroImage 26, 673–684.
Savadjiev, P., Campbell, J.S.W., Pike, G.B., Siddiqi, K., 2006. 3D curve inference for diffusion MRI regularization and fibre tractography. Medical Image Analysis 10, 799–813.
Schneider, W., 2009. Pittsburgh Brain Connectivity Competition 2009: Mapping the Human Connectome. In: Proceedings of the 15th Annual Meeting of the Organization for Human Brain Mapping, p. S28.
Sethuraman, J., 1994. A constructive definition of Dirichlet priors. Statistica Sinica 4, 639–650.
Sijbers, J., den Dekker, A.J., 2004. Maximum likelihood estimation of signal amplitude and noise variance from MR data. Magnetic Resonance in Medicine 51, 586–594.
Strelitz, R.A., 1989. Computing the normalization factor in Bingham distributions. Journal of Statistical Computation and Simulation 32, 193–200.
Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain, Thieme.
Tournier, J.-D., Calamante, F., Gadian, D.G., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage 23, 1176–1185.
Tournier, J.-D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. NeuroImage 35, 1459–1472.
von dem Hagen, E.A.H., Henkelman, R.M., 2002. Orientational diffusion reflects fiber structure within a voxel. Magnetic Resonance in Medicine 48, 454–459.
Watson, G.S., 1965. Equatorial distribution on a sphere. Biometrika 52, 193–201.
Yeh, F.-C., Wedeen, V.J., Tseng, W.-Y.I., 2011. Estimation of fiber orientation and spin density distribution by diffusion deconvolution. NeuroImage 55, 1054–1062.