NeuroImage 184 (2019) 140–160

Contents lists available at ScienceDirect

NeuroImage

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

Sparse wars: A survey and comparative study of spherical deconvolution algorithms for diffusion MRI

Erick Jorge Canales-Rodríguez [a,b,c,d,*], Jon Haitz Legarreta [e], Marco Pizzolato [b], Gaëtan Rensonnet [b,f], Gabriel Girard [b], Jonathan Rafael- Patino [b], Muhamed Barakovic [b], David Romascano [b], Yasser Alemán-Gómez [g,a,h], Joaquim Radua [c,d,i,j,l], Edith Pomarol-Clotet [c,d], Raymond Salvador [c,d], Jean-Philippe Thiran [a,b], Alessandro Daducci [k]

[a] Department of Radiology, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland
[b] Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[c] FIDMAG Germanes Hospitalàries, Sant Boi de Llobregat, Barcelona, Spain
[d] Mental Health Research Networking Center (CIBERSAM), Madrid, Spain
[e] Departamento de Física Matemática y Fluidos, Facultad de Ciencias, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain
[f] ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[g] Center for Psychiatric Neuroscience, Department of Psychiatry, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland
[h] Medical Image Analysis Laboratory (MIAL), Centre d'Imagerie BioMédicale (CIBM), Lausanne, Switzerland
[i] Centre for Psychiatric Research and Education, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[j] Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[k] Computer Science Department, University of Verona, Verona, Italy
[l] Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

ARTICLE INFO

Keywords:
Diffusion MRI
Spherical deconvolution
Sparse regression
LASSO
Non-negative least squares

ABSTRACT

Spherical deconvolution methods are widely used to estimate the brain's white-matter fiber orientations from diffusion MRI data. In this study, eight spherical deconvolution algorithms were implemented and evaluated. These included two model selection techniques based on the extended Bayesian information criterion (i.e., best subset selection and the least absolute shrinkage and selection operator), iteratively reweighted l_2- and l_1-norm approaches to approximate the l_0-norm, sparse Bayesian learning, Cauchy deconvolution, and two accelerated Richardson-Lucy algorithms. Results from our exhaustive evaluation show that there is no single optimal method for all different fiber configurations, suggesting that further studies should be conducted to find the optimal way of combining solutions from different methods. We found l_0-norm regularization algorithms to resolve more accurately fiber crossings with small inter-fiber angles. However, in voxels with very dominant fibers, algorithms promoting more sparsity are less accurate in detecting smaller fibers. In most cases, the best algorithm to reconstruct fiber crossings with two fibers did not perform optimally in voxels with one or three fibers. Therefore, simplified validation systems as employed in a number of previous studies, where only two fibers with similar volume fractions were tested, should be avoided as they provide incomplete information. Future studies proposing new reconstruction methods based on high angular resolution diffusion imaging data should validate their results by considering, at least, voxels with one, two, and three fibers, as well as voxels with dominant fibers and different diffusion anisotropies.

1. Introduction

Diffusion magnetic resonance imaging (MRI) is currently the main technique for studying the spatial organization of the brain's white matter non-invasively and *in vivo*. It is based on the application of strong magnetic field gradients to characterize the displacement of water molecules within intra- and extracellular spaces. The diffusion process in white matter is unequal along different orientations due to the spatial

* Corresponding author. Department of Radiology, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland.
E-mail address: erick.canalesrodriguez@epfl.ch (E.J. Canales-Rodríguez).

https://doi.org/10.1016/j.neuroimage.2018.08.071
Received 7 February 2018; Received in revised form 10 August 2018; Accepted 29 August 2018
Available online 5 September 2018

organization of axons. Specifically, the movement of water molecules is more restricted in directions perpendicular to axons than along their longitudinal axes. The diffusion MRI signal measured along different directions is used to determine, via inverse modeling, relevant properties of the imaged tissue.

A number of reconstruction methods have been proposed for characterizing the intra-voxel orientational heterogeneity using diffusion MRI data. These methods allow the detection of the principal orientations of fibers within a voxel, which can be used to reconstruct the main neuronal bundles of the brain by means of fiber tracking algorithms (Iturria-Medina et al., 2007). Among the different techniques, spherical deconvolution (SD) algorithms have received special attention due to their ability to resolve fiber crossings on data acquired within clinically feasible scan times (Alexander, 2005; Behrens et al., 2003; Canales-Rodríguez et al., 2015; Cheng et al., 2014; Daducci et al., 2014b; Dell'Acqua et al., 2010, 2007; Descoteaux et al., 2009; Jeurissen et al., 2014; Jian and Vemuri, 2007; Kaden et al., 2007; Kaden and Kruggel, 2012; Melie-García et al., 2008; Patel et al., 2010; Ramirez-Manzanares et al., 2007; Scherrer et al., 2016; Sotiropoulos et al., 2012; Tournier et al., 2007, 2004; Yeh and Tseng, 2013). Notably, as shown by (Jian and Vemuri, 2007) many of these reconstruction methods can be formulated in a linear unified deconvolution framework, where the coefficients of the Fiber Orientation Density (FOD) function are computed by solving a system of linear equations. This approach often leads to an under-determined system (i.e., if the number of diffusion measurements at each voxel is lower than the number of coefficients to be determined). In this situation, no algorithm can be expected to reliably find a solution because the system does not have a unique, well-defined solution. For instance, the standard least squares algorithm produces negative volume fractions (Jian and Vemuri, 2007). A common strategy to avoid such physically meaningless results is to add additional prior information to the system. In practice, this is done by minimizing a global cost function with two different terms: one term summarizes the discrepancy between the measured data and the values predicted by the model, and the other term is a regularization function that helps stabilizing the inversion and limiting the model complexity. As a result, the optimization algorithm searches for solutions that are compatible with the observed data while satisfying some additional requirements. Various choices for the regularization function exist in the literature, each one promoting solutions with different desirable properties. For instance, the solution (i.e., FOD) can be constrained to be non-negative (Jian and Vemuri, 2007). This choice is motivated by physical reasons: the FOD function expresses the probability to have a fiber oriented along any direction, or alternatively, it defines the volume fractions of all fibers, hence it cannot be negative. The optimization problem could also be complemented with other regularization functions aimed at reducing the complexity of the FOD. In fact, at the imaging resolution available nowadays it is commonly accepted that diffusion MRI is sensitive to detect only the major fiber bundles and that it can be used to disentangle up to 3 different fiber populations per voxel (Jeurissen et al., 2013; Schultz, 2012). Hence, the FOD can reasonably be considered sparse, either explicitly (i.e., by assuming that only a few elements are higher than zero) or implicitly (i.e., by assuming that the FOD is a smooth surface that possesses a low number of lobes) (Daducci et al., 2014b). The assumption of a smooth FOD is very useful in probabilistic tractography, where the smoothness is considered as a measure of uncertainty in the estimated fiber directions due to the experimental noise, or alternatively, as an index of fiber dispersion (Behrens et al., 2007). The assumption of a sparse FOD, on the other hand, is useful in applications using deterministic tractography, where the streamlines are reconstructed using the principal fiber directions (Wedeen et al., 2008). In the theory section, we will provide more details about the regularization functions used to promote such behaviors, but for now, we anticipate that they are often defined in terms of the l_p-norm, where a value $0 \leq p \leq 1$ promotes sparsity.

Since many of the proposed SD methods are based on different assumptions and reconstruction algorithms, various studies have been carried out to compare their performance. For instance, in their work (Jian and Vemuri, 2007) investigated different algorithms promoting sparsity and stability. Their results indicated that the non-negative least squares (NNLS) method (Lawson and Hanson, 1974) should be the choice for the multi-fiber reconstruction problem. However, in that early study, the other tested methods were implemented without the non-negativity constraint. Hence, the potential benefit of adding such a constraint to other regularized methods was not explored. In a study based on the results of the "HARDI reconstruction challenge 2012" (organized in the context of the "ISBI, 2012" conference), 20 algorithms for recovering the intra-voxel fiber structure were compared (Daducci et al., 2014a). That study contains the most extensive evaluation of local reconstruction methods published to date. It encompassed a mixture of classical techniques, including Q-Ball Imaging (QBI) (Tuch, 2004), Diffusion Spectrum Imaging (DSI) (Tuch, 2002; Wedeen et al., 2005) and Constrained Spherical Deconvolution (CSD) (Tournier et al., 2008), as well as alternative sparse SD methods inspired by compressed sensing theory (Candès et al., 2006; Donoho, 2006). Overall, the method DSI_{LR} (i.e., Deconvolved DSI) (Canales-Rodríguez et al., 2010) was the most accurate and stable. Nevertheless, non-negative sparse regression methods using l_1-norm regularization showed accurate reconstructions that were comparable to those obtained from DSI but using faster acquisition schemes with a lower number of measurements (Daducci et al., 2014a). Interestingly, the sparse methods outperformed the methods promoting smoothness. Various l_1-norm SD methods have been proposed over the past few years for the analysis of diffusion MRI data with reduced datasets (Aranda et al., 2015; Landman et al., 2012; Rathi et al., 2011; Tristán-Vega and Westin, 2011; Yeh and Tseng, 2013).

Recently, it was pointed out that current state-of-the-art sparse SD techniques based on l_2- and l_1-norms are suboptimal, and a better performance was reported by enforcing the solution to be more sparse by means of the l_0-norm (Daducci et al., 2014b). In line with this result, a new Sparse Bayesian Learning (SBL) algorithm (Pisharady et al., 2017a, 2017b) was proposed, which showed outstanding results in the same dataset employed in the HARDI reconstruction challenge 2012. The SBL formulation is strongly linked to the l_0-norm optimization problem. In fact, it was shown that under some conditions the global minimum of the SBL problem is only achieved at the minimum of the l_0-norm solution (Wipf and Nagarajan, 2008). The SBL cost function possesses fewer local minima than that of the l_0-norm, thus it has the potential to produce fewer convergence errors (Wipf and Rao, 2004). In contrast, the latest evaluation study (Mastropietro et al., 2017) found that in some cases the damped Richardson-Lucy method (Dell'Acqua et al., 2007), which yields non-sparse solutions, produced better results than the l_0-norm approach implemented in (Daducci et al., 2014b). This result seems to question the effectiveness of sparse methods and in particular the optimality of the l_0-norm. However, that study only examined a single sparse algorithm, and thus, results cannot be generalized to a broader context.

In the statistics and signal processing literature, many of the available sparse methods show heterogeneous performance depending on the type of dataset and application (Lyu et al., 2013; Schmidt et al., 2007; Zhang et al., 2015; Zhang and Rao, 2011). Thus a priori choosing a well-tailored method for diffusion MRI based only on theoretical considerations is difficult. Therefore, it is fundamental to evaluate various of these algorithms empirically in the more specific context of spherical deconvolution. Moreover, the validation approach employed in a number of studies has been mainly focused on testing voxels with two fiber populations having similar volume fractions. This is limiting since the best algorithm for reconstructing fiber crossings with two fibers might not be optimal in the case of a different number of fibers (Parker et al., 2013). A more comprehensive validation should account for voxels with different numbers of fibers and volume fractions.

In this work, we address the limitations of previous studies by implementing a number of sparse and non-sparse linear regression algorithms to solve the spherical deconvolution inverse problem, and by validating them across various synthetic datasets. In particular, we test

different levels of intra-voxel orientational heterogeneity by considering voxels with one, two, and three fiber populations, as well as fiber crossings with very dominant fibers. The implemented algorithms include two model selection techniques based on the extended Bayesian information criterion (i.e., best subset selection and the least absolute shrinkage and selection operator), iteratively reweighted l_2- and l_1-norm approaches to approximate the l_0-norm, sparse Bayesian learning, Cauchy deconvolution, and two accelerated Richardson-Lucy algorithms. All the algorithms were implemented using the same generative diffusion model and the same linear framework, thus results entirely depend on the chosen objective functions and optimization approaches. As a result of this work, we identify the optimal algorithms and provide a list of recommendations for future validation studies.

The remainder of the manuscript is organized as follows. An overview of the different theoretical concepts central to the study is provided in the Theory section. In the Evaluated Methods section, we provide details about the algorithms implemented in this study. In Material and Methods, we describe the metrics designed to assess the performance of the reconstructions as well as the computer simulations and datasets. Relevant findings are described in the Results section and finally pondered upon in the Discussion and Conclusions sections. Further technical details about the evaluated algorithms are provided in the Appendix section.

2. Theory

2.1. The deconvolution problem

A general expression for the intra-voxel diffusion MRI signal in the white matter including contributions from both the anisotropic and isotropic compartments is given by (Anderson, 2005; Dell'Acqua et al., 2007)

$$S_i(b_i, \hat{\mathbf{v}}_i) = S_0\left(\int FOD(\hat{\mathbf{r}}) H(\hat{\mathbf{r}}, b_i, \hat{\mathbf{v}}_i) d\hat{\mathbf{r}} + f_{CSF} \exp(-b_i D_{CSF}) \right), \quad (1)$$

where the integral is taken on the surface of the unit sphere; *FOD* denotes the Fiber Orientation Density function fully describing the intra-voxel orientation heterogeneity; H is the single-fiber response function, i.e., the signal originated in a compartment filled by a regular array of parallel fibers oriented along the unit-vector $\hat{\mathbf{r}}$; b_i is the experimental b-value and $\hat{\mathbf{v}}_i$ is a unit-vector denoting the orientation of the diffusion-sensitizing gradient; S_0 is the signal when $b_i = 0$; D_{CSF} is the self-diffusion coefficient of water in the cerebrospinal fluid (CSF) and f_{CSF} designates its volume fraction. The sum of all relative volumes obeys the equality

$$\int_{\Omega} FOD(\hat{\mathbf{r}}) d\hat{\mathbf{r}} + f_{CSF} = 1. \quad (2)$$

The problem of estimating *FOD* from the measured data is known as spherical deconvolution, an inverse problem that has spurred a large research effort over the past decades and is still being actively investigated.

2.2. Deconvolution on spherical grids

A very popular approach to solve the spherical deconvolution problem is to transfer the continuous *FOD* function into its discrete counterpart, such that this is evaluated only at the discrete points defining the spherical mesh $\Omega = \{\hat{\mathbf{r}}_j, j = 1, \ldots, \mathrm{m}\}$. Formally, this can be expressed as $FOD(\hat{r}) = \sum_{j=1}^{m} f_j \delta(\hat{r}_j - \hat{r})$, where m is the number of points, $\delta(\hat{\mathbf{r}}_j - \hat{\mathbf{r}})$ is a delta function defined to one if $\hat{\mathbf{r}} \equiv \hat{\mathbf{r}}_j$ and zero otherwise, thus $f_j = FOD(\hat{\mathbf{r}}_j)$. The main advantage of this simplification is that Eq. (1) can be linearized (Patel et al., 2010; Ramirez-Manzanares et al., 2007)

$$S_i(b_i, \hat{\mathbf{v}}_i) = S_0\left(\sum_{j=1}^{m} f_j H(\hat{\mathbf{r}}_j, b_i, \hat{\mathbf{v}}_i) + f_{CSF} \exp(-b_i D_{CSF}) \right). \quad (3)$$

The resulting system of equations can be written in matrix form as

$$\mathbf{s} = \mathbf{H}\mathbf{f}, \quad (4)$$

where $\mathbf{s}$ is a column-vector of length n containing the measured signals for different sampling parameters (i.e., b_i and $\hat{\mathbf{v}}_i$, $i = 1, \ldots, \mathrm{n}$); $\mathbf{H}$ is a $\mathrm{n} \times (\mathrm{m}+1)$ matrix whose first m columns contain the single-fiber responses corresponding to fibers oriented along the vectors $\hat{\mathbf{r}}_j$ while the last column contains the isotropic signal from the CSF component; $\mathbf{f}$ is a column-vector of length m+1 containing the m *FOD* values and the CSF weight f_{CSF} to be estimated.

The generative model that will be used in this study is the cylindrically symmetrical diffusion tensor model: $\mathbf{H}_{ij} = \exp\left(-b_i \hat{\mathbf{v}}_i^T \mathbf{D}_j \hat{\mathbf{v}}_i\right)$, where $\mathbf{D}_j$ denotes a diffusion tensor with the main eigenvector $\hat{\mathbf{e}}_1 = \hat{\mathbf{r}}_j$ and sorted diffusivities, i.e., $\lambda_1 > \lambda_2 = \lambda_3$. For a detailed list of other possible candidate models see (Ferizi et al., 2014; Panagiotaki et al., 2012).

2.3. A survey of sparse deconvolution

The solution to the inverse problem given by Eq. (4) is not straightforward because the resulting system of linear equations has more unknowns than measurements (i.e., $\mathrm{m} \gg \mathrm{n}$) and some of the columns of $\mathbf{H}$ are highly correlated. This ill-posed problem can lead to numerical instabilities and physically meaningless results (Jian and Vemuri, 2007). A common strategy to avoid such instabilities is to use algorithms that search for solutions compatible with the observed data but which also satisfy some additional constraints. For instance, Eq. (4) can be solved via non-negative least squares (Jian and Vemuri, 2007)

$$\tilde{\mathbf{f}} = \underset{\mathbf{f} \geq 0}{\operatorname{argmin}} \|\mathbf{s} - \mathbf{H}\mathbf{f}\|_2^2, \quad (5)$$

where $\| \cdot \|_2$ denotes the l_2-norm. In contrast to the standard least squares case, Eq. (5) has no analytical solution in closed form. Fortunately, there are efficient solvers to find the optimal solution. A popular choice is the Lawson and Hanson algorithm (Lawson and Hanson, 1974) which is very efficient for small-to-medium scale problems and when the solution is sparse. Notably, it has been established that non-negativity alone can be a powerful constraint in sparse recovery (Slawski and Hein, 2011). However, in noisy data, the algorithm tends to identify false components due to overfitting (Slawski and Hein, 2011).

A powerful approach to stabilize the inversion and avoid overfitting is to promote sparsity. The sparsest possible non-negative solution can be obtained by solving the following l_0-norm optimization problem

$$\operatorname{argmin} \|\mathbf{f}\|_0, \text{ subject to } \mathbf{s} = \mathbf{H}\mathbf{f} \text{ and } \mathbf{f} \geq 0, \quad (6)$$

where $\|\mathbf{f}\|_0$ refers to the number of non-zero elements in $\mathbf{f}$ (i.e., $\|\mathbf{f}\|_0 = \#\{j : \mathrm{f}_j \neq 0\}$). In order to take into account the noise in the measurements, this problem is usually replaced by

$$\tilde{\mathbf{f}} = \underset{\mathbf{f} \geq 0}{\operatorname{argmin}} \|\mathbf{s} - \mathbf{H}\mathbf{f}\|_2^2, \text{ subject to } \|\mathbf{f}\|_0 \leq k_{\max}, \quad (7)$$

which leads to the non-negative best-subset selection (BSS) in constrained form (Miller, 1984), where $k_{\max}$ is the maximum number of assumed components. The brute-force search method to solve this problem requires fitting the $\sum_{k=1}^{k_{\max}} p!/k!(p-k)!$ models containing $k = 1, 2, \ldots, k_{\max}$ predictors, where p is the number of elements in f. Then, for each value of k, the best model is the one having the smallest residual sum of squares. Finally, a single best model is identified using model selection techniques. In practical situations, the exhaustive combinatorial search of all possible subsets of variables is computationally

prohibitive. Consequently, various approximated algorithms have been proposed to efficiently solve this problem with high probability. They are based on smart approaches for sampling the solution space (Miller, 1984) or on brute-force search over a reduced set of variables, obtained after applying a dimension reduction step (Hong et al., 2014).

Alternatively, Eq. (7) can be written as a penalized least-square formulation

$$\tilde{\mathbf{f}} = \underset{\mathbf{f}\geq 0}{\operatorname{argmin}} \|\mathbf{s}-\mathbf{H}\mathbf{f}\|_2^2 + \lambda_0 \|\mathbf{f}\|_0, \tag{8}$$

where λ_0 is a non-negative regularization parameter. This form is convenient for simultaneous parameter estimation and variable selection. The main difficulty in directly minimizing Eq. (8) is related to the fact that the l_0-norm diversity measure is discrete, non-convex and non-differentiable. Thus, no closed-form analytic solution is available. Various approximated methods have been proposed to solve this problem, some of which are evaluated in this study. Two popular variants are iteratively reweighted l_1-norm (IRL1) and l_2-norm (IRL2) minimization. Before describing these methods, it is instructive to introduce the Least Absolute Shrinkage and Selection Operator (LASSO) based on the l_1-norm penalty,

$$\tilde{\mathbf{f}} = \underset{\mathbf{f}\geq 0}{\operatorname{argmin}} \|\mathbf{s}-\mathbf{H}\mathbf{f}\|_2^2 + \lambda_1 \|\mathbf{W}\mathbf{f}\|_1, \tag{9}$$

where $\mathbf{W} = diag(\mathbf{w})$ is a diagonal matrix with elements $\mathbf{W}_{jj} = \mathbf{w}_j$, and the term $\|\mathbf{W}\mathbf{f}\|_1 = \sum_j \left|\mathbf{w}_j\mathbf{f}_j\right|$. The original LASSO corresponds to the case where $\mathbf{W}$ is the identity matrix (Tibshirani, 2011), otherwise, it is termed weighted LASSO (Zou, 2006). Although this method does not produce the sparsest possible solution, in some cases it is equal to the solution of the l_0-norm. A strong point in favor of this formulation is that it is a convex optimization problem, which solution can be efficiently obtained by standard optimization tools (Bruckstein et al., 2009). Several algorithms have been proposed in the past decade to solve this problem, which may be grouped into different categories such as active-set methods, interior point methods, and entire regularization path methods. One appealing advantage of methods in the latter group is that they automatically consider multiple values of λ_1 and provide the related optimal solutions at a low computational cost (Friedman et al., 2010).

In the IRL1 minimization framework, the l_0-norm problem is approximated at iteration k+1 by solving the following weighted LASSO

$$\mathbf{f}^{(k+1)} = \underset{\mathbf{f}\geq 0}{\operatorname{argmin}} \|\mathbf{s}-\mathbf{H}\mathbf{f}\|_2^2 + \lambda \left\|\mathbf{W}^{(k)}\mathbf{f}\right\|_1, \tag{10}$$

where the weighting matrix in Eq. (10) takes the form $\mathbf{W}^{(k)} = \left(\mathbf{T}^{(k)}\right)^{-1}$, $\mathbf{T}^{(k)}$ is a diagonal matrix defined by $\mathbf{T}_{jj}^{(k)} = \mathbf{f}_j^{(k)} + \varepsilon$, and $\varepsilon > 0$ is a fixed application-dependent constant (Candès et al., 2008). The reader should note that this problem can be rewritten in the standard LASSO form by making the change of variables $\mathbf{q} = \left(\mathbf{T}^{(k)}\right)^{-1}\mathbf{f}$ and $\mathbf{X} = \mathbf{H}\mathbf{T}^{(k)}$, resulting in

$$\tilde{\mathbf{q}} = \underset{\mathbf{q}\geq 0}{\operatorname{argmin}} \|\mathbf{s}-\mathbf{X}\mathbf{q}\|_2^2 + \lambda \|\mathbf{q}\|_1. \tag{11}$$

This transformation is convenient since many LASSO solvers only implement the standard form. In summary, the IRL1 scheme allows us to approximate the l_0-norm non-convex problem given in Eq. (8) by means of a sequence of l_1-norm convex problems computed with LASSO.

Enforcing sparsity via l_p-norm regularization with $p \in [0,1]$ has gained much interest in the last few years. A state-of-the-art method for solving this problem is the regularized FOCal Underdetermined System Solver (FOCUSS) (Gorodnitsky and Rao, 1997), which minimizes the following cost function

$$\tilde{\mathbf{f}} = \operatorname{argmin} \|\mathbf{s}-\mathbf{H}\mathbf{f}\|_2^2 + \lambda_p \|\mathbf{f}\|_p^p, \tag{12}$$

where $\|\mathbf{f}\|_p^p = \sum_j \left|\mathbf{f}_j\right|^p$. Depending on the chosen value of p, this problem is somehow intermediate between the l_0-norm and l_1-norm. Equation (12) can be solved using Iterative Reweighted Least-Squares based on the l_2-norm (IRL2) (Rao et al., 2003), where the solution at iteration k+1 is computed by solving the following regularized least squares problem

$$\mathbf{f}^{(k+1)} = \operatorname{argmin} \|\mathbf{s}-\mathbf{H}\mathbf{f}\|_2^2 + \lambda_p \left\|\mathbf{T}^{(k)}\mathbf{f}\right\|_2^2, \tag{13}$$

where $\mathbf{T}^{(k)} = \left(\mathbf{W}^{(k)}\right)^{-1}$, $\mathbf{W}^{(k)}$ is a diagonal matrix defined by $\mathbf{W}_{jj}^{(k)} = \left|\mathbf{f}_j^{(k)}\right|^{1-p/2} + \varepsilon$ and ε is a small non-negative constant that provides stability to the solution. The main advantage of this approximation is that the new problem has an analytical solution in the form

$$\mathbf{f}^{(k+1)} = \left(\mathbf{H}^T\mathbf{H} + \lambda_p\left(\mathbf{T}^{(k)}\right)^T\mathbf{T}^{(k)}\right)^{-1}\mathbf{H}^T\mathbf{s} = \mathbf{W}^{(k)}\left(\mathbf{A}^{(k)}\right)^T\left(\mathbf{A}^{(k)}\left(\mathbf{A}^{(k)}\right)^T + \lambda_p\mathbf{I}\right)^{-1}\mathbf{s}, \tag{14}$$

where $\mathbf{A}^{(k)} = \mathbf{H}\mathbf{W}^{(k)}$. By starting with some feasible approximated initial solution, $\mathbf{f}^{(0)}$, the optimization gradually reinforces some of the already prominent entries while suppressing the rest until they become zeros. The repeated weighting procedure concentrates the solution in the minimal active regions that are essential for accurately reproducing the measurements (Rao et al., 2003). In the limit $p \to 0$ the diversity measure $\|\mathbf{f}\|_p^p$ becomes equal to the l_0-norm, i.e., $\lim_{p\to 0}\sum_j \left|\mathbf{f}_j\right|^p = \|\mathbf{f}\|_0$. Therefore, this approach provides another alternative algorithm to solve the l_0-norm optimization problem. In practice, this is done by using $\mathbf{W}_{jj}^{(k)} = \left|\mathbf{f}_j^{(k)}\right| + \varepsilon$ in Eq. (14).

The sparse regression problem can also be formulated using Bayesian statistics. In this work, we will study two different methods: Cauchy deconvolution and sparse Bayesian learning. Cauchy deconvolution is quite popular for seismic deconvolution (Sacchi, 1997) and sparse spectrum estimation (Sacchi and Ulrych, 1996). The formulation is based on assuming a Gaussian Likelihood model and a Cauchy prior distribution over $\mathbf{f}$. As the Cauchy distribution is heavy-tailed, it promotes solutions that are usually more sparse than the LASSO. When the prior and the likelihood are combined, the cost function to be minimized becomes (Sacchi, 1997)

$$\mathrm{J} = \|\mathbf{s}-\mathbf{H}\mathbf{f}\|_2^2 + 2\sigma^2\sum_{i=1}^{m+1}\log\left(1+\frac{\mathbf{f}_i^2}{2\sigma_c^2}\right), \tag{15}$$

where the scale parameter σ_c controls the level of sparsity that can be attained by the inversion and σ^2 is the noise variance. The maximum a posteriori solution to this non-linear system of equations can be found by iteratively computing the solution to the Tikhonov regression (Sacchi and Ulrych, 1996)

$$\mathbf{f}^{(k)} = \left(\mathbf{H}^T\mathbf{H} + \lambda_c\mathbf{Q}^{-1}\right)^{-1}\mathbf{H}^T\mathbf{s}, = \mathbf{Q}\mathbf{H}^T\left(\mathbf{H}\mathbf{Q}\mathbf{H}^T + \lambda_c\mathbf{I}\right)^{-1}\mathbf{s}, \tag{16}$$

where $\mathbf{f}^{(k)}$ denotes the solution at the k-th iteration, $\lambda_c = \sigma^2/\sigma_c^2$, and $\mathbf{Q}$ is a diagonal matrix updated at each iteration with elements given by

$$\mathbf{Q}_{ii} = 1 + \frac{1}{2}\left(\frac{\mathbf{f}_i^{(k-1)}}{\sigma_c}\right)^2. \tag{17}$$

The last method surveyed is Sparse Bayesian learning (SBL). It has been used to obtain parsimonious representations in the context of regression, classification and basis selection (Tipping, 2001). The SBL formulation is based on assuming a Gaussian Likelihood model and a zero-mean Gaussian prior distribution over $\mathbf{f}$, where a different variance α_j moderates the strength of the prior on each entry $\mathbf{f}_j$. Such prior encourages models with few non-zero weights. As shown by (Wipf and

Nagarajan, 2010) the maximum a posteriori SBL solution satisfies

$$\tilde{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \|\mathbf{s} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda g_{SBL}(\mathbf{f}), \tag{18}$$

where

$$g_{SBL}(\mathbf{f}) = \min_{\boldsymbol{\alpha}\geq 0} \{\mathbf{f}^T\mathbf{A}^{-1}\mathbf{f} + \log|\mu I + \mathbf{H}\mathbf{A}\mathbf{H}^T|\}, \tag{19}$$

$\mathbf{A}$ is a diagonal matrix with values $\mathbf{A}_{jj} = \alpha_j$ and $\mu = \lambda$. It has been shown that other values of μ may lead to even better performance (Wipf and Nagarajan, 2010). Indeed, in the limit $\mu \to 0$ the global minimum of Eq. (18) is only achieved at the minimum of the l_0-norm solution given by Eq. (8) (Wipf and Nagarajan, 2008). Notably, the resulting cost function possesses far fewer local minima than that of the l_0-norm, thus yielding fewer convergence errors (Wipf and Rao, 2004). Hence, this approach provides another important method to obtain l_0-norm-like sparse solutions.

For readers interested in learning more about sparse regression techniques we recommend the following survey lectures and review articles (Bruckstein et al., 2009; Daubechies et al., 2010; Tropp and Wright, 2010; Vidaurre et al., 2013; Wipf and Nagarajan, 2009; Wipf and Rao, 2004; Zhang et al., 2015; Zou et al., 2007).

3. Evaluated methods

In this manuscript, we implemented the sparse algorithms described in the previous section, as well as two other non-sparse methods. A qualitative overview of the main properties of the evaluated techniques, including the assumed noise model, regularization function, and optimization algorithm is provided in Table 8. Here we summarize the main aspects of their implementations and further technical details are provided in the Appendix section.

(1) Best-subset selection: a brute-force search over a reduced set of variables was implemented. In a first step, the NNLS algorithm was used to reduce the dimension of the data, from which all the variables with a value equal to zero were discarded. Our choice was motivated by previous experimental results indicating that the NNLS algorithm is mainly affected by false positives but not by false negatives (Slawski and Hein, 2011). Then, a combinatorial search is carried out over the retained variables to create all possible individual models comprising $k = 1$, 2 and 3 fibers, plus the isotropic compartment if considered. Finally, all models are ranked according to the extended Bayesian information criterion (EBIC) (Chen and Chen, 2008) and the model with the lowest EBIC is selected. This algorithm was termed NNLS-BSS-EBIC. To the best of our knowledge, this is the first time this approach is used to address the non-negative SD problem.

Table 1
Parameters used for the different reconstructions algorithms. All methods used the same parameters for different b-values, except for AdRL-SD and ARUMBA-SD, where it was necessary to modify the number of iterations N_{iter} for optimal results.

(1) NNLS-BSS-EBIC	$\gamma = 1$, $\sigma = 1/20$
(2) LASSO-EBIC	$\gamma = 1$, $\sigma = 1/20$
(3) Cauchy-SD	$\sigma^2 = 7 \times 10^{-4}$, $\sigma_c^2 = 10^{-6}$, $N_{iter} = 30$
(4) RFOCUSS	$\lambda_0 = 1.2 \times 10^{-3}$, $\varepsilon = 10^{-6}$, $N_{iter} = 30$
(5) IRL1	$\lambda = 1.5 \times 10^{-3}$, $\varepsilon = 10^{-3}$, $N_{iter} = 30$
(6) SBL	$\mu = 10^{-6}$, $c = 0.85$, $N_{iter} = 30$
(7) AdRL-SD	$\nu = 8$, $\eta = 0.06$
	$b = 3000$ mm^2/s, $N_{iter} = 50$
	$b = 1500$ mm^2/s, $N_{iter} = 140$
	$b = 1000$ mm^2/s, $N_{iter} = 210$
(8) ARUMBA-SD	$b = 3000$ mm^2/s, $N_{iter} = 50$
	$b = 1500$ mm^2/s, $N_{iter} = 140$
	$b = 1000$ mm^2/s, $N_{iter} = 210$

Table 2
Quantitative performance of all methods in synthetic voxels with a single fiber. Reported values are the average of each metric. The best method for each metric is highlighted in bold. A ranking of global performance was included based on the *GRP* measure.

$b = 3000$ s/mm^2	θ	Δf	n$^+$	n$^-$	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	3.70	0.03	0.08	0.00	0.92	1.88	2
(2) LASSO-EBIC	**3.52**	0.05	0.22	0.00	0.81	3.27	4
(3) Cauchy-SD	5.26	0.14	0.32	0.00	0.68	6.50	5
(4) RFOCUSS	5.72	0.17	0.38	0.00	0.63	9.59	8
(5) IRL1	5.87	0.18	0.39	0.00	0.63	8.29	7
(6) SBL	4.82	0.10	0.23	0.00	0.78	7.15	6
(7) AdRL-SD	3.54	0.02	0.12	0.00	0.89	2.07	3
(8) ARUMBA-SD	3.57	**0.01**	**0.04**	0.00	**0.96**	**1.25**	**1**
$b = 1500$ s/mm^2	θ	Δf	n$^+$	n$^-$	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	3.94	0.05	0.20	0.00	0.81	3.95	4
(2) LASSO-EBIC	3.60	0.05	0.20	0.00	0.82	3.70	3
(3) Cauchy-SD	4.75	0.10	0.23	0.00	0.78	5.12	6
(4) RFOCUSS	5.00	0.12	0.26	0.00	0.75	5.76	7
(5) IRL1	5.76	0.16	0.38	0.00	0.68	7.55	8
(6) SBL	4.65	0.08	0.19	0.00	0.81	4.42	5
(7) AdRL-SD	**3.18**	**0.00**	**0.00**	0.00	**1.00**	**0.75**	**1**
(8) ARUMBA-SD	3.19	0.00	0.00	0.00	1.00	0.75	2

Table 3
Quantitative performance of all methods in synthetic voxels with two fibers. Reported values are the average of each metric. The best method for each metric is highlighted in bold. A ranking of global performance was included based on the *GRP* measure.

$b = 3000$ s/mm^2	θ	Δf	n$^+$	n$^-$	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	7.45	0.20	0.10	0.38	0.62	4.37	5
(2) LASSO-EBIC	6.97	0.19	0.19	0.36	0.57	4.91	6
(3) Cauchy-SD	6.80	0.17	0.10	0.25	0.71	3.64	2
(4) RFOCUSS	6.75	0.17	0.12	0.23	**0.72**	3.71	3
(5) IRL1	6.78	0.17	0.15	**0.22**	0.69	3.91	4
(6) SBL	**6.34**	**0.16**	**0.07**	0.27	0.71	**3.42**	**1**
(7) AdRL-SD	9.09	0.27	0.24	0.64	0.46	6.90	8
(8) ARUMBA-SD	8.45	0.26	0.17	0.56	0.49	6.00	7
$b = 1500$ s/mm^2	θ	Δf	n$^+$	n$^-$	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	7.60	0.19	0.19	0.35	0.59	5.08	5
(2) LASSO-EBIC	6.98	0.19	0.20	0.33	0.56	5.11	6
(3) Cauchy-SD	6.72	0.19	0.05	0.27	0.72	3.48	3
(4) RFOCUSS	6.67	0.19	0.06	0.25	**0.72**	3.47	2
(5) IRL1	6.81	**0.18**	0.17	**0.23**	0.66	4.28	4
(6) SBL	**6.40**	0.18	**0.05**	0.28	0.70	**3.40**	**1**
(7) AdRL-SD	8.78	0.27	0.14	0.58	0.50	6.08	8
(8) ARUMBA-SD	8.50	0.26	0.17	0.53	0.50	6.00	7

(2) LASSO & EBIC: the non-negative LASSO problem (see Eq. (9)) was solved using the modified least-angle regression (LARS) solver (Efron et al., 2004) included in the SparseLab Toolbox (https://sparselab.stanford.edu/). This algorithm allows obtaining the entire regularization path (i.e., the solutions for many regularization values λ_1). The optimal result is searched among the many available solutions using model selection techniques (Zou et al., 2007). Specifically, the optimal λ_1 was chosen according to the EBIC criterion. Finally, before the EBIC evaluation the coefficients are 'debiased', i.e., the solution was recomputed using a standard least squares method over the identified non-zero coefficients. This method was termed LASSO-EBIC. It is similar to the l_1-norm regularization approach introduced in (Landman et al., 2012). However, we extended that earlier work by providing an automatic way to select the optimal regularization parameter and by removing the estimation bias that was highlighted in (Daducci et al., 2014b). Preliminary results suggest that the performance of this implementation is superior to the standard approach based on selecting a fixed λ_1 for all the voxels.

Table 4
Quantitative performance of all methods in synthetic voxels with two fibers, where one fiber is very dominant. Reported values are the average of each metric. The best method for each metric is highlighted in bold. A ranking of global performance was included based on the *GRP* measure.

$b = 3000$ s/mm^2	θ	Δf	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	18.75	0.22	0.49	0.83	0.53	4.46	3
(2) LASSO-EBIC	**14.26**	**0.16**	**0.42**	**0.54**	**0.62**	**3.37**	**1**
(3) Cauchy-SD	25.74	0.28	0.87	1.00	0.23	6.40	8
(4) RFOCUSS	23.79	0.27	0.86	0.87	0.26	5.99	6
(5) IRL1	25.60	0.27	0.91	0.93	0.21	6.32	7
(6) SBL	20.19	0.23	0.60	0.79	0.46	4.82	5
(7) AdRL-SD	17.22	0.21	0.46	0.77	0.56	4.15	2
(8) ARUMBA-SD	18.62	0.23	0.49	0.84	0.53	4.50	4
$b = 1500$ s/mm^2	θ	Δf	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	11.59	**0.10**	**0.34**	**0.33**	**0.70**	**2.49**	**1**
(2) LASSO-EBIC	**11.54**	0.11	0.35	0.34	0.68	2.61	2
(3) Cauchy-SD	21.64	0.25	0.77	0.67	0.30	5.41	6
(4) RFOCUSS	21.92	0.24	0.75	0.68	0.34	5.31	5
(5) IRL1	18.05	0.22	0.72	0.51	0.38	4.68	4
(6) SBL	16.56	0.18	0.47	0.50	0.58	3.70	3
(7) AdRL-SD	30.69	0.44	0.85	1.69	0.15	8.35	8
(8) ARUMBA-SD	27.62	0.39	0.75	1.50	0.25	7.45	7

Table 5
Quantitative performance of all methods in synthetic voxels with three fibers. Reported values are the average of each metric. The best method for each metric is highlighted in bold. A ranking of global performance was included based on the *GRP* measure.

$b = 3000$ s/mm^2	θ	Δf	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	10.77	0.27	0.29	1.12	0.43	5.35	6
(2) LASSO-EBIC	10.67	0.25	0.34	1.06	0.46	5.33	5
(3) Cauchy-SD	10.96	0.21	0.27	1.01	0.40	5.02	4
(4) RFOCUSS	10.16	0.19	0.25	0.89	0.46	4.57	3
(5) IRL1	9.45	**0.18**	0.21	**0.77**	**0.53**	**4.03**	**1**
(6) SBL	**9.28**	0.21	**0.20**	0.87	0.50	4.27	2
(7) AdRL-SD	12.27	0.36	0.54	1.53	0.37	7.20	8
(8) ARUMBA-SD	11.20	0.34	0.36	1.30	0.42	6.11	7
$b = 1500$ s/mm^2	θ	Δf	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	10.55	0.24	0.34	1.00	0.47	5.38	6
(2) LASSO-EBIC	10.87	0.25	0.29	0.96	0.48	5.19	5
(3) Cauchy-SD	10.91	0.23	0.23	0.99	0.41	5.06	4
(4) RFOCUSS	10.29	0.21	0.20	0.89	0.45	4.64	3
(5) IRL1	9.40	**0.18**	0.16	**0.69**	**0.58**	**3.80**	**1**
(6) SBL	**9.27**	0.22	**0.14**	0.83	0.52	4.12	2
(7) AdRL-SD	12.23	0.36	0.40	1.41	0.38	6.80	8
(8) ARUMBA-SD	11.56	0.33	0.37	1.27	0.41	6.33	7

(3) Cauchy deconvolution: this algorithm was implemented using Eqs. (16) and (17). As the standard algorithm does not include non-negativity constraints, we propose a parameterization for obtaining non-negative solutions within the same reweighted Tikhonov regularization framework. For more details see the Appendix. This method was termed Cauchy-SD.

(4) Non-negative regularized FOCUSS: this algorithm was implemented using Eq. (14). Our preliminary numerical experiments indicated optimal reconstructions for $p = 0$ (i.e., l_0-norm regularization) so we focused only on that limit. As the standard algorithm does not include non-negativity constraints, we adapted it by using the same parameterization used in the Cauchy-SD method. The resulting algorithm was termed Non-Negative Regularized FOCUSS (RFOCUSS).

(5) Non-negative iterative reweighted l_1 minimization: this algorithm was implemented using Eq. (11). At each iteration, the LASSO problem was solved using the algorithm FISTA (Beck and Teboulle, 2009), which is included in the open-source SPArse Modeling Software (SPAMS: http://spams-devel.gforge.inria.fr/) (Mairal et al., 2014), which by construction constrains the solution to be non-negative.

(6) Sparse Bayesian Learning: the SBL cost function given in Eqs. (18) and (19) was minimized using an alternating iterative scheme based on the Expectation Maximization (EM) algorithm developed in (Wipf and Nagarajan, 2010). The same parameterization used in Cauchy-SD and RFOCUSS was applied here to obtain non-negative solutions. For more details about the equations and the optimization algorithm, see the Appendix section.

(7) Accelerated Richardson-Lucy deconvolution: the damped RL-SD method (dRL-SD) proposed in (Dell'Acqua et al., 2010) was implemented. This algorithm transforms a perfectly smooth initial estimate into sharper estimates, with sharpness increasing with the number of iterations. Hence, the number of iterations can be considered as a regularization parameter controlling the smoothness of the final estimate. In this study, a new variant to accelerate the convergence of this technique is proposed. It is based on applying ideas from Nesterov's accelerated gradient descent method (Nesterov, 2004) to the multiplicative updates and on the adaptive restart strategy proposed by (O'Donoghue and Candès, 2015). The resulting algorithm was termed AdRL-SD.

(8) Accelerated RUMBA-SD: the robust and unbiased model-based spherical deconvolution (RUMBA-SD) method developed in (Canales-Rodríguez et al., 2015) was included in the evaluation. A

Table 6
Quantitative performance of all methods in the structured field phantom from the HARDI reconstruction workshop 2012. Reported values are the average of each metric. All methods were evaluated using a signal with 60 diffusion sensitizing gradients and $b = 3000$ s/mm^2, except DSI_{LR} which employed a DSI sampling scheme with 257 diffusion gradients and $b_{max} = 8000$ s/mm^2. The best method for each metric is highlighted in bold. A ranking of global performance was included based on the *GRP* measure.

SNR=30	θ	Δf	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	5.35	0.10	0.13	**0.30**	0.63	5.01	6
(2) LASSO-EBIC	5.21	0.10	0.09	0.31	0.66	4.43	3
(3) Cauchy-SD	5.09	0.09	0.08	0.35	0.66	4.38	2
(4) RFOCUSS	5.09	0.09	0.10	0.34	0.64	4.62	4
(5) IRL1	5.15	**0.09**	0.12	0.33	0.63	4.77	5
(6) SBL	**5.04**	0.09	**0.05**	0.35	**0.68**	**3.98**	**1**
(7) AdRL-SD	6.39	0.17	0.10	0.59	0.62	6.13	11
(8) ARUMBA-SD	5.97	0.15	0.07	0.49	0.64	5.30	8
(9) CSD	5.95	0.17	0.08	0.57	0.62	5.83	10
(10) Ball & Stick	5.30	0.11	0.10	0.44	0.64	5.03	7
(11) DSI_{LR}*	5.78	0.13	0.10	0.54	0.65	5.52	9
SNR=20	θ	Δf	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	6.43	0.12	0.07	0.39	0.65	4.66	5
(2) LASSO-EBIC	6.27	0.12	0.07	0.36	0.66	4.53	4
(3) Cauchy-SD	5.88	0.10	0.08	0.36	0.66	4.38	2
(4) RFOCUSS	5.81	0.10	0.09	0.34	0.65	4.45	3
(5) IRL1	**5.79**	**0.10**	0.11	**0.33**	0.64	4.66	6
(6) SBL	5.88	0.10	**0.06**	0.37	**0.67**	**4.20**	**1**
(7) AdRL-SD	7.22	0.17	0.12	0.57	0.63	6.23	11
(8) ARUMBA-SD	6.70	0.15	0.07	0.49	0.65	5.21	7
(9) CSD	6.67	0.17	0.09	0.55	0.62	5.81	10
(10) Ball & Stick	6.41	0.12	0.09	0.52	0.63	5.26	8
(11) DSI_{LR}*	6.44	0.13	0.11	0.55	0.65	5.61	9
SNR=15	θ	Δf	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	7.46	0.13	0.09	0.46	0.65	4.57	4
(2) LASSO-EBIC	7.33	0.14	0.10	0.44	0.64	4.64	6
(3) Cauchy-SD	6.79	0.11	0.13	0.39	0.63	4.50	3
(4) RFOCUSS	6.72	0.11	0.15	0.37	0.62	4.58	5
(5) IRL1	**6.72**	**0.10**	0.14	**0.34**	0.63	4.43	2
(6) SBL	6.84	0.11	**0.10**	0.38	**0.65**	**4.21**	**1**
(7) AdRL-SD	7.87	0.18	0.17	0.55	0.58	5.89	10
(8) ARUMBA-SD	7.56	0.16	0.13	0.51	0.62	5.23	8
(9) CSD	7.32	0.17	0.17	0.51	0.57	5.72	9
(10) Ball & Stick	8.26	0.16	0.18	0.70	0.59	6.20	11
(11) DSI_{LR}*	7.31	0.14	0.13	0.55	0.65	5.04	7

Table 7
Quantitative performance of all methods in the 3D phantom from the HARDI reconstruction workshop 2013. Reported values are the average of each metric. All methods were evaluated using a signal with 60 diffusion sensitizing gradients and $b = 3000$ s/mm^2. The best method for each metric is highlighted in bold. A ranking of global performance was included based on the *GRP* measure.

SNR=30	θ	*Δf*	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	14.33	0.16	**0.05**	0.20	0.82	**3.68**	**1**
(2) LASSO-EBIC	12.79	0.14	0.09	0.19	0.82	3.87	2
(3) Cauchy-SD	29.28	0.32	0.06	0.42	0.64	7.04	10
(4) RFOCUSS	23.09	0.25	0.07	0.34	0.71	5.80	9
(5) IRL1	20.72	0.23	0.07	0.35	0.73	5.50	8
(6) SBL	19.85	0.22	0.09	0.30	0.74	5.40	6
(7) AdRL-SD	7.78	**0.12**	0.24	**0.15**	0.78	5.02	5
(8) ARUMBA-SD	9.46	0.12	0.14	0.16	0.82	3.97	3
(9) SS3T-CSD	18.75	0.33	0.06	0.29	0.74	5.44	7
(10) Ball & Stick	**6.67**	0.37	0.11	0.15	**0.87**	4.28	4
SNR=20	θ	*Δf*	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	18.43	0.20	**0.05**	0.27	0.78	**3.99**	**1**
(2) LASSO-EBIC	15.94	0.17	0.13	0.24	0.76	4.28	2
(3) Cauchy-SD	30.37	0.34	0.07	0.44	0.62	6.54	10
(4) RFOCUSS	23.41	0.26	0.07	0.35	0.70	5.25	8
(5) IRL1	21.13	0.23	0.08	0.35	0.73	4.94	6
(6) SBL	19.74	0.21	0.10	0.30	0.74	4.74	5
(7) AdRL-SD	**8.95**	**0.13**	0.32	**0.17**	0.74	5.02	7
(8) ARUMBA-SD	9.98	0.13	0.24	0.18	0.77	4.39	3
(9) SS3T-CSD	24.47	0.39	0.09	0.35	0.66	6.12	9
(10) Ball & Stick	9.45	0.37	0.16	0.22	**0.82**	4.71	4
SNR=10	θ	*Δf*	n^+	n^-	*SR*	*GRP*	Ranking
(1) NNLS-BSS-EBIC	29.30	0.31	**0.11**	0.41	0.62	4.97	6
(2) LASSO-EBIC	24.84	0.27	0.23	0.37	0.60	4.86	5
(3) Cauchy-SD	34.41	0.39	0.14	0.48	0.55	6.01	10
(4) RFOCUSS	22.93	0.24	0.20	0.35	0.66	4.41	3
(5) IRL1	19.78	0.18	0.20	0.34	0.70	**3.92**	**1**
(6) SBL	18.91	0.18	0.33	0.28	0.61	4.24	2
(7) AdRL-SD	**12.74**	0.18	0.82	**0.21**	0.38	5.61	8
(8) ARUMBA-SD	13.04	**0.17**	0.72	0.22	0.44	5.20	7
(9) SS3T-CSD	21.19	0.29	0.59	0.27	0.39	5.97	9
(10) Ball & Stick	16.21	0.39	0.28	0.37	**0.71**	4.82	4

new variant, termed ARUMBA-SD, was implemented to accelerate the convergence. In particular, the same acceleration approach included in the AdRL-SD method was adopted. See the Appendix for more details.

4. Materials and methods

4.1. Evaluation metrics

Local peaks from each reconstructed *FOD* were identified as those vertices in the spherical grid with higher values than their adjacent neighbors within a tolerance cone of 15°. Only the highest three peaks were considered and only those with amplitudes exceeding at least one-tenth of the amplitude of the highest peak (i.e., $0.1 \cdot f_{max}$) were retained as in (Parker et al., 2013).

The performance of the algorithms was quantified by comparing the resulting peaks against the true fiber directions used to generate the synthetic signals. We adopted some evaluation metrics widely used in the literature. Specifically, we used the angular error, defined as the average minimum angle between the extracted peaks and the true fiber directions (Tuch et al., 2002):

$$\theta = \frac{1}{M_{true}} \sum_{k=1}^{M_{true}} \min_m \left\{ \arccos\left(\left| \mathbf{e}_m^T \mathbf{v}_k \right| \right) \right\}, \tag{20}$$

where M_{true} is the true number of fiber populations, $\mathbf{e}_m$ is the unitary vector along with the m-th detected fiber peak and $\mathbf{v}_k$ is the unitary vector along the k-th true fiber direction. As usual, the angular error between each pair of fibers was measured by comparing each true fiber with the closest estimated fiber. As the number of detected peaks can be lower than the number of true fibers, some considerations are worth bearing in mind when evaluating Eq. (20). Specifically, when considering simulated signals with two or three fibers, if only one peak is detected, then the second and third missing fibers are assumed to be aligned along the same direction of the detected peak. Similarly, when comparing results for signals generated using three fibers, if only two peaks are detected, then the third missing fiber is considered to be aligned to the same direction as the detected peak that is the most aligned to the unrecovered fiber (Canales-Rodríguez et al., 2008).

The volume fraction error of the estimated fiber compartments was assessed by means of the mean absolute error between the estimated and the actual peak amplitudes:

$$\Delta f = \frac{1}{M_{true}} \sum_{k=1}^{M_{true}} |f_m - f_k|, \tag{21}$$

where f_m is the normalized height of the m-th detected fiber peak and f_k is the volume fraction of the k-th true fiber. This metric was evaluated by matching the true fibers with the closest estimated fibers as above.

The mean success rate (*SR*) was defined as the proportion of voxels in which the algorithm estimated the right number of fiber compartments. Specifically, we assigned a value of 1 or 0 to each voxel depending on whether or not the reconstruction was considered successful. The solution was considered successful only when the number of detected peaks was equal to the number of true fibers and the angle between each peak and the closest fiber was lower than 25°. This allowed distinguishing between the estimated peaks that are "close enough" to the true fibers, i.e. true positives, and those which are spurious, i.e. false positives (Daducci et al., 2014a). To discriminate the different factors leading to an erroneous estimation, the mean number of over-estimated n^+ and under-estimated n^- fiber populations were also computed. Accordingly, fibers outside the tolerance cone were considered as spurious fibers for the evaluation of n^+, and the absence of a peak within the tolerance cone was considered as a false negative for the evaluation of n^-.

Finally, we propose a new global relative performance (*GRP*) measure to quantify each method in relation to the average performance with respect to the proposed metrics

$$GRP(i) = \frac{\theta_i}{\langle \theta \rangle} + \frac{\Delta f_i}{\langle \Delta f \rangle} + \frac{n_i^+}{\langle n^+ \rangle} + \frac{n_i^-}{\langle n^- \rangle} + \frac{1 - SR_i}{1 - \langle SR_i \rangle}, \tag{22}$$

where x_i denotes the value of metric x for method i and $\langle x \rangle$ is the mean value of x for all methods $i = \{1, \ldots, N\}$, respectively. The *GRP* allows us to rank all methods by taking into account their performance in terms of all the individual metrics. Methods yielding lower *GRP* values should be recommended over methods with higher values.

The *GRP* was also used to find the optimal parameters for each individual method. This was possible by changing the meaning of the variables. Now x_i was considered to be the value of metric x for a given set of parameters (i.e., regularization value, the number of iterations, etc.) denoted by index i, and $\langle x \rangle$ was the mean value of x resulting from all sets of parameters.

4.2. Synthetic datasets

This work focuses on four different types of fiber configurations, which includes the cases of voxels containing a single fiber, two fibers, two fibers with a very dominant fiber and three fibers. For each considered case, three different datasets were produced, each one corresponding to a different b-value: $b = 3000$ s/mm^2, $b = 1500$ s/mm^2 and $b = 1000$ s/mm^2. For each b-value and fiber configuration, the diffusion signal in 1000 voxels was generated using the multi-tensor model (Tuch et al., 2002). A clinical sampling scheme with $N = 64$ diffusion gradients and a single $b = 0$ was employed. In all voxels, a constant $S_0 = 100$ was assumed.

Table 8
A qualitative overview of the main properties of the evaluated techniques: noise model, non-negativity, regularization function, optimization algorithm and computation time. In order to guide the choice for a technique, suited to the goals of researchers in the field, we highlight a set of experimental conditions (i.e., *b*-values and fiber configurations) where the methods showed optimal and sub-optimal performance.

Method	Noise model	Non-negative constraint	Regularization function	Optimization algorithm	Average computation time per voxel (seconds) ***	Optimal in the following cases	Sub-optimal in the following cases
NNLS-BSS-EBIC	Gaussian-Rician*	Yes	l_0-norm & extended Bayesian information criterion (EBIC)	non-negative least squares & combinatorial search	0.024	b ≥ 1000; single fiber population; fiber crossing with a dominant fiber	fiber crossing with three fibers
LASSO-EBIC	Gaussian-Rician*	Yes	l_1-norm & EBIC	Least-angle regression (entire regularization path)	0.009	b ≥ 1000; single fiber population; fiber crossing with a dominant fiber	fiber crossing with three fibers
Cauchy-SD	Gaussian	Yes	The logarithm of Cauchy distribution	Iterative reweighted Tikhonov regularization (IRTR)	0.006	b ≥ 1000; fiber crossing with inter-fiber angle <50º	single fibers; fiber crossing with a dominant fiber
RFOCUSS	Gaussian	Yes	l_0-norm	Focal underdetermined system solver: Iteratively reweighted l_2-norm	0.005	b ≥ 1000; fiber crossing with inter-fiber angle <50º	single fibers; fiber crossing with a dominant fiber
IRL1	Gaussian	Yes	l_0-norm	Iteratively reweighted l_1-norm using FISTA	0.16	b ≥ 1000; fiber crossing with inter-fiber angle <50º; fiber crossing with three fibers	single fibers; fiber crossing with a dominant fiber
SBL	Gaussian	Yes	SBL cost function**	An alternating scheme using type-II maximum likelihood (Expectation-Maximization) & IRTR	0.13	b ≥ 1000; fiber crossing with inter-fiber angle >30º; fiber crossing with three fibers	single fibers
AdRL-SD	Gaussian	Yes	None	Richardson-Lucy (multiplicative updates)	0.0018	b > 1500; single fiber population; fiber crossing with inter-fiber angle >60º	b < 1500; fiber crossing with inter-fiber angle <60º; fiber crossing with three fibers
ARUMBA-SD	Rician	Yes	None	Richardson-Lucy (multiplicative updates)	0.001	b > 1500; single fiber population; fiber crossing with inter-fiber angle >55º	b < 1500; fiber crossing with inter-fiber angle <55º; fiber crossing with three fibers

*Gaussian approximation of a Rician distribution.
**Regularization function defined in Eq. (19).
***Implemented in Matlab®: the computation was done using an Intel Core i7-7700HQ, 2.8 GHz processor and the reconstruction was not parallelized.

In order to simulate the structural heterogeneity observed in real data and to avoid over-optimistic results, different volume fractions and diffusivities were considered for each fiber compartment. The resulting signals were contaminated by Rician noise with signal-to-noise ratios (i.e., $SNR = S_0/\sigma$) close to those reported in clinical data. The following subsections provide further details about the synthetic data generation procedure.

4.2.1. Single fiber population

Fiber orientations, diffusivities, and SNRs were different for each voxel. The fiber orientation was chosen as a random unit vector on the sphere. The SNR was generated from a uniform distribution on the interval [15, 30], i. e, $SNR \sim U(15, 30)$. The diffusivities were generated from uniform distributions with $\lambda_1 \sim U(1.4\,10^{-3}, 1.8\,10^{-3})$ mm^2/s, $\lambda_2 \sim U(0.1\,10^{-3}, 0.5\,10^{-3})$ mm^2/s and $\lambda_3 = \lambda_2$ (Canales-Rodríguez et al., 2009) and the synthetic signals were generated for the acquisition protocol described in the previous section.

4.2.2. Two-fibers

Various synthetic datasets with different inter-fiber angles were generated. The inter-fiber angle was gradually modified from 1 to 90°, in steps of 5°. The volume fraction of the first fiber was generated from a uniform distribution on the interval [0.3, 0.7], i.e., $f_1 \sim U(0.3, 0.7)$, and the second fiber was fixed at $f_2 = 1 - f_1$. The diffusivities of each fiber compartment and the noise level were generated from uniform distributions as described in the single fiber experiment.

4.2.3. Two-fibers with a very dominant fiber

These datasets were created to test the performance of the methods to detect fiber crossings consisting of a dominant fiber and a fiber with a very small volume fraction. The volume fraction of the non-dominant fiber was generated from a uniform distribution on the interval [0.1, 0.3], i.e., $f_1 \sim U(0.1, 0.3)$, and $f_2 = 1 - f_1$. The inter-fiber angle was gradually modified from 50 to 90°, applying 5-degree increases. We did not employ lower inter-fiber angles to focus on the impact of unbalanced volume fractions rather than small inter-fiber angles. For each voxel, the noise level and the diffusivities of each fiber compartment were generated from uniform distributions using the same range of values described in previous sections.

4.2.4. Three-fibers

The ability of the different methods to detect crossings of three fibers was tested by simulating voxels containing three fiber bundles in different orientations. The three fibers were generated to be equidistant from one another. The inter-fiber angles were gradually modified from 1 to 90°, in steps of 5°. The volume fractions were generated as follows: $f_1 \sim U(0.25, 0.3)$, $f_2 \sim U(0.3, 0.35)$ and $f_3 = 1 - (f_1 + f_2)$. As in previous sections, the diffusivities and noise levels were generated from uniform distributions.

4.3. Phantoms

4.3.1. HARDI reconstruction workshop 2012

The *structured field phantom* released in the "HARDI Reconstruction

Challenge 2012" workshop organized as part of the International Symposium on Biomedical Imaging (ISBI) 2012 conference (Daducci et al., 2014a) was also considered. The 3D phantom consists of $16 \times 16 \times 5$ voxels comprising five different fiber bundles with different diffusivities and fractional anisotropy. It was designed to simulate realistic non-planar configurations of bending, crossing and kissing tracts. The proportion of voxels containing one, two and three fibers was 35%, 35%, and 30% respectively. In voxels with multiple fibers, the inter-fiber angles were distributed as follows: 27% smaller than 30º, 35% in the range 31º-60º and 38% above 61º. For more details see (Daducci et al., 2014a) and the workshop webpage (http://hardi.epfl.ch/static/events/2012_ISBI/).

From the available datasets, we have chosen the same HARDI data used to evaluated the CSD method (Tournier et al., 2008), which was created using a single shell with $b = 3000\,\mathrm{s/mm^2}$ and $N = 60$ diffusion sensitizing gradients. Three different datasets corresponding to different noise levels were included in the analysis, i.e., SNR = 30, 20 and 15. Results from the $\mathrm{DSI_{LR}}$ (i.e., Diffusion Spectrum Imaging with deconvolution) method (Canales-Rodríguez et al., 2010) were also reported even though it employed a sampling scheme with many more diffusion gradients (i.e., $N = 257$) and a higher maximum b-value (i.e., $b_{max} = 8000\,\mathrm{s/mm^2}$). $\mathrm{DSI_{LR}}$ was included as a reference since it yielded the most accurate and stable reconstructions in the workshop (Daducci et al., 2014a). In addition, we also included two state-of-the-art SD techniques: the CSD method (Tournier et al., 2008) implemented in MRtrix3 (http://www.mrtrix.org/) and the Ball & Stick model (Behrens et al., 2007) implemented in FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT).

4.3.2. HARDI reconstruction workshop 2013

The implemented algorithms were tested as well on the synthetic diffusion MRI phantom developed for the "HARDI Reconstruction Challenge 2013" workshop, organized during the ISBI 2013 conference. This phantom comprises a set of 27 fiber bundles with fibers of varying radii and geometry which connect different areas of a 3D image with $50 \times 50 \times 50$ voxels. It contains a wide range of configurations including branching, crossing and kissing fibers, together with the presence of isotropic compartments. The proportion of voxels containing one, two and three fibers was 88.7%, 9.8%, and 1.5% respectively. The proportion of fiber crossings with dominant fibers (i.e., >0.7) was 51.6%.

The intra-voxel diffusion MRI signal was generated using $N = 64$ gradient sensitizing directions with constant $b = 3000\,\mathrm{s/mm^2}$ plus one additional image with $b = 0$. In voxels belonging to single-fiber white matter bundles, the signal measured along the q-space unit direction $\hat{\mathbf{q}} = \mathbf{q}/|\mathbf{q}|$ was generated by a mixture of signals from intra- and extra-axonal compartments: $S = \mathrm{f_{int}} s_{\mathrm{int}}(\mathbf{q}, \hat{\mathbf{v}}, \tau, L, R) + \mathrm{f_{ext}} s_{\mathrm{ext}}(\mathbf{q}, \hat{\mathbf{v}}, b, \lambda_1, \lambda_2)$, where $\hat{\mathbf{v}}$ denotes the local fiber orientation. The intra-axonal signal s_{int} was created following the theoretical model of a restricted diffusion process inside a cylinder of length $L = 5\,mm$ and radius $R = 5\,\mu m$ at a diffusion time $\tau = 20.8\,\mathrm{s}$. The extra-axonal signal s_{ext} was generated using a diffusion tensor model with cylindrical symmetry (i.e., $\lambda_1 = 1.7\ 10^{-3}\,\mathrm{mm^2/s}$, $\lambda_2 = \lambda_3 = 0.2\ 10^{-3}\,\mathrm{mm^2/s}$). Mixture fractions were fixed to $\mathrm{f_{int}} = 0.6$ and $\mathrm{f_{ext}} = 0.4$. The dataset is available at http://hardi.epfl.ch/static/events/2013_ISBI/.

Like in our HARDI phantom 2012 evaluation, the Ball & Stick model and the CSD method were included as a reference. In this case, we choose the single-shell 3-tissue CSD (SS3T-CSD) method (Dhollander et al., 2016; Dhollander and Connelly, 2016) implemented in MRtrix3 because this phantom simulates different tissues, and thus it may provide a better fitting to the data. In fact, we verified that results from SS3T-CSD were more accurate than those from CSD in this dataset.

4.4. Practical implementation details

The eight methods were implemented using the same generative diffusion model and the same linearization framework (i.e., leading to the same dictionary $\mathbf{H}$). In the case of the experiments considering the four different types of fiber configuration reported in sections 4.2.1-4.2.4, the dictionary was created using the signal generative model given in Eqs. (3) and (4) with diffusivities $\lambda_1 = 1.6\ 10^{-3}\ \mathrm{mm^2/s}$ and $\lambda_2 = 0.3\ 10^{-3}\ \mathrm{mm^2/s}$. These values were chosen for being in the middle of the range of diffusivities used to generate the synthetic signals. Hence, the assumed diffusivities could be lower or higher than the "true" diffusivities, allowing us to incorporate the possibility of model misspecification. The dictionary was created using m = 724 fiber orientations distributed on the unit sphere (362 unit vectors on the hemisphere with antipodal symmetry) with a mean angular separation between adjacent neighbor vertices of 8.36°, and a standard deviation of 1.18°.

In experiments involving the phantom datasets from both HARDI Reconstruction Challenges a different dictionary was created for each data with different SNR. In a first step, a diffusion tensor model was fitted to the data. Then, all voxels with fractional anisotropy higher than 0.85*max (FA) were labeled as voxels containing a single fiber. Finally, the mean values of the diffusivities (i.e., λ_1 and λ_2) in these voxels were used to build the dictionary.

The initial solution $\mathbf{f}^0$ for Cauchy-SD, RFOCUSS, SBL, AdRL-SD, and ARUMBA-SD was set as a non-negative iso-probable spherical function. In the IRL1 method, the diagonal matrix $\mathbf{T}$ was initialized as $\mathbf{T}_{jj}^{(0)} = \varepsilon$. In order to select the optimal parameters for each method, we conducted preliminary experiments on the synthetic data for voxels with two fibers. The estimated values are listed in Table 1. They were selected as those minimizing the *GRP* measure given by Eq. (22). The only methods that required adjusting the number of iterations for each b-value were AdRL-SD and ARUMBA-SD. The reported values in Table 1 were employed in all experiments described in next sections.

In both datasets from the HARDI Reconstruction Challenges, the Ball & Stick model was fitted using default parameters in FSL. A maximum number of $k = 3$ fibers per voxel and a Rician noise model was specified. Likewise, SS3T-CSD was evaluated on the HARDI Challenge 2013 phantom using standard reconstruction options in MRtrix3. The CSD method was initially calculated on the HARDI Challenge 2012 phantom using the default regularization parameters suggested by the software itself (i.e., $l_{\max} = 8$, $\lambda_{neg} = 1$ and $\lambda_{norm} = 1$, where $l_{\max}$ is the maximum spherical harmonic order, λ_{neg} is the regularization parameter that controls the strength of the non-negativity constraint and λ_{norm} is a regularization parameter controlling the norm of the solution). However, the performance of this method was affected by a higher number of spurious fibers compared to the other evaluated techniques. For this reason, we explored its performance as a function of these parameters and we found better results when using the following values: $l_{\max} = 8$, $\lambda_{neg} = 4$ and $\lambda_{norm} = 0$. In order to place the method in its best conditions, we only report results corresponding to this set of parameters optimized for this dataset.

5. Results

5.1. Single fiber population

The performance of all algorithms in the experiment considering synthetic voxels with a single fiber is summarized in Table 2. Results corresponding to $b = 1000\,\mathrm{s/mm^2}$ are reported in Table 1 of the Supplementary Material document.

From these results, we can draw the following conclusions. In the first place, the Richardson-Lucy algorithms ARUMBA-SD and AdRL-SD provided the best global reconstructions. While ARUMBA-SD performed better for $b = 3000\,\mathrm{s/mm^2}$, at lower b-values AdRL-SD was slightly better. The performance of the other methods also depended on the b-values. For instance, NNL-BSS-EBIC ranked #2 for $b = 3000\,\mathrm{s/mm^2}$ and #4 for $b = 1500\,\mathrm{s/mm^2}$ and $b = 1000\,\mathrm{s/mm^2}$, respectively. At low b-values, LASSO-EBIC was the third best method. The performance of iterative reweighted methods like Cauchy-SD, RFOCUSS, IRL1, and SBL was

affected by a higher number of over-estimated fibers n^+. These methods tend to produce an average of 20–40 false fibers every 100 voxels.

5.2. Two-fibers

The quantitative performance of all algorithms is depicted in Fig. 1 and Table 3. Fig. 1 corresponds to results for $b = 3000$ s/mm^2 and Table 3 summarizes results for $b = 3000$ and 1500 s/mm^2. Table 2 of the Supplementary Material shows results for $b = 1000$ s/mm^2.

A set of patterns can be drawn from these results. First, iterative reweighted methods provided the highest mean *SR* values, followed in descending order by the model-selection methods NNL-BSS-EBIC and LASSO-EBIC, and finally by ARUMBA-SD and AdRL-SD (see Fig. 1). This trend is explained by the superior performance of reweighted methods to resolve fiber crossings with small inter-fiber angles. Second, at high inter-fiber angles, ARUMBA-SD and AdRL-SD were able to resolve fiber crossings with higher *SR* than all iterative reweighted methods, except SBL. Finally, SBL performed well over the whole range of inter-fiber angles. It was the only method showing a stable recovery (i.e., high and constant *SR* value) from 35 to 90°. For lower angles, its performance decreased nearly linearly. The above patterns were also observed in the analyses on datasets with lower *b*-values.

When taking into account results from all metrics together, as defined by the *GRP* measure (see Tables 3 and 2 in the Supplementary Material), the SBL method showed the best global reconstructions for the three *b*-values investigated. SBL showed a top performance for all individual metrics: it always produced the lowest n^+ and θ, and was in the top three in terms of Δf and *SR*. Cauchy-SD and RFOCUSS completed the global podium.

In order to visually confirm the high-quality results provided by SBL at high and low *b*-values, the distribution of detected fibers as a function of the inter-fiber angle is plotted in Fig. 2. Notably, the resolution power of SBL was not significantly affected when using a *b*-value as low as $b = 1000$ s/mm^2. Results from both *b*-values were nearly similar for inter-fiber angles higher than 40°. Although at lower inter-fiber angles results for $b = 3000$ s/mm^2 were better, at $b = 1000$ s/mm^2 SBL was able to detect fiber crossings at 30° with a probability higher than 50%.

5.3. Two-fibers with a very dominant fiber

Table 4 summarizes the results of the evaluation for $b = 3000$ and 1500 s/mm^2. Results corresponding to $b = 1000$ s/mm^2 are shown in Table 3 of the Supplementary Material. In this experiment, the model selection techniques NNL-BSS-EBIC and LASSO-EBIC were remarkably the best methods. LASSO-EBIC was the best for $b = 3000$ s/mm^2 and the second best for lower *b*-values. NNL-BSS-EBIC showed superior performance at $b = 1500$ s/mm^2 and $b = 1000$ s/mm^2 and ranked in the top 3 at $b = 3000$ s/mm^2. AdRL-SD and ARUMBA-SD were competitive at $b = 3000$ s/mm^2 but failed to produce good results at lower *b*-values. Among the iterative reweighted methods, SBL was the only one producing relatively good results: it ranked #5 for $b = 3000$ s/mm^2 and #3 for the lower *b*-values. Overall, the performance of all methods was inferior to that reported in previous experiments, suggesting the difficulty of this benchmark test. The main source of error was related to the detection of a single fiber with orientation located between the two simulated fibers, thereby producing higher n^-, n^+ and θ, and lower *SR* values.

5.4. Three-fibers

Findings from the experiments simulating voxels with three fibers and different inter-fiber angles are reported in Fig. 3 and Table 5. Fig. 3 shows results for $b = 1500$ s/mm^2 and Table 5 summarizes results for $b = 3000$ and 1500 s/mm^2. Table 4 of the Supplementary Material depicts results for $b = 1000$ s/mm^2.

IRL1 produced the best global reconstructions for the three *b*-values investigated (see Table 5 and Table 4 in the Supplementary Material), systematically obtaining the lowest n^- and Δf metrics and the highest *SR* while securing the first or second best performance in terms of θ and n^+. The global performance of all methods was stable across *b*-values. The other iterative reweighted methods, namely SBL and RFOCUSS, were the second and third best, respectively.

When analyzing performance as a function of the inter-fiber angle we obtained the following main results. First, model selection methods like LASSO-EBIC and NNL-BSS-EBIC have a good performance in resolving

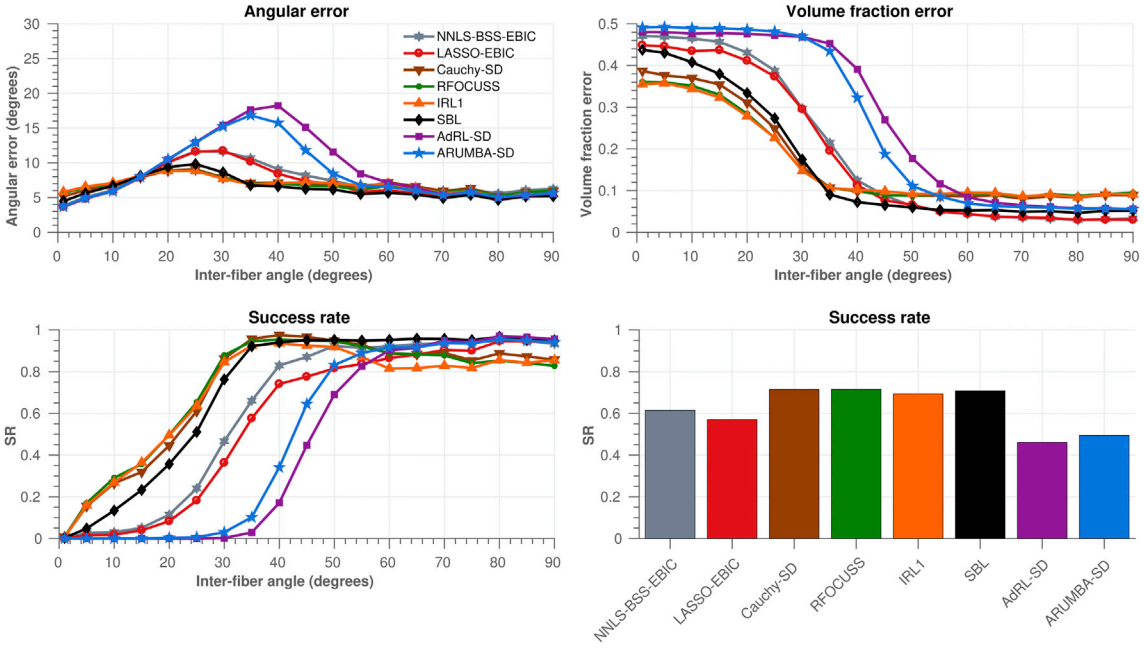

Fig. 1. Reconstruction accuracy of all methods in synthetic data simulating voxels with two fibers at different inter-fiber angles. The upper panel shows results from the angular error and the volume fraction error metrics. The lower panel depicts the success rate (SR), as a function of the inter-fiber angle on the left, and the mean *SR* over the whole range of angles on the right. Continuous lines in each plot represent the mean values for each method. Analyses correspond to the dataset generated with $b = 3000$ s/mm^2.

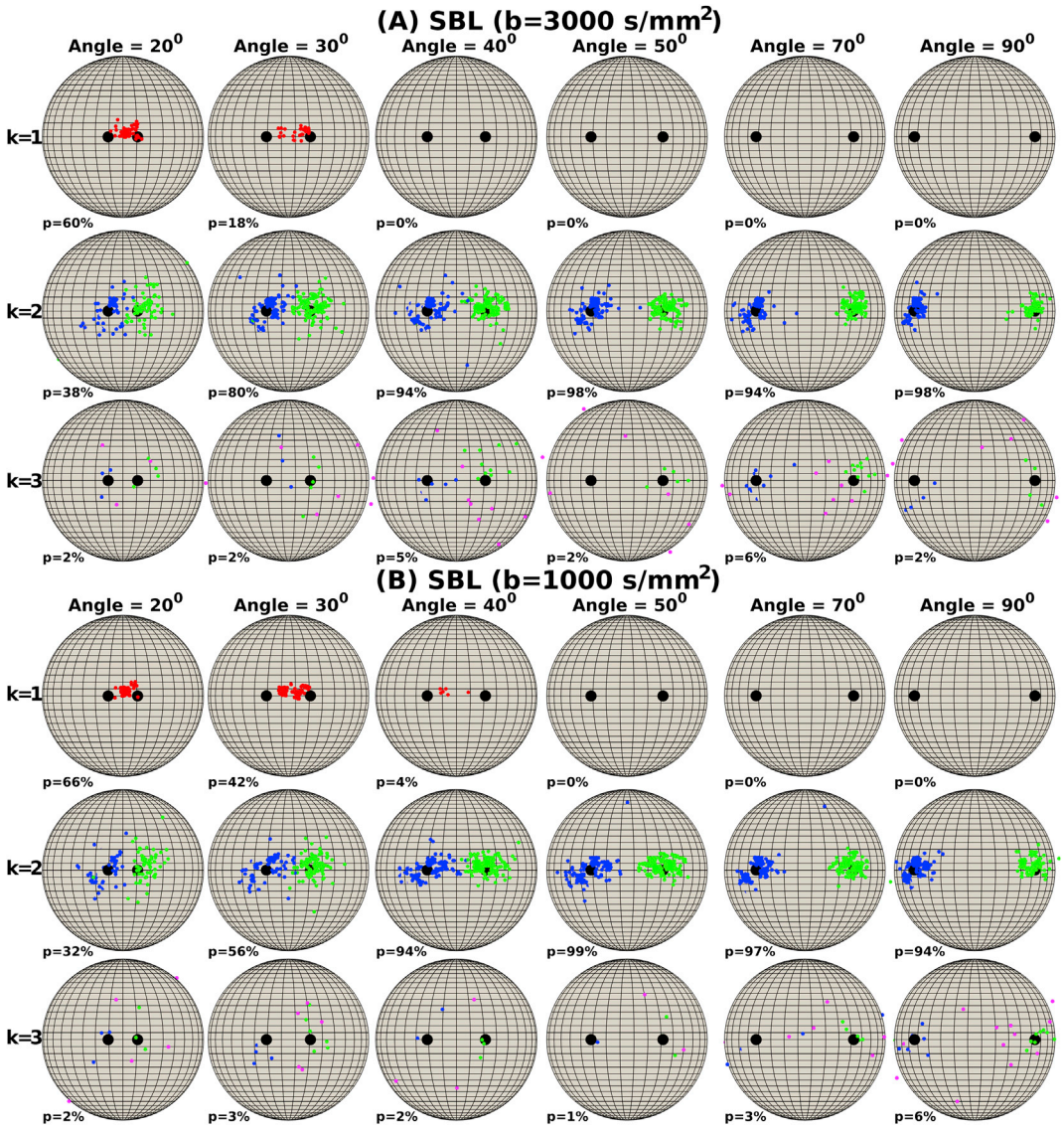

Fig. 2. Distribution of fibers on the sphere detected by SBL at different inter-fiber angles in voxels with two fibers. To improve clarity in the images, 200 voxels were randomly chosen and a subset of representative angles (from left to right, 20, 30, 40, 50, 70 and 90°) was selected. The upper panel (A) shows results for $b = 3000$ s/mm^2 and the lower panel (B) for $b = 1000$ s/mm^2. Each panel includes three rows. The upper one shows the peaks in voxels where a single fiber, k = 1, was detected (red color). The middle row corresponds to voxels where two fibers, k = 2, were detected. A different color was assigned to each fiber (i.e., blue or green) as a function of the distance to each true fiber. Finally, the lower row shows the peaks from voxels where three fibers, k = 3, were detected. The two peaks closest to the true fibers were colored in blue and green (as in the middle row) and the third spurious fiber is shown in magenta. For each inter-fiber angle, the percentage p of voxels with k = 1, 2 and 3 peaks is shown. As a reference, the true fibers are plotted as thicker black markers.

fiber crossings with small inter-fiber angles (see Fig. 3). Second, ARUMBA-SD and AdRL-SD were able to resolve fiber crossings with high *SR* at the higher inter-fiber angles. Third, the performance of all methods, except IRL1 and SBL, significantly declined at $b = 1000$ s/mm^2 yielding mean probabilities of correct recovery below 40% (i.e., $SR < 0.4$). IRL1 was the method showing higher *SR* for inter-fiber angles below 55° and SBL exhibited better *SR* scores at higher angles.

Fig. 4 shows the distribution of fibers detected by IRL1 and SBL as a function of the inter-fiber angle at $b = 1000$ s/mm^2. For inter-fiber angles higher than 40° both methods were able to identify the true fibers with a high probability. Note that although for the higher angles both methods detected k = 3 fibers in 100% of cases, the actual *SR* values were lower than 1 because some fibers were located outside the tolerance cone.

5.5. Phantom from the HARDI reconstruction workshop 2012

The performance of all algorithms in the HARDI 2012 3D phantom is summarized in Table 6 for each of the three noise levels investigated (i.e., SNR = 30, 20 and 15). SBL produced the best global reconstructions for the three SNRs, closely followed by the other iterative reweighted methods (i.e., Cauchy-SD, RFOCUSS, and IRL1). The superior performance of SBL in terms of *SR* was due to a lower n^+ score. In contrast, the other reweighted methods depicted slightly lower n^- values. The global ranking of the methods changed for each SNR value. Cauchy-SD showed the second best performance at SNR = 30 and 20, and IRL1 at SNR = 15.

Notably, despite the fact that DSI_{LR} was a top method in the HARDI reconstruction workshop 2012 and that it employs a higher number of measurements and higher b-values, it ranked among the lower three-five in our new experiments. Indeed, DSI_{LR}, CSD, Ball & Stick, AdRL-SD and ARUMBA-SD, which are among the best current intra-voxel techniques, produced the worst results for this phantom. This suggests that the sparse methods implemented in this study can be considered as state-of-the-art techniques, and may provide similar or even better reconstructions than some of the available intra-voxel algorithms. At SNR = 30, the *SR* values

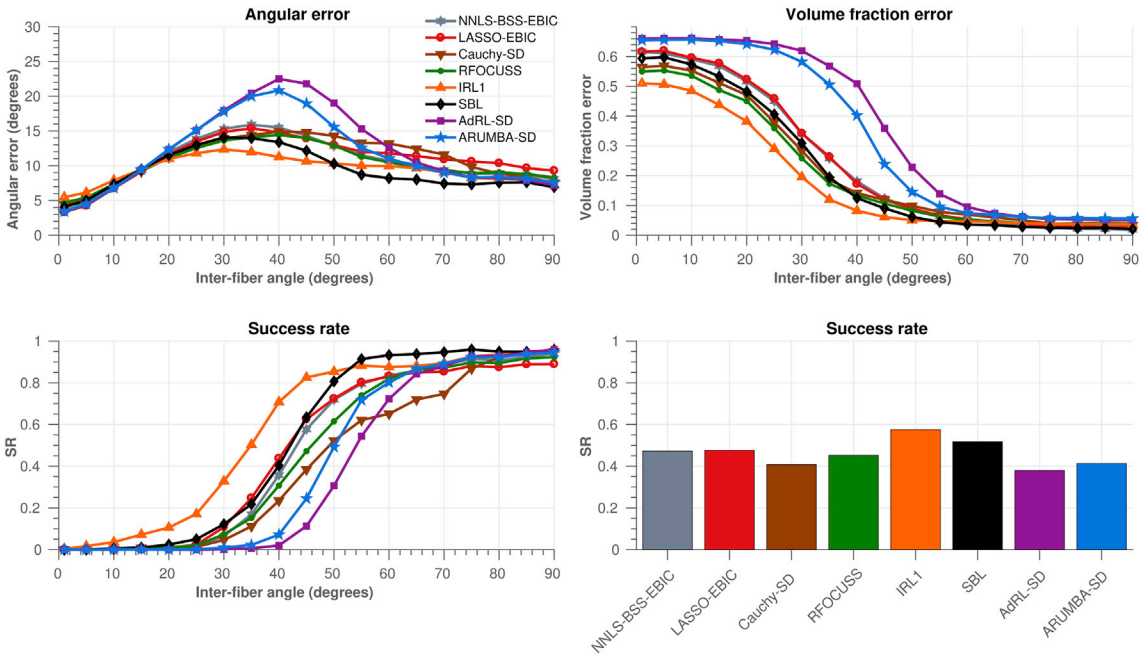

Fig. 3. Reconstruction accuracy of all methods in synthetic data simulating voxels with three fibers at different inter-fiber angles. The upper panel shows results from the angular error and the volume fraction error metrics. The lower panel depicts the success rate (SR), as a function of the inter-fiber angle on the left, and the mean *SR* in the whole range of angles on the right. Continuous lines in each plot represent the mean values for each method. Analyses correspond to the dataset generated with $b = 1500$ s/mm^2.

reported for CSD (Tournier et al., 2007) and QBI-CSA (Aganj et al., 2010) in the reconstruction workshop were lower than 0.55 (see Fig. 2 in (Daducci et al., 2014a)). All the methods included in our study produced higher *SR* values even at higher noise levels (see Table 6). The reader should note that in our evaluation CSD obtained higher *SR* values than those reported earlier (i.e., 0.57–0.62 in Table 6) because this method was evaluated using a set of regularization parameters specially optimized for this phantom.

Fig. 5 shows the true fibers and the fibers detected by SBL, RFOCUSS, and DSI_{LR} in one slice of the phantom with SNR = 20. The arrow in orange highlights a single-fiber region. All methods were able to detect this fiber bundle producing a very low number of spurious fibers. The black arrow emphasizes a fiber crossing region with inter-fiber angles smaller than 45°. In this region, both SBL and RFOCUSS detected the two fibers in nearly all voxels, while DSI_{LR} succeeded only in a few voxels. The upper blue arrow points to a more complex fiber crossing region containing three fibers, where the angle between two of them was small. In these voxels, all methods failed to recover the true configurations (although in some cases RFOCUSS detected the three fibers). In most voxels they identified two fibers: the fiber crossing with smaller inter-fiber angle was detected as a single fiber. Finally, the central red arrow highlights a fiber crossing region with very small inter-fiber angles. In that area, both SBL and RFOCUSS were able to detect fiber crossings in a larger number of cases in comparison to DSI_{LR}. We visually confirmed that, overall, SBL detected a lower number of spurious fibers than RFOCUSS, but also a higher number of under-estimated fibers.

5.6. Phantom from the HARDI reconstruction workshop 2013

Table 7 summarizes the performance of all algorithms in the HARDI 2013 phantom for each noise level (i.e., SNR = 30, 20 and 10). At SNR = 30 and 20, NNLS-BSS-EBIC provided the best global reconstructions, followed by LASSO-EBIC. NNLS-BSS-EBIC produced the lowest n^+ values for all SNRs. Notably, the lowest θ and n^- values were achieved by AdRL-SD, except at SNR = 30, where Ball & Stick produced the lower θ. However, AdRL-SD also produced the highest n^+. A more equilibrated balance between n^+ and n^- was achieved by ARUMBA-SD, which was the third best global method at SNR = 30 and 20. For these SNR values, the reweighted methods produced the worst reconstructions. Among the class of reweighted methods, SBL was the most accurate. In contrast, the performance of all methods significantly declined for the data with SNR = 10, except for IRL1, which produced the best results. The performance of the Ball & Stick method was stable, it was the fourth best global method for all SNRs, while SS3T-SD ranked in the lower half in this experiment.

Fig. 6 depicts the true fibers and the fibers detected by NNLS-BSS-EBIC, SBL, and ARUMBA-SD in a region of interest of the phantom with SNR = 30. The arrow in orange highlights a single-fiber region where all methods were able to detect the fiber bundle. The upper black arrow points to a fiber crossing region with inter-fiber angles of 35–40°. The three methods were successful in this region, although SBL provided the most accurate estimates. The red and blue arrows emphasize a complex fiber crossing region containing dominant and non-dominant fibers. SBL was the only method able to detect fiber crossings with very small inter-fiber angles (see voxels close to the red arrow) however, it failed at identifying the smallest fibers (i.e., in blue). In contrast, NNLS-BSS-EBIC and ARUMBA-SD detected the non-dominant fibers in most cases, being NNLS-BSS-EBIC the method with the best recovery. The superior global performance showed by NNLS-BSS-EBIC in Table 7 is consistent with the visual inspection of the results.

6. Discussion

In this study, we have continued the quest towards optimal algorithms for spherical deconvolution by implementing and testing eight different techniques. It is important to recall that all methods were implemented using the same generative diffusion model, the same linear deconvolution framework, and the same dictionary **H**. Hence, results entirely depend on the chosen objective/regularization functions and optimization algorithms.

From our exhaustive evaluation, we can outline some general conclusions. First, none of the methods outperformed the others in all

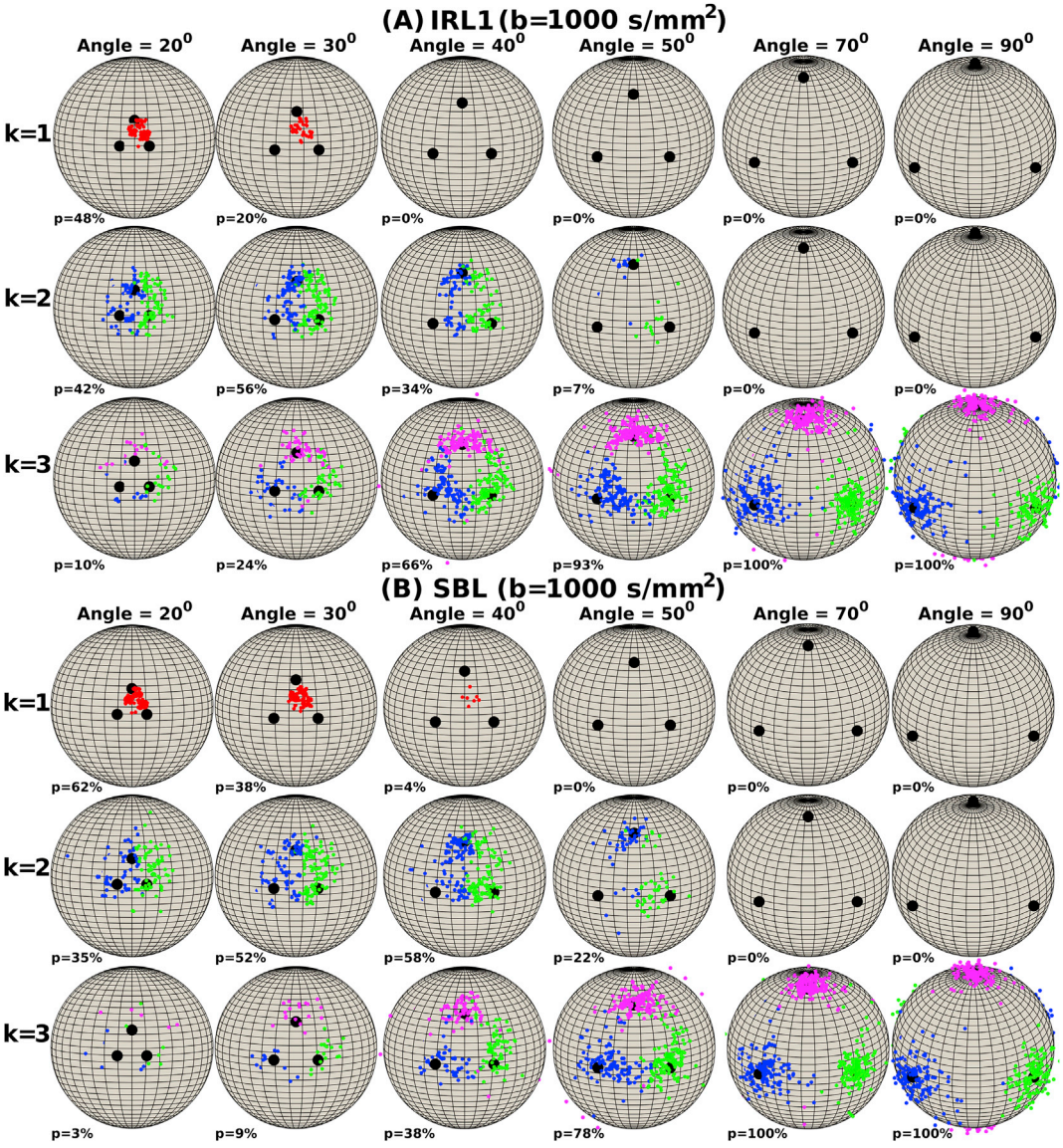

Fig. 4. Distribution of fibers on the sphere detected by IRL1 and SBL at different inter-fiber angles in voxels with three fibers and b=1000 s/mm^2. To improve clarity in the images, 200 voxels were randomly chosen and a subset of representative angles (from left to right, 20, 30, 40, 50, 70 and 90°) was selected. The upper panel (A) shows results for IRL1 and the lower panel (B) for SBL. Each panel includes three rows. The upper row shows the peaks in voxels where a single fiber, k = 1, was detected (red color). The middle row corresponds to voxels where two fibers, k = 2, were detected. Finally, the lower row shows the peaks from voxels with three fibers, k = 3. The three peaks closest to the true fibers were colored in blue, green and magenta, respectively, as a function of the distance to each true fiber. For each inter-fiber angle, the percentage p of voxels with k = 1, 2 and 3 peaks is shown. As a reference, the true fibers are plotted as thicker black markers.

experimental conditions. Second, by looking at the results for different fiber configurations we concluded that different classes of methods were optimal in each case. Specifically, the accelerated multiplicative Richardson-Lucy algorithms ARUMBA-SD and AdRL-SD were the best methods in voxels with a single fiber. This result is consistent with a previous finding reporting a better performance for dRL-SD compared to CSD in single fiber regions (Parker et al., 2013). In voxels with two fibers, the family of reweighted algorithms produced superior results, and among them, SBL produced the best reconstructions. In contrast, in voxels with two fibers where one fiber was very dominant, the model selection techniques NNLS-BSS-EBIC and LASSO-EBIC outperformed the rest of the methods. In voxels with complex configurations of three fibers, IRL1 was the best method followed by SBL. It is interesting to note that the latter finding is not in agreement with results from a recent comparison between the algorithms dRL-SD and IRL1 (Mastropietro et al., 2017). In that study, the authors reported a better performance for dRL-SD than for IRL1 in voxels with three fibers. A possible explanation for this discrepancy is the use of a different IRL1 algorithm in our study, i.e., the regularized version.

The evaluation of the two datasets from the HARDI challenges revealed that the performance of all methods was different in each case. Notably, in the HARDI 2012 phantom which mainly contains fiber crossings with two or three fibers with similar volume fractions, SBL produced the best results. In contrast, in the HARDI 2013 phantom where 89% of voxels contains a single fiber and the proportion of fiber crossings with dominant fibers is high (i.e., about 50%), NNLS-BSS-EBIC provided the most accurate reconstructions. This result, together with the results described above, suggests that in practical applications the optimal algorithm depends on the intravoxel complexity of the sample, and more specifically on the proportion of voxels with one, two or more fibers and the proportion of voxels with very dominant fibers. Therefore, future studies should be conducted to find the optimal way of combining

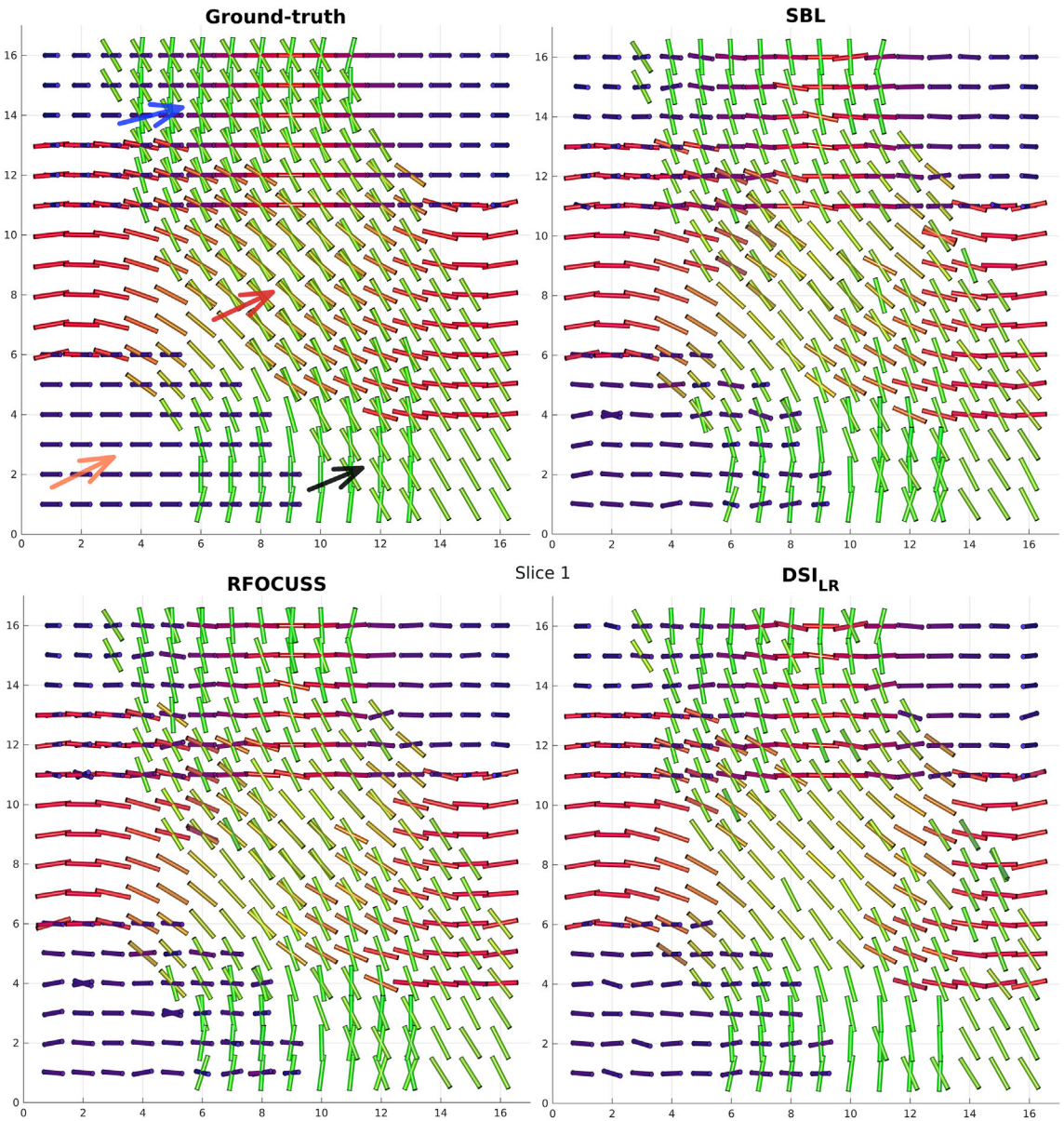

Fig. 5. True fibers and the fibers detected by SBL, RFOCUSS, and DSI$_{LR}$ in the HARDI 2012 phantom. These results correspond to the slice #1 from the dataset with SNR = 20. SBL and RFOCUSS were evaluated using a sampling scheme with 60 diffusion sensitizing gradients and $b = 3000$ s/mm^2. A DSI acquisition protocol with 257 diffusion sensitizing gradients and $b_{max} = 8000$ s/mm^2 was used to evaluate DSI$_{LR}$. Main peaks extracted from the estimated *FOD*s are visualized as cylinders. The orange arrow (bottom left) highlights a single-fiber region; the black arrow (bottom right) emphasizes a fiber crossing region with inter-fiber angles smaller than 45°; the blue arrow (top) points to a more complex fiber crossing region containing three fibers, and the red arrow (center) highlights a fiber crossing region with very small inter-fiber angles.

solutions from different methods, or to identify the most reliable algorithm for each brain region, for instance, by using machine learning techniques.

Table 8 provides a qualitative overview of the main properties of the eight evaluated techniques, including the assumed noise model, regularization function, optimization algorithm used to minimize the resulting cost function, and reconstruction time per voxel. The table also reports a brief description of the experimental situations in which a specific method showed the most prominent advantages and disadvantages. Based on our results, we recommend the use of the following algorithms: SBL, NNLS-BSS-EBIC, IRL1, and ARUMBA. SBL showed a stable performance for all the fiber configurations and *b*-values, particularly in fiber crossings with two and three fibers. As discussed above, NNLS-BSS-EBIC excelled at solving fiber crossings with dominant fibers; IRL1 was the best method to solve fiber crossings with three fibers and ARUMBA-SD produced a very low number of spurious fibers in voxels with a single fiber population. Although other methods also performed well for each one of these fiber configurations, those mentioned above were more stable across all experimental conditions.

This is not the first time that an SBL algorithm is implemented to solve the spherical deconvolution inverse problem. A recent work proposed a data fusion framework based on SBL that requires diffusion datasets sampled at different spatial resolutions (Pisharady et al., 2017a). The implementation of that method, termed BusineX, is based on a Reversible Jump Gibbs sampler similar to that introduced by (Melie-García et al., 2008). Interestingly, the authors reported a superior performance of BusineX, in terms of the *SR* metric, with respect to algorithms evaluated in the HARDI 2012 challenge. Specifically, BusineX provided *SR* values of

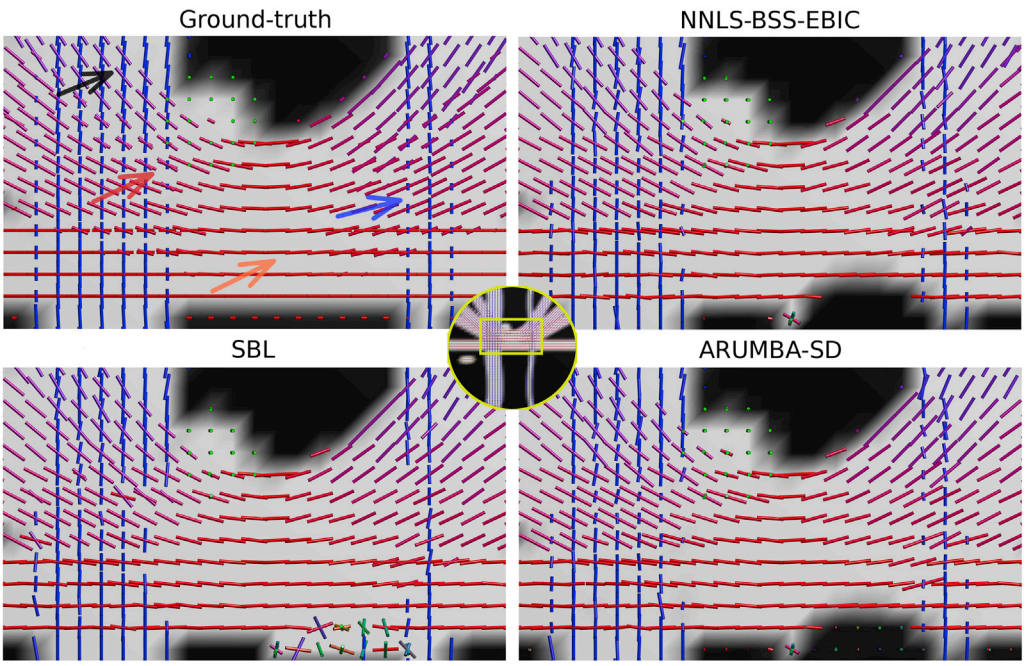

Fig. 6. True fibers and the fibers detected by NNSL-BSS-EBIC, SBL, and ARUMBA-SD in a region of interest of the HARDI 2013 phantom. These results correspond to the dataset with SNR = 30. Main peaks extracted from the estimated *FODs* are visualized as cylinders. The orange arrow (bottom) highlights a single-fiber region; the black arrow (top) points to a fiber crossing region with inter-fiber angles of 35–40°; and the red and blue arrows (center) emphasize a complex fiber crossing region containing dominant and non-dominant fibers, respectively.

82.99% and 80.36% for SNRs of 24.3 and 10, respectively (Pisharady et al., 2017a). As the reported *SR* metric is based on a different definition than that used in our study, we recomputed this metric for our own SBL implementation and obtained *SR* values of 88.33%, 87.13% and 85.51% for the datasets with SNRs of 30, 20 and 15, respectively. Results from both studies are not directly comparable because the evaluations were performed using different SNRs and sampling schemes (i.e., $N = 50$ and $b = 1500$ s/mm^2 in BusineX; $N = 60$ and $b = 3000$ s/mm^2 in SBL). However, this analysis shows that our SBL implementation is competitive.

The performance of each method was quantified in terms of individual standard metrics, including the angular error between the true and estimated fibers, the volume fraction error, the success rate and the number of overestimated and underestimated fibers. In order to characterize the global performance of all methods and identify the optimal ones, we combined all indicators and defined a new overall multimetric index (i.e., the global relative performance (*GRP*)). A limitation of that approach is related to the fact that different multimetric measures could be defined, and each one may produce a slightly different global ranking of methods. Yet, little is known about the relative impact of these metrics to predict optimal fiber tracking results. For this reason, in the definition of the *GRP* metric, the same weight was assigned to each individual metric. Future studies should be carried out to characterize the importance of each individual metric in different applications. Despite this limitation, we observed that methods with better *GRP* scores also showed consistently better individual metrics and better visual solutions. In any case, as all the considered metrics are available in Tables 2–7, the reader could use this information to find the optimal algorithm for each individual metric or to create a new global metric based on particular needs.

In the phantoms from the HARDI reconstruction challenges, various sparse techniques implemented in this study performed better than some popular reconstruction methods, including CSD, Ball & Stick, and dRL-SD. However, as CSD and Ball & Stick are based on different generative models (i.e., different single-fiber response functions) it is not possible to know to which extent our results are more influenced by discrepancies between the generative models and the ones used to create the synthetic signals, or by differences in the cost functions and optimization algorithms. Therefore, these observations cannot be generalized to other datasets. The purpose of including CSD and Ball & Stick in the analyses is to provide the reader with a reference about the performance of the implemented methods in relation to these state-of-the-art techniques that are being extensively used in clinical applications.

Whether these results could be generalized to fiber tracking in clinical data remains to be established. The main difficulty is related to the fact that tractography is also an ill-posed problem (Daducci et al., 2016) and current algorithms are not accurate (Thomas et al., 2014). For instance, in the synthetic data produced for the ISMRM 2015 tractography challenge various fiber tracking algorithms reconstructed a high number of false positives tracts even when they used the ground truth field of fiber orientations (Maier-Hein et al., 2017). Another reason is related to the complexity of the considered fiber configurations. In this study, it was not possible to simulate the variability and richness of the patterns observed in real brains, including the microscopic orientation dispersion and undulation of axons, as well as the fanning and bending fiber geometries at the voxel level. Indeed, to the best of our knowledge, these factors have not been previously studied in the context of spherical deconvolution. Nevertheless, it is important to recall that our synthetic data were generated using a range of diffusivities and fractional anisotropies similar to those measured in different brain regions. As these diffusivities depend on the fiber dispersion and other microscopic features, our simulated datasets were implicitly affected by these factors. Nonetheless, future studies should be conducted to study the effects of dispersion and microstructure in a more controlled manner. This could be done by using advanced microstructure diffusion MRI generative models and realistic axon distributions measured from histology. Additional evaluation metrics suited for such complex configurations will be also required.

It is important to remark that the results presented in this work are based on numerical simulations. Therefore, there is no guarantee that these observations hold true to the same extent in human brain data. For instance, in fiber bundles with a high degree of dispersion, SD methods promoting distributed solutions (e.g., CSD, Ball & Stick, AdRL-SD, and ARUMBA-SD) may show a better performance than methods promoting sparsity. The sparse methods will return a sparse approximation of the fiber orientations, and thus, it may be necessary to re-adjust the algorithms for such configurations (e.g., the sparsity term can be attenuated by decreasing the regularization parameter). On the other hand, the analyses performed in this work are based on single-shell diffusion MRI data, which allowed us to model the white matter compartments but not the contamination from other tissue types, like the partial volume effects

that occur at the cortical and subcortical gray and white matter boundaries. Despite these limitations, this study allowed us to spot the strengths and weaknesses of each approach and to draw interesting considerations on the optimal performance for different classes of algorithms. This information may be used to improve the current reconstruction techniques, and as a guide for identifying the most adequate methods for a particular application.

7. Conclusions

In summary, we implemented and evaluated eight regression algorithms to solve the SD problem. Results from our exhaustive evaluation show that there is no single optimal method for all different fiber configurations, suggesting that further studies should be conducted to find the optimal way of combining solutions from different methods. We found l_0-norm regularization algorithms to resolve more accurately fiber crossings with small inter-fiber angles. However, in voxels with very dominant fibers, algorithms promoting more sparsity encountered more problems in detecting smaller fibers. The best algorithm to reconstruct fiber crossings with two fibers did not perform optimally in voxels with one or three fibers. Therefore, the simplified validation system employed in a number of previous studies, where only two fibers with similar volume fractions were tested, should be avoided as it provides incomplete information. Future studies proposing new SD methods based on HARDI data should validate their results by considering, at least, voxels with one, two, and three fibers, as well as voxels with dominant fibers and different diffusion anisotropies and dispersion. The techniques implemented in this study will be available at https://github.com/ejcanalesr/sparse-wars.

Acknowledgments

This work was supported by the Catalonian Government (2017-SGR-1271 to FIDMAG) and several grants from the Plan Nacional de I+D+i 2008–2011 and 2013–2016, and the Instituto de Salud Carlos III and co-funded by European Union (ERDF/ESF, "Investing in your future"): Miguel Servet Research Contracts (CPII13/00018 to RS, MS14/00041 to JR and CPii16/00018 to EP-C) and Research Project Grants (PI14/01151 to RS, PI14/01148 to EP-C, PI14/00292 and CP14/00041 to JR, and PI15/00277 to EC-R). Gaëtan Rensonnet is a research fellow of the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS). This work will be included as part of Jon Haitz Legarreta's doctoral thesis.

Appendix

In this section, we provide further technical details about the implemented algorithms.

Best-subset selection

Here we propose a practical implementation of the non-negative best-subset selection problem. In a first step, the NNLS algorithm is used to reduce the dimension of the data, from which all the variables with a value equal to zero are discarded. Then, a combinatorial search is carried out over the retained variables to create all possible individual models M_k comprising $k = 1, 2$ and 3 fibers, plus the isotropic compartment, if it is selected in the first step. Each M_k model is created by considering every time a different subset of k columns of $\mathbf{H}$, and the corresponding parameters are estimated using a standard least squares method. As we are searching for non-negative solutions, all models with negative parameters are excluded from the comparison. Finally, all models are ranked according to the extended Bayesian information criterion (EBIC) (Chen and Chen, 2008) and the model with the lowest EBIC is selected. The EBIC functional is given by (Chen and Chen, 2012)

$$EBIC(M_k) = n \log\left(\frac{1}{n} RSS(M_k)\right) + k \log(n) + 2k\gamma \log(P), \tag{23}$$

where $\gamma = [0, 1]$ is a parameter controlling the penalization on complex models and RSS is defined by

$$RSS(M_k) = \sum_{i=1}^{n} \left(\mathbf{s}_i - \sqrt{\left(\sum_{j \in M_k} \mathbf{H}_{ij} \widehat{\mathbf{f}}_j\right)^2 + \sigma^2} \right)^2, \tag{24}$$

which results from the Gaussian approximation of a Rician distribution (Andersen, 1996). The vector $\widehat{\mathbf{f}}$ contains the k estimated coefficients, σ^2 is the expected noise variance and P is equal to the total number of tested models.

Model selection via the LASSO

The non-negative LASSO problem was solved using the modified least-angle regression (LARS) algorithm (Efron et al., 2004) included in the SparseLab Toolbox (https://sparselab.stanford.edu/), which computes the entire regularization path. The optimal λ_1 was chosen according to the EBIC criterion reported in Eqs. (23) and (24). It is important to note that the EBIC functional depends on k but not on λ_1. However, it is possible to establish a one-to-one mapping between each value of λ_1 and the model M_k defined by the k predictors (i.e., non-zero coefficients) detected in the solution for that regularization parameter (Zou et al., 2007). Before the EBIC evaluation the coefficients were 'debiased', i.e., the solution is recomputed using a standard least squares method over the identified non-zero coefficients. The regularization path was determined by the software in the predefined range of values: $\lambda_1 \in [\lambda_{\min}, \lambda_{\max}]$, with $\lambda_{\min} = 10^{-4}\lambda_{\max}$ and $\lambda_{\max} = \left\|\mathbf{H}^T\mathbf{s}\right\|_\infty$ (the upper limit $\lambda_1 = \lambda_{\max}$ produces the trivial solution $\mathbf{f} = 0$).

Cauchy deconvolution

The formulation is based on assuming a Gaussian Likelihood model

$$P(\mathbf{S}|\mathbf{f}, \sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{s} - \mathbf{H}\mathbf{f}\|_2^2 \right), \tag{25}$$

and a Cauchy prior distribution over $\mathbf{f}$

$$P(\mathbf{f}|\sigma_c) = \prod_{i=1}^{m+1} \frac{1}{2\pi\sigma_c^2} \frac{1}{\left(1+\frac{\mathbf{f}_i^2}{2\sigma_c^2}\right)}. \tag{26}$$

The scale parameter σ_c controls the level of sparsity that can be attained by the inversion. When the prior and the likelihood are combined, the cost function to be minimized is given by Eq. (15). The maximum a posteriori estimate is obtained by differentiating the cost function with respect to $\mathbf{f}$ and equating the derivative to zero (Sacchi and Ulrych, 1996). The solution to the resulting non-linear system of equations can be found by iteratively computing the solution given in Eqs. (16) and (17).

This algorithm does not include non-negativity constraints, so we propose to solve the problem in terms of a new auxiliary vector $\boldsymbol{\mu}$ that satisfies $\mathbf{f}_j = \boldsymbol{\mu}_j^2, j = 1,...m+1$. This substitution allows us to obtain non-negative estimates. The latter can be written in matrix form as $\mathbf{f} = \mathbf{U}\boldsymbol{\mu}$, where $\mathbf{U} = diag(\boldsymbol{\mu})$ is a diagonal matrix (i.e., $\mathbf{U}_{jj} = \boldsymbol{\mu}_j$). The above parameterization was proposed by (Ilya, 2014) for obtaining non-negative solutions within the Tikhonov regularization framework. The final modified reweighted algorithm is given by the updates

$$\begin{aligned}
&\boldsymbol{\mu}^{(k+1)} = \left(\mathbf{U}^{(k)}\right)^{-1}\mathbf{Q}\mathbf{H}^T(\mathbf{H}\mathbf{Q}\mathbf{H}^T + \lambda_c\mathbf{I})^{-1}\mathbf{s},\\
&\mathbf{U}^{(k+1)} = diag\left(\boldsymbol{\mu}^{(k+1)}\right),\\
&\mathbf{f}^{(k+1)} = \mathbf{U}^{(k+1)}\boldsymbol{\mu}^{(k+1)},\\
&\mathbf{Q} = diag\left(1+\frac{1}{2}\left(\frac{\mathbf{f}^{(k+1)}}{\sigma_c}\right)^2\right).
\end{aligned} \tag{27}$$

Non-negative regularized FOCUSS

The original FOCUSS described by Eq. (14) does not include non-negativity constraints, so we adapted it by using the same parameterization used in the previous section (i.e., $\mathbf{f} = \mathbf{U}\boldsymbol{\mu}$). The resulting Non-Negative Regularized FOCUSS (RFOCUSS) algorithm, for $p = 0$, is given by the updates

$$\begin{aligned}
&\boldsymbol{\mu}^{(k+1)} = \left(\mathbf{U}^{(k)}\right)^{-1}\mathbf{W}^{(k)}\left(\mathbf{A}^{(k)}\right)^T\left(\mathbf{A}^{(k)}\left(\mathbf{A}^{(k)}\right)^T + \lambda_0\mathbf{I}\right)^{-1}\mathbf{s},\\
&\mathbf{U}^{(k+1)} = diag\left(\boldsymbol{\mu}^{(k+1)}\right),\\
&\mathbf{f}^{(k+1)} = \mathbf{U}^{(k+1)}\boldsymbol{\mu}^{(k+1)},\\
&\mathbf{W}^{(k+1)} = diag\left(\mathbf{f}^{(k)} + \varepsilon\right),\\
&\mathbf{A}^{(k+1)} = \mathbf{H}\mathbf{W}^{(k+1)}.
\end{aligned} \tag{28}$$

Sparse Bayesian Learning

The SBL formulation is based on assuming a Gaussian Likelihood model, see Eq. (25), and a zero-mean Gaussian prior distribution over $\mathbf{f}$

$$P(\mathbf{f}|\alpha) = \prod_{j=1}^{m+1}\left(2\pi\alpha_j\right)^{-\frac{1}{2}}\exp\left(-\frac{\mathbf{f}_j^2}{2\alpha_j}\right), \tag{29}$$

with Gamma distributions for the hyperparameters $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{m+1}]$ and σ^2 (Tipping, 2001). Each independent α_j moderates the strength of the prior on each entry $\mathbf{f}_j$. Our implementation is motivated by previous findings (Wipf and Nagarajan, 2010) showing that the maximum a posteriori SBL solution can be computed using Eqs. (18) and (19). The hyperparameters are computed via type-II maximum likelihood using the Expectation-Maximization algorithm (Wipf and Nagarajan, 2010)

$$\boldsymbol{\alpha}^{(k+1)} = \tilde{\mathbf{f}}^2 + \boldsymbol{\varepsilon}, \quad \text{where} \boldsymbol{\varepsilon} = \boldsymbol{\alpha}^{(k)} - \left(\boldsymbol{\alpha}^{(k)}\right)^2 \circ diag\left(\mathbf{H}^T(\mu I + \mathbf{H}\mathbf{A}\mathbf{H}^T)^{-1}\mathbf{H}\right)^{-1}, \tag{30}$$

where '$\circ$' stands for the Hadamard component-wise multiplication. The regularization term is defined as $\lambda = c\sigma^2$, with c being a scale constant and σ^2 is computed as in (Wipf and Rao, 2004)

$$\sigma^2 = \frac{1}{n}\left\|\mathbf{s} - \mathbf{H}\tilde{\mathbf{f}}\right\|_2^2 + \frac{1}{n}\sigma^2\sum_{j=1}^{m+1}\left(1 - \frac{\boldsymbol{\varepsilon}_j}{\boldsymbol{\alpha}_j^{(k+1)}}\right). \tag{31}$$

As all $\tilde{\mathbf{f}}$, $\boldsymbol{\alpha}$ and σ^2 depend on each other, they are computed using an alternating iterative scheme. In line with the non-negative parameterization used in previous sections, our non-negative SBL implementation is defined by the updates

$$\begin{aligned}
&\boldsymbol{\mu}^{(k+1)} = \left(\mathbf{U}^{(k)}\right)^{-1}\mathbf{A}\mathbf{H}^T\left(\mathbf{H}\mathbf{A}\mathbf{H}^T + \lambda^{(k)}\mathbf{I}\right)^{-1}\mathbf{s},\\
&\mathbf{U}^{(k+1)} = diag\left(\boldsymbol{\mu}^{(k+1)}\right),\\
&\mathbf{f}^{(k+1)} = \mathbf{U}^{(k+1)}\boldsymbol{\mu}^{(k+1)},\\
&\boldsymbol{\varepsilon} = \boldsymbol{\alpha}^{(k)} - \left(\boldsymbol{\alpha}^{(k)}\right)^2 \circ diag\left(\mathbf{H}^T\left(\mu I + \mathbf{H}\mathbf{A}\mathbf{H}^T\right)^{-1}\mathbf{H}\right)^{-1},\\
&\boldsymbol{\alpha}^{(k+1)} = \left(\mathbf{f}^{(k+1)}\right)^2 + \boldsymbol{\varepsilon},\\
&\mathbf{A} = diag\left(\boldsymbol{\alpha}^{(k+1)}\right),\\
&\lambda^{(k+1)} = \frac{c}{\mathrm{n}}\left\|\mathbf{s} - \mathbf{H}\mathbf{f}^{(k+1)}\right\|_2^2 + \frac{c}{\mathrm{n}}\lambda^{(k)}\sum_{j=1}^{m+1}\left(1 - \frac{\varepsilon_j}{\alpha_j^{(k+1)}}\right).
\end{aligned} \tag{32}$$

In addition to the global iterations, we have found much better estimates when also including local iterations to update $\boldsymbol{\varepsilon}$, $\boldsymbol{\alpha}$ and $\mathbf{A}$ for each solution $\mathbf{f}^{(k+1)}$. In our implementation, we fixed the number of local and global iterations to be equal.

Variations on the multiplicative Richardson-Lucy algorithm

The Richardson-Lucy (RL) deconvolution algorithm can be constrained to be non-negative without the need for including additional penalization functions or re-parameterizations in the estimation process. A modified version of this algorithm adapted to Gaussian noise was used to estimate $\mathbf{f}$ from diffusion data (Dell'Acqua et al., 2007). The estimation process is based on multiplicative updates, which converge to the NNLS solution. This algorithm was later improved to avoid the amplification of small spurious fibers by adding a damping factor, which resulted in the damped RL-SD method (dRL-SD) (Dell'Acqua et al., 2010)

$$\begin{aligned}
&\mathbf{f}^{(k+1)} = \mathbf{f}^{(k)} \circ \left[1 + \mathbf{u} \circ \left(\frac{\mathbf{H}^T\mathbf{s} - \mathbf{H}^T\mathbf{H}\mathbf{f}^{(k)}}{\mathbf{H}^T\mathbf{H}\mathbf{f}^{(k)}}\right)\right],\\
&\mathbf{u} = \left(1 - \mu\left(1 - \frac{\left(\mathbf{f}^{(k)}\right)^{\nu}}{\left(\mathbf{f}^{(k)}\right)^{\nu} + \eta^{\nu}}\right)\right),
\end{aligned} \tag{33}$$

where $\mu = \max(0, 1 - 4std(\mathbf{s}))$ is a parameter that depends on the standard deviation of the vector of measurements and the Hadamard multiplication and the division operators are applied component-wise to the vector's elements. As the initial estimate $\mathbf{f}^0$ is fixed to a positive uniform density distribution, the subsequent estimates remain positive as well, and the algorithm always produces reconstructions with non-negative elements. From a numerical point of view, the algorithm is stable as it does not involve any matrix inversion (Dell'Acqua et al., 2007).

Recently, it was demonstrated that the RL estimation framework can be extended to more realistic MRI noise models. When applied to Rician distributed noise it led to the following iterative estimation process (Canales-Rodríguez et al., 2015)

$$\begin{aligned}
&\mathbf{f}^{(k+1)} = \mathbf{f}^{(k)} \circ \frac{\mathbf{H}^T\left[\mathbf{s} \circ \frac{I_1\left(\mathbf{s} \circ \mathbf{H}\mathbf{f}^{(k)}/\sigma^2\right)}{I_0\left(\mathbf{s} \circ \mathbf{H}\mathbf{f}^{(k)}/\sigma^2\right)}\right]}{\mathbf{H}^T\mathbf{H}\mathbf{f}^{(k)}},\\
&\sigma^2 = \frac{1}{n}\left\{\frac{\mathbf{s}^T\mathbf{s} + \left(\mathbf{f}^{(k+1)}\right)^T\mathbf{H}^T\mathbf{H}\mathbf{f}^{(k+1)}}{2} - \mathbf{1}^T\left[\left(\mathbf{s} \circ \mathbf{H}\mathbf{f}^{(k+1)}\right) \circ \frac{I_1\left(\mathbf{s} \circ \mathbf{H}\mathbf{f}^{(k+1)}/\sigma^2\right)}{I_0\left(\mathbf{s} \circ \mathbf{H}\mathbf{f}^{(k+1)}/\sigma^2\right)}\right]\right\},
\end{aligned} \tag{34}$$

where $\mathbf{1}$ denotes a column vector of ones with n elements, I_1 and I_0 are the modified Bessel functions of the first kind of order 1 and 0 respectively. The ratio $I_1(x)/I_0(x)$ is efficiently evaluated in terms of the Perron continued-fraction approximation (Gautschi and Slavik, 1978)

$$\frac{I_1(x)}{I_0(x)} = \frac{x}{2 + x - \frac{2x(1+1/2)}{3+2x-\frac{2x(1+3/2)}{4+2x-\frac{2x(1+5/2)}{5+2x}}}}, \tag{35}$$

The resulting algorithm was termed 'Robust and Unbiased Model-Based Spherical Deconvolution (RUMBA-SD)' (Canales-Rodríguez et al., 2015).

In this study, a new variant to accelerate the convergence of these techniques is proposed. It is based on applying ideas from Nesterov's accelerated gradient descent method (Nesterov, 2004) to the multiplicative RL updates. As the RL algorithms described above could be equivalently rewritten as gradient descent methods with additive updates, the strategy developed by Nesterov (and later included in FISTA (Beck and Teboulle, 2009)) can also be incorporated into these algorithms to obtain new solutions computed as

$$\mathbf{b}^{(k+1)} = \left(1 - \gamma^{(k)}\right)\mathbf{f}^{(k+1)} + \gamma^{(k)}\mathbf{f}^{(k)}, \tag{36}$$

where the solution $\mathbf{b}^{(k+1)}$ of the accelerated algorithms depends on the estimates $\mathbf{f}^{(k+1)}$ and $\mathbf{f}^{(k)}$ computed by the original algorithms (i.e., Eq. (33) or Eq. (34)) and the sequence of parameters $\gamma^{(k)}$ is computed as

$$\begin{cases} t^{(1)} = 1 \\ t^{(k+1)} = \left(1 + \sqrt{1 + 4\left(t^{(k)}\right)^2}\right) \Big/ 2 \\ \gamma^{(k)} = \left(1 - t^{(k)}\right) \Big/ t^{(k+1)} \end{cases} \tag{37}$$

Despite the effective acceleration provided by this approach, the resulting algorithm is not descent. It is not uncommon to observe oscillatory patterns, called Nesterov ripples, in the values of the objective function being minimized. To overcome this drawback, O'Donoghue and Candès suggested an adaptive restart strategy that dramatically improved the convergence rate of the algorithm (O'Donoghue and Candès, 2015). It is based on resetting the parameter $t^{(k+1)}$ to 1 and disabling the acceleration step in cases where the iteration does not decrease the objective function. To preserve the non-negativity, the acceleration step can be modified to project to zero any negative value, i.e., $\mathbf{b}^{(k+1)} = \max(\mathbf{b}^{(k+1)}, 0)$. In our study, we have used this strategy with the resulting modified algorithms being termed Accelerated dRL-SD (AdRL-SD) and Accelerated RUMBA-SD (ARUMBA-SD). The accelerated versions produce reconstructions similar to those from the original algorithms but using a much lower number of iterations (e.g., they required approximately 1/4 of the iterations to obtain similar solutions).

Although these algorithms can produce sparse solutions when the number of iterations is chosen to be large, the best performance is obtained with a low number of iterations. This gives rise to a smooth *FOD* profile of amplitudes with a small number of lobes. As usual, a sparse representation is obtained by extracting the peaks from these lobes.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.08.071.

References

Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N., 2010. Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle. Magn. Reson. Med. 64, 554–566. https://doi.org/10.1002/mrm.22365.

Alexander, D.C., 2005. Maximum entropy spherical deconvolution for diffusion MRI. In: Information Processing in Medical Imaging, pp. 76–87. https://doi.org/10.1242/jcs.02651.

Andersen, A.H., 1996. On the Rician distribution of noisy MRI data. Magn. Reson. Med. 36, 331–333.

Anderson, A.W., 2005. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. Magn. Reson. Med. 54, 1194–1206. https://doi.org/10.1002/mrm.20667.

Aranda, R., Ramirez-Manzanares, A., Rivera, M., 2015. Sparse and Adaptive Diffusion Dictionary (SADD) for recovering intra-voxel white matter structure. Med. Image Anal. 26, 243–255. https://doi.org/10.1016/j.media.2015.10.002.

Beck, A., Teboulle, M.F.I.S.-T.A., L.I.P, 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. 2, 183–202. https://doi.org/10.1137/080716542.

Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088. https://doi.org/10.1002/mrm.10609.

Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage 34, 144–155. https://doi.org/10.1016/j.neuroimage.2006.09.018.

Bruckstein, A.M., Donoho, D.L., Elad, M., 2009. From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. 51, 34–81. https://doi.org/10.1137/060657704.

Canales-Rodríguez, E.J., Daducci, A., Sotiropoulos, S.N., Caruyer, E., Aja-Fernández, S., Radua, J., Mendizabal, J.M.Y., Iturria-Medina, Y., Melie-García, L., Alemán-Gómez, Y., Thiran, J.-P., Sarró, S., Pomarol-Clotet, E., Salvador, R., 2015. Spherical deconvolution of multichannel diffusion MRI data with non-Gaussian noise models and spatial regularization. PLoS One 10, e0138910. https://doi.org/10.1371/journal.pone.0138910.

Canales-Rodríguez, E.J., Iturria-Medina, Y., Alemán-Gómez, Y., Melie-García, L., 2010. Deconvolution in diffusion spectrum imaging. Neuroimage 50, 136–149. https://doi.org/10.1016/j.neuroimage.2009.11.066.

Canales-Rodríguez, E.J., Melie-García, L., Iturria-Medina, Y., 2009. Mathematical description of q-space in spherical coordinates: exact q-ball imaging. Magn. Reson. Med. 61, 1350–1367. https://doi.org/10.1002/mrm.21917.

Canales-Rodríguez, E.J., Melie-García, L., Iturria-Medina, Y., Martínez-Montes, E., Alemán-Gómez, Y., Lin, C.-P., 2008. Inferring multiple maxima in intravoxel white matter fiber distribution. Magn. Reson. Med. 60, 616–630. https://doi.org/10.1002/mrm.21673.

Candès, E.J., Romberg, J., Tao, T., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theor. 52, 489–509. https://doi.org/10.1109/TIT.2005.862083.

Candès, E.J., Wakin, M.B., Boyd, S.P., 2008. Enhancing sparsity by reweighted L1 minimization. J. Fourier Anal. Appl. 14, 877–905. https://doi.org/10.1007/s00041-008-9045-x.

Chen, J., Chen, Z., 2012. Extended BIC for small-n-large-P sparse GLM. Stat. Sin. 22, 555–574. https://doi.org/10.5705/ss.2010.216.

Chen, J., Chen, Z., 2008. Extended Bayesian information criteria for model selection with large model spaces. Biometrika 95, 759–771. https://doi.org/10.1093/biomet/asn034.

Cheng, J., Deriche, R., Jiang, T., Shen, D., Yap, P.T., 2014. Non-negative spherical deconvolution (NNSD) for estimation of fiber orientation distribution function in single-/multi-shell diffusion MRI. Neuroimage 101, 750–764. https://doi.org/10.1016/j.neuroimage.2014.07.062.

Daducci, A., Canales-Rodriguez, E.J., Descoteaux, M., Garyfallidis, E., Gur, Y., Lin, Y.C., Mani, M., Merlet, S., Paquette, M., Ramirez-Manzanares, A., Reisert, M., Rodrigues, P.R., Sepehrband, F., Caruyer, E., Choupan, J., Deriche, R., Jacob, M., Menegaz, G., Prckovska, V., Rivera, M., Wiaux, Y., Thiran, J.P., 2014a. Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI. IEEE Trans. Med. Imag. 33, 384–399. https://doi.org/10.1109/TMI.2013.2285500.

Daducci, A., Palú, A.D., Descoteaux, M., Thiran, J.P., 2016. Microstructure informed tractography: pitfalls and open challenges. Front. Neurosci. 10 https://doi.org/10.3389/fnins.2016.00247.

Daducci, A., Van De Ville, D., Thiran, J.P., Wiaux, Y., 2014b. Sparse regularization for fiber ODF reconstruction: from the suboptimality of $\ell 2$ and $\ell 1$ priors to $\ell 0$. Med. Image Anal. 18, 820–833. https://doi.org/10.1016/j.media.2014.01.011.

Daubechies, I., Devore, R., Fornasier, M., Güntürk, C.S., 2010. Iteratively reweighted least squares minimization for sparse recovery. Commun. Pure Appl. Math. 63, 1–38. https://doi.org/10.1002/cpa.20303.

Dell'Acqua, F., Rizzo, G., Scifo, P., Clarke, R.A., Scotti, G., Fazio, F., 2007. A model-based deconvolution approach to solve fiber crossing in diffusion-weighted MR imaging. IEEE Trans. Biomed. Eng. 54, 462–472. https://doi.org/10.1109/TBME.2006.888830.

Dell'Acqua, F., Scifo, P., Rizzo, G., Catani, M., Simmons, A., Scotti, G., Fazio, F., 2010. A modified damped Richardson-Lucy algorithm to reduce isotropic background effects in spherical deconvolution. Neuroimage 49, 1446–1458. https://doi.org/10.1016/j.neuroimage.2009.09.033.

Descoteaux, M., Deriche, R., Knösche, T.R., Anwander, A., Knosche, T.R., Anwander, A., 2009. Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Trans. Med. Imag. 28, 269–286. https://doi.org/10.1109/TMI.2008.2004424.

Dhollander, T., Connelly, A., 2016. A novel iterative approach to reap the benefits of multi-tissue CSD from just single-shell (+ b = 0) diffusion MRI data. Proc. Intl. Soc. Mag. Reson. Med. 24, 3010.

Dhollander, T., Raffelt, D., Connelly, A., 2016. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image. In: ISMRM Workshop on Breaking the Barriers of Diffusion MRI, p. 5.

Donoho, D.L., 2006. Compressed sensing. IEEE Trans. Inf. Theor. 52, 1289–1306. https://doi.org/10.1109/TIT.2006.871582.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Ann. Stat. 32, 407–451.

Ferizi, U., Schneider, T., Panagiotaki, E., Nedjati-Gilani, G., Zhang, H., Wheeler-Kingshott, C.A.M.M., Alexander, D.C., 2014. A ranking of diffusion MRI compartment models with in vivo human brain data. Magn. Reson. Med. 72, 1785–1792. https://doi.org/10.1002/mrm.25080.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Software 33, 1–22. https://doi.org/10.18637/jss.v033.i01.

Gautschi, W., Slavik, J., 1978. On the computation of modified Bessel function ratios. Math. Comput. 32, 865–875.

Gorodnitsky, I.F.I.F., Rao, B.D.B.D., 1997. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. IEEE Trans. Signal Process. 45, 600–616. https://doi.org/10.1109/78.558475.

Hong, S., Kim, Y., Park, T., 2014. Practical issues in screening and variable selection in genome-wide association analysis. Canc. Inf. 13, 55–65. https://doi.org/10.4137/CIN.S16350.

Ilya, F., 2014. A novel method based on the tikhonov functional for non-negative solution of a system of linear equations with non-negative coefficients. Int. J. Comput. Meth. 11, 1350071. https://doi.org/10.1142/S0219876213500710.

Iturria-Medina, Y., Canales-Rodríguez, E.J., Melie-García, L., Valdés-Hernández, P.A., Martínez-Montes, E., Alemán-Gómez, Y., Sánchez-Bornot, J.M., 2007. Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. Neuroimage 36, 645–660. https://doi.org/10.1016/j.neuroimage.2007.02.012.

Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K., Sijbers, J., 2013. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. Hum. Brain Mapp. 34, 2747–2766. https://doi.org/10.1002/hbm.22099.

Jeurissen, B., Tournier, J.D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. Neuroimage 103, 411–426 doi:S1053-8119(14)00644-2 [pii]10.1016/j.neuroimage.2014.07.061.

Jian, B., Vemuri, B.C., 2007. A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE Trans. Med. Imag. 26, 1464–1471. https://doi.org/10.1109/TMI.2007.907552.

Kaden, E., Knösche, T.R., Anwander, A., 2007. Parametric spherical deconvolution: inferring anatomical connectivity using diffusion MR imaging. Neuroimage 37, 474–488. https://doi.org/10.1016/j.neuroimage.2007.05.012.

Kaden, E., Kruggel, F., 2012. Nonparametric Bayesian inference of the fiber orientation distribution from diffusion-weighted MR images. Med. Image Anal. 16, 876–888. https://doi.org/10.1016/j.media.2012.01.004S1361-8415(12)00012-6 pii).

Landman, B.A., Bogovic, J.A., Wan, H., ElShahaby, F.E.Z., Bazin, P.L., Prince, J.L., El Zahraa ElShahaby, F., Bazin, P.L., Prince, J.L., 2012. Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. Neuroimage 59, 2175–2186. https://doi.org/10.1016/j.neuroimage.2011.10.011S1053-8119(11)01172-4 pii).

Lawson, C.L., Hanson, R.J., 1974. In: Upper Saddle River, N.P.H. (Ed.), Solving Least-squares Problems, p. 161.

Lyu, Q., Lin, Z., She, Y., Zhang, C., 2013. A comparison of typical ℓ_p minimization algorithms. Neurocomputing 119, 413–424. https://doi.org/10.1016/J.NEUCOM.2013.03.017.

Maier-Hein, K.H., Neher, P.F., Houde, J.C., Côté, M.A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.C., Lin, Y.C., Ji, Q., Reddick, W.E., Glass, J.O., Chen, D.Q., Feng, Y., Gao, C., Wu, Y., Ma, J., Renjie, H., Li, Q., Westin, C.F., Deslauriers-Gauthier, S., González, J.O.O., Paquette, M., St-Jean, S., Girard, G., Rheault, F., Sidhu, J., Tax, C.M.W., Guo, F., Mesri, H.Y., Dávid, S., Froeling, M., Heemskerk, A.M., Leemans, A., Boré, A., Pinsard, B., Bedetti, C., Desrosiers, M., Brambati, S., Doyon, J., Sarica, A., Vasta, R., Cerasa, A., Quattrone, A., Yeatman, J., Khan, A.R., Hodges, W., Alexander, S., Romascano, D., Barakovic, M., Auría, A., Esteban, O., Lemkaddem, A., Thiran, J.P., Cetingul, H.E., Odry, B.L., Mailhe, B., Nadar, M.S., Pizzagalli, F., Prasad, G., Villalon-Reina, J.E., Galvis, J., Thompson, P.M., Requejo, F.D.S., Laguna, P.L., Lacerda, L.M., Barrett, R., Dell'Acqua, F., Catani, M., Petit, L., Caruyer, E., Daducci, A., Dyrby, T.B., Holland-Letz, T., Hilgetag, C.C., Stieltjes, B., Descoteaux, M., 2017. The challenge of mapping the human connectome based on diffusion tractography. Nat. Commun. 8 https://doi.org/10.1038/s41467-017-01285-x.

Mairal, J., Bach, F., Ponce, J., 2014. Sparse modeling for image and vision processing. Found. Trends® Comput. Graph. Vis. 8, 85–283. https://doi.org/10.1561/0600000058.

Mastropietro, A., Scifo, P., Rizzo, G., 2017. Quantitative comparison of spherical deconvolution approaches to resolve complex fiber configurations in diffusion MRI: ISRA-based vs L2L0 sparse methods. IEEE Trans. Biomed. Eng. 64, 2847–2857 doi:10.1109/TBME.2017.2676980.

Melie-García, L., Canales-Rodríguez, E.J., Alemán-Gómez, Y., Lin, C.P., Iturria-Medina, Y., Valdés-Hernández, P.A., 2008. A Bayesian framework to identify principal intravoxel diffusion profiles based on diffusion-weighted MR imaging. Neuroimage 42, 750–770. https://doi.org/10.1016/j.neuroimage.2008.04.242.

Miller, A.J., 1984. Selection of subsets of regression variables. J. R. Stat. Soc. Ser. A 147, 389. https://doi.org/10.2307/2981576.

Nesterov, Y., 2004. Introductory Lectures on Convex Optimization. A Basic Course. Springer.

O'Donoghue, B., Candès, E., 2015. Adaptive restart for accelerated gradient schemes. Found. Comput. Math. 15, 715–732. https://doi.org/10.1007/s10208-013-9150-3.

Panagiotaki, E., Schneider, T., Siow, B., Hall, M.G., Lythgoe, M.F., Alexander, D.C., 2012. Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. Neuroimage 59, 2241–2254 doi:S1053-8119(11)01156-6 [pii]10.1016/j.neuroimage.2011.09.081.

Parker, G.D., Marshall, D., Rosin, P.L., Drage, N., Richmond, S., Jones, D.K., 2013. A pitfall in the reconstruction of fibre ODFs using spherical deconvolution of diffusion MRI data. Neuroimage 65, 433–448. https://doi.org/10.1016/j.neuroimage.2012.10.022.

Patel, V., Shi, Y., Thompson, P.M., Toga, A.W., 2010. Mesh-based spherical deconvolution: a flexible approach to reconstruction of non-negative fiber orientation distributions. Neuroimage 51, 1071–1081. https://doi.org/10.1016/j.neuroimage.2010.02.060.

Pisharady, P.K., Sotiropoulos, S.N., Duarte-Carvajalino, J.M., Sapiro, G., Lenglet, C., 2017a. Estimation of white matter fiber parameters from compressed multiresolution diffusion MRI using sparse Bayesian learning. Neuroimage. https://doi.org/10.1016/j.neuroimage.2017.06.052.

Pisharady, P.K., Sotiropoulos, S.N., Sapiro, G., Lenglet, C., 2017b. A sparse Bayesian learning algorithm for white matter parameter estimation from compressed multi-shell diffusion MRI. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2017. Lecture Notes in Computer Science, 10433. Springer, Cham, pp. 602–610. https://doi.org/10.1007/978-3-319-66182-7_69.

Ramirez-Manzanares, A., Rivera, M., Vemuri, B.C., Carney, P., Mareci, T., 2007. Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry. IEEE Trans. Med. Imag. 26, 1091–1102. https://doi.org/10.1109/TMI.2007.900461.

Rao, B.D., Engan, K., Cotter, S.F., Palmer, J., Kreutz-Delgado, K., 2003. Subset selection in noise based on diversity measure minimization. IEEE Trans. Signal Process. 51, 760–770. https://doi.org/10.1109/TSP.2002.808076.

Rathi, Y., Michailovich, O., Setsompop, K., Bouix, S., Shenton, M.E., Westin, C.F., 2011. Sparse multi-shell diffusion imaging. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 58–65. https://doi.org/10.1007/978-3-642-23629-7_8.

Sacchi, M.D., 1997. Reweighting strategies in seismic deconvolution. Geophys. J. Int. 129, 651–656 doi:10.1111/j.1365-246X.1997.tb04500.x.

Sacchi, M.D., Ulrych, T.J., 1996. Estimation of the discrete Fourier transform, a linear inversion approach. Geophysics 61, 1128–1136. https://doi.org/10.1190/1.1444033.

Scherrer, B., Schwartzman, A., Taquet, M., Sahin, M., Prabhu, S.P., Warfield, S.K., 2016. Characterizing brain tissue by assessment of the distribution of anisotropic microstructural environments in diffusion-compartment imaging (DIAMOND). Magn. Reson. Med. 76, 963–977. https://doi.org/10.1002/mrm.25912.

Schmidt, M., Fung, G., Rosales, R., 2007. Fast optimization methods for L1 regularization: a comparative study and two new approaches. In: Machine Learning: ECML 2007. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 286–297. https://doi.org/10.1007/978-3-540-74958-5_28.

Schultz, T., 2012. Learning a reliable estimate of the number of fiber directions in diffusion MRI. Med. Image Comput. Comput. Interv. - MICCAI 15, 493–500. https://doi.org/10.1007/978-3-642-33454-2_61, 2012.

Slawski, M., Hein, M., 2011. Sparse recovery by thresholded non-negative least squares. In: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, 24. Curran Associates, Inc., pp. 1926–1934

Sotiropoulos, S.N., Behrens, T.E., Jbabdi, S., 2012. Ball and rackets: inferring fiber fanning from diffusion-weighted MRI. Neuroimage 60, 1412–1425. https://doi.org/10.1016/j.neuroimage.2012.01.056S1053-8119(12)00073-0 pii).

Thomas, C., Ye, F.Q., Irfanoglu, M.O., Modi, P., Saleem, K.S., Leopold, D.A., Pierpaoli, C., 2014. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. Proc. Natl. Acad. Sci. Unit. States Am. 111, 16574–16579. https://doi.org/10.1073/pnas.1405672111.

Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 273–282. https://doi.org/10.1111/j.1467-9868.2011.00771.x.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244. https://doi.org/10.1162/15324430152748236.

Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. Neuroimage 35, 1459–1472. https://doi.org/10.1016/j.neuroimage.2007.02.016.

Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. Neuroimage 23, 1176–1185. https://doi.org/10.1016/j.neuroimage.2004.07.037.

Tournier, J.D., Yeh, C.H., Calamante, F., Cho, K.H., Connelly, A., Lin, C.P., 2008. Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. Neuroimage 42, 617–625. https://doi.org/10.1016/j.neuroimage.2008.05.002.

Tristán-Vega, A., Westin, C.-F., 2011. Probabilistic ODF estimation from reduced HARDI data with sparse regularization. Med Image Comput Comput Assist Interv 14, 182–190.

Tropp, J.A., Wright, S.J., 2010. Computational methods for sparse solution of linear inverse problems. Proc. IEEE 98, 948–958. https://doi.org/10.1109/JPROC.2010.2044010.

Tuch, D.S., 2004. Q-ball imaging. Magn. Reson. Med. 52, 1358–1372. https://doi.org/10.1002/mrm.20279.

Tuch, D.S., 2002. Diffusion MRI of Complex Tissue Structure. PhD Thesis. MIT.

Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Wedeen, V.J., 2002. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn. Reson. Med. 48, 577–582. https://doi.org/10.1002/mrm.10268.

Vidaurre, D., Bielza, C., Larrañaga, P., 2013. A survey of L1 regression. Int. Stat. Rev. 81, 361–387. https://doi.org/10.1111/insr.12023.

Wedeen, V.J., Hagmann, P., Tseng, W.Y., Reese, T.G., Weisskoff, R.M., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. 54, 1377–1386. https://doi.org/10.1002/mrm.20642.

Wedeen, V.J., Wang, R.P., Schmahmann, J.D., Benner, T., Tseng, W.Y.I.I., Dai, G., Pandya, D.N., Hagmann, P., D'Arceuil, H., de Crespigny, A.J., D'Arceuil, H., de Crespigny, A.J., 2008. Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. Neuroimage 41, 1267–1277. https://doi.org/10.1016/j.neuroimage.2008.03.036.

Wipf, D., Nagarajan, S., 2010. Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions. IEEE J. Sel. Top. Signal Process. 4, 317–329. https://doi.org/10.1109/JSTSP.2010.2042413.

Wipf, D., Nagarajan, S., 2009. Solving sparse linear inverse problems: analysis of reweighted l1 and l2 methods. J. Sel. Top. Signal Process. (Special Issue Compressive Sensing) 4.
Wipf, D.P., Nagarajan, S.S., 2008. A new view of automatic relevance determination. In: Advances in Neural Information Processing Systems, 20, pp. 1625–1632.
Wipf, D.P., Rao, B.D., 2004. Sparse Bayesian learning for basis selection. IEEE Trans. Signal Process. 52, 2153–2164. https://doi.org/10.1109/TSP.2004.831016.
Yeh, F.C., Tseng, W.Y.I., 2013. Sparse solution of fiber orientation distribution function by diffusion decomposition. PLoS One 8. https://doi.org/10.1371/journal.pone.0075747.
Zhang, Z., Rao, B.D., 2011. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. IEEE J. Sel. Top. Signal Process. 5, 912–926. https://doi.org/10.1109/JSTSP.2011.2159773.
Zhang, Z., Xu, Y., Yang, J., Li, X., Zhang, D., 2015. A survey of sparse representation: algorithms and applications. IEEE Access 3, 490–530. https://doi.org/10.1109/ACCESS.2015.2430359.
Zou, H., 2006. The adaptive lasso and its oracle properties. J. Am. Stat. Assoc. 101, 1418–1429. https://doi.org/10.1198/016214506000000735.
Zou, H., Hastie, T., Tibshirani, R., 2007. On the “degrees of freedom” of the lasso. Ann. Stat. 35, 2173–2192. https://doi.org/10.1214/009053607000000127.