

# Dictionary Learning on the Manifold of Square Root Densities and Application to Reconstruction of Diffusion Propagator Fields<sup>\*</sup>

Jiaqi Sun<sup>1</sup>, Yuchen Xie<sup>2</sup>, Wenxing Ye<sup>3</sup>, Jeffrey Ho<sup>1</sup>, Alireza Entezari<sup>1</sup>,  
Stephen J. Blackband<sup>4</sup>, and Baba C. Vemuri<sup>1,\*\*</sup>

<sup>1</sup> CISE Department, University of Florida, Gainesville, FL 32611, USA

<sup>2</sup> Qualcomm Inc., San Diego, CA 92121, USA

<sup>3</sup> Google Inc., Mountain View, CA 94043, USA

<sup>4</sup> Department of Neuroscience, McKnight Brian Institute, University of Florida,  
Gainesville, FL 32611, USA  
vemuri@cise.ufl.edu

**Abstract.** In this paper, we present a novel dictionary learning framework for data lying on the manifold of square root densities and apply it to the reconstruction of diffusion propagator (DP) fields given a multi-shell diffusion MRI data set. Unlike most of the existing dictionary learning algorithms which rely on the assumption that the data points are vectors in some Euclidean space, our dictionary learning algorithm is designed to incorporate the intrinsic geometric structure of manifolds and performs better than traditional dictionary learning approaches when applied to data lying on the manifold of square root densities. Non-negativity as well as smoothness across the whole field of the reconstructed DPs is guaranteed in our approach. We demonstrate the advantage of our approach by comparing it with an existing dictionary based reconstruction method on synthetic and real multi-shell MRI data.

**Keywords:** Dictionary learning, Manifold, DW-MRI, Diffusion propagator reconstruction.

## 1 Introduction

Diffusion weighted MRI, as a non-invasive imaging technique, helps explore the complex micro-structure of fibrous tissues through sensing the Brownian motion of water molecules [1]. Water diffusion is fully characterized by the diffusion Probability Density Function (PDF) called the diffusion propagator (DP) [2]. Under the narrow pulse assumption, the diffusion propagator denoted by

---

<sup>\*</sup> This research was in part funded by the NIH grant NS066340 to Baba C. Vemuri, and the following grants AFOSR FA9550-12-1-0304, ONR N000141210862, NSF CCF-1018149 to Alireza Entezari.

<sup>\*\*</sup> Corresponding author.

$P(\mathbf{r})$  and the diffusion signal attenuation  $E(\mathbf{q})$  are related through the Fourier transform[2]:

$$P(\mathbf{r}) = \int E(\mathbf{q}) \exp(-2\pi i \mathbf{q} \cdot \mathbf{r}) d\mathbf{q} \quad (1)$$

where  $E(\mathbf{q}) = S(\mathbf{q})/S_0$ ,  $S_0$  is the diffusion signal with zero diffusion gradient.

Given the diffusion MRI data, reconstructing the DP is one of the most important problems in the field. Numerous techniques have been proposed to this end [3–5]. For further reading, we refer the interested reader to a recent survey [6]. Most of these methods either assume a model in which case, the basis functions for reconstruction are predefined or in the case of model free approaches, they have to explicitly enforce the positivity constraints on the DP, which in some works was not done. In our work, we take a fresh approach to this problem namely a dictionary learning approach. This approach will move away from the requirement of pre-specifying the basis functions and instead learns it from the data and hence is data adaptive. This is a more flexible approach over fixing the basis. Further, by using the square root density representation of the DP, we make use of the intrinsic structure of the manifold of square root densities in the reconstruction process without having to resort to explicit enforcement of non-negativity constraint on DP reconstruction. In the following, we first present a brief review of relevant dictionary learning techniques and then present a review of state-of-the-art in DP reconstruction from multi-shell diffusion MRI.

## 1.1 Dictionary Learning on Riemannian Manifolds: Literature Review

Sparse coding which calls for modeling data as a linear combination of a small number of elements from a collection of atoms, i.e., the dictionary, has been proven very effective in many image processing tasks [7]. In these tasks, learning a dictionary that adapts well to the data is of great significance for a good performance of the sparse representation. Therefore, considering the geometric structure of the data space is critical to the success of dictionary learning. Most existing dictionary learning algorithms often assume that the data points and the atoms are vectors in a Euclidean space, and the dictionary is learned based on the vector space structure of the input data. However, the data involved in many image analysis tasks often reside on Riemannian manifolds such as the space of, symmetric positive definite (SPD) matrices [8] and square root densities. Therefore, the existing extrinsic approaches which overlook the potentially important intrinsic geometric structure of the data are inadequate in the context of such applications. Recently, this inadequacy was addressed by a few researchers [9] leading to the generalization of dictionary learning to manifolds, specifically, to the manifold of SPD matrices. However, most of these methods seek to transform the problem to a simpler space and solve it there, instead of respecting the geometric structure of the SPD matrix manifold. Needless to say, none of them truly incorporated the intrinsic geometry implied by the data as is done in this paper.

In general, dictionary learning in the Euclidean setting can be formulated as  $\min_{c_1, \dots, c_n, D} \sum_{i=1}^n \|s_i - Dc_i\|^2 + \mathbf{Sp}(c_i)$ , where  $s_1, \dots, s_n$  is the given collection of data points,  $D$  is the matrix with columns composed of the atoms  $a_i$ ,  $c_i$  the sparse coding coefficients and  $\mathbf{Sp}(c_i)$  the sparsity promoting term. When generalizing it to a Riemannian manifold  $\mathcal{M}$ , one of the key difficulties that needs resolution is to make sure that the collection of atoms as well as the approximation of data points generated using the atoms still lie on the manifold. The reason being, in Euclidean space, it is the global linear structure that guarantees the data synthesized from the atoms is contained in the same space, whereas, on Riemannian manifolds the Riemannian geometry provides only local linear structures through the Riemannian exponential and logarithmic maps. Yet, by taking advantage of this diversity of linear structures it is possible to formulate the dictionary learning in a data specific way. Details regarding this formulation will be discussed in subsequent sections. It suffices to say that we employ the *log* and *exp* maps along with an affine constraint to achieve this goal.

## 1.2 DP Reconstruction from Multi-shell Acquisitions: Literature Review

We now present a brief review of DP reconstruction from multi-shell diffusion MRI data. Various techniques have been proposed to reconstruct the DP from multi-shell acquisitions of the diffusion signal [5, 10], which, compared to single shell acquisitions provide additional information about the radial signal decay. Most of them assume a particular model for the diffusion signal as in q-ball imaging (QBI) [11]. As an alternative, another category of methods place weak assumptions about the diffusion signal and therefore are capable of generating relatively unbiased reconstruction results, such as diffusion spectrum imaging (DSI) proposed in [12] and the tomographic reconstruction methods in [13–15]. These methods interpolate the spherical domain data samples onto a dense regular lattice and then reconstruct the DP using the Fourier transform relationship between  $P(\mathbf{r})$  and  $E(\mathbf{q})$ . This idea is also adopted in our proposed method.

However, all of the aforementioned multi-shell methods solve the reconstruction problem in a voxel-wise manner, thereby always lead to a noisy reconstruction across the field. This gives us a strong motivation for applying dictionary learning to the reconstruction of DP fields, because the globally defined dictionary plays an implicit role in regularizing the reconstructions over the entire field. In recent years, a few dictionary learning based DP reconstruction methods have been proposed. In [16], Bilgic et al. applied adaptive dictionaries to accelerate the DSI method for estimating the DPs. In [17], Merlet et al. proposed a parametric dictionary learning framework obtaining a closed form DP and ODF modeling from diffusion MRI data. An over-complete dictionary based reconstruction of DP fields from single shell acquisition was presented in [18]. Nevertheless, due to the absence of explicit use of the geometric structure of the data space itself, none of these methods can guarantee the non-negativity of the reconstructed propagators, an intrinsic and basic property of the DP. Accordingly, these methods are prone to higher numerical errors.

Recently several approaches that guarantee the non-negativity of the reconstructed DP or ODF were proposed. For instance, in [19] authors used the Spherical Harmonic (SH) representation for ODF and enforced non-negativity on the continuous domain by enforcing the positive semi-definiteness of Toeplitz-like matrices constructed from the SH representation. Cheng et al. in [20] proposed to reconstruct ODFs (DPs) by estimating the square root of ODF (DP) called the wave-function directly from diffusion signals, ensuring non-negativity. The idea of taking advantage of the square root parameterization of DPs is also adopted in our proposed method. However, unlike dictionary based methods, the two methods discussed above are not guaranteed to yield a smooth reconstruction across the field and the reconstruction basis are pre-specified.

In this paper, we propose to apply the dictionary learning method generalized to the manifold of square root densities to the reconstruction of DP fields. As the nature of a globally learned dictionary indicates, our method will yield a smooth reconstruction which is desirable in real applications. By taking into consideration the intrinsic geometric structure of the manifold formed by the square root of DPs, our method performs better than the reconstruction techniques based on dictionary learning in a Euclidean setting. Furthermore, the non-negativity of the reconstructed DPs is naturally guaranteed due to the adoption of the square root representation and use of the intrinsic geometry of this space.

Rest of the paper is organized as follows. Section 2 contains brief background on Riemannian manifolds and the dictionary learning formulation. We present application to reconstruction of DP fields in section 3 and provide several examples in section 4. Finally section 5 contains conclusions.

## 2 Theory

### 2.1 Relevant Basics of Riemannian Manifolds

In this subsection, we briefly go over some fundamentals of Riemannian geometry, details of which can be found in [21]. A manifold  $\mathcal{M}$  of dimension  $d$  is a topological space that is locally homeomorphic to open subsets of the Euclidean space  $\mathbb{R}^d$  at each point. With a globally defined differential structure, manifold  $\mathcal{M}$  becomes a differentiable manifold. The global differential structure allows one to define the globally differentiable tangent space. The tangent space at  $p \in \mathcal{M}$  denoted by  $T_p\mathcal{M}$  is a vector space that contains all the tangent vectors to  $\mathcal{M}$  at point  $p$ . A Riemannian manifold is a differentiable manifold on which each tangent space  $T_p\mathcal{M}$  at point  $p$  is equipped with a differentiable varying inner product  $\langle \cdot, \cdot \rangle_p$ . The family of the inner products is called a Riemannian metric. Let  $p_i, p_j$  be two points on manifold  $\mathcal{M}$ , the geodesic curve  $\gamma : [0, 1] \rightarrow \mathcal{M}$  is a smooth curve with the minimum length connecting  $p_i$  and  $p_j$ . Let  $v \in T_p\mathcal{M}$  be a tangent vector to the manifold at point  $p$ , there exists a unique geodesic  $\gamma_v$  satisfying  $\gamma_v(0) = p$  with initial tangent vector  $v$ . The exponential map  $exp_p : T_p\mathcal{M} \rightarrow \mathcal{M}$  of  $v$  is defined as  $exp_p(v) = \gamma_v(1)$ . Logarithmic map, as the inverse of the exponential map, is denoted as  $log_p : \mathcal{M} \rightarrow T_p\mathcal{M}$ . Given two points  $p_i, p_j \in \mathcal{M}$ ,  $log_{p_i}$  maps point  $p_j$  to the unique tangent vector at  $p_i$  that is the initial velocity of

the geodesic  $\gamma$  with  $\gamma(0) = p_i$  and  $\gamma(1) = p_j$ . The geodesic distance between  $p_i$  and  $p_j$  is computed by  $dist(p_i, p_j) = \|\log_{p_i}(p_j)\|_{p_i}$ .

### 2.2 Dictionary Learning on Riemannian Manifolds: Formulation

In the Euclidean setting, given a collection of signals  $s_1, \dots, s_n \in \mathbb{R}^d$ , classical dictionary learning methods seek to find a dictionary  $D \in \mathbb{R}^{d \times m}$  whose columns consist of  $m$  atoms such that each signal  $s_i$  can be approximated as a sparse linear combination of these atoms  $s_i \approx Dc_i$ , where  $c_i \in \mathbb{R}^m$  is the coefficient vector. Using  $l_1$  regularization on  $c_i$ , the dictionary learning problem can be formulated as:

$$\min_{c_i, D} \sum_{i=1}^n (\|s_i - Dc_i\|_2^2 + \lambda \|c_i\|_1) \tag{2}$$

where  $\lambda$  is a regularization parameter.

In the Riemannian manifold setting, denote  $s_1, \dots, s_n \in \mathcal{M}$  as a collection of  $n$  data points on the manifold  $\mathcal{M}$ , and  $a_1, \dots, a_m \in \mathcal{M}$  as atoms of the learned dictionary  $\mathcal{D} = \{a_1, \dots, a_m\}$ . Due to the local linear geometric structure of  $\mathcal{M}$ , it is improper to use the linear combination of atoms  $\hat{s}_i = \sum_{j=1}^m c_{ij} a_j$  to approximate the data  $s_i$ , since there is no guarantee that  $\hat{s}_i$  is on the manifold. Instead, by using the geodesic linear interpolation on  $\mathcal{M}$ ,  $s_i$  can be estimated by  $\hat{s}_i = \exp_{s_i} \left( \sum_{j=1}^m c_{ij} \log_{s_i}(a_j) \right)$  where,  $\exp_{s_i}$  and  $\log_{s_i}$  are exponential and logarithmic map at  $s_i$  respectively, and  $c_{ij} \in \mathbb{R}$  are the coefficients. Intuitively, in order to approximate data point  $s_i$ , we project all the atoms in the dictionary to the tangent space at  $s_i$  and perform linear combination  $v_i = \sum_{j=1}^m c_{ij} \log_{s_i}(a_j)$  on the tangent vector space  $T_{s_i}\mathcal{M}$ , then the approximation  $\hat{s}_i$  is obtained by taking the exponential map of  $v_i$  at  $s_i$ .

Our goal is to build a dictionary that minimizes the sum of reconstruction error for each data point. Define

$$E_{\text{data}} = \sum_{i=1}^n \text{dist}(s_i, \hat{s}_i)^2 = \sum_{i=1}^n \|\log_{s_i}(\hat{s}_i)\|_{s_i}^2 = \sum_{i=1}^n \left\| \sum_{j=1}^m c_{ij} \log_{s_i}(a_j) \right\|_{s_i}^2. \tag{3}$$

By using the  $l_1$  sparsity regularization, the dictionary learning problem on the manifold  $\mathcal{M}$  can be formulated as the following optimization problem

$$\min_{\mathbf{C}, \mathcal{D}} \sum_{i=1}^n \left\| \sum_{j=1}^m c_{ij} \log_{s_i}(a_j) \right\|_{s_i}^2 + \lambda \|\mathbf{C}\|_1, \text{ s.t. } \sum_{j=1}^m c_{ij} = 1, i = 1, \dots, n \tag{4}$$

where  $\mathbf{C} \in \mathbb{R}^{n \times m}$  and the  $(i, j)$  entry of  $\mathbf{C}$  is written as  $c_{ij}$ . A similar data term was used in [22] but the atoms were assumed fixed. The affine constraint implies that we are using affine subspaces to approximate the data instead of the usual subspaces, which are simply affine subspaces based at the origin. Generalizing from vector spaces to Riemannian manifolds, there is no corresponding notion of

the origin that can be used to define subspaces, and this geometric fact requires the abandonment of the usual subspaces in favor of general affine subspaces. We can also introduce other regularizations in our framework instead of the  $l_1$  norm but that will be a topic for future research. Similar to classical dictionary learning methods, we use the iterative method to solve this optimization problem:

1. **Sparse coding step:** fix the dictionary  $\mathcal{D}$  and optimize with respect to the coefficients  $\mathbf{C}$ .
2. **Codebook optimization step:** fix  $\mathbf{C}$  and optimize with respect to  $\mathcal{D}$ .

The first step is a regular sparse coding problem that can be easily solved by many existing fast algorithms. However the second subproblem is much more challenging, since the optimization methods in Euclidean space are not appropriate for atoms on manifolds.

We developed a line search based algorithm on Riemannian manifold to update the dictionary  $\mathcal{D}$ . Let the cost function to be minimized be denoted by  $f(a_1, \dots, a_m)$ . First, we need to initialize the atoms in the dictionary. One possible choice of initialization is the  $m$  clusters of the data  $s_1, \dots, s_n$  generated by a K-means algorithm applied to all the data on  $\mathcal{M}$ . Then, a line search on the manifold is used to optimize  $f(a_1, \dots, a_m)$ . Intuitively, the idea is to find a descent direction  $v$  on the tangent space, and then walk a step along the geodesic  $\gamma$  whose initial velocity is  $v$ . The details are listed in Algorithm 1. The convergence analysis of the line search method on manifold is discussed in [23].

---

### Algorithm 1. Line search on Riemannian manifold

---

**Input:** A set of data  $\mathcal{S} = \{s_1, \dots, s_n\}$  on the manifold  $\mathcal{M}$ , coefficients  $\mathbf{C} \in \mathbb{R}^{n \times m}$  and initial dictionary atoms  $a_1^0, \dots, a_m^0$ .

**Output:** The optimal dictionary atoms  $(a_1^*, \dots, a_m^*)$  that minimize the cost function  $f(a_1, \dots, a_m)$ .

1. Set scalars  $\alpha > 0$ ,  $\beta, \sigma \in (0, 1)$  and initialize  $k = 0$ .
2. Compute  $\text{grad}f(a_1^k, \dots, a_m^k) = (\frac{\partial f(a_1^k)}{\partial a_1}, \dots, \frac{\partial f(a_m^k)}{\partial a_m})$
3. Pick  $\eta^k = (\eta_1^k, \dots, \eta_m^k) = -\text{grad}f$ , where  $\eta_i^k \in T_{a_i^k} \mathcal{M}$ .
4. Find the smallest  $t$  such that

$$f(\exp_{a_1^k}(\alpha\beta^t\eta_1^k), \dots, \exp_{a_m^k}(\alpha\beta^t\eta_m^k)) \leq f(a_1^k, \dots, a_m^k) - \sum_{i=1}^m \sigma\alpha\beta^t \|\eta_i^k\|_{a_i^k}.$$

5. Set  $a_i^{k+1} = \exp_{a_i^k}(\alpha\beta^t\eta_i^k)$ ,  $i = 1, \dots, m$ .
  6. Stop if  $f$  does not change much, otherwise set  $k = k + 1$  and go back to step 2.
- 

## 2.3 Manifold of Square Root Densities

In this section, without loss of generality we restrict the analysis to PDFs defined on the interval  $[0, T]$  for simplicity:  $\mathcal{P} = \{p : [0, T] \rightarrow \mathbb{R} | \forall s, p(s) \geq 0, \int_0^T p(s) ds = 1\}$ .

In [24], the Fisher-Rao metric was introduced to study the Riemannian structure formed by the statistical manifold. For a PDF  $p_i \in \mathcal{P}$ , the Fisher-Rao metric is defined as  $\langle v_j, v_k \rangle = \int_0^T v_j(s)v_k(s)\frac{1}{p_i(s)}ds$ , where  $v_j, v_k \in T_{p_i}\mathcal{P}$ . The Fisher-Rao metric is invariant to reparameterizations of the functions. In order to facilitate easy computations when using Riemannian operations, the square root density representation  $\psi = \sqrt{p}$  was used in [25]. The space of square root density functions is defined as  $\Psi = \{\psi : [0, T] \rightarrow \mathbb{R} | \forall s, \psi(s) \geq 0, \int_0^T \psi^2(s)ds = 1\}$ . As we can see,  $\Psi$  forms a convex subset of the unit sphere in a Hilbert space. Then the Fisher-Rao metric can be obtained as  $\langle v_j, v_k \rangle = \int_0^T v_j(s)v_k(s)ds$ , where  $v_j, v_k \in T_{\psi_i}\Psi$  are tangent vectors. Given any two functions  $\psi_i, \psi_j \in \Psi$ , the geodesic distance between these two points is given in closed form by  $\text{dist}(\psi_i, \psi_j) = \cos^{-1}(\langle \psi_i, \psi_j \rangle)$ , which is just the angle between  $\psi_i$  and  $\psi_j$  on the unit hypersphere. The geodesic at  $\psi_i$  with a direction  $v \in T_{\psi_i}\Psi$  is defined as  $\gamma(t) = \cos(t)\psi_i + \sin(t)\frac{v}{|v|}$ . Then, the exponential map can be represented as  $\text{exp}_{\psi_i}(v) = \cos(|v|)\psi_i + \sin(|v|)\frac{v}{|v|}$ . To ensure the exponential map is a bijection, we restrict  $|v| \in [0, \pi)$ . The logarithmic map is then given by  $\text{log}_{\psi_i}(\psi_j) = u \cos^{-1}(\langle \psi_i, \psi_j \rangle) / \sqrt{\langle u, u \rangle}$ , where  $u = \psi_j - \langle \psi_i, \psi_j \rangle \psi_i$ .

Using the expressions for  $\Psi$  discussed above, we can perform the dictionary learning on square root density functions. Let  $s_1, \dots, s_n \in \Psi$  be a collection of square root density functions, and  $a_1, \dots, a_m \in \Psi$  be atoms in the dictionary  $\mathcal{D}$ .  $\mathbf{C}$  is a  $n \times m$  matrix. If we use  $l_1$  regularization, our dictionary learning framework becomes

$$\min_{\mathbf{C}, \mathcal{D}} \sum_{i=1}^n \left\| \sum_{j=1}^m c_{ij} \cos^{-1}(\langle s_i, a_j \rangle) \frac{u_{ij}}{|u_{ij}|} \right\|_{s_i}^2 + \lambda \|\mathbf{C}\|_1, \text{ s.t. } \sum_{j=1}^m c_{ij} = 1, i = 1, \dots, n. \tag{5}$$

where  $u_{ij} = a_j - \langle s_i, a_j \rangle s_i$ . Note that in this formulation, the normalization on atoms  $a_i$  is not needed. Because by incorporating the manifold structure of the square root densities, the atoms we learned are always on the hypersphere, while traditional dictionary learning (Equation (2)) needs the normalization to guarantee the unique solution. This optimization problem can be efficiently solved using the algorithm presented in section 2.2.

### 3 Application to Reconstruction of DP Fields

As mentioned in the introduction section, in the DP reconstruction problem, we aim to reconstruct a smooth field of DPs  $P(\mathbf{r}, \mathbf{x})$  from a given field of multi-shell diffusion weighted MRI data  $E(\mathbf{q}, \mathbf{x})$ , where  $\mathbf{x}$  represents the spatial locations. We propose to solve this problem in two steps. Briefly speaking, the first step is to acquire a rough estimation of the DP at each voxel through the Fourier transform relationship between the signal  $E(\mathbf{q})$  and the DP,  $P(\mathbf{r})$ , specified in Equation (1). In the second step, in order to get a smooth reconstruction of the DPs over the entire field, we apply the proposed dictionary learning algorithm

on the set of square root densities obtained by taking the square root of the DPs estimated from step 1. The implementation details are given below.

Despite the simple relationship between the diffusion signal and the DP described in Equation (1), it is often infeasible to reconstruct the DP from the diffusion signal directly through Fourier transform. The reason is that in practice, the diffusion signal is sampled in the  $\mathbf{q}$  space following some pre-specified sampling scheme, which might not be uniform and regular. One solution is to define a regular lattice in  $\mathbf{q}$  space and estimate the values on this lattice through interpolation. Inspired by the work of Ye et al.[15], we choose Body Centered Cubic (BCC) lattice to be our regular lattice for interpolation and apply Fourier transform to the interpolated values to get an estimate of the DPs. As demonstrated in [15], in 3-D, BCC lattice is the optimal lattice for  $\mathbf{q}$  space sampling because its reciprocal lattice (i.e. the FCC lattice) is the densest sphere packing lattice.

Specifically, given  $N$  sample measurements  $E(\mathbf{q}_n)$  on multiple spherical shells, the desired  $K$  values  $E(\mathbf{x}_k)$  on the BCC lattice  $\mathbf{x}_k \in \mathcal{L}$  can be estimated by solving the following linear system  $E(\mathbf{q}_n) = \sum_{\mathbf{x}_k \in \mathcal{L}}^{1 \leq k \leq K} E(\mathbf{x}_k) \text{sinc}_{\mathcal{L}}(\mathbf{q}_n - \mathbf{x}_k)$ ,  $n = 1, \dots, N$ , where  $\text{sinc}_{\mathcal{L}}(x)$  is the ideal interpolation function that depends on the sampling lattice  $\mathcal{L}$ . The  $\text{sinc}_{\mathcal{L}}(x)$  for the BCC lattice is computed by Ye et. al in [15]. Once the  $K$  estimates  $E(\mathbf{x}_k)$  on the lattice are obtained, we get a continuous representation of  $E(\mathbf{q})$  as  $E(\mathbf{q}) = \sum_{\mathbf{x}_k \in \mathcal{L}}^{1 \leq k \leq K} E(\mathbf{x}_k) \text{sinc}_{\mathcal{L}}(\mathbf{q} - \mathbf{x}_k)$ . Taking Fourier transform on this equation, we get  $P(\mathbf{r}) = \text{box}(\mathbf{r}) \sum_{\mathbf{x}_k \in \mathcal{L}}^{1 \leq k \leq K} E(\mathbf{x}_k) \exp(-2\pi i \mathbf{x}_k \cdot \mathbf{r})$ , where  $\text{box}(\mathbf{r})$  is the Fourier transform of  $\text{sinc}_{\mathcal{L}}(\mathbf{q})$ .

According to the definition of DP,  $P(\mathbf{r})$  is a PDF defined on a 3-D displacement  $\mathbf{r}$  space. Therefore, by adopting the square root parameterization, we are able to map the estimated DPs from the space of PDFs to the manifold of square root densities. Let  $\psi(\mathbf{r}) = \sqrt{P(\mathbf{r})}$  denote the square root of the DP at a single voxel, we apply the proposed dictionary learning algorithm on the set of  $\psi(\mathbf{r})$  over the entire field. After solving for the globally defined dictionary  $\mathcal{D}$  and the coefficient matrix  $\mathbf{C}$  over the field, the reconstructed DPs can be obtained by solving a weighted mean problem on the hypersphere as described in section 2.2.

## 4 Experiments

In this section, we evaluate our reconstruction method by comparing it to a traditional dictionary learning based DP reconstruction method on both synthetic and real data sets.

### 4.1 Synthetic Data

We synthesized a  $32 \times 32$  field of diffusion signals simulating two straight fibers crossing in the center. The signals were generated using a mixture of two Gaussian functions. The data was sampled on multiple  $\mathbf{q}$  shells using the interlaced scheme described in [15]. Note that this sampling scheme is not a necessity for the application of our proposed method, we used it simply due to its high reso-

lution in  $\mathbf{q}$  space. Rician noise with level  $\delta$  varying from 0.05 to 0.3 was added on the generated data.

Next we give the parameter settings in our reconstruction framework. In the first step, we chose a BCC lattice to interpolate the signals onto, which consists of two staggered Cartesian lattices of size  $(11 \times 11 \times 11)$  and  $(12 \times 12 \times 12)$  respectively. The  $\|\mathbf{r}\|$  value to evaluate  $P(\mathbf{r})$  on was set to be 18. Then in the process of dictionary learning on the square root density manifold, we set the dictionary size to be 100.

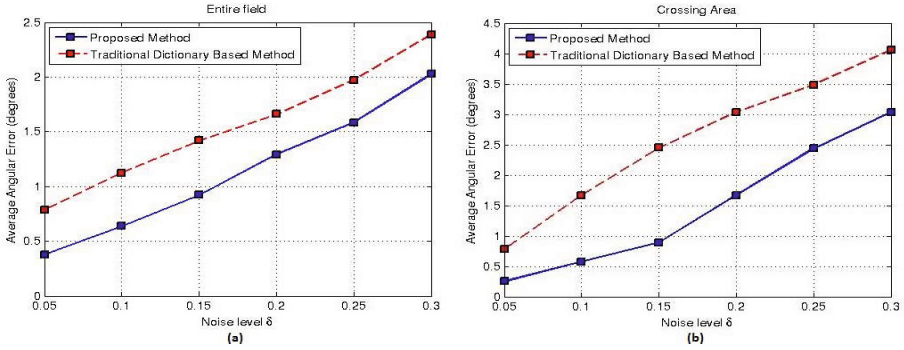
In order to demonstrate the advantage of incorporating the manifold structure into the reconstruction, we compared our method with the method in [18]. Since in [18] the authors adopted an adaptive kernel framework to model the signal in  $\mathbf{q}$  space, which assumes a single  $b$  value in the signal acquisition, their framework can not handle data acquired on multiple  $\mathbf{q}$  shells. Therefore, for the purpose of comparison we generalized it to make it applicable to multi-shell data, by using a tensor product of two 1-D splines in place of the 1-D spline in the Kernel they used. Also, we removed the NLM-based term in the cost function of [18] to achieve fair comparisons with our method.

We applied both methods to the synthetic data we generated with varying noise levels. The accuracy of the reconstruction was evaluated in terms of the average angular error over the entire field as well as within the crossing area. The angular error was computed based on the reconstructed  $P(\mathbf{r})$  value at  $\|\mathbf{r}\| = 18$ . The quantitative comparisons of the two methods are given in Fig.1. The plot shows that our method has a higher accuracy in the reconstruction over the traditional dictionary learning based method. Note that the scales of the Average Angular Error axis are different in these two graphs. By comparing them we can see that the advantage of our scheme over the other one is more significant in the crossing area, where an accurate reconstruction is more difficult to achieve. This good performance is due to the incorporation of the intrinsic geometric structure of the Riemannian manifold in our reconstruction process.

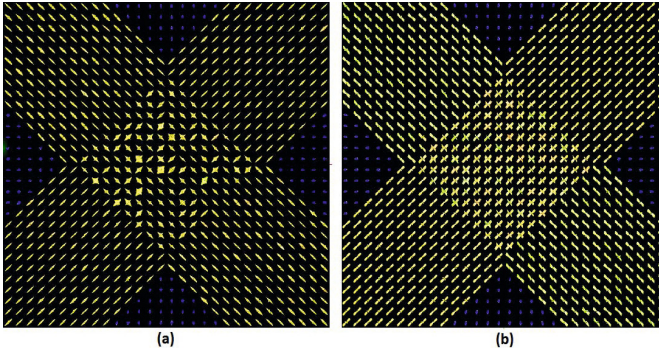
In addition to the numerical comparison, a visual comparison is also shown above. The plots of the reconstructed  $P(\mathbf{r})$  field from synthetic data with noise level  $\delta = 0.25$  for the two methods are displayed side by side in Fig. 2. As is shown in the image, our method yields a smoother reconstruction over the entire field, especially in the crossing region. Furthermore, in our result(b), the fiber directions can be easily identified at each voxel in the crossing area whereas in the other one(a), the information is lost at some locations.

## 4.2 Real Data

In this section we present experimental results on real diffusion MRI data acquired from two different regions of a mouse brain: 1) a sagittal set through the midline and 2) a coronal set at the level of the corpus callosum. All magnetic resonance imaging was performed on a  $600MHz$  Bruker imaging spectrometer, using a conventional diffusion weighted spin echo sequence. Parameters of the data acquisition for the two data sets are as follows. Dataset 1 was acquired with: *slicethickness* =  $0.35mm$ ,  $1.8 \times 0.9cm^2$  field-of-view,  $256 \times 128$  data matrix,



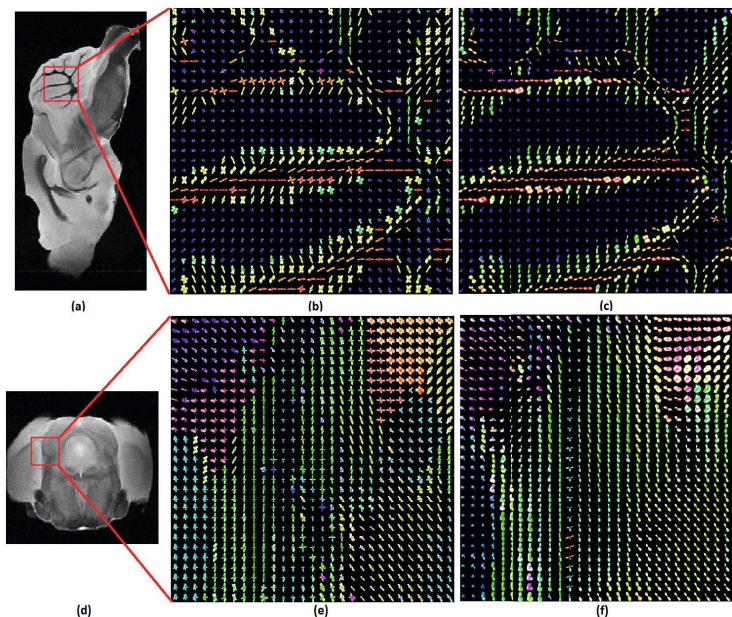
**Fig. 1.** The average angular error on the synthetic data set with varying noise levels. (a) Over the entire field. (b) Within the crossing area.



**Fig. 2.** Reconstruction of  $P(\mathbf{r})$  on synthetic data set with noise level  $\delta = 0.25$ . (a) The traditional dictionary learning based method. (b) The proposed method.

and  $70.3\mu\text{m}$  in-plane resolution. Diffusion parameters include: diffusion time,  $\Delta = 12\text{msec}$ , diffusion gradient duration,  $\delta = 1\text{msec}$ , and b-values of 187, 750, 1687 and  $3000\text{s/mm}^2$ . Dataset 2: *slicethickness* =  $0.3\text{mm}$ ,  $1.2 \times 1.2\text{cm}^2$  field-of-view,  $192 \times 192$  data matrix, and  $62.5\mu\text{m}$  in-plane resolution, the diffusion parameters are identical to the ones in Dataset 1. The sampling scheme was the same as used in the synthetic experiments.

The  $P(\mathbf{r})$  reconstruction results for both data sets are displayed in Fig.3. The images in the first row correspond to Dataset 1 while the second row Dataset 2. In each row, from left to right, the images are respectively the S0 image of the entire image plane including the ROI (the region in the red box), the reconstruction result using traditional dictionary learning based method and the one given by our proposed method. It is obvious from the visualization that our method performs better at smoothing out the noise from both data



**Fig. 3.** Reconstruction results on real data. (a) The  $S_0$  image of the entire field of Dataset 1, where the ROI is indicated by the red box. (b)  $P(\mathbf{r})$  Reconstruction using traditional dictionary learning based method on Dataset 1. (c)  $P(\mathbf{r})$  Reconstruction using the proposed method on Dataset 1. (d)(e)(f) are the corresponding images for Dataset 2.

sets and therefore yields smoother DP fields. Furthermore, the fiber orientations estimated in our method are in accordance with expectations. As we can see in the ROI of Fig 3(a), which is part of the mouse cerebellum, the orientations of the DPs in the white matter (corresponding to the dark region in the  $S_0$  image) are more consistent in (c) than in (b).

## 5 Conclusions

In this paper, we generalized the traditional dictionary learning methods on Euclidean space to Riemannian manifolds. Specifically, we proposed a novel dictionary learning framework for data on the manifold of square root densities and applied it to the reconstruction of DP fields from multi-shell diffusion MRI data. Through multiple synthetic and real data experiments, we showed that our reconstruction method performs well in comparison to the traditional dictionary learning based DP reconstruction methods, hence, justifying the incorporation of the geometric structure of the data space (square root density Riemannian manifold in our case) into our reconstruction.

## References

1. Basser, P., Mattiello, J., Lebihan, D.: Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance* (1994)
2. Callaghan, P.T.: Principles of nuclear magnetic resonance microscopy. Oxford University Press (1991)
3. Ozarslan, E., Shepherd, T.M., Vemuri, B.C., Blackband, S.J., Mareci, T.H.: Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). *Neuroimage* (2006)
4. Jian, B., Vemuri, B.C., Ozarslan, E., Carney, P.R., Mareci, T.H.: A novel tensor distribution model for the diffusion-weighted MR signal. *NeuroImage* (2007)
5. Descoteaux, M., Deriche, R., Bihan, D.L., Mangin, J., Poupon, C.: Multiple q-shell diffusion propagator imaging. *MIA* (2011)
6. Asselmlal, H., Tschumperle, D., Brun, L., Siddiqi, K.: Recent advances in diffusion MRI modeling: Angular and radial reconstruction. *MIA* (2011)
7. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* (2006)
8. Fletcher, P., Joshi, S.: Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing* (2007)
9. Sra, S., Cherian, A.: Generalized dictionary learning for symmetric positive definite matrices with application to nearest neighbor retrieval. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) *ECML PKDD 2011, Part III. LNCS (LNAI)*, vol. 6913, pp. 318–332. Springer, Heidelberg (2011)
10. Caruyer, E., Deriche, R.: Diffusion MRI signal reconstruction with continuity constraint and optimal regularization. *MIA* (2012)
11. Tuch, D.S.: Q-ball imaging. *MRM* (2004)
12. Wedeen, V.J., Hagmann, P., Tseng, W.Y., Reese, T.G., Weisskoff, R.M.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *MRM* (2005)
13. Pickalov, V., Basser, P.: 3D tomographic reconstruction of the average propagator from MRI data. In: *ISBI* (2006)
14. Wu, Y., Alexander, A.: Hybrid diffusion imaging. *NeuroImage* (2007)
15. Ye, W., Portony, S., Entezari, A., Blackband, S.J., Vemuri, B.C.: An efficient interleaved multi-shell sampling scheme for reconstruction of diffusion propagators. *IEEE TIP* (2012)
16. Bilgic, B., Setsompop, K., Cohen-Adad, J., Wedeen, V., Wald, L.L., Adalsteinsson, E.: Accelerated diffusion spectrum imaging with compressed sensing using adaptive dictionaries. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS*, vol. 7512, pp. 1–9. Springer, Heidelberg (2012)
17. Merlet, S., Caruyer, E., Deriche, R.: Parametric dictionary learning for modeling EAP and ODF in diffusion MRI. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS*, vol. 7512, pp. 10–17. Springer, Heidelberg (2012)
18. Ye, W., Vemuri, B.C., Entezari, A.: An over-complete dictionary based regularized reconstruction of a field of ensemble average propagators. In: *ISBI* (2012)
19. Schwab, E., Afsari, B., Vidal, R.: Estimation of non-negative ODFs using the eigenvalue distribution of spherical functions. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II. LNCS*, vol. 7511, pp. 322–330. Springer, Heidelberg (2012)

20. Cheng, J., Jiang, T., Deriche, R.: Nonnegative definite EAP and ODF estimation via a unified multi-shell HARDI reconstruction. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 313–321. Springer, Heidelberg (2012)
21. Spivak, M.: A comprehensive introduction to differential geometry. Publish or Perish, Berkeley (1979)
22. Cetingul, H.E., Vidal, R.: Sparse Riemannian manifold clustering for HARDI segmentation. In: ISBI (2011)
23. Absil, P., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press (2008)
24. Rao, C.R.: Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. (1945)
25. Srivastava, A., Jermyn, I., Joshi, S.: Riemannian analysis of probability density functions with applications in vision. In: CVPR (2007)