Fast Non-rigid Whole Brain Streamline Matching based on Fast Fiber k-NN and Group-wise MDS

Junyan Wang (wonjy8573@hotmail.com)
Zhejiang Lab

Research Article

Keywords: DWI, tractography, fast fiber k-NN algorighm, white matter topography analysis, whole-brain fiber matching

Posted Date: April 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1505524/v1

Fast Non-rigid Whole Brain Streamline Matching based on Fast Fiber k-NN and Group-wise MDS

Junyan Wang[1*]

[1*]Research Center for Healthcare Data Sciences,Interdisciplinary Innovation Research Centers, Zhejiang Lab, Kechuang Avenue, Zhongtai Sub-District, Yuhang District, Hangzhou City, 311121, Zhejiang Province, P.R. China.

Corresponding author(s). E-mail(s): wonjy8573@hotmail.com;

Abstract

Non-rigid whole-brain fiber streamline matching is a highly challenging technical problem in brain white matter analysis, largely due to the lack of reasonable vector-space representation and efficacious streamline distance measurement in the original curve space is computationally intractable. To address the issue of fast and accurate streamline distance measurement, we propose a *fast fiber k-NN algorithm* based on a strong relationship between point-wise K-NN and fiber-wise k-NN. Additionally, we propose a distance-preserving vector-space representation of the fiber streamlines based on a novel group-wise multi-dimensional scaling (MDS) technique. Based on the fast fiber k-NN and group-wise MDS, we propose a novel computationally tractable framework for non-rigid whole-brain white matter fiber matching. In this framework, our fast fiber k-NN algorithm is used to initialize the point-cloud matching in the vector space of the streamlines computed by the group-wise MDS. In our experiments, we show that our fast fiber k-NN algorithm reasonably approximates the exhaustive fiber k-NN search at a significantly reduced computational cost, the group-wise MDS effectively computes the vector representation in the aligned vector space, and our fast fiber matching method achieves higher accuracy for large displacement non-rigid whole-brain streamline matching compared to conventional image registration based streamline matching.

Keywords: DWI, tractography, fast fiber k-NN algorighm, white matter topography analysis, whole-brain fiber matching

1 Introduction

Diffusion-weighted magnetic resonance imaging (dMRI) based tractography is currently the only method for *in vivo* and non-invasive measurement of structural connectivity of the brain. In recent years, we have witnessed remarkable progress made to improve the quality of tractography [16, 17, 21, 22, 26]. In this work, we are focused on analyzing the fiber tracts, a.k.a. tractograms, generated by tractography.

1.1 Non-rigid whole brain white matter fiber matching

Identifying corresponding fibers across different fiber bundles [3, 10, 20, 25, 33, 35] is a known critical step in many group-wise white matter analysis tasks, such as group-wise tract segmentation, group-wise tractometry and tract atlas reconstruction. Therefore, it is useful as a standalone tool for, such as, the fiber-based longitudinal and cross-sectional studies. However, due to the well-known challenges, e.g. large variability of the fiber curves, ambiguity in fiber comparison as well as the computational intractability, the research on non-rigid whole-brain white matter fiber matching is still limited to date. Generally, only linear, i.e rigid or affine, transformations are considered to model the variability of the fiber bundles for direct fiber-wise registration [9, 10, 19, 35]. More recently, the fiber matching problem has been cast as a graph matching problem [20]. Although this idea is mathematically sound, this work deals with a relatively small set of whole brain fibers in the matching, and it cannot be applied to large-scale whole-brain structural connectivity analysis in which tens of thousands of fibers or even more are often needed. This is perhaps due to that graph matching is computationally demanding and defining the matching cost requires exhaustive fiber distance computation. Alternatively, a nonrigid volumetric registration may be performed and the fiber-wise matching can be done subsequently. A drawback of this method is that it is generally assumed that no topological changes occur in the images to be registered, such that a voxel in one image can always be matched to a voxel in the other image. Since the topology of a fiber tract varies considerably across different subjects, it is arguable that the volumetric registration is applicable to white matter fiber matching. Besides, upon completion of the fiber warping, exhaustive fiber-wise distance computation is still needed to establish the fiber-wise correspondences between the target and warped fiber bundles.

We consider the difficulty in nonrigid whole-brain streamline matching is largely due to the lack of reasonable vector-space representation for which streamline distance/similarity computation is straightforward, and computing streamline distance/similarity in the original curve space is often intractable.

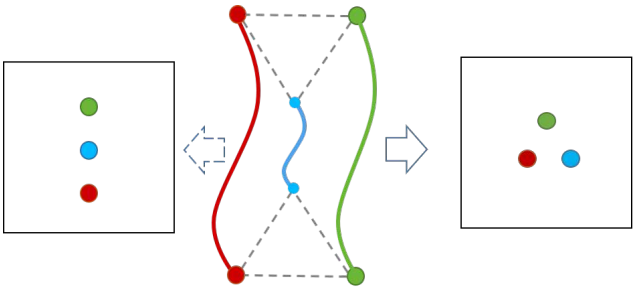


Fig. 1: Inaccurate fiber distances due to fiber re-sampling. The end points on the three fiber curves are the re-sampled points. The dashed lines represent the point-wise distances between the re-sampled points on the different fiber curves. The green, blue and red dots in the two other boxes are virtual representations of the fibers in the same colors.

1.2 Fiber k-NN computation in white matter analysis

Since it is often intractable to compute the k-NN fiber distances for massive fiber bundles across multiple subjects, alternatively, fibers of different individual subjects are usually clustered and representative fibers of each cluster are selected, and the cross-subject analysis is carried out with those representative fibers. A typical choice of the representative fiber is the cluster centroid. Since the fiber clusters and their centroids are coarse representation of fiber bundles, the specificity and accuracy of the resultant analysis are limited.

A common practice to compute fiber k-NN at affordable computational cost is to re-sample the fibers of varied lengths and topologies to the same relatively small number of points [8, 10, 18], such that the fiber distance computation can be significantly reduced. The rationale of this method lies in the fact that the re-sampled fibers represent the original fiber to certain degree. A popular technique is the minimum average direct-flip (MDF) distance used in QuickBundle tractogram simplification [8]. However, this method can be inaccurate. As shown in Figure 1, we illustrate one simple example of inaccurate distance quantification due to re-sampling of fibers of unequaled lengths to same number of points. This example shows that the apparent representations of the fibers, visualized as dots in a two dimensional virtual plane, form an equilateral triangle if only two extreme points are used to quantifying their mutual distances, despite the fact that the fibers are better represented by the dots lying on a virtual straight line based on their closeness.

Besides the problem of streamline matching, k-nearest-neighbor (k-NN) fiber distance computation is usually an inevitable step in other problems of large-scale white matter fiber analysis, including tractogram segmentation[18, 19, 23], white matter atlas generation [34], and tractometry [25].

We propose a *fast fiber k-NN algorithm* that approximates the original exhaustive algorithm without resampling the fibers to equal number of points,

which is the first attempt toward efficent approximation of fiber k-NN search to the best of our knowledge.

1.3 Vector space representation of tractogram streamlines

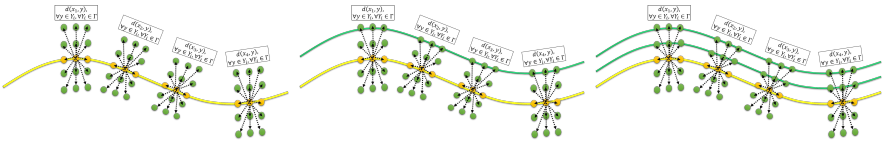
Vector space representation of tractograms based on spectral embedding of the pairwise streamline affinity matrix has been proposed for tractogram segmentation literature [18]. The resultant embedding vector space represents a sparse pairwise fiber local affinity. Motivated by the study of brain topography [2, 13, 15, 30–32], which aims to understand the functional and structural organization of the brain, we propose to represent the fiber streamlines in a vector space while preserving the entire geometric construct unaltered. Thus, the topographic organization of the original streamline bundle can be perfectly preserved in the vector space. Accordingly, we propose to adopt the multidimensional scaling formulation, which is known to be distance-preserving. This idea has been proposed in the tractogram colorization literature for individual tractograms [14]. We additionally observe that for different sample sets, the solution of MDS can be misaligned due to the non-uniqueness of the matrix factorization. To address this problem, we propose a group-wise MDS that produces aligned vector representations for different whole brain tractograms.

1.4 Method overview

i) *Our fast fiber k-NN algorithm* is motivated by the observation that the fiber-wise k-NN distance is strongly related to point-wise K-NN distance. In the remainder of this paper, we use point-wise K-NN to denote point-wise K-nearest-neighbor and fiber wise k-NN to denote fiber-wise k-nearest-neighbor. We used upper and lower case k in the two nearest neighbor systems to emphasize that the sizes of the neighborhoods are generally different. In this work, we first establish the theoretical relationship between the point-wise K-NN and fiber-wise k-NN in the context of fiber distance computation. Accordingly, we propose a *fiber k-NN likeliness measure* which can be computed efficiently with fast point-wise K-NN algorithms, such as the celebrated k-d tree algorithm [4]. The main idea here is to use the fiber k-NN likeliness measure to approximately compute the fiber-wise k-NN.

ii) *Our group-wise MDS* is based on the notion of isometric embedding of fiber curves using multidimensional scaling (MDS) [6], and it maps each curve to a vector in a finite dimensional vector space. Since the mapping is isomorphic and nearly-isometric, we consider that the resultant finite dimensional embedding space and the fiber curve space are equivalent in their geometrical construct, which is an ideal property for studying brain organization with streamline-based white matter analysis. We call the resultant vector space representation of fiber streamlines as topographic vectors.

iii) *Our non-rigid whole-brain streamline matching algorithm* is an application of the proposed fast fiber k-NN algorithm and white matter topography



(a) (b) (c)

Fig. 2: Illustration of the fast fiber k-NN algorithm. (a) Point-wise K-NN is computed for all points on all fibers. (b) We can identify the fibers having points belonging to the point-wise K-NN of a fiber. (c) We can repeat step (b) and count the number of the K-NN points on each fiber.

measure. Our main idea is that with our topographic vectors, the fiber matching problem can be simplified to a point-set matching problem. A technical problem raised here is that the solution to the MDS based fiber embedding is non-unique and brain is highly symmetrical, which causes large ambiguity in the matching. In this work, we further remove the ambiguity by initializing the matching across different whole-brain fibers using our fast fiber k-NN algorithm.

The preliminary work was presented in [28, 29]. In this work, we propose a new framework for non-rigid whole-brain fiber streamline matching based on the fast fiber k-NN and MDS based fiber streamline representation proposed in [28, 29].

1.5 Organization

The rest of this paper is organized as follows. In Section 2, we present the theory and details of the fast fiber k-NN algorithm. In Section 3, we present the workflow for non-rigid whole-brain fiber matching. In Section 4, we present the experimental results for the fast fiber k-NN algorithm, white matter topography analysis and the non-rigid whole-brain fiber matching. Finally, we conclude our work with discussions in Section 5.

2 Fast approximate k-nearest-neighbor fiber distance

Generally speaking, we will need to compute all pairwise distances between fibers in order to compute the k-NN fibers. Thus, we can obtain the pairwise fiber distances by summing over all points on each fiber. The whole process is computationally demanding, and it does not scale well. Our idea is that instead of exhaustively searching for the neighboring fibers based on the fiber distances, we propose to identify the neighboring fibers based on a novel easy-to-compute measure of likeliness of fiber neighborhood proposed in this work.

2.1 Theoretical relationship between the K-NN point distance and the k-NN fiber distance

In what follows, we treat a fiber as a discrete set of points, and we consider each fiber bundle, namely the set of fibers, as a set of multiple point sets each of which corresponds an individual fiber. Our main idea is to efficiently approximate the k-NN fiber distance computation, and we establish its rationale below.

Theorem 1 *Suppose we are given a collection of point sets, denoted as* Γ*, and two point sets* X, Y *belonging to* Γ*, i.e.* $X, Y \in \Gamma$*, and suppose* $N_\Gamma^K(y)$ *is the set of the K-NN points of a point* $y \in Y$ *within* Γ *and the point-wise K-NN is defined by point-wise distance* $d(\cdot,\cdot)$*, additionally if* $N_\Gamma^K(y) \cap X \neq O$*, then*

$$d_X(y) = \min_{x \in X} d(x,y) = \min_{x \in N_\Gamma^K(y) \cap X} d(x,y) \tag{1}$$

This theorem suggests that we can use pointwise K-NN distances to approximate the fiber-wise k-NN. The proof is deferred to the appendix.

A common fiber distance of interest is the following:

$$d_H(X,Y) = \max_{y \in Y} \min_{x \in X} d(x,y) = \max_{y \in Y} \min_{x \in N_\Gamma^K(y) \cap X} d(x,y) \tag{2}$$

which is known as the one-sided Hausdorff distance. The *min* distance is crucial to establish the point-wise correspondences when comparing unordered point sets. Various outer measures, such as *max*, *mean*, *std*, can also be used to define the fiber-wise distance.

Moreover, $\forall y \in Y,\ N_\Gamma^K(y) \cap X \neq O$ means that X and Y are *close-by*. This can be stated formally as

Proposition 2 *Suppose* $\forall y \in Y,\ N_\Gamma^K(y) \cap X \neq O$

$$d_H(X,Y) \leq \max_{y \in Y} R(N_\Gamma^K(y)) \tag{3}$$

where $R(N_\Gamma^K(y)) = \max_{z \in N_\Gamma^K(y)} d(z,y)$ *may be called the radius of* $N_\Gamma^K(y)$*.*

Proposition 2 can be proven by upper-bounding Eq. (2) based on the definition of $N_\Gamma^K(y)$.

2.2 Fast fiber k-NN approximation

In this section, we present our method for fast approximating the fiber k-NN. Our idea is that we can naturally relate the size of $N_\Gamma^K(y) \cap X$ to the closeness of fibers using an easy-to-compute *fiber k-NN-likeness measure* based on the

size of $N_\Gamma^K(y) \cap X$ defined below:

$$\delta^K(X\|Y) = \frac{1}{\|Y\|} \sum_{y \in Y} \|N_\Gamma^K(y) \cap X\| \tag{4}$$

where $\|\cdot\|$ is the size of a set. X with larger $\delta^K(X\|Y)$ is considered closer to Y, which gives rise to our definition of the Approximate fiber k-NN:

Definition 1 (Approximate fiber k-NN) We define the approximate fiber k-NN in Γ for fiber Y as the set $\mathfrak{N}_\Gamma^k(Y) = \{X^1, X^2, \cdots, X^k\}$, if $\exists K$ such that $\delta^K(X\|Y) \geq \delta^K(X'\|Y)$ for $\forall X \in \mathfrak{N}_\Gamma^K(Y)$ and $\forall X' \notin \mathfrak{N}_\Gamma^K(Y)$.

Note that the k in the fiber k-NN is usually much smaller than the K value in the point-wise K-NN.

In the following, we show that Eq. (4) can be computed efficiently. First, we show that the intersection $N_\Gamma^K(y) \cap X$ is easy to compute with a *point-to-tract mapping* defined below.

Definition 2 (Point-to-tract mapping) For fiber bundle $T = \{t^1, t^2, \cdots, t^N\}$, where each fiber is defined by a point set as $t^i = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \cdots, \mathbf{x}_{m_i}^i\}$, if we denote the set of all points on all fibers as $P = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$, the injection mapping $PT : i \in \{\forall i' \| \mathbf{x}_{i'} \in P\} \mapsto j \in \{\forall j' \| t_{j'} \in T\}$ is called a **point-to-tract mapping**.

This definition is consistent with the benchmark file formats of fiber bundles, such as the `trk` and `tck` files[1], in which the fiber bundles are stored as point sets, and this point-to-tract mapping is inherently provided with the tract files.

Based on the above definition, we can see that it is straightforward to compute $\|N_\Gamma^K(y) \cap X\|$ in Eq. (4), since

$$\|N_\Gamma^K(y) \cap X\| = m_{PT(N_\Gamma^K(y))}(i(X)) \tag{5}$$

where $m_A(x)$ is the multiplicity of x in the multiset[2] A, and $i(X)$ is the index of X.

Furthermore, we show that it is also convenient to identify the fibers giving non-zero $\delta^K(X\|Y)$ via the following proposition.

Proposition 3 (Indicator of tract neighbors) *Suppose $N_\Gamma^K(y)$ is a set of point indices and $N_\Gamma^K(Y)$ denotes the union of the indices of tracts containing points belonging to*

[1] trk is the file format of the tract files for TrackVis (http://www.trackvis.org/) and tck is the file format of the tract files for MRtrix http://www.mrtrix.org/.

[2] A multiset is a modification of the concept of a set that, unlike a set, allows for multiple instances for each of its elements https://en.wikipedia.org/wiki/Multiset.

Algorithm 1 Fast fiber k-NN algorithm.

Require: **Tractogram 1** $T_1 = \{t_1^1, t_1^2, t_1^3 ..., t_1^{N_1}\}$, **Tractogram 2** $T_2 = \{t_2^1, t_2^2, t_2^3 ..., t_2^{N_2}\}$, **Point to Tract mapping** PT_2 **for** T_2, **Point Neighborhood size**, K, **Fiber Neighborhood Size**, k ($k \ll K$)
Ensure: **k-NN fiber distance**, $d^k(T_1, T_2)$

while $1 \leq i \leq N_1 1$ **do**
  $t_1 \leftarrow T_1(i)$;
  $D_i^{\|t_1\| \times K}, ID_i^{\|t_1\| \times K} \leftarrow$ `knnsearch` $(T_2, t_1, k = K)$; % point-wise K-NN producing K-NN distances D_i and the neighbor set ID_i;
  $N_{T_2}^K(t_1) \leftarrow PT_2(ID_i)$;
  while $j \in \mathcal{N}_{T_2}^K(t_1)$ **do**
    $\delta^K(j \| t_1) \leftarrow$ Eq. (4);
  end while State $\mathfrak{N}_{T_2}^k(t_1) \leftarrow$ `sort`$(\delta^K(j \| t_1),$ `top` $k)$;
  while $t_2 \in \mathfrak{N}_{T_2}^k(t_1)$ **do**
    while $1 \leq j \leq \|t_1\|$ **do**
      $N_{T_2}^K(t_1) \cap t_2 \leftarrow$ `find`$(PT_2(ID_i(j,:)) == t_2)$;
      $\hat{d}(j) \leftarrow \min(D_i(j, N_{T_2}^K(t_1) \cap t_2)$;% Eq. (1)
    end while
    $d^k(i, t_2) \leftarrow \max_{1 \leq j \leq \|t_1\|} \hat{d}(j)$; %max is used as a matter of choice
  end while
end while

the K-NN of certain $y \in Y$*, i.e.* $N_\Gamma^K(Y) = \bigcup_{y \in Y} PT(N_\Gamma^K(y))$*. We have*

$$\delta^K(X \| Y) > 0 \text{ iff } X \in N_\Gamma^K(Y) \tag{6}$$

To prove this proposition we only need to apply definition of $\delta^K(X\|Y)$, definition of $N_\Gamma^K(y)$ and Definition 2. This proposition shows that we can easily identify the union set $N_\Gamma^K(Y)$ and compute Eq. (4) for this set only.

Our fast fiber k-NN algorithm is presented in Algorithm 1. The idea of the fast fiber k-NN algorithm is illustrated in Figure 2.

3 Non-rigid whole-brain fiber matching

Non-rigid whole-brain fiber matching is a critical emerging problem in white matter fiber analysis. In this work, we propose to apply the fast fiber k-NN algorithm to deal with this challenging problem. The main idea is that, instead of matching the fiber curves directly, we propose to match the *feature vector* of each fiber, which we call whole-brain topographic vectors.

3.1 White matter topography characterized by isometric Euclidean embedding

Measuring white matter topography is by itself an important problem in brain research. We propose to measure fiber-wise white matter topography based on the notion of isometric embedding, and specifically we adopt multidimensional scaling (MDS) [6]. This embedding space faithfully represents the Euclidean metric geometry of the original data.

3.1.1 Basic formulations of MDS

The cMDS embedding can be obtained as follows [6]:

$$\begin{array}{ll} (a) & \mathbf{B} = P \cdot \left[-0.5\mathbf{D}^{(2)}\right] \cdot P \\ (b) & \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T = \mathtt{svd}(\mathbf{B}) \\ (c) & \mathbf{Z}_p = \mathbf{E}_p \cdot \mathbf{\Lambda}_p^{0.5} \end{array} \tag{7}$$

where `svd` is the singular value decomposition, $\mathbf{D}^{(2)}$ is the input squared distance matrix, P is known as the centering matrix defined as $P_{ij} = 1 - 1/n, \forall i = j$ and $P_{ij} = -1/n, \forall i \neq j$ where n is the number of samples, $\mathbf{Z}_p$ and $\mathbf{E}_p$ are the first p columns of $\mathbf{Z}$ and $\mathbf{E}$, and $\mathbf{\Lambda}_p$ is the top-left $p \times p$ block matrix of $\mathbf{\Lambda}$ where p is the embedding dimension. We use all the eigenvectors with positive eigenvalues for embedding.

3.1.2 The geometry of fibers in the embedding space

The MDS formulation implies that we can see more contrast on the values of embedding vectors for fibers geometrically more apart, rendering the values intuitive measures of the fiber topography. In addition, the significance of each dimension of the embedding vectors in terms of its contribution to forming the target metric geometry is naturally characterized by the corresponding eigenvalues:

Theorem 4

$$\|\mathbf{B} - \mathbf{z}_i \mathbf{z}_i^T\|_F = \sqrt{\sum_{j \neq i} \lambda_j^2} \tag{8}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ...)$ *are the eigenvalues corresponding to the eigenvectors* $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, ...]$.

The proof is straightforward, we defer it to the appendix.

In conclusion, the embedding space preserves the metric-distance geometry of the original fibers. Furthermore, each dimension of the vectors coincide with the classic yet vague notion of topographic gradients in the neuroscience literature, and, here, we call them the *topographic vectors*.

3.2 Topographic vector alignment by Group-wise MDS

We consider matching the whole-brain topographic vectors as a problem of *point set registration*. Compared with direct fiber matching or registration, the computational complexity of point set registration is significantly smaller. However, due to the brain symmetry and non-uniqueness of the MDS, the topographic vectors of different fiber bundles may be globally misaligned. Fortunately, by leveraging the inter-set fiber distances computed by our proposed fast fiber k-NN algorithm, we can remove the global misalignment. We call this technique the group-wise MDS.

We observe that the global misalignment of topographic vectors is caused by the non-uniqueness of the MDS solution. Mathematically, we can express this as $\mathbf{B} = \mathbf{Z}\mathbf{Z}^T = \mathbf{Z}\mathbf{R}\mathbf{R}^T\mathbf{Z}^T$, where $\mathbf{R}^{p\times p}$ is an arbitrary orthogonal matrix. Therefore, we propose to align the topographic vectors by finding the optimal orthogonal transformation $\mathbf{R}$.

Without loss of generality, here we consider the case for two subjects. Let the two fiber bundles be denoted as T_1 and T_2. Their topographic vectors are denoted as $\mathbf{Z}_1^{N^1\times p}$ and $\mathbf{Z}_2^{N^2\times p}$. We propose to use the fiber correspondences based on the k-NN fiber distance to calibrate the pose of the topographic vectors. Suppose we consider T_1 as a fixed reference, we can find the subset of $\mathbf{Z}_2$ corresponding to $\mathbf{Z}_1$ based on the fiber-wise k-NN. The subset is denoted as $\mathbf{Z}_{21}^{N^1\times p}$. From this, our formulation for *global topographic vector alignment* can be written as follows:

$$\mathbf{Z}_{21}^* = \mathbf{Z}_{21}\mathbf{R}^*,\ \mathbf{R}^* = \underset{\mathbf{R}\mathbf{R}^T=\mathbf{I}}{\arg\min} \|\mathbf{Z}_{21}\mathbf{R} - \mathbf{Z}_1\|_F^2 \tag{9}$$

When the orthogonality constraint is removed, this problem can be solved by normal equation. To solve this problem with the constraint we first rearrange the objective function as

$$\begin{aligned} \|\mathbf{Z}_{21}\mathbf{R} - \mathbf{Z}_1\|_F^2 &= \|\mathbf{U}_{21}\mathbf{\Lambda}_{21}\mathbf{V}_{21}^T\mathbf{R} - \mathbf{Z}_1\|_F^2 \\ &= \|\mathbf{R} - \mathbf{V}_{21}\mathbf{\Lambda}_{21}^{-1}\mathbf{U}_{21}\mathbf{Z}_1\|_F^2 \end{aligned} \tag{10}$$

which is in the typical form of orthogonality-constrained least-square problem and it admits a closed-form solution [11]:

$$\mathbf{Z}_{21}^* = \mathbf{Z}_{21}\mathbf{R}^*,\ \mathbf{R}^* = \mathbf{U}^*\mathbf{V}^{*T} \tag{11}$$

where $[\mathbf{U}^*, \mathbf{D}^*, \mathbf{V}^*] = \texttt{svd}(\mathbf{V}_{21}\mathbf{\Lambda}_{21}^{-1}\mathbf{U}_{21}\mathbf{Z}_1)$, and `svd` is singular value decomposition operation. Note that since the optimal transformation is orthogonal, this alignment preserves the within-tractogram pairwise distances.

After the above global alignment, we propose to remove the local misalignment by iteratively registering the high dimensional topographic vectors across different sets using high-dimensional iterative closest point algorithm [5]. Our

Algorithm 2 Non-rigid whole-brain fiber matching.

Require: **Reference topographic vectors** $\mathbf{Z}_0 = \{\mathbf{z}_0^1, \mathbf{z}_0^2, \mathbf{z}_0^3 ..., \mathbf{z}_0^{N_1}\}$,
Moving topographic vectors $\mathbf{Z}_1 = \{\mathbf{z}_1^1, \mathbf{z}_1^2, \mathbf{z}_1^3 ..., \mathbf{z}_1^{N_1}\}$,
Indices of 1-NN fibers in T_1 **to** T_0 $\texttt{idx}_{1o}^1$
Ensure: **Matching indices of** T_1 **to** T_0, $\mathcal{M}_{10}$
$D_i \leftarrow \texttt{fullpairwisedistance}(T_i)$, $i = 0, 1$;
$\mathbf{Z}_i \leftarrow \texttt{MDS}(D_i)$, $i = 0, 1$;
$\mathbf{Z}_{1o} \leftarrow \mathbf{Z}_1(\texttt{idx}_{1o}^1)$; % a subset of Z_1
$T_{1o}^* \leftarrow \texttt{icp}(\mathbf{Z}_{1o}, \mathbf{Z}_o)$; % high-dimenstional iterative-closest-point registration
$\mathbf{Z}_1^* \leftarrow T_{1o}^*(\mathbf{Z}_1)$; % transforming the original point set to the reference set
$\mathcal{M}_{10} \leftarrow \texttt{knnsearch}(\mathbf{Z}_0, \mathbf{Z}_1, k = 1)$

algorithm is summarized in Algorithm 2. The source code of this work will be released soon.

4 Experimental results

4.1 Materials

In our experiment, we use the minimally preprocessed DWI data from the Human Connectome Project (HCP) Young Adult dataset [12].

We generate the fiber orientation density (FOD) maps for all the DWI data using the method proposed in [27]. The order of spherical harmonics is set to 12. Afterwards, we generate a whole brain tractogram of 10,000 fibers for each subject by using the anatomically constrained tractography implemented in MRtrix3 with default parameter setting [24, 26].

4.2 Performance of the fast fiber k-NN algorithm

In this subsection, we compare the accuracy and computational efficiency of our fast fiber k-NN distance computation against exhaustive pairwise distance computation. The latter also produces the ground-truth distances. We only evaluate the accuracy based on the one-sided Hausdorff distances defined by Eq. (2) for simplicity. The point-wise neighborhood sizes in our method are set to $K = 100, 200, ..., 1000$. The neighborhood sizes for the exhaustive fiber distance computation are $k_2 = 10, 20, \cdots, 100$, and the tract-wise neighborhood size for our method is $k_1 = 10$. We first evaluate our method in terms of the number of overlapped fiber sets in the tract neighborhood from both methods. The histograms of fiber k-NN overlap for $K = 10$ to 100 are shown in Figure 3. The results show that our method gives high neighborhood overlap ratio with the ground-truth fiber k-NN when $k_2 = 10$ with a large K. Particularly,

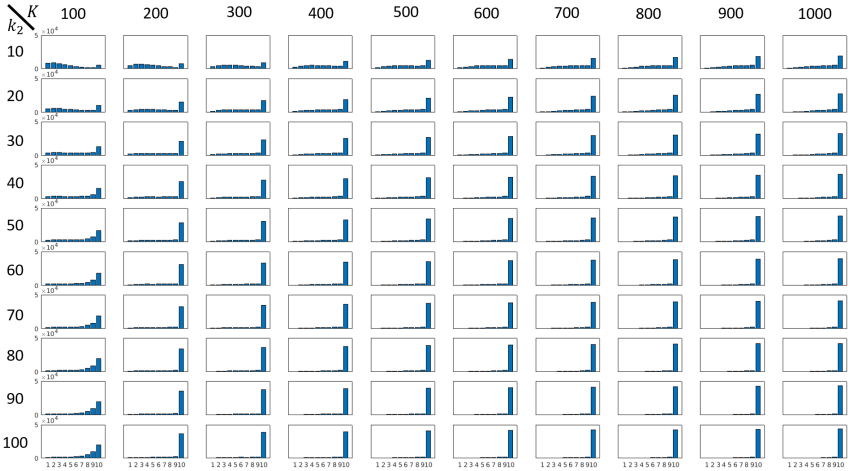


Fig. 3: Histograms of neighborhood overlap between our method and ground truth. The x axis of the histograms, i.e. the size of the neighborhood overlap, ranges from 1 to 10.

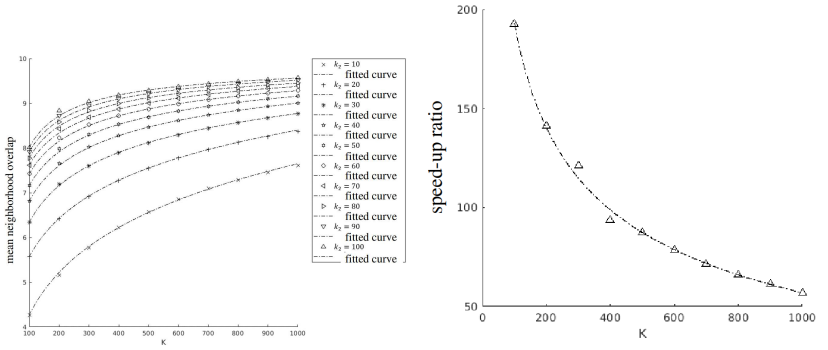
when $K = 1000$, our method gives almost perfect overlap with the ground-truth fiber k-NN, meaning that our fast fiber 10-NN almost surely lies in those ground-truth fiber 20-NN sets.

In addition, we study how the computational cost and the accuracy of our method in different parameter settings are related. We calculate the mean neighborhood size overlap (MNSO) for different K values, and we also calculate the speed-up ratio of our method, denoted by r, which is defined as the computation time of exhaustive fiber distance computation divided by the computation time of our method. We present the results in Figure 4 (a). To further quantify the performance of the fiber k-NN approximation, we fit the scatter points of MNSO and speed-up ratio r to two curves against k_2 and K using least-absolute-residuals (LAR) robust fitting, as shown in Figure 4 (b). The statistics are summarized in Table 1.

For fixed parameters K=500 and $k_1 = k_2$=20, the average time for computing the baseline exhaustive fiber distance is 3768.04 secs in MATLAB on a server PC with Intel(R) Xeon(R) Gold 6230N CPU of 32 threads and 64GB ram, while our method requires an average of 730.06 secs to achieve a satisfactory approximation.

Table 1: Statistics of performance curves.

Curves	k_2	RMSE	R-square
MNSO(K)	10	0.0264	0.9995
MNSO(K)	20	0.00928	0.9999
MNSO(K)	30	0.0108	0.9998
r	-	2.054	0.9982



(a) (b)

Fig. 4: Performance curves of the fast fiber k-NN algorithm. (a) is the neighborhood overlap of our method. (b) is the computational speed-up ratio of our method against the exhaustive pairwise fiber distance computation. The dashed curves are LAR robust fitting of the scatter points.

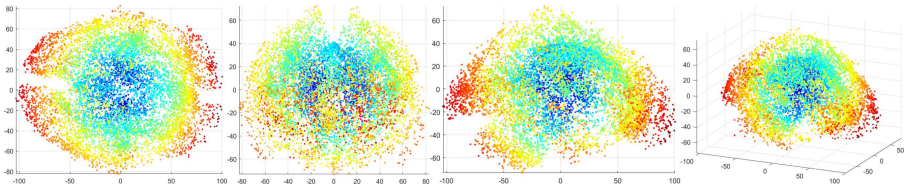


Fig. 5: Top-3 whole-brain white matter topographic vectors of subject 100206. The colors of the points correspond to the norm of the vectors.

4.2.1 Performance of group-wise MDS

For the individual analysis using MDS, we require the full pairwise distance matrix for each tractogram. Note that the tract "distance" $d_X()$ defined by Eq. (2) is asymmetric and it is not a metric distance. In this work, we symmetrize the fiber distance measure in the following form: $d_{min}(X,Y) = \min(d_H(X,Y), d_H(Y,X))$, by which short fibers are considered to be part of the close-by long fibers. Note that our fast fiber k-NN algorithm is readily applicable to distance measures based on the one-sided Hausdorff distance and it can be generalized to many other measures based on the form of $d_X(y)$ defined in Eq. (1). We chose the maximum embedding dimensions to be those with positive eigenvalues. The dimensions of the topographic vectors are arranged in descending order by the magnitude of the eigenvalues. The linear correlation between the ground-truth distance and the embedding distance is 0.98 ± 0.0032, meaning that the embedding is nearly isometric and stable.

To evaluate the validity of the group-wise MDS, we propose to compare the ground-truth aligned topographic vectors with the aligned topographic

vectors by our method. Both alignments are found by solving the optimal R in the group-wise MDS formulation. We consider the solution found by a ground-truth pairwise distance to be the ground-truth alignment. On the other hand, the solution found by our approximated fast fiber k-NN distance is our solution. Besides, we propose to evaluate the reproducibility of our method by comparing our aligned topographic vectors with the reference topographic vectors directly. The fidelity of the ground-truth alignment is established by comparing the ground-truth alignment with the reference topographic vectors.

We adopt three measures to quantify the performance: a) element-wise Pearson linear correlation of all topographic vectors denoted as `corr`, b) mean RMSE (M-RMSE) of the topographic vectors for single dimension, and c) the point-wise variability (PWV) of the embedding performance, which is defined as the standard deviation of the RMSE for the two sets of embedding vectors in comparison. The statistics are summarized in Table 3, according to which our method produces topographic vectors almost identical to the ones given by exhaustive fiber distance computation in terms of linear correlation and point-wise deviation in the topographic vector space. Furthermore, the reproducibility of our method is also high and stable. The imperfect fidelity similarity between the ground-truth alignment and the reference shows that there are differences between the whole brain white matter streamlines, caused by either individual differences or randomness of probabilistic tractography.

4.3 Performance of non-rigid whole-brain fiber matching

In this experiment, we generate 100 pairs of whole-brain tractograms: one in the subject native T1 space, the other is warped to a reference subject. To achieve this, we first compute the tract-density image (TDI) [7] for each tractogram in the subject native space and then we register the TDI image to a reference TDI image using symmetric diffeomorphic image registration algorithm SyN with ANTs [1]. We adopt the implementation by Dipy [9]. Afterwards, each tractogram is warped to the reference using the nonrigid warp field generated in the last step. In the generated 100 tractogram pairs, one in each pair is in the subject native space, the other is in the reference subject space, and every fiber streamline is corresponded in each pair.

We then apply our nonrigid fiber matching method presented in Algorithm 2 to match each pair of tractograms. In our method, we use $k_1 = 10$ and $K = 500$ in the fast fiber k-NN algorithm for initializing group-wise MDS and linear ICP based point cloud registration. Afterwards, we can apply point-wise k-NN on the aligned topographic vectors to extract the fiber correspondences. An example visual result of the fiber matching in the embedding space is shown in Figure 6. From the results, we cam observe that the topographic vectors from different subjects obtained by individual MDS can be largely misaligned, as shown in the top rows of Figure 6, yet our method effectively removes the ambiguity caused by the non-uniqueness of MDS. Note that the point cloud registration algorithm in our framework can be nonrigid. However, we find that the linear point cloud registration combined with point matching in

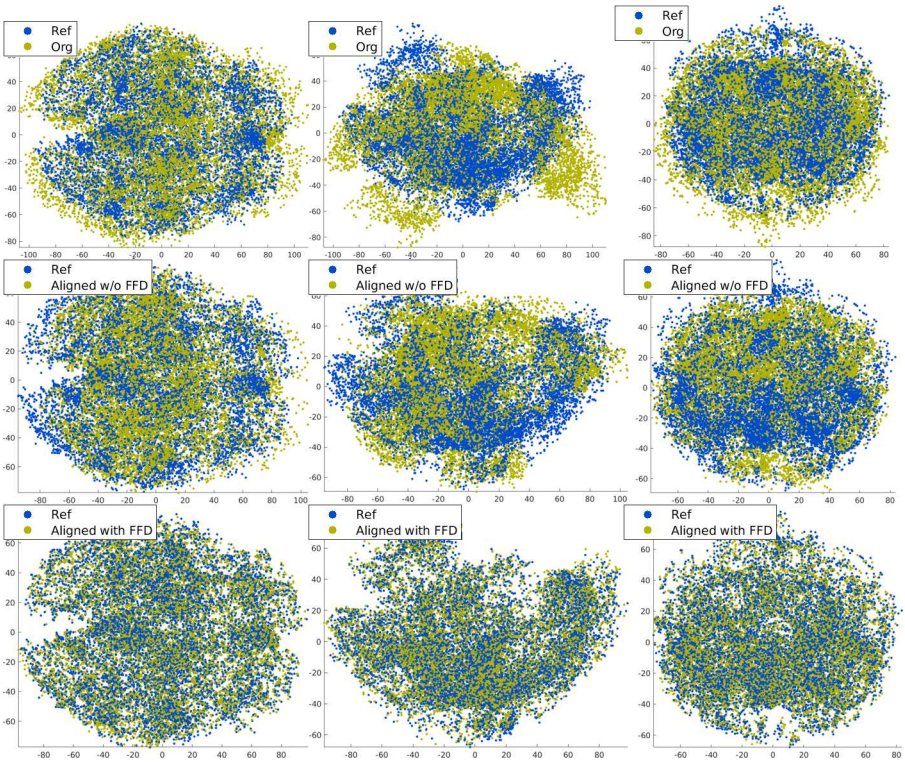


Fig. 6: Visualization of fiber matching in embedding space. The top row shows the initial topographic vectors of the original and reference tractogram. The mid row shows the results of aligning the original topographic vectors to the reference topographic vectors without using our FFD for initialization. The bottom row shows the alignment with FFD for initialization.

the topographic vector space achieves highly accurate matching for nonrigidly deformed tractograms.

We further compare the final results of fiber streamline matching against the streamline matching obatined by TDI image registration implemented in Dipy [9]. The image registration based fiber streamline matching proceeds as follows. For the input pair of streamlines we generate their TDIs, and perform image registration based on SyN algorithm. Afterwards, we use the warpfield generated by the last step to warp the tractograms to obtain the registered tractograms. Finally, we use fiber streamline distance computation to establish the streamline correspondences. The streamline distance can be computed by either exhaustive method or our proposed fast fiber k-NN algorithm. We choose the fast fiber k-NN to implement this step in this experiment.

The direct comparison of errors of fiber streamline matching are shown in Figure 7. From the visual results, we observe that our method compares favorably to the image registration based method.

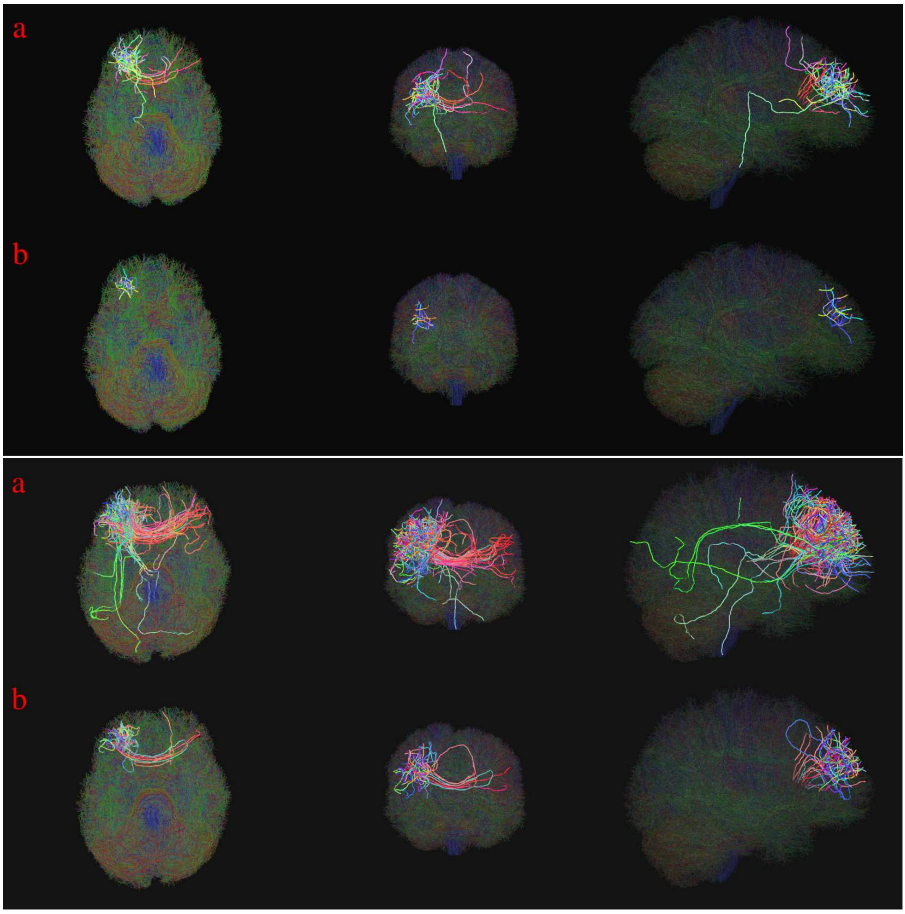


Fig. 7: Comparison of fiber matching results. The incorrectly matched fibers are highlighted, and the rest of the fibers are darkened. Rows labeled with a are the results of fiber matching by volumetric image registration; rows labeled with b are the results of fiber matching by our fast topographic vector matching algorithm.

Since the fibers in the pairs of tractograms in this experiment are corresponded, we are able to quantitatively measure the matching errors. The matching error can be defined as the average Hamming distance between the ground-truth correspondences and the matching results. We additionally calculate the point-wise drift error in the topographic vector space for reference. The statistics of the matching errors are summarized in Table 2. Furthermore, we calculate the histograms of the matching errors of our fast fiber matching algorithm and the image registration based fiber matching. The result is shown in Fig. 8. From both the histograms and statistics we observe that our

method achieves significant improvement over the image registration based fiber matching.

Table 2: Performance of fiber matching.

Measures	Original	Linear ICP with FFkNN	Fiber matching by image registration
Point-wise err.	12.50 ± 2.92	0.061 ± 0.016	N/A
Matching err.	0.99 ± 0.10	**0.0045 ± 0.0025**	0.013 ± 0.0045

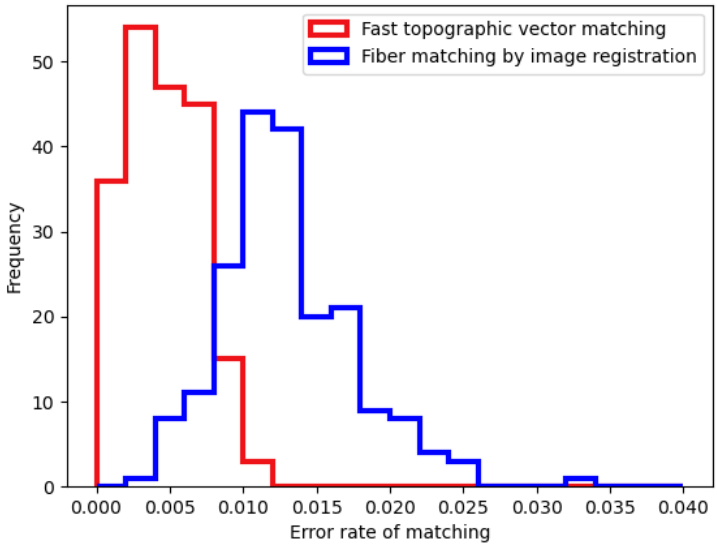


Fig. 8: Performance of fiber matching: image registration based matching against our method

5 Conclusion and future work

In this work, we proposed a novel non-rigid whole-brain fiber streamline matching framework based on a novel fast fiber k-NN algorithm and a novel group-wise MDS algorithm. We extensively validated different modules of our method as well as the whole streamline matching framework using a large dataset from HCP.

The proposed fast fiber k-NN algorithm can be applied to approximate any streamline distance based on point-wise proximity. However, it may fail if the two fiber bundles in comparison are spatially distant and the point-wise K-NN

Table 3: Performance of topographic vector alignment.

Measures	FFkNN vs.GTD	Original vs.GTD	Reproducibility (FFkNN vs. Ref)	Fidelity (GTD vs. Ref)
`corr`	0.989±0.0027	0.0002±0.004	0.968±0.0034	0.972±0.0028
M-RMSE	2.43±0.24	24.71±0.23	4.39±0.25	4.14±0.20
PWV	1.02±0.11	6.65±0.14	1.27±0.09	1.21±0.07

FFkNN: fast fiber k-NN algorithm (Our method); GTD: ground-truth distance (Exhaustive fiber distance computation); Original: Target topographic vectors without alignment

would fail to identify sufficient number of nearby points on any pair of fiber streamlines. Hence, we recommend to use the image registration based tractogram alignment to provide the initial input to our method. The proposed group-wise MDS is currently a one-way mapping from the streamline space to vector space. The continuous latent structure of the streamline space has not been established. A possible future direction is to derive the reversable mapping between the streamline space and vector space. A well-known limitation of the classic MDS is that it requires the full pariwise distance matrix that can be intractable for massive tractograms containing millions of streamlines. The current framework works well for up to around twenty thousand streamlines corresponding to a full distance matrix of 1.5GB. A possible remedy to the problem is to adopt iterative method, such as the SMACOF [6], instead of the closed-form MDS. The proposed streamline matching framework has not addressed the warp field generation based on the streamline correspondence. However, the point-wise error in the topographic vector space may be mapped to the original voxel space along each streamline to derive a voxel-wise registration cost. The proposed algorithms may be applied to improve the resolution of the existing tract-based white matter analysis and hopefully yielding higher specificity of the result.

Information Sharing Statement. Data used in this paper were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Acknowledgement. This work was supported by the Research Initiation Project of Zhejiang Lab (No. 111008-PI2101).

Appendix A Proof of Theorem 1

Proof of Theorem 1 This theorem is proven by showing

$$x^* = \underset{x \in N_\Gamma^K(y) \cap X}{\arg\min} \; d(x, y) = \underset{x \in X}{\arg\min}\, d(x, y) \tag{A1}$$

Since $N_\Gamma^K(y)$ is the set of point-wise K-nearest points of y in Γ and Γ contains point set X, and since $N_\Gamma^K(y) \cap X \neq O$, we can define two subsets of X as X_y and $\overline{X}_y$ satisfying $N_\Gamma^K(y) \cap X = X_y$ and $\overline{X}_y = X \setminus X_y$. Based on the definition of the point-wise K-nearest neighbor, we know that

$$\forall x^\# \in X_y, \forall x^* \in \overline{X}_y, d(x^\#, y) \leq d(x^*, y) \tag{A2}$$

Therefore,

$$x^* = \arg\min_{x \in X} d(x, y) = \arg\min_{x \in X_y} d(x, y) \tag{A3}$$

which completes the proof. □

Appendix B Proof of Theorem 2

Proof of Theorem 2 The L.H.S. of Eq. (8) can be rewritten as

$$\begin{aligned} &\|\mathbf{E}^T\mathbf{B}\mathbf{E} - \mathbf{E}^T\mathbf{z}_i\mathbf{z}_i^T\mathbf{E}\|_F \\ =&\|\mathbf{\Lambda} - \underbrace{[\delta_1, \delta_2, \cdots, \delta_n]^T}_{\delta_i=1, \delta_j=0, \forall j \neq i} \lambda_i [\delta_1, \delta_2, \cdots, \delta_n]\|_F \\ =&\sqrt{\sum_{i=1}^{n} (\lambda_i - \delta_i \lambda_i)^2} \end{aligned} \tag{B4}$$

which completes the proof. □

References

[1] Avants BB, Yushkevich P, Pluta J, et al (2010) The optimal template effect in hippocampus studies of diseased populations. Neuroimage 49(3):2457–2466

[2] Aydogan DB, Shi Y (2016) Probabilistic tractography for topographically organized connectomes. Springer International Publishing, pp 201–209

[3] Bells S, Cercignani M, Deoni S, et al (2011) Tractometry-comprehensive multi-modal quantitative assessment of white matter along specific tracts. Proceedings of the International Society for Magnetic Resonance in Medicine 19(2009):678. URL http://cds.ismrm.org/protected/11MProceedings/files/678.pdf

[4] Bentley JL (1975) Multidimensional Binary Search Trees Used for Associative Searching. Commun ACM 18(9):509–517. https://doi.org/10.1145/361002.361007

[5] Bergström P, Edlund O (2014) Robust Registration of Point Sets Using Iteratively Reweighted Least Squares. Comput Optim Appl 58(3):543–561. https://doi.org/10.1007/s10589-014-9643-2, URL https://doi.org/10.1007/s10589-014-9643-2

[6] Borg I, Groenen PJF (2005) Modern multidimensional scaling: Theory and applications. Springer Science & Business Media

[7] Calamante F, Tournier JD, Jackson GD, et al (2010) Track-density imaging (TDI): super-resolution white matter imaging using whole-brain track-density mapping. Neuroimage 53(4):1233–1243

[8] Garyfallidis E, Brett M, Correia M, et al (2012) QuickBundles, a Method for Tractography Simplification. Frontiers in Neuroscience 6:175. https://doi.org/10.3389/fnins.2012.00175, URL https://www.frontiersin.org/article/10.3389/fnins.2012.00175

[9] Garyfallidis E, Brett M, Amirbekian B, et al (2014) Dipy, a library for the analysis of diffusion MRI data. Frontiers in Neuroinformatics 8:8. https://doi.org/10.3389/fninf.2014.00008, URL https://www.frontiersin.org/article/10.3389/fninf.2014.00008

[10] Garyfallidis E, Ocegueda O, Wassermann D, et al (2015) Robust and efficient linear registration of white-matter fascicles in the space of streamlines. NeuroImage 117:124–140

[11] Gibson WA (1962) On the least-squares orthogonalization of an oblique transformation. Psychometrika 27(2):193–195

[12] Glasser MF, Sotiropoulos SN, Wilson JA, et al (2013) The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80:105–124

[13] Jbabdi S, Sotiropoulos SN, Behrens TE (2013) The topographic connectome. Current Opinion in Neurobiology 23(2):207–215

[14] Jianu R, Demiralp C, Laidlaw D (2009) Exploring 3d dti fiber tracts with linked 2d representations. IEEE Transactions on Visualization and Computer Graphics 15:1449–1456. https://doi.org/10.1109/TVCG.2009.141

[15] Lambert C, Simon H, Colman J, et al (2017) Defining thalamic nuclei and topographic connectivity gradients in vivo. Neuroimage 158:466–479

[16] Maier-Hein, Others (2017) The challenge of mapping the human connectome based on diffusion tractography. Nature communications 8(1):1349

[17] Nath V, Schilling KG, Parvathaneni P, et al (2020) Tractography Reproducibility Challenge With Empirical Data (TraCED): The 2017 ISMRM Diffusion Study Group Challenge. Journal of Magnetic Resonance Imaging 51(1):234–249

[18] O'Donnell LJ, Westin CF (2007) Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE transactions on medical imaging 26(11):1562–1575

[19] O'Donnell LJ, Wells WM, Golby AJ, et al (2012) Unbiased groupwise registration of white matter tractography. In: International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, pp 123–130

[20] Olivetti E, Sharmin N, Avesani P (2016) Alignment of tractograms as graph matching. Frontiers in neuroscience 10:554

[21] Poulin P, Jörgens D, Jodoin PM, et al (2019) Tractography and machine learning: Current state and open challenges. Magnetic Resonance Imaging 64:37–48

[22] Schilling KG, Nath V, Hansen C, et al (2019) Limits to anatomical accuracy of diffusion tractography using modern approaches. NeuroImage 185:1–11. https://doi.org/https://doi.org/10.1016/j.neuroimage.2018.10.029, URL https://www.sciencedirect.com/science/article/pii/S1053811918319888

[23] Siless V, Chang K, Fischl B, et al (2018) AnatomiCuts: Hierarchical clustering of tractography streamlines based on anatomical similarity. Neuroimage 166:32–45

[24] Smith RE, Tournier JD, Calamante F, et al (2012) Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. NeuroImage 62(3):1924–1938. https://doi.org/https://doi.org/10.1016/j.neuroimage.2012.06.005

[25] St-Jean S, Chamberland M, Viergever MA, et al (2019) Reducing variability in along-tract analysis with diffusion profile realignment. NeuroImage 199(May):663–679. https://doi.org/10.1016/j.neuroimage.2019.06.016, https://arxiv.org/abs/1902.01399

[26] Tournier J, Calamante F, Connelly A, et al (2012) MRtrix: diffusion tractography in crossing fiber regions. International Journal of Imaging Systems and Technology 22(1):53–66

[27] Tran G, Shi Y (2015) Fiber Orientation and Compartment Parameter Estimation From Multi-Shell Diffusion Imaging. IEEE Transactions on Medical Imaging 34(11):2320–2332

[28] Wang J, Shi Y (2018) GIFE: Efficient and Robust Group-Wise Isometric Fiber Embedding. In: Wu G, Rekik I, Schirmer MD, et al (eds) Connectomics in NeuroImaging. Springer International Publishing, Cham, pp 20–28

[29] Wang J, Shi Y (2019) A Fast Fiber k-Nearest-Neighbor Algorithm with Application to Group-wise White Matter Topography Analysis. In: International conference on Information Processing in Medical Imaging (IPMI)

[30] Wang J, Aydogan DB, Varma R, et al (2017) Topographic Regularity for Tract Filtering in Brain Connectivity. In: International conference on Information Processing in Medical Imaging (IPMI)

[31] Wang J, Aydogan DB, Varma R, et al (2018) Modeling topographic regularity in structural brain connectivity with application to tractogram filtering. NeuroImage 183:87–98

[32] Wedeen VJ, Rosene DL, Wang R, et al (2012) The Geometric Structure of the Brain Fiber Pathways. Science 335(6076):1628–1634. https://doi.org/10.1126/science.1215280

[33] Xia Y, Shi Y (2020) Groupwise track filtering via iterative message passing and pruning. NeuroImage 221(February):117,147. https://doi.org/10.1016/j.neuroimage.2020.117147, URL https://doi.org/10.1016/j.neuroimage.2020.117147

[34] Zhang T, Chen H, Guo L, et al (2014) Characterization of U-shape streamline fibers: Methods and applications. Medical image analysis 18(5):795–807

[35] Zvitia O, Mayer A, Shadmi R, et al (2010) Co-registration of white matter tractographies by adaptive-mean-shift and Gaussian mixture modeling. IEEE Transactions on Medical Imaging 29(1):132–145