

Optimal Experimental Design for Diffusion Kurtosis Imaging

Dirk H. J. Poot*, Arnold J. den Dekker, Eric Achten, Marleen Verhoye, and Jan Sijbers

Abstract—Diffusion kurtosis imaging (DKI) is a new magnetic resonance imaging (MRI) model that describes the non-Gaussian diffusion behavior in tissues. It has recently been shown that DKI parameters, such as the radial or axial kurtosis, are more sensitive to brain physiology changes than the well-known diffusion tensor imaging (DTI) parameters in several white and gray matter structures. In order to estimate either DTI or DKI parameters with maximum precision, the diffusion weighting gradient settings that are applied during the acquisition need to be optimized. Indeed, it has been shown previously that optimizing the set of diffusion weighting gradient settings can have a significant effect on the precision with which DTI parameters can be estimated. In this paper, we focus on the optimization of DKI gradient settings. Commonly, DKI data are acquired using a standard set of diffusion weighting gradients with fixed directions and with regularly spaced gradient strengths. In this paper, we show that such gradient settings are suboptimal with respect to the precision with which DKI parameters can be estimated. Furthermore, the gradient directions and the strengths of the diffusion-weighted MR images are optimized by minimizing the Cramér–Rao lower bound of DKI parameters. The impact of the optimized gradient settings is evaluated, both on simulated as well as experimentally recorded datasets. It is shown that the precision with which the kurtosis parameters can be estimated, increases substantially by optimizing the gradient settings.

Index Terms—Diffusion gradient settings, diffusion kurtosis imaging (DKI), diffusion-weighted (DW) magnetic resonance imaging (MRI), experimental design, optimization.

I. INTRODUCTION

DIFFUSION-WEIGHTED (DW) magnetic resonance imaging (MRI) is the only method available that measures noninvasively diffusion properties of tissues. Knowledge

of these diffusion properties allows the characterization of intrinsic features of tissue microdynamics and microstructure, such as cell permeability [1], [2]. The diffusion of water molecules within a voxel is characterized by a statistical distribution describing the random displacement of these molecules during a fixed-time diffusion process. A popular model to describe this distribution is the Diffusion tensor (DT) model, which assumes the diffusion distribution to be Gaussian.

Previously, it has been reported that the diffusion distribution in the human brain is generally non-Gaussian [2]–[4], due to diffusion restriction by cell membranes and compartments of different sizes present in the neural tissue. Since the DT imaging (DTI) model is limited to Gaussian diffusion only, the model can generally not describe these diffusion profiles accurately. Diffusion kurtosis imaging (DKI) was recently proposed as an extension to the Gaussian DT model. It was shown that DKI allows a better detection and characterization of changes in various white and gray matter structures [5]. In addition to the second central moment of the diffusion distribution, DKI also measures the kurtosis excess of that distribution. The kurtosis excess is defined as the fourth central moment of the distribution divided by the square of the variance minus 3 [6]. The “minus 3” term is often explained as a correction to make the kurtosis zero for a Gaussian distribution. Hence, compared to a DTI model, the inclusion of the kurtosis excess allows a more accurate description of the diffusion properties of neural tissues [3], [7].

Commonly, DKI data are acquired using a standard set of diffusion-weighting gradients with fixed directions and with regularly spaced gradient strengths. As shown in this paper, such imaging settings are suboptimal with respect to the precision with which DKI parameters can be estimated from the DW MR images. This precision strongly depends on the directions and strengths (b -values) of the diffusion-weighting gradients during the DW MR acquisition. A lower bound on the variance (precision) of any unbiased estimator is given by the Cramér–Rao lower bound (CRLB) [8]. In this paper, the gradient settings of the DKI experiments are optimized by minimizing the CRLB of DKI parameters of interest.

Previously, several studies were published that optimized DW MRI settings for estimating DTI parameters such as the fractional anisotropy (FA) and the mean diffusivity (MD) [9], [10]. In [9], e.g., the CRLB of various DT parameters was minimized with respect to the b -values of the DW MR images. In this paper, the study is extended to the optimization of DKI gradient settings. Furthermore, since the DKI model has significantly more parameters than the DTI model, a new numerical optimization strategy is developed. Extensive simulation experiments vali-

Manuscript received October 30, 2009; accepted November 22, 2009. Current version published March 03, 2010. This work was supported by the Institute for Science and Technology, Belgium. The work of D. H. J. Poot was supported by the agency for Innovation by Science and Technology (I.W.T.), Belgium. *Asterisk indicates corresponding author.*

*D. H. J. Poot is with the IBBT-Vision Lab, University of Antwerp, B-2610 Antwerp, Belgium (e-mail: dirk.poot@ua.ac.be).

A. J. den Dekker is with Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: a.j.dendekker@dcsc.tudelft.nl).

E. Achten is with Ghent Institute for Functional and Metabolic Imaging, Ghent University Hospital, B-9000 Ghent, Belgium (e-mail: rik.achten@ugent.be).

M. Verhoye is with the BioImaging Laboratory, University of Antwerp, B-2610 Antwerp, Belgium (e-mail: marleen.verhoye@ua.ac.be).

J. Sijbers is with the IBBT-Vision Lab, University of Antwerp, B-2610 Antwerp, Belgium (e-mail: jan.sijbers@ua.ac.be).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2009.2037915

dated with real experiments show that using the optimized gradient settings allows estimation of DKI parameters with a substantially higher precision.

The paper is organized as follows. Section II describes the DKI signal model, the DKI parameters, the CRLB for estimating these parameters, and the optimization method. Next, in Section III, simulations and real experiments are presented, which investigate the robustness and improvement of the performance of the optimized gradient settings, compared to the traditional gradient settings. Finally, in Section V, the conclusions are drawn.

II. METHODS

To obtain the most precise DKI parameter estimates, the directions and b -values of the diffusion-weighting gradients need to be optimized. First, in Section II-A, the kurtosis imaging model is explained. Next, Section II-B describes the computation of the CRLB for estimating kurtosis estimators. In Section II-C, various kurtosis parameters are introduced. After that, Section II-D elaborates on the optimization of the gradient settings. Finally, Sections II-E and II-F describe an efficient optimization strategy.

A. Diffusion Kurtosis Imaging

The diffusion of hydrogen atoms in a voxel can be characterized by a 3-D symmetric probability density function (PDF) $f(\mathbf{x}, t)$, where the random variable \mathbf{x} denotes the random displacement of molecules during a diffusion process in a time interval t . It depends on the microstructure of the voxel, which is generally different for each voxel.

The DW MRI does not measure the diffusion PDF directly. The gradients that are applied during the diffusion weighting introduce a change in the phase of the precessing and diffusing hydrogen atoms, which leads to a reduction of the magnitude of the DW image when compared to the unweighted image. The magnitude of a DW image depends on the diffusion in the direction of the applied diffusion-weighting gradient, specified by \mathbf{q} . The direction of the q -space vector \mathbf{q} is given by the unit length gradient direction vector \mathbf{g} , and its magnitude is given by

$$q = \gamma \delta G \quad (1)$$

where γ [rad/s, T] is the gyromagnetic ratio, δ [s] is the duration of the pulsed gradients, and G [T/m] is the magnitude of the applied diffusion-weighting gradient. To take the duration of the diffusion gradient pulses into account, the b -value is usually defined as [11]

$$b = q^2 t [\text{s}/\text{m}^2] \quad (2)$$

where $t = (\Delta - \delta/3)$ in which Δ [s] is the time separation between the leading edges of the diffusion gradient pulses [11].

Let $x = \mathbf{g}^T \mathbf{x}$ be the component of a displacement vector \mathbf{x} in the direction of \mathbf{g} . The PDF of \mathbf{x} in the direction of \mathbf{g} is then given by

$$f_{\mathbf{g}}(x; t) = \int_{\mathbf{g}^T \mathbf{x} = x} f(\mathbf{x}; t) d\mathbf{x}. \quad (3)$$

That is, $f_{\mathbf{g}}(x; t)$ is $f(\mathbf{x}; t)$ integrated over the two dimensions orthogonal to \mathbf{g} .

The diffusion coefficient $D_{\mathbf{g}}$ in the direction of the gradient \mathbf{g} , which is the variance of the diffusion in that direction, is given by

$$D_{\mathbf{g}} = \frac{1}{2t} \mathbb{E}_{f_{\mathbf{g}}} [x^2] \quad (4)$$

where $\mathbb{E}_{f_{\mathbf{g}}}[\cdot]$ is the expectation operator with respect to $f_{\mathbf{g}}$. The excess kurtosis $K_{\mathbf{g}}$ of the diffusion in the direction \mathbf{g} is given by [6]

$$K_{\mathbf{g}} = \frac{\mathbb{E}_{f_{\mathbf{g}}} [x^4]}{\mathbb{E}_{f_{\mathbf{g}}} [x^2]^2} - 3. \quad (5)$$

The phase shift induced by the diffusion-weighting gradients along the direction \mathbf{g} is a linear function of x and q . Therefore, the magnitude of the noise-free MR signal after diffusion weighting with the gradient \mathbf{q} is given by

$$\mathcal{A}(\mathbf{q}) = A_0 \mathbb{E}_{f_{\mathbf{g}}} [e^{iqx}] = A_0 \int_{x=-\infty}^{\infty} e^{iqx} f_{\mathbf{g}}(x; t) dx \quad (6)$$

where A_0 is the non-DW signal intensity. Note that (6) is equal to the characteristic function of $f_{\mathbf{g}}(x; t)$, multiplied by A_0 .

An approximate parametric model of (6) can be derived from a Taylor series expansion around $q = 0$ of the natural logarithm of $\mathcal{A}(\mathbf{q})$ [7]

$$\begin{aligned} \ln \mathcal{A}(\mathbf{q}) &= \ln A_0 - b \sum_{i,j=1}^3 g_i g_j D_{ij} \\ &+ \frac{b^2}{6} \left(\sum_{i=1}^3 \frac{D_{ii}}{3} \right)^2 \sum_{i,j,k,l=1}^3 g_i g_j g_k g_l W_{ijkl} + O(q^6) \end{aligned} \quad (7)$$

in which g_i is the i th component of \mathbf{g} , D_{ij} is the ij th element of the second-order DT \mathbf{D} , and W_{ijkl} is the $ijkl$ th element of the fourth-order kurtosis tensor \mathbf{W} . A detailed derivation of (7) can be found in the Appendix. The elements D_{ij} and W_{ijkl} are defined as

$$D_{ij} = \frac{1}{2t} \mathbb{E}_f [x_i x_j] \quad (8)$$

and

$$\begin{aligned} W_{ijkl} &= \frac{9}{(\mathbb{E}_f [\mathbf{x}^T \mathbf{x}])^2} (\mathbb{E}_f [x_i x_j x_k x_l] - \mathbb{E}_f [x_i x_j] \mathbb{E}_f [x_k x_l] \\ &- \mathbb{E}_f [x_i x_k] \mathbb{E}_f [x_j x_l] - \mathbb{E}_f [x_i x_l] \mathbb{E}_f [x_j x_k]) \end{aligned} \quad (9)$$

respectively, where $\mathbb{E}_f[\cdot]$ is the expectation operator with respect to f and with x_i the i th component of \mathbf{x} [7]. Note that both \mathbf{D} and \mathbf{W} are fully symmetric with respect to an interchange of indexes.

From (7), the following approximate parametric DKI model of the magnitude of the noise-free DW MR signal can then be obtained as

$$A(\mathbf{q}; \boldsymbol{\theta}) = A_0 \exp \left(-b \sum_{i,j=1}^3 g_i g_j D_{ij}^{\text{app}} + \frac{b^2}{6} \left(\sum_{i=1}^3 \frac{D_{ii}^{\text{app}}}{3} \right)^2 \right. \\ \left. \times \sum_{i,j,k,l=1}^3 g_i g_j g_k g_l W_{ijkl}^{\text{app}} \right) \quad (10)$$

where the diffusion and kurtosis tensors \mathbf{D} and \mathbf{W} are replaced by the apparent diffusion and kurtosis tensors \mathbf{D}^{app} and \mathbf{W}^{app} , respectively [12]. It is known that, for short duration δ of the diffusion gradient pulse, the apparent diffusion and kurtosis tensors approach the true diffusion and kurtosis tensors given by (8) and (9), respectively [7]. In (10), $\boldsymbol{\theta}$ denotes a 22×1 parameter vector composed of the following scalar valued parameters: A_0 , 6 parameters representing the independent elements of the symmetric tensor \mathbf{D}^{app} , and 15 parameters representing the independent elements of the fully symmetric tensor \mathbf{W}^{app} .

B. CRLB of the Kurtosis

The goal of this paper is to optimize the experimental design of a diffusion-weighting acquisition scheme such that diffusion kurtosis parameters can be estimated as precisely as possible. For this purpose, we will employ the CRLB framework. The CRLB provides a lower bound on the variance of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of the parameters $\boldsymbol{\theta}$ of a statistical model of measurements. It is well known that the maximum-likelihood (ML) estimator is asymptotically unbiased and efficient [8], [13]. Hence, its variance will attain the CLRb asymptotically, i.e., for an increasing number of observations. Experiments showed that the number of observations available in typical DKI measurements is sufficient for the asymptotic properties of the ML estimator to be valid.

Suppose that the joint PDF of a set of measurements $\tilde{\mathbf{A}}$ is given by $p(\tilde{\mathbf{A}}; \boldsymbol{\theta})$, which is parametric in $\boldsymbol{\theta}$. Then, the Fisher information matrix \mathbf{I} of these measurements is given by

$$\mathbf{I}(\boldsymbol{\theta}_0) = \mathbb{E} \left\{ \left[\frac{\partial \ln p(\tilde{\mathbf{A}}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\}^2 \quad (11)$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. Then, the Cramér–Rao inequality is given by

$$\text{cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{I}^{-1}(\boldsymbol{\theta}_0) \quad (12)$$

where $\text{cov}(\hat{\boldsymbol{\theta}})$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$. The right-hand side of (12) is known as the CRLB.

For a model $A_i(\boldsymbol{\theta})$, $i \in [1, \dots, N]$ of N independent Rician distributed data with the same noise parameter σ , such as corre-

sponding voxels in N magnitude DW MR images, this matrix is given by

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_{i=1}^N \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^T} I_{\text{rice}}(A_i, \sigma) \quad (13)$$

where I_{rice} is the Fisher information of a Rician distributed variable. This I_{rice} is given by [14] and [15]

$$I_{\text{rice}}(A_i, \sigma) = \int_{y=0}^{\infty} \frac{y}{\sigma^2} e^{-\frac{y^2 + A_i^2}{2\sigma^2}} I_0 \left(\frac{y A_i}{\sigma^2} \right) \\ \times \left(\frac{y I_1 \left(\frac{A_i y}{\sigma^2} \right)}{\sigma^2 I_0 \left(\frac{A_i y}{\sigma^2} \right)} - \frac{A_i}{\sigma^2} \right)^2 dy \quad (14)$$

where I_0 and I_1 are modified Bessel functions of the first kind of order 0 and 1, respectively. The integral in (14) cannot be solved analytically. However, since I_{rice} essentially only depends on A_i/σ , it can easily be tabulated for fast evaluation. We use a lookup table with polynomial interpolation to obtain values with high accuracy.

Let $m(\boldsymbol{\theta})$ be a kurtosis parameter, given as function of the DKI model parameters $\boldsymbol{\theta}$. Then, the CRLB of $m(\boldsymbol{\theta})$ is given by

$$\text{cov}(m(\hat{\boldsymbol{\theta}})) \geq I_m(\boldsymbol{\theta})^{-1} \quad (15)$$

where I_m is given by

$$I_m(\boldsymbol{\theta})^{-1} = \frac{\partial m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \mathbf{I}(\boldsymbol{\theta})^{-1} \frac{\partial m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \quad (16)$$

The gradient settings of the DW MR images can be optimized by minimizing the CRLB of a well-chosen parameter $m(\boldsymbol{\theta})$.

C. Kurtosis Parameters

In this section, the kurtosis parameters that will be considered in the experimental design and analysis are described.

1) *Mean Kurtosis*: The mean kurtosis is given by

$$MK = \frac{1}{4\pi} \iint_{\mathbb{S}^2} K(\mathbf{n}) d\mathbb{S}_{\mathbf{n}}^2 \quad (17)$$

where $K(\mathbf{n})$ is the parameterized excess kurtosis given by

$$K(\mathbf{n}) = \frac{\left(\sum_{i=1}^3 \frac{D_{ii}}{3} \right)^2 \sum_{i,j,k,l=1}^3 n_i n_j n_k n_l W_{i,j,k,l}}{\left(\sum_{i,j=1}^3 n_i n_j D_{i,j} \right)^2} \quad (18)$$

integrated over the unit sphere \mathbb{S}^2 , with $\mathbf{n} = [n_1 \ n_2 \ n_3] \in \mathbb{S}^2$. The derivation of (18) is given in the Appendix. Note that the definition of the mean kurtosis in (17) differs from a previous definition in [16, eq. (2)], where the MK is computed by averaging the kurtosis in the sampled gradient directions. This requires the same gradient directions to be sampled at multiple b -values. We prefer the definition in terms of the integral since this allows free selection of the gradient directions and b -values and allows accurate mean kurtosis values, even when the gradient directions are not uniformly distributed on \mathbb{S}^2 .

2) *Radial Kurtosis*: The radial kurtosis is the mean of the kurtosis in the directions orthogonal to the direction of main diffusion

$$K_{\perp} = \int_{\phi=0}^{2\pi} K(\mathbf{v}_2 \cos \phi + \mathbf{v}_3 \sin \phi) d\phi \quad (19)$$

where \mathbf{v}_i is the i th eigenvector of the DT \mathbf{D} , sorted by decreasing eigenvalue. This definition differs slightly from the radial kurtosis defined in [16], where the kurtosis along the second and third eigenvectors is averaged. The radial kurtosis is an interesting parameter, since the diffusion is restricted mainly in the radial direction. Therefore, it can be expected that the kurtosis, which is nonzero due to the restricted diffusion, is most pronounced in the radial direction.

3) *Kurtosis Anisotropy (KA)*: In [16, eq. (8)], the KA was defined as

$$FA_K = \sqrt{\frac{2}{3} \frac{(K_1 - \bar{K})^2 + (K_2 - \bar{K})^2 + (K_3 - \bar{K})^2}{K_1^2 + K_2^2 + K_3^2}} \quad (20)$$

with $K_i \equiv K(\mathbf{v}_i)$ and $\bar{K} = (K_1 + K_2 + K_3)/3$, the mean of the kurtosis in the DT eigenvector directions. This definition is in direct analogy to the FA in DTI. The original motivation for the diffusion FA was that it is a coordinate system invariant, dimensionless characterization of the differences between diffusion in the different directions. However, in contrast to the diffusion eigenvalues, the kurtosis itself is dimensionless and thus does not need to be normalized to obtain a dimensionless value. In our opinion, a KA parameter should not scale with the mean kurtosis, which might be zero, but should only be based on the variability in the kurtosis. Also, note that (20) only uses the kurtosis in the three directions specified by the diffusion eigenvectors. Since the kurtosis is specified by a higher order tensor, this might not capture all kurtosis variability.

We propose a different KA parameter, which, in our opinion, is more in line with the important characteristics of the FA. This new KA parameter is given by the standard deviation of the kurtosis

$$KA = \sqrt{\frac{1}{4\pi} \iint_{\mathbb{S}^2} (K(\mathbf{n}) - MK)^2 d\mathbb{S}_{\mathbf{n}}^2} \quad (21)$$

D. Optimization of the Gradient Settings

Let the settings of all diffusion weighting gradients during a DKI experiment, in which N DW MR images are acquired, be defined by $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_N]^T$. Here each \mathbf{q}_i specifies the gradient settings, i.e., the b -value and the direction of the diffusion-weighting gradient, of a DW MR image. Then, \mathbf{Q} can be optimized by minimizing the CRLB of the model parameters, $\mathbf{I}(\boldsymbol{\theta})^{-1}$. However, optimization methods need a scalar function to optimize. Therefore, a scalar objective function of the CRLB of the model parameters is required. This scalar objective function should evaluate the overall quality of the gradient settings \mathbf{Q} . Our objective function is the CRLB of a kurtosis parameter, $I_m(\boldsymbol{\theta})^{-1}$, given in (16). This function depends on the actual

tissue properties $\boldsymbol{\theta}_0$. These properties $\boldsymbol{\theta}_0$ are generally different in each voxel. Since the brain images contain many voxels and different tissues in which one might be interested, the acquisition scheme should be optimal for a distribution of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$. The optimal \mathbf{Q} can then be found by minimizing the objective function, weighted with the prior distribution $p(\boldsymbol{\theta})$

$$\hat{\mathbf{Q}} = \arg \min_{\mathbf{Q}} \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}) I_m(\boldsymbol{\theta})^{-1} d\boldsymbol{\theta}. \quad (22)$$

In practice, the prior distribution $p(\boldsymbol{\theta})$ can be approximated with experimental data. However, it is difficult to evaluate the 22-dimensional integral in (22). Therefore, it often is much more convenient to approximate the integral by drawing M samples from the prior distribution $p(\boldsymbol{\theta})$ and evaluating the objective function on this set only. These M samples $\boldsymbol{\theta}_i$ from $p(\boldsymbol{\theta})$, which should be fixed during the optimization, can be collected in a set $\Theta = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M\}$. Then, the integral in (22) can be approximated by

$$H(\mathbf{Q}; \Theta) = \frac{1}{M} \sum_{i=1}^M I_m(\boldsymbol{\theta}_i)^{-1} \quad (23)$$

which is the mean CRLB of the kurtosis parameter of the elements of Θ . The optimal gradient settings are then given by

$$\hat{\mathbf{Q}} = \arg \min_{\mathbf{Q}} H(\mathbf{Q}; \Theta). \quad (24)$$

E. Efficient Implementation of the Optimization

The optimal set \mathbf{Q} is in principle given by (24). However, due to the large number of DW images in a typical DKI experiment, it is not trivial to find this optimum. The actual function to optimize is a sum of scalar functions, each of which depends on the CRLB of a parameter vector $\boldsymbol{\theta}$, which is a 22×22 matrix. Most common optimization techniques for large problems use the analytic or numerically computed derivatives of the function to be optimized. However, the derivative of the inverse Fisher matrix \mathbf{I} with respect to \mathbf{Q} is difficult and computationally expensive to compute. There are general optimization routines that do not use the gradient of the function, e.g., `fminsearch` in MATLAB (The MathWorks, Inc.), but they typically require a huge number of function evaluations, and it was observed that the final gradient set found by `fminsearch` was not close to optimal. This might be due to local minima and/or almost flat parts in the objective function. To be able to overcome local minima, simulated annealing was chosen as alternative optimization method [17]. The simulated annealing method iteratively updates the diffusion-weighting gradients, one at a time. When a gradient is modified, the Fisher information matrix \mathbf{I} needs to be updated to evaluate the change in objective function. As follows from (13), updating the magnitude and direction of one diffusion-weighting gradient is only a rank two update of \mathbf{I} . Therefore, the Woodbury identity [18]

$$(A + CBC^T)^{-1} = A^{-1} - A^{-1}C(B^{-1} + C^T A^{-1}C)^{-1}C^T A^{-1} \quad (25)$$

can be used to efficiently update the inverse Fisher matrix.

F. Efficient Integration on the Sphere

The integral in (17) and (21) over \mathbb{S}^2 is numerically approximated by sampling the integrand on a weighted set of n , approximately uniformly distributed, sample points $\mathbf{p}_1, \dots, \mathbf{p}_n$ in \mathbb{S}^2 . The weights $\mathbf{w} = [w_1, \dots, w_n]$ of the sample points are chosen to maximize the accuracy of the approximation of the integrals. For this, first note that any function on \mathbb{S}^2 can be expanded in the spherical harmonics basis. Since the integral of all non-zeroth-order spherical harmonics is zero, integrating over \mathbb{S}^2 determines the magnitude of the zeroth-order spherical harmonic of the integrand, multiplied by the area of \mathbb{S}^2 . As the zeroth-order spherical harmonic is constant, accurate evaluation of the integral requires that $\sum_i w_i = 4\pi$. Smooth functions, such as the integrands considered in (17) and (21), generally have a decaying spectrum, i.e., the energy of the functions is mainly concentrated in low-order spherical harmonics. To avoid contamination of the evaluated integral by the energy in these low-order spherical harmonics, the weights \mathbf{w} are orthogonalized to a finite number of spherical harmonics. To minimize the influence of energy of the integrands in arbitrary higher order harmonics, the two-norm of \mathbf{w} is minimized. It can be proven that the weight vector \mathbf{w} that has these properties is the solution of a least squares problem with only 1 nonconfound

$$\mathbf{w} = [4\pi, 0, \dots, 0](\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \quad (26)$$

where the components of \mathbf{S} are given by

$$S_{i,j} = Y_{l_j}^{m_j}(\mathbf{p}_i) \quad (27)$$

where $Y_{l_j}^{m_j}$ is the m_j th real-valued spherical harmonic of order l_j , and $l_1 = 0, m_1 = 0$. Usually, all $(L+1)^2 < n$ unique combinations of $|m_j| \leq l_j \leq L$ have to be included, where L is the maximum order of spherical harmonics to which \mathbf{w} is orthogonal. However, when both the set of points and the integrand are symmetric around 0, the odd l_j and a symmetric half of the points \mathbf{p}_i do not need to be included in the computations. The integrands in (17) and (21) are symmetric around 0, so by selecting a symmetric set of points, this property can be used.

III. EXPERIMENTS

As described in Section II-D, the gradient settings are optimized by minimizing the CRLB of a kurtosis parameter, evaluated on a set Θ obtained from a prior distribution of DKI model parameters. In practice, samples from this distribution are obtained from prior DKI measurements. In this paper, the prior DKI measurements were obtained from a human and small animal DKI experiments of which the details are described in Section III-A. In Section III-B, several aspects that are important for the selection of Θ are discussed. Then, in Section III-C, the sets Θ that were used for the optimizations are specified. Finally, in Section III-D, traditional gradient settings are reviewed and the settings of the optimized gradients are described.

A. Acquisition of DKI Data

Human DKI data were used for the construction of Θ . These data were acquired after approval of the Institutional Review Board and after informed consent was obtained from the healthy volunteer. The dual spin echo DW 2-D Echo Planner Imaging (EPI) images were acquired with a Siemens 3.0-T MRI scanner. The volume was recorded with 45 slices with an acquisition matrix of 82×82 . The voxel dimensions were 2.7-mm isotropic and the echo time was $TE = 96$ ms. A 30-channel head coil was used and the Generalized Autocalibrating Partial Parallel Acquisitions (GRAPPA) acceleration factor was 2, with 24 reference lines. The bandwidth was 1356 Hz/pixel and the transverse relaxation (TR) = 5.900 s. The maximum b -value of the set of recorded DW MR images was 2800 s/mm^2 . The SNR of the gray matter in the MR images with a b -value of zero was 12. To evaluate the performance and robustness of the optimized gradient settings, a second volunteer was scanned with the optimized settings on a different Siemens 3.0-T MRI scanner. This dataset was recorded with 55 slices with an acquisition matrix of 128×128 . The voxel dimensions were 2.5-mm isotropic and the echo time was $TE = 104$ ms. A 30-channel head coil was used and the GRAPPA acceleration factor was 2, with 24 reference lines. The bandwidth was 1955 Hz/pixel and the $TR = 7.700$ s. The maximum b -value of the set of recorded DW MR images was 2800 s/mm^2 . Furthermore, DKI data of a rat were acquired with a Bruker 7T small animal scanner. This dataset was acquired with 50 slices with an acquisition matrix of 96×64 , reconstructed to an image size of 128×64 . The slice thickness was 0.37 mm, excluding the gap of 0.10 mm between slices. The in-plane resolution was 0.37 mm and the echo time was $TE = 24$ ms. The bandwidth was 8333 Hz/pixel and the $TR = 11.000$ s. The images were recorded with two shot EPI and mono polar diffusion-weighting gradients with a maximum b -value of 2800 s/mm^2 , obtained with the diffusion times $\delta = 5$ ms and $\Delta = 12$ ms. From these datasets, the model parameters were estimated with a ML estimator [13].

B. Prerequisites for the Selection of Prior DKI Model Parameters

There are several aspects that influence the selection of Θ .

- 1) For elements of Θ with a large positive kurtosis, the magnitude of the DW MR images, predicted with the DKI model (10), will grow strongly for large q . This is caused by the fourth power of q inside the exponent function and indirectly, it is a result of the finite region in which the series expansion used for the DKI model is accurate. The high SNR of these anomalously high predicted DW magnitudes reduce the CRLB of the kurtosis parameters, which might cause the optimization procedure to increase some gradients to unrealistically high magnitudes. This can be avoided by limiting the maximum q value that is allowed in the optimization or by selecting elements for Θ without large positive kurtosis.
- 2) The DKI parameters will depend on the tissue type under study. Therefore, representative parameter vectors of the

TABLE I
DIFFUSION EIGENVALUES FOR Θ_1 IN mm^2/s AND FA VALUE

$\lambda_1 \times 10^3$	$\lambda_2 \times 10^3$	$\lambda_3 \times 10^3$	FA
0.955	1.076	1.287	0.1507
0.940	1.103	1.309	0.1639
0.525	0.687	2.813	0.7514
0.631	1.250	1.839	0.4528
0.522	0.680	2.570	0.7286
0.525	0.782	2.317	0.6711
4.057	4.349	4.877	0.0936
3.217	3.441	3.944	0.1050

different tissues should be included in Θ . When the set Θ is too small or does not contain elements from all relevant tissue types, the optimized gradient settings might be good for the kurtosis parameters in the test set, but not for all the various brain tissue types.

- 3) The computation time required for the optimization depends almost linearly on the number of elements in Θ . Therefore, to limit the computation time required by the optimization, Θ should not contain excessively many elements. In our experiments, optimizations are performed with several hundreds of test tensors in Θ .
- 4) Some kurtosis parameters, such as the radial kurtosis, depend on the intrinsic DT coordinate system. When the optimization is performed with respect to these kurtosis parameters, the coordinate system should be well defined for all elements of Θ . This can be established by selecting Θ from sufficiently anisotropic tissues, such as the white matter structures.

C. Selection of DKI Model Parameter Sets

This section describes the sets Θ . These sets Θ contain the samples from the prior distribution $p(\theta)$ that were selected for the optimization experiments. Each element of a Θ contains the parameters of the DKI model, from which the kurtosis parameters can be evaluated. In order to investigate the sensitivity of the optimization of the kurtosis parameters to different Θ , the optimization is performed for the following three sets.

- Θ_1 To avoid unrealistically large b -values due to large positive kurtosis, the set Θ_1 was constructed to have **zero kurtosis** and a range of realistic DT eigenvalues. The diffusion eigenvalues were typical for the gray matter, white matter, and the cerebrospinal fluid (CSF) present in the human DKI dataset. To avoid indeterminacy of DT eigenvectors, the eigenvalues were chosen sufficiently different. The DT eigenvalues are given in Table I and were manually selected from the DKI dataset (cf., Section III-A). Furthermore, to make the gradient settings to be optimized (approximately) rotationally invariant, 60 DTs were generated from each set of eigenvalues by rotating the first eigenvector (FE) toward the 20 corners of a dodecahedron and subsequently rotating the second eigenvector in steps of 120° .

- Θ_2 The second set was obtained by randomly selecting 400 DK tensors from the **white and gray matter** of the DKI dataset. The probability to be included in the set was equal for each gray and white matter voxel. Since the estimated parameter vectors are noise corrupted, the diffusion eigenvalues might occasionally be unrealistically low or the estimated kurtosis might be large in some directions. Therefore, the lowest diffusion eigenvalues were adjusted to be at least $1/b_m$, with $b_m = 3000 \text{ s}/\text{mm}^2$ and for all directions \mathbf{g} in which the kurtosis was positive, the kurtosis was decreased as long as $q_m^2 t < b_m$, with $q_m = \arg \min_q A(q\mathbf{g})$, with A from (10).

- Θ_3 The third set was obtained by randomly selecting 400 DK tensors from **white matter only**, $\text{FA} > 0.4$, of the DKI dataset. The probability of a voxel to be included in the selection was proportional to the FA value of the voxel. The elements of Θ_3 were adjusted with the same procedure as the elements of Θ_2 .

D. Optimized and Traditional Gradient Settings

The optimized gradient settings were compared with a “traditional” set T of diffusion gradient settings for DKI [3].

- T Specifies DW MR images with diffusion weighting gradients in 30 directions, with 5 different b -values, 500 to $2500 \text{ s}/\text{mm}^2$ in steps of $500 \text{ s}/\text{mm}^2$ and 10 images with a b -value of zero. So in total T specifies 160 DW MR images.

For fair comparisons, the optimized sets used the same number of DW MR images as the traditional set T . The optimized sets of gradient settings were as follows.

- $O_{i,m}^{b_j}$ These DKI gradient settings were optimized with Θ_i , with a maximum b -value b_j allowed in the optimization, with $b_1 = 2800 \text{ s}/\text{mm}^2$ and $b_2 = 2500 \text{ s}/\text{mm}^2$. The optimization was performed with respect to the CRLB of the mean kurtosis of the elements of Θ_i . The maximum allowed b -value was limited to avoid the selection of excessively large b -values due to the breakdown of the DKI model for very high b -values.

- $O_{i,r}^{b_j}$ The same as $O_{i,m}^{b_j}$, but now optimized with respect to the CRLB of the radial kurtosis obtained from Θ_i .

IV. RESULTS AND DISCUSSION

In this section, the results of the experiments are discussed. First, in Section IV-A, the optimized and traditional gradient settings are compared on kurtosis parameters of the sets Θ . Next in Section IV-B, a good optimized set of gradient settings is reviewed. Finally, in Section IV-C, recorded DKI data are used

TABLE II
 NORMALIZED VALUE OF THE OBJECTIVE FUNCTION H OF THE (a) MEAN KURTOSIS, (b) RADIAL KURTOSIS, OR (c) KA. THIS OBJECTIVE FUNCTION H IS THE AVERAGE CRLB OF THE MEAN OR RADIAL KURTOSIS OF THE GRADIENT SETS FOR THE DIFFERENT TEST SETS Θ_i . NORMALIZATION WAS PERFORMED SUCH THAT THE LOWEST VALUE IN EACH COLUMN WAS 1

(a)				(b)				(c)		
	Θ_1	Θ_2	Θ_3		Θ_1	Θ_2	Θ_3		Θ_2	Θ_3
T	151	1.406	1.911	T	77	1.571	2.541	T	1.253	1.717
$O_{1,m}^{b_1}$	1	2.299	4.737	$O_{1,m}^{b_1}$	1	1.966	7.173	$O_{1,m}^{b_1}$	1.744	3.716
$O_{1,r}^{b_1}$	1	1.455	1.880	$O_{1,r}^{b_1}$	1	1.398	2.267	$O_{1,r}^{b_1}$	1.319	1.569
$O_{2,m}^{b_1}$	496	1.000	1.186	$O_{2,m}^{b_1}$	332	1.410	1.379	$O_{2,m}^{b_1}$	1.000	1.177
$O_{2,r}^{b_1}$	116	1.261	1.437	$O_{2,r}^{b_1}$	80	1.000	1.480	$O_{2,r}^{b_1}$	1.110	1.316
$O_{3,m}^{b_1}$	203	1.367	1.000	$O_{3,m}^{b_1}$	260	1.719	1.112	$O_{3,m}^{b_1}$	1.616	1.000
$O_{3,r}^{b_1}$	453337	3.983	1.137	$O_{3,r}^{b_1}$	220969	3.696	1.000	$O_{3,r}^{b_1}$	4.515	1.113
$O_{2,m}^{b_2}$	211	1.061	1.396	$O_{2,m}^{b_2}$	112	1.393	1.758	$O_{2,m}^{b_2}$	1.048	1.345
$O_{2,r}^{b_2}$	125	1.296	1.653	$O_{2,r}^{b_2}$	73	1.042	1.842	$O_{2,r}^{b_2}$	1.130	1.444
$O_{3,m}^{b_2}$	326	1.584	1.242	$O_{3,m}^{b_2}$	410	1.720	1.517	$O_{3,m}^{b_2}$	1.803	1.210
$O_{3,r}^{b_2}$	460150	2.623	1.357	$O_{3,r}^{b_2}$	226377	2.517	1.406	$O_{3,r}^{b_2}$	2.957	1.282

to compare the performance of the optimized gradient settings with the traditional gradient settings.

A. Results of the Optimization

In this section, the performance of the different gradient settings is compared.

Table II shows the relative $H(Q; \Theta_i)$ (23). That is, Table II shows the normalized mean CRLB of the mean [see Table II(a)] or radial [see Table II(b)] kurtosis from the three sets Θ_i , for all gradient sets Q . The values are normalized by dividing with the lowest $H(Q; \Theta_i)$ in each column. This table clearly shows that the gradient settings influence the precision with which the kurtosis parameters can be estimated, as the normalized mean CRLB of kurtosis parameters of the elements of Θ_i is different for each gradient set. Also, Table II shows that only the gradient settings optimized for Θ_1 , i.e., $O_{1,r}^{b_1}$ and $O_{1,m}^{b_1}$, have a low relative mean CRLB on kurtosis parameters computed from Θ_1 . The relative mean CRLB of the other gradient settings $O_{2,j}^{b_2}$ and $O_{3,j}^{b_2}$ on the estimation of kurtosis parameters of Θ_1 is much larger.

Further inspection showed that these other gradient settings had a substantially higher CRLB on the elements of Θ_1 that modeled CSF. This is caused by the very high diffusivity of CSF, which is not present in the gray or white matter from which Θ_2 and Θ_3 were selected. Since one is usually not interested in the CSF, this is not a problem for the use of the other optimized gradient settings, but clearly shows the importance of the selection of the elements of Θ .

Furthermore, Table II shows that the traditional gradient settings T has a higher mean CRLB, especially for Θ_3 , which contains parameter vectors of white matter structures. The mean CRLB of the radial kurtosis of T is 2.5 times larger than the mean CRLB of the radial kurtosis of the best gradient settings.

Table II further shows that the gradient settings $O_{2,m}^{b_1}$, optimized for the mean kurtosis based on Θ_2 , generalizes well. That is, the mean CRLB of the mean and radial kurtosis of Θ_2 and Θ_3 is close to minimal, when the gradient settings are specified by $O_{2,m}^{b_1}$. Since Θ_2 was randomly selected from the gray and white matter, it contains mostly gray, but also white matter

voxels, which might explain the relatively low mean CRLB of the kurtosis parameters on the white matter only set Θ_3 .

The sets of gradient settings $O_i^{b_2}$ were limited to a maximum b -value of 2500 s/mm². When the performance of these sets is compared to $O_i^{b_1}$, it is clear that the mean CRLB for the set Θ_i , which they are optimized on is only slightly increased. In particular, the mean CRLB of the mean kurtosis of Θ_2 is only increased by 6.1% for $O_{2,m}^{b_2}$, compared to $O_{2,m}^{b_1}$. However, the mean CRLB of different kurtosis parameters or model parameters is increased substantially by this lower b -value of 2500s/mm². This can, e.g., be seen by comparing the relative mean CRLB of the radial kurtosis of Θ_3 of the gradient settings $O_{2,m}^{b_2}$ with that of $O_{2,m}^{b_1}$, i.e., 1.758/1.379, which is larger than the 6.1% increase in mean CRLB of the mean kurtosis of Θ_2 for these gradient settings.

Summarized Table II(a) and (b) show the following.

- 1) The mean CRLB can be substantially decreased by optimizing the set of gradients.
- 2) For the optimization, it is important that the selection of samples from a prior distribution of tensors matches the diffusion and kurtosis properties found in the relevant tissues under study.
- 3) Gradient sets with good performance for both gray and white matter structures and for both the mean and radial kurtosis can be found, such as our set $O_{2,m}^{b_1}$.

B. Optimized Set of DW Gradients

From the results presented in the previous section, it can be concluded that the optimized gradient set $O_{2,m}^{b_1}$ produced the best results overall. Therefore, Fig. 1 shows the optimized gradients $O_{2,m}^{b_1}$, which are optimized with respect to the mean kurtosis of Θ_2 . Fig. 1(a) shows the sorted b -values of the optimized gradient set. The b -values automatically separate in, more or less, distinct levels. Fig. 1(c) shows the gradient directions and magnitudes of the individual directions, plotted on a sphere. The density of gradients is indicated by the gray level of the sphere. As can be seen in Fig. 1(c), the gradient directions are approximately isotropically distributed, which is a result that is obtained automatically by the optimization.

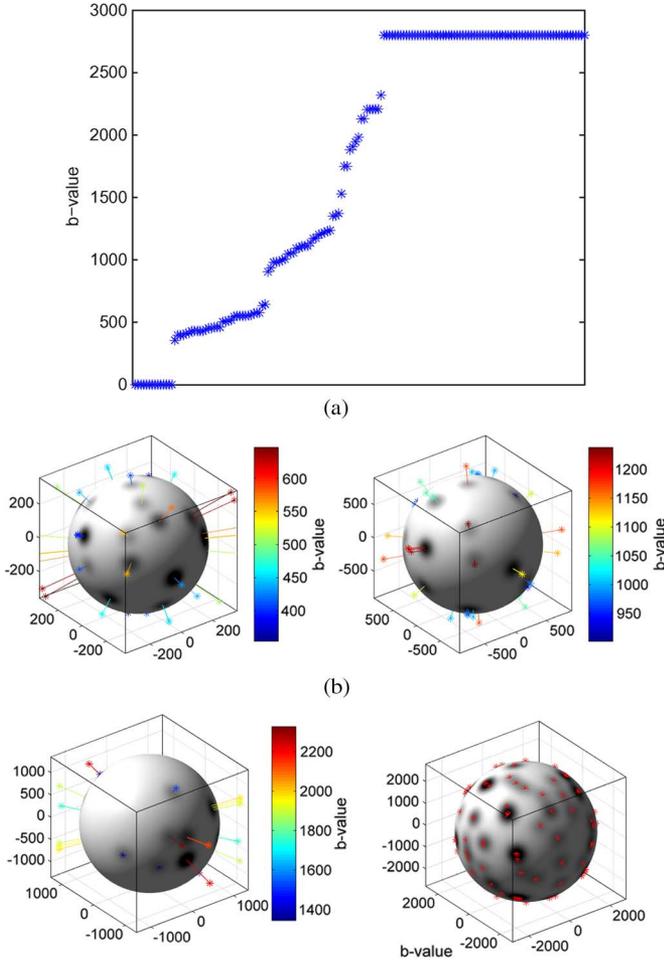


Fig. 1. Magnitude and direction of optimized gradients $O_{2,m}^{b_1}$.

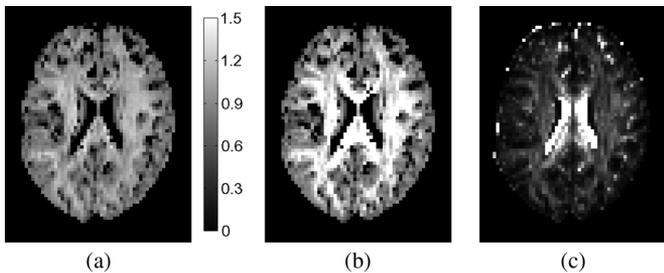


Fig. 2. (a) Mean kurtosis, (b) radial kurtosis, and (c) KA.

As in the acquired DKI dataset, the maximum b -value allowed in the optimization was 2800 s/mm^2 . Fig. 1(a) shows that a substantial number of gradients are located at this maximum b -value. This suggests that the precision might be improved by increasing the maximum allowed b -value even further. However, as the model (10) is based on a series expansion, it is not suited to extrapolate the magnitude of DW MR images acquired with higher b -values.

C. Comparison of the Precision

This section compares the CRLB of kurtosis parameters of the optimized gradient set $O_{2,m}^{b_1}$ with the traditional gradient set T . Fig. 2(a) and (b) shows the mean and radial kurtosis of the

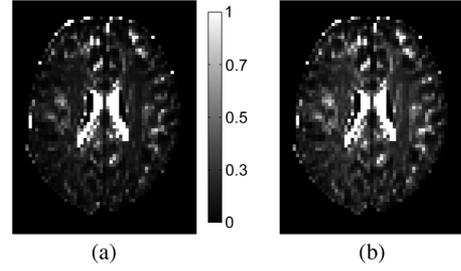


Fig. 3. $\sqrt{\text{CRLB}}$ of the mean kurtosis, when estimated with T or $O_{2,m}^{b_1}$. (a) $\sqrt{\text{CRLB}(T)}$. (b) $\sqrt{\text{CRLB}(O_{2,m}^{b_1})}$.

human DKI dataset, respectively. As is clearly visible, the radial kurtosis is substantially larger in the white matter structures, compared to the mean kurtosis. This indicates that the deviations from the Gaussian distribution are strongest in the radial direction. Fig. 3 shows the square root of the CRLB of the mean kurtosis, evaluated with the traditional and the optimized gradient sets, respectively. Thus, this figure displays a lower bound on the standard deviation of the mean kurtosis. In Fig. 3, it is clearly visible that the kurtosis parameters of the CSF cannot be precisely estimated by both gradient settings, as the square root of the CRLB of the mean kurtosis is high compared to the mean kurtosis of the CSF. The CRLB of the other tissues is lower, indicating more precise estimates. In Fig. 3, the differences between the CRLB of the mean kurtosis of the gradient sets are difficult to see. Therefore, the precision of a kurtosis parameter estimate obtained with the gradient sets T and $O_{2,m}^{b_1}$ was compared by evaluating the logarithm of the ratio of the CRLB of that kurtosis parameter

$$R(\boldsymbol{\theta}; T, O_{2,m}^{b_1}) = \ln \left(\frac{I_m^{-1}(\boldsymbol{\theta}, T)}{I_m^{-1}(\boldsymbol{\theta}, O_{2,m}^{b_1})} \right). \quad (28)$$

The value of R is 0 when both gradient sets have an equal CRLB of the kurtosis parameter in that voxel. Positive values indicate that the CRLB of the kurtosis parameter obtained with gradient settings T is larger than that obtained with the optimized gradient set $O_{2,m}^{b_1}$. Fig. 4 shows R of the mean kurtosis, radial kurtosis, and KA parameters. It is clearly visible that the gradient set $O_{2,m}^{b_1}$ improves the CRLB for all brain structures, except for the CSF. It was observed that the median reduction of the CRLB of the mean kurtosis in the gray matter was a factor 2.1. The factor is even larger than the value obtained in the simulation experiment with Θ_2 , which consists of a selection of gray and white matter voxels. Fig. 5 shows $R(\boldsymbol{\theta}, T, O_{2,m}^{b_1})$ for the diffusion parameters. The median of the CRLB of the MD in the gray matter is 12% larger for the traditional gradient settings compared with $O_{2,m}^{b_1}$. The precision of the FA and direction of the FE is almost equal for these two sets of gradient settings. Fig. 4 also shows the difference in performance between the gradient sets optimized with respect to the mean and radial kurtosis. As could be expected, the CRLB of the mean kurtosis is higher for $O_{2,r}^{b_1}$ than for $O_{2,m}^{b_1}$. The median CRLB of the mean kurtosis in the gray matter increases by 29%, the radial kurtosis is approximately the same and the CRLB of the KA increases by 19%.

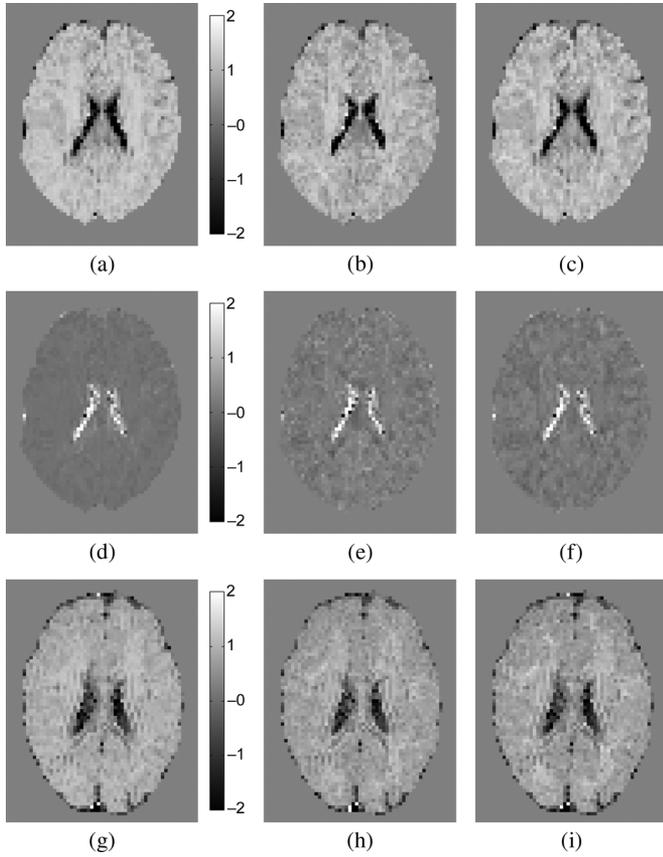


Fig. 4. $R(28)$ with (a), (d), (g) mean kurtosis; (b), (e), (h) radial kurtosis; and (c), (f), (i) KA of the diffusion kurtosis tensors of a human DKI dataset. The subparts (a), (b), and (c) are from a dataset recorded with the traditional gradient settings, compared with $O_{2,m}^{b1}$, $R(\theta, T, O_{2,m}^{b1})$. In (d), (e), and (f), the gradient settings optimized with respect to the mean kurtosis and radial kurtosis are compared by $R(\theta, O_{2,m}^{b1}, O_{2,r}^{b1})$, in terms of the precision of the mean kurtosis, radial kurtosis, and KA. The subparts (g), (h), and (i) are from a dataset recorded with the optimized gradient settings $O_{2,m}^{b1}$, compared with the traditional gradient settings $R(\theta, T, O_{2,m}^{b1})$.

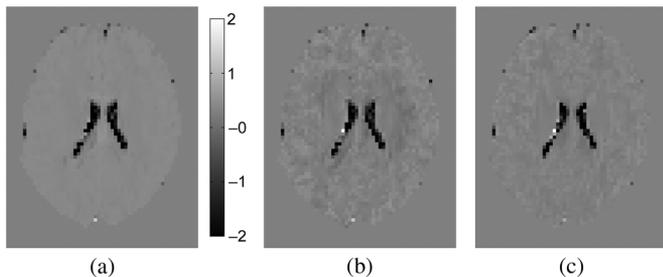


Fig. 5. $R(\theta, T, O_{2,m}^{b1})$ of the first dataset, for the (a) MD, (b) fractional diffusion anisotropy (FA), and (c) direction of the FE. As is visible in this figure, the precision of the DT parameters in the white and gray matter is not substantially changed when the traditional gradient settings are replaced by the gradient settings $O_{2,m}^{b1}$.

This is a further indication that the gradient settings $O_{2,m}^{b1}$ are better than $O_{2,r}^{b1}$.

The last row of Fig. 4 shows $R(\theta, T, O_{2,m}^{b1})$ of the second dataset, recorded with the optimized gradient settings $O_{2,m}^{b1}$ (accidentally) rotated 90° in the slice direction. These figures are

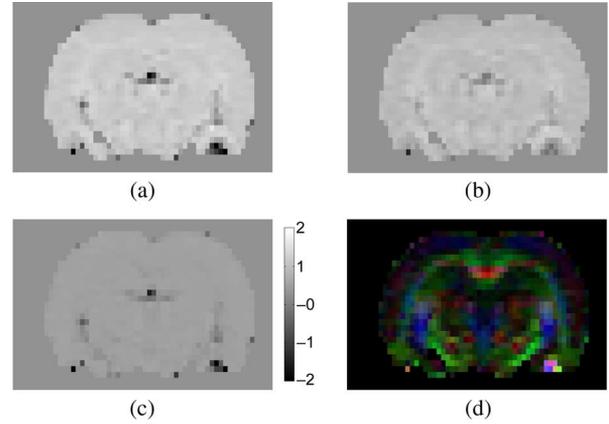


Fig. 6. Subparts (a), (b), and (c) show the relative performance of the traditional gradient set T , $O_{2,m}^{b1}$ and O_{rat} for the mean kurtosis. For localization, an FA map colored with the direction of the FE (FEFA) is provided in subpart (d). (a) $R(\theta, T, O_{rat})$; (b) $R(\theta, T, O_{2,m}^{b1})$; (c) $R(\theta, O_{2,m}^{b1}, O_{rat})$; and (d) FEFA map.

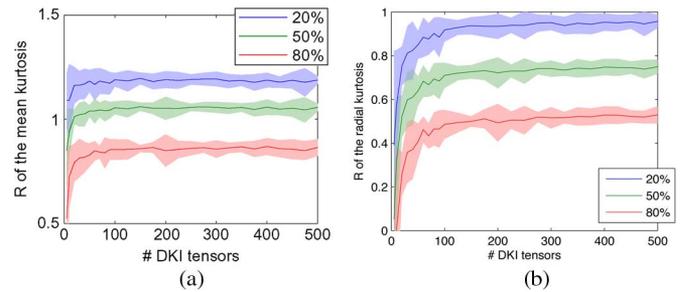


Fig. 7. Subparts (a) and (b) show the relative performance of the optimized gradient settings as function of the number of elements in the set Θ that is used during the optimization. All gradient settings were optimized with respect to the mean kurtosis. The performance of each optimized settings O is measured with $R(\theta, T, O)$ [(28)] for the (a) mean kurtosis and (b) radial kurtosis. For each set of optimized gradient settings, R is computed for θ of each of the 20 k voxels not used in any optimization. The three curves, 20%, 50% (= median), and 80%, give the value of R for which the indicated percentage of tested voxels has a larger value. The shaded areas indicate the 95% confidence interval of a single optimization, obtained by repeating the selection of voxels and optimization ten times for each number of voxels in the selection.

quite similar to the first row. This demonstrates that the optimized gradient settings are robust to differences between subjects and not very sensitive to changes in the acquisition parameters and orientation, as the acquisitions differ substantially. The median reduction of the CRLB of the mean kurtosis in the gray matter of this dataset was a factor of 1.9, which, considering the differences in acquisition, is close to the original improvement factor.

To further study the influence of changing recording settings, the DKI recording of a rat was made and the results are presented in Fig. 6. This figure shows that the gradient settings $O_{2,m}^{b1}$, which are optimized for the human brain, improve the precision of the mean kurtosis, compared to the traditional gradient settings T . For the entire brain, the median $R(\theta, T, O_{2,m}^{b1}) = 0.72$, while the gradient settings optimized for this acquisition have a median $R(\theta, T, O_{rat}) = 0.96$. So from these datasets, we find that the gradient settings specific for the rat brain improve the CRLB of the median kurtosis by 27%, compared to the gradient settings found for the human

brain. This relatively small difference, with respect to the substantially different MR system and subject, that the optimized gradient settings can be applied to slightly different acquisitions without substantial loss in precision of the diffusion and kurtosis parameters.

Finally, Fig. 7 shows the performance, as measured by $R(\boldsymbol{\theta}, T, *)$, of the optimized gradient settings $*$ as function of the number of elements in $\boldsymbol{\Theta}$. The value of R was computed for each $\boldsymbol{\theta}$ of the 20 k voxels from the first dataset that were not used in any optimization. As is clearly indicated by this figure, this performance quickly levels off and is essentially constant above 350 elements. Hence, 400 elements are sufficient to optimize the gradient settings.

V. CONCLUSION

In summary, this paper presents a novel method to optimize the diffusion-weighting gradient settings. This method is based on the minimization of the CRLB for estimating kurtosis parameters. The results show that the increase in precision that can be obtained, compared to a traditionally used set of gradients, is substantial. For the mean kurtosis estimated in gray matter voxels, an improvement of the CRLB with a factor of 2.1 was observed. Alternatively, when the required precision is already achieved using standard gradient settings, optimizing the gradient settings allows one to achieve the same precision from a reduced number of DW MR images or from a set of DW MR images with a reduced SNR (allowing a higher resolution).

The gradient settings are optimized using a prior distribution of DKI model parameters. This prior distribution can be obtained from DW MR images. In this paper, it was shown that the prior distribution substantially influences the precision of the kurtosis parameters. However, it was also shown that the performance of the optimized gradient settings is not substantially reduced when a different subject is scanned or other parameters of the acquisition are (slightly) changed.

APPENDIX

DERIVATION OF A PARAMETRIC MODEL OF THE DW MR IMAGES AND EXCESS KURTOSIS

This appendix gives a derivation of the expressions (7) and (18). First, it is shown that the first terms of the Maclaurin series (Taylor expansion around $q = 0$) of the logarithm of the DW image intensity $\mathcal{A}(q)$ (6) lead to simple expressions in terms of the diffusion and kurtosis coefficients. Next, the formula that computes the kurtosis in any direction from the diffusion and kurtosis tensors is derived.

Since magnitude DW MR images are recorded, and by construction $\mathcal{A}(q) = \mathcal{A}^*(-q)$, the magnitude is symmetric in q . With that, the following expression can be derived:

$$\left. \frac{\partial^2 \ln \mathcal{A}}{\partial q^2} \right|_{q=0} = \left[-\frac{\left(\frac{\partial \mathcal{A}}{\partial q} \right)^2}{\mathcal{A}^2} + \frac{\partial^2 \mathcal{A}}{\partial q^2} \right]_{q=0} \quad (29)$$

$$= 0 - \mathbb{E}_{f_{\mathbf{g}}}[x^2] = -2tD_{\mathbf{g}}. \quad (30)$$

Furthermore, the fourth derivative gives

$$\frac{\partial^4 \ln \mathcal{A}}{\partial q^4} = \mathbb{E}_f[x^4] - 3\mathbb{E}_f[x^2]^2 = 4t^2 D_{\mathbf{g}}^2 K_{\mathbf{g}}. \quad (31)$$

With these derivatives in terms of the diffusion and kurtosis coefficients, the series expansion of the logarithm of the DW signal can be given by

$$\ln \mathcal{A}(q) = \ln A_0 - D_{\mathbf{g}} q^2 t + \frac{1}{6} D_{\mathbf{g}}^2 K_{\mathbf{g}} q^4 t^2 + \mathcal{O}(q^6). \quad (32)$$

In general, the 3-D second derivative of $\ln A$ can be described by a symmetric tensor of rank 2

$$\frac{\partial^2 \ln \mathcal{A}}{\partial \mathbf{q}^2} = \frac{\partial^2 \ln \mathcal{A}}{q_i \partial q_j} = -2t D_{i,j} \quad (33)$$

from which the diffusion in the direction \mathbf{g} can be computed by

$$D_{\mathbf{g}} = \sum_{i,j=1}^3 g_i g_j D_{i,j}. \quad (34)$$

The 3-D fourth derivative of $\ln A$ can be described by a fully symmetric tensor of rank 4

$$\begin{aligned} \frac{\partial^4 \ln \mathcal{A}}{\partial \mathbf{q}^4} &= \frac{\partial^4 \ln \mathcal{A}}{q_i \partial q_j \partial q_k \partial q_l} \\ &= M_{i,j,k,l} = 4t^2 \left(\sum_{i=1}^3 \frac{D_{ii}}{3} \right)^2 W_{i,j,k,l}. \end{aligned} \quad (35)$$

With this fourth derivative and $D_{\mathbf{g}}$, the kurtosis in any direction can be computed by

$$\begin{aligned} K_{\mathbf{g}} \equiv K(\mathbf{g}) &= \frac{\sum_{i,j,k,l=1}^3 g_i g_j g_k g_l M_{i,j,k,l}}{4t^2 \left(\sum_{i,j=1}^3 g_i g_j D_{i,j} \right)^2} \\ &= \frac{\left(\sum_{i=1}^3 \frac{D_{ii}}{3} \right)^2 \sum_{i,j,k,l=1}^3 g_i g_j g_k g_l W_{i,j,k,l}}{\left(\sum_{i,j=1}^3 g_i g_j D_{i,j} \right)^2}. \end{aligned} \quad (36)$$

Note that, due to the division by the diffusion in the direction of \mathbf{g} , the kurtosis itself cannot be represented by a rank 4 tensor.

ACKNOWLEDGMENT

The data were acquired at the Institute for Functional and Metabolic Imaging, Belgium, and the Bio Imaging Laboratory, University of Antwerp, Belgium.

REFERENCES

- [1] D. C. Alexander, "Modelling, fitting and sampling in diffusion MRI," in *Visualization and Processing of Tensor Fields*. Berlin, Germany: Springer-Verlag, Mar. 2009, pp. 3–20.

- [2] D. C. Alexander, "Axon radius measurements in vivo from diffusion mri: A feasibility study," in *Proc. Comput. Vis., 2007. ICCV 2007. IEEE 11th Int. Conf.*, Oct., pp. 1–8.
- [3] H. Lu, J. H. Jensen, A. Ramani, and J. A. Helpert, "Three-dimensional characterization of non-Gaussian water diffusion in humans using diffusion kurtosis imaging," *NMR Biomed.*, vol. 19, pp. 236–247, Mar. 2006.
- [4] E. Fieremans, "Validation methods for diffusion weighted magnetic resonance imaging in brain white matter," Ph.D. dissertation, Univ. Gent. Faculteit Ingenieurswetenschappen, Gent, Belgium, 2008.
- [5] M. M. Cheung, E. S. Hui, K. C. Chan, J. A. Helpert, L. Qi, and E. X. Wu, "Does diffusion kurtosis imaging lead to better neural tissue characterization? A rodent brain maturation study," *NeuroImage*, vol. 45, no. 2, pp. 386–392, 2009.
- [6] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 2nd ed. New York: Hafner, 1967, vol. 2.
- [7] J. H. Jensen, J. A. Helpert, A. Ramani, H. Lu, and K. Kaczynski, "Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging," *Magn. Reson. Med.*, vol. 53, no. 6, pp. 1432–1440, Jan. 2005.
- [8] A. van den Bos, *Parameter Estimation for Scientists and Engineers*. New York: Wiley, Aug. 2007.
- [9] O. Brihuega-Moreno, F. P. Heese, and L. D. Hall, "Optimization of diffusion measurements using Cramer-Rao lower bound theory and its application to articular cartilage," *Magn. Reson. Med.*, vol. 50, pp. 1069–1076, 2003.
- [10] D. K. Jones, M. A. Horsfield, and A. Simmons, "Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging," *Magn. Reson. Med.*, vol. 42, no. 3, pp. 515–525, 1999.
- [11] S. Mori, *Introduction to Diffusion Tensor Imaging*. New York: Elsevier, 2007.
- [12] S. Mori, R. Itoh, J. Zhang, W. E. Kaufmann, P. C. Van Zijl, M. Solaiyappan, and P. Yarowsky, "Diffusion tensor imaging of the developing mouse brain," *Magn. Reson. Med.*, vol. 46, no. 1, pp. 18–23, 2001.
- [13] J. Sijbers, A. J. den Dekker, E. Raman, and D. Van Dyck, "Parameter estimation from magnitude MR images," *Int. J. Imag. Syst. Technol.*, vol. 10, no. 2, pp. 109–114, 1999.
- [14] O. T. Karlsen, R. Verhagen, and W. M. Bovée, "Parameter estimation from Rician-distributed data sets using a maximum likelihood estimator: Application to T_1 and perfusion measurements," *Magn. Reson. Med.*, vol. 41, no. 3, pp. 614–623, 1999.
- [15] J. Sijbers and A. J. den Dekker, "Maximum likelihood estimation of signal amplitude and noise variance from MR data," *Magn. Reson. Med.*, vol. 51, no. 3, pp. 586–594, 2004.
- [16] E. S. Hui, M. M. Cheung, L. Qi, and E. D. Wu, "Towards better MR characterisation of neural tissues using directional diffusion kurtosis analysis," *NeuroImage*, vol. 42, pp. 122–134, Apr. 2008.
- [17] E. H. L. Aarts and J. H. M. Korst, *Simulated Annealing and Boltzmann Machines*. New York: Wiley Chichester, 1987.
- [18] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.