# A Sparse Tensor Decomposition with Multi-Dictionary Learning Applied to Diffusion Brain Imaging

Cesar F. Caiafa
PBS - Indiana University, (47401) Bloomington, IN, USA
IAR - CONICET/CICPBA (CC5 1894) V. Elisa, ARGENTINA
Email: ccaiafa@indiana.edu

Andrzej Cichocki
LABSP - RIKEN
(351-0198), Wako, Japan
Email: cia@brain.riken.jp

Franco Pestilli
PBS - Indiana University
(47401) Bloomington, IN, USA
Email: franpest@indiana.edu

***Abstract*—We use a multidimensional signal representation that integrates diffusion Magnetic Resonance Imaging (dMRI) and tractography (brain connections) using sparse tensor decomposition. The representation encodes brain connections (fibers) into a very-large, but sparse, core tensor and allows to predict dMRI measurements based on a dictionary of diffusion signals. We propose an algorithm to learn the constituent parts of the model from a dataset. The algorithm assumes a tractography model (support of core tensor) and iteratively minimizes the Frobenius norm of the error as a function of the dictionary atoms, the values of nonzero entries in the sparse core tensor and the fiber weights. We use a nonparametric dictionary learning (DL) approach to estimate signal atoms. Moreover, the algorithm is able to learn multiple dictionaries associated to different brain locations (voxels) allowing for mapping distinctive tissue types. We illustrate the algorithm through results obtained on a large in-vivo high-resolution dataset.**

## I. Extended Abstract

Diffusion Magnetic Resonance Imaging (dMRI) allows to estimate structural brain connections (fibers) in-vivo by measuring the diffusion of water molecules within a volume (voxels $v = 1, 2, 3, \ldots, N_v$) at different magnetic field gradient directions ($\boldsymbol{\theta} = \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{N_\theta}$). Fibers describe the putative position and orientation of the neuronal axons bundles traveling within the living human brain [1] (**Fig. 1(a)**).

Diffusion signals can be modeled as non-negative linear combinations of the signals associated to overlapped fibers [2]. Several parametric signal models have been used in the literature for fitting dMRI signals, for example, under specific conditions (e.g. small $b$-values), the water molecules diffusion process follows a Gaussian distribution which leads to the Diffusion Tensor (DT) model [1]. However, under more general conditions (e.g. large $b$-values), this simple model does no longer hold [3].

We propose a non-parametric learning algorithm to estimate fiber response functions from a given dMRI dataset, which is based on a recently proposed multidimensional signal representation that encodes diffusion dMRI and tractography (brain connections) into a sparse tensor decomposition [4]. Previous DL methods [5], [6] focussed only on the dMRI signal within individual voxels. Instead, we incorporate tractography data into the model. So whereas previous contributions learn typical diffusion patterns, including crossing fibers, with the objective to reduce the number of measurements and apply denoising, we learn atoms that correspond to single fibers with different orientations. Our approach combines what is generally referred to as microstructure (dMRI signals within individual voxels) and macrostructure (the anatomical properties described by the tractography data).

**Single-dictionary model:** We propose the following sparse Tucker model [7] for a full-brain diffusion signal (demeaned[1]) $\mathbf{Y} \in \mathbb{R}^{N_{\boldsymbol{\theta}} \times N_v}$ with $N_{\boldsymbol{\theta}}$ gradient directions and $N_v$ voxels (**Fig. 1(b)**):

$$\mathbf{Y} \approx \hat{\mathbf{Y}} = \underline{\mathbf{\Phi}} \times_1 \mathbf{D} \times_3 \mathbf{w}^T, \tag{1}$$

where $\underline{\mathbf{\Phi}} \in \mathbb{R}^{N_a \times N_v \times N_f}$ is a sparse tensor whose non-zero entries, $\underline{\mathbf{\Phi}}(a, v, f)$, indicate the orientation of fiber $f$ in voxel $v$, which is approximated by atom $a$, "$\times_i$" is the tensor-by-matrix product in mode-$i$, $\mathbf{D} \in \mathbb{R}^{N_{\boldsymbol{\theta}} \times N_a}$ is a dictionary of diffusion predictions whose columns (atoms) correspond to fiber orientations, and $\mathbf{w} \in \mathbb{R}^{N_f}$ is a vector of nonnegative fibers weights.

**Multi-dictionary model:** In our experiments, the single-dictionary model was not able to provide good approximations to the data in all voxels so, here, we propose a multi-dictionary model as follows:

$$\mathbf{Y} \approx \hat{\mathbf{Y}} = [\underline{\mathbf{\Phi}}_1 \times_1 \mathbf{D}_1, \underline{\mathbf{\Phi}}_2 \times_1 \mathbf{D}_2, \ldots, \underline{\mathbf{\Phi}}_N \times_1 \mathbf{D}_N]_{[2]} \times_3 \mathbf{w}^T, \tag{2}$$

where $[\underline{\mathbf{A}}_1, \underline{\mathbf{A}}_2, \ldots, \underline{\mathbf{A}}_N]_{[n]}$ is the mode-$n$ concatenation of tensors $\underline{\mathbf{A}}_i$ ($i = 1, 2, \ldots, N$). In other words, voxels are grouped into $N$ subsets, each one having a single-dictionary model, i.e. $\mathbf{Y}_n \approx \underline{\mathbf{\Phi}}_n \times_1 \mathbf{D}_n \times_3 \mathbf{w}^T$ with $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N]$.

**The proposed learning algorithm:** We fix the support of tensor $\underline{\mathbf{\Phi}}$ using the output of any available tractography[2] algorithm [8].

Our algorithm starts with the single dictionary model of eq. (1) and iterates between two optimization steps:

- **STEP 1:** Given a vector $\mathbf{w}$, find the dictionary $\mathbf{D}$ and the nonzero values in tensor $\underline{\mathbf{\Phi}}$ that minimize the error $\|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2$. By defining $\mathbf{B} = \underline{\mathbf{\Phi}} \times_3 \mathbf{w}^T \in \mathbb{R}^{N_a \times N_f}$, the objective function can be written as $\|\mathbf{Y} - \mathbf{D}\mathbf{B}\|_F^2$, thus the dictionary $\mathbf{D}$ can be updated as in the *codebook update stage* of the K-SVD algorithm [9] using nonnegative constraints on $\mathbf{B}$. Then, the values of nonzero entries in $\underline{\mathbf{\Phi}}$ are updated such that equation $\mathbf{B} = \underline{\mathbf{\Phi}} \times_3 \mathbf{w}^T$ holds.
- **STEP 2:** Given $\mathbf{D}$ and $\underline{\mathbf{\Phi}}$, find the optimal nonnegative vector $\mathbf{w}$ that minimizes the error $\|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2$. By defining $\underline{\mathbf{M}} = \underline{\mathbf{\Phi}} \times_1 \mathbf{D}$, the objective function can be written as $\|vec(\mathbf{Y}) - \mathbf{M}_{(3)}^T \mathbf{w}\|_F^2$, where $\mathbf{M}_{(3)}$ is the mode-3 unfolding matrix of tensor $\underline{\mathbf{M}}$. This problem can be solved by applying a nonnegative least squares (NNLS) algorithm, e.g. [10].

Once this two-steps iteration converges, we evaluate the obtained errors in each voxel and split the tensor $\underline{\mathbf{\Phi}} = [\underline{\mathbf{\Phi}}_1, \underline{\mathbf{\Phi}}_2]_{[2]}$ such that $\underline{\mathbf{\Phi}}_1$ ($\underline{\mathbf{\Phi}}_2$) contains the voxels with lowest (highest) errors. We set the threshold as the median of the error distribution so the tensor $\underline{\mathbf{\Phi}}$ is divided into two equal sized subtensors. Then, we repeat the above described optimization steps but restricting the DL (STEP 1) to each subset of voxels. We repeat this procedure until convergence is achieved and a predefined maximum number of dictionaries is reached (see Algorithm 1).

[1] We model the signal after the mean diffusion is removed, i.e. $\mathbf{Y}(\boldsymbol{\theta}, v) = S(\boldsymbol{\theta}, v) - \frac{1}{N_{\boldsymbol{\theta}}} \sum_{\boldsymbol{\theta}_i} S(\boldsymbol{\theta}_i, v)$, where $S(\boldsymbol{\theta}, v)$ is the measured diffusion signal.

[2] Tractography methods typically identify potential fibers by looking at the directions of maximal diffusivity voxel by voxel.

**Results:** Our algorithm was tested on a full-brain high-resolution dataset ($1.25mm$, $b = 2,000s/mm^2$) from the Human Connectome Project (HCP)[11]. We used fibers of twenty major human white-matter tracts [12] obtained with the probabilistic tractography method [8]. We fit the model to the diffusion data and analyzed the obtained dictionaries by grouping atoms in subsets of similar diffusion response functions (RFs) (**Fig. 2(a)**)-left. It is highlighted that a diverse set of RFs are obtained, ranging from low to high Fractional Anisotropy (FA, [1]), as displayed in light-yellow (top) and dark-blue (bottom), respectively. In **Fig. 2(a)**-right, the spatial distribution of the obtained RFs are shown for the case of one important human white-matter tract: the Corticospinal tract. In order to compare the obtained atoms against theoretical DTI-based models, in **Fig. 2(b)**, three examples of DTI diffusion RFs are shown. Finally, in **Fig. 2(c)**, the RF obtained empirically based on voxels with no crossing fibers using MRTRIX 0.2 software [8] is shown.

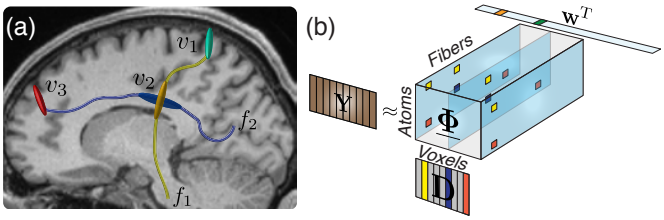


Fig. 1. **The Single-Dictionary Sparse Tucker Model for dMRI: (a)** Illustration of two white matter fibers ($f_1$ and $f_2$) and three voxels ($v_1$, $v_2$ and $v_3$). The diffusion signal at voxel $v_2$ is a nonnegative linear combination of the signals associated to $f_1$ and $f_2$. **(b)** The full-brain diffusion signal $\mathbf{Y} \in \mathbb{R}^{N_{\boldsymbol{\theta}} \times N_v}$ is decomposed as the product of a very large, but sparse, core tensor $\underline{\mathbf{\Phi}} \in \mathbb{R}^{N_a \times N_v \times N_f}$, and factors $\mathbf{D} \in \mathbb{R}^{N_{\boldsymbol{\theta}} \times N_a}$ (mode-1, dictionary matrix) and $\mathbf{w}^T \in \mathbb{R}^{1 \times N_f}$ (mode-3, vector of weights).

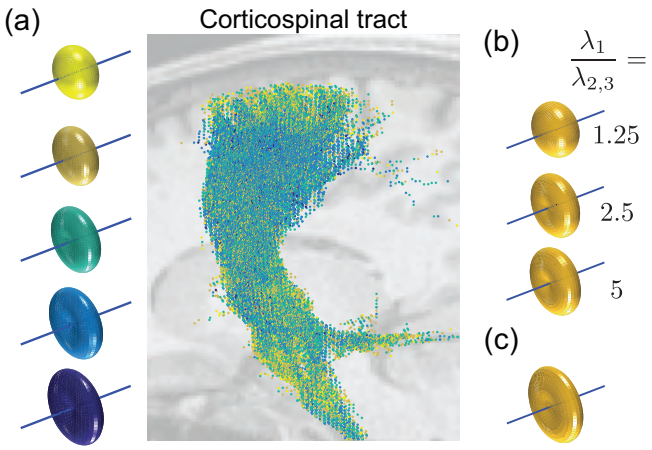


Fig. 2. **Experimental results:** *Dataset:* a full-brain dataset having $N_v = 224,675$ voxels, $N_{\boldsymbol{\theta}} = 90$ gradient directions and $N_f = 58,143$ fibers (probabilistic Tractography, $L_{max} = 10$). We set the number of dictionaries $N_{max} = 4$ and the number of atoms (orientations) $N_a = 1,981$. The obtained relative error is $\|\mathbf{Y} - \hat{\mathbf{Y}}\|_F / \|\mathbf{Y}\|_F = 0.093$. **(a)**-left: Centroids of five clusters of obtained atoms. In order to provide a high-resolution 3D visualization of atoms, we computed their Spherical Harmonics representation using 25 coefficients ($L_{max} = 4$, relative error = $0.10 \pm 0.08$). The diffusion direction is displayed as a blue line for each atom. **(b)** Visualization of the theoretical Response Function (RF) of a fiber using the classical DT model: $s(\boldsymbol{\theta}) \propto e^{-b\boldsymbol{\theta}^T diag[\lambda_1,\lambda_2,\lambda_3]\boldsymbol{\theta}}$ [1] for three different sets of parameter values $(\lambda_1, \lambda_2, \lambda_3)$. **(c)** The RF estimated by using the MRTRIX software [8] which is based on a detection of voxels with maximum Fractional Anisotropy (FA) as an indication of collinear fiber voxels.

**Algorithm 1 : Multi-Dictionary Tensor Decomposition**

**Require:** Diffusion signal $\mathbf{Y} \in \mathbb{R}^{N_{\boldsymbol{\theta}} \times N_v}$, support of $\underline{\mathbf{\Phi}} \in \mathbb{R}^{N_a \times N_v \times N_f}$, initial weights $\mathbf{w} \in \mathbb{R}^{N_f}$, dictionary $\mathbf{D} \in \mathbb{R}^{N_{\boldsymbol{\theta}} \times N_a}$, number of dictionaries $N_{max}$.
**Ensure:** Tensor decomposition with $N \leq N_{max}$ dictionaries as in eq. (2).

$N = 1$; Start with the single-dictionary model
$[\underline{\mathbf{\Phi}}_1, \mathbf{D}_1]$ = update_Phi_D($\mathbf{Y}, \underline{\mathbf{\Phi}}, \mathbf{D}, \mathbf{w}$); See Algorithm 2
**while** not convergence **do**
  **while** not convergence **do**
    $\underline{\mathbf{M}} = [\underline{\mathbf{\Phi}}_1 \times_1 \mathbf{D}_1, \underline{\mathbf{\Phi}}_2 \times_1 \mathbf{D}_2, \ldots, \underline{\mathbf{\Phi}}_N \times_1 \mathbf{D}_N]_{[2]}$;
    $\mathbf{w} = argmin_{\mathbf{w} \geq 0} \|\mathbf{Y} - \mathbf{M}_{(3)}^T \mathbf{w}\|_F^2$; optimize fiber weights
    **for** $n = 1$ **to** $N$ **do**
      $[\underline{\mathbf{\Phi}}_n, \mathbf{D}_n]$ = update_Phi_D($\mathbf{Y}, \underline{\mathbf{\Phi}}_n, \mathbf{D}_n, \mathbf{w}$); See Algorithm 2
    **end for**
  **end while**
  **if** $N \leq N_{max}$ **then**
    **for** $n = 1$ **to** $N$ **do**
      Sort and split dataset into high and low voxel errors:
      $[\mathbf{Y}_{2n-1}, \mathbf{Y}_{2n}] = \mathbf{Y}_n$; $[\underline{\mathbf{\Phi}}_{2n-1}, \underline{\mathbf{\Phi}}_{2n}] = \underline{\mathbf{\Phi}}_n$;
    **end for**
    $N = 2 \times N$; number of dictionaries is doubled
  **end if**
**end while**
**return** $\underline{\mathbf{\Phi}}_n$,$\mathbf{D}_n$,$\mathbf{w}$;

**Algorithm 2 : update_Phi_D($\mathbf{Y}, \underline{\mathbf{\Phi}}, \mathbf{D}, \mathbf{w}$)**

**Require:** Diffusion signal $\mathbf{Y} \in \mathbb{R}^{N_{\boldsymbol{\theta}} \times N_v}$, support of tensor $\underline{\mathbf{\Phi}} \in \mathbb{R}^{N_a \times N_v \times N_f}$, fiber weights $\mathbf{w} \in \mathbb{R}^{N_f}$ and dictionary $\mathbf{D} \in \mathbb{R}^{N_{\boldsymbol{\theta}} \times N_a}$.
**Ensure:** Updated $\underline{\mathbf{\Phi}}$ and $\mathbf{D}$

$\mathbf{B}_0 = \underline{\mathbf{\Phi}} \times_3 \mathbf{w}^T$;
$[\mathbf{B}, \mathbf{D}] = argmin_{\mathbf{B},\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{B}\|_F^2$ s.t. $supp(\mathbf{B}) = supp(\mathbf{B}_0)$ and $\mathbf{B} \geq 0$; e.g. using the *codebook update stage* of the K-SVD algorithm [9] with nonnegative constraints on the entries of $\mathbf{B}$.
Update nonzero entries of $\underline{\mathbf{\Phi}}$ s.t. $\mathbf{B} = \underline{\mathbf{\Phi}} \times_3 \mathbf{w}^T$ is held;
**return** $\underline{\mathbf{\Phi}}$,$\mathbf{D}$;

## REFERENCES


[1] J. Li, Y. Shi, and A. W. Toga, "Mapping Brain Anatomical Connectivity Using Diffusion Magnetic Resonance Imaging: Structural connectivity of the human brain," *IEEE Signal Processing Magazine*, vol. 33, no. 3, pp. 36–51, Apr. 2016.

[2] F. Pestilli, J. D. Yeatman, A. Rokem, K. N. Kay, and B. A. Wandell, "Evaluation and statistical inference for human connectomes," *Nat Meth*, vol. 11, no. 10, pp. 1058–1063, Sep. 2014.

[3] E. Özarslan, C. G. Koay, T. M. Shepherd, M. E. Komlosh, M. O. Irfanoglu, C. Pierpaoli, and P. J. Basser, "Mean apparent propagator (MAP) MRI: A novel diffusion imaging method for mapping tissue microstructure," *Human Brain Mapping Journal*, vol. 78, no. C, pp. 16–32, Sep. 2013.

[4] C. F. Caiafa and F. Pestilli, "Multidimensional encoding of brain connectomes," *bioRxiv*, p. 107607, Feb. 2017.

[5] S. Merlet, E. Caruyer, A. Ghosh, and R. Deriche, "A computational diffusion MRI and parametric dictionary learning framework for modeling the diffusion signal and its features," *Medical Image Analysis*, vol. 17, no. 7, pp. 830–843, Oct. 2013.

[6] A. Gramfort, C. Poupon, and M. Descoteaux, "Denoising and fast diffusion imaging with physically constrained sparse dictionary learning," *Medical Image Analysis*, vol. 18, no. 1, pp. 36–49, Jan. 2014.

[7] C. F. Caiafa and A. Cichocki, "Computing Sparse representations of multidimensional signals using Kronecker bases." *Neural Computation*, pp. 186–220, Dec. 2012.

[8] J.-D. Tournier, F. Calamante, and A. Connelly, "MRtrix: Diffusion tractography in crossing fiber regions," *Int. J. Imaging Syst. Technol.*, vol. 22, no. 1, pp. 53–66, Feb. 2012.

[9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.

[10] D. Kim, S. Sra, and I. S. Dhillon, "A non-monotonic method for large-scale non-negative least squares," *Optimization Methods and Software*, vol. 28, no. 5, pp. 1012–1039, Oct. 2013.

[11] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, K. Ugurbil, and f. t. W.-M. H. Consortium, "The WU-Minn Human Connectome Project: An overview," *NeuroImage*, vol. 80, no. C, pp. 62–79, Oct. 2013.

[12] J. D. Yeatman, R. F. Dougherty, N. J. Myall, B. A. Wandell, and H. M. Feldman, "Tract profiles of white matter properties: automating fiber-tract quantification," *PLoS ONE*, vol. 7, no. 11, p. e49790, Nov. 2012.