Denoising and fast diffusion imaging with physically constrained sparse dictionary learning

Alexandre Gramfort, Cyril Poupon, Maxime Descoteaux

Denoising and fast diffusion imaging with physically constrained sparse dictionary learning

A. Gramfort[1,2,3,*], C. Poupon[4], M. Descoteaux[5]

Abstract

Diffusion-weighted imaging (DWI) allows imaging the geometry of water diffusion in biological tissues. However, DW images are noisy at high b-values and acquisitions are slow when using a large number of measurements, such as in Diffusion Spectrum Imaging (DSI). This work aims to denoise DWI and reduce the number of required measurements, while maintaining data quality. To capture the structure of DWI data, we use sparse dictionary learning constrained by the physical properties of the signal: symmetry and positivity. The method learns a dictionary of diffusion profiles on all the DW images at the same time and then scales to full brain data. Its performance is investigated with simulations and two real DSI datasets. We obtain better signal estimates from noisy measurements than by applying mirror symmetry through the q-space origin, Gaussian denoising or state-of-the-art non-local means denoising. Using a high-resolution dictionary learnt on another subject, we show that we can reduce the number of images acquired while still generating high resolution DSI data. Using dictionary learning, one can denoise DW images effectively and perform faster acquisitions. Higher b-value acquisitions and DSI techniques are possible with approximately 40 measurements. This opens important perspectives for the connectomics community using DSI.

Keywords:
diffusion-weighted imaging, diffusion spectrum imaging (DSI), orientation distribution function (ODF), denoising, undersampling, dictionary learning, sparse coding.

*Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, 37-39 Rue Dareau, 75014 Paris, France
Email address: alexandre.gramfort@telecom-paristech.fr (A. Gramfort)
[1]Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, Paris, France
[2]INRIA, Parietal team, Saclay, France
[3]NeuroSpin, CEA Saclay, Bat. 145, 91191 Gif-sur-Yvette Cedex, France
[4]LRMN/NeuroSpin, CEA Saclay, Bat. 145, 91191, Gif-sur-Yvette, cedex France
[5]Sherbrooke Connectivity Imaging Laboratory, Computer Science Department, Université de Sherbrooke

1. Introduction

Diffusion-weighted imaging (DWI) is able to non-invasively image the diffusion of water molecules in biological tissues. DWI was rapidly made popular by several clinical applications using apparent diffusion coefficient (ADC) imaging and diffusion tensor imaging (DTI) (Basser et al., 1994). However, the diffusion tensor is an over-simplified *Gaussian* view of the local diffusion phenomenon happening in each imaging voxel. The holy grail of DWI is to recover the full tridimensional (3D) probability distribution describing the local diffusion phenomenon. This is often called the ensemble average propagator (EAP) formalism (Tuch, 2002; Wedeen et al., 2005; Descoteaux et al., 2011), which provides a powerful framework to describe and predict the diffusion behavior in complex materials. The EAP contains the full 3D information about the water molecule diffusion within the imaging voxel, which goes beyond principal directions that can be used for tractography (Merlet et al., 2012b). The EAP can serve to estimate parameters that reflect the microstructural environment, such as axonal diameter in recent works (Assaf et al., 2008; Ozarslan et al., 2013).

EAP imaging can be long and demanding in terms of acquisition requirements (Descoteaux et al., 2011). Hence, the last 10 years have seen the emergence of numerous techniques to reconstruct the angular information of the EAP, the orientation distribution function (ODF) or other such angular distributions (Seunarine and Alexander, 2009; Descoteaux and Poupon, 2013, *in press*) from a reduced sampling scheme. These new techniques are most often restricted to a single shell in q-space with N uniform measurements for a single b-value (typically $b \in [1000, 3000]$ s/mm^2). This spanned the rich literature of high angular resolution diffusion imaging (HARDI), from compartment modeling to model-free and deconvolution techniques. These works are well covered in the following two book chapters (Seunarine and Alexander, 2009; Descoteaux and Poupon, 2013, *in press*).

In the last 2-4 years, Diffusion Spectrum Imaging (DSI) and 3D DWI have regained popularity, because of two applications. First, several works have shown that the radial information of the DWI signal is important and can be sensitive to white-matter anomalies caused by demyelination or brain damage (Assaf et al., 2008; Alexander, 2008). Consequently, new modeling and anisotropy measures from the EAP have appeared in the literature to better capture both the radial and angular information contained in the diffusion signal. Second, the recent fame of DSI (Wedeen et al., 2012) combined with connectomics studies (Hagmann et al., 2008; Honey et al., 2009), as well as the human brain connectome project[6] have made DSI a central acquisition protocol, despite the issue of long scanning

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

time.

No matter what diffusion imaging protocol is used, be it scalar DWI, DTI, HARDI, or DSI, there is always a trade-off between quality of the data and acquisition time. Indeed, the higher the number of acquired images, the better the estimation of the diffusion signal will be. A common way of improving the signal-to-noise ratio (SNR), which is particularly poor for large b-values (see Fig. 6 for an example), is to repeat the acquisition of the same signal with the same sequence parameters and average them. However, for clinical requirements and applications, and considering the reduction of the risk of motion artifacts, an acquisition time between 3 and 15 minutes is the limit. A first challenge is therefore to be able to improve the SNR of a single acquisition of DWI with denoising algorithms. A second challenge is the ability to reduce the number of acquired images while offering the high resolution data required to estimate complex white matter structures, such as fiber crossing configurations, and microstructural features, such as axonal diameter (Assaf et al., 2008; Alexander, 2008). This paper addresses both of these challenges, providing validation and performance quantification using denoising metrics. The motivation for the use of a denoising benchmark is to compare the results obtained from undersampled data, with full resolution data after denoising using some well established methods. The experimental section focuses on DSI data, as it is a protocol with a dense sampling scheme using high b-value images. Two questions are of particular interest: Can we obtain DSI data with the same number of DWIs required for single b-value HARDI? How much can we subsample the q-space while keeping high spectral resolution in diffusion images?

The intuition behind this paper is that the signal measured by multiple DWIs over the q-space is redundant and shares an underlying structure: the DSI acquisition on a 258 points half-space or on the full 515 points sampling contains redundant information that one can learn and then use to denoise or reduce the number of acquisitions. We show that a dictionary estimated from DSI data captures the geometry of white matter brain structures and can thus be used in 2 different setups: i) intra-subject studies, for denoising purposes and ii) inter-subject studies, to perform super-resolution of q-space data. The latter is done by acquiring subsampled DSI data (low resolution) and using a high resolution dictionary of DSI profiles learnt on another subject in order to recover the full DSI. This inter-subject validation setup was earlier proposed in our previous work (Gramfort et al., 2012), and by Bilgic et al. (2012a), although using an alternative non-physically constrained dictionary learning formulation (see next section). The key contribution of this work is to use sparse coding to estimate a dictionary of prototypical diffusion profiles constrained by physical properties of the signal. We indeed enforce symmetry and positivity for the atoms in the dictionary taking into account the structure of the signal present in multiple DWIs. See for example (Tournier et al., 2007), for a previous demonstration of the relevance of non-negativity constraints. As for symmetry, the physics of dMRI tells us that the measured

local diffusion signal must also be symmetric (Tuch, 2002, Sections 3.3 and 8.3.2.3).

Results are presented on a publicly available simulation dataset and on two real DSI datasets, one from the Pittsburgh Brain Competition 2009 Challenge and one from our institute. The results of the proposed method are compared to the SNR improvements obtained by applying mirror symmetry through the q-space origin, Gaussian denoising and state-of-the-art non-local means denoising. A preliminary version of this work was presented at the MICCAI 2012 international conference (Gramfort et al., 2012). This manuscript complements it with more details on the method, an extensive simulation study and results on a new dataset (Pittsburgh Brain Competition 2009).

2. Theory

Diffusion-Weighted Imaging and Diffusion Spectrum Imaging. Under the narrow pulse assumption (Stejskal and Tanner, 1965), there is a Fourier relationship between the measured DWI signal and diffusion propagator, $P(\mathbf{R})$,

$$P(\mathbf{R}) = \int_{\mathbf{q}\in\Re^3} E(\mathbf{q}) e^{-2\pi i \mathbf{q}\cdot\mathbf{R}} d\mathbf{q}, \tag{1}$$

with $E(\mathbf{q}) = S(\mathbf{q})/E_0$, where $S(\mathbf{q})$ is the diffusion signal measured at position $\mathbf{q}$ in q-space, and E_0 is the baseline image acquired without any diffusion sensitization ($q = 0$). We denote $q = |\mathbf{q}|$ and $\mathbf{q} = q\mathbf{u}$, $\mathbf{R} = r\mathbf{r}$, where $\mathbf{u}$ and $\mathbf{r}$ are 3D unit vectors. The wave vector $\mathbf{q}$ is $\mathbf{q} = \gamma\delta\mathbf{G}/2\pi$, with γ the nuclear gyromagnetic ratio of water molecules and $\mathbf{G} = g\mathbf{u}$ the applied diffusion gradient vector. The norm of the wave vector, $\mathbf{q}$, is related to the diffusion weighting factor (the b-value), $b = 4\pi^2 q^2 \tau$, where $\tau = \Delta - \delta/3$ is the effective diffusion time with δ the duration of the applied diffusion sensitizing gradients and Δ the time between the two pulses. Note that the Fourier relationship between the EAP and the diffusion signal of Eq. 1 is strictly valid only if the narrow pulse assumption is met, which is rarely the case in *in vivo* 3D q-space MRI. Nonetheless, we can measure the approximation of the average diffusion propagator by taking the ensemble average over the imaging voxel, hence the name Ensemble Average Propagator, EAP (Tuch, 2002).

The current state-of-the-art acquisition technique to reconstruct the 3D diffusion propagator is DSI. The original DSI protocol (Wedeen et al., 2005) measured $S(\mathbf{q})$ on a Cartesian grid restricted to a sphere of radius 5, resulting in 515 q-space discrete measurements $S(\mathbf{q})$. Then, a simple 3D inverse Fast Fourier Transform (FFT) is applied to recover the EAP at every imaging voxel. Finally, the diffusion ODF, Ψ, can be extracted by numerically com-

puting the radial integral over the discrete DSI grid, $r \in [0, 5]$, as

$$\Psi(\mathbf{u}) = \int_0^5 P(r\mathbf{u})r^2 dr. \tag{2}$$

DSI acquisition is a long process. A typical full brain coverage acquisition with 60 axial slices, 2 mm isotropic voxels, parallel imaging, a repetition time of approximately TR=11 s, a full DSI grid with 515 directions and b-values from 0 to 6,000 s/mm^2 or so, takes 1h45 min of acquisition (Descoteaux et al., 2011). Because diffusion is symmetric (Tuch, 2002), one can reduce acquisition time by half if only the half-space is acquired, resulting in 257 directions (Hagmann et al., 2008). The missing half is then obtained by applying mirror symmetry through the q-space origin.

Dictionary learning. Sparse coding, equivalently referred to as dictionary learning, applied to diffusion weighted images such as DSI data reveals the latent structure of the diffusion in white matter voxels. However, sparse coding is not compressed sensing (CS). Compressed sensing consists of three ingredients: a linear sensing process, a linear transformation to the data that generates sparsity and is incoherent to the sensing basis, and a solver used for signal recovery that promotes sparse estimates, *e.g.* using ℓ_1 norm or ℓ_0 non-linear pseudo-norm Donoho (2006). There are further conditions on the sensing and the transformation for CS to work. Our technique is not a CS setup, nor does it involve any of the theoretical properties required by CS. Sparse coding offers a way to learn a code book, a dictionary, that can be used to accurately approximate the diffusion signal with a few dictionary elements, also called atoms. This dictionary forms a data-driven model for the diffusion signal. A parallel can be made between such a dictionary and parametric decompositions such as wavelets, short time Fourier transforms or spherical harmonics often used for, respectively, images, audio signals and data defined on a sphere. A relevant example of data defined on the sphere is the angular distribution estimated by HARDI. Here however, the dictionary is data-driven. Given a model of the signal, such as one given by a parametric or data-driven dictionary, it is possible to denoise it or solve inverse problems such as deconvolution or super-resolution, which amounts to performing undersampled, and therefore faster, acquisitions.

Background. Several recent works have applied sparse techniques with reduced sampling to DWI (Lee and Singh, 2010; Merlet and Deriche, 2010; Cheng et al., 2011; Aboussouan et al., 2011; Michailovich et al., 2011; Menzel et al., 2011; Rathi et al., 2011; Dolui et al., 2011; Landman et al., 2012; Bilgic et al., 2012b; Mani et al., 2012; Paquette and Descoteaux, 2012; Ye et al., 2012). These techniques can be separated into methods for DTI and multi-tensor techniques (Landman et al., 2012; Mani et al., 2012; Paquette and De-

scoteaux, 2012), methods for single-shell HARDI (Cheng et al., 2011; Michailovich et al., 2011; Dolui et al., 2011) and the rest of methods designed for DSI (Lee and Singh, 2010; Merlet and Deriche, 2010; Aboussouan et al., 2011; Menzel et al., 2011; Bilgic et al., 2012b) and multiple-shell HARDI (Rathi et al., 2011; Cheng et al., 2011). Although these works are very preliminary, it seems to be possible to obtain good quality DSI data from undersampled measurements.

Regarding denoising, which is the other focus of this paper, groups in the DWI denoising community have first denoised DW data assuming a Gaussian noise on each separate DWI channel (Manjòn et al., 2008). These techniques were then extended to take into account the Rician noise nature of the DWI signal (Coupé et al., 2010; Descoteaux et al., 2008; Aja-Fernàndez et al., 2008; Tristán-Vega and Aja-Fernández, 2010; Aja-Fernández et al., 2011; Brion et al., 2011) and, recently, the non-central Chi-squared distribution in the case where parallel imaging is used (Aja-Fernández et al., 2011; Brion et al., 2011). However, only the technique of (Tristán-Vega and Aja-Fernández, 2010) performs denoising across the DWI channels, i.e. considering all DW images within a certain angular cone around the DW image being denoised. This was done using non-local (NL) means and linear minimum mean square error (LMMSE). Our approach is different from the aforementioned denoising techniques. Our technique considers all DWI channels, searches over all voxels to learn the atoms that best describe the underlying structure. We simultaneously use all directions and b-values to estimate the underlying structure of the data. We also succeed in combining multiple images corrupted by different noise levels in a common estimation procedure, thanks to a proper whitening of the data.

Dictionary learning has recently started to appear and be used in dMRI (Merlet et al., 2012a; Ye et al., 2012; Bilgic et al., 2012b,a). Previous contributions proposed using solvers inspired by K-SVD. The original K-SVD relies on orthogonal matching pursuit (OMP), for which efficient implementations are freely available, and singular value decompositions (SVD). In (Merlet et al., 2012a), the atoms learnt are parametric diffusion profiles. The dictionary updates are therefore not done with an efficient SVD but a non-linear optimization over the parameter space of a family of profiles. This makes the estimation particularly slow. In (Ye et al., 2012), K-SVD is used for dictionary learning jointly with a non-local means for spatial smoothing and regularization. The joint estimation of the diffusion profiles and the spatial field is of interest, although the paper does not address the issue of hyperparameter selection, as well as computation time to scale to full brain data. In (Bilgic et al., 2012b), a K-SVD type method is employed with the sparse coding step performed using iterative least squares (FOCUSS). The estimated dictionary is then used to accelerate DSI acquisition using an undersampled q-space. The work of (Bilgic et al., 2012b) is in close spirit to ours. However, our contribution improves over this work on four points: first, our problem statement is truly based on the physical constraints of DSI acquisition (posi-

tivity and symmetry); second, we use a state-of-the-art online sparse coding (Mairal et al., 2010) technique that outperforms K-SVD in terms of computation time and memory usage; third, we justify the quadratic data fit term used in the minimized cost function thanks to a preprocessing whitening procedure; and fourth, we propose a principled model selection procedure based on cross-validation in order to avoid overfitting. One should mention the recent contribution in Ho et al. (2013), that tackles the problem of dictionary learning in non-euclidean spaces, such as the Riemanian manifold of symmetric positive-definite matrices. This work is adapted to the estimation of a code book on the reconstructed diffusion signal, rather than the raw DSI as we propose here.

Sparse coding has been used to speed up structural MRI acquisition (Ravishankar and Bresler, 2011). However, to the best of our knowledge, this is the first time sparse coding is used on raw DWI with a model that takes into account symmetry and positivity. Hence, our approach makes several contributions. First, we show how we can improve the quality of DSI data by estimating better antipodal q-space signals compared to the ones obtained by symmetry of the acquired DSI data, as classically done in public software such as *TrackVis (Diffusion Toolkit)*, public datasets as provided by the *Pittsburgh Brain Competition 2009* and in several connectomics publications (Hagmann et al., 2008; Honey et al., 2009). The field of ODFs is then denoised and its angular structure is enhanced. Second, we show that once a DSI dictionary is learnt for a particular acquisition protocol, it can be used on other subjects with highly undersampled DSI acquisitions. We show that using only 40 q-space measurements performs as well as state-of-the-art NL-means denoising, which already improves over naive half-space symmetric acquisition. Therefore, using the learnt DWI structure, our technique fills-in the missing q-space measurements. In the computer vision and image processing community, this is called *inpainting* and can also be presented as DSI *super-resolution*.

3. Methods

Notations. We write vectors in bold, $\mathbf{a} \in \mathbb{R}^n$, matrices with capital bold letters, $\mathbf{A} \in \mathbb{R}^{n \times n}$. A scalar a is positive if $a \in \mathbb{R}_+$. We denote $\|\mathbf{A}\|_{\text{Fro}}$ the Frobenius norm, $\|\mathbf{A}\|_{\text{Fro}}^2 = \sum_{i,j=1}^n \mathbf{A}_{ij}^2$, and $\|\mathbf{A}\|_1 = \sum_{i,j=1}^n |\mathbf{A}_{ij}|$ the ℓ_1 norm. Column i of a matrix is written $\mathbf{A}^i$. If $\mathcal{I}$ is a list of $|\mathcal{I}|$ indices, $\mathbf{A}_{\mathcal{I}}$ is the matrix $\mathbf{A}$ restricted to the rows in $\mathcal{I}$. $\mathbf{I}$ stands for the identity matrix. Quantities estimated from the data are written $\widehat{\mathbf{A}}$. A matrix with non-negative elements is denoted $\mathbf{A} \geq 0$.

3.1. Dictionary learning for DWI

We start by describing the dictionary learning model and its assumptions when applied to dMRI data. We then discuss the strategy employed to calibrate model parameters.

Problem statement. A dataset for one subject consists of d DWI volumes containing p voxels. Each volume corresponds to a direction and a b-value in the q-space. Only voxels within the brain volume are considered. Let $\mathbf{S} \in \mathbb{R}_+^{d \times p}$ denote the data. DWI have positive values.

Generative model. We consider the following model for the signal at voxel i:

$$\mathbf{S}^i = \mathbf{D}\mathbf{W}^i + \mathbf{e}^i,\ 1 \leq i \leq p\ , \tag{3}$$

with dictionary $\mathbf{D} \in \mathbb{R}_+^{d \times k}$ and the coefficients $\mathbf{W} \in \mathbb{R}_+^{k \times p}$. The integer k is the number of dictionary elements, also called *atoms*. The noise $\mathbf{e}^i \in \mathbb{R}_+^d$ is known to have a Rician distribution, for single channel imaging, or a non-central Chi-squared distribution when parallel imaging (Aja-Fernández et al., 2011; Brion et al., 2011) is used. However, in this work, the noise will be assumed to be Gaussian with mean $\mu \in \mathbb{R}^d$ and diagonal covariance $\mathbf{\Sigma} = \text{diag}((\sigma_j)_{j=1,\ldots,d}) \in \mathbb{R}^{d \times d}$ because we have a high enough SNR in our synthetic and real datasets. We denote it $\mathbf{e}^i \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. This modeling assumption is discussed and justified below.

The estimation procedure detailed below uses the standard and computationally convenient ℓ_2 norm for the data fit. Such a norm makes the implicit assumption of an additive noise that is Gaussian with unit variance (see examples of Maximum a Posteriori (MAP) estimation with Gaussian white noise). To meet this constraint, given our modeling assumption on $\mathbf{e}^i$, we define the whitened data $\mathbf{S}_w^i = \mathbf{\Sigma}^{-1/2}(\mathbf{S}^i - \mu)$ so that $\mathbf{S}_w^i \sim \mathcal{N}(\mathbf{D}_w\mathbf{W}^i, \mathbf{I})$ where $\mathbf{D}_w = \mathbf{\Sigma}^{-1/2}\mathbf{D}$. In practice, μ and $\mathbf{\Sigma}$ are estimated from voxels in the background of the images. This whitening step is crucial to estimate a joint model of data with both high and low b-values that have very different signal to noise ratios (SNR).

In order to learn the factorization $\mathbf{DW}$, one needs to set priors on both $\mathbf{D}$ and $\mathbf{W}$.

- Sparse $\mathbf{W}$ *i.e.,*: the signal in each voxel can be modeled by a linear combination of a few atoms.
- Positivity constraint $\mathbf{D} \geq 0$ and $\mathbf{W} \geq 0$: each atom representing a diffusion profile is positive and the linear combination is only cumulative (no compensation with negative weights).
- Unit norm constraint on the columns of $\mathbf{D}$ to avoid scaling ambiguity.

Given a set of DSI volumes, a model satisfying the constraints above can be estimated

by minimizing the following cost function:

$$(\widehat{\mathbf{D}}_w, \widehat{\mathbf{W}}) = \underset{\mathbf{D},\mathbf{W}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{S}_w - \mathbf{D}_w\mathbf{W}\|_{\text{Fro}}^2 + \lambda\|\mathbf{W}\|_1 \quad (4)$$
$$\text{s.t. } \|\mathbf{D}^k\|_2^2 \leq 1, \mathbf{D} \geq 0, \mathbf{W} \geq 0$$

The parameter λ balances the reconstruction error and the ℓ_1 regularization term. A high value of λ will promote sparser estimates of $\widehat{\mathbf{W}}$. Following (Mairal et al., 2010), we use an online cyclic descent to minimize Eq. 4. The optimization algorithm requires setting two extra parameters: the batch size and the number of iterations, which corresponds to the number of times the algorithm goes over the data. The batch size refers to the number of voxels that are simultaneously fed to the solver before updating the dictionary. The benefit of this online procedure is that the solver can work without storing the entire dataset in memory, only storing small batches of voxels instead. This makes the solver very scalable, as required for full brain data. In practice, each batch contains a random set of voxels to avoid biaising the solver with neighbouring voxels, which can be too similar in their structure. Experimental tests showed that both parameters, if high enough, did not have a significant influence on the results. On full brain data, a batch size set to 500 voxels and a number of iterations set to 100 turned out to be a good compromise between computation time and stability of the estimated dictionary. For visualization, the estimated dictionary $\widehat{\mathbf{D}}$ will be unwhitened $\widehat{\mathbf{D}} = \mathbf{\Sigma}^{1/2}\widehat{\mathbf{D}}_w$.

Denoising and subsampling. Once the dictionary $\mathbf{D}$ is learnt, one can use it to decompose a new set of data, eventually using a subsampling of the directions. Let us denote $\mathcal{I}$ a list of sampling directions and $n_{\mathcal{I}}$ its cardinal. Given a set of whitened subsampled data $\mathbf{S}_{\mathcal{I}w}$, the coefficients $\mathbf{W}$ can be obtained by solving:

$$\widehat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}} \frac{1}{2n_{\mathcal{I}}}\|\mathbf{S}_{\mathcal{I}w} - \mathbf{D}_{\mathcal{I}w}\mathbf{W}\|_{\text{Fro}}^2 + \nu\|\mathbf{W}\|_1 \quad (5)$$
$$\text{s.t. } \mathbf{W} \geq 0$$

where $\nu > 0$ balances the reconstruction error and the ℓ_1 regularization term. The optimization is done using the same ℓ_1 solver used in the dictionary learning phase. The full signal can then be obtained as: $\widehat{\mathbf{S}} = \widehat{\mathbf{D}}\widehat{\mathbf{W}} \in \mathbb{R}^{d \times p}$.

Given a data-driven, high resolution dictionary, one can therefore perform DSI estimation and ODF reconstruction from undersampled q-space data. Note that $\mathbf{W}$ estimated in (4) is different from $\mathbf{W}$ in (5), as here we have access to a limited number of DWIs. We therefore do not reestimate a full DSI dictionary, and only minimize over $\mathbf{W}$. The rationale

for the use of a different regularization parameter, here denoted ν, is because its setting depends in practice on the number of DWIs and the noise level. In the following experiments the q-space subsampling is not optimized, although the dMRI literature teaches us that the sampling strategy is crucial for optimal reconstructions (Merlet et al., 2011; Caruyer and Deriche, 2012). It is beyond the scope of this paper to explore such optimal sampling strategies in the context of the proposed technique. In the following results, we undersampled the Cartesian directions with a regular interval between neighboring indices in $\mathcal{I}$. Deciding to use half of the directions would amount to take every even index. In practice, the subsampling was not always done with an integer division (see below). The consequence of this strategy is that it preserves an approximately uniform Cartesian sampling.

Figure 1 illustrates several subsampling of measurements used in our experiments. The samples are projected onto the sphere and colored according to their b-value in q-space. Although the scope of this work is not to find an optimal subsampling strategy, one can see that the samples remain relatively uniform on the sphere and across b-values.

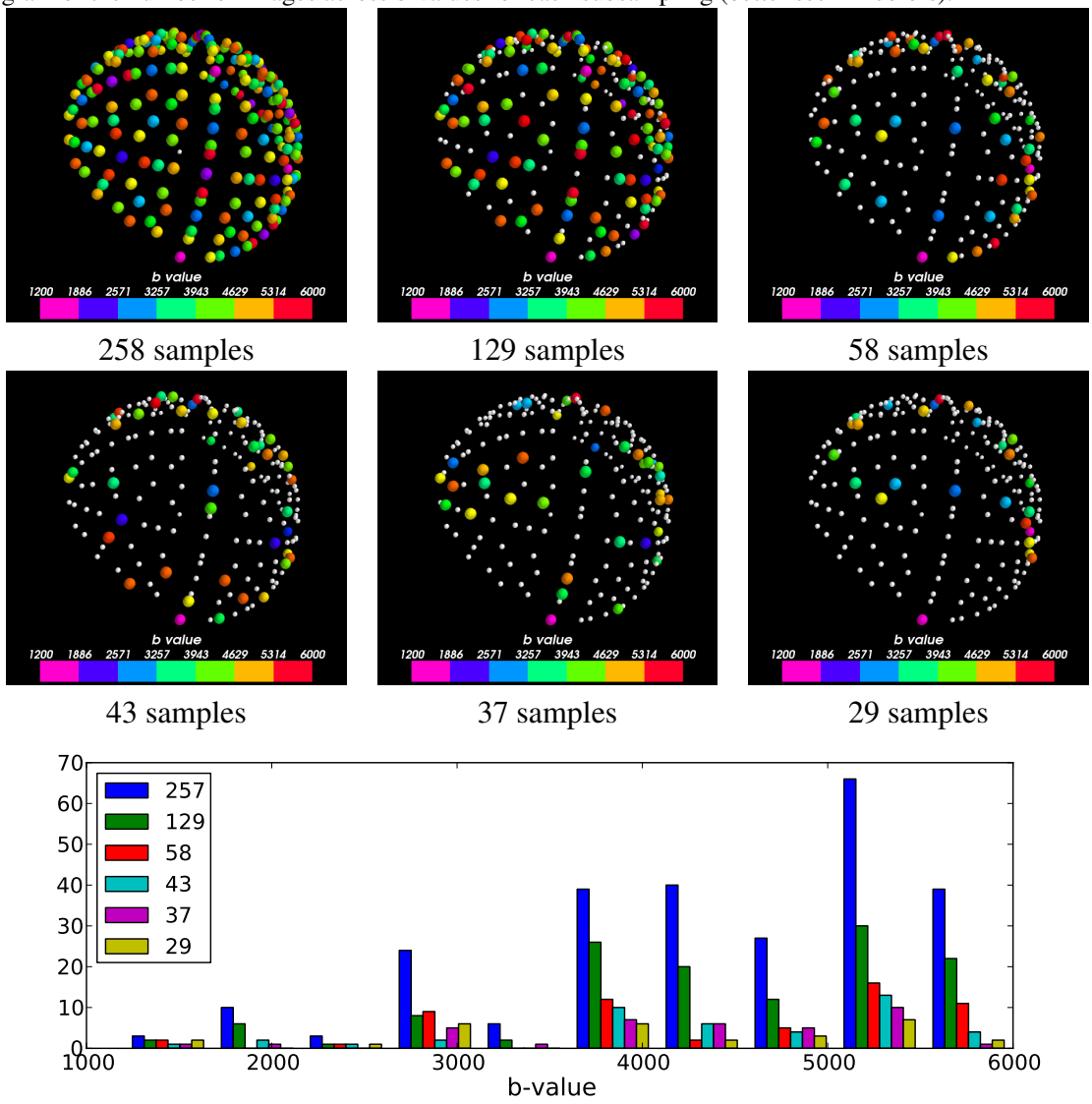
Noise parameters. The noise parameters $(\mu, \mathbf{\Sigma})$ are obtained from regions, where the diffusion signal is assumed to be only noise. On real data, we use a ROI in the background of the images (outside the brain) and have made sure that there are at least 1,000 voxels in the ROI. For the following simulation results, we used extra simulated voxels with no diffusion.

Model selection and parameters estimation. The estimation procedure involves some parameters, namely: λ to learn the dictionary (4) and ν to estimate the weights given a learnt dictionary (5). A principled way to tune these parameters is to use cross-validation. The idea is to learn a model on a fraction of the data and see how well it can explain, or fit, the rest of the data. Testing on data unused in the model estimation is done to avoid overfitting.

A natural way of splitting the data is to only use some voxels for model estimation, and evaluate the model on the left-out voxels. However, the physical properties of the problem give some insights on properties of the signal that can be exploited for better model selection. The diffusion process is symmetric, meaning that two DWI recorded with opposite gradient directions should be the same in the absence of noise. This suggests that a model estimated on half of the directions $\mathcal{H}$ (d=258 directions) can be used to estimate a full dictionary of $2d - 1 = 515$ directions. The minimization of Eq. 4 restricted to the data $\mathbf{S}_\mathcal{H}$ gives $\widehat{\mathbf{D}}_\mathcal{H} \in \mathbb{R}^{d \times k}$, which can be used to generate $\widehat{\mathbf{D}}_\mathcal{H} \in \mathbb{R}^{2d-1 \times k}$ by applying mirror symmetry through the q-space origin.

In order to assess the quality of the model without overfitting, the model selection involves two other sets of directions: $\mathcal{L}$ to learn $\widehat{\mathbf{W}}$ and $\mathcal{T}$ to test the reconstruction error, as illustrated in Fig. 2. The best parameters λ and ν minimize this error. It is quantified

Figure 1: Sampling schemes represented over a single shell where color codes for the b-values and a histogram of the number of images across b-values for each subsampling (better seen in colors).



with the average root mean square error (RMSE):

$$\text{RMSE} = \frac{1}{p|\mathcal{T}|}\|\mathbf{S}_{\mathcal{T}} - \widehat{\mathbf{D}}_{\mathcal{T}}\widehat{\mathbf{W}}\|_{\text{Fro}}^2. \tag{6}$$

Assuming the dictionary is available, the model selection only involves the estimation of ν, which only requires splitting the available data in 2 sets: $\mathcal{L}$ and $\mathcal{T}$.

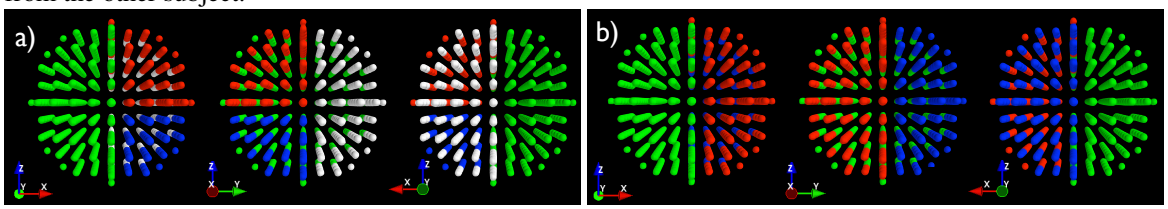
In the following experiments, the parameter λ was chosen in a range of five values $(1, 0.1, 0.01, 0.001, 0.0001)$ and ν in a logarithmic grid of 15 values between 0 and $1e^{-6}$. In all the experiments, we made sure that the estimated λ and ν were inside these parameter ranges. In order to quantify the performance of our method, we use as baseline the solution that consists in applying a simple mirror symmetry to the data, as done classically (Hagmann et al., 2008; Honey et al., 2009). We denote the error obtained as RMSE_{sym} . We report the quality of our solution as a ratio between the two quantities:

$$\rho_{RMSE} = \frac{\text{RMSE}_{sym}}{\text{RMSE}} \tag{7}$$

A ratio above 1 indicates an improvement with respect to a symmetrization.

The following results involve two setups. An intra-subject denoising procedure and an inter-subject procedure. We use the term *intra* when a dictionary is learnt on a subject and this dictionary is used on this *same* subject. We use the term *inter* when a dictionary from a subject is used to reconstruct the DSI signal of a *different* subject while undersampling.

Figure 2: The 258 directions $\mathcal{H}$ used to estimate a dictionary are in green. a) For intra-subject cross-validation, the directions $\mathcal{L}$ used to estimate the coefficients are in red, and the left out directions $\mathcal{T}$ used to evaluate the model parameters are in blue. The white directions $\mathcal{V}$ are only used for validation. b) For inter-subject cross-validation the same color code applies. There is no validation set in this case, as the validation data is obtained from the other subject.



3.2. Synthetic Data Simulation

The simulation data consists of the 3D structured field presented at the *HARDI reconstruction challenge 2012*[7]. This structured field synthetic phantom consists in a 3D volume made up of 5 slices of 16x16 voxels, and contains 3 different fiber bundles: two fiber bundles crossing in the diagonals of the 16x16 slice at 90 degrees crossed by a third circling

[7] http://hardi.epfl.ch/static/events/2012_ISBI/

and bending bundle, as seen on the ground truth ODFs of Fig. 3. In every voxel of the dataset, the diffusion signal corresponding to the underlying fiber configuration is simulated according to the same gradient list as our real DSI data. The diffusion signal, $S(\mathbf{q})$, is simulated using the classical Gaussian mixture model (Tuch, 2004; Descoteaux et al., 2007; Canales-Rodríguez et al., 2009):

$$S(\mathbf{q}) = \frac{E(\mathbf{q})}{E_0} = \sum_{i=1}^{M} f_i \exp(-b\,\mathbf{u}^T \mathbf{D}_i \mathbf{u}) + \epsilon, \tag{8}$$

M is the number of fiber compartments, each one characterized by a self-diffusion tensor $\mathbf{D_i}$ and volume fraction of f_i, such that $\sum_{i=1}^{M} f_i = 1$. In the contest data, $E_0 = 1$ without loss of generality and diffusivities were generated using symmetric tensors, $\mathbf{D_i} = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$, in the range of $\lambda_1 \in [1, 2] \times 10^{-3}$ mm^2/s and $\lambda_2 = \lambda_3 \in [0.1, 0.6] \times 10^{-3}$ mm^2/s, as done in (Canales-Rodríguez et al., 2009). The datasets were corrupted with additive Rician noise, ϵ. We investigated two noise scenarios similar to the $b = 0$ image of our real datasets; one with high SNR and another with mid-range SNR (SNR $= 1/\epsilon$), namely 36 and 18, which are realistic noise levels similar to our real data acquisitions. Noiseless data is presented in Fig. 3.

Performance of the method on simulated data was quantified by different measures: the difference in the number of fiber compartments (DNC) and the angular error (AE) with respect to the known ground truth. To compute the DNC and the AE, we extract the maxima on a discrete grid with $N = 4000$ uniform points (Descoteaux et al., 2007) for the estimated ODFs and compare them to the ground truth maxima. Then, the DNC becomes the mean difference between the number of maxima extracted on the estimated ODFs and the true number of maxima, and the AE is computed between the maxima extracted on the estimated ODFs and the respective maxima within the ground truth. The ODF maximas were defined as points maximal inside a 20° radius and with a value ≥ 0.5 on the min-max normalized ODF. The AE is computed by finding the best match between the detected peaks and the true peaks. This approach is slighltly biased towards underestimation of the AE when DNC is not 0. A full comparison and discussion on this issue can be found in the *HARDI reconstruction challenge 2012*[8].

The accuracy of the maxima reveals the ability of the method to extract the high spatial frequencies that are particularly useful for fiber tracking, while a measure such mean squared error reveals the ability to capture the full DSI signal, both its radial and angular parts.

[8] http://hardi.epfl.ch/static/events/2012_ISBI/

3.3. DSI Data Acquisitions

Two datasets were used for performance assessment and validation. The first consists of two subjects with full DSI acquisitions (515 DWIs) that were used to learn dictionaries of DSI atoms. The second is the publicly available Pittsburgh Brain Competition 2009 DSI datasets, which are used for the denoising experiment only.

DSI data used for dictionary learning. The data consists of two subjects. A standard DSI acquisition mimicking the original DSI protocol (Wedeen et al., 2005) was done in a 3 T Trio MR Siemens system, equipped with a whole body gradient (40 mT/m and 200 T/m/s) and a 32 channels head coil. Isotropic 2 mm spatial resolution and $d = 515$ DW measurements were acquired, comprising q-space points of a cubic lattice within the sphere of five lattice units in radius (see (Wedeen et al., 2005)). TE/TR=147 ms/11.5 s, BW=1680 Hz/pixel, 96x96 matrix, 60 axial slices with a parallel reduction factor of 2, δ and Δ were 41 and 45 ms, $b_{\max}$ = 6000 s/mm^2. The 515 q-encoding values were set in the middle of a cubic lattice of size 17x17x17, with values outside the acquired 515 values padded to zero. The total time for this acquisition was 1h40 minutes.

The SNR of the $b = 0$ image was 36 and the SNR of the DWI for the $b = 960, 3360,$ and 6000 s/mm^2 datasets were estimated to 12, 7.5, and 6.5 respectively. These SNR values were computed using two manually selected regions of interests (ROI). The first region was selected in the background of the DWI, and the second was segmented in the white matter and corresponded to the corpus callosum (CC). The SNR was then computed as SNR=mean(signal_CC)/std(noise), where mean(signal_CC) corresponds to the mean signal in the CC ROI, taken in the DW image which corresponds to the gradient direction aligned with the CC fiber direction (left-right). std(noise) is the standard deviation of the noise in the ROI taken in the background. In a sense, this corresponds to a worst case SNR because it is in the CC region that the diffusion signal is most attenuated (see last row of Fig. 6). We see that the lowest SNR is greater than 6, which implies that the Rician noise distrubution is well approximated by a Gaussian distribution in this dataset. This procedure is also described in (Descoteaux et al., 2011; Jones et al., 2012).

Pittsburgh Brain Competition 2009 datasets. In 2009, there was a fiber clustering competition held at the Human Brain Mapping (HBM) conference. This competition, known as the Pittsburgh Brain Competition (PBC)[9], provided the community with three DSI datasets (*brain0, brain1, brain2*), amongst numerous other raw and processed datasets. These datasets were also generated in a 3 T MR Siemens system with a WIP pulse program, equipped with a 32 channels head coil. The specific imaging parameters were TE/TR=160ms/9.12 s,

[9] http://pbc.lrdc.pitt.edu/

96x96 matrix, isotropic 2.4 mm resolution, 50 axial slices with a parallel reduction factor of 2 and a maximum b-value of $b_{\max} = 7000$ s/mm^2. These DSI datasets include 515 measurements, but only half the cubic lattice is acquired with duplicates (half the Cartesian hemi-plane with 257 grid points acquired twice and one $b = 0$ at the origin). Hence, the datasets contain only half the grid point measurements of our other datasets. The data is thus symmetrized and then averaged to reduce noise effects. Hence, the PBC data cannot be used for learning. It will only be used for denoising experiments.

Symmetry and denoising comparisons. For the rest of this paper, acquisition of 258 directions with simple mirror symmetry will be abbreviated HALF, as opposed to the FULL acquisition using 515 measurements. Different symmetry completion procedure are also included for comparison, either using Gaussian smoothing with $\sigma = 0.35$ (optimal σ in our experiments which corresponds to 2.5 mm FWHM) and state-of-the-art non-local (NL) means denoising (Descoteaux et al., 2008; Coupé et al., 2010) (11x11x11 search volume and 26 neighbors local neighborhood). The denoising techniques are common and there are publicly available implementations that one can run on the raw diffusion data[10]. Using such denoising techniques, the experiment will consist in denoising half of the data and then symmetrizing it. The better the fit is with the other half of the data, the better the denoising will be.

Visualization. To visualize the reconstructed DSI profiles, we show the raw diffusion signal along a particular DW direction (as in Fig. 6) or the diffusion ODF computed from the DSI signal (Wedeen et al., 2005; Descoteaux et al., 2011), only for voxels in a white matter mask as computed from a thresholded FA map at 0.1 (Figs. 6-8 in particular). ODFs are finally visualized as deformed spheres, with radius deformed proportional to its ODF value. Fields of ODFs are usually visualized with a classical RGB (red-green-blue) colormap to highlight patient left-right (red), anterior-posterior (green), and inferior-superior (blue) directions. Moreover, these fields of ODFs are always overlaid on a slice of generalized fractional anisotropy (GFA) to highlight the preservation of the anisotropy and high spatial frequencies while undersampling. GFA = STD(Ψ)/RMS(Ψ), where STD and RMS stand for the standard deviation and root mean square of the ODF (Tuch, 2004).

4. Results

4.1. Simulation results

ODF reconstruction. Figure 3 qualitatively illustrates results on the simulation data. The ground truth ODFs are shown in the right of the top row and the noise-free full DSI ODFs

[10] https://www.irisa.fr/visages/benchmarks/

are shown on the left of the second row. In fact, this full noise-free DSI ODF is our gold standard and is what our reconstruction based on sparse coding is competing against. The motivation here is that we assess the error due to the subsampling of the q-space DSI grid, and not the approximation error due to the DSI itself. Note that the full DSI data does not perfectly recover all angular profiles. The ground truth has a better angular resolution and most of the lower volume fraction crossing compartments (at the boundary of the middle square structure) are missed by the gold standard.

The top left subfigure of Fig. 3 shows the diffusion ODF of the dictionary learnt using 49 atoms on the mid-range SNR 18 simulation. Atoms are ordered from left to right, starting at the bottom left corner based on the variance they explain on the data. We note important single-fiber profiles covering the 6 diagonals and horizontals orientations. We also note several pure 90 degrees crossing ODF profiles, and a smaller number of lower angle crossing or wider single-fiber profiles arising. The dictionary clearly captures the angular information contained in the gold standard.

The second and third rows of Fig. 3 show the reconstruction obtained with sparse coding at different level of undersampling (258, 58, 37, 29 and 18 measurements) with mid-range noise SNR 18. The subsampling schemes are shown in Fig. 1. We can observe that most of the ODF profiles are well recovered all the way down to 18 measurements, with the exception of the challenging crossings, which are already lost on the gold standard. However, note that in these complex crossing configurations, our approach with 37 or more measurements seem to perform better than the gold standard at reconstructing some of the lower volume fraction crossings.

Quantitative experiment. Figure 4 shows the results in terms of angular information for high and low SNRs using 49 atoms in the dictionary. In Fig. 4a) and b), errors are shown in terms of DNC as a function of the number of DWI images available for full DSI estimation, for different SNRs and number of atoms in the dictionary. We report the percentage of voxels for which the number of maxima recovered is incorrect. For the correctly recovered maxima, the error can be further quantified in terms of AE. The average AE when compared to the AE obtained with all noiseless 515 DSI is below 1.5 degrees when using more than 43 DWIs at SNR 36 and when using more than 58 DWIs at SNR 18. In other words, the subsampling deteriorates the angle estimation by less than 1.5 degrees on average.

In Fig. 4 c), black pixels illustrate perfect agreement between ground truth phantom and the noiseless DSI, whereas pink pixels show where the noiseless DSI underestimates one maxima with respect to the ground truth (single fiber detected instead of a two fibers crossing). This is visually confirmed in Fig. 3. In d)-h), the comparison is between our sparse coding approach with different undersamplings and the noiseless DSI. The noiseless DSI acts as our gold standard. Here, pink and red pixels mean that our approach has

Methods	Gaussian	NL means	SC DSI for various values of k					
	$\sigma = 0.35$		$k = 100$	169	225	400	900	1600
Subject 1	1.16	1.19	1.31	1.31	1.31	1.31	1.31	1.31
Subject 2	1.13	1.16	1.28	1.25	1.23	1.30	1.30	1.29

Table 1: Intra-subject denoising. ρ_{RMSE} (Eq. 7) between simple DSI symmetry, Gaussian smoothing, NL means and our sparse coding (SC) based denoising using k atoms. Reconstruction obtained with sparse coding gives the best performance on the validation data.

detected a single fiber instead of a two or three fibers crossings respectively, whereas blue pixels show where our approach with undersampling is actually *better* than the noiseless DSI. There, we recover the correct two bundles crossing, as in the ground truth, whereas the noiseless DSI had just found a single fiber (seen in c)). This explains the gain in maxima detection performance and angular resolution of our approach.

4.2. Real DSI data

Figure 5 shows the ODFs corresponding to the dictionary learnt with 100 atoms on both subjects. Atoms are ordered from left to right, starting at the bottom left corner based on the variance they explain on the data. We see that the most important atoms are smooth profiles and mostly single fiber structures. After approximately 30 atoms, crossing profiles appear. At the end of the dictionary, more complex ODF profiles are also present. This behavior of the learnt dictionary is similar if we increase its size k. It is interesting to observe that dictionaries obtained on two different subjects reconstruct some ODF profiles that look surprisingly similar. The similarity between dictionaries is however not only visual, as confirmed by the quantitative performance evaluation below.

Intra-subject denoising. Table 1 shows how the reconstruction obtained with sparse coding is able to accurately reconstruct the 257 unmeasured directions. The quantity represented is ρ_{RMSE} (cf. Eq. 7), for which a value larger than one means that the DWIs reconstructed on the other half of the the q-space are more accurate than the ones obtained with the usual symmetrization.

One can observe that ρ_{RMSE} is consistently above one for all denoising methods, but that our approach based on dictionary learning outperforms Gaussian smoothing and NL means for all dictionary sizes and for both subjects. Increasing the number of atoms in the dictionary only slightly improves the accuracy on subject 2. Hence, for the rest of the experiments, we use dictionaries with 100 atoms. These dictionaries are illustrated in Fig. 5.

In order to visualize how the raw DWI data is denoised by the different techniques, Fig. 6 shows a comparison of results on a region of interest chosen in the centrum semio-

vale, in mid-coronal slice. It can be observed that NL means (row 3) and Gaussian (row 4) denoising tend to over-smooth and blur the anatomical structures of the raw data, as opposed to our dictionary learning approach (row 1) that succeeds in denoising, but also enhancing the structure by preserving the high spatial frequencies (see red box).

Finally, Fig. 7 overlays ODFs to GFA maps in the centrum semiovale, where the corpus callosum (CC) crosses with the corticospinal tract (CST) in-plane and the superior longitudinal fasciculus (SLF) out of plane. The cingulum (Cg) is also visible with a green single fiber population out of the plane. Single, two and three fibers crossings can be seen. One can observe how Gaussian smoothing and NL Means over-smooth GFA maps, but also appreciate how our dictionary learning method is able to recover ODF profiles as sharp as the FULL raw DSI.

Inter-subject undersampling and sparse coding. One can push the sparse coding approach one step further than raw DSI denoising. If one has a dictionary estimated on a subject, it is possible to undersample the q-space when acquiring data for another subject, and subsequently use the dictionary to reconstruct a full resolution DSI signal. Figs. 8 and 9 show the reconstructed ODFs and the reconstruction errors as a function of the number of measurements using different dictionary/subject pairs, when using sparse coding.

First, Fig. 8 shows ODFs estimated from an undersampled reconstruction for subject 2, using the dictionary learnt on subject 1 with 100 atoms (Fig. 5 left). It is impressive to see that a DSI dictionary learnt on a subject can be used to perform undersampled DSI on a different subject.

Figure 9 shows the RMSE ratio between simple HALF DSI with symmetry and our sparse coding approach as a function of the number of measurements. The baseline naive symmetry is illustrated with the pink line, whereas we also show Gaussian smoothing and NL means error ratios for comparison.

Of course, as the number of samples decreases, the overall field of ODFs seems more noisy, but the overall RMSE remains acceptable, as seen in Fig. 9. At a total of 37, 29, and 21 respective measurements, the results become worse than NL means, Gaussian smoothing and simple symmetry DSI respectively, in terms of RMSE. On the other hand, we observe that ODF profiles are degraded more quickly as a function of undersampling. Note that the structured voxels with a single fiber orientation in the Cg, CC, CT and SLF are well preserved all the way down to 29 measurements. However, although crossings are found in all undersampling schemes, ODF peaks in crossing areas become less accurate below approximately 43 measurements.

Pittsburgh Brain Competition datasets denoising and undersampling. Finally, Fig. 10 shows that undersampled sparse DSI using a dictionary learnt on subject 2 with 100 atoms can

reconstruct the DSI signal and diffusion ODFs of *brain0* Pittsburgh Brain Competition. ODFs are again shown in a region of the centrum semiovale, where we see crossing configurations between the CC, CST and SLF, as well as single fiber profiles from the Cg. One can see that our data-driven approach based on sparse coding preserves crossings, angular resolution and structure of the ODF field while considerably undersampling. As a consequence, one also sees that the underlying GFA maps are preserved while undersampling. Qualitatively, the 37 measurements reconstruction is of similar quality as the original DSI data. Moreover, as before, at 29 measurements, single fiber ODFs are well-preserved but crossings become more disorganized and angular resolution is lost. Although not shown here, similar qualitative results are obtained on *brain1* and *brain2*, with dictionaries learnt on subjects 1 and 2.

5. Discussion

When applied to diffusion MRI data, sparse coding, a.k.a. dictionary learning, reveals the latent structure of the white matter voxels. It is important to realize that the results in this paper focused on DSI data, but the technique is applicable to diffusion-weighted imaging in general. The same technique could be used on a DTI acquisition, single-shell HARDI scheme or more advanced multiple-shell HARDI or radial sampling schemes for DWI acquisitions. Here, we have extensively studied the impact of our data-driven sparse coding approach on DSI data because there is a huge potential gain in denoising, especially at high b-values, but also a large gain in undersampling, since the classically used 515 measurements share a lot of structure.

In this work, we showed that sparse coding can be used to accurately approximate the diffusion signal with a few dictionary elements. Here, the dictionary is non-parametric and estimated from tens of thousands of voxels on full brain data, taking into account the specificities of diffusion MRI data. Proper dictionary learning was made possible on all DWIs at the same time by taking into account the varying SNR across b-values, via a dedicated whitening procedure. Positivity was enforced to model the fact that the diffusion signal in a voxel is the sum of the contributions from the different structures present in that voxel.

Exploiting the estimated model of the signal, we showed how one can denoise data and solve inverse problems such as super-resolution of q-space data. We furthermore showed that the signal model, *i.e.* the dictionary, could be learnt on another subject without degrading the signal estimation. This was highlighted through careful synthetic and real data experiments. The proposed technique is attractive thanks to its small modeling assumptions and its limited number of parameters, which are automatically calibrated using a cross-validation scheme that is based on the physical properties of the diffusion signal, namely

its symmetry and positivity. Note that, in the context of tractography, using post-processing techniques of the EAP profile, one can also produce asymmetric angular profile that reflect local Y-shaped branching present in underlying structures (Barmpoutis et al., 2008).

The key benefit of our method is its ability to perform denoising across all the DWI channels at the same time, consequently enhancing the image quality, in particular for noisy high b-values. While the technique of (Tristán-Vega and Aja-Fernández, 2010) uses DW images within a certain cone around the DW image being denoised, we propose estimating the underlying structure from all directions and b-values. This is made possible by a proper whitening of the data, which allows multiple images corrupted by different noise levels to be combined in the estimation. However, note that the learning and denoising is done voxel-by-voxel. As for non-local means denoising (Coupé et al., 2010) or more recent estimation techniques using a spatial regularization term such as total variation in the minimization problem (see [11] for several recent techniques), we could further improve the method by using a spatial model on the signal. We chose to remain at the voxel level to perform fair comparisons with standard DSI, and we leave the use of spatial regularization to the tractography algorithms.

Results have showed that using just half of the data (258 DWI), we can better predict the other 257 DWI than the classic mirror symmetry procedure. This statement also holds even if we use as little as 40 q-space measurements, as illustrated in our root mean squared error curves experiments (Fig. 9) and our good angular reconstructions (Fig. 4). Our sparse coding technique performs better than symmetrizing, Gaussian denoising or state-of-the-art NL means denoising. This is an important message for the DSI and connectomics (Hagmann et al., 2008; Honey et al., 2009) communities that are using half-space DSI acquisitions.

Finally, beyond denoising, we have showed that a dictionary learnt from one subject can be used to reconstruct a full DSI dataset from an undersampled acquisition of a different subject. Accuracy was quantified based on MSE on the diffusion data as well as orientation information extracted from ODF reconstructions. We have done so using the same MR imaging system (Siemens in this work). We expect that a different dictionary would have to be learnt for each MR system, but this learning only needs to be done once. In a way, this would act as a "calibration" step, which would not necessarily be limited to a single subject. Indeed, the dictionary learning algorithm employed is very scalable and could be done on more voxels originating from a population of subjects, healthy subjects or subjects with a condition. Hence, from now on, we could imagine acquiring approximately 40 measurements on new subjects, and use our previously learnt dictionaries to reconstruct a full DSI dataset. Authors in (Bilgic et al., 2012b) come up with similar conclusions.

[11] http://hardi.epfl.ch/static/events/2012_ISBI/workshop_program.html#proceedings

Consequently, we can have fast acquisitions and obtain high resolution DSI data. DSI can be estimated with the same amount of data and scanning time as HARDI.

The challenges that one might face in the exploration of the proposed technique, *e.g.* in clinical settings, is the difficulty to use for patients a dictionary learnt on healthy subjects. It might require a dedicated dictionary learning stage or further development, to take for example into account the issue of movements such as head rotations between scanning sessions. Future work will be dedicated at optimizing the dictionary learning parameters and the q-space subsampling scheme, in order to enhance other criteria such as ODF reconstruction. Indeed a DSI sampling on a grid might not be optimal should the angular information be the information of interest. Work will also be done to enhance the computation of other radial EAP metrics such as return-to-origin probability, mean squared displacement and kurtosis. Currently, our model selection criteria is based on a mean squared error over all signal measurements, but one can think of a different measure or weighting to better recover and preserve high frequencies needed for higher angular resolution of the ODF. Future work will also be dedicated to finding optimal measurement strategies and to finding a way to inject the noise distribution, if it is different from Gaussian. One could hope performing the subsampling in k-space to remain in the Gaussian noise regime, but this would mean that we would also lose the symmetry and positivity constraint that are crucial in our current problem statement. This remains a rich field for future investigation.

6. Conclusion

We have showed that it is possible to learn the latent structure of white matter based on the symmetry and positivity physical constraints of diffusion-weighted imaging. Using online dictionary learning, we can learn the structure of raw DWI on full brain data to either denoise the data or use the learnt dictionary to perform undersampled acquisitions, hence fast imaging.

In particular, we have showed that we can perform robust and fast DSI reconstructions from only 40 measurements. Our results show that one can learn the DWI structure from one subject and use it to denoise or perform fast imaging on a different subject. The performance of our new sparse DWI technique was investigated with a publicly available simulation dataset and publicly available DSI datasets from the Pittsburgh Brain Competition 2009 Challenge. This work opens perspectives for better denoising and faster diffusion imaging techniques.

Acknowledgments

The authors are thankful to the organizers of the Pittsburgh Brain Competition 2009 for making public three raw DSI datasets, as well as K-H. Cho, Y-P. Chao and Pr. C-

P. Lin for their participation in the DSI data acquisition. We would also like to thank Michael Paquette and Samuel St-Jean for their contributions. This work was supported by *Programme Hubert Curien Orchid 2008 franco-taiwanais* and *NSERC Discovery Grant*.

References

Aboussouan, E., Marinelli, L., Tan, E., 2011. Non-cartesian compressed sensing for diffusion spectrum imaging, in: International Society for Magnetic Resonance in Medicine (ISMRM), p. 1919.

Aja-Fernàndez, S., Niethammer, M., Kubicki, M., Shenton, M.E., Westin, C.F., 2008. Restoration of dwi data using a rician lmmse estimator. IEEE Transactions Medical Imaging 27, 1389–1403.

Aja-Fernández, S., Tristán-Vega, A., Hoge, W.S., 2011. Statistical noise analysis in GRAPPA using a parametrized noncentral Chi approximation model. Magnetic resonance in medicine 65, 1195–206.

Alexander, D.C., 2008. A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features. Magnetic Resonance in Medicine 60, 439–448.

Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y., Basser, P.J., 2008. Axcaliber: A method for measuring axon diameter distribution from diffusion mri. Magnetic Resonance in Medicine 59, 1347–1354.

Barmpoutis, A., Vemuri, B.C., Howland, D., Forder, J.R., 2008. Extracting tractosemas from a displacement probability field for tractography in DW-MRI. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 11, 9–16.

Basser, P., Mattiello, J., LeBihan, D., 1994. Estimation of the effective self-diffusion tensor from the nmr spin echo. Journal of Magnetic Resonance B 103, 247–254.

Bilgic, B., Setsompop, K., Cohen-adad, J., Wedeen, V., 2012a. Accelerated Diffusion Spectrum Imaging with Compressed Sensing using Adaptive Dictionaries, in: MICCAI 2012, Nice, France. pp. 1–9.

Bilgic, B., Setsompop, K., Cohen-Adad, J., Yendiki, A., Wald, L.L., Adalsteinsson, E., 2012b. Accelerated diffusion spectrum imaging with compressed sensing using adaptive dictionaries. Magnetic Resonance in Medicine 68, 1747–1754.

Brion, V., Poupon, C., Riff, O., Aja-Fernández, S., Tristán-Vega, A., Mangin, J.F., Bihan, D.L., Poupon, F., 2011. Parallel mri noise correction: An extension of the lmmse to non central distributions, in: Fichtinger, G., Martel, A.L., Peters, T.M. (Eds.), MICCAI (2), Springer. pp. 226–233.

Canales-Rodríguez, E.J., Melie-García, L., Iturria-Medina, Y., 2009. Mathematical description of q-space in spherical coordinates: exact q-ball imaging. Magnetic Resonance in Medicine 61, 1350–1367.

Caruyer, E., Deriche, R., 2012. A computational framework for experimental design in diffusion MRI, in: CDMRI - MICCAI Workshop on Computational Diffusion MRI, Nice, France.

Cheng, J., Merlet, S., Ghosh, A., Caruyer, E., Jiang, T., Deriche, R., 2011. Compressive sensing ensemble average propagator estimation via l1 spherical polar fourier imaging, in: International Society for Magnetic Resonance in Medicine (ISMRM), p. 1924.

Coupé, P., Manjòn, J.V., Gedamu, E., Arnold, D., Robles, M., Collins, D.L., 2010. Robust Rician noise estimation for MR images. Medical Image Analysis 14, 483–493.

Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R., 2007. Regularized, fast, and robust analytical q-ball imaging. Magnetic Resonance in Medicine 58, 497–510.

Descoteaux, M., Deriche, R., Bihan, D.L., Mangin, J.F., Poupon, C., 2011. Multiple q-shell diffusion propagator imaging. Medical Image Analysis 15, 603–621.

Descoteaux, M., Poupon, C., 2013, *in press*. Diffusion-Weighted MRI. In Comprehensive Biomedical Physics, Elsevier.

Descoteaux, M., Wiest-Daesslé, N., Prima, S., Barillot, C., Deriche, R., 2008. Impact of rician adapted non-local means filtering on hardi, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 122–130.

Dolui, S., Michailovich, O.V., Rathi, Y., 2011. Compressed sensing of diffusion mri data using spatial regularization and positivity constraints, in: ISBI, IEEE. pp. 1597–1601.

Donoho, D., 2006. Compressed sensing. IEEE Transactions on Information Theory 52, 1289–1306.

Gramfort, A., Poupon, C., Descoteaux, M., 2012. Sparse DSI : Learning DSI Structure for Denoising and Fast Imaging, in: MICCAI 2012, Part II, LNCS 7511, pp. 288–296.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. PLoS biology 6, e159.

Ho, J., Xie, Y., Vemuri, B., 2013. On A Nonlinear Generalization of Sparse Coding and Dictionary Learning, in: Proceedings of The 30th International Conference on Machine Learning, pp. 1480–1488.

Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. Proceedings of the National Academy of Sciences of the United States of America 106, 2035–40.

Jones, D.K., Knösche, T.R., Turner, R., 2012. White Matter Integrity, Fiber Count, and Other Fallacies: The Do's and Don'ts of Diffusion MRI. NeuroImage , in press.

Landman, B.A., Bogovic, J.A., Wan, H., Elshahaby, F.E.Z., Bazin, P.L., Prince, J.L., 2012. Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. NeuroImage 59, 2175–2186.

Lee, N., Singh, M., 2010. Compressed sensing based diffusion spectrum imaging, in: International Society for Magnetic Resonance in Medicine (ISMRM), p. 1697.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2010. Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research 11, 19.

Mani, M., Jacob, M., Guidon, A., Liu, C., Song, A., Magnotta, V., Zhong, J., 2012. Acceleration of high angular and spatial resolution diffusion imaging using compressive sensing, in: IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain.

Manjòn, J.V., Carbonell-Caballero, J., Juan J. Lull a, G.G.M., Mart-Bonmatì, L., Robles, M., 2008. Mri denoising using non-local means. Medical Image Analysis 12, 514–523.

Menzel, M.I., Tan, E.T., Khare, K., Sperl, J.I., King, K.F., Tao, X., Hardy, C.J., Marinelli, L., 2011. Accelerated diffusion spectrum imaging in the human brain using compressed sensing. Magnetic Resonance in Medicine 66, 1226–1233.

Merlet, S., Caruyer, E., Deriche, R., 2011. Impact of radial and angular sampling on multiple shells acquisition in diffusion mri, in: MICCAI, Springer. pp. 113–121.

Merlet, S., Caruyer, E., Deriche, R., 2012a. Parametric Dictionary Learning for Modeling EAP and ODF in Diffusion MR, in: MICCAI 2012, Nice, France. pp. 10–17.

Merlet, S., Deriche, R., 2010. Compressed Sensing for Accelerated EAP Recovery in Diffusion MRI, in: MICCAI 2010 Computational Diffusion MRI workshop, Beijing, China. pp. 14–26.

Merlet, S., Philippe, A.C., Deriche, R., Descoteaux, M., 2012b. Tractography via the ensemble average propagator in diffusion mri, in: Ayache, N., Delingette, H., Golland, P., Mori, K. (Eds.), MICCAI (2), Springer. pp. 339–346.

Michailovich, O., Rathi, Y., Dolui, S., 2011. Spatially regularized compressed sensing for high angular resolution diffusion imaging. IEEE transactions on medical imaging 30, 1100–15.

Ozarslan, E., Koay, C.G., Shepherd, T.M., Komlosh, M.E., Irfanoglu, M.O., Pierpaoli, C., Basser, P.J., 2013. Mean apparent propagator (map) mri: A novel diffusion imaging method for mapping tissue microstructure. NeuroImage 78, 16 – 32.

Paquette, M., Descoteaux, M., 2012. Sparse l1-l1 multi-tensor imaging at the price of dti, in: ISBI HARDI reconstruction workshop, Barcelona, Spain.

Rathi, Y., Michailovich, O.V., Setsompop, K., Bouix, S., Shenton, M.E., Westin, C.F., 2011. Sparse multi-shell diffusion imaging, in: MICCAI, Toronto, Canada. pp. 58–65.

Ravishankar, S., Bresler, Y., 2011. MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE transactions on medical imaging 30, 1028–41.

Seunarine, K.K., Alexander, D., 2009. Multiple fibres: beyond the diffusion tensor. In T.E.B. Behrens and H. Johansen-Berg (eds). Diffusion MRI. Elsevier.

Stejskal, E., Tanner, J., 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. Journal of Chemical Physics 42, 288–292.

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, 1459–72.

Tristán-Vega, A., Aja-Fernández, S., 2010. Dwi filtering using joint information for dti and hardi. Medical Image Analysis 14, 205 – 218.

Tuch, D., 2004. Q-ball imaging. Magnetic Resonance in Medicine 52, 1358–1372.

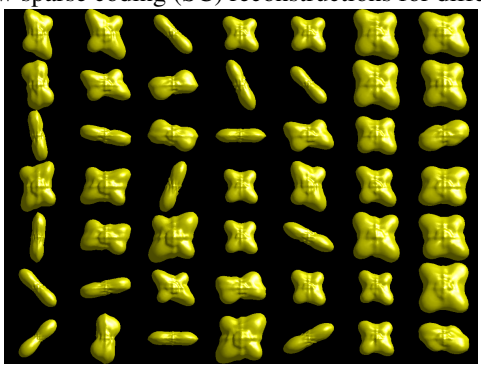
Tuch, D.S., 2002. Diffusion MRI of Complex Tissue Structure. Ph.D. thesis. Harvard University and Massachusetts Institute of Technology.

Wedeen, V.J., Hagmann, P., Tseng, W.Y.I., Reese, T.G., Weisskoff, R.M., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magnetic Resonance in Medicine 54, 1377–1386.

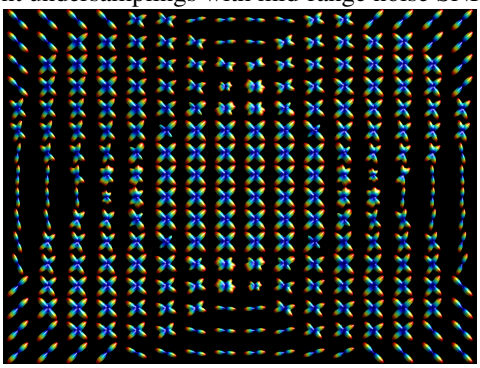
Wedeen, V.J., Rosene, D.L., Wang, R., Dai, G., Mortazavi, F., Hagmann, P., Kaas, J.H., Tseng, W.Y.I., 2012. The geometric structure of the brain fiber pathways. Science (New York, N.Y.) 335, 1628–34.

Ye, W., Vemuri, B.C., Entezari, A., 2012. An over-complete dictionary based regularized reconstruction of a field of ensemble average propagators. Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro. IEEE International Symposium on Biomedical Imaging 9, 940–943.

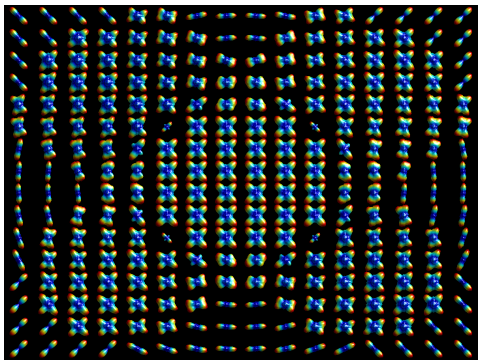
Figure 3: One slice of ground truth ODFs and DSI noiseless data exhibiting the complex geometry with 1, 2, 3 fiber crossings of the simulation data (http://hardi.epfl.ch/static/events/2012_ISBI/download.html). Rows 2-4 show sparse coding (SC) reconstructions for different undersamplings with mid-range noise SNR 18.



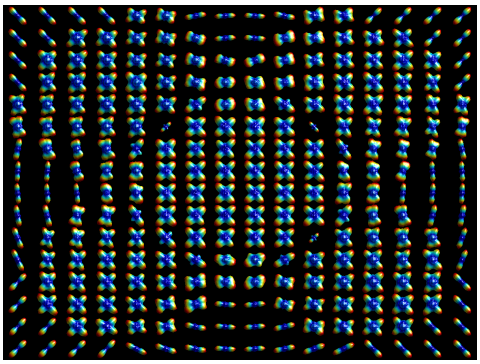
49 atoms from SNR18



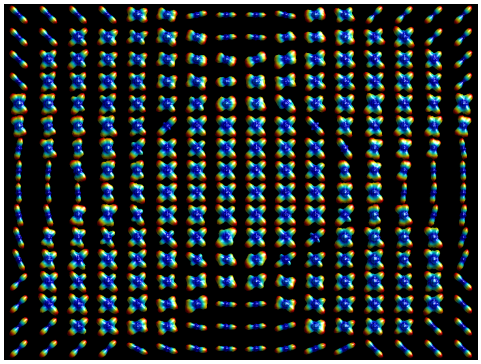
Ground truth



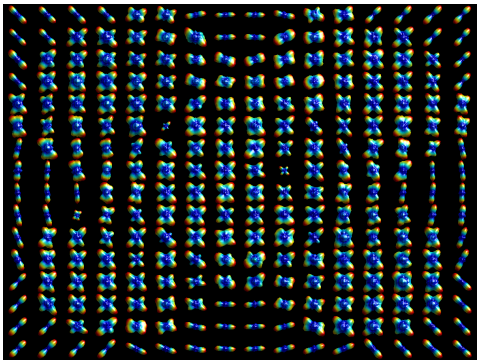
Gold standard (DSI full 515)



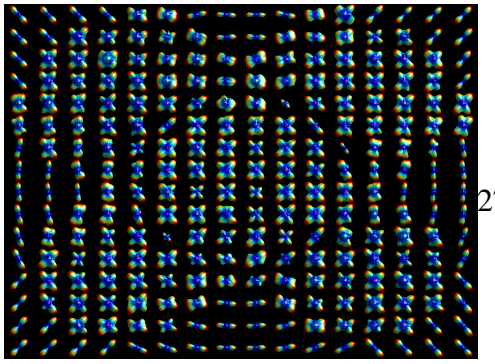
SC DSI 258



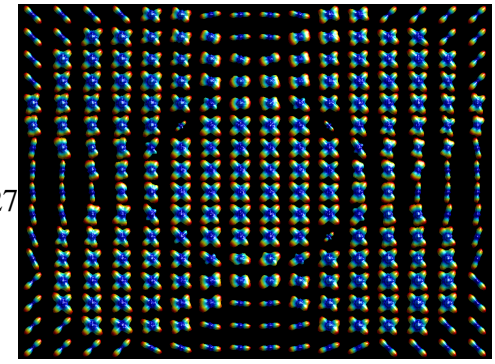
SC DSI 58



SC DSI 37

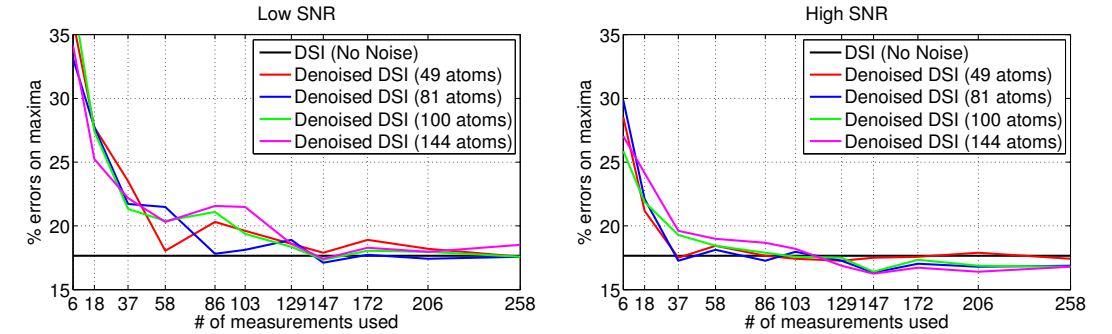


SC DSI 29

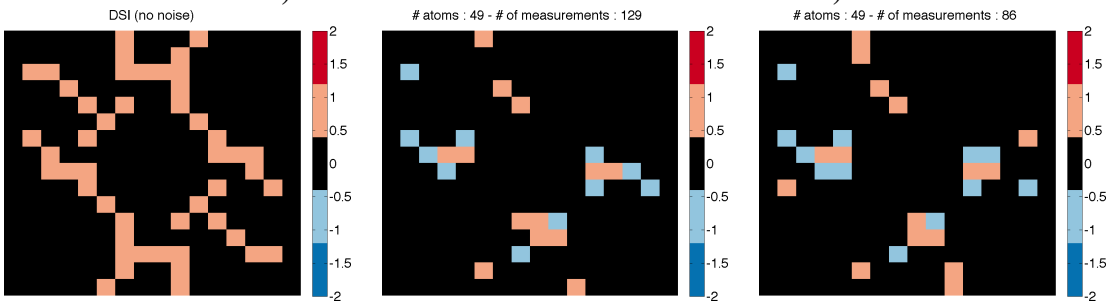


SC DSI 18

Figure 4: Error in maxima detection. In a) and b), errors are shown in terms of percentage of voxels with incorrect number of estimated maxima. Results are presented for various number of DWI images available for DSI estimation, and for different SNRs and number of atoms in the dictionary. In c), black pixels illustrate perfect agreement between ground truth phantom and DSI no noise whereas pink pixels show where DSI no noise underestimates 1 maxima with respect to the ground truth (single fiber detected instead of a two fibers crossing). In d)-h), the comparison is between sparse coding with different undersamplings and the noiseless DSI. The noiseless DSI acts as our gold standard. Here, pink and red pixels mean that our sparse coding method has detected a single fiber instead of a 2 or 3 fiber crossings respectively, whereas blue pixels show where our approach with undersampling is actually *better* than DSI with no noise if this pixel is in pink in c).



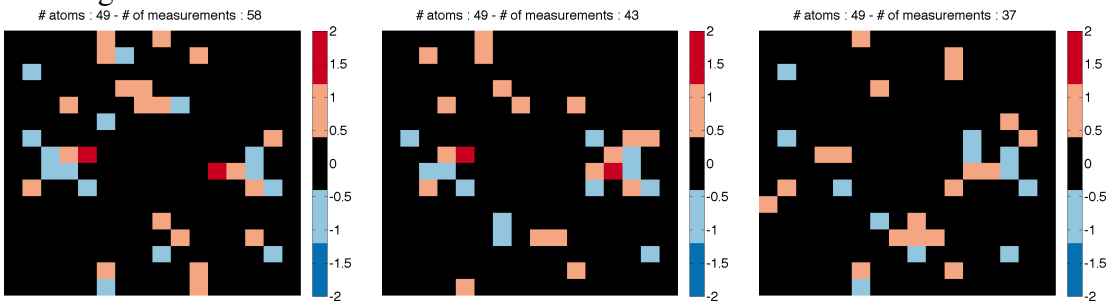
a) b)



c) DSI no noise vs ground truth

d) SC DSI 129 vs DSI no noise

e) SC DSI 86 vs DSI no noise



f) SC DSI 58 vs DSI no noise

g) SC DSI 43 vs DSI no noise

h) SC DSI 37 vs DSI no noise

Figure 5: ODFs computed from the learnt dictionaries on the 2 subjects (100 atoms and inter-subject cross-validation). Top is for subject 1 and bottom for subject 2. Here, the ODF colormap goes from blue to red, for mininal to maximal ODF values on the sphere to better highlight the 3D nature of ODF glyphs.

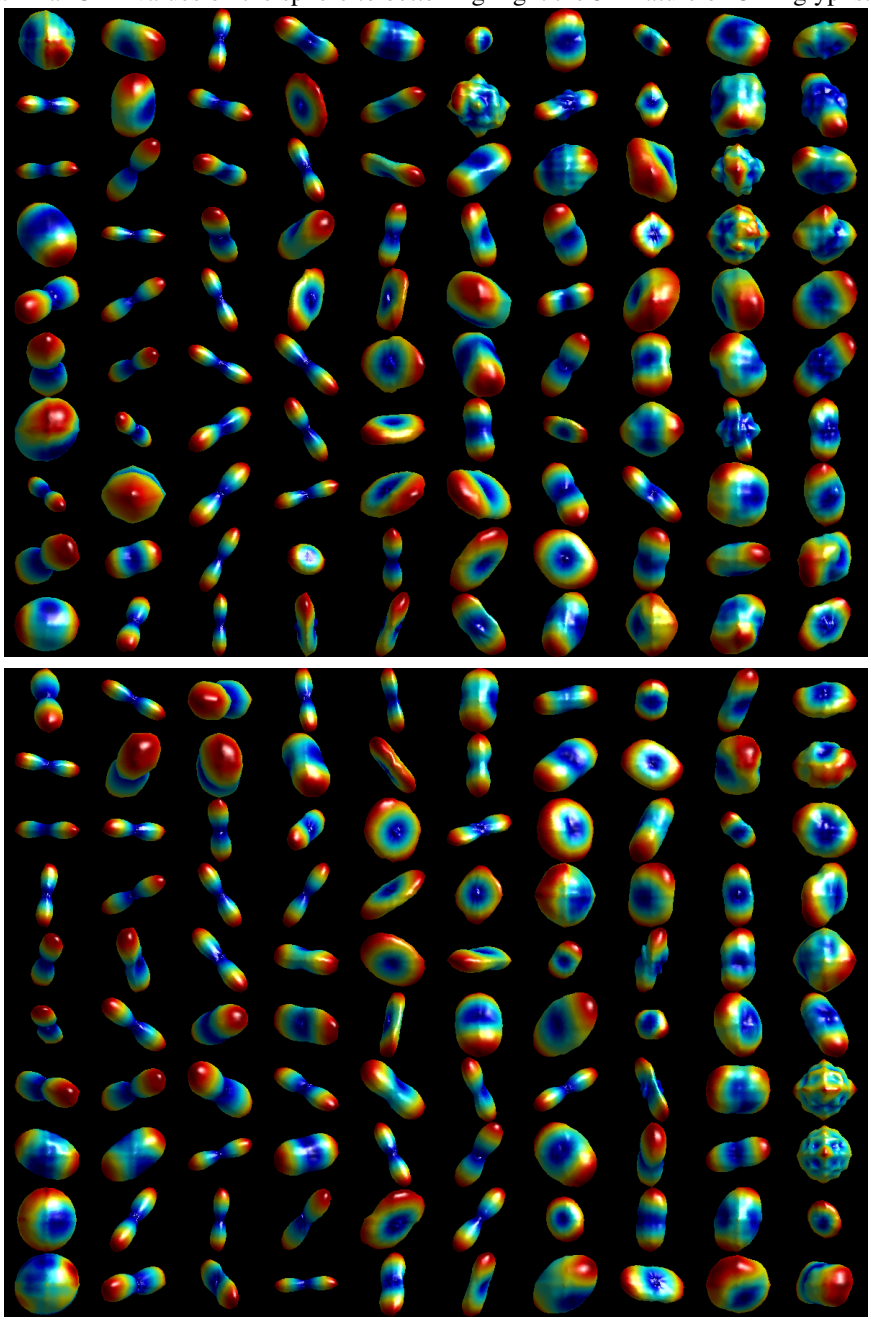
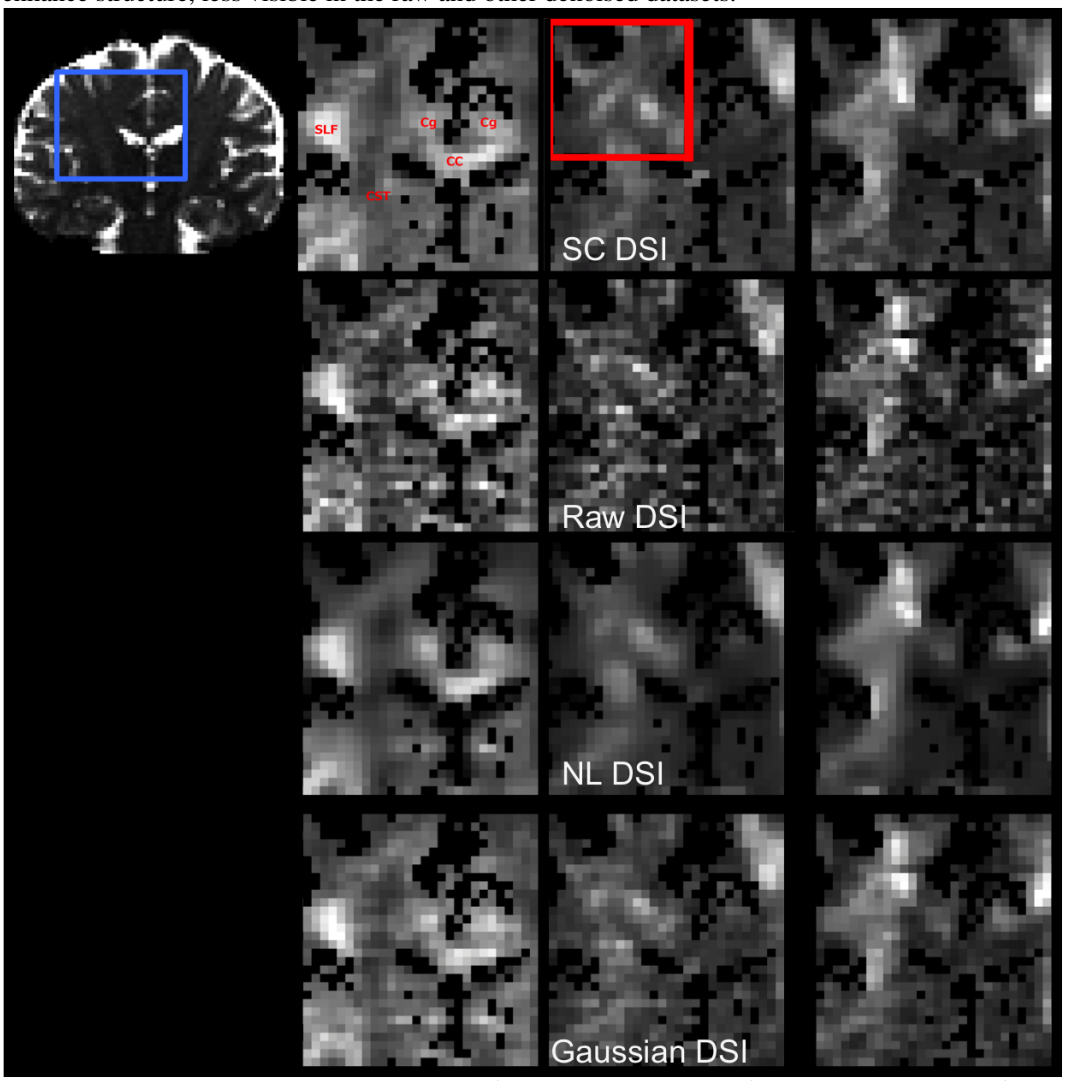


Figure 6: Denoising the raw DSI data with our sparse coding (SC) technique versus state-of-the-art non-local means (NLM) and Gaussian (optimal $\sigma = 0.35$) denoising. The region of interest is chosen in the centrum semiovale, in mid-coronal slice, as seen in an approximate blue square in the top left figure. In this region, we see the corpus callosum (CC), the cingulum (Cg), the corticospinal tract (CST) and the superior longitudinal fasciculus (SLF). One can appreciate in the red box how the dictionary learning approach is able to highlight and enhance structure, less visible in the raw and other denoised datasets.



$b = 960$ s/mm^2 $\mathbf{q} = (0, 0, -2)$ | $b = 3360$ s/mm^2 $\mathbf{q} = (-3, -2, -1)$ | $b = 6000$ s/mm^2 $\mathbf{q} = (-5, 0, 0)$

Figure 7: Full ($d = 515$) DSI vs. Half DSI ($d = 258$) with respect to simple symmetry, Gaussian ($\sigma = 0.35$), NL means and our sparse coding (SC) based denoising of subject 1 (k atoms) and 258 measurements. Underlying GFA reflects the spatial smoothing done by the different techniques, where we see that white and gray areas have less contrast (voxels all blended together) in the Gaussian and NL means cases. Hence, Gaussian and NL means ODFs lose some angular sharpness, compared to full DSI and sparse coding, especially seen in the red box area.

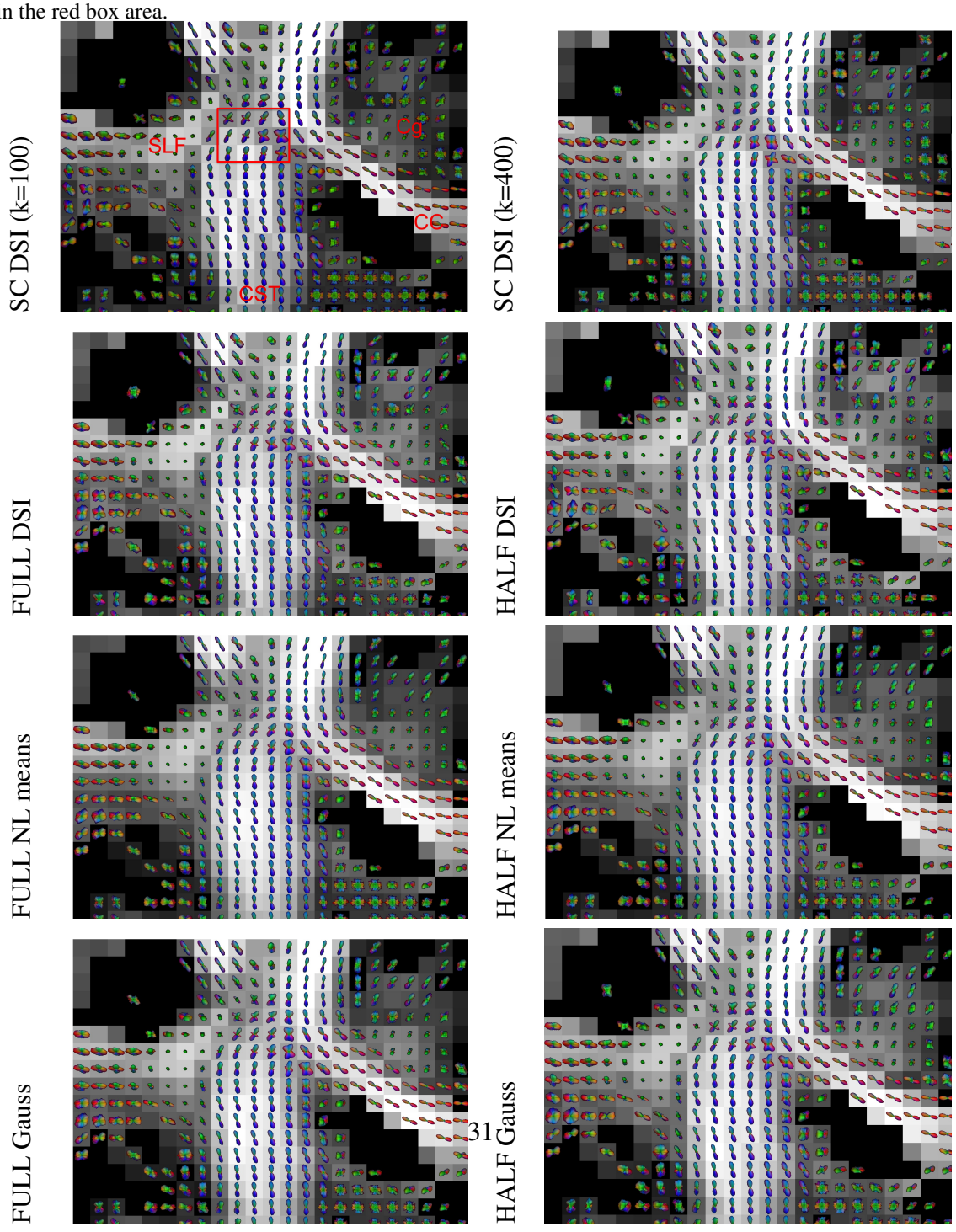
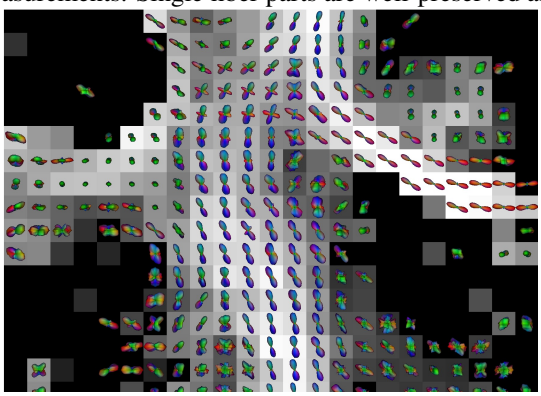
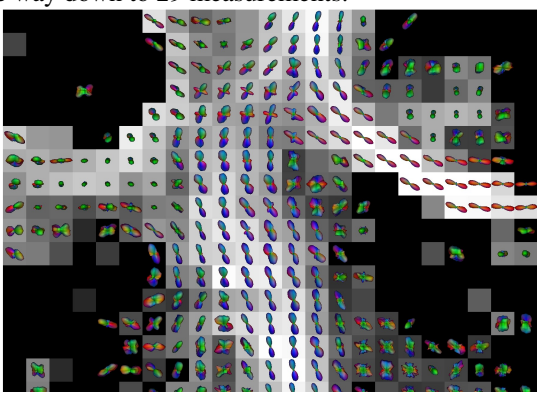


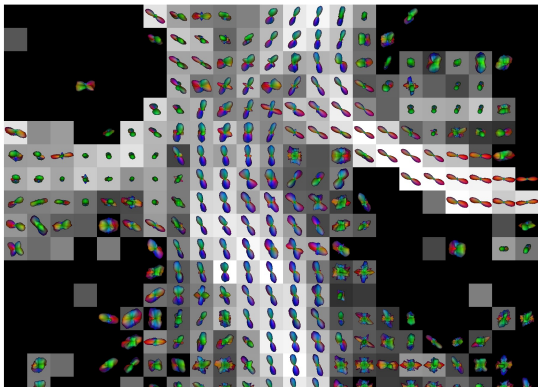
Figure 8: Undersampled reconstruction using the dictionary learnt on subject 1 ($k = 100$ atoms) to reconstruct DSI signal and diffusion ODFs of subject 2 in the centrum semiovale similar to Fig. 7. We note that the angular profiles in the crossing area are preserved through undersampling but become noisier below 43 measurements. Single fiber parts are well-preserved all the way down to 29 measurements.



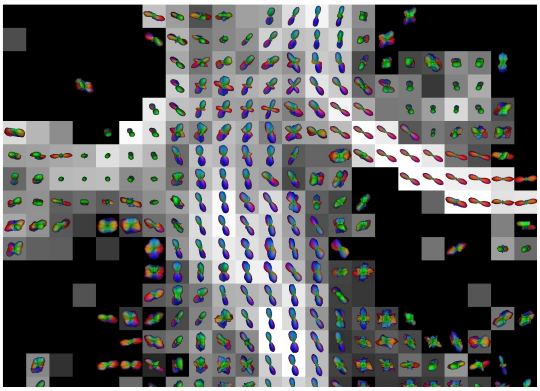
258 measurements



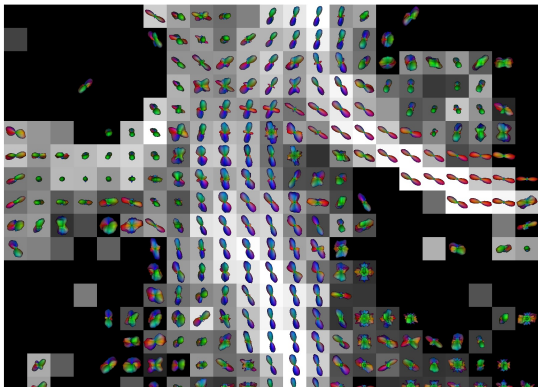
129 measurements



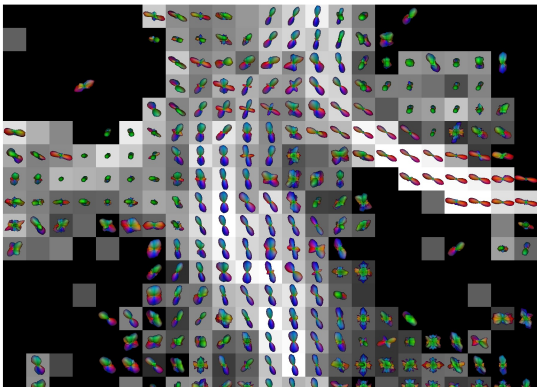
58 measurements



43 measurements



37 measurements



29 measurements

Figure 9: Reconstruction error ratios for intra and inter-subject settings as a function of the number of measurements (k=100 atoms). brain1/brain1 is for the intra-subject case, while brain1/brain2 is the inter-subject case, i.e atoms learnt on subject 1 and used to estimate the full DSI data of subject 2 using only a few measurements.

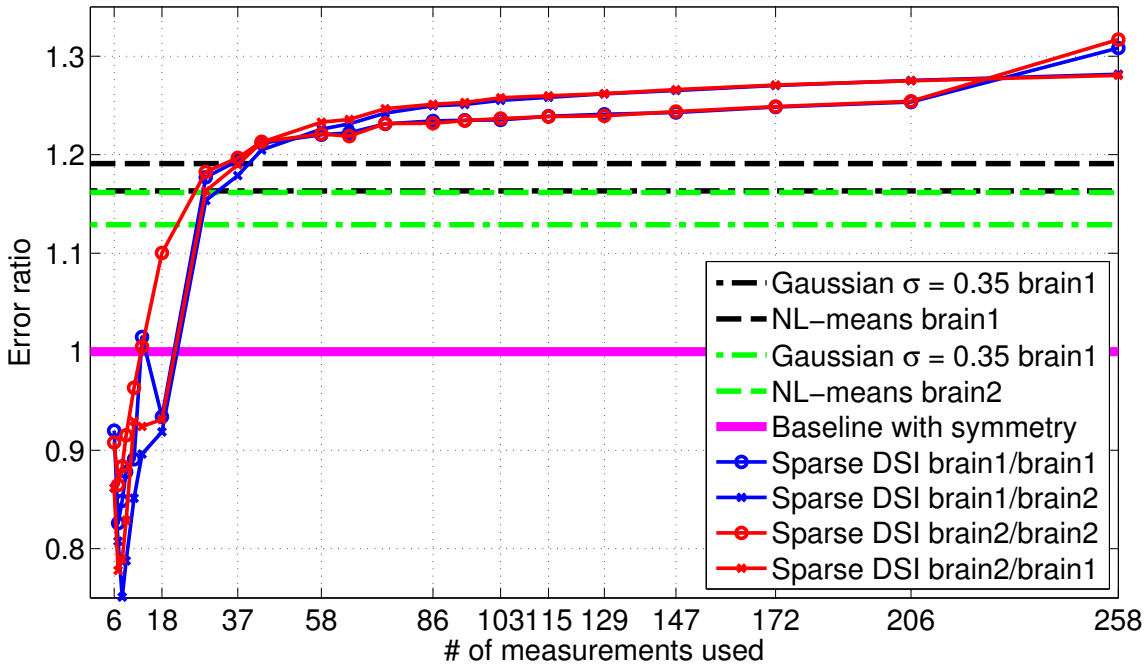
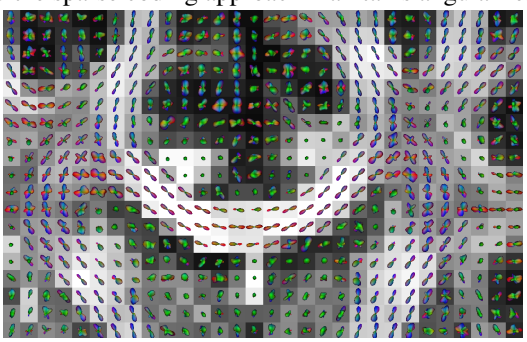
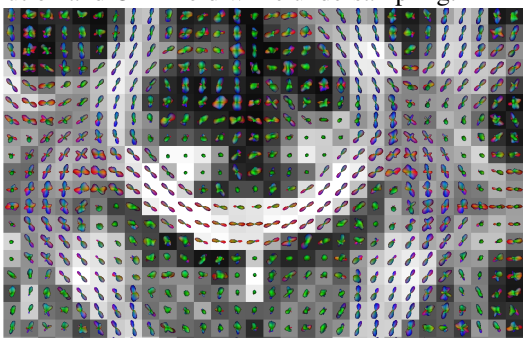


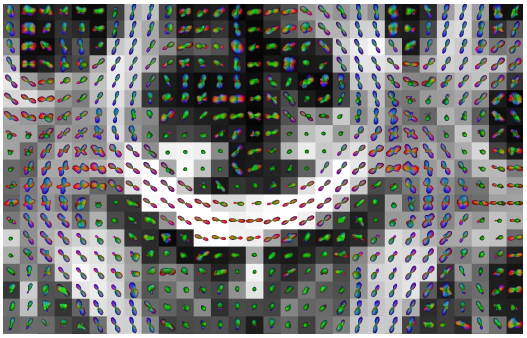
Figure 10: Undersampled reconstruction using a dictionary learnt with sparse coding (SC) on subject 2 with 100 atoms to reconstruct DSI signal and diffusion ODFs of *brain0* of the Pittsburgh Brain Competition. Note how the sparse coding approach maintains angular resolution and ODF field while undersampling.



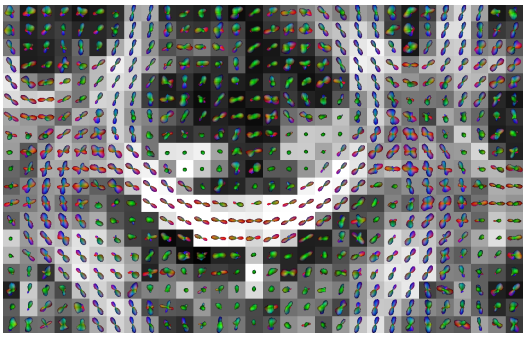
ODFs from original
258 measurements (no average)



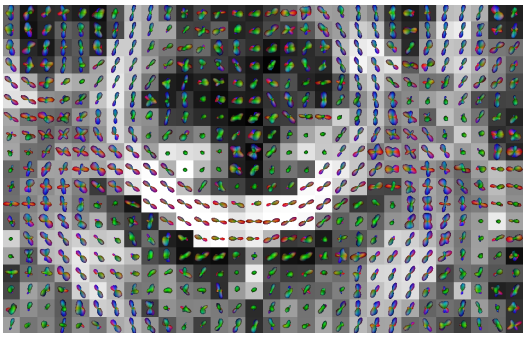
ODFs from original
515 measurements (2 averages)



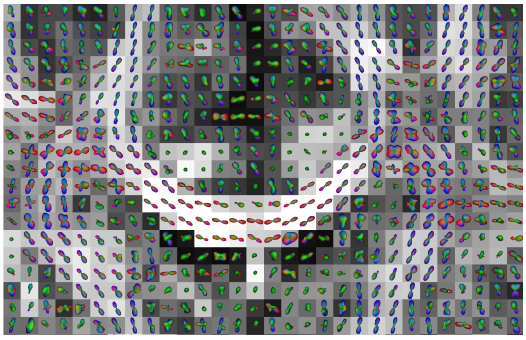
SC DSI 258 measurements



SC DSI 58 measurements



SC DSI 37 measurements



SC DSI 29 measurements