



**HAL**  
open science

# A computational diffusion MRI and parametric dictionary learning framework for modeling the diffusion signal and its features

Sylvain L. Merlet, Emmanuel Caruyer, Aurobrata Ghosh, Rachid Deriche

► **To cite this version:**

Sylvain L. Merlet, Emmanuel Caruyer, Aurobrata Ghosh, Rachid Deriche. A computational diffusion MRI and parametric dictionary learning framework for modeling the diffusion signal and its features. *Medical Image Analysis*, 2013, pp.MEDIMA779. 10.1016/j.media.2013.04.011 . hal-00820817

**HAL Id: hal-00820817**

**<https://hal.science/hal-00820817>**

Submitted on 7 May 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A computational diffusion MRI and parametric dictionary learning framework for modeling the diffusion signal and its features

Sylvain Merlet<sup>a</sup>, Emmanuel Caruyer<sup>b</sup>, Aurobrata Ghosh<sup>a</sup>, Rachid Deriche<sup>a</sup>

<sup>a</sup>*Athena Project-Team, INRIA Sophia Antipolis - Méditerranée, France*

<sup>b</sup>*Section of Biomedical Image Analysis Department of Radiology University of Pennsylvania*

---

## Abstract

In this work, we first propose an original and efficient computational framework to model continuous diffusion MRI (dMRI) signals and analytically recover important diffusion features such as the Ensemble Average Propagator (EAP) and the Orientation Distribution Function (ODF). Then, we develop an efficient parametric dictionary learning algorithm and exploit the sparse property of a well-designed dictionary to recover the diffusion signal and its features with a reduced number of measurements. The properties and potentials of the technique are demonstrated using various simulations on synthetic data and on human brain data acquired from 7-T and 3-T scanners. It is shown that the technique can clearly recover the dMRI signal and its features with a much better accuracy compared to state-of-the-art approaches, even with a small and reduced number of measurements. In particular, we can accurately recover the ODF in regions of multiple fiber crossing, which could open new perspectives for some dMRI applications such as fiber tractography.

**Keywords:** Dictionary Learning, Sparse Reconstruction, Compressive Sensing, MRI, Diffusion MRI, Orientation Distribution Function, Ensemble Average Propagator, White Matter.

---

## 1. Introduction

Diffusion MRI (dMRI) assesses the integrity of brain anatomical connectivity and is very useful for examining and quantifying white matter (WM) microstructure and organization not available with other imaging modalities. dMRI determines the WM structure by exploiting the way the water molecules diffuse. The first diffusion images were obtained in the mid-1980s [23], and was based on the pioneering work of [38], who introduced the pulsed gradient spin-echo (PGSE) sequence. It allows the quantification of the water diffusion by estimating the displacement of water particles from the phase change that occurs during the acquisition process. More importantly, under the so called narrow pulse assumption, we can show that the normalized signal attenuation  $E(\mathbf{q})$  is written as the Fourier transform of the Ensemble Average Propagator (EAP)  $P(\mathbf{R})$

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

where  $\mathbf{q}$  and  $\mathbf{R}$  are both 3D-vectors that respectively represent the effective gradient direction and the displacement direction. We can decompose them as  $\mathbf{q} = q\mathbf{u}$  and  $\mathbf{R} = R\mathbf{r}$ , where  $\mathbf{u}$  and  $\mathbf{r}$  are 3D unit vectors.

Diffusion Tensor Imaging (DTI) [7, 6] method characterizes the diffusion by a Gaussian distribution, and is known to be a limited model. In particular DTI is not able to resolve crossing fibers. Resolving crossing fibers helps to disambiguate between several possible tracts in regions of crossing fibers and reconstruct more accurate anatomical connectivity through

fiber tractography. Recently, more complex models appeared to overcome this limitation. Nevertheless, these techniques often require many acquisitions in particular when High Angular Resolution Diffusion Imaging (HARDI) [44, 4, 43, 15, 20, 1] or Diffusion Spectrum Imaging (DSI) [46] are used. HARDI techniques allow the estimation of the Orientation Distribution Function (ODF) [44, 4, 43, 15, 1], which gives the probability that a water molecule diffuses in a given direction. Several authors (Aganj et al., Tristan-Vega et al., Wedeen et al.) express the ODF  $\Upsilon(\mathbf{r})$  as the integration of the EAP over a solid angle, i.e.

$$\Upsilon(\mathbf{r}) = \int_0^\infty P(R, \mathbf{r}) R^2 dR. \quad (2)$$

In [40, 3], the authors propose to estimate the fiber orientation distribution called the fiber ODF (fODF). The fODF is able to resolve up to 30 degrees crossings consistently [41], which makes this model a promising resource to estimate fiber orientation. In our work, we reconstruct the ODF as described in Eq. 2 and we do not aim to compare our approach with the family of method estimating the fODF. A review of method reconstructing the ODF and the fODF can be found in [21]. Another HARDI technique has been proposed in [20], where the authors characterize the diffusion signal by a Wishart distribution. Jian et al. shows improvements over the classical DTI technique and present an estimation scheme for the fiber orientation and EAP. Among the HARDI techniques, [48] introduces Neurite Orientation Dispersion and Density Imaging (NODDI), which allows the estimation of the microstructural complexity of dendrites and axons. Diffusion Spectrum Imaging (DSI) was developed

in parallel to the HARDI techniques [46]. In DSI, the EAP  $P(\mathbf{R})$  is directly obtained by taking the inverse Fourier transform of the normalized signal  $E(\mathbf{q})$  measured in the  $\mathbf{q}$ -space (see Eq. 1). However, the high resolution EAP obtained with DSI requires many measurements. HARDI and DSI are impractical for clinical use in MRI systems commonly found in hospital. Accelerated acquisitions, relying on a smaller number of sampling points, are thus very welcome to efficiently estimate the complex features of the diffusion process.

Sparse reconstruction approaches were found to successfully reduce the number of acquisitions in dMRI [29, 24, 33, 36, 42, 9, 19, 26, 30, 47]. These techniques are usually based on a  $l_1$  minimization of the diffusion signal with respect to a sparse representation. Merlet and Deriche, Menzel et al. [29, 24] combine the Compressive Sensing (CS) theory and DSI to accelerate the acquisition. Merlet and Deriche [30] use orthonormal bases to sparsely describe the diffusion signal. In [33] and in [42], the authors elegantly design dictionaries for sparse modeling in dMRI. They provide an overcomplete dictionary computed from a discretized version of predefined functions, i.e. the Spherical Ridgelets in [33] (see [36] for the multiple shells version) and the Spherical Wavelets in [42]. Learning a dictionary provides an alternative way to design sparse dictionaries [9, 19, 26, 47].

Some approaches have been recently proposed in order to design dictionaries that enable sparse representations (A good overview can be found in [2]). For instance, Bilgic et al., Gramfort et al. [9, 19] learn dictionaries from DSI like acquisitions and use it to either denoise full DSI data or to perform under-sampled DSI acquisitions and reconstructions. In particular, Gramfort et al. nicely exploit the symmetry of the signal in order to assess free parameters of the dictionary learning problem. However, these two latter works lead to non-parametric dictionaries, which does not provide continuous representations of the diffusion signal nor allow the determination of analytical formulae for diffusion features. The strength of the parametric dictionary learning approach, as the one we propose in this article, lies in its ability to address these weaknesses. A work regarding parametric dictionary learning was published in [47], in which the dictionary atoms are formed by a weighted combination of  $3^{rd}$  order B-splines. It proved that the method is efficient on synthetic data simulated with 81 gradient directions. The work of [47] appears promising in reconstructing the diffusion signals, and further enhancement could be done regarding the development of analytical formulae to estimate other diffusion features. This would make this work a good resource in the context of dictionary learning. More recently, we proposed in [26] to learn a dictionary where each atom is constrained to be a parametric function. In [26], this parametric function is a combination of a radial part and an angular part represented by the symmetric and real Spherical Harmonics (SH) [15]. The radial part is a polynomial weighted by an exponential. 50 measurements were sufficient to reconstruct very good quality diffusion signals, ODFs and EAPs. However, this approach essentially handles the learning of the radial part, i.e. the polynomial coefficients and a scale parameter in the exponential, whereas we observed (see [27]) that the angular part could make the dictio-

nary much sparser if we adequately combine several SH functions instead of only one.

In this work, we present a method, which exploits the sparse property of a well designed dictionary based on a computational dMRI framework, in order to recover the diffusion signal with a reduced number of measurements. This framework enables a continuous modeling of the diffusion signal and leads to analytical formulae to estimate important diffusion features, namely the ODF and the EAP. To improve our previous work in [26], we modify the parametric function, describing the atoms, to learn both the radial and the angular part, which provide a very sparse representation of diffusion signals and further reduce the number of measurements (15 measurements are found to be sufficient to start recovering the EAP and some derived diffusion features whereas 50 measurements are used in [26]). Furthermore, we extend the experimental part of [26] by learning and validating our approach on the synthetic data proposed in the HARDI contest at ISBI 2012 <sup>1</sup>, and on real data acquired from both 3T and 7T scanners. A preliminary work [27] regarding the learning of both the radial part and the angular part of the diffusion signal was published in the proceedings of the HARDI contest at ISBI 2012 and we obtained the best results in our category. Our approach presented in this paper indicates an increase in terms of reconstruction accuracy compared to the results presented in [27].

The article is structured as follows : we start by introducing the dMRI framework together with the proposed dictionary, then we focus on the parametric dictionary learning algorithm and finally we conclude with an experimental part illustrating the added-value of our approach with promising results showing how our approach allows the accurate reconstruction of the diffusion signal and some of its features. This experimental part is completed by a comparison with state of the art approaches, and is performed on synthetic and real data from 3T and 7T scanners.

## 2. A computational framework for the recovery of the complete diffusion MRI process

In this section, we introduce a new dMRI framework for modeling the diffusion signal. From this continuous representation, we derive analytical formulae that enable the estimation of important diffusion features such as the Ensemble Average Propagator (EAP), the Orientation Distribution Function (ODF). We give full derivations for these formulae in the appendixes.

### 2.1. Continuous diffusion modeling with a constrained dictionary

We propose to design an overcomplete dictionary  $\Psi = \{\Psi_k\}_{k=0,\dots,K}$ , such that the diffusion signal  $E$  is expressed as a truncated linear combination of  $K$  3D atoms  $\Psi_k$ , i.e

---

<sup>1</sup><http://hardi.epfl.ch/>

$$E(\mathbf{q}\mathbf{u}) = \sum_{k=0}^K c_k \Psi_k(\mathbf{q}\mathbf{u}), \quad (3)$$

with  $c_k$  the transform coefficients,  $q$  the norm of the effective gradient and  $\mathbf{u}$  an unitary vector.

In diffusion multiple shells imaging [5, 34, 17], previous works have proposed the modeling of the diffusion signal in bases where each atom  $\Psi_k$  is a combination of a radial part and an angular part. These works primarily handle the question of the radial part. For instance, Descoteaux et al. [17] describes the radial part of the atom as a combination of two rational functions coming from the total solution of the Laplace equation, whereas Ozarslan et al., Assemlal et al. [34, 5] use polynomials weighted by an exponential. Merlet et al. [26] increased the sparsity of the representation in [34] by designing a dictionary where the polynomial coefficients and the scale parameters in the exponentials are learned from a training data set [26]. Although the radial attenuation of the diffusion signal is more or less well fitted with these bases/dictionaries, the accurate and sparse estimation of the full diffusion signal is still limited by the angular part described by the real and symmetric Spherical Harmonic basis functions (SH). The SH have been proved useful in many settings but are not sparse enough in modeling the directional features of the diffusion process [32]. However, our preliminary work published in [27] shows that a well-chosen combination of SH could sparsely represent the angular part. Therefore, we propose to model this angular part with such a combination of SH. As for the radial part, we model it with a combination of exponential functions weighted by a monomial in order to ensure the continuity of the function at zero [11]. The complete description of each atom  $\Psi_k$  of the dictionary is given by :

$$\begin{aligned} \Psi_k(\mathbf{q}) &= \Psi_k(\mathbf{q}\mathbf{u}) = \frac{1}{\sqrt{\chi_k}} \sum_{i=0}^I \alpha_{ki} \exp(-v_{ki}q^2) \sum_{j=0}^J \beta_{kij} q^{l(j)} Y_j(\mathbf{u}) \\ &= \frac{1}{\sqrt{\chi_k}} \sum_{i=0}^I \sum_{j=0}^J \gamma_{kij} \exp(-v_{ki}q^2) q^{l(j)} Y_j(\mathbf{u}) \\ &= \Psi_k(\boldsymbol{\gamma}_k, \mathbf{v}_k, \mathbf{q}\mathbf{u}), \end{aligned} \quad (4)$$

with  $\mathbf{q}$  the 3D effective gradient,  $\mathbf{u}$  an unitary vector and  $q$  the norm of the effective gradient such that  $\mathbf{q} = \mathbf{q}\mathbf{u}$ .  $I$  and  $J$  are, respectively, the radial order and the angular order of the dictionary.  $J$  also corresponds the total number of SH taken into account in the modeling not to be confused with the maximal SH order  $L$ . Indeed,  $J$  is directly related to the maximal SH order  $L$  as  $J = (L+1)(L+2)/2$ .  $Y_j(\mathbf{u})$  is the SH of order  $l(j) = 0$  for  $j = 1$ ,  $l(j) = 2$  for  $j \in \{2, \dots, 6\}$ ,  $l(j) = 4$  for  $j \in \{7, \dots, 15\}$  ... .  $\boldsymbol{\gamma}_k = \{\gamma_{kij}\}_{i=0\dots I, j=0\dots J}$  and  $\mathbf{v}_k = \{v_{ki}\}_{i=0\dots I}$  are two vectors of parameters, which will be set during the learning process. The term  $q^{l(j)}$  ensures the continuity of  $\Psi_k$  at zero.  $\chi_k$  is a constant, which ensures the normalization of  $\Psi_k$  for the  $\ell_2$  norm, i.e  $\sqrt{\int_{\mathbb{R}^3} \Psi_k^2(\mathbf{q}) d\mathbf{q}} = 1$ , and is expressed as

$$\chi_k = \sum_{i'=0}^I \sum_{i=0}^I \sum_{j=0}^J \frac{\gamma_{kij} \gamma_{ki'j}}{2(v_{ki} + v_{ki'})^{\ell(j)+3/2}} \Gamma\left(\ell(j) + \frac{3}{2}\right), \quad (5)$$

with  $\Gamma$  the gamma function. We derive eq. 5 in Appendix A.

If we consider the simple case where all the coefficients  $\gamma_{kij}$  and  $v_{kj}$  are zero but  $\gamma_{k00}$  and  $v_{k0}$ , we get  $\Psi_k(\mathbf{q}\mathbf{u}) = \frac{1}{\sqrt{\chi_k}} \frac{\gamma_{k00}}{2\sqrt{\pi}} \exp(-v_{k0}q^2)$ . This mono-exponential representation of the atom  $\Psi_k$  does not depend on the angular direction  $\mathbf{u}$  and is, for instance, appropriate to describe isotropic diffusion configuration.

Note that in [27] each atom is described by a combination of SHORE basis functions [34] with a predefined scale parameter. The main advantage of the atom description in Eq. 4 lies in the possibility to learn the scale parameters  $v_{ki}$ , which provides a sparser signal estimation than in [27]. We see in the experiment part that this new framework leads to an increase in terms of reconstruction accuracy compared to the results presented in [27], which were already the best in their category.

## 2.2. Closed formulae for diffusion features

Using the dictionary,  $\Psi = \{\Psi_k\}_{k=0,\dots,K}$ , proposed in the previous section to reconstruct the diffusion signal (see Eq. 4), we derive important and analytical closed formulae for estimating the EAP and the ODF. We describe these formulae in the following.

### 2.2.1. The Ensemble Average Propagator

The EAP, denoted  $P(\mathbf{R}\mathbf{r})$ , represents the full 3D displacement probability function of water molecules in every voxel and underlies the derivation of the ODF. It is the inverse Fourier transform of the normalized diffusion signal, denoted  $E(\mathbf{q}\mathbf{u})$ ,

$$P(\mathbf{R}\mathbf{r}) = \int_{q=0}^{\infty} \int_{\mathbf{u} \in S^2} E(\mathbf{q}\mathbf{u}) \exp(+2\pi i \mathbf{q}\mathbf{R}\mathbf{u}\cdot\mathbf{r}) d\mathbf{u} q^2 dq. \quad (6)$$

$q$  and  $R$  are, respectively, the norm of the effective gradient and the radius of the 3D location in every voxel,  $\mathbf{u}$  and  $\mathbf{r}$  are unit vectors. From eq. 3 and 6, we derive in Appendix B the following expression for the EAP :

$$P(\mathbf{R}\mathbf{r}) = \sum_{k=0}^K \frac{c_k}{\sqrt{\chi_k}} \sum_{i=0}^I \sum_{j=0}^J \gamma_{kij} (-1)^{l(j)/2} \left(\frac{\pi}{v_{ki}}\right)^{l(j)+3/2} R^{l(j)} \exp\left(\frac{-(\pi R)^2}{v_{ki}}\right) Y_j(\mathbf{r}) \quad (7)$$

If we consider the special case of the mono-exponential representation of the atom  $\Psi_k$ , we get  $P(\mathbf{R}\mathbf{r}) = \sum_{k=0}^K \frac{c_k}{\sqrt{\chi_k}} \frac{\gamma_{k00}}{2\sqrt{\pi}} \left(\frac{\pi}{v_{k0}}\right)^{3/2} \exp\left(\frac{-(\pi R)^2}{v_{k0}}\right)$ , i.e an isotropic propagator described by an mono-exponential decay similar in every direction.

### 2.2.2. Solid angle ODF

The ODF represents the full angular distribution of  $P(\mathbf{R}\mathbf{r})$ . One relies on the ODF to perform fiber tractography [16], then an accurate and fast computation of this diffusion feature is very appreciated. From Eq. 2, we derive in Appendix C the following closed form for the ODF :

$$\Upsilon(\mathbf{r}) = \sum_{k=0}^K \frac{c_k}{\sqrt{\chi_k}} \sum_{i=0}^I \sum_{j=0}^J \gamma_{kij} (-1)^{l(j)/2} \left(\frac{\pi}{v_{ki}}\right)^{l(j)+1} \frac{(l(j)+1)!!}{2 \binom{2\pi^2}{v_{ki}}^{l(j)/2+1}} Y_j(\mathbf{r}) \quad (8)$$

Considering the precedent example, the ODF is represented by a scalar, i.e.  $\Upsilon(\mathbf{r}) = \sum_{k=0}^K \frac{c_k}{\sqrt{\chi_k}} \frac{\gamma_{k00}}{8\pi^{3/2}}$ , meaning that the ODF has the same value for every direction.

### 3. A parametric dictionary learning for sparse dMRI

Here, we introduce a parametric dictionary learning (PDL) method that enables a sparse representation of any diffusion signal from continuous and parametric functions. There are four advantages to consider a parametric approach for dictionary learning:

- A parametric dictionary is defined by a set of parameters ( $\boldsymbol{\gamma}_k$  and  $\boldsymbol{\nu}_k$  in Sec. 2), which gives a continuous representation of each atom and, thus, enables a continuous modeling of the diffusion signal. This is suitable for data interpolation and extrapolation.
- Analytical formulae can be derived to estimate important diffusion features as the EAP and the ODF.
- PDL is acquisition independent, i.e. the sampling scheme used for learning the dictionary does not have to be the same as the sampling scheme used for reconstructing the signals.
- PDL enables one to reduce the dimensionality of the dictionary atoms.

These four advantages, together with the quality of the obtained results (see Sec. 4), makes our parametric dictionary approach very attractive compared to non parametric methods [9, 19].

Concerning the development of our algorithm, we started by considering the K-SVD algorithm [2] as a model for our own method. Although the K-SVD method appears powerful in designing sparse dictionaries, this technique as described in [2] designs only non-parametric dictionaries, which do not present the advantages described above. Hence, we developed our own algorithm, which overcomes the limitation of the K-SVD algorithm. Our algorithm alternates between 2 steps: a sparse coding step and a dictionary update step, where the vectors of parameters  $\boldsymbol{\gamma}_k$  and  $\boldsymbol{\nu}_k$  (see Sec. 2) are estimated for every atom  $\mathbf{d}_k$  of the dictionary, using the non linear Levenberg-Marquardt (LM) algorithm. The section 3.1 presents our dictionary learning algorithm and the section 3.2 describes the method we use

to reconstruct any diffusion signal using the dictionary previously learned.

#### 3.1. Dictionary learning algorithm

**Notation :** Suppose the training data set consists in  $M$  observations  $\{\mathbf{s}_i\}_{i=1}^M$  (i.e.  $M$  voxels). For each observation  $\mathbf{s}_i$  we have  $m_s$  samples in the  $q$ -space, i.e.  $\mathbf{s}_{1..M} \in \mathbb{R}^{m_s}$ . We represent  $\{\mathbf{s}_i\}_{i=1}^M$  in matrix form  $\mathbf{S} \in \mathbb{R}^{m_s \times M}$  where  $\mathbf{s}_i$  is the  $i^{\text{th}}$  column. The algorithm searches for the dictionary  $\mathbf{D} \in \mathbb{R}^{m_s \times K}$ , that enables the sparsest representation for every column of  $\mathbf{S}$ . The dictionary consists in  $K$  atoms  $\{\mathbf{d}_k\}_{k=1}^K$  with  $\mathbf{d}_k \in \mathbb{R}^{m_s}$  a column of  $\mathbf{D}$ . We constrain  $\mathbf{d}_k$  to be an instance of the 3D function  $\Psi_k(\boldsymbol{\gamma}_k, \boldsymbol{\nu}_k, \mathbf{q}\mathbf{u})$  in Eq. 4. Here, we do not try to directly estimate  $\mathbf{d}_k$  but the vectors of parameters  $\boldsymbol{\gamma}_k$  and  $\boldsymbol{\nu}_k$ , that characterize the atom  $\mathbf{d}_k$ . For each observation  $\mathbf{s}_i$ , we define a coefficient vector  $\mathbf{c}_i \in \mathbb{R}^K$ , which forms the  $i^{\text{th}}$  column of the coefficient matrix  $\mathbf{C} \in \mathbb{R}^{K \times M}$ .

**Problem statement:** Given a training data set  $\mathbf{S}$ , we search for the dictionary  $\mathbf{D}$  that gives the sparsest representation of this set (i.e. for each column  $\mathbf{s}_i$  of  $\mathbf{S}$ ). Mathematically, the problem is to find the dictionary  $\mathbf{D}$  and the vectors  $\mathbf{c}_i$  in  $\mathbf{C}$  by solving :

$$\arg \min_{\mathbf{c}_i, \mathbf{D}} \{\|\mathbf{S} - \mathbf{D}\mathbf{C}\|_2^2\} \text{ subject to } \forall_i \|\mathbf{c}_i\|_1 \leq \epsilon \quad (9)$$

with  $\epsilon$  a small real defining the degree of sparsity of the dictionary. The minimization of the first term in Eq. 9 enables the signal estimation  $\mathbf{D}\mathbf{C}$  to remain close to the training data set  $\mathbf{S}$  and the constraint imposes the sparsity of each signal representation  $\mathbf{c}_i$  with respect to the dictionary  $\mathbf{D}$ . The method to solve Eq. 9 is described in the following and a summary of the algorithm is given in Alg. 1. This algorithm iteratively alternates between sparse signal estimations (i.e.  $\{\mathbf{c}_i\}_{i=1}^M$ ) and updates of the dictionary (i.e.  $\mathbf{D}$ ) so to better fit the training data set (i.e.  $\mathbf{S}$ ).

**First step (Sparse signal estimation):** In the first step, the estimation of the column vector  $\mathbf{c}_i$  is performed separately for each signal  $\mathbf{s}_i$ , i.e for each column of  $\mathbf{S}$ . Sparse estimation is achieved by solving the LASSO (Least Absolute Shrinkage and Selection Operator) problem [39]. It consists in minimizing the following objective function

$$\min_{\mathbf{c}_i} \|\mathbf{s}_i - \mathbf{D}\mathbf{c}_i\|_2^2 + \lambda \|\mathbf{c}_i\|_1. \quad (10)$$

$\lambda$  is a constant that controls the degree of sparsity of the coefficients estimated. Note that we relax the constraint in Eq. 9 by using a Lagrangian multiplier ( $\lambda$ ). There exist numerous iterative algorithms for efficiently solving such kind of constrained problems. These include coordinate descent, least-angle regression (LARS) [18], fast iterative thresholding shrinkage algorithm (FISTA) [8], etc. A number of these methods are available under the Python<sup>TM</sup> library Scikit-learn [35]. We use a Python<sup>TM</sup> implementation [35] of coordinate descent to solve Eq. 10.

**Second step (Dictionary update):** In the second step, we update the dictionary  $\mathbf{D}$ . For this purpose, we compute an absolute averaged coefficient vector  $\hat{\mathbf{c}} \in \mathbb{R}^K$ , such that  $\hat{\mathbf{c}} =$

$1/M \sum_i |c_i|$  ( $|\cdot|$  denotes the absolute value of each vector component), and find the atoms associated with the non zeros values of  $\hat{\mathbf{c}}$ . It gives a rough idea of which atoms are used for modeling the signal and enables one to discard some unnecessary atoms and, thus, to enforce sparsity. Then, in this set of atoms, we update one atom at a time, while fixing all the others. This process is repeated for all the atoms associated with the non-zero coefficients of  $\hat{\mathbf{c}}$ .

The in-update atom is denoted  $\mathbf{d}_{k_0}$ . To update this atom, we begin by decomposing the error term in eq. 9 as in [2], i.e.

$$\begin{aligned} \|\mathbf{S} - \mathbf{D}\mathbf{C}\|_{\ell_2}^2 &= \left\| \mathbf{S} - \sum_{k=1}^K \mathbf{d}_k \mathbf{c}_k^r \right\|_{\ell_2}^2 \\ &= \left\| \left( \mathbf{S} - \sum_{k \neq k_0} \mathbf{d}_k \mathbf{c}_k^r \right) - \mathbf{d}_{k_0} \mathbf{c}_{k_0}^r \right\|_{\ell_2}^2 \\ &= \|\mathbf{E}_{k_0} - \mathbf{d}_{k_0} \mathbf{c}_{k_0}^r\|_{\ell_2}^2, \end{aligned} \quad (11)$$

where  $\mathbf{c}_k^r$  is the  $k^{\text{th}}$  row of  $\mathbf{C}$ . The error matrix, denoted  $\mathbf{E}_{k_0}$ , contains the error between each observation  $\mathbf{s}_i$  (the  $i^{\text{th}}$  column of  $\mathbf{S}$ ) and its respective estimation with the dictionary where the  $k_0^{\text{th}}$  atom is removed. We could directly use the LM algorithm in order to fit the atom  $\mathbf{d}_{k_0}$  to the error matrix  $\mathbf{E}_{k_0}$ . However, because it takes into account all the observations  $\mathbf{s}_i$ , this dictionary update would not impose sparsity. Instead, we enforce the sparsity by constraining the atom  $\mathbf{d}_{k_0}$  to fit only a subset of observations and not the entire data set. For this purpose, we define the group of observations that use the atom  $\mathbf{d}_{k_0}$ , i.e.  $w_{k_0} = \{i, 1 \leq i \leq M, c_i(k_0) \neq 0\}$ . In other words, they are the observations whose coefficients, associated with the atom  $\mathbf{d}_{k_0}$ , are non zeros. Then, we compute the error matrix  $\mathbf{E}_{w_{k_0}} \in \mathbb{R}^{m_s \times \text{card}(w_{k_0})}$ . It corresponds to the estimation error between the observation vector  $\{\mathbf{s}_i\}_{i \in w_{k_0}}$  that forms the columns of  $\mathbf{S}_{w_{k_0}} \in \mathbb{R}^{m_s \times \text{card}(w_{k_0})}$  and the signal estimated for the group of observation  $w_{k_0}$  (The  $k_0^{\text{th}}$  atom is still removed from the dictionary), i.e.  $\hat{\mathbf{S}}_{w_{k_0}} = \sum_{k \neq k_0} \mathbf{d}_k \mathbf{c}_k^r(i), i \in w_{k_0}$ . Mathematically speaking, we have  $\mathbf{E}_{w_{k_0}} = \mathbf{S}_{w_{k_0}} - \sum_{k \neq k_0} \mathbf{d}_k \mathbf{c}_k^r(i), i \in w_{k_0}$ . Finally, we estimate the vector of parameters  $\boldsymbol{\gamma}_k$  and  $\mathbf{v}_k$  by constraining  $\mathbf{d}_{k_0}$  to fit the error matrix  $\mathbf{E}_{w_{k_0}}$ . This part is performed via the non linear Levenberg-Marquardt algorithm (LMA). The atom update procedure is repeated for every atom  $\mathbf{d}_{k_0}$  associated with the non zeros coefficients of  $\hat{\mathbf{c}}$ .

The method is given in Algorithm 1 as a whole.

**Convergence :** The sparse coding step (Eq. 10) is well known to be convex and the coordinate descent algorithm allows one to converge to the unique solution specific to the current dictionary  $\mathbf{D}$ . The dictionary update step, where  $\boldsymbol{\gamma}_k$  and  $\mathbf{v}_k$  are estimated using the LMA may converges to a local minima, depending on the initial solutions. Then, Eq. 9 is convex for  $\mathbf{c}$  and converge for  $\mathbf{D}$ , which do not ensure a convergence to a global minimum. Nevertheless, in our experiments, a stationary point has been reached after few iterations and the resulting dictionaries were proved very good experimentally.

**Initialization :** The problem in Eq. 9 admits local minima, and the solution may vary depending on the initial parameters.

---

### Algorithm 1 Semi-parametric dictionary learning

---

1. Initialize the dictionary by fixing its dimension  $K$  and the vectors of parameters  $\boldsymbol{\gamma}_k$  and  $\mathbf{v}_k$  for  $k = 1 \dots K$  as random.
2. Sparse estimation of the observations  $\{\mathbf{s}_i\}_{i=1}^M$ . We use the coordinate descent algorithm to solve for  $\mathbf{c}_i$  associated to each observation :

$$\min_{\mathbf{c}_i} \|\mathbf{s}_i - \mathbf{D}\mathbf{c}_i\|_2^2 + \lambda \|\mathbf{c}_i\|_1.$$

3. Updating the dictionary. Compute the absolute averaged coefficients vector  $\hat{\mathbf{c}} = 1/M \sum_i |c_i|$ . Repeat until all the atoms of the dictionary, with non zeros value in  $\hat{\mathbf{c}}$ , have been scanned :

- Let note the current atom, the  $k_0^{\text{th}}$ .
- Define the group of observation that use this atom :  $w_{k_0} = \{i, 1 \leq i \leq M, c_{k_0}(i) \neq 0\}$ .
- Compute the error matrix  $\mathbf{E}_{w_{k_0}} = \mathbf{S}_{w_{k_0}} - \sum_{k \neq k_0} \mathbf{d}_k \mathbf{c}_k^r(i), i \in w_{k_0}$ .  $\mathbf{S}_{w_{k_0}}$  contains the observation  $\mathbf{s}_i, i \in w_{k_0}$ .
- Apply the Levenberg - Marquardt algorithm to estimate the vectors of parameters  $\boldsymbol{\gamma}_{k_0}$  and  $\mathbf{v}_{k_0}$ , which constrain  $\mathbf{d}_{k_0}$  to best fit  $\mathbf{E}_{w_{k_0}}$ .
- Update  $\mathbf{d}_{k_0}$  according to  $\boldsymbol{\gamma}_{k_0}$  and  $\mathbf{v}_{k_0}$ .

4. Go back to the step 2 unless the overall error does not vary anymore
- 

We tried several ways to initialize the algorithm among which random initialization and initialization from signals selected at random in the training data set.. However, these kinds of initialization were not satisfactory since the corresponding solutions were too different between several attempts to build the dictionary. After many experiments, we finally selected each initial atom as a random combination of several training signals. This gave us the best satisfactory results, with the smallest sensitivity to the initialization.

### 3.2. Signal estimation via the learned dictionary

The purpose of section 3.1 was to design a parametric dictionary  $\mathbf{D}$ . Now, we are able to recover any sparse representation  $\mathbf{c}$  of diffusion signal  $\mathbf{s}$  regarding the dictionary  $\mathbf{D}$  by solving the LASSO problem :

$$\min_{\mathbf{c}} \|\mathbf{s} - \mathbf{D}\mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1. \quad (12)$$

We use the same algorithm as in the learning step to solve Eq. 10, i.e. the coordinate descent algorithm.

## 4. Experiments on synthetic data

We first train and validate our parametric dictionary on synthetic data. We assume the normalized diffusion signal  $E(\mathbf{q})$  is generated from the multi-tensor model for  $F$  fibers,

$$E(q\mathbf{u}) = \sum_{f=1}^F p_f \exp(-4\pi^2 \tau q^2 \mathbf{u}^T \mathbf{T}_f \mathbf{u}), \quad (13)$$

where a fibre  $f$  is defined by a tensor matrix  $\mathbf{T}_f$  and weight  $p_f$ , such that  $\sum_f p_f = 1$ .  $q$  denotes the norm of the effective gradient and  $\mathbf{u}$  is a unitary vector in Cartesian coordinate.

The analytical ground truth of the EAP for any radius  $R$  is then given by

$$P(R\mathbf{r}) = \sum_{f=1}^F p_f \frac{1}{\sqrt{(4\pi\tau)^3 |\mathbf{T}_f|}} \exp\left(\frac{-R^2 \mathbf{r}^T \mathbf{T}_f^{-1} \mathbf{r}}{4\tau}\right), \quad (14)$$

with  $\mathbf{r}$  a unitary vector in Cartesian coordinate.

We can also derive the ODF feature using the solid angle closed form expression [1, 43],

$$\psi(r) = \sum_{f=1}^F p_f \frac{1}{4\pi |\mathbf{T}_f|^{\frac{1}{2}} (\mathbf{r}^T \mathbf{T}_f^{-1} \mathbf{r})^{\frac{3}{2}}}. \quad (15)$$

In the remainder of this section, we describe the steps required to correctly design our parametric dictionary, i.e. the choice of the dictionary radial and angular orders and the learning phase, then we validate the learned dictionary on synthetic data using the analytical formulae we have just described.

#### 4.1. Which radial and angular order for the dictionary ?

We need to fix the angular and radial order, respectively denoted  $J$  and  $I$  in Eq. 4, for the dictionary generation.

We begin by defining the dictionary angular order  $J$ , which is related to the SH order  $L$  as  $J = (L + 1)(L + 2)/2$ . For this purpose, we generate synthetic ODFs (see Eq. 15) corresponding to two fibers crossing at different degrees :  $0^\circ$ ,  $40^\circ$ ,  $60^\circ$ ,  $90^\circ$ . Then, we fit each ODF with the Spherical Harmonic basis of order  $L = 0, 2, 4, 6, 8, 10, 12$  using a least squares fitting technique, and we compute the Normalized Mean Square Error (NMSE) between the original synthetic ODFs ( $\psi$ ) and the estimated ODFs in terms of SH ( $\psi_e$ ), i.e.  $NMSE = \frac{\|\psi - \psi_e\|_2^2}{\|\psi\|_2^2}$ . The resulting NMSEs are shown in Fig. 1. Regarding this figure, we find that a SH order of  $L = 8$  is sufficient to correctly estimate an ODF (In particular for fibers crossing at  $40^\circ$  or more). This leads to set the dictionary angular order to  $J = 45$ .

The radial order is related to the number of fiber crossing in the voxel of interest. We assume a maximum of three crossing fibers in each voxel. This is the case in the region where the corticospinal, corpus callosum and superior longitudinal fasciculus fiber bundles are crossing. Voxel with more than three fibers are considered as part of the noisy background. Consequently, setting the radial order at three appears to be the best choice and was found experimentally satisfactory. We show in Fig. 2 some examples of radial attenuation in arbitrary chosen directions simulated using a multi-Gaussian model mimicking three crossing fibers (plain line). We fit each of these radial attenuation with the radial part of the model in Eq. 4 and radial

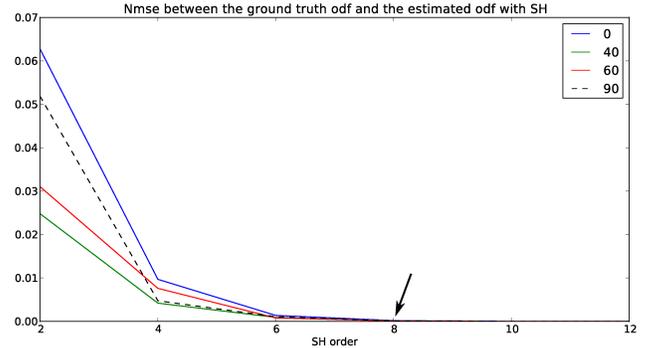


Figure 1: NMSE between the ground truth ODFs and the estimated ODFs in terms of SH. The arrow indicates the SH order corresponding to a NMSE considered as sufficiently close to zero.

orders  $I = 1, 2, 3$ , using the Levenberg-Marquardt algorithm (The angular order  $J$  is set to zero). Fig. 2 indicates that we need a radial order  $I = 3$  to accurately estimate radial attenuation characterized by three crossing fibers.

#### 4.2. Training phase

We train the dictionary on multi-Gaussian signals, used for the HARDI contest at ISBI 2012<sup>2</sup>. The contest was organized with the aim to provide a way for different groups to propose their own reconstruction algorithms and to fairly compare their methods against the others on a common set of ground-truth data.

Our training data set  $S_{train}$  is composed of the first  $M = 5000$  instances of diffusion signal contained in the file *TestingIV*, in which the multi-Gaussian synthetic signals are generated with parameters taken at random (number of fibers, fractional anisotropy related to a fiber and crossing angle between these fibers). The dictionary angular and radial orders are respectively set to  $J = 45$  and  $I = 3$  (see Sec. 4.1). We take  $m_s = 1000$  q-space samples for each instance of signal spread between  $b_{min} = 0$  and  $b_{max} = 10000$  s/mm<sup>2</sup>.

One difficulty in dictionary learning is the choice of the regularization parameter  $\lambda$  in Eq. 9. In order to assess  $\lambda$ , we use a cross validation (CV) procedure. For this purpose, we consider another set of signals  $S_{val}$ , called the validation data set, and composed of 1000 signals, which have not been used for training, and we repeat the following procedure for a range of  $\lambda$ ,

1. Design a dictionary  $\mathbf{D}_\lambda$  using Algorithm 1 with regularization parameter  $\lambda$  and the training data set  $S_{train}$ .
2. Using  $\mathbf{D}_\lambda$ , solve the LASSO problem for  $\mathbf{c}$  (see Eq. 12) with  $S_{val}$  as entry, and compute the validation error  $\epsilon_{val} = \frac{\|S_{val} - \mathbf{D}_\lambda \mathbf{c}\|_2^2}{\|S_{val}\|_2^2}$

We keep  $\lambda$  that minimizes  $\epsilon_{val}$ . This procedure is repeated after adding Rician noise, with SNR=10, 20 and 30, to the validation set. Rician noise is added in the following way

<sup>2</sup><http://hardi.epfl.ch/>

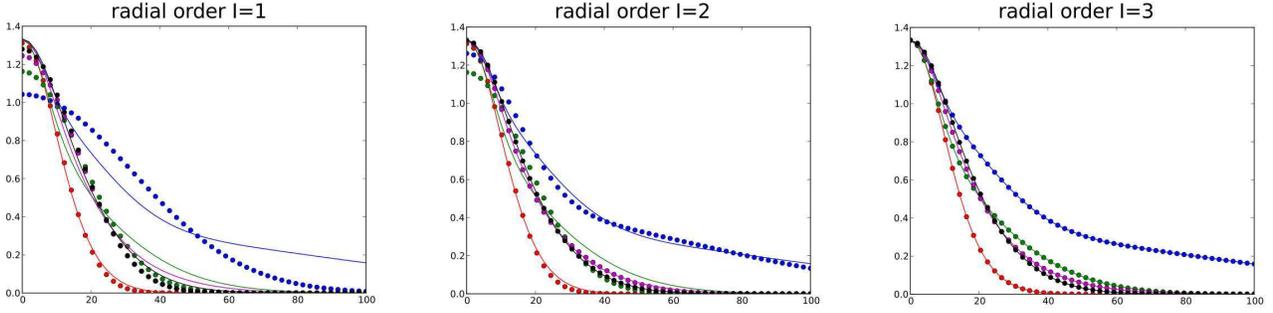


Figure 2: Examples of radial attenuation in arbitrary chosen directions simulated using a multi-Gaussian model mimicking three crossing fibers (plain line). We fit each of these radial attenuation with the radial part of the model in Eq. 4 and radial orders  $I = 1, 2, 3$ , using the Levenberg-Marquardt algorithm (The angular order  $J$  is set to zero).

:  $S_{\text{val, noisy}} = \sqrt{(S_{\text{val}} + \epsilon_1)^2 + \epsilon_2^2}$ , where  $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, \sigma)$  with  $\sigma = 1/\text{SNR}$ . Validating  $\lambda$  on noisy data enables one to prevent from overfitting.

We show in Fig. 3,  $\epsilon_{\text{val}}$  for  $\lambda$  in the range  $[1 \cdot 10^{-5}, 1 \cdot 10^{-4}]$ , in case of noisy and noiseless validation data. We also show the training error (blue curve in Fig. 3), i.e.  $\epsilon_{\text{train}} = \|\mathcal{S}_{\text{train}} - \mathbf{D}_\lambda \mathbf{c}\|_2^2 / \|\mathcal{S}_{\text{train}}\|_2^2$  where  $\mathbf{c}$  is the solution of the LASSO problem with  $\mathcal{S}_{\text{train}}$  as entry. We observe, in Fig. 3, four different  $\lambda$  ( $2.9 \cdot 10^{-5}, 4.4 \cdot 10^{-5}, 4.8 \cdot 10^{-5}, 5.8 \cdot 10^{-5}$ ) that minimize the validation error depending on the amount of noise we add to the validation data set. Because the noiseless case is not observed in practice, we discard the corresponding value of  $\lambda$  and consider the average value of ( $4.4 \cdot 10^{-5}, 4.8 \cdot 10^{-5}, 5.8 \cdot 10^{-5}$ ), which gives  $\lambda = 5 \cdot 10^{-5}$ . This value is used to generate a new dictionary.

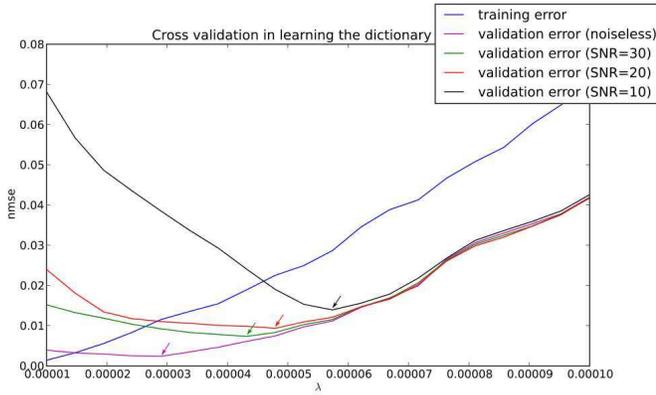


Figure 3: Training and validation error ( $\epsilon_{\text{train}}$  and  $\epsilon_{\text{val}}$ ) computed for  $\lambda$  in the range  $[1 \cdot 10^{-5}, 1 \cdot 10^{-4}]$ . The blue curve represents the training error and the other curves represent the validation error in a noiseless case (purple curve), after adding Rician noise to the validation data set with SNR=30 (green curve), SNR=20 (red curve), SNR=10 (black curve). The arrows indicate the minimum of each curve corresponding to the validation error.

We obtain a dictionary containing 659 atoms. Fig. 4 shows the first 200 ODFs of these atoms. The atom ODFs are sorted in decreasing energy order from left to right and top to bottom. We observe various shape ranging from single fiber structures

to more complex fiber configurations.

### 4.3. Validation

We validate the dictionary on the reconstruction of noisy multi-Gaussian signals, used for the HARDI contest at ISBI 2012<sup>3</sup>.

We consider 1000 signals, which have not been used for training the dictionary. Our preliminary and promising results on parametric dictionary learning (PDL) were published in the proceedings of this event [27]. At this stage, we obtained the best results in our category. Here, we enrich these previous results with a comparison of three different sampling schemes (displayed in Fig. 5) :

- A single shell sampling scheme with 64 measurements uniformly spread on one shell at a b-value  $b = 3000 \text{ s} \cdot \text{mm}^{-2}$  (Fig. 5a).
- A multiple shells sampling scheme with 15 measurements spread on 2 shells at b-values  $b = 1500, 2500 \text{ s} \cdot \text{mm}^{-2}$  (Fig. 5b).
- A multiple shells sampling scheme with 64 measurements spread on 2 shells at b-values  $b = 1500, 2500 \text{ s} \cdot \text{mm}^{-2}$  (Fig. 5c).

To obtain the single shell (SS) sampling scheme, we use the algorithm given in [22, 14] to uniformly distribute points on a sphere. For the multiple shells (MS) sampling schemes, we use the algorithm given in [10] by setting the parameters in such a way that the the number of points on each shell is proportional to  $q^1$ . These particular parameters have been proved efficient in [10, 25]. An important advantage of this algorithm is that the points from each shell have staggered directions and follow a near-optimal uniform distribution. You can generate and download sampling schemes for multiple Q-shell diffusion MRI with this web application : <http://www-sop.inria.fr/members/Emmanuel.Caruyer/q-space-sampling.php>.

<sup>3</sup><http://hardi.epfl.ch/>

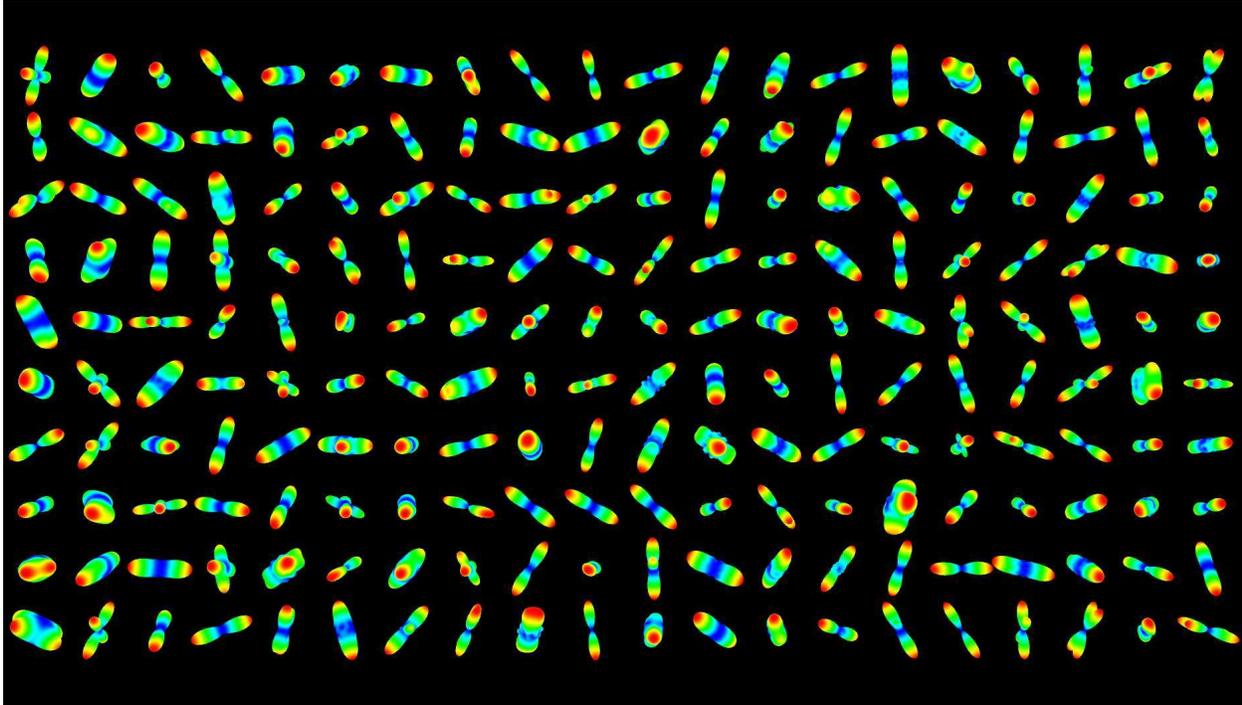


Figure 4: First 200 ODFs of the dictionary atoms. The atoms are sorted in decreasing energy order from left to right and top to bottom.

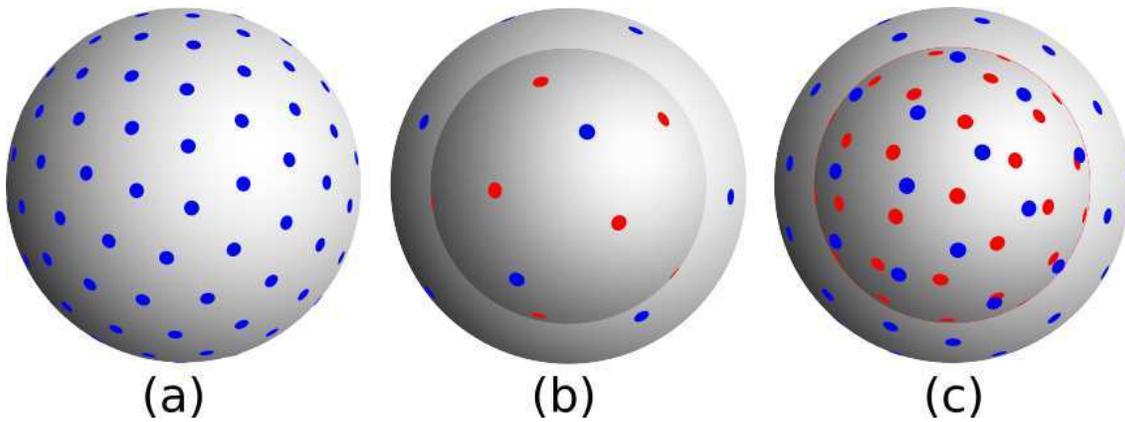


Figure 5: Sampling schemes used for validation. (a) A single shell sampling scheme with 64 measurements uniformly spread on one shell at a b-value  $b = 3000 \text{ s} \cdot \text{mm}^{-2}$ . (b) A multiple shells sampling scheme with 15 measurements spread on 2 shells at b-values  $b = 1500, 2500 \text{ s} \cdot \text{mm}^{-2}$ . (c) A multiple shells sampling scheme with 64 measurements spread on 2 shells at b-values  $b = 1500, 2500 \text{ s} \cdot \text{mm}^{-2}$ .

We perform the experiments with two metrics used in the contest, i.e. the weighted difference in the number of fiber compartments (DNC) and the mean angular error (AE) at each voxel. For these two metrics we extract the maxima on the estimated ODFs and compare them to the ground truth maxima. Then, the DNC becomes the difference between the number of maxima extracted on the estimated ODFs  $M_e$  and the true number of maxima  $M_{gt}$ , weighted by the true number of maxima at each voxel, i.e.,  $DNC = \frac{|M_e - M_{gt}|}{M_{gt}}$ . The AE is the mean angular error between the maxima extracted on the estimated ODFs and the respective maxima within the ground truth. We also enrich the results with a comparison of two other features, which have not been used in the contest validation, i.e. the diffusion signal and the Ensemble Average Propagator (EAP). To compare these features, we compute the Normalized Mean Square Error (NMSE) between the ground truth feature  $x$  and its estimation  $x_e$  given by  $NMSE = \frac{\|x - x_e\|_2^2}{\|x\|_2^2}$ . The DNC, AE and NMSE are, then, averaged on all the voxels. We add Rician noise to the normalized diffusion signal in the following way :  $E_{noisy} = \sqrt{(E + \epsilon_1)^2 + \epsilon_2^2}$ , where  $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, \sigma)$  with  $\sigma = 1/SNR$ .

The three following sections present the results for the three proposed sampling schemes, i.e. the SS sampling scheme with 64 measurements, the MS sampling scheme with 15 measurements, and the MS sampling scheme with 64 measurements. We also compare with state of the art techniques such that QBI, using the solid angle ODF [1], and the SHORE reconstruction using a sparse prior ( $\ell_1$ -SHORE) [34, 12, 28, 31, 26]. We show quantitative results in Tab. 1, 2, 3 and 4, and qualitative results in Fig. 6, 7 and 8. In each table, we write in blue letters the best score for a given SNR and metric.

#### 4.3.1. Single shell sampling scheme with 64 measurements

In these experiments, we first use the SS sampling scheme, i.e. 64 measurements uniformly spread on a shell at a  $b$ -value  $b = 3000 \text{ s} \cdot \text{mm}^{-2}$ , and we compare our dictionary based ODF estimation (D-ODF) to the solid angle ODF via QBI (SA-ODF) developed in [1]. For the SA-ODF, we set a SH order equal to the one used for the dictionary construction, i.e.  $L = 8$ . We adjust the Laplace-Beltrami regularization parameter [15, 1] using the generalized cross validation algorithm [13]. For our PDL approach, because we deal with  $\ell_1$  norm and not  $\ell_2$  norm, we use a simple cross validation procedure [45] to find the regularization parameter  $\lambda$  in Eq. 12.

Overall, in Fig. 6 the D-ODFs are sharper than the SA-ODFs. Furthermore, the SA-ODFs appear more sensitive to noise than the ODFs based on our PDL estimation. Indeed, we observe that the D-ODFs are very robust to noise, even at  $SNR=10$  where they are still correctly aligned with the underlying structure shown by the ground truth (On the right of Fig. 6), whereas the maxima extracted from the SA-ODFs mostly give corrupted or completely false fiber orientation estimation even for simple configuration as single fibers.

The quantitative results, in Tab. 1, confirms our previous remarks. In particular, at  $SNR=10$ , the SA-ODFs are not able

to provide proper diffusion directions. The DNC mean value is higher than 0.5 meaning that, in average, more than half of the maxima in each voxel are not detected. The results regarding our PDL approach give more accurate diffusion directions at every SNR.

Besides the good directional information given by the ODFs estimated with our PDL approach, we can also estimate the EAP and interpolate/extrapolate the diffusion signal on the entire  $q$ -space, whereas QBI only estimates the ODFs. More results regarding the estimation of the latter features are given in Tab. 4, and we will discuss these results in Sec. 4.3.3.

#### 4.3.2. Multiple shells sampling scheme with 15 measurements

In these experiments, we compare a SHORE reconstruction using a sparse prior ( $\ell_1$ -SHORE) to our dictionary reconstruction. The SHORE basis has been introduced by [34] and was used in [28, 31, 26] in the context of sparse recovery. The  $\ell_1$ -SHORE method consists in solving the LASSO problem (see Eq. 12) using coordinate descent while replacing the dictionary  $\mathbf{D}$  by the SHORE basis [34]. In order to provide a fair comparison a SH order of  $L = 8$  is used for the generation of the SHORE basis. In both methods ( $\ell_1$ -SHORE and our PDL approach) we use cross validation [45] to assess the regularization parameter  $\lambda$  in Eq. 12. [34] and [12] respectively provide closed formulae to estimate the ODF and the EAP when the diffusion signal is modeled in the SHORE basis.

We see in Tab. 2 that our PDL approach outperforms the  $\ell_1$ -SHORE reconstruction in terms of angular error, difference in the number of compartments (DNC), signal NMSE and EAP NMSE.

Fig. 7 shows the ODFs estimated via our dictionary approach (D-ODF) and the ODFs estimated via the  $\ell_1$ -SHORE method (SHORE-ODF). It qualitatively indicates an improvement of the angular information given by the D-ODFs over the SHORE-ODFs.

Regarding the PDL approach proposed at the ISBI contest, we obtained an angular error equal to  $14.5^\circ$  at  $SNR=10$ , equal to  $11^\circ$  at  $SNR=20$ , equal to  $9.5^\circ$  at  $SNR=30$ . The results based on our new proposed framework, shown in Tab. 2, indicates an improvement on the accuracy of the maxima estimation compared to the results obtained in the contest. Note that the PDL approach proposed in the contest was already the best in its category.

In the following we show the impact of an increase of the number of samples while keeping the two shells at  $b$ -values  $b = 1500, 2500 \text{ s} \cdot \text{mm}^{-2}$ .

#### 4.3.3. Multiple shells sampling scheme with 64 measurements

Now, we use the MS sampling scheme with 64 measurements and we compare once again the  $\ell_1$ -SHORE method to our PDL approach. Our dictionary approach still outperforms the SHORE reconstruction in terms of angular error, DNC, signal NMSE and EAP NMSE. Fig. 8 shows the ODFs estimated via our PDL approach (D-ODF) and the ODFs estimated via the  $\ell_1$ -SHORE method (SHORE-ODF). We observe that the D-ODFs give a very accurate estimation of the underlying fiber structure where the SHORE-ODFs fail to provide coherent fiber

Single shell sampling scheme with 64 measurements	Angular error		DNC	
	D-ODF	SA-ODF	D-ODF	SA-ODF
SNR 30	4.9398	6.6509	0.2068	0.2329
SNR 20	5.6386	7.1436	0.2102	0.2524
SNR 10	8.2530	12.419	0.2500	0.5993

Table 1: Dictionary based ODF estimation (D-ODF) versus solid angle ODF via QBI (SA-ODF) using a single shell sampling scheme with 64 measurements. We added rician noise from SNR=10 to 30. Two metrics are shown : the angular error and the difference in the number of compartments (DNC).

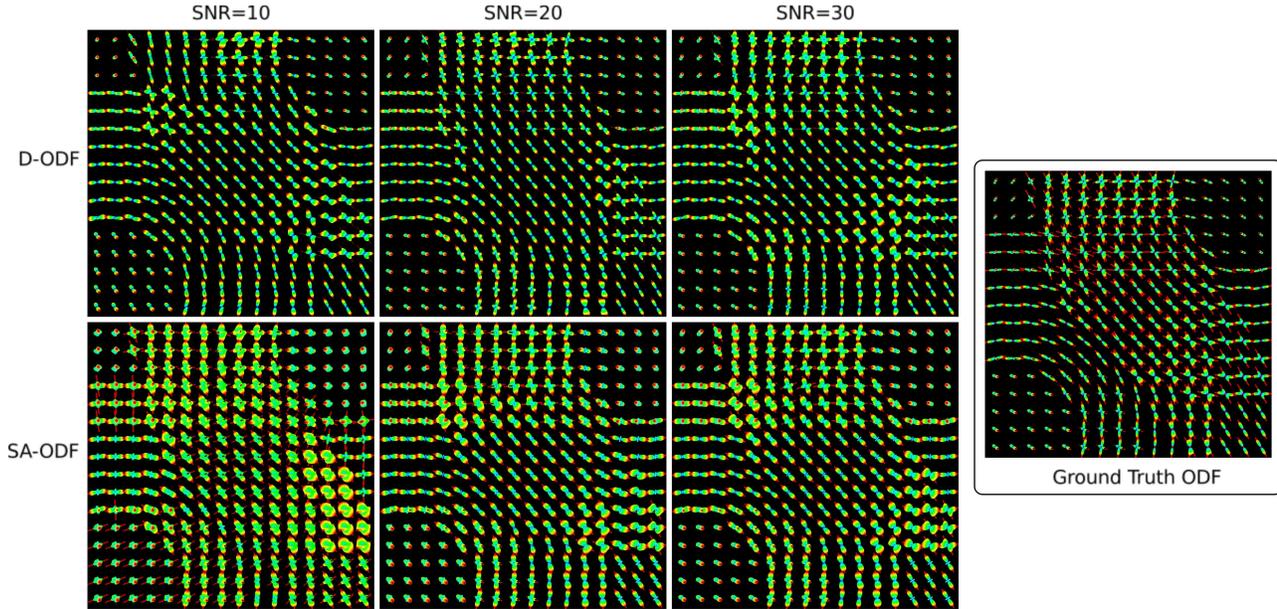


Figure 6: Dictionary based ODF estimation (D-ODF, top) versus solid angle ODF via QBI (SA-ODF, bottom). We added rician noise from SNR=10 to 30 (left to right). We also show the maxima extracted from the ODFs. On the right, we show the ground ODFs and their respective maxima.

MS sampling scheme with 15 measurements	Angular error		DNC		Signal NMSE		EAP NMSE	
	Dictionary	$\ell_1$ -SHORE	Dictionary	$\ell_1$ -SHORE	Dictionary	$\ell_1$ -SHORE	Dictionary	$\ell_1$ -SHORE
SNR 30	8.6066	14.670	0.2472	0.4010	0.0104	0.0433	0.0176	0.1040
SNR 20	9.7626	16.313	0.2540	0.4463	0.0170	0.0578	0.0251	0.1122
SNR 10	13.344	22.354	0.2734	0.4836	0.0343	0.1027	0.0422	0.1350

Table 2: Dictionary based reconstruction versus  $\ell_1$ -SHORE based reconstruction using a multiple shells sampling scheme with 15 measurements. We added rician noise from SNR=10 to 30. Four metrics are shown : the angular error, the difference in the number of compartments (DNC), the signal NMSE and the EAP NMSE.

MS sampling scheme with 64 measurements	Angular error		DNC		Signal NMSE		EAP NMSE	
	Dictionary	$\ell_1$ -SHORE	Dictionary	$\ell_1$ -SHORE	Dictionary	$\ell_1$ -SHORE	Dictionary	$\ell_1$ -SHORE
SNR 30	5.6233	8.8950	0.2187	0.3106	0.0035	0.0368	0.0066	0.0732
SNR 20	6.3080	9.6641	0.2309	0.3401	0.0054	0.0386	0.0104	0.0746
SNR 10	8.3224	13.126	0.2511	0.3995	0.0144	0.0752	0.0222	0.0825

Table 3: Dictionary based reconstruction versus  $\ell_1$ -SHORE based reconstruction. We added rician noise from SNR=10 to 30. Four metrics are shown : the angular error, the difference in the number of compartments (DNC), the signal NMSE and the EAP NMSE.

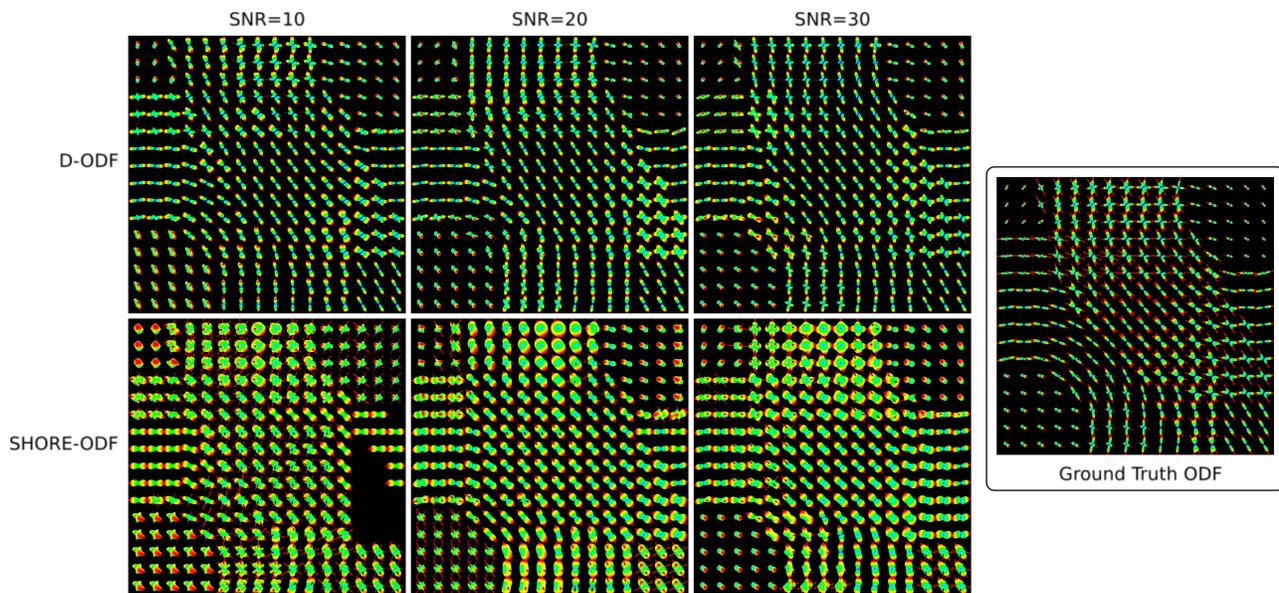


Figure 7: Dictionary based ODF estimation (D-ODF, top) versus  $\ell_1$ -SHORE based ODF estimation (SHORE-ODF, bottom). We added rician noise from SNR=10 to 30 (left to right). We also show the maxima extracted from the ODF. On the right, we show the ground ODFs and their respective maxima.

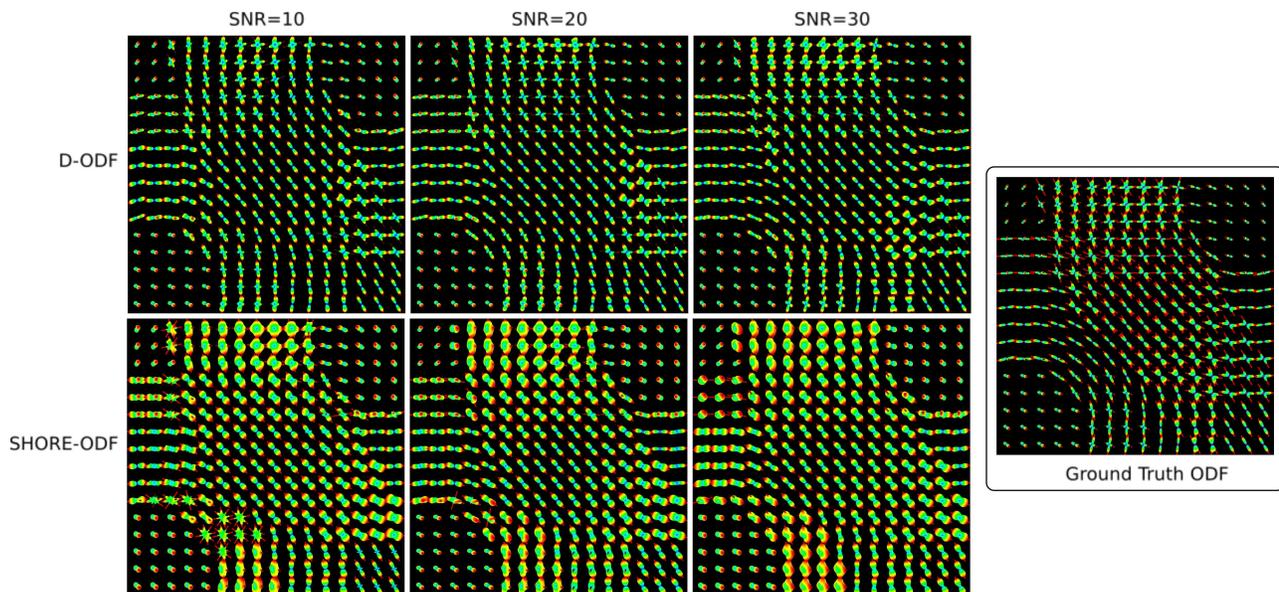


Figure 8: Dictionary based ODF estimation (D-ODF, top) versus  $\ell_1$ -SHORE based ODF estimation (SHORE-ODF, bottom) using a multiple shells sampling scheme with 64 measurements. We added rician noise from SNR=10 to 30 (left to right). We also show the maxima extracted from the ODFs. On the right, we show the ground ODFs and their respective maxima.

direction estimation, especially at SNR=10. We confirm this by the quantitative results shown in Tab. 3.

We also compare the results of our PDL approach using this MS sampling scheme, i.e. 2 shells at b-values  $b = 1500, 2500 \text{ s} \cdot \text{mm}^{-2}$  with 64 measurements, to our PDL approach using the SS sampling scheme studied in Sec. 4.3.1, i.e. one shell at b-value  $b = 3000 \text{ s} \cdot \text{mm}^{-2}$  with 64 measurements. All the results are given in Tab. 4. Regarding the directional results (angular error and DNC), we observe a slight advantage with the SS sampling scheme. Moreover, if we want to reconstruct the full diffusion signal and the EAP, we notice that a MS sampling scheme is more adequate. This is because the radial information of the diffusion process is better considered when using 2 shells instead of only one shell.

#### 4.4. Discussion on experiments with synthetic data

We studied the choice of the free parameters in the learning process, i.e. the dictionary angular and radial orders (4.2), and  $\lambda$  in Eq. 9 (4.1). Our study led to a very good estimation of the diffusion direction (via the computation of the ODF), the diffusion signal and the EAP.

In particular, our PDL approach was shown to better estimate the diffusion directions than the solid angle ODF via QBI does and to compute the diffusion signal and the EAP in a more accurate way than a SHORE reconstruction (using a sparse prior) does.

Regarding the sampling scheme, we observe a slight advantage for the SS sampling scheme considering the directional features, but we found that the MS sampling at b-values  $b = 1500, 2500 \text{ s} \cdot \text{mm}^{-2}$  is more appropriate, because it enables one to deal with the radial information in addition to the angular information. In Sec. 4.3.3 the dictionary reconstruction combined with a multiple shells sampling scheme was proved to efficiently approximate the diffusion signal, the EAP and the ODF.

Regarding the minimal number of measurements required before a large decrease of the reconstruction accuracy, we also perform a last experiment on synthetic data, in which we study the impact of the number of samples  $N$  on the error metrics presented in this experiment part with synthetic data. For this purpose, we still consider the same set of signals (different from the training data set) and add Rician noise with  $SNR = 20$ . We use the MS sampling scheme with 2 shells at b-values  $b = 1500, 2500 \text{ s} \cdot \text{mm}^{-2}$  and vary the number of samples between  $N = 5$  and  $N = 100$ . Fig. 9 shows the resulting values. In this figure, we also plot a vertical line, which represents the number of samples  $N$  where the metric errors show a large increase of their values (which means that the estimations are not correct anymore). Overall, we define this bound to  $N_{min} \approx 15$ . In Fig. 9, we also represent the variance of each metric error. We observe from the variance, that the estimation of the diffusion signal and the EAP are robust to noise. However, the estimation of the diffusion direction is more sensitive to noise when we reduce the number of samples.

In conclusion, these synthetic experiments show that our PDL approach can sparsely model multi-fiber compartments

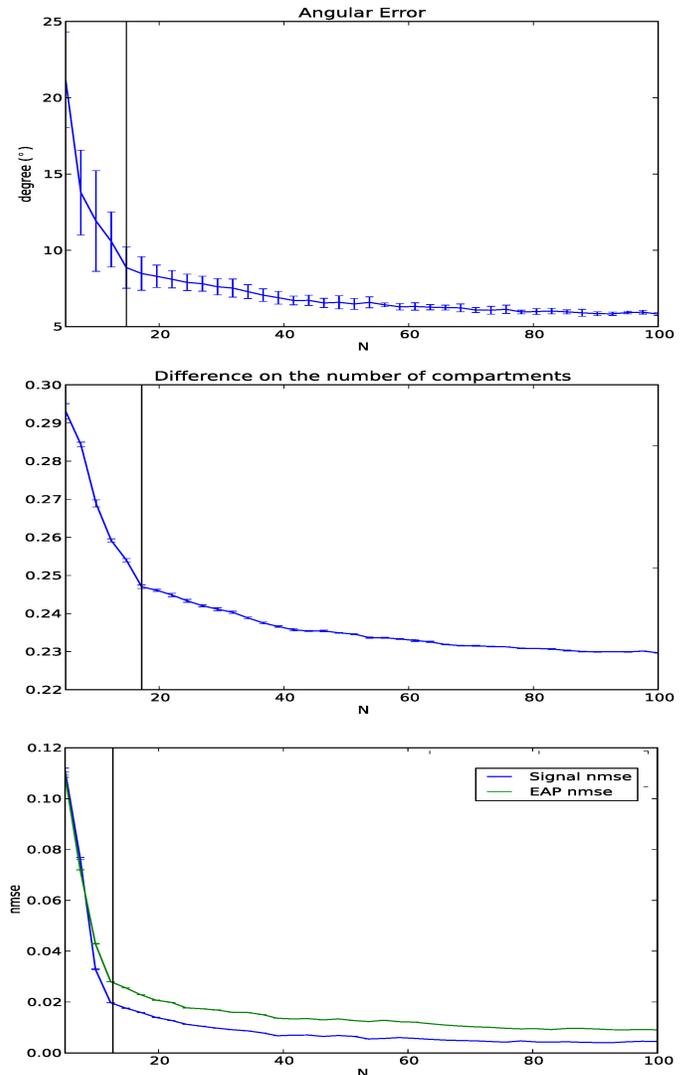


Figure 9: Evolution of the angular error, the DNC, the signal NMSE and the EAP NMSE in function of the number of samples  $N$ . We added Rician noise with SNR=20.

	Angular error		DNC		Signal NMSE		EAP NMSE	
	MS-SC	SS-SC	MS-SC	SS-SC	MS-SC	SS-SC	MS-SC	SS-SC
SNR 30	5.6233	<b>4.9398</b>	0.2187	<b>0.2068</b>	<b>0.0035</b>	0.0207	<b>0.0066</b>	0.0298
SNR 20	6.3080	<b>5.6386</b>	0.2309	<b>0.2102</b>	<b>0.0054</b>	0.0239	<b>0.0104</b>	0.0321
SNR 10	8.3224	<b>8.2530</b>	0.2511	<b>0.2500</b>	<b>0.0144</b>	0.0433	<b>0.0222</b>	0.0526

Table 4: Comparison between the multiple shells sampling scheme (MS-SC) with 64 measurements and the single shell sampling scheme (SS-SC) with 64 measurements, on reconstruction using the PDL approach. Four metrics are shown: the angular error, the difference in the number of compartments (DNC), the signal NMSE, the EAP NMSE. The simulations are performed with Rician noise at SNR=10, 20 and 30.

signals with the assumption of mono-exponential signal decay with b-value. In addition, it overcomes the preliminary PDL approach presented at the HARDI contest at ISBI 2012, already ranked first in its category.

## 5. Experiments on real data

In this section, we propose to validate our parametric dictionary learning (PDL) method on real data from human brains. For this purpose, we acquired three distinct sets of data :

- A first set of measurements coming from a 7T scanner, used both to learn the dictionary and to validate it.
- A second set of measurements coming from a 3T scanner, used for the learning process.
- A third set of measurements coming from a 3T scanner, used to validate the dictionary learned on the previous 3T scanner data.

### 5.1. Learning and reconstruction on a 7T scanner data

Training data was acquired on a 7T whole-body MR scanner (MAGNETOM, Siemens Healthcare) equipped with Siemens-AC072 whole body gradient coils, and an 24 channel phased array coil (Nova Medical). 12 axial slices were acquired with a 2D single shot DW-STEAM-EPI (TR/TE/ $\Delta/\delta = 3000/58/120/15$  ms) sequence giving 2 mm isotropic resolution. The echo time (TE), diffusion time ( $\Delta$ ), and gradient duration ( $\delta$ ) were optimized to provide maximum SNR for a maximum b-factor. 8 different b-values  $b=500/1000/2000/3000/4000/5000/6000/7000$  s/mm<sup>2</sup> are considered with 70 orientations at each b-value, and an imaging matrix of 96x96x12. The measurement locations are distributed using the algorithm given in [10] by setting the parameters in such way that there are a constant number of measurements per b-value. We use 11 slices as training data set, in which only the voxels corresponding to a Fractional Anisotropy (FA) superior to 0.20 are taken into account. The FA is computed from the diffusion tensor estimated with the whole set of measurements.

We validate the reconstruction, based on the previously learned dictionary, on the twelfth slice. We compare the ODFs estimated via our PDL approach (D-ODF) to the ODFs estimated with the  $\ell_1$ -SHORE method (SHORE-ODF) on three different sampling schemes. The three sampling schemes consider measurements at b-values  $b=1000/2000$  s/mm<sup>2</sup>, and we change the number of measurements as N=15, 30 and 60. To distribute

the samples between the two b-values, we use the algorithm given in [10] by setting the parameters in such a way that the number of samples on each shell is proportional to  $q^1$  [10, 25].

We choose a region of interest and show the estimated ODFs (see Fig. 10) along with the extracted maxima. This region contains several crossing configurations with different degree of crossing, and thus is appropriate for ODF validation. In Fig. 10, the middle corresponds to the D-ODFs and the bottom to the SHORE-ODFs. From the left to the right, we see the results for N=15, 30 and 60 samples. We consider the ground truth as the estimated signal using the  $\ell_1$ -SHORE method when all the measurements are taken into account, i.e. 70x8 measurements. The corresponding ground truth ODF are shown at the top of Fig. 10.

Overall, the SHORE-ODFs lead to more false maxima than the D-ODFs. For instance, at  $N = 60$  in the region A (in red), the SHORE-ODFs are not able to correctly resolve the crossing fiber configuration (erroneous number of detected maxima). This phenomenon is emphasized when the number of measurements decreases, whereas our dictionary estimation still provides a coherent map of ODFs.

### 5.2. Learning and reconstruction on a 3T scanner data

We also train our dictionary on data from a 3T Verio (MAGNETOM, Siemens Healthcare) scanner equipped with a 32-channel head coil. The data were acquired at a spatial resolution of 2 mm<sup>3</sup> isotropic, for 6 different b-values  $b=500/1000/2000/3000/5000/7000$  s/mm<sup>2</sup>, 70 orientations at each b-value, and an imaging matrix of 128x128x60. The dictionary is learned from all the measurements on the axial slices 25 to 35, for the voxels with  $FA \geq 0.20$ . The FA is computed from the diffusion tensor estimated with the whole set of measurements.

For the reconstruction, we use data from a 3T scanner (Philips Achieva) equipped with a 8-channel SENSE coil. The data were acquired at a spatial resolution of 2 mm<sup>3</sup> isotropic, for 6 different b-values  $b=500/1000/2000/4000/6000/8000$  s/mm<sup>2</sup>, 70 orientations at each b-value, and an imaging matrix of 128x128x60. We consider three sampling schemes with N=15, 30 and 60 samples and distribute them proportionally to  $q^1$  on two b-values  $b=1000/2000$  s/mm<sup>2</sup> [10, 25].

In Fig. 11, the top corresponds to the D-ODFs and the bottom to the SHORE-ODFs, on a selected region of interest. From the left to the right, we see the results for N=15, 30 and 60 samples. We also consider the ground truth as the estimated signal using the  $\ell_1$ -SHORE method when all the measurements are

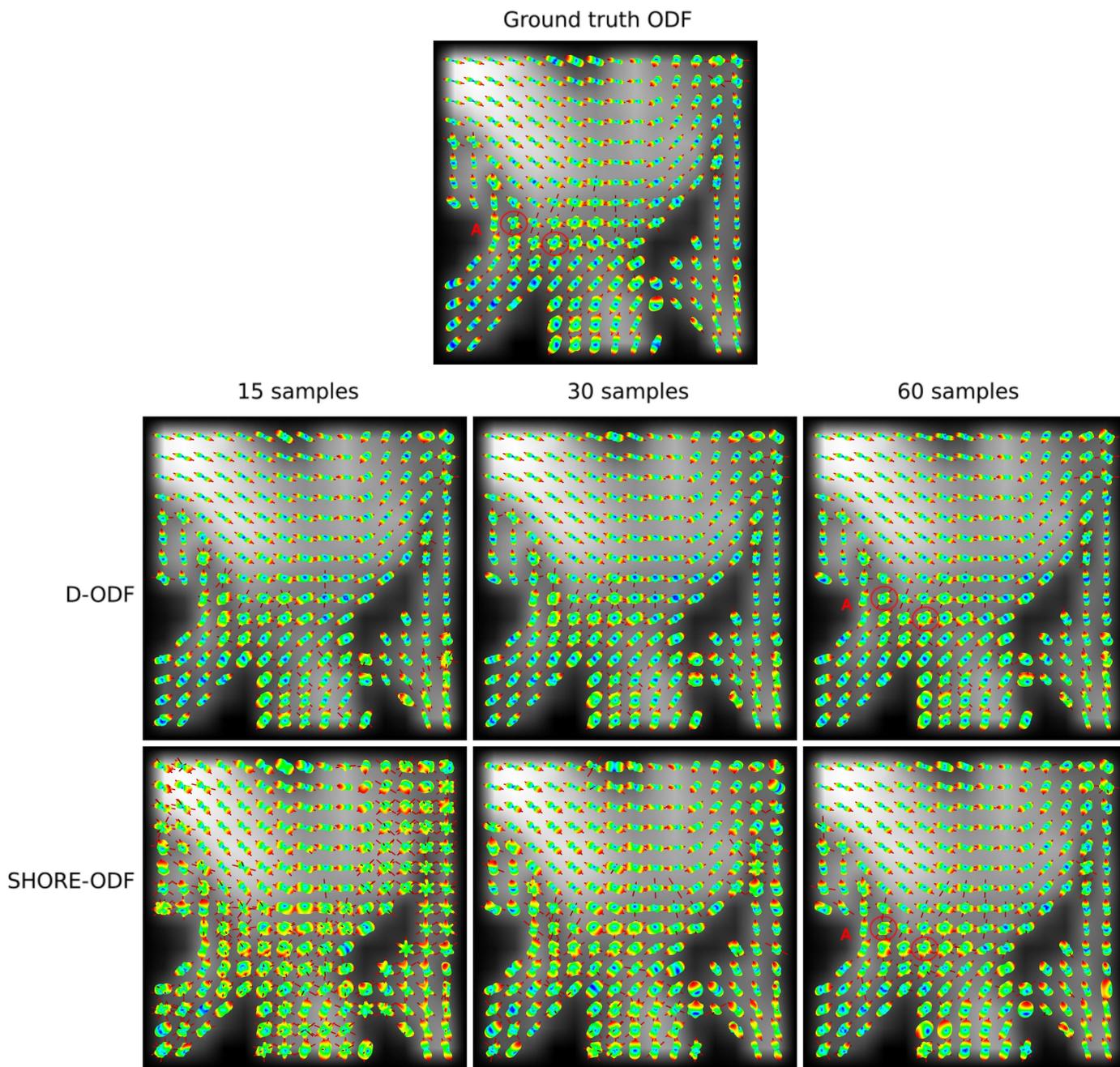
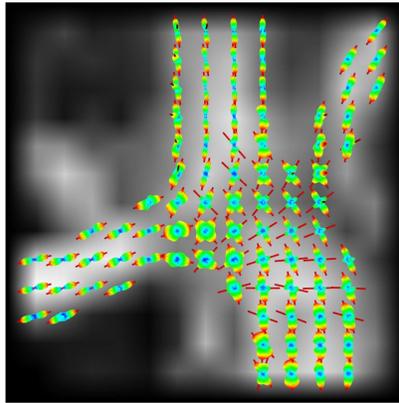


Figure 10: ODFs estimated from a 7T scanner data via our PDL approach (D-ODF, middle) and via the SHORE technique (SHORE-ODF, bottom).  $N=15, 30$  and  $60$  samples are considered (left to right). We show the ground truth ODF at the top.

Ground truth ODF

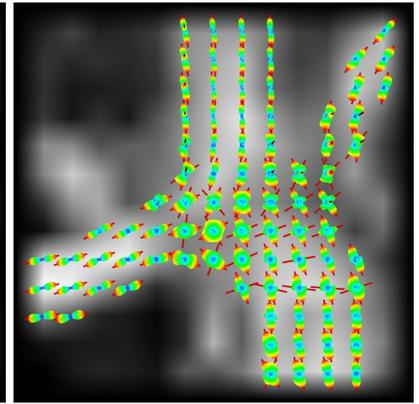
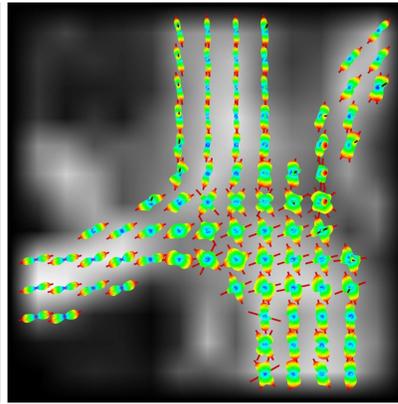
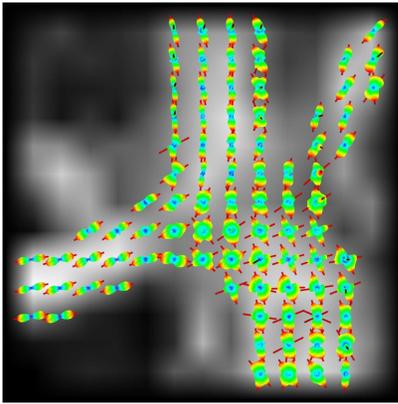


15 samples

30 samples

60 samples

D-ODF



SHORE-ODF

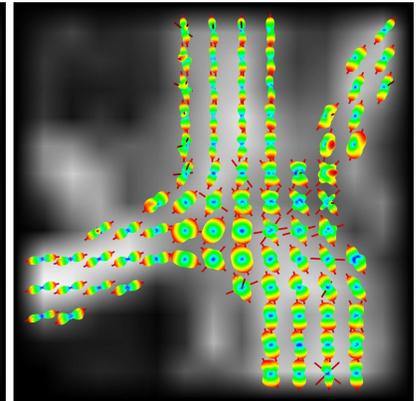
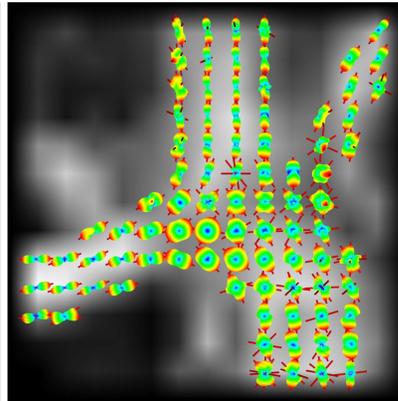
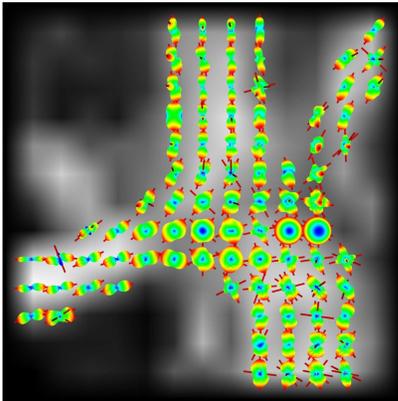


Figure 11: ODFs estimated from a 3T scanner data via our PDL approach (D-ODF, middle) and via the SHORE technique (SHORE-ODF, bottom). N=15, 30 and 60 samples are considered (left to right). We show the ground truth ODF at the top.

taken into account, i.e. 70x6 measurements. The corresponding ground truth ODFs are shown at the top of Fig. 11.

We obtain sharper ODFs estimation with our PDL approach on the selected region of interest. Again, we see that our dictionary is still able to model the fiber crossing configuration with only 15 samples, and provide less noisy ODFs than the ODFs estimated with SHORE.

## 6. Conclusion

We have proposed an original and efficient computational framework to model continuous diffusion MRI (dMRI) signal and to recover its important features such as the ODF and the EAP with a reduced number of measurements. The idea, we implemented, has been to use a parametric dictionary learning algorithm and to exploit the sparse property of a well designed dictionary to recover the diffusion signal and its features. Numerous experimental results have been carried out for validation on synthetic and human brain data acquired from 7-T and 3-T scanners. We have shown that we can clearly recover the diffusion signals and its features with a much better accuracy compared to state-of-the-art approaches and can accurately recover ODF in regions of multiple fiber crossing, even with a small and reduced number of measurements. This opens new perspectives for some dMRI applications, including for example tractography, where the improved characterization of the fiber orientations is likely to greatly and quickly help tracking through regions with and/or without crossing fibers

## Acknowledgments

We would like to express our thanks and gratitude to A. Anwander and D. Bibeck for the 7T whole-body MR scanner data, E. Banner who is part of the Neurinfo Imaging Platform (INRIA, INSERM, FEDER, Région Bretagne et Rennes metropole funding) for the 3T Magnetom Verio scanner data and G. Gilbert for the 3T scanner (Philips Achieva) data. This work has been partially supported by the NucleiPark research project (ANR Program Maladies Neurologique et maladies Psychiatriques) and the France Parkinson Association. Part of this work has been done within the framework of the France-Quebec Cooperation and thanks to the support of "Service de Cooperation et dAction Culturelle du Consulat General de France a Quebec"

## Appendix A. Derivation of the normalization constant

We define a dictionary of  $M$  functions  $f_m$ , from  $\mathbb{R}^3$  to  $\mathbb{R}$ . We want to normalize the atoms of the dictionary, for the classical  $\ell_2$  norm:

$$\|f_m\|_2^2 = \int_{\mathbb{R}^3} f_m(\mathbf{q})^2 d\mathbf{q} = 1. \quad (\text{A.1})$$

The functions  $f_m$  are constructed from elementary functions  $g_{i,j}^{(m)}$  (in what follows, we drop the index  $m$  for the sake of clarity):

$$f_m = \sum_{i=1}^I \sum_{j=1}^J g_{i,j}(\mathbf{q}), \quad (\text{A.2})$$

where

$$g_{i,j}(\mathbf{q} \cdot \mathbf{u}) = \gamma_{i,j} e^{-v_i q^2} q^{\ell(j)} Y_j(\mathbf{u}). \quad (\text{A.3})$$

The normalization in Eq. A.1 rewrites as

$$\begin{aligned} 1 &= \int_{\mathbb{R}^3} \left( \sum_{i,j} g_{i,j}(\mathbf{q}) \right)^2 d\mathbf{q} \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{i'=1}^I \sum_{j'=1}^J \gamma_{i,j} \gamma_{i',j'} \int_0^\infty e^{-v_i q^2} e^{-v_{i'} q^2} q^{\ell(j)} q^{\ell(j')} q^2 dq \\ &\quad \int_{S^2} Y_j(\mathbf{u}) Y_{j'}(\mathbf{u}) d^2\mathbf{u} \end{aligned} \quad (\text{A.4})$$

Provided that the spherical harmonics functions form an orthonormal basis, we have

$$\int_{S^2} Y_j(\mathbf{u}) Y_{j'}(\mathbf{u}) d^2\mathbf{u} = \delta_{j,j'}. \quad (\text{A.5})$$

Therefore, the normality constraint rewrites as

$$1 = \sum_{i=1}^I \sum_{j=1}^J \sum_{i'=1}^I \gamma_{i,j} \gamma_{i',j} \int_0^\infty e^{-(v_i+v_{i'})q^2} q^{2\ell(j)+2} dq. \quad (\text{A.6})$$

Let's use the substitution  $x = (u_i + u_{i'})q^2$  in the above integral, we have

$$1 = \sum_{i=1}^I \sum_{j=1}^J \sum_{i'=1}^I \frac{\gamma_{i,j} \gamma_{i',j}}{2(v_i + v_{i'})^{\ell(j)+3/2}} \int_0^\infty e^{-x} x^{\ell(j)+1/2} dx \quad (\text{A.7})$$

$$= \sum_{i=1}^I \sum_{j=1}^J \sum_{i'=1}^I \frac{\gamma_{i,j} \gamma_{i',j}}{2(v_i + v_{i'})^{\ell(j)+3/2}} \Gamma\left(\ell(j) + \frac{3}{2}\right) \quad (\text{A.8})$$

## Appendix B. Derivation of the Ensemble Average Propagator

The EAP  $P(\mathbf{R}\mathbf{r})$  is defined as

$$P(\mathbf{R}\mathbf{r}) = \int_{q=0}^\infty \int_{\mathbf{u} \in S^2} E(\mathbf{q}\mathbf{u}) \exp(+2\pi i \mathbf{q}\mathbf{R}\mathbf{u} \cdot \mathbf{r}) d\mathbf{u} q^2 dq. \quad (\text{B.1})$$

We use the spherical plane wave expansion

$$\exp(\pm 2\pi i \mathbf{q}\mathbf{R}\mathbf{u} \cdot \mathbf{r}) = \frac{2\pi}{\sqrt{qR}} \sum_{j=0}^\infty (\pm i)^{\ell(j)} J_{\ell+1/2}(2\pi qR) Y_j(\mathbf{u}) Y_j(\mathbf{r}), \quad (\text{B.2})$$

where  $J_{\ell+1/2}$  is the standard bessel function of the first kind and order  $\ell + 1/2$ . Then we get

$$\begin{aligned} P(\mathbf{R}\mathbf{r}) &= \int_{q=0}^\infty \int_{\mathbf{u} \in S^2} \left[ q^2 \sum_{k=0}^K c_k \frac{1}{\sqrt{\chi_k}} \sum_{i=0}^I \sum_{j=0}^J \gamma_{kij} \exp(-v_{ki} q^2) q^{\ell(j)} Y_j(\mathbf{u}) \right] \\ &\quad \left[ \frac{2\pi}{\sqrt{qR}} \sum_{j'=0}^\infty (\pm i)^{\ell(j')} j_{\ell}(2\pi qR) Y_{j'}(\mathbf{u}) Y_{j'}(\mathbf{r}) \right] d\mathbf{u} dq, \end{aligned} \quad (\text{B.3})$$

We shorten this expression using the orthonormal property of the spherical harmonic basis, i.e.  $\int_{\mathbf{u} \in S^2} Y_j(\mathbf{u}) Y_{j'}(\mathbf{u}) d\mathbf{u} = \delta_{jj'}$ . Then formula (B.3) becomes

$$P(R\mathbf{r}) = \sum_{k=0}^K c_k \frac{2\pi}{\sqrt{R\chi_k}} \sum_{i=0}^I \sum_{j=0}^J \gamma_{kij} (-1)^{\ell(j)/2} Y_j(\mathbf{r}) \underbrace{\int_{q=0}^{\infty} \exp(-v_{ki}q^2) q^{\ell(j)+3/2} J_{\ell(j)+1/2}(2\pi qR) dq}_{I_{kij}(R)}, \quad (\text{B.4})$$

We use the formula from [37], i.e.  $\int_{q=0}^{\infty} x^{\nu+1} \exp(-\alpha x^2) J_{\nu}(\beta x) = \frac{\beta^{\nu}}{(2\alpha)^{\nu+1}} \exp\left(-\frac{\beta^2}{4\alpha}\right)$

$$I_{kij}(R) = \int_{q=0}^{\infty} \exp(-v_{ki}q^2) q^{\ell(j)+3/2} J_{\ell(j)+1/2}(2\pi qR) dq \quad (\text{B.5})$$

$$= \frac{(2\pi R)^{\ell+1/2}}{(2v_{ki})^{\ell+3/2}} \exp\left(-\frac{(2\pi R)^2}{4v_{ki}}\right) \quad (\text{B.6})$$

Finally, we get a closed form for the propagator :

$$P(R \cdot \mathbf{r}) = \sum_{k=0}^K \frac{c_k}{\sqrt{\chi_k}} \sum_{i=0}^I \sum_{j=0}^J \gamma_{kij} (-1)^{\ell(j)/2} \left(\frac{\pi}{v_{ki}}\right)^{\ell(j)+3/2} R^{\ell(j)} \exp\left(-\frac{(\pi R)^2}{v_{ki}}\right) Y_j(\mathbf{r}) \quad (\text{B.7})$$

### Appendix C. Derivation of the Orientation Distribution Function

The ODF is given by

$$\Upsilon(\mathbf{r}) = \int_0^{\infty} P(R \cdot \mathbf{r}) R^2 dR. \quad (\text{C.1})$$

We insert (7) in (C.1) and gather all the  $R$ -dependant terms within the integral to get

$$\Upsilon(\mathbf{r}) = \sum_{k=0}^K c_k \frac{(2\pi)^{3/2}}{\sqrt{\chi_k}} \sum_{i=0}^I \sum_{j=0}^J \gamma_{kij} (-1)^{\ell(j)/2} \left(\frac{\pi}{v_{ki}}\right)^{\ell(j)+3/2} Y_j(\mathbf{r}) \underbrace{\int_0^{\infty} R^{\ell(j)+2} \exp\left(-\frac{(\pi R)^2}{v_{ki}}\right) dR}_{I(v_{ki}, \ell(j))} \quad (\text{C.2})$$

We use the formula from [37], i.e.  $\int_0^{\infty} x^{2n} \exp(-\rho x^2) dx = \frac{(2n-1)!!}{2(2\rho)^n} \sqrt{\frac{\pi}{\rho}}$ . Then,

$$I(v_{ki}, \ell(j)) = \frac{(\ell(j)+1)!!}{2\left(\frac{2\pi^2}{v_{ki}}\right)^{\ell(j)/2+1}} \sqrt{\frac{v_{ki}}{\pi}} \quad (\text{C.3})$$

And,

$$\Upsilon(\mathbf{r}) = \sum_{k=0}^K \frac{c_k}{\sqrt{\chi_k}} \sum_{i=0}^I \sum_{j=0}^J \gamma_{kij} (-1)^{\ell(j)/2} \left(\frac{\pi}{v_{ki}}\right)^{\ell(j)+1} \frac{(\ell(j)+1)!!}{2\left(\frac{2\pi^2}{v_{ki}}\right)^{\ell(j)/2+1}} Y_j(\mathbf{r}) \quad (\text{C.4})$$

### References

- [1] Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N., 2010. Reconstruction of the ODF in single and multiple shell q-ball imaging within constant solid angle. *Magnetic Resonance in Medicine* 64, 554–566.
- [2] Aharon, M., Elad, M., Bruckstein, A., 2006. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on* 54, 4311–4322.
- [3] Alexander, D., 2005. Maximum entropy spherical deconvolution for diffusion mri, in: Christensen, G., Sonka, M. (Eds.), *Information Processing in Medical Imaging, Springer Berlin Heidelberg*. pp. 76–87.
- [4] Anderson, A., 2005. Measurements of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine* 54, 1194–1206.
- [5] Assemlal, H., Tschumperle, D., Brun, L., 2009. Efficient and robust computation of pdf features from diffusion mr signal. *Medical Image Analysis* 13, 715–729.
- [6] Basser, P.J., Mattiello, J., Le Bihan, D., 1994a. Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance B*, 247–254.
- [7] Basser, P.J., Mattiello, J., Le Bihan, D., 1994b. Mr diffusion tensor spectroscopy and imaging. *Biophysical Journal* 66, 259–267.
- [8] Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* 2, 183–202.
- [9] Bilgic, B., Setsompop, K., Cohen-Adad, J., Yendiki, A., Wald, L., Adalsteinsson, E., 2012. Accelerated diffusion spectrum imaging with compressed sensing using adaptive dictionaries. *Magnetic Resonance in Medicine*.
- [10] Caruyer, E., Cheng, J., Lenglet, C., Sapiro, G., Jiang, T., Deriche, R., 2011. Optimal Design of Multiple Q-shells experiments for Diffusion MRI, in: *MICCAI Workshop on Computational Diffusion MRI - CDMRI'11, Toronto, Canada*.
- [11] Caruyer, E., Deriche, R., 2012. Diffusion MRI signal reconstruction with continuity constraint and optimal regularization. *Medical Image Analysis* 16, 1113–1120.
- [12] Cheng, J., Jiang, T., Deriche, R., 2011. Theoretical analysis and practical insights on eap estimation via a unified hardi framework, in: *MICCAI workshop on Computational Diffusion MRI*.
- [13] Craven, P., Wahba, G., 1985. Smoothing noisy data with spline functions. *Numerische Mathematik* 31, 377–403.
- [14] Deriche, R., Calder, J., Descoteaux, M., 2009. Optimal real-time q-ball imaging using regularized kalman filtering with incremental orientation sets. *Medical image analysis* 13, 564–579.
- [15] Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R., 2007. Regularized, fast, and robust analytical q-ball imaging. *Magnetic Resonance in Medicine* 58, 497–510.
- [16] Descoteaux, M., Deriche, R., Knosche, T., Anwander, A., 2009. Deterministic and probabilistic tractography based on complex fibre orientation distributions. *Medical Imaging, IEEE Transactions on* 28, 269–286.
- [17] Descoteaux, M., Deriche, R., Le Bihan, D., Mangin, J.F., Poupon, C., 2011. Multiple q-shell diffusion propagator imaging. *Medical Image Analysis* 15, 603–621.
- [18] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. *The Annals of statistics* 32, 407–499.
- [19] Gramfort, A., Cyril, P., Descoteaux, M., et al., 2012. Sparse dsi: Learning dsi structure for denoising and fast imaging, in: *MICCAI*.
- [20] Jian, B., Vemuri, B., Özarslan, E., Carney, P., Mareci, T., 2007. A novel tensor distribution model for the diffusion-weighted mr signal. *NeuroImage* 37, 164–176.

- [21] Johansen-Berg, H., Behrens, T.E.J. (Eds.), 2009. Diffusion MRI : from quantitative measurement to in-vivo neuroanatomy. Elsevier/Academic Press, Amsterdam, Boston. Nlmcopyc.
- [22] Jones, D., Horsfield, M., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine* 42, 515 – 525.
- [23] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M., et al., 1986. Mr imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology* 161, 401.
- [24] 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.
- [25] 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.
- [26] Merlet, S., Caruyer, E., Deriche, R., 2012a. Parametric dictionary learning for modeling eap and odf in diffusion mri, in: Lecture Notes in Computer Science series MICCAI 2012, Nice, France. p. 7512.
- [27] Merlet, S., Caruyer, E., Ghosh, A., Deriche, R., 2012b. Parametric dictionary learning in diffusion mri, in: HARDI Contest, ISBI - International Symposium on Biomedical Imaging, Alessandro Daducci and Jean-Philippe Thiran and Yves Wiaux, Barcelona, Spain.
- [28] Merlet, S., Deriche, Rachid, D., Whittingstall, K., Descoteaux, M., 2012c. Diffusion and multiple orientations from 1.5 mr systems with limited gradient tables, in: ISMRM 20th Annual Meeting (2012), Melbourne, Australia. p. 1922.
- [29] Merlet, S., Deriche, R., 2010. Compressed sensing for accelerated eap recovery in diffusion mri, in: Proceedings Computational Diffusion MRI - MICCAI Workshop.
- [30] Merlet, S., Deriche, R., 2012. Continuous diffusion signal, eap and odf estimation via compressive sensing in diffusion mri. *Medical Image Analysis* .
- [31] Merlet, S., Paquette, M., Deriche, R., Descoteaux, M., 2012d. Ensemble average propagator reconstruction via compressed sensing: Discrete or continuous bases ?, in: ISMRM 20th Annual Meeting (2012), Australia. p. 2277.
- [32] Michailovich, O., Rathi, Y., 2010. On approximation of orientation distributions by means of spherical ridgelets. *Image Processing, IEEE Transactions on* 19, 461–477.
- [33] Michailovich, O., Rathi, Y., Dolui, S., 2011. Spatially regularized compressed sensing for high angular resolution diffusion imaging. *Medical Imaging, IEEE Transactions on* 30, 1100–1115.
- [34] Ozarslan, E., Koay, C., Shepherd, T., Blackband, S., Basser, P., 2009. Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space mri., in: ISMRM 17th Annual Meeting and Exhibition, Honolulu., p. 1396.
- [35] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research* 12, 2825–2830.
- [36] Rathi, Y., Michailovich, O., Setsompop, K., Bouix, S., Shenton, M., Westin, C., 2011. Sparse multi-shell diffusion imaging, in: MICCAI, Springer. pp. 58–65.
- [37] Ryzhik, I., Jeffrey, A., Zwillinger, D., 2007. Table of integrals, series and products. Academic Press.
- [38] 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.
- [39] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* , 267–288.
- [40] Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion mri: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* 35, 1459 – 1472.
- [41] Tournier, J.D., Yeh, C.H., Calamante, F., Cho, K.H., Connelly, A., Lin, C.P., 2008. Resolving crossing fibres using constrained spherical deconvolution: Validation using diffusion-weighted imaging phantom data. *NeuroImage* 42, 617 – 625.
- [42] Tristan-Vega, A., Westin, C., 2011. Probabilistic odf estimation from reduced hardi data with sparse regularization. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011* , 182–190.
- [43] Tristan-Vega, A., Westin, S., Aja-Fernandez, S., 2009. Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging. *NeuroImage* 47, 638–650.
- [44] Tuch, D., 2004. Q-ball imaging. *Magnetic Resonance in Medicine* 52, 1358–1372.
- [45] Ward, R., 2009. Compressed sensing with cross validation. *Information Theory, IEEE Transactions on* 55, 5773 –5782.
- [46] Wedeen, V., Hagmann, P., Tseng, W., Reese, T., Weisskoff, R., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine* 54, 1377–1386.
- [47] Ye, W., Vemuri, B., Entezari, A., 2012. An over-complete dictionary based regularized reconstruction of a field of ensemble average propagators, in: Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, IEEE. pp. 940–943.
- [48] Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C., 2012. Noddi: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage* 61, 1000 – 1016.