Estimation of White Matter Fiber Parameters from Compressed Multiresolution Diffusion MRI using Sparse Bayesian Learning

Pramod Kumar Pisharady\textsuperscript{a}, Stamatios N Sotiropoulos\textsuperscript{b}, Julio M Duarte-Carvajalino\textsuperscript{a}, Guillermo Sapiro\textsuperscript{c,d}, Christophe Lenglet\textsuperscript{a}

\textsuperscript{a}CMRR, Radiology, University of Minnesota, Minneapolis, Minnesota, USA  
\textsuperscript{b}FMRIB, John Radcliffe Hospital, University of Oxford, Headington, UK  
\textsuperscript{c}Electrical and Computer Engineering, Duke University, Durham, North Carolina, USA  
\textsuperscript{d}Biomedical Engineering and Computer Science, Duke University, Durham, North Carolina, USA

Abstract

We present a sparse Bayesian unmixing algorithm \textbf{BusineX: Bayesian Unmixing for Sparse Inference-based Estimation of Fiber Crossings (X)}, for estimation of white matter fiber parameters from compressed (under-sampled) diffusion MRI (dMRI) data. BusineX combines compressive sensing with linear unmixing and introduces sparsity to the previously proposed multiresolution data fusion algorithm RubiX, resulting in a method for improved reconstruction, especially from data with lower number of diffusion gradients. We formulate the estimation of fiber parameters as a sparse signal recovery problem and propose a linear unmixing framework with sparse Bayesian learning for the recovery of sparse signals, the fiber orientations and volume fractions. The data is modeled using a parametric spherical deconvolution approach and represented using a dictionary created with the exponential decay components along different possible diffusion directions. Volume fractions of fibers along these directions define the dictionary weights. The proposed sparse inference, which is based on the dictionary representation, considers the sparsity of fiber populations and exploits the spatial redundancy in data representation, thereby facilitating inference from under-sampled q-space. The algorithm improves parameter estimation from

\textit{Email address: pramodkp@umn.edu (Pramod Kumar Pisharady)}
dMRI through data-dependent local learning of hyperparameters, at each voxel and for each possible fiber orientation, that moderate the strength of priors governing the parameter variances. Experimental results on synthetic and in-vivo data show improved accuracy with a lower uncertainty in fiber parameter estimates. **BusineX resolves a higher number of second and third fiber crossings.** For under-sampled data, the algorithm is also shown to produce more reliable estimates.

**Keywords:** Sparse Bayesian learning, compressive sensing, linear unmixing, diffusion MRI, fiber orientation, sparse signal recovery

1. Introduction

1.1. White Matter Parameter Estimation

Multi-compartment models are used to represent the diffusion MR signal from the brain white matter and to estimate microstructure features of the imaged tissue (Behrens et al., 2003; Panagiotaki et al., 2012; Daducci et al., 2015). Estimation of orientations and volume fractions of anisotropic compartments in these models helps infer the white matter fiber anatomy (Behrens et al., 2007). Accurate estimation of these parameters is challenged by the relatively limited spatial resolution of diffusion MRI (dMRI) data, which may lead to increased partial volume artifacts. Advances in magnetic field strength have significantly improved spatial resolution (Vu et al., 2015), although it may lead to increased noise and scanning time. One effective way to mitigate the effects of noise is the multiresolution data fusion approach introduced in RubiX (Sotiropoulos et al., 2013), which combines high SNR characteristics of low resolution (LR) data with high spatial specificity of high resolution (HR) data. It allows combining images with different diffusion contrast at different spatial resolutions and finding the right trade-off between SNR and resolution. This method was further extended recently for fusion of data acquired at different magnetic field strengths, combining the benefits of high spatial and angular resolutions (k / q-space complementarity) (Sotiropoulos et al., 2016). In parallel to these approaches, the
recent developments in compressive sensing (Ji et al., 2009; Otazo et al., 2015; Paquette et al., 2015; Duarte-Carvajalino et al., 2014; Michael Lustig and Pauly, 2007; Seeger et al., 2010; Ramirez Manzanares et al., 2007) are effective ways to deal with the increased scan time, which result in fewer measurements (diffusion gradients) within a voxel.

Prior efforts in estimation of white matter fiber parameters include both parametric approaches (Tuch et al., 2002; Behrens et al., 2003; Anderson, 2005; Kaden et al., 2007; Sotiropoulos et al., 2008, 2013; Coupe et al., 2013; Scherrer et al., 2016) and non-parametric approaches (Tournier et al., 2004, 2007; Ozarslan et al., 2006; Dell’Acqua et al., 2007; Aganj et al., 2010). These methods exploit multiple diffusion measurements with a large number of diffusion gradients. Considering the fact that the number of crossing fiber bundles within a voxel is limited, we propose a novel sparse signal recovery algorithm for improved inference from data with under-sampled q-space (i.e., data acquired with lower number of diffusion encoding directions). We introduce sparsity based representation and inference into the data fusion approach of RubiX, combining the benefits of regularized noise and reduced scan time.

1.2. Compressive Sensing and Sparse Bayesian Learning

Compressive sensing approaches exploit the sparsity for optimal acquisition and recovery of signals (Ji et al., 2009; Michael Lustig and Pauly, 2007). Compressive sensing is used for reconstruction from accelerated imaging techniques across different MRI modalities; in structural MRI (Otazo et al., 2015; Michael Lustig and Pauly, 2007; Seeger, 2010; Seeger et al., 2010), functional MRI (Zong et al., 2014), and dMRI (Duarte-Carvajalino et al., 2014; Ramirez Manzanares et al., 2007; Rathi et al., 2011; Tristan-Vega A, 2011; Aranda et al., 2015). A comparison of sampling strategies and sparsifying transforms to improve compressive sensing in diffusion spectrum imaging pointed out the importance of joint optimization of the sampling scheme and the sparsifying transform (Paquette et al., 2015).

The seminal work by Tipping on sparse Bayesian learning (SBL)
using automatic relevance determination (ARD) (MacKay, 1994) provides a framework for obtaining sparse solutions to regression and classification problems. The sparsity of parameters is enforced by selection of appropriate prior probability distributions for the parameters to be estimated. Relevance learning is done in SBL by using a mixture of zero-mean Gaussian distributions with individual hyperparameters for variance prior distributions. The hyperparameters associated independently with every weight moderate the strength of the prior and govern the variances of the Gaussian scale mixture, adapting to the data. Considering the success of SBL for sparse signal recovery in fields such as computer vision and machine learning (Wright et al., 2010), in this work we use SBL for the recovery of sparse fiber parameters from dMRI.

1.3. Linear Unmixing

Data sample vectors are assumed to be composed of a mix of endmembers in linear unmixing algorithms (Dobigeon et al., 2008). Linear unmixing algorithms estimate both the number of endmembers and their individual contributions. These algorithms are mostly used in the unmixing of component spectra of hyperspectral imagery in remote sensing signal processing (Dobigeon et al., 2008; Bioucas-Dias et al., 2012; Tang et al., 2015; Iordache et al., 2011; Pardo and Sapiro, 2001; Castrodad et al., 2011).

In this work, we consider the multiple anisotropic components (corresponding to fibers) and the single isotropic component in the diffusion model as the endmembers in unmixing problem, and recover these endmembers using an SBL based linear unmixing approach. Previous study (Daducci et al., 2014b) has shown that $l_1$ norm minimization based approaches for promoting sparsity, which are widely used in spherical deconvolution based methods, have the drawback of inconsistency with the sum-to-one constraint (i.e., the physical constraint that the volume fractions of anisotropic and isotropic compartments within a voxel sum to unity) (Panagiotaki et al., 2012). They addressed the issue using a constrained formulation between the data and a sparsity prior bounding the $l_0$ norm of the fiber orientation distributions (the number of fibers). We
demonstrate that sparse Bayesian learning within a linear unmixing framework
is another way to address the sum-to-one and non-negativity (volume fractions
$\geq 0$) constraints, simultaneously promoting sparsity. The approach in SBL is
typically much sparser as it is based on the notion of setting weights to zero
(rather than constraining them to small values), and as it offers probabilistic
predictions without the need to set additional regularization parameters (Tip-
ning and Faul, 2002).

1.4. Proposed Method: Bayesian Unmixing for Sparse Inference-based Estima-
tion of Fiber Crossings (BusineX)

The above mentioned works on compressive sensing in dMRI utilized basis-
based transforms and exploited the sparsity in the basis representation. Our
approach in BusineX is different in several aspects. A major difference is the
Bayesian linear unmixing formulation with SBL based relevance learning. The
unmixing formulation makes the Bayesian inference hard, but it helps in re-
covering fiber parameters with better accuracy, especially when the number of
diffusion measurements are reduced. The SBL framework identifies relevant
fiber orientations by enforcing sparsity, and it further enhances accuracy in
estimation of multi-fiber volume fractions and orientations.

ARD has been used for data-adaptive estimation of fiber parameters (Behrens
et al., 2007), avoiding data unsupported model complexities. The relevance
learning in the proposed approach, which explicitly models sparsity, enhances
the relevance determination by tuning the variance prior hyperparameters in-
dividually and independently for each possible fiber orientation. The non-
negativity and sum-to-one constraints, which make the sparse representation
and inference challenging, are addressed using the linear unmixing framework.
The proposed BusineX algorithm exploits the spatial redundancy in data rep-
resentation, and it improves the estimation of fiber parameters. The number
of fibers that best fit the observed data is estimated by the automatic detec-
tion of number of endmembers using a reversible jump Markov chain Monte
Carlo (MCMC) sampler. In addition to the improvement in no-acceleration
In this paper we extend the framework by introducing SBL based relevance learning. The paper also presents detailed experimental results and analysis, including comparisons of estimated fiber volume fractions in addition to the fiber orientations and diffusivity.

2. Methods

This section details the proposed BusineX algorithm. The dictionary representation of the HR data using compartment model (ball & stick) is described in Subsection 2.1. The representation of LR data using a spatial partial volume model is briefly discussed in Subsection 2.2. The sparse Bayesian learning approach and sparsity based linear unmixing algorithm are described in Subsections 2.3 and 2.4 respectively. Subsection 2.5 explains the MCMC sampling procedure.

2.1. Dictionary Representation of High Resolution Data

The HR data is represented using a dictionary containing exponential decay component vectors in the compartment model of diffusion. The measured dMRI signal at each HR voxel is first modeled using the ball & stick (1) model (Behrens et al., 2003; Panagiotaki et al., 2012),

\[
S_{HR}^k = S_{HR}^0 \left[ \left( 1 - \sum_{n=1}^{N} f_n \right) e^{-b_k d} + \sum_{n=1}^{N} f_n e^{-b_k d (g_k^T v_n)^2} \right] \tag{1}
\]

where,
- \(S_{HR}^k\) is the signal at HR voxel after application of \(k^{th}\) diffusion-sensitizing gradient with direction \(g_k\) and b-value \(b_k\),
- \(S_{HR}^0\) is the HR signal without diffusion gradient applied,
- \(f_n\) is the volume fraction of anisotropic compartment with orientation \(v_n\), and
- \(d\) is the apparent diffusivity.
The measured signal at an HR voxel is the sum of the attenuation signal and measurement noise (2),

\[ y_{HR}^k = \frac{S_{HR}^k}{S_{HR}^0} + \eta_{HR}^k. \]  

Based on (1) and (2), the measured signal along all \( K \) diffusion-sensitizing directions can be written in a dictionary form (3) as

\[
y_{HR} = \begin{pmatrix} e^{-b_1 d} & e^{-b_1 d(g_1^T v_1)^2} & \ldots & e^{-b_1 d(g_N^T v_N)^2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-b_K d} & e^{-b_K d(g_K^T v_1)^2} & \ldots & e^{-b_K d(g_K^T v_N)^2} \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_N \end{pmatrix} + \eta_{HR},
\]

where

\[ f_0 = \left(1 - \sum_{n=1}^{N} f_n\right), \quad f_n \geq 0. \]

Hence,

\[ y_{HR} = Ef + \eta_{HR}. \]  

In Equation (4), \( E \) represents the local dictionary matrix (5) for the HR diffusion data and \( f \) is the sparse vector representation of the HR data in this dictionary \( E \). The non-zero entries in \( f \) define the number and volume fractions of fibers \( (sticks) \) in a voxel.

\[
E = \begin{pmatrix} e^{-b_1 d} & e^{-b_1 d(g_1^T v_1)^2} & \ldots & e^{-b_1 d(g_N^T v_N)^2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-b_K d} & e^{-b_K d(g_K^T v_1)^2} & \ldots & e^{-b_K d(g_K^T v_N)^2} \end{pmatrix}.
\]

The possible orientations of anisotropic components in the dictionary (second column onwards) are pre-specified and formed using a 5th order icosahedral tessellation of the sphere with 10242 points. The estimated orientation is approximated to the nearest pre-specified orientation during the dictionary update process.

With the above dictionary formulation, the problem of finding the number of fibers, its volume fractions and orientations reduces to accurately estimating the
sparse vector $f$. The estimation of the sparse vector $f$ is detailed in Subsections 2.3 and 2.4. Another unknown parameter in the model, the apparent diffusivity $d$, is estimated through the Bayesian inference that maximizes the joint posterior probability of the HR and LR data, as per the multiresolution graphical model in RubiX (Sotiropoulos et al., 2013), as detailed in Appendix A. The partial volume model used to represent the LR data is discussed in the next Subsection (Subsection 2.2).

### 2.2. Partial Volume Representation of Low Resolution Data

The significance of LR data is its high SNR, which is useful for regularizing the noise in HR data, as in the RubiX framework (Sotiropoulos et al., 2013). In the RubiX framework, the LR data and HR data are collected from the same subject through two scans at different spatial resolutions (voxel sizes). The two datasets are aligned (if necessary) using rigid body transformations. Once the data is aligned, the LR data can be represented using corresponding HR data (data that correspond to the same physical location, but at a different spatial resolution grid), with a partial volume model (Sotiropoulos et al., 2013). The model calculates attenuation signal at an LR voxel as a linear combination of the signals at overlapping $M$ HR voxels:

$$S_{LR}^k S_{0LR} = \sum_{m=1}^{M} w_m S_{0HR}^m S_{HR}^m, \quad w_m = e^{-\|r_m - r_0\|^2 \gamma^2}. \quad (6)$$

The HR signal contributes to the LR signal via a discretized Gaussian distance weighing function (DWF) with weights $w_m$ given by the normalized Euclidean distance between the DWF center $r_0$ at LR voxel and the spatial position of each HR voxel $r_m$, and the unknown standard deviation of the DWF, $\gamma$. $\gamma$ is same for $M$ HR voxels overlapped by an LR voxel, but can be different for different LR voxels.

### 2.3. Hierarchical Bayesian Inference and Sparse Bayesian Learning

The volume fractions and fiber orientations are estimated using a semi-supervised hierarchical Bayesian linear unmixing approach, an extension of...
sparse Bayesian inference dealing with constraints \cite{Araki2009,Tipping2001}. A hierarchical Bayesian framework (Fig. 1) is utilized for the sparse inference. In Bayesian inference, prior probability distributions, namely priors, are defined for constraining the parameters to be estimated \cite{Jaynes1968}. In SBL a mixture of zero-mean Gaussian distributions with individual hyperparameters controlling the variances is used as the prior on the parameter to be estimated (volume fractions here). Gamma distributions are used as hyperpriors, which form the priors over the hyperparameters. The mixture of Gaussians with hyperparameters associated independently with every weight was shown equivalent to using a product of Student-\textit{t} priors, once the hyperparameters are integrated out \cite{Tipping2001}. This hierarchical formulation leads to a sparse solution.

Mathematically the prior over volume fractions is given by,

\[
p(f|\alpha) = \prod_{n=1}^{N} \mathcal{N}(f_n|0, \alpha_n^{-1}),
\]  

Figure 1: The hierarchical Bayesian network used in BusineX. \( y^k_{LR} \) and \( y^k_{HR} \) are the measured signals along diffusion gradient direction \( k \), at LR and HR voxels respectively. \( f_n \) is the \( n \)-th component of the anisotropic volume fractions vector and \( \alpha_n \) is the hyper-parameter in the prior distribution of \( f_n \). The influence of the parameters on LR data is through the spatial partial volume model \cite{6}. 

\( \mathcal{N} \)
where the hyper-parameter $\alpha_n$ controls the variance of individual Gaussians. The update procedure for $\alpha_i$ (detailed in Subsection 2.4) is such that many of the $\alpha$ are pushed to higher values, adapting to the data. The variance $1/\alpha$ of the corresponding Gaussians are pushed towards zero which forces the corresponding weights to be zero (or negligibly small), leading to a sparse solution.

The proposed sparse approach and independent tuning of the hyperparameters for each voxel and for each possible fiber orientation promote adaptation of the estimated number of fibers to the data. Thus the complexity of fiber patterns (single fiber vs. multiple fibers) at each voxel is decided better by the data (refer to Section 4.1 for a discussion).

2.4. Sparsity based Bayesian Linear Unmixing Inference

Finding the volume fractions $\mathbf{f}$ in (4) with a large number of possible fiber orientations ($N$) is an ill-posed problem. We introduce sparsity in the dictionary and estimation process, to propose an efficient algorithm for volume fraction and fiber orientation estimations. The non-negativity and sum-to-one constraints of volume fractions make the sparse representation and inference especially difficult. We fix the sparsity level (the number of fibers, which is the same as the number of non-zero anisotropic components) to a small number $n_0$ ($n_0 << N$). The problem is then formulated as a linear unmixing inference where the diffusion signals correspond to a mixture of the dictionary components with positive weights $\mathbf{f}$. We follow a semi-supervised hierarchical Bayesian linear unmixing approach \cite{Dobigeon et al., 2008} for sparsity-based inference of fibers. The method is semi-supervised because the dictionary is known for a given diffusivity, gradient directions, b-values, and possible fiber orientations, but we don’t know the values of diffusivity, fiber orientations, or the volume fractions within each compartment.

Assuming Gaussian noise\footnote{We implemented Rician noise model \cite{Henkelman, 1985} as well, which provided identical results. Gaussian noise model is presented here for simplicity, as our focus is on the sparse} the likelihood function of the HR data can be
expressed as \( \mathcal{N}(\mathbf{y}_{HR} | \mathbf{E}, \sigma^2) \).

\[
p(y_{HR} | f, \alpha, \sigma^2) = \left( \frac{1}{2\pi\sigma^2} \right)^K \exp \left( - \frac{\|y_{HR} - \mathbf{E}f\|^2}{2\sigma^2} \right),
\]

where \(\sigma^2\) corresponds to the variance of the error in representation of \(y_{HR}\) using dictionary \(\mathbf{E}\) and volume fractions \(f\). Let \(f^+ = [f_1, \ldots, f_{n_0}]^T\) be the volume fractions with \(n_0\) non-zero anisotropic components, then \(f^+\) belongs to a simplex \(S\).

\[
S = \left\{ f^+ | f_n > 0, \forall n = 1, \ldots, n_0, \sum_{n=1}^{n_0} f_n \leq 1 \right\}.
\]

Once the prior and likelihood are defined, Bayesian inference proceeds by calculating the posterior using Bayes’ rule. The generative model of RubiX is adapted here with a novel inference algorithm for the volume fractions and fiber orientations. The proposed algorithm introduces an additional layer of adaption to the data adaptive ARD framework, to enhance the automatic detection of the number of fibers, by tuning the volume fraction variance for each possible fiber orientation and by explicitly modeling sparsity to improve the relevance determination.

The volume fractions posterior is given by (10) (Tipping, 2001)

\[
p(f^+, \alpha, \sigma^2 | y_{HR}) = \frac{p(y_{HR} | f^+, \alpha, \sigma^2) p(f^+, \alpha, \sigma^2)}{p(y_{HR})}.
\]

We cannot compute (10) as the normalizing integral (11) cannot be computed analytically.

\[
p(y_{HR}) = \int p(y_{HR} | f^+, \alpha, \sigma^2) p(f^+, \alpha, \sigma^2) df^+ d\alpha d\sigma^2.
\]

Instead the posterior (10) is decomposed as

\[
p(f^+, \alpha, \sigma^2 | y_{HR}) = p(f^+ | y_{HR}, \alpha, \sigma^2) p(\alpha, \sigma^2 \mid y_{HR}),
\]

where

\[
p(f^+ | y_{HR}, \alpha, \sigma^2) = \frac{p(y_{HR} | f^+, \sigma^2) p(f^+ | \alpha)}{p(y_{HR} | \alpha, \sigma^2)}.
\]

Bayesian unmixing inference.
We can compute (13) as its normalizing integral (14) is a convolution of Gaussians (Tipping, 2001),

\[ p(y_{HR}|\alpha, \sigma^2) = \int p(y_{HR}|f^+, \sigma^2) p(f^+|\alpha) df^+. \]  

(14)

We now introduce the linear unmixing framework to the sparse inference. Blind unmixing under positivity constraints was introduced by Moussaoui et al. (Moussaoui et al., 2006). Dobigeon et al. (Dobigeon et al., 2008) further extended this by including the sum-to-one constraint, attempting to resolve the scale indeterminacy inherent in blind source separation problems. We introduce these linear unmixing constraints to the posterior computation in (13), to propose sparsity based linear unmixing inference (15),

\[ p(f^+|y_{HR}, \alpha, \sigma^2) \sim e^{-\left(f^+-\mu_f\right)^T \Lambda_f^{-1} \left(f^+-\mu_f\right)} 1_S(f^+), \]  

(15)

where

\[ \Lambda_f = \left[ \sigma^{-2} \left( E_{n0}^+ - e_0 u^T \right)^T \left( E_{n0}^+ - e_0 u^T \right) + A \right]^{-1}, \]  

(16)

and

\[ \mu_f = \sigma^{-2} \Lambda_f \left( E_{n0}^+ - e_0 u^T \right)^T (y_{HR} - e_0), \]  

(17)

with \( u \) is a \( 1 \times n_0 \) vector, \([1, \ldots, 1]^T\), and \( A = \text{diag}(\alpha_0, \alpha_1, \ldots, \alpha_N) \). \( E_{n0}^+ \) contains the columns of \( E \) that correspond to \( n_0 \) non-zero coefficients in \( f^+ \) (effective dictionary) and \( e_0 \) is the column corresponding to the isotropic compartment (ball in the HR model). \( 1_S(f^+) \) in (15) is 1 if \( f^+ \in S \) and 0 otherwise.

Each hyper-parameter \( \alpha_n \) in \( A \) are updated iteratively (Tipping, 2001) as per (18),

\[ \alpha_n^{\text{new}} = \gamma_n / \mu_{n^2}, \]  

(18)

where \( \gamma_n = 1 - \alpha_n * \Lambda_{nn} \), and \( \Lambda_{nn} \) is the \( n^{th} \) diagonal element of the posterior volume fractions covariance (16). The noise variance \( \sigma^2 \) is updated as per (19),

\[ (\sigma^2)^{\text{new}} = \frac{\| (y_{HR} - e_0) - (E_{n0}^+ - e_0 u^T) \mu_f \|^2}{K - \sum_n \gamma_n}. \]  

(19)

The priors we used for volume fractions are a mixture of Gaussians with variances controlled by the hyperparameters \( \alpha_n \), as detailed in Subsection 2.3.
We followed the rest of the parameter priors and the inference procedure, including the estimation of diffusivity $d$, as in RubiX (Sotiropoulos et al., 2013). The priors used for $S^0$ and $\sigma$ are unconditional and non-informative (uniform). Conditional priors are used for orientation and diffusivity and are defined as a mixture of Watson distributions with non-informative hyper-parameter for orientation and normal distribution with informative hyper-parameter for diffusivity.

2.5. Hybrid Metropolis-Within, Reversible Jump Gibbs Sampler for Detecting the Number of Fibers

The generation of samples according to (15-17) is accomplished using a Gibbs sampler (Algorithm 1). It proceeds by repeated application of (18) and (19) and the corresponding updates of posterior statistics $\Lambda_f$ and $\mu_f$ from (16) and (17). Each column in the effective dictionary $E_{n0}^+$ can be switched at random with another to test a different fiber orientation.

In order to find the number of fibers that best fits the data automatically we used a metropolis-within reversible jump Gibbs sampler (Dobigeon et al., 2008) which kills or generates fibers as per the death and birth probabilities (Denison et al., 2002), respectively (Algorithm 1). The following 3 cases can occur in each iteration:

- **CASE 1** - Add a new fiber through $BIRTH$ move: The volume fraction of the new fiber is drawn from a Beta distribution, $Beta(1, n_0)$. The other anisotropic volume fractions are scaled so that anisotropic and isotropic volume fractions sum to one.

- **CASE 2** - Remove a fiber through $DEATH$ move: The remaining anisotropic volume fractions are scaled so that anisotropic and isotropic volume fractions sum to one.

---

2See Appendix A for the detailed inference procedure.
Algorithm 1 Hybrid Metropolis-Within Reversible Jump Gibbs Sampler

1: procedure Initialization
2:    Initialize $E$; \% dictionary, using ball & stick model fitting \[1\]
3:    Initialize $f^+$; \% volume fractions, using ball & stick model fitting \[1\]
4:    Initialize prior probabilities;
5:    Initialize $\alpha^+$;
6: end procedure
7: procedure Iterations \% we used 1,500 iterations
8:    for i=1 to \# iterations do
9:        Calculate $\Lambda_f$ and $\mu_f$; \% posterior covariance & mean, \[16\] \& \[17\]
10:       Update $\alpha^+$; \% variance of volume fractions prior \[18\]
11:       Update $\sigma^2$; \% noise variance \[19\]
12:       Switch (p) \% p is random $[0-1]$, cases $p \leq 1/3$, $1/3 < p \leq 2/3$, $p > 2/3$
13:          CASE 1: Propose BIRTH move, $n_0 = n_0 + 1$;
14:          CASE 2: Propose DEATH move, $n_0 = n_0 - 1$;
15:          CASE 3: Propose SWITCH move, $n_0 = n_0$;
16:     End
17:     Accept / Reject BIRTH / DEATH / SWITCH move;
18:        propose new $f^+$; \% new volume fractions proposal
19:    end for
20: end procedure
• CASE 3 - Maintain the number of fibers through SWITCH move: Neither a fiber is added nor removed. The inference is proceeded by switching the fiber orientations, columns in the effective dictionary $E_{n_0}^+$. 

In our experiments, we limited the maximum number of anisotropic components (the number of fibers), $n_0^{max}$ to 3. No BIRTH move is allowed when $n_0 = n_0^{max}$ and no DEATH move is allowed when $n_0 = 1$. In all iterations, all possible cases (from the above 3 cases) are kept equally likely, i.e. probability of 1/3 when all the 3 cases are possible and 1/2 when only 2 cases are possible.

The update in the number of fibers is accepted or rejected based on the move acceptance probability. The acceptance probability for a BIRTH move $\rho_b$ is given by $\rho_b = \min\{1, A_b\}$, where $A_b$ is the acceptance ratio. The acceptance probability for a DEATH move $\rho_d$ is given by $\rho_d = \min\{1, A_d\}$, where $A_d$ is the rejection ratio (refer Appendix B for the derivation of $A_b$ and $A_d$). SWITCH moves are accepted or rejected using Metropolis sampling criterion.

3. Experiments and Results

We conducted experiments using 2 sets of synthetic data and one set of in-vivo data. The datasets and the results are detailed in this section.

3.1. Synthetic Data from HARDI Reconstruction Challenge

The first synthetic data we used is simulated from the 2 structured field phantoms used to evaluate algorithms in the HARDI reconstruction challenge organized as part of the ISBI 2012 conference (Daducci et al., 2014a). We used this challenging dataset with complex fiber configurations to test the general performance of the algorithm and to compare it with other existing methods. We used 50 diffusion gradients to simulate the data, using a multi-Tensor model.

---

3We limited the maximum number of fibers $n_0^{max}$ to 3 to have a reasonable convergence time, and since more than 3 fibers is not expected. The algorithm is generic with respect to $n_0^{max}$, with increased convergence time for higher $n_0^{max}$. 

15
Daducci et al. [Daducci et al. 2014a] reported the results of the challenge and compared 20 algorithms used for recovering the intra-voxel fiber structures. Here we compare the performance of BusineX using these reported results.

There are 2 phantoms in the structured field dataset, a simple one (the dataset released with the challenge announcement, the training set, hereafter called ISBI dataset-1) and a complex one (the dataset used to evaluate algorithms to decide the winner, the testing set, hereafter called ISBI dataset-2). Both phantoms had a size of $16 \times 16 \times 5$ voxels. We copied the 5th slice of the phantom and added it as an additional slice to generate the dMRI data with an image size $16 \times 16 \times 6$, using the simulation algorithm the challenge organizers released [http://hardi.epfl.ch/static/events/2012_ISBI/download.html#testingdata]. We added the 6th slice as we need to simulate the LR dataset with half the resolution with an image size $8 \times 8 \times 3$, which we did by down-sampling the HR data to LR image size (averaging the signals at groups of $2 \times 2 \times 2$ HR voxels). We report the average results from all the voxels (6 slices) making the comparisons fair with the average results (5 slices) reported previously [Daducci et al. 2014a]. The HR data is simulated at an SNR 10 (with Rician noise) [Daducci et al. 2014a]. A factor of $8/\sqrt{2}$ is maintained in the ratio of SNR of LR to that of HR signal [Sotiropoulos et al. 2013].

Fig. 2 shows a visualization of the orientations and the sum of anisotropic volume fractions (upper panels) estimated from ISBI dataset-1 (SNR=10), which has the same sum of volume fractions (unity) at every voxel. The histograms of corresponding sum of anisotropic volume fractions (lower left panel) and orientation error (lower right panel) for all the voxels show the improved estimations in BusineX.

We further evaluated and compared the performance of BusineX using 2 criteria, the correct assessment of the number of fiber populations expressed with success rate and the error in orientation estimation expressed with angular precision [Daducci et al. 2014a]. These measures are reported for the testing set, the ISBI dataset-2 (complex dataset).
Figure 2: Comparison showing volume fraction and orientation estimations from the ISBI dataset-1. The SNR of the data is 10. Color coded orientation estimates from BusineX (upper left panel) and RubiX (upper right panel) are shown with the corresponding sum of anisotropic volume fractions in the background. Lower panels show the normalized histograms of sum of anisotropic volume fractions (left) and orientation error (right) for both cases, for all the 6 slices. The comparisons show the improved volume fraction and orientation estimations in BusineX (ground truth for the sum of anisotropic volume fractions is one at every voxel).
3.1.1. Success Rate and Angular Precision

The success rate and angular precision is calculated as below (20 and 21) (Daducci et al., 2014a).

**Success Rate (SR)**

\[
SR = \left( 1 - \frac{|M_{true} - M_{estimated}|}{M_{true}} \right) \times 100,
\]

where \(M_{true}\) and \(M_{estimated}\) are, respectively, the true and estimated number of fiber compartments inside a voxel.

**Angular Precision (AP)**

\[
AP = \frac{180}{\pi} \arccos(|d_{true} \cdot d_{estimated}|),
\]

where \(d_{true}\) and \(d_{estimated}\) are a pair of true and estimated fiber orientation vectors in a voxel.

The reported results are the mean SR and AP across all voxels and fibers (Table 1). The better performance of BusineX is evident in these results. In particular, for dataset-2 (the dataset used to evaluate algorithms to decide the winner of the challenge), BusineX provided an SR of 80.36%, 11.13% better than RubiX (69.23%) and 16.36% better than the algorithm reported as top in SR (64%) (Rodriguez et al., 2010), in the comparison provided in (Daducci et al., 2014a) (Refer Fig. 9, page 396). This shows the benefit of BusineX in detecting fibers more accurately, which is made possible through the explicit calculation of volume fractions posterior probability (15), as detailed in Section 2.4. Table 1 also reports the interquartile range representing the dispersion in AP and also the number of diffusion measurements used by each method.

The results reported in Table 1 are obtained using the uncorrected SNRs from (Daducci et al., 2014a). The SNR corrected for the variable echo-time (TE) is 24.3 in our case (b-values 1500, see Table II in (Daducci et al., 2014a)). We also did experiments with a corrected SNR of 24.3 (instead of 10). The corresponding SR and AP are 82.99% and 4.88 degrees respectively.

In order to study whether the estimated fiber populations are close enough to the real ones, we calculated the SR using the tolerance cone approach (Daducci et al., 2014a). An estimated fiber is considered resolved only if the corresponding
Table 1: Mean Success Rate and Angular Precision - ISBI dataset-2

<table>
<thead>
<tr>
<th></th>
<th>Success Rate (%)</th>
<th>Mean Angular Precision (Degrees)</th>
<th>Interquartile Range (Degrees)</th>
<th># diffusion measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>BusineX</td>
<td>80.36</td>
<td>5.27</td>
<td>6.8</td>
<td>50</td>
</tr>
<tr>
<td>RuhiX</td>
<td>69.23</td>
<td>6.98</td>
<td>7.9</td>
<td>50</td>
</tr>
<tr>
<td>DSI</td>
<td>64.00</td>
<td>7.50</td>
<td>5.4</td>
<td>257</td>
</tr>
<tr>
<td>Rodriguez et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN-L2</td>
<td>60.30</td>
<td>6.50</td>
<td>7.33</td>
<td>48</td>
</tr>
<tr>
<td>Ramiru Mananares et al. 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOT</td>
<td>57.06</td>
<td>11.22</td>
<td>8.2</td>
<td>60</td>
</tr>
<tr>
<td>Ozarslan et al. 2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2-L2</td>
<td>55.29</td>
<td>7.84</td>
<td>6.35</td>
<td>37</td>
</tr>
<tr>
<td>Canales Rodriguez et al. 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QBI</td>
<td>52.06</td>
<td>15.54</td>
<td>12.9</td>
<td>60</td>
</tr>
<tr>
<td>Aganj et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSD</td>
<td>49.12</td>
<td>11.08</td>
<td>8.1</td>
<td>60</td>
</tr>
<tr>
<td>Tournier et al. 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSI</td>
<td>37.65</td>
<td>14.19</td>
<td>17.9</td>
<td>257</td>
</tr>
<tr>
<td>Wedeen et al. 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DTI</td>
<td>34.7</td>
<td>16.48</td>
<td>13.75</td>
<td>6</td>
</tr>
<tr>
<td>Basser et al. 1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* These results are the best approximations from the graphical plot in Fig. 9, page 396, (Daducci et al. 2014a).

DTI- diffusion tensor imaging, DSI- diffusion spectrum imaging, DSI\textsubscript{LR}- DSI Lucy-Richardson, DOT- diffusion orientation transform, QBI\textsubscript{CSA}- Q-ball imaging constant solid angle, CSD- constrained spherical deconvolution.
orientation falls within a tolerance cone of 20° around the real fiber population. This measure, which is reported as SR\textsubscript{\textast}} (Daducci et al., 2014a) is 70.95% for BusineX and 63.86% for RubiX.

3.2. Synthetic Data Simulated using Camino

To study the effect of under-sampling and to have uniform ground truth values for volume fractions, we simulated a second synthetic dataset using the Camino toolbox (Cook et al., 2006). A Tensor-Cylinder-Sphere model is used to simulate single and crossing fiber structures (with 2 and 3 fibers) with image size $10 \times 10 \times 2$ (LR) and $20 \times 20 \times 4$ (HR), at different under-sampling factors. The diffusivity value used to simulate the data is $1.7 \times 10^{-9} m^2/s$. To make the fiber pattern in the image continuously varying, the orientation of fibers at each HR voxel is selected such that it varies across each dimension by 1 degree/voxel. A minimum crossing angle of 45 degrees is maintained in this case. Diffusion signals are simulated along 200 uniformly distributed directions, with a b-value of 1500 $s/mm^2$. The noise free LR signal is created by down-sampling the HR data to LR image size (averaging the signals at groups of $2 \times 2 \times 2$ HR voxels). Rician noise is added to both LR and HR images by adding zero-mean Gaussian signal in quadrature. A factor of $8/\sqrt{2}$ is maintained in the ratio of SNR of LR to that of HR signal (lower noise in LR data) (Sotiropoulos et al., 2013). We simulated HR data with two SNRs, 15 and 25. Under-sampling of diffusion directions is done by a factor of up to four to simulate acceleration in image acquisition.

The algorithm performance is compared with the ball & stick model applied to the HR dataset (using BedpostX tool [Behrens et al., 2007] in FSL) and RubiX (Sotiropoulos et al., 2013) applied to HR and LR datasets. Both RubiX and BusineX are applied to HR and LR datasets, with the first 100 measurements forming the no-acceleration data. This is done in order to approximately match the acquisition time, making the comparisons fair (as BedpostX uses data at only HR resolution). The 100 measurements are under-sampled again up to a factor of four to simulate accelerations. We used the protocol proposed by
Caruyer et al. (Caruyer et al., 2011) for under-sampling, which makes any first N samples isotropic.

3.2.1. Fiber Orientation Estimation

Fig. 3 shows the mean error and standard deviation in fiber orientation estimation, and the variations with acceleration, for 1, 2, and 3 fiber cases, with SNR 15 and 25. On comparison, BusineX provided better estimation accuracy, at a slightly lower uncertainty. The variation in estimation error with acceleration is lower in BusineX.

Figure 3: Comparison of fiber orientation estimation error (mean across 1600 voxels) and its variation with acceleration factor (under-sampling in number of diffusion measurements). Three data points with no under-sampling (1), under-sampling by a factor of 50% (2) and under-sampling by a factor of 75% (4) are shown. Y-axis represents mean error across voxels and across fibers in 2 and 3 fiber cases. The error bars shown represent the standard deviation in estimation (scaled by 50% for all the methods, for better visualization), representing the estimation uncertainty.
3.2.2. Fiber Volume Fraction Estimation

Fig. 4 presents the histogram results of volume fraction estimation in the 2 fiber case with SNR 15. The true value of both fiber 1 and 2 volume fractions is 0.3. The comparisons show the improved volume fraction estimation in both first and second fiber cases. A similar comparison in the 3 fiber case is shown in Fig. 5, in which each fiber has equal volume fractions of 0.25.

![Figure 4](image1.png)  ![Figure 5](image2.png)

Figure 4: Comparison of volume fraction estimation in the 2 fibers case (SNR=15, no-acceleration). (a) & (b) are histograms of estimated volume fractions of fiber #1 & #2 respectively. Each of the fibers have a true volume fraction of 0.30, marked with red cross on the x-axis. The total number of voxels having a volume fraction of 0.3 is 1600.

Figure 5: Comparison of volume fraction estimation in the 3 fibers case (SNR=15, no-acceleration). (a)-(c) are histograms of estimated volume fractions of fiber #1, #2, & #3 respectively. Each of the fibers have a true volume fraction of 0.25, marked with red cross on the x-axis. The total number of voxels having a volume fraction of 0.25 is 1600.

The above comparisons (Fig. 4 and 5) are on estimation from data without
acceleration. Fig. 6 shows a comparison under different accelerations (under-sampling factors) in 2 fibers case with SNR 15.

![Figure 6: Comparison of volume fraction estimation of fiber 1 (row 1) and fiber 2 (row 2) at different under-sampling factors. Left column shows the case with no under-sampling, middle column shows the case with 50% under-sampling, and the right column shows the case with 75% under-sampling. Each of the fibers have a true volume fraction of 0.30, marked with red cross on the x-axis. The total number of voxels having a volume fraction of 0.3 is 1600. The SNR of the data is 15.](image)

3.2.3. Diffusivity Estimation

We used a diffusivity value of $1.7 \times 10^{-9} \text{m}^2/\text{s}$ for the data simulation using Camino. The estimated mean diffusivity (mean of the diffusivity across 1600 voxels) is $1.6868 \times 10^{-9} \text{m}^2/\text{s}$ with a variance $1.02 \times 10^{-21}$. RubiX provided identical results: mean diffusivity of $1.6813 \times 10^{-9} \text{m}^2/\text{s}$ with a variance $1.14 \times 10^{-21}$. BedpostX provided slightly lower accuracy in diffusivity estimation: $1.6296 \times 10^{-9} \text{m}^2/\text{s}$ with a variance of $5.14 \times 10^{-21}$. 

23
3.2.4. Estimation Uncertainty

Table 2 provides mean span of 95% cones of uncertainty, which is a measure of the width of estimated distributions, representing the uncertainty in estimation. The estimation uncertainty in BusineX is slightly better than that in RubiX, and the estimation uncertainty in BedpostX is approximately two times that in BusineX.

Table 2: Comparison of estimation uncertainties - Mean span of 95% cones of orientation uncertainty (in degrees)

<table>
<thead>
<tr>
<th>No under-sampling</th>
<th>50% under-sampling</th>
<th>75% under-sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiber 1</td>
<td>Fiber 2</td>
<td>Fiber 1</td>
</tr>
<tr>
<td>BusineX</td>
<td>2.55</td>
<td>2.62</td>
</tr>
<tr>
<td>RubiX</td>
<td>2.65</td>
<td>2.74</td>
</tr>
<tr>
<td>BedpostX</td>
<td>4.88</td>
<td>5.24</td>
</tr>
</tbody>
</table>

3.3. In-vivo Data Acquired using 3T Siemens Prisma Scanner

We acquired in-vivo data from a healthy subject using 3T Siemens Prisma scanner. For HR acquisitions the acquisition matrix was $140 \times 140 \times 92$ voxels with a resolution of $1.5 \times 1.5 \times 1.5 \, \text{mm}^3$. For LR acquisitions the resolution was reduced to $3 \times 3 \times 3 \, \text{mm}^3$ for an acquisition matrix size $70 \times 70 \times 46$ voxels. Diffusion weighting was applied in 200 evenly spaced directions with a b-value of 1500 $s/mm^2$. Twenty one volumes without diffusion weighting are equally interleaved in the dataset.

3.3.1. In-vivo Data Results

We report 4 sets of results from the in-vivo experiments showing, a) the stability of fiber orientation and volume fraction estimates with acceleration (Fig. 7), b) improved estimation of orientation and volume fractions (Fig. 8, 9 & 10), c) improved detection of second and third fiber crossings (quantitative,
Fig. 11), and d) improved diffusivity estimation (Fig. 12). We used the Connectome Workbench (Marcus et al., 2011) from the Human Connectome Project for visualizing the results (Fig. 7-10).

Fig. 7 shows a representative comparison of fiber orientation and sum of volume fractions estimates at different accelerations (under-sampling factors in number of diffusion scans). The region shown is the centrum semiovale area, where commissural fibers (the corpus callosum, CC) and association fibers (the superior longitudinal fasciculus, SLF) crosses the projection fibers (the corticospinal tract, CST). The comparison shows the robustness of fiber orientation and volume fraction estimates with acceleration.

We compare our in-vivo results to the reconstructions provided by RubiX (Sotiropoulos et al., 2013), BedpostX (Behrens et al., 2007), and Constrained Spherical Deconvolution (CSD) (Tournier et al., 2007). The implementations of BedpostX available in FSL (Jenkinson et al., 2012a) and CSD available in MRtrix3 (Tournier et al., 2012) are used for the experiments. The method used for calculation of the response function is tourneir and for fiber orientation distribution is csd. The spherical harmonic order ($l_{\text{max}}$) used is the default value in MRtrix3 ($l_{\text{max}} = 8$). The visualization tool in MRtrix (mrview) is used to visualize the estimated Orientation Distribution Functions (ODFs).

Fig. 8 provides representative comparisons from the centrum semiovale area, showing improved detection of crossing fibers by BusineX, as compared to RubiX. Fig. 9 compares performance of BusineX with that of RubiX, BedpostX, and CSD, showing improved estimations of the fibers crossing the pons. The highlighted regions show improved detection of crossing fibers by BusineX. We calculated the mean sum of volume fractions of first, second, and third fibers from the region of interest (ROI) highlighted in red in Fig. 9. BusineX, RubiX, and BedpostX provided values of 0.469, 0.399, and 0.408 respectively. The

---

4The processing in Workbench estimates bingham distributions from the set of estimated posterior fiber parameter samples, for each fiber orientation in a voxel which is labeled a structure identifier (Marcus et al., 2011).
Figure 7: Comparison showing the stability of fiber orientation and volume fraction estimates with acceleration (under-sampling). Upper panel shows color coded orientation estimates at the region near the centrum semiovale, highlighted in the coronal view in lower panel. The background to the orientation estimates is the sum of anisotropic volume fractions. In the upper panel left column shows the case with no under-sampling, middle column shows the case with 50% under-sampling, and the right column shows the case with 75% under-sampling.
Figure 8: Comparison between BusineX and RubiX, showing improved detection of crossing fibers by BusineX. Panels 2 and 3 show color coded orientation estimates at the region near right and left centrum semiovale, highlighted in the coronal view in uppermost and lowermost panels, respectively. The background to the orientation estimates is the sum of anisotropic volume fractions estimated by each method. The estimation is done from HR and LR datasets under-sampled by a factor of 2 (100 diffusion directions).
higher mean sum of volume fractions obtained from BusineX may correspond to improved detection of crossing fibers.

Fig. 10 shows another comparison of the performance of BusineX with that of RubiX, BedpostX, and CSD, showing improved parameter estimates near SLF. We also analyzed the performance improvement quantitatively, by counting the number of second and third fiber crossings in the white matter, as well as in several specific ROIs. Fig. 11 shows the number of second and third fiber crossings in the white matter, near the pons in the ROI from Fig. 9, in the left/right SLF and in the left/right posterior corona radiata (PCR). It can be noticed that, while RubiX and BedpostX tend to recover fewer second and third fiber crossings as the under-sampling factor increases, BusineX performs equally well even with only a quarter of the original diffusion gradients. Lastly we have shown a map of the estimated mean diffusivity in Fig. 12. By comparison with DTI, both BusineX and BedpostX provide diffusivity estimates with improved contrast. Compared to BedpostX, the estimate from BusineX also appears to be less noisy.

4. Discussion

4.1. Complexity of Fiber Patterns

The use of SBL in BusineX enhances the variance adaptation of fiber volume fractions to the data, at each voxel and for each possible fiber orientation. This is made possible by moderating the strength of priors through associated hyperparameters. In other words, the proposed framework performs relevance learning by tuning the variance hyperparameters spatially (across voxels) and angularly (across possible fiber orientations), which is the main novelty of our approach. Contrary to earlier approaches in fiber parameter estimation, which utilizes a fixed ARD weight for all voxels and fibers (Behrens et al., 2007; Sotiropoulos et al., 2013), this variance adaption through relevance determination improves adaption of the fiber parameters to the data. This is an important cause for
Figure 9: Comparison between BusineX, RubiX, BedpostX and CSD showing improved detection of fibers crossing the pons. Upper and middle panels show color coded orientation estimates (ODF in the case of CSD) at the pons region highlighted in the coronal view in lower left panel. The background is the sum of anisotropic volume fractions estimated by BedpostX for all the methods. A comparison of the areas highlighted in the lower right panel shows improved detection of CST fibers at the level of the pontine crossing tract. The estimation from BedpostX and CSD is done from the full HR dataset with 200 diffusion directions. The estimation from BusineX and RubiX is done from HR and LR datasets with 100 diffusion directions per dataset. We use a total of 200 directions in all experiments to approximately match the acquisition time (LR acquisitions can be done faster than HR). BedpostX and CSD use only one dataset (HR), whereas BusineX and RubiX use two datasets (HR and LR) with half of the diffusion measurements.
Figure 10: Comparison between BusineX, RubiX, BedpostX, and CSD showing improved detection of crossing fibers by BusineX. Upper panels show color coded orientation estimates in the SLF region highlighted in the sagittal view in lower right panel. Lower left panel shows the color coded ODF estimated using MRtrix. The background is the sum of anisotropic volume fractions estimated by BedpostX for all the methods. A comparison of the areas highlighted in the lower middle panel shows improved detection of association fibers through the SLF. BusineX resolves the association fibers in both highlighted areas. RubiX and BedpostX do not resolve all the fibers at the yellow area, and CSD does not resolve all the fibers at the magneta area. The estimation from BedpostX and CSD is done from the full HR dataset with 200 diffusion directions. The estimation from BusineX and RubiX is done from HR and LR datasets with 100 diffusion directions per dataset. We use a total of 200 directions in all experiments to approximately match the acquisition time (LR acquisitions can be done faster than HR). BedpostX and CSD use only one dataset (HR), whereas BusineX and RubiX use two datasets (HR and LR) with half of the diffusion measurements.
Figure 11: Variation in number of second and third fiber crossings (with volume fractions greater than 5%) in the white matter and in five selected ROIs.

Figure 12: Estimated mean diffusivity maps from BusineX (left), BedpostX (middle), and DTI (right).
the improved performance of BusineX, as illustrated in our simulation as well as in-vivo results.

Another reason for the possible improvement in detection of crossing fibers is the automatic tuning of fiber complexity using the hybrid Metropolis-within reversible jump Gibbs sampler (Section 2.5), which helps fiber complexity adaptation by accepting or rejecting addition and deletion of fibers. Also, the rescaling of volume fractions after every addition (BIRTH) and deletion (DEATH) move, as per the non-negativity and sum-to-one constraints, further facilitates improvement in parameter estimation.

4.2. Local and Spatial Diffusion Models

The proposed algorithm is generic with respect to the local diffusion model and the corresponding model parameters. The algorithm can be applied to any model which can represent the data in dictionary form (Section 2.1). We chose the ball & stick model to approximate the diffusion signal, but it can be replaced with more complex models such as a non-monoexponential decay model to approximate multi-shell data (Jbabdi et al., 2012). In order to support the flexibility of the proposed method, we used different models for synthetic data simulation: A Tensor-Cylinder-Sphere model for the Camino dataset and a Multi-Tensor model for the ISBI HARDI dataset.

To verify our spatial model, the assumption that the attenuation signal at an LR voxel can be approximated by a weighted linear combination of attenuation signals at corresponding overlapping HR voxels (6), we calculated the mean and standard deviation of the root mean square error (RMSE) of estimated LR attenuation signal from the actual LR attenuation data. The mean RMSE for our in-vivo data is 0.067 with a standard deviation of 0.048, which justifies our assumption (RMSE close to zero).

The HR local model (1) and the LR spatial partial volume model (6) (Sotiropoulos et al., 2013) adapted in the proposed approach were recently modified to improve the generalizability and performance (Sotiropoulos et al., 2016): the local model is based on the multi-shell non-monoexponential decay (Jbabdi et al.)
Moreover the weighted sum of signal attenuations in the spatial partial volume model is replaced with the ratio of weighted sums of diffusion-weighted signals and non-diffusion-weighted signals. Improvement in estimation, particularly in volume fractions and diffusivities at tissue boundaries, is reported. In our current implementation, we maintained the models as in RubiX [Sotiropoulos et al., 2013] to make the comparisons fair and as our objective is to show the benefits of modeling sparsity. These model changes can be easily adapted to the presented framework.

The elements of our over-complete dictionary (5) created from the HR local model are obtained from an icosahedral tessellation of the sphere. We use a 5th order tessellation limiting the number of possible fiber orientations to 10242, as a compromise between orientation accuracy and computational expense. The worst-case discretization error due to this approximation is 1.18 degrees. The order of tessellation can be increased for slightly improved orientation estimation accuracy, at the expense of computational time.

4.3. Multiple Resolutions, Benefits for Data at Single Resolution

We have introduced a novel method for mapping white matter fiber parameters by combining information from data at high and low spatial resolutions through a sparse linear unmixing framework. The algorithm works on any combination of voxel sizes provided the LR voxel size is an integer multiple (e.g. 2x) of the HR voxel size. The resolutions used in our experiments are 1.5 mm (HR) and 3.0 mm (LR). The idea of combining multiple resolutions for noise regularization was previously presented, discussing the specific aspects and issues related to combining multiple resolutions [Sotiropoulos et al., 2013].

The fiber pattern in the HARDI reconstruction challenge dataset mainly varies across two dimensions (in-plane) and has limited variation in the third dimension (slices). Both our algorithm and RubiX might have benefitted from this spatial consistency to some extent, as these algorithms use the same priors for all HR voxels overlapped by an LR voxel (2 × 2 × 2 = 8 HR voxels overlapped by one LR voxel). In spite of this, BusineX performs better than RubiX (11.13
% in SR and 1.71 Degrees in AP) which is attributable to the volume fractions posterior computation using the proposed sparse Bayesian learning algorithm.

It is straightforward to apply the algorithm for inference from data at a single resolution. The sparse Bayesian unmixing inference procedure detailed in Section 2 remains the same. However, the multiresolution inference as in RubiX, detailed in Appendix A and the partial volume model for LR data, detailed in Section 2.2 need modifications (or deletions). The improvement in accuracy due to the sparse formulation, as well as the benefit of lower number of diffusion measurements may remain similar to the proposed approach, though the inference from data at single resolution may affect the noise regularization behavior of the algorithm (which is originally a benefit of fusing information from LR data).

4.4. Current Limitations and Future Work

Estimation using the BusineX algorithm requires the acquisition of two scans of the same subject at different spatial resolutions, which may lead to specific challenges in the pre-processing steps. The combined inference from two scans may be more sensitive to distortions (motion, B0 inhomogeneity, and eddy current). These EPI distortions can be different for the two scans, as the spatial resolutions are different. We corrected these distortions independently using FSL (Jenkinson et al., 2012b; Andersson and Sotiropoulos, 2015) before aligning the two datasets, to minimize their effects in the inference.

The use of LR data in BusineX is to regularize and mitigate the noise in HR data, by defining the priors and hyper-priors. To overcome the above-mentioned limitations caused by two scans, we plan to explore the possibility of learning these priors from an atlas which can be registered to the single resolution (HR) data. This would allow reconstruction and estimation by acquiring images at a single spatial resolution. We also plan to develop a multi-shell version of BusineX since such dMRI data has been shown to improve fiber orientation mapping and tractography, for example, the multi-shell multi-tissue CSD (Jeurissen et al., 2014).
The current implementation of our algorithm is computationally intensive, mainly due to the MCMC iterations. The algorithm takes about 8 seconds to process one voxel with a CPU speed of 2.6 GHz. To speed-up the processing, we parallelized the algorithm using OpenMP. It takes an average time of 250 milliseconds / voxel on a server with 32 processors. The computational performance of the algorithm can be further improved using GPU/CUDA.

4.5. Concluding Remarks

Reducing acquisition time and maintaining SNR are two challenging goals in dMRI acquisition. We proposed a sparse Bayesian algorithm, namely BusineX, to achieve these goals simultaneously, extending and improving an existing multiresolution approach (RubiX) by efficiently introducing sparsity. BusineX is useful for reconstruction of fiber parameters from accelerated dMRI data. The results from simulation and in-vivo experiments have shown detection of more number of second and third fiber crossings, with improved accuracy and lower estimation uncertainty, for data under-sampled by a factor of up to four. The near linear behavior of the orientation estimation error as well as the number of detected fiber crossings with acceleration shows the potential of the proposed approach for application in shortening the acquisition time of dMRI.

Our main motivation for this work is to demonstrate improvements in the estimation of white matter parameters through explicit modeling of sparsity using sparse Bayesian learning. As discussed above, the main limitation of the proposed algorithm is the need to acquire data at two different spatial resolutions. Several single resolution algorithms are available in the literature (for example the algorithms we compared in Table 1), which can also achieve good angular precision. Our future work will focus on extending BusineX for fiber parameter estimation from single resolution multi-shell data.

Appendix A. Bayesian Inference

The application of Bayes rule with the complete Bayesian inference procedure is briefed in this section. We modified the procedure in RubiX Sotiropou-
\[ p(\Omega/Y) \propto p(Y/\Omega) p(\Omega), \quad (A.1) \]

where \( Y = (Y_{LR}, \{Y^m_{HR}\}) \) represents both HR and LR data, and \( \Omega = (f_n, v_n, d, S^0_{HR}, \eta_{HR}, S^0_{LR}, \eta_{LR}) \) is the set of all parameters to be estimated.

As the priors are conditional on hyperparameters \( C \),

\[ p(\Omega/C/Y) \propto p(Y/\Omega) p(\Omega/C) p(C), \quad (A.2) \]

where

\[
p(Y/\Omega) = p(Y_{LR}/\Omega) \prod_{m=1}^M p(Y^m_{HR}/\Omega) = \prod_{k=1}^K p(Y^k_{LR}/\Omega) \prod_{m=1}^M \prod_{k=1}^K p(Y^m_{HR}/\Omega),
\]

\[ p(\Omega/C) = p(S^0_{LR}) p(\sigma_{LR}) p(\gamma) \prod_{m=1}^M p(S^0_{HR}) p(\sigma^m_{HR}) p(d^m/C_d) p(f^m/C_F) \prod_{n=1}^N p(v_n/C_{v_n}), \quad (A.3) \]

and

\[ p(C) = p(C_d) p(C_F) \prod_{n=1}^N p(C_{v_n}). \quad (A.5) \]

The priors and hyper-priors (other than that for volume fractions, Equation (7)) used are as detailed next.

Priors:

\[ p(S^0_{LR}) = p(S^0_{HR}) = U(0, \infty), \]

\[ p(\sigma_{LR}) = 1/\sigma_{LR}, \quad p(\sigma_{HR}) = 1/\sigma_{HR}, \]

\[ p(\gamma) = U(0, \infty), \]

\[ p(d/C_d) = p(d/\mu_d, \sigma_d) = N(\mu_d, \sigma^2_d), \]

\[ p(v_n/C_{v_n}) = p(v_n/\mu_r, k_r) = |\sin(\theta)| \sum_{r=1}^R c(k_r) e^{-k_r(\mu_r^2 v_n^2)}. \quad (A.6) \]
Hyper-priors:

\[ p(\mu_d) = \Gamma(a, c), \ p(\sigma_d^2) = U(0, \infty), \]
\[ p(C_v) = \prod_{r=1}^{R} p(k_r) p(\mu_r), \]
\[ p(k_r) = U(0, \infty), \ p(\mu_r) = U(S^2). \quad (A.7) \]

Appendix B. Acceptance Probability for \textit{BIRTH} and \textit{DEATH} Moves

Consider a \textit{BIRTH} move from the state \{\(f^{+}(t), E_{n_0}^{+}(t), n_0^t\}\) to a new state \{\(f^{++}, E_{n_5}^{++}, n_5^*\)\}. The acceptance probability \(\rho_b\) for \textit{BIRTH} move is \(\rho_b = \min\{1, A_b\}\), where \(A_b\) is the acceptance ratio given by (B.1) (Green, 1995; Denison et al., 2002),

\[ A_b = P_p \times Pr_p \times T_p \times \left| J(f^*) \right|, \quad (B.1) \]

where

\[ P_p = \frac{p(f^{++}, E_{n_5}^{++}, n_5^*)}{p(f^{+}(t), E_{n_0}^{+}(t), n_0^t)}, \] the ratio of the posterior probabilities,

\[ Pr_p = \frac{q(f^{+}(t), E_{n_0}^{+}(t) | f^{++}, E_{n_5}^{++})}{q(f^{++}, E_{n_5}^{++} | f^{+}(t), E_{n_0}^{+}(t))}, \] the ratio of proposal distributions,

\[ T_p = \frac{dR^{*}}{b_R(t)}, \] the ratio of transition probabilities, and

\[ \left| J(f^*) \right| = \text{the Jacobian of the transformation.} \quad (B.2) \]

The Jacobian \(\left| J(f^*) \right|\) accounts for the change in scale when moving between models of different dimensions. The ratio of transition probabilities is 1 in most of the cases, when the birth and death moves are equally likely. The ratio of the proposal distributions \(Pr_p\) is given by (Dobigeon et al., 2008)

\[ Pr_p = \frac{1}{g_{1,n_0^t(f^{++})}} \frac{n_5^{max} - n_0^t}{n_0^t + 1}, \quad (B.3) \]
where \( g_{a,b}(\cdot) \) denotes the pdf of a Beta distribution \( Be(a, b) \). The posterior ratio \( P_p \) can be written as the product of the likelihood ratio and prior probability ratios of volume fractions, dictionary, and number of fibers (B.4),

\[
P_p = \frac{p(f^*, E^*, n_0^*)}{p(f(t), E(t), n_0^*)} = \frac{p(y|f^*, E^{+*}, n_0^*)}{p(y|f(t), E^{+(t)}, n_0^*)} \\
\times \frac{p(f^*|n_0^*)}{p(f(t)|n_0^*)} \times \frac{p(E^{+*}|n_0^*)}{p(E^{+(t)}|n_0^*)} \times \frac{p(n_0^*)}{p(n_0^*)}.
\]

(B.4)

The prior ratio of volume fractions is given by (B.5)

\[
\frac{p(f^*|n_0^*)}{p(f(t)|n_0^*)} = \prod_{n=0}^{n_0^*} \mathcal{N}(f_{n+0}|0, 1/\alpha_{n+0}^+) = \left( \frac{\alpha_{n+0}^+}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{\alpha_{n+0}^+ f_{n+0}^2}{2}}.
\]

(B.5)

The prior ratio of the dictionary is given by (B.6) (Denison et al., 2002)

\[
\frac{p(E^{+*}|n_0^*)}{p(E^{+(t)}|n_0^*)} = \frac{n_0^{(t)} + 1}{n_0^{\max} - n_0^{(t)}}.
\]

(B.6)

The prior associated to the number of fibers is uniform, and so the prior ratio of number of fibers is 1.

Substituting the values of \( P_p \), \( Pr_p \), and \( T_p \) in B.1 the acceptance ratio \( A_b \) is given by

\[
A_b = e^{-\left( \frac{1}{2} ||y_{HR} - E_{n_0^*}^+ f^{+*}||^2 + \frac{1}{2} ||y_{HR} - E_{n_0^*}^{+(t)} f^{+(t)}||^2 \right)} \\
\times \frac{d_{R^*}}{b_{R(t)}} \times \frac{1}{g_{1,n_0^*(f^{+*})}} \\
\times \left( \frac{\alpha_{n_0^*}}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{\alpha_{n_0^*} f_{n_0^*}^2}{2}}.
\]

(B.7)

The acceptance probability \( \rho_d \) for DEATH move is \( \rho_d = min\{1, A_d\} \), where \( A_d \) is the rejection ratio. The derivation and calculation of \( A_d \) is similar to the calculation of \( A_b \) except that the ratio of transition probabilities \( T_p \) is \( \frac{b_{R(t)}}{g_{R(t)}} \) (Denison et al., 2002) and prior ratio of volume fractions is \( \left( \frac{\alpha_{n_0^{(t)}}}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{\alpha_{n_0^{(t)}} f_{n_0}^2}{2}} \).

Acknowledgements

This work was supported by NIH grants P41 EB015894, P30 NS076408, and the Human Connectome Project (U54 MH091657). The work of GS on sparse modeling is also supported by NSF, ARO, ONR, and NGA.
We would like to thank Sudhir Ramanna for his help with the acquisition of the in-vivo dataset used in this paper.

References


Behrens TE, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, et al. . Characterization and propagation of uncer-


Denison DGT, Holmes CC, Mallick BK, Smith AFM. Bayesian Methods for Nonlinear Classification and Regression. Chichester, U K: Wiley 2002:.


Moussaoui S, Brie D, Mohammad-Djafari A, Carteret C. Separation of non-negative mixture of non-negative sources using a Bayesian approach and


