Author(s): Liu, Jia; Gasbarra, Dario; Railavo, Juha

Title: Fast Estimation of Diffusion Tensors under Rician noise by the EM algorithm

Year: 2016

Version:

Fast Estimation of Diffusion Tensors under Rician noise by the EM algorithm

Jia Liu*

Department of Mathematics and Statistics, University of Jyväskylä, P.O.Box (MaD) FI40014 Finland

Dario Gasbarra

Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68 FI00014 Finland

Juha Railavo

Helsinki University Hospital

Abstract

Diffusion tensor imaging (DTI) is widely used to characterize, in vivo, the white matter of the central nerve system (CNS). This biological tissue contains much anatomic, structural and orientational information of fibers in human brain. Spectral data from the displacement distribution of water molecules located in the brain tissue are collected by a magnetic resonance scanner and acquired in the Fourier domain. After the Fourier inversion, the noise distribution is Gaussian in both real and imaginary parts and, as a consequence, the recorded magnitude data are corrupted by Rician noise.

Statistical estimation of diffusion leads a non-linear regression problem. In this paper, we present a fast computational method for Maximum Likelihood estimation (MLE) of diffusivities under the Rician noise model, based on the Expectation Maximization (EM) algorithm. By using data augmentation, we are able to transform a non-linear regression problem into the the generalized linear modeling framework, reducing dramatically the computational cost. The Fisher-scoring method is used for achieving fast convergence of the tensor parameter. The new method is implemented and applied using both synthetic and real data in a wide range of *b*-amplitudes up

---

*Corresponding author

Email addresses: `jia.liu@jyu.fi` (Jia Liu), `Dario.Gasbarra@helsinki.fi` (Dario Gasbarra), `juha.railavo@elisanet.fi` (Juha Railavo)

to 14000 s/mm^2. Higher accuracy and precision of the Rician estimates are achieved compared with other log-normal based methods. In addition, we extend the maximum likelihood (ML) framework to the maximum a posteriori (MAP) estimation in DTI under the aforementioned scheme by specifying the priors. We will describe how close numerically are the estimators of model parameters obtained through MLE and MAP estimation.

1. Introduction

Diffusion tensor imaging (DTI) is a powerful tool to detect, in vivo, the white matter anatomy and structures of the brain. The raw MR-data are collected by a magnetic resonance scanner and consist of spectral measurement from the displacement distribution of water molecules constrained into cellular structures. Diffusion anisotropy characterizes the nervous fibers.

After the Fourier inversion, the MR-signals are corrupted by a complex Gaussian noise, and consequently, the recorded measurement magnitudes, referred as diffusion weighted magnetic resonance imaging (DW-MRI) data, will follow the Rician distribution. The complex noise is composed of two components, where the real and imaginary parts are still independently Gaussian [2, 3, 4]. The simplest method for diffusion tensor estimation (DTE) is based on the linearized log-normal regression model, where the residual variance is assumed to be either constant (the Least Squares) or depending on the signal amplitude (the Weighted Least Squares). These Gaussian noise models fail to fit the high frequency data, which carry information about the higher order diffusion characteristics. In the existing literature [5, 6, 7] on the ML-estimation of diffusion tensors under the Rician noise, the maximization algorithm involves repeated computation of modified Bessel functions. By using data augmentation we are able to replace the Rician likelihood by a Poisson likelihood which is standard in the generalized linear modeling (GLM) framework.

Such simplification reduces dramatically the computational burden of the Fisher-

2

scoring maximization algorithm. This applies also at high *b*-amplitudes, where in the low signal regime measurements below a threshold are customarily coded as zeros. In the standard LS or WLS approaches, zero-measurements are problematic since they cannot be fitted by a log-normal distribution, and simply discarding them induces selection bias. The appropriately modeled noise level provides capability of data correction in further insights, e.g. removing artefacts from the raw data.

This paper is structured as follows. Section 2 describes the noise in MRI and data augmentation, specifying the statistical model for DTE. In Section 3 we discuss the implementation of the EM and the Fisher-scoring algorithms in the DTI context. In addition, we also specify priors for the parameters and discuss the computation of the Maximum a Posteriori Estimator (MAPE) under the same scheme. Section 4 illustrates the results from both synthetic and real data. In Section 6 we conclude with an overview of the methods and the undergoing developments. Theoretical details are left for the appendices.

2. GLM for MRI observations

2.1. Rician noise in MRI

In magnetic resonance imaging (MRI), we usually need to take the noise in the raw MR-acquisitions into account. The complex valued noise ε is composed of two *i.i.d.* Gaussian random variables with zero mean and variance σ^2, one for the real and the other one for the imaginary component. After the Fourier inversion, the signal intensity $S \geq 0$ is corrupted by a complex Gaussian noise, and $Y = |S + \varepsilon|$ will be observed. Consequently, the observed MR-signal magnitudes follow a Rician distribution resulting in the likelihood function

$$p_{S,\sigma^2}(y) = \frac{y}{\sigma^2} \exp\left(-\frac{y^2 + S^2}{2\sigma^2}\right) I_0\left(\frac{yS}{\sigma^2}\right),$$ (1)

where I_α is the α-order modified Bessel function of first kind. For $\alpha = 0$ it has also the following representation in terms of Gaussian hypergeometric series [8]:

$$I_0(2\tau) = {}_0F_1(1, \tau^2) = \sum_{n=0}^{\infty} \frac{\tau^{2n}}{(n!)^2}.$$ (2)

3

Let $t = S^2/(2\sigma^2)$, then Eq. (1) gives

$$P_{t,\sigma^2}(Y \in dy) = \frac{y}{\sigma^2} \exp\left(-t - \frac{y^2}{2\sigma^2}\right) I_0\left(\frac{y}{\sigma}\sqrt{2t}\right) dy \qquad (3)$$

with $\tau = yS/(2\sigma^2) = \sqrt{2t}y/(2\sigma)$.

2.2. Data augmentation

We follow the strategy presented in [9] implementing augmented data N from a Poisson distribution with mean $t > 0$. The likelihood for the observed data can be transformed from the Rician likelihood Eq. (3) to a joint augmented density

$$P_{t,\sigma^2}(N = n, Y^2 \in dy^2) = P_{t,\sigma^2}(N = n, X \in dx) \qquad (4)$$
$$= P_t(N = n)P_{\sigma^2}(X \in dx|N = n) = \frac{(tx)^n}{(n!)^2(2\sigma^2)^{n+1}} \exp\left(-t - \frac{x}{2\sigma^2}\right) dx \,,$$

where X is from the conditional distribution Gamma$(N + 1, 1/(2\sigma^2))$ given N. Eq. (4) provides a transformation from a non-linear regression problem to the GLM framework

$$f_{\xi,\phi}(z) = c(z,\phi) \exp\left(\frac{z\xi - a(\xi)}{\phi}\right) \qquad (5)$$

with z corresponding to the response in general, see [10] for more details.

3. Method

3.1. DW-MRI and parametrization

In DW-MRI, the signal is modeled as the first equality

$$S(\mathbf{q}) = S_0 \exp\left(-bd(\mathbf{g})\right) = S_0 \exp\left(Z\theta\right), \qquad (6)$$

where the control vector $\mathbf{q} \in \mathbb{R}^3$ is determined by the sequence of gradient pulses, $b = |\mathbf{q}|^2$, and $\mathbf{g} = \mathbf{q}/|\mathbf{q}| \in \mathscr{S}^2$ is a vector of unit length. The MR-signal decays exponentially with respect to the b-amplitude. Depending on the gradient direction $\mathbf{g}$ the decay is modeled by the reflection symmetric diffusivity function $d : \mathscr{S}^2 \to \mathbb{R}^+$.

Great efforts have been devoted to modeling the diffusivity, and in general we can have parametrization as the second equality in Eq. 6. In the simplest model the diffusivity is expressed by a symmetric and positive definite rank-2 tensor $D \in \mathbb{R}^{3\times3}$,

4

giving

$$\log S(\mathbf{q}) = \log S_0 - b\mathbf{g}^\top D\mathbf{g} = \log S_0 + Z\theta \ ,$$

where in the left hand side the diffusion tensor is parametrized as

$$\theta = (\theta_1,\dots,\theta_6)^\top := \left(D_{xx},D_{yy},D_{zz},D_{xy},D_{xz},D_{yz}\right)^\top$$

with a design matrix

$$Z = Z(\mathbf{q}) = -b\left(\mathbf{g}_x^2,\mathbf{g}_y^2,\mathbf{g}_z^2,2\mathbf{g}_x\mathbf{g}_y,2\mathbf{g}_x\mathbf{g}_z,2\mathbf{g}_y\mathbf{g}_z\right) \ .$$

In high angular resolution models (HARDI) (see e.g. [11]), the diffusivity is modeled with a totally symmetric Cartesian tensor D of order $n \in \mathbb{N}$, as

$$d(\mathbf{g}) := \sum_{\ell_1=1}^{3} \sum_{\ell_2=1}^{3} \cdots \sum_{\ell_{2n}=1}^{3} D_{\ell_1,\ell_2,\dots,\ell_{2n}} g_{\ell_1} g_{\ell_2} \cdots g_{\ell_{2n}} \ .$$

3.2. EM in MLE

In the optimization of the likelihood, we employ the EM (Expectation - Maximization) algorithm, which is one among the iterative methods in the MLE or in the maximum a posteriori estimation (MAPE). The EM algorithm proceeds in two steps and shortens the computational complexity by using augmented data. In terms of our case, in the E-step we calculate the expectation of the log-likelihood w.r.t the conditional distribution of N given by the observations and other parameters with fixed values. In the M-step, we find the ML parameter of S_0^2 and σ^2 by maximizing the augmented log-likelihood quantities. The computational details are listed in Appendix A.

Note that the data are obtained by given different b values and gradients in the experiment, they are discrete complex numbers, therefore, we use sums instead of integrals in the algorithms. The log-likelihood from Eq. (4) then is expressed as

$$Q := \log\left(p_{t,\sigma^2}(N=n,Y)\right) = c(Y,N) + N\log(t) - (N+1)\log(\sigma^2) - t - \frac{Y^2}{2\sigma^2}, \quad (7)$$

where $c(Y,N) = N\log(Y^2) - 2\log(N!) - (N+1)\log(2)$ does not depend on (t,σ^2) and will be omitted in the M-step. From Section 3.1, we have $t = S_0^2\exp(2Z\theta)/2\sigma^2$.

In the EM-iteration, given the current parameter estimates $(\theta^{(k)}, S_0^{2\,(k)}, \sigma^{2\,(k)})$, we update the conditional expectation of the augmented data by

$$\langle N \rangle^{(k)} := E_{t^{(k)}, \sigma^{2(k)}}\left(N \big| Y\right) = \frac{\tau^{(k)}\, I_1\left(2\tau^{(k)}\right)}{I_0\left(2\tau^{(k)}\right)} \quad \text{with} \quad \tau^{(k)} = \frac{Y S_0^{(k)} \exp(Z\theta^{(k)})}{2\sigma^{2(k)}}.$$

In the M-step we update σ^2 and S_0^2 by the recursions

$$(\sigma^{(k+1)})^2 = \left(\sum_{i=1}^{m}\left((S_0^{(k)})^2 \exp\left(2Z_i\theta^{(k)}\right) + Y_i^2\right)\right)\bigg/\left(2m + 4\sum_{i=1}^{m}\langle N_i\rangle^{(k)}\right) \qquad (8)$$

and

$$(S_0^{(k+1)})^2 = 2(\sigma^{(k)})^2\left(\sum_{i=1}^{m}\langle N_i\rangle^{(k)}\right)\bigg/\left(\sum_{i=1}^{m}\left(\exp\left(2Z_i\theta^{(k)}\right)\right)\right), \qquad (9)$$

where m is the number of acquisitions at each voxel.

For the tensor parameter θ, we employ a stabilized Fisher scoring method: given the stabilizing parameter $\alpha \in [0, 1]$, we iterate the recursion

$$\theta \to \theta + \left((1-\alpha)J(\theta) + \alpha \mathscr{S}(\theta)^\top \mathscr{S}(\theta)\right)^{-1} \mathscr{S}(\theta), \qquad (10)$$

until convergence to a fixed point [12]. In Eq. (10) the score $\mathscr{S}(\theta)$ is given by

$$\mathscr{S}(\theta) = 2\sum_{i=1}^{m} Z_i \langle N_i\rangle^{(k)} - \left(S_0^{(k)}/\sigma^{(k)}\right)^2 \sum_{i=1}^{m}\exp(2Z_i\theta)Z_i^\top,$$

and the corresponding Fisher information is

$$J(\theta) = 2\left(S_0^{(k)}/\sigma^{(k)}\right)^2 \sum_{i=1}^{m}\exp(2Z_i\theta)Z_i^\top Z_i.$$

The initials of the EM algorithm can be obtained through the least squares (LS) from a truncated dataset with the diffusion weighting ranging from $0 \sim 1000 s/mm^2$ in order to fit the Gaussian model (see [13], [14]). To pursue higher quality of the initials, we could further apply the weighted least squares (WLS) described in [4]. In the Appendix B we compare the differences between our EM algorithm and the direct optimization of the Rician likelihood in Eq. (1), which is commonly used to compute the MLE in DTI. It should be noted that the well-known EM algorithm is needed because of the latent augmented variables; it does not decrease the marginal likelihood of the data.

6

3.3. EM in MAPE

In the Bayesian framework, the maximum a posteriori estimation (MAPE) aims to obtain the point estimates by maximizing the posterior density. The advantage of MAPE over the likelihood approach is that the prior knowledge of the unknown parameters of interest with respect to (w.r.t.) the observed measurements can be transferred into the modeling framework by the prior distribution. Specifically, we can include restrictions to the parameters in terms of probability distributions, for instance regularization can be simultaneously included into the model by adding the knowledge of tuning parameters. Compared with the likelihood approach, Bayesian strategy typically yields less uncertainty and better knowledge of the parameters (the posterior) as it is analysing the probability distribution of every parameter of interest. The difference between MLE and MAPE in this scenario is in the prior probability $\pi(\xi)$. Given the data y, the normalizing constant in the posterior density $\pi(\xi|y)$ does not depend on the parameter ξ. We find the MAPE by maximizing the joint density $\pi(\xi)p_\xi(y)$, and this is achieved by iterating the EM-recursion

$$\xi^{(k+1)} = \arg\max_{\xi \in \Xi} \left\{ E_{\xi^{(k)}} \left(\log p_\xi(z,y) \big| y \right) + \log \pi(\xi) \right\} \tag{11}$$

with the penalization $\log \pi(\xi)$ until convergence to a fixed point. The log-prior penalization term has a regularizing effect, which vanishes asymptotically as the sample size increases [7].

In DTE, we can assign conjugate priors in light of Section 3.2 for σ^2 and S_0^2. Since we have only weak knowledge of the tensor parameter θ, we may choose non-informative priors which are either scale- or shift-invariant [15]. A simple Bayesian hierarchical model is obtained after the following choices:

- σ^2 has scale invariant improper prior with density $\pi(\sigma^2) \propto 1/\sigma^2$,

- $S_0^2 \sim \text{Gamma}(c_1, c_2)$, where c_1, c_2 are very small.

- $\theta \in R^d$ has the isotropic centered Gaussian prior $\mathcal{N}(0, \Omega^{-1})$, where Ω is a $d \times d$ precision matrix.

7

The penalized EM-updates for MAPE are given by

$$(\sigma^{(k+1)})^2 = \left(\frac{1}{2}\sum_{i=1}^{m}\left((S_0^{(k)})^2\exp(2Z_i\theta^{(k)})+Y_i^2\right)\right)\bigg/\left(\sum_{i=1}^{m}(2\langle N_i\rangle^{(k)}+1)+1\right) \quad (12)$$

and

$$(S_0^{(k+1)})^2 = \left(\sum_{i=1}^{m}\langle N_i\rangle^{(k)}+c_1\right)\bigg/\left(\frac{1}{2(\sigma^{(k)})^2}\sum_{i=1}^{m}\left(\exp(2Z_i\theta^{(k)})+c_2\right)\right). \quad (13)$$

Additionally, this gives the modified score and Fisher scoring

$$\tilde{\mathscr{S}}(\theta) = \mathscr{S}(\theta) - \Omega\theta, \quad \text{and } \tilde{J} = J(\theta) + \Omega, \quad \text{respectively.}$$

Under our Bayesian model with weak priors the MAP estimation Eq. (12) and Eq. (13) are similar as the ML updates Eq. (8) and Eq. (9). Indeed, usually $\sum_{i=1}^{m}\langle N_i\rangle \gg 1$, and we can omit the difference between Eq. (8) and Eq. (12). Then when c_1 and c_2 are small enough, the difference between the likelihood and posterior mode of S_0, expressed in Eq.(9) and Eq. (13) respectively, can also be ignored. The only difference when updating θ is that we have considered the correction between the elements of a tensor represented by the prior distribution, the inverse covariance matrix, Ω. Such correction may be ignorable.

Remark: By the normalized likelihood, the MLE can be treated as a special case of the MAPE where the precision of the parameters depend on the chosen prior. If the effects of the priors are weak enough to be ignored, then the posterior distribution is asymptotically approximated by the likelihood. The consequence is that numerically the MAP tend to the ML estimates numerically. Such remark is not unusual (see [16]) but nearly has never appeared in the DTI literature.

4. Results

4.1. Synthetic Data

Experiment 1. We first simulate four datasets by choosing a positive tensor of 2nd order and of 4th order, respectively from the same voxel with fixed S_0 (5.4595 in logarithmic level) and two different noise variance σ^2. The synthetic data in the experiment

arise from models with parameter values (the same gradients, *b* values and the number of replication which had been used to collect a real human dataset) resembling the real scenario. Each dataset contains 1440 ($32 \times 15 \times 3$) measurements corresponding to 32 distinct gradients and 15 distinct increasing *b* values (knots), and then being repeated three times. Furthermore, the *b* knots gradually increase every 32 gradients up to $14000s/mm^2$ with in total 480 experimental parameters. The ground truth (GT) of high (H-) and low (L-) Rician noise σ are 93.0405 and 12.8821, respectively. Thus we get the (nondiffusion weighted) non-dw SNR ($:= S_0/\sigma$) being 2.5256 and 18.2408, respectively, which fall into the wide range of clinic settings (< 25) [6]. Firstly, we give an overview of the data which are used in this experiment under the signal 2nd order tensor model in Fig. 1. Fig. 1a and Fig. 1b describe the generated data and the corresponding true signal intensities under the Rician noise model from the low and the high noise case, respectively, where we only take the first replication (480 measurements) as an example due to the similar behaviour of the other two repeats. From Fig. 1b, we can see that the data depicted by the blue curve is much more noisy than that in Fig. 1a. The corresponding diffusion profile of the 2nd order tensor is shown in Fig.
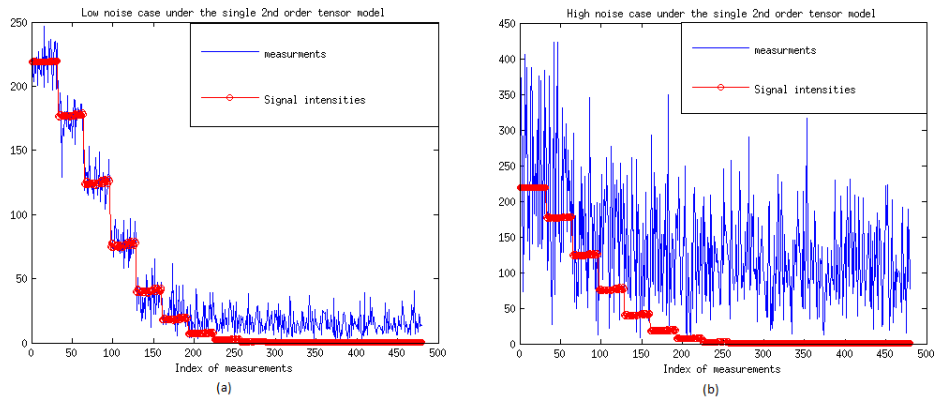

Figure 1: The thick curve represents the generated data and the red curve gives the corresponding true signal intensities. Fig. 1a and Fig. 1b describe the generated data and the corresponding true signal intensities under the Rician noise model from the low and the high noise case, respectively.

2, where the diffusion profile under the signal 2nd order tensor model represented as an ellipsoid can somehow explain the extent of the departure from normality in the the

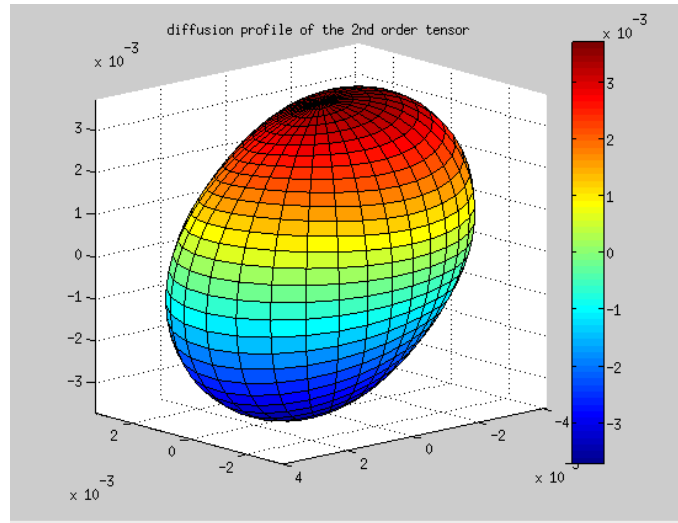movements of water molecule. In addition, we plot the corresponding diffusion profile

Figure 2: Scatter plot of the diffusion profile under the selected 2nd order tensor.

of the 4th order tensor in this experiment in Fig. 3, which is also considered to account for possible departures of the observed diffusion from normality.
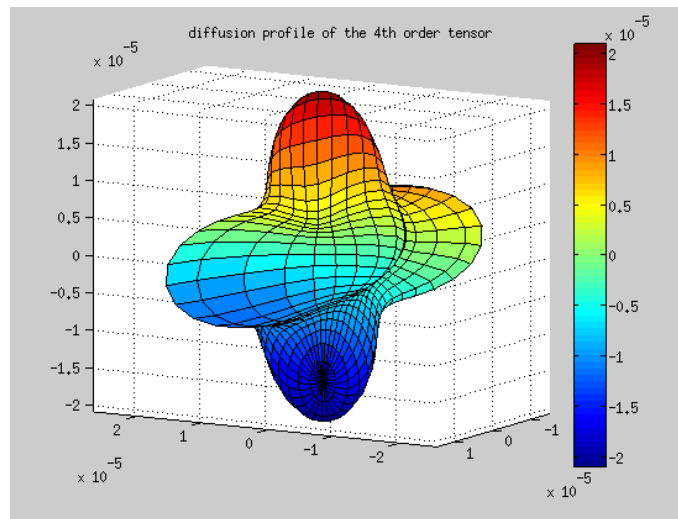

Figure 3: Scatter plot of the diffusion profile under the selected 4th order tensor.

To compare the performance, we plot the ML estimated signals and the correspond-

10

ing GT as a function of b values shown in Fig. 4, where we only consider the first 480 measurements as an illustration. The signals are calculated by averaging the 32 gradients for each distinct b value. In Fig.4a the signals of ground truth are from the 4th order tensor. However, in reality the ground truth should be unique no matter what choice of angular resolution of the tensor is. Actually in this experiment the signals of ground truth from the 2nd and the 4th order have very small difference (the max modulus (m.) deviation in logarithmic scale is less than 0.1, and the mean m. deviation is 0.0374). In order to distinguish the results from different datasets, we plot the results in log scale in Fig.4b, where we legend the logarithmic signals from the 4th order tensor as GT due to the very small differences mentioned above. The GT are displayed by the thick blue line. As Fig. 4b points out, the results from the dataset under the single 4th order tensor model at the high noise level has 'large' deviation from the GT, but the estimates from the other cases fit the GT quite well. Furthermore, we calculate the empirical signal to noise ratio (SNR $:= \frac{S}{SD_v} = \frac{mean(S_g(b,g))}{\sigma}$), and only consider one replication. Here instead of averaging the signal intensities of the whole acquisitions as defined in [17], we average the 32 gradients (g) at each distinct b values for representing the changes of the SNR when b value is increasing. To distinguish the difference, we again plot the results from the first 480 measurements in logarithmic level depicted in Fig.5. It is shown that in the high-noise level case, the results under the single 4th order tensor model have a bit larger bias when $b \geq 3000 s/mm^2$.

Experiment 2. For comparison of the methods, we generate 100 datasets from the high (Fig. 6, Fig. 7 Fig. 8) noise case and another 100 datasets from the low (Fig.9, Fig.10) noise case under the same 4th order tensor as in Experiment 1 and compare the sample means of SNR (SNR $:= \frac{S}{SD_v} = \frac{mean(S_{g,r}(b,g,r))}{\sigma}$) of the whole 1440 measurements in each sample data with the corresponding GT from the different methods, where the mean of the signals in the numerator is calculated by averaging the 32 gradients (g) and the total number of replications (r) from the whole measurements. Note that here we also average the number of replication in each dataset. Fig.6 represents the results from the datasets generated by the high-noise level, where "$*$" denotes that only the low frequencies (b values less than $1000 s/mm^2$) are considered in the estimation. This figure

11
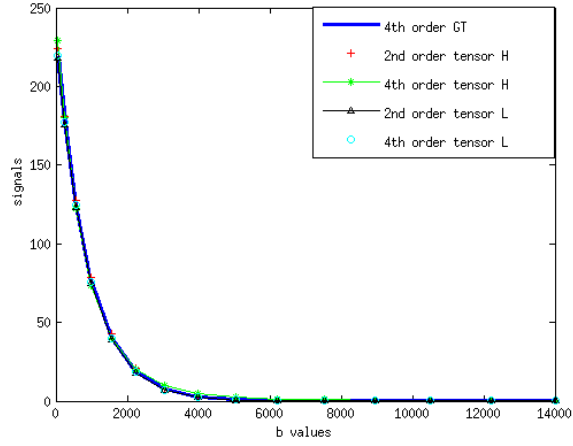
(a)
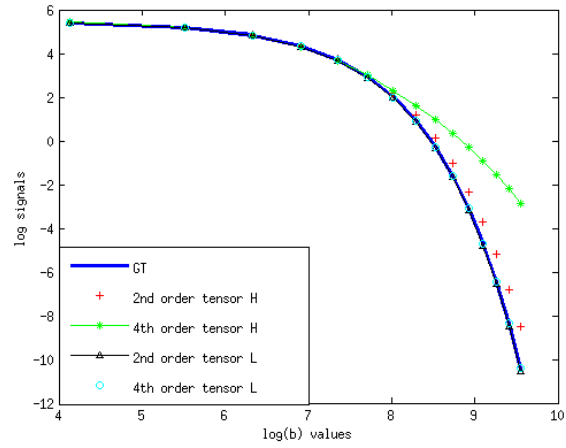
(b)

Figure 4: Fig. 4a. represents the signals $S(b) = S_0 exp(Z\theta)$ calculated from the estimated diffusion profile by the proposed MLE method. The thick-blue line depicts the signal intensities of the GT from the 4th order tensor. The green-start line and the cyan circles show the results under the 4th order tensor model from the datasets of the high- and low- noise levels, respectively. The triangular-black line and the red crosses are the results under the single 2nd order tensor model from the datasets of the high- and the low- noise levels, respectively. Fig. 4b is the corresponding results in log scale.
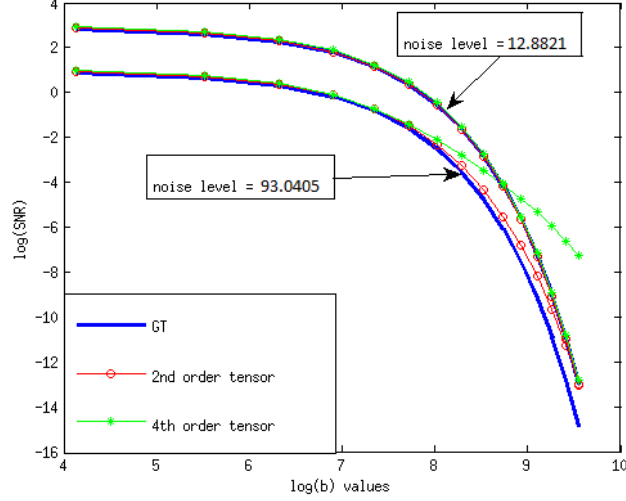
12

Figure 5: Empirical logarithmic SNR as a functions of $\log(b)$ values. The GT are represented by the thick-blue lines, of which the upper curve is from the low non-dw SNR corresponding the low noise level with $\sigma = 12.8821$, while the bottom one has the high noise level with $\sigma = 93.0405$. The red-circle lines are the fitted profile under the single 2nd order tensor model, and the green-star lines show the empirical SNR under the single 4th order tensor model.

reveals that the fitting profile by our method is the best, while the WLS results from the whole data space are much worse than the others. To compare the further performance, we compute the sample mean of signal intensities, and as an example we pick up from the first replication those intensities with a low b value. The result is in Fig. 7, from which, we can see that our results are slightly over estimated from the high-noise level data, but still being the best. The results from the other two methods are under estimated. In addition, we compute the sample mean of signal decay $S(b)/S(0) := \exp(Z\theta)$ by the tensor coefficients averaging by the gradients for obtaining the mean square errors. Fig. 8 describes the mean square error of signal decay in log level as a function of b values. Note that the results by the LS* and the WLS* are extrapolated to the high-frequency region by using the same design matrix Z and their tensor estimates. This figure reveals that even in the region of low b values ($b = 800 - 1000s/mm^2$), our method still performs better than the others.

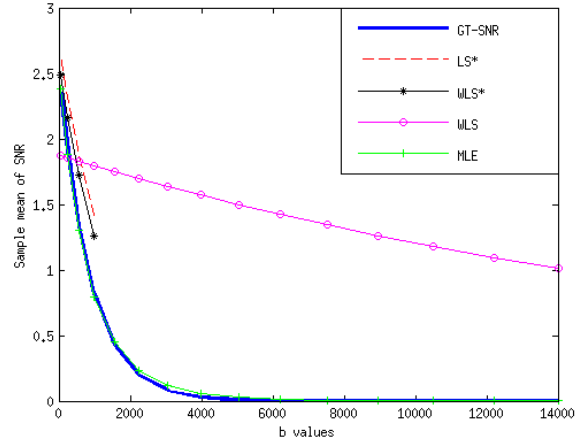Fig. 9 and Fig. 10 correspond with Fig. 7 and Fig. 8 from the 100 sample data

13

Figure 6: Sample mean of SNR as a function of b values. The sample means are calculated from 100 simulated datasets. The SNR are calculated from the estimates estimated by the different methods. The thick-blue curve represents the SNR of the GT. The red-dash line and the black-star line are the estimators by the LS and the WLS with the truncated datasets, respectively. The cyan-circle line is the results through the WLS, and the green-cross line is empirical values by our MLE method.
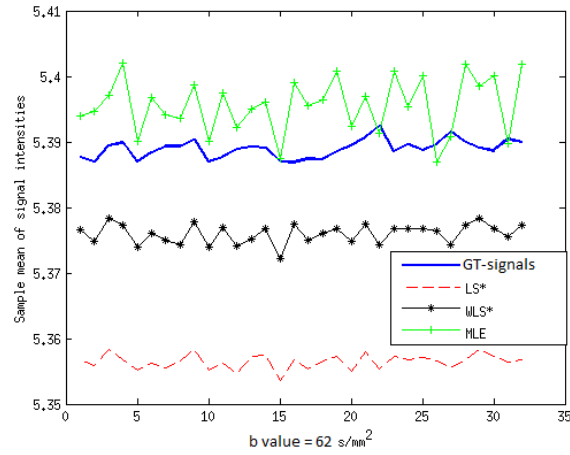

Figure 7: Sample mean of signal intensities. Again the thick-blue curve represents the GT. The red-dash line and the black-star line are the results by the LS and the WLS methods with the truncated datasets, respectively. The green-cross line show the results by our MLE method. We did not show the results by the WLS from the whole dataset as the bad performance in Fig. 6.
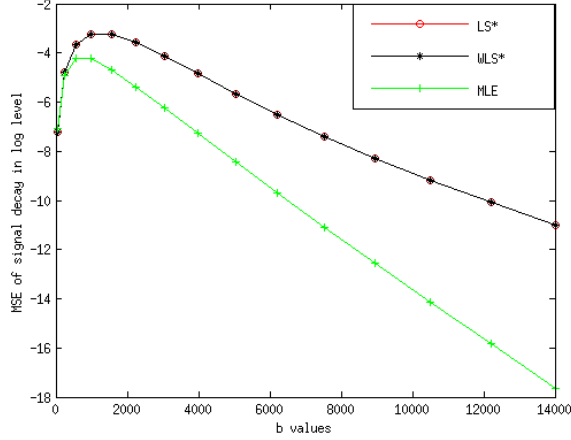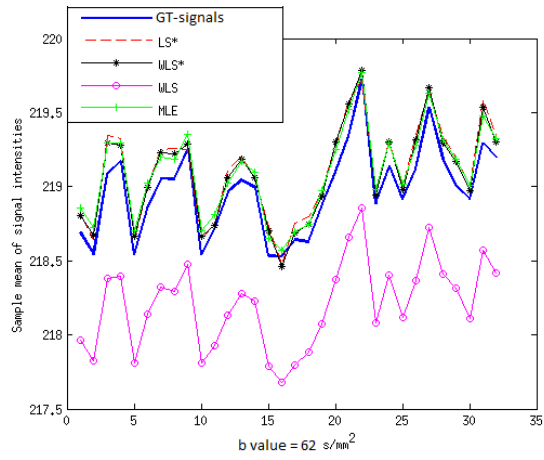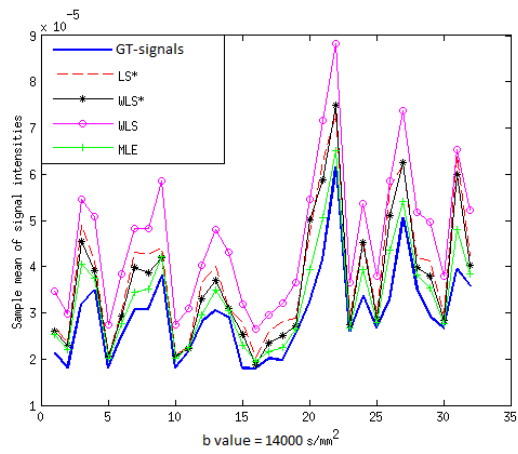
14

Figure 8: MSE of sample mean of averaged signal decay as a function of the distinct *b* values from the first 480 measurement. The red-circle line and the black-star line the results by the LS and the WLS methods with the truncated datasets ($b \leq 1000s/mm^2$), respectively. They are almost overlapping. The green-cross line shows the results by our MLE method. We did not show the WLS results from the whole dataset due to the bad performance in Fig. 6.

generated by the low noise. Fig. 9a reveals that the estimated signal intensities from our method are roughly similar than the results from the LS* and the WLS* when the *b* value equals to $62s/mm^2$. In Fig. 9b again the signal intensities by the LS* and the WLS* are extrapolated to the high-frequency region by the estimated diffusion profile, and our method shows a better fitted profile than the others. Fig. 10 describes the mean square error of signal decay as a function of *b* values. Since the difference is visible, we do not need rescale the results in the log level. This figure reveals at $b \leq 1000s/mm^2$, the LS*, WLS* and the WLS methods perform quite similarly, and the results by our method represent the smallest MSE in the whole region of the frequencies.

All the synthetic experiments were carried out on a 64-bit 4 core computer with 16 Gb RAM, and the CPU of each core is 3.40GHz with MATLAB. The average computational time of the aforementioned MLE method under the 4th order tensor model is 0.5435 seconds (an example record from the 100 datasets under low noise case), which is extremely shorter than the minutes running time per voxel from the current standard methods such as MATLAB Nelder-Mead based or gradient-based estimators

15

(a) low *b* value, $b = 62s/mm^2$

(b) high *b* value, $b = 14000s/mm^2$

Figure 9: Sample mean of signals intensities. The plots illustrate the means of signal intensities at $b = 62$ and $14000s/mm^2$, respectively, of each gradient from the first replication estimated by the four methods. The red-dash line and the black-star line are the results by the LS and the WLS methods with the truncated datasets, respectively. The green-cross line show the results by our MLE method, and cyan-circle line is the results through the WLS.
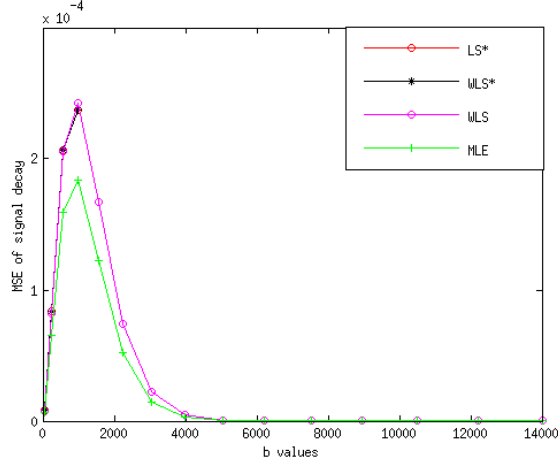
(see [18, 19]).

16

Figure 10: MSE of sample mean of averaged signal decay as a function of the distinct *b* values from the first 480 measurement. The red-circle line and the black-star line the results by the LS and the WLS methods with the truncated datasets ($b \leq 1000 s/mm^2$), respectively. They are almost overlapping with the results by the WLS. The green-cross line show the results by our MLE method.

4.2. Real Data

The data consist of 4596 diffusion MR-images of the brain of an healthy human volunteer, taken from four 5*mm*-thick consecutive axial slices, and measured using a Philips Achieva 3.0 Tesla MR-scanner. The image resolution is 128×128 pixels of size 1.875×1.875 *mm²*. After masking out the skull and the ventricles, we remain with a region of interest (ROI) containing 18764 voxels. In the protocol, we used all the combinations of the 32 gradient directions with the *b*-values varying periodically in the range $0 - 14000 s/mm^2$, with $2 - 3$ repetitions, for a total of 23 323 644 data points. The average computational cost per voxel by our method under the 4th order tensor model from this dataset is 1.8331 seconds. We illustrate the results mainly under the 4th order tensor model. Fig.11 shows the mean diffusivity (MD) and the fractional anisotropy (FA) of diffusion from two consecutive slices, where FA is computed from the results under the 2nd order tensor model, which is given by

$$\text{FA} = \frac{\sqrt{3((\lambda_1 - E[\lambda])^2 + (\lambda_2 - E[\lambda])^2 + (\lambda_3 - E[\lambda])^2)}}{\sqrt{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}. \qquad (14)$$

17

The average values of FA from these two ROI are 0.2769 and 0.2861, respectively. The color in FA represents the orientations of the fibers. Under the 4th order tensor model, MD is expressed as

$$MD = \frac{1}{5}(D_{1111} + D_{2222} + D_{3333} + 2D_{1122} + 2D_{1133}$$
$$+2D_{2233}) = \frac{1}{5}\text{trace}(D). \tag{15}$$

The average values of MD from Slice 3 and 4 are 6.248e-03 mm^2/s, 6.045e-03 mm^2/s, respectively, and we have the same estimated values of MD under the 2nd order tensor model.

We also plot the Rician noise map of σ from the two consecutive slices shown in Fig. 12, where the artefacts are clearly depicted by white color representing very high noise, which reveal the true scenario from the raw MR images, and are confirmed independently by our estimation.

Visualization of angular resolution of DTI data under different tensor models from the region of interest (ROI) of two consecutive slices are displayed in Fig. 13, where the ROI is near the hippocampus and the empty spaces inside of left parts of the diffusion profiles (DP) are the masked ventricle. DP under the 4th order tensors provide detailed information of diffusion through the higher angular resolution. In addition, the colors represents the principle orientations of diffusion at each voxel. These tensor profiles are plotted by MATLAB fanDTasia toolbox [21]. We also conduct the experiment with the real data on the 64-bit 4 core computer with 16 Gb RAM, and the CPU of each core is 3.40GHz with MATLAB. The total running time is 2.9733e+04 and 3.4395e+04, equally 1.5846 and 1.8331 seconds per voxel in average under the 2nd and 4th order tensor model, respectively.

Note that the algorithms presented in this work are under the assumption of voxel independence, therefore, the algorithms are parallelizable across voxels. The code related to the proposed method and the above results is available by request, which can also work on the cluster by parallel computation pixel by pixel.
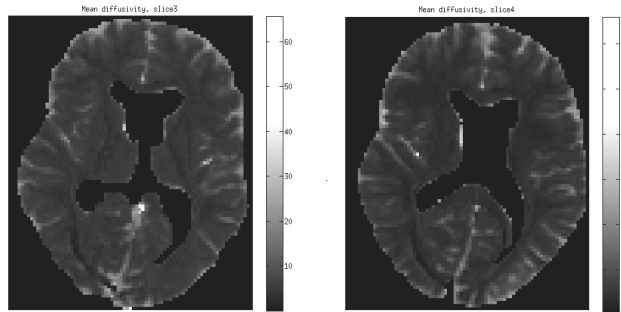
(a) Mean diffusivity (MD)
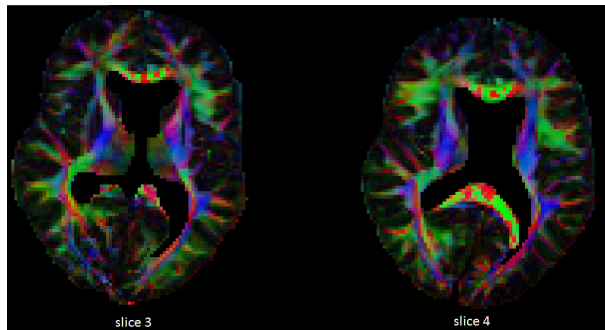
(b) Fractional anisotropy (FA)

Figure 11: MD and FA maps from two consecutive slices, where the estimated FA are computed under the 2nd order tensor model. The color in FA represents the orientations of the fibers: Red, left-right; Green, anterior-posterior; Blue, superior-inferior. The color coded FA maps are drawn by using the software ExploreDTI [20]. The corresponding MD maps are from the results under the 4th order tensor model, where the white spots corresponding to the corrupted data (artefacts) with measured magnitudes increasing to high b values.

5. Method Comparisons

Comparison between our EM method and the traditional MLE [7]. In this section, we discuss the differences between our data-augmentation based on the EM algorithm and on the typical MLE method through direct maximization at the Rician log-likelihood Q_r. Detailed calculation can be found in Appendix B.

1. We do not need to calculate all the elements of the Hessian as we can directly find
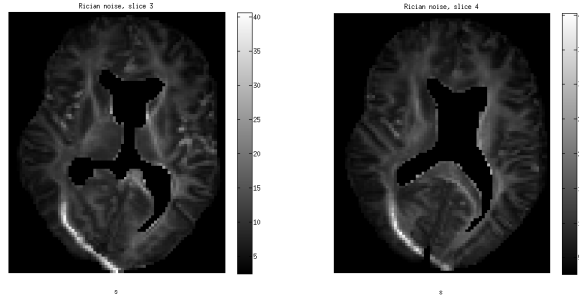
Figure 12: Rician noise map from two consecutive slices. The white curves in the left bottom of the slices depict the artefacts corresponding to very high noise.

the modes of S_0^2 and σ^2 by data augmentation. A small improvement appears in the reparametrization of S_0 or $\log S_0$ by S_0^2.

2. In the E-step we compute

$$\langle N_i \rangle = E_{\theta^{(k)}, \sigma^{2(k)}, S_0^{2(k)}}(N_i | Y_i), \qquad (16)$$

which does not depend on the parameters θ, σ^2 and S_0^2. In the M-step we use Eq. 16, the recursive values from $\theta^{(k)}, \sigma^{2(k)}, S_0^{2(k)}$, instead of solving the intractable formula w.r.t those parameters. This dramatically reduces the computation of the score from Eq.(B.2,B.1,B.3) to Eq.(A.3,A.2,A.4), respectively.

3. The EM algorithm allows us to use empirical values from Eq.(16) to compute the Fisher information. Our Fisher information $J(\theta)$ which fits the whole range of SNR and is slightly bigger than the approximated one, $\mathscr{I}_r(\theta)$, expressed in (Eq. (B.4)), which requires heavy mathematical calculations to deal with different expectations (see [7] for more details). In addition, when computing the score of θ in Eq. (10), we do not need to update the items containing N_i as they are fixed values from Eq.(16). All those lead to reduced computation in practice.

Comparison between our EM method and the EM method described e.g. in [22, 1].
Firstly, the theory part of the comparing EM method can be found in Appendix C.

1. In terms of the EM algorithm, both methods are likely in computation. Since the augmented data are calculated in the E-step by the knowns and parameters in the
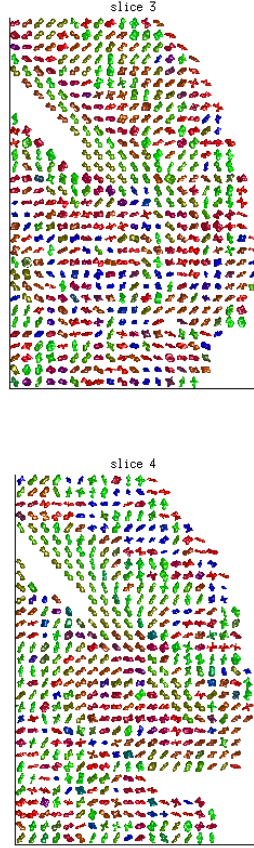
20

Figure 13: Visualization of the 4th order diffusion tensor profiles from two consecutive slices of a ROI. The color-code represents the main principal direction of the diffusion: Red, left-right; Green, anterior-posterior; Blue, superior-inferior.

current iteration given by, respectively

$$\langle N \rangle^{(k)} := E_{S^{(k)}, \sigma^{2(k)}} \left(N \middle| Y \right) = \frac{\tau^{(k)} I_1 \left(2\tau^{(k)} \right)}{I_0 \left(2\tau^{(k)} \right)} \, ,$$

$$\langle \cos(\varphi) \rangle^{(k)} = E_{S^{(k)}, \sigma^{2(k)}} \left(\cos(\varphi) \middle| Y \right) = \frac{I_1 \left(2\tau^{(k)} \right)}{I_0 \left(2\tau^{(k)} \right)},$$

$$\text{with} \quad \tau^{(k)} = \frac{Y S_0^{(k)} \exp(Z \theta^{(k)})}{2\sigma^{2(k)}}.$$

In the M-step, we calculate the partial derivative of Q w.r.t. σ^2 and S_0. Such

21

derivatives are straightforward to compute as presented in [1]. Then the computation till now from both methods should be roughly similar. The difference is that, in our EM algorithm, we update θ, the tensor parameter by a stabilized Fisher scoring method.

2. In theory, the augmentation in the two EM algorithms have essential difference, that is, they are working in different space. The implemented augmentation is in the natural integer space, while the introduced augmentation in Appendix C works on the phase data space.

3. In terms of Bayesian strategy, both methods can be totally different, because we can include the prior knowledge of the argument data through the prior distributions, then N will be generated from the reinforced Poisson distribution (see [9]) and $\cos(\phi)$ will be obtained from the Von Mises distribution given in Eq. (C.1).

6. Discussion

Our method substantially differs from the previous ones in the literature and the advantages are summarized by the following points: 1) We implement the recently developed data augmentation method [9], which allows the non-linear regression problem to be transformed into the GLM framework in DTE. 2) Subsequently, the computation is dramatically reduced due to the tractable modes of parameters of interest in the sense of point estimation. In addition, when employing Fisher-scoring scheme we simplify the complexity of the Fisher information. 3) Our Rician noise model can be combined with any tensor model in different representation, such as spheric harmonic expansion, by reparametrization. 4) Either ML or MAP estimation yields more accurate estimates than the LS and the WLS do. In addition, high frequencies from the low SNR data and the zero measurements are also included into the estimation. These data are known to contain detailed anatomical information of the complex tissue in vivo. 5) Our method leads to significantly less biased estimates of the noise level, which plays key role in denoising the MRI and cleaning the artefacts.

Positivity constraints. The physical feature of diffusion requires the tensor to be positive definite. Our model allows to check the positivity of diffusivity in the tensor

22

updates under the scheme of Fisher-scoring method. For the rank-2 tensor model, the constraining is fairly easy to do by computing the eigenvalues of the tensor matrix *D*. For HARDI, Barmpoutis et al. [11] propose the Gram matrix approach, using the quartic form to guarantee the positivity. Other methods such as [23] address the constraint by calculating the Z-eigenvalue polynomials.

MLE VS MAPE. In this work, we did not list the results from MAPE but we emphasize the differences between these two methods. Bayesian methods have advantages in the learning process, meaning that they may gain extra information from the prior knowledge. When the prior is weak, like in our case, we learn things from the data, what we actually do when approaching the problem through frequentist statistical modeling. In order to learn the uncertainty of the diffusion parameters, a fully Bayesian approach is highly recommended to characterize the posterior parameter distributions rather than point estimation.

7. Acknowledgement

Appendix

A. MLE by the EM algorithm in DTI

We consider the Rician noise model with the Poissonian data augmentation of Section 2. The latent augmented variable N conditionally on X, Z is given by

$$p_{t,\sigma}(N = n|X,Z) = \frac{1}{I_0(2\tau)} \frac{\exp(-2\tau)\tau^{2n}}{(n!)^2}, n \in \mathbb{N}, \text{ with } \quad \tau = \sqrt{\frac{Xt}{2\sigma^2}} \text{ and } \quad X = Y^2 .$$

It follows [9] that this discrete distribution is referred as reinforced Poisson distribution with parameter τ.

In the EM algorithm we need to compute the conditional expectation of N conditionally on X and the design matrix Z. Given the current values $t^{(k)}$, $\sigma^{2(k)}$, then

$$\langle N \rangle^{(k)} := E_{t^{(k)}, \sigma^{2(k)}} \left(N | X, Z \right) = \sum_{n=1}^{\infty} n p_{t,\sigma} (N = n | X, Z)$$

$$= \tau^{(k)} / 2 \frac{d}{d\tau^{(k)}} \log {}_0F_1(1, (\tau^{(k)})^2) = \tau^{(k)} / 2 \frac{d}{d\tau^{(k)}} \log J_0(2\tau^{(k)}\sqrt{-1})$$

$$= \frac{\tau^{(k)} J_{-1}(2\tau^{(k)}\sqrt{-1})}{J_0(2\tau^{(k)}\sqrt{-1})} = \frac{\tau^{(k)} I_1(2\tau^{(k)})}{I_0(2\tau^{(k)})} \ ,$$

with

$$t^{(k)} = t(S_0^{2(k)}, \theta^{(k)}, \sigma^{2(k)}) = \frac{S_0^{2(k)} \exp(2Z\theta^{(k)})}{2\sigma^{2(k)}}, \tau^{(k)} = \frac{\sqrt{X_i}}{2\sigma^{2(k)}} \exp(Z_i \theta^{(k)}) S_0^{(k)}.$$

Note that ${}_0F_1(1, \tau^2) = J_0(2\tau\sqrt{-1}) = I_0(2\tau)$, where $J_0(z)$ is the zero-order Bessel function of first kind, $I_0(z)$ is the zero-order modified Bessel function of first kind, which satisfies

$$J_{\nu}'(x) = J_{\nu-1}(x) - \frac{\nu}{x} J_{\nu}(x),$$

and

$$J_{-n}(x) = (-1)^n J_n(x), \qquad I_n(z) = i^{-n} J_n(zi).$$

In the M-step, we maximize the parameters of the augmented log-likelihood Q from Eq. (4) w.r.t $(\theta, \sigma^2, S_0^2)$. Omitting the items not depending on these parameters, Q can be expressed as

$$\sum_{i=1}^{m} \left(\log(S_0^2) - 2\log(\sigma^2) + 2Z_i\theta \right) \langle N_i \rangle^{(k)} - m\log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{m} \left(S_0^2 \exp(2Z_i\theta) + X_i \right).$$

(A.1)

It is easy to see in Eq. (A.1) that the log likelihood w.r.t σ^2 and S_0^2 are inverse Gamma and Gamma distributions, respectively. Hence, we update these two parameters by their modes:

$$\widehat{\sigma^2}_{ML} := \arg\max_{\sigma_g^2}(Q) = \frac{\sum_{i=1}^{m} (X_i + \exp(2\hat{\theta}Z_i)\hat{S_0}^2)}{2\sum_{i=1}^{m} (2\langle N_i \rangle + 1)}$$

(A.2)

24

and

$$\widehat{S^2_{0ML}} := \arg\max_{S_0^2}(Q) = \frac{2\widehat{\sigma^2}_{ML} \sum_{i=1}^m \langle N_i \rangle}{\sum_{i=1}^m \exp(2Z_i\hat{\theta})}. \tag{A.3}$$

To apply the Fisher scoring method, the score of θ is

$$\mathscr{S}(\theta) = 2\sum_{i=1}^m \langle N_i \rangle Z_i - \frac{\widehat{S^2_{0ML}}}{\widehat{\sigma^2}_{ML}} \sum_{i=1}^m \exp(2Z_i\theta)Z_i, \tag{A.4}$$

and the Fisher-information is given by

$$J(\theta) = E\left[-\frac{\partial^2 Q}{\partial \theta_h \partial \theta_k}\right] = \frac{\widehat{S^2_{0ML}}}{\widehat{\sigma^2}_{ML}} \sum_{i=1}^m \exp(2Z_i\theta)Z_iZ_i^T. \tag{A.5}$$

B. Maximization of Rician Log-likelihood

Without data agumentation, we have to directly maximize the Rician log-likelihood Q_{Rician}, in short Q_r thereafter, by using some typical MLE method, such as gradient descent. Then the first (the score) and second derivatives of Q_r are usually required. The log-likelihood Q_r is

$$Q_r = \text{const.} - m\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^m \left(Y_i^2 + \exp(2Z_i\theta)S_0^2\right)$$

$$+ \sum_{i=1}^m \log I_0\left(\frac{Y_i\exp(Z_i\theta)\sqrt{S_0^2}}{\sigma^2}\right),$$

where $I_k(\tau)$ are modified Bessel functions of first kind satisfying

$$I_0'(\tau) = I_1(\tau), \qquad I_0''(\tau) = I_1'(\tau) = (I_0(\tau) + I_2(\tau))/2.$$

The score of σ^2 and S_0^2 are respectively given by

$$\frac{\partial Q_r}{\partial \sigma^2} = -\frac{m}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^m \left(Y_i^2 + \exp(2Z_i\theta_i)S_0^2\right) \tag{B.1}$$

$$- \frac{1}{\sigma^4}\sum_{i=1}^m g\left(Y_i\exp(Z_i\theta)S_0\sigma^{-2}\right)Y_i\exp(Z_i\theta)S_0$$

and

$$\frac{\partial Q_r}{\partial S_0^2} = -\frac{1}{\sigma^2}\sum_{i=1}^m \exp(2Z_i\theta_i) + \frac{1}{2\sigma^2\sqrt{S_0^2}}\sum_{i=1}^m g\left(Y_i\exp(Z_i\theta)S_0\sigma^{-2}\right)Y_i\exp(Z_i\theta). \tag{B.2}$$

25

The score of θ is given by

$$\frac{\partial Q_r}{\partial \theta_k} = -\frac{S_0^2}{\sigma^2} \sum_{i=1}^{m} \exp(2Z_i\theta_i)Z_{ik} + \frac{1}{\sigma^2} \sum_{i=1}^{m} g\left(Y_i\exp(Z_i\theta)S_0\sigma^{-2}\right)Y_i\exp(Z_i\theta)S_0Z_{ik}.$$

(B.3)

The Hessian of θ is

$$\frac{\partial Q_r^2}{\partial \theta_h \partial \theta_k} = -\frac{2S_0^2}{\sigma^2}\sum_{i=1}^{m}\exp(2Z_i\theta_i)Z_{ih}Z_{ik} + \frac{S_0}{\sigma^2}\sum_{i=1}^{m}Y_i\exp(Z_i\theta)Z_{ik}Z_{ih}$$

$$\left\{g\left(Y_i\exp(Z_i\theta)S_0\sigma^{-2}\right) + g'\left(Y_i\exp(Z_i\theta)S_0\sigma^{-2}\right)\frac{Y_i\exp(Z_i\theta)S_0}{\sigma^2}\right\}$$

$$= \sum_{i=1}^{m} Z_{ih}Z_{ik}\left(-4t_i^2 + \tau_i(g(\tau_i) + \tau_i g'(\tau_i))\right) = \sum_{i=1}^{m} Z_{ih}Z_{ik}\left(-4t_i^2 + \tau_i^2 - \tau_i^2\left(\frac{I_1(\tau_i)}{I_0(\tau_i)}\right)^2\right).$$

where we use

$$t_i = \frac{S_0^2\exp(2Z_i\theta_i)}{2\sigma^2}, \qquad \tau_i = \frac{Y_i\exp(Z_i\theta)S_0}{2\sigma^2}, \qquad g(\tau) = \frac{d}{d\tau}\log I_0(\tau) = \frac{I_1(\tau)}{I_0(\tau)},$$

$$g'(\tau) = \frac{d^2}{d\tau^2}\log I_0(\tau) = \frac{1}{2}\left(1 + \frac{I_2(\tau)}{I_0(\tau)}\right) - \left(\frac{I_1(\tau)}{I_0(\tau)}\right)^2 = 1 - \frac{I_1(\tau)}{\tau I_0(\tau)} - \left(\frac{I_1(\tau)}{I_0(\tau)}\right)^2$$

with

$$I_2(\tau) = I_0(\tau) - \frac{2I_1(\tau)}{\tau}.$$

For SNR > 10, the corresponding Fisher-information matrix is approximated by

$$\mathscr{I}_r(\theta) = E\left[-\frac{\partial Q_r^2}{\partial \theta_h \partial \theta_k}\right] \approx \sum_{i=1}^{m} Z_{ih}Z_{ik}\left(\frac{S_0^2}{\sigma^2}\exp(2Z_i\theta) - \tfrac{1}{2}\right),$$

(B.4)

where (see[7])

$$E\left[\tau_i^2\left(\frac{I_1(\tau_i)}{I_0(\tau_i)}\right)^2\right] \approx \left(\frac{S_0^2}{\sigma^2}\exp(2Z_i\theta)\right)^2 + \frac{S_0^2}{\sigma^2}\exp(2Z_i\theta) - \frac{1}{2}.$$

C. *Theory of the EM algorithm by the phase data*

Consider the Rician noise model in Eq. (1), and define the phase

$$\varphi := \arg\left(S + \varepsilon_1 + i\varepsilon_2\right) \in [0, 2\pi)$$

such that

$$S + \varepsilon_1 = Y\cos(\varphi), \quad \varepsilon_2 = Y\sin(\varphi).$$

It follows from the Bayes formula that the joint density of φ and Y for fixed S and σ^2 is given by

$$
\begin{aligned}
p_{S,\sigma^2}(y,\varphi) &= \frac{y}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(Y\cos(\varphi) - S)^2 - \frac{1}{2\sigma^2}Y^2\sin(\varphi)^2 \right) \\
&= \frac{y}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(Y^2 + S^2 - 2SY\cos(\varphi)) \right) \\
&= p_{S,\sigma^2}(y)p_{S,\sigma^2}(\varphi|Y),
\end{aligned}
$$

or alternatively, similar formula can be found in [24] without using the Bayes theorem. Here the conditional density

$$
p_{S,\sigma^2}(\varphi|Y) = \frac{1}{2\pi I_0(SY/\sigma^2)} \exp\left(\frac{SY}{\sigma^2}\cos(\varphi) \right), \quad \varphi \in [0,2\pi), \tag{C.1}
$$

is an instance of the symmetric Von Mises distribution on the circle. See 4.3.2. in [25]. Note also that if the data $Y = 0$, we get we get a Gaussian likelihood

$$
p_{S,\sigma^2}(\varepsilon_r = -S, \varepsilon_i = 0) = \frac{y}{2\pi\sigma^2} \exp\left(-\frac{S^2}{2\sigma^2} \right),
$$

and in such a case the augmentation is not needed.

D. EM with latent phase measurements in multicompartment models

[1] introduces a related EM algorithm based on data augmentation with the complete complex-valued measurements $\mathbf{Y} = (Y_{ij} : 1 \le i \le N, 1 \le j \le M)$ for the individual diffusion compartments, and incomplete magnitude measurements $S_i = \left| \sum_j Y_{ij} \right|$. The E-step gives

$$
\begin{aligned}
Q(\Theta|\Theta^{(k)}) &= E\left[\ell(\Theta|Y)\big|S,\Theta^{(k)} \right] = \int \ell(\Theta|\mathbf{Y})p(\mathbf{Y}|S,\Theta^{(k)})d\mathbf{Y} \\
&= \text{const.} - (M+1)N\log(\sigma^2) + \frac{M+1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=0}^{M} E\left[2v_{ij}\Re(Y_{ij}) - |Y_{ij}|^2 - v_{ij}^2 \Big| S_i, \Theta^{(k)} \right]
\end{aligned}
$$

where $v = v(\Theta)$ and $\Re(z)$ denotes the real part of a complex z. Since

$$
\begin{aligned}
E\left[\Re(Y_{ij})\big|S_i,\Theta^{(k)} \right] &= E\left[E\left[\Re(Y_{ij})\big|\Re(Y_i) \right]\Big|S_i,\Theta^{(k)} \right] \\
&= v_{ij}^{(k)} + \frac{1}{M+1}E\left[\Re(Y_i)\big|S_i,\Theta^{(k)} \right] - \frac{v_i^{(k)}}{M+1}, \quad \text{and}
\end{aligned}
$$

$$
E\left[\Re(Y_i)\big|S_i,\Theta^{(k)} \right] = \frac{S_i I_1\left(S_i v_i^{(k)}/\sigma^2\right)}{I_0\left(S_i v_i^{(k)}/\sigma^2\right)},
$$

27

where $v^{(k)} = v(\Theta^{(k)})$, we obtain up to additive and multiplicative constants which do not depend on Θ, we obtain equation (7) in [1]:

$$Q(\Theta|\Theta^{(k)}) = \text{const.} + \text{const.} \sum_{i,j} \left\{ 2v_{ij} \left(\frac{S_i}{M+1} \frac{I_1\left(S_i v_i^{(k)}/\sigma^2\right)}{I_0\left(S_i v_i^{(k)}/\sigma^2\right)} - \frac{v_i^{(k)}}{M+1} + v_{ij}^{(k)} \right) - v_{ij}^2 \right\}.$$

(D.1)

In the M-step it is used the gradient of (D.1), given by

$$\frac{\partial Q(\Theta|\Theta^{(k)})}{\partial \theta} = \text{const.} \sum_{i,j} \left(\frac{S_i}{M+1} \frac{I_1\left(S_i v_i^{(k)}/\sigma^2\right)}{I_0\left(S_i v_i^{(k)}/\sigma^2\right)} - \frac{v_i^{(k)}}{M+1} + v_{ij}^{(k)} - v_{ij} \right) \frac{\partial v_{ij}(\Theta)}{\partial \theta}$$

We note that one could use simply the EM algorithm with for a single component $(M = 0)$ with latent data $(Y_i : i = 1, \ldots, n)$, optimizing in the M-step

$$Q(\Theta|\Theta^{(k)}) = \text{const.} + \text{const.} \sum_i \left\{ \frac{2v_i S_i I_1\left(S_i v_i^{(k)}/\sigma^2\right)}{I_0\left(S_i v_i^{(k)}/\sigma^2\right)} - v_i^2 \right\}$$

with gradient

$$\frac{\partial Q(\Theta|\Theta^{(k)})}{\partial \theta} = \text{const.} \sum_i \left\{ \frac{S_i I_1\left(S_i v_i^{(k)}/\sigma^2\right)}{I_0\left(S_i v_i^{(k)}/\sigma^2\right)} - v_i \right\} \frac{\partial v_i(\Theta)}{\partial \theta}$$

Since the phase augmentation under the single compartment model is quite similar in the computation by applying the proposed EM-MLE scheme, therefore, it is straightforward to extend our methods to the multiple compartment case.

References

[1] R. M. Henkelman, Measurement of signal intensities in the presence of noise in mr images, Medical physics 12 (2) (1985) 232–233.

[2] C. G. a. Koay, A signal transformational framework for breaking the noise floor and its applications in mri.

[3] H. Zhu, H. Zhang, J. G. Ibrahim, B. S. Peterson, Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data, Journal of the American Statistical Association 102 (480) (2007) 1085–1102.

[4] J. Rajan, B. Jeurissen, M. Verhoye, J. Van Audekerke, J. Sijbers, Maximum likelihood estimation-based denoising of magnetic resonance images using restricted local neighborhoods, Physics in medicine and biology 56 (16) (2011) 5221.

[5] J. Veraart, W. Van Hecke, J. Sijbers, Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a rician noise model, Magnetic Resonance in Medicine 66 (3) (2011) 678–686.

[6] J. L. Andersson, Maximum a posteriori estimation of diffusion tensor parameters using a rician noise model: why, how and but, Neuroimage 42 (4) (2008) 1340–1356.

[7] A. Jeffrey, D. Zwillinger, Table of integrals, series, and products, Academic Press, 2007.

[8] D. Gasbarra, J. Liu, J. Railavo, Data augmentation in rician noise model and bayesian diffusion tensor imaging, Preprint, 2014, p. arXiv:0935897.

[9] P. McCullagh, J. A. Nelder, Generalized linear models.

[10] A. Barmpoutis, M. S. Hwang, D. Howland, J. R. Forder, B. C. Vemuri, Regularized positive-definite fourth order tensor field estimation from dw-mri, NeuroImage 45 (1) (2009) S153–S162.

[11] K. Lange, Optimization, 2nd edition, Vol. 95, Springer Texts in Statistics, 2013.

[12] D. K. Jones, P. J. Basser, squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted mr data, Magnetic Resonance in Medicine 52 (5) (2004) 979–993.

[13] P. Barber, D. Darby, P. Desmond, Q. Yang, R. Gerraty, D. Jolley, G. Donnan, B. Tress, S. Davis, Prediction of stroke outcome with echoplanar perfusion-and diffusion-weighted mri, Neurology 51 (2) (1998) 418–426.

[14] E. T. Jaynes, Probability theory: the logic of science, Cambridge university press, 2003.

[15] G. Sparacino, C. Tombolato, C. Cobelli, Maximum-likelihood versus maximum a posteriori parameter estimation of physiological system models: the c-peptide impulse response case study, Biomedical Engineering, IEEE Transactions on 47 (6) (2000) 801–811.

[16] L. Griffanti, F. Baglio, G. M. Preti, P. Cecconi, M. Rovaris, G. Baselli, M. M. Lagana, Signal-to-noise ratio of diffusion weighted magnetic resonance imaging: Estimation methods and in vivo application to spinal cord, Biomedical Signal Processing and Control 7 (3) (2012) 285–294.

[17] A. Ghosh, T. Milne, R. Deriche, Constrained diffusion kurtosis imaging using ternary quartics & mle, Magnetic Resonance in Medicine 71 (4) (2014) 1581–1591.

[18] B. Landman, P.-L. Bazin, J. Prince, Diffusion tensor estimation by maximizing rician likelihood, IEEE, 2007, pp. 1–8.

[19] A. Leemans, B. Jeurissen, J. Sijbers, D. K. Jones, Exploredti: a graphical toolbox for processing, analyzing, and visualizing diffusion mr data, in: 17th Annual Meeting of Intl Soc Mag Reson Med, 2009, p. 3537.

[20] A. Barmpoutis, B. C. Vemuri, T. M. Shepherd, J. R. Forder, Tensor splines for interpolation and approximation of dt-mri with applications to segmentation of isolated rat hippocampi, Medical Imaging, IEEE Transactions on 26 (11) (2007) 1537–1546.

[21] V. Solo, J. J. Noh, An em algorithm for rician fmri activation detection, in: From Nano to Macro, ISBI 2007. 4th IEEE International Symposium on Biomedical Imaging, IEEE, 2007, pp. 464–467.

[22] X. Zhu, Y. Gur, W. Wang, P. Fletcher, Model selection and estimation of multi-compartment models in diffusion mri with a rician noise model, in: Information Processing in Medical Imaging, volume 7917 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 644–655.

375 [23] L. Qi, G. Yu, E. X. Wu, Higher order positive semidefinite diffusion tensor imaging, SIAM Journal on Imaging Sciences 3 (3) (2010) 416–433.

[24] C. G. Koay, P. J. Basser, Analytically exact correction scheme for signal extraction from noisy magnitude mr signals, Journal of Magnetic Resonance 179 (2) (2006) 317–322.

380 [25] N. I. Fisher, T. Lewis, B. J. Embleton, Statistical analysis of spherical data, Cambridge university press, 1987.