White Matter Fiber Representation using Continuous Dictionary Learning

Guy Alexandroni, Yana Podolsky, Hayit Greenspan, Tal Remez, Or Litany, Alex M. Bronstein, and Raja Giryes

Faculty of Engineering, Tel Aviv University, Tel Aviv ,Israel

Abstract. With increasingly sophisticated Diffusion Weighted MRI acquisition methods and modeling techniques, very large sets of streamlines (fibers) are presently generated per imaged brain. These reconstructions of white matter architecture, which are important for human brain research and pre-surgical planning, require a large amount of storage and are often unwieldy and difficult to manipulate and analyze. This work proposes a novel continuous parsimonious framework in which signals are sparsely represented in a dictionary with continuous atoms. The significant innovation in our new methodology is the ability to train such continuous dictionaries, unlike previous approaches that either used pre-fixed continuous transforms or training with finite atoms. This leads to an innovative fiber representation method, which uses Continuous Dictionary Learning to sparsely code each fiber with high accuracy. This method is tested on numerous tractograms produced from the Human Connectome Project data and achieves state-of-the-art performances in compression ratio and reconstruction error.

Keywords: WM fibers, MRI, HARDI, sparsity, compression, similarity, Continuous Dictionary Learning

1 Introduction

Diffusion Weighted Magnetic Resonance Imaging (DWMRI) is a powerful tool for creating extremely detailed 3D representations of White Matter (WM) architecture. It maps the structure of neural connections by measuring the local orientation of water molecule diffusion [6]. With the development of DWMRI, models are built to delineate the underlying anatomy of the WM tissue based on the diffusion data. Since water diffusion orientation correlates with the main neuron bundles trajectories, a natural approach is to track consistent diffusion pathways. This approach is called tractography (Fig.1) and it creates a set of fibers (termed tractogram) representing the brain's main neuron bundles [9]. Each fiber is represented by an array of 3D points sampled using fixed step size.

Analyzing tractograms may pose several challenges. First, neural pathways are continuous objects that vary in length significantly (between 10mm and 100mm). In representing the fibers as a set of 3D points, one can use a fixed sampling frequency, ending up with signals of different dimension; Alternatively,
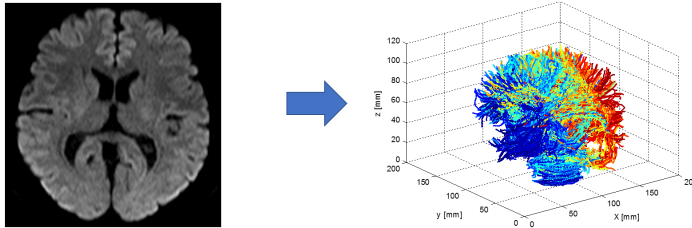
Fig. 1. DWMRI image (source: https://mrimaster.com/characterise image dwi .html) and a tractogram modeling the WM architecture (colors for visualization only).

fixed number of points can be used, leading to an inadequate representation of long fibers. To overcome this, alternative representation schemes are proposed, e.g. in [8] fibers and fiber bundles are presented as Gaussian Processes and this representation enables using an inner-product based metric for fiber similarity.

Modern tractography tools often produce highly-dense tractograms, containing millions of fibers, resulting in extremely large files. In population studies, multiple brains are analyzed and compared, therefore the storage issue is further exacerbated. The challenges above show the need to find more efficient representations for fibers, as well as fiber compression schemes. There have been some efforts in recent years to address this matter. For example, [7] suggested a compression framework for fibers which includes optimization on the number and location of samples along a given fiber in addition to quantization and encoding steps, which reduce the memory needed to store a full tractogram.

An alternative approach is the one of sparse coding that provides a better-compressed representation of the data compared to classic methods such as PCA and LDA and was shown to be beneficial in various applications [4]. In this scheme, each data point is assumed to be decomposed of a small number of columns (known as atoms) of a given dictionary, which may be either predefined (e.g., DCT or wavelet), or alternatively, can be inferred from the input data using dictionary learning methods. The latter tends to improve the sparse representation. While many dictionary learning strategies exist such as MOD and K-SVD [4], they are mainly designed for the case where the data lies on the same grid or has a fixed limited size. Therefore, the existing strategies are less suitable for dealing with data points of varying sizes, or with continuous data. Indeed, continuous dictionaries have been used before such as in the case of data admitting sparsity in the Fourier domain [3]. Yet, these dictionaries have not been learned, which limits their effectiveness in many applications.

Contribution: In this work we propose a novel continuous dictionary learning (CDL) scheme, which allows processing data with samples that do not necessarily lie on a grid or have a fixed size. This strategy provides a novel fiber representation, which is uniform in size and approximates with high accuracy fibers of any length. In addition, we motivate the use of the suggested representation by presenting a compatible metric that allows the calculation of inter-fiber similarities. Such calculations are the basic building blocks of frequently performed tasks in tractograms analysis, such as clustering, classification, and registration.

2   Methods

Standard Dictionary Learning (SDL): The SDL strategies have been designed for data with a fixed size (e.g, image patches). Given a data point $f \in \Re^d$ that admits a sparse representation $x \in R^K$ in a given dictionary $D \in \Re^{d \times K}$ (typically $K > d$), i.e., $f = Dx$, one may recover x from f by minimizing

$$\min_x \|x\|_0 \ \ s.t. \ f = Dx. \tag{1}$$

If D is unknown, it may be learned from a given set of data points f_i by solving:

$$\min_{D,X} \|F - DX\|_F^2 \ \ s.t. \ \forall i \ \|x_i\|_0 \leqslant T_0, \tag{2}$$

where $F \in \mathbb{R}^{d \times N}$ and $X \in \mathbb{R}^{K \times N}$ contain in their columns the data points $f_i \in \mathbb{R}^d$ and their sparse representations $x_i \in \mathbb{R}^K$ respectively; T_0 is the target sparsity; and $\|\cdot\|_F$ is the Frobenius norm. While both Eqs. 1 and 2 are NP hard problems, many approximation methods have been proposed for them such as orthogonal matching pursuit (OMP) for the first and K-SVD for the second [4].

In order to use SDL such as K-SVD for data which is continuous or of varying size, one needs to sample (perhaps with the use of an interpolation) the data points with a fixed size and only then train a dictionary for the sampled versions. This kind of coding is clearly suboptimal for fibers and many other types of data in which the length of the data points varies significantly and therefore many of them might be overly or underly sampled. Thus, a continuous dictionary that may be sampled in accordance to a given data sample is much needed.

Continuous Dictionary Learning (CDL): A possible way for coding data points of varying size is to use a continuous dictionary (CD), in which each atom is a continuous function. Then, given a certain data point, the functions of the dictionary can be sampled at the same sampling locations of this point. Then any standard sparse coding strategy may be used to recover the representation of this signal. Notice that this method provides us with a representation of the same size (as the number of the functions in the CD) for all the data points.

As SDL improves the sparse coding, it is expected that the same would happen also with CDL. Yet, it is impossible to learn the columns of the CD directly as is done in the discrete case. Inspired by the double sparsity model for dictionary learning [4], we propose to learn a parametric dictionary $D = \Phi A$ such that each of its atoms is a linear combination (defined by $A \in \mathbb{R}^{K \times K}$) of the CD Φ. In this case, the number of parameters to be learned is finite but results with a CD. Given a set of N continuous functions f_i, the CDL problem becomes

$$\min_{A, x_i} \sum_{i=1}^{N} \int_{t \in \Omega_i} (f_i(t) - \Phi(t) A x_i)^2 dt \ \ s.t. \ \forall i \ \|x_i\|_0 \leqslant T_0, \tag{3}$$

where Ω_i is the domain of f_i (we assume it is included in the domain of the functions in Φ). The integration in Eq. 3 may be approximated by a finite sum

over d_i samples of f_i and the functions in Φ. Assuming a uniform sampling, this approximation results with the following minimization problem

$$\min_{A,x_i} \sum_{i=1}^{N} \sum_{t=1}^{d_i} (f_i(t) - \Phi(t)Ax_i)^2 \ s.t. \ \forall i \ \|x_i\|_0 \leqslant T_0. \tag{4}$$

For calculating A we use stochastic gradient descent with mini-batches of size b and a varying learning rate. At iteration n, given the previous approximation A_{n-1}, we re-sample the dictionary ΦA_{n-1} for all the data points of the current mini-batch in their corresponding locations and calculate their sparse representation using OMP. Then we calculate A_n by taking a gradient step based on the current batch and its representation in A_{n-1}.

Fiber Representation: In the context of fibers, we initialize A to be the identity matrix and Φ to be randomly selected fibers. We get a continuous version for each by using a cubic spline interpolation, which both promotes smoothness and preserves the locations of the fibers' sampled points. To improve Φ, we may select a smaller number of atoms for it at the beginning of the training and then every several iterations add to it as an atom the data point that has the worst representation. The matrix A is expanded accordingly (the new entries are initialized to zero except the corner, which is set to be 1 to select the new atom to the dictionary). This technique improves the training.

Sparse Inter-Fiber Similarity: We base our suggested similarity measure on the known Cosine similarity which was found to be beneficial for fibers [10]:

$$sim^{cos}(f_i, f_j) = \frac{\langle f_i, f_j \rangle}{\|f_i\| \|f_j\|} \tag{5}$$

where $\langle \cdot \rangle$ stands for inner product and $\| \cdot \|$ for L_2 norm. Cosine similarity cannot be applied to signals of different length. It also cannot be used directly on the sparse representations due to the non-orthogonal nature of the dictionary.

We propose a modified cosine similarity measure that suits the suggested representation. This is a modification of the method suggested in [2]. The non-orthogonality is treated by the addition of weighting matrix that contains the similarity measure between atoms. The approximation of the cosine similarity measure is achieved by replacing the original fibers with the reconstructed fibers:

$$sim^{cos}(f_i, f_j) \cong sim^{cos}\left(\tilde{f}_i, \tilde{f}_j\right) = \frac{\langle \Phi A x_i, \Phi A x_j \rangle}{\|\Phi A x_i\| \|\Phi A x_j\|} = \frac{x_i^T A^T \Phi^T \Phi A x_j}{\|\Phi A x_i\| \|\Phi A x_j\|}, \tag{6}$$

We strive to avoid adding unnecessary reconstruction error and to perform the similarity computation without going back to the original space. Therefore, we propose to substitute the denominator' norms by the norms of the original fibers, which can be saved before the sparse coding phase and later saved as part of the compressed data. This step would add one additional value to the representation, which does not have a significant impact on the representation length while allowing to use the original fiber norm for similarity approximation. The

modified compressed representation $\widehat{x}$ will contain the value: $norm_i = \|f_i\|$, concatenated to the sparse coefficients vector x_i. In addition, we define the similarity matrix $S = A^T \Phi^T \Phi A \in \mathbb{R}^{K \times K}$, whose entries measure the similarity between the atoms in ΦA. Based on the above, we propose a new similarity measure between fibers, the Cosine using Continuous Dictionary Similarity (CCDS), which relies only on the compressed representation and can be calculated efficiently:

$$sim^{CCDS}\left(\widehat{x_i}, \widehat{x_j}\right) = \frac{x_i^T S x_j}{\|norm_i\|\|norm_j\|} \tag{7}$$

3 Experiments and Results

Data: The proposed method was evaluated on MRI datasets of 5 healthy volunteers from the Human Connectome Project (HCP). Acquisition protocol included 3 shells (b-values of 1000, 2000, 3000), 96 unique directions for each, isotropic resolution of 1.25 mm and imaging matrix of $144 \times 168 \times 111$ pixels [1]. Diffusion data was preprocessed using the HCP diffusion pipelines [5]. WM fibers were obtained using Q-Space Diffeomorphic Reconstruction (QSDR) and streamline tractography using DSI-Studio[1] software. Tractography was terminated upon reaching 1M fibers. In all the following experiments, we used three types of sets per brain, randomly selected from the full tractogram: 20,000 fibers used as the training-set (TrS) for the dictionaries, 2,000 used as the validation-set (VS) and 10,000 other fibers served as the testing set (TS).

Sparse Coding of Fibers using Different Types of Dictionaries: For each of the 5 tractograms, we created 3 different dictionaries based on the TrS, using the CDL and the other two alternatives described in Section 2: CD and SDL. The training was performed with the following configurations for each technique:
CDL: The sparsity constraint, T_0, was set to 7 non-zero coefficients. Initial dictionary of size 500 was chosen to be the best out of 5 random initializations, evaluated on the VS. Every 10 iterations, a new atom was added to Φ, increasing it to a final size of 700 atoms. These parameters were found in our experiments to provide a reasonable compromise between error rate and compression rate. Batch size was set to $b = 500$. A Learning rate (LR) of $LR(n) = min(10^{-6}, 6 * 10^{-6}/log(n))$ was used, where n is the iteration number. The training process was stopped after $4k$ iterations, as the average reconstruction error on the VS was no longer descending.
_CD:_ The dictionary is the initial CD selected for CDL. _SDL:_ The fibers under evaluation were interpolated and sampled at 20 3D points along the fiber's path, equidistantly. This choice of m was found to be good in [11]. Instead of learning the dictionary on an entire fiber set, a representative subset of fibers is used in order to reduce the learning time. This subset was selected by randomly choosing 10000 fibers out of the full set. Such a train-set is diverse and empirically well represent the full tractogram. The K-SVD learning process was initiated with a

---

[1] developed by Fang-Cheng Yeh from the Advanced Biomedical MRI Lab, NTU Hospital, Taiwan, and made available at http://dsi-studio.labsolver.org/Download/
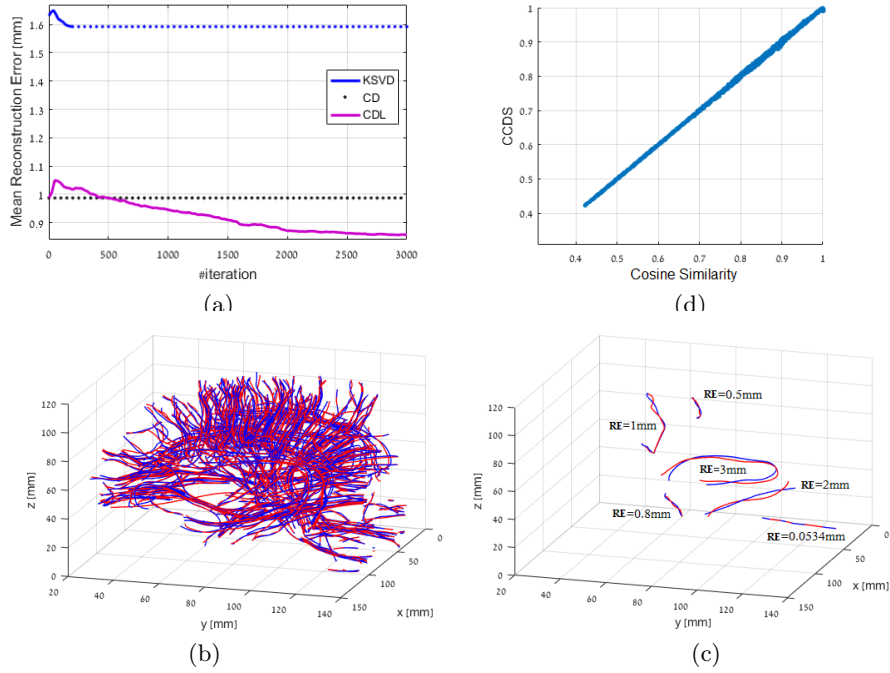
Fig. 2. (a) RE convergence of SDL (blue) and CDL (magenta) and CD fixed value (dashed-black); (b) 500 original fibers (blue) and their matching reconstructions (red); (c) sample pairs of fibers with their RE; (d) Cosine vs. CCDS similarities for one brain.

set of K fibers representing the centroids of K cluster of the TrS, computed using K-means. The dictionary size was set to 700, T_0 was set to 7 and the process was terminated after 200 K-SVD iterations, where no significant improvement was recorded in the VS reconstruction error.

Performance Evaluation: In order to evaluate the accuracy of the new representation relative to the original data from the tractography process, we need to reconstruct each fiber f_i. This is done by sampling the continuous dictionary in the same sampling locations t of the original fiber and applying $\hat{f}_i(t) = \Phi(t)Ax$. Next, the distance between the two sets of 3D points, $f_i, \hat{f}_i$ is evaluated. We assume that a good approximation for the distance between the two curves can be inferred from the distances between compatible pairs of sampling points from the two curves. We computed two measurements: average Euclidean distance over all point pairs, and maximal Euclidean distance among all point pairs. For a set of fibers, we computed mean and median values of the above two measurements, over all fibers. The aforementioned evaluation scheme was slightly modified to match the SDL framework: Using the SDL, one can reconstruct fibers with a prefixed number of samples as opposed to the original representation that had a varying number of sampling points. To solve this we interpolated the SDL reconstructed output fibers using spline interpolation. Next, the fibers were sampled equidistantly with the same number of sampling points as was in the original

tractogram. Figure 2(a) presents the reconstruction error (RE) measured as the mean distance in the VS as a function of the learning iteration of SDL and CDL. For the CD, a fixed value is plotted as no learning is performed. Figure 2(b) and (c) visualizes original fibers from TS and their reconstructions using SDL. Table 1 shows the RE of the TS for all 5 brains. Notice that even the CD without learning provides a more accurate coding than the SDL. The addition of learning further improves the representation.

Table 1. Performance evaluation: Reconstruction error in mm using the suggested method (CDL) compared to SDL and CD, tested on the TS of 5 different brains.

RE[mm]	mean of mean			median of mean			mean of max			median of max		
Method	SDL	CD	CDL	SDL	CD	CDL	SDL	CD	CDL	SDL	CD	CDL
$Brain_1$	1.650	1.068	**0.859**	1.395	0.849	**0.747**	3.335	2.731	**2.234**	2.839	2.094	**1.845**
$Brain_2$	1.575	0.989	**0.801**	1.338	0.815	**0.683**	3.170	2.595	**2.113**	2.724	2.044	**1.707**
$Brain_3$	1.709	1.084	**0.882**	1.469	0.857	**0.748**	3.425	2.858	**2.362**	2.934	2.161	**1.911**
$Brain_4$	1.662	1.012	**0.816**	1.367	0.833	**0.696**	3.400	2.689	**2.196**	2.838	2.055	**1.758**
$Brain_5$	1.596	1.011	**0.808**	1.361	0.858	**0.717**	3.217	2.669	**2.202**	2.750	2.147	**1.805**
Avg	1.638	1.033	**0.833**	1.380	0.842	**0.718**	3.309	2.708	**2.221**	2.817	2.100	**1.805**

Approximation of cosine similarity using CCDS: The proposed CCDS similarity measure was evaluated by comparing the distances between the fibers in their original representation (Eq. 5) and the distances between the same fibers as calculated in the sparse representation using CCDS (Eq. 7). The distances were evaluated on 10,000 randomly selected fibers from the TS. Figure 2(d) shows the correlation graph of cosine vs. CCDS similarities for one of the brains.

4   Discussion and Conclusions

This work proposes a novel continuous dictionary learning scheme, which allows training a dictionary on diverse training sets such that samples do not need to lie on a grid or have a fixed size. The significance of the new methodology is the ability to train continuous dictionaries. This strategy provides a novel fiber representation, which is uniform in size and approximates with high accuracy fibers of any length.

We have tested the suggested method on 5 tractograms that were created from HCP data, leading to very satisfying results. With only 7 non-zero coefficients per fiber, we achieved average reconstruction error of $0.83mm$ and max reconstruction error of $2.22mm$. Based on the average number of points per fiber in the original representation, our compression ratio is comparable with the state-of-the-art (95.6%) [7] that relies on a tractography with two times shorter step, which means initial representation two times larger. According to Table 1, CDL clearly outperformed the two alternative examined methods. Omitting the learning stage gives about 25% higher average reconstruction error. This emphasizes the importance of the learning process. Using K-SVD on a fixed size version of the data led to the least satisfying results. We believe K-SVD did not

succeed due to the varying length of the WM fibers. The experiments conducted use constrained nonzero coefficients number, $T_0 \leq 7$, and a dictionary size of 700. Additional tests (not included herein due to lack of space), suggest that if more accuracy is needed, it may be achieved with higher T_0 and a bigger dictionary, obviously with additional computational cost.

The suggested representation is advantageous over other compression schemes also due to the ability to estimate inter-fiber similarity using the proposed CCDS measure, directly in the compressed form, thus allowing common operations on fibers without decompressing the data. In addition, common metrics that can be used on signals of different length are usually computationally expensive, as opposed to CCDS, which is highly efficient. The legitimacy of CCDS was verified by achieving high correlation to the known Cosine similarity on the original representations (Fig. 2(d)). The proposed continuous dictionary learning methodology is general and can be applied to in other application with similar data characteristics.

References

1. Andersson, J.L., Sotiropoulos, S. N.: An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage 125, 1063–1078 (2015)
2. Anonymous: Sparse Representation for White Matter Fiber Compression and Calculation of Inter-Fiber Similarity. Workshop Paper In Press (2017).
3. Candes, E. J., Fernandez-Granda, C.: Towards a Mathematical Theory of Super-resolution. Comm. Pure Appl. Math., 67(6), 906–956 (2014).
4. Elad, M.: Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer (2010)
5. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J. L., Van Essen, D.C.: The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80 105–124 (2013)
6. Prkovska, V., Descoteaux, M., Poupon, C., ter Haar Romeny, B. M., Vilanova, A.: Classification Study of DTI and HARDI Anisotropy Measures for HARDI Data Simplification. New Developments in the Visualization and Processing of Tensor Fields. Springer Berlin Heidelberg 229–251 (2012).
7. Presseau, C., Jodoin, P.M., Houde, J.C., Descoteaux, M.: A new compression format for fiber tracking datasets. NeuroImage 109, 73–83 (2015)
8. Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., Deriche, R.: Unsupervised white matter fiber clustering and tract probability map generation: Applications of a Gaussian process framework for white matter fibers. NeuroImage, 51(1), 228–241 (2010).
9. Yamada, K., Sakai, K., Akazawa, K., Sachiko, Y. U. E. N., Nishimura, T.: MR tractography: a review of its clinical applications. Magnetic resonance in medical sciences, 8(4), 165–174 (2009).
10. Zimmerman-Moreno, G., Alexandroni, G., Greenspan, H.: White Matter fiber set simplification by redundancy reduction with minimum anatomical information loss. In Computational Diffusion MRI. Springer International Publishing 171–182 (2016).
11. Zvitia, O., Mayer, A., Shadmi, R., Miron, S. and Greenspan, H.K., Co-registration of white matter tractographies by adaptive-mean-shift and Gaussian mixture modeling. IEEE Transactions on Medical Imaging, 29(1), pp.132-145 (2010).