Multi-Tissue Decomposition of Diffusion MRI Signals via ℓ Sparse-Group Estimation

Pew-Thian Yap[‡,†] **[Senior Member, IEEE]**,

Department of Radiology and Biomedical Research Imaging Center (BRIC), University of North Carolina at Chapel Hill, NC, U.S.A

Yong Zhang[‡,†]**, and**

Department of Psychiatry & Behavioral Sciences, Stanford University, U.S.A

Dinggang Shen[‡] **[Senior Member, IEEE]**

Department of Radiology and Biomedical Research Imaging Center (BRIC), University of North Carolina at Chapel Hill, NC, U.S.A. Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

Abstract

Sparse estimation techniques are widely utilized in diffusion MRI (DMRI). In this paper, we present an algorithm for solving the ℓ sparse-group estimation problem and apply it to the tissue signal separation problem in DMRI. Our algorithm solves the ℓ problem directly, unlike existing approaches that often seek to solve its relaxed approximations. We include the mathematical proofs showing that the algorithm will converge to a solution satisfying the first-order optimality condition within a finite number of iterations. We apply this algorithm to DMRI data to tease apart signal contributions from white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) with the aim of improving the estimation of the fiber orientation distribution function (FODF). Unlike spherical deconvolution (SD) approaches that assume an invariant fiber response function (RF), our approach utilizes a response function group (RFG) to span the signal subspace of each tissue type, allowing greater flexibility in accounting for possible variations of the response function throughout space and within each voxel. Our ℓ algorithm allows for the natural groupings of the RFs to be considered during signal decomposition. Experimental results confirm that our method yields estimates of FODFs and volume fractions of tissue compartments with improved robustness and accuracy. Our ℓ algorithm is general and can be applied to sparse estimation problems beyond the scope of this paper.

Index Terms

Diffusion MRI; fiber orientation distribution function (FODF); sparse-group approximation; ℓ regularization

---

[‡]Corresponding authors: P.-T. Yap, Y. Zhang, and D. Shen.
[†]P.-T. Yap and Y. Zhang contributed equally to this work.

I. Introduction

Brain tissue changes such as demyelination and axonal loss are common pathological features of many neurological diseases. Precise quantification of these changes can improve the accuracy of disease diagnosis and is essential for monitoring treatment response and effective patient management. Diffusion magnetic resonance imaging (DMRI) [1] is an ideal imaging modality for this purpose owing to its unique ability to extract microstructural information by utilizing restricted and hindered diffusion to probe compartments that are much smaller than the voxel size.

Sparse estimation techniques are widely utilized in DMRI [2]–[5]. However, most techniques to date are predominantly based on ℓ_1 regularization. This choice is partly due to the challenges involved in solving ℓ_0-regularized problems, which are in nature non-convex, non-smooth, and discontinuous. While iterative hard thresholding algorithms [6]–[8] and greedy algorithms such as matching pursuit [9], [10], orthogonal matching pursuit [11], and subspace pursuit [12] have been developed for solving ℓ_0-regularized least-squares problems, how they can be extended for solving more general ℓ_0 minimization problems, such as those involving grouping [13], [14], is not immediately apparent. We present in this paper an algorithm for solving the ℓ_0 sparse-group problem, which is the ℓ_0 counterpart of the ℓ_1 sparse-group LASSO problem presented in [14]. We include the mathematical proofs to show that our algorithm converges to a solution that satisfies the first order optimality condition within a finite number of iterations.

We apply our algorithm to the DMRI tissue signal separation problem. In DMRI, each imaging voxel is in the order of 10 mm^3 in size and thus contains thousands of cells and tissue components. The diffusion of water molecules in each compartment (e.g., axons, dendrites, extracellular space, and cell soma) is affected by local viscosity, composition, geometry, and membrane permeability [15]. While the most commonly used diffusion tensor imaging (DTI) provides scalar indices such as fractional anisotropy (FA) and mean diffusivity (MD) for quantifying white matter integrity, they are limited by their non-specificity in identifying the exact causes of white matter changes [16], [17]. For example, the reduction in FA may be caused by any or a combination of the following factors that reduce the diffusion barriers: demyelination, fiber orientation variability, larger axon diameter, and lower packing density. For determining the more specific causes of changes, the signal contributions of the different white matter microstructural compartments need to be identified so that confounding information can be removed and information pertaining specifically to axonal injury and myelin damage can be teased out.

The general goal of diffusion compartment analysis is to identify signal contributions from constituent compartments and extract information that is most relevant to our problem at hand. Such approach increases specificity by removing confounding information. For example, it has been shown in [18] that by simply removing the diffusion signal that can be attributed to free diffusion [19], the sensitivity to the differences between stable and converting mild cognitive impairment (MCI) subjects can be significantly improved. In [20], the authors reported that by utilizing a model characterizing both intra- and extra-cellular diffusion, good estimations of neurite density and orientation dispersion can be obtained. It

has been reported in [21] that by proper consideration of signal contributions from gray matter and cerebrospinal fluid, false positive local orientations can be significantly reduced. All these studies point to the importance of compartmentalized analysis of DMRI data.

We employ our ℓ_0 algorithm to tease apart, at each voxel, the signal contributions of white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) so that the fiber orientation distribution function (FODF) can be estimated with greater accuracy. The concept of FODF [4], [22]–[25] was first introduced by Tournier and his colleagues in [22]. Assuming that all WM fiber bundles in the brain share identical diffusion characteristics, the diffusion-attenuated signal can be expressed as the spherical convolution of a fiber response function (RF), i.e., the signal profile of a single coherent fiber bundle, with the FODF. The FODF, which represents the distribution of fiber orientations within the voxel, can therefore be recovered using spherical deconvolution (SD) [22]. The SD technique proposed in [22] was later improved in [23] by introducing constrained spherical deconvolution (CSD) for better conditioning of the inverse problem, greater robustness to noise, and less susceptibility to negative artifacts that are physically impossible. Unlike the multi-tensor approach [26], SD-based methods do not require one to specify the number of tensors to fit to the data. A discrete version of the formulation, coupled with sparse regularization, was investigated in [2], [3], [27]. Tracing FODFs across space allows one to gauge connectivity between brain regions [28] and provides *in vivo* information on white matter pathways for neuroscience studies involving development, aging, and disorders [29]–[34].

It has been recently reported that a mismatch between the kernel used in CSD and the actual fiber RF can cause spurious peaks in the estimated FODF [35]. Although CSD has been recently extended in [21] to include RFs of not only the white matter (WM) but also the gray matter (GM) and the cerebrospinal fluid (CSF), these RFs are still spatially fixed and similar shortcomings as reported in [35] remain a problem.

In this paper, we propose to estimate the WM, GM, and CSF volume fractions and the WM FODF by using response function groups (RFGs). Each RFG is a collection of exemplar RFs catering to the variations of the actual RFs. Unlike the conventional approach of using fixed RFs, the utilization of RFGs will allow tolerance to RF variations and hence minimize estimation error due to the mismatch between the RF and the data. The FODF and compartmental volume fractions for each voxel are estimated by solving the ℓ_0 sparse-group problem that takes into account the natural groupings of the RFs. Our work is an integration of concepts presented in [21], [36], [37] with a novel ℓ_0 estimation framework.

II. Related Works

Recently, Daducci et al. [38] linearize the fitting problem of microstructure estimation techniques, such ActiveAx [39] and NODDI [20], to drastically speed up axon diameter and density estimation by a few orders of magnitude. They achieve this by solving an ℓ_1 minimization problem with a dictionary containing instances of the respective biophysical models generated with a discrete sampling of the diffusion parameters. Their work was recently extended to account for crossing fibers [40], [41]. Similar to [38], we observed that by using RFGs containing RFs of varying parameters, the data can be explained with greater

fidelity than by using RFs with a single fixed set of parameters. Dissimilar to [38], our work (i) solves a cardinality penalized problem instead of the ℓ_1 problem, and (ii) explicitly considers the natural coupling between RFGs via sparse-group estimation, which is similar but not identical to [14].

The use of ℓ_0 penalization is motivated by the observations reported in [42], where the authors show that the commonly used ℓ_1-norm penalization [27] conflicts with the unit sum requirement of the volume fractions and hence results in suboptimal solutions. To overcome this problem, the authors propose to employ a reweighted ℓ_1 minimization approached described by Candes et al. [43] to obtain solutions with enhanced sparsity, approximating solutions given by ℓ_0 minimization. However, despite giving improved results, this approach is still reliant on the suboptimal solution of the unweighted ℓ_1 minimization problem that has to be solved in the first iteration of the reweighted minimization scheme.

Another issue with ℓ_1 minimization is that it biases and attenuates coefficient magnitudes [44], resulting in the erosion of FODF peaks. While this can potentially be corrected by debiasing [44] — a post-processing approach where the coefficients are re-estimated without regularization using the support of the solution identified by ℓ_1 minimization — the desirable denoising effect of ℓ_1 penalization might be undone [44].

Besides ℓ_1-norm relaxation, other relaxation approaches have been proposed recently to substitute the ℓ_0-"norm"[1] by an ℓ_p-"norm" for some $p \in (0, 1)$ (see for example [45]–[47]). In general, these approaches do not necessarily give high-quality solutions. See [48] for examples showing that, when $p \in (0, 1]$, the solutions can deviate significantly from the true sparse solutions.

III. Contributions

In this work, we propose to directly minimize the ℓ_0 penalized problem instead of resorting to reweighted minimization. By doing so, we (i) overcome the suboptimality of reweighted minimization, and (ii) improve estimation speed greatly by avoiding solving the ℓ_1 minimization problem (especially the sparse-group problem [14]) multiple times to gradually improve sparsity. We will describe in this paper an algorithm based on iterative hard thresholding (IHT) [49] to effectively and efficiently solve the ℓ_0 penalized sparse-group problem.

The RFG framework affords the following advantages: (i) The exact RF does not need to be specified and can be determined automatically from the RFGs based on the data — this is akin to blind deconvolution for FODF estimation; (ii) The RF is allowed to vary across voxels as well as across fiber populations within each voxel; and (iii) The signal for each fiber population can be explained using a combination of RFs, making possible the modeling of signal with non-monoexponential decay.

---

[1]The quotation marks warn that ℓ_0-"norm" is not a proper norm.

Part of this work has been reported in our recently published conference paper [5]. Herein, we provide additional examples, results, derivations, and insights that are not part of the conference publication.

IV. Proposed Approach

In what follows, we will describe the concept of response function groups (RFGs), the estimation problem, the optimization framework, and some implementation issues.

A. Response Function Groups (RFGs)

DMRI acquires several diffusion-sensitized images, probing water diffusion in various directions and at various diffusion scales. At each voxel location, the diffusion-weighted signal $S(b, \hat{\mathbf{g}})$, measured for diffusion weighting b and at gradient direction $\hat{\mathbf{g}}$, can be represented as a mixture of N models:

$$S(b, \hat{\mathbf{g}}) = \sum_{i=1}^{N} f_i S_i(b, \hat{\mathbf{g}}) + \varepsilon(b, \hat{\mathbf{g}}), \quad (1)$$

where f_i is the volume fraction associated with the i-th model S_i and $\varepsilon(b, \hat{\mathbf{g}})$ is the fitting residual. A wide variety of microstructural models described in [50] can be employed to capture the diffusion patterns of intra- and extra-axonal diffusion compartments. In the current work, we are interested in distinguishing signal contributions from WM, GM, and CSF and we choose to use the tensor model $S_i(b, \hat{\mathbf{g}}) = S_i(0) \exp(-b\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{D}_i\hat{\mathbf{g}})$, where $S_i(0, \hat{\mathbf{g}}) = S_i(0)$ is the baseline signal with no diffusion weighting and $\mathbf{D}_i$ is a diffusion tensor. This model affords great flexibility in representing different compartments of the diffusion signal. Setting the diffusion tensor $\mathbf{D} = \lambda\mathbf{I}$, the model represents isotropic diffusion with diffusivity λ. When $\lambda = 0$ and $\lambda > 0$ the model corresponds to the dot model and the ball model, respectively [50]. Setting $\mathbf{D} = (\lambda_{\parallel} - \lambda_{\perp})\hat{\mathbf{v}}\hat{\mathbf{v}}^{\mathrm{T}} + \lambda_{\perp}\mathbf{I}$, $\lambda_{\parallel} > \lambda_{\perp}$, the model represents anisotropic diffusion in principal direction $\hat{\mathbf{v}}$ with diffusivity $\lambda_{\parallel}$ parallel to $\hat{\mathbf{v}}$ and diffusivity $\lambda_{\perp}$ perpendicular to $\hat{\mathbf{v}}$. When $\lambda_{\perp} = 0$ and $\lambda_{\perp} > 0$ the model corresponds to the stick model and the zeppelin model, respectively [50]. In our case, each RF is represented using the tensor model with a specific set of diffusion parameters.

Spherical deconvolution (SD) of the white matter (WM) diffusion-weighted signal with a fiber RF has been shown to yield high-quality estimates of FODFs [23]. According to [2], the solution to SD can be obtained in discretized form by including in the mixture model (1) a large number of anisotropic tensor models uniformly distributed on the unit sphere with fixed $\lambda_{\parallel}$ and $\lambda_{\perp}$ and solving for $\{f_i\}$ via minimizing the residual in the least-squares sense with some appropriate regularization, giving us the FODF. In this work, instead of fixing $\lambda_{\parallel}$ and $\lambda_{\perp}$, we allow each of them to vary across a range of values. Therefore, each principal direction is now represented by a group of RFs, i.e., a RFG. Additional isotropic RFGs are included to account for signal contributions from GM and CSF. Formally, the representation can be expressed as

$$S(b, \widehat{\mathbf{g}}) = \underbrace{\left[\sum_{i=1}^{N_{\text{WM}}} S_i^{\text{WM}}(b, \widehat{\mathbf{g}}) \right] + S^{\text{GM}}(b, \widehat{\mathbf{g}}) + S^{\text{CSF}}(b, \widehat{\mathbf{g}})}_{\widehat{S}(b, \widehat{\mathbf{g}})} + \varepsilon(b, \widehat{\mathbf{g}}), \quad (2)$$

where we have N_{WM} WM RFGs, a GM RFG, and a CSF RFG with

$$S_i^{\text{WM}}(b, \widehat{\mathbf{g}}) = \sum_{j=1}^{K_{\text{WM}}} f_{i,j}^{\text{WM}} S_{i,j}^{\text{WM}}(0) \exp(-b\widehat{\mathbf{g}}^{\text{T}} \mathbf{D}_{i,j}^{\text{WM}} \widehat{\mathbf{g}}),$$

$$S^{\text{GM}}(b, \widehat{\mathbf{g}}) = \sum_{j=1}^{K_{\text{GM}}} f_j^{\text{GM}} S_j^{\text{GM}}(0) \exp(-b\widehat{\mathbf{g}}^{\text{T}} \mathbf{D}_j^{\text{GM}} \widehat{\mathbf{g}}), \quad (3)$$

$$S^{\text{CSF}}(b, \widehat{\mathbf{g}}) = \sum_{j=1}^{K_{\text{CSF}}} f_j^{\text{CSF}} S_j^{\text{CSF}}(0) \exp(-b\widehat{\mathbf{g}}^{\text{T}} \mathbf{D}_j^{\text{CSF}} \widehat{\mathbf{g}}),$$

and

$$\mathbf{D}_{i,j}^{\text{WM}} = (\lambda_{\parallel,j}^{\text{WM}} - \lambda_{\perp,j}^{\text{WM}}) \widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i^{\text{T}} + \lambda_{\perp,j}^{\text{WM}} \mathbf{I},$$

$$\mathbf{D}_j^{\text{GM}} = \lambda_j^{\text{GM}} \mathbf{I}, \quad (4)$$

$$\mathbf{D}_j^{\text{CSF}} = \lambda_j^{\text{CSF}} \mathbf{I}.$$

Given the signal $S(b, \widehat{\mathbf{g}})$, we need to estimate the volume fractions $\{ f_{i,j}^{\text{WM}} \}$, $\{ f_j^{\text{GM}} \}$, and $\{ f_j^{\text{CSF}} \}$. If we take the Fourier transform of (2), it is easy to see that these volume fractions are in fact the weights that decompose the ensemble average propagator (EAP) of the overall signal into a weighted sum of the EAPs of the individual tensor models. The *overall* volume fractions associated with WM, GM, and CSF are respectively

$$f^{\text{MW}} = \sum_{i=1}^{N_{\text{WM}}} \underbrace{\sum_{j=1}^{K_{\text{WM}}} f_{i,j}^{\text{WM}}}_{f^{\text{WM}_i}}, \ f^{\text{GM}} = \sum_{j=1}^{K_{\text{GM}}} f_j^{\text{GM}}, \ f^{\text{CSF}} = \sum_{j=1}^{K_{\text{CSF}}} f_j^{\text{CSF}}. \quad (5)$$

The set $\{ f^{\text{WM}_i} | i = 1, \dots, N_{\text{WM}} \}$ gives the WM FODF. For the sake of notational simplicity, we group the volume fractions into a vector

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}^{\mathrm{WM}_1} \\ \vdots \\ \mathbf{f}^{\mathrm{WM}_{N_{\mathrm{MW}}}} \\ \mathbf{f}^{\mathrm{GM}} \\ \mathbf{f}^{\mathrm{CSF}} \end{bmatrix}, \quad (6)$$

where

$$\mathbf{f}^{\mathrm{WM}_i} = \left[f_{i,1}^{\mathrm{WM}}, \ldots, f_{i,K_{\mathrm{WM}}}^{\mathrm{WM}} \right]^{\mathrm{T}},$$

$$\mathbf{f}^{\mathrm{GM}} = \left[f_1^{\mathrm{GM}}, \ldots, f_{K_{\mathrm{GM}}}^{\mathrm{GM}} \right]^{\mathrm{T}}, \quad (7)$$

$$\mathbf{f}^{\mathrm{CSF}} = \left[f_1^{\mathrm{CSF}}, \ldots, f_{K_{\mathrm{CSF}}}^{\mathrm{CSF}} \right]^{\mathrm{T}}.$$

See Figure 1 for an illustration of how the signal can be represented using the RFGs. Note that N_{WM} is not the number of fiber populations, but is instead a sufficiently large number that results in a dense coverage of possible fiber directions. The fiber populations are unknown a priori and are automatically estimated by solving the sparse-group problem described next.

B. Estimation of Volume Fractions

To estimate the volume fractions, we solve the following optimization problem:

$$\min_{\mathbf{f} \geq 0} \left\{ \phi(\mathbf{f}) = \| S(b, \mathbf{g}) - \widehat{S}(b, \mathbf{g}) \|_2^2 + \alpha\gamma \|\mathbf{f}\|_0 + (1-\alpha)\gamma \quad (8) \right.$$

$$\left[\sum_{i=1}^{N_{\mathrm{WM}}} \mathscr{I}(\| \mathbf{f}^{\mathrm{WM}_i} \|_2) + \mathscr{I}(\| \mathbf{f}^{\mathrm{GM}} \|_2) + \mathscr{I}(\| \mathbf{f}^{\mathrm{CSF}} \|_2) \right] \right\},$$

where $\mathscr{I}(z)$ is an indicator function returning 1 if z 0 or 0 otherwise. The ℓ_0-"norm" gives the cardinality of the support, i.e., $\|\mathbf{f}\|_0 = |\mathrm{supp}(\mathbf{f})| = |\{k : f_k \quad 0\}|$. Parameters $\alpha \in [0, 1]$ and $\gamma > 0$ are for penalty tuning, analogous to those used in sparse-group LASSO [14]. Note that $\alpha = 1$ gives the ℓ_0 fit, whereas $\alpha = 0$ gives the group ℓ_0 fit. The problem can be written more succinctly in matrix form:

$$\min_{\mathbf{f} \geqslant 0} \left\{ \phi(\mathbf{f}) = \|\mathbf{A}\mathbf{f} - \mathbf{s}\|_2^2 + \alpha\gamma\|\mathbf{f}\|_0 + (1-\alpha)\gamma \sum_{g \in \mathcal{G}} \mathcal{I}(\|\mathbf{f}_g\|_2) \right\}, \quad (9)$$

where $\mathbf{f}_g$ denotes the subvector containing the elements associated with group $g \in \mathcal{G} = \{\text{WM}_1, ..., \text{WM}_{N_{\text{WM}}}, \text{GM}, \text{CSF}\}$. If the signal vector $\mathbf{s}$ is acquired at $P(b, \mathbf{g})$-points, $\mathbf{A} = [\mathbf{A}^{\text{WM}}|\mathbf{A}^{\text{GM}}|\mathbf{A}^{\text{CSF}}]$ is a $P \times N$ matrix ($N = N_{\text{WM}}K_{\text{WM}} + K_{\text{GM}} + K_{\text{CSF}}$) with columns containing all the individual WM, GM, and CSF tensor models sampled at the corresponding P points. We describe next our algorithm for the solution to this problem.

C. Optimization

The problem we are interested in solving has the following form:

$$\min_{\mathbf{f} \geqslant 0} \left\{ \phi(\mathbf{f}) = l(\mathbf{f}) + r(\mathbf{f}) \right\}. \quad (10)$$

In our case, the loss term $l(\mathbf{f}) = \|\mathbf{A}\mathbf{f} - \mathbf{s}\|_2^2$ is smooth, convex, and has a gradient that is Lipschitz continuous. The regularization term $r(\mathbf{f}) = \alpha\gamma\|\mathbf{f}\|_0 + (1-\alpha)\gamma\sum_{g\in\mathcal{G}}\mathcal{I}(\|\mathbf{f}_g\|_2)$ is non-convex, non-smooth, and discontinuous. Note that the solution is trivial, i.e., $\mathbf{f}^* = 0$, when $\gamma \geq \|\mathbf{s}\|_2^2$. We introduce here an algorithm called non-monotone iterative hard thresholding (NIHT), which we developed based on [6], [49], [51]–[53], to solve this problem. Proof of convergence is provided in the appendix.

1) Non-Monotone Iterative Hard Thresholding—The solution is outlined in Algorithm 1. The algorithm seeks the solution via gradient descent using a majorization-minimization (MM) [54] formulation of the problem. Step 1a minimizes the majorization of the objective function $\phi(\cdot)$ at $\mathbf{f}^{(k)}$. The minimization involves a gradient descent step with step size $1/L^{(k)}$ (more on this in the next section). The parameters $L_{\min}$ and $L_{\max}$ constrain the step size so that it is neither too aggressive nor too conservative (Step 3). We choose the initial step size $1/L_0^{(k)}$ as proposed in [55], using a diagonal matrix $1/L_0^{(k)}\mathbf{I}$ to approximate the inverse of the Hessian matrix of $l(\mathbf{f})$ at $\mathbf{f}^{(k)}$ (Step 3). A suitable step size is determined via backtracking line search, where the step size is progressively shrunk by a factor of $1/\tau$ (Step 1c). Parameter η ensures that the line search leads to a sufficient change in the value of the objective function. Since the problem is non-monotone, i.e., the value of the objective function is not guaranteed to decrease at every iteration, we require the value of the objective function to be slightly smaller than the largest value of the objective function in M previous iterations (Step 1b). For $M > 0$, the algorithm may increase the value of the objective function occasionally but will eventually converge faster than the monotone case with $M = 0$ [52]. Parameter ε controls the stopping condition (Step 2). We divide $|\phi(\mathbf{f}^{(k+1)}) - \phi(\mathbf{f}^{(k)})|$ by $\max(\phi(\mathbf{f}^{(k+1)}), 1)$ to compute the relative change or the absolute change of the value of the objective function, whichever is smaller. In this work, the following parameters were used: $L_{\min} = 1 \times 10^{-9}$, $L_{\max} = 1 \times 10^9$, $\eta = 1 \times 10^{-4}$, $\tau = 2$, $M = 10$, and $\varepsilon = 1 \times 10^{-6}$.

Algorithm 1

Non-Monotone Iterative Hard Thresholding (NIHT)

Data: Signal vector **s**; RF matrix **A**.

Parameters: Grouping $\mathcal{G}$; sparsity γ; balance parameter α.

Parameters: Factor $\tau > 1$; step size constants $L_{\min} < L_{\max}$; line search constant $\eta > 0$; stopping tolerance $\varepsilon > 0$; non-monotone parameter M.

Initialization: $L_0^{(0)} = 1$; initial solution $\mathbf{f}^{(0)}$.

Output: Volume fraction vector **f**.

/* Main Steps */

1 Set $L^{(k)} = L_0^{(k)}$.

  1a Solve subproblem (see Section IV-C2)

$$\mathbf{f}^{(k+1)} \in \underset{\mathbf{f} \geq 0}{\text{Arg min}} \left\{ l(\mathbf{f}^{(k)}) + \nabla l(\mathbf{f}^{(k)})^{\mathrm{T}}(\mathbf{f} - \mathbf{f}^{(k)}) + \frac{L^{(k)}}{2}\|\mathbf{f} - \mathbf{f}^{(k)}\|_2^2 + r(\mathbf{f}) \right\}.$$

$$(11)$$

  1b If

$$\phi(\mathbf{f}^{(k+1)}) \leq \max_{[k-M]^+ \leq i \leq k} \phi(\mathbf{f}^{(i)}) - \frac{\eta}{2}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2 \quad (12)$$

  is satisfied, then go to Step 2.

  1c Set $L^{(k)} \leftarrow L^{(k)}$ and go to Step 1a.

2 If

$$\frac{|\,\phi(\mathbf{f}^{(k+1)}) - \phi(\mathbf{f}^{(k)})\,|}{\max(\phi(\mathbf{f}^{(k+1)}), 1)} < \varepsilon$$

  is satisfied, then return $\mathbf{f}^{(k+1)}$ as a solution. Otherwise, go to Step 3.

3 Set

$$L_0^{(k+1)} = \max\left\{ L_{\min}, \ \min\left\{ L_{\max}, \frac{\Delta l^{\mathrm{T}}\Delta \mathbf{f}}{\|\Delta \mathbf{f}\|_2^2} \right\} \right\}, \quad (13)$$

  where $\Delta\mathbf{f} = \mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}$ and $\Delta l = \nabla l(\mathbf{f}^{(k+1)}) - \nabla l(\mathbf{f}^{(k)})$.

4 Set $k \leftarrow k + 1$ and go to Step 1.

---

2) Solution to Subproblem—The subproblem is group separable and can be solved by tackling separately the problem associated with each group $g \in \mathcal{G}$. With some algebra, the subproblem associated with group g can be shown to be

$$\mathbf{f}_g^{(k+1)} \in \underset{\mathbf{f}_g \geq 0}{\text{Arg min}} \left\{ \phi_g(\mathbf{f}_g) = \|\mathbf{f}_g - \mathbf{z}_g^{(k)}\|_2^2 + \frac{2}{L^{(k)}} r_g(\mathbf{f}_g) \right\}, \quad (14)$$

where

$$\begin{aligned} \mathbf{z}^{(k)} &= \mathbf{f}^{(k)} - \nabla l(\mathbf{f}^{(k)})/L^{(k)} \\ &= \mathbf{f}^{(k)} - 2\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{f}^{(k)} - \mathbf{s})/L^{(k)} \end{aligned} \quad (15)$$

and $\mathbf{f}_g$ and $\mathbf{z}_g^{(k)}$ are subvectors of $\mathbf{f}$ and $\mathbf{z}^{(k)}$ associated with group g. Note that (15) is a gradient descent with step size $1/L^{(k)}$. If we let $\tilde{\gamma}_1^{(k)} = 2\alpha\gamma/L^{(k)}$, and $\tilde{\gamma}_2^{(k)} = 2(1-\alpha)\gamma/L^{(k)}$, the solution to the problem can be obtained by hard thresholding [6]. That is, if $\tilde{\mathbf{z}}_g = \text{hard}_+(\mathbf{z}_g^{(k)}, \tilde{\gamma}_1^{(k)})$,

where

$$[\text{hard}_+(\mathbf{z}, \beta)]_i = \begin{cases} z_i & \text{if } z_i < \sqrt{\beta}, \\ 0 & \text{otherwise,} \end{cases} \quad (16)$$

the solution to the subproblem (14) is

$$\mathbf{f}_g^{(k+1)} = \begin{cases} \tilde{\mathbf{z}}_g & \text{if } \|\tilde{\mathbf{z}}_g\|_2^2 > \tilde{\gamma}_1^{(k)} \|\tilde{\mathbf{z}}_g\|_0 + \tilde{\gamma}_2^{(k)}, \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

See the appendix for the proof. Note that the sparse-group LASSO [14] can be implemented in a similar fashion by replacing the above solution with a soft-thresholding version.

D. Implementation Issues

1) Normalization—The columns of $\mathbf{A}$ and the signal vector $\mathbf{s}$ are normalized to have unit ℓ_2-norm before solving (10). After obtaining the solution vector, it is rescaled back to the original range.

2) Tuning Parameter—If we replace $r(\mathbf{f}) = \alpha\ \gamma\|\mathbf{f}\|_1 + (1-\alpha)\ \gamma\Sigma_{g\in\mathcal{G}}\|\mathbf{f}_g\|_2$ in (10), the thresholding operation in (16) has to be replaced by soft-thresholding at level $\beta/2$. According

to [56]–[58], in this case the parameter γ should be set according to the *universal penalty level*

$$\gamma = 2\sigma\sqrt{2 \log N}, \quad (18)$$

where σ is the noise standard deviation that can be measured from the background signal. To set the threshold in the case of hard-thresholding (see (16)) at the same level, we let

$$\gamma = 2\sigma^2 \log N. \quad (19)$$

3) Initialization—To speed up computation, we warm start the optimization using the solution provided by a subspace pursuit algorithm [12], modified to use ℓ_1 sparse projection [59].

V. Experimental Results

We compared the proposed method (**L200**) with the following methods:

- **L0** and **L1**: ℓ_0 and ℓ_1 minimization using a single RF each for WM, GM, and CSF [21]. The diffusion parameters of the WM, GM, and CSF RFs were determined respectively based on regions in the corpus callosum, cortex, and ventricles.

- **L211**: Sparse-group LASSO [14] using RFGs identical to the proposed method.

Similar to [42], and according to [43], we executed sparse-group LASSO multiple times for **L1** and **L211**, each time reweighing the ℓ_{21}-norm and the ℓ_1-norm so that they eventually approximate their ℓ_0 counterparts. The tuning parameter γ was set according to Section IV-D2. We set $\alpha = 0.5$ so that the effects for both group and individual sparsity are balanced. We will also report results of the above-mentioned methods with GM and CSF RFs removed, retaining only the WM RFs. In other words, single-tissue (ST) models are used. The ST variants are denoted respectively as **L200-ST**, **L211-ST**, **L0-ST**, and **L1-ST**.

The parameters of the RFGs were set to cover the typical values of the diffusivities of the WM, GM, and CSF voxels in the datasets described in the next sections: $\lambda_{\parallel}^{\mathrm{WM}} = 1 \times 10^{-3}$ mm²/s, $\lambda_{\perp}^{\mathrm{WM}} = [0.2:0.05:0.3] \times 10^{-3}$ mm²/s, $\lambda^{\mathrm{GM}} = [0.0:0.1:0.8] \times 10^{-3}$ mm²/s, and $\lambda^{\mathrm{CSF}} = [1.3:0.1:1.5] \times 10^{-3}$ mm²/s. The notation $[a:s:b]$ denotes values from a to b, inclusive, with step s. Note that in practice, these ranges do not have to be exact but should however cover the range of parameter variation. We set, in (3), $S_{i,j}^{\mathrm{WM}}(0) = S^{\mathrm{WM}}(0), S_{j}^{\mathrm{GM}}(0) = S^{\mathrm{GM}}(0)$, and $S_{j}^{\mathrm{CSF}}(0) = S^{\mathrm{CSF}}(0), \forall i, j$, according to the typical values of non-diffusion-weighted signals in the corpus callosum, cortex, and ventricles, respectively. The WM RFGs are distributed evenly on 321 points of a hemisphere, generated by subdivision of an icosahedron [60].

A. Synthetic Data

For quantitative evaluation, we generated a synthetic dataset using a mixture of four tensor models [26]. Two of which are anisotropic and represent two WM compartments that are at an angle of 45°, 60°, or 90° with each other. The other two are isotropic and represent the GM and CSF compartments. The generated diffusion-weighted signals therefore simulate the partial volume effects resulting from these compartments. The volume fractions and the diffusivities of the compartments were allowed to vary in ranges that mimic closely the real data discussed in the next section. The diffusion weightings and gradient directions were set according to the real data. Various levels of noise (SNR = 20, 30, 40, measured with respect to the signal value at $b = 0$ s/mm^2) was added. The performance statistics (see next paragraph) for the data generated using a combination of WM/GM/CSF diffusivities and volume fractions were computed. The statistics over a set of combinations of diffusivities and volume fractions were then averaged and reported. This procedure was repeated 100 times.

We evaluated the accuracy of the estimated volume fractions and FODFs based on the ground truth data. The accuracy of the estimated volume fractions was evaluated by computing the root mean square (RMS) error between the estimated volume fractions (5) and the ground truth volume fractions. The accuracy of the estimated FODFs [22] was evaluated by comparing their peaks (local maxima) with respect to the directions of the WM tensor models used to generate the synthetic data. The orientational discrepancy (OD) measure defined in [61] was used as a metric for evaluating the accuracy of peak estimates. The peaks were estimated based on the method described in [4].

The results shown in Figure 2 indicate that, compared with the ℓ_1 methods, the ℓ_0 methods improve the estimation accuracy of both volume fractions and FODFs for different SNRs and crossing angles. For the ST methods, it is not surprising that the estimated volume fractions are incorrect. More important, without proper modeling of the different tissue compartments, errors are also introduced in the FODFs; see for examples the OD values for 90° crossing angle.

Figure 3 shows the results for synthetic data generated using only the WM compartments. In this case, the estimates of the ST methods are improved. The accuracy of the volume fraction estimates of the multi-tissue methods are slightly worse than the single-tissue methods, but the errors remain reasonably low. The FODF estimates of the ℓ_0 methods are consistently better than the ℓ_1 methods.

B. Real Data

For reproducibility, diffusion-weighted imaging data from the Human Connectome Project (HCP) [62] were used. The $1.25 \times 1.25 \times 1.25$ mm^3 data were acquired with diffusion weightings $b = 1000, 2000, 3000$ s/mm^2, each applied in 90 non-collinear directions. 18 baseline images with low diffusion weighting $b = 5$ s/mm^2 were also acquired. All images were acquired with reversed phase encoding for correction of EPI distortion.

In this evaluation, we focus only on the multi-tissue methods. Figures 4 and 5 indicate that the WM, GM, and CSF volume fraction maps match quite well with the anatomy shown by

the T_1-weighted image. Figure 4 indicates that L200 gives the best structural clarity in regions such as the cerebellum. L211 and L1 over-estimate the GM volume fractions in WM regions. Figure 5 confirms that L200 gives the best structural clarity and that in the brainstem L200 and L0 yield volume fraction estimates that are more spatially consistent. L200 also captures greater amount of cortical GM details. Figure 6 shows the WM FODF glyphs at two regions that contain a mixture of signals from WM and GM—the cortex and the deep GM. The left two columns of the figure show the WM FODF glyphs of the cortical region. In this region, the tissue type progressively transits from WM to GM and then CSF. The results indicate that for pure WM region, the results are relatively consistent across methods. Differences start to show up in regions with GM partial volume. The right two columns of the figure show results in deep GM with WM and GM partial voluming. At the cortex, the ℓ_0 methods are able to extract more directional information than the ℓ_1 methods, as indicated by the larger glyphs that penetrate deeper into the GM. In the deep GM, the larger glyphs yielded by the ℓ_0 methods again confirm that they are able to extract a greater amount of directional information. The ℓ_0 methods can therefore provide more information for tractography algorithms in studies that need to investigate neuronal connections associated with the cortex and the deep GM.

C. Computation Speed

Our implementation of the ℓ_0 sparse-group solver is 8 times faster than the reweighted LASSO sparse-group solver implemented using a similar framework but by replacing hard thresholding with soft thresholding.

VI. DISCUSSION

Our method relies on a discrete dictionary consisting of RFs with parameters sampled discretely from the parameter space. When greater accuracy is needed, the parameter sampling density has to be increased. This increases the computational cost and when dictionary is large the computation can be intractable. This is especially the case when more sophisticated models, such as those used for microstructural estimation, need to be fitted to the data.

There are a number of remedies to this problem. First, we can use an iterative screening approach recently proposed in [59] to remove unrelated RFs during estimation. That is, during optimization of the ℓ_0 sparse-group problem, a subspace containing the solution can be estimated in each iteration and the minimization problem has to be solved only in this subspace, which is much smaller than the original space. This subspace is iteratively refined as better solution candidates are obtained. This approach can greatly reduce the computational load and at the same time promote robustness to local minima.

Alternatively, we can use a dictionary learning method to progressively refine the dictionary, as described in [63]. With such approach, the parameters of the RFs in the dictionary can be progressively refined in the continuous parameter space and hence obviates the need for including a large amount of RFs in the dictionary. Optimization is performed by alternating between solving the sparse problem and parameter adjustment.

Another approach is to use a Markov chain Monte Carlo (MCMC) based optimization technique, as reported in [64], [65]. This approach progressively narrows down the support of the solution with the goal of maximizing the posterior probability.

The estimation of the FODF can be further improved by imposing spatial regularity, similar to what was done in [66], [67]. The assumption is that the signals of neighboring voxels are highly correlated and, by encouraging spatially similar estimation outcomes, one can improve robustness to noise and yield greater estimation accuracy with a lesser number of samples [66], [67]. To account for larger changes at structural boundaries, regularization terms based on total variation (TV) [68] or wavelet [69] can be employed for edge preservation.

VII. CONCLUSION

We have shown that instead of restricting ourselves to one RF per compartment, it is possible to employ a group of RFs per compartment to cater to possible data variation across voxels. The use of ℓ_0 penalization is motivated by the following observations: (i) ℓ_1 penalization conflicts with the unit sum requirement of the volume fractions, as noted in [42]; (ii) ℓ_1 penalization is biased and attenuates coefficient magnitudes [52], resulting in the erosion of FODF peaks. Our results support that our ℓ_0 sparse-group estimation improves both volume fraction and FODF estimates.

Our method provides the flexibility of including different diffusion models in different groupings for robust microstructural estimation. Potential future work includes incorporating complex diffusion models [50] for estimation of subtle properties such as axonal diameter. We will also apply our method to the investigation of pathological conditions such as edema and also to applications such as tissue segmentation [70].

Acknowledgments

References

1. Johansen-Berg H, Behrens TE, editorsDiffusion MRI — From Quantitative Measurement to In vivo Neuroanatomy. Elsevier; 2009.

2. Jian B, Vemuri BC. A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE Transactions on Medical Imaging. Nov; 2007 26(11): 1464–1471. [PubMed: 18041262]

3. Landman BA, Bogovic JA, Wan H, ElShahaby FEZ, Bazin P-L, Prince JL. Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. NeuroImage. 2012; 59:2175–2186. [PubMed: 22019877]

4. Yap PT, Shen D. Spatial transformation of DWI data using nonnegative sparse representation. IEEE Transactions on Medical Imaging. 2012; 31(11):2035–2049. [PubMed: 22711770]

5. Yap P-T, Zhang Y, Shen D. Diffusion compartmentalization using response function groups with cardinality penalization. Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015; 9349:183–190. [PubMed: 30101230]

6. Blumensath T, Davies ME. Iterative thresholding for sparse approximations. Journal of Fourier Analysis and Applications. 2008; 14(5–6):629–654.

7. Blumensath T, Davies ME. Iterative hard thresholding for compressed sensing. Applied and Computational Harmonic Analysis. 2009; 27(3):265–274.

8. Blumensath T. Accelerated iterative hard thresholding. Signal Processing. 2012; 92:752–756.

9. Mallat SG, Zhang Z. Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing. 1993; 41(12):3397–3415.

10. Tropp JA. Greed is good: Algorithmic results for sparse approximation. IEEE Transactions on Information Theory. 2004; 50(10):2231–2242.

11. Cai TT, Wang L. Orthogonal matching pursuit for sparse signal recovery with noise. IEEE Transactions on Information Theory. 2011; 57(7):4680–2011.

12. Dai W, Milenkovic O. Subspace pursuit for compressive sensing signal reconstruction. IEEE Transactions on Information Theory. May; 2009 55(5):2230–2249.

13. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society. 2007; 68:49–67. vol. Series B.

14. Simon N, Friedman J, Hastie T, Tibshirani R. A sparse-group lasso. Journal of Computational and Graphical Statistics. 2013; 22(2):231–245.

15. Assaf Y, Basser PJ. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. NeuroImage. 2005; 27:48–58. [PubMed: 15979342]

16. Riffert TW, Schreiber J, Anwander A, Knösche TR. Beyond fractional anisotropy: Extraction of bundle-specific structural metrics from crossing fiber models. NeuroImage. 2014; 100:176–191. [PubMed: 24936681]

17. Dubois J, Dehaene-Lambertz G, Perrin M, Mangin J-F, Cointepas Y, Duchesnay E, Bihan DL, Hertz-Pannier L. Asynchrony of the early maturation of white matter bundles in healthy infants: Quantitative landmarks revealed noninvasively by diffusion tensor imaging. Human Brain Mapping. 2007; 29:14–27.

18. Fritzsche KH, Westin C-F, Meinzer H-P, Stieltjes B, Pasternak O. Free-water correction reveals wide spread differences between stable and converting MCI subjects. International Society for Magnetic Resonance in Medicine (ISMRM). 2013

19. Pasternak O, Shenton ME, Westin C-F. Estimation of extracellular volume from regularized multi-shell diffusion MRI. Medical Image Computing and Computer Assisted Intervention (MICCAI). 2012

20. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage. 2012; 61:1000–1016. [PubMed: 22484410]

21. Jeurissen B, Tournier J-D, Dhollander T, Connelly A, Sijbers J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. NeuroImage. 2014

22. Tournier JD, Calamante F, Gadian DG, Connelly A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage. 2004; 23(3):1176–1185. [PubMed: 15528117]

23. Tournier JD, Calamante F, Connelly A. Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. NeuroImage. 2007; 35(4):1459–1472. [PubMed: 17379540]

24. Cheng J, Deriche R, Jiang T, Shen D, Yap P-T. Non-negative spherical deconvolution (NNSD) for estimation of fiber orientation distribution function in single-/multi-shell diffusion MRI. NeuroImage. 2014; 101:750–764. [PubMed: 25108182]

25. Tran G, Shi Y. Fiber orientation and compartment parameter estimation from multi-shell diffusion imaging. IEEE Transactions on Medical Imaging. 2015

26. Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW, Wedeen VJ. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine. 2002; 48:577–582. [PubMed: 12353272]

27. Ramirez-Manzanares A, Rivera M, Vemuri BC, Carney P, Mareci T. Diffusion basis functions decomposition for estimating white matter intra-voxel fiber geometry. IEEE Transactions on Medical Imaging. 2007; 26(8):1091–1102. [PubMed: 17695129]

28. Yap PT, Wu G, Shen D. Human brain connectomics: Networks, techniques, and applications. IEEE Signal Processing Magazine. 2010; 27(4):131–134.

29. Jeurissen B, Leemans A, Jones DK, Tournier JD, Sijbers J. Probabilistic fiber tracking using the residual bootstrap with constrained spherical deconvolution. Human Brain Mapping. 2011; 32(3): 461–479. [PubMed: 21319270]

30. Yap PT, Fan Y, Chen Y, Gilmore J, Lin W, Shen D. Development trends of white matter connectivity in the first years of life. PLoS ONE. 2011; 6(9):e24678. [PubMed: 21966364]

31. Wee CY, Yap PT, Li W, Denny K, Brownyke J, Potter G, Welsh-Bohmer K, Wang L, Shen D. Enriched white-matter connectivity networks for accurate identification of MCI patients. NeuroImage. 2011; 54(3):1812–1822. [PubMed: 20970508]

32. Wee CY, Yap PT, Zhang D, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D. Identification of MCI individuals using structural and functional connectivity networks. NeuroImage. 2012; 59(3):2045–2056. [PubMed: 22019883]

33. Wee CY, Wang L, Shi F, Yap PT, Shen D. Diagnosis of autism spectrum disorders using regional and interregional morphological features. Human Brain Mapping. 2014; 35(7):3414–3430. [PubMed: 25050428]

34. Jin Y, Wee C-Y, Shi F, Thung K-H, Yap P-T, Shen D. Identification of infants at high-risk for autism spectrum disorder using multi-parameter multi-scale white matter connectivity networks. Human Brain Mapping. 2015

35. Parker G, Marshall D, Rosin P, Drage N, Richmond S, Jones D. A pitfall in the reconstruction of fibre ODFs using spherical deconvolution of diffusion MRI data. NeuroImage. 2013; 65:433–448. [PubMed: 23085109]

36. Wang Y, Wang Q, Haldar JP, Yeh F-C, Xie M, Sun P, Tu T-W, Trinkaus K, Klein RS, Cross AH, Song S-K. Quantification of increased cellularity during inflammatory demyelination. Brain. 2011; 134:3590–3601. [PubMed: 22171354]

37. White NS, Leergaard TB, D'Arceuil H, Bjaalie JG, Dale AM. Probing tissue microstructure with restriction spectrum imaging: Histological and theoretical validation. Human Brain Mapping. 2013; 34:327–346. [PubMed: 23169482]

38. Daducci A, Canales-Rodríguez EJ, Zhang H, Dyrby TB, Alexander DC, Thiran J-P. Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data. NeuroImage. 2015; 105:32–44. [PubMed: 25462697]

39. Alexander DC, Hubbard PL, Hall MG, Moore EA, Ptito M, Parker GJ, Dyrby TB. Orientationally invariant indices of axon diameter and density from diffusion MRI. NeuroImage. 2010; 52:1374–1389. [PubMed: 20580932]

40. Auría A, Romascano D, Canales-Rodriguez E, Wiaux Y, Dirby TB, Alexander D, Thiran J, Daducci A. Accelerated microstructure imaging via convex optimisation for regions with multiple fibers (AMICO$_X$). International Conference on Image Processing (ICIP). 2015:1673–1676.

41. Auría A, Canales-Rodriguez E, Wiaux Y, Dirby TB, Alexander D, Thiran J, Daducci A. Accelerated microstructure imaging via convex optimization (AMICO) in crossing fibers. The Annual Meeting of the International Society of Magnetic Resonance in Medicine (ISMRM). 2015

42. Daducci A, Ville DVD, Thiran J-P, Wiaux Y. Sparse regularization for fiber ODF reconstruction: From the suboptimality of ℓ_2 and ℓ_1 priors to ℓ_0. Medical Image Analysis. 2014; 18:820–833. [PubMed: 24593935]

43. Candès EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted ℓ_1 minimization. Journal of Fourier Analysis and Applications. 2008; 14(5):877–905.

44. Figueiredo M, Nowak R, Wright S. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE Journal of Selected Topics in Signal Processing. 2007; 1(4):586–598.

45. Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Processing Letters. 2007; 14(10):707–710.

46. Chen X, Xu F, Ye Y. Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. SIAM Journal on Scientific Computing. 2010; 32(5):2832–2852.

47. Chen X, Zhou W. Tech Rep. Department of Applied Mathematics, The Hong Kong Polytechnic University; 2010. Convergence of reweighted ℓ_1 minimization algorithms and unique solution of truncated ℓ_p minimization.

48. Zhang Y. PhD dissertation. Simon Fraser University; 2014. Optimization methods for sparse approximation.

49. Lu Z. Iterative hard thresholding methods for ℓ_0 regularized convex cone programming. Mathematical Programming. 2013:1–30.

50. Panagiotaki E, Schneider T, Siow B, Hall MG, Lythgoe MF, Alexander DC. Compartment models of the diffusion MR signal in brain white matter: A taxonomy and comparison. NeuroImage. 2012; 59:2241–2254. [PubMed: 22001791]

51. Birgin EG, Martínez JM, Raydan M. Nonmonotone spectral projected gradient methods on convex sets. SIAM Journal on Optimization. 2000; 10(4):1196–1211.

52. Wright SJ, Nowak RD, Figueiredo MAT. Sparse reconstruction by separable approximation. IEEE Transactions on Signal Processing. 2009; 57(7):2479–2493.

53. Gong P, Zhang C, Lu Z, Huang JZ, Ye J. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. International Conference on Machine Learning. 2013

54. Lange K. Optimization. 2. Springer; 2013. ser. Springer Texts in Statistics

55. Barzilai J, Borwein J. Two point step size gradient methods. IMA Journal of Numerical Analysis. 1988; 8:141–148.

56. Donoho DL. De-noising by soft-thresholding. IEEE Transactions on Information Theory. 1995; 41(3):613–627.

57. Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. SIAM Review. 2001; 43(1):129–159.

58. Zhang CH. Nearly unbiased variable selection under minimax concave penalty. The Annals of Statistics. 2010; 38(2):894–942.

59. Yap P-T, Zhang Y, Shen D. Iterative subspace screening for rapid sparse estimation of brain tissue microstructural properties. Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015; 9349:223–230. [PubMed: 30101231]

60. Leech J. [Online; accessed 28-June-2016] Approximating a sphere by recursive subdivision. http://www.neubert.net/Htmapp/SPHEmesh.htm

61. Yap P-T, Chen Y, An H, Yang Y, Gilmore J, Lin W, Shen D. SPHERE: SPherical Harmonic Elastic REgistration of hardi data. NeuroImage. 2011; 55(2):545–556. [PubMed: 21147231]

62. Essen DCV, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K. The WU-Minn human connectome project: An overview. NeuroImage. 2013; 80:62–79. [PubMed: 23684880]

63. Aranda R, Ramirez-Manzanares A, Rivera M. Sparse and adaptive diffusion dictionary (SADD) for recovering intra–voxel white matter structure. Medical Image Analysis. 2015; 26(1):243–255. [PubMed: 26519793]

64. Elad M, Yavneh I. A plurality of sparse representations is better than the sparsest one alone. IEEE Transactions on Information Theory. 2009; 55(10):4701–4714.

65. Crandall R, Dong B, Bilgin A. Randomized iterative hard thresholding: A fast approximate MMSE estimator for sparse approximations.

66. Goh A, Lenglet C, Thompson PM, Vidal R. Estimating orientation distribution functions with probability density constraints and spatial regularity. Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2009; 5761:877–885.

67. Michailovich O, Rathi Y, Dolui S. Spatially regularized compressed sensing for high angular resolution diffusion imaging. IEEE Transactions on Medical Imaging. 2011; 30(5)

68. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D. 1992; 60:259–268.

69. Zhang Y, Dong B, Lu Z. ℓ_0 minimization for wavelet frame based image restoration. Mathematics of Computation. 2013; 82:995–1015.

70. Yap P-T, Zhang Y, Shen D. Brain tissue segmentation based on diffusion mri using ℓ_0 sparse-group representation classification. Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015; 9351:132–139. [PubMed: 30035276]

71. Rockafellar RT, Wets RJ-B. Variational Analysis. Springer; 1998.

72. Iusem AN. On the convergence properties of the projected gradient method for convex optimization. Computational & Applied Mathematics. 2003; 22:37–52.

Biographies

Pew-Thian Yap is an Assistant Professor of Radiology and Biomedical Engineering at the University of North Carolina at Chapel Hill. He is also a faculty member of the Biomedical Research Imaging Center (BRIC). His research interest lies in diffusion MRI and its application to the characterization of brain white matter connectivity. He has made vital contributions to advancing innovative neurotechnologies for human brain mapping by developing novel algorithms for studying white matter pathways in the human brain. He has published more than 140 peer-reviewed papers in major international journals and conference proceedings. He is a senior member of IEEE.

Yong Zhang received his Ph.D. at Simon Fraser University in 2014. He then took up a postdoctoral research position at Stanford University. His major interests involve applying various optimization methods to solving problems arising in computer vision, machine learning, and medical image analysis. He has published more than 10 peer-reviewed papers in major international journals and conference proceedings.

Dinggang Shen is a Professor of Radiology, Biomedical Research Imaging Center (BRIC), Computer Science, and Biomedical Engineering at the University of North Carolina at Chapel Hill. He is currently directing the Center for Image Analysis and Informatics, the Image Display, Enhancement, and Analysis (IDEA) Lab in the Department of Radiology, and also the medical image analysis core in the BRIC. He was a tenure-track assistant professor at University of Pennsylvania and a faculty member at Johns Hopkins University. His research interests include medical image analysis, computer vision, and pattern recognition. He has published more than 700 papers in international journals and conference proceedings. He serves on the editorial boards of six international journals. He was on the Board of Directors of the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society from 2012 to 2015.

APPENDIX

We provide here a detailed convergence analysis of the non-monotone iterative hard thresholding (NIHT) algorithm. We show in Theorem 1 that the number of iterations for the inner loop is bounded and in Theorem 2 that NIHT converges to a local minimizer of (10).

Theorem 1

For each $k \geq 0$, the termination criterion of the inner loop (12) is satisfied after a finite number of iterations.

Proof

Since $\mathbf{f}^{(k+1)}$ is a minimizer of problem (14), we have

$$\nabla l(\mathbf{f}^{(k)})^{\mathrm{T}}(\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}) + \frac{L^{(k)}}{2}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2 + r(\mathbf{f}^{(k+1)}) \leq r(\mathbf{f}^{(k)}). \quad (20)$$

Since ∇l is Lipschitz continuous with constant L_l, we have (from Proposition 12.60 in [71])

$$l(\mathbf{f}^{(k+1)}) \leq l(\mathbf{f}^{(k)}) + \nabla l(\mathbf{f}^{(k)})^{\mathrm{T}}(\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}) + \frac{L_l}{2}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2. \quad (21)$$

Combining these two inequalities, we obtain

$$l(\mathbf{f}^{(k+1)}) + r(\mathbf{f}^{(k+1)}) \leq l(\mathbf{f}^{(k)}) + r(\mathbf{f}^{(k)}) - \frac{L^{(k)} - L_l}{2}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2 \quad (22)$$

and hence

$$\phi(\mathbf{f}^{(k+1)}) \leq \phi(\mathbf{f}^{(k)}) - \frac{L^{(k)} - L_l}{2} \|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2. \quad (23)$$

It follows that

$$\phi(\mathbf{f}^{(k+1)}) \leq \max_{[k-M]^+ \leq i \leq k} \phi(\mathbf{f}^{(i)}) - \frac{L^{(k)} - L_l}{2} \|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2. \quad (24)$$

Hence, (12) holds whenever $L^{(k)} \quad L_l + \eta$. This implies that $L^{(k)}$ needs to be updated only a finite number of times and hence the conclusion holds.

Theorem 2

Let $\{\mathbf{f}^{(k)}\}$ *be generated by NIHT when solving* (10). *Then these hold:*

    i.   *If the support set* $supp(\mathbf{f}) = \{i : f_i \quad 0\}$, *then* $supp(\mathbf{f}^{(k)})$ *changes only finitely often.*

    ii. $\mathbf{f}^{(k)}$ *converges to local minimizer* $\mathbf{f}^*$ *of problem* (10). *Moreover,* $supp(\mathbf{f}^{(k)}) \rightarrow$ $supp(\mathbf{f}^*)$, $r(\mathbf{f}^{(k)}) \rightarrow r(\mathbf{f}^*)$, $\phi(\mathbf{f}^{(k)}) \rightarrow \phi(\mathbf{f}^*)$, *and*

$$\mathbf{f}^* \in \underset{\mathbf{f} \geqslant 0}{Argmin} \ \{l(\mathbf{f}) : f_i \neq 0, i \in supp(\mathbf{f}^*)\}. \quad (25)$$

Proof

Let. $L^{(k)}$ denote the final value of $L^{(k)}$ at the kth iteration. From the proof of Theorem 1, we know that $\bar{L}^{(k)} \in [L_{\min}, \tilde{L})$, where $\tilde{L} = \tau(L_l + \eta)$. Let $\{\mathbf{f}^{(k)}\}$ be the series of solutions to (14), then for all $k \quad 0$,

$$|f_i^{(k+1)}| \geq \delta = \sqrt{2\alpha\gamma/\tilde{L}} > 0, \text{ if } f_i^{(k+1)} \neq 0. \quad (26)$$

Suppose $supp(\mathbf{f}^{(k)}) \quad supp(\mathbf{f}^{(k+1)})$ for some $k \quad 1$. Then some i exists such that ($f_i^{(k)} \neq 0$ but $f_i^{(k+1)} = 0$) or ($f_i^{(k)} = 0$ but $f_i^{(k+1)} \neq 0$), which together with (26) implies that $|f_i^{(k)}| \geq \delta$ or $|f_i^{(k+1)}| \geq \delta$ and we have

$$\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2 \geq |f_i^{(k+1)} - f_i^{(k)}| = \max(|f_i^{(k+1)}|, |f_i^{(k)}|) \geq \delta. \quad (27)$$

Since $\{\phi(\mathbf{f}^{(k)})\}$ is bounded below and $\phi(\mathbf{f}^{(k)}) - \phi(\mathbf{f}^{(k+1)}) \geq \frac{\eta}{2}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2$, $\{\phi(\mathbf{f}^{(k)})\}$ converges to a finite value as $k \rightarrow \infty$ and

$$\lim_{k \rightarrow \infty} \|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2 = 0. \quad (28)$$

This contradicts (27) and implies that $\text{supp}(\mathbf{f}^{(k)})$ does not change when k is sufficiently large.

(ii) It follows from (i) above that there exist some $K \geq 0$ such that $\text{supp}(\mathbf{f}^{(k)})$ stops changing and is fixed at a certain $\text{supp}^* \triangleq \text{supp}(\mathbf{f}^{(K)})$ for all $k \geq K$. Then one can observe from (14) that $\forall k > K$,

$$\mathbf{f}^{(k+1)} = \arg\min_{\mathbf{f} \geqslant 0} \left\{ l(\mathbf{f}^{(k)}) + \nabla l(\mathbf{f}^{(k)})^T(\mathbf{f} - \mathbf{f}^{(k)}) + \frac{\bar{L}^{(k)}}{2}\|\mathbf{f} - \mathbf{f}^{(k)}\|_2^2 : f_i \neq 0, i \in \text{supp}^* \right\}. \quad (29)$$

Using $\phi(\mathbf{f}^{(k)}) - \phi(\mathbf{f}^{(k+1)}) \geq \frac{\eta}{2}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|^2$ and the fact that $r(\mathbf{f}^{(k)}) = r(\mathbf{f}^{(k+1)})$ when $k > K$, we have

$$l(\mathbf{f}^{(k)}) - l(\mathbf{f}^{(k+1)}) \geq \frac{\eta}{2}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|^2. \quad (30)$$

It follows from Theorem 1 in [72] that $\mathbf{f}^{(k)} \rightarrow \mathbf{f}^*$, where

$$\mathbf{f}^* \in \operatorname*{Argmin}_{\mathbf{f} \geqslant 0} \left\{ l(\mathbf{f}) : f_i \neq 0, i \in \text{supp}^* \right\}. \quad (31)$$

Now consider a small perturbation vector $\Delta\mathbf{f}$. By using (31) we have

$$l(\mathbf{f}^*) \leq l(\widetilde{\mathbf{f}}^*), \quad (32)$$

where $\widetilde{f}_i^* = 0$ for $i \notin \text{supp}^*$ and $\widetilde{f}_i^* = f_i^* + \Delta f_i$ for $i \in \text{supp}^*$. In addition, based on the conclusion in part (i), we have $l(\widetilde{\mathbf{f}}^*) \leq l(\mathbf{f}^* + \Delta\mathbf{f})$ (otherwise, $\text{supp}(\mathbf{f}^{(k)})$ can change). Thus, $\mathbf{f}^*$ is a local minimizer of (10). In addition, we know from (26) that $f_i^{(k)} \geq \delta$ for $k > K$ and $i \in \text{supp}^*$. This yields $f_i^* \geq \delta$ for $i \in \text{supp}^*$ and $f_i^* = 0$ for $i \notin \text{supp}^*$. Hence, $\text{supp}(\mathbf{f}^{(k)}) = \text{supp}(\mathbf{f}^*) = \text{supp}^*$ for all $k > K$, which clearly implies that $r(\mathbf{f}^{(k)}) = r(\mathbf{f}^*)$ for every $k > K$. By continuity of l, we have $l(\mathbf{f}^{(k)}) \rightarrow l(\mathbf{f}^*)$. It then follows that

$$\phi(\mathbf{f}^{(k)}) = l(\mathbf{f}^{(k)}) + r(\mathbf{f}^{(k)}) \rightarrow l(\mathbf{f}^*) + r(\mathbf{f}^*) = \phi(\mathbf{f}^*). \quad (33)$$

Finally, (25) holds due to (31) and the relation $\mathrm{supp}(\mathbf{f}^*) = \mathrm{supp}^*$.

We show here the proof for the solution to subproblem (14).

Proof

The subproblem (14) can be written as

$$\mathbf{f}_g^{(k+1)} \in \underset{\mathbf{f}_g \geq 0}{\mathrm{Argmin}} \left\{ \phi_g(\mathbf{f}_g) = \|\mathbf{f}_g - \mathbf{z}_g^{(k)}\|_2^2 + \widetilde{\gamma}_1^{(k)}\|\mathbf{f}_g\|_0 + \widetilde{\gamma}_2^{(k)}\mathcal{I}(\|\mathbf{f}_g\|_2) \right\}. \quad (34)$$

Now let us consider the following two cases:

I. $\mathbf{f}_g = 0$. Then the cost is $\phi_g(0) = \|\mathbf{z}_g^{(k)}\|_2^2$.

II. $\mathbf{f}_g \neq 0$, i.e., at least one entry of $\mathbf{f}_g$ is nonzero. Then we obviously have $\widetilde{\gamma}_2^{(k)}\mathcal{I}(\|\mathbf{f}_g\|_2) = \widetilde{\gamma}_2^{(k)}$. According to [6], we have

$$\widetilde{\mathbf{z}}_g = \mathrm{hard}_+(\mathbf{z}_g^{(k)}, \widetilde{\gamma}_1^{(k)}) \in \underset{\mathbf{f}_g \geq 0}{\mathrm{Argmin}} \left\{ \phi_g(\mathbf{f}_g) = \|\mathbf{f}_g - \mathbf{z}_g^{(k)}\|_2^2 + \widetilde{\gamma}_1^{(k)}\|\mathbf{f}_g\|_0 \right\}, \quad (35)$$

The cost for this solution is $\|\widetilde{\mathbf{z}}_g - \mathbf{z}_g^{(k)}\|_2^2 + \widetilde{\gamma}_1^{(k)}\|\widetilde{\mathbf{z}}_g\|_0 + \widetilde{\gamma}_2^{(k)}$.
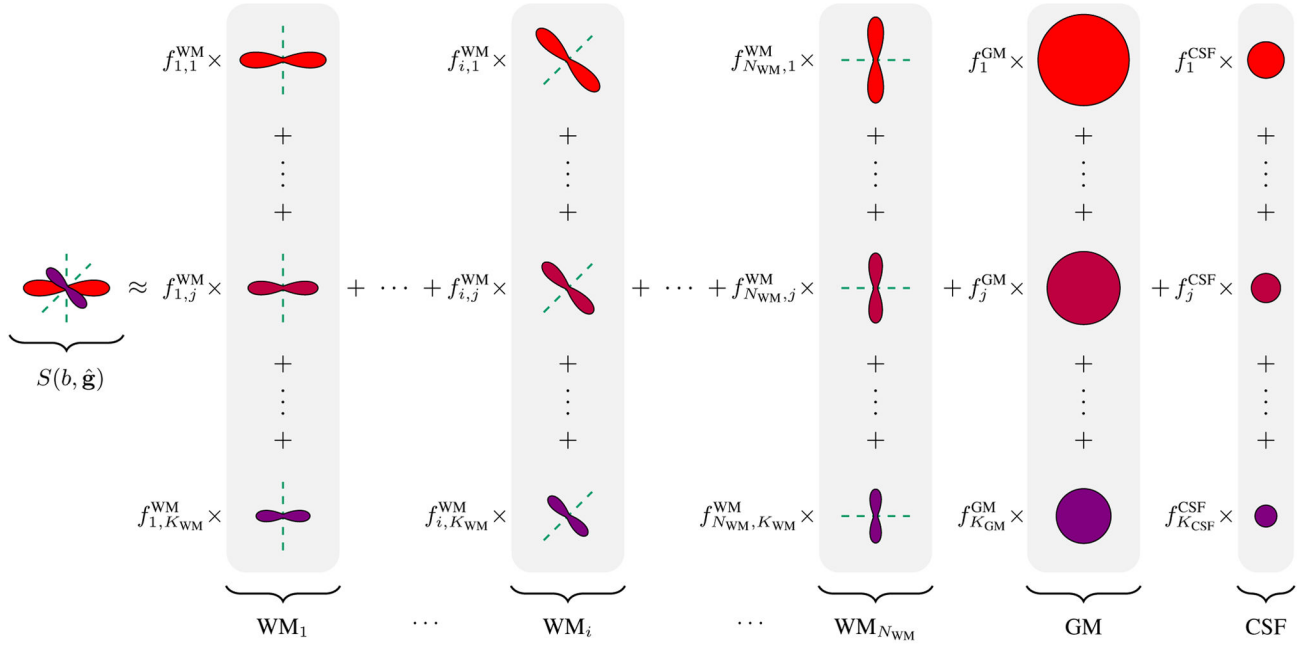
Comparing the costs of the two cases above, the solution for problem (14) can be obtained as (17). That is, the solution is non-zero when

$$\|\mathbf{z}_g^{(k)}\|_2^2 > \|\widetilde{\mathbf{z}}_g - \mathbf{z}_g^{(k)}\|_2^2 + \widetilde{\gamma}_1^{(k)}\|\widetilde{\mathbf{z}}\|_0 + \widetilde{\gamma}_2^{(k)} \quad (36)$$

or equivalently

$$\|\widetilde{\mathbf{z}}_g\|_2^2 > \widetilde{\gamma}_1^{(k)}\|\widetilde{\mathbf{z}}_g\|_0 + \widetilde{\gamma}_2^{(k)}. \quad (37)$$

In addition, from (9) we observe that if $\mathbf{f} = 0$, $\phi(0) = \|\mathbf{s}\|_2^2$. Thus, the optimal solution is $\mathbf{f}^* = 0$ when $\alpha\gamma + (1-\alpha)\gamma = \gamma \geq \|\mathbf{s}\|_2^2$.

**Fig. 1.**
The signal, $S(b, \hat{\mathbf{g}})$, can be represented using response function groups (RFGs) associated with white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). The volume fractions $\{ f_{i,j}^{\text{WM}} \}$, $\{ f_{j}^{\text{GM}} \}$, and $\{ f_{j}^{\text{CSF}} \}$ are estimated using a sparse-group estimation framework. The groupings are shaded in gray. Note that each white matter RFG WM_i is associated with a direction $\hat{\mathbf{v}}_i$. See (2), (3), and (4) for the mathematical details.
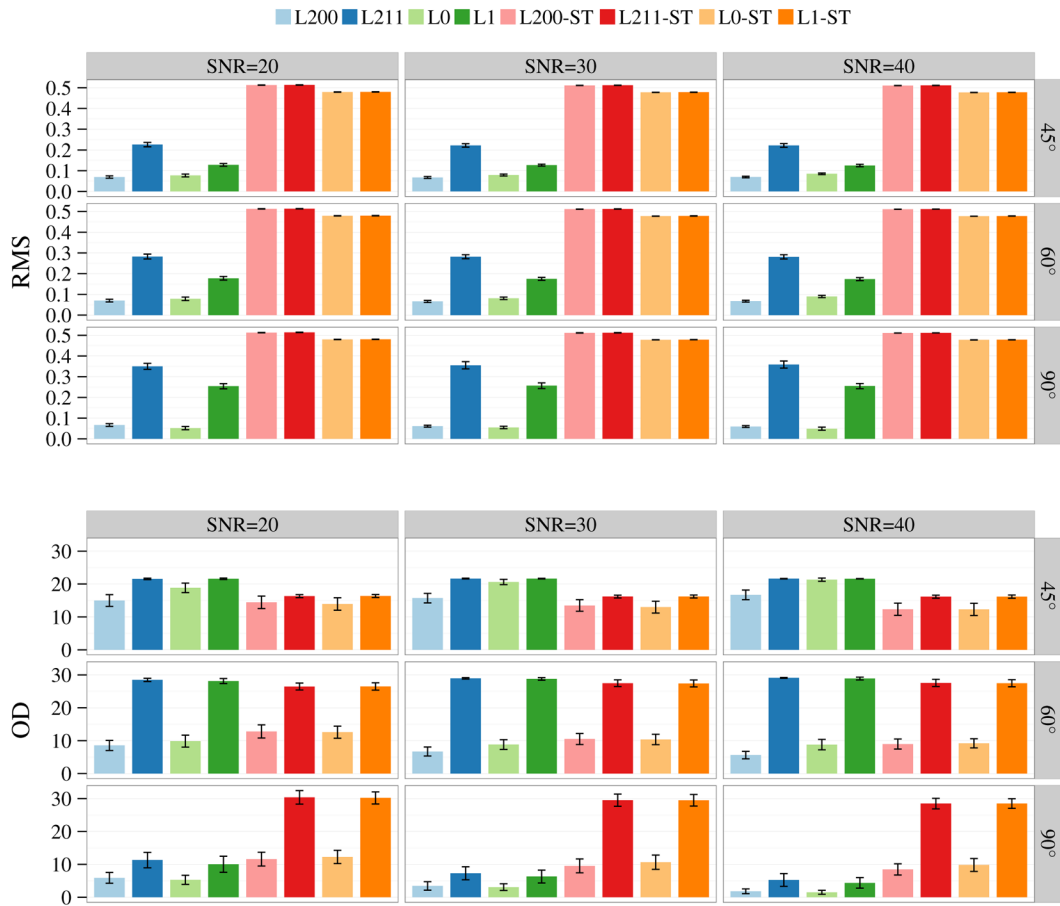
Fig. 2.
Estimation accuracy for volume fractions in terms of RMS error and FODFs in terms of OD. The bar plots show the mean performance statistics for SNR = 20, 30, 40 and crossing angle = 45°, 60°, 90°. The error bars indicate the standard deviations.

Fig. 3.
Estimation accuracy for volume fractions in terms of RMS error and FODFs in terms of OD.
Only the WM compartments were used to generate the data. The bar plots show the mean
performance statistics for SNR = 20, 30, 40 and crossing angle = 45°, 60°, 90°. The error
bars indicate the standard deviations.
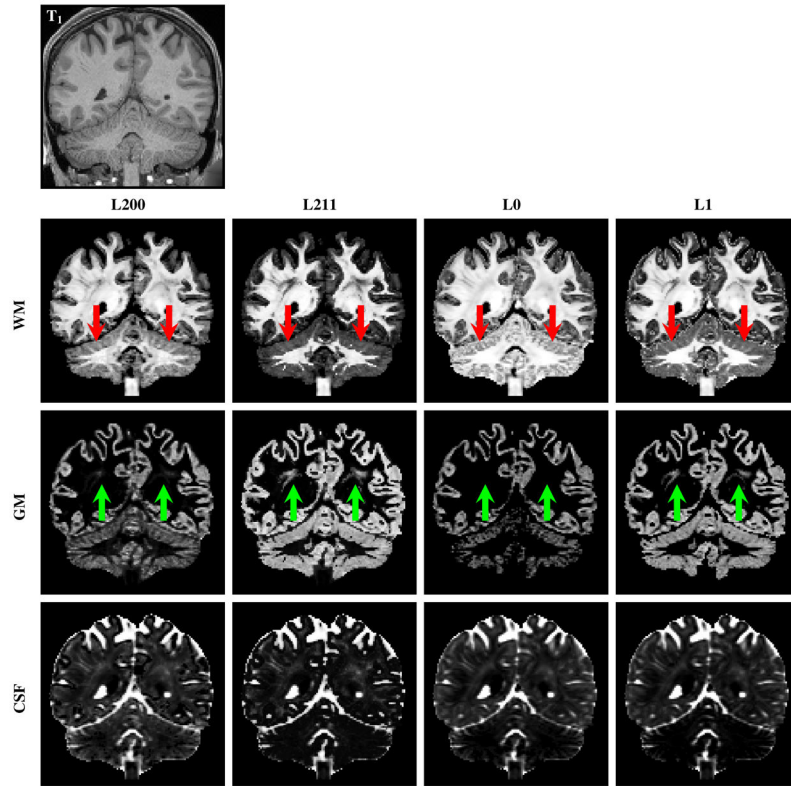
Fig. 4.
The WM, GM, and CSF volume fraction images obtained by L200, L211, L0, and L1. The T_1-weighted image at the top is shown for reference. L200 gives the best structural clarity in regions such as the cerebellum (red arrows). L211 and L1 over-estimate the GM volume fractions in WM regions (green arrows).
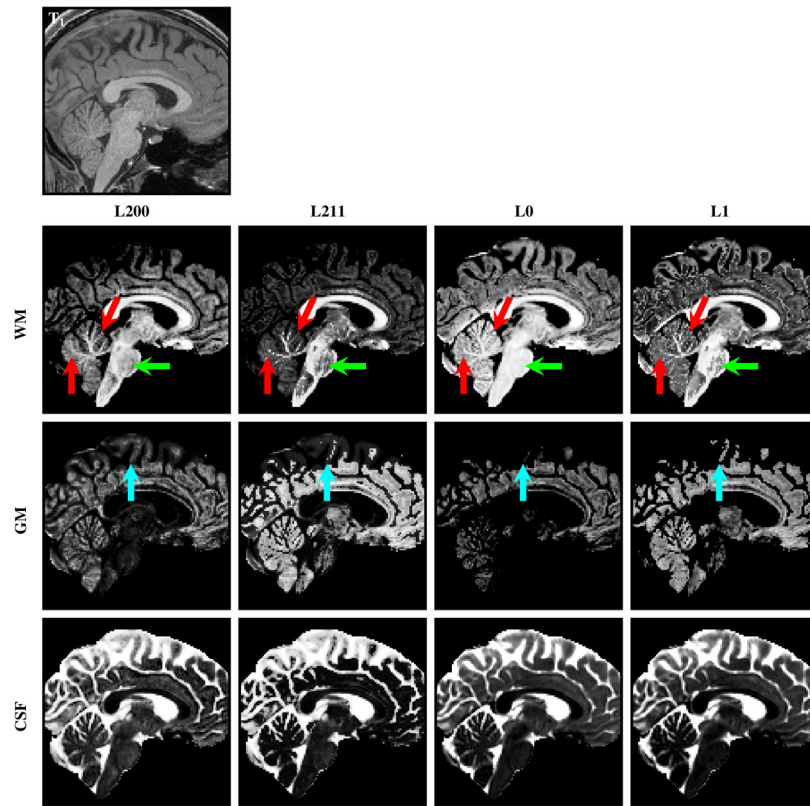
Fig. 5.
The WM, GM, and CSF volume fraction images obtained by L200, L211, L0, and L1. The T_1-weighted image at the top is shown for reference. L200 gives the best structural clarity in regions such as the cerebellum (red arrows). In the brainstem, L200 and L0 yield volume fraction estimates that are more spatially consistent (green arrows). L200 also captures greater amount of cortical GM details (cyan arrows).
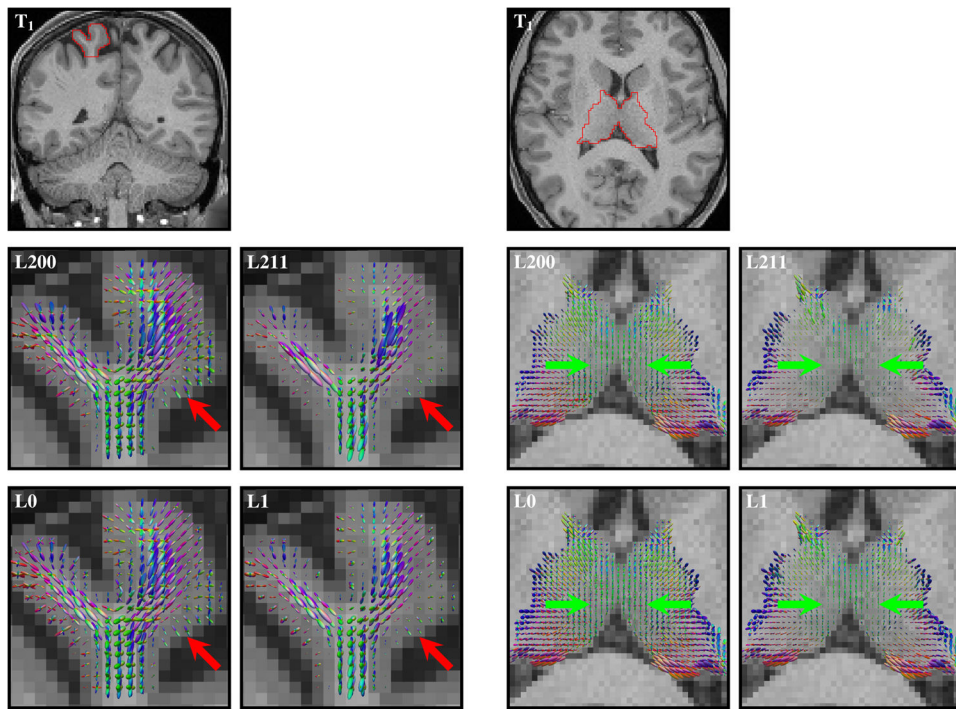
Fig. 6.
WM FODF glyphs, scaled according to the WM volume fractions, at (**left**) a cortical region and (**right**) a deep GM region. The red and green arrows mark regions where ℓ_0 methods yield a greater amount of WM directional information than the ℓ_1 methods.