Comprehensive Maximum Likelihood Estimation of Diffusion Compartment Models Towards Reliable Mapping of Brain Microstructure

A. Stamm[1,2], O. Commowick[3], S.K. Warfield[2], and S. Vantini[1]

[1]MOX, Dept. of Mathematics, Politecnico di Milano, Italy
[2]CRL, Boston Children's Hospital, Harvard Medical School, MA, USA
[3]VISAGES, INSERM U746, CNRS UMR6074, INRIA, Univ. of Rennes I, France
Contact: `aymeric.stamm@polimi.it`

Abstract. Diffusion MRI is a key in-vivo non invasive imaging capability that can probe the microstructure of the brain. However, its limited resolution requires complex voxelwise generative models of the diffusion. Diffusion Compartment (DC) models divide the voxel into smaller compartments in which diffusion is homogeneous. We present a comprehensive framework for maximum likelihood estimation (MLE) of such models that jointly features ML estimators of (i) the baseline MR signal, (ii) the noise variance, (iii) compartment proportions, and (iv) diffusion-related parameters. ML estimators are key to providing reliable mapping of brain microstructure as they are asymptotically unbiased and of minimal variance. We compare our algorithm (which efficiently exploits analytical properties of MLE) to alternative implementations and a state-of-the-art strategy. Simulation results show that our approach offers the best reduction in computational burden while guaranteeing convergence of numerical estimators to the MLE. In-vivo results also reveal remarkably reliable microstructure mapping in areas as complex as the centrum semi-ovale. Our ML framework accommodates any DC model and is available freely for multi-tensor models as part of the ANIMA software[4].

1 Introduction

Diffusion MRI has raised a lot of interest over the past two decades as it provides an in-vivo non invasive mean for investigating the brain microstructure with great hopes of improved diagnosis, understanding and treatment of brain disorders. Current MR technologies however are limited in spatial resolution (finest resolution achieved so far in-vivo in diffusion is ~ 1.25 mm^3 [11]). Going further yields too long scans and too noisy data. Hence, brain microstructure is only accessible through careful modeling of the diffusion from which the MR signal arises. The coarser the resolution, the more heterogeneous the microstructure within the voxel and the more complex the voxelwise diffusion modeling.

Recently, there has been a growing interest in diffusion compartment models (DCM) [4, 6]. Their strength lies in their biological interpretability in that each

---

[4] `https://github.com/Inria-Visages/Anima-Public/wiki`

voxel compartment can be matched to an homogeneous biological substrate. Assuming Gaussian compartmental diffusion, the most complete microstructure mapping is given by the multi-tensor (MT) model [6], out of which several simplifications were devised in [4]. In essence, each compartment in the MT model is characterized by its diffusion tensor (DT). Current publicly available toolboxes for diffusion MRI processing often include multi-tensor ML estimation routines[5]. However, little attention has been paid to the actual numerical convergence to the ML estimate. We here propose a comprehensive maximum likelihood (ML) framework for the estimation of DCMs that aims at filling this gap. In effect, assuming identifiability of the unknown DCM, its ML estimator is guaranteed to be unbiased and of minimal variance as sample size increases. Our ML framework assumes that measurement error is modeled by white noise, which is fair for high signal-to-noise ratio (SNR) areas. Our ML framework is still valid for low SNR but a number of analytic properties of the likelihood used to improve time-efficiency do not hold anymore, which might lead to lengthier computations.

The contribution of this work is a comprehensive numerically convergent and massively fast framework for MLE of the MT model, provided that Gaussian homogeneous noise can be assumed. In Section 2.1, we briefly recall the definition of the MT model, propose a novel parametrization that features parameter-independent constraints, and formulate the derived maximization problem on a constrained domain. In Section 2.2, we provide analytic expressions of the ML estimators of the noise variance, the baseline MR signal and the compartment weights given the DTs and the number of compartments. In Section 2.3, we describe a time-efficient strategy for dealing with the constraints introduced in Section 2.1. In Section 2.4 we give the analytic expression of the log-likelihood depending on the unknown DTs exclusively and we derive the analytic expression of its Jacobian, which is key for time efficiency. Next, in Section 3, we present a simulation study whose goals are two-fold: (i) a mutual comparison of multiple algorithms for solving the maximization problem defined in Section 2.4 and (ii) a comparison with a reference estimation strategy [6]. We also describe our experiment on real data using a healthy subject from the Human Connectome Project (HCP) database [11]. Finally, Section 4 presents the results of both the simulation study and the experiment on real data and discuss our contribution.

2   ML Estimation of the Multi-Tensor Model

2.1   Description of the Problem

In the following, each voxel is considered independently. $\mathbf{y}$ is a set of N measured signals for pairs of b-values b_i and diffusion gradient directions $\mathbf{g}_i$, $\boldsymbol{\mu}$ is the set of corresponding signals generated by the DCM and τ^2 the noise inverse variance. **The Log-Likelihood.** The Gaussian log-likelihood function ℓ reads:

$$\ell(\boldsymbol{\mu}, \tau^2; \mathbf{y}) = \frac{N}{2}\ln\left(\frac{\tau^2}{2\pi}\right) - \frac{\tau^2}{2}\left\|\mathbf{y} - \boldsymbol{\mu}\right\|^2. \tag{1}$$

The MT Model. If the MT model with $C+1$ compartments holds, the MR signal for a b-value b_i and direction $\mathbf{g}_i$ can be written in matrix form as:

$$\boldsymbol{\mu}(S_0, \mathbf{w}, \mathbf{D}) = S_0(\mathbf{a}(D_0) + \varPhi(\mathbf{D})\mathbf{w}), \tag{2}$$

where S_0 is the baseline signal, $\mathbf{w} \in \mathbb{R}^C$ are the compartment weights, $\mathbf{D} \in \left[\mathcal{S}\left(\mathbb{R}^3\right)\right]^{C+1}$ are the DTs, $\mathbf{a}$ is an N-dimensional vector s.t. $a_i(D_j) = e^{-b_i \mathbf{g}_i^\top D_j \mathbf{g}_i}$ and $\varPhi$ is an $N \times C$ matrix s.t. $\varPhi_{ij} = a_i(D_j) - a_i(D_0)$.

Parametrization. We propose the following parametrization:
- S_0 is left as is with the constraint $S_0 \geq 0$ since MR signals are positive,
- τ^2 is left as is with the constraint $\tau^2 \geq 0$ since a variance is positive,
- $\mathbf{w}$ are fundamentally proportions, hence subject to $\mathbf{w} \geq \mathbf{0}$ and $\mathbf{1}^\top \mathbf{w} \leq 1$,
- $\mathbf{D}$ are 3D covariance matrices, i.e. positive definite symmetric matrices. Hence, we propose the following parametrization:

$$D_j = (d_{1j} + d_{2j})\mathbf{e}_{1j}\mathbf{e}_{1j}^\top + d_{2j}\mathbf{e}_{2j}\mathbf{e}_{2j}^\top + d_{3j}I_3, \tag{3}$$

where $\mathbf{e}_{kj}$ $(k = 1, 2, 3)$ are the 3 eigenvectors of D_j, parametrized by Euler angles $\theta_j \in [0, \pi]$, $\phi_j \in [0, 2\pi]$ and $\alpha_j \in [0, 2\pi]$ and respectively associated with the eigenvalues $\lambda_{1j} \geq \lambda_{2j} \geq \lambda_{3j} > 0$ such that $d_{1j} = \lambda_{1j} - \lambda_{2j} \geq 0$, $d_{2j} = \lambda_{2j} - \lambda_{3j} \geq 0$ and $d_{3j} = \lambda_{3j} > 0$.

Objective. To find the ML estimators of the parameters $\mathbf{w}$, $\mathbf{D}$, S_0, and τ^2, i.e., to maximize eq. (1) plugged-in with eq. (2) subject to the constraints above.

2.2 Complete Solution for a Known Number of Compartments

We propose to maximize eq. (1) in a stepwise fashion solving maximization sub-problems over a subset of the parameters in terms of the others:

Estimation of τ^2. Solving the partial derivative equation (PDE) in τ^2 yields:

$$\widehat{\tau^2}^{-1}(\mathbf{w}, \mathbf{D}, S_0) = N^{-1} \|\mathbf{y} - S_0\mathbf{a}(D_0) - S_0\varPhi(\mathbf{D})\mathbf{w}\|^2. \tag{4}$$

Since the right hand-side is always positive, eq. (4) defines the MLE of τ^2.

Estimation of $\mathbf{w}$. Plugging eqs. (2) and (4) into eq. (1) turns the problem into a least squares problem linear in the compartment weights. As in [1], we resort to the method of variable projection (VP) to get the weights estimates:

$$\widehat{\mathbf{w}}(\mathbf{D}, S_0) = \left[\varPhi^\top(\mathbf{D})\varPhi(\mathbf{D})\right]^{-1} \varPhi^\top(\mathbf{D})\left[S_0^{-1}\mathbf{y} - \mathbf{a}(D_0)\right], \tag{5}$$

In this section, the number of compartments is known which implies, in particular, that there are no pairs of compartments with equal diffusion distributions. Hence, the $C \times C$ matrix $\varPhi^\top\varPhi$ is invertible and eq. (5) is the MLE of $\mathbf{w}$, provided that it satisfies the constraints $\widehat{\mathbf{w}} \geq \mathbf{0}$ and $\mathbf{1}^\top \widehat{\mathbf{w}} \leq 1$.

Estimation of S_0. Plugging eqs. (2), (4) and (5) into eq. (1) and solving the PDE in S_0 yields:

$$\widehat{S_0}(\mathbf{D}) = \frac{< \mathbf{a}(D_0), P_\varPhi^\perp(\mathbf{D})\mathbf{y} >}{< \mathbf{a}(D_0), P_\varPhi^\perp(\mathbf{D})\mathbf{a}(D_0) >}, \tag{6}$$

where $P_\Phi^\perp = I_N - \Phi(\Phi^\top \Phi)^{-1}\Phi^\top$ is the $N \times N$ projector on the orthogonal complement of the column space of Φ and $<\cdot, \cdot>$ is the inner product in $\mathbb{R}^N$. Equation (6) is the MLE of S_0 which is automatically positive since $\mathbf{y} \geq \mathbf{0}$.

Estimation of D. Maximizing eq. (1) finally boils down to maximizing numerically the following log-likelihood function w.r.t. the DTs (obtained by plugging eqs. (2) and (4) to (6) into eq. (1)):

$$\ell(\mathbf{D}; \mathbf{y}) = -\frac{N}{2}\left[1 + \ln\left(\frac{2\pi}{N}\left(\left\| P_\Phi^\perp(\mathbf{D})\mathbf{y} \right\|^2 - \frac{<\mathbf{a}(D_0), P_\Phi^\perp(\mathbf{D})\mathbf{y}>}{\left\| P_\Phi^\perp(\mathbf{D})\mathbf{a}(D_0) \right\|^2} \right) \right) \right]. \quad (7)$$

At this point, it is worth making a few observations:

1. Equation (5) provides the MLE of $\mathbf{w}$ only if they lie inside the constrained domain of the original maximization problem, which is not guaranteed by the equations alone. This will be the object of section 2.3. In addition, eqs. (4) to (6) define the MLEs of τ^2, $\mathbf{w}$ and S_0 for a given set $\mathbf{D}$ of DTs. Yet, they are the solution of the original maximization problem if and only if we can find the MLE of $\mathbf{D}$, which requires a careful numerical maximization of eq. (7). This will be the object of section 2.4.
2. The VP method was already introduced in [1] as a powerful tool for DCM estimation. Our contribution is instead a comprehensive framework that provides the MLE of all the DCM parameters in a remarkably fast computation time, and automatically guaranteeing asymptotic unbiasedness and efficiency of the estimated DCMs. Estimators proposed in [1] are instead not the MLE of the corresponding parameters, which makes it difficult to guarantee unbiasedness and efficiency in all circumstances. In addition, notwithstanding S_0 and τ^2 are not diffusion parameters, their accurate knowledge is critical for a reliable estimation of the diffusion parameters. In [1], there are no guidelines pertaining to their estimation. In the present work instead, we compute their MLE analytically. In addition, we propose clear indications to deal with compartment weights constraints in our MLE framework.

2.3   Handling compartment weights constraints

Equation (4) always provides estimates of τ^2 within the constrained domain. This is not necessarily the case for eqs. (5) and (6). When the equations provide estimates that violate the constraints, we must search for the maximum on the boundary of the constrained domain. Our strategy relies on few key observations:

1. The sum-to-one constraint for compartment weights can be easily handled by expressing one of the weights as a linear combination of the other ones.
2. The positivity constraints on the remaining ones can be efficiently handled due to some analytical properties of the log-likelihood:
   – The symmetry properties of the log-likelihood function which is indeed invariant to the re-labeling of the fascicle compartments.
   – Along the domain boundary for $\mathbf{w}$ (i.e., when one or more compartment weights are constrained to 0) the log-likelihood function coincides with the log-likelihood of a reduced model with fewer compartments.

As a result, search on the boundary (when required) can be efficiently carried out by applying the strategy described in section 2.2 on a MT model with one or several compartments removed. This boundary search guarantees, in the end, compartment weights estimates always satisfying the constraints.

2.4 Solving for the MLE of the diffusion tensors

Our objective is to find a suitable optimization algorithm that features (i) short computation times and (ii) convergence to the maximum of the log-likelihood function defined in eq. (7). It is well-known that knowledge of the analytic Jacobian is of great help in achieving both. Let $x_k \in \{\theta_j, \phi_j, \alpha_j, d_{1j}, d_{2j}, d_{3j}\}_{j=0,...,C}$ be one of the parameters the MLE of which has to be obtained numerically. It is possible to analytically compute the derivative of eq. (7) w.r.t. x_k, which formally defines the analytic Jacobian. After some algebra, one can show that:

$$
\begin{aligned}
\partial_k \ell = -\frac{N}{2} & \left(\left\| P_{\Phi}^{\perp}(\mathbf{D})\mathbf{y} \right\|^2 - \frac{<\mathbf{a}(D_0), P_{\Phi}^{\perp}(\mathbf{D})\mathbf{y}>}{\left\| P_{\Phi}^{\perp}(\mathbf{D})\mathbf{a}(D_0) \right\|^2} \right)^{-1} \Big[<\mathbf{y}, \partial_k P_{\Phi}^{\perp}(\mathbf{D})\mathbf{y}> \\
& - 2\widehat{S_0} \left(<\partial_k \mathbf{a}(D_0), P_{\Phi}^{\perp}(\mathbf{D})\mathbf{y}> + <\mathbf{a}(D_0), \partial_k P_{\Phi}^{\perp}(\mathbf{D})\mathbf{y}> \right) \\
& + \widehat{S_0}^2 \left(<\partial_k \mathbf{a}(D_0), P_{\Phi}^{\perp}(\mathbf{D})\mathbf{a}(D_0)> + <\mathbf{a}(D_0), \partial_k P_{\Phi}^{\perp}(\mathbf{D})\mathbf{a}(D_0)> \right) \Big],
\end{aligned}
\tag{8}
$$

where ∂_k denotes the partial derivative operator w.r.t. x_k. Given the results presented in sections 3 and 4, we recommend the use of the Levenberg-Marquardt (LM) algorithm [3], which achieves convergence in remarkably short times. This algorithm solves unconstrained problems only. Since our parametrization makes x_k bound-constrained only, we can easily turn our constrained maximization problem into an unconstrained one by suitable mappings of x_k.

3   Material & Methods

The following sections pertain to the experimental study. The goals are two-fold: (i) a comparison of 4 algorithms for solving the constrained maximization problem defined in section 2.4 and (ii) a comparison with the estimation strategy used in [6], which is one of the reference papers on MT models. The former comparison is carried out in a simulation study while the latter one also in a case study on real data. All computations were performed with a Xeon 2.6 GHz processor, using 15 cores. Sections 3.1 and 3.2 details the simulation study and case study respectively, while section 3.3 provides a brief description of the compared algorithms/methods and section 3.4 depicts the evaluation metrics.

3.1   Simulation Study

We built an in-house diffusion phantom using subsets of the 6 compartments defined in Table 1 to generate purely isotropic areas (Area 0F) and areas with

Table 1. In-House Diffusion Phantom Compartments.

Compartment	Characteristics (10^{-3} mm^2/s for diffusivities)
Free Water (FW)	$\lambda_1 = \lambda_2 = \lambda_3 = 3.0$
Stationary Water (SW)	$\lambda_1 = \lambda_2 = \lambda_3 = 10^{-5}$
Isotropic Restricted Water (IRW)	$\lambda_1 = \lambda_2 = \lambda_3 = 1.0$
Circular Fascicle (CF)	Orient. $[0°, 360°]$, $\lambda_1 = 1.8$, $\lambda_2 = 0.3$, $\lambda_3 = 0.2$
Vertical Fascicle (VF)	Orient. $90°$, $\lambda_1 = 1.6$, $\lambda_2 = 0.5$, $\lambda_3 = 0.4$
Diagonal Fascicle (DF)	Orient. $-45°$, $\lambda_1 = 1.7$, $\lambda_2 = 0.2$, $\lambda_3 = 0.16$

1 (Area 1F), 2 (Area 2F) and 3 (Area 3F) fascicle compartments. The 3 isotropic compartments are always included. Then, we set spatially varying compartment weights within biologically feasible ranges. We set baseline signals in each area to realistic values using the mean baseline signals in grey matter, WM and cerebro-spinal fluid from case #153227 of the HCP database [11]. We added Gaussian noise as estimated in the same subject assuming a Rayleigh distribution of the background signals, which led to an average SNR ~ 23 dB in diffusion-weighted MR images. Finally, we used the HCP diffusion gradient table to generate $N = 288$ diffusion-weighted MR images. Comparisons based on the simulated phantom are made at known number of compartments to avoid confounding model selection errors. Isotropic diffusivities are assumed to be known.

3.2   Real Data

Ground truth microstructure parameters and the actual number of compartments in each voxel is missing for real data. Hence, we visually compare the estimated MT models in the corpus callosum (CC) and in the centrum semiovale (CSO, where decussation of 3 brain circuits happens), provided by our framework (method A1) and the strategy in [6], which is a reference on MT models (method B), both methods being tuned to run in the same amount of time. Model selection was performed by comparing models with 0, 1, 2 and 3 fascicle compartments using the unbiased AIC criterion [8]. We used the same HCP subject that helped building our phantom. The 3 isotropic compartments defined in section 3.1 were included in the model as well.

3.3   Compared Methods

In this section, we provide a brief description of the various methods compared in this experimental section. All methods of type A* correspond to algorithmic variants of our MLE framework while method B is a reference strategy.

Method A1. We use the Levenberg-Marquardt (LM) algorithm [3] with analytic Jacobian as implemented in ITK[6].

Method A2. We use the LM algorithm with numerical Jacobian.

Method A3. We use the globally convergent *conservative convex separable approximation* (CCSA) algorithm [10] implemented in the NLOpt library [2].

Method A4. We use the derivative-free *bounded optimization by quadratic approximations* (BOBYQA) algorithm [5] also implemented in NLOpt.

---

[6] http://www.itk.org

Method B. We use the strategy proposed in [6], with no spatial regularization, which performs brute-force estimation of all parameters using BOBYQA.

We did not include the estimation strategies proposed in [4, 1] in the comparison, since the former does not apply to DCMs with more than two compartments and the latter is difficultly reproducible due to the lack of documentation.

3.4   Evaluation Metrics

The comparison of the 5 strategies defined in Section 3.3 focuses on the trade-off between error w.r.t. the ground truth model and computation time. We naturally quantify this error by means of the MSE. For a given estimator $\widehat{\theta}$ of a parameter θ, its MSE is defined as $\mathbb{E}[d^2(\widehat{\theta}, \theta)]$, where d is a selected metric. Hereafter, we compute the MSEs of both the weights estimator $\widehat{\mathbf{w}}$ and the tensors estimator $\widehat{\mathbf{D}}$ separately in areas 0F, 1F, 2F and 3F of the phantom. We use the Euclidean distance for weights and the log-Euclidean distance for tensors. The MSEs are estimated as the averaged squared distances over the voxels of each area. This procedure led to 7 sets of performance curves for each area (4) and each estimator (2) that show MSE variations as a function of computation time.

4   Results & Discussion

Figure 1 shows the variations in MSE of both the weight and tensor estimators as defined in Section 3.4 as we let more time to the algorithm for estimating the MT models. Comparing method A1 (i.e. our framework) to the others at fixed MSE emphasizes – in the 4 simulated scenarios – multiple advantages of our proposed MLE framework: (i) there is an extra time cost in approximating the Jacobian (A1 vs A2), (ii) the LM algorithm is faster than other convergent gradient-based (GB) optimizers (A1 vs A3), (iii) GB optimization is faster than derivative-free (A1 vs A4) and (iv) our algorithm outperforms one of the reference approaches to MT model estimation (A1 vs B).

Next, Figure 2 shows the estimated MT models using both method A1 and method B in a fixed computation time of 30s (corresponding to 0.13s/voxel/core) for the crop on the CC and for the crop on the CSO. Visual inspection shows the ability of method A to provide more spatially coherent estimates of the MT model with less artifacts, mostly visible in three fascicle areas.

In summary, we have set up a novel and comprehensive framework for the efficient ML estimation of the MT model featuring massive reduction of computational burden. The framework generalizes to any DCM with known analytic Jacobian of the generative model. Future works will investigate its performances for estimating other important DCMs such as DIAMOND [7], DDI [9] or NODDI [12] and the possibility of embedding model selection in the framework.

References

1. Farooq, et al.: Brain Microstructure Mapping from diffusion MRI using Least Squares Variable Separation. CDMRI (MICCAI Workshop). pp. 1–9 (2015)
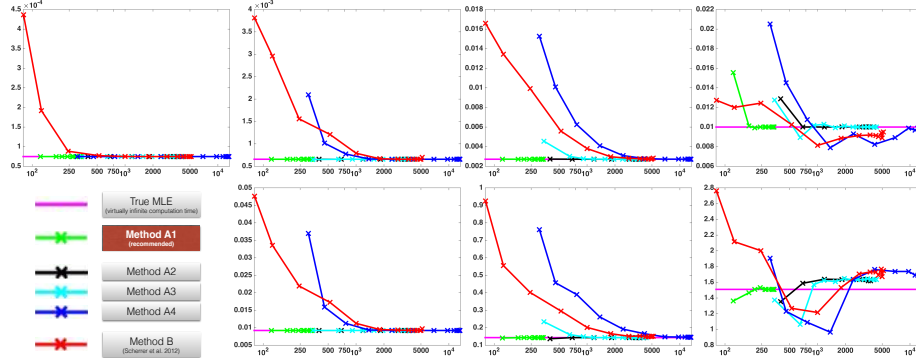
Fig. 1. Performance curves. MSE variations as a function of computation time (in sec). 1st row, $\widehat{\mathbf{w}}$; 2nd row, $\widehat{\mathbf{D}}$. Columns match areas 0F, 1F, 2F, 3F from left to right.
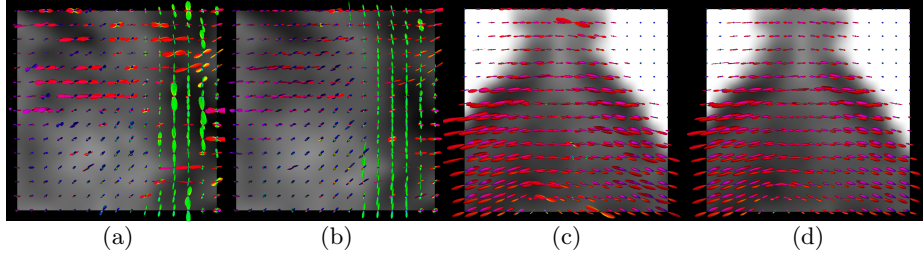

Fig. 2. Estimated Multi-Tensor Models In-Vivo. In the CSO (a,b), in the CC (c,d); using method B (a,c) or method A1 (b,d).

2. Johnson, S.: The NLOpt package. `http://ab-initio.mit.edu/nlopt`
3. Marquardt, D.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters. J. Soc. Ind. Appl. Math. 11(2), 431–441 (1963)
4. Panagiotaki, et al.: Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. Neuroimage 59(3), 2241–54 (2012)
5. Powell, M.: The BOBYQA algorithm for bound constrained optimization without derivatives. Tech. rep., University of Cambridge (2009)
6. Scherrer, B., Warfield, S.: Parametric representation of multiple white matter fascicles from cube and sphere diffusion MRI. PLoS One 7(11), e48232 (2012)
7. Scherrer et al.: Characterizing brain tissue by assessment of the distribution of anisotropic microstructural environments in DCI (DIAMOND). MRM (2015)
8. Stamm, et al.: Fast identification of optimal fascicle configurations from standard clinical diffusion MRI using Akaike information criterion. ISBI. pp. 238–41 (2014)
9. Stamm, A., Pérez, P., Barillot, C.: A new multi-fiber model for low angular resolution diffusion MRI. In: ISBI. pp. 936–939 (2012)
10. Svanberg, K.: A Class of Globally Convergent Optimization Methods Based on Conservative Convex Separable Approximations. SIAM J.O. 12(2), 555–573 (2002)
11. Van Essen, et al.: The WU-Minn Human Connectome Project: an overview. Neuroimage 80, 62–79 (2013)
12. Zhang, et al.: NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 61(4), 1000–16 (2012)